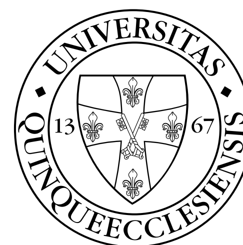


PHARMACEUTICAL BIOLOGY

RÉKA DUDÁS, LAURA NAGY, EDINA PANDUR, VIKTOR SOMA POÓR, KATALIN SIPOS

„Az élettudományi-klinikai felsőoktatás gyakorlatorientált és hallgatóbarát korszerűsítése
a vidéki képzőhelyek nemzetközi versenyképességének erősítésére”
TÁMOP-4.1.1.C-13/1/KONV-2014-0001

University of Pécs, Medical School
Department of Pharmaceutical Biology – 2015

SZÉCHENYI 2020

MAGYARORSZÁG KORMÁNYA

Európai Unió
Európai Szociális Alap

BEFEKTETÉS A JÖVŐBE

Manuscript completed: October 2015

Publisher: University of Pécs

Editor: Katalin Sipos

Lectors: Réka Dudás, Laura Nagy, Edina Pandur, Viktor Soma Poór, Katalin Sipos

Language assistant: Vilmos Wartha

Technical editor: Szilvia Czulák

Length: 282 oldal

CONTENT

1. Introduction to Pharmaceutical Biology 17

2. Pro- and Eukaryotic Cells 21
   - 2.1. Classification of pro- and eukaryotic organisms 21
   - 2.2. Differences between eukaryotic and prokaryotic cells 21
     - Cellular organelles 21
     - Genetic material 21
     - Transcription and translation 22
     - The endosymbiotic theory 22
     - Cell wall: types and structure 22

3. Stem Cells 25
   - 3.1. Definition of stem cells 25
   - 3.2. Types of stem cells 25
   - 3.3. Source of stem cells 25
   - 3.4. Dangers of stem cell therapy 26
   - 3.5. Stem cell therapy in current medical routine 26

4. The Structure of the cell 27
   - 4.1. Cellular and nuclear membranes 27
     - Introduction 27
     - Plasma membrane 27
     - Nuclear membrane 31
   - 4.2. The nucleus 33
     - Structure of the nucleus 33
     - Function of the nucleus 33
   - 4.3. Cellular organelles: Endoplasmic reticulum, Golgi complex, vesicles, endosomes, lysosomes 34
     - Introduction 34
     - The endomembrane system - Vesicles 34
     - The endoplasmic reticulum (ER) 35
     - Chemical composition of the ER 35
     - Major types of the ER 35
     - Biogenesis of the endoplasmic reticulum 36
     - The Golgi complex 36
     - Functions of Golgi complex 37
     - Endosomes 37
     - Lysosomes 38

Cellular organelles – Mitochondrion 38
Other cellular organelles 40
4.4. The cytoskeleton 42
Introduction 42
Functions of the cytoskeleton 43
Structure and composition of microtubules 43
Microtubule-associated proteins 43
Functions of microtubules 43
Axonal transport 44
Motor proteins of the microtubules 44
Kinesins- structure and function 44
Dyneins- structure and function 45
Microtubule organizing centers 45
Microtubule nucleation and synthesis 46
Intermediate filaments 46
Microfilaments 47
Motor proteins of microfilaments 47
Actin-binding proteins 48

5. Structure of DNA, Genes, Chromosomes, Chromatin, Genome 51
5.1. Introduction 51
5.2. The structure of DNA 51
The base composition of DNA 51
The Watson-Crick proposal 51
DNA supercoiling 52
Topoisomerases 52
5.3. Genes and chromosomes 52
Telomeres 52
Centromere 53
5.4. Chromatin 53
Condensation of DNA into chromosomes 54
Characteristics of histone and nonhistone proteins 54
Further condensation of chromatin 54
5.5. Genome 55
The complexity of the genome 55
Genes, gene families, pseudogenes 56
Genome and gene expression 56

6. Human Genome Project - The largest mission of the 20th century's last decade 59
6.1. Preliminaries 59
6.2. The origins – the hierarchical method 59
6.3. Celera Genomics 60
Results: February, 2001 60
Results 60
Structure of genome 61
“Junk DNA” 62
ENCODE Project: Encyclopedia of DNA Elements – examine these sections 62

7. Replication 65
7.1. Features of replication 65
7.2. Prokaryotic replication 65
7.3. Differences between the prokaryotic and eukaryotic replication 66
7.4. Telomeres and telomerases 66

8. DNA repair 67
8.1. Introduction 67
8.2. Types of DNA damage 67
8.3. Repair mechanisms 67
8.4. Medical importance of DNA repair 68

9. Transcription 69
9.1. In general 69
Types of RNA 69
Transcription in prokaryotes 69
9.2. Eukaryotic transcription 70
Initiation of transcription 71
Elongation 71
Termination of transcription 72
9.3. Synthesis and processing of ribosomal and transfer RNAs 72
Processing of pre-rRNAs 72
Processing of tRNA transcripts 72
9.4. RNA processing 73
In general 73
Processing of mRNA 73
9.5. Regulation of transcription 75
Alternative splicing 76
Half-lives of RNAs 76
Degradation of foreign RNAs (RNA interferencia) 76

10. Transcription factors ... 79
10.1. Introduction ... 79
10.2. Classification of transcription factors ... 79
10.3. Functional classification of transcription factors ... 79
10.4. The structure of transcription factors ... 80
10.5. Transcription factor motifs ... 80
10.6. DNA sites involving in regulating transcription ... 81
10.7. Promoter identification ... 81
10.8. The role of enhancers ... 81
10.9. Coactivators ... 82
10.10. Poised polymerases ... 82
10.11. Transcriptional repression ... 82

11. The genetic code ... 85
11.1. In general ... 85
Characteristics of the genetic code ... 87
11.2. Structure and function of the ribosome ... 87
Introduction ... 87
Prokaryotes ... 87
Importance of the rRNA ... 88
Eukaryotes ... 88
Ribosome biogenesis ... 89
Types of the ribosomes ... 90
RNA-binding sites of the ribosome ... 90
Transfer RNAs ... 90
Amino acid activation ... 91
11.3. Phases and regulation of translation ... 91
Steps of protein synthesis ... 91
Polyribosomes ... 92
Prokaryotic translation ... 92
Eukaryotic translation ... 94
The eIF2-cycle ... 95

12. Types of proteins ... 99
12.1. Structure ... 100
12.2. Shape ... 100
12.3. Solubility of proteins ... 101
12.4. Charge of proteins ... 101
12.5. Peptides ... 101

12.6. Proteins and proteids 102
12.7. Proteins and diseases 102
Prion 102
Protein aggregates 102
Therapeutic proteins 104
Important recombinant therapeutic proteins 104

13. Posttranslational modifications and Protein degradation 105
13.1. Posttranslational modifications 105
Proteolytic modifications 105
Chaperones 105
Disulfide bridges 105
Glycosylation 106
Reversible binding of small groups 106
Activating inactive precursor protein 107
13.2. Protein degradation 107
Proteasomal ubiquitin-dependent pathway 108

14. Gene regulation 111
14.1. Gene regulation in prokaryotes 111
The lactose operon 111
Trp repressor 112
14.2. Gene regulation of eukaryotes 112
Eukaryotic and prokaryotic cells: differences 113
Nucleosome structure and chromatin remodeling 113
Chromatin structure, euchromatin, heterochromatin 114
Chromatin remodeling complex (CRC) 114
Histone modifications 114
Histone acetylation 114
Histone methylation 114
Histone code hypothesis 114
DNA methylation 115
Gene regulation during transcription 115
Enhancers 116
Silencers 116
Transcription factors 116
Introns, splicing and alternative splicing 117
RNA maturation – degradation control 117
RNA maturation – RNA editing 117

RNA interference (RNAi) 117
Dicer 118
RISC 118
RNA interference in medicine 118
Protein degradation and posttranslational modifications 118

15. Intracellular protein transport 119
15.1. Protein sorting and transport 119
Nuclear protein transport 119
Peroxisomal protein transport 120
Protein translocation into the mitochondria 120
Protein transport into the endoplasmic reticulum (ER) 121
15.2. Endomembrane system, vesicular transport 121
Generation of vesicles 121
Fusion of the vesicle with the target membrane 122
Golgi system 122
Endocytosis 123

16. Viruses 125
16.1. Discovery 125
16.2. Features 125
16.3. Structure 126
16.4. Classification 126
Life forms of viruses 126
According to the type of nucleic acid 126
Accodring to the symmetry 127
According to the host 128
16.5. Infection and reproduction 128
Adsorption 128
Penetration 128
Decapsidation (if it is necessary) 129
Virus synthesis 129
16.6. Some medically important viruses 130
Oncogen viruses 130
AIDS 130
Influenza viruses 131
Chicken pox 131

16.7. Defense: vaccinations, medicines, natural defense ... 131
Variolisation or variolation ... 131
Protection with drugs ... 132
16.8. Genetic benefits of viruses ... 133

17. Antibiotics ... 135
17.1. Antibiotics ... 135
17.2. Little history – Important steps toward to the discovery of therapeutic antibiotics ... 135
17.3. The first available antibiotic drugs ... 138
17.4. Classification according to the production ... 138
17.5. Antibiogram ... 139
17.6. Pros and cons ... 139
17.7. Antibiotic interactions ... 140
17.8. Resistance ... 140
17.9. Cross-resistance ... 141
17.10. Superbacteria=superbug ... 141
17.11. MRSA ... 141
17.12. Researches nowdays ... 142
17.13. Primycin ... 142
17.14. Uses of antibiotics beyond medication ... 142
17.15. Effect mechanism ... 143
17.16. Groups according to the mode of action ... 143
17.17. Cell wall synthesis inhibitors ... 143
Beta-lactam antibiotics ... 143
Penicillins and cephalosporins belong to this group ... 144
Resistance to beta-lactam antibiotics ... 144
Aminglicosid antibiotics ... 144
Glycopeptides ... 145
17.18. Enzyme inhibitors ... 145
Gyrase inhibitors ... 145
Folic acid antagonists ... 145
17.19. Ribosome-acting antibiotics: Inhibitors of protein synthesis ... 146
Tetracyclines ... 146
Macrolide antibiotics ... 146
Chloramphenicol ... 146
Lincosamide ... 147

17.20. Inhibitors of Membrane Function ... 147
Polymyxins ... 147
Polymyxin A,B,C,D,E ... 147
17.21. Summary ... 148

18. The cell cycle ... 149
18.1. Definition and phases of the cell cycle ... 149
18.2. Regulation of cell cycle: check points ... 149
18.3. Regulation of cell cycle: the cyclin/Cdk complexes ... 150
18.4. Regulation of Cdk activity ... 150
18.5. External activators of the cell cycle ... 151
18.6. Relationship of cell cycle regulation and tumorgenesis ... 151

19. Mitosis and meiosis ... 153
19.1. Mitosis ... 153
Errors in mitosis ... 154
19.2. Meiosis ... 154
Errors in meiosis ... 154
19.3. Comparison of mitosis and meiosis ... 155

20. Signal transduction ... 157
20.1. Introduction ... 157
20.2. Elements of the cell signalling systems ... 157
Types of signalling through extracellular messengers ... 157
The major types of signalling pathways - signalling elements ... 158
The major types of signalling pathways –general signal termination possibilities ... 158
Extracellular messengers ... 159
Receptors ... 159
Second messenger systems ... 160
20.3. Communication between cell signalling pathways ... 161
20.4. G protein-coupled receptors ... 162
Classification of GPCRs ... 162
GPCR interacting proteins (GIPs) ... 163
G proteins ... 163
20.5. Cell signalling through GPCRs ... 166
cAMP and protein kinase A ... 167
Ca- signalling ... 168
Signal termination of GPCRs ... 170
Amplification of extracellular signals through second messengers ... 171

20.6. Tyrosine-kinase receptors 171
Monomeric G proteins 172
20.7. Cytokine receptors and JAK-STAT signalling 176
20.8. TGFβ/BMP-Smad signalling 177

21. Apoptosis 179
21.1. Process of apoptosis 180
Initiation 180
Intrinsic pathway 180
Extrinsic pathway: Death receptors 182
p53 183
Execution phase 184
Endonucleases 184
Transglutaminases 184
Caspases 185
Inhibitor of apoptosis: IAP 185
Apoptosis vs. cell division 185

22. Molecular biology of tumors 187
22.1. Introduction to tumors 187
22.2. Classification of tumors 187
22.3. Development of tumors 187
22.4. Characteristics of Cancer Cells 188
22.5. Factors causing cancer 188
22.6. The genetics of cancer 189
22.7. Oncogenes 189
Examples for oncogenes 191
22.8. Tumor-suppressor genes 191
Classification of tumor-suppressor genes 191
Examples 192
New Strategies for Combating Cancer 193

23. Genetics 195
23.1. Introduction to genetics 195
The genetic code 195
Regulation of protein expression 195
Terminology 195
Chromosomes 196
Linkage of genes 196
The human genome 197

Benefits and concerns of genome research on different fields 197
Genetic disorders 198
Mutations 199
23.2. Chromosomal abnormalities 201
Classification of chromosomal abnormalities 201
Numerical abnormalities: aneuploidies 203
Structural abnormalities 205
23.3. Sex-determining chromosomes or allosomes 208
Chromosome Y 208
Chromosome X 208
Pseudoautosomal regions 208
X-Y sex determination system 209
X chromosome inactivation 209
Numerical abnormalities 209
Genetic diseases linked to allosomes 209
23.4. Mendelian (single-gene) inheritance 210
Terminology 210
Autosomal dominant inheritance 210
Autosomal recessive inheritance 211
X-linked recessive inheritance 212
Other types of Mendelian inheritance 213
Atypical Mendelian inheritance 214
Mitochondrial inheritance 215
Genetic linkage 215
23.5. Mitochondrial diseases 216
Structure of mitochodria 216
Usual symptoms of mitochondrial diseases 216
Genetic classification of mt diseases 216
Examples of mitochondrial diseases 217
Heredity of mitochondrial diseases 217
Treatment 217
23.6. Epigenetics 217
Definition of epigenetics 217
DNA methylation 218
Histone modifications 218
microRNA 218
X-chromosome inactivation 218

23.7. Polygenic inheritance, complex diseases 218
Terminology and definitions 218
Identification of multifactorial disorders 219
Oligogenic and polygenic inheritance 220
Multifactorial inheritance 220
The way to find susceptibility genes 221
Examples of multifactorial disorders 221
Susceptibility testing for multifactorial disorders 225
23.8. Pedigree analysis 225
Introduction 225
Symbols 226
Autosomal recessive traits 226
Autosomal dominant traits 227
X-linked recessive traits 228
X-linked dominant traits 228
Y-linked traits 229
Mitochondrial inheritance 229
Maternal effect 229
Sex-influenced characteristics 230
Sex-limited characteristics 230
23.9. Population genetics 231
Introduction 231
Inheritance influencing factors 231
Genetic variation 232
Genotypic frequencies 232
Allelic frequencies 232
The Hardy-Weinberg Law 234
Extensions of the Hardy-Weinberg law 235
Nonrandom mating 235
One form of nonrandom mating – inbreeding 235
Natural selection 236
Mutation 237
Genetic drift 237
23.10. Developmental genetics 238
History 238
Unity of type 238
Body plan 240
Toolkit genes 240

Non-coding sections 240
Allometric growing 241
Hox genes 242
Positional information 242
Cellular mechanisms 242
Cell specification 243
Human development 243
Homeotic mutations 244
Avatism 244
Summary 245
23.11. Gene therapy 245
Introduction 245
Methods 245
Vectors of gene therapy 245
Dangers of gene therapy 246
Gene therapy in practice 247

Appendix – Methods of Molecular Biology and Genetics 249
Nucleic acids 249
Isolation 249
Detection 250
Electrophoresis of nucleic acids 253
Cloning 254
Analysation of nucleic acids 256
Nucleic acid interactions: Hybridizations, protein interactions: footprinting, EMSA, ChIP and RNA interference 256
Proteins 259
Protein isolation 259
Separations 260
Detection of proteins 266
Analysis of proteins 267
Interactions of proteins 271
Expression of recombinant proteins 272
Model organisms 273
Examples of model organisms 273
Genetically modified organisms or transgenic organisms 274
Imaging techniques 275
Microscope 275
Electron microscopes 277

Cultured cells ... 277
- Primary cultures ... 277
- Secondary cell cultures ... 277
- Cell strains ... 278
- Cell lines ... 278
- Cell banks ... 278
- Transfection of cell cultures ... 278
- siRNAs ... 279

Special methods in genetics ... 279
- Chromosome number and structure, labeling techniques ... 279
- Cytogenetics ... 279
- Chromosome morphology ... 280
- Molecular cytogenetics ... 281

Special methods in pharmaceutical research ... 281
- Structure based drug design ... 281
- Combinatorial chemistry ... 282
- High-throughput screening (HTS) ... 282

1. INTRODUCTION TO PHARMACEUTICAL BIOLOGY

This subject provides you the basic information about a living cell, its organization, main structural elements, and regulation at the molecular level.

A cell is separated from its environment, but it is in constant exchange of materials and information associated with its surroundings. The two major kingdoms of organisms are the **unicellular and multicellular** categories. In the latter case, cells are organized into tissues and organs which contain terminally differentiated cells. Body parts are involved in the complex information of sharing with one another.

Cells can be **prokaryotic or eukaryotic**, but both types have **common characteristics** including the following:

- *cell membrane* with the potential to import and export molecules;
- *energy source* for the different activities of the cell (synthetic reactions, motion, reactivity, secretion of molecules, interactions with other cells, to cite just a few examples);
- *genetic information* for reproduction (the cell or the whole organism), which has to be loyal and maintained including the possibility for correction, if and when it is necessary.

Cells can be attached to a surface, to other molecules, or can be found freely within a solution. Some specialized cells carry out different functions at the different sides of the cell (apical-basal, enterocytes, kidney cells). Prokaryotic cells have to react to the changes of the environment fast and effectively. They lack most of the subcellular compartments which are found in the eukaryotic cells (see lecture about the comparison of the two cell types). **Compartmentalization** helps the regulation of the molecular biological and biochemical processes of a cell.

The cell **membrane** has special structure, impermeable for most of the molecules (please see: lecture on membranes, Protein targeting and transport, and Transporters and channels – in Biochemistry). Special transporter systems make it possible for material exchange with the environment. Membrane receptors sense the messages from the extracellular environment, and through intracellular signaling transduction pathways the cell reacts to these signals. All these activities help to maintain the cellular internal conditions stable. The cell is constantly absorbing **energy** providing molecules, and releasing break down products (these enzymatic reactions are the subject of Biochemistry).

The cell's **genetic information** is reserved in its DNA content. In eukaryotic cells, it is organized into chromosomes. The products of the genes are proteins and RNAs. **Gene expression regulation** determines specifically the region of the DNA genome is used for RNA synthesis at the relative time within the cell. There are more levels of gene expression regulation: structure of the chromosomes, DNA and histone modifications, transcription (RNA synthesis from DNA template), processing (maturation) of RNA, degradation of RNA, translation (protein synthesis), posttranslational processing of the proteins and protein degradation. The higher structural level of the eukaryotes (endomembrane system, compartmentalization, separation of transcription and translation) is a basis for the more complex regulation of gene expression.

Every cell contains the same basic elements and building blocks: water, organic macromolecules made up from common building blocks (see: table on the lecture). **Water** is the special environment of the living cell, but more than just a liquid compound, the majority of the water is organized within the cell because of its special structure, it is important component of the higher structure of the macromolecules, and is involved in the enzymatic reactions. As a polar molecule water is able to create a large number of hydrogen bonds not limited only to water molecules, but with other polar macromolecules (proteins and carbohydrates).

The **four major groups** of biologically important macromolecules are:

- nucleic acids (DNA, RNA)
- proteins (the largest amount in the cell)
- lipids (according to the size they are not really macromolecules)
- carbohydrates (rather: polysaccharides).

They are built up from **small repeated units**: nucleotides, amino acids, fatty acids and sugars, respectively. Because of this common basic components it is assumed, of which there must have been an ancient cell which served in the role of the "parent" of all living cells today, known as LUCA, **last universal common ancestor**.

If we examine the elements of the living cell, we see H, C, N and O provide more than 90% of the cellular mass. Those elements, which are found in only small amount in the cell, play an important role as the dominant ones: they are structural components (P), or important cofactors in enzymatic reactions (Mg, Se), or are part of the ionic milieu of the cell, taking part in the generation of action potential or muscle contraction, or in signal transduction (K, Na, Ca).

The molecular and atomic components of the cell are held together by bonds possessing different strengths: covalent and non-covalent interactions. Covalent bonds are much stronger and they are created and broken up by enzymatic reactions. Non-covalent bonds are:

- ionic interactions (salt bridges)
- hydrogen bonds
- van der Waals forces.

Details of these biologically important forces can be found in the lecture material.

The **significance of the non-covalent bonds** refers to their relatively weak interactions, but notably, in large numbers, they generate strong forces, however, eventually, one by one they can be easily created and broken up. This feature is very important in replication, transcription for example, when hydrogen bonds which keep the double helix together of the DNA must be released. The other important feature of the non-covalent bonds defines how it provides the specificity of binding between two molecules and if they gather to a special surface in a relatively large number.

Nucleic acids (DNA and RNA) are built up of **nucleotides**: sugar (ribose), phosphate and organic base. The latter one can be purine or pyrimidine bases: adenine (A), guanine (G), and thymine (T - in DNA), cytosine (C) or uracil (U – in RNA), which are connected to the ribose with glicosidic bonds. Nucleotides join to one another with phosphodiester bonds between the 3' OH of ribose and 5' phosphate of the other nucleotide. To label the sequence of a polynucleotide chain, only the abbreviated names of the bases are used (AAGTCCT) from the 5' ("free" phosphate) to the 3' ("free" OH) end. Another significant **difference between the RNA and DNA** is the structure of ribose: in DNA there is 2' deoxy-ribose. The presence of this 2' oxygen in RNA gives rise to some consequences: RNA cannot form a regular double helix, like DNA, though RNA has complex secondary and tertiary structures with extensive H-bonding. Moreover, RNA can be easily degraded. In RNA, especially in tRNA and rRNA, there are numerous base modifications, some of them are conserved, meaning they are kept during the evolution because of important structural/functional roles.

Distinctively, RNA has many functions (heritable information – in viruses, enzymatic activity – ribozymes, higher structure, ribonucleotide coenzymes in biochemical reactions) and has led scientists to conclude, it is highly likely, originally, there was a so-called **RNA world.** Later, with the development of DNA (more stable for faithful heritage) and proteins (more diversity for the large number of specific functions) these molecules acquired the majority of the functions.

Proteins are built of **amino acids** which join to one another with peptide bonds. The atoms involved in the peptide bond are located in one planar, which is one of the reasons for the special secondary and tertiary **structures** of the polypeptides (see: Types of proteins). In mammalian cells there are 20 protein building amino acids, which provide a great variability in proteins. The side chains of amino acids can be modified extensively, some of these changes are permanent, yet others are reversible. They have structural or functional consequences (see: Posttranslational protein modifications). Among the numerous **functions** of proteins some important are: enzymes, structural components of the cell and the extracellular matrix, membrane channels and transport, motor proteins, immune functions and components of the blood (transport, osmotic pressure).

Special structural units of the proteins are the motif, fold and domain. **Motif**s are small units with an arrangement of secondary structures found in different proteins, usually with similar functions. **Fold** is generated with motif(s) with structural arrangement of secondary structures. **Domain** contains all the tertiary structural elements of a region of a protein – in which the region can fold separately, has some special function, and may be coded by a unit of a gene.

Sugars and **carbohydrates** are **essential components** of the nucleic acids (pentose), glycoproteins and glycolipids. They provide important **energy** stores (glycogen) and **structural** elements in plants (cellulose) and animals (chitin). Large oligosaccharide units can be attached to lipids and proteins, which results in the change of solubility, antigenity, and can be labels for specific functions.

Lipids are not always macromolecules, but many times they are combinations of fatty acids, a backbone molecule (glycerol or sphingosine), phosphate and some alcohol molecule. As an **energy store** (triacylglycerol), they are neutral, but in the form of phospholipids they carry polar groups, and are involved in the **membrane** structure. Some representatives of the lipids are the components of **signal transduction** (see: Signal transduction pathways). **Cholesterol** also belongs to the lipids. It is essential in membrane structures and provides a large number of important derivatives (steroid hormones, bile acids).

2. PRO- AND EUKARYOTIC CELLS

2.1. Classification of pro- and eukaryotic organisms

The living can be divided into three domains: archaeabacteria, eubacteria and eukaryotes. The former two groups are called together prokaryotes. While the archaebacteria are interesting, we discuss mostly the eubacteria here due to their medical importance.

2.2. Differences between eukaryotic and prokaryotic cells

Cellular organelles

The presence or lack of endomembrane organelles is the most striking difference between the two groups. The eukaryotic organisms can separate spaces within their cytoplasm, resulting optimal environmental conditions for certain reactions (e.g. lysosomes).The following table contains the list of organelles

	prokaryotes	eukaryotes
Nuclear envelope	no	yes
Golgi apparatus	no	yes
Endoplasmic reticulum	no	yes
Lysosome	no	yes
Cytoskeleton	(no)	yes
Mitochondrion	no	yes
Chloroplast	no	yes/no

Genetic material

Apart from the lack of nucleus, there are several differences between the genetic material of the two groups.

Usually only one circular chromosome can be found in the prokaryotes (haploid organisms), while the eukaryotes contain several linear chromosomes (diploid or polyploid organisms).

The prokaryotic chromosome contains minimal amounts of noncoding sequences, there are no introns. Gene expression is generally regulated in groups (operon, see chapter of prokaryotic gene expression regulation).

The eukaryotic genomes carry huge amounts of non-coding DNA. Transcription of genes are regulated independently.

	Prokaryotes	Eukaryotes
Nucleus	no	yes
Chromosomes	one circular + plasmids	more than one linear
Ploidy	haploid	usually diploid
Recombination	partial during sex	during meiosis
DNA binding structural proteins	few proteins	histones and non-histones
Unit of gene expression regulation	operon	gene
Introns	no	yes
Non-coding DNA	few	a lot of

Transcription and translation

In eukaryotic cells transcription and translation are separated, while in prokaryotic cells they happen parallel. (See its importance in the chapter about prokaryotic gene expression regulation)

	Prokaryotes	Eukaryotes
mRNA maturation	no	▪ 5' CAP ▪ introns ▪ polyA tail
localisation of ribosomes	free	ER + free
size of ribosomes	70S	80S

The endosymbiotic theory

The mitochondria and the chloroplasts share many common features with the prokaryotic organisms:

These organelles contain circular DNA, what resembles of prokaryotic chromosomes. They have prokaryotic like enzyme and transport systems. Translation is performed by 70S ribosomes.

These facts suggest, that the mitochondrion and the chloroplast evolved from phagocyted and partially digested prokaryotes. The emerging symbiosis was mutually beneficial for both the eukaryotic and prokaryotic cell.

Cell wall: types and structure

Several pro- and eukaryotic organism bear with cell wall. The structure of cell wall is different varies among the taxonomical groups, but the main functions are the same: mechanical and osmotic protection, filtration of materials reaching the cell membrane.

Bacterial cell walls

Bacteria are divided into two groups by the staining of their cell walls. Gram positive bacteria are stained purple, while Gram-positive bacteria stained red or pink.

Gram-positive bacteria: The cell membrane is covered by a thick peptidoglycan layer. Examples are the *Staphylococcus*, *Streptococcus* and *Bacillus* genuses.

Gram-negative bacteria: An outer cell membrane covers the cell's outer side, what is followed by a thin peptidoglycan layer. After that there is a periplasmic space, then the inner cell membrane. Porins are important proteins of the outer cell membrane, they regulate which molecules can reach the inner cell membrane. Other important molecules of the outer membrane are the lipopolysaccharides (LPS), what are important antigens.

Plant cell walls

There are primary and secondary plant cell walls. The primary cell wall is characteristic to the dividing cells, main components are the cellulose, hemicellulose and pectin. Non-dividing cells develop secondary cell walls, to provide structural stability (e.g. tree trunk). It contains cellulose, xylane and lignin.

Fungal cell wall

The unique component of fungal cell walls in the chitin. Most of the antifungal antibiotics act by inhibition of this cell wall synthesis.

3. STEM CELLS

3.1. Definition of stem cells

Stem cells are found in all **multicellular** organisms that can divide and **differentiate** into diverse specialized cell types and can **self-renew** to produce more stem cells.

Nowadays stem cells are in limelight due to the potential in stem cell therapy.

3.2. Types of stem cells

Stem cells can be classified based on their ability to differentiate (potency). The higher the potency, the stem cell can differentiate into more types of cells.

- Totipotent (omnipotent) stem cell: can differentiate into embryonic AND extraembryonic cells. These cells can make a complete organism. Example: Early stage zygote.
- Pluripotent stem cell: (plurimus: very, many) can differentiate into any of the three germ layers and cells but not extraembryonic cells. Example: embryonic stem cells.
- Multipotent stem cell: stem cells with limited potency. They cannot differentiate into gonads but can be the source of several types of cells of the same group. For example: somatic stem cells.
- Unipotent stem cell: can differentiate into only one type of cell, but can renew itself. Example: muscle stem cells.

3.3. Source of stem cells

The source of stem cells is one of the most important and highly debated question of their medical use.

- Embryonic stem cell (ES): Isolated from blastocysts. (In human embryos it is four-five days after impregnation.) They are pluripotent.
- Somatic stem cells: Isolated from adult organisms. Their potency and maximal number of division is lower than the potency of ES cells. Usually multipotent cells. Their number decreases as the organism ages.
- Umbilical cord derived stem cells: It's possible to isolate embryonic stem cells from the umbilical cord blood after childbirth. (for additional information see the lecture)
- Induced pluripotent stem cells (iPSC): When treated with the necessary transcription factors even differentiated somatic cells can be reprogrammed into pluripotent stem cells. Medical use of iPSCs is very promising. Shinya Yamanaka and sir John Gurdon received the Nobel prize for medicine in 2012 for their achievements on iPSCs.

Therapeutic stem cells can be of autologous or heterologous source: the former means that the patient receives his own stem cells, while in the case of the latter the stem cells come from someone else.

3.4. DANGERS OF STEM CELL THERAPY

Stem cell therapy has a huge potential, it might offer a cure for many severe diseases (for example: neurodegenerative diseases, spinal cord injuries, DM1), but are not free of risks.

- Quackery: many tries to utilise the last hopes of patients and getting rich with nonsensical or downright dangerous 'treatments"
- Teratoma: the implanted stem cells might form malignant tumours
- Recurring disease: in the case of autologous stem cell treatments, the genome of stem cells still carries the mutations responsible for the treated disease

3.5. STEM CELL THERAPY IN CURRENT MEDICAL ROUTINE

Stem cells can give hope treating many diseases, but up to date the only routinely used stem cell therapy is the haematopoietic stem cell transplantation (often called bone marrow transplantation). Diseases of the hematopoietic system (e.g. leukaemia, multiple myeloma, SCID (see: gene therapy lecture)) can be cured. Bone marrow is the most common source of stem cells, but peripheral or cord blood can also yield hematopoietic stem cells. The stem cells must be isolated from the donor, while the recipient's bone marrow is eradicated. Then the stem cells can be transplanted. This treatment is not without risk (graft-versus-host reaction, risk of infections, lifelong immunosuppressive therapy), so it is only used when there are no other options for cure.

4. THE STRUCTURE OF THE CELL

4.1. CELLULAR AND NUCLEAR MEMBRANES

Introduction

The evaluation of the dual layer phospholipid membrane is essential in the formation of a single cell. This surrounding membrane is a flexible layer made up of phospholipids and proteins. The intracellular membranes are essential in the formation of the cellular organelles. At the same time they create a permeable barrier which separates and also interconnects the cells.

In 1957 J. D. Robertson described the structure of the cell membrane, seen for the first time in an electron microscope. Although in electron microscopes, the lipid membrane seems like a trilaminar structure, the plasma membrane is a lipid bilayer, with the outer and inner edges of the membrane and a thin space between the two leaflets composed of the apolar head groups of the lipids (See figure from lecture).

Plasma membrane

Functions of the plasma membrane

1. The plasma membrane separates the cells from each other but it has an essential role in the interactions between cells. So the plasma membrane forms a *selective barrier* and surrounds the cell, which is able to separate the charged molecules but also lets them move selectively through the two leaflets.

2. The *selective molecule transport* happens through channels and pores formed by different proteins. The solute transport is strictly regulated although some molecules are able to pass the plasma membrane by passive transport and by diffusion. Other molecules (e.g. amino acids, sugars) pass through the plasma membrane by the action of ATP, active transport.

3. The plasma membrane has the ability to *respond extracellular signals*. The extracellular signal molecules are able to bind special membrane proteins called receptors and send information into the cell by the action of receptor proteins. This way they are able to change the life of the cell (growth, metabolic activity etc.). Some of the signal molecules can diffuse through the plasma membrane, as they are apolar molecules, and act inside the cell (usually on cytoplasmic or nuclear receptors). The overall signaling process is called signal transduction (more details in the cell signaling chapter).

4. Although the cells are separated by the plasma membrane they capable of communicating with each other, even with signaling molecules or with special proteins and lipids found in the outer leaflet of the plasma membrane. The cells are able to recognize each other by the action of these membrane molecules and they can transmit signals to the adjacent cells (*intercellular interactions*).

5. The intracellular membranes have the same basic structure as the plasma membrane. They enclose special regions of the cytoplasm; they form compartments in the cytosol. The result of *compartmentalization* is the formation of the cellular organelles (nucleus, mitochondrion, chloroplast-only in plants and some bacterial cells, ribosomes, endoplasmic reticulum, Golgi complex, peroxisomes, plant vacuoles).

6. The membranes form distinct compartments and this way they separate the different biochemical activities of the cells. They *provide a scaffold* for these reactions (for the enzymes and substrates) maintain the active (stable) form of the participating molecules.

7. The mitochondrial and the chloroplast's membranes have a special role in the cell's *energy transduction*. These membranes have different machineries (transport molecules and enzymes) which have the capability to convert light energy to chemical energy (chloroplast) or convert chemical energy to ATP (mitochondrion).

The lipid bilayer

In 1970 L. D. Frye and Michael Edidin used fluorescent antibodies to show that plasma membrane molecules can diffuse in the plane of the membrane, indicating that cell membranes are fluid. In 1972 S. Jonathan Singer and Garth Nicolson described the structure of the plasma membrane and proposed a new model, the *fluid mosaic model,* which is the central dogma of membrane biology.

The lipid bilayer is the basic structure for all cell membranes. It consists of different proteins and lipids. The membrane is in a fluid state as the lipids are able to move laterally in the leaflets. So the membrane is a dynamic structure, in which the different molecules can move and create interactions.

The lipid bilayer consists of two leaflets, the outer and the inner leaflet. The main components of the membrane are phospholipid molecules. These molecules are spontaneously capable of forming bilayers by aggregation. The reason for this event is very simple. The lipids are amphiphilic molecules with polar (hydrophilic) and nonpolar (hydrophobic) groups. As the polar head groups are water soluble they turn outwards of the leaflets, and the nonpolar parts of the lipids turn inside the plane of the membrane (See figure from lecture).

On the external surface of the membrane, glycolipids and glycoproteins are found (lipids and proteins with short chains of sugars). They play a role in signal transduction, recognition, adhesion, channel formation or they can also act as transporters. Some of the polypeptides have nonpolar alpha-helices and they extend through the membrane. They act as receptors or transport molecules.

Membrane lipids

(You will learn about the synthesis and functions of lipids in detail in Pharmaceutical Biochemistry)

There are three main types of membrane lipids: phosphoglycerides, sphingolipids, and cholesterol (See figures from lecture).

1. *Phosphoglycerides*: they contain a phosphate group and a glycerol backbone; only two hydroxyl groups are esterified to fatty acids (diglycerides), the third is esterified to the phosphate group. The basic molecule is phosphatidic acid (not found in membranes).
   Types of phosphoglycerides (based on the additional group):
   a) phosphatidylcholine- neutral
   b) phosphatidylethanolamine- neutral
   c) phosphatidylserine- negatively charged
   d) phosphatidylinositol- negatively charged

2. *Sphingolipids*: the basic molecule is ceramide consisting of sphingosine linked to a fatty acid molecule. They are amphipatic molecules and similar to the phosphoglycerides.
   Types of sphingolipids (based on the additional group):
   a) Sphingomyelin (phosphorylcholine):
      - the only membrane phospholipid that is not built with a glycerol backbone

b) Glycolipids (carbohydrate):
   - cerebroside: substitution is a simple sugar
   - ganglioside: substitution is a small cluster of sugars

3. *Cholesterol*: the hydrophilic head group is faced to the outer side of the membrane while the remaining part is embedded in the bilayer.

The role of lipids in the membrane/Membrane fluidity

The lipid composition of the cellular membranes is quite different. There are variations in the types and abundance of lipids (See table from lecture). The lipids determine the physical state of the membranes and influence the movement and activity of membrane proteins. Due to the lipids in the membrane, their structure is continuous and flexible; they are able to change their shapes, play a role in locomotion. The membranes are able to fuse with each other, form vesicles, thus they also play a special role in solute transport.

The two leaflets of the membranes also have different lipid composition. This feature of the membranes, the *lipid asymmetry,* is due to the divergent functions of the two leaflets (See figure from lecture). Glycolipids are found in the outer leaflet and they usually serve as receptors. Phosphatidylethanolamine and phosphatidylserine are found in the inner leaflet. The first one play a role in membrane fusion and membrane budding, while the latter is able to bind positively charged amino acids (channel formation).

The lipid composition of the membrane is responsible for the main characteristic feature of the membrane, the *fluidity* (viscosity). The lipids maintain the fluid state of the membrane, but this characteristic depends on the temperature. Normally (~37C^o), the lipids are able to move laterally, or rotate in the membrane. This state of the membrane is called the *liquid crystalline* phase. If the temperature lowers, the membrane lipids are not able to make movements anymore; the liquid crystalline phase is converted to *frozen crystalline gel*. The process is called *phase transition* and the temperature at which transition occurs is called *transition temperature* (See figure from lecture). The transition temperature depends on the presence of unsaturated fatty acids. The greater the degree of unsaturated fatty acids of the bilayer, the *lower* the temperature before the bilayer forms gel. Cholesterol degree is crucial in membrane fluidity. As it is a rigid molecule, it makes the membrane more durable while decreasing its permeability. Cholesterol is able to interact with phospholipids and decreases their mobility in the lipid bilayer. So cholesterol maintains an intermediate fluidity.

The question is how the cells can maintain the membrane fluidity in a changing environment. The answer is remodeling. The cells are able to change the phospholipid composition of the membrane even with changing the percentage of different lipids or changing the degree of saturation of fatty acids. For the latter mechanism, the cell desaturates single bonds in the fatty acid chains to double bonds, reshuffles the different fatty acid chains and forms new unsaturated phospholipids. This process is carried out by special enzymes: desaturases, phosholipases and acyltransferases.

Membrane proteins

There are many different proteins in the plasma membrane which depends on the type of the cell. Some of the proteins are only found on the external surface of the cell (receptor proteins, junctions of the cells), while others are found on the inner surface of the membrane (send signals into the cytoplasm). The membrane proteins are classified into three groups:

1. Integral or transmembrane proteins (e.g. Glycophorin A, See figure from lecture):
   - They pass through the plasma membrane.
   - Their structural elements are made up of extracellular domain, transmembrane domain/domains (single pass or multipass transmembrane proteins) and intracellular domain. The transmembrane domain has an alpha helical structure.

- Their functions are the following: receptors, channels and transporters (ion, solute and electron transfer)
- They are amphipathic molecules.
- They are in direct contact with the phospholipids (barrier).
- They are able to move in the leaflets.

2. Peripheral membrane proteins:
   - They are found even on the extracellular or intracellular leaflet.
   - They are bound to the membrane by weak electrostatic bonds.
   - They are anchors for integral membrane proteins.
   - Other functions: enzymes, transmit signals

3. Lipid anchored membrane protein:
   - They are found even on the extracellular or intracellular surface.
   - They are linked to a lipid molecule in the membrane (lipid anchored).
   - GPI-anchored proteins (glycosyl-phosphatidylinositol anchored) are found on the outer surface of the membrane: receptors, enzymes, adhesion molecules.
   - Cytoplasmic lipid anchored proteins are attached by long hydrocarbon chains embedded in the inner leaflet: enzymes, signaling molecules.

Plasma membrane has a dynamic structure

The plasma membrane is a fluid membrane in which the molecules (lipids and proteins) are able to move. The mobility of proteins depends on the phospholipid composition of membrane, thus the lipids maintain the fluidity.

1. Movements of the lipids (See figure from lecture)
   - Lateral shift: they diffuse laterally in one leaflet rapidly.
   - Transverse diffusion: flip-flop movement between the two leaflets at slow rate. This process uses enzymes such flippases that catalyze the flipping of specific phospholipases (movement of phosphatidylethanolamine, phosphatidylserine from the extracellular surface to the cytoplasmic leaflet) in the plasma membrane. Flippases use the energy of ATP hydrolysis to flip the lipids.

2. Movements of the membrane proteins (See figure from lecture)
   - Random movements within the membrane at moderate rate (reduced diffusion).
   - Directed movements in the membrane.
   - Some of the proteins are immobilized.

3. Patterns of movement of integral membrane proteins (See figure from lecture)
   - Random movement/diffusion
   - Immobilized protein due to the interactions with the cytoskeletal elements
   - Immobilized protein by the other intermembrane proteins
   - Immobilized protein by the fences of cytoskeletal proteins
   - Immobilized protein by extracellular matrix elements
   - Movement in one direction due to the interaction with a motor protein

Glycocalyx

The glycocalyx is a carbohydrate rich zone on the extracellular leaflet. Some of the proteins are glycosylated and the carbohydrates coat the cell surface. These carbohydrates are bound to integral membrane proteins and lipids to form glycoproteins, glycolipids and *proteoglycan* molecules. Proteoglycans are formed by long polysaccharide chains linked to a protein core. The protein core sometimes passes through the plasma membrane or attached to GPI anchor. Formally, the glycocalyx is a part of the extracellular matrix.

Functions of the glycocalyx:

- Provides protection against mechanical and chemical damages,
- Maintains the interactions between the cells,
- Keeps the adjacent cells at a distance to inhibit protein-protein interactions.

Permeability of phospholipid bilayer

(You will learn about the transporters, ion channels and the types of transport in detail in Pharmaceutical Biochemistry)

Different processes are known by which substances (ions, sugars, amino acids etc.) move across the plasma membrane.

1. Diffusion through the lipid bilayer: a molecule moves across the membrane from the region of high concentration towards the region of low concentration

2. Diffusion through an aqueous, protein-lined channel

3. Facilitated diffusion: a facilitated transporter protein binds the molecule, undergoes a conformational change and helps the solute go through the channel. On the opposite side of the channel the solute diffuses down from the protein.

4. Active transport: requires an energy-driven protein, that is able to hydrolyze ATP to ADP+P and uses the energy to “pump” the moving substances against a concentration gradient.

Nuclear membrane

A eukaryotic nucleus contains the chromatin Chromosomal DNA), one or more nucleoli (ribosomal RNA synthesis), a nucleoplasm and a nuclear matrix (fibrillar network).

Nuclear envelope

(See figure from lecture)

1. The nucleus is separated from the cytoplasm by two membranes: *inner nuclear membrane* and *outer nuclear membrane*. Each consists of a phospholipid bilayer and associated proteins. Between the two membranes there is a 10-50 nm wide *intermembrane space* which is continuous with the endoplasmic reticulum. The main function of nuclear membrane is to form a barrier for ions, solutes and macromolecules.

2. Inner nuclear membrane: contains specific proteins which anchor the chromatin fibers and *nuclear lamina*.

3. Outer nuclear membrane: contains ribosomes on its surface, it is continuous with the membrane of endoplasmic reticulum.

4. Nuclear lamina: a protein network provides mechanical support for nuclear envelope. It consists of lamins which serve as intermediate filaments in the cytoplasm. Assembly of lamins is regulated by phosphorylation and dephosphorylation.

Nuclear pore complex

(See figures from lecture)

Nuclear pores are responsible for molecule transport across the nuclear membrane. This process is quite different from the transport mechanisms of cellular membranes. Nuclear pore complex fills the nuclear pore, projects into the cytoplasm and nucleoplasm and regulates the molecule transport across the nuclear pore.

Structure and characteristics of the nuclear pore complex

1. A mammalian cell contains 3000-4000 NPCs.
2. Consists of 30 different NPC proteins (~100 copies) called nucleoporins.
3. It has octagonal symmetry.
4. Transport rate: 500 macromolecule/sec
5. Direction of transport: both directions at the same time
6. One or more aqueous channels: small water soluble molecules can pass by diffusion
7. The *nuclear basket* is formed by the distal ends of long filaments projecting into the nucleoplasm; the filaments are interlinked with a ring–like structure.
8. *Cytoplasmic filaments* extend from the cytoplasmic side of the NPC into the cytosol.
9. *Proximal filaments* are attached to the central transporter proteins and together they form the channel.
10. The central channel is formed by nucleoporins rich in phenylalanine-glycine repeats.
11. The NPC is embedded into the nuclear membrane.
12. The *cytoplasmic and nucleoplasmic rings* hold the NPC at its place in the nuclear membrane.
13. The *outer and the inner spoke rings* interlink the cytoplasmic and nucleoplasmic rings and maintain the structure of NPC.
14. Diffusion of small (5000 Daltons or less) water-soluble molecules
15. Large molecules (over 60000 Daltons): active transport
16. Ribosomal subunits, RNA and DNA polymerases (over 100000-200000 Daltons) are ferried by specific receptor molecules

4.2. THE NUCLEUS

Nucleus is the largest organelle containing the genetic material of the cell (chromatin). Bounded by the nuclear envelope, the nucleus protects the DNA of the cell, and also provides spatial separation and more sophisticated regulation of the gene expression.

Structure of the nucleus

(for figure, see lecture slides)

The **nuclear envelope** is a double membrane surrounding the nucleus. The lumen between the two membranes is contagious with the endoplasmic reticulum. The nuclear pore complexes (NPCs) perform the nucleocytoplasmic transport of several molecules (see Chapter “cell membrane, nuclear membrane”).

The **nucleolus** is a dynamic, non-membrane bound, robust structure within the nucleus. In late telophase, the nucleolus assembles around the NOR (nucleolus organizing region), which contains several copies of the rRNA genes. Then it disassembles when the cell enters mitosis – this is the reason why nucleolus is said to be dynamic. Nucleoli are not membrane enclosed, still it is easy to recognize them by phase contrast microscopy due to the higher density of the structure. The main function of the nucleolus is the ribosomal RNA transcription as well as ribosome subunit biogenesis and assembly. Recent research revealed additional functions like RNA editing, DNA damage repair and tRNA processing.

The **nucleoplasm** is the place of transcription and post-translational modifications (mRNA processing), like splicing. Processed mRNA then is transported to the cytosol through NPCs, where it is bound to ribosomes to start translation: the way from mRNA to peptide chain. There are several different non-membrane-bound subcompartments within the nucleoplasm, like speckles and Cajal-bodies.

The **chromatin** is the genetic material of the cell found in the nucleus. In electron-micrographs, chromatin can be seen in two configurations. **Heterochromatin** refers to condensed DNA, which is not actively transcribed, while **euchromatin** shows a loose structure, containing expressed genes. Euchromatin is decondensed and free from histones and other proteins to make DNA available for RNA polymerases. Different cell types have a different pattern of expression, therefore different chromatin structure.

The **nuclear lamina** is a two-dimensional net-like structure under the nuclear envelope made up by lamin intermedier filament proteins. It gives mechanical support to the nucleus, and also helps chromatin organization by binding to the matrix attachment region (MAR) on the DNA. It is also involved in the regulation of the cell cycle. In the prophase and prometaphase of the mitosis, the nuclear envelope, the lamina and the NPCs are disorganized so that the mitotic spindle can reach the chromosomes and attach them to the kinetochores. At the end of mitosis, in the anaphase and telophase the nuclear reorganization is regulated in time. First, the “skeleton” proteins are organized on the surface of the partially condensed chromosomes, then the envelope is reorganized. After, new NPCs are formed, through which lamins are actively transported. See protein transport in Chapter “Protein transport in the cell”.

Function of the nucleus

There are many advantages of having a nucleus. DNA, especially euchromatin is very sensitive, and spatial separation from different cellular processes is an additional protection beyond histones. The NPCs selectively transport big molecules in and out of the nucleus, acting as a barrier to unwanted others. In eukaryotes, the steps of gene expression are spatially separated: The transcription of genes

to mRNA is done in the nucleus, followed by RNA processing. The mature RNA is then transported out of the nucleus, and the translation of the mRNA to peptides happens on the cytoplasmic ribosomes. The proteins then can be selectively transported back to the nucleus. Ribosomal subunits are assembled in the nucleus from RNAs synthetized in the nucleus, plus proteins imported from the cytoplasm. The assembled ribosome is then exported, and function in the cytoplasm. The spatial separation of the processes of gene expression gives a possibility to a very sophisticated regulation, which is absent in Prokaryotes.

For further information on gene expression and its regulation, see Chapter “Genome and gene expression”.

4.3. Cellular organelles: Endoplasmic reticulum, Golgi complex, vesicles, endosomes, lysosomes

Introduction

In the late 40’s (after the development of electron microscope), it became evident that the cytoplasm of the eukaryotic cells was subdivided into distinct compartments bounded by membranes. These membranous compartments in the cytoplasm formed different organelles. These organelles form an endomembrane system, in which the individual components function as part of a coordinated unit (see figure from lecture).

The endomembrane system - Vesicles

The cellular organelles are the part of an integrated network, in which molecules (proteins, lipids etc.) are transported by the action of transport vesicles. The transport vesicles are membrane bound organelles that bud from the donor membrane and fuse with the membrane of the acceptor organelle. The transport vesicles can be formed from the ER, Golgi network and plasma membrane. Based on the membrane components, they can be classified into three groups:

1. COPI coated vesicles
2. COPII coated vesicles
3. Clathrin coated vesicles

The three groups of vesicles transport molecules in different directions. The protein coat determines the type of molecules that will be transported from one organelle to another (see Protein trafficking for more details).

In the eukaryotic cell several distinct biochemical pathways were indentified that were responsible for material synthesis and transport (see figure from lecture):

1. Biosynthetic pathway: synthesis, modification and transport of the proteins
2. Secretory pathway: proteins are discharged from the cell
   a) constitutive secretion-continuous
   b) regulated secretion-in a response to a stimulus (see Protein trafficking for more details)
3. Endocytic pathway: the movement of materials from the extracellular space to early endosomes where sorting occurs. Endocytosis can be divided into two categories:
   a) *pinocytosis* (uptake of a particular fluid)
   b) *receptor mediated endocytosis* (uptake of specific extracellular ligands following their binding to receptors) (see Endosomes and Protein trafficking for more details).

The endoplasmic reticulum (ER)

The endoplasmic reticulum is the most dynamic and morphologically variable of all membranous organelles. The ER utilizes a cytoskeleton scaffold, associated motor proteins, and undergoes constant rearrangement, while it maintains its structure. The ER is a continuous network of interconnected tubules, cisternae and highly organized lamellar sheets. These structures are the building blocks of the different types of the ER, which include rough (RER), smooth (SER), transitional (tER or exit sites), sarcoplasmic reticulum (SR) and the nuclear envelope. In eukaryotic cells the ER consists of several interconnected branching tubular membranes that protrude from the nuclear envelope and extend to the periphery of the plasma membrane (see figure from lecture).

The size of the ER is relatively constant, although the lipids and proteins (newly synthesized membrane and luminal proteins) move into and out of the compartment. The proteins are transported from the ribosomes to the ER, from the ER to the other organelles or to the plasma membrane (secretion), from the Golgi to the ER, etc. (see protein trafficking for more details).

Chemical composition of the ER

It has higher protein concentration (60-70%) and lower phospholipid concentration (30-40%) than the plasma membrane. The phospholipids of ER are mostly phosphatidylcholines (55%) and phosphatidylethanolamines (25%) and they have very low concentration of sphingomyelin (5%) compared to the plasma membrane.

Major types of the ER

The endoplasmic reticular membranes are thinner than the plasma membrane and they are more stable and less fluid than the plasma membrane. The endoplasmic reticulum is divided into two subcompartments: 1) the rough endoplasmic reticulum (RER) and 2) the smooth endoplasmic reticulum (SER). Both types of ER consist of a system of membranes that encloses a space, or lumen, that is separated from the surrounding cytosol. The composition of the luminal/cisternal space inside the ER membranes is quite different from that of the surrounding cytosol. The SER and RER have important structural and functional differences.

Structural differences:

1. Rough endoplasmic reticulum
   a) It is studded with electron dense particles called ribosomes on its cytosolic surface. Ribosomes are bound to ER membrane by ribophorins that add some structural rigidity to the membranes.
   b) It is composed of a network of flattened sacs interconnected by short tubular segments (see figure from lecture).
   c) It is continuous with the outer membrane of the nuclear envelope.
2. Smooth endoplasmic reticulum
   a) It lacks associated ribosomes.
   b) Its membranous elements are highly curved and tubular, forming an interconnecting system.
   c) It has three specialized forms:
      - Lamellar form: exists as excessive lamellae of long, flattened sacs bounded by membranes.
      - Vesicular form: exists as small vesicles, round structures bounded by membranes.
      - Tubular form: exists as elongated tubules bounded by membranes.

The three forms of SER are freely convertible, suggesting that ER is a highly dynamic pleomorphic organelle. All three forms of SER are also found in RER.

Different types of cells contain markedly different ratios of the two types of ER, depending on the activities of the cell. The ER is the primary site of membrane and secretory protein synthesis, translocation, and maturation in the cell. Approximately one-third of all cellular protein is translocated into the membrane and/or the oxidizing lumen of the ER. The rough endoplasmic reticulum can be converted to SER if it is needed.

Functional differences:

1. Rough endoplasmic reticulum
   a) Synthesis of the integral membrane proteins of plasma membrane
   b) Synthesis of the secreted proteins
   c) Synthesis of the soluble proteins of the organelles

   (The other types of proteins, e.g. cytosolic proteins, peripheral membrane proteins, nuclear proteins, proteins incorporated into chloroplasts, mitochondria and peroxisomes are synthesized on free ribosomes.)

2. Smooth endoplasmic reticulum
   a) Steroid hormone production in testes, ovaries and adrenal gland
   b) Detoxification in the liver
   c) Storage and release of calcium ions

Biogenesis of the endoplasmic reticulum

Membranes do not arise *de novo*, membranes arise from preexisting membranes. Membranes can only grow when the newly synthesized proteins and lipids are inserted into existing membranes in the ER. As the membrane moves from one compartment to the next, its proteins and lipids are modified by enzymes that reside in the cell's various organelles. These modifications make the membrane compartment unique. The membrane asymmetry evolves first in the ER, and this asymmetry remains with the movement of the membranes (domains on the cytosolic surface remain on the cytosolic part and the luminal domains remain on the luminal surface).

Most membrane lipids are synthesized entirely within the endoplasmic reticulum. There are two exceptions 1) synthesis of sphingomyelin and glycolipids begins in the ER and it is completed in the Golgi complex), and 2) some of the unique lipids of the mitochondrial and chloroplast membranes (synthesized by enzymes that reside in those membranes). Newly synthesized phospholipids are inserted into the half of the bilayer facing the cytosol. The lipids can move from the cytosolic side to the luminal side by the action of flippases. The ER membrane contains enzymes that are able to change the lipid composition of the membranes; they can convert a phospholipid to another one.

The Golgi complex

The Golgi complex is typically organized as a series of three to eight flattened cisternae (0.5-1 µm diameter) arranged as a stack with surrounding tubules and vesicles. The flattened nature of the cisternae produces a large surface-to-volume ratio that is thought to facilitate the activity of resident Golgi enzymes. The Golgi is positioned centrally in the transport route between the ER and plasma membrane. The Golgi membrane is about 7.5 nm thick, and the lamellae are separated by gaps. Each lamella contains fenestrations, small vesicles and tubules. The Golgi stacks in mammalian cells are interconnected by membranous tubules to form a single, large ribbon-like complex situated adjacent to the cell's nucleus (see figure from lecture).

The stack of cisternae of Golgi complex exhibits a cis-to-trans polarity. Newly synthesized membrane and secretory components enter the stack at its cis face (cis-Golgi), comprised of cisternae and associated tubules/vesicles in close vicinity to the ER. The vesicles then pass through cisternae in the middle of the stack (medial-Golgi) and leave the Golgi at the trans face (trans-Golgi), which is at the opposite end of the stack (see figure from lecture). The cis face of Golgi is called forming face and the inner face is called maturing face. New membrane components are added to the forming face from ER and old membranes and vesicles bud from the maturing face.

Two working models of Golgi exist. The first one is the cisternal maturation model that claims that the cisternae physically move from the cis face to the trans face, so each cisterna matures into the next one. The second model presents that the materials are transported through the Golgi complex by transport vesicles and the cisternae remain in place (see figure from lecture).

Two major sites of Golgi complex play a role in protein sorting in different directions (ER, Golgi, plasma membrane, lysosome). These tubular structured parts are the cis-Golgi network and trans-Golgi network (see figure from lecture).

Functions of Golgi complex

(see Protein trafficking for more details)

Protein synthesis: proteins are transported from the ER to Golgi apparatus forming face in transport vesicles.

Glycosylation of proteins by glycosyl transferases

Protein transport: posttranslationally modified proteins are transported to other organelles by membrane coated vesicles.

Membrane synthesis continues in Golgi and new membranes move to and fuse with the plasma membrane.

Endosomes

(see Protein trafficking for more details)

Endosomes are produced from the plasma membrane by endocytosis through which the cell uptakes different materials. The materials are said to be internalized and the vesicles (chlatrin coated vesicles) are transported to a dynamic network of tubules and vesicles known as endosomes (see figure from lecture). The fluid in the lumen of endosomes is acidic by the action of H-ATPase (proton pumps) in the boundary membrane.

Endosomes are divided into two classes:

- early endosomes: located near the peripheral region of the cell,
- late endosomes: typically located closer to the nucleus.

Early endosomes progressively mature into late endosomes. During the maturation process the pH decreases, protein exchange occurs (Rab proteins) and morphology of the structure changes. Late endosomes are also called multivesicular bodies (MVB, see figure from lecture). Late endosomes get lysosomal enzymes from the trans-Golgi network and they begin the digestion of the materials. Then the content of the late endosomes is transferred to a lysosome. If there are some molecules that can be reused in the cell (e.g. receptor proteins), a new vesicle will bud from the sorting compartment (between the early and late endosomes; see figure from lecture). The recycling compartment (vesicle)

transports these proteins back to the plasma membrane (see Protein trafficking for more details) and the cell reuses them.

Lysosomes

Lysosome is the “stomach” of the eukaryotic cells. It contains approximately 50 different enzymes, acid hydrolases that can digest every type of biological molecules (see table from lecture). The enzymes found in lysosomes are produced in the rough endoplasmic reticulum and transported to the lysosome by membrane coated vesicles. Acid hydrolases can act only in acidic environment. The pH optimum of these enzymes is 4.6. The acidic pH is maintained by high proton concentration which is produced by H-ATPase proton pumps found in the boundary membrane of the lysosome. Highly glycosylated proteins found in the lysosomal membranes form protecting layer that protects the membranes from enclosed hydrolases. The shape and electron density of the lysosomes are different, thus it is very difficult to identify them in the cells (see figure from lecture).

Lysosomes play a special role in organelle turnover. Organelle turnover is the regulated destruction of the organelles in the cytoplasm. This process is called autophagy (see figure from lecture).

Steps of autophagy:

1. The organelle is surrounded by a double membrane, autophagosome is produced.
2. The autophagosome is fused with a lysomome and forms an autophagolysosome.
3. Lysosomal enzymes digest the organelle and the breakdown products form a residual body.
4. The residual body is eliminated by exocytosis or it is retained in the cytoplasm as lipofuscin granule (aging process).

Due to autophagy the components of an organelle become available to the cell (deprivation), as the cell acquires energy for its life. Autophagy play an important role in the neuroprotection. It maintains homeostasis or normal functioning by protein degradation and turnover of the destroyed cell organelles for new cell formation. The failure of autophagy is thought to be one of the main reasons for the accumulation of cell damage and aging.

Cellular organelles – Mitochondrion

Mitochondria are described as oval-**shape**d bodies with a size of 0.5 – 10 μm. In real life, they are intensively dynamic cellular components: they are able to move within the cells (neuronal axons) or fixed to the cytoskeleton (muscle cells) and are able to carry out both fission and fusion. The **number** of mitochondria per cell varies depending on the cell type and the activity of the specific cell: no mitochondria are found in the mature red blood cells, while hundreds or thousands in liver, muscle, heart cells.

Structure: Mitochondrion is surrounded by a double membrane system. The two membrane layers differ in composition and function, as well. The **outer membrane** contains porins and is freely permeable to ions and small molecules (less than 5000 Da). Its composition is similar to the cell membrane. The **inner membrane** creates **cristae** to increase the membrane surface. It is impermeable to even small charged molecules, but contains several transport systems which are indispensable for the biochemical processes taking place in mitochondrion. Between the two membrane layers is the **intermembrane space**, which among other roles, is important in apoptosis (see in: *Apoptosis*). For more in depth information in reference to the structure (including figures), please see the *lectures*.

Within the **matrix** of the mitochondrion, is the location of the large number of **enzymes** (biochemical reactions characteristic of mitochondrion: in *Biochemistry studies*) involved in diverse biochemical

reactions, the most predominant and outstanding of which is the generation of **ATP**, the basic energy-providing molecule of every cell.

Functions of mitochondrion:

- providing energy (citric acid cycle, lipid oxydation, ATP synthesis)
- urea cycle (eliminating N)
- iron-sulfur cluster and hem biosynthesis
- involvement in metabolic regulation
- apoptosis and cell death

In addition to the enzymes, mitochondrion carries its own **genome** and transcriptional and translational machineries. The number of DNA molecules per mitochondrion is variable (see consequences in: *Mitochondrial diseases*). The human mitochondrial genome is circular, and encodes for 13 mRNAs, 22 tRNAs and 2 rRNAs. Messenger RNAs produce different, highly hydrophobic subunits of the oxidative phosphorylation system (this is synthetizing ATP with the use of molecular oxygen).

Further characteristics of the **human mitochondrial genome**:

- double stranded with a L (light) and H (heavy) chain containing 16569 bp (base pairs).
- its mutational rate is relatively high (low fidelity of the DNA polymerase, and low efficiency of the DNA repair system)
- almost no non-coding DNA sequences: missing regulatory DNA regions
- special amino acid codons.

Comparison of human nuclear and mitochondrial genomes: on *lectures*.

The **limited autonomy of mitochondrion** refers to the mitochondrial genome which is not enough to generate all the necessary mitochondrial proteins (enzymes, proteins of the transcription, translation, replication and repair), so a large number of proteins and RNAs have to move into the mitochondria. They are transported into the organelle through special protein (TIM and TOM) or other transport complexes (RNAs). The protein transport into mitochondria is unidirectional: from the cytosol to the organelle (outer, inner membrane, matrix or intermembrane space – please see more in the notes: *Protein targeting and transport*), except in apoptosis, when cytochrome c is leaving mitochondria. According to new theories, mitochondrion retains the synthesis of those proteins, which are too hydrophobic to be transported from the cytosol (they are membrane-bound proteins and part of the oxidative phosphorylation system).

The endosymbiont hypothesis: following the appearance of oxygen in the atmosphere, it became possible to merge two different types of cells, one of them providing nucleus, and the other (an aerobic bacterium) providing the source of mitochondrion. In time, the latter one lost a large portion of its DNA, relying on the nuclear genome for proper functioning, but gained additional biochemical processes. This is part of regulation, called **compartmentalization**, which is important not only in cellular biological activities, but also in the regulation of metabolism.

Proofs of the endosymbiont theory:

- structure of the mtDNA (shape, missing histons and introns)
- structure of mt (mitochondrial) ribosomes
- translation starts with *N*-formylmethionine (see in: *Translation*)
- sensitivity to certain antibiotics (see: *Antibiotics*)
- ability of mitochondria for fission and fusion (both mechanisms require the energy of GTP and the mitochondrial proton motive force).

Other cellular organelles

Peroxisomes

Structure

Peroxisomes (microbodies) are small, single membrane-bound organelles found in all eukaryotic cells. They are morphologically similar to lysosomes but assembled from proteins synthetized on free ribosomes and then transferred to the peroxisome as completed polypeptide chains. Peroxisomes replicate by division but they don't have their own genomes.

Function

Peroxisomes contain more than 50 different enzymes, which are involved in different biochemical pathways. Their name comes from the originally defined function of the organelle: performing oxidation reactions with hydrogen peroxide as product. Hydrogen peroxide is harmful to the cell, so peroxisomes readily decompose it by the enzyme **catalase**. Many substrates are broken down by such oxidative reactions, including uric acid, amino acids and fatty acids. Fatty acid oxidation (**beta oxidation**) is particularly important, as it is the major source of metabolic energy. In animal cells, fatty acids are oxidized in mitochondria also but in yeasts and plants fatty acid oxidation is restricted to peroxisomes.

Peroxisomes play a role in lipid biosynthesis: they synthetize **cholesterol** in animal cells, and they are involved in bile acid synthesis from cholesterol in liver cells. **Plasmalogens**, the important membrane phospholipids found in brain and heart, are also synthetized by peroxisomal enzymes.

Glyoxisomes are special peroxisomes found only in plant cells. Their name comes from the main function of this compartment, the **glyoxylate cycle**, which is the variant of the citric acid cycle. In plant seeds, the fatty acids are converted to carbohydrates by the glyoxylate cycle.

In leaves, during photosynthesis, CO2 is converted to carbohydrates by the **Calvin cycle**. In this process, occasionally a two-carbon phosphoglycolate is generated, which is recovered by peroxisomal enzymes.

Vacuoles

Structure

Vacuoles are found mainly in plant and fungal cells. The **tonoplast** or vacuolar membrane surrounds the organelle. These organelles contain water solutions of inorganic and organic molecules, including enzymes. Occasionally engulfed solids can be found in some vacuoles. The shape, size and structure vary depending on the needs of the cell. Plant cells, for instance, can have such a big vacuole that occupies even 80% of the cell volume. Vacuoles are evolved by the fusion of vesicles.http://en.wikipedia.org/wiki/Vacuole - cite_note-3

Function

The function of vacuoles varies according to cell type. The main function is storage: harmful materials are isolated or exported from the cell, and metabolic wastes can safely be stored within vacuoles. Vacuoles can also act as deposit for water and small molecules. **Protein bodies** are special vacuoles storing proteins needed for germination in plant seeds. Vacuoles also have a role in turgor (hydrostatical pressure) and pH maintenance, and also in the mechanical support of structures like flowers by means of the central vacuolar pressure.

Plastids

Structure

Plastids are double membrane-bound organelles found only in plant and algae cells. They contain pigments for photosynthesis, which make the cell colourful. Plastids synthetize and store molecules that are used by the cell. They have their own genome: a circular, double-stranded DNA molecule ('ctDNA' or 'cpDNA' or '**plastome'**). Proplastids (the progenitors of plastids) and young chloroplasts can divide by fission.

The main parts of a chloroplast are its three membrane systems (outer membrane, inner membrane, thylacoids), and the gel-like stroma within the inner membrane. There is also an intermembrane space between the inner and outer membranes.

Outer membrane

The outer membrane is a semi-permeable membrane which is permeable to small molecules and ions but acts as a barrier to large proteins. Chloroplast polypeptides are transported through by the TOC complexes (translocases of the outer membrane), similarly as in mitochondria.

Membranes of chromoplasts and amyloplasts can protrude into the cytoplasm, forming a **stromule** (stroma-containing tubule), increasing the surface area, or branching with the endoplasmic reticulum.

Inner membrane

This membrane borders the stroma, and regulates transport in and out of the chloroplast. The TIC complexes (translocase of the inner membrane) transport polypeptides into the stroma. An additional function of the inner membrane is the synthesis of fatty acids, lipids and carotenoids.

Stroma

The **stroma** is a gel-like, protein-rich fluid inside the inner membrane. The main biochemical process taking place in the stroma is the Calvin cycle (carbon fixation into sugar). Several other structures can be found floating in the stromal space, like the thylakoids, ribosomes, nucleoids and granules.

Chloroplast ribosomes synthesize part of the chloroplast proteins by transcribing mRNAs from the chloroplast DNA and translating them into proteins.

Starch granules are non-membrane bounded accumulations of starch in the stroma, growing, while sugar is synthetized, and consumed at night for respiration and sugar transport.

The most important stromal protein is the **Rubisco**. It is the enzyme responsible for CO_2 fixation into sugar molecules.

Thylakoid system

Thylakoids are membrane-bound sack-like structures in the stroma, containing chlorophyll. This is the place of light reactions of the photosynthesis. **Granal thylakoids** are flat sacks stacked in grana, linked by helical **stromal thylakoids**. Each **granum** contains 10-20 thylakoids, with continuous thylakoid space enclosed in them. Important enzyme complexes of the photosynthetic light reaction are embedded in the thylakoid membranes. **Photosystem II and I** (PS II and PS I) are light-harvesting complexes containing chlorophylls and carotenoids. These complexes absorb light energy and use it to energize electrons. Proteins in the thylakoid membrane use these electrons to pump hydrogen ions into the thylakoid space, decreasing its pH. **ATP synthase**, a large protein complex, uses the H^+ concentration gradient to generate ATP (as H^+ flows into the stroma). There are various photosynthetic pigments in the photosystems, taking part in light absorption and transfer. The main pigment found in

all chloroplasts is the **chlorophyll a**, but there are other chlorophyll forms such as accessory pigments (chl b, c, d and f). Another group of accessory pigments is the **carotenoids**, like *phycoerythrin* giving red colour to the chloroplasts of red algae, helping the transfer and dissipation of excess energy. Other carotenoids, the red-orange β-carotene or the yellowish *xanthophylls* (eg. *zeaxanthin*), are also common accessory pigments. The third group of pigments includes **phycobilins** with pigments of all colours.

Function

The main function of plastids is performing photosynthesis. Other functions include storage of starch, or synthetize fatty acids. There are many different plastid variants, all differentiated from proplastids:

1. **Chloroplast**: green plastid performing photosynthesis
2. **Chromoplast**: different coloured plastids synthetizing and storing pigments
3. **Gerontoplast**: playing a role in the degradation of the photosynthetic apparatus during senescence
4. **Leucoplast**: colourless plastids, performing monoterpene synthesis or differentiating to other specialized forms:
   a) **Amyloplast**: starch storage
   b) **Elaioplast**: fat storage
   c) **Proteinoplast**: protein storage and modification
   d) **Tannosome**: tannin and polyphenol production

Plastid genome

The genome of a plastid is called '**plastome'**, ranging from 75–250 kilobase in length. Many similarities can be found between the characteristics of the plastome and the mitochondrial genome. Plastome exists as a circular DNA molecule in a variable copy number per plastid. Its 100 genes encode rRNAs, tRNAs, proteins acting in photosynthesis or plastid gene transcription and translation. However, most plastid proteins are encoded by nuclear genes, and plastid gene expression is heavily regulated by proteins coded by nuclear genes. Note that this is similar to what is seen in mitochondria: mitochondria have their own genome but many mitochondrial proteins are encoded by nuclear genes (see "Restricted autonomy of the mitochondrial genome" in Chapter 'Mitochondrion'). Plastid DNA is associated with proteins in so called "**plastid nucleoids"**, a structure similar to the nucleoid of Prokariotes, attached to the inner envelope membrane. A single plastid nucleoid can contain more than 10 DNA copies. During plastid development or conversion, morphology, size and location of the nucleoid change as well.

4.4. The cytoskeleton

Introduction

The cytoskeleton has a key role in the support and in the movements of the cells (cilia, flagella). The cytoskeleton is composed of three well-defined filamentous structures such as microtubules, microfilaments, and intermediate filaments that together form an elaborate interactive network. The cyctoskeletal filaments are composed of protein subunits that are attached to each other by noncovalent bonds (see table from lecture).

1. Microtubules are long, hollow, unbranched tubes composed of tubulin monomers.
2. Microfilaments are solid, thin structures, composed of actin.
3. Intermediate filaments are tough, ropelike fibers composed of a variety of related proteins.

Functions of the cytoskeleton

The cytoskeleton is found in every eukaryotic cell. The functions of the cytoskeleton can be varied by cell type (see figure from lecture).

1. Serves as a scaffold providing structural support and maintaining cell shape.
2. Serves as an internal framework to organize organelles within the cell.
3. Directs cellular locomotion and the movement of materials within the cell.
4. Provides anchoring site for mRNA.
5. Serves as a signal transducer.
6. An essential component of the cell's division machinery.
7. Force-generating apparatus (moves cell from one place to another by the action of cilia and flagella).

Structure and composition of microtubules

(see figure from lecture)

1. Hollow, cylindrical structures
2. Composed of globular proteins, tubulin heterodimers (α- and β-tubulin)
3. The tubulin dimers organized as a linear array
4. The α-tubulin monomer has a bound GTP which is not hydrolyzed and nonexchangeable
5. The β-tubulin has a bound GDP which is exchanged for a GTP for the assembly of the filament
6. Outer diameter is 25 nm
7. Wall thickness is 4 nm
8. Tubulins are arranged into protofilaments
9. 13 protofilaments form a microtubule by noncovalent bonds
10. Protofilaments are asymmetric
11. Protofilaments have a polarity
12. Microtubules have plus and minus ends

Microtubule-associated proteins

MAPs comprise a heterogeneous group of proteins that plays a role in the building up of a microtubular network. They attach to the surface of microtubules to increase their stability and promote their assembly. One domain of the MAP is attached to the side of a microtubule and to another domain that projects outwards as a filament from the microtubule surface (see figure from lecture). Some MAPs form cross-bridges connecting microtubules to each other, thus maintaining their parallel alignment. The microtubule-binding activity is controlled by the phosphorylation and dephosphorylation of particular amino acid residues.

Tau is a microtubule-associated protein that is found in neurons. The overphosphorylation of this protein leads to the formation of neurofibrillary tangles, which play a role in the development of Alzheimer's disease. These phosphorylated tau proteins are not able to bind microtubules.

Functions of microtubules

(see figure from lecture)

1. determining and maintaining the shape of the cell
2. maintaining the internal organization of the cell (maintains the positions of the cellular organelles)

3. axonal transport
4. axonal growth in embryogenesis
5. plant cells: maintaining the shape by influencing the synthesis of the cell wall (cellulose and microtubules are coaligned)

Axonal transport

The transport of the materials in membrane coated vesicles depends on the microtubular network. The neurotransmitters are transported from the ER and the Golgi apparatus in vesicles (from the cell's body towards the periphery). The axon is the projection of a neuron which play a role in the communication between the neurons by synapses. Axons are filled with cytoskeletal elements, including bundles of microfilaments, intermediate filaments (called neurofilaments) and microtubules (see figure from lecture). The vesicles filled with neurotransmitters move along the microtubules of an axon, either towards or away from the cell body. These movements are mediated primarily by microtubules, which serve as tracks for a variety of motor proteins that generate the forces required to move objects within a cell (see figure from lecture).

The transport vesicles move in two directions. If they travel from the cell body towards the terminals of a neuron, they are said to move in an anterograde direction. While, if they move from the axon terminals (synapse) towards the cell body, it is called retrograde transport.

Motor proteins of the microtubules

Motor proteins generate the energy for material transport on the surface of the microtubules or microfilaments. As they are molecular motors, they can convert the chemical energy (from ATP) into mechanical energy. This way, they can generate force to move the cargo attached to the motor protein. Molecular motors move unidirectionally along their cytoskeletal track in a stepwise manner. Motor proteins can be classified into three groups:

1. kinesins- microtubules
2. dyneins- microtubules
3. myosins- microfilaments

There are no known motor proteins that use intermediate filaments as tracks. The motor proteins undergo a series of conformational changes that constitute a mechanical cycle. The steps of the mechanical cycle are coupled to the steps of the chemical or catalytic cycle, which provide the energy necessary to the motor activity.

Chemical cycle:

a) binding of an ATP molecule
b) hydrolysis of ATP
c) release of ADP+Pi products
d) binding a new ATP molecule

Kinesins- structure and function

The first kinesin, kinesin-1 protein, was discovered in 1985 by Ronald Vale. It is only one member of the kinesin-like protein superfamily that contains at least 45 different KLPs. Some members of the KLP superfamily are not motor proteins but adaptor proteins which play a role in making interactions between the kinesin and the vesicle. Kinesin is a tetramer constructed from two identical heavy chains and two identical light chains (see figure from lecture). A kinesin molecule contains a pair of globular heads that are responsible for bindig of a microtubule and an ATP molecule. The head (or motor do-

main) hydrolyses the ATP, producing energy for the movements. The head group is connected to a neck, a rodlike stalk, and a fan-shaped tail. The tail (made by two heavy chains and two light chains) binds cargo.

Kinesin is a plus end directed microtubular motor protein, which means that it moves towards the plus end of the microtubule facing to the plasma membrane. The kinesin molecule moves along a single protofilament of a microtubule at a velocity proportional to the ATP concentration (the maximal speed is 1 μm per second). The protein moves in distinct steps on the surface of the filament. Each step is approximately 8 nm in length, which is equal to the length of one tubulin dimer. One step requires the hydrolysis of an ATP molecule.

The movement is *processive*, which means that the motor protein moves along an individual microtubule for a long distance without falling off (~1μm). The two heads of a kinesin molecule work together in a coordinated manner, so that they are always present at different stages in their chemical and mechanical cycles at any given time (see figure from lecture):

1. one head binds to the microtubule
2. conformation changes in the adjacent neck region of the motor protein
3. the other head moves forward and towards the next binding site on the protofilament
4. the first head finds its precise binding site on the protofilament
5. the head and neck portions of a monomeric kinesin heavy chain are associated with a microtubule.
6. hydrolysis of ATP leads to a swinging movement of the neck

Dyneins- structure and function

Dynein is the first microtubule-associated motor protein that was discovered in 1963. This protein is responsible for the movement of cilia and flagella. The cytoplasmic form of dynein was discovered 20 years later. Cytoplasmic dynein is present in animal eukaryotes but not in plants. Cytoplasmic dynein is a huge protein (1.5 million daltons) composed of two identical heavy chains and a variety of intermediate and light chains (see figure from lecture). The heavy chains consist of a large globular head with an elongated projection (stalk). The dynein’s head is the catalytic core of the protein: it binds and hydrolyses the ATP molecule. The stalk binds the microtubule by a special binding site situated at its tip. The longer projection in the dynein is the stem (or tail) that binds the intermediate and light chains, the functions of which are not known.

Cytoplasmic dynein is a minus end directed motor protein, which means that the vesicles are transported towards the cell body or nucleus. It plays special roles in positioning the mitotic spindle, cetrosome and Golgi, segregating the chromosomes and moving organelles within the cell. Cytoplasmic dynein does not interact directly with membrane-bounded vesicles but requires an intervening multisubunit adaptor, called dynactin. Dynactin also regulates dynein activity and helps bind the motor protein to the microtubule (see figure from lecture).

Microtubule organizing centers

These centers are responsible for microtubule nucleation.

1. Basal bodies: only found in cells with cilia or flagella. They have nine peripheral triplets of microtubules, and the ciliary doublet (in the centre) terminates where the cilia joins the basal bodies. Striated boundles of fibers (rootlets) anchor basal bodies to the cytoplasm.
2. Centrosome: contains two barrel-shaped centrioles surrounded by pericentriolar material (PCM). Centrioles are usually found in pairs. They are responsible for the initiation and organization of the microtubular cystoskeleton. Microtubules terminate in the PCM. Each

centriole has nine peripheral microtubule triplets (A, B and C tubule) but lacks a central pair. The pricentriolar satellites radiate away from the triplets (see figure from lecture).

The polarity of the microtubules is always the same: the minus end is associated with the centrosome, and the plus (or growing) end is situated at the opposite tip The growing end of a microtubule contains specific proteins that help attach the microtubule to a particular target, such as an endosome or Golgi cisterna or a condensed chromosome in a mitotic cell.

Microtubule nucleation and synthesis

The MTOC controls the number of microtubules, their polarity, the number of protofilaments that build up their walls as well as the time and location of their assembly. The MTOCs have a common protein component, γ- tubulin molecules which form a ring complex in the PMC (see figure from lecture). The insoluble fibers of the PCM serve as attachment sites for ring-shaped structures that have the same diameter as microtubules have (25 nm) and contain γ-tubulin. The γ-tubulin ring is the minus end of the microtubule.

The microtubule structure composed of more than one protofilament allows the ends to be dynamic and the filaments to be resistant to thermal breakage. Tubulin dimmers are attached to the plus end of the microtubule. For the attachment of one tubulin dimer, hydrolysis of one GTP molecule is necessary. The GTP that is bound to the α-tubulin monomer is trapped at the dimer interface and is never hydrolyzed or exchanged. So the GTP is the integral part of the heterodimer. The nucleotide on the β-tubulin, being either the GTP or the GDP form, is exchangeable. The assembly of tubulin dimers requires a GTP molecule to be bound to the β-tubulin subunit. It is known that GTP hydrolysis is not required for the incorporation of the dimer onto the end of a microtubule. On the contrary, the GTP is hydrolyzed to GDP shortly after the dimer is incorporated into a microtubule, and the resulting GDP remains bound to the assembled polymer. When a dimer is released from a microtubule during disassembly and enters the soluble pool, the GDP is replaced by a new GTP (see figure from lecture). Thus, the tubulin dimer can be reused in the synthesis of the microtubules.

Disassembly of microtubules is affected by temperature, elevated Ca concentration and chemicals (colchicine, vinblastine, vincristine, nocodazole, and podophyllotoxin). The dynamic instability of microtubule depends on the replacement of GDP to a new GTP. If the microtubule is closed, GDP- β-tubulin dimers remain at the plus end of the tubule. The GDP bound tubulin has a curved conformation, so it is less able to fit into the straight protofilament. Thus the protofilament curves outward and undergoes a catastrophic shrinkage (see figure from lecture).

Intermediate filaments

They are solid, unbranched filaments their diameter is 10–12 nm. To date, intermediate filaments (IFs) have only been identified in animal cells. These filaments are strong, flexible ropelike fibers, they provide mechanical strength to cells against physical stress in neurons, muscle cells and the epithelial cells. Intermediate filaments are a chemically heterogeneous group of structures that, in humans, are encoded by approximately 70 different genes. The polypeptide subunits of IFs can be divided into five major classes based on the type of cell in which they are found (see table from lecture). Type I and type II IFs are the acidic and basic keratins, type III IFs are found in mesenchymal cells, muscle cells, neurons and glial cells, type IV IFs are found only in neurons, called neurofilaments and type V IFs are the lamins found in the nuclear lamina.

All IFs have similar structure and function; all of them contain a central rod-shaped, α-helical, fibrous domain. This central domain play a crucial role in the assembly of the intermediate filaments.

Assembly of IFs (see figure from lecture):

1. Two monomers (monomer has two globular, α-helical domains) are attached to each other in parallel orientation and form a dimer
2. Two dimers are connected to each other in antiparalel orientation and form a tetramer that lacks polarity
3. Eight tetramers attach to each other in a lateral arrangement to form a 60 nm long filament
4. These 60 nm long units bind to each other and form a polymer called intermediate filament

The assembly of an intermediate filament does not require energy (ATP/GTP). The intermediate filaments are less sensitive to chemicals and they are insoluble due to their structure. The assembly and disassembly of the IFs are regulated by the phosphorylation and dephosphorylation of the subunits. The intermediate filaments form a dynamic network in the cytoplasm of the cell (see figure from lecture). They are interconnected to the other types of cytoskeletal fibers. Connections to microtubules and microfilaments are primarily established by members of the plakin family of proteins, such as the dimeric plectin molecule.

Microfilaments

They are very long filaments with a diameter of 8 nm. They are composed of G-actin (globular actin monomers) subunits and they form an F-actin (filament). Microfilaments interact with heavy meromyosin. The cortex of most cells contains a thick bundle of microfilaments. Microvilli and cilia on the surface of the cells also contain microfilaments. Microfilaments are abundant in muscle cells called thin filaments, together with myosins (thick filament) they are responsible for muscle contraction. Microfilaments play a role in locomotion, movements of the cilia and flagella and cytokinesis (see figure from lecture).

An actin subunit consists of four domains (see figure from lecture). The ATP binding cleft is in the middle of each domain. In the presence of ATP, actin monomers polymerize to form a flexible, helical filament. As a result of its subunit organization an actin filament is essentially a two-stranded structure with two helical grooves running along its length. Because each actin subunit has polarity, the entire microfilament also has polarity. The two ends of an actin filament, plus and minus ends, have different structures and dynamic properties.

Assembly of actin filament (see figure from lecture):

1. nucleation: actin monomers form a complex and G-actins incorporate at both ends of the complex, but the plus end binds monomers 10 times faster then the other end.
2. elongation: if ATP concentration drops the monomers bind only to the plus end of the filament.
3. Disassembly of the filament happens on the minus end of the filament.

In an *in vitro* assembly the polymerization and depolymerization reach a steady state (when ATP concentration reaches of 0.3 μM) which means that equal amount of actin monomer builds into the plus end of the filament that dissociates from the minus end (see figure from lecture). Cells maintain the equilibrium between the monomeric and polymeric forms of actin, so the production of microfilaments is a dynamic process.

Motor proteins of microfilaments

Myosin was first isolated from mammalian skeletal muscle tissues but it is also found in protists, plants, nonmuscle cells of animals, and vertebrate cardiac and smooth muscle tissues. All myosins share a characteristic motor (head) domain. The head contains a binding site for actin filaments and a

catalytic site that binds and hydrolyzes ATP to generate energy for movements. The tail domains are highly divergent: contains a variety of light chains. Myosins are generally divided into two groups:

1. Conventional (type II) myosins:
   a) They generate force in muscles and some nonmuscle cells.
   b) Each myosin II is composed of two heavy chains, two light chains, and two globular heads (catalytic sites).
   c) All of the machinery required for motor activity is contained in a single head.
   d) The tail portion plays a structural role allowing the protein to form filaments.

2. Unconventional (type I) myosins:

   a) They have only a single head and are unable to assembly into filaments *in vitro*.
   b) Myosin I's precise role in cellular activities is unclear.
   c) Myosin V is involved in organelle transport.
   d) Several of them are associated with cytoplasmic vesicles and organelles.

The human genome encodes 16 different myosin II heavy chains, three of which function in nonmuscle cells. Among their nonmuscle activities, myosins are required for splitting a cell in two during cell division, generating tension at focal adhesions, cell migration, and the turning behavior of growth cones. Type II myosins consist of a pair of globular heads that contain the catalytic site of the molecule, a pair of necks, each consisting of a single, uninterrupted α-helix and two associated light chains, and a single, long, rod-shaped tail (see figure from lecture). The unconventional myosins were discovered in 1973. Although the function of myosin I is unclear myosin V is a special motor protein that plays a role in vesicular transport. It is a dimeric protein with two globular heads that moves processively along actin filaments (see figure from lecture). Processivity is due to the high affinity of the myosin heads for the actin filament. Myosin V has a special long neck (23 nm), it is about three times longer than that of myosin II. Because of its long neck, it can take very large steps.

Several unconventional myosins (myosin I, V, and VI) are associated with various types of cytoplasmic vesicles and organelles. Some vesicles contain microtubule-based motors (kinesins and/or cytoplasmic dynein) and microfilament-based motors (unconventional myosins) too and, the two types of motors are linked to one another. The cytoskeletal elements cooperate with each other. The motor proteins are able to pass one vesicle to another motor protein and this way the transport of the vesicle is continuous even to anterograde or retrograde directions (see figure from lecture).

Actin-binding proteins

Actin filaments in nonmuscular cells are organized into a variety of patterns, including bundles, thin (two-dimensional) networks, and complex three-dimensional gels. The organization and behavior of actin filaments are determined by more than 100 different actin-binding proteins that affect the localized assembly or disassembly of the actin filaments, their physical properties, and their interactions with one another and with cellular organelles.

Major types of actin-binding proteins:

1. Nucleating proteins: promote nucleation of actin filaments. Arp2/3 complex contains two actin-related proteins, the two Arps adopt a conformation that provides a template to which actin monomers can be added
2. Monomer sequestering proteins: thymosins are proteins that bind to actin-ATP monomers and prevent them from polymerizing. They shift the monomer–polymer equilibrium in a cell and determine whether polymerization or depolymerization is favored at the time.
3. End-blocking (capping) proteins: regulate the length of actin filaments by binding to one or the other end of the filaments, forming a cap that blocks both loss and gain of subunits.

4. Monomer polymerizing proteins: Profilin is a small protein that binds to one side of the G-actin and it catalyzes the dissociation of its bound ADP, which is rapidly replaced with ATP.
5. Actin filament depolymerizing proteins: cofilin, ADF, and depactin bind to actin-ADP subunits. They fragment actin filaments, and promote their depolymerization. These proteins play a role in the rapid turnover of actin filaments.
6. Crosslinking proteins: alter the 3D organization of the actin filaments. These proteins have two or more actin-binding sites and therefore can crosslink two or more separate actin filaments (filamin).
7. Filament severing proteins: bind to the side of an existing filament and break it in two. Severing proteins (e.g., gelsolin) also can promote the incorporation of actin monomers by creating additional free barbed ends.
8. Membrane binding proteins: interlink the actin filaments to the plasma membrane indirectly, by attaching them to peripheral membrane proteins. The plasma membrane moves together with the contractile filaments and with the cytoskeleton (e.g. locomotion, Shape of a cell, phagocytosis).

5. STRUCTURE OF DNA, GENES, CHROMOSOMES, CHROMATIN, GENOME

5.1. INTRODUCTION

In the late 40's the researchers discovered that DNA is the genetic material which is responsible for inheritance. It was very difficult to reveal how DNA could transmit traits as it is composed of only four subunits. The X-ray diffraction analysis of DNA showed that the polymer was composed of two strains and formed a helix. The detailed and proper structure of DNA was described in 1953 by Watson and Crick.

5.2. THE STRUCTURE OF DNA

Deoxyribonucleic acid molecules consist of four nucleotide bases: adenine (A), thymine (T), cytosine (C), guanine (G). Adenine and guanine are the purine bases, cytosine and thymine are the pyrimidine bases (see figure from lecture).

A nitrogenous base with a ribose is called nucleoside. The nucleoside with a phosphate group is called the nucleotide (see figure from lecture). The way, in which the nucleotide subunits are linked together, gives a DNA strand a chemical polarity. The 5' end of the nucleotide (phosphate group) is attached to the 3' end –OH group of the sugar molecule by phosphodiester bond.

The base composition of DNA

In 1950 Erwin Chargaff discovered that the base composition and the ratios of the four component bases were quite different from one type of organism to another. Chargaff found out that the number of adenines was equal to the number of thymines and the number of cytosines equaled the number of guanines. However, the sum of A and T was not equal to the sum of C and G.

The Watson-Crick proposal

(see figure from lecture)

1. The molecule is composed of two chains of nucleotides.
2. The two chains spiral around each other to form a pair of right-handed helices.
3. The two chains comprising one double helix run in opposite direction so they are antiparallel. One chain is aligned in the 5' to 3' direction; its partner must be aligned in the 3' to 5' direction.
4. The sugar–phosphate backbone of each strand is located on the outside of the molecule with the two sets of bases projecting towards the center.
5. The phosphate groups give the molecule a negative charge.
6. Hydrophobic interactions and van der Waals forces between the bases provide stability for the entire DNA molecule.
7. The two strands are held together by hydrogen bonds between each base of one strand and an associated base on the other strand (A=T; G≡C).
8. The distance from the phosphorus atom of the backbone to the center of the axis is 1 nm.
9. The width of the double helix is 2 nm.
10. A pyrimidine base is always paired with a purine base (A-T; C-G).
11. The spaces between adjacent turns of the helix form two grooves of different width: a wider major groove and a narrower minor groove. Proteins that bind to DNA often contain domains that fit into these grooves.

12. The double helix makes one complete turn every 10 residues (3.4 nm).
13. The nucleotide sequences of the two strands are always fixed relative to one another. The two chains of the double helix are complementary to one another.
14. DNA is responsible for inheritance. It contains the genetic information and stores codes for the right order of amino acids. It determines the amino acid sequences of the proteins. DNA must contain the information for synthesis of new DNA strands. It is also a director of cellular activity. It is responsible for the expression of the genetic message.

DNA supercoiling

DNA molecules in the nucleus are in compact forms. The molecules are twisted upon themselves. This state of DNA is the supercoiled state. This way DNA is more compact, it can fit in the nucleus and move faster under the extracellular effects (electric field, centrifugal force).

1. Relaxed DNA: has standard number of 10 base pairs per turn of the helix.
2. Underwound DNA: has a greater number of base pairs per turn of the helix.

DNA is more stable with 10 base pairs per turn, thus it tends to twist upon itself and forms supercoiled DNA. An underwound molecule spontaneously assumes a negatively supercoiled conformation (see figure from lecture). DNA is referred to as *negatively supercoiled* when it is underwound and *positively supercoiled* when it is overwound. The negatively supercoiled DNA exerts a force that helps separate the two strands of the helix, which is required during both replication and transcription.

Topoisomerases

The enzymes that are responsible for supercoiling are called topoisomerases. They have two types:

1. Type I. topoisomerases: they make a transient brake in strand of the DNA duplex. The enzyme cleaves one strand of the DNA and then allows the intact, complementary strand to undergo a controlled rotation, which relaxes the supercoiled molecule. It is essential for DNA replication and transcription.
2. Type II. topoisomerases: they make two transient breaks in both strands of DNA duplex. The action of the enzyme is seen in figure from lecture. The functions of topoisomerase II:
   a) supercoiling-relaxation
   b) catenation-decatenation (interlink)
   c) knotting-unknotting (tie) (see figure from lecture)

5.3. Genes and Chromosomes

(See Introduction to genetics for more details)

Characteristics of organisms are regulated by units of inheritance called genes. Each trait is controlled by two forms of a gene called *alleles*. Alleles could be identical or nonidentical. When alleles are nonidentical, the dominant allele masks the recessive allele. The genes are found on chromosomes. The chromosomes were first discovered in dividing cells by German researchers in 1880. Later, in 1903, Walter Sutton discovered that the chromosomes are the physical carriers of the genetic factors. Chromosomes consist of *chromatin* fibers, composed of DNA and associated proteins. Each chromosome contains a single, continuous DNA molecule.

Telomeres

(See figure from lecture)

Telomer is a specific region of a chromosome which is found at the end of each DNA molecule. It consists of unusual repeated DNA sequences (from 500 to 5000 copies). Telomeres have conserved functions in vertebrates. Several special DNA-binding proteins are able to attach specifically to the telomer region of the chromosome. The overhanging 3' end of the DNA turns back and forms a loop that protects the telomer (See DNA replication for more details). In 1984 Elizabeth Blackburn and Carol Greider discovered a new enzyme called telomerase which was able to put new nucleotides/repeats to the 3' end of the telomer. This way the DNA polymerase uses the newly synthesized repeats as template for the 5' end of the complementary DNA strand. So the telomerase enzyme is a reverse transcriptase that synthesizes DNA from RNA templates. Usually the enzyme itself holds the RNA template for synthesis. The telomer region has different functions:

1. complete replication of the chromosome
2. protects the chromosome from nucleases
3. prevents from fusing with another DNA molecule

Without telomerases the telomer region of the chromosome would become shorter and shorter with each cell division. As the activity of telomerase enzyme becomes lower and lower in each cell division, the cell will reach that point when the length of the telomer sequence is not enough (can not be shorter) for the next cell division, thus the cell dies (aging).

Centromere

The centromere region of a chromosome is the site where DNA is indented (See figure from lecture). It contains tandemly repeated (171 base pairs) DNA sequences (satellite DNA- See in An introductions to genetics) and a protein containing structure called kinetochore. The kinetochore is the attachment site of spindle fibers (microtubules) during mitosis and meiosis.

5.4. Chromatin

(See An introduction to genetics for more details)

The chromatin is the complex of DNA and proteins. Its structure differs in the different stages of the cell cycle: less compact during DNA synthesis (S phase) and the most compact during cell division (M phase). During the interphase two types of chromatin can be determined: euchromatin and heterochromatin.

1. Euchromatin:
    - active DNA
    - transcription occurs here
    - exhibits normal pattern of condensation and relaxation during cell cycle

2. Heterochromatin:
    - simple sequence DNA
    - centromere and telomere regions
    - often associated with the nuclear envelope
    - inactive DNA/permanently silenced

    b) Constitutive heterochromatin:
        - never expressed throughout the life cycle
        - repeated sequences

c) Facultative heterochromatin:
d) sometimes expressed
    - inactivated during special phase of life cycle or
    - certain types of differentiated cells

Condensation of DNA into chromosomes

The condensation of chromatin into chromosome requires specific proteins. These proteins are the *histones* that form the basic element of condensation and *nonhistone proteins* that help in the further compaction of DNA. The histone proteins have five major classes: H1, H2A, H2B, H3, and H4 based on their size and amino acid composition (See table from lecture). The primary level of organization is the *nucleosome* (See figure from lecture). The nucleosome is a complex of eight histone proteins (two each of H2A, H2B, H3, H4). Each nucleosome has a nucleosome core particle consisting of 146 base pairs of supercoiled DNA sequence. Type H1 histone resides outside the nucleosome and binds near the site where DNA exits the nucleosome. So H1 is a linker between the nucleosome core particle and a linker DNA fragment that interlinks the nucleosome with the next one. The length of linker DNA is 53 base pairs.

Characteristics of histone and nonhistone proteins

1. Hstone proteins:
   a) positively charged, basic (lysine and arginine amino acids)
   b) abundant
   c) found in all cell types
   d) conserved function
   e) provide the first level of packaging (nucleosome)

2. Nonhistone proteins: associated with chromosomes
   a) several Types
   b) amount of them is variable with time and cell's type
   c) interact with DNA or with other proteins (e.g. histones)
   d) negatively charged, acidic

Further condensation of chromatin

The second level of condensation is the "beads on a string" form. In this structure the nucleosomes are connected to each other with linker DNA (see figure from lecture). Then the nucleosomes are packaged into 30 nm chromatin fiber by the action of H1 protein. The exact role of this protein is not understood, but its presence is necessary for the further compaction of DNA. The nucleosomes are coiled and form solenoids (solenoid model). A solenoid contains six nucleosomes and three solenoids giving the 30 nm wide fiber (see figure from lecture). The linker DNA gently curves as it connects consecutive core particles, which are organized into a single, continuous helical structure containing about 6–8 nucleosomes. The other theory for DNA condensation is the zig-zag model. In this model the linker DNA is present in a straight, extended state that criss-crosses back and forth between consecutive core particles, which are organized into two separate stacks of nucleosomes.

The condensation continues with the loop domain formation. The loop domains form a 300 nm wide chromatin from the 30 nm wide fibers. A human chromosome contains approximately 2000 loop domains. The DNA loops are attached to the nonhistone protein scaffold and, on the other hand, special DNA regions, matrix attachment regions (MARs), are interlinked to the nuclear matrix (see figure from lecture). Type II topoisomerases are also found in the scaffold and they regulate the level DNA of supercoiling. When cells prepare for mitosis, the loops become further compacted into mitotic chromosomes (metaphase chromosome) (see figure from lecture).

5.5. GENOME

(See An introduction to genetics and Human genome project for more details)

The complexity of the genome

The important properties of the DNA duplex are the abilities to separate and reanneal. The separation is called *denaturation*. The DNA *melting* is the process of denaturation due to the elevating temperature. The denaturation of DNA can be monitored by following the absorbance of UV light at 260 nm wavelength (see figure from lecture). The temperature at which the shift in absorbance is half completed is termed the melting temperature (Tm): the higher the GC content of the DNA, the higher the Tm. This increased stability of GC-containing DNA reflects the presence of three hydrogen bonds between the bases as compared with AT pairs.

The other property of the DNA is that the two separated strands are able to reassociate as the complementary base pairs find each other in the two chains. This process is called reannealing or *renaturation*. The renaturation can be monitored by photometry. This feature of the DNA led to the development of hybridization techniques that became the basis of such methods like DNA sequencing, *in situ* hybridization, northern blot, southern blot etc.

The complexity of a genome can be determined by the denaturation and renaturation of the DNA strands. These processes are influenced by the ionic strength of the solution, the incubation temperature, the concentration of DNA, the period of incubation, and the size of the interacting molecules. If we compare the renaturation curves of different organisms e.g. viruses and bacterial cells, we can determine that the renaturation takes more time, if the genome is larger (see figure from lecture): the smaller the genome, the faster the renaturation.

The eukaryotic genome is more complex. If we produce fragments from DNA duplexes, and determine the renaturation curves, we can see that the curves are not symmetrical similarly to prokaryotes or viruses. The reason is that the various nucleotide sequences in a preparation of eukaryotic DNA fragments are present at markedly different concentrations. The reannealing curve of a eukaryotic genome shows three distinct steps that correspond to the three classes of DNA: highly repeated fraction, moderately repeated fraction and nonrepeated fraction (see figure from lecture).

1. Highly repeated fraction:
   a) satellite DNAs
   b) minisatellite DNAs
   c) microsatellite DNAs

2. Moderately repeated fraction:
   a) retrotransposons (RNA)
   b) DNA transposons
   c) repeated DNA sequences with coding functions
   d) (rRNA, histones)
   e) repeated DNA sequences lacking coding functions:
      - SINE (short interspersed nuclear elements),
      - LINE (long interspersed nuclear elements),
      - LTR (long terminal repeats)

3. Nonrepeated fraction: - protein coding genes

Genes, gene families, pseudogenes

(see Human genome project and An introduction to genetics for more details)

If we compare the total chromosomal DNA per cell in various species it is evident that much of the DNA in certain organisms does not encode RNA or have any apparent regulatory or structural function. Most genes (90-95%) encode proteins. However, there are a significant number of RNA-only genes, and recent work has shown that RNA genes are far more important than previously thought. The total number of human protein-coding genes is estimated to be 30,000–35,000:

1. Solitary genes: represented only once in the haploid genome
2. Duplicated or diverged genes in gene families: genes have similar but not identical amino acid sequences
3. Tandemly repeated genes encoding rRNAs, tRNAs, snRNAs and histones

Pseudogenes are the nonfunctional copies of genes. They contain most of the gene's sequence, but have stop codons or frameshifts in the middle, or they lack promoters, or are truncated or are just fragments of genes:

1. Non-processed pseudogenes: result of tandem gene duplication or transposable element movement
2. Processed pseudogenes: come from mRNA that has been reverse-transcribed and then randomly inserted into the genome

Genome and gene expression

(See Human genome project and Transcription for more details)

For the transmission of the genome (genetic information) the cell has to copy its whole DNA content without making any mistakes. After DNA synthesis the cell has to segregate the nucleic acid molecules into the daughter cells. For the mechanisms of transmission and maintenance of the genome many proteins and processes are needed: DNA polymerases, replication of the DNA, control of replication and segregation during mitosis/meiosis. For the proper segregation of nucleic acids, DNA is tightly packed together with proteins (histones) into chromosomes. Genes are arranged linearly on the chromosomes, with intergenic DNA spacers.

If we examine the genome size of the different organisms, we can see that there is no simple correlation between the genome size and the complexity of organisms. The reason is that relatively few genes correspond to a considerable number of proteins, and the gene density in the genome is quite different in different organisms (see table from lecture). For example in humans, the number of base pairs is about 3 billion that corresponds to 29 thousand of genes, while the rice has only half billion base pairs but 59 thousand of genes.

When the cell needs proteins for its metabolism, the protein coding genes on the chromosomal DNA are copied into RNA. This process is the transcription of DNA to messenger RNA (see figure from lecture). These RNA molecules are used as templates for protein synthesis (translation). The flow of genetic information in cells is therefore from DNA to RNA to protein (central dogma of molecular biology). In the past few decades it was revealed that proteins are also able to regulate the expression of genes or the maturation of mRNA molecules (RNA splicing). Moreover, there are enzymes that can transcribe back the RNA into DNA. So the flow of the genetic information is bidirectional. Regulation of the gene expression is possible at each level:

- Chromosome
- Transcription

- RNA movement and processing
- RNA life-time
- Translation
- Post-translational processing
- Protein life time

The expression of the genes can be regulated in time and space and/or during development in different tissues or with different circumstances (nucleus vs mitochondrion).

6. HUMAN GENOME PROJECT - THE LARGEST MISSION OF THE 20TH CENTURY'S LAST DECADE

6.1. PRELIMINARIES

1953: Discovery of DNA structure (James Watson and Francis Crick)

1973: First published DNA sequence (24 bp, lac operator)

1982: Starting of GeneBank

1983: Development of PCR

1987: 1st automated sequencer

1990: Human Genome Project began

The 2nd period of the HGP is now in process. The first, commonly referred to as the pregenomic period, was completed in 2003 highlighting the sequencing of the last chromosomes. Today, in this post-genomic era, we try to understand the information. The aim of the program was to describe the sequence of the 3 billion base pairs (bp) of the human genome: from the first chromosomes short arms telomere to the end of the long arm of Y chromosome.

Genome is best defined as the haploid or diploid cells entire genetic material, the entire DNA content, the DNA in the nuclei and the DNA in the plastids (mitochondria or chloroplasts), however, it was not the aim to sequence the genome of mitochondria, because it was performed about 10 years prior to the Human Genome Project, in 1981.

The program launched in 1990, and at the time it was estimated researchers will require 30 years and 3 billion dollars to complete the project. The cost, in USD, to sequence 1 bp was $1.00 and researchers could sequence 100.000 bps within one year). The scientists planned the first phase of the project up to 15 years, because the development of the science first needed to be adequately designed and indeed, estimates were correct!

The main sponsor was the Department of Energy and the National Institutes of Health (NIH) of the United States of America. Other sponsors included the WelcomeTrust, French, German, Japanese and Chinese research institutions. Research centers were able to financially support the immense operation through the use of intelligent researchers and professional, reliable scientific equipment. They were identified cohesively as "Consortium".

The first director of the program was James Watson, the man credited with the description of the double helix structure of DNA.

The project included nine human genome to accurately sequence: eight men and one woman. Only 1 woman was included, because a women's genomic library contains two copies of the X chromosomes and does not have any information about the Y chromosome. In support of ethical measures, members of the research team were excluded as potential donors.

6.2. THE ORIGINS – THE HIERARCHICAL METHOD

First, the 100-200 bp marker sequences are identified in each 150.000 bps long sequence. Notably, this marker sequence must be unique and should occur only once in the genome.

Following identification, they are disseminated out to the genome into 150.000 bp long, overlapping pieces. These sections were cloned into BAC libraries, or bacterial artificial chromosomes. They selected the redundant/completely identical clones from the others. Next, they divided the 150.000 bp long segments further and sequenced the pieces.

They assembled the pieces together according to the overlapping sections and then by the markers. They knew the location and orientation of these BAC clones according to the marker sequences.

In 1998: the first high-performance sequencing was developed and established as a practice. It changed the meaning and essence of sequencing.

6.3. Celera Genomics

In 1998, a private company (Celera Genomics) was founded and it too began a “Human Genome Project”, however, the new group differed and therefore did not support the consortium’s operations, preferring to implement its own, independent research.

In the beginning, the consortium’s Human Genome Project had stated and considered highly relevant, the project cannot be profit-oriented, and the results should be taken into the public domain. In contrast, Celera intended to promote the information meaning, if and when Celera is prepared in advance, it intends to market the results, thus gaining financially. The President of the United States at the time, Former President Bill Clinton, intervened on the question: what is marketable and what should not be marketable. He believed the sequence far too valuable and therefore must not be marketable in any way, as it was considered public property.

Craig Venter, the manager of Celera Genomics in 2002, publicly remarked one of the samples in their research was his DNA, which was eventually combined with 3/5 of the entire sample.

Results: February, 2001

Due to the race between the Consortium and Celera Genomic, both companies published partial results. Two journals published the results and the data was considered nearly 95% complete. Nature published the data from the international consortium and Science published the data from Celera Genomics.

The interesting question was poised, ‘How could Celera Genomics perform identical scientific research in a fraction of the time, 1990-2001 vs. 1998-2001?’ The science clearly had developed. Celera Genomics did not use the conventional and time-consuming hierarchical system, they used the “shotgun method”, meaning they cut the whole genome into small pieces and sequenced both ends. Instead of using biomarkers, they used high-performance bioinformatics to assemble the overlapping short sequences.

The consortium completed the project in 2003.

Up until the 13th of December, 2013, this human genome was the human reference genome, however, at the end of 2013 a new assembly was released by the “Genome Reference Consortium”.

Results

The first surprising result: they found "only" about 20-25 thousand genes! This is only an estimate, since it is only sequences without functions. They expected in the beginning, at the very least at or about 80,000 genes. Today, we estimate the number may be more closer to 25,000. The number of

genes is only two times that of the Drosophila's. The average length of exons is 100-200 bps, the average length of introns is 1000-4000 bps. An average gene concludes 7-9 exons.

The "useful" part of the result is just 10%, but the actual protein encoding portion is just approx. 1.5-2%. The difference between persons are very high, 0.1%.

An average chromosome size is 130 Mb-s: 50-250Mbs, depending on the particular chromosome. The density of genes is not the same along the chromosomes, moreover, it is different on each chromosome, for example, the gene density of the 19th and the 22nd is much larger than the gene density of 4th or 18th chromosomes. The region under the telomere (the subtelomeric region) is rich in genes, but the region around the centromere and the Y chromosome long arm are free from genes.

The largest known human gene is the gene of dystrophin, 2.4 Mb with 79 exons.

It was found that many genes seem to be duplicated and modified during the evolution, and these genes are similar to one another and also produce gene families, e.g. Histones, Globins, Immunglobulins and the MHCII family. Additionally, non-coding pseudogenes may arise with duplication: a number of non-coding sequences were found, which were very similar to coding regions. They lost their function by a mutation after duplication.

The length of encoding genes is primarily determined by the length of introns. The exon-intron ratio may vary among genes, e.g. the ratio of introns is 99.4% in the gene of distrophin, 77% in the gene of ApoB (ApolipoproteinB). There are genes that do not contain introns in the human genome too, for example histones, most tRNAs, some hormone (dopamine, serotonin and α2 adrenergic receptor) genes.

The average distance between genes is 30 Kb (in contrast to prokaryotes where it is nearly 1-5 kb).

Structure of genome

For additional information, please see DNA structure, chromosome chromatine, genome chapter.

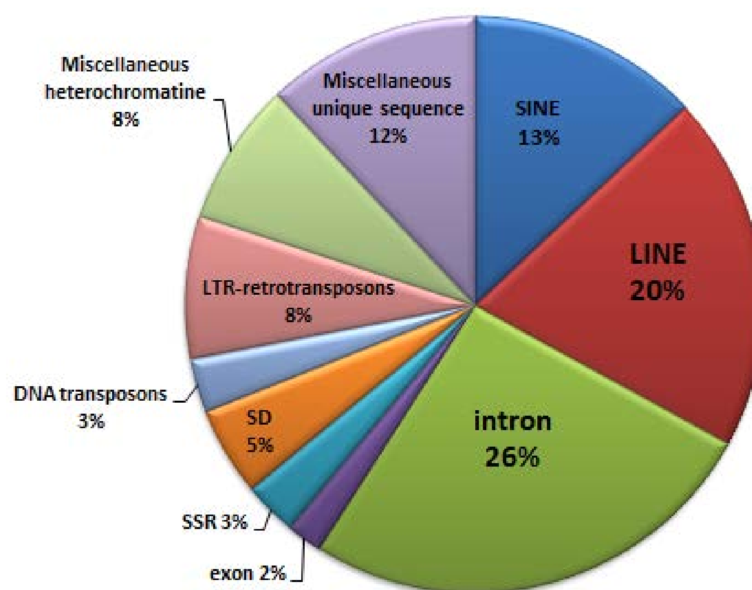


Figure 6-1 Structure of Genome
SINE: short interspersed nuclear elements; **LINE**: long interspersed nuclear elements; **SSR**: simple sequence repeats=microsatellites; **SD**: segmental duplications

Transposons = interspersed repetitive sequences: they are capable of displacement, so they are considerably instable DNA-s, which can change their location in the genome. They are classified due to the length of the repeating units.

SINE (short interspersed nuclear elements) are characterized by a large number of repeating units, for example, Alu sequences (the name originates from the AluI restriction enzyme and is what we use in detecting these sections). It is about 280 bp long, but it occurs in 700,000 -1,000,000 copies in the human genome.

LINE (= long interspersed nuclear elements) are 1.4-6.1 Kb in length, and they occur in 60,000 - 100,000 copies.

LTR-retrotransposons (=retroviral-like transposons) contain long terminal repeats (LTRs), encode reverse transcriptase and are similar to retroviruses. Length: 100bp-5kb, half a million copies.

Tandem repetitive sections: blocks formed by a limited number of repetitive units occurring only in certain chromosomal regions. It has three groups: satellite DNA, minisatellite DNA and microsatellite DNA.

Satellite DNA: it was named after the cesium chloride gradient centrifugation pattern: following the centrifugation of the whole genome, satellite DNAs made its own band, wide by comparison to other DNAs. They are small DNA fragments which constituted the majority of the non-coding heterochromatic regions around the centromere. The length of the repeating units is 5-171 bps, the size of the total block is 100Kb-1Mb. It does not translate. **They have an important role in cell division, notably, the connection of the drawstring in the centromeres.**

Minisatellite DNA: it does not translate, they are usually scattered in the genome. The length of the repeating section is only 6 -25 bps, while the assembly of these blocks can be about 0.1 to 20 kbs. They can be found in the telomeres and around the telomeres primarily.

Microsatellite DNA: The repeating unit is 1-4 bp long and make 150 bps or shorter blocks. The tri-and tetranukleotide repeating units are rare, but they are important in some diseases (trinucleotide repeat mutations and diseases - trinucleotide expansions as Huntington's disease). While the most mononucleotide repeat rows are A or T bases, and their length can be up to 10 Mbps, the dinucleotides are formed from CT/AG structure. The function of microsatellite DNAs is unknown, but many molecular biological methods are based on their use, for example DNA fingerprint, VNTR (variable number = Variable Number Tandem Repeats tandem repeats).

"Junk DNA"

After the first period of the HGP, Francis Crick (the describer of the structure of DNA) stated, in reference to nearly 99% of the genome of which is considered "*little better than junk*". (2003)

Today, we know it is a considerable failure and how immense differences these sections may cause within the gene expression. Following the recognition of the fact these sections are indeed important, the new goal is to understand these sections with genome-wide association studies.

ENCODE Project: Encyclopedia of DNA Elements – examine these sections

Hypothetical and proven functions of these sequences:

- Regulation of gene expressions
- Transcription factor sites
- Protection of genome

- Genetic “switches” (in gene expression, when and where)
- Enhancers
- Silencers
- Promoters
- Operators
- Insulators

7. REPLICATION

Replication is the process in which the genetic material of the cell gets duplicated prior a cell division.

7.1. FEATURES OF REPLICATION

- covers the whole genome
- efficient
- has high fidelity
- semiconservative
- bidirectional

7.2. PROKARYOTIC REPLICATION

Replication is highly conserved among organisms. Most steps are very similar in prokaryotic and eukaryotic cells.

Replication starts at an AT rich sequence called replication origin (or origin of replication). Bacterial genome origins and bacterial plasmid origins differ from each other. Origin recognition proteins initiate the replication process.

While the DNA is in double stranded, helical form, most of the enzymes can not bind to it. Topoisomerase I enzyme unwinds the helix structure. The tension is released from the coil by single strand cuts.

Helicase moves unidirectionally along the backbone of the DNA and separates the two strands by breaking the hydrogen bonds between the nucleotides. Single Strand Binding proteins (SSBs) attach to the separated strands inhibiting the renaturation.

Because all DNA polymerases need 3' end OH group to start the polymerisation, the primase enzyme places one RNA primer (oligonucleotide) on the leading strand and several on the lagging strand.

The actual synthesis of the new DNA strands is done by the DNA polymerase III holoenzyme. This enzyme featured by high processivity (it can synthesize long stretches of DNA at once, 5'-3' elongation activity and 3'-5' exonuclease activity. DNA synthesis of the leading strand is heading towards the replication fork. It is continuous. Synthesis of the leading strand is heading away from the replication fork. It is discontinuous, DNA polymerase III can elongate the DNA strand only until it reaches the previous RNA primer of the following Okazaki fragment.

Due to its 5'-3' exonuclease activity, DNA polymerase I can cleave the RNA primer, then it is able to synthesize the DNA strand due its 5'-3' elongation activity.

The DNA ligase joins the neighbouring Okazaki fragments by synthesizing the missing phosphodiester bond.

DNA topoisomerase II nicks the finished DNA strands and untangles the two circular chromosomes.

7.3. Differences between the prokaryotic and eukaryotic replication

While most of the processes are highly similar in eukaryotic and prokaryotic replication, there are several differences as well:

- Prokaryotes have only one, circular chromosome, what means that with only one replication origin it is possible to duplicate the whole genome. The eukaryotic chromosomes contain a lot of replication origins, the replication start at many places at once. This is necessary for two reasons: the eukaryotic genome contain more than one chromosome and the size of the whole eukaryotic genome is bigger by at least one magnitude than the prokaryotic genome.
- The histone proteins of the eukaryotic genome should be duplicated as well.
- Telomeres are present at the end of the linear eukaryotic chromosomes.
- Eukaryotic DNA polymerases lack 5’-3’ exonuclease activity

7.4. Telomeres and telomerases

The eukaryotic chromosomes are linear, primase enzyme cannot place the primer at the very end of the 5’ end of the chromosome. Due to this the chromosome shortened by every replication. To protect the gene coding sequences of the chromosomes, there are repetitive sequences at the end of the chromosomes, called telomeres. After each replication the telomeres will get shortened. After a given number of replications the telomeres are gone and the cell start to loose genes, what leads to cell death. This is one of the mechanisms behind physiological aging.

In reproductive cells and embryonic stem cells telomerase is active, it rebuilds the telomeres. Other dividing cells lack telomerase activity.

Malignant cells almost always gain mutations what activates the telomerase enzyme, making the immortal.

8. DNA REPAIR

8.1. INTRODUCTION

Our genome is always changing, but because most of the mutations are harmful, many DNA repair systems are present in our cells. DNA damage can be caused by radiation, genotoxic chemicals, errors in replication or by just random deterioration of the DNA molecule. The Nobel Prize in Chemistry 2015 was awarded jointly to Thomas Lindahl, Paul Modrich and Aziz Sancar “for mechanistic studies of DNA repair”.

8.2. TYPES OF DNA DAMAGE

Deamination: Due to random deamination of the base, cytosine can convert into uracil. During the next replication uracil will pair with adenine, changing the original CG to TA.

Depurination: A nucleotide can lose its purine base, resulting in random base substitution or deletion on the opposite strand.

Alkylation: Guanine can receive a methyl group converting it into a O^6-methylguanine, which will pair with T instead of C. During the next replication, the original GC will be changed into AT.

Pyrimidine dimers: UV irradiation forms covalent bonds between adjacent thymine nucleotides. It distorts the conformation of the double helix that inhibits replication.

Double strand breaks: High energy gamma or ionizing radiation breaks the backbone of both DNA strands.

Mismatch: During replication the DNA polymerase incorporates a non-complement nucleotide.

8.3. REPAIR MECHANISMS

Photoreactivation In many prokaryotic, fungal and animal cells pyrimidine dimers can be reverted directly into separate pyrimidine nucleotides by the photolyase enzyme. This reaction is dependent on the external energy of visible blue light, hence its name photoreactivation. In humans photolyase is not active, nucleotide excision repair will correct the pyrimidine dimers.

Base excision repair (BER) Base excision repair removes altered, single nucleotides which distort the structure of the double helix. DNA glycosylase enzyme recognizes the erroneous nucleotide and cleaves the base from the sugar. There are different DNA glycosylase enzymes specific for different alterations. After the cleavage of the base, an endonuclease removes the ‘empty’ sugar from the DNA backbone, then a DNA polymerase fills the gap by inserting a new nucleotide which is complementary to the undamaged strand. The missing phosphodiester bond is made by a DNA ligase enzyme.

Nucleotide excision repair (NER) NER repairs bulky errors, such as thymine dimers or chemically altered nucleotides. There are pathways for checking these types of errors:

- Transcription-coupled pathway which repairs frequently transcribed (thus important) genes.
- Global genomic pathway which removes errors from the rest of the genome, but checks each location less frequently.

Endonucleases cut the DNA strand on both sides of the error, then the damaged strand is removed by a helicase enzyme. The gap is filled by a DNA polymerase, then the missing phosphodiester bond is made by a DNA ligase enzyme.

Mismatch repair (MMR) Mismatch is an error made by the DNA polymerase during DNA replication. The challenge of this type of repair is that while the mismatching nucleotides distort the double helix, but both of these nucleotides are chemically sound. The original, supposedly error-free strand is recognised by its DNA methylation pattern, which the newly synthesized strand lacks. Endonucleases and helicases remove the DNA strand containing the erroneous nucleotide, than a DNA polymerase fills the gap and a DNA ligase seals the backbone.

Double strand break repair by homologous recombination This repair requires an unbroken, homologous DNA strand to act as a template. In diploid organisms the homologous chromosome acts as the template. The repaired DNA strand is without a gap. The Crispr/Cas9 system is a promising tool for directional genome editing.

Double strand break repair by non-homologous end joining In the case of the lack of a homologous sequence, a special DNA ligase will fix the broken DNA strands. During this process, nucleotides are cleaved from both broken strands, results a ‘gap’ in the sequence.

SOS response A ‘last chance’ pathway in prokaryotic cells. This mechanism does not repair the mutations per se, but let’s finish the replication.

8.4. MEDICAL IMPORTANCE OF DNA REPAIR

Any error in one of the repair mechanisms itself leads to increased mutation rate. A classic example is xeroderma pigmentosum, an error in the nucleotide excision repair. These patients suffer from the consequences of UV radiation induced DNA damages. Their skin burns very quickly in direct sunlight, and covered with freckles. The chance of melanoma formation is increased in these patients.

9. TRANSCRIPTION

9.1. IN GENERAL

Transcription is the process in which the information stored in DNA in the form of nucleotide order is copied into RNA molecules. RNA has the same "language" i.e. nucleotide order as DNA.

There are notable similarities and differences between **replication (DNA synthesis) and transcription**.

Similarities:

- DNA serves as a template
- RNA polymerase which is the enzyme of transcription carries out a similar enzymatic reaction than DNA polymerase
- RNA and DNA polymerases are processive enzymes

Differences:

- Only one strand of DNA is the template
- Only certain regions of DNA are copied
- RNA polymerase exhibits only moderate proofreading activity
- The structure of RNA polymerase differs from the structure of DNA polymerase
- No need for a primer in support of the synthesis

Types of RNA

Major categories are: mRNA, rRNA and tRNA. mRNA (**messenger**) carries the information of the DNA to the ribosomes, the site of protein synthesis (translation). Distinctively, rRNAs (**ribosomal**) are the structural and functional nucleic acid components of the ribosomes. Additionally, tRNA (**transfer**) provides a connection between mRNA codes and amino acid order of a newly synthetized protein (future classes of RNAs are in the lecture).

Transcription features **three stages**: initiation, elongation and termination. They have similar and dissimilar features in prokaryotes and eukaryotes, the main differences are in the recognition of the initiation and termination signals. Transcription is **regulated** intensively in every cell, at different levels, especially at the initiation stage.

Transcription in prokaryotes

RNA polymerase

Prokaryotes have only one type of **DNA dependent RNA polymerase**, this enzyme synthetizes all of the RNA transcripts. The enzyme has five subunits (2 α, 1 β, 1 β', 1 Ω) to make up the **core enzyme**. This core enzyme can associate with the DNA double helix only loosely. Site specific, strong interaction with DNA is provided with the RNA polymerase **holoenzyme**, in which a sigma (σ) factor joins the core enzyme. The function of the **sigma factor** is to find the promoter DNA sequence, where holoenzyme can bind and start to transcribe a specific RNA. Other regions of the sigma factor prevents binding of a sole sigma factor to the DNA. A prokaryotic cell has many different sigma factors, which recognize promoters of certain gene "groups". Not every sigma factor is expressed permanently, their expression level depends on environmental and other factors. The **primary sigma** factor binds to the promoters of the house keeping genes (the products of these genes are essential for the everyday life and the growing of the cell), while th**e alternative sigma** factors are expressed at specific cases.

Initiation

The promoter sequences recognized by the sigma factors belong to the so called consensus sequences. **Consensus sequence** is a kind of statistical probability, not a genuine existing sequence, but an ideal one. The more similar an actual sequence is to the consensus one, the stronger is the promoter. Strong promoter is best defined as the frequently starting transcription (high level of expression of the gene). The components of the bacterial promoter recognized by the primary sigma factor are: -35, -10 and occasionally the UP elements.

Phases of the initiation

Closed complex: RNA polymerase holoenzyme associates with the DNA at the promoter region. The double stranded DA has to open up, unwind around the promoter, so in the **open** complex RNA polymerase can achieve close contact with the template DNA strand. In the **initiation** complex a short RNA strand is synthetized, which creates an RNA-DNA hybrid. When it is long enough, RNA polymerase moves forward on the DNA template, leaving behind the promoter region and releasing the sigma factor (**promoter clearance**).

Elongation

RNA polymerase proceeds with the synthesis of the new RNA chain in a **processive** manner, with a relatively high speed (20-50 nucleotide/sec). Since the transcriptional bubble is also moving, positive and negative supercoiling develops within the DNA, which are revealed by **topoisomerase** enzymes. The transcriptional complex is held together by the 10-14 base pair long RNA-DNA hybrid, and by the conformation of the RNA polymerase. Occasionally because of specific structures in the DNA or the RNA transcript, **pausing or transcriptional arrest** may occur, however, the latter one needs additional factors for transcription continuation. Transcriptional pausing and arrest are part of the gene expression regulation.

Termination

Prokaryotic transcriptional terminator sequences can be of two types: rho-factor dependent or **intrinsic** (rho-factor independent). In case of the intrinsic terminator, the DNA template contains an inverted repeat sequence followed by a run of adenosines (A). This sequence is coding a **stem-loop** structure in the nascent RNA, with a **run of uridines (U)**. The result of this structure will be, in which the stem-loop structure is shortened the RNA-DNA hybrid, while the base pairing between the A and U sequences will be considerably weak. The two components together weaken the interaction between the DNA template and the RNA transcript to a degree, and as a result, the polynucleotide chains will be separated.

In the **Rho-dependent** type of transcription termination a protein factor, named Rho recognizes a specific site on the nascent RNA transcript. Rho is a hexameric protein with **ATPase** activity. Rho is pulling away the RNA from the template DNA strand with the help of the ATP energy.

9.2. Eukaryotic transcription

Transcription in eukaryotic cells has the same **basic phases**, as in prokaryotes: initiation, elongation and termination. Though eukaryotic cells have **three RNA polymerases**, the core enzyme structure, and the elongation reactions are very similar to the prokaryotic ones. In eukaryotic cells the **chromatin structure** hinders the movement of the RNA polymerase, so nucleosome remodeling enzymes and histone chaperons are essential for the transcription process. Besides, many posttranslational histone modifications are part of the gene expression regulation, just like covalent modifications of the DNA (See: in Eukaryotic gene expression regulation).

RNA polymerases:

- RNA polymerase I transcribes the rRNA genes,
- RNA polymerase II all the mRNAs and small regulatory RNAs,
- RNA polymerase III produces tRNAs, 5S rRNA and snRNA.

They all are large, **multisubunit enzymes**, the most described is the RNA Pol II. This contains 12 subunits, 5 of them creating the core enzyme, similarly to the prokaryotic polymerase. RNA Pol I and III have more subunits than the RNA Pol II, probably the additional polypeptide chains carry specific functions. The largest subunit of RNA Pol II, named **Rpb1, has a special C-terminal domain (CTD)** with repeated units of 7 amino acids. Many of these CTD amino acids can be phosphorylated. They are substrates of different kinases in a timely order, which is part of the transcription process regulation (initiation-elongation phase procession) and the maturation of the mRNA transcript (cap and poly(A) tail synthesis and splicing).

Initiation of transcription

The recognition of the promoters requires additional factors for RNA polymerases. These are called **general transcription factors**: they guide the RNA polymerases to the right promoter sequences. For the three polymerases they are different (labeled TF and the number of the polymerase, and an identification, like: TFIIA), but there is one common subunit among them, the TATA binding protein: TBP. RNA Pol II has the following TFs: TFIIA, B, D, E, F, H, most of them having more subunits (see table at the lecture). RNA Pol II with the general transcription factors assembles at the promoter to generate the **pre-initiation complex (PIC**). TBP and associated factors (TAFs) are components of TFIID, they bind the consensus sequence called **TATA-box**. Additional promoter sequences of polymerase II may be the **BRE element** (recognized by TFIIB), the **initiator element (INR)** and the **downstream promoter element (DPE**), the latter two are bound by the TFIID. TFIID initiates the gathering of the PIC, TFIIB follows, binding the BRE element, and making the contact with the DNA stronger. After TFIIA binds, TFIIF carries the core RNA polymerase to the complex. TFIIE and TFIIH are the last factors to join PIC. Subunits of TFIIH have several notable **essential functions**: the helicase activity promotes the unwinding of the double stranded DNA, other activities are involved in DNA repair, there are cyclin and Cdk subunits (see cell cycle), and kinase activity which is involved in the regulatory phosphorylation of CTD of Rpb1 (the large subunit of RNA Pol II). General transcription factors are usually not enough for regulated transcription of a gene, **activators and repressors** have effect through a large protein complex called mediator, as well as chromosome remodeling and histone modifying enzymes.

Elongation

Transcriptional elongation is the phase which **follows the promoter clearance**. CTD of Rpb1 is getting phosphorylated by TFIIH, RNA polymerase changes its conformation, holds the DNA template firmly, and releases the general transcription factors. Because of certain structural interactions of the DNA template and the RNA transcript, **transcriptional pausing** may occur. Polymerase can move forward by alone, or needs the help of **elongation factors.** Transcriptional arrest cannot be solved without the help of elongation factors. This phenomenon is part of the gene regulatory system.

Phosphorylation of CTD of Rpb1 is a signal for **cap synthetizing enzymes**, while negative elongation factors trigger pausing. The 5’ cap structure attracts other kinases, which phosphorylate further the CTD: elongation proceeds and other processing enzymes gather. In case of pausing and arrest, **cleavage factors** (TFIIS) may help to resume elongation. During elongation nucleosomes have to be disassembled and reassembled with the help of **histone chaperones.** The function of topoisomerases at the elongation phase is to remove generated supercoiling of DNA upstream and downstream of the synthetizing RNA polymerase.

Termination of transcription

The signal for termination of RNA polymerase I is a U-rich sequence, but also additional DNA binding proteins are involved in the process. Transcriptional termination of RNA polymerase II is coupled to the **3' polyadenylation**. As the synthesis of the poly(A) tail starts after the cleavage of the primary transcript, this step is the signal for termination, as well. According to the **allosteric model**, the cleavage will cause a conformational change in the RNA polymerase which will result in the release of the template and dissociation off the complex. The other **model (torpedo)** states that the important component after the cleavage is the RNA which is not released from the polymerase. While the polymerase continues the synthesis, 5'-3' exonuclease will digest the RNA chain until it reaches the RNA polymerase and obliges it to dissociate from the DNA.

9.3. Synthesis and processing of ribosomal and transfer RNAs

Eukaryotic cells possess three **types of RNA polymerases**. RNA Pol I synthetizes the larger sized rRNAs (28S, 18S and 5.8S). The products of RNA Pol II are mRNAs, snRNAs, snoRNAs and many of the miRNAs. tRNAs , 5S rRNA and some snRNAs are made by RNA Pol III. Considering the amount within a cell rRNAs together with the tRNAs provide the majority of ribonucleic acids (80%). All three RNA polymerases have the 5 **core subunits** which are similar to the prokaryotic core enzyme. Polymerase I and III have more subunits than polymerase II, the additional ones have specific functions. In this section we will consider only the differences of rRNA and tRNA synthesis compared to the production of mRNAs.

Both polymerase I and III will recognize **different promoter sequences** with the help of different transcription factors than polymerase II. But there is a common feature among them: all three enzymes are using **TBP** (TATA-binding protein) as part of a complex. Promoters of polymerase III have unusual features: some of the basic promoter regions are found downstream related to the transcriptional start site.

The signal for transcription **termination** for polymerase I and III is a run of uridines in the RNA, but additional proteins are also required.

Processing of pre-rRNAs

RNA polymerase I synthetizes a long primary transcript which includes the 28S, 18S and 5.8S rRNAs. This way it is guaranteed that the ratio of these rRNAs will be same in the cell. The pre-rRNA is **cleaved** by RNase III, and the rRNAs with the final sizes are released after exonuclease digestions. Some rRNAs in lower organisms may contain introns which are removed by self-splicing (ribozymes). Mature rRNAs contain large number of **modified nucleotides**, the most frequent ones are the methyl-ribose and the pseudouridine. They are located to regions in the molecule where they are involved in important functions of the ribosome. The above mentioned modifications are carried out by specific enzymes which are guided to the right location of the rRNA by the small nucleolar ribonucleoprotein **(snoRNP) complexes**. The RNA components of the complexes can base pair with the target region of the rRNA.

Processing of tRNA transcripts

tRNAs are synthetized as longer precursors. The 3' end is cut by **RNase D**, while the 5' end of tRNA is cleaved by the **RNase P** endonuclease. The latter one contains an RNA component, which can act as a ribozyme in bacteria. The **introns** found in many pre-tRNAs are removed by an ATP requiring splicing step. The 3' end of tRNA will accept a **CCA sequence**, which is added without the use of template. This sequence is universal as this binds the amino acid during the process of translation. Nucleotides

of tRNAs are extensively **modified chemically**. The role of these modifications can be to increase the structural stability of the tRNA or to help the interactions of tRNA with other molecules, including mRNA in the translation.

9.4. RNA PROCESSING

In general

RNA processing involves all the steps when **primary RNA transcripts or precursor RNAs or pre-RNAs** (tRNA, rRNA and mRNA) are modified to get mature, final products. **These steps** can be: the cleavage of the pre-RNA, either at the ends, or removing introns (non-coding sequences), editing of the RNA product by insertion or deletion or modification of nucleotides, and in case of mRNA 5' cap and 3' polyadenylate synthesis. These processing steps have **multiple functions**: their speed and efficiency contribute to the *gene expression regulation* – an RNA product can be used only when it is complete. Also the above mentioned modifications can protect the RNA from *degradation* – half-life of an RNA product is an important regulatory possibility. Processing provides a *tool for diversity*: alternative splicing or RNA editing for example may give rise to more protein products from a single mRNA. The process of maturation is the subject of a well-developed *quality control* system: only properly processed RNA can reach its final destination in the cell. Other RNAs are degraded, as well as exogenous RNAs (RNA interference, see in Gene expression regulation).

Another important feature of the majorities of the maturation reactions is that many times **ribonucleoproteins (RNPs),** complex of RNA and protein molecules are involved in the reactions. The role of the RNA in these complexes can be *catalytic* (ribozymes) or *guiding* the actual catalytic component to the proper site with base-pairing. The large complexes which carry out the processing steps are often in close connections with each other, as one step follows the other meanwhile the reactions are controlled as well. This **cooperation** is helped by the localization of these complexes they can be close to each other in the nucleus or nucleolus, for example.

Here we discuss the relatively complicated maturation process of mRNA. The processing of the primary tRNA and rRNA is found in the Transcription of tRNA and rRNA chapter.

Processing of mRNA

5' capping

The 5' end of the pre-mRNA is modified in the following way: first a 5' phosphate is removed by a *phosphatase*, then a *guanyl transferase* joines a GMP molecule there with an unusual 5'-5' linkage. Finally the added guanine is *methylated*, as occasionally the 2' O in the second or even in the third nucleotide of the newly synthetized, only 20-30 nucleotide long transcript. This cap structure has **functions** in the elongation and termination of transcription, in the further processing of the RNA, in the nucleo-cytoplasmic transport of the RNA and in the translation initiation process. The cap also provides protection from exonucleases.

3' polyadenylation

Every mRNA has at least one polyadenylation site, but some have more which gives an opportunity to synthetize more proteins from one mRNA. The steps of polyadenylation are the following: **cleavage** at a specific sequence of the mRNA, then **poly(A) polymerase** synthetizes a long poly(A) tail (200 nucleotides) which will be covered by poly(A)-binding proteins. The **functions** of this poly(A) tail are protection against exonucleases as well as helping initiation of translation.

5’ capping, 3’ polyadenylation and intron splicing are **coordinated in time and space** with the help of the phosphorylated CTD of the large subunit (Rpb1) of RNA polymerase II. Processed, mature mRNA has to leave the nucleus, as translation occurs in the cytoplasm. After polyadenylation signals initiate the release of the nuclear restricted proteins from the mRNA. Other proteins remain attached to the mRNA, and follow it to the cytoplasm. There could be **localization elements** found on the 3’ end of the mRNA which help the RNA to move to a specific region of the cytoplasm where translation will happen (for example translation of special neuronal proteins close to the synapses). These localization elements are bound by proteins, which help translation as well.

RNA splicing

Introns are found in every type of pre-RNA (transfer, ribosomal, messenger). Though they are removed by different mechanisms, the basic reactions are the same, **series of transesterifications.** The majority of introns are non-coding, and they are degraded after splicing. Exceptions are some mRNA introns, which code snoRNAs or miRNAs. Splicing of introns is much more frequent in higher organisms. The appearance of introns and the consecutive possibilities of exon shuffling and alternative splicing played important role in evolution and in the increase of the number of (protein) gene products in eukaryotes.

The basic of intron splicing is two transesterification reactions. This way there is no need for direct use of energy (ATP), and the reaction is reversible (insertion of introns into dsDNA). In eukaryotes intron splicing occurs with a complex of RNP particles, but in lower organisms two types of **self-splicing** take place. They are called group I and group II introns. **In group I a free guanosine** helps the first cleavage at the 3’-exon 5’-intron junction, then the 3’-OH on the exon end initiates the second transesterification reaction, the result of which will be the joining of the two exons and the release of the intron. **Group II introns** differ from the previous one in the first cleavage reaction where they do not need an additional factor but rather they use the 2’-OH of a special A (**branch-point adenosine**) within the intron sequence. This will result in a lariat-formed intron released. This structure is very similar to the one generated by the spliceosomes. Group II introns may behave as mobile genetic elements, and insert themselves into dsDNA.

mRNAs of higher eukaryotes contain introns with larger sizes than their exons. During processing of pre-mRNA these introns are removed and exons are connected with each other by large ribonucleoprotein (RNP) complexes, the **spliceosomes**. The reaction is somewhat similar to the group II intron splicing, as these introns contain a specific adenosine and the **lariat-structure intron** will be produced.

There at least **three basic sequences** which have to be recognized for efficient splicing: the 5’, the 3’ splice sites and the branch-point nucleotide in the intron. (The role of other sequence elements is discussed in the Regulation at the transcriptional level section.)

The spliceosome has been built up of **five small nuclear RNPs** (snRNPs – U1, U2, U4, U5, U6), each of them carrying a short RNA (snRNA) and a number of proteins. The RNA components of the snRNPs help the spliceosome to locate splice sites with base pairing to the mRNA or to each other. Other sequences are recognized by protein factors, and some proteins act as chaperones to generate the proper conformation of the whole complex. Additional proteins are needed to provide energy by ATP hydrolysis for **structural development** of the complex, or for correct localization of the spliceosome.

It is important to note, that all processing events are in close connection with each other in time and space as well as with the process of transcription itself. 5’ capping is essential for elongation of transcription, while 3’ polyadenylation is the signal for transcriptional termination, and splicing occurs during the RNA synthesis. The organizer of these events is the **CTD of Rpb1** of RNA polymerase II.

When mRNA is maturated completely after a **quality check** it is transported into the cytoplasm where translation will take place. **Some of the proteins** which are bound to the mRNA will accompany it to the cytoplasm, like poly(A)-binding proteins, exon junction complex and cap binding proteins. They have roles in the translation process.

9.5. REGULATION OF TRANSCRIPTION

Transcription regulation is part of the gene expression regulation. By regulating transcription, the types and amounts of actually transcribed RNA is determined in a cell, depending on the developmental status, environmental factors, and signal transduction pathways activities. Every phase of transcription can be regulated, including initiation (the best controlled step), elongation, termination of transcription, maturation of RNAs, transport of RNAs into the cytosol and degradation of RNAs. Transcriptional control mechanisms differ in prokaryotes and eukaryotes.

Promoter strength can be a factor which determines how frequently a gene is transcribed. The more resembles a promoter sequence the consensus sequence, the stronger the promoter is. Other possibility for regulation is the relative distance of **regulatory and promoter sequences**. Regulatory sequences bind activators or repressors (increasing or decreasing the frequency of transcription initiation from a promoter, respectively). If an activator can bind a regulatory sequence in a proper position to a promoter, it helps initiation of transcription. If the regulatory sequence is in overlapping position with the promoter, the bound protein (repressor) will inhibit the initiation of transcription. These types of regulations are found mainly in prokaryotes.

Transcriptional **pausing and arrest** steps cause delay in the elongation. Elongation factors may be needed to overcome these types of obstructions. Eukaryotic RNAs are processed and are useful for the cell only in their mature state. The speed of the **processing** is a type of regulation. **RNA structure** may promote or inhibit the movement of the RNA polymerase. Additional (protein) factors may be needed to initiate structural modifications of the RNA.

In eukaryotes **enhancer** sequences bind regulatory proteins. These sequences can be located upstream or downstream from the promoter, even at large distances. In this case DNA-looping proteins help the **regulatory protein** to reach the transcription initiation complex. Regulatory proteins do not act by themselves: they attract additional (**co-activator and co-repressor**) factors. These regulatory proteins can be **modified** allosterically (by binding small molecule to the allosteric site, different from the active site of the molecule) and covalently (could be the end result of a signaling cascade). In both cases after the modifications the conformation (3D structure) of the regulatory protein will change, which has influence on its activity (DNA binding or interaction with other proteins).

Regulator and co-regulator proteins are the members of **large complexes** which gather around the promoter region. These complexes contain a multisubunit **Mediator** complex which is essential for high rate transcription initiation by RNA polymerase II. This complex creates interactions among activators, co-activators, transcription factors, RNA polymerase, and even chromosome remodeling and modifying complexes (gene expression regulation by chromosome modifications: in Gene regulation of eukaryotes). **Major ways** how an activator may have effect on transcription: helps to bind additional, necessary activators; assists RNP polymerase promoter binding; stimulates promoter clearance and elongation from pause. A repressor may compete for a binding site with an activator; may block the binding region of an activator; may bind directly to the transcription factors; or may have effect on chromatin structure indirectly.

Alternative splicing

As we saw there are relatively small number of genes in the human genome, but the number of active proteins is much higher. There are more possibilities for generating diversity during transcription and translation. Alternative splicing and RNA editing are among these opportunities.

Alternative splicing provides more possibilities for distribution of exons of a given pre-mRNA which process will result different mRNAs from a single primary transcript. There are constitutive (always present in mature mRNA) and regulated (present in mRNA just at special circumstances) exons. To generate this variability, **alternative 5' or 3' splice sites** have to be recognized by the spliceosomes. With alternative transcriptional start sites or alternative polyadenylation sites, the diversity of the final gene products can be increased even further. This large number of variations can be seen in mammalian central nervous systems.

Two theories for splicing control exist: the exon and the intron definition models. Known fact is, that splicing sites are recognized co-transcriptionally. There are **protein factors** which mark the splicing sites (SR proteins in exons) or cover the cryptic splice sites (hnRNP complexes). To determine splice sites, the CTD of RNA polymerase II large subunit and snRNPs of the splicing complex have to be involved as well. In the pre-mRNA specific splicing enhancer or silencer sequences are found, which are able to bind specific proteins. After splicing occurred, the exon-exon junction sequences are labeled by protein complexes (**EJC: exon junction complex**). This step is necessary for the mRNA transport from the nucleus into the cytoplasm, as well as for efficient translational initiation.

RNA editing mechanisms include modifications, deletions or insertions of nucleotides of the otherwise mature mRNA. Some of these editing actions can be found frequently, others are rare. Occasionally almost half of a complete mRNA sequence can be edited, other times only one or a few nucleotides are changed. The two most common nucleotide modifications are the results of **deamination reactions**: from adenosine, inosine will be produced, which base pairs with guanosine, and from cytidine, uridine will be the product. The former one is more general. An example for the latter reaction is the editing of the mRNA of apolipoprotein B in the intestine. After deamination of a specific cytidine, a stop codon will be created, by which step a shorter form of the protein is synthetized in the intestine compared to the apolipoprotein B of the liver cells (see this example in details in the lecture). The addition or deletion of uridine(s) happen with the help of **guide RNAs**.

Half-lives of RNAs

One way of gene expression regulation is to destroy the mRNA. Different mRNAs have different half lives in cell, but usually much shorter than rRNAs. The factors which influence the stability of RNAs are the presence of the 5' cap and the 3' poly(A) tail, as well as the structure of the RNA (double stranded regions, hairpins). The ARE (AU-rich elements) located at the 3' UTR (untranslated region) of certain mRNAs are considered to increase the instability of the mRNA (example in the lecture). The general scheme for a eukaryotic mRNA degradation is the following: it starts usually with the breakdown of the poly(A) tail (**deadenylase**), followed by the action of a 3' to 5' **exonuclease** (as part of the exosome complex). Or after deadenylation a **decapping** enzyme will remove the 5' cap structure, providing a substrate for a 5' to 3' exonuclease. Rare nucleotides of RNA degradation will be recycled for synthetic reactions.

Degradation of foreign RNAs (RNA interferencia)

The signal for this protective activity of the cell is usually a **double stranded RNA** piece. This is cleaved by an enzyme called **Dicer** into small ds fragments (short interfering RNAs), which are loaded onto a protein complex (**RISC: RNA induced silencing complex**, which includes an Argonaute protein).

Similar mechanisms are found in the action of gene expression regulating **micro RNAs (miRNAs**), but in this case a single stranded pri-miRNA is synthetized which takes a hair-pin structure. After Dicer cleavage one strand of the short ds RNA fragments is removed (passenger strand), and the other (guide) strand will hybridize to the target RNA sequence blocking its translation.

10. TRANSCRIPTION FACTORS

10.1. Introduction

The regulation of gene expression is controlled by specific proteins called transcription factors. Regulation of gene expression in eukaryotic cells occurs primarily at three distinct levels (see Eukaryotic transcription for more details):

1. Transcriptional-level control
2. Processing-level control
3. Translational-level control

Transcriptional control is regulated by a large number of proteins, called transcription factors. Transcription factors can act either as *transcriptional activators* that stimulate transcription of the adjacent gene or as *transcriptional repressors* that inhibit its transcription.

The control of gene transcription is influenced by the affinity of transcription factors for particular DNA sequences and the ability to bind and interact with other proteins. A transcription factor can act on different genes on the same recognition DNA sequence, and different transcription factors can act on the same gene. Approximately 5 to 10 percent of the human genes encode transcription factors, and the expression of different transcription factors varies between different cells, tissues, stage of development, physiologic state.

10.2. Classification of transcription factors

(see figure from lecture)

General transcription factors: they help to position eukaryotic RNA polymerase correctly at the promoter, aid in pulling apart the two strands of DNA to allow transcription to begin. They are involved in the formation of the preinitiation complex (TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH).

Upstream transcription factors: they bind somewhere upstream of the initiation site to stimulate or repress transcription

Specific transcription factors: they bind to the recognition sequences are present in the proximity of the gene

10.3. Functional classification of transcription factors

1. *Constitutively-active* transcription factors: they are present in all cells at all times; general transcription factors Sp1, NF1, CCAAT

2. *Conditionally-active transcription factors*: they require activation
   a) *Developmental or cell specific transcription factors*: their expression is tightly controlled, but, once expressed, require no additional activation; GATA, HNF, MyoD, Hox
   b) *Signal-dependent transcription factors*: require external signal for their activation
      - I*Extracellular ligand* (endocrine or paracrine) *dependent transcription factors*: nuclear receptors
      - *Intracellular ligand* (autocrine) *dependent transcription factors*: activated by small intracellular molecules; SREBP, p53, orphan nuclear receptors
      - *Cell membrane receptor-dependent*: second messenger signaling cascades resulting in the phosphorylation of the transcription factor

(1) *Resident nuclear factors*: reside in the nucleus regardless of activation state; CREB, AP-1, Mef2
(2) *Latent cytoplasmic factors*: inactive forms residing in the cytoplasm, but, when activated, are translocated into the nucleus; STAT, R-SMAD, NF-κB, Notch

10.4. THE STRUCTURE OF TRANSCRIPTION FACTORS

The 3D structure of numerous DNA binding proteins has been determined by X-ray crystallography and NMR spectroscopy. It has been revealed that transcription factors contain different domains that correspond to different functions of the protein. Transcription factors typically contain three domains (see figure from lecture):

1. DNA-binding domain attached to specific sequences of DNA called response elements.
2. Activation domain regulating transcription by interacting with other proteins. Many transcription factors contain a surface that promotes the binding of the protein with another protein of identical or similar structure to form a dimer.
3. Signal sensing domain sensing external signals and transmitting them to the rest of the transcription complex.

10.5. TRANSCRIPTION FACTOR MOTIFS

Binding of the protein to the DNA is accomplished by a specific combination of van der Waals forces, ionic bonds, and hydrogen bonds between amino acid residues and various parts of the DNA. The transcription factors usually bind to the major groove of the DNA double helix. The most common motifs that occur in eukaryotic DNA-binding proteins are the a) zinc finger, b) helix–loop–helix, and c) leucine zipper.

a) Zinc finger motif:
- the largest class of mammalian transcription factors
- the zinc ion of each finger is coordinated to two cysteines and two histidines (see figure from lecture)
- fingers can act independently of one another
- the first discovered zinc-finger protein, TFIIIA, has nine zinc fingers (see figure from lecture)
- the motif provides the structural framework for a wide variety of amino acid sequences that recognize a diverse set of DNA sequences

b) Helix-loop-helix motif (HLH)(see figure from lecture):
- motif is characterized by two α-helical segments separated by an intervening loop.
- the HLH domain has a special group of highly basic amino acids whose positively charged side chains contact the DNA and determine the sequence specificity of the transcription factor.
- proteins with basic-HLH (or bHLH) motif always occur as dimers (MyoD)
- the two subunits of the dimer are encoded by different genes, so the protein is a heterodimer.
- they play a key role in the differentiation of certain types of tissues
- participate in the control of cell proliferation

c) Leucine zipper motif (see figure from lecture):
- leucines occur every seventh amino acid along a stretch of α- helix.
- leucines along the stretch of polypeptide face the same direction.
- Two α- helices are capable of zipping together to form a coiled coil.

- Form dimers
- it contains a set of basic amino acids on one side of the leucine-containing α- helix.
- the basic segment of leucine zipper called bZIP motif
- stretch of basic amino acids allows the protein to recognize a specific nucleotide sequence in the DNA
- AP-1 is a heterodimer whose two subunits (Fos and Jun) are encoded by the genes FOS and JUN.
- Both of these genes play an important role in cell proliferation
- Mutations in these genes that prevent the proteins from forming heterodimers also prevent the proteins from binding to DNA

10.6. DNA SITES INVOLVING IN REGULATING TRANSCRIPTION

(see Eukaryotic transcription for more details)

TATA box is a major element of the gene's promoter which is a region upstream from a gene that regulates the initiation of transcription. The region that begins from the TATA box to the transcription start site is called the *core promoter*. The core promoter is the site of assembly of a preinitiation complex consisting of RNA polymerase II and several general transcription factors. The two other promoter sequences, called the CAAT box and GC box, are located upstream from the gene. The CAAT and GC boxes bind transcription factors, and regulate frequency of gene expression. The TATA, CAAT, and GC boxes are located within 100–150 base pairs upstream from the transcription start site.

10.7. PROMOTER IDENTIFICATION

(see Molecular biology methods for details)

1. Deletion mapping
2. DNA footprinting
3. Genome-wide localization analysis/Chromatin immunoprecipitation

10.8. THE ROLE OF ENHANCERS

(see Eukaryotic transcription for more details)

The enhancers are DNA elements that can regulate the expression of different genes. An enhancer typically extends about 200 base pairs in length and contains multiple binding sites for sequence specific transcriptional activators. They can be moved from one place to another within a DNA molecule, or even inverted, without affecting the binding ability and the effect of a transcription factor. Deletion of an enhancer can decrease the level of transcription. The enhancers usually scattered in the DNA and they bind different transcription factors.

There are specific DNA sequences on a promoter that can regulate the binding of a transcription factor to the enhancer region. These DNA components are the insulators that can bind to proteins of the nuclear matrix and the DNA segments between insulators correspond to the loop domains (see figure from lecture). The insulators are also called enhancer-blocking elements.

10.9. Coactivators

Coactivators are large complexes that consist of numerous subunits. They can be divided into two groups:

1. those that interact with components of the basal transcription machinery: bring in components of the basal transcription complex needed for transcription to occur
2. those that act on chromatin: by using *histone acetyltransferases* (HATS, CPB) or by using *chromatin remodeling complexes* (SWI/SNF)

Transcription factors, bind to the DNA and recruit coactivators, which facilitate the assembly of the transcription preinitiation complex. The histone acyltranferase (coactivator) binds to the responsive element-transcription factor complex. The enzyme transfers acetyl groups from an acetyl CoA donor to the amino groups of specific lysine residues. Histones of the nucleosome both upstream and downstream from the TATA box are being acetylated. The acetylated histones bind SWI/SNF, which is a chromatin remodeling complex. The two coactivators change the structure of the chromatin to a more open, accessible state. This way the basal transcription factors and TATA-binding protein are able to join to relaxed DNA part. Coactivators and basal transcription factors disrupt additional nucleosomes to allow transcription to be initiated. The remaining nucleosomes of the promoter are acetylated by histone acetylase and the RNA polymerase attaches to the promoter, and transcription is about to begin (see figure from lecture).

10.10. Poised polymerases

RNA polymerase has two critical phosphorylation sites that control its activity, both involving the amino acid serine. In full-on transcription mode, both serine-2 and serine-5 are phosphorylated. When only ser-5 is phosphorylated, the polymerase is “poised,” inactive but ready to go. RNA polymerases are also bound to “transcriptionally silent” genes that initiate transcription but do not transition to elongation, so that a full-length primary transcript is never generated. The poised polymerases require some additional regulatory event to push them into the productive phase in which the gene is fully transcribed. Poised polymerases may play an important role in facilitating the rapid activation of genes.

10.11. Transcriptional repression

There are also enzymes to remove the acetyl groups from histones. Removal of acetyl groups is accomplished by histone deacetylases (HDACs). Histone acetyltranferase enzymes are associated with transcriptional activation; HDACs are associated with transcriptional repression.

HDACs are present as subunits of larger complexes described as *corepressors*. Corepressors are similar to coactivators, except for that if they bind to specific genetic loci by the action of transcriptional factors (repressors) that cause the targeted gene to be silenced.

A repressor binds to its binding site on an active promoter, then the corepressor (HDAC) is recruited to this repressor molecule. The HDAC removes acetyl groups from the histone tails. A separate protein containing histone methyltransferase activity adds methyl groups to the K9 residue of H3 histone tail. The loss of acetyl groups accompanied by the addition of methyl groups lead to chromatin inactivation and gene silencing (see figure from lecture).

The second way of transcriptional repression is DNA methylation (see Epigenetics for more details). This process is carried out by DNA methyltransferases. This chemical modification occurs on carbon 5 of a cysteine. The methylation is an epigenetic mark that allows certain regions of the DNA to be identified and utilized differently from other regions. Methylation is thought to maintain these elements in an inactive state. Abnormal DNA methylation patterns are often associated with diseases. DNA meth-

ylation has not been found in yeast or nematodes. Plant DNA is often heavily methylated, and studies on cultured plant cells indicate that, as in animals, DNA methylation is associated with gene inactivation.

Genes are said to be imprinted according to their parental origin. Imprinting can be considered an epigenetic phenomenon because the differences between alleles are inherited from one's parents but are not based on differences in DNA sequence (see Epigenetics for more details). The mammalian genome contains at least 80 imprinted genes located primarily in several distinct chromosomal clusters. Genes are thought to become imprinted as the result of selective DNA methylation of one of the two alleles. Active and inactive versions of imprinted genes differ in their methylation patterns. Disturbances in imprinting patterns have been implicated in a number of rare human genetic disorders.

11. THE GENETIC CODE

11.1. IN GENERAL

Messenger RNA is an intermediate in the flow of information from DNA to protein. The genetic code is the *nucleotide base sequence* on DNA (and subsequently on mRNA by transcription) which will be translated into a sequence of amino acids of the protein to be synthesized.

George Gamow proposed that each amino acid in a polypeptide was encoded by *three sequential nucleotides*. In other words, the code words, or codons, for amino acids were ***nucleotide triplets***. The triplet nature of the code was soon (1961) verified in a number of insightful genetic experiments conducted by Francis Crick and Sydney Brenner. They worked with T4 bacteriophages and the rIIB gene. They inserted or deleted one to four nucleotides into/from the wild type gene (frameshift mutation-see An introduction to genetics for details). It was revealed that the translated protein was only active if three nucleotides were deleted from the DNA sequence. This proved that the genetic code uses a ***codon of three DNA bases*** that corresponds to an amino acid.

The codons for amino acids are ***non-overlapping triplets*** of nucleotides which means that one triplet corresponds to only one amino acid. It is ***degenerate*,** some of the amino acids are specified by more than one codon. The degeneracy of the code was originally predicted by Francis Crick. He revealed the great range in the base composition in the DNAs of various bacteria. It was clear that the G and C content could range from 20 percent to 74 percent of the genome, whereas the amino acid composition of the proteins from these organisms showed little overall variation. This suggested that the same amino acids were being encoded by different base sequences, which would make the code degenerate.

Identification of the codons was carried out by Marshall Nirenberg and Heinrich Matthaei. They developed a **cell-free system** by rupturing E. coli bacterial cells. This system was capable of building up proteins from amino acids. The first message they tested was a polyribonucleotide consisting exclusively of uridine; the message was called poly(U). When poly(U) was added to a test tube containing a bacterial extract with all 20 amino acids and the materials necessary for protein synthesis, the system manufactured a polypeptide. The assembled polypeptide was analyzed and found to be polyphenylalanine—a polymer of the amino acid phenylalanine. So the codon UUU specifies phenylalanine (see figure from lecture).

In 1964 Nirenberg and Leder developed a ribosome binding assay in which very short artificial RNA sequences (three nucleotides) were used in the cell-free systems. The system allowed the ribosome to bind with the type of tRNA molecule that was complementary to the one codon. They labeled one type of amino acid at a time and then put the mixture through a filter. The filter allowed unbound tRNAs but not the ribosomes with the bound triplet to pass through. There was radioactivity found on the filter the corresponding amino acid was added. It was possible to determine the triplet mRNA sequence that coded for each amino acid (see figure from lecture).

This universal decoder chart lists each of the ***64 possible mRNA codons*** and the corresponding amino acid specified by that codon. To use the chart to translate the codon UGC, for example, find the first letter (U) in the indicated row on the left. Follow that row to the right until you reach the second letter (G) indicated at the top; then find the amino acid that matches the third letter (C) in the row on the right (see table from lecture). Similar amino acids are encoded by similar codons, thus the amino acids with similar properties are clustered in the decoder chart (basic, acidic, hydrophobic, polar). Usually the first two letters are enough for decoding the amino acid. This phenomenon of the genetic code is called ***the tRNA wobble*** (see tRNAs for more details).

First letter \ Second letter	U	C	A	G	Second letter / Third letter
U	phenylalaline (UUU)	serine (UCU)	tyrosine (UAU)	cysteine (UGU)	U
	phenilalaline (UUC)	serine (UCC)	tyrosine (UAC)	cysteine (UGC)	C
	leucine (UUA)	serine (UCA)	**STOP** (UAA)	**STOP** (UGA)	A
	leucine (UUG)	serine (UCG)	**STOP** (UAG)	tryptophan (UGG)	G
C	leucine (CUU)	proline (CCU)	histadine (CAU)	arginine (CGU)	U
	leucine (CUC)	proline (CCC)	histadine (CAC)	arginine (CGC)	C
	leucine (CUA)	proline (CCA)	glutamine (CAA)	arginine (CGA)	A
	leucine (CUG)	proline (CCG)	glutamine (CAG)	arginine (CGG)	G
A	isoleucine (AUU)	theorine (ACU)	asparagine (AAU)	serine (AGU)	U
	isoleucine (AUC)	theorine (ACC)	asparagine (AAC)	serine (AGC)	C
	isoleucine (AUA)	theorine (ACA)	lysine (AAA)	arginine (AGA)	A
	methione (AUG) **START CODON**	theorine (ACG)	lysine (AAG)	arginine (AGG)	G
G	valine (GUU)	alanine (GCU)	aspartate (GAU)	glycine (GGU)	U
	valine (GUC)	alanine (GCC)	aspartate (GAC)	glycine (GGC)	C
	valine (GUA)	alanine (GCA)	glutamate (GAA)	glycine (GGA)	A
	valine (GUG)	alanine (GCG)	glutamate (GAG)	glycine (GGG)	G

The colinearity of genes and proteins was discovered by Charles Yanofsky. He studied the gene tryptophan synthase which was responsible for the synthesis of tryptophan amino acid. This enzyme is essential for the bacteria for growing in Trp free environment. He prepared several substitution mutants (auxotrophs), which required Trp in the medium in order to grow. Using genetic recombinations (see Meiosis for more details), he developed a map of the various mutations (see figure from lecture). Using technology available in the 1950s he sequenced the mutant proteins being synthesized to correlate the mutation in the DNA sequence with the corresponding change in the amino acid sequence. He revealed that the changes were colinear. Each mutation resulted in only one amino acid change, thus each nucleotide was part of only one codon: the genetic code was non-overlapping. Different point mutations in the same position (based on his power of resolution) could result in different amino acids in the product.

Characteristics of the genetic code

1. Specificity: the genetic code is specific, that is a specific codon always code for the same amino acid.
2. Universality: the genetic code is universal, that is, the same codon is used in all living organisms, prokaryotics and eukaryotics (see exceptions in the starred section*)
3. Degeneracy: the genetic code is degenerate i.e. although each codon corresponds to a single amino acid, one amino acid may have more than one codons, e.g. arginine has 6 different codons.
4. It is a triplet code: each three-nucleotide codon in the mRNA specifies one amino acid
5. It is comma-free: mRNA is read three bases at a time without skipping any bases.
6. It is non-overlapping/non-ambiguous: each nucleotide is part of only one codon and is read only once during translation.
7. The code has start and stop signals: AUG is the usual start signal and defines the open reading frame.
8. Stop signals are codons with no corresponding tRNA: the nonsense or chain-terminating codons; generally three stop codons: UAG, UAA, and UGA

11.2. Structure and Function of the Ribosome

Introduction

Ribosomes in the cytoplasm or on the surface of the rough endoplasmic reticulum are responsible for the translation of mRNA into amino acid sequence, the synthesis of proteins.

Ribosomes are made of RNA and protein. Ribosomes contain several different types of RNAs and a collection of different proteins. All ribosomes are composed of two subunits of different size.

Prokaryotes

The large (or 50S) ribosomal subunit of bacteria contains two molecules of RNA and approximately 34 different proteins. The small (or 30S) ribosomal subunit of bacteria contains one molecule of RNA and

* Universality
The first exceptions to the universality of the genetic code were found to occur in the codons of mitochondrial mRNAs. For example, in human mitochondria, UGA is read as tryptophan rather than stop, AUA is read as methionine rather than isoleucine, and AGA and AGG are read as stop rather than arginine (see figure from lecture). The genetic code of the plants is universal. More recently, exceptions have been found, in the nuclear DNA codons of protists and fungi. For example the start codon can be other than AUG, but the first amino acid is always methionine (see figure from lecture).

21 different proteins (see figure from lecture). The sedimentation coefficient is measured in Svedberg units (S) which is the rate of sedimentation of a component in a centrifuge is related both to the molecular weight and the 3D shape of the component. The sedimentation coefficient of prokaryotic ribosome is 70S.

Type of RNA	Approximate number of nucleotides	Subunit location
16S	1,542	30S
5S	120	50S
23S	2,904	50S

Functional 30S bacterial subunits can be made by mixing the 21 purified proteins of the small subunit with purified small-subunit ribosomal RNA. The components of the small subunit contain all the information necessary for the assembly of the entire subunit. At least one of the proteins of the small subunit appears to function solely in ribosome assembly; deletion of this protein from the reconstitution mixture greatly slowed the assembly process but did not block the formation of fully functional ribosomes. Reconstitution of the large subunit of the bacterial ribosome is also possible from its components in an *in vitro* environment. E. coli needs seven copies of its rRNA genes to meet the cell's need for ribosomes.

Importance of the rRNA

The genes encoding rRNA are the most conserved genes in all cells. Due to the differences between prokaryote and eukaryote ribosomes it was possible to create antibiotics to destroy a bacterial infection without damaging the cells of an infected person (see Antibiotics for more details).

Eukaryotes

A eukaryotic cell may contain millions of ribosomes composed of two subunits. The large subunit is the 60S consists of 49 ribosomal proteins, 28S rRNA, 5,8S rRNA molecules. The small subunit is the 40S consists of 33 ribosomal proteins and 18S rRNA molecules (see figure from lecture). Ribosomes are so numerous that more than 80 percent of the RNA in most cells consists of ribosomal RNA. The DNA sequences encoding rRNA are normally repeated hundreds of times. Human cells contain about 200 rRNA gene copies per haploid genome, spread out in small clusters on five different chromosomes. Regulation of ribosome assembly hinges on the regulation of rRNA synthesis and maturation (see Transcription for more details). The rRNA encoding DNA, called rDNA, is typically clustered in one or a few regions of the genome. The human genome has five rDNA clusters; these clusters of rDNA are gathered together in nuclear structures, called nucleoli. The bulk of a nucleolus is composed of nascent ribosomal subunits that give the nucleolus a granular appearance.

Type of rRNA	Approximate number of nucleotides	Subunit location
18S	1,900	40S
5S	120	60S
5.8S	156	60S
28S	4,700	60S

A 5S rRNA is 120 nucleotides long, is present as part of the 60S subunit. In eukaryotes, the 5S rRNA molecules are encoded by a large number of identical genes that are separate from the other rRNA genes and are located outside the nucleolus. The 5S rRNA is transported to the nucleolus to join the other components involved in the assembly of ribosomal subunits. The 5S rRNA genes are transcribed by RNA polymerase III. RNA polymerase III is unusual among the three polymerases in that it can bind to a promoter site located within the transcribed portion (internal promoter) of the target gene (see Transcription for more details).

The eukaryotic ribosome cannot be made up of its components in vitro. It means that the information that is needed to the reconstitution for the ribosome is no longer available in the mature components. Only the precursor rRNA molecules possess the information.

Ribosome biogenesis

The ribosome consists of rRNA molecules and proteins. The protein components of the ribosomes are synthesized in the cytoplasm. These proteins are transported into the nucleus through the nuclear pores then they move into the nucleolus and associate to the matured rRNA molecules to form the ribosomal subunits. The rRNAs are transcribed at the nucleolus except for the 5S rRNA, which is transcribed outside the nucleolus. After its synthesis the 5S rRNA molecules are transported into the nucleolus. The nascent RNA transcripts contain associated particles. These particles consist of RNA and protein that work together to convert the rRNA precursors to their final rRNA products and assemble them into ribosomal subunits. The rRNA is transcribed as a 45S pre-rRNA containing internal and external transcribed spacers (ITS/ETS) in the nucleolus (see figure from lecture). In the next step RNA modifying enzymes attach to their recognition sites by interaction with guide RNAs (form RNA-RNA duplex). These guide RNAs are called small nucleolar RNAs (snoRNA) which form complex with proteins and exist as small-nucleolar-ribonucleoproteins (snoRNP). The bound snoRNA guides an enzyme (methylase or a pseudouridylase) and forms the snoRNP to modify a particular nucleotide in the pre-rRNA. There are approximately 200 different snoRNAs, one for each site in the pre-rRNA that is ribose-methylated or pseudouridylated (see Transcription for more details). The ribosomal subunits assemble in the nucleus, and then exported into the cytoplasm. Ribosome assembly happens in the cytoplasm.

Types of the ribosomes

(see Cellular organelles and Protein trafficking for details)

1. Free ribosome:
   - can move only in the cytoplasm
   - proteins from free ribosomes are released into the cytosol and used within the cell
   - proteins containing disulfide bonds cannot be formed in free ribosomes (the cytosol is a reducing environment)
2. Membrane-bound ribosome:
   - rough ER
   - polypeptide chains with ER-targeting signal sequence
   - transported to their destinations (organelles, plasma membrane)
   - proteins undergo exocytosis- secretory pathway

RNA-binding sites of the ribosome

A ribosome contains four binding sites for RNA molecules: one is for the mRNA and three called the A-site (aminoacyl), the P-site (peptidyl), and the E-site (exit) for tRNAs (see figure from lecture). A tRNA molecule is bound tightly at the A- and P-sites only if its anticodon forms base pairs with a complementary codon on the mRNA molecule. The A- and P-sites are close enough together for their two tRNA molecules to be forced to form base pairs with adjacent codons on the mRNA molecule. This feature of the ribosome maintains the correct reading frame on the mRNA.

The A site in the ribosome binds to an aminoacyl-tRNA (a tRNA bound to an amino acid). The NH2 group of the aminoacyl-tRNA containing the new amino acid attacks the carboxyl group of peptidyl-tRNA (contained within the P site), which contains the last amino acid of the growing chain called peptidyl transferase reaction. The tRNA that is hanging on the last amino acid is moved to the E site, and what used to be the aminoacyl-tRNA becomes the peptidyl-tRNA (see Translation for more details).

Transfer RNAs

(see Transcription and Processing of RNAs for more details)

The tRNA molecules transport the activated amino acids to the ribosome. The ribosome binds the tRNA the anticodon of which corresponds to the codon of mRNA.

1. Transfer RNA molecules are encoded by repeated DNA sequences on the chromosomes (one type of tRNA is encoded by several genes scattered around the chromosomes).
2. The human genome contains about 1300 tRNA genes.
3. These genes are found in small clusters scattered around the genome.
4. A single cluster contains multiple copies of different tRNA genes, and the DNA sequence encoding a given tRNA is typically found in more than one cluster.
5. The tDNA consists largely of nontranscribed spacer sequences with the tRNA coding sequences.
6. tRNAs are transcribed by RNA polymerase III (internal promoter)
7. The primary transcript of a transfer RNA molecule is larger than the final product

8. (tRNA maturation)
9. Several bases are modified. Unsual bases: pseudouridine, ribothymidine, methylinosine, inosine, dimethylguanosine, methylguanosine, dihydrouridine
10. pre-tRNA processing: endonuclease called ribonuclease P
11. The amino acid is attached to the 3’ end of tRNA
12. The length of tRNAs is between 73 and 93 nucleotides
13. Unusual bases are the results from enzymatic modifications of one of the four standard bases (posttranscriptionally).
14. All tRNAs have sequences of nucleotides that are complementary to sequences located in other parts of the molecule → secondary structure resembles to cloverleaf (see figure from lecture)
15. tRNAs have the triplet sequence CCA at their 3’ end.
16. In eukaryotes CCA is added enzimatically to the 3’ end
17. Tertiary structure of tRNA is L-shaped (see figure from lecture)
18. The anticodon is located in the middle loop of the tRNA molecule (see figure from lecture). The middle loop composed of seven nucleotides; the middle three constitute the anticodon. Amino acid attaches to the opposite end of the tRNA
19. tRNA wobble: tRNA can recognize codons with variable third bases (more than one codon can use the same tRNA; less tRNA molecules are needed for the 61 different codons- see figure from lecture)

Amino acid activation

Amino acids are covalently linked to the 3’ ends of their cognate tRNA(s). This bond is formed by an enzyme called an aminoacyl-tRNA synthetase (aaRS) (see figure from lecture). Humans have 20 different aminoacyl-tRNA synthetases, one for each of the 20 amino acids. Each synthetase is capable of binding the appropriate amino acid to the tRNA. The aminoacyl-tRNA synthetase is able to recognize all of tRNAs, while, at the same time, discriminates against all of the tRNAs for other amino acids. The first step of the binding process is hydrolysis of an ATP that activates the amino acid by formation of an adenylated amino acid, which is bound to the enzyme. The second step is the transfer of the amino acid to the 3’end of tRNA molecule. The PPi produced in the first reaction is subsequently hydrolyzed to Pi, further driving the overall reaction towards the formation of S aminoacyl-tRNA and AMP. The specificity of the amino acid activation is as critical for the translational accuracy as the correct matching of the codon with the anticodon. The reason is that the ribosome only sees the anticodon of the tRNA during translation. Thus, the ribosome will not be able to discriminate between tRNAs with the same anticodon but be linked to different amino acids. If the tRNA holds the improper amino acid, it will be incorporated into the growing amino acid chain as it would be the correct one.

11.3. Phases and Regulation of Translation

Steps of protein synthesis

Initiation

- mRNA attaches to the 30S
- First tRNA comes and binds to mRNA where the nucleotide code matches
- This triggers 50S binding to 30S.

Elongation

- First tRNA is connected to A site.
- It moves to P site as another tRNA approaches

- Second tRNA binds to A site
- Peptide bond is formed between amino acids of tRNAs
- The tRNA moves to P site
- Ribosome is ready to accept another tRNA and continue process
- Each tRNA adds another amino acid to the growing peptide chain

Termination

- Ribosome moves along nucleotide triplets one by one
- Ribosome reaches “stop codon,” peptide chain is finished
- Last tRNA leaves the ribosome, leaving behind completed peptide chain

Polyribosomes

A complex of ribosomes and mRNA is called a polyribosome or polysome (see figure from lecture). Each ribosome initially assembles from its subunits at the initiation codon and then moves from that point towards the 3’ end of the mRNA until it reaches a termination codon. When a ribosome moves towards the second codon on the mRNA, another ribosome binds to the initiator AUG codon and begins the translation. Polyribosomes increase the rate of protein synthesis as one mRNA molecule is translated simultaneously on more than one ribosome. The polysome is organized at the surface of the ER membrane into a circular loop or spiral.

Prokaryotic translation

Protein synthesis is the most complex activity of the cell. Information contained in the mRNA is decoded through the formation of base pairs between complementary sequences in the transfer and messenger RNAs. The protein synthesis requires tRNAs with activated amino acids, ribosomes, mRNA, different factors (proteins), cations and GTP molecules that provide the energy for translation.

Initiation:

- Translation begins at the initiation codon, AUG, which then puts the ribosome in the proper reading frame.
- The small subunit (30S) recognizes the start codon, AUG on the mRNA
- The Shine-Delgarno sequence helps in finding the start codon
- The Shine-Delgarno element is found at the 5' side of each initiator AUG codon in prokaryotic polycistronic mRNAs. This element is complementary to sequences present near the 3'-end of the 16S rRNA of the prokaryotic ribosome (see figure from lecture).
- AUG is the only codon for methionine
- methionine is always the first amino acid at the N-terminus of a nascent polypeptide chain.
- In bacteria, the initial methionine is formylated methionine called *N*-formylmethionine. The formyl group is added to the amino group of methionine. This process is catalyzed by the enzyme methionyl-tRNA formyltransferase. This modification is done after methionine has been loaded onto $tRNA^{fMet}$ by aminoacyl-tRNA synthetase.
- The first amino acid is removed enzymatically from the proteins by the action of methionine aminopeptidase.
- There is a special tRNA for formylmethionine and another one for the methionine residues
- Initiation requires Initiation Factors (IFs)
- Bacterial cells require three initiation factors, IF1, IF2, and IF3
- IFs attach to the 30S subunit
- IF1 promotes attachment of the 30S subunit to the mRNA and may prevent the aa-tRNA from entering the wrong site on the ribosome.

- IF2 is a GTP-binding protein required for attachment of the first aminoacyl-tRNA.
- IF3 prevents the large (50S) subunit from joining prematurely to the 30S subunit.

Initiation process (see figure from lecture):

1. association of the 30S ribosomal subunit with the mRNA at the AUG initiation codon, it requires IF1 and IF3
2. IF2-GTP binds to the small subunit
3. the formylmethionyl-tRNAfMet binds to the mRNA by binding to IF2-GTP
4. the 50S subunit joins the complex, IF2 and IF3 are released, GTP is hydrolyzed, and finally IF2-GDP is released
5. the initiator tRNA enters the P site of the ribosome
6. subsequent tRNAs enter the A site

Elongation (see figure from lecture):

- the initiator tRNA binds within the P site,
- the ribosome is available for the second aminoacyl-tRNA into the vacant A site,
- two elongation factors are needed, EF-Tu and EF-G are GTP-binding proteins
- EF-Tu-GTP delivers aminoacyl-tRNAs to the A site of the ribosome.
- conformational changes occur within the ribosome that causes the tRNA to remain bound to the mRNA in the decoding center
- the GTP is hydrolyzed and the Tu-GDP complex is released
- the A and P sites are in close proximity within the peptidyl transfer site of the subunit (see figure from lecture)
- the formation of a peptide bond between the two amino acids (A and P-site)
- the amine nitrogen of the aa-tRNA in the A site carries out a nucleophilic attack on the carbonyl carbon of the amino acid bound to the tRNA of the P site
- peptide bond formation occurs spontaneously without the need of energy
- the reaction is catalyzed by peptidyl transferase enzyme which is the enzyme component of the large subunit
- tRNA becomes deacetylated at the P-site
- the ribosome moves three nucleotides (one codon) along the mRNA in the 5→3 direction
- EF-G-GTP binds to the large subunit
- hydrolysis of GTP
- translocation of the dipeptide: the movement of the deacylated tRNA and peptidyl-tRNA into the E and P sites
- the ribosome, binds to the second codon of the mRNA by hydrogen bonds
- EFG-GDP leaves the ribosome
- the deacylated tRNA leaves the ribosome, and a new aminoacyl tRNA enters the A site int he next elongation cycle
- the nascent polypeptide exits through the channel within the large subunit (see figure from lecture)

Termination (see figure from lecture):

- the ribosome reaches the stop codon (UAA, UAG, or UGA)
- there is no corresponding tRNA for stop codon
- release factors are needed
- release factors can be divided into two groups
- class I RFs recognize stop codons in the A site of the ribosome
- class II RFs are GTP-binding proteins
- bacteria have two class I RFs: RF1 recognizes UAA and UAG stop codons, and RF2 recognizes UAA and UGA stop codons

- RF1 or 2 enters the A site of the ribosome, and interacts directly with the stop codon
- the ester bond linking the nascent polypeptide chain to the tRNA is then hydrolyzed
- the completed polypeptide is released
- hydrolysis of the GTP bound to the class II RF called RF3 leads to the release of the class I RF from the A site.
- release of the deacylated tRNA from the P site
- dissociation of the mRNA from the ribosome
- disassembly of the ribosome into its large and small subunits

Eukaryotic translation

Translation of the eukaryotes is more complex than in prokaryotes. It requires more regulating factors that help in the overall process to be more precise.

Initiation:

- the first codon on the messenger RNA is the code for methionine (AUG)
- eukaryotes: initiator AUG is generally the first encountered by the ribosome
- usually a specific sequence context surrounding the initiator AUG aids ribosomal discrimination
- this context is A/GCCA/GCCAUGA/G in most mRNAs
- the first amino acid is transported by initiator tRNA that differs from tRNA that carries the methionine residues to the mRNA
- initiator tRNA binds to the small subunit
- eukaryotic cells require at least 12 initiation factors several of these eIFs (e.g., eIF1, eIF1A, eIF5, and eIF3) bind to the 40S subunit, which prepares the subunit for binding to the mRNA (see factors in the tables below).

Initiation factor	Activity
eIF-1	repositioning of met-tRNA to facilitate mRNA binding
eIF-2	heterotrimeric G-protein composed of α, β, and γ subunits; formation of ternary complex consisting of eIF2-GTP + initiator methionine tRNA (met-tRNAimet); AUG-dependent met-tRNAimet binding to 40*S* ribosome
eIF-2B (also called GEF) guanine nucleotide exchange factor, composed of 5 subunits: α, β, γ, δ, ε	GTP/GDP exchange during eIF-2 recycling; genes encoding the subunits identified as EIF2B1–EIF2B5, mutations in any one of which causes the severe autosomal recessive neurodegenerative disorder called leukoencephalopathy with vanishing white matter
eIF-3, composed of 13 subunits	ribosome subunit dissociation by binding to 40*S* subunit; eIF-3e and eIF-3i subunits transform normal cells when overexpressed, eIF-3A overexpression has been shown to be associated with several human cancers
initiation factor complex often referred to as eIF-4F composed of 3 primary subunits: eIF-4E, eIF-4A, eIF-4G and at least 2 additional factors: PABP, Mnk1	mRNA binding to 40*S* subunit, ATPase-dependent RNA helicase activity, interaction between polyA tail and cap structure

Initiation factor	Activity
PABP: polyA-binding protein	binds to the polyA tail of mRNAs and provides a link to eIF-4G
Mnk1 and Mnk2 eIF-4E kinases	phosphorylate eIF-4E increasing association with cap structure
eIF-4A	ATPase-dependent RNA helicase
eIF-4E	5' cap recognition; frequently found overexpressed in human cancers, inhibition of eIF4E is currently a target for anti-cancer therapies
4E-BP (also called PHAS) 3 known forms	when de-phosphorylated 4E-BP binds eIF-4E and represses its' activity, phosphorylation of 4E-BP occurs in response to many growth stimuli leading to release of eIF-4E and increased translational initiation
eIF-4G	acts as a scaffold for the assembly of eIF-4E and -4A in the eIF-4F complex, interaction with PABP allows 5'-end and 3'-ends of mRNAs to interact
eIF-4B	stimulates helicase, binds simultaneously with eIF-4F
eIF-5	release of eIF-2 and eIF-3, ribosome-dependent GTPase
eIF-6	ribosome subunit dissociation

- the initiator tRNA enters the P site of the subunit in association with eIF2-GTP
- the small ribosomal subunit with its associated initiation factors (eIF1, 2 and 3) and charged tRNA which together form a 43S preinitiation complex
- preinitiation complex finds the 5’ end of the mRNA
- mRNA with initiation factors also form a complex: eIF4E binds to the 5’ cap of the eukaryotic mRNA; eIF4A (helicase) moves along the 5’ end of the mRNA removing any double-stranded regions that would interfere with the movement of the 43S complex along the mRNA; eIF4G serves as a linker between the 5’ capped end and the 3’ polyadenylated (poly A tail) end of the mRNA (see figure from lecture).
- 43S complex recognizes the AUG initiation codon
- eIF2-GTP is hydrolyzed
- eIF2-GDP with the other initiation factors are released from small subunit
- large subunit (60S) joins to the small subunit (eIF5)
- the eIF3 complex is responsible for the formation of the ternary complex (eIF-2-GTP-met-tRNAimet) and the formation of the 43S preinitiation complex (PIC)
- the energy for the formation of the 80S initiation complex comes from the hydrolysis of the GTP bound to eIF-2→ the eIF2-cycle

The eIF2-cycle

(see figure from lecture)

- the cycle involves the regeneration of GTP-bound eIF-2 following the hydrolysis of GTP during translational initiation
- the GTP bound to eIF-2 is hydrolyzed providing energy for association of the whole ribosome

- in order for additional rounds of translational initiation to occur, the GDP bound to eIF-2 must be exchanged for GTP
- this is the function of eIF-2B which is also called guanine nucleotide exchange factor (GEF)
- the cycle is regulated by phosphorylation (eIF2α kinases)
- kinase phosphorylates the serine residues in the α-subunit of eIF-2
- when eIF-2 is phosphorylated the GDP-bound complex is stabilized and exchange for GTP is inhibited (initiation is inhibited)
- eIF-2B binds more tightly to the phosphorylated eIF2A, slowing the rate of nucleotide exchange
- there are four related kinases known to exist in mammalian cells
- each eIF-2α kinase contains unique regulatory domains that interact with various inducing agents in response to different stress-related conditions (endoplasmic reticulum stress, nutrient stress, viral infection, and in erythrocytes as a consequence of limiting heme)
- the kinases regulate the global translation of the eukaryotic cells

Elongation:

- the initiator tRNA binds within the P site
- the ribosome is available for the second aminoacyl-tRNA into the vacant A site
- two elongation factors are needed, eEF1 and eEF2 which are GTP-binding proteins
- elongation occurs in a cyclic manner
- each incoming aminoacyl-tRNA is brought to the ribosome by an eEF1α-GTP complex
- tRNA is deposited into the A site the GTP is hydrolyzed and the eEF1α-GDP complex dissociates
- GDP must be exchanged for GTP, carried out by eEF1βγ
- the peptide attached to the tRNA in the P site is transferred to the amino group at the aminoacyl-tRNA in the A site
- peptide bond formation is catalyzed by peptidyltransferase, termed transpeptidation
- the A site needs to be freed in order to accept the next aminoacyl-tRNA
- translocation is catalyzed by eEF2 coupled to GTP hydrolysis (conformational changes in the factor)
- the ribosome moves along the mRNA toward the next codon of the mRNA
- following translocation eEF2 is released from the ribosome
- translocation is regulated by the phosphorylation of the eEF2, when phosphorylated the enzyme is inhibited
- phosphorylation of eEF2 is catalyzed by the enzyme eEF2 kinase (eEF2K)
- eEF2Ks are also regulated by phosphorylation, but in this case the phosphorylation activates the enzyme (kinase → active eEF2K → eEF2 inactive)
- eEF2K kinases: mTOR (mammalian target of rapamycin), AMPK (AMP-activated protein kinase)
- for each cycle of elongation, at least two molecules of GTP are hydrolyzed: one during aminoacyl-tRNA selection and one during translocation
- final step: the deacylated tRNA leaves the ribosome (free E site)

Termination:

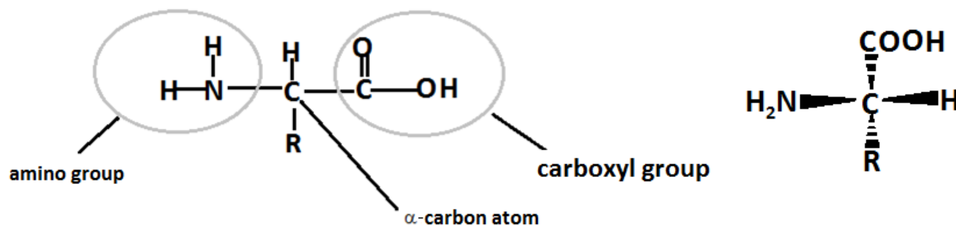
- requires release factors
- eukaryotes have a single class I RF, eRF1, which recognizes all three stop codons (see prokaryotes for more details)
- eRF1 enters the A site of the ribosome, and interacts directly with the stop codon
- the ester bond linking the nascent polypeptide chain to the tRNA is then hydrolyzed
- the completed polypeptide is released

- hydrolysis of the GTP bound to the class II RF called eRF3 leads to the release of the eRF1 from the A site.
- release of the deacylated tRNA from the P site
- dissociation of the mRNA from the ribosome
- disassembly of the ribosome into its large and small subunits

12. TYPES OF PROTEINS

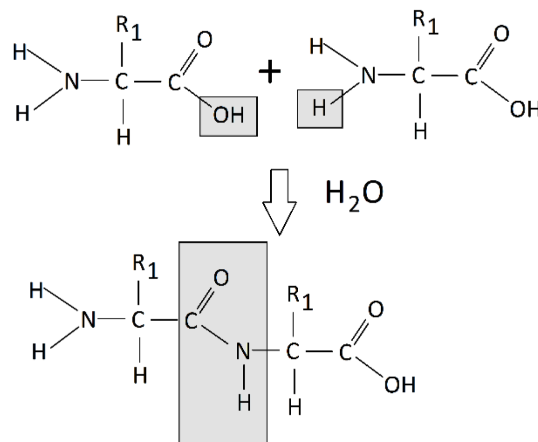
Proteins are the assemblies of α-amino acid molecules and therefore, are considered polymers of these amino acids. In a living organism, there are 25 types of amino acids and 20 of these cannot be created de novo, or by the organism.

A general formula can be written characterizing amino acids: there is an α carbon atom, a carboxyl group, an amino group and an R-marked carbohydrate chain. Amino acids differ from one another in the R side chain.



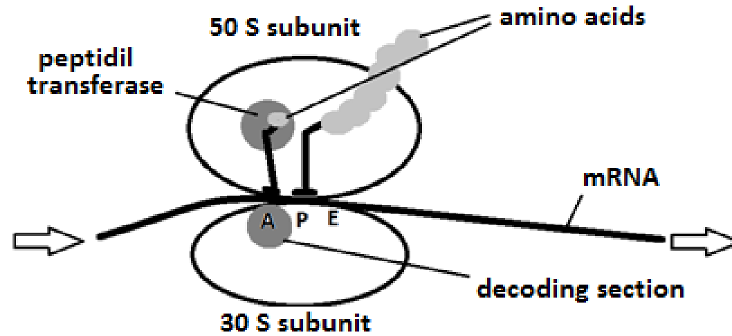
Each of the amino- and carboxyl groups can protonate and deprotonate. The amino acids possess amphoter properties, and because of these groups, they often act as weak bases in acidic solutions and as weak acids in base solutions. We define this structure „zwitterionic" when carboxyl and amino groups are charged.

Animo acids bind together with peptide bond: a water molecule leaves the two amino acids.



In cells, this bond is created by ribosomes. RNA-s can act like enzymes (ribozymes) in ribosomes, peptidyl transferase RNA enzymes make the peptide bond. Some antibiotics can inhibit the functioning of this enzyme in bacteria, e.g. macrolides and chloramphenicol.

The prokaryotic 23S rRNA of the 50 S ribosome subunit and the eukaryotic 28S rRNA of the 60S sub-unit exhibit this ribozyme function (see translation).

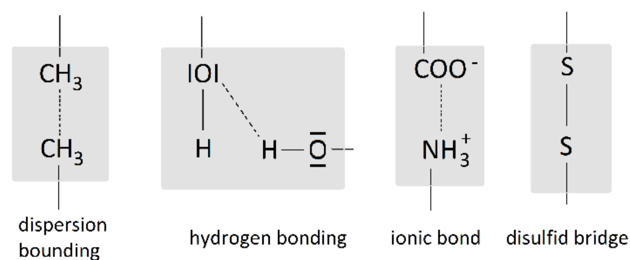


12.1. STRUCTURE

Primary structure: the sequence of the amino acids

Secondary structure: the general 3-dimensional structure of local segments of polypeptides: α-helix, β-sheet, β-turn, random sections

Tertiary structure: tree-dimensional structure of the polypeptide chain. It describes the specific atomic position in 3D space. Various bounds stabilize the chains:



The strongest bond is the disulfide bridge.

Quaternary structure: exist only if and when more polypeptides build up the protein: this describes the relative position of the connected polypeptide chains.

12.2. SHAPE

(Memo: Tertiary structure of the protein refers to the geometric shape.)

Globular: spherical shape, the polypeptide chain(s) curled up, especially enzymes and transport proteins belong to this group (e.g. globin of hemoglobin).

Fibrillar proteins: elongated shape, they form yarns in couples or in groups of three. They have rather static functions and provide structural support, form the cell shape and supply mechanical protection (e.g. α-keratin in nails, hair, skin, feathers, hooves; collagen in tendons or fibroin fibers in silk). Fibrillar proteins are poorly soluble in water.

12.3. Solubility of proteins

The solubility of proteins depends on the R side chains of the amino acids on the surface of the protein: the distribution and the amount of hydrophobic and hydrophilic side chains will determine this property. Proteins with more hydrophobic amino acids on the surface are less soluble in water, charged and polar groups increase the solubility of the protein. Generally highly soluble / dissolved proteins are in high amount in the blood, for example fibrinogen (the molecule that build up the fibrin fiber). Insoluble proteins are mainly fibrillar proteins, such fibrin fibers.

Proteomics: a descriptive scientific discipline which describes the expression, localization, function, quantity and interaction of proteins in an observed cell at a specified age and under specified conditions. Proteomics is the study of the proteome (note omics).

Proteome: the whole protein content of a cell or organism at a specified age under specified environmental conditions.

12.4. Charge of proteins

Classification according to the charge of the R side chain.

Glycine could belong both to the apolar and polar group.

Apolar amino acids with hydrophobic side chains: side chains of apolar amino acids ends with a carbohydrate group

Uncharged amino acids with polar side chain: Side chain of polar amino acids may participate in hydrogen bonds.

Negatively charged amino acids with acidic side chains: Negatively charged amino acids contain carboxyl group in the R side chain

Positively charged amino acids with basic side-chain: Positively charged amino acids contain amino group in the R side chain.

12.5. Peptides

Dipeptides: two amino acids with a single peptide bond

Tripeptides: 3 amino acid, 2 peptide bonds

Tetrapeptides: 4 amino acids

Oligopeptides: made of less than 10 amino acids

Polypeptides: 10 <amino acids

In practice, we usually refer to polypeptides as „proteins”, when the number of amino acids reaches 100.

Even small molecules may have important biological impact. For example, Aspartame (the artificial sweetener) is a dipeptide, glutathion (the antioxidant) and melanostatin (pituitary hormone) are tripeptides.

12.6. Proteins and proteids

Simple proteins are built up from amino acids, in other words, we can hydrolyze them only into amino acids.

Complex proteins (proteids) contain not only amino acids but also other constituents. We can classify them according to the type of these groups (note these proteins have quaternary structure).

Metalloproteins: contain metal ions, for example: cytochrome contains Cu^{2+}, alcohol dehydrogenase contains Zn^{2+}.

Heme proteins: contain iron-porphyrin moiety, for example, hemoglobin, myoglobin and

cytochrome c.

Phosphoproteins: contain phosphate groups, for example, Casein.

Lipoproteins: contain lipid moieties, for example, triglycerides and cholesterol.

Glycoproteins: contain carbohydrate moiety, for example, globulins.

Flavoproteins: contain flavinnucleotid moiety, for example, succinate dehydrogenase.

Nucleoproteins: contain nucleic acid (see viruses), for example, Tobacco Mosaic Virus (TMV), and ribosomes.

12.7. Proteins and diseases

Prion

Prion is a disease-causing protein and is similar to a viral infection. The name originates from the combined words, “protein” and “infection” and is considered the shorter form of “proteinaceous infectious particle” phrase. All known prion diseases affect the neural system in mammals and are are always fatal. A prion is a misfolded protein molecule which can convert properly folded proteins into the misfolded prion form as a bad template. The misfolded prions are extremely stable, and therefore, led to the change in the formerly recommended disinfecting procedure.

Prions create diseases by making so-called amyloid plaques (aggregates) in the central nervous system which destroys the normal tissue structure.

Some human prion diseases are as follows: Creutzfeldt-Jakob Disease, variant Creutzfeldt-Jakob Disease, Kuru and Fatal Familial Insomnia. Some better known animal prion diseases are Bovine Spoingioform Encephalopathy, Chronic Wasting Disease and Scrapie.

Creutzfeld-Jakob disease: rare and lethal disease, central nervous system atrophy, this is the human occurring form of "mad cow disease" (Bovine spongiform encephalopathy). Unfortunately, we discovered the virulence of this disease too late, as there were a variety of operations in which patients contracted prion disease during surgery.

Protein aggregates

Protein aggregates can form not only in the effect of the intake of prions, but also when two or more of the same protein molecule entangles, forming a large protein aggregate. The cause of the formation can be encoded in the gene of the protein or it can be formed because of the environment of the pro-

tein. The biologic activity of these proteins decrease or it can simply cease in its functionality, and this is particularly what may often cause problems. These proteins can form deposits in various cell components or in various tissues, which understandably also causes problems, too.

It is important to detect and analyze these aggregates to diagnose some diseases, e.g. Huntington's disease, Parkinson's disease, in Alzheimer's disease and in various neurodegenerative diseases.

Disease	Protein aggregates
Alzheimer’s disease	β-amyloid (plaque),tau protein (neurofibrillary tangles)
Parkinson’s disease	α-synuclein/ubiquitin
Lewy body dementia	α-synuclein
Huntington’s disease	Poly-glutamine/ubiquitin
Creutzfeldt–Jakob disease	Prion protein
Pick’s disease	Tau (Pick-bodies)

Alzheimer’s disease: there are two processes. Beta-amyloid proteins clump together creating plaques in the brain. The small clumps may block cell-to-cell signaling at synapses and eventually, the neurons die.

Tau proteins stabilize microtubules normally, however, in Alzheimer’s disease, this protein twists and cannot function properly. Because of this, neurons undergo necrosis.

There is a hypothesis in which only the abnormal function of Tau protein can cause Alzheimer’s disease. Once the process begins, the diseased cells will „infect” normal ones.

Parkinson’s disease: there is a mutation in the gene of alpha-synuclein (we can diagnose this disease according to this mutation), this mutation causes an abnormal structure in this protein. These abnormal proteins are prone to aggregate and potentially kill the affected neurons. If we analyze the brain of a patient with Parkinson’s disease, we can see these alpha-synuclein aggregates in certain parts of the brain and what is now considered as the Lewy-bodies.

Lewy body dementia: more common in men, in Lewy bodies we often see the effects throughout the entirety of the brain, similar to Parkinson’s disease.

Symptoms are very similar to Alzheimer’s disease, however, patients with Lewy body dementia are more prone to hallucinations. According to the symptoms, the patients die in 6-12 years from the onset of the disease and diagnosis.

A Huntington’s disease is one type of the polyglutamine diseases, meaning that polyglutamine accumulate and aggregate in the cells. Some other polyglutamin diseases are fragile-X syndrome, cerebellar ataxia and some muscle deteriating diseases.

Pick disease: degeneration of frontal and temporal lobes. As nerve cells destroy, Tau protein containing Pick’s body replace the cells.

Therapeutic proteins

Therapeutic proteins are proteins used for pharmaceutical purposes and design in laboratories. The first marketed therapeutic protein was insulin in 1920.

Today we produce therapeutic proteins in different cell cultures (bacteria, yeast and mammalian cells) with genomic engineering techniques, and we use recombinant techniques and create recombinant proteins. The first recombinant therapeutic protein was the insulin molecule commonly produced today in yeast cultures or in *Escherichia coli* cultures. During the design process of these drugs, protein modification strives to attain a higher response yet avoid the potential risk of side effects. For example, today we use insulin analog molecules (insulin lispro) to treat people living with type 1 and type 2 diabetes. This insulin lispro is produced in *E. coli* and has some important modifications: lysine and proline residues on the penultimate position on the C terminal end of the B chain are reversed, this can inhibit the formation of insulin dimers and hexamers but it has no effect on the receptor binding and because of this, there will be more active molecules in the injection.

Important recombinant therapeutic proteins

Insulin: Hunalog (market name): produced in *E. coli*

Erythropoietin: Epiao (market name), it increases the formation of red blood cells for renal patients and for doping, particular in sports.

Growth Hormone (GH): If growth hormone deficient patients are supplemented with GH, they eventually reach normal stature. Long ago, the required amount of GH was obtained from postmortem cerebral extracts, which is precisely what raised several moral and safety concerns. Many patients who could not afford this therapy– but today, with recombinant techniques, it considered affordable and safe

Interferons: we use them against viral infections and virus-induced tumors.

Vaccines: with recombinant techniques we can design safe vaccines against pathogens, today we have recombinant protein vaccines against cholera (Dukoral: recombinant cholera toxin B with inactivated whole bacteria), Diphtheria, Hepatitis B (since 1981, it is a surface antigen produced in yeast) and Tetanus.

Antitoxins: „antisera”: against snake venoms, experiments with botulin antitoxins.

Antibodies: some important examples include the following: Avastin, which prevents new blood vessel formation (anti-VEGF antibody) and prevents malignant tumor formation. Remicade and Humira are important in the treatment of Crohn's disease, chronic inflammatory disease, rheumatoid arthritis (TNF-blocker) and lastly, Lucentis, which is used against macular degeneration.

13. POSTTRANSLATIONAL MODIFICATIONS AND PROTEIN DEGRADATION

13.1. POSTTRANSLATIONAL MODIFICATIONS

- Not all proteins are functional in the form of a polypeptide chain produced by the ribosome. Many proteins need to undergo post-translational modifications in order to achieve:
- proper folding (to acquire the correct, active conformation)
- regulating activity (to be active when and where it is needed)
- degradation/stabilization (degrade misfolded proteins and maintain balance of concentration)

Proteolytic modifications

Proteolytic enzymes, like proteases and peptidases, cleave the polypeptide chain at a specific place. Proteolytic modifications have a role in:

- Cleavage of N-terminal amino acid: Following translation, most proteins have an N-terminal methionine, which must be first cut in the developmental stages of protein maturation.
- Cleavage of signal sequence: Once a protein is transported to its intended place, the signal sequence, which helps find the correct location, is not required any more, therefore it is cleaved.
- Precursor protein activation: Many proteins leave the ribosome as non-functional pro- or pre-pro form. The proteolytic enzyme converts the protein into its active form. For example, inactive chymotripsinogen produced by pancreatic cells, is cleaved in the gut to develop into its active form, chymotrypsin.

Chaperones

Chaperones, like heat-shock proteins (Hsp) help newly synthetized polypeptide chains to acquire their correct 3D structure (non-covalently) by:

- Help correct timing by preventing early folding
- Help correct folding by protecting polypeptide from environmental effects
- Correcting misfolded proteins: unfold misfolded proteins to raise the possibility of folding again correctly
- Preventing protein aggregation (role in neurodegenerative diseases)

Chaperones recognize hydrophobic surfaces of proteins, and protect them during folding until they acquire stability within the protein. Heat shock proteins earned their name by their capability to prevent protein denaturation on elevated temperature, and to refold unfolded/misfolded proteins.

Disulfide bridges

Disulfide bonds stabilize the conformation (3D structure) with covalent bonds. Disulfide bonds are formed between two sulfur-containing cysteine residues. Commonly, secreted proteins contain disul-phide bridges. Covalent disulphide bonds are generated in the oxidizing environment of the ER by sulfhydryl oxidation (de novo disulfide bond formation). In the reducing environment of the cytoplasm, by means of disulphide bond isomerization, the bridges can swap to the correct place to achieve the stable folded form of the protein. In eukaryotic cells, both processes are carried out by PDIs (Protein Disulphide Isomerases).

Glycosylation

Covalent binding of carbohydrate chains (glycans) to the protein generates **glycoproteins**. Glycans vary widely in complexity from monosaccharides to hundreds of molecules in linear or branched chain. Glycoproteins are almost exclusive for eukaryotes, and the complexity of the glycan chains increases as an organism proceeds up the evolutionary tree.

Carbohydrates are localized on the outer side of proteins in hydrophilic environment. Many secreted and membrane bound proteins have covalently bound glycans.

The roles of glycosylation:

- to increase the proteins' solubility (prevent from aggregation)
- small molecules: intracellular signal transduction
- to protect secreted proteins from degradation
- to provide specificity of protein-protein interactions
- to improve cellular adhesion (recognition by other cells): (eg. Leukocyte binding to the site of inflammation; influenza virus recognizes sialic acid on the surface of the host cell, making the host cell internalize the virus).
- to increase the molecular volume of proteins

There are two main types of glycosylation: N-glycosylation, whereby the carbohydrate is attached to the protein through a nitrogen atom and O-glycosylation, in which the covalent linker is an oxygen atom.

- N-glycosylation: The carbohydrate chain is covalently attached to a N-atom originating from an asparagine residue. The process occurs in the ER, and the precursor oligosaccharide includes glucose, mannose and N-acetyl-glucosamine. The protein contains a glycosylation signal sequence where the oligosaccharide binds. The carrier in these procedures is a dolichol pyrophosphate (DPP). Mutations in any of the 15 genes involved in these processes results in developmental disorders.
- O-glycosylation: The precursor here is N-acetyl-galactoseamine, attached to an O-atom originating from a Ser/Thr residue, in the Golgi apparatus. O-linked carbohydrates are less complex than N-linked appendages. O-glycosylation is characteristic for the extracellular matrix, secreted proteins (mucus) and IgA1 and IgD.

A different type of O-glycosylation exists in cytoplasmic and nuclear proteins, termed O-GlcNAc modification. The reversible O-GlcNAc modification may have a role in controlling phosphorylation, as the site for these two types of modifications is identical. Therefore, O-GlcNAc can block both phosphorylation and acetylation.

The donor molecule for the reversible O-glycosilation is UDP-GlcNAc, whose level is controlled by the nutritional state of the cell. The reversible monoglycosylation responds to nutrition by controlling protein phosphorylation.

Reversible binding of small groups

This modification type includes phosphorylation, methylation and acetylation. By being reversible, these processes are quick and makes the cell capable of rapidly responding to changes and stimuli: these modifications possess a role in enzyme activity regulation, signaling and chromatin structure. Various modifications modulate the same protein's function differently, therefore one protein can have **diverse functions**. As these are reversible processes, there are two enzyme systems possessing the modifications of different directions. Reversibility raises the possibility for a protein to function as a **chemical switch**.

Phosphorylation / dephosphorylation

These modifications are catalyzed by protein kinases and phosphatases. The donor of the phosphate groups is ATP, target aminoacids are Ser, Thr, Tyr. Phosphorylation changes the protein conformation by the negative charge of the phosphate group, which affects the activity. Also, phosphorylation gives rise to new recognition sites, important in protein binding. See also Jak-STAT signaling pathway.

Acetylation / deacetylation

During this process, acetyl transferases attach an acetyl group to an N-terminal Lysine of the protein from an acetyl-CoA donor molecule. Tubulin is acetylated in the axons of neurons, which gives them a longer half-life than for cytoplasmic ones. Acetylation of histone proteins plays a key role in the regulation of the gene expression by changes in chromatin structure.

Methylation / demethylation

During this process, methyl (CH_3) groups are attached to Arg or Lys residues by methyl transferases, from an S-adenosine-methionine (SAM) donor molecule. More than one methyl group can be added to the side chain, unlike for phosphorylation. Methyl groups are chemically stable, the process is typical for histone proteins (epigenetic regulation, but do not mix with DNA methylation).

Activating inactive precursor protein

Many polypeptides leave the ribosome as precursors. These need to be activated in order to function. In many cases this happens through proteolytic modifications (see “precursor protein activation” above).

- Enzyme activation: e.g., pepsinogen is produced by chief cells in the stomach and is activated by an autocatalytic process when it enters the acidic environment in the stomach. In the acidic environment, the pepsinogen unfolds and cleaves itself to produce active pepsin at the end.
- Prohormon-hormon conversion e.g. the inactive preproinsulin carries a signal sequence, which is needed for the transport to the endoplasmic reticulum. Here, in a first step of activation, the signal sequence is cut off and three disulfide bonds are generated. This proinsulin needs another step of activation, the work of three proteases cleaves the prohormone to give rise to the active form of insulin.
  Another example is hepcidin;
  - preproinsulin - insulin
  - prohepcidin - hepcidin
- Retroviral polyprotein cleavage: For many retroviruses, including HIV, proteins are generated with polyprotein synthesis as first step. This is cleaved by proteases encoded by retroviral genes. In pharmaceutics, these can be used as targets of antiretroviral therapy.

13.2. Protein Degradation

The life of a protein ends by irreversible degradation. If a protein is not needed any more, the cell must eliminate it by destruction. Misfolded and damaged proteins are also eliminated by means of degradation. Eliminating part of the enzyme pool is a way to control cellular metabolism, and in many cases, active proteins are generated by the cleavage of the precursor. Another aim of protein degradation is the recycling of amino acids, towards generating new proteins at the end. Notably, the destruction of protein systems within the cell is tightly controlled.

For a clearer view, here are the aims and objectives of protein degradation:

- Elimination of misfolded or damaged proteins
- Regulation of cellular metabolism (decreasing the number of enzymes)
- Generation of active proteins (cleavage of precursors)
- Amino acid recycling

Protein degradation can happen in two pathways:

- Lysosomal proteolysis: non-selective (except of starvation). See lysosomes in chapter 'Cytoplasmic organelles'.
- Proteasomal ubiquitin-dependent pathway: Selective degradation of Ub-tagged proteins.

Proteasomal ubiquitin-dependent pathway

The role of proteasomes

The first step of degradation is the unfolding of the protein, which is an energy-dependent process. The proteolytic machineries – the proteasomes – have two main functions: ATPase function to provide energy for the other function, the peptidase activity. Eukaryotic proteasomes are barrel-shaped structures, including about 30 different polypeptides, oligomerized meaning the proteasome has an internal space with active protease site. Bacteria feature a less elaborate method in the degradation of protein. ATP is needed to the translocation of the protein to the central cavity, where the degradation happens quickly. Thirty to eighty ATP molecules are needed to denature a protein sequence of nearly 100 amino acids. The proteasome generates 5 – 20 amino acid-long peptides. Other steps of cleavage can happen, including the departure of short peptide fragments from the proteasome.

The recognition of proteins for degradation

Protein degradation is highly selective, cells must recognize the proper proteins requiring degradation.

Bacteria use diverse tags to mediate substrate recognition. Tags are unstructured peptide signals, however, it is likely a primary recognition sequence, all located on the N- and C-termini of the proteins. Adaptor proteins can also help the binding of the substrate to the degradation machinery.

In Eukaryotes, ubiquitination helps in targeting the degradation of protein. Ubiquitin is a 76 aminoacid peptide, which can covalently bind to Lysine side chains of proteins. Beyond single ubiquitin molecules, whole ubiquitin chains can bind to a protein. The most important role of ubiquitination is to target the modified protein to degradation. This process is also known as "death tagging" or "death ticketing". Ubiquitination is a reversible process, whereas deubiquitination is catalyzed by deubiquitinating enzymes. Ubiquitination is a three-step process involving three enzymes:

- E1 – Ub activating enzyme (C-term. of Ub covalently attached to Cys in active site of E) ATP consuming process
- E2 – Ub conjugating enzyme (E1-Ub interacts with E2, Ub is transferred to active-site Cys of E2)
- E3 – Ub-protein ligase (transfer of Ub from E2-Ub to amino group of Lys of target protein)

E3 can catalyze the attachment of Ub to another Ub, generating polyubiquitin chains. By acknowledging the number of Lys residues on Ub is 7, one can imagine the diversity of different polyubiquitin products. This is catalyzed by the many different E3 types, as one type tends to catalyze binding of Ub to only one residue.

Rates of protein degradation

Rate of proteolysis differs depending on the type of protein. Abnormal proteins degrade quickly, whereas normal proteins' half-life moves on a wide range from 11 minutes to one month. According to the N-end rule, the N-terminal amino acid can determine the half-life of a protein (e.g., Val → 100 h, Gln → 0.8 h in mammals). The rate of protein degradation in the cell can change due to physiological changes, e.g., during starvation, the proteolysis speeds up to provide free amino acids for the cell.

Unfolded Protein Response (UPR)

UPR is a stress response within the cell. When there is a high amount of unfolded or misfolded proteins in the ER, the cell stops translation and increases the rate of protein degradation, and increases the number of chaperones in order to achieve balance. If this is unsuccessful, apoptosis begins. An error in UPR is responsible for Creutzfeld-Jakob disease, Alzheimer's disease, Parkinson's disease and Huntington's disease.

Diseases related to proteolysis

Abnormal proteasomal activity causes abnormal proteins to accumulate and generate aggregates or inclusions, leading to diseases like Alzheimer's disease, Parkinson's disease, Huntington's disease or ALS.

14. GENE REGULATION

14.1. GENE REGULATION IN PROKARYOTES

According to the central dogma, genes are first transcribed to mRNA, then translated to polypeptide, in which, following posttranslational modifications, serve in the role as enzymes or structure proteins in the cells. The entire process is known as **gene expression**. There are many points within the whole process that can be regulated, meaning repressed or enhanced. Comprehensively, these regulations are best summarized as gene expression regulation or in the shorter sense, gene **regulation**.

In bacteria, genes can be constitutively expressed, meaning they are constantly active, or inducible, inferring that inducers are required for gene expression to begin. Protein production uses up resources and energy, so it is obvious the cell efforts to avoid producing proteins are not necessary at the moment. Nonetheless, it is crucial to maintain the potential to initiate gene expression when environmental or physiological conditions change.

Generally, **transcriptional regulators** locate their target sequences in the genome, and in bacteria, these regulators lie close to the promoter. Regulatory proteins are capable of influencing the transcription of the *downstream* gene. Regulators increasing protein expression are called **activators**, while those decreasing the level of expressed protein, are **repressors**. Several mechanisms have developed in bacteria to regulate transcription initiation. Some of these are shown here through specific examples:

The lactose operon

An **operon** is an integration of genes meaning they are transcribed (and regulated) as a unit. The lac operon is an inducible operon, modulated by the relative availability of glucose and lactose. For E. coli, glucose is the preferred energy source, although lactose can also be used. The lac operon includes the genes having a role in lactose metabolism. These genes (therefore the lac operon) are repressed when glucose is freely available, as there is no need to utilize lactose. However, once the glucose level is suboptimal and there is enough lactose available for the cell, the expression of lactose metabolism genes is induced.

The adjacent lactose metabolism genes are the following:

- *lacZ* encodes β-galactosidase (lactose cutting)
- *lacY* encodes lactose permease (lactose uptake)
- *lacA* encodes galactoside O-acetyltransferase (protection against toxic ß-galactosides)

Upstream to the lactose metabolism genes, the *lacI* gene is located, encoding the **Lac repressor** protein, which is one of the two effectors in the system. This protein binds to the lac operator sequence upstream of the structural genes and represses the transcription only in the absence of allolactase (which is present when lactose is actively metabolized). When allolactase binds to the Lac repressor, it is inactivated, therefore the transcription of the structural genes is activated. Another effector is **CAP** (catabolyte gene-activator protein), which is a receptor of cAMP, is present only when glucose level is low. CAP, binding to the CAP operator sequence upstream to the promoter, activates the transcription of the lactose metabolism genes. Once the genes of an operon are transcribed, the transcript is referred to as ‘polycistronic mRNA’.

The system operation is described below:

In a normal situation, the glucose level is high, while the level of lactose is low. Notably, when there is a high level of glucose, cAMP is not present, therefore CAP is inactive and doesn’t activate expres-

sion. Low lactose level means there is low allolactase as well, thus the Lac repressor protein binds to the lac operator sequence, blocking the expression of the lactose metabolism genes.

As the level of glucose decreases and lactose level increases, cAMP appears in the cell causing the CAP protein to bind to the CAP operator sequence, activating transcription. Lactose presence causes allolactase to appear, which binds to the Lac repressor protein, prohibiting it to bind to a lac operator sequence. Lac operator sequence is freed from the inhibition, therefore the transcription altogether is enhanced.

Jacob and Monod received the Nobel Prize in support of the description of the operon model in 1965.

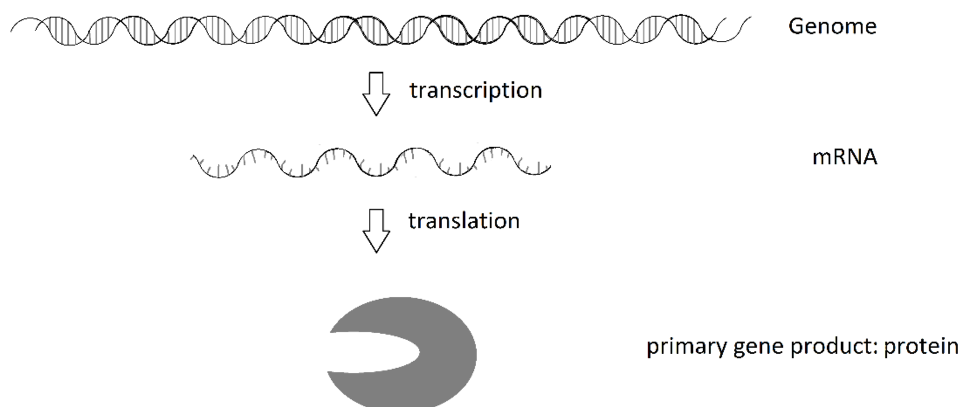
Trp repressor

E. coli can synthetize tryptophane, and the genes encoding the proteins needed for the synthesis are located in an operon. If trp is present in an optimal level, it is not necessary to synthetize it, therefore normally the operon is silenced by repressing its activation. The operator sequence is downstream to the promoter, overlapping with it. Trp is an allosteric effector of the **Trp repressor** protein. When trp is present in the environment, it binds to the Trp repressor protein activating it. This active Trp repressor protein binds to the operator sequence, consequently the binding of the RNA polymerase to the promoter sequence is blocked. Thus, the transcription cannot start.

In the absence of trp, Trp repressor protein remains inactive, doesn't bind to the operator, so the RNA polymerase is free to bind to the promoter and transcription starts. The polycistronic mRNA is then translated to proteins synthetizing tryptophan, which is crucial for the cell to survive.

In Prokaryotes, translation starts immediately following transcription, as there is no spatial separation of these processes. This is the primary difference between prokaryotic and eukaryotic gene translation.

14.2. Gene regulation of eukaryotes



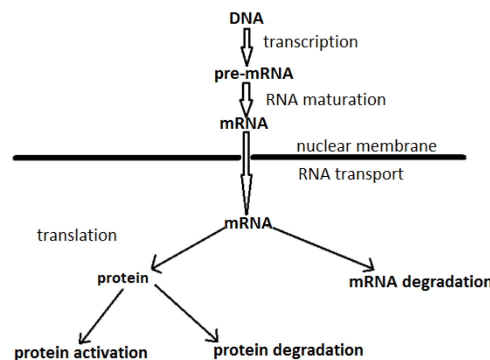
Every cell contains the entire genome and the relevant information of all proteins. However, there are immense differences between a neuron and a red blood cell. Genes are not continuously expressed yet genes are expressed in distinct patterns and under rigorous control. The expression pattern of the genes is specific and characteristic; it depends on the type, the age and the condition of the cell. Gene

regulation is important in commitment and the fate of cells during development and in response to different environmental effects.

Eukaryotic and prokaryotic cells: differences

Generally speaking, we can state genes in prokaryotes are usually expressed and the regulation mechanisms focus on repressing these perpetual and expressing genes. The ground state emphasizes the expression of genes. In eukaryotic cells, most of the genes are usually not expressed, gene expression regulators help to express genes in different levels. These differences are developed because prokaryotes are unicellular organisms and they require all of the encoded cellular mechanisms in this particular single cell in which they exist. In a complex multicellular organism, the functions are shared between the cells throughout the tissues and organs. Therefore, we need gene expression regulation within the human body.

Due to some differences between eukaryotic and prokaryotic cells, there are new potential regulator mechanisms in eukaryotic cells. Gene expression control works on every level of the protein synthesis from the DNA to the active protein.



New possibilities to control gene expression in eukaryotes:

Most of the prokaryotes do not have histones, the control of unfolding of the DNA to make a new opportunity in regulating gene expression. In eukaryotes, transcription and translation occur in separate locations, therefore, transport of mRNA can be regulated. The genes of prokaryotes are organized into operons, but in the eukaryotic cells each gene has an own promoter. Prokaryotes have only one or few chromosomes, in which the genes does not contain introns, and so, they are continuous, by contrast, eukaryotes have more, species-specific number of chromosomes and they contain noncoding (intron) regions.

Nucleosome structure and chromatin remodeling

The structure of the chromosomes influence the functionality of the genes in eukaryotes. The condensed chromatin structure inhibit the gene expression as transcription factors, activators and RNA polymerases cannot access the DNA strand. Local loosening is necessary prior to the transcription.

Chromatin structure, euchromatin, heterochromatin

See more in Chromosome, chromatine, genome chapter

Please note the also lecture slides, too!

While heterochromatin is a tightly packed form, euchromatin structure is a lightly packed form of chromatin. Euchromatin regions are rich in genes and they are often under active transcription. Euchromatin and heterochromatin regions are separated by barriers which prevents continuation of the gene expression. Barrier sequences separate euchromatin and heterochromatin regions.

Histone octamers contain H2A, H2B, H3 and H4 histones (2 from each). This octamer is wrapped into a DNA strand (147 bp) and an H1 histone stabilize this structure. N-terminal tails of histones in the octamer protrude from the structure. While DNA is negatively charged, histone tails are basically positively charged, so they have electrostatic interactions and it can stabilize or destabilize the nucleosome.

Chromatin remodeling complex (CRC)

ATP-dependent chromatin-remodeling complex regulate gene expression by moving, ejecting and restructuring nucleosomes. During hydrolysis of ATP, this protein complex utilizes its energy to reposition (slide, twist or loop) nucleosomes along the DNA, remove histones or exchange different histone variants and establishes nucleosome-free regions on the DNA for transcription.

Histone modifications

Covalent modification of histone N-terminal tails change the charge pattern and the stability of nucleosome structure. These modifications are primarily acetylation, methylation, phosphorylation and ubiquitination.

Histone acetylation

Histone acetyltransferase (HAT) enzyme binds an acetyl group to the histones (to the amino group of lysine). Acetylation reduces the positive charge of the histone tails, and as a result, relaxes chromatin condensation. The potential for gene expression increases, due to the structure of the nucleosome breaking loose and then forming euchromatin in a considerably easier fashion. Distinctively, Deacetylase enzyme can remove the acetyl group.

Some transcription factors possess their own acetyltransferase activity, while others are capable to bind acetyltransferase enzyme (e.g. CBP can bind HATs on the promoters).

Histone methylation

Histone methyl-transferase (HMT) usually binds a methyl group to the lysine amino groups. This is typically the methylation of the 4th lysine of the H3 histone, abbreviation: H3K4me3.

The methylated histones which are created nearer to the transcription sites start most frequently. It prevents the relaxing of the nucleosome and so, prevents the expression of the genes.

Histone code hypothesis

According to the histone code hypothesis, the commonly referred histone-code is a pattern of covalent histone modifications (such as acetylation, methylation, phosphorylation, ubiquitination or polyribosyla-

tion) which has specific meaning. These modification patterns are capable to activate or inhibit the transcription of the DNA. Distinctively, small changes can have immense biological impact. This histone code is inheritable. This modifications result a change in the gene expression but it is not written in the DNA. (Epigenetics)

DNA methylation

In higher eukaryotes (such as humans), methylation is the only covalent modification directly on the DNA strand. In this event, pyrimidine rings in cytosine nucleotides of DNA can methylate. The methyl group binds at a CpG dinucleotide sequence in somatic cells, and it is nearly exclusive.

In early embriotic cells, there are a number of non-methylated CpG islands and there are methylated CpA dinucleotides. Currently, there is no known reason.

The methylated regions include less functional genes, the methylation inhibits transcription.

The inactive X chromosome has many methylated CpG islands.

Gene regulation during transcription

Please note, below is only a brief summary. Please refer to additional chapters for more information: transcription factors and regulation of transcription!

The eukaryotic RNA polymerase II is not enough to begin the transcription, the initiation requires additional proteins, those which we define as transcription factors.

There is a minimal eukaryotic promoter which is very week in itself, thus, the transcription requires additional activators.

The minimal promoters (core promoters) are:

- TATA box and initiator region (INR), OR
- INR and DPE (downstream promoter element).

Activator sequences can found upstream towards the 5’ end direction from the TATA box within 100 to 200 bp:

- CAAT-box: the first known activator sequence, it enhances the efficiency of the promoter and is non-specific
- GC-box: GGGCGG sequence, often in multiple copies (these are also effective in reverse orientation)

There are two types of regulatory elements which act to the transcription:

1. Cis-regulatory elements: specific (non-coding) DNA sequences in determined chromosomes, they act only on the adjacent/nearby genes.
2. Trans-regulatory elements: DNA sequences encode transcription factors, which may regulate the expression of distant genes.

Enhancers

Enhancer are short regions of DNA (50-150bp) and can be bound with proteins (transcription factors) to enhance transcription levels of genes.

Enhancers are primarily cis-acting elements but they may all three of the orientation described below:

- in the 5’ direction from the gene (from the promoter) ,
- within the gene
- After the gene in 3’ direction

They can act also far from the gene (kilobases) in contrast to the core (base) promoter, which have a relatively fixed position near the gene. Their sequence is conserved, but the effect is independent from the orientation (reversed orientation does not affect the function). More sequence details are important (multiple base mutation observed reduction in effectiveness) as by the basic promoter sequence.

Silencers

Silencers are DNA sequences capable to bind repressors (transcription factors). When a repressor protein binds to the silencer region, RNA polymerase cannot bind to the promoter region (blocks the transcription) so in contrast with enhancers, silencers have a negative effect on transcription.

Transcription factors

Transcription factors are intense DNA-binding proteins regulating gene transcription. Additionally, they both have an activator or repressor function.

They have three basic domains which builds up their structure: DNA-binding domain (attaches to a specific DNA sequence), trans-activating domain (can bind other proteins such as coregulators) and signal sensing domain (capable to up-regulate or down-regulate the gene expression according to external signals).

Some important transcription factor families according to the DNA binding domain:

Helix-turn-helix proteins

- Three large α-helical region (helixes) forms the binding region, these are separated with short, flexible sections (turns)
- They are important in some of the ontogenetic stages, as commonly referred homeotic genes (they regulate many other genes)

Zinc Finger Proteins

- The DNA binding region has loops which are identical to the fingers of a glove
- They bind to the recognition site as dimers
- Steroid receptors also belong to this protein group

Amphipathic helixproteins

- Their names originate from the dimerization domain
- Leucine zipper proteins also belong to this protein group

Introns, splicing and alternative splicing

Introns interrupt the eukaryotic genes encoding sections, which splice out from the genes during the maturation of mRNA. The maturation of mRNA needs consensus sequences, for example at the 5' and 3' points in the sequence, and the junction sequences.

It is often possible to form different combinations because there are more consensus sequences and the sections between these sequences can be combined in the precursor mRNA. The implementation of the various options is the alternative splicing. This process is regulated and therefore, results in the coding of multiple proteins by a single gene. Some variations can be defined as abnormal because they may be implicated in some diseases, genetic disorder and or cancer.

By alternative splicing, exons can be skipped from the pre-mRNA, introns can remain or the existing exons can combine alternatively.

RNA maturation – degradation control

In eukaryotes, in spite of degradation of the mRNA it matures, receives a CAP and a polyA-tail.

The concentration of mRNA in the cytoplasm depends not only on the synthesis rate, but also depends on the stability of the mRNA. Without the polyA tails, mRNAs quickly degrade. In special cases, other factors can affect the stability of the mRNA, for example: hormones, low molecular weight compounds and 3'-UTR (untranslated) sequences.

RNA maturation – RNA editing

During RNA editing, specific nucleotides can be converted to other nucleotides. This can create changes within the amino acid sequence, new STOP codons, or new splice sites. Cytidine (C) can be converted to uridine (U) and adenosine (A) to inosine (I).

Examples

C→U

The Apo B100 and ApoB48 proteins are expressed from the same gene however, within the intestines the mRNA enter an editing mechanism, specific cytidines are deaminated, resulting in a CAA sequence which will be edited to be UAA and or a STOP codon. As a result, the gene product in the intestines will be shorter (the DNA sequence does not change).

A→I

Another example is the deamination of adenosine and is the primary form of RNA editing in mammals. Inosine can behave as G bases. This conversation can be specific (only determined adenosine bases are converted) or indiscriminate (where up to 50% of the A bases are edited). They are very important in neural cells.

Note: Additionally, there are chemical agents which can inhibit the translation, such as neomycin (antibiotic), ricin (toxin) or cycloheximide (a protein produced by bacterial species).

RNA interference (RNAi)

RNA interference is a process where transcribed mRNAs degrade in a sequence specific way. The enzyme and an enzyme complex is important in this process.

Dicer

It is an enzyme with RNase III activity and can recognize double stranded RNAs and cut them into small (21-22 nucleotide long) fragments: siRNAs (small interfering RNA). These small fragments twist out and separate into two pieces: a guide and a passenger RNA strand. The passenger RNA strand degrades and the guide builds it into a RISC complex.

One part of the Dicer recognizes the double stranded RNAs and another part cuts them into small pieces. The distance between these two parts define the length of these siRNAs.

RISC

There is an enzyme complex known as the RISC complex=RNA-induced silencing complex, as formerly discussed: the guide strand of RNA builds into this. This RNA containing complex can bind other RNAs that are complementary to the incorporated strand. The RISC degrades the bound complementary RNA strands. (The RISC complex does not degrade DNA strands!)

RNA interference in medicine

We can generate double stranded RNA to trigger RNA interference and gene silencing. We can induce this process with single stranded RNAs or special small RNAs that have double stranded sections in their structure such as small hairpin RNAs (shRNA) and microRNAs (miRNA). If these RNAs contain a complementary section to the targeted gene, the dicer enzyme will initiate the process. We can use this technique against incorrect folded proteins in neurodegenerative diseases (against the creation of protein aggregates) or against virus proteins. All viruses create an RNA somewhere during the multiplication (see the picture in Virus lesson) and we can make antisense RNAs against these molecules.

Protein degradation and posttranslational modifications

Gene expression can be regulated even after the protein synthesis. Most of the proteins need post-translational modifications to reach their functional form. Without these modifications the protein is not considered useful. If there is an unneeded or damaged protein, the cell can label the protein with the addition of ubiquitin which will activate the degradation of the protein.

Types of post-translational modifications:

- Phosphorylation
- Glycosylation
- Ubiquitination
- S-Nitrosylation
- Methylation
- N-Acetylation
- Lipidation

15. INTRACELLULAR PROTEIN TRANSPORT

15.1. PROTEIN SORTING AND TRANSPORT

There are two major protein transport pathways:

- in/out of the cell
- within the cell.

All the proteins are synthetized in the cytosol on ribosomes. If there are no extra signals in their sequence, they remain in the cytosol. If the proteins contain ***targeting (or sorting, or signal) sequences***, they will be transported to their final destination, into one of the cellular organelles: mitochondrion, nucleus, peroxisomes, or ER. From the ER, the proteins may have more choices: they can stay in the ER, move into the Golgi, transported into the lysosomes, built into the cell membrane or secreted from the cell. Signal sequences are ***sufficient and necessary*** for proper protein movement. Some proteins may contain more than one sorting sequences. What is important for the recognizing receptor is the steric structure or the physical properties of the signal sequence region of the protein, rather than the actual amino acid order.

These **protein movements** may occur like gated, transmembrane or vesicular transport, can be uni- or bidirectional. They are all facilitated/receptor mediated transports and notably, the energy used is not intended for the direct transport. Considering the translational status and the higher structure of the moving protein, the transport can be co- or posttranslational, and in the folded or unfolded state of the cargo protein. The signal sequence may remain on the protein after transport (nuclear – shuttle proteins), or can be digested after transport (mitochondrion, ER).

Every transmembrane protein movement within the cell has **basic structural elements**: these are the targeting sequence of the protein, a receptor, specific for this sequence, and some type of translocon, the latter one is different depending on the subcellular organelle (nucleus, mitochondrion, peroxisome and ER). It is important to note, prior to intracellular vesicular transport (towards lysosome, cell membrane, Golgi, secretion) the protein first must enter the ER via transmembrane transport.

Nuclear protein transport

The **main *features*** of nuclear protein transport are: bidirectional, proteins are transported posttranslationally in a folded state, targeting sequence is not cleaved (examples on targeting sequences: in lectures) and gated transport. Protein movement occurs through the nuclear pore complexes (NPC) (see structure in: Membranes). Small molecules may diffuse passively through the NPC. Larger molecules must carry targeting sequences which are recognized by specific import or export receptors. There are molecules which shuttle back and forth between the cytoplasm and the nucleus. One type of gene expression regulation is a protein modification which makes a targeting sequence available for the receptor. There is an extensive nucleo-cytoplasmic RNA movement. RNAs can be carried by receptors which recognize bound proteins, or by specific RNA receptors (see more in Transcription, and Biogenesis of ribosomes – in Translation).

The **basic *components*** of nuclear protein transport are: protein with localization signal (NLS: nuclear localization signal for import, NES: nuclear export signal for export from the nucleus), Ran: small GTP binding protein (active with GTP-bound form), Ran-GAP: GTPase activating protein – in cytosol (producing Ran-GDP from Ran-GTP), Ran-GEF: guanine exchange factor – in nucleus (producing GTP bound form of Ran), NTF2: taking Ran-GDP back to nucleus. Because of the different localizations of the GAP and GEF factors, Ran-GTP concentration is much higher in the nucleus, than in the cytosol.

Protein with NLS in cytosol will be bound by ***importin***, a soluble NLS receptor. Importin also recognizes certain proteins of the NPC, and the cargo (transportable protein) and receptor complex will move into the nucleus. In this movement, the FC (phenylalanine-glycine rich) regions of the **nucleoporins** play an essential role (see: in Nuclear membrane). In the nucleoplasm, the high level of **Ran-GTP** will cause to dissociate the importin-cargo complex. Importin will return to the cytoplasm bound to Ran-GTP. In the cytoplasm Ran-GAP will help the hydrolysis of GTP into GDP.

In contrary, the high nuclear Ran-GTP concentration will promote the association of proteins with NES and the receptor, ***exportin***, and in complex with Ran-GTP transported into the cytosol. There Ran-GAP initiates hydrolysis of GTP, resulting in the release of the cargo and exportin. The latter one will return into the nucleus.

Peroxisomal protein transport

Protein transport into peroxisomes is similar to the nuclear one: proteins are moved posttranslationally, and are fully folded. But the movement is unidirectional: from cytosol (the place of synthesis) into the final destination (peroxisome). Two types of ***peroxisomal targeting signals*** (PTS1 and PTS2) are recognized by different ***receptors*** (Pex5p and Pex7p), which also bind the components of the docking complex in the peroxisomal membrane. The docking complex together with associated other membrane proteins (RING complex) promote the protein translocation into the peroxisome. PTS receptors will return into the cytosol after the event.

Protein translocation into the mitochondria

Main **characteristics** of the mitochondrial protein transport: posttranslational, protein is translocated in an unfolded state, signal sequence may be cleaved (matrix) or not (outer membrane). The protein may contain more than one targeting signal depending on the final intraorganellar destination (matrix, inner or outer membrane, intermembrane space) and the type of movement of the protein (see later).

There are more types of translocase complexes in the inner and the outer membrane of the mitochondria which contain the receptors for the targeting signals as well as the translocator. The most important ones are the **TOM and TIM complexes** (translocase of the outer or inner membrane, respectively). Additional complexes play distinctive roles in the protein movement, too: the SAM complex helps the proteins to find proper arrangement in the outer membrane, and the OXA complex helps to insert mitochondrially synthetized proteins into the inner membrane.

The protein is transported **unfolded** from the cytosol, cytosolic chaperones (Hsp70 family) help the protein to remain in an unfolded state. This step and the release of the protein by the chaperone including the binding of the targeting sequence by the membrane receptor require the energy of the ATP hydrolysis. After the signal sequence has passed through the translocator of the TOM complex, it is going to bind the TIM complex. In order to pass through the TIM complex requires the membrane potential of the inner mitochondrial membrane which has been built by the electron transport system (see later in Biochemistry). This membrane potential makes the matrix side of the inner membrane negatively charged, helping the signal sequence with positive AA side chains getting through the channel. Once the protein is in the mitochondrial matrix, the N-terminal signal sequence is cleaved by a **signal peptidase**. The remaining polypeptide is bound by mitochondrial Hsp70 and Hsp60 chaperone proteins, to provide the final conformation of the protein. These steps need the energy of ATP hydrolysis.

In summary: the components of the protein transport into the mitochondrial matrix are – protein with targeting sequence, cytosolic and mitochondrial chaperones, TOM and TIM complexes containing the receptors and the translocators, energy of ATP (indirectly), and membrane potential.

Protein transport into the membranes and the intermembrane space of the mitochondria may occur via four different mechanisms (see more in the Lectures).

Protein transport into the endoplasmic reticulum (ER)

Main **characteristics** of the transport are: cotranslational, protein in unfolded state, unidirectional (exception: protein is not folded correctly in the ER lumen). The cargo protein carries an ER targeting sequence in its N-terminal region. As soon as this sequence can be recognized, it is bound by the receptor, the **signal recognition particle (SRP),** a GTP-binding polypeptide-RNA complex. The SRP-ER signal sequence interaction slows down the translation of the cargo protein. The SRP-protein-ribosome complex is attached to the **SRP receptor**, a two-subunit, GTP-binding, ER membrane embedded protein. The function of the SRP receptor is to transfer the cargo protein to the **translocon** (channel) protein, while the GTP molecules on SRP and SRP receptor is hydrolyzed to GDP and phosphate and SRP dissociates off the complex. After signal sequence and the side chains of the translocon channel protein interact with one another, translocation of the cargo protein begins. A signal peptidase removes the signal sequence in the lumen of the ER, the complete protein will fold and is then modified, and the translating ribosome will be released.

Those proteins, which will be inserted into the **ER membrane**, contain a signal anchor sequence. This sequence is not digested, and is recognized by the interior of the translocon channel. The direction of the translocation of the nascent protein depends on the final orientation of the mature protein in the membrane (see details in the lecture).

Proteins in the ER lumen can be target of extensive **posttranslational modifications**, including folding, the generation of disulfide bridges, glycosylation (see: Posttranslational modifications of proteins). Modified and folded proteins go pass through a quality control system, and only properly processed proteins can leave the ER. Misfolded or large amount of synthetized proteins initiates a so called unfolded protein response, when with decreased general translational rate and increased expression of chaperons the cell tries to correct the mistake. In the unfortunate event when folding or modification faults cannot be adjusted, misfolded proteins are translocated back into the cytosol where they are degraded by the proteasome system (ERAD: ER-assisted degradation).

15.2. Endomembrane system, vesicular transport

Cellular membrane is a semipermeable membrane, which is freely permeable for hydrophobic and small neutral molecules (O_2, CO_2), and partially to water, but impermeable for ions, polar organic molecule, and large molecules, like proteins. As we saw, proteins are transported into the nucleus via the nuclear pore complexes, into mitochondria through translocators, into the ER with the help of translocons. All these movements use receptors and specific targeting sequences.
Proteins move from the ER into the Golgi system and the lysosomes with vesicles. Exocytosis and endocytosis occur via **vesicular transport** as well. Those proteins which are located in the lumen of the vesicles will be released from the cell (exocytosis) or transported into the lumen of ER, Golgi or lysosomes. Proteins sitting in the vesicular membrane will be located in the membrane of the cell or one component of the endomembrane system. Not only proteins are transported with vesicles, but large membrane pieces as well, so the transport cannot be unidirectional.

Generation of vesicles

A vesicle is **budding** off from a donor compartment, targeting and fusing to the recipient (acceptor) compartment. Cargo proteins are carried in the lumen of the vesicle, usually bound to receptors inserted in the vesicular membrane, or in the membrane, which will remain in the recipient membrane. This is true for the exocytic vesicles as well, in which case the final destination of the membrane pro-

teins is the cell membrane, while the luminal proteins are secreted. **Secretion** pathways can be constitutive (continuous – like secretion of albumin in the liver) or regulated (vesicular fusion happens after specific signals – release of insulin or neurotransmitters for example). The **endocytic** pathway starts in the plasma membrane, and proceeds through endosomes and lysosomes.
The **components** of a vesicle are the following: coat proteins (COPI, COPII, clathrin), adapter proteins (connection between coat and receptor/cargo proteins), cargo proteins (soluble and transmembrane). (See details of vesicle components and structure in the lecture.)
For generation of a budding vesicle at a certain membrane region the **phospholipid** structure of the membrane has to be changed. This activates a membrane bound GEF (GTP exchange factor), the substrate of this factor is a soluble **small GTPase** protein. It is activated by GTP binding and attached to the membrane. This change will gather **adaptor proteins** then **coat proteins** to the forming (curving membrane) vesicle together with the cargo proteins. When the vesicle reaches its final shape, it is released from the donor membrane. (More details see in the lecture.)

Fusion of the vesicle with the target membrane

Travelling of the vesicle to the target membrane happens after uncoating (recycling the coat proteins) and with the help of the cellular microtubular system. There are helper factors to find for the vesicle the correct **fusion** site on the recipient membrane: small GTP-binding proteins (Rab) on the vesicle and/or the target membrane, Rab effector proteins (motor, tethering), SNARE complexes (v-vesicular, t-target membrane).

Steps of vesicle mediated transport:

- Selection of the cargo molecules – either membrane proteins or soluble proteins.
- Budding and release of the vesicle.
- Uncoating and transport of the vesicle.
- Tethering, docking and fusion with the target membrane.
- Disassembly of the SNARE complex.

Golgi system

Proteins synthetized on the rough endoplasmic reticulum **enter the ER** cotranslationally. There are different **modifications** occurring on the newly synthetized proteins in the ER lumen: actions of signal peptidase (removing signal sequence), of *N*-oligosaccharide transferase (glycosylation of the protein), of transamidase, of disulfide isomerase (PDI) and of chaperonines to provide secondary and tertiary structures and posttranslational modifications of the protein. Naturally not all the modifications take place on every polypeptide chain, only those to which there are sequence instructions in the protein.

To move further the protein has to enter the **Golgi apparatus** (see structure of Golgi apparatus in the Endomembrane system) via COPII coated vesicles. There are enzymes in the Golgi system for further **protein modifications**, mainly addition and removal of oligosaccharide complexes. Some proteins which reside in the Golgi remain there, some are retrieved back to the ER via COPI coated vesicles. These proteins either were carried to the Golgi by mistake, or they had to be modified in the Golgi, and then returned back to their final residence (the retrieval mechanism is explained in the lecture in details).

The majority of proteins move from the *trans*-Golgi network (TGN) **after sorting**: retrograde to Golgi, to lysosomes or to the plasma membrane. The sorting signal can be some posttranslational protein modifications, protein aggregation, or the presence of lipid rafts in the TGN membrane. Those proteins which are secreted by the cell leave the cell via exocytosis.

Endocytosis

The **functions** of endocytosis are multiple: to take up certain essential materials (iron), to remove dangerous or toxic molecules from the cell's environment, to change the cell surface according to the needs of the cell. Cargo is bound by membrane receptors, adaptor proteins and **clathrin** generate a vesicle which after decoating will fuse with an endosome. The membrane receptors can be digested or reused after recycling. The separation of the cargo and the receptor occurs in the endosome which has a special, low pH internal milieu. This pH of the lumen of the endosome is changing; this is the base of the distinguishment of **early, late endosomes, and lysosomes**.

Lysosomes are packed with acidic hydrolytic enzymes, what can digest almost any kind of complex molecules. Note that those specific enzymes are transported from the ER to the lysosome. The targeting sequence of these proteins is a special **mannose-6-phosphate (M6P)** created in the Golgi system, where it is recognized by its receptor, and via clathrin coated vesicles the proteins are transported into the lysosomes.

16. VIRUSES

16.1. DISCOVERY

The word “virus” is derived from the Latin language, it means, poison. In 1892, Dimitriy Ivanovskiy demonstrated by experimenting these are not poisonous agents: he diluted a suspension of Tobacco Mosaic Virus and put it upon Tobacco leaves: notably, the effect was similar to all the leaves resulting in the development of necrotic spots. The concentration was not an affecting factor.

1899: Martinus Beirnick described the viruses as: such a pathogen that are not visible under microscope and can pass through the bacteria filter.

Finally, they were described and characterized following the discovery of electron microscope.

In 1953, Salvador Luria and his collegues wrote, that viruses are such organisms that are not visible under microscope, can infect living cells and reproduce themselves in those living cells.

In 1967, Luria and Jim Darnell wrote down that viruses are such organisms that have nucleic acid (DNA or RNA), can replicate in living cells using directly its processes to build up their own particles which also contain viral genom so they can infect other cells.

16.2. FEATURES

Living or not? It is a continous scientific question. They are „only” nucleoproteids (nucleic acid + protein) and canot replicate themseves without the host cells. Viruses do not have ribosomes: so they are not capable to replicate themselves, they need a host cell for these processes. Viruses can be regarded as intracellular parasites, because they cannot grow or multiply without using reproduction apparatus of the host cells.

The average size of viruses are between 20-300 nanometers (nm): the smallest known viruses are Parvoviruses (20 nm), the largest viruses are the Poxviruses (400 nm). The length of Filoviruses are 10000 nm, but the diameter of theirs is just 80 nm. Viruses are usually haploid=n, except members of the Retroviridae/Retrovirus family, which are diploid: 2n. The containing nucleic acid can be circular or linear, and this nucleic acid could be segmented, and potentially present in several pieces, for example the RNA of the Influenza virus comprises eight strands. Each virus contains just one type of nucleic acid, but we can find every nucleic acid type:

- dsDNA: eg. Herpes viruses, Poxviruses, Adenoviruses
- ssDNA: eg. Parvoviruses, Circoviruses
- dsRNA: eg. Reoviruses
- (+) ssRNS: eg. Enteroviruses, Rhinoviruses, Hepatoviruses, Cardio viruses
- (-) ssRNS: eg. Paramyxoviruses, Rhabdoviruses, Filoviruses, Delta virus, Bunyaviruses, Hantavirus, Influenza A, B, C
- ssRNA (retroviruses=RT)
- dsRNA RT

The viral genome size is approximately 100-1000 kb. MS2 bacteriophage virus have the smallest genome, its size is 4kb and there are only 4 genes on it. Viruses could have bidirectional promoters (eg. Geminiviruses) when both 5'-3 'direction and 3'-5' direction could be meaningful.

The double-stranded DNA genome of viruses can be linear (Adenoviruses, Herpes viruses and a number of phage), while others have a circular genome (Baculovviruses, Papovaviruses).

16.3. Structure

We can study these organisms with the electron microscopes, magnified at least 1000x.

3 Parts of a virus:

1. **Core**: nucleic acid core: contains only one type of nucleic acid
2. **Capsid**: the coat
   2 part of it:
   a) Nucleocapsid: it closely related to the core
   b) Outer coat from homologous proteins
3. **Envelope (=peplon)**:
   a) Lipid (from the host cell)
   b) Proteins: (from the virus)

Capsomer subunits (monomers) build up the capsid, which is the coat protein of the virus. This capsomer create the nucleocapsid with the nucleic acid. The virus may have a peplon (not always), this peplon is a double membrane phospholipid which is derived from the host cell. Viral proteins can attach to the double membrane.

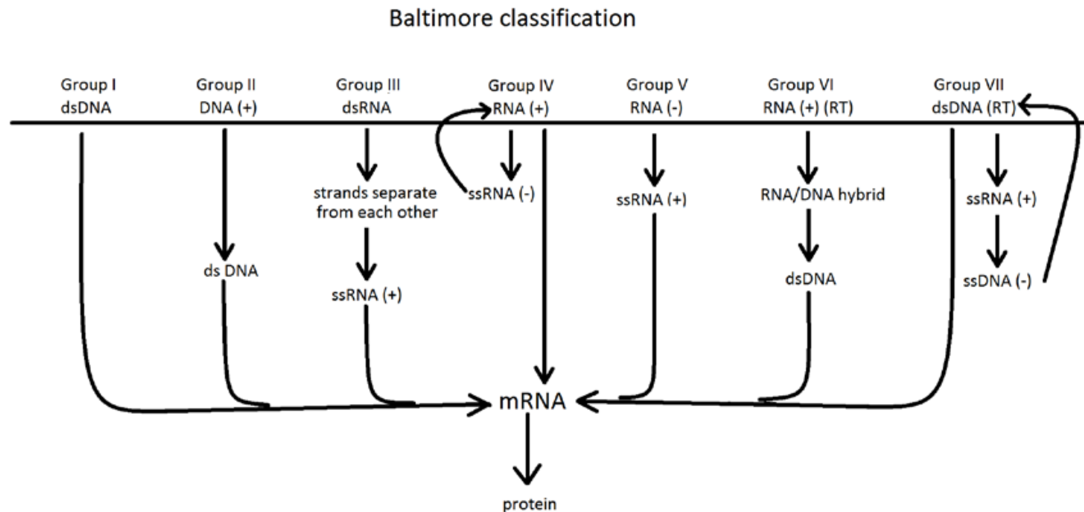
16.4. Classification

Life forms of viruses

1. Virion: extracellular life form. In this phase, viruses do not show any symptoms of life outside the host cell. In this condition we call viruses „virion”. In ideal conditions they retain their infectivity.
2. Vegetative virus: intracellular life form, inside the host cell. In this phase, viruses are able to multiply, thus, they are considered to be alive.
3. Provirus = integrated virus: the DNA of the virus is integrated into the host cell: not all viruses are able to transform to this form, only retroviruses. (There are so called prophages=integrated bacteriophages.)

According to the type of nucleic acid

Baltimore classification
Group I: dsDNA viruses
Group II: ssDNA viruses: positive strand or sense DNA strand
Group III: dsRNA viruses
Group IV: (+)ssRNA viruses: positive strand or sense RNA
Group V: (−)ssRNA viruses: negative strand or antisense RNA
Group VI: ssRNA-RT viruses: positive strand or sense RNA with DNA intermediate in life-cycle: Retroviruses
Group VII: dsDNA-RT viruses



Accodring to the symmetry

We can distinguish these shapes only during the virion state of the virus.

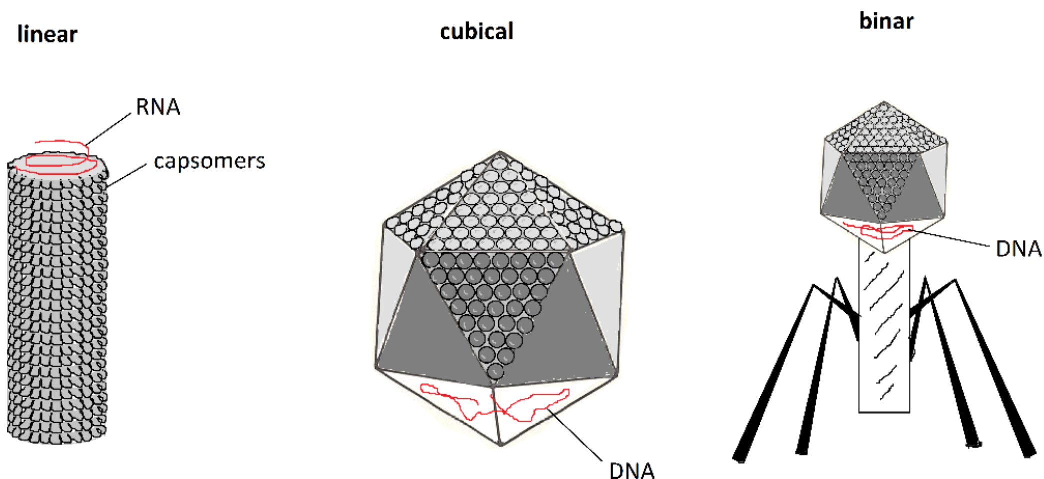
1. **Helical**: subtypes:
   a) Rod shaped, e.g.: Tobacco mosaic virus (TMV)
   b) Bullet shape: e.g.: Lyssavirus
   c) Thread-like e.g.: Filovirus

2. **Cubical or icosahedral** : e.g.: Herpesviruses, Adenoviruses
   It is composed from 20 equilateral triangles: it has 30 edges and 12 vertices. The subunits on the apexes are pentamers, the subunits on the edges are hexamers. There are 5 and 6 neighboring capsomers around these.

3. **Binary**: e.g.: bacteriophages
   Two types of symmetry are combined in this virus: the cubical and the helical simmetry: the head is cubical and the tail is helical.

4. **Pleiomorf**: there are not any capsids, there is only envelope

5. **Complex**: virus that does not belong into any of previous listed groups, for example Pox-viruses (Poxviruses are oval and they cause smallpox) and we classfy here the HIV too.



According to the host

1. Plant infecting viruses: Tobacco mosaic virus (TMV), cauliflower mosaic virus (CMV)
2. Bacteriophages: T2 bacteriophage, T4 phage, P1 phage, P22 phage
3. Animals (and human) infecting viruses: varicella, rubella, foot and mouth disease

16.5. Infection and Reproduction

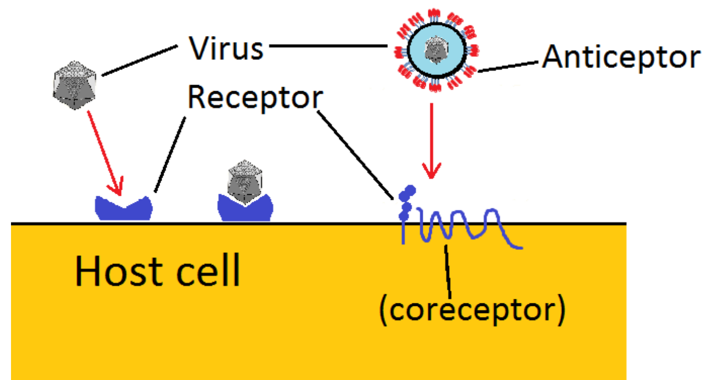
There are large differences in the infection and reproduction between the different types of viruses but we can identify some of the steps describing the primary process.

Adsorption

Adsorption of viruses can occur in several different ways:

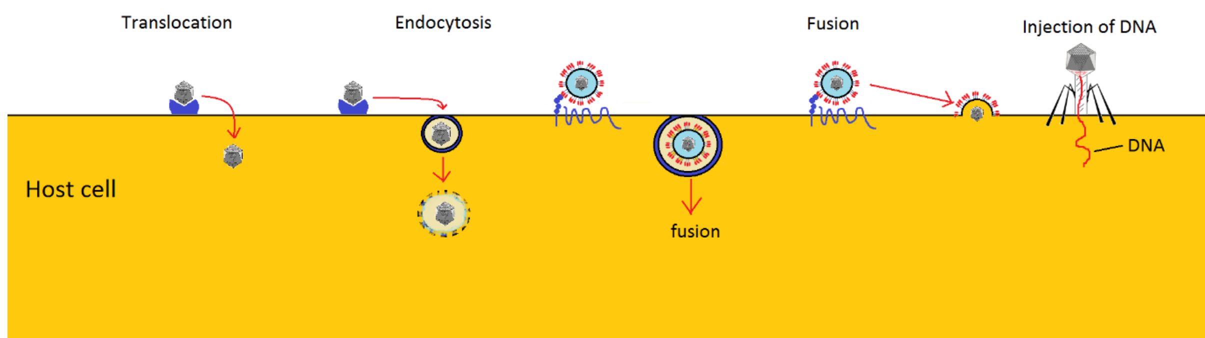
- with ionic linkage: double positive ions (e.g.: Magnesium ions (Mg^{2+})) are provided to ensure it happens
- receptor-anticeptor binding: the virus links to the host cells receptor with its anticeptor

The change of the receptor could inhibit the adsorption (for example the mutation of CCR5 (this is the coreceptor of CD4 on the T-helper cells) could cause HIV „resistance”).



Penetration

The virus passes through the membrane during this phase. It can be **translocation**, **endocytosis**, or **fusion** between the virus envelope and the host cells plasmamembrane.



Decapsidation (if it is necessary)

Decapsidation can occur in many distinctive ways. The capsid could detach during the penetration, but this can stay on the virus genome and accompany it to the nuclear membrane, where it helps the nuclei acid to pass through the nuclear pore resulting in degradation immediately following this process. It is possible only a part of the capsid detaches (eg. Reoviruses) and it is also likely enzymes will digest this capsid (eg. Poxviruses).

Virus synthesis

Replication of nucleic acids

During this phase, all of the nucleic acids replicate: every nucleic acid in its own way: according to their nucleic acid type (ss DNS, ssRNS, …).

Transcription of mRNA

mRNA synthesize from the viral genome and the host cell begin to translate the viral proteins.

Translation

Early translated proteins are defined as enzymes important in the reproduction Following production, late proteins begin to express to generate structure proteins. These proteins synthesize in much larger quantities, when compared with early proteins.

Assembly

Following the replication of nucleic acid and the expression of the structural proteins, the virus particles begin to assemble. If more viruses infect the same cell, it is possible the virus will not be packed into its own coating! (In this case the „descendants” of these wrongly-packed viruses are normal because the DNA is not altered)

Cubical viruses can create empty capsids (capsids without nuclei acids) because the nucleic acids do not participate in the built-up, they are going to be packed into the capsid afterwards.

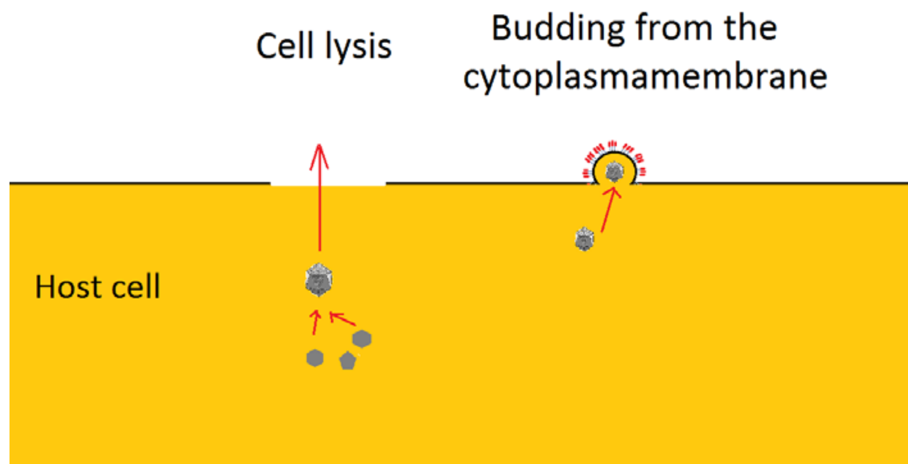
Helical viruses: the capsomer and the nucleic acid create spontaneously the virion, so helical viruses cannot create empty capsids.

Escape

Viruses can escape out from cells in several different ways: according to the coat or peplon structure.

- **Viruses without peplon**: the cell lysate and the viruses escape
- **Viruses with peplon**: with budding: the peplomer build into the host cell membrane, the nucleocapsid bind to this and then they escape while budding. The cells do not lysate in this way.

Herpesviruses are exceptions: the nucleocapsid assembles in the cell nucleus, the maturation and the peplon-creating occur in the nuclar membrane and it buds from it to the cytoplasma, after it viruses bud from the cytoplasma membrane



16.6. Some medically important viruses

Oncogen viruses

Rarely do we acknowledge the oncogenic viruses but they play a major role in cancer diseases and in malignus tumors. There are some researchers who believe 40% of cancers could be caused by viruses.

The first evidence viruses have role in human cancer were provided by Denis Burkitt, a British surgeon. The Burkitt's lymphoma was named after him in 1956-58. The highly malignant tumor is prevalent particularly in African children on the head and neck, it starts from the lymph nodes. Epstein and Barr demonstrated in 1964 a herpes virus caused this cancer. It was named after the two researchers.

Today the most widely known cancer-caused viruses are:

- the Hepatitis B and C that may cause liver cancer;
- Epstein-Barr virus may induce blood and lymphatic tumors
- Human Papilloma Virus can create a Merkel-cell skin cancer,
- XMRV may induce prostate cancer.
- There are retroviruses too!: HTLV-1 may produce adults human T-cell lymphoma

The research with the Hepatitis B changed the disinfection methods of medical devices: this virus is much more resistant to heat and disinfectants and it can spread with needles in the case of surgeries, drug abuse or making tattoos.

AIDS

AIDS=Acquired Immune Deficiency Syndrome, that is caused by HIV virus (=Human Immunodeficiency Virus)

It was indentified in 1981 but it was known for a long time. HIV is a retrovirus and causes a lifelong infection. It is 100% lethal.

The diameter of the virion is 100 nm, its symmetry is complex. The genome is a positive strand single stranded RNA with 9200 nucleotide, the viral genes are in 3 groups. It attacks the T lymphocytes, often attaching to the CD4 receptor. The infection incubation period is usually 5 years (perhaps more) Infected humans can infect during the incubation period!

There is no current known cure, but antiviral agents can slow down the course of the infection and disease drammatically.

A mutation can protect! There is a mutation in the CCR5 receptors gene preventing the adsorption of the virus to its anticeptor. This mutations rate is 0,2-11% in the Caucasian population and just 2% in the African population.

The virus can infect infants during the birth and during breast-feeding.

Influenza viruses

There are three types of influenza viruses: A, B and C

Its genome is segmented and negative single stranded RNA, the virus contain 8 segments.

It has a lipid-containing envelope and from this envelope neuraminidase (NA) and haemagglutinin (HA) glycoproteins stands out. (These two materials can be recognized and detected by the host immune system, so they are important in vaccine production.) We can classify and denominate these viruses according to these two molecules: the haemagglutinin has 16 subtypes: (H1-H16) and the neuraminidase has 9 (N1-N9). For example, H1N1 means this virus contain the H1 subtype of haemagglutinin and N1 subtype of neuraminidase.

When these surface glycoproteins change, the immune system no longer recognizes them, which is precisely why we should create newer vaccines.

New versions of viruses can be established with point mutations (antigenic drift) or if more viruses infect the same cell they can recombinate (antigenic shift – because of the segmented genome can mix during the assembly the parts/segments and in this way new versions can be estabilished from these segments).

The pandemic H1N1 virus in 2009 was derived from swine-, bird- and human viruses (with antigenic shift).

Chicken pox

The name of the virus is Varicella simplex or varicella-zoster virus (VZV or HHV-3). It is a Herpes virus. It can infect only once, but the virus can re-fracture and cause shingles at the next eruption. Today, we have vaccines against this virus and even adults can give this vaccine. The vaccine was developed in Japan in 1988 by Takashi and has been operational in Hungary since 1998 meaning a majority of the population have not received it. The vaccine is a live and attenuated virus.

Warning! During chickenpox it is not recommended to administer use aspirin or medications containing aspirin under 18 years of age as it increases the risk of the Reye' syndrome (this is a life-threatening inflammation and swelling of the brain, it can cause liver disease, too).

16.7. Defense: vaccinations, medicines, natural defense

Variolisation or variolation

The history of immunization dates far back into history: the first described variolisation was in China in the 1st century! They used dried and pulverized smallpox rashes: they absorbed it in their nose. This method was used for a very long time including in Europe. Mention and use of this was first noted in the USA in 1790.

Edward Jenner discovered the smallpox immunity caused cowpox immunity too: The first vaccine was created in 1970. The word “vaccinae” derived from these experiments, it means: from cow-> vaccine. Jenner discovered those who undergone the cowpox (milkmaids), acquired immunity against small-pox. Following the discovery, Jenner initiated experiments with the diabled son (5 years old) of his servant and successfully treated the patient meaning, he did not contract smallpox. Jenner’s second trial patient was himself.

Types of vaccines against virus:

- Live attenuated vaccine (LAV) (e.g.: polio, tuberculosis, yellow fever): they derived from dis-ease-causing pathogens which have boon weakened. They will cause little or no disease.
- Inactivated vaccines/killed pathogen (e.g.: influenzea, cholera, hepatitis A, B, bubonic plague). Inactivated vaccines are made from viruses which may have been killed by physical or a chemical processes, e.g.: with formaldehyde (but the vaccine itself contains only traces of formaldehyde) These vaccines cannot cause disease, they are safer and more stable than live attenuated vaccines, but they have less strong immune response compared to live vac-cines.
- Subunit vaccines: It contains viral protein subunits which are separated and purified or can be produce by recombinant genetic engineering (produced by genetically modified organ-isms). The subunits immunize poorly, but sometimes it is considered the most safest meth-od.

Protection with drugs

Viruses themselves do not carry out own metabolism without the host cell: they do not absorb agents from the environment (for example drugs, antibiotics or any other materials). As a result, over a length period of time, no medicine was available to vombat these viruses

Today, the number of the antiviral agents grow constantly. The most widely used medicine is acyclovir: it is a selective antiviral pharmaceutical activated only in infected cells by the viralthymidine kinase (TK) enzyme by phosphorylation.

Contemporary, widely preffered antiviral picalsharmaceu: Acyclovir (eg Virokill, Herpesin, Aciclosan, Virolex, Zovirax, Ciklovir, Telviran, Aciclovir AL), ribavirin (Rebetol, Virazole), ganciclovir (Cymevene), famciclovir (Famvir), valacyclovir (Valtrex), valganciclovir (Valcyte), foscarnet (Foscavir), ritonavir (Norvir), didanosine (Videx), stavudine (Zerit), lamivudine (Zeffix), zanamivir (Relenza Rotadisk), osel-tamivir (Tamiflu), nevirapine (Viramun), inosine, pranobex, isoprinosine, indinavir (Crixivan) , emtricita-bine (Emtriva)

Those of you watching the popular TV drama series, “House M.D.”, likely heard about interferons. Researchers have been experimenting with them since the second part of the 20th century, but there are many serious side effects. Interferons are produced naturally too in infected cells, they inhibit the infection of surrounding cells.

16.8. GENETIC BENEFITS OF VIRUSES

“Vector” means, gene delivery organisms. Many viruses are used as vectors of gene delivery into the host cells. There are retro-viruses integrating into the genome but other viral vectors are used, too.

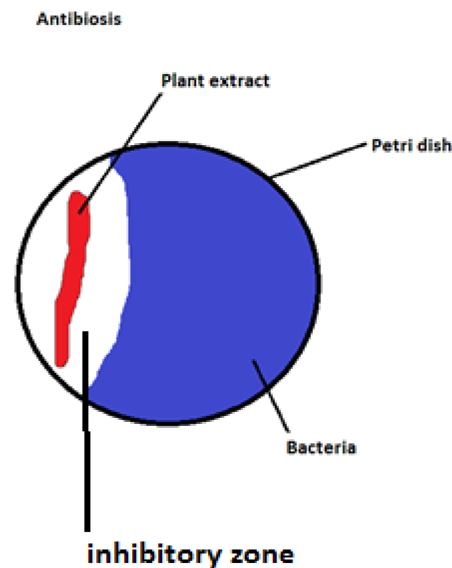
Natural genetic engineers:

- **SV40 virus vector**: it has 5243 basepairs and one circular DNA with specific endonuclease cleavige points: each restriction endonulcease can cut it only at one specified DNA sequence. We can cut it and build a DNA piece in the viral genome with ligase enzyme.
- **Adenoviruses**: they cause very mild disease in humans and thus, are considered safe. In 1999, an 18 year old male patient died due to the adenovirus infection during the gene therapy. To date, it is the only known and proven fatality due to gene therapy.
  Adenoviruses do not integrate into the genome but this gives the highest expression of the delvered genes which is why it is the most commonly used vector in the human gene therapy.
- **Vaccinia viruses**: they can insert relatively large regions into the host cells.
- **Various type of retroviruses**: they can infect only dividing cells virtually 100% probability. They integrate the genes into the host cells genome.

DNA viruses may insert a longer DNA strand into the cells but this DNAs will not be integrated into the genome, they replicate in the cytoplasm, far from the chromosomes as episomes, completely independently from the chromosomes and from the nuclei. Only some of the viruses can integrate into the host cells genome, only retroviruses.

17. ANTIBIOTICS

Antibiosis: a biological interaction between two or more organisms often disadvantageous to at least one of the participants; it can also be an antagonistic association **between an organism and the metabolic substances produced by another**. We can examine this phenomenon, for example, in plants, fungi, bacteria, algae and corals. (This interaction can also be observed between antibiotics and bacteria.)



17.1. ANTIBIOTICS

Accodring to the German *W. Pschyrembel: Clinical Dictionary:* "Umbrella term, metabolits of molds, bacteria or Streptomyces or their semi-synthetic derivatives (...) which inhibit the proliferation of viruses, bacteria, fungi, protozoa or they can even destroy them. In a narrower sense, only those medical agents known as antibiotics which act against bacteria and are not effective against viruses, and fungi."

Today, we use this word (antibiotics) only for anti-bacterial agents and fully synthetic materials are also defined as antibiotics (synthetic materials were called chemotherapeutics before). We can identify the agents most effective against viruses such as antiviral pharmaceuticals and the agents which are generally effective against fungi known as antifungals.

Disinfectants are not antibiotics.

17.2. LITTLE HISTORY – IMPORTANT STEPS TOWARD TO THE DISCOVERY OF THERAPEUTIC ANTIBIOTICS

Many plant materials are defined as bacteriostatic or bactericidal materials meaning there are a plethora of plant materials known to combat bacteria in folk medicine, such as chamomile.

History bestows upon Sir Alexander Fleming as the first and foremost association to antibiotics. He was presented the Nobel Prize for his efforts relating to antibiotics in 1945 along with Ersnt Boris Chain and Sir Howard Walter Florey. Yet, Sir Fleming was not the first scientist to have discovered the phenomenon of antibiosis and the influential effectivenss of molds against bacteria.

Classic antibiotics, derived from fungi and bacteria, are not considered entirely new, as they reportedly were described some 4000 years ago used in the production of ointments made from moldy soybeans and used to treat skin infections in China

The American Indians made pastes from moldy corn and used them for similar purposes.

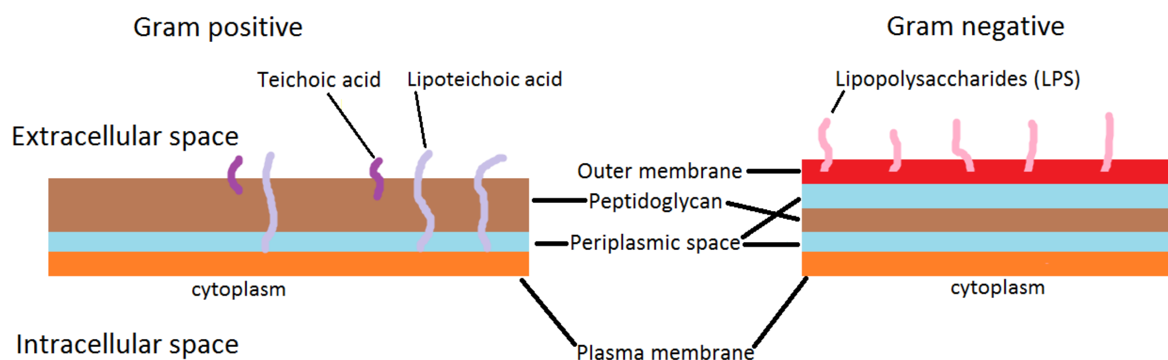
In 1867, Joseph Lister used and offered an antiseptic carbolated solution spray to his surgeon colleagues. This carbolated solution was in the center of antibacterial research for a long time.

Robert Koch developed the culture method for microorganisms, using first red algae extract, also known as agar. He systematized and divided the cultured bacteria into groups (he cultured bacteria from tissue samples, blood samples, urine samples and salvia). He discovered that bacteria are capable of producing materials preventing the growth of other bacteria, defined as a phenomenon of antibiosis.

John Tyndall described the anti-bacterial properties of some fungi in 1875.

1880: Paul Ehrlich (who also discovered the mast cells) was looking for staining procedures and dyes, which are specific for bacteria so he searched drugs which could be linked specifically only to bacteria. He worked primarily with inorganic substances. In this way, he discovered Salvarsan, which became the first synthetic antibiotic medicine and as a result, now consider Ehrlich the father of chemotherapy. (Today you can find this pharmaceutical as Arzefamin.) It was already introduced in 1910, and it was able to cure syphilis.

In 1884, Hans Christian Gram invented the Gram Staining Method, used to classify bacteria into two groups, Gram+ and Gram- bacteria)



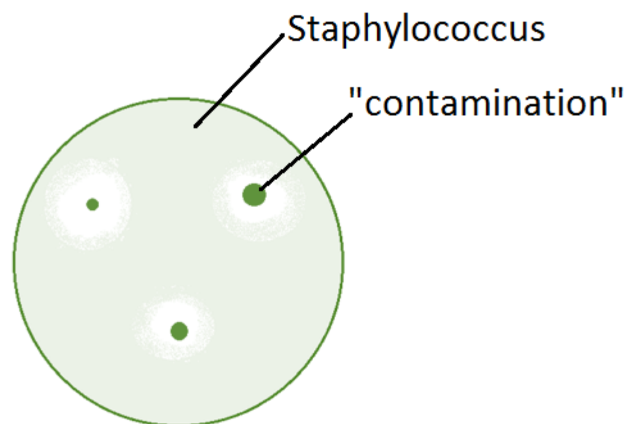
In 1896, a French medical student, Ernest Duchesne described when observing a yeast it produced substances, which can destroy bacteria. He wrote his doctoral thesis about in reference to this bacterial destruction that was made by fungal species. Additionally, he tested his theories on animals (guinea pigs) too. Mostly, he experimented with species of Penicillium, but his work was not accepted and so he did not recieve a doctoral degree.

In 1893, Bartolomeo Gosia effectively crystallized mycophenolic acid from molds of Penicillium, resultin inhibition in the growth of Bacillus anthracis. Unfortune, this experiment remained unrecognized and Gosia did not recieve any credit for this.

Sir Alexander Fleming worked as an assistant during World War I, serving on the medical team of Almroth Wright. This team endeavored to discover antibacterial agents combating soldiers' wound

infections. They cultured the bacteria on agar medium (formerly discovered by Koch), these bacteria were derived from wound samples. Notably, they implemented a variety of disinfectants yet the most promising and widely used was favored by Dr. Lister, which, although toxic and often destroyed helpful leukocytes (in faster rate, as the pathogen died), so it caused the death of wounded soldiers too many times. In 1921, Dr. Fleming discovered lysozyme, which demaged the bacterial cell wall within the human body. Interestingly, lyzozym is often used in tooth paste too. Dr. Fleming continued to experiment with lysozyme but was unable to use it in appropriate concentration, required in the healing process. This was the beginning of the scientific carreer of Fleming.

Reportedly, in 1928, ten years following WWI (according to a legend), Dr. Fleming errored when not ensuring to properly close a Petri dish containing a sample of Staphylococcus prior to leaving for his vacation. Upon return to his laboratory, Dr. Fleming discovered his sample was contaminated by air and as result, additional microorganisms infested the Petri dish destroying the surrounding bacteria (Staphylococcus). Far from the contamination the bacteria could grow nicely. Dr. Fleming discovered the infection was a mold, classified as Penicillium notatum (= Penicillium chrisogenum), so he named this substance penicillin that can kill bacteria. Dr. Flemind began experiments featuring this particular mold, in effort to describe and heighten the potential benefits and effectiveness of penicillin.



Howard Florey, Ernst Chain and Norman Heatley, all researchers at Oxford, began started to optimize the medical use of penicillin as genuine product medicine. The thrww researchers developed more isolation procedure in 1940-41, but unfortunately, vital European factories were destroyed during the war. Production of pharmaceuticals started in 1942 in the United States.

Interestingly, it was difficult to determine the best possible strain used in the development of antibiotic medicines, one which ideally produced the proper amount of antiobiotics. Towards a means of expediting the search, research centers in the United States asked the public for help in the form of requesting anyone in the possession of moldy kitchen dishes to please consider them to the appropriate laboratories for analysis. The winner was an elderly woman who sent in a moldy cantaloupe to a participating laboratory in Illinios.

In 1945, Sir Alexander Fleming, Ernst Boris Chain and Sir Howard Walter Florey recieved Nobel Prize for their work in the field of penicillin.

17.3. The first available antibiotic drugs

The date of the appearance of antibiotic pharmaceuticals on the market is shown below. Notably, Salvarsan spent 20 years waiting to be officially introduced onto the market whereas Protesil only took three years.

Salvarsan by Paul Ehrlich - 1910: effective against syphilis and known today as Arzefamin. It is only effective against spirochetes and nowadays other pharmaceuticals are considered when treating syphilis.

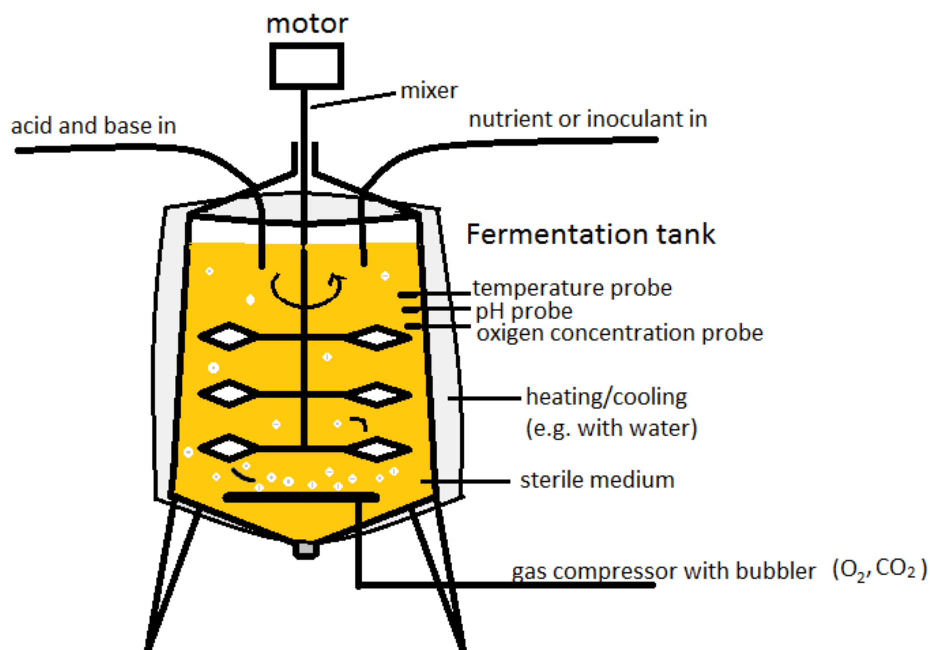
Prontosil by Gerhard Domagk et al, 1935 (Bayer): In 1939, Domagk received the Nobel Prize for his efforts in the development of Prontosil. It is effective against the immense group of Gram + cocci, but is not effective against Enterbacteriaceae.

Gramicidin by Rene Dubos: it was the first commercialized antibiotic used during WW II, 1939-1940, although others only were used in hospitals prior to WW II.

Penicillin by Fleming: 1942: Penicillin was only used by the British and their allies during WW II, including every antibiotic drugs developed in Great Britain until the end of the war.

17.4. Classification according to the production

1. **Fermentated antibiotics**: microorganism (bacteria and fungi) produce the antibiotics as secondary metabolites and is precisely why they are often referred to as natural antibiotics. They are grown in liquid growth medium in large containers (fermenters). Conditions (concentration of oxigen and carbon dioxide, pH, temperature and nutrient levels) are monitored and adjusted to attain the optimal growth and the effective production of antibiotics. Organisms may develope resistance faster to the fermentated antibiotics due to the agent pre-exposure in its natural environment. These antibiotics are more often toxic than synthetic antibiotics.



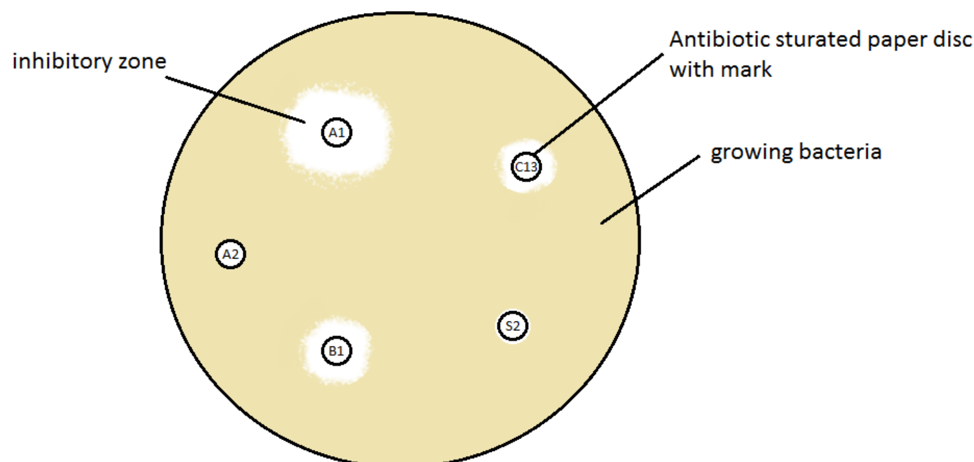
2. **Semi-synthetic agents**: chemically modified agents which are derived from bacteria or fungi by fermentation (e.g. Semicillin=semi-synthetic Penicillin derivates). These drugs are developed to decrease toxicity, decrease possibility of resistance, or increase effectiveness.

3. **Chemotherapeutic agents / chemotherapeutics**: chemically synthesized antibiotics. Chemotherapeutics feature an advantage meaning the bacteria are not exposed to the compounds until they are released on the market so bacteria usually do not have resistance against these drugs. These drugs are also designed to have even greater effectiveness and less toxicity than natural antibiotics.

17.5. Antibiogram

Antibiogram is an antiobiotic sensitivity test to select the most effective antiobioticn treating a patient. We culture bacteria derived from a patient's sample (blood, wound, saliva, etc.) on an agar medium and administer antibiotic saturated paper discs on the surface of the medium. Every paper disc contains different antibiotic. After culturing we generally note inhibitory zones surrounding the discs. We can measure the size of the inhibitory zones and choose a proper antiobiotic. It is necessary if the antibiotic treatment is not effective in treating the patient. Since the breeding and testing requires a long time (several days to one week), it is important to develop new methods (the patient might die during this time). To make decision about the treatment, we should think about the side effects and about the price too. If we choose an effective antibiotic with a narrower spectrum, it will not harm the normal flora so much.

In using this technique, we can monitor the evolution of bacteria and thus treat a new patient with potentially higher rates of success. Notably, if a commonly preferred antibiotic is used, we can easliy see if thge antibiotic loses its effectiveness or do new strains appear.



17.6. Pros and cons

Without the use of antibiotics a number of diseases may be fatal or may have severe complications (including a tooth infection). Today, an increasing number of research suggests as bacterial infections develop, or participate in the creation of different type of cancers, gastric ulcer, gastric cancer, certain types of coronary heart disease, it may be possible one day to prevent these diseases through treatment of antibiotics.

Strikingly, today there is a notable increase in the number of operational resistant strains and it is now characterized as a race between bacteria and reseachers/pharmaceutical corporations.

Antibiotics may upset the delicate balance of beneficial bacteria within the body and often these antibiotics serve to decrease the body's the natural defense.

17.7. Antibiotic interactions

Certain substances within the body can increase the time spent of antibiotics: by not allowing them decay or allow to absorbe much more from the drug. Other substances can influence the absorption of the antibiotics. Some toxic antibiotics, such tetracyclines, may cause demage resulting even death, if another mechanism increase the amount of the drug within the blood. Additionally, other substances may increase the metabolism (breakdown) of other molecules (they rapidly deflate from the body and therefore, are not effective due to the relative shortness of time exposed).

Some of the antibiotics may enhance the effects of caffeine (palpitations, dizziness and headache). In practice, a very large group of antibiotics are not easily deflated or degraded when exposed to caffeine and as a result, caffeine appears to have a continuous effect upon interacting cells. Thus, during antibiotic treatment it is not recommended to consume caffeine-containing foods and beverages. Since caffeine can enhance the effect of some pain relief medication (e.g. paracetamol), several manufacturers add caffein into pain killers and antipyretics with the active substance. The use of these painkillers should be avoided during antibiotics treatment! (E.g. Béres Trinell Pro)

Calcium can impair the absorption of certain antibiotics, e.g. Ciprofloxacin, and Doxycylin, Norfloxacyn, and do not interact with other antibiotics, such as Penicillin and Erythromycin.

Each leaflet features the warning alcohol should not be consumed during the treatment of antibiotics. Alcohol and its interaction with antibiotics may delay the degradation rate of one another. Some antibiotics prevent the degradation of alcohol, such as Metronidazole, Tinidazole, Cephamandol, Latamoxef, Cefoperazone, Cefmenoxim and Furazolidone. Alcohol can reduce the effect of antibiotics e.g. in the case of Doxycycline and Erythromycin.

Tetracyclines can form complexes featuring an immense amount of molecules, thus they should not be taken together with calcium, magnesium, aluminum-containing agents, iron containing agents and antacids.

Certain antibiotics may reduce the effectiveness of oral contraceptives. Overall, we can say the use of antibiotics can cause decay in the natural intestinal flora, so the adsorption of other subjects can change during the antibiotics treatment period until the balance is restored.

17.8. Resistance

Many antibiotics are produced by bacteria which are resistant to their own antibiotics (They reduce the chances of growing other species in the same environment: antibiosis) Resistance can be obtained either from other bacteria even from distantly related bacteria through conjugation (horizontal gene transfer).

It is enough to have some random mutation by the receptor-acting antibiotic, which causes a change in the receptor –it may create resistance. A newly acquired resistance can suddenly cause a problem by known strains too: as we first detect the pathogen we can not use an appropriate treatment on the patient, the patient can die during the time the results of the the antibiogram arrive

The resistance can be disadvantageous for the bactria too, if it needs additional energy (e.g. production of an extra protein). In this time as the antibiotic medicine is not in the environment of the bacteria, the sensitive bacteria can grow faster. (The resistance can disappear without the antibiotic.)

If we do not develope more and more new antibiotics, our future will be the same as it was before the discovery of antibiotics, we may not be defended against bacteria and people would die in a smaller or larger incision, surgery, or childbirth.

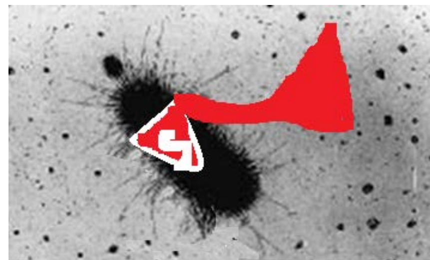
17.9. Cross-resistance

If a mutation formed, this property that prevents the bacteria against an antibiotic will be effective against similar antibiotics too: this is the effect of cross-resistance.

By chemically similar compounds, the resistance can be formed not only against an active substance but also against other similar compound or against the whole group of molecules. Cross-resistance exists generally at similar-acting drugs.

If the resistance is formed against the whole group of antibiotics we talk about "full cross-resistance". If it formed against only some parts of the group, we speak about "partial cross-resistance".

17.10. Superbacteria=superbug



Superbug: a pathogenic bacteria strain that has developed resistance to the medications normally used against it, so this bacteria will be resistant to one or more antibiotic(s) that would normally treat it.

The number of multi-resistant bacteria increases, which is due to the fact that bacterial genes can be replaced by a conjugation process. By conjugation bacteria can replace genetic material (plasmids) by direct cell-to-cell contact or by a bridge-like connection. Genes can replace between different species too in this way! This is the so called horizontal gene transfer. (Bacteria have so called plasmids (circular DNA), and the resistance genes take place here.)

Moreover, the bacteria should also take DNA from their environment, but this can retain only if this DNA has a homologous section and can be integrated into the genome. Genes can carry bacteriophages too, because they integrate into the bacterial genome and during the replication they can take away a part from the bacterial genome.

17.11. MRSA

MRSA (Methicillin-resistant Staphylococcus aureus) is responsible for the most multiresistant infection. This superbacteria is a “descendant” of the generally known S. aureus (it does not often require treatment). This resistant bacteria often cause infections in hospitals, at patients with weakened immune systems (intensive care, surgery). (Hospital infections are called: nosocomial infections). They primarily cause skin infection, which can be fatal.

It was often described to be caught from beach sand, which is an excellent medium for bacterial multiplication.

MRSA infections caused more deaths than AIDS in the United States today.

17.12. Researches nowdays

Fewer and fewer new antibiotics will be released because of resistances and cross-resistances, that is why we need antibiotics with entirely new point of attack. New research especially examine the bacterial amino acid synthesis and fatty acid synthesis in attacking bacteria. It is an important research topic the connection between the bacterial cell and the host cell.

It is an increasing point of view that by mild to moderate disease, we shoul use only bacteriostatic agents, in this case less of super multi-resistant bacteria can be created.

Much researchers pay attention to the body's immune system and would like to support it, for example with large doses of vitamin B3.

There is an ongoing research with bacteriophages too.

17.13. Primycin

(It has several names in the market, e.g. Ebrimycin gel)

The only antibiotic medicine which was discovered and isolated in Hungary, Vályi-Nagy Tibor discovered it in 1949. Thermopolyospora galeriensis fungal species produced it and there are research projects at our university (PTE) with primicyn.

It belongs to the family of macrolide antibiotics, it has bactericidal activity.

Its spectrum is very wide: it is effective against Gram-positive bacteria (and against Mycobacktérium strains). It has an effect against yeasts and filamentous fungi, vibrios, some algae, protozoa and macroviruses too. At higher concentrations it is effective against Gram-negative bacterial strains, including those resistant and poly-resistant strains.

We can use it locally, it is not absorbed into the skin or the intestinal tract.

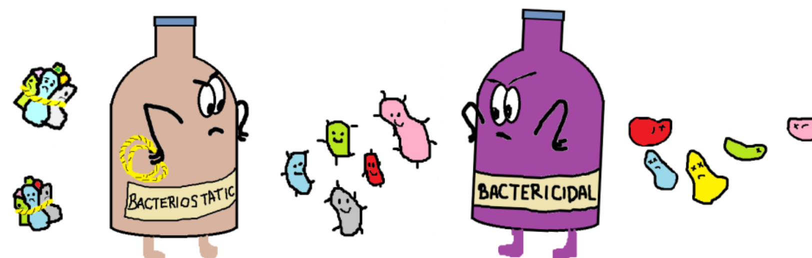
17.14. Uses of antibiotics beyond medication

We use antibiotics in the molecular biology and genetic researchers as selection tools of bacteria. For example: we can check the cloning with resistance genes. We link the resistance gene to another gene which we would like to integrate into the genome. After the cloning we culture the cells on antibiotic media: the resistant cells can grow: this contain the genes what I would like to integrate into the genome, the cells which do not contain the genes can not grow. (They are sensitive to the antibiotic.) Antibiotic media can prevent the grow of bacteria that lose the plasmid during the reproduction, because it is a continuous selection pressure in the medium.

17.15. Effect mechanism

A key principle: the more specific effect of the drug is on the pathogen, the smaller is the effect on humans (or other animals).

There exist **bacteriostatic agents**, which inhibite the growth of bacteria and there exist **bactericidal compounds**, which can destroy the bacteria. (Some antibiotics have both effect, for example according to the concentration: in bigger dose/concentration it may have bacteriocidal. Some antibiotics could have different effect on different bacterial strains.)



17.16. Groups according to the mode of action

In general there are some critical points in the bacterial metabolism what we can attack with antibiotics:

Cell wall synthesis inhibitors: act on cells that are being divided, eg. β-lactams (penicillins, cephalosporins, carbapenems) glycopeptide antibiotics. These antibiotics obstruct the formation of bacterial cell wall during multiplication, without the cell wall bacteria cells are less resistant to osmotic conditions. (Good selective target: humans do not have cell wall (peptidoglycan)) These cell wall synthesis inhibitors are able to destroy Gram-positive pathogens, but they are not effective or not so effective against Gram-negatives because they have different cell wall structure.

Enzyme inhibitors:_they can disturb metabolic processes and nucleic acid replication, e.g. DNA gyrase inhibitors.

Ribosome-acting drugs: Inhibitors of protein synthesis: e.g. aminiglikozids, macrolides, chloramphenicol, tetracycline. These drugs have effect on human ribosomes too. They can bind to our ribosomes but usually with less effectivity. Most human cells synthetize proteins at a much slower rate than bacteria, but since it is a common process these drugs have mire toxic side effects.

Inhibitors of cell membrane function: certain antibiotics (e.g.: polymyxins) act like detergents and dissolve bacterial cell membranes.

17.17. Cell wall synthesis inhibitors

Beta-lactam antibiotics

They damage the cell wall (inhibite the cell wall synthesis): they attack the bacterial cell wall, which is significantly different from the cells of the human body, so they have relatively few side effects. They link to the penicillin-binding protein and inhibit in this way the formation of peptidoglycan structure, so they act only on dividing cells. They are **bactericidal**, so bacteria-killer agents.

Penicillins and cephalosporins belong to this group

Disadvantage: the inactive bacteria survive the antibiotic therapy, so later the patient may fall back. They are ineffective against intracellular pathogens (they do not enter the human or animal cells!)

They decompose quickly, so they can spread over many doses. (e.g. 3-6 portions in a day)

Penicillins

They bind to bacterial transpeptidase enzymes, that is how they inhibite the cell wall formation (peptidoglycan formation). They are non-toxic. (It does not mean that they have no side effects! Severe side effects are relatively rare by this medicine.)

Its spectrum is relatively wide, we use it especially in ear, nose, pharynx, esophagus, and infectious diseases of the respiratory system.

Cephalosporins

They also link to the transpeptidase enzyme. They are less sensitive against conventional ß-lactamases.

We use it by infections in respiratory system, kidneys, urinary tract infections and by various wound infections. We often use them by patients who are allergic to penicillin.

They have relatively few acute side effects, rare allergic reactions or kidney damages. Coagulation disorders may occur at older patients .

They are non-toxic, but of course there may form allergy against them like by any other antibiotics.

They are also effective only against gram-negative bacteria.

Resistance to beta-lactam antibiotics

Bacteria may produce beta-lactamase enzyme which break down the drug (cleave the β-lactam ring of the antibiotic). ß-lactamases are found in all bacteria naturally but in variable amounts and with varying levels of activity. The β-lactams bind to the penicillin-binding receptor and it can activate the production of beta-lactamase enzyme, and therefore we should use inhibitor substances which inhibite the activity of the enzyme. We can use these β-lactamase inhibitors (BLI) as adjuvants with ß-lactam antibiotics: clavulanic acid (in Amoxicillin, Ticarcillin), sulbactam (inAmoxicillin), tazobactam (in Piperacillin)

Another form of resistance is when there is a change in the cell wall structure in such a way that the beta-lactam can not pass through it (protein or mucinous layer on the surface).

Aminglicosid antibiotics

They are highly efficient antibiotics, we use them especially by severe infections.

They have bactericidal and bacteriostatic effects: first, it adhere to the surface of the bacterial cells, this disturbances the permeability, after that they get into the cell and disturb the protein synthesis by binding to the ribosomes irreversibly. Because this processes we can classify these antibiotics to the group of “cell wall synthesis inhibitors” and to the group of “ribosome-acting drugs”.

They influence the formation of critical bacterial proteins: they bind to the 30S ribosome's S12 protein, thus they inhibite the binding of N-formyl-methionine on this, so the protein synthesis can not start on these ribosomes (since this would be the first amino acid binding).

We use them primarily by injection or use them topically by ointments in mouth, and skin infections (due to the poorly absorbed tract) or we can use them in eye drops too.

There is a relatively large risk of heart, kidney and neurological damage if we use it in injection. It can not reach the central nervous system by meningitis (so we can not use it by meningitis).

Pregnant women should not take aminglicosides because the baby's nervous system can damage!

Glycopeptides

The glycopeptide antibiotics can also inhibit the formation of the bacterial cell wall, but also changed the bacteria ribonucleotide synthesis. They are effective only against Gram-positive bacteria. They are bactericides (so they kill the bacteria).

They can not adsorbe from the tract, and therefore we only use them by infusion. They should only be administered slowly because they irritating all tissues. If we use them by intramuscular injections, it may cause tissue necrosis. They can demage renal function, and can cause hearing damage.

They do not penetrate into meninges, but by meningitis (inflammation of meninges) they can be transferred to the liquor.

Resistance: if the binding site (receptor) change (e.g. by mutations), it will not be able to bind, so there will be no effect on the bacteria.

17.18. Enzyme inhibitors

Gyrase inhibitors

Gyrase inhibitors act on a DNA enzyme, which may specifically be found only in bacteria and it prevents cell division in this way. We usually used them when the patient is allergic to other antibiotics.

They are particularly effective against the infections of the urinary tract, or in the respiratory tract and skin infections. They can kill even Helicobacter pylori, which is known to be involved in the most recent gastrointestinal ulceration.

Side effects: nausea, stomach pain, diarrhea. Rare but there can occur a relatively severe form of headache, dizziness, depression, insomnia and agitation.

It is not allowed to take these drugs during pregnancy, lactation and before the end of the growth, because it could form disorder in developing bones and joints.

Folic acid antagonists

Sulfonamide is the most well-known representative of this group. Prontosil (which was put to the market in 1935) also belongs to this group.

They have a bacteriostatic effect. Their structure is similar to the para-aminobenzoic acid, so they can inhibite the incorporation of this molecule in the cell wall, so the folate acid synthesis and the cell wall formation can be cancelled. They can inhibite the synthesis of folic acid by more ways and by more

points. Folic acid is essential for the synthesis of adenine and thymine, two of the four nucleic acids that make up our DNA. (Humans do not synthesize folic acid that is why it is a good selective target.)

We do not use them often, because now there are too many resistant strains, especially by trachoma (it is a special eye-infection).

We use them combinated with other antibiotics, for the treatment of pneumonia, and toxoplasmosis and also against malaria

17.19. Ribosome-acting antibiotics: Inhibitors of protein synthesis

Tetracyclines

They have bacteriostatic effects, so they inhibite the proliferation. They bind to the 30S ribosomal subunit and prevent the binding of bacterial aminoacyl-tRNA to the ribosome-mRNA complex.

They have relatively few side effects, although they may be toxic in large quantities!

They absorbed well from the tract, so they can be applied in the form of tablets.

Originally they had a very broad spectrum of activity, but it has now dropped due to the growing number of tetracycline resistance micro-organisms.

We use it mainly in the therapy of bronchitis, pneumonia, and chronic skin inflammation

Side effects: discoloration on developing children's teeth, and rarely inflammation of liver. (They extreted mainly in the liver and in the bile.)

Macrolide antibiotics

Originally Stretomyces erithreus produced them.

They inhibite protein synthesis: they bind to the 50S ribosomal subunit reversibly. They are bacteriostatic, but may be bactericidal according to the concentration and certain derivatives can be bactericidal too!

Erythromycin belong to this group, we use it often to replace penicillin, or cephalosporin (e.g. if the patient is allergic to Penicillin).

They are used especially for lung diseases if other antibiotics are not effective.

We can use them by Toxoplasmosis too (it is a neural parasite that infects primarily infants and fetuses, we can use this drug in the second half of pregnancy, to avoid the damage of the nervous system)

Other side effects: hepatic dysfunction and hearing loss. (It excreted mainly in the bile from the body, it accumulates in large quantities in the liver, so we can use them because this by hepatic inflammatory well.)

Resistance: by decreasing of permeability or changing in receptors or due inactivity of ribosomes

Chloramphenicol

They disturb the creation of proteins in bacteria, but they disturb the protein synthesis in human cells too: they bind to the 50 S bacterial ribosomal subunit reversibly and inhibite the protein synthesis.

They can bind to the 80S subunit of mammalian cells, and the 70S subunit in mitochondria too but with less effectivity.

They have usually bacteriostatic effects, and in some cases can only elicit a bactericidal effect (by Nesseria and Streptococcus pneumoniae).

They absorbed well from the digestive system.

They can passes into the cerebrospinal fluid too (the volume in the cerebrospinal fluid is the half of the volume in the serum)

They have serious side effects, that is why we use them only in cases of typhoid fever, or if we can not use any other drugs (they damage mainly the bone marrow and hematopoietic organs). We should not use them at the therapy newborns.

Lincosamide

Lincomycin and the 10x stronger Clindamycin is the best known representatives of this group.

They damage the natural intestinal flora more than other antibiotics, they often cause serious inflammation, so we use them only by severe cases, when other drugs have not worked, mainly in joints and infection of genitals.

They are bacteriostatics. They bind to the 23S part of the 50S subunit of bacterial ribosomes, so they prevent protein synthesis.

(They are effective against malaria, toxoplasmosis too.)

We can use them in the so called „toxic shock syndrome” because they block „M protein” of the bacteria.

Resistance: similar as by macrolids: with the change of the biding site.

They are antagonist of macrolids so we do not use them together.

17.20. Inhibitors of Membrane Function

Polymyxins

Polymyxin A,B,C,D,E

These medicines bind to the membrane and act like cationic detergent with positively charged free amino groups: interact with lipopolysaccharides and phospholipids and increased the permeability. Because the integrity of the cell membrane demaged, macromolecules and ions escape from the cells that results huge demage or death of the cell. (Some antifungal drugs use this targets too: plyenes)

All have bactericidal effects

Originally they derived from Bacillus polymyxa

They are effective against Gram negative rods (e.g. Pseudomonas aeruginosa)

17.21. Summary

Cell wall synthesis inhibitors	Beta-lactams
	Aminglycosides
	Glycopeptides
Enzime inhibitors	Gyrase inhibitors
	Folic acid antagonists
Ribosome-acting antibiotics: Inhibitors of protein synthesis	Tetracyclines
	Macrolide antibiotics
	Chloramphenicol
	Lincosamide
	(Aminglycosides)
Inhibitors of Membrane Function	Polymyxins

Antibiotic group	Effect mechanism 1	Effect mechanism 2
Beta-lactams	Cell wall synthesis inhibitors	bactericides
Aminglycosides	Cell wall synthesis inhibitors and Ribosome-acting drugs	Bactericides (cell wall inh.) and bacteri-ostatic (ribosome-acting)
Glycopeptides	Cell wall synthesis inhibitors	bactericides
Tetracyclines	Ribosome-acting drugs	bacteriostatic
Macrolide antibiotics	Ribosome-acting drugs	Bacteriostatic (their derivatives and high concentration could be bactericide)
Girase enzyme inhibitors	Enzyme inhibitors	bactericides
Folic acid antagonists	Cell wall synthesis inhibitors	Bacteristatics
Chloramphenicol	Ribosome-acting drugs	Bacteristatic (certain strains are bacteri-cid)
Lincosamides	Ribosome-acting drugs	Bacteristatic
Polymyxins	Inhibitors of Membrane Function	Bactericids

18. THE CELL CYCLE

18.1. DEFINITION AND PHASES OF THE CELL CYCLE

A new cell can be generated only by the division of an existing one. **Cell cycle is** the regulated chain of events which results in duplication of the material of a cell dividing this content into two identical cells (a parent cell produces two daughter cells). The primary event of this process is the faithful duplication of the genetic material (DNA) of the cell, and equal distribution into the two daughter cells.

In unicellular organisms this means the generation of two new cells. In **multicellular** creatures this event is more complicated: a vast number of controlled cell divisions produce a new organism. In the developed multicellular organisms, some cells will never divide again, while others divide only during certain conditions (extracellular signals, for example) and others retain their ability for constant division.

Cell cycle can be divided into phases. There are **five phases**, G1, S, G2 (the three together sometimes are called **interphase**), M and G0 (zero). Each phase has its own characteristics and distinctive role. In the G (gap) phases, mainly synthetic reactions (protein, membrane synthesis) occur. The two most dramatic phases are the S in which replication of the genetic material (DNA **s**ynthesis) takes place, and the M phase, which means **m**itosis, with nuclear and cytoplasmic divisions. The duration of these phases are alter depending on the organism and/or tissue, but in general, mammalian cells possess a 24 hour cell cycle, 10-12 hours of the S and 1 hour of the M phase.

G1 and G2 phases are **g**ap phases. G1 phase is between M and S, G2 is between S and M. In the gap phases, not only cell growth occurs, but also the controlling of the process: judgment of intra- and extracellular conditions before further transitions of phases (see later: **checkpoints** of cell cycle). G0 is the so called **resting state**, which can be temporary because of certain conditions, or permanent, if the cell is not dividing anymore. Cells may re-enter into G1 at changing environment or special signals (see later: mitogens). Obviously, cells carry out their basic functions during the G0 phase.

Although the length and events of the phases may differ from organism to organism, the basic regulatory mechanisms are very similar, and the protein components of this regulation are conserved. This is the primary reason why the control mechanisms of the cell cycle can easily be studied in many different **model organisms,** from yeasts cells as unicellular eukaryotes to the fruit fly and the frog. Human cell cultures can provide additional information about the process of the cell cycle. It is important to note, in many instances, tumors are developed because of mistakes in the control of cell cycle (see in: Biology of tumors).

18.2. REGULATION OF CELL CYCLE: CHECK POINTS

The cell is quite vulnerable during the cell cycle processes: the cell rounds up, loses contact with other cells, DNA is firmly packaged and cannot be used for transcription and notably, in the replication of the entire DNA, a large number of proteins must act in precise order.

The **cell cycle control system** reacts to the signs of the cycle itself: cycle will proceed only in case of completeness and the control system provides delay in the process if a repair has to be carried out. At certain points, (**checkpoints)** within the cycle often feature important transitions between phases. The characteristics of these transitions are the following:

- the cycle may or may not proceed (on/off), in an irreversible way
- the control system is reliable and effective
- the system is reacting to internal or external signals.

The four major regulatory transition points (checkpoints) of the cell cycle are:

- restriction point (or start point in yeast) in late G1
- S phase checkpoint
- G2/M transition
- metaphase/anaphase transition.

At these points, the cell cycle control system makes decisions, whether to initiate essential processes in the cycle or not – after considering the status of the cell and environmental signals. The **restriction point** is a decision-making of the beginning of a cell cycle – after this point, the cell has to carry out the entire cell cycle, or it may just halt temporarily if problems arise, and how and if this problem is going to be solved.

18.3. Regulation of cell cycle: the cyclin/Cdk complexes

Cyclins are proteins and are defined as the levels of which change throughout the cell cycle. The amount of **Cdks (cyclin dependent kinases**) is stable, but their activity changes depending on the amount of cyclins and other regulatory factors (see later: Regulation of Cdk activity). In humans, there are 4 classes of cyclins, and more than 20 types of Cdks. The special type of cyclins (D,E,A,B) appear at certain phase of the cell cycle: there are G1/S, S, M and G1 cyclins (see figure from lecture). Cyclins bind to special Cdks, and activate them by this binding. Cyclins can substitute one another to a certain degree. In humans there are at least 12 different cyclin/Cdk complexes.

When cyclin binds to a Cdk, it activates it, of which, Cdk due to its **kinase activity,** phosphorylates a specific set of proteins resulting in cell cycle specific events and activities. This can be posttranscriptional activity, but cell-cycle control may occur at the transcriptional level. Also a phosphorylated protein can change its activity, or may become the substrate or the target of another protein.

18.4. Regulation of Cdk activity

Usually cyclin binding to a Cdk is necessary, but not a sufficient step required for complete Cdk activation. This is true especially in the regulation of passing through the restriction point, in which the decision has to be controlled in multiple ways.

Other possibilities for Cdk activity regulation:

1. **Phosphorylation**
   - activating phosphorylations by Cdk-activation kinases (CAKs)
   - inhibitory phosphorylation (Wee1)
   - phosphatases (Cdc25 removes the inhibitory P of Wee1 – control of M-Cdk activity)
2. Binding of **inhibitory proteins**
   - Cdk inhibitor proteins (CKIs): they bind to Cdks or cyclin/Cdk complexes (G1/S and S-Cdk activity regulation)
3. Intracellular **localization**
   - Both cyclins and Cdks are synthetized in the cytosol, and have to move to the nucleus to interact. (At G2/M checkpoint phosphorylated cyclin B accumulates in the nucleus.)
4. **Protein degradation**
   - It happens due to the ubiquitin attachment method. The target can be cyclins, Cdks, inhibitor proteins or proteins with other functions. One example is the regulation of the metaphase/anaphase transition in mitosis. **Two major pathways** in which protein degradation is involved within cell cycle regulation are highlighted in the following: either the target protein of the E3 ubiquitin ligase is phosphorylated, or the ubiquitin ligase is phosphorylated and activated in this way.

18.5. External activators of the cell cycle

Mitogens and growth factors can initiate cell cycle by changes which lead to overcome the restriction point. These factors have an effect on a cell via signal transduction. Most of the signal transduction events happen in the G1 phase (See: Signal transduction for further details). The majority of these pathways involve the MAP kinase activity, and many of the transcriptional targets are components of the cell cycle regulatory system (for example, certain cyclins).

18.6. Relationship of cell cycle regulation and tumorgenesis

There are many possibilities whereby cell cycle regulation disturbances can lead to abnormal cell proliferation:

- The external signal transduction pathway may be uncontrollably activated
- Mutations: activation of proliferative signal, or loss of degradation of proteins (cyclins)
- Loss of regulatory mechanisms

See more in: Tumorgenesis.

Examples of the role of tumor suppressor genes (p53 and Rb protein) in cell cycle regulation are found within the lecture material.

Other examples of cell cycle regulatory mechanisms include the following: activation of prereplication complex, interaction of cell cycle regulation and DNA repair and regulation of the anaphase-promoting complex (APC).

19. MITOSIS AND MEIOSIS

Both mitosis and meiosis follow the replication. (see chapters of cell cycle and replication).

19.1. Mitosis

Mitosis is the process of eukaryotic cell division, the cell separates into two identical daughter cells.

Prophase During prophase the loose chromatin fibres condense into discrete chromosomes. The nucleolus disappears.

Centrosomes move to the opposite poles of the cell. Centrosome is a pair of centrioles, the Microtubule Organizing Centre (MTOC). Each centriole is a cylindrical structure formed by 9 microtubule triplet. The centrioles are duplicated during the interphase.

Prometaphase The nuclear membrane disintegrates (called open mitosis).

The mitotic spindle forms during prometaphase. The mitotic spindle consists of the centrosomes and the attaching microtubules. Its function is the movement of the chromosomes and the cytokinesis. One end of the microtubules always attach to the centrosome.

- Kinetochore microtubules: attach to the chromosomes and move the toward the equatorial plane of the dividing cell
- Interpolar microtubules: they connect the two centrosomes
- Astral microtubules: attach to the polar region of the cell

Chromosomes form kinetochores at the centromere regions Kinetochore is a protein complex what act as a platform for the microtubules. Kinetochores stabilize the attached microtubules.

Metaphase The formation of the mitotic spindle finishes in the metaphase. The chromosomes are aligned to the equatorial plane by the kinetochore microtubules.

Mitotic spindle checkpoint or spindle assembly checkpoint (SAC). SAC proteins check the correct kinetochore-kinetochore microtubule attachments. If a kinetochore is not attached to a microtubule, SAC will suspend the process of mitosis by keeping the MPF level high.

Anaphase There are two parallel processes in anaphase.

Anaphase A Sister chromatids are hold together by protein complexes, protease enzymes (separases) hydrolyse these cohesins. The separated sister chromatids are pulled apart by the shortening kinetochore microtubules. This shortening is due to the depolymerisation of the microtubules.

Anaphase B Centrosomes move to the polar regions of the cell driven by the elongation and slide of polar microtubules and shortening of astral microtubules.

Telophase Telophase is the ‚reversal' of metaphase and prophase. Poles pulled more apart. The nuclear envelope is reorganized. Nucleoli appear again. Condensed chromatin relaxes. Golgi complex and ER reassembles.

Cytokinesis starts at the same time as telophase. Animal cells are separated by a contractile actin-miosin ring formed at the equatorial plane. In plants a new cell wall is synthesized between the two sister cells.

Errors in mitosis

Errors are rare because it is a strictly regulated process and the consequences are often lethal for the cell.

Chromosomal nondisjunction: sister chromatids do not separate what result aneuploidy (abnormal chromosome numbers)

Cellular nondisjunction: failed cytokinesis result binucleated cells.

19.2. MEIOSIS

Meiosis is a special type of cell division as it decreases the number of chromosomes by half. Meiosis produce haploid gametes, in humans: oocytes (egg cells) and zoocytes (sperm cells) containing 23 chromosomes.

Prior meiosis, the genetic material of the cell is replicated in the S phase, the same ways as prior to the mitosis.

The process of meiosis can be divided into two distinct parts: Meiosis I (prophase I, metaphase I, anaphase I and telophase I) and Meiosis II I (prophase II, metaphase II, anaphase II, telophase II and cytokinesis). The following will discuss only the most important steps, for details see the lecture!

Prophase I: After the meiotic S-phase, in the prophase of the meiosis I the homologous chromosomes pair with each other. The paired homologous chromosomes exchange homologous DNA fragments, which is called homologous recombination. Genetic recombination (crossing over) involves nicks of the DNA strands, then exchanging of the homologous strands. Genetic recombination is an important mechanism of increasing genetic variation by producing new combination of alleles.

Metaphase I and anaphase I: The homologous chromosome pairs are lined up in the equatorial plane by the kinetochore microtubules. The chromosome pairs are separated in the anaphase, but the sister chromatids are left attached together. The assortment of the chromosomes between the two daughter cells is random (independent assortment), what also increases the offsprings' genetic variation.

Telophase I: Nuclear membrane is formed again. Chromosomes unfold into chromatin.

Meisosis II follows meiosis I **without DNA replication**.

Prophase II: Nuclear envelope and nucleolus disappears. Chromatin condense into chromosomes.

Metaphase II: Chromosomes (**not chromosome** pairs) are lined up in the equatorial plane by kinetochore microtubules.

Anaphase II: Cohesins are cleaved between the sister chromatids. These chromatids are then separated and moved towards the polar region of the cell.

Telophase II and cytokinesis: Nuclear membrane is formed again. Chromosomes unfold into chromatin. Sister cells separate from each other.

Errors in meiosis

The most important error is the chromosomal non disjunction. For details see the chromosome aberrations chapter.

19.3. Comparison of Mitosis and Meiosis

	Mitosis	Meiosis
Number of cell divisions	1	2
Number of sister cells	2	4 (1)
Chromosome number	2n -> 2n	2n -> 1n
Homologous recombination	no	yes
Sister cells identical	yes	no

20. SIGNAL TRANSDUCTION

20.1. INTRODUCTION

Cells which are capable of recognizing extracellular signal molecules on their surface by receptors or the signal molecules. Considering their size is small and hydrophobic, enter through the plasma membrane and act on intracellular receptors. This mechanism is the signal transduction with which the cells are able to react to the stimulus in an adequate means. They respond suitably to the extracellular signals. The process maintains the activity of the cells, and the cells can survive even within a changing environment.

20.2. ELEMENTS OF THE CELL SIGNALLING SYSTEMS

Types of signalling through extracellular messengers

Cells communicate with each other through extracellular messenger molecules. These extracellular messengers can travel a short distance and stimulate cells that are in close proximity to the origin of the message, or they can travel throughout the body, stimulating cells distanced from the source (see figures from lecture).

1. Autocrine signalling: the cell produces and releases the messenger molecules into the extracellular space. The messenger molecules bind back to the receptors on the same cell. Consequently, cells releasing the message will stimulate themselves.

2. Paracrine signalling: A group of cells produces the messenger molecules which travel only short distances through the extracellular space to cells in close proximity to the cells (usually in the same organ) generating the message. These messenger molecules are usually unstable, or they are degraded by enzymes, or they bind to the extracellular matrix.

3. Endocrine signalling: messenger molecules are secreted into the blood and transported toward the target organ. They travel long distances from the site of production to the site of action. Endocrine messengers are also called hormones.

4. Juxtacrine signalling: a cell produces a membrane bound hormone precursor (inactive molecule) interacting with a receptor on an adjacent cell. The messenger is activated on the receptor of the adjacent cell and transmits signal into the cytoplasm of the target cell.

5. Neurocrine signalling: It is similar to paracrine signalling but involves neurons. The signals can be secreted by neurons, as neurotransmitters. The neurotransmitters travel through the synaptic cleft and bind to their receptors on the postsynaptic membrane.

6. Intacrine signalling: an endocrine inactive precursor (Hi) is generated in one organ. When the messenger reaches the target organ it is transformed into an endocrine active compound (Ha) and acts on the target cells.

7. Cryptacrine signalling: It is similar to paracrine signalling. The signalling molecule is synthesized and secreted into a closed cellular environment. The messenger cannot diffuse out from the crypt and only acts on the cells found within this environment.

8. Pherocrine signalling: chemical messengers called pheromones are secreted into the environment targeting individuals of the same species. Characteristically, their role supports mating during reproduction.

9. Photocrine signalling: daylight affects structures within the eye and signals are transmitted to the brain influencing the daily rhythm of the body.

The major types of signalling pathways - signalling elements

A cell signalling pathway is always initiated with the release of a messenger molecule. The messenger molecules originating from the extracellular space are called **first messengers**. Cells possess cell surface **receptors** with which they are able to respond to the extracellular stimulus/messenger. The receptors can recognize and bind ligand molecules; they are said to be specific for their own ligands. The messenger molecule binds to the extracellular receptor and usually causes the **conformational change** of the receptor, thus the signal is relayed across the plasma membrane to the intracellular part of the receptor.

There are **two major pathways** by which the extracellular signal is transmitted into the cell, where it elicits the appropriate response (see figure from lecture).

1. The extracellular ligand, the first messenger binds to the receptor leading to the conformational change of the intracellular part of the receptor. The receptor itself **has no enzymatic activity**, thus the active receptor recruits a **signalling molecule** (G protein), which activates an **effector molecule**, usually an enzyme producing **second messenger** molecules. Second messengers activate different signalling proteins which can be enzymes e.g., kinases or transcription factors. The active enzymes triggers an **activation cascade** of target proteins that may lead to short term or long term changes in the activity of the cell (e.g. transcription, survival, protein synthesis, movement, metabolic changes).

2. The extracellular ligand, the first messenger binds to the receptor leading to the conformational change of the intracellular part of the receptor. The intracellular domain of the receptor is an **enzyme** usually a kinase that **phosphorylates** itself on specific amino acid residues, and recruits **signalling proteins** to the phosphorylated amino acids. The activated signalling proteins form a signalling cascade which ends in the activation of cytoplasmic enzymes or transcription factors. The target proteins can act in the cytoplasm (short term change, e.g., increased enzyme activity) or in the nucleus (long term change, e.g., changes in the transcriptional rate of distinct genes).

Depending on the type of cell and message, the response initiated by the target protein may involve a change in gene expression (transcription), an alteration of the activity of metabolic enzymes, a reconfiguration of the cytoskeleton, an increase or decrease in cell mobility, a change in ion permeability, activation of DNA synthesis, or even the death of the cell.

A signalling pathway consists of a series of proteins. Once the target protein is activated it alters the **conformation** of the next protein in the protein cascade. The conformational change is usually triggered by **phosphorylation**. Target proteins receive signals from the active receptor and alter the activity of the target cell. The overall process which consists of the extracellular messenger, receptor, signalling proteins (effector), second messengers (if they are produced), and target proteins, is called **signal transduction**.

The major types of signalling pathways –general signal termination possibilities

1. The first possibility to terminate the signal is the elimination of the extracellular messenger molecules by **extracellular enzymes** which destroy the first messengers.
2. **Dephosphorylation** of the receptor kinase inhibits the binding of the signalling proteins to the receptor.

3. **Internalization** of the receptor by the action of ubiquitin ligase. The enzyme puts ubiquitin molecules to the intracellular part of the receptor. **Ubiquitinilation** leads to receptor invagination into a clathrin coated vesicle. Then the coat is then released and the endosome is fused with a lysosome. The receptor protein sometimes is recycled from the endosome back to the plasma membrane while the ligand is degraded in the lysosome.
4. The phosphorylation cascade is reversed by the action of **phosphatases**. Whereas protein kinases typically work as a single subunit, many protein phosphatases contain a catalytic and a key regulatory subunit aiding towards determining substrate specificity (see figure from lecture).

Extracellular messengers

There are several extracellular molecules acting as extracellular messengers (first messengers) on the target cells either on cell surface receptors or intracellular (cytoplasmic/nuclear) receptors.

1. Small molecules such as **amino acids** and their derivatives acting as neurotransmitters or hormones (e.g. dopamine, adrenaline, noradrenaline, acetylcholine, thyroid hormones). Thyroid hormones have cytoplasmic receptors as amines can pass through the plasma membrane.
2. **Gases** such as NO or CO. Gases can diffuse through the plasma membrane so they act intracellularly. They can diffuse away very fast from the extracellular space they act only paracrine way.
3. **Steroids** which are synthesized from cholesterol (e.g., sex steroids, mineralocorticosteroids, glycocorticosteroids). Steroids also able to pass through the plasma membrane due to they are hydrophobic, they have nuclear receptors in the nuclear envelope.
4. **Polypeptides** and proteins such as secreted proteins, membrane proteins, components of the extracellular matrix.
5. **Eicosanoids** which are 20-carbon atom containing fatty acid derivatives (arachidonic acid derivatives). They have a role in inflammation, pain, blood pressure and blood clotting.

Receptors

Receptors are found to be cell surface or intracellular receptors. Receptors can bind the extracellular ligands and transmit the signals into the cell and the cell responds with changes of cellular activities.

1. **GPCRs (also called 7-pass receptors)**-G protein coupled receptors: the receptor is coupled with a specific GTPase protein which may activate the effector enzyme. The effector enzyme produces second messengers which transmit the signal toward downstream signalling proteins.

2. **Enzyme-coupled receptors:** These receptors are found on the extracellular surface of the cell, they have three major domains, an extracellular ligand binding domain, a transmembrane domain and an intracellular domain usually with enzymatic activity.

   a) **RTKs**-Receptor protein-tyrosine kinases (or tyrosine kinase receptors): the receptor recruits signalling proteins (docking and adaptor proteins) which are activated by the kinase domain of the receptor. Adaptor proteins recruit new signalling proteins transmitting the signal toward the target proteins by phosphorylation.
   b) **Tyrosine-kinase-associated receptors:** have no intrinsic enzyme activity but directly recruit cytoplasmic tyrosine kinases to relay the signal.
   c) **Receptor serine/threonine kinases** directly phosphorylate specific serines or threonines on themselves and on latent gene regulatory proteins with which they are associated.

d) **Histidine-kinase-associated receptors** activate a two-component signalling pathway in which the kinase phosphorylates itself on histidine and then immediately transfers the phosphoryl group to a second intracellular signalling protein.
e) **Receptor guanylyl cyclases** directly catalyze the production of cyclic GMP in the cytosol.
f) **Receptor-like tyrosine phosphatases** remove phosphate groups from tyrosines of specific intracellular signalling proteins. They are only "like" because their ligands have not been discovered.

3. **Ligand-gated channels**: They conduct ion flow through the plasma membrane (e.g. cGMP-gated cation channels) or in the endomembrane system (e.g. IP3-gated Ca2+ channels). They can change the membrane potential (depolarization or hyperpolarisation) or can change the activity of different cytoplasmic enzymes (CaM-kinases).

4. **Steroid hormone receptors** (ligand-regulated transcription factors): Steroid hormones are able to diffuse across the plasma membrane as they are apolar molecules. Their receptors are found on the nuclear envelope. The receptor-hormone complex is translocated into the nucleus and binds to specific hormone responsive elements (HRE) in the promoter region of particular genes. The receptor-hormone complex changes the transcriptional rate (increases or decreases) of specific genes which have a role in sexual differentiation, pregnancy, carbohydrate metabolism, and excretion of sodium and potassium ions.

5. **B-cells and T-cells receptors**: They act by unique mechanisms, as they have a role in the immune system and fight against the infections. They can recognize foreign antigens and couples with cytoplasmic protein tyrosine kinases.

Second messenger systems

Second messenger molecules are produced by receptor activated effector enzymes. The activation process is indirect because a G protein transmits the signal from the active receptor to the effector. The second messengers can be classified into three major groups:

1. **Hydrophilic molecules:**
   - water soluble
   - located in the cytosol
   - cAMP, cGMP, IP3, Ca^{2+}

2. **Hydrophobic molecules:**
   - water-insoluble
   - bound to the intracellular leaflet of the plasma membrane or
   - diffuse into the intermembrane space
   - DAG-diacylglycerol
   - PIs-Phosphatidylinositols

3. **Gases:**
   - diffuse across the plasma membrane
   - diffuse through cytosol
   - NO, CO, H2S

There are five second messenger systems in eukaryotic cells (see GPCRs for more details).

1. **cAMP system**:
   - Ligands are neurotransmitters (epinephrine, acetylcholine) and
   - hormones (ACTH, ANP, FSH, glucagon)
   - Transducer molecules: stimulatory (G_S) and inhibitory (Gi) G-
   - proteins
   - Effector enzyme: Adenylyl cyclase
   - Second messenger: cAMP
   - Secondary effector: protein kinase A

2. **cGMP system**:
   - Ligands are gases (NO), hormones (ANP) or light (photon)
   - Transducer molecules: Gt-transducin (only in retina)
   - Effector enzyme: Guanylyl cyclase
   - Second messenger: cGMP
   - Secondary effector: protein kinase G, cGMP-dependent cation
   - channels

3. **Tyrosine kinase system**:
   - Ligands are growth factors (PDGF, EGF, IGF) or hormones
   - (insulin)
   - Transducer molecule: kinase domain of the receptor tyrosine
   - kinase
   - Second messenger: protein phosphatases

4. **Phosphoinositol system**:
   - Ligands are neurotransmitters (acetylcholine, epinephrine)
   - and hormones (TRH, GHRH, Oxytocin)
   - Transducer molecule: G_q
   - Effector enzyme: Phospholipase C
   - Second messenger: IP3 (inositol-triphosphate) and DAG
   - (diacylglycerol)
   - Second effectors: Ca^{2+} and protein kinase C

5. **Arachidonic acid system**:
   - Ligand: histamine
   - Transducer: G protein
   - Effector: Phospholipase A
   - Second messenger: arachidonic acid
   - Secondary effector: lipoxygenase, cycloxygenase

20.3. Communication between cell signalling pathways

Cell signalling pathways can converge, diverge and crosstalk, they can regulate each other by common signalling proteins. In case of **convergence**, different cell surface receptors regulate the same effector molecule in the cytoplasm. Signals from one ligand can **diverge** to activate different effectors and thus activation of different signalling pathways lead to the activation of a huge number of signalling proteins and result in short term or long term changes in the cell's life. Signals can be passed back and forth between pathways as a result of **crosstalk.** Signal transducers from different pathways can regulate or influence the activity of the effector enzymes, the synthesis of intracellular messengers or the activation of signalling proteins.

An example of *convergence* is the communication between a GPCR and RTK on the cell's surface. A GPCR is activated by its ligand that leads to the activation of G_q-protein. G_q activates the effector enzyme called phospholipase C which generates IP3 and DAG second messengers. The IP3 molecules activate IP3 receptors on the smooth endoplasmic reticulum and increase the cytosolic Ca^{2+} concentration and DAG activates protein kinase C. Additionally, an RTK is activated by its ligand and undergoes a conformational change leading to the activation of the intracellular kinase domain. The receptor is responsible for the activation of three different effectors (divergence). One of them is phospholipase C that also generating the same second messengers as GPCR activated effector enzyme (see figure from lecture).

Another example for *convergence* is the activation of three different receptors (Integrin, GPCR, RTK) by three different ligands. The activated receptors produce a docking site for the same docking protein Grb2 which binds the same Sos/GEF adaptor protein responsible for the activation of Ras GTPase (see figure from lecture).

During *divergence* one signal is able to activate distinct cell signalling pathways. The IP3-kinase phosphorylates PIP2 molecules in the plasma membrane and forms PIP3 molecules which provide docking sites for proteins containing PH-domain (pleckstrin homology domain). PDK-1 (phosphatidyl dependent kinase-1) and PKB/Akt (protein kinase B) bind to the PIP3 molecules. PDK-1 phosphorylates PKB, and PKB autophosphorylates its kinase domain. The phosphorylation leads to the activation of PKB. PKB regulates different pathways in the cell e.g. protein synthesis, glycogen synthesis, glucose uptake.

An example involving cAMP illustrates the importance of *crosstalk* between signal transduction pathways (see figure from lecture). Epidermal growth factor receptor is activated by EGF. The kinase domain of the receptor undergoes a conformational change leading to autophosphorylation on the tyrosine residues. Phosphorylated tyrosines bind Grb2- Sos/GEF proteins. Sos activates Ras-MAP kinase cascade which phosphorylates the transcription factor CREB on the serine residue, activating the transcription factor and allowing it to bind to specific sites on the DNA. Cyclic AMP acts by activating the cAMP dependent kinase PKA, to block the transmission of signals from Ras to Raf, which inhibits the activation of the MAP kinase cascade. Additionally, PKA phosphorylates the identical transcription factor CREB. These two pathways intersect at an important signalling effector, the transcription factor CREB.

20.4. G PROTEIN-COUPLED RECEPTORS

The name of GPCRs originates from the interaction between the receptor and a heterotrimeric G protein. The receptor itself does not possess enzymatic activity. The function of the G protein is to interlink the active receptor and the effector enzyme. There are more than 700 GPCRs in humans, and in mice there are about 1000. All GPCRs have a similar structure. The basic structure of the GPCR consists of three domains: extracellular, transmembrane (TM) and intracellular domains. The extracellular part of the receptor is built up by three loops which together form the ligand binding site. The major type of GPCRs passes through the plasma membrane seven times with **seven transmembrane α-helices**. After the ligand binding the TM domains move closer to each other leading to conformational changes in the arrangement of the intracellular loops. This change is responsible for G protein binding.

Classification of GPCRs

Classification of GPCRs is based upon sequence homologies and evolutionary relationships. The GPCRs are divided into six classes:

Class A: Rhodopsin-related receptors (e.g. olfactory receptors, visual rhodopsin, prostaglandin receptor, amine receptors (adrenaline, serotonin, dopamine), melatonine receptor and tachykinins)

Class B: Secretin-related receptors (e.g. secretin, glucagon, growth hormone releasing hormone (GHRH), vasoactive intestinal peptide (VIP))

Class C: Glutamate-related receptors

Class D: Fungal pheromone receptors

Class E: cAMP receptors

Class F: Archea (archeabacterial) opsin receptors

The last three classes are not found in mammalian cells, but there are two additional receptor families in mammals which cannot be classified into the above mentioned classes:

1. Adhesion receptor family
2. Taste receptor family

GPCR interacting proteins (GIPs)

These proteins can **modulate the receptor function** by changing its structure or its ligand binding site:

- alter the ligand affinity by binding to its ligand binding site
- inhibit the dimerization or oligomerization of the receptor
- control of receptor localization, including transfer to or removal from the plasma membrane
- promote close association with other signal proteins inhibiting the action of the receptor on the G protein

G proteins

The two major classes of G proteins are the heterotrimeric G proteins and monomeric G proteins (see later at *tyrosine kinase receptors*). The heterotrimeric G proteins are obligate mediators of GPCRs, as they are responsible for transmitting the signal from the receptor to the effector enzyme.

Properties of heterotrimeric G proteins

(see figure from lecture)

- consist of three subunits: α, β and γ
- all three subunits are bound to the plasma membrane
- α subunit is the catalytic subunit with GTPase activity (it binds and hydrolyzes GTP to GDP+Pi)
- βγ subunits are the regulatory subunits
- when regulatory subunits bind to the α subunit holding a GDP molecule in its GTP binding pocket, the G protein is inactive
- the nucleotide binding site in Gα consists of loops extending from the edge of a 6-stranded β-sheet
- three switch domains and five G regions have been identified, altering position when GTP substitutes for GDP on Gα
- **G-1 region** encompasses the glycine-rich pyrophosphate binding loop (P-loop) and contains a conserved lysine followed by a serine
- **G-2 region** is located in switch I. It has a central threonine which has a function in GTP hydrolysis-driven conformational rearrangements

- **G-3 region** is part of the switch II region: the main chain amide of the conserved glycine co-ordinates the γ-phosphate
- **G-4 region** is partially responsible for binding the guanine base
- **G-5 region** aids in guanine base recognition and binding
- GTP hydrolysis occurs by **nucleophilic attack** of a water molecule on the terminal phosphate of GTP (see figure from lecture)
- **Switch domain II** of Gα includes a conserved glutamine residue aiding in positioning the attacking water molecule adjacent to GTP at the active site
- **Switch III** stabilizes the extended helix of switch II in the GTP-bound conformation, which is aligned for effector recognition. In the GTP-bound conformation, switch III residues form interactions with cognate residues in switch II.
- β subunit maintains the **stability** of the G protein
- β subunit has a special **β-propeller** structure which is formed from multiple repeats of a sequence called the WD-repeat (see figure from lecture)
- β-propeller is a common structural motif for protein domains involved in protein-protein interaction

Action of heterotrimeric G proteins

(see figure from lecture)

1. GPCR binds its ligand and becomes active
2. G protein binds to the intracellular part of the active receptor
3. α subunit is released from the regulatory subunits
4. α subunit replaces the GDP to GTP, and becomes active
5. the active catalytic subunit binds to the effector enzyme and activates it
6. Effector enzyme produces second messengers until the α subunit hydrolyzes GTP
7. α subunit cleaves GTP and comes off the effector
8. Effector enzyme is inactivated
9. α subunit binds back to the regulatory subunits
10. heterotrimeric G protein is ready for further activation

Heterotrimeric G protein isoforms

Heterotrimeric G proteins feature five predominant families: Gs, Gi, Gq, G_{12}, $G_{\beta\gamma}$. Further classification of G proteins is based on the activity of α subunit: Gs, Golf, Gi, Go, Gt, Gq, G_{12} are the most important types (see table below).

Family	Subunit isoform	Tissue distribution	Effector enzymes and channels
Gs	α_s	Ubiquitous	Adenylyl cyclase ↑
	α_{olf}	Brain/olfactory neurons	Adenylyl cyclase ↑
Gi	α_i	Wide distribution incl. neurons	Adenylyl cyclase ↓ Ca^{2+} channels ↓ K^+ channels ↑
	α_o	Neurons, brain, heart	Adenylyl cyclase ↓ Ca^{2+} channels ↓ K^+ channels ↑
	α_z	Adrenal chromaffin cells, neurons, platelets	Ca^{2+} channels ↓ K^+ channels ↑
	α_t	Retina	cGMP phosphodiester-ase ↑
Gq	α_q	Ubiquitous	Phospholipase Cβ ↑ p63-RhoGEF ↑ Tyrosine kinase ↑ K^+ channels ↑
	α_{11}	Ubiquitous	p63-RhoGEF ↑
	α_{14}	Stromal, epithelial cells	Tyrosine kinase ↑
	α_{15}	Myeloid cells	K^+ channels ↑
	α_{16}	Myeloid cells	
G_{12}	α_{12}	Ubiquitous	Phospholipase D ↑
	α_{13}	Ubiquitous	Phospholipase Cε ↑
Gβ/γ			Phospholipase Cβ2 Adenylyl cyclase ↑

G protein regulators

(see figures from lecture)

There are various proteins capable of **binding to the G proteins**. Some of these proteins inhibit or activate the G proteins.

1. RGS/GAP: Regulators of G protein signalling or GTPase activating proteins are negative regulators of G proteins. When a G protein binds a GDP it is inactive and when GDP is re-placed by GTP it is active. GAP proteins can increase the rate of GTP hydrolysis, in which they accelerate the inactivation of G proteins. GAP proteins are specific to the α-subunit. There are about 25 RGS proteins encoded in the human genome, each of which interacts

with a particular set of G proteins. Frequently a (+) charged **arginine** residue of a GAP inserts into the active site of catalytic subunit and helps to stabilize the transition state by interacting with (-) charged O atoms of the terminal phosphate of GTP during hydrolysis.

2. GEF: Guanine nucleotide exchange factors promote GDP/GTP exchange. Therefore GEF proteins are positive regulators of G proteins. Due to the action of GEFs G proteins can replace GDP to a new GTP faster and G proteins remain active. GEF proteins usually act indirectly; it interlinks the G protein with the active receptor. GEF proteins have a crucial role in the activation of monomeric G proteins (see later).

3. AGS: Activators of G protein signalling can act on heterotrimeric G proteins independently of GPCRs. Some of these proteins possess GEF activity.

20.5. Cell signalling through GPCRs

1. Gs- The ligand (e.g. catecholamines, glucagon) binds to the receptor which leads to conformational changes on the intracellular part of the receptor. Gs-protein binds to the receptor and the α-subunit is released from the regulatory subunits. α subunit replaces GDP to GTP and binds to the effector enzyme, **adenylyl cyclase**. Adenylyl cyclase is a large multipass transmembrane protein with its catalytic domain on the cytosolic side of the plasma membrane. There are eight isoforms in mammals, most of which are regulated by both G proteins and Ca2+. The enzyme synthesizes **cAMP** second messengers from ATP. cAMP molecules activate the second effector, **protein kinase A** (see later) which phosphorylates target enzymes in the cytosol and transcription factors in the nucleus.

2. Gi- This type of G proteins inhibits the enzymatic activity of adenylyl cyclase therefore decreases the cytosolic cAMP concentration. More over the regulatory subunits of Gi directly control **cation channels** in the plasma membranes. For example acetylcholine released by the vagus nerve reduces both the rate and strength of heart muscle cell contraction. This effect is mediated by muscarinic acetylcholine receptors (GPCRs) activating the Gi protein. The βγ subunits bind to potassium channels in the heart muscle cell plasma membrane and aid towards effectively opening them. The opening of these potassium channels makes it harder to depolarize the cell and thereby contributes to the inhibitory effect of acetylcholine on the heart.

3. Gq- Receptors that act through Gq activate the inositol phospholipid signalling pathway (see figures from lecture). Gq activates **phospholipase C-β** which generates two types of second messengers: **inositol 1,4,5-trisphosphate** (IP3) and **diacylglycerol** (DAG). Both are generated by the cleavage of PIP2 (phosphatidyl inositol 4,5-bisphosphate). IP3 is a water-soluble molecule that acts as a small intracellular mediator. It leaves the plasma membrane and diffuses through the cytosol. When it reaches the endoplasmic reticulum (ER), it binds to and opens **IP3-gated calcium channels** (IP3 receptors). Ca2+ stored in the ER rushes through the channels, raising the concentration of Ca2+ in the cytosol. The increase in cytosolic calcium level propagates the signal by influencing the activity of Ca2+-sensitive intracellular proteins (see later). DAG remains in the plasma membrane and activates the calcium dependent **protein kinase C** (PKC) molecules. PKCs translocate from the cytosol to the cytoplasmic face of the plasma membrane. There they are activated by the combination of calcium, diacylglycerol, and the membrane phospholipid phosphatidylserine. PKC phosphorylates target proteins that vary depending on the cell type. There are three major classes of PKCs: **conventional PKCs** are activated by Ca2+ and diacylglycerol, **novel PKCs** are activated by dicylglycerol and the others are called **atypical PKCs** have other activators than Ca2+ or DAG. PKCs are bound to anchoring proteins or scaffold proteins in different compartments of the cell. Therefore, these proteins determine the sub-

strates which will be phosphorylated. Targets of PKCs are glucagon receptors, EGF receptor, insulin receptor, β-adrenergic receptor etc. DAG molecules can be further cleaved to **arachidonic acid**, which can either act as a signalling molecule or be used in the synthesis of other small lipid signal molecules called **eicosanoids** (e.g. prostaglandins).

4. Gt- Transducin is special G protein which is only found in the retina, and it has a crucial role in vision. The GPCR of transducin is **rhodopsin**. Rhodopsin consists of two parts: cis-retinal and opsin. Rhodpsin's ligand is not a molecule, it is best defined as light (photons). As rhodopsin absorbs a photon, cis-retinal is transformed to trans-retinal leading to conformational change of opsin and the cleavage of the interlinking ionic bonds between the TM-domains. The rearrangement of TM-domains changes the conformation of the intracellular loops now positioned for binding transducin. The α subunit of transducin activates **cyclic GMP phosphodiesterase**. The phosphodiesterase then hydrolyzes cyclic GMP, resulting in decreased cyclic GMP levels in the cytosol. This drop in cyclic GMP concentration decreases the amount of cyclic GMP bound to **cGMP-gated cation channels**, allowing them to close. The closing of cation channels leads to the hyperpolarization of the plasma membrane. The visual transduction response is the *fastest* G protein-mediated response known in vertebrates. The phototransduction apparatus is in the outer segment of the rod (responsible for noncolor vision), which contains a stack of discs, each formed by a closed sac of membranes featuring many embedded photosensitive rhodopsin molecules. The plasma membrane surrounding the outer segment contains many cyclic-GMP-gated cation channels (see figure from lecture). These channels are directly controlled by the cytosolic cGMP concentration depending on the activity of transducin and cyclic phosphodiesterase. **Hyperpolarization** is due to the light-induced activation of rhodopsin in the disc membrane decreasing the cyclic GMP concentration and closes the cation channels in the surrounding plasma membrane. The visual response is regulated by negative feedback. **Rhodopsin kinase** *(RK)* phosphorylates the cytosolic tail of activated rhodopsin on multiple serines of which this mechanism partially inhibits the ability of the rhodopsin to activate transducin. **Arrestin** then binds to the phosphorylated rhodopsin and initiates its internalization. **RGS protein** is also capable to bind to activated transducin and increases the rate of GTP hydrolysis, therefore transducin returns to its inactive state. Lastly, the cation channels which close in response to light are not only permeable to sodium but **calcium** too. When the cation channels close, the normal influx of calcium is inhibited, causing the calcium concentration in the cytosol to decrease. This stimulates guanylyl cyclase to resynthesize the cyclic GMP molecules. In the activation of **guanylyl cyclase** a specific calcium-sensitive protein (not calmodulin!) has a special role. This protein is inactive when calcium is bound to it and active when it is calcium-free. It therefore stimulates the enzyme when calcium levels fall in the cytosol of photoreceptors.

5. Golf- specific GPCRs called **olfactory receptors** are able to recognize odours. These receptors are found only on the surface of the modified cilia extending from each cell (see figure from lecture). The receptors act through an olfactory specific G protein, Golf. Golf activates adenylyl cyclase and increases the cAMP concentration. Cyclic AMP opens **cyclic-AMP-gated cation channels**, allowing an influx of sodium ions depolarizing the olfactory neuron generating a nerve impulse (action potential). The nerve impulse then travels along the axon to the brain.

cAMP and protein kinase A

Cyclic AMP (cAMP) acts as a small intracellular messenger molecule in prokaryotic and eukaryotic cells. Its normal concentration in the cytoplasm is about 10–7 M, but an extracellular signal on GPCRs can increase this concentration more than twentyfold in seconds (see figure from lecture). Cyclic AMP is synthesized from **ATP** by a plasma-membrane-bound enzyme **adenylyl cyclase** by cyclization re-

action and it is rapidly and continuously destroyed by **cyclic AMP phosphodiesterases** that hydrolyze cAMP to adenosine 5'-monophosphate (5-AMP) (see figure from lecture). The cyclization reaction removes two phosphate groups as pyrophosphate. A **pyrophosphatase** enzyme hydrolyzes pyrophosphate to phosphate. Cyclic AMP is unstable and it is hydrolyzed by specific phosphodiesterases. Different cell types respond differently to an increase in cyclic AMP concentration, and the response of one cell type depends on the type of the activated GPCR (see table from lecture). The actual cAMP concentration of the cell depends on the number of active Gs and Gi proteins, though Gs activates while Gi inhibits adenylyl cyclase (see figure from lecture).

Cyclic AMP acts by the activation of **cAMP-dependent protein kinase A** (PKA) enzyme. This kinase phosphorylates specific serines or threonines on target proteins, including intracellular signalling proteins and effector proteins. The target proteins differ from one cell type to another, which is primarily why cAMP acts differently within various cells. In the inactive state, PKA consists of a complex of two **catalytic subunits** and two **regulatory subunits** (RI and RII). Inactive PKA holoenzymes differ in subcellular localization. RI-containing PKA is mainly cytosolic whereas a large portion of RII-containing PKA is attached to membranes and organelles. The binding of cyclic AMP to the regulatory subunits alters their conformation, causing them to dissociate from the complex (see figure from lecture). The released catalytic subunits are thereby activated to phosphorylate specific target proteins. The RII regulatory subunits of PKA (also called A-kinase) are important for localizing the kinase inside the cell. **Cyclic AMP-dependent protein kinase anchoring proteins (AKAPs**) are anchored to membranes, microtubules and other subcellular structures and interlink the regulatory subunits to a component of the cytoskeleton or a membrane of an organelle (see figure from lecture). Some AKAPs also bind other signalling proteins, forming a complex functioning as a signalling module. RII-type PKA is more effective because of the close proximity of substrates and AKAP also binds **phosphodiesterases** which are capable of rapid elimination of cAMP and abort the signalling process. In this regard, the response to cAMP develops a strong, brief and local pulse of PKA activity. RII-PKA has a lower affinity for cAMP than RI-PKA. RI-PKA activates at cAMP levels that are 2-8-fold lower than those required for RII-PKA. Distinctly, that RII contains autophosphorylatable pseudosubstrate linker sequence between its dimerisation domain and cAMP module. Catalytic subunits of PKA phosphorylate the RII domains and RII is not able to rebind PKA unless it is first dephosphorylated by phosphatases.

The effect of somatostatin on transcription - an example for PKA activity

(see figure from lecture)

The binding of an extracellular signal molecule to its GPCR activates adenylyl cyclase via Gs and increases cyclic AMP concentration in the cytosol. The rise in cyclic AMP concentration activates PKA, and the released catalytic subunits of PKA can then enter the nucleus, where they phosphorylate the gene regulatory protein CREB. Once phosphorylated, ***CREB*** *recruits the coactivator CBP, which stimulates gene transcription. The regulatory region of the* ***somatostatin*** *gene contains a short recognition DNA sequence, called the* ***cyclic AMP response element*** *(CRE), which is also found in the regulatory region of many other genes activated by cyclic AMP. CREB recognizes this responsive element on the promoter region and binds to it. Thus, CREB can transform a short cyclic AMP signal into a* ***long-term change*** *in a cell, a process where by, for example within in the brain, plays a vital role in the formation of learning and memory.*

Ca- signalling

Ca signal is generated by **IP3** second messenger molecules produced by the cleavage of PIP2 molecules in the plasma membrane. IP3 molecules are water soluble they travel along the cytoplasm and bind to the **IP3 receptors** on the membrane of endoplasmic reticulum. IP3 receptors are **calcium channels**, when they are open Ca2+ ions rush out from the ER into the cytosol generating a Ca signal.

Ca2+ ions act as a signal because their concentration in the cytosol is normally very low (~10^{-7} M), whereas it's concentration in the extracellular fluid (~10^{-3} M) and in the lumen of the ER is high. Thus, there is a **large gradient** driving Ca2+ into the cytosol across both the plasma membrane and the ER membrane. The elevating Ca2+ concentration in the cytosol activates **Ca2+-responsive proteins** in the cell.

In addition to IP3 receptors Ca2+ may enter the cytosol through **ryanodine receptors** (identified due to their sensitivity to the plant alkaloid ryanodine). Ryanodine receptors are activated by Ca2+ binding and amplify the Ca2+ signal. Ca2+ also activates IP3 receptors but only in the presence of IP3, so Ca signalling possesses a **positive feedback** loop aiding towards the increase in cytosolic Ca2+ concentration. A very high concentration of Ca2+ can inactivate the receptors/channels thus stops the further Ca2+ release from the ER.

There are several mechanisms maintaining the low concentration of Ca2+ in the cytosol (see figures from lecture):

1. **Ca2+-pump** in the plasma membrane utilises the energy of ATP hydrolysis to pump Ca2+ out of the cytosol.
2. Muscle and nerve cells, which have an additional Ca2+ transport protein, a **Na+-driven Ca2+ exchanger** in their plasma membrane that couples the efflux of Ca2+ to the influx of Na+.
3. A **Ca2+ pump** in the ER membrane: this Ca2+-pump enables the ER to take up large amounts of Ca2+ from the cytosol against a steep concentration gradient, even when Ca2+ levels in the cytosol are low.
4. **Low-affinity, high-capacity Ca2+ pump** in the inner mitochondrial membrane has an important role in limiting the Ca2+ signal and in terminating it.
5. **Calcium binding proteins** binding Ca2+ and transmit the signal to target proteins.

If we monitor the cytosolic Ca2+ concentrations of individual cells it reveals the initial Ca2+ signal appears small and localized to one or more regions of the cell. These Ca2+ peaks reflect the local opening of individual or small groups of Ca2+ channels in the ER. If the extracellular signal is sufficiently strong and continuous, this localized Ca2+ signal can spread through the cytosol. A Ca2+ spike is often followed by a series of further spikes. Ca2+ **oscillations** can persist for as long as receptors are activated at the cell surface. The oscillations depend on the binding of Ca2+ on both the **IP3 receptors and ryanodine receptors**. The released Ca2+ initially stimulates more Ca2+ release from both receptors, when the Ca2+ concentration increases to the necessary level, the Ca2+ inhibits further release (see figure from lecture). The frequency of the Ca2+ oscillations reflects the strength of the extracellular stimulus, and this frequency can be translated into a frequency-dependent cell response (see figure from lecture).

Cytosolic Ca2+-binding proteins transmit Ca-signals to target proteins. The most important is **calmodulin**, which is found in all eukaryotic cells and its amount can reach as much as 1% of the total protein mass. Calmodulin acts as a multipurpose intracellular Ca2+ receptor. It consists of a highly conserved, single polypeptide chain with four high-affinity Ca2+-binding sites (see figure from lecture). The molecule has a dumbbell shape, with two globular heads, which can bind to many target proteins. The **globular heads** are connected by a long α-helix, this helps to adopt different conformations depending on the target protein. Each globular head has two Ca2+-binding domains. When it binds Ca2+, it undergoes a conformational change. Two or more Ca2+ ions are needed for the conformational change, the protein responds to the increasing Ca2+ concentration. Ca2+/calmodulin **doesn't have enzymatic activity** itself but instead acts by binding to and activating other proteins. Active Ca2+/calmodulin binds to its target protein, the calmodulin further changes its conformation. Calmodulin regulates enzymes and membrane transport proteins (e.g. plasma membrane Ca2+-pump).

Though calmodulin doesn't have enzymatic activity its role in Ca signal is indirect. Ca2+/Calmodulin activates serine/threonine protein kinases called **Ca2+/calmodulin-dependent kinases** (CaM-kinases). Some CaM-kinases phosphorylate gene regulatory proteins (e.g. CREB), and in this way activate or inhibit the transcription of specific genes.

CaM-kinase II is found in most animal cells but is especially enriched in the nervous system. It reaches 2% of the total protein concentration in some regions of the brain, and it is highly concentrated in synapses. CaM-kinase II has two important functions:

1. It is a molecular memory device. It switches to an active state when exposed to Ca2+/calmodulin and then remains active even after the Ca2+ signal has ceased.
2. The enzyme can use its intrinsic memory mechanism to act as a frequency decoder of Ca2+ oscillations. This property is especially important at a nerve cell synapse, where changes in intracellular Ca2+ levels in a postsynaptic cell as a result of neural activity can lead to long-term changes in the subsequent effectiveness of the specific synapse.

CaM-kinase II is a large protein complex of 12 subunits. In the absence of Ca2+/calmodulin, the enzyme is inactive as the result of an interaction between an inhibitory domain and the catalytic domain. CaM-kinase **phosphorylates** itself (autophosphorylation) after Ca/calmodulin binding. The autophosphorylation occurs on the **inhibitory domain** of the enzyme, thus the inhibitory domain cannot rebind to the catalytic subunit even if Ca/calmodulin has released from the kinase. The enzyme remains active even in the absence of Ca2+ (**Ca2+ independent form**) prolonging the duration of the kinase activity. The enzyme maintains this activity until **serine/threonine protein phosphatases** dephosphorylates and shuts the kinase off (see figure from lecture). CaM-kinase II has a role in some types of **memory** and learning in the vertebrate nervous system.

CaM-kinase II also acts as a **frequency decoder** of Ca2+ oscillations. During low frequencies of Ca2+ spikes, the enzyme becomes inactive after each spike, as enzyme's activity is not long enough for the enzyme to remain active in the absence of Ca2+ spike. At higher frequencies, autophosphorylation maintains the activity of CaM-kinase between the Ca2+ spikes.

Signal termination of GPCRs

To prevent overstimulation, receptors have to be blocked from continuing to activate G- proteins. To maintain sensitivity to extracellular stimuli, the receptor, the G protein, and the effector must all be returned to their inactive state.

1. **Desensitization** (see figure from lecture): is the process that blocks active receptors from turning on additional G- proteins. The cytoplasmic domain of the activated GPCR is phosphorylated by a specific type of kinase, called G protein-coupled receptor kinase (GRK). GRK is serine-threonine protein kinases recognizing activated GPCRs. After phosphorylation of the GPCR a special protein, called **arrestin** binds to GPCRs and compete for binding with heterotrimeric G proteins. Arrestin binding prevents the further activation of additional G- proteins. Arrestin molecules are also capable of binding to clathrin molecules promoting the uptake of phosphorylated GPCRs into the cell by **endocytosis** (internalization of the receptor into clathrin coated pits). The receptor may be recycled from the endosome into the plasma membrane after dephosphorylation or it is degraded in the lysosome. β-arrestin may also bind a **cytosolic phosphodiesterase**, bringing this enzyme close to where cAMP is being produced, contributing to signal turn off.

2. Gα **hydrolyzes** GTP to GDP + Pi (GTPase). Gα subunit possesses a weak GTPase activity, which allows them to slowly hydrolyze the bound GTP and inactivate themselves. Termination of the response is accelerated by regulators of G protein signalling (RGSs). The

interaction with an RGS protein increases the rate of GTP hydrolysis by the catalytic subunit. The presence of GDP on Gα causes it to rebind to the inhibitory βγ complex.

3. **Protein phosphatases** catalyze removal of phosphate groups (by hydrolysis) from the active proteins that were phosphorylated by protein kinase A. Inactivation of the signalling proteins leads to the termination of signal transduction.

4. **Phosphodiesterases** are activated even without the arrestin molecules and cleave cAMP molecules to 5'AMP molecules stopping further activation of protein kinase A enzymes.

Amplification of extracellular signals through second messengers

The first stage in signal amplification is generated from the activating interaction of receptor with G protein and G protein couplings with the effector enzyme, adenylyl cyclase. Based on the relative enzymatic activity of Gs, during 15 seconds the activated adenylyl cyclase molecule produces 225 cAMP molecules before Gs hydrolyzes GTP. There is a 10,000-fold signal amplification in case of the activation of a single G protein-coupled receptor.

In the visual transduction, a single activated rhodopsin molecule catalyzes the activation of hundreds of molecules of transducin at a rate of about 1000 transducin molecules per second. Each activated transducin molecule activates a molecule of cGMP phosphodiesterase, each of which hydrolyzes about 4000 molecules of cGMP per second. This results in a significant decrease in the concentration of cGMP closing hundreds of cation channels in the plasma membrane. Distinctly, a single rod cell responds to a single photon in a manner which facilitates a change in the signals from the photoreceptor to the brain (see figure from lecture). In the inositol phospholipid signalling pathway, a nanomolar change in the concentration of an extracellular signal can induce micromolar changes in the concentration of small intracellular mediators. A single extracellular signal molecule can alter many thousands of protein molecules within the target cell due to the amplification cascade.

20.6. Tyrosine-kinase receptors

Enzyme-coupled receptors consist of three domains: extracellular ligand binding domain, one transmembrane domain and an intracellular domain which possesses enzymatic activity. Receptor tyrosine kinases are more diverse than GPCRs. Their dominant mode of signal transmission is the formation of **oligomeric signal transduction particles** formed on the activated RTK dimer. RTKs are epidermal growth factor receptor (EGFR), platelet-derived growth factor receptor (PDGFR), fibroblast growth factor receptors (FGFRs), hepatocyte growth factor receptor (HGFR), insulin receptor, insulin-like growth factor-1 receptor (IGF1R), vascular endothelial growth factor receptor (VEGFR), receptors for macrophage-colony-stimulating factor (MCSF), ephrin receptors, and the neurotrophins, including nerve growth factor (NGF) also bind to RTKs.

There are about 60 genes encoding human RTKs. These receptors can be classified into more than 16 structural subfamilies (see figures from lecture). The functional structural domains of most of RTKs are the cysteine-rich, immunoglobulinlike, fibronectin-type III-like domains, EGF domain, leucin-rich domain, cadherin, discoidin, kringle, tyrosine kinase and SAM domains. Some RTKs contain a large insert in the tyrosine kinase domain (e.g. PDGFR).

In the signalling process of RTKs, the extracellular ligand binds to the receptor binding site resulting in the **dimerization** or **oligomerization** of the receptor and a conformational change in the structure of the intracellular domains. Thus, the intracellular tyrosine kinase domain is activated and **phosphorylates** selected tyrosine side chains, both on the receptor proteins themselves and on intracellular signalling proteins that subsequently bind to the phosphorylated tyrosines on the receptors. Due to dimerization the two kinase domains move closer to each other, they can become activated and cross-

phosphorylate each other on multiple tyrosines, a process referred to as **transautophosphorylation** (see figure from lecture). Phosphorylation of tyrosines within the kinase domain increases the kinase activity of the enzyme and phosphorylation of tyrosines outside the kinase domain creates high-affinity docking sites for the binding of specific intracellular signalling proteins. Different signalling proteins bind to specific phosphorylated tyrosines on the activated receptors by phosphotyrosine-binding domains (SH2, PTB domains, see later). Transautophosphorylation triggers the transient assembly of an intracellular signalling complex, which can then relay the signal to various destinations in the cell (see figure from lecture). Because different RTKs bind different combinations of these signalling proteins, they activate different responses.

Highly conserved phosphotyrosine-binding domains:

1. **SH2** domain: Src homology region, SH2-containing proteins form non-covalent complexes with tyrosine-phosphorylated proteins. Some SH2 proteins have no enzymatic activity so called adaptor proteins. SH2 domains can also bind the phosphate groups of phosphatidyl-inositol lipid groups.

2. **PTB** domain: phosphotyrosine binding module, have conserved arginines to coordinate and charge-neutralize the negatively charged phosphotyrosine residues. For example IRS-1, insulin receptor substrate is a docking protein which has no enzymatic activity.

Some signalling proteins are composed almost entirely of SH2 and SH3 domains and function only as **adaptors** to couple tyrosine-phosphorylated proteins to other signalling proteins without their own SH2 domains (see figure from lecture). For example, the insulin receptor binds its ligand then the activated receptor phosphorylates itself on tyrosines residues. **Phosphotyrosines** recruit a docking protein called insulin receptor substrate-1 (IRS1) via a **PTB domain**. Another special structural domain, the **PH domain** (pleckstrin homology domain) of IRS1 also binds to specific phosphoinositides on the inner surface of the plasma membrane. The activated receptor phosphorylates IRS1 also on tyrosines, and one of these phosphotyrosines recruits the adaptor protein (Grb2) via an **SH2 domain**. Grb2 uses one of its two SH3 domains to bind to a proline-rich region of the monomeric GTPase-activating protein called Sos (see figure from lecture), which also binds to phosphoinositides in the plasma membrane by its PH domain. Grb2 is able to bind **scaffold proteins** through **SH3 domains**. Scaffold proteins transmit the signal from the activated insulin receptor toward other signalling proteins in the cytosol.

Monomeric G proteins

Monomeric G proteins are composed of only one subunit, **α-subunit** which is homologous with the α-subunit of the heterotrimeric G proteins. Monomeric G proteins are **GTPases**, they are active when binding GTP and inactive when binding GDP. Ras-like superfamily of GTPases are involved in signal transduction. The monomeric small G proteins can be grouped into five subfamilies:

1. Ras group (Ras, Rap, Ral): regulators of mitogenic response, gene expression and chemotaxis
2. Rab group (more than 60 members): regulators of vesicular trafficking
3. Rho group (Rho, Rac, Cdc42): regulators of the actin cytoskeleton and gene transcription
4. Ran group: regulators of microtubules and transport of Nucleocytoplasmic proteins
5. Arf group (Arf1-6, Arl1-7, Sar): regulators of vesicular trafficking and endocytosis

Ras-GTPases

Humans possess three major Ras proteins: H-, K-, and N-Ras. These GTPases have different functions, but work in the same way, and they have the same effect on target proteins in the signalling pathways. Ras GTPases contain one or more covalently attached **lipid groups** assisting in anchoring

the protein to the cytoplasmic face of the membrane. The proteins relay signals from the membrane to other parts of the cell. Ras is often required for stimulating **cell proliferation** or **differentiation**.

Ras proteins are active signal transducers, but only when GTP occupies their guanine nucleotide binding site. Ras GTPases have only a very low enzymatic activity. In order to function, Ras needs input from accessory proteins, **GTPase activating proteins** (GAPs) and **guanine nucleotide exchange factors** (GEFs).

1. **GAPs**: interaction with a GAP is essential for signal termination. It **increases** the rate of GTP hydrolysis therefore inactivates Ras.
2. **GEFs**: they enable Ras-GDP to dock with receptor-recruited adaptors (Grb2/Sos) and they are essential in Ras **activation**, forcing exchange of GDP to GTP.
3. **GDIs**: guanine nucleotide dissociation inhibitor protein acts on Rho and Rab proteins. GDIs allow the G protein to detach from the membrane to cytosol and to move to other compartments. Moreover, they **inhibit** the exchange/release of GDP to a new GTP molecule and as a result, inactivate the G proteins.

Monomeric G proteins have a catalytic site and the binding surfaces for regulators and effectors. The **G-domain** (G-boxes or G-regions) is highly conserved and contains the key residues for GTP hydrolysis. The N-terminal domain is a tightly folded structure consisting of a **six-stranded β-sheet** surrounded by five **α-helices**. The β-sheet and α-helices are connected by ten polypeptide loops. Five of these loops form the catalytic centre and this is the place where the switch mechanism is triggered.

G-1 region encompasses the glycine-rich pyrophosphate-binding loop (**P-loop**) and contains a conserved lysine and an adjacent conserved serine. The glycine-rich turn allows the loop to wrap around the phosphates and the amide nitrogens provide a positively charged field to accommodate the negatively charged phosphates. The conserved lysine helps to **stabilize the transition state** during hydrolysis of GTP by neutralizing charges at the γ-phosphate.

G-2 region is the **switch I domain**, it is located in the loop between the first α- helix and second β-strand. The G-2 box contains an essential threonine that coordinates the **Mg^{2+}** which has a role in GTP hydrolysis. Switch I (like switch II) is a **mobile element** involving the binding of regulators and effectors.

G-3 region is found at the N-terminal part of the second α-helix and forms the part of **switch II domain** in all G proteins (both monomeric and heterotrimeric). There is an invariant aspartate coordinating the magnesium ion via a water molecule, and the conserved glycine coordinates the γ-phosphate. The whole switch II domain runs from the C-terminal part of the third β-strand through fourth loop to the second α-helix.

G-4 region is located in the loop between the fifth β-strand and fourth α-helix and it is responsible for binding the guanidine base.

G-5 region runs between the sixth β-sheet and the fifth α-helix and helps in guanidine base recognition and binding.

The switch mechanism is a hydrolysis-driven conformational change in Ras. The conformational change is triggered by the contacts made with the γ-phosphate of GTP, the essential Magnesium ion, and the key residues in the switch regions (see figure from lecture).

The mitogen-activated protein kinase (MAPK) cascade

Both the tyrosine phosphorylations and the activation of Ras triggered by activated RTKs are usually short-lived. To stimulate cells to proliferate or differentiate, these short-term signalling events must be converted into long-term ones often relaying the signal downstream to the nucleus to alter the pattern

of gene expression. The **mitogen-activated protein kinase module** (MAP kinase module) is capable to receive, maintain and relay the signal from the active receptor toward the nucleus of the cell. The three components of this system together form a functional signalling module (see figure from lecture).

A characteristics of growth factor signalling pathways is the downstream activation of kinases, referred to **MAP kinase kinase kinase** (MAPKKK) e.g., Raf, activating a **MAP kinase kinase** (MAPKK) by phosphorylation. Their targets are the **MAP kinases**, which are cytoplasmic serine/threonine protein kinases translocated from the nucleus when they are activated and activate **transcription factors** in the cytosol.

In the mammalian Ras-MAP-kinase signalling pathway, these three kinases are known by shorter names: **Raf** (MAPKKK), **Mek** (=MAPKK) and **Erk** (MAPK). Erk MAP kinase enters the nucleus and phosphorylates one or more components of a gene regulatory complex. This activates the transcription of a set of immediate early genes. These genes usually encode gene regulatory proteins, which turn on other genes (see figure from lecture). Raf is the beginner of the MAPK cascade and is the prime effector for Ras signals. Raf is a serine/threonine kinase present as three human isoforms with partially overlapping functions:

1. ubiquitous Raf-1 or C-Raf
2. A-Raf
3. B-Raf: high levels in neuronal tissue, lower levels elsewhere

All three forms share conserved homology regions, known as CR1, CR2 and CR3. The C-terminal part of the proteins contains a **serine/threonine kinase domain** which corresponds to CR3. The N-terminal half of the protein comprises a complex regulatory region containing CR1. CR1 contains the **Ras binding domain** and an adjacent cysteine-rich domain, which is a **zinc finger motif**. In resting state, Raf is cytosolic and only moves to the membrane in response to the appearance of Ras-GTP. The serine phosphorylation sites in the CR1 and CR2 regions are inhibitory sites while the phosphorylation sites in the CR3 region are activators of Raf. The CR1 and CR2 serines are phosphorylated by PKA or Erk kinase.

The classic idea of the action of a MAPK cascade of soluble enzymes implies an interaction with each other through diffusion or collision. MAPK pathway enzymes are present as two- or three-membered cassettes. In this way, the Ras–MAP-kinase signalling pathway conveys signals from the cell surface to the nucleus and alters the pattern of gene expression.

Many factors serve to influence the **duration** of the signalling response, including **positive and negative feedback loops** in which MAP kinases also participate. These influencing kinases may combine to provide responses that are either graded or switch-like and either brief or long lasting.

There are five families of MAPK enzymes; all of which require dual phosphorylation on special tyrosine residues:

1. Extracellular signal-regulated kinase (Erk1 and Erk2)
2. Stress-activated protein kinase (SAPK), also known as c-Jun N-terminal kinase (Jnk1, Jnk2 and Jnk3)
3. p38 kinase homologous (p38α, p38β, p38γ and p38δ)
4. Erk3/4
5. Erk5

Scaffold

Three-component MAP kinase signalling modules operate in all eukaryotic cells. The different modules mediate different responses in the same cell. Some of these MAP kinase modules use one or more of the same kinases and yet manage to activate different effector proteins and hence different respons-

es. One possibility to avoid cross-talk between the different parallel signalling pathways is to use scaffold proteins (see figure from lecture). For example, when scaffold mechanism occurs in yeasts. Budding yeast have at least six three-component MAP kinase cascades. A mating response and the response to high osmolarity in these cells work with two different scaffold protein avoiding cross-talk. The mating response is triggered when a mating factor secreted by yeast of opposite mating type binds to a GPCR. This activates a G protein, from which the regulatory subunits activate the MAP kinase kinase kinase (kinase A). Kinase A relays the response onward. Another MAP kinase (kinase C) activates several proteins by phosphorylation mediating the mating response, in which the yeast cell stops dividing and prepares for fusion. The three different kinases in this module are bound to scaffold protein 1. Once a yeast cell is exposed to a high-osmolarity environment, it is induced to synthesize glycerol to increase its internal osmolarity. This response is mediated by an osmolarity-sensing receptor which activates a different MAP kinase cascade with a second scaffold protein. Even when both pathways use the same MAP kinase kinase kinase, no cross-talk occurs between them, due to the MAPKKK molecules binding to different scaffold proteins. Scaffolds bind all or some of the kinases in each MAP kinase module to form a complex and thereby help to ensure response specificity.

Mammalian cells also use **scaffold strategy** to inhibit cross-talk between different MAP kinase modules. There are at least 5 parallel MAP kinase modules which work in a mammalian cell. These modules are built up with 12 MAP kinases, 7 MAP kinase kinases, and 7 MAP kinase kinase kinases. Two of these modules identified JNK and p38 cascades are activated by different kinds of cell stresses, such as UV irradiation, heat shock, and osmotic stress, as well as by inflammatory cytokines (see figure from lecture). The scaffold strategy reduces the opportunities for amplification and spreading of the signal to different parts of the cell.

PI3-kinase pathway

Phosphoinositide 3-kinase (PI3-kinase) binds to the plasma membrane and can be activated by receptor tyrosine kinases or GPCRs. This kinase phosphorylates inositol phospholipids rather than proteins. It plays a central role in promoting cell survival and growth.

Phosphatidylinositol (PI) is membrane lipid which can undergo reversible phosphorylation at multiple sites on its inositol head group to generate several types of phosphorylated lipids called phosphoinositides (see figure from lecture). When activated, PI is phosphorylated by PI kinase then **PI(4)P** is phosphorylated by **PIP kinase** generating **PI(4,5)P2**. **PI3-kinase** catalyzes phosphorylation at the third position of the inositol ring to generate **PI(1,4,5)P3** molecules (see figure from lecture). PI(4,5)P2 is cleaved by PLCβ, activated by GPCR and Gq protein, or PLCγ activated by RTKs to generate soluble IP3 and membrane-bound diacylglycerol. PI(3,4,5)P3 is not cleaved, it remains in the plasma membrane until specific phosphoinositide phosphatases (e.g PTEN phosphatases) dephosphorylate it. The PI(3,4,5)P3 molecules in the plasma membrane serve as **docking sites** for various intracellular signalling proteins. Class I PI3-kinases are heterotrimers, composed of a catalytic subunit and different regulatory subunits. The regulatory subunit can act as an adaptor protein that binds to two phosphotyrosines on activated RTKs through its two SH2 domains. PI3-kinases which are activated by GPCRs usually bind to the regulatory subunits of the G protein.

PI(3,4,5)P3 docking sites are occupied by target proteins that have specific interaction domain like **PH** (pleckstrin homology) domain. PH domains occur in about 200 human proteins, including the Ras-GEF Sos.

Akt/Protein kinase B

The most important PH-domain-containing protein is the **serine/threonine protein kinase** Akt. This signalling pathway is the major pathway activated by the hormone **insulin**. It also plays a key part in promoting the survival and growth of many cell types.

The extracellular ligands bind to specific RTKs, which activate PI3-kinase to produce PI(3,4,5)P3. The PIP3 recruits two protein kinases to the plasma membrane via their PH domains: Akt or PKB and phosphoinositide-dependent protein kinase 1 (PDK1). Binding of Akt to PIP3 leads to the activation of the enzyme, more over, the bound PDK1 is also ale to activate Akt by the phosphorytaion of a serine residue on Akt. Akt phosphorylates various target proteins at the plasma membrane, as well as in the cytosol and nucleus (see figure from lecture). The effect of Akt on most of the known targets is to inactivate them by phosphorylation.

The effect of PI3-kinase–Akt pathway on cells depends on a large, serine/threonine protein kinase called **TOR** (target of rapamycin). In mammalian cells, this protein is called mTOR. TOR exists in cells in two functionally distinct multiprotein complexes. In mammalian cells, **mTOR complex 1** contains the protein **raptor**. This complex is sensitive to rapamycin, and it stimulates cell to growth by stimulating protein synthesis. The **mTOR complex 2** contains another protein called **rictor** and is insensitive to rapamycin. It helps to activate Akt by phosphorylation, and it regulates the actin cytoskeleton via Rho family GTPases (see figure from lecture). The mTor complex 1 is activated by growth factors via the PI-3-kinase–Akt pathway. Akt activates mTOR in complex 1 by phosphorylating and inhibiting, a GAP called Tsc2. Tsc2 acts on a monomeric Ras-related GTPase called **Rheb** and maintains its inactive state. Rheb in its active form activates mTOR. Akt activates mTOR and promotes cell growth.

Rho GTPases

Rho family monomeric GTPases regulate the **actin** and **microtubule** cytoskeletons. They control cell shape, polarity, motility, and adhesion. Additionally, they also regulate cell cycle progression, gene transcription, and membrane transport. The three best-characterized members are **Rho** itself, **Rac**, and **Cdc42**. Likewise as Ras-MAP kinase cascade, **GEFs activate** and **GAPs inactivate** the Rho family GTPases. Some of the GEFs and GAPs are specific for one particular member of Rho family, whereas others are less specific. Inactive Rho family GTPases are often bound to **guanine nucleotide dissociation inhibitors** (GDIs) in the cytosol, which inhibit the GTPases interacting with their Rho-GEFs at the plasma membrane.

Ephrin receptor tyrosine kinase is found on the surface of motor neurons aiding in the guidance of the migrating apex of the axon (called growth cone) to its muscle target (see figure from lecture). The cell surface **ephrin** protein binds to and activates the Eph receptor, causing the growth cones to collapse. The response depends on a Rho-GEF known as **ephexin**, which is associated with the cytosolic tail of the Eph receptor. Once ephrin activates the Eph receptor, the receptor activates a **cytoplasmic tyrosine kinase** phosphorylating **ephexin** on a tyrosine residue, enhancing the ability of ephexin to **activate** the Rho protein **RhoA**. The activated RhoA binds a GTP in its catalytic site and regulates effector proteins causing the growth cone to collapse by stimulating the myosin-dependent contraction of the actin cytoskeleton. When ephrin is not bound to the Eph receptor, ephexin activates three different Rho family members (Cdc42, Rac, and RhoA), promoting the forward advance of the growth cone.

20.7. CYTOKINE RECEPTORS AND JAK-STAT SIGNALLING

The cytokine receptors can bind local mediators like **cytokines** and **hormones,** such as prolactin or growth hormone. These receptors do not possess enzymatic activity and therefore, are often associated with **cytoplasmic tyrosine kinases,** known as **Janus kinases** (JAKs). Janus kinases phosphorylate and activate gene regulatory proteins called **signal transducers and activators of transcription** (STATs). STAT proteins are latent gene regulatory proteins and are located in the cytosol. They only migrate into the nucleus and regulating gene transcription following activation.

Cytokine receptors are dimers or trimers and are predominantly associated with one or two of the four known JAKs (JAK1, JAK2, JAK3, and Tyk2). Cytokine binding alters the intracellular part of the recep-

tor and the conformational change brings two JAKs into close proximity enabling them to transphosphorylate each other. The JAKs then phosphorylate tyrosines on the cytokine receptors. These phosphotyrosines are the docking sites for STATs (see figure from lecture). There are at least six STAT proteins in mammals. Each has an **SH2** domain through which STAT proteins can bind to the **phosphotyrosine docking sites** on an activated cytokine receptor. The JAKs phosphorylate the STAT on tyrosines, causing the STAT to dissociate from the receptor. The released STATs bind to each other through SH2 domains and form heterodimers. The **STAT dimers** are translocated to the nucleus, where, in combination with other gene regulatory proteins, it binds to a specific DNA response element in the promoter region of different genes and stimulates their transcription. Different cytokines and receptors can activate different combination of JAK-STAT proteins and therefore mediate different cellular response (see table from lecture).

Regulation of JAK-STAT signalling happens by negative feedback mechanisms:

1. The STAT dimers activate genes encoding inhibitory proteins in the assistance of shutting off the response.
   a) the inhibitory proteins can bind to and inactivate phosphorylated JAKs and their associated phosphorylated receptors.
   b) other inhibitory proteins bind to phosphorylated STAT dimers and prevent them from binding to their DNA targets.

2. Inactivation of the activated JAKs and STATs by dephosphorylation of their phosphotyrosines. This process is performed by protein tyrosine phosphatases. Some of these phosphatases have dual-specificity, they also able to dephosphorylate serine and threonine residues on target proteins. Therefore these phosphatases are responsible for making tyrosine phosphorylation is short-lived.

20.8. TGFB/BMP-SMAD SIGNALLING

The transforming growth factor-β (TGFβ) superfamily consists of the **TGFβ/activin** family and the larger **bone morphogenetic protein** (BMP) family. They act either as hormones or, as local mediators to regulate e.g., pattern formation, proliferation, specification and differentiation, extracellular matrix production, and cell death. They are involved in tissue repair and in immune regulation. Their receptors are enzyme-coupled receptors known as singlepass transmembrane proteins featuring a **serine/threonine kinase domain** on the cytosolic side of the plasma membrane. The two classes of these receptors, **type-I and II**, are homodimers. TGFβ and BMP receptors are formed by different combinations of type-I and type-II dimers. The formation of the dimers brings the kinase domains close proximity so type-I phosphorylates and activates type-II receptor. The receptor dimers form an active **tetrameric receptor** complex.

The activated type-I receptor directly binds and phosphorylates a latent gene regulatory protein of the **Smad** family (Sma in C. elegans and Mad in Drosophila). TGFβ/activin receptors phosphorylate Smad2 or Smad3, while BMP receptors phosphorylate Smad1, Smad5, or Smad8. The activated Smads are called **R-Smads**. Once they have been phosphorylated, they dissociate from the receptor. The active Smad protein binds to Smad4 (**co-Smad**), which can form a complex with any of the five R-Smads. The Smad complex then translocates into the nucleus, and regulates the transcription of specific target genes (see figure from lecture).

The signalling pathway is regulated by negative feedback mechanisms:

1. Inactivation of the receptor by ubiquitylation and further internalization and degradation.

2. Secreted inhibitory proteins bind directly to the signal molecules and prevent them from activating their receptors on target cells. For example noggin and chordin inhibit BMPs, and follistatin inhibits activins.

3. Inhibitory Smads, Smad6 or Smad7 bind to the activated receptor and inhibit its signalling ability
   a) they compete with R-Smads for binding sites on the receptor
   b) they recruit a ubiquitin ligase, known as Smurf, which ubiquitylates the receptor, leading to receptor internalization and degradation
   c) they recruit a protein phosphatase that dephosphorylates and inactivates the receptor
   d) the inhibitory Smads bind to the co-Smad, Smad4, and inhibit it, either by preventing its binding to R-Smads or by promoting its ubiquitylation and degradation

21. APOPTOSIS

The development and maintenance of multicellular organisms is dependent not only on the rate of cell division but also on the destruction of specific cells. Within the body of a mature healthy human being, cells die at almost the same rate as they proliferate. Cells die due to damage or infection and therefore serve to protect the health and well-being of the organism. Notably, the demise of these cells is not considered random, it occurs by a programmed sequence of molecular events, in which the cells destroy themselves from within and will be eaten by the surrounding cells, leaving no trace. „Apoptosis" is a Greek word best defined and refers to the "fall of the leaves".

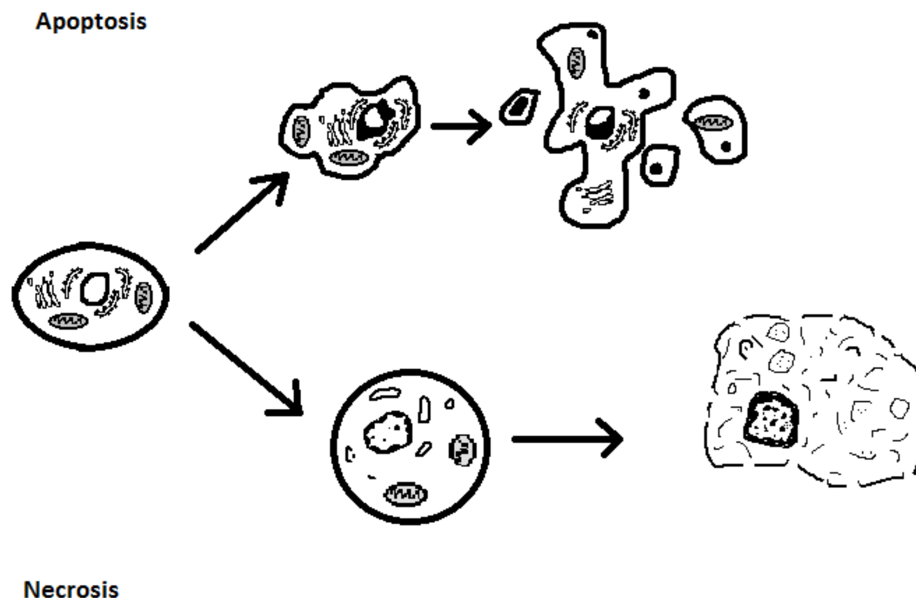
Apoptosis is often used as a synonym of programmed cell death, but the two words do not mean exactly the same; apoptosis is only one way how cells can die in a controlled way. Apoptosis was the first described programmed cell death process, however, we now know there are more cell death mechanisms belonging to this group.

Apoptosis was discovered in *Caenorhabditis elegans*, when researchers detected it during the development of the hermaphrodite worm. Notably, 1090 cells are initially formed and it is always the same 131 cells which die throughout the mature development (an adult hermaphrodite worm contains 959 cells). Following the discovery of this process, researchers correctly assumed, the death mechanism itself is not a random event.

The term "apoptosis" was used first in 1972 by John F. Kerr and his colleagues.

To compare the apoptotic process with random cell death mechanisms, it is important to understand the concept of necrosis. Necrosis is a means of cell death which is caused by a harmful physical or chemical effect. The morphological changes are typical, the cells swell and the chromatin condensed within the nucleus and dispersed in clumps. The membrane is considerably damaged, and as a result, the contents of the cell leak into the extracellular space surrounding the cell. This creates inflammation whereby macrophages are summoned to the area and phagocytize the damaged cells.

Morphological changes of apoptosis differ from the above described processes in a variety of steps. Internal and external factors can initiate this process (see below). First, the cell loses water and shrinks in size. The nuclear chromatin condenses under the nuclear membrane and DNA endonucleases cleaves DNA into small pieces. Enzymes destroy the cytoskeleton, it collapses and the external cell surface ripples (blebs) and so-called apoptotic bodies detach from the cell. The neighboring cells phagocytize these apoptotic bodies. In this way, macromolecules or cell debris, does not exit into the environment, and as a result, does not induce any immune response. One of the earliest signs of initiation of apoptosis is that phosphatidylserine rearranges to the external membrane from the internal membranes, which then will induce phagocytosis by other cells.



21.1. PROCESS OF APOPTOSIS

Initiation

Apoptosis is provided to kill the unnecessary cells. Due to apoptosis, cells live only as they are considered useful. Cells often die because something changed in them, (e.g. the DNA is damaged) or cells can die because they are no longer considered useful and the organism revokes the survival factors (e.g. growth factors and hormones). These survival factors prevent the cells against their own self-destructive system. Without these survival factors, cells die due to apoptosis.

We classify the initiative signals into two groups:

1. External stimuli:
   a) Death receptors (e.g. tumor necrosis factors α (TNFα) receptor)
   b) Revocation of survival factors (e.g.: growth factors and hormones)

2. Internal stimuli:
   a) DNA damage (radiation, chemical modifier substances)
   b) Poisons and medicines intracellularly

Intrinsic pathway

Mitochondria derived pathway

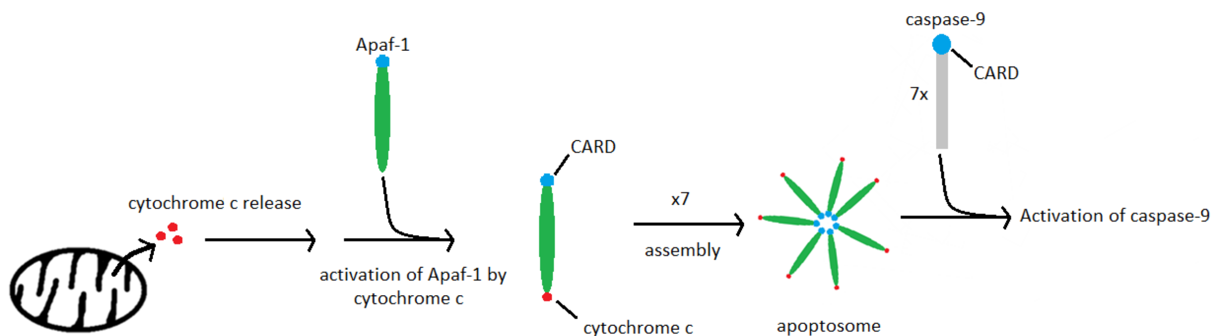
There are several molecules inside the mitochondria which can initiate apoptosis if and when they are released from the mitochondria into the cytoplasm, e.g. cytochrome-c, SMAC. These substances often enter the cytoplasm, if the mitochondrial membrane becomes more permeable or if specific membrane channels are open. These molecular events can occur due to the effect of an internal stimuli or cell damage.

Cytochrome-c

Cytochrome-c may be released from the mitochondria impact of an intracellular apoptotic stimuli. Cytochrome-c can activate the caspase cascade system with an adapter protein Apaf-1 and this also activates the effector mechanisms of the caspases, and as a result, the cell will die due to apoptosis.

Apaf-1 protein partially unfolds as cytochrome-c activates it. Seven activated Apaf-1 proteins form a ring complex called **apoptosome**. This apoptosome can bind caspase-9 molecules with the so-called caspase recruitment domain (CARD) (every Apaf-1 molecule bind a caspase-9 molecule). This process activates the capsase-9 molecules, and these molecules will induce the caspase cascade.

Specific proteins often influence the permeability of the mitochondria membrane to the cytochrome-c (e.g.: Bcl-2 decrease the permeability).



SMAC

SMAC is a group of proteins; called second mitochondria-derived activators of caspases. If SMAC proteins exit the mitochondria into the cytoplasm, they bind to apoptosis-inhibitor proteins (IAP) and serve to obstruct their functionality.

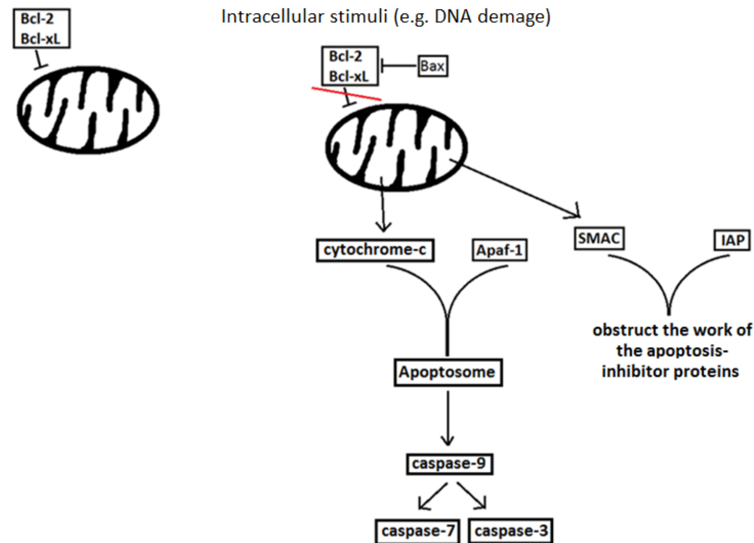
Bcl-2

Bcl-2 is a group of apoptosis inhibitor proteins, of which, some serve to prevent the cells from the initiation of apoptosis and yet, others may even stop and halt the process (most important Bcl-2 family members: Bcl-2 and Bcl-x_L proteins).

The Bcl-2 and Bcl-x_L proteins are located in the outer membrane of the mitochondria and act on specific channels: they obstruct the efflux of cytochrome-c from the mitochondria.

Once the cells suffer damage, proapoptotic BAX proteins serve to obstruct the Bcl-2 and Bcl-x_L proteins: so they cannot prevent the release of molecules from the mitochondria. Cytochrome-c exit the mitochondria and create apoptosome with the Apaf-1 proteins. Apoptosomes activate the caspase cascade: first the caspase-9, this activate the caspase-7 and 3. The caspase-7 and 3 will be the executioner caspases.

Summary



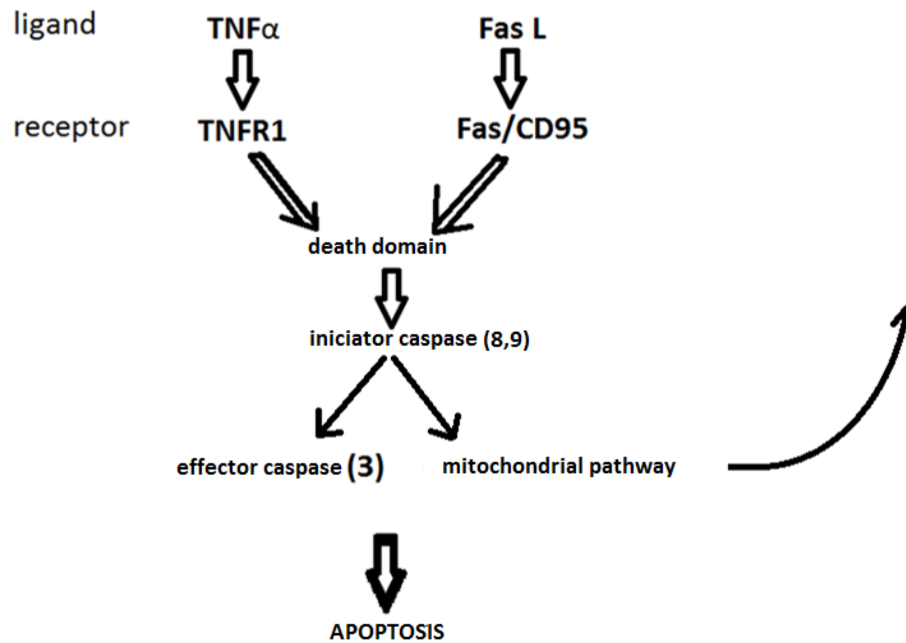
Extrinsic pathway: Death receptors

Apoptosis can be initiated through the activation of death receptors (Fas, TNFαR, DR3, DR4, and DR5). Death receptors usually oligomerize as the adequate ligand binds, and begins to recruit specific adaptor proteins and activates the caspase cascade. This is how T cells kill the infected cells (Fas-FasL).

Binding of FasL (=Fas Ligand) induces Fas trimerisation, TNFα induces TNFR1 trimerisation, which recruits initiator caspase-8, or 9 via adaptor proteins, also known as death domains. Caspase-8 or 9 then activate two processes:

1. It will induce cytochrome-c release from the mitochondria (initiates the intrinsic pathway too)
2. Activate caspase-3, which activates the caspase cascade

Recommended video: https://www.youtube.com/watch?v=9KTDz-ZisZ0



p53

P53 receives its name from its molecular weight (53 kDa).

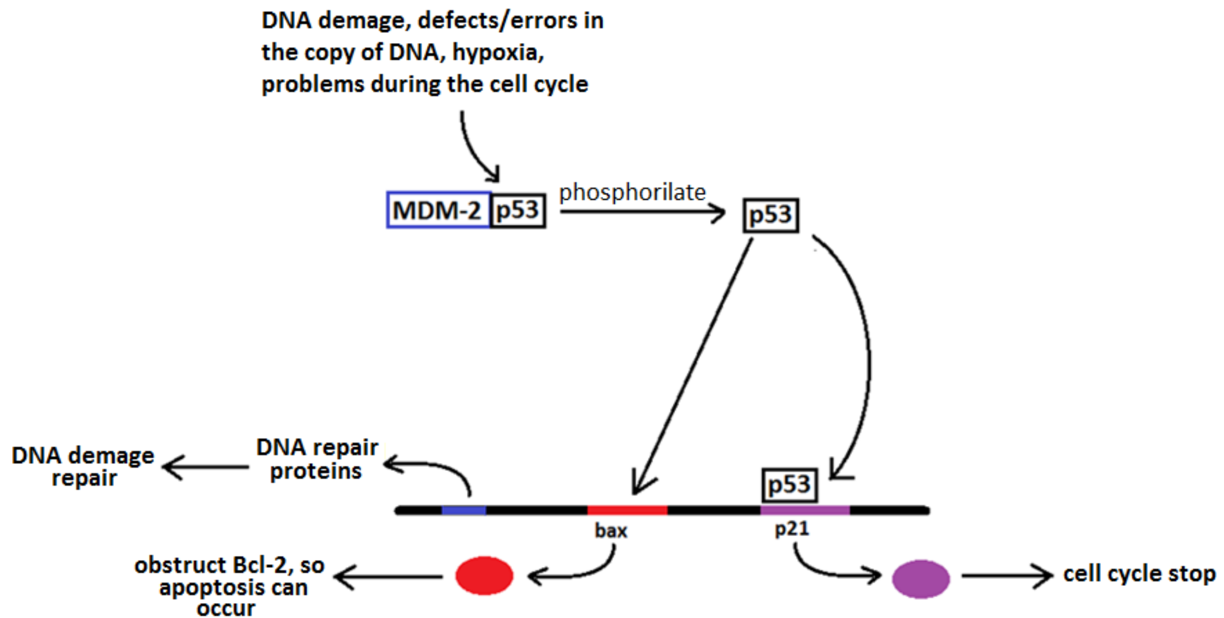
P53 protein is a check-point at the G1/S phase in the cell division. If the DNA is damaged, or there are errors in replication, the p53 phosphorylate is detached from the MDM-2 protein which can activate the transcription of other genes, due to its activity as it acts like a transcription factor.

P53 can initiate the transcription of p21 protein: p21 will obstruct the cyclins and cyclin dependent kinases, which are important for the continuation of the cell division. As a result, the cell division will stop and the cell will not transmit the improper and damaged DNA.

P53 can initiate the repair of the damaged DNA if it activates the transcription of repair proteins, but if the damage is too intense, the process will shift to the order of apoptosis.

Start of the apoptosis: the p53 binds to the promoter of the BAX protein and helps transcribe the protein. BAX protein will obstruct Bcl-2 proteins (see intrinsic pathway).

The mutation and dysfunction of p53 often causes unlimited proliferation and tumor formation (in more than half of human tumors, p53 is not active or is missing).



So, p53 stops the cell cycle and may activate additional pathways: DNA damage can be repaired, or apoptosis may start.

Execution phase

In this phase, effector molecules (enzymes) create the morphological changes of the apoptosis, for example, effector caspases cleave the cellular proteins, endonucleases cleave the DNA leading to the destruction of the nucleus, fragmentation of DNA and condensation of chromatin. The cell begins to shrink, membrane blebb and apoptotic bodies become detached. Participating enzymes:

- endonucleases
- transglutaminases
- caspases

Endonucleases

Endonucleases are Ca^{2+} -dependent enzymes, and distinctively, enhanced Ca^{2+} levels can activate them. Their most important function is the fragmentation of DNA in two steps: first they cut DNA into 50-300 kbp long sections, then into 180 bp long sections. If we can detect these small sections (e.g. with gel electrophoresis, in association with Tunel reactions that mark the free ends of DNA), then we can recognize apoptosis.

Transglutaminases

Transglutaminases make crosslinks between proteins, creating protein aggregates in this way. These crosslinks are irreversible. These enzymes are Ca^{2+}-dependent, too

(they are important in other processes too, e.g. by the coagulation, they link the fibrin fibers to one another and they are important by the cornification of keratinocytes.)

Caspases

Caspases are proteases that have a cysteine within their structure (in the active site) and cleave their target proteins at specific aspartic acids; this is because they are called caspases (c from cysteine and asp from aspartic acid).

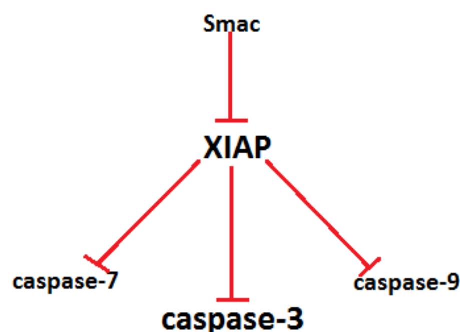
Caspases have two families, one of them has a role in the apoptosis (Ced-3), the other (ICE) has a role related to the inflammatory processes. Cleaving caspases have aspartate in their structure, too, they can cleave each other and activate each other in this way, in an amplifying proteolytic cascade. This caspase-activation process is the so-called caspase cascade.

Initiator caspases begin the process (caspase- 2, 8, 9, 10). At the end of the line there are the effector caspases=executioner caspases (caspase-3, 6, 7), which cleave the proteins. The effectors are activated by the initiators. The activation of effector caspases is universal. Initiator caspases became active specifically (the apoptotic pathways may differ from each other, but the final effector caspase is generally a caspase-3).

Inhibitor of apoptosis: IAP

They inhibit enzymes which take part in apoptotic pathways (endonucleases, transaminases and caspases) primarily. There can be IAPs in viruses, too, which are derived from higher organisms: notably, during infections, cells are not able to kill the infected cells.

XIAP: the most effective known inhibitor group, which inhibits the initiators and the effector caspases too.



Apoptosis vs. cell division

Apoptosis is normally not a harmful process, it has an important role in the normal ontogeny and in other positive process (see lecture slides). 10^{11} cells die in an adult human everyday, which is considered normal (unnecessary cells, autoreactive T cells, etc.,). Apoptosis plays a role in the maintenance of the normal cell number. The proliferation of the cells without brakes, or, without the normal cell , can initiate abnormal processes. Cell proliferation and apoptosis should operate in balance and are effectively controlled in the human body. Because over- and underactivity of apoptosis can cause serious diseases (e.g. autoimmune diseases and tumor formation), it is important to discover effective and strategically used apoptosis inducers and inhibitors for therapeutic purposes.

22. MOLECULAR BIOLOGY OF TUMORS

22.1. INTRODUCTION TO TUMORS

Cancer results from alterations in the DNA of a somatic cell. The mutation leads to altered gene expression. Cancer cells are characterized with **uncontrolled proliferation**, leading to malignant tumor formation. As long as the tumor remains localized, the cancer can be potentially stopped and cured by the surgical removal of the tumor. However, malignant tumors often *metastasize* (establish secondary tumors), and these cases are difficult to treat (sugery alone, is not enough as spawn cells circulate within the body and are capable of tumor generation in nearly every organ).

Cancer is in the focus of research, with a variety of studies whose strategic aim is towards the development of finding an effective cure and also to understand the basic functionality of the disease.

22.2. CLASSIFICATION OF TUMORS

Tumors are classified based on the origin of the tissue:

- Carcinoma: the most frequent type, derived from endo- or ectodermal tissue (lung, colon, breast, prostate, stomach, pancreas, skin)
- Sarcoma: originates from mesodermal tissue
- Leukemia and lymphoma: malignant proliferation of haemopoietic (blood-forming) cells
  - Leukemia: originates from one cell
  - Lymphoma: malignant proliferation of the lymphatic tissue

22.3. DEVELOPMENT OF TUMORS

Cancer formation is a multi-step process, many steps including both genotypic and phenotypic changes. Each step represents a breakthrough of one of the many antitumor-defense barriers of the body. Any mutation can be caused by exposure to mutagenic agents or environment. The steps of tumor development are as follows:

1. **Activation of oncogenes: Proto-oncogenes** promote normal cell division. Mutation of proto-oncogenes generates oncogenes, which speeds up proliferation within an environment where cells should be in resting state.
2. **Inactivation of tumor-suppressor genes:** Tumor-suppressor genes act as „**dividing stop**" signals. When a mutation inactivates these genes, the cell is released from growth inhibition. This adds to the effect of oncogene activation.
3. **Loss of the ability of suicide (= no apoptosis):** Apoptosis is a hard-wired process to eliminate misbehaving cells. A successful tumor-cell needs to turn off its apoptotic machinery to form a large population of malignant daughter-cells. The steps of the process include the following:
   - activation of pro-apoptotic genes
   - neglect death signals
   - neglect absence of surviving signals
4. **Angiogenesis**: Tumors may stimulate the growth of blood vessels into the tumor tissue, thereby increasing the accessibility of O_2 and nutrients in support of potential tumor expan-

sion. Angiogenesis begins with tumor cells secreting **angiogenic factors**. This is the initial stage of rapid tumor expansion.

5. **Reaching immortality:** Healthy cells possess limited proliferation, meaning they have a limited lifetime. This is regulated by the shortening of telomeres within every cell cycle, and hence, reaching to crisis and apoptosis (telomeres: regions on the end of certain chromosomes). Cancer cells demonstrate significant levels of **telomerase expression**, which is absent in differentiated cells. Telomerase elongates the telomeres of cancer cells, thereby subverting the generational clock towards immortality.

6. **Invasion – activation of metastatic genes:** Primary tumors may expand for years within the same organ, however, once when cancer cells break through the barriers and enter the blood stream, they often migrate to other organs and then colonize. This process, referred to as **metastasis**, is the most life-threatening characteristic of cancer.

7. Avoid the immune system – **blocking and hiding**

22.4. Characteristics of Cancer Cells

- Loss of growth control
- Malignant cells are not responsive to influences causing normal cells to stop growth and division. The capacity (rate) for growth and division is similar to normal cells. Unlike normal cells, malignant cells continue to develop in the absence of growth factors or when cells contact surrounding cells.
- Multilayer growth pattern
- Normal cells in culture grow as monolayer, whereas cancer cells grow in multilayered clumps, or foci. Additionally, malignant cells do not require attachment to a substrate in order to grow.
- Aberrant chromosome numbers (aneuploidy)
- Normal cells are diploid, while cancer cells are genetically unstable and often have highly aberrant chromosome complements, which may occur as a result of defects in the mitotic checkpoint or the presence of an abnormal number of centrosomes.
- Fail to elicit apoptosis
- Once the chromosome content of a normal cell becomes disturbed, apoptosis is activated. Cancer cells fail to elicit apoptotic response.
- High metabolic requirements
- Cancer cells often depend on glycolysis, which is an anaerobic metabolic pathway. This property may reflect the high metabolic requirements of cancer cells and an inadequate blood supply within the tumor. In hypoxia (reduced O_2), cancer cells activate a transcription factor called HIF that inducing the formation of new blood vessels and promoting the migratory properties of the cells, which may contribute to the spread and expansion of the tumor.

22.5. Factors causing cancer

- Carcinogens: cancer-causing agents: Agents damaging DNA
  - Mutagenic chemicals (present in high numbers typically in cigarettes and urban soot)
  - Ionizing radiation (X-ray, UV)

- DNA and RNA tumor viruses: carry genes whose products interfere with cell growth regulation.
  - DNA viruses: polyoma virus, simian virus 40 (SV40), adenovirus, herpes-like viruses.
  - RNA viruses (retroviruses): similar in structure to HIV

- Causes of most cancers is still unknown
- Diet can influence the potential risk for cancer
  - E.g., fat and alcohol increases the risk considerably
  - compounds in fruits, vegetables, tea reduce the risk
  - drugs often demonstrate a preventive effect (aspirin, indomethacin)

22.6. THE GENETICS OF CANCER

The development of a malignant tumor (**tumorigenesis**) is a multistep process. Every step of tumor development (discussed above) originates from genetic causes, tumorigenesis occurs by a cumulative progression of genetic alterations. Cancer results from the uncontrolled proliferation of a single wayward cell, therefore, cancer is said to be **monoclonal**.

The human body contains billions of cells which undergo cell division everyday. These cells may have the potential to mutate into a malignant tumor. Distinctively, the primary reason this only occurs in about one-third of the human population during an entire lifetime is due to a malignant transformation, which requires more than a single genetic alteration. However, there are a few types of mutations we can inherit making humankind much more likely to develop cancer. The genes we inherit (**germline mutation**) have a significant influence on our risks of developing cancer, but the most significant impact derives from genes which undergo an alteration in our lifetime (**somatic mutation**).

The genetic changes leading to cancer occur **gradually**, therefore the cells within the line:

- become increasingly less responsive to regulation
- and, are better equipped to invade normal tissues

Accordingly, in concept, tumorigenesis requires the cell responsible for initiating the cancer, the capability to possess a large number of cell divisions. This requirement has focused an immense attention upon the types of cells present in a tissue which may have the potential to develop into a tumor.

The first step is the formation of a **benign tumor**: a tumor composed of cells proliferating uncontrollably and cannot metastasize to other sites.

The products of the genes involved in carcinogenesis are usually responsible for:

a) cell cycle regulation
b) cell adhesion
c) DNA repair

The sequence in which genes are mutated influences the development of cancer.

- Genetic mechanisms: Point mutation / Translocation / Amplification → Abnormal control → Increased proliferation rate → Tumor
- Non-genetic mechanisms: Functional alterations of proto-oncogenes due to specific factors (inflammation or viral infection) or epigenetic processes

22.7. ONCOGENES

Oncogenes encode proteins promoting the loss of growth control and the conversion of a cell to a malignant state. Most oncogenes accelerate cell proliferation, may lead to genetic instability, prevent apoptosis, or promote metastasis. Oncogens act **dominantly**. Oncogenes were first discovered in the genomes of transforming viruses as genes responsible for tumorigenesis. An oncogene carried by a

virus is referred as **viral oncogene** (*v-onc*). Some examples for transforming viruses and their oncogenes:

- Rous sarcoma virus - RSV (oncogene: src)
- Human papilloma virus - HPV (oncogenes: E6, E7)
- Eppstein-Barr virus - EBV (oncogene: LMP1)
- Kaposi sarcoma-associated herpes virus - KSHV (oncogene: FGF4)

Retroviral oncogenes are derived from cellular proto-oncogenes (*c-onc*), such as the case of RSV (therefore v-src and c-src are distinguished). When a virus meets its host cell, they aim to infect it. A host cell can be *productively infected* (permissive cell), being subject of a lytic cycle. Virus DNA is replicated in the cytosol, new virus particles are assembled and as the cell dies, new virus particles are released. In the rare case of transformation, occurring in nonpermissive cells, an *abortive infection* happens, meaning the viral DNA integrates to the host genome. This genetic change triggers phenotypic changes, like alterations in shape and growth. The viral genom integrates randomly, and when it mutates a proto-oncogene, it becomes activated and the first step of tumorigenesis is complete.

Cellular oncogenes are derived from **proto-oncogenes**, which are normal cellular genes distinguishable in possessing a role in normal cell proliferation, or in association with genes bearing no known function. Included below are several **proto-oncogenes** with known functions:

- Growth factors
- Growth factor receptors
- Signal transduction proteins
- Transcription factors
- Regulators of cell cycle
- MicroRNAs

Any proto-oncogene can be subject of different **structural alterations** to generate activated oncogenes:

a) structural mutation
b) regulatory mutation
c) epigenetic alterations

There are different **pathways** of proto-oncogene activation:

1. **Loss of normal function**: The gene can be mutated altering the properties of the gene product so that the gene no more functions normally.
2. **Gene amplification**: The gene can become duplicated one or more times, resulting in excess production of the encoded protein.
3. A **chromosome rearrangement** can bring a distant DNA sequence into close proximity of the proto-oncogene, which can either alter the expression of the gene or the nature of the gene product.

In order for a cell to become malignant, both alleles of a tumor-suppressor gene must be lost, and a proto-oncogene must be converted into oncogene.

Examples for oncogenes

Different oncogenes become activated in different types of tumors, which reflects variations in the signaling pathways operating in diverse cell types.

1. Oncogenes encoding **growth factors** or their receptors include the following:
   - Simian sarcoma virus: contains the oncogene ***sis***, derived from a cellular gene which codes for PDGF. Overexpression of PDGF has been implicated in the development of brain tumors (gliomas).
   - Oncogene of avian erythroblastosis virus ***erbB*** directs the formation of an altered EGF receptor stimulating the cell in the presence of EGF. This altered version of the receptor stimulates the cell constitutively, meaning, regardless of whether or not the growth factor is present in the medium.
   - Some malignant cells contain more **surface receptors** compared to normal cells, which makes them sensitive to low concentrations of growth factors → overstimulated division.

2. Oncogenes encoding cytoplasmic **protein kinases** include the following:
   - **Raf**, a serine-threonine protein kinase (MAP kinase cascade) mutation → oncogene → the enzyme is always "on"
   - **Src** protein tyrosine kinase → phosphorylates signal transduction proteins → control of the cytoskeleton, and cell adhesion.

3. Oncogenes encoding nuclear **transcription factors**
   - **Myc** protein: stimulates cells to reenter cell cycle from G_0 stage. Overexpression → uncontrollable proliferation

4. Oncogenes encoding products affecting **apoptosis** include the following:
   - Overexpression of ***Bcl-2*** gene → suppression of apoptosis → abnormal cell proliferation → tumor development

22.8. TUMOR-SUPPRESSOR GENES

Most of the proteins encoded by tumor-suppressor genes act as **negative regulators of cell proliferation**, and consequently, their elimination promotes uncontrolled cell growth. They also help maintaining genetic stability, and this may be the reason for aberrant karyotypes in cancer cells.

Oncogenes, as mentioned above, arise from proto-oncogenes as the result of gain-of-function mutations (mutations causing the gene product to exhibit new functions leading to malignancy). Tumor-suppressor genes, in contrast, suffer **loss-of-function mutations and/or epigenetic inactivation** rendering them unable to restrain cell growth. One can think of oncogenes as accelerators and tumor-suppressor genes as **brakes for cell growth and proliferation**. Unlike for oncogenes, tumor-suppressor gene mutations act recessively (both chromosomes need to carry the mutation for the effect to manifest)

Classification of tumor-suppressor genes

1. Cell surface molecules (TGFβ receptor, DCC)
2. Signal transduction proteins (NF-1, GAPs, APC)
3. Transcription factors (RB, p53)
4. DNA repair genes (BRCA1, BRCA2, XP genes)
5. MicroRNAs

Examples

1. **Intracellular signal transduction proteins**
   a) **Neurofibromatosis: NF-1**
   → Ras GTPase actvation, inhibition of cell proliferation by Ras
   b) **Adenomatosis polyposis coli (APC)**
   → Stimulation of β-catenine degradation; β-catenine increases cell proliferation

2. **Transcription factors**
   a) **Retinoblastoma**
   RB gene was the first tumor-suppressor gene to be discovered. It is associated with a rare childhood cancer of the retina, called retinoblastoma. The disease can be inherited or occurs sporadically. In case of inherited retinoblastoma the trait running in the family is the disposition towards developing retinoblastoma, genotypically the deletion in one copy of the RB gene. The development of the disease requires both copies of RB to be altered or eliminated. In fact, a second sporadic mutation is the ultimate reason for disease manifestation. People who suffer from the inherited form of retinoblastoma are also at high risk of developing other types of tumors later in life, particularly soft-tissue sarcomas (tumors of mesenchymal rather than epithelial origin). Cells from these tumors in vitro can be “treated” by the reintroduction of a wild-type RB gene, which suppresses the cancerous phenotype, indicating that the loss of this gene function contributes significantly to tumorigenesis. The protein encoded by the RB gene (pRB) regulates the G_1 to S transition. The transcription factors of E2F family are targeted by pRB. During G1, E2F proteins are normally bound to pRB, which prevents them from activating a number of genes encoding proteins required for S-phase activities (e.g., cyclin E and DNA polymerase). A cell that loses pRB activity as the result of RB mutation would be expected to lose its ability to inactivate E2F, thereby removing certain restraints over the entry to S phase. The arrest of the cell cycle in G_1, required for normal cell differentiation, is directed by pRB.
   b) **p53 protein**
   Distinctively, p53 is said to be the “Guardian of the genome”, it suppresses the formation of tumors and maintains genetic stability. p53 gene may be the most important tumor suppressor gene within the human genome (TP53 is the most commonly mutated gene in human cancers). Notably, p53 acts as a transcription factor. In case of genetic damage, p53 inhibits the G_1-S transition to give time to DNA repair. Additionally, p53 can trigger apoptosis to eliminate cells with genetic damage. Failure to repair DNA damage leads to the production of abnormal cells possessing the potential to become malignant. Proper functioning of the p53 protein is very sensitive to even slight changes in the amino acid sequence.

3. **Other tumor-suppressor genes**
   Mutations of tumor-suppressor genes other than RB or p53 are detected in only a few types of cancer.
   a) Colon cancer: inherited deletion in a tumor-suppressor gene called *APC*.
   b) Inherited breast cancer: mutations in BRCA tumor-suppressor genes (BRCA: acts as transcription factor, has a role in DNA repair).

4. **MicroRNAs**: A New Player in the Genetics of Cancer
   MicroRNAs (miRNA) are tiny regulatory RNAs that negatively regulate the expression of target mRNAs. Some miRNAs inhibit the expression of mRNAs encoding proto-oncogenes. In the absence of the miRNAs, the oncogenic protein is overexpressed, which promotes the development of cancer. Because these miRNAs inhibit tumorigenesis, they can be thought of as tumor suppressors. Some miRNAs act more like oncogenes than tumor suppressors. One specific cluster of miRNA genes, for example, is overexpressed during the

formation of certain lymphomas. The abnormal expression of miRNAs has also been implicated as a causal factor in tumor cell invasiveness and metastasis.

New Strategies for Combating Cancer

Conventional cancer therapies fight tumor cells with the potential side-effect of damage in healthy cells. Currently, medical research endeavors to replace these procedures with **targeted therapies** using our knowledge about the molecular basis of malignancy, and for the same reason it is prompt to further study these processes. A therapy is said to be targeted, if it only attacks cancer cells, and leaves healthy cells unharmed. Another way to interpret targeted therapy is the therapy targets the protein whose inactivation leaves the cancer cells unable to grow or survive. Therapy can also be targeted to the cancer cells of a particular patient based on their unique pattern of somatic mutations. Several new and current established treatments are considerably successful utilizing targeted therapy, and as a result, the optimistic attitude to these studies is indeed welcome.

New strategies include:

1. Antibodies against tumor cells.
2. Inhibition of cancer-promoting proteins.
3. Preventing the growth of blood vessels responsible in the nourishment of the tumor.

Immunotherapy

- **Passive immunotherapy** uses antibodies made outside the body (i.e., in the laboratory) and administered to patients providing immunity against the disease. Passive immunotherapies do not stimulate a patient's immune system to "actively" respond to a disease in the way a vaccine does. The most widely used form is **monoclonal antibody** therapy. Monoclonal antibodies are considered targeted therapy, meaning it binds only to a single cancer-cell specific target, or it is directed at a cancer specific enzyme or protein. Binding of the monoclonal antibody initiates the apoptotic pathway of the cancer cell.
  - Herceptin: antibody against growth factors stimulating the proliferation of breast cancer cells (25% of breast cancers are sensitive to this therapy)
  - Rituxan: binds to B-cell surface proteins of non-Hodgkin’s lymphomas (95% success)
  - Vectibix: directed against the EGF receptor in colon cancer
  - Arzerra: likely to be approved for treatment of chronic lymphocytic leukemia
  - Under development: a radioactive atom or a toxic compound conjugated to the antibody targets the cancer cell, and kills the targeted cell.

- **Active immunotherapy** triggers the patient’s own immune system to fight against malignant cells. Cancer vaccines are mostly preventative vaccines, these are administered to a patient pior to illness. Therapeutic vaccines, like the prostate cancer vaccine, are given to a person who already has the disease. Cancer vaccines are targeted because they cause the immune system to attack the cancer cells, honing in on one or more specific tumor antigens.
  - Dendritic cells (immune cells) taken from cancer patients → made to display tumor proteins → injected back into the patient → tumor can be rejected
  - Another approach is to develop cancer vaccines against telomerase.
  - Other personalized treatments are being developed.

- **Immunotoxins** are human-made proteins consisting of a targeting portion (antibody) linked to a toxin. The toxin amplifies the action of the antibody, while the antibody provides specificity for the treatment. Protein binds to target (cancer) cell → endocytosis → toxin kills the cell. Immunotoxins do not generate allergic reactions, still toxicity on target cells is retained. In conclusion, immunotoxins destroy only the target (cancer) cells, with no side-effects. Exam-

ples for such toxins that can be bound to antigens and used as a considerable cancer treatment.

- ricin (lectin of castor bean *Ricinus communis*)
- mellitin (toxic component of the bee venom)

- **Inhibiting the Activity of Cancer-Promoting Proteins**
  The abnormal behavior of cancer cells is due to proteins present at abnormal concentration or display abnormal activity. Tumor cells are said to be dependent on the continued activity of deviant proteins (oncogene addiction). Blocking the activity of these particular proteins may kill the malignant cells, or stop the growth or invasive properties of the cell. Recently, researchers identified an arsenal of small molecular weight compounds found to inhibit the activity of deviant proteins. Test results demonstrate moderate success to date, however, the agents perhaps are not targeting the appropriate cells within the tumor.

- **Inhibiting the Formation of New Blood Vessels (Angiogenesis)**
  A growing tumor can undergo the “angiogenic switch”, a point when the chance for healing drops due to metastasis formation. An angiogenesis inhibitor denies the tumor to access to nutrients and oxygen required for growth, therefore these agents are beneficial in the fight against cancer (e.g., Avastin, inhibitor of VEGF growth factor).

23. GENETICS

23.1. INTRODUCTION TO GENETICS

The genetic code

The information flow in the cell begins with the genetic material of the cell; the DNA molecule, consisting of a chain of nucleotides of four different types. The first conversion of the genetic code is the transcription of DNA nucleotides to RNA; a process producing shorter, complementary RNA molecules by RNA polymerases. The second code conversion step is the translation of the RNA codons on ribosomes to peptides, which will make up polypeptide chains. In a very simplistic way;

$$DNA \xrightarrow{transcription} RNA \xrightarrow{translation} Protein$$

Regulation of protein expression

1. Protein expression can be regulated on different levels.
   a) Level of transcription
   b) Genes in the genome can be present in many copies, and the higher the **copy number** is, the higher the expression of the particular gene.
   c) The transcription begins when transcription factors bind to the promoter region, and changes in **promoter activity** cause changes is known as gene expression.
   d) Also, when factors binding to **upstream regulatory sequences** (enhancers, silencers) can alter gene expression.
   e) **DNA methylation** also alters the expression level.

2. Level of translation
   a) The **lifetime of mRNA molecules** affects the gene expression: longer lifetime means higher expression level.
   b) **Codon usage** also affects expression (see degenerate genetic code in “The Genetic Code”).
   c) Due to **alternative splicing**, many different mRNA transcripts may originate from a sequence; see “mRNA processing”.
   d) **RNA interference** can occur, with the silencing of mRNAs by small RNAs binding to them and not leaving them to be translated.

3. Post-translational modification
   a) **Chemical modification** of proteins alter their function. Localization is facilitated by target sequence addition. Also, formation of disulphide bonds is important for correct folding. Hydroxylation of aminoacids may be required, as well as phosphorylation.
   b) **Protein turnover** affects the functionality of proteins
   c) **Inhibition, or allosteric changes** also alter protein function.

Terminology

- The genome is the entirety of the hereditary information (DNA) of an organism.
- The **gene** is the unit of hereditary information; generally one gene carries the information to produce one gene product (polypeptide or RNA). Therefore, genes are also referred as the basic units of genetics.

- The number and morphology of species-specific chromosomes is known as **karyotype**, or the chromosome pattern.
- The **gene map** depicts the location of the genes on the chromosomes.
- **Cytogenetics** studies the structure and the inheritance of chromosomes.
- **Somatic cells** are diploid, having two sets of chromosomes (in humans they number 46), of which 22 pairs are autosomes, plus one pair of sex chromosomes named X and Y (males have XY, females XX)
- **Germline cells** are monoploid; they have one set of chromosomes (22 autosomes and one sex chromosome – X or Y).
- Homologous chromosomes or **homologues** are members of a pair of chromosomes, carrying the same genes on the same sequence (**locus)**, but at any specific locus they have identical or slightly different forms of the same gene (**alleles).**
- The **phenotype** is the collection of the observable features (**traits**) of an individual, while **genotype** means the complement of the genes of an individual.
- **Homozygote** individuals have two identical alleles at a given locus, while **heterozygotes** have two different alelles at the given locus.

Chromosomes

The human karyotype consists of 46 chromosomes, of which 44 are autosomes, and 2 sex chromosomes (X and Y). Chromosome morphology is described by the position of the **centromere** (the position where the sister chromatids are joined in the metaphase of mitosis), which determines the relative length of the arms (p for short arm, q for long arm). In **metacentric** chromosomes, the centromer is in the middle, so the arms are similar in length. The centromer of **acrocentric** chromosomes is at the end of an arm, making one arm really short, while the other arm is long. **Submetacentric** chromosomes have a longer and a shorter arm.

Chromosomes are classified based on the length and centromere position into 7 groups.

Using Giemsa-staining, the **karyogram** can be obtained by arranging the chromosomes in pairs and numbering according to size. **Idiogram** is the schematic view of the karyogram, showing the characteristic **G-band pattern** of the chromosomes, where one band covers about 50 genes. The numbering of the bands is achieved outwards from the centromere, the first number refers to region, the second refers to the band, after which a sub-band number can occur). For example, the human Major Histocompatibility Complex (MHC) locus is located on the short arm of the chromosome 6 at 6p21.3 ('six P two one point three').

Linkage of genes

(See also "Genetic linkage")

*During the prophase I of meiosis, **crossing overs** (exchange of genetic material) may occur between chromosome pairs, resulting in **recombinant chromosomes**. When matching chromosomal regions (obviously on homologues) break, they reconnect to the other chromosome. Two adjacent loci are likely to be located on the same sister chromatid. The probability of a crossover to occur between them (the **co-inheritance**) is proportional to the distance of the genes (loci). Closer genes are more likely to be inherited together. The distance of two genes is expressed in **centimorgan** (cM – in honor of geneticist Thomas Hunt Morgan). Note that genetic 'distance' doesn't refer to true physical distance.*

The human genome

The human genome consists of 46 nuclear chromosomes (plus the mitochondrial genome, see "Mitochondria): 22 pairs of autosomes and a pair of allosomes (sex chromosomes). Egg cells and sperm cells have haploid genomes, consisting of 3 billion base pairs of DNA. Somatic, diploid cells have six billion base pairs of genetic material. Differences between human genomes are on the order of 0.1%, while the difference between genomes of humans and closest relatives, the chimpanzees, is around 4%. Immense efforts have been made to reveal the sequence of the human genome: in 2012, the Human Genome Project managed to sequence thousands of complete human genomes. For further information on Human Genome Project, see Chapter "Human Genome Project", and lecture notes.

Human genome consists of 20,000-25,000 protein-coding genes, which gives only a small, 1.5% fraction of the genome. The rest is **non-coding DNA**, including

- genes for noncoding RNA (e.g. tRNA and rRNA)
- **pseudogenes**
- **introns**
- untranslated regions of mRNA
- regulatory DNA sequences
- repetitive DNA sequences
- sequences related to mobile genetic elements.

Pseudogenes are non-protein-coding genes and appeared during evolution via duplication and divergence. Some can have regulatory functions (promoters, splice sites) while others have lost their function and are not expressed in the cell. Pseudogenes have a common ancestor with a functional gene, therefore can be used to study evolutionary relationships. Their location corresponds to their developmental expression. As an example, note how the exon-intron pattern of the alpha-globin gene is conserved in many vertebrate globin genes (human neuroglobin and plant globin genes both have 3 introns). The series of individual development repeats the development during evolution.

Gene families are sets of similar genes, born by duplication of the original gene. For example the ten genes for human haemoglobin subunits are located in two clusters on different chromosomes, known as the α-globin and β-globin loci.

Benefits and concerns of genome research on different fields

Molecular medicine: Knowing an individual's DNA sequence improves the **diagnosis** of several diseases, even the potentiality can be tested for some diseases. Huntington's disease is a fine example of severity testing. The cause of the neuronal dysfunction within the disease is the accumulation of toxic protein fragments cut from abnormally elongated huntingtin proteins. The length of the protein is increased due to an enhanced number of a CAG segment repeat: 10-35 is the normal copy number; humans with 36 to 39 copies may not develop the signs of the disease, while those with 40-120 repeats nearly always develop the disorder. A repeat number higher than 60 usually associates to a juvenile onset of the disease, while a number below 60 is the cause among the elderly.

Pharmacogenomics designs "custom drugs" based on individual genetic profiles. Risk assessment evaluates the health risks faced by individuals who may be exposed to radiation and to cancer-causing chemicals and toxins.

In the fields of Bioarchaeology, Anthropology, Evolution, and Human Migration, understanding genomics will help research understand **evolution** (germline mutations in lineages) and **migration** (maternal inheritance). Today, one can study mutations on the Y chromosome to trace lineage and migration of males.

DNA identification: **Forensics** can compare suspect's DNA to evidence, exonerate persons who have been wrongly accused, identify crime and catastrophe victims, establish paternity and other family relationships based on genome research methods.

Ecology uses genome research to identify endangered and protected species, or detect microorganisms as pollutants.

During organ transplantation, the search for donors becomes easier in the use of current genome research methods. The circle for donor search can be expanded, as not only family relatives can match the criteria, and this can be effectively and quickly tested using genomic methods.

Agriculture and Bioprocessing can apply these methods to reduce the costs of agriculture and provide consumers with more nutritious, pesticide-free foods, increase outputs and reduce waste. Vaccine can be incorporated into foods, e.g., to immunize wild animals against pathogens. Alternate uses for crops such as tobacco have been found including a genetically engineered tobacco plant which produces a bacterial enzyme effective in the breakdown of explosives, such as TNT and dinitroglycerol.

Microbial Genomics offers rapid and efficient detection of pathogens, and also new energy sources (biofuels). GMO bacteria can produce medicine, detergents and even plastic.

Concerns about genome research include issues on the *privacy* of genetic information. Additionally, there is growing concern in reference to discrimination based on genetic information. Genetic testing may always bear the risk of *uncertainty*, thereby diminishing the reliability of testing below 100%. Lastly, the *commercialization* of the products (the tests) is often controversial and serves to raise questions about its reliability.

In *reproductive issues,* it is often uncertain if the parents have comprehensively understood the limitations and risks of genetic technology, yet, it may benefit the parents in making a difficult decision based on the otherwise reliable but not 100% test results.

In *clinical issues,* the tests require an intensive evaluation concerning accuracy and reliability. Also, healthcare professionals must prepare themselves for upcoming, new genetic methods. Once again, it is uncertain if patients possess the information in its entirety, and understand its limits to effectively enable them and their family members, to make the desired choice, albeit a difficult one.

Philosophical questions regarding human responsibility, i.e., free will vs. genetic determinism, may also exact scrutiny.

In reflecting on *health and environmental* issues, one may wonder if the manufacture of certain foods, among a host of products, are truthfully considered safe for human consumption and are indeed not a threat to the environment.

Genetic disorders

Genetic disorders are illnesses or abnormalities caused by abnormalities of the genome. Some are present at birth (congenital disorders), others manifest later (even at elderly age, such as Alzheimer's Disease). If the reason for the disorder is an abnormality of one or more chromosomes, it is characteristically defined as **chromosomal disorders**. Some genetic disorders are purely genetically determined, while others have a non-genetical "trigger" beyond the genetic background to manifest (**multifactorial** diseases). Disorders caused by the abnormality of a single gene are called **single gene defects** (Mendelian inheritance), while those requiring many abnormal genes to manifest are best known as "**polygenic disorders**".

Chromosomal disorders are classified based on the chromosome concerned, and the nature of the mutation causing the symptoms. See Chapter "Chromosomal disorders".

1. **Autosomal aneuploidy** syndromes affect autosomes, which have an abnormal number (other than 2): trisomy disorders; Down-syndrome (trisomy 21), Patau-syndrome (trisomy 13) and Edwards-syndrome (trisomy 18).
2. **Sex chromosome aneuploidy** syndromes affect allosomes, possessing an abnormal number. These can originate from a maternal non-disjunction in meiosis. A woman having only one X and no Y chromosome, has Turner-syndrome (45,X), while those men having two X and a Y chromosome have Klinefelter-syndrome (47, XXY) .
3. **Sex chromosome triploidy** is responsible for triple X syndrome (47, XXX) .
4. Common **deletion syndromes** are caused by a part of a chromosome which is missing, and or caused by an error in crossover during meiosis, which can be visualized by karyotyping. Deletion of a part of the short arm of chromosome 5 results in Cri du chat syndrome.
5. **Microdeletion syndromes** are caused by chromosomal deletions but too small to be visualized by karyotyping, whereby FISH (Fluorescent In Situ Hybridization) is then used in detection. Angelman-syndrome is caused by deletion or inactivation of genes on the maternally inherited chromosome 15 while the paternal copy, which may be of normal sequence, is imprinted and therefore silenced. Prader-Willi syndrome is caused by a similar loss of paternally inherited genes and maternal imprinting on the same chromosome.
6. **Triplet repeat expansions** demonstrate the instability of abnormal expansions of DNA-triplets. Such phenomenon is the primary reason Huntington’s disease occurs when the number of triplet repeats reaches 35. The abnormal HTT protein expressed causes the symptoms. In the case of Fragile X-syndrome, a repeat expansion above 200 results in the methylation of the CGG repeat expansion and *FMR1* promoter, leading to the silencing of the *FMR1* gene, the product of which is required for normal neural development.

Disorders with **single-gene (Mendelian) inheritance** are caused by a mutation in one single gene. See also Chapter “Single-gene (Mendelian) inheritance”.

1. **Autosomal dominant** inheritance: Even heterozygotes are affected, one mutant allele manifests, e.g., Achondroplasia, Huntington disease and Marfan-syndrome.
2. **Autosomal recessive** disorders: Only homozygotes are affected, two mutant alleles are needed for manifestation. e.g., Alkaptonuria and Cystic fibrosis. If both alleles are mutant, yet the mutations are different, the individual is a compound heterozygote: as an example, in some cases of beta-thalassemia. If an individual is a carrier for two mutations (of different loci), this is known as a double heterozygote.
3. **X-linked recessive disorders** are caused by a mutation on the X chromosome. Mostly, it affects males exclusively, through a healthy carrier mother: Duchenne muscular dystrophy, Haemophylia and Red-green color blindness.

Polygenic inheritance, complex diseases: More than one gene accounts for the manifestation of the disease. Many times, several genes (polygenic), or even environmental factors (complex) act together in the manifestation of the disease. See also Chapter “Polygenic inheritance, complex diseases”. Several examples include Alzheimer’s disease, Diabetes mellitus and Hypertension.

Mitochondrial inheritance refers to an atypical Mendelian inheritance pattern typical in disorders caused by a mutation on a mitochondrial gene. Clearly, maternal inheritance pattern characterizes these type of diseases, e.g., Leber hereditary optic neuropathy (LHON). For further information on mitochondrial inheritance, see Chapter “Mitochondrial inheritance” and “Mitochondria”.

Mutations

A mutation is a structural change in the DNA, which is inherited to daughter cells. According to the proportion of affected cells within an organism, germline and somatic mutations are distinguishable.

Germline mutations are present in gametes and are then fertilized, leading to every cell of the descendant organism to carry the mutation. **Somatic** mutations occur in one cell during individual development, therefore these are only present in a proportion of the cells.

Classification of the mutations can be achieved either on structural or functional base.

Structural classification:

1. Chromosome mutations: change in all or part of a chromosome
2. Gene mutations: change in a small segment of genomic DNA

   a) Point mutations = nucleotide substitution: change of one single nucleotide
   - Subclasses based on type of substitution:
     (1) Transition: purine for purine (A→G or G→A), or pyrimidine for pyrimidine (C→T or T→C)
     (2) Transversion: purine for pyrimidine, or pyrimidine for purine
   - Subclasses based on the effect on protein product:
     (1) Silent mutation: doesn't alter the encoded aminoacid (the genetic code is degenerated)
     (2) Missense mutation: causes aminoacid-change
       (a) Conservative: no effect on protein function
       (b) Non-conservative: altered protein function (A→T causes sickle-cell disease by Glu→Val substitution)
     (3) Nonsense mutation: a new STOP codon is generated, causing premature transition termination
     (4) Regulatory mutation = transcription mutation: mutation in a regulatory sequence, causing altered mRNA production
     (5) RNA processing mutation: mutations affecting mRNA processing (splicing, 5'-capping or 3' polyadenylation)
       (a) Splice-site mutations may occur at the beginning or end of an intron, causing exon skipping (splice acceptor sites are mutated)
       (b) Cryptic splice-site activation can occur by a mutation of a silent splice-site, causing a shortened exon

   b) Deletion/insertion: also include duplication, loss or gain of more than one nucleotides.
   - Small deletions/insertions: few nucleotide are deleted/inserted, may be a result of 'slippage': In replication, a mispairing occurs between complementary strands.
     (1) Frameshift mutation: The number of deleted/inserted nucleotides is not a multiple of three, therefore the mutation disturbs the reading frame, result in shortened proteins.
     (2) In-frame mutation: The number of mutated nucleotides is a multiple of three, therefore the reading frame is intact.
   - Large deletions/insertions: 22 bp to 10 Mb fragments are involved. Larger mutations can be analyzed with light microscope and are considered as chromosome mutations.
     (1) Unequal crossing-over: If a misalignment occurs between nearby sequences with close homology, an unequal crossing-over between these causes a deletion on one chromatid and duplication on the other.
     (2) Retrotransposition: Rarely large insertions consist of transposable elements like SINES or LINES (short and long interspersed sequences), moving from one region (deletion) of the genome to another (insertion).

- Unstable triplet repeat expansions: A triplet repeat size increase reaching a critical size causes either gene suppression (Fragile X syndrome), or production of abnormal protein (Huntington's disease).

Functional classification:

1. Loss-of-function (LOF): the quantity or activity of the gene product is reduced. In conditions having autosomal recessive inheritance, the heterozygous carrier is healthy, as the 50% of the normal gene product is enough to function normally. If the inheritance is autosomally dominant, due to a loss of 50% of the product and is therefore, insufficient, , the situation is best defined as **haploinsufficiency**.
2. Gain-of-function (GOF): As a result of the mutation, the activity or quantity of the gene product is increased, or new activity of the gene product occurs. Typically, Huntington's disease is referred to as a suitable example, with the more common form of dominant inheritance for GOF mutations.
3. Dominant negative effect: The product of the mutated allele interferes with the product of a normal allele and as a result, the net effect is a decrease in functionality.

Special terms and phenomena beneficial in the study of genetics

Mosaicism: An individual having more than one cell line, which are different genetically, as a mutation occurred during the early stages of individual development. Therefore, the individual has a proportion of cells with different genome.

Genomic imprinting: Different gene expression dependent upon the origin of the parent

Epigenetic effect: Alteration of the gene expression not caused by a mutation.

Homozygote: individual having identical alleles of a gene on both homologues.

Heterozygote: individual having different alleles of a gene on the homologues.

Compound heterozygote: Individual having two different, mutated alleles on the homologues. A common situation in cystic fibrosis and b-thalassaemia.

Hemizygous: individual having only one allele of a gene (or chromosome – typical in males).

Pleiotropic: A gene affecting multiple one traits, seemingly unrelated. For example, phenylketonuria (PKU), caused by the mutation of the gene coding phenylalanine hydroxylase, affecting the mental state, pigmentation and posture.

23.2. Chromosomal abnormalities

Chromosome mutations are changes which most often are caused by a problem during meiosis or by mutagens (e.g. chemicals and or radiation). The effects of these changes to the individual are often unpredictable.

Note below the analysis describing chromosome mutations utilizing the tools of cytogenetics (*see Chapter "Special methods in genetics"*).

Classification of chromosomal abnormalities

We can classify chromosomal abnormalities according to the nature of the change: it can be a **change in the number (aneuploidy) or change in the structure of chromosomes**. We can classify chro-

mosomal abnormalities according to the affected chromosomes and based on this, we can distinguish **abnormalities affecting sex chromosomes** and others which **impact autosomes**.

Chromosome abnormalities can be **inherited** from a parent or can be created “**de novo**”. Healthy parents can inherit diseases too, for example, trisomic symptoms can form because of a translocation or isochromosome of a completely healthy parent.

Most chromosome abnormalities which form during the development occur as an accident in the egg or sperm. Therefore, the abnormality will present itself within every cell of the body. Some abnormalities are generated only after conception, resulting mosaicism, where only some cells will contain the abnormality.

A child's risk of chromosomal abnormalities increase with the age of the mother. The phenomenon can be factually explained since women are born with all the eggs she will ever possess and so as a result, a woman who is 30 years old also has eggs which are 30 years old. Genetic mutations and problems can accumulate over time. In addition to the **maternal age, environmental factors** may be important too, such as exposure to radiation and mutagen chemical agents. Some studies infer the age of the father may not be completely considered insignificant, although the father’s age is believed to possess less impact on the risk of abnormalities.

Age of the mother	Risk of Down syndrome	Risk of any chromosomal abnormailty
20	0.060%	0.190%
22	0.070%	0.200%
24	0.080%	0.210%
26	0.085%	0.210%
28	0.095%	0.230%
30	0.105%	0.260%
32	0.130%	0.310%
34	0.200%	0.420%
36	0.340%	0.641%
38	0.571%	0.980%
40	0.943%	1.515%
42	1.563%	2.381%
44	2.632%	3.846%
46	4.348%	6.250%
48	7.143%	10.000%

Data based on information in Hook EB: Rates of chromosome abnormalities at different matemal ages. Obsttetrics and Gynecology 58:282-285, 1981; and Hook EB, Cross PK, Schreinemachers DM: Chromosomal abnormality rates at amniocentesis and in live-born infants. Journal of the American Medical Association 249(15):2003-2038, 1983.

Numerical abnormalities: aneuploidies

Aneuploidy: the deviation from the normal number of chromosomes.

In describing the number of chromosomes, we form words with the label “ploidy”. These words determine how many copies of chromosomes there are in each cell. In humans, the difference in these numbers may cause serious disease and is often lethal. However, we know many organisms which can live with different ploid states (mostly plants and insects) and we also know some useful aneuploidies within nature.

Polyploidy: more than 2 copies of a chromosome set.

We see useful polyploid organisms within our environment, too: many of our cultivated plants are polyploid. Triploid watermelon is seedless, bananas in markets are triploids, Persian lime is triploid and strawberries can be tetraploid or even octaploid. *For more: see the lecture slides.*)

Monosomy is defined as the lack of a chromosome. In humans, only the lack of X chromosome is compatible with life (Turner syndrome); in autosomes only “partial monosomies”=**deletion syndromes** can result in a viable human. The symptoms of deletion syndromes are created by the effect of haploinsufficiency or haploid insufficiency, meaning it is a phenomenon in which a single copy of the gene is not enough for the creation of wild-type phenotype.

Some important deletion syndromes:

- 1q21.1 deletion syndrome=TAR syndrome
- Wolf-Hirschhorn syndrome
- Cri du chat
- Williams syndrome
- Jacobsen syndrome
- Miller–Dieker syndrome
- DiGeorge syndrome

(*For more details about these diseases see lecture slides.*)

Trisomy: three copies from the particular chromosome

In generally trisomy means there is third copy from a chromosome in a cell.

Other conditions can form trisomic symptoms, too: where only a part of the chromosome exists in all three copies. In this way, we can describe more types of trisomic diseases, more phenomenon that can form trisomic symptomes: whole trisomy, mosaic form of trisomy and translocational trisomy. In the entire trisomy, there is an extra chromosome in every cell of the body. In the mosaic form, only a part of the cells in the organism contain the extra chromosome. In translocation trisomy, only a part of the chromosome is present in all three copies. The intensity of the symptoms are different according to the affected organs/tissues/body parts and the amount of tripled part and portion of the chromosome.

Some important trisomies:

- Down syndrome
- Edwards syndrome
- Patau syndrome
- Trisomy 9
- Trisomy 8=Warkany syndrome
- Trisomy 22=Cat eye syndrome
- Trisomy 16

(*For more details about these diseases see lecture slides.*)

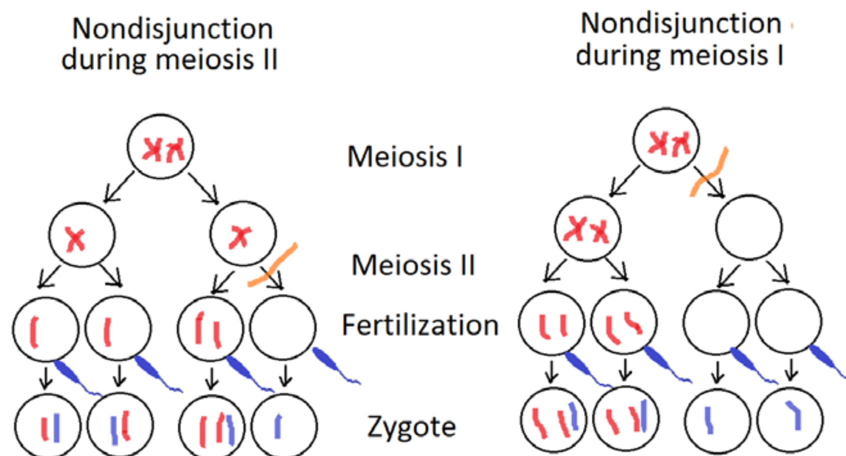
Tetra- and pentasomies are rare by autosomes in humans; but by sex chromosomes there were viable examples of XXXX, XXYY and even XXXXX, XXXXY and XYYYY.

Aneuploidy can cause by:

- Nondisjunction
- Anaphase lag
- Endoreplication/endoreduplication

Nondisjunction is a process where the sister chromatids or homologous chromosomes do not separate from each other during cell division. This can be caused by a random error and faulty form of drawstrings in principle, or other problems (we can induce it artificially with colchicine).

According to the place of nondisjunction, different descendants can form.

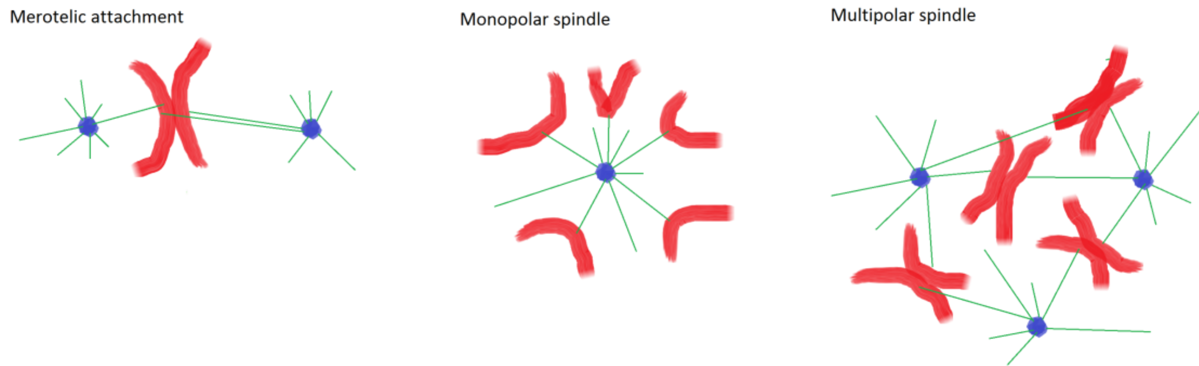


Nondisjunction can form in more ways: if a checkpoint is weakened, it may result daughter cells that will fail during the separation of chromosomes. Incorrect attachment and spindles (merotelic attachment, monopolar and multipolar spindles) can form nondisjunction, too.

Merotelic attachment means one kinetochore is attached to both mitotic spindle poles.

Multipolar spindles: more than two spindle poles form.

Monopolar spindle: only a single spindle pole forms. This produces a single daughter cell with doubled number of chromosomes.



Anaphase lag is a process where the migration of chromosomes slows down during anaphase and the result is one in which chromosomes can be excluded from one of the daughter cells.

Endoreplication is the replication of DNA (during the S phase) without cytokinesis.

The inadequate separation of chromosomes can lead to micronucleus formation: some of the genetic material excluded from the newly formed nuclei.

Mosaicism

Mosaic or mosaicism in genetic means the presence of two or more genotypes in one individual who developed from a single fertilized egg. (Mosaicism is not chimerism where two or more genotype arises from the fusion of more fertilized zygotes in the early stage of embryonic development.)

Mosaicism can form by more ways. For example a new mutation appears in a cell during the development and it spreads. Nondisjunction can occur after the fertilization during the development that leads to mosaicism.

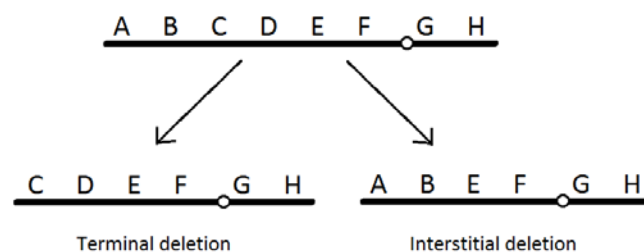
Structural abnormalities

(See Chapter XYZ for more information)

Deletions: a portion of the chromosome which is missing or deleted.

We can classify deletions according to the place and the size of the deletion:

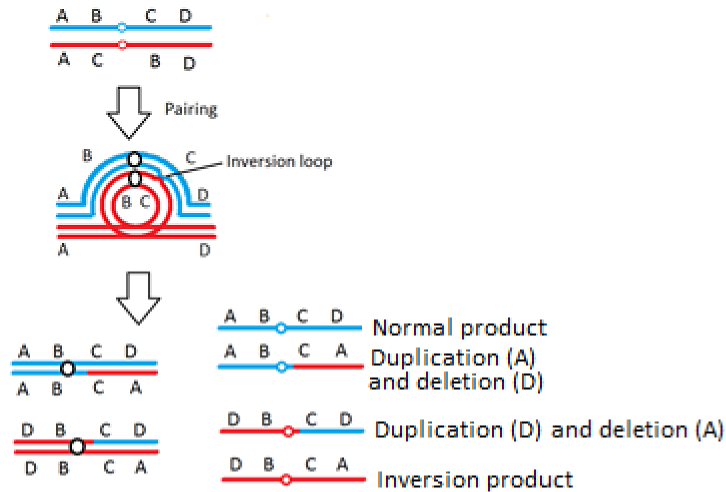
According to the place we distinguish interstitial and terminal deletions, according to the size, we can describe intragenic and multigenic deletions.



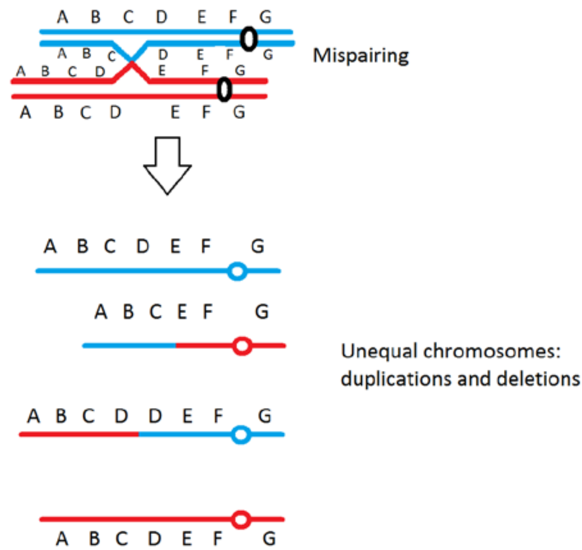
There are some deletions that are related to some tumors: specific types of neuroblastoma, melanoma and lung small cell carcinoma (*see lecture slides*).

Deletions can establish, through breaking without rejoining, losses from translocation, during chromosomal crossovers within a chromosomal inversion or by unequal crossing over.

A chromosomal inversion makes crossing over more difficult, the created new strands will have new variations as seen below (inversions may create duplications and deletions during crossing over).



Deletions can be created by unequal crossing over too, where homologous sequences do not pair precisely.



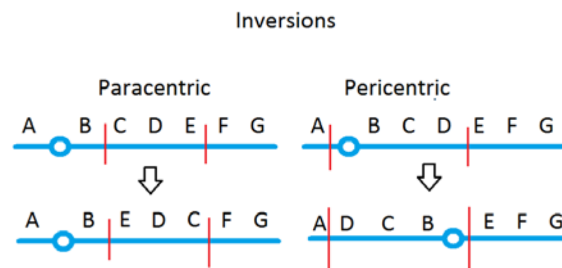
Duplications: A portion of the chromosome is doubled (extra genetic material)

There are two types of duplications: tandem and reverse duplications, according to the order of new genetic material.

Duplications can form as shown before via unequal crossing over and during chromosomal crossovers within a chromosomal inversion.

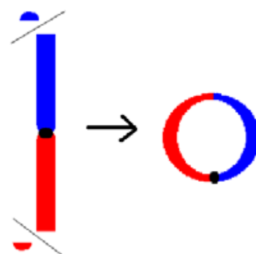
Inversions: a portion of the chromosome has broken off, turned upside down and reattached

There are two types of inversion according to the place of the break. If inversion occurs outside the centromere, then inversion is defined as paracentric inversion, if the inversion occurs in the region containing centromere, then we refer to it as a pericentric inversion.



As it is described earlier, inversions can make other chromosomal rearranges during crossing over.

Ring formation: a portion of the chromosome has broken off and formed a circle. During ring formation, both ends are typically lost, but it also happen when only one end is lost. Ring chromosomes can form from nearly all of the human chromosomes and this change in the structure can be compatible with life. The symptoms of a developing disease depends on the section that broken away from the chromosome. However, in some cases it develops results in only mild problems.



Isochromosome: a chromosome that has lost one of its arms and replaced it with an exact copy of the other arm. The result is a chromosome with 2 identical arms. Isochromosomes can have role in specific tumor formations and can lead to Turner's syndrome, too.

Translocations: a portion of the chromosome is transferred to another chromosome

There are two main types of translocations:

Reciprocal translocation: segments from two different chromosomes have been exchanged.

Robertsonian translocation: when two chromosomes fuse, usually at the centromere and thus creating a translocation. Only certain chromosomes (acrocentric chromosomes) are able to participate in this kind of process - in humans these only occur with chromosomes 13, 14, 15, 21, and 22. A human with robertsonian translocation is healthy, but the offspring may have different aneuploidic problems because these chromosomes cannot separate during the egg or sperm formation.

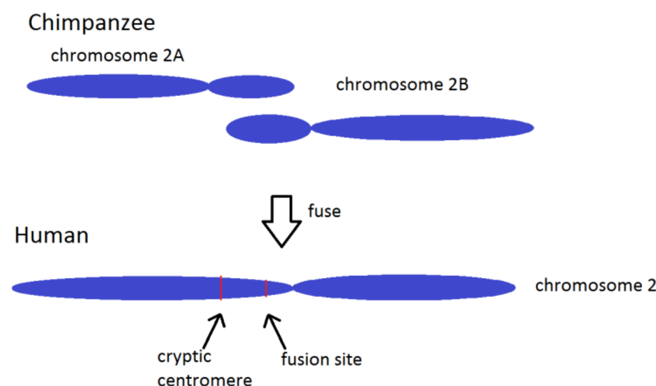
During translocations genes can be destroyed by the breaking sites. More studies demonstrate that specific translocations have a role in different types of solid tumor formation (*see lecture slides*).

Chromosomal rearrangements can be called **balanced** if rearrangements change the chromosomal gene order or place but do not remove or duplicate any of the DNA of the chromosomes (the amount of genetic material do not change). If the rearrangements affect the amount of genes/DNA, we refer to it as **imbalanced rearrangements**.

Balanced translocations can be creating normal phenotype but carrier parents can inherit loss or extra amount of genetic material resulting in the deletion syndromes or trisomic syndromes (*see lecture slides*).

Evolution – chromosomal level

The set of human and mouse chromosomes can be built into one another with deletions, inversions and translocations just as in how we use them as if they were Legos (*see lecture slides)* and it is right for every mammal (if we are generous and see the chromosome set global). For example, the largest difference between the chromosome set and genetic material of chimpanzees and humans is a chromosome fusion.



23.3. SEX-DETERMINING CHROMOSOMES OR ALLOSOMES

In contrast with the homologous somatic chromosomes (autosomes), X and Y chromosomes differ from each other. In mammals and birds they play a crucial role in biological gender determination.

Chromosome Y

Chromosome Y is significantly smaller than chromosome X. It is only 59 million base pairs and contains about 200 genes. This chromosome evolves the fastest. In humans (and in most of the mammals) chromosome Y determine the biological gender. Patrilineal inheritance: the father inherits it to all his sons.

Chromosome X

The chromosome x contains around 153 million base pairs and code 2000 genes. Human females have one pair of X chromosomes but one of it is silenced (see X chromosome inactivation). Human males have one X chromosome inherited from their mothers.

Pseudoautosomal regions

Chromosome X and Y probably evolved from one pair autosomes. These two chromosomes are so different now that only acrosomal regions (pseudoautosomal region, PAR) are able to pair and recombine. These regions are highly similar on both chromosomes, they make possible the pairing and segregation of these chromosomes during meiosis.

X-Y sex determination system

The presence or lack of Y chromosome will define the biological gender. The key gene is called SRY, what codes the testis-determining factor (TDF or sex determining region protein Y (SRY)). TDF is a transcription factor that upregulates other transcription factors what will differentiate the gonad into testis.

X chromosome inactivation

Because female cells contain twice as many copies of X chromosomes than male cells, there must be a mechanism present to compensate of this difference. This mechanism is the X chromosome inactivation. As a result one of the X chromosomes gets inactivated, the transcription is inhibited on that chromosome. (See: calico cat from the lecture)

From one of the randomly chosen X chromosomes a long, non-coding RNA transcribed (Xist), what covers that chromosome (Xi chromosome). As a result the transcription from that in silenced, while the other chromosome remains active (Xa).

Xi forms condensed heterochromatin and attaches to the nuclear membrane. The observation of these Barr bodies were an important tool in biological gender determination (see: Sex and Olympics at lecture).

The pseudoautosomal regions remain transcriptionally active on both X chromosomes.

Numerical abnormalities

As seen before, the loss or gain of a chromosome has catastrophic results. Numeric abnormalities of autosomes are almost always lethal for the cells, with only a handful of exceptions (Down-syndrome, malignant cells). However there are a few diseases what are caused by abnormalities of the allosomes.

Klinefelter-syndrome (XXY): Affected men carry on extra X chromosome. Symptoms: tall stature, feminine characteristics, hypogonadism, often infertility.

Turner-syndrome (X0): The cells of the affected women lack one X chromosome (and the Barr body). Symptoms: short stature, low hairline, webbed neck, wide chest, broad hands and legs. Possible problems: learning difficulties and gonadal dysfunction. The latter causes amenorrhea and infertility.

Triple XXX syndrome (XXX): Trisomy X. The affected women might be taller than the average, but there are no significant physical signs of this condition. The might be menstrual disturbances, premature ovarian failure.

XYY-syndrome: The carrier men has one extra Y chromosome. There are no significant physical symptoms, they might face learning difficulties.

Genetic diseases linked to allosomes

Y-linked diseases

Because there are no vital genes on chromosome Y, there are no severe diseases linked to it. (Except: disorders of SRY region.)

X-linked diseases

See: topic on genetic diseases

23.4. Mendelian (single-gene) inheritance

Several human traits and disorders are consequences of an error in one single gene. The number of such traits is over 12 000. Gregor Mendel, in the late 19th century, performed an enormous, 8-year study on peas to observe their inheritance patterns. His study resulted in the establishment of the fundamental laws of inheritance. The *law of segregation* says that individuals have two copies of a gene for each trait and only one of the alleles is inherited to the offspring. The *law of independent assortment* says that the copies are separated independently. The term "Mendelian" is used for single-gene inheritance as respect to the work done by the Czech researcher.

Terminology

Autosomal genes are located on autosomes (other than sex chromosomes). Autosomes are in pairs, thus genes (copies of a gene) exist in pairs. **X-linked** or **Y-linked genes** are located on sex chromosomes (allosomes).

The **genotype** is the genetic constitution, while the **phenotypes** refers to the trait observed on the individual.

A **locus** is the position of the gene on the chromosome. **Alleles** are the alternative forms of the same gene.

Heterozygotes have different alleles at the same locus, while **homozygotes** have identical genes at the same locus.

Heterogeneity is synonym to diversity, **locus heterogeneity** suggests that a trait can be caused by mutations in genes at different loci, while allelic heterogeneity indicates a trait resulted by several different mutations at the same locus.

Dominant phenotypes can be caused by a single copy of an abnormal gene, while for recessive phenotypes both copies should be abnormal for the manifestation of the illness.

Pedigree analysis is a technique applied on genetic counselling. The construction of the family tree of the affected individual(s) helps us determine the inheritance pattern of the illness.

Autosomal dominant inheritance

An error in one of the copies of a gene is enough for the disorder to manifest, meaning that a person will be affected with one mutant copy of a gene (the other allele can be normal). Affected individuals' children inherit the mutant allele with 50% chance. As the child is affected with this single copy, the overall average probability of being affected is 50%. The main characteristics of autosomal dominant inheritance:

- Heterozygotes are affected (homozygotes too).
- Both sex are affected and both inherit the illness to children of either sex.
- Chance for affected individual's children to be affected is 50%.
- Disorder passes through the family, showing **vertical transmission**.

Some examples for disorders showing autosomal dominant inheritance: Huntington-disease, retinoblastoma, familial hypercholesterolemia, Marfan-syndrome.

Variable expression/expressivity

For some autosomal dominant disorders the severity of the disease varies within the affected family. This means that the severity of symptoms is different between individuals. For example, the symptoms of neurofibromatosis type1 spread from pigmented areas of skin (café-au-lait patches) in childhood to scoliosis and even malignancy.

Reduced penetrance

Some autosomal dominant disorders' pedigree shows a lack of affected individuals in a generation, called non-penetrance. **Penetrance** is the proportion (expressed as proportion of 1 or percentage) of heterozygotes showing symptoms. Examples of reduced penetrance are retinoblastoma, with approx. 90% penetrance.

Homozygosity

For autosomal dominant disorders, homozygotes can rarely occur with an elevated severity (related to what is seen for heterozygotes). This can be seen in the case of early coronary artery disease in a homozygote for familial hypercholesterolaemia. Mating of heterozygotes can be the cause of homozygosity of the child. As an exception, the example of the Huntington disease shows that homozygotes are not more severely affected than heterozygotes.

Codominance

Two or more alleles that are expressed equally are called codominant. An example is the inheritance of the AB0 blood group, where the child with an AB genotype expresses both A and B phenotypes equally (both A and B alleles are dominant over the 0 allele).

Autosomal recessive inheritance

A trait or disorder manifested only if it is genotypically homozygous – that is, having two mutant alleles – is said to be inherited in an autosomal recessive pattern. Each allele was inherited from a heterozygous (carrier) parent. In case two heterozygote carriers have children, in average 50% will be carriers, 25% will be homozygous unaffected and 25% homozygous affected. The main characteristics of the autosomal recessive inheritance:

- Only homozygotes are affected (and compound heterozygotes – see later), heterozygotes are carriers.
- Both sexes are affected.
- **Horizontal inheritance** pattern: members of a single sibship are affected.
- If heterozygotes have children, the risk for the sibling to be affected is 25%.
- If heterozygotes have children, the risk for a future sibling to be carrier is 50%, while the risk that an existing unaffected sibling is a carrier is 67%.

Many disorders show autosomal recessive inheritance. These include cystic fibrosis, galactosaemia, phenylketonuria (PKU), sickle-cell disease and thalassemias.

Consanguinity

If two individuals having a common ancestor are in relationship, we call it **consanguineous mating**. Offsprings have an increased risk for autosomal recessive disorders, compared to the risk for the general population. The risk increases in proportion of the degree of relation between the parents. Most individuals carry at least one autosomal recessive mutation, but as they are heterozygous, they will not manifest. The chance that a child of first cousins will be homozygous for its grandfather's mutant allele is 1.5% (same is true for grandmother's mutant allele), so the overall risk of the child for having an

autosomal recessive disorder is 3%. In an extreme and illegal case of first degree relatives having children, the risk for abnormality is 50%, (if the parents are both heterozygote)!

Pseudodominant inheritance

Sometimes there are many carriers of an autosomal recessive disease within the extended family, thus many mating will occur between heterozygotes and homozygotes. This results in the autosomal recessive disorder to show autosomal dominant pattern in the pedigree. This can happen in small inbred communities or as a consequence of assortative mating (choosing partner by similar disability).

Compound and double heterozygosity

If an autosomal recessive disorder is caused by different mutations in the two alleles, the affected person is called a **compound heterozygote**. An example to such phenomenon occurs in cases of cystic fibrosis and b-thalassemia. These disorders showed allelic heterogeneity, still recessively inherited.

An individual carrying different autosomal recessive disorders is a **double heterozygote**. These individuals are unaffected by both disorders (but carriers for both), while a compound heterozygote shows symptoms.

X-linked recessive inheritance

Mutations in a gene located on the X chromosome lead to disorders with X-linked recessive inheritance, and usually affect males exclusively. Males are said to be **hemizygous**, which means they have only one X chromosome (thus one copy of the gene). If they have only one mutant allele, it is obvious that they will be affected. Females have two X chromosomes, therefore two copies of the gene, and the healthy copy compensates the effect of the mutant one. Therefore females are rarely affected by X-linked recessive disorders. A heterozygous female is carrier of the disease, which will manifest in 50% of her sons, while 50% of her daughters will be carriers of the disorder. An affected male will transmit the mutant allele to all of his daughters (who will be obligate carriers), and none of his sons (as no mutant allele occurs on Y chromosome). The main characteristics of the X-linked recessive inheritance:

- Only directly related males are affected through a female line.
- The chance for a carrier female to have affected sons is 50%, for daughters it is 50% chance that they are carriers.
- Daughters of an affected male will all be carriers, sons will all be unaffected.
- There is no male-to-male transmission.
- Affected females (who have symptoms) are extremely rare – only if a daughter gets a mutant allele from the (carrier) mother and the (affected) father as well.

The number of known X-linked recessive diseases is over 200, including androgen insensitivity, Duchenne muscular dystrophy, haemophilias and red-green color blindness.

X-chromosome inactivation

As men have only one X-chromosome and women have two, theoretically women should have double the level of X-chromosome encoded proteins. This is not true, the protein levels are similar because of the **dosage compensation** (the male and the female X-chromosome activities are equal). One of the X-chromosomes is inactivated in 1-2-week-old female embryos, caused by the Xist (X inactivation specific transcript) inactivation signal, which is expressed only on one of the X chromosomes. Inactivation signal is present as the methylation of the 5’ region of genes, blocking the expression of the gene. Inactivation excludes the **pseudoautosomal region**, a small area on the tip of the short arm.

X-chromosome inactivation occurs randomly in every cell: on average half of the female embryo's cell expresses maternal X chromosome, the rest expresses male X-chromosome. Daughter cell of these embryonic cells will transmit the inactivation pattern of the parent cell. Consequently, adult females will have half of their cells with maternal X-chromosome expressed, whereas the other half expresses paternal X-chromosome. This process is referred to as **Lyonization**, after Dr Mary Lyon who first described it.

Female interphase cells show a so called **Barr body** in the nucleus, which is a small mass of sex chromatin representing the inactivated X-chromosome.

Note: If a female has three X chromosomes (47,XXX), two of them will be inactivated (two Barr bodies can be seen), and for males with two X chromosomes (47,XXY), one X chromosome will be inactivated (there will be a Barr body in these males' cells).

Females with an X-linked recessive disorder

In a very rare situation, females can also be affected by X-linked recessive disorders. The reasons can be the following:

Homozygosity

In the rare case that a female is homozygote to a recessive mutation, she is similarly affected as a hemizygous affected male. A female can inherit the mutant allele from both parents, or one mutation comes from one parent, the other can be a new mutation. A relatively common example of such a case is red-green color blindness (1/12 males and 1/144 females).

Turner syndrome

The absence of the second sex chromosome (45,X0) is called Turner syndrome. The expression of the (only) mutant allele is similar to normal males with only one X chromosome (46,XY) as neither has a normal homologous allele.

Androgen insensitivity

The insensitivity to testosterone is a very rare disorder affecting a chromosomally male but phenotypically female individual. Such a 'female' can have an X-linked recessive mutation with an effect seen for males.

Skewed X-chromosome inactivation

X chromosome inactivation occurs randomly: in each cell inactivation of maternal or paternal X chromosome is a matter of chance. Rarely this normally 50:50 ratio can be asymmetric and in case the active X chromosome holds a recessive mutation, the woman will show phenotypical signs. Asymmetry can be the result of preferential inactivation of abnormally shaped X-chromosomes.

Other types of Mendelian inheritance

X-linked dominant inheritance

It is an X-chromosome mutation, which is phenotypically manifested in heterozygous females. Examples include incontinentia pigmenti, Rett syndrome, fragile X syndrome. Features of the X-linked dominant inheritance are as follows:

- Both sexes are affected, usually females are less severely (mosaic involvement if skin is affected).

- Chance that a child of an affected female will be affected is 50%. Some disorders are lethal in males (only unaffected males can survive, therefore the sex ratio of children of an affected mother is 2:1 daughters: sons with half of the daughters affected)
- There is no male-to-male transmission, all the daughters and none of the sons of an affected male will be affected.

Very rarely affected females are unable to reproduce, eg. in Rett syndrome, every affected woman is an 'isolated' case.

Y-linked inheritance

Holandric disorders are transmitted from father to son only, as genes on Y chromosome are involved. Most of these genes play role in the determination of male sex and spermatogenesis. If a mutation of a Y chromosomal gene causes infertility, it will be transmitted to all the sons and next generation males.

Pseudoautosomal inheritance

A mutation in a gene of the pseudoautosomal region causes **partial sex-linkage**. Crossover can happen between these regions, therefore a mutation can cause a disorder which seems to be X-linked in some parts of the pedigree and Y-linked in others. An example is Leri-Weil syndrome, caused by a mutation in the SHOX gene, resulting in short stature and skeletal abnormalities.

Atypical Mendelian inheritance

Some single-gene disorders may show unexpected patterns of inheritance. Here we mention some mechanisms that account for such abnormalities.

Anticipation

For some single-gene disorders the disease occurs earlier or with more severe manifestation in sequential generations. This phenomenon can be understood when symptoms are caused by triplet repeat mutations expanding from generation to generation. Triplet repeats can be small, as in Huntington's disease, or large, as in fragile X syndrome.

New mutations, germline mosaicism

When an individual in a family turns out to have a single-gene disorder as an isolated case, the disease is a consequence of a new mutation during any step of gametogenesis. As it cannot be determined, which stage the mutation arose, genetic counselling on a future sibling is a difficult issue. If the mutation arose in a late mitotic division, or in meiosis, the risk for a future sibling is very low, but if it occurred during an early mitotic division in a germline stem cell, the parent is likely to produce other gametes carrying mutation (**germline mosaicism**).

Digenic and triallelic inheritance

Some diseases only manifest if two mutations are present together, the additive effects of which account for the symptoms. For example, in retinitis pigmentosa, an individual with mutations in both the *ROM1* and *peripherin* genes develops severe vision loss, while a mutation, only in one of these genes, leaves the individual unaffected. If the expression of a disease requires three mutations (both a dominant susceptibility mutation and recessive mutations on another locus), it is called **triallelic** inheritance.

Genomic imprinting

Despite the Mendelian principles that suggest identical alleles to express equally, some genes show different expression patterns. During **genomic imprinting**, these genes are tagged with a molecular message in the meiosis (depending on the parent of origin). The process involves methylational inactivation of regions regulating transcription, and as this alteration of a gene's expression does not change its structure, it is said to be an **epigenetic** phenomenon. Parental imprint does not change during replication and cell division (meaning it is the same throughout the life), but it is removed during gametogenesis, then it is re-established.

Uniparental disomy

According to the Mendelian law of segregation, chromosomes segregate independently. This law is usually true, but in few cases the individual gets both homologous chromosomes from the same parent. **Uniparental disomy** (UPD) can arise by a nondisjunction during meiosis in one of the parents, which results in a disomic gamete. If such a disomic gamete is fertilized by a normal gamete, the result will be a trisomic zygote. After this, a disposal of one of the three chromosomes can make the zygote to survive (**trisomy rescue**).

Uniparental heterodisomy refers to the case when a non-disjunction during meiosis I leads to the condition when both parental copies are present. **Uniparental isodisomy** indicates the presence of two copies of a parental chromosome, suggesting a nondisjunction during meiosis II.

UPD can be masked (unnoticed), if there are no recessive mutations carried by the chromosome involved in UPD. If UPD involves genomically imprinted areas, diseases like Angelmann syndrome or Prader-Willi-syndrome can occur.

Mitochondrial inheritance

Mitochondria, located in the cytoplasm, contain DNA, called mitochondrial DNA. Any individual receives mitochondria from the mother only, as sperms have few mitochondria, which are selectively eliminated after fertilization. Any disorder caused by a mutation in mitochondrial genes is maternally inherited, showing **matrilineal pattern** of transmission, called **cytoplasmic** or **mitochondrial inheritance**. Two examples are **MELAS** (mitochondrial encephalopathy, lactic acidosis, and stroke-like episodes) and Leber hereditary optic neuropathy (LHON).

Predicting the risk for the offsprings is difficult, because we cannot know if a few or a lot of mitochondria carry the mutation. These are referred to as **heteroplasmia** and **homoplasmia**, respectively. These explain the wide range of variability in these disorders.

For more detailed information on mitochondrial inheritance, see Chapter 'Mitochondria'.

Genetic linkage

According to observations, alleles of two adjacent loci are more often inherited together, than it would be expected by chance. This phenomenon is referred to as genetic linkage. During meiosis, crossovers can happen between homologues. The probability of a crossover between two distant alleles will be high. For closer loci there is low probability for crossover to happen between them. Linked alleles of the same chromosome are called **coupled** alleles, while alleles on the opposite chromosomes are said to be in **repulsion**.

The frequency of crossover to occur between two loci is the **recombination fraction**, denoted as Θ (greek theta). Its value can be obtained by investigating the genotypes in families. If 1 out of 10 children has inherited a recombinant chromosome, the recombination fraction is 0.1. The value of 0.5

would show segregation by single chance, thus the loci are not linked. For values under 0.5, the two alleles are proven to be in linkage.

Linkage and physical distance

The genetic distance refers to the probability of a crossover to happen between the two genes of interest. Genetic distance is measured in **centimorgans** (cM), or map units. 1 cM means that 1 recombination occurs in 100 meiosis. Therefore a recombination fraction of 0.01 equals to 1 cM. Genetic linkage is not a proportion of the physical distance. The latter one is measured in base pairs, and the relationship is not linear, as at so called recombination hot spots, there is higher likelyhood of a crossover to happen.

LOD scores

LOD stands for Log of the odds. This is a method to prove or disprove linkage of two loci. When LOD is above 3, the linkage is proven, while an LOD under -2 shows that the loci are not linked.

Linkage disequilibrium

The chance of association of two alleles at linked loci can be calculated theoretically. For two linked loci A and B, both with two alleles (A1, A2 and B1, B2), the frequency is 0.5 for each allele. Each possible haplotype has a frequency of 0.25 as an expection (A1B1; A1B2; A2B1; A2B2 with a frequency of 0.5 X 0.5 = 0.25). If these frequencies are seen in the population, the loci are in linkage equilibrium. However, if studying the population shows significant deviation from the expected value, we deal with linkage disequilibrium.

23.5. MITOCHONDRIAL DISEASES

Structure of mitochodria

See the chapter on cellular organelles

Usual symptoms of mitochondrial diseases

The main function of mitochondria is to generate ATP through oxidative phosphorylation, therefore failure of these organelles will result ATP shortage. Consequently the most affected tissues will be the ones with the highest energy consumption, e.g.: heart muscle, skeletal muscle, retina, central nervous system. The symptoms are usually not specific what make the diagnosis of mitochondrial diseases especially hard to diagnose. The symptoms might be: ataxia, myopathy, cardiomyopathy, seizures, loss of sight or hearing, lactate acidosis.

Genetic classification of mt diseases

Homoplasmy: all copies of mitochondrial DNA (mtDNA) share the same sequence within the cell.

Heteroplasmy: there are more than one mtDNA population present within the cell. In the cases of diseases: some copies will carry mutation while others will be wild type. The symptoms of the disease manifest only if the ratio of the faulty copies above a given threshold.

Examples of mitochondrial diseases

- Leber hereditary optic neuropathy (LHON): A first identified mitochondrial disease, which is the mutation of the gene coding the NADH dehydrogenase enzyme. The usual symptom is bilateral loss of sight in early adulthood.
- MERRF syndrome (or Myoclonic Epilepsy with Ragged Red Fibers): Caused by the mutation of tRNA-Lys. Symptoms: myoclonic epilepsy, loss of hearing, lactate acidosis. The histologic picture of the muscle fibres shows the accumulated, defective mitochondria in the subsarcolemmal region.
- Leigh disease: Necrotising encephalomyelopathy with various genetic background. The first symptoms appear early (3 months - 2 years): problems of coordination, seizures, and lactate acidosis. The disease is always lethal.

Heredity of mitochondrial diseases

The heredity of mitochondrial diseases depends on the localisation of the mutation.

- Mitochondrial heredity: the mutation occurred in the mtDNA. All the children (boys and girls alike) of the affected mother will inherit the disease, while an affected father's child will remain healthy.
- Nuclear DNA mutation: the heredity will follow the Mendelian rules.

Treatment

Unfortunately, the treatment is only alleviation of the symptoms. The diagnosis of mitochondrial diseases is possible by sequencing the affected genes.

As a pharmacist you must be aware the fact that some drugs might damage the mtDNA. Azido thymidine (AZT) is one of these compounds. AZT was the first drugs used successfully in AIDS patients. It acts by blocking the activity of the viral reverse transcriptase. Many of its adverse side effects is due the mitochondrial damage. These days, the combined drugs use minimal concentration of AZT to minimize the side effects.

Even if it is not possible to treat mitochondrial diseases, pronuclear transfusion can prevent the inheritance of the disease. During this process, the nucleus of the affected mother's oocyte is transfused to an enucleated oocyte from healthy donor. This way the mtDNA in not inherited (see lecture)

23.6. Epigenetics

Definition of epigenetics

Epigenetic refer to changes of the genome which alter gene expression, but there are no changes in the nucleotide sequence. Epigenetics changes may or may not inheritable.

The main mechanisms of epigenetics:

- DNA methylation
- histone modifications
- microRNA
- X-chromosome inactivation

DNA methylation

The most prominent methylation of DNA is the methylation of cytosine at the 5 position. The methylation process is performed by DNA-methyl-transferase enzymes (DNMT) and the result is the 5-methylcysteine. Usually the cytosines upstream of guanines (CpGs) are methylated. High concentrations of CpGs can be found around promoter regions, which are called CpG islands. These CpG island regulate gene expression. The hyper methylation of a promoter region will cease the gene expression. It is done by Methyl CpG Binding protein 2 (MeCP2) what binds to hypermethylated DNA segments, physically blocking the transcription factors.

DNA methylation pattern is inherited to the daughter cells of cell division. Shortly after replication DNA methyltransferases copy the methylation pattern of the the methylated strand to the newly synthesized, unmethylated strand. The time gap between replication and methylation is crucial in mismatch repair (MMR, see DNA repair chapter)

Genomic imprinting is the process when certain alleles inherited from one of the parents are silenced. The mechanism of silencing is mainly DNA methylation, so there is no change in the DNA sequence. There are some diseases (e.g. Angelman syndrome/Prader-Willi syndrome) where genomic imprinting has a fundamental role determining the clinical symptoms (see lecture).

Histone modifications

Histone proteins can undergo several posttranslational modifications, involving the addition of small groups (methylation, phosphorylation, acetylation, ubiquitination, sumoylation). These modifications can alter the structure of the nucleosome, thus changing expression of the involved genes.

microRNA

MicroRNA (miRNA) molecules are short (17-25 nucleotides long), non-coding RNA. They play a role in mRNA silencing, by pairing to complementary mRNA strands and causing its degradation either by the cleavage of the mRNA, decreasing its stability or decreasing the effectivity of translation.

X-chromosome inactivation

See chapter on sex-determining chromosomes

23.7. Polygenic inheritance, complex diseases

Terminology and definitions

The topic of Medelian inheritance describes the distinct characteristics of single-gene inheritance. **Non-Mendelian inheritance**, in comparison, features traits inheriting in an oligo- or polygenic manner. In the discussion of multifactorial disease, environmental factors often act in addition to the polygenic inheritance of susceptible genes.

A large group of common or **complex diseases** exhibits **familial aggregation** (Diabetes mellitus, Alzheimer disease etc.). These cannot be explained by Mendelian inheritance, thus diagnosis and therapy for most of these diseases is often difficult, therefore, these diseases are currently a hot topic of genetics research

Multifactorial traits (and also disorders) are determined by interactions between genes at different loci, with consideration to the effect of environmental factors and influences. If only a few genes are

known to demonstrate their role at the onset of the disease, it may be concluded or defined as **oligogenic inheritance**, whereas if there are many genes involved, it is best described as **polygenic disorder**. **Complex disease** is defined when environmental factors interact with genetic inheritance. Complex disorders can be categorized distinctly on the onset of the disorder: some can be present at birth (e.g. Hirschprung disease), others occur in later life, like Alzheimer's disease.

Identification of multifactorial disorders

Notably, multifactorial disorders cannot be directly correlated to Mendelian inheritance as pedigree analysis cannot prove multifactorial inheritance. Therefore, to acquire evidence in support of both genetic and environmental components in association to disease, researchers are endeavored to establish conclusions based on data originated from the combination of the following studies:

Family studies

Family relatives share a proportion of their genes depending on the degree of relationship (eg. first degree relatives share 50% of their genes, second degree relatives share 25% etc.). If a trait is determined by multifactorial inheritance, relatives often demonstrate the trait in proportion of their genetic similarity.

In family studies families with one or more affected individuals are identified, and the incidence of the disorder in other family members is studied. Following the analysis of the data, the **relative risk factor** is determined: λ_R is the risk to relative R as compared to the population risk. For example, if λ_S equals 10, it means the risk for a sibling to be affected by the disorder is 10 times higher than the risk for the general population.

The verification a disorder demonstrates familial aggregation is not adequate to surely prove multifactorial inheritance. Some diseases are distinctly relative to environmental factors and influence or habits typically shared by family relatives such as, nutrition, well-being, physical training, proneness to infection and disease, etc.

Twin studies

Analyzing the concordance rates in twins provides information on the contribution of the genetic and environmental factors causing a disorder. **Concordant** twins share the same condition, and in **discordant** twins, only one twin is affected, affected while the other is generally unaffected. Monozygotic or identical twins (MZ) share the entire genetic material, whereas dizygotic or fraternal twins (DZ) share only an average of 50 % of their genes, such as siblings.

- For traits or conditions 100% genetically determined, concordance rates for MZ twins is 100%, but much lower for DZ twins.
- If a condition is multifactorial, the concordance for MZ twins is higher when compared with DZ twins, but does not reach 100% (as even monozygotic twins exist within different environments!).
- In the event a condition is entirely determined by the environment, the concordance rates for monozygotic and dizygotic twins are generally equal to one another.

The limitation of twin studies is the relative low number of twins found within the general population. In reference to disease, concordant twins are more often ascertained and defined when compared to discordant twins, and monozygotic twins may demonstate an increase in identical environmental factors when compared with dizygotic twins. A potential solution to these dilemas may be in the study of **twins separated at birth**, however the number of such twins found with this characteristic is extremely low.

Adoption studies

To distinguish between the environmental effects and genetic factors causing a disease, adoption studies have been accomplished implementing different strategies:

- Child adopted from affected biological parent (change in environment, genes from affected parent) → study of the incidence of the disorder in child
- Child adopted from family and onset of disease occuring in new family → study of the biological parents and family members
- Child 1 adopted from affected parents to unaffected parents vs. Child 2 adopted from unaffected parents to affected parents → comparison of the incidence in the children

Higher incidence rates in biological relatives compared to unrelated adoptive relatives demonstrate genetic factors involved in the disease. If the third type of studies features a distinctive elevated incidence rate for children adopted to an affected family (Child 2) that fact clearly demonstrates the distinct role and influence of environmental factors.

Oligogenic and polygenic inheritance

Continuous phenotypes

Many human characteristics follow **normal distribution**, like height, skin color, blood pressure and intelligence. If the number of individuals demonstrating a particular quantitative value is plotted, one sees a genuine 'bell-shaped curve' which is characteristic. Normally distributed characteristics are defined as **continuous phenotypes**, which are studied by **quantitative genetics**. The loci contributed to such phenotypes are generally referred to as **quantitative trait loci**.

Oligo- and polygenic inheritance

The concept of **polygenic inheritance** proposes quantitative traits are caused by the additive effect of a variable number of genes (**polygenes**). **Oligogenic inheritance** assumes other quantitative traits are caused by a relatively low number of genes and also allows one of the genes to exert more influence on the phenotype when compared to the others. The locus of such a gene is defined as the **major susceptibility locus**. It is dominant over the oligogenic background, influencing the expression of other genes. This phenomenon is described as **epistasis**.

Heritability

Heritability demonstrates the proportion of total variance caused by additive genetic effects. It is denoted as H and expressed as a fraction of 1 (eg. 0.7) or in percentage. A high H value shows a high genetic determination of the trait.

Multifactorial inheritance

Most multifactorial disorders occur as a discontinuous phenotype (i.e. affected or unaffected), despite the underlying continuous susceptibility. The model explaining this phenomenon is referred to as **'liability-threshold model'**. **Liability** (the susceptibility to the disease, including both genetic and environmental factors) is normally distributed in a population, thus it can be represented as a Gaussian curve. A specific parameter of this curve is the **threshold** on the right side, showing the liability value, beyond which all individuals are affected. For a population exclusively including the relatives of affected individuals, the curve is shifted to the right (because relatives share both genes and environment). Therefore, the number of individuals under the curve, and consequently the potential risk to relatives increases.

The way to find susceptibility genes

Identifying genes contributing to multifactorial disorders is one of the most intensive research areas today. Finding these susceptibility genes can help us understand the underlying biochemical effects and even lead to the further development of tests for early detection or new approaches to prevention or treatment. Unfortunately the progress to date is considerably slower than desired and very little research is useful towards the goals described above. The main strategies are the following:

Linkage analysis

Genetic linkage (see Mendelian inheritance!) analysis involving siblings' genomes, is described as a search for areas shared by affected siblings more often than it would be expected by chance. Target areas can be polymorphic microsatellites or SNPs. This method can effectively detect large segments of genome which contain susceptibility loci. These segments are too large, so the identification of the relevant genes is impossible. Therefore, after this step, one or both the following studies should be performed.

Association studies

When an allele at a particular locus occurs more often within affected individuals when compared with an unaffected control population, it indicates the association of the allele with the disease. Target areas are microsatellites, SNPs, often using microarray technique. An advantage of this approach is that families including only one affected individual can be analyzed as well. A major disadvantage is the association found may be irrelevant. Therefore linkage disequilibrium studies are recommended following this step.

Linkage disequilibrium

The definition of linkage disequilibrium is "the association of specific alleles at linked loci more or less often than it would be expected by chance". Most affected individuals of a population inherit their susceptibility allele from a common ancestor, so if the mutation occurs in the near past, closely linked alleles adjacent to the susceptibility allele will also be identical. In practice, if an approximate location has been identified, this location is analyzed in depth using polymorphic DNA markers (usually SNPs), which map to the region and narrow down the area of interest. With the benefit of the information gained by the Human Genome Project, researchers can locate the susceptibility genes.

Examples of multifactorial disorders

Mainly, multifactorial disorders can be classified based on the onset time: **congenital** diseases are present at birth or onset in early infancy, while **acquired** disorders have an onset in childhood or adult life. Congenital ones are like cardiac defects, cleft lip, Hirschprung disease, neural tube defects, Acquired diseases include Alzheimer disease, coronary artery disease, diabetes mellitus (types 1 and 2) and schizophrenia.

Alzheimer disease

Alzheimer disease (AD) is the most common cause of dementia, with a prevalence of 20% in individuals over 80 years of age. Symptoms are progressive memory loss, emotional disturbance, and loss of intellectual skills. Pathological signs comprise of neurofibrillary tangles of tau protein, that are neurotoxic plus senile plaques comprised of amyloid fibers derived from amyloid precursor protein (APP).

Relative risk for first-degree relatives is 3-4-fold higher compared to controls. Concordance for monozygotic twins (MZ) varies between 30-80%, and 10-40% for dizygotic twins (DZ). Heritability is 0.44-0.8. The data reveal the multifactorial inheritance of the disease.

A small subset of families show pre-senile onset (before 65 years), with an **autosomal dominant** inheritance. A gene first implicated was *APP*, encoding the *amyloid precursor protein*. The mutation of this gene causes an APP production increase, leading to overproduction of amyloid fibres, giving birth to the senile plaques. Another susceptibility gene was identified: *Presenilin-1* (*PS1*) and *Presenilin-2* (*PS2*). As a result, these genes enhance APP processing and the subsequent mutations of these genes benefit APP accumulation.

The inheritance pattern in nearly all families with AD is more characteristic for **polygenic AD**. Most cases occur after the age of 65, and risk to offsprings is much lower than 50%. A polymorphism of *APOE* locus demonstrates association with the late-onset AD. *Apolipoprotein E* (*ApoE*) is involved in lipid metabolism, and has three common alleles: *E2*, *E3* and *E4*. The frequency of *E4* allele is increased in AD patients (for *E4/E4* homozygotes the risk is 12-fold higher than for *E3/E3*). At least 4 other genes are involved as susceptibility genes for late-onset AD.

In brief: Polygenic AD is in association with – amongst others – the APoE gene alleles. The E4 allele confers the highest risk to developing late-onset AD. There is another, early-onset form, with autosomal dominant inheritance, with APP and presenilin genes involved.

Coronary Artery Disease

Coronary artery disease (CAD) is the leading cause of death in the USA. It is six times more common in Americans when compared with Japanese. The case of Japanese immigrants – who triple their original rate of CAD – shows the importance of environmental factors (smoking, diet, lack of exercise).

The heritability of CAD is 0.5–0.6 for **premature CAD**, which is an early-onset form, occurring before 55 years of age. Men are more susceptible than women, MZ concordance rates are 40-65% and DZ rates are 15-30%.

The evolution of atherosclerotic lesions in the coronary artery begins with a fatty streak in the intima of the artery evolving into a fibrous plaque containing smooth muscle, lipid and fibrous tissue. These plaques become vascular and may bleed, ulcerate and calcify, thereby causing severe vessel narrowing and provide ground for thrombosis, which can lead to complete occlusion and myocardial infarction.

CAD is a complication of single-gene disorders. These include hypercholesterolaemia and other forms of hyperlipidaemia. Over 20 genes are candidates for polygenic CAD: genes controlling lipid metabolism, blood pressure, clotting, and fibrinolysis. An insertion/deletion in *ACE* gene, encoding the *angiotensin-converting enzyme*, has been studied in depth. DD homozygotes have higher levels of the enzyme than those with ID or II genotype, and they also show higher incidence of CAD.

Low-cholesterol diet and the use of statins is widely recommended in older men and in those with strong family history. Environmental factors also play a role in influencing the risk of CAD are diet, physical activity and smoking. Susceptibility of CAD is a complex multifactorial condition, having many genetic and environmental factors playing a role in the development or prevention of the disease.

In brief: CAD is a complication of single-gene disorders, including hyperlipidaemias. Over 20 genes are canditates of risk factors, eg. ACE gene, where DD homozygotes show higher risk for CAD. There is a high proportion of environmental factors affecting the onset of the disease, like smoking, diet, physical activity.

Diabetes mellitus

The two primary types of diabetes mellitus (DM) both demonstrate multifactorial inheritance.

Type 1 DM or 'insulin-dependent DM' (IDDM)

IDDM is a relatively rare form, usually occurring in childhood or early adult life as the consequence of **autoimmune** destruction of the pancreas b-cells which produce insulin.

The risk for siblings is 5-6%. MZ concordance is 30-40%, DZ concordance is 5-10%.

Two susceptibility genes are known by time. The **HLA system** accounts for 30-40% of total genetic susceptibility. *IDDM1* locus products *HLA-DR3* and *HLA-DR4* antigens occuring in 95% of the cases, but only 50% in the general population. *IDDM2* locus includes the *insulin gene* plus an upstream region of a microsatellite repeat sequence. A low number of repeats carries susceptibility to IDDM by reducing the expression of the insulin gene in the fetal thymus, thus reducing immunological tolerance to insulin and insulin-producing cells.

Type 2 DM or 'maturity onset DM' or 'non-insulin-dependent DM' (NIDDM)

NIDDM occurs in 5% of adults above 45 years of age. Symptoms are rather the consequences of insulin resistance than relative insulin deficiency. Treatments can include diet or oral hypoglycemic agents.

Risk to first-degree relatives is 10-15% (2-3-times the general population risk), the concordance for MZ twins is 90%, which means the genetic inheritance is strong.

Efforts have been made to find genes which convey susceptibility to NIDDM, however these studies proved unsuccessful.

'Maturity-onset diabetes of the young' (MODY),

MODY is a rare type of DM, with clearly autosomal dominant inheritance.

The mutations causing MODY are transcription factors regulating insulin gene expression or pancreas development: *MODY1* (*hepatocyte nuclear factor-4-a* gene), *MODY2* (*glucokinase* gene), *MODY3* (*hepatic transcription factor-1* gene), and *MODY4* (*islet duodenum homeobox-1* - *IDX1* - gene).

In brief: Type 1 DM or IDDM is a rare form caused by an autoimmune destruction of insulin-producing cells. Two loci have been identified in association to IDDM. The common Type 2 DM or NIDDM is caused by insulin resistance. To date no susceptibility loci could be assigned. The third type of DM, MODY, is an autosomal dominant disease with 4 genes affecting susceptibility.

Hirschprung disease (HSCR)

Hirschprung disease is a congenital multifactorial disease usually present at birth or in early infancy with acute intestinal obstruction and abdominal distension.

The disease is caused by the absence of ganglion cells (nerve-endings) in the distal colon and the rectum, thus the lack of peristalsis in these areas. The long segment type (L-HSCR) is a more severe condition than the short segment type (S-HSCR), as the disease is present also proximal to the sigmoid colon (for L-HSCR). About 30% occurs as part of a multiple malformation syndrome (e.g. Down's Syndrome). The remaining 70% are isolated, non-syndromal cases with multifactorial inheritance.

Research revealed the underlying molecular pathogenesis: the protooncogene **RET** is a major susceptibility locus, with heterozygous mutations exerting a loss-of-function effect. RET encodes a *transmembrane tyrosine-kinase receptor* which moderates cell signaling in the embryonic enteric nervous

system. Other mutations recently identified include RET ligands *GDNF* and *NTN*, the *endothelin* gene (*EDNRB*) and its ligand *EDN3*.

Hirschprung disease is the only multifactorial disease for which the major underlying genetic pathways have been determined. Since relatively few genes are involved, the disease is categorized as oligogenic.

In brief: HSCR is a rare example of a complex disease with well revealed genetic susceptibility. It is characterized by oligogenic inheritance with one major susceptibility locus – the RET gene – and a small number of epistatic loci.

Neural tube defects

The defects under this term occur during early embryogenesis: the neural tube fails to close completely. Serious symptoms can be present, such as anencephaly, occipital encephalocoele and spina bifida. Some of these occur as part of a multiple malformation syndrome, but most of them are isolated, non-syndromal disorders.

The heritability value (H=0.6) shows the strong role of genetic factors. Studies to date revealed only one association with a polymorphism: the *MTHFR* gene involving a C→T mutation shows 2-4-fold increase in the risk (if both the mother and the baby are homozygous). The enzyme encoded by *MTHFR* gene is *methylenetetrahydrofolate reductase*. Homozygotes for the upper mentioned mutation is associated with reduced red cell folate levels.

Periconceptional folic acid supplementation can prevent 80% of all cases. This is a nice example for a mainly genetically disorder to be treated by altering an environmental factor, in this case nutrition.

In brief: Neural tube defects are congenital malformations with polygenic inheritance. Only one association could be revealed: the mutation of the MTHFR gene confers high risk for these defects. However, as an elegant prevention, periconceptional folic acid supplementation – a change in an environmental factor! – can prevent 80% of the cases.

Schizophrenia

Schizophrenia is a mental illness affecting 1% of the world’s population. It is characterized by an onset in late adolescence or young adulthood, abnormalities in thought, emotion and social relationships, often with delusional thinking and disordered mood.

Strong evidence based on family and twin studies demonstrates the major role of genetic factors causing schizophrenia. First-degree relatives have a risk of 10-12%, this number is 2-4% for second degree relatives and 2% for third-degree relatives. Concordance for MZ twins is 50-70% and 10-15% for DZ twins. Heritability falls between 0.8 and 0.85.

Two candidates to susceptibility genes are *DTNBP1* and *NRG1*. *DTNBP1* encodes the protein *dysbindin*, which plays a role in synaptic function in the brain. *NRG1* encodes *neuroreglin 1*, a protein inducing myelination. These genes have been identified by independent linkage and association/linkage disequilibrium studies, but no specific pathogenic mutations could be revealed.

In brief: A few genes carrying risk for this mental illness – dysbindin and neuroreglin 1 – have been assigned by independent studies.

Susceptibility testing for multifactorial disorders

Predictions based on genomic tests

The fact that some susceptibility loci have been identified, can make one think that susceptibility testing for common diseases is now possible and may offer offer a new approach toward prevention and/or therapy. Several corporations in the USA and UK offer **'predictive genomic profiling'** for susceptibility genes of CAD, asthma, metabolism of toxins, etc. Clients cen get information about their lipid- or toxin-metabolism, and can get guidance in living a healthy, full life.

However, the interpretation of the results may be overly optimistic. For example, consider the association of *ApoE4* with Alzheimer disease. The risk for AD for an *E4*/*E4* homozygote is 35% for men and 50% for women, while those having no *E4* allele at all, these numbers are 4.5% and 9.3%. The risk corresponds to a late onset of the disease, but it is impossible to predict the time when the symptoms will start. Another big concern about the prediction is the statistical data demonstrating that 50% of all *E4*/*E4* homozygotes will never develop the disease, and 5-10% of the population develop the disease without having any *E4* alleles. These data decrease in the accuracy and reliability of such tests.

Empiric recurrence risk

In multifactorial diseases, it is typically not possible to derive recurrence risks based on theory. If there is an inheritance determined by a normally distributed liability, then risks can be derived based on statistics, knowing the normal population incidence. Instead of theoretical risk, the **empiric risks** are utilized for counselling clients exhibiting multifactorial disorders. Empiric risks are based on observation (incidence patterns in relatives obtained from family studies).

23.8. PEDIGREE ANALYSIS

Introduction

Today, thanks to enormous strides in research, we know more about human anatomy, physiology, biochemistry and molecular biology. For many families, we have detailed records and information throughout the generations, and the genetic background of human diseases emphasizes the importance of pedigree analysis.

The study of human genetic characteristics presents some major obstacles:

1. Controlled matings are not possible: specific crosses cannot be achieved otherwise deemed crucial in the determination of an individual featuring a dominant trait, whether homozygous or hetrozygous.
2. Specifically, humans have a long generation time (about 20 years): in the event science could control human crosses, research likely will require an average of 40 years to observe with clarity, the F2 progeny.
3. Human family size is generally small: in a family having only 2 children, it is impossible to detect Mendelian 3:1 ratio. Even an extremely large family with 10 to 15 children would not permit the recognition of a dihybrid 9:3:3:1 ratio.
4. Failure to truthfully identify parentage

Due to the above mentioned reasons humans are the best and the worst of all organisms, in reference to genetic study.

Distinctively, an important technique to study human inheritance is the pedigree. A pedigree is a pictorial representation of a family history, a family tree that outlining the inheritance of one or more characteristics.

Based on pedigree analysis, the determination can be made in reference to the mode of inheritance: autosomal recessive, autosomal dominant, sex-linked recessive, sex-linked dominant, mitochondrial, maternal effect. Importantly, risk of disease can be, to some degree of accuracy, predictable and useful in examining the future offspring within a given family. This is commonly referred today as Genetic Counselling.

Symbols

(see figures from seminar)

1. Males in a pedigree are represented by squares, females by circles
2. A horizontal line drawn between two symbols representing a man and a woman indicates a mating
3. Children are connected to their parents by vertical lines extending below the parents
4. Persons who exhibit the trait of interest are represented by filled circles and squares
5. Unaffected persons are represented by open circles and squares.
6. Each generation in a pedigree is identified by a Roman number
7. Within each generation family members are assigned Arabic numbers
8. Children in each family are listed in birth order from left to right
9. Deceased family members are indicated by a slash through the circle or square
10. Twins are represented by diagonal lines extending from a common point
11. The person from whom the pedigree is initiated is identified as the **proband** and is usually designated by an arrow

Autosomal recessive traits

(see figure from seminar)

Autosomal recessive traits normally appear with equal frequency in both sexes

They appear only when a person inherits two alleles for the trait

If the trait is uncommon, most parents carrying the allele are heterozygous and unaffected

The trait appears to skip generations: a recessive allele may be passed for a number of generations without the trait appearing in a pedigree

If the parents are heterozygous, approximately 25% of the offspring are expected to express the trait

If both parents are affected by an autosomal recessive trait (homozygous recessive), all the offspring will be affected

Persons from outside the family are usually homozygous for the normal allele: once an affected person mates with someone outside the family (AA), often the result is none of the offspring are expected to express the trait, but all will become carriers (Aa).

A recessive trait is more likely to appear in a pedigree when consanguinity occurs (mating between closely related people)

Autosomal recessive diseases

Tay-Sachs disease (see figure from seminar):

- children appear healthy at birth but become listless and weak at about 6 months of age

- their physical and neurological conditions, increase in severity leading to blindness, deafness, and eventually death at 2 to 3 years of age.
- cause of the disease is the accumulation of a lipid called GM2 ganglioside in the brain
- GM2 ganglioside cannot be destroyed destroyed due to the omission of an enzyme known as hexosaminidase A
- excessive amount of GM2 ganglioside causes swelling and neurological symptoms
- Heterozygotes are healthy due to one normal copy of the hexosaminidase A allele is enough for producing about half the normal amount of the enzyme, which maintains normal break down of GM2 ganglioside
- Phenylketonuria (see figure form seminar):
- an autosomal recessive disease resulting from a defect in an enzyme typically metabolizing the amino acid phenylalanine.
- without the enzyme high level of phenylalanine is developed and causes
- brain damage
- children born to women with PKU (who are not on a phenylalanine-restricted diet) frequently exhibit a low birth weight, developmental abnormalities, and mental retardation

Autosomal dominant traits

(see figure from seminar)

1. Autosomal dominant traits appear in both sexes with equal frequency
2. Both sexes are capable of transmitting these traits to their offspring
3. Every person with a dominant trait inherits the allele from at least one parent
4. Autosomal dominant traits do not skip generations
5. Exceptions: trait is a result of new mutation
6. If an autosomal dominant allele is rare, affected persons are usually heterozygous
7. If one parent is affected and heterozygous (Aa) and the other parent is unaffected (aa), approximately 50% of the offspring will be affected
8. If both parents are affected and heterozygous (Aa), approximately 75 % of the children will be affected

Autosomal dominant diseases

Familial hypercholesterolemia:

- blood cholesterol is highly elevated due to a defect in cholesterol transport
- cholesterol is a lipid (a nonpolar, or uncharged, compound), it is not readily soluble in the
- blood (a polar, or charged, solution), therefore it must be transported throughout the body in small soluble particles called lipoproteins (LDL, HDL, VLDL, IDL, chylomicrons)
- cause of the disease is the defect in the gene encodeing the LDL receptor
- result: too little cholesterol is removed from the blood, leading to increased risk of coronary artery disease (heart attack)
- heterozygous: blood LDL level is twice as normal
- homozygous dominant: two defective LDL receptor alleles, no functional LDL receptor, cholestereol level is six times higher than normal, may suffer a heart attack as early as age 2
- Huntington's disease (see figure from seminar):
- symptoms: jerky, involuntary movements, subtle, consisting of mild behavioral and neurological changes; as the disease progresses, speech is impaired, walking becomes difficult, and psychiatric problems develop often leading to insanity
- the condition typically appears in middle age
- patients with Huntington's disease live for 10 to 30 years from the onset of the disease.
- Huntington's disease appears with equal frequency in males and females

- it rarely skips generations
- if one parent has the disorder, approximately half of the children will be affected
- Achondroplasia (see figure from seminar):
- it causes diminished growth in the long bones of the legs, leading to dwarfism
- If two persons who heterozygous for achondroplasia marry, chances are that 50% of their children will also be heterozygous and dwarfs
- one child in every four born to the couple will not inherit the achondroplasia gene and will be of average height, and one child in four will be homozygous for the gene
- homozygosity for this gene is lethal

X-linked recessive traits

(see figure from seminar)

1. These traits appear more frequently in males
2. Males inherit only a single copy of the allele to display the trait,
3. Females must inherit two copies of the allele, one from each parent, to be affected.
4. A male inherits his X chromosome from his mother
5. Affected males are usually born to unaffected mothers who are heterozygous.
6. It tends to skip generations: the trait is passed from unaffected female to affected male to unaffected female
7. When a woman is heterozygous, approximately 50% of her sons will be affected and of her daughters will be unaffected carriers
8. They are not passed from father to son, because a son inherits his father's Y chromosome, not his X.
9. All daughters of an affected man will be carriers
10. When a woman displays an X-linked trait, she must be homozygous for the trait, and all of her sons will also display the trait

X-linked recessive disease

Hemophilia A (see figure from seminar):

- this disease results from the absence of a protein necessary for blood to clot, factor VIII
- the gene for factor VIII is located on the tip of the long arm of the X chromosome
- symptom: excessive bleeding
- spontaneous bleeding occurs in joints which produces pain, swelling, and erosion of the bone
- treatment: administration of factor VIII

X-linked dominant traits

(see figure from seminar)

1. X-linked dominant traits appear in males and females, although they often affect more females than males
2. A male inherits an X-linked dominant trait only from his mother
3. The trait is not passed from father to son
4. A female gets an X chromosome from both her mother and father, so females can receive an X-linked trait from both parents.
5. Each child with an X-linked dominant trait must have an affected parent
6. X-linked dominant traits do not skip generations
7. Affected men pass the trait on to all their daughters and none of their sons

8. Affected women (if heterozygous) pass the trait on to half of their sons and of their daughters

X-linked dominant disease

Hypophosphatemia/familial vitamin D-resistant rickets:

- symptoms: bone deformities, stiff spines and joints, bowed legs, and mild growth deficiencies.
- it results from the defective transport of phosphate
- people with this disorder excrete large amounts of phosphate in their urine, resulting in low levels of phosphate in the blood and reduced deposition of minerals in the bone
- males with hypophosphatemia are often more severely affected than females

Y-linked traits

(see figure from seminar)

1. Only males are affected, and the trait is passed from father to son
2. If a man is affected, all his male offspring should also be affected
3. Y-linked traits do not skip generations
4. Dominance is irrelevant: there is only 1 copy of each Y-linked gene (hemizygous).
5. Few human traits are Y linked: *hairy ears* (hair on the ear pinna)
6. Y-linked disorder causes infertility (*Male Infertility Disorder*)
7. Y-linked disease: *retinitis pigmentosa* which is the most common form of inherited retinal degeneration (inheritance also can be autosomal recessive, autosomal dominant, X-linked and maternally acquired)

Mitochondrial inheritance

(see figure from seminar)

1. Mitochondria are only inherited from the mother
2. If a female has a mitochondrial trait, all of her offspring inherit it
3. If a male has a mitochondrial trait, none of his offspring inherit it
4. Only 1 allele is present in each individual, so dominance is not a consideration
5. An example for mitochondrial disease: *MELAS syndrome* (Mitochondrial Enchephalomyopathy, Lactic Acidosis and Stoke-like episodes) affects the brain and nervous system and muscles; Its symptoms are muscle weakness, pain, stroke-like episodes, brain damage, vision problems, dementia, lactic acidosis, ataxia.

Maternal effect

(see figure from seminar)

1. The phenotype of the offspring is determined by the genotype of the mother: the genes are inherited from both parents, but the offspring's phenotype is not determined by its own genotype but by the genotype of its mother
2. In cytoplasmic inheritance, the genes for a characteristic are inherited from only one parent, usually the mother
3. Genetic maternal effect occurs when substances present in the cytoplasm of an egg are essential in early development

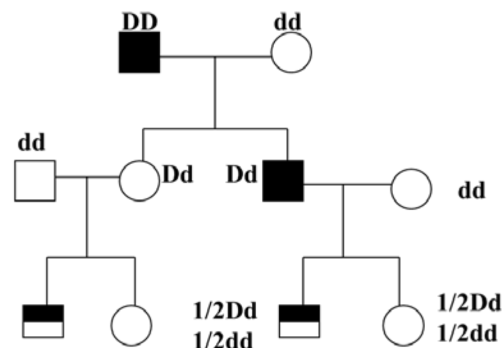
4. The phenotype of the progeny is not necessarily the same as the phenotype of the mother, 5. The progeny's phenotype is determined by the mother's genotype, not her phenotype
5. Neither the male parent's nor the offspring's own genotype has any role in the offspring's phenotype
6. Genes that exhibit genetic maternal effect are therefore transmitted through males to future generations.

Sex-influenced characteristics

1. Sex influenced characteristics are determined by autosomal genes and are inherited according to Mendel's principles
2. Autosomal genes are expressed differently in males and females.
3. A particular trait is more readily expressed in one sex

Baldness – a sex-influenced trait

- hair is lost prematurely from the front and the top of the head
- pattern baldness is an autosomal character believed to be dominant in males and recessive in females
- men and women can inherit baldness from either their mothers or their fathers.
- men require only a single allele for baldness to become bald
- women require two alleles for baldness
- pattern baldness is expressed weakly in women
- the expression of the allele for pattern baldness is enhanced by the presence of male sex hormones



Sex-limited characteristics

1. A sex-limited characteristic is encoded by autosomal genes that are expressed in only one sex
2. The trait has zero occurrences in the other sex
3. A sex-limited characteristic can be inherited from either parent

Male-limited precocious puberty – a sex-limited trait

- it results from an autosomal dominant allele that is expressed only in males
- females with the gene are normal in phenotype.
- males undergo puberty at an early age, usually before the age of 4
- affected males are fully fertile
- the trait is rare, affected males are usually heterozygous

- a male with precocious puberty who mates with an unaffected woman will transmit the allele for precocious puberty to 50% of the children but it will be expressed only in the sons
- if the female is heterozygous and mates with a male who has normal puberty, 50% of the sons will exhibit precocious puberty
- disease affects the receptor for luteinizing hormone
- during normal puberty in males, high levels of LH stimulate the production of testosterone which stimulates the anatomical and physiological changes
- defective LH receptor stimulates testosterone production even in the absence of LH
- boys with this allele produce high levels of testosterone at an early age
- defective LH receptors are also found in females who carry the precocious-puberty gene, but their presence does not result in precocious puberty, because FSH and estrogens are also required along with LH.

Dd dd

dd

dd

1/2Dd
1/2dd
All normal

1/2Dd (affected)
1/2dd (normal)

23.9. POPULATION GENETICS

Introduction

Population genetics studies the genetic makeup of groups of individuals and focuses with distinct changes within their *genetic composition*. Population geneticists generally focus their attention on an *ideal population*, which is a group of interbreeding, sexually reproducing individuals featuring a common set of genes, the *gene pool*. Hence, the gene pool is best defined as the entirety of the alleles of the individuals within a population.

As a population evolves by changes in its gene pool, population genetics, therefore, consequently examines *evolution*. Population genetics also studies the *variations in alleles* within and between groups and the evolutionary forces responsible for determining the *patterns* of genetic variations.

Inheritance influencing factors

Inheritance pattern of a population can be influenced by several factors:

1. Size of the population: the population is large enough not to exhibit the effects of a random loss of genes through chance events e.g., there is **no genetic drift.**
2. Mutations: the mutation rate at the locus of the gene being studied is not significantly high.
3. Mating between individuals is random (**panmictic population**).
4. New individuals are not gained by **immigration** or lost by **emigration**
   e.g., there is **no gene flow** between neighbouring populations of same species
5. Neighbouring populations: the gene's allele has **no selective advantage or disadvantage.**

The genetic drift, mutation, mating choice, migration and natural selection together or alone can affect the transmission of genes from generation to generation. When none of these factors is operational, the relative proportions of the alleles (**gene frequencies**) will be constant. Therefore the population is in *genetic equilibrium*.

Genetic variation

Genetic variability is maintained by three mechanisms:

1. random fertilization
2. random segregation of the chromosomes
3. crossing over/recombination

Genetic variations may exhibit phenotypic variations. Many variations exists at the molecular level owing to the redundancy of the genetic code, which allows different codons to specify the same amino acids. Thus, two individuals may produce the same protein even if their DNA sequences are different. Genetic variation is the basis of evolution, and the extent of genetic variation within a population affects its potential to adapt to environmental change.

Genotypic frequencies

(see figure from lecture)

A frequency is simply a proportion or a percentage, usually expressed as a decimal fraction. The genotypic and allelic frequencies of the population are used to represent the gene pool of the population. To calculate a genotypic frequency, we add up the number of individuals possessing the genotype and divide by the total number of individuals in the sample (N). For a locus with three genotypes AA, Aa, and aa, the frequency (f) of each genotype is:

$$f(AA) = \frac{\text{number of } AA \text{ individuals}}{N}$$

$$f(Aa) = \frac{\text{number of } Aa \text{ individuals}}{N}$$

$$f(aa) = \frac{\text{number of } aa \text{ individuals}}{N}$$

The sum of all the genotypic frequencies always equals 1.

Allelic frequencies

(see figure from lecture)

The gene pool of a population can also be described in terms of the allelic frequencies. There are always fewer alleles than genotypes. The genotypes are temporary assemblages of the alleles. The individual alleles are passed to the next generation through the gametes.

Allelic frequencies can be calculated from the numbers or the frequencies of the genotypes. To calculate the allelic frequency from the numbers of genotypes, we count the number of copies of a particular allele present in a sample and divide by the total number of all alleles in the population:

$$\text{frequency of an allele} = \frac{\text{number of copies of the allele}}{\text{number of copies of all alleles at the locus}}$$

For a locus with two alleles (A and a), the frequencies of the alleles are represented by the symbols p and q (see later), and can be calculated as follows:

$$p = f(A) = \frac{2n_{AA} + n_{Aa}}{2N}$$

$$q = f(a) = \frac{2n_{aa} + n_{Aa}}{2N}$$

Symbols: n_{AA}, n_{Aa}, and n_{aa} represent the numbers of AA, Aa, and aa individuals, and N represents the total number of individuals in the population. We divide by 2N because each diploid individual has two alleles at a locus.

When determining the frequencies of alleles for loci featuring more than two alleles, the allelic frequencies are calculated from the number of genotypes. It is essential to total the number of copies of an allele by adding twice the number of homozygotes to the number of heterozygotes possessing the allele, then dividing this sum by twice the number of individuals within the population.

The three alleles are A^1, A^2 and A^3, from which the possible six genotypes are A^1A^1, A^1A^2, A^1A^3, A^2A^2, A^2A^3, A^3A^3.

$$p = f(A^1) = \frac{2n_{A^1A^1} + n_{A^1A^2} + n_{A^1A^3}}{2N}$$

$$q = f(A^2) = \frac{2n_{A^2A^2} + n_{A^1A^2} + n_{A^2A^3}}{2N}$$

$$r = f(A^3) = \frac{2n_{A^3A^3} + n_{A^1A^3} + n_{A^2A^3}}{2N}$$

Once the allelic frequencies of an *X-linked gene* are effectively calculated, the same principles are applied. However, it is imperative to note a female possesses two X chromosomes, whereas a male has only a single X chromosome. Females can be either homozygous (X^AX^A or X^aX^a) or heterozygous (X^AX^a). All males are hemizygous (X^AY or X^aY). To determine the frequency of the X^A allele (p), we first count the number of copies of X^A: we multiply the number of X^AX^A females by two and add the number of X^AX^a females and the number of X^AY males. Next, the sum is divided by the total number of alleles at the locus, which is twice the total number of females plus the number of males.

$$p = f(X^A) = \frac{2n_{X^AX^A} + n_{X^AX^a} + n_{X^AY}}{2n_{\text{females}} + n_{\text{males}}}$$

$$q = f(X^a) = \frac{2n_{X^aX^a} + n_{X^AX^a} + n_{X^aY}}{2n_{\text{females}} + n_{\text{males}}}$$

The Hardy-Weinberg Law

(see examples from lecture)

The Hardy-Weinberg law is one of the most important principles of population genetics. The law was formulated independently by both Godfrey H. Hardy and Wilhelm Weinberg in 1908. The law is a mathematical model evaluating the effect of reproduction on the genotypic and allelic frequencies of a population. For an autosomal locus with two alleles, the Hardy-

Weinberg law highlights the following statements:

1. If a population is large, randomly mating, and not affected by mutation, migration, or natural selection.
2. The allelic frequencies of a population do not change.
3. The genotypic frequencies stabilize after one generation. It means they can change in the first generation after random mating, because one generation of random mating is required to produce Hardy-Weinberg proportions of the genotypes. Afterward, the genotypic frequencies, like allelic frequencies, do not change as long as the population continues to meet the assumptions of the Hardy-Weinberg law.

The Hardy-Weinberg law indicates when the assumptions are met, reproduction alone does not alter allelic or genotypic frequencies and the allelic frequencies determine the frequencies of genotypes.

Of what we should remember:

- p = frequency of A
- q = frequency of a
- p = AA + ½ Aa = frequency of A
- q = aa + ½ Aa = frequency of a
- $p + q = 1$
- We get: $q = 1-p$ and $p = 1-q$
- 2pq = heterozygous genotype frequency
- p^2 = homozygous dominant genotype frequency
- q^2 = homozygous recessive genotype frequency
- $p^2 + 2pq + q^2 = 1$

When genotypes are in the expected proportions of p^2, 2pq, and q^2, the population is said to be in **Hardy-Weinberg equilibrium**. When a population is in Hardy-Weinberg equilibrium, the genotypic frequencies are determined by the allelic frequencies.

The Hardy-Weinberg principle is only can be utilized if

- the organisms are diploid
- only sexual reproduction occurs
- generations are non overlapping
- mating is random
- population size is infinitely large
- allele frequencies are equal in males and females
- there is no migration, mutation or selection

The Hardy-Weinberg principle is able to calculate the gene frequencies from the genotype frequencies. It uses the calculated gene frequency to predict the expected genotypic frequencies in the next generation. It is able to verify the present population is in genetic equilibrium. We can use it to determine carrier frequencies (autosomal recessive and X-linked recessive inheritance).

Extensions of the Hardy-Weinberg law

(see examples from lecture)

The Hardy-Weinberg law can also be applied to multiple alleles and X-linked alleles. In general, the genotypic frequencies expected at equilibrium are the square of the allelic frequencies. We can also use the square of the allelic frequencies to calculate the equilibrium frequencies for a locus with multiple alleles. An autosomal locus with three alleles, A^1, A^2, and A^3, has six genotypes: A^1A^1, A^1A^2, A^1A^3, A^2A^2, A^2A^3, and A^3A^3. According to the Hardy-Weinberg law, the frequencies of the genotypes at equilibrium depend on the frequencies of the alleles. If the frequencies of alleles A^1, A^2, and A^3 are p, q, and r, respectively, then the equilibrium genotypic frequencies will be the square of the allelic frequencies:

$$p^2 = f(A^1A^1)$$

$$2pq = f(A^1A^2)$$

$$q^2 = f(A^2A^2)$$

$$2pr = f(A^1A^3)$$

$$2qr = f(A^2A^3)$$

$$r^2 = f(A^3A^3)$$

For an X-linked locus with two alleles, X^A and X^a, there are five possible genotypes: X^AX^A, X^AX^a, X^aX^a, X^AY, and X^aY. Females possess two X-linked alleles, thus the expected proportions of the female genotypes can be calculated by using the square of the allelic frequencies. The frequencies of X^A and X^a are p and q. The equilibrium frequencies of the female genotypes are $(p+q)^2 = p^2$

Males have only a single X-linked allele, and so the frequencies of the male genotypes are p (frequency of X^AY) and q (frequency of X^aY). The frequency of an X-linked recessive trait among males is q, whereas the frequency among females is q^2. When an X-linked allele is uncommon, therefore the trait will be much more frequent in males than in females.

Nonrandom mating

One of the assumptions of the Hardy-Weinberg law is that the mating is random. Nonrandom mating affects the genotype frequencies. Nonrandom matings feature two types:

1. Positive assortative mating: refers to a tendency for like individuals to mate e.g., tall people mate preferentially with other tall people; short people mate preferentially with other short people.
2. Negative assortative mating: refers to a tendency for unlike individuals to mate e.g., tall and short people generally prefer to mate with one another.

One form of nonrandom mating – inbreeding

1. mating between related individuals
2. it affects all genes
3. it causes a departure from the Hardy-Weinberg equilibrium frequencies
4. it leads to an increase in the proportion of homozygotes and a decrease in the proportion of heterozygotes in a population

5. it is determined by the inbreeding coefficient (F= 0 to 1)

When inbreeding occurs, the frequency of the genotypes will be:

$$f(AA) = p^2 + Fpq$$

$$f(Aa) = 2pq - 2Fpq$$

$$f(aa) = q^2 + Fpq$$

Natural selection

Natural selection takes place when the populations produce more offspring, than environmental resources are capable of maintaining. Thus, they must compete for survival. If the adaptive traits have a genetic basis, they are inherited by the offspring and appear with greater frequency in the next generation. A trait providing a reproductive advantage thereby increases with time, enabling populations to better adapt to their environments. Natural selection promotes adaptation. Individuals exhibiting the best adaptations survive and successfully reproduce.

Environmental factors act as selecting agents (biotic and abiotic). The effect of natural selection on the gene pool of a population depends on the **fitness (W)** values of the genotypes within the population. Fitness is defined as the relative reproductive success of a genotype.

- A fundamental concept in evolutionary theory is “fitness”: the ability to survive and reproduce.
- Reproduction is key: to be evolutionarily fit, an organism must pass its genes on to future generations.
- Basic idea: the more fit individuals contribute more to future generations than less fit individuals.
- the genes found in more fit individuals ultimately take over the population.

Differential fitness among genotypes over time leads to changes in the frequencies of the genotypes, which, in turn, leads to changes in the frequencies of the alleles making up the genotypes. We can calculate the **selection coefficient** (s), which is the relative intensity of selection against a genotype, from fitness (W). The selection coefficient is equal to 1 - W.

Types of selection:

1. Selection against a recessive allele: heterozygotes still hold the recessive allele, selection is a long process
2. Selection against a dominant allele: next generation does not show the dominant phenotype
3. Selection against an incompletely dominant allele
4. Overdominance
5. Underdominance

Two types of selection (types 4 and 5) are special situations leading to equilibrium, where there is no further change in allelic frequency. With overdominance, both alleles are favored in the heterozygote, and neither allele is eliminated from the population. With underdominance the heterozygote has lower fitness than both homozygotes. Underdominance leads to an unstable equilibrium.

Summary: Natural selection changes allelic frequencies; the direction and magnitude of change depends on the intensity of selection, the dominance relations of the alleles, and the allelic frequencies. Directional selection favors one allele over another and eventually leads to fixation of the favored al-

lele. Overdominance leads to a stable equilibrium with maintenance of both alleles in the population. Underdominance produces an unstable equilibrium due to the lower fitness within the heterozygote compared with those of the two homozygotes.

Mutation

All genetic variants ultimately arise through mutation. Mutation is the source of new alleles in the population. Usually it converts one allelic form of a gene to another. It may be neutral, detrimental or advantageous, depending on the environment. Alleles resulting from unfavourable mutations are selected against, and only remain in the gene pool if they are recessive (remain „hidden" in the heterozygotes). The mutations affect on allelic frequencies. They can *influence the rate* at which one genetic variant increases at the expense of another. When the only evolutionary force acting upon a population is mutation, allelic frequencies change with the passage of time due to some alleles mutating into others. Eventually, these allelic frequencies reach equilibrium and are determined only by the forward and reverse mutation rates. The mutation rates for most genes are low; so change in allelic frequency due to mutation in one generation is very small, and long periods of time are required for a population to reach mutational equilibrium. Thus, the effect of typical mutation rates on Hardy-Weinberg equilibrium is negligible.

Migration

Migration is defined as the movement of individuals from one population to another. It has two types:

1. Immigration: individuals migrate *into* the population
2. Emmigration: individuals migrate *out* of the population

Genetic migration is about gene movement, rather than actual movement of organisms, and is referred to as *gene flow*. Gene flow or migration may cause changes in the allelic frequency of a population by introducing alleles from other populations. The magnitude of change due to migration depends on both the extent of migration and the difference in allelic frequencies between the source and the recipient populations.

The overall effect of migration is twofold:

1. It prevents genetic divergence between populations.
2. It increases genetic variation within populations.

If populations are large, migration may have little or no effect on allele frequency. If populations are small, migration may have a big impact on allele frequency. Migration increases the effective size of populations, and may prevent allelic fixation.

Genetic drift

Genetic drift refers to *changes in gene frequency due to random events* occurring from generation to generation. When the population is large, mating is random and the environment is stable, the frequency of alleles remains stable from generation to generation. Distinctively, in the event of an effective population size, the number of adults (including the number of breeding females plus the number of breeding males) and their gametes contribute to the next generation.

One kind of deviation from an expected ratio (gene frequencies) due to limited sample size is referred to as *sampling error.* Sampling error occurs when gametes unite to produce progeny. When population is diminished in size, a limited number of gametes unite to produce the individuals of the next generation. Chance influences which alleles are present in this limited population and, in this way, sampling error may lead to changes in allelic frequency, which is called genetic drift. Once the population is

decreased, allele frequencies often increase, decrease, or completely lost. When there are few copies of an allele, the effect of genetic drift is larger, and when there are many copies the effect is smaller. The genetic drift is influenced by the sex ratio: half of the genes in the gene pool come from males and half come from females. Other factors influencing effective population size include variation between individuals in reproductive success, fluctuations in population size, the age structure of the population, and whether mating is random.

Founder effect

(see figure from lecture)

- Occurs when a small group of individuals (*founder population*) colonizes a geographically isolated area.
- When a small group in a population splinters off from the original population and forms a new one.
- There is a reduced genetic variation within the population.
- It may lead to allele fixation and/or speciation.
- An example: Amish migration to Pennsylvania: Ellis-van Creveld syndrome.

Bottleneck effect

(see figure from lecture)

- A drastic reduction occurs in a population (volcanoes, earthquakes, landslides etc.).
- The range of alleles decreases, and the frequency of alleles change.
- There is a reduced genetic variation.
- Smaller population may not be able to adapt to new selection pressures.

23.10. DEVELOPMENTAL GENETICS

Note: The multicellularity of plants developed separately from the animal kingdom, and during this session you will learn about the developmental mechanisms of animals.

To understand this section, lecture slides are essential.

History

Developmental biology evolved from embryology. About 1894 Wilhelm Roux created the experimental embryology (see lecture slides) and testified there are great parallels between ontogeny and evolution. He believed science will reveal parallels over time. He said the basis of the evolution is the heritable changes in the organism's development. The final shape of morphological structures is formed by growing, so we can understand the morphological evolution with the evolution of development. This is very similar to what Darwin testified in reference to evolution, „descent with modification”.

Unity of type

There was discovered long ago, meaning, there are certain structures of organism which are very similar. Étienne Geoffroy Saint-Hilaire defined it in such terms, "God created organism based on a plan and the adaptations are secondary” - morphological evolution describes these secondary adaptations. Unity of type theory describe homologies and emphasizes there is a basic body plan modified by some regulatory elements of the certain organisms. For example, pentadactyl limbs, such as the fingers of a human, flipper of a seal, or wing if a bird (see lecture slides).

Carl Ernst von Baer (1791-1876) compared embryos and observed very similar stages formed during development by animals which are relatives (belong to the same strain). In this early stage of the embryonic development (after gastrulation and neurulation), the embryos of fishes, amphibians, reptiles birds and mammals enter a very similar embryonic stage but as time goes by the developmental pathways separate and embryos became increasingly recognizable where they belong (order, family, strain). He created laws according to his observations. Notably, these laws continue to have significant impact.

Laws

“Law #1 - the more general characters of a large group appear earlier in the embryo than compared with the more special characters.”

What are these common characters by mammals and humans? The neural tube on the dorsal side of the organism, notochord, foregut with gills gap, heart on the ventral side, eye vesicles, enclosed nervous system (which is widened at the beginning, where the brain will develop). These parts develop and appear earlier during the development than the characteristics that are considerably different, for example the limbs, plume, etc.

“Law #2 - From the most general forms the less general are developed, and so on, until finally the most special arises.”

“Law #3 - Every embryo of a given animal form, instead of passing through the other forms, rather becomes separated from them.”

“Law #4 - Fundamentally, therefore, the embryo of a higher form never resembles any other form, but only its embryo.”

Summary of von Baer's laws: “More general structures, that are common to all members of a group of animals evolved earlier when compared to the specific traits distinguishing the various parts of the group."

In 1983, Klaus Sander identified this stage/these stages as the phylotypic stage. He observed this particularly unique stage of the development in which the different classes of vertebrates are very similar and have very similar shape. In vertebrates, we know this as the phylotypic stage and/or the Pharyngula stage.

The preceding stages have significant differences even among close relatives and the progress flows into a very similar stage. Von Baer's laws are valid only if the starting point is the phylotypic stage.

Haeckel biogenetical law,

„ontogeny, briefly recapitulates phylogeny”. This does not mean ontogeny recapitulates the phenotype of the embryo’s predecessors, but rather the developmental history of its predecessors with modifications and is notably truncated (see von Baer’s laws). This law speaks describes the “terminal addition”, too. Terminal addition in developmental biology means that during a strain evolution, new sections of the development are added to the old ones and the phenotype differs in this way. Distinctively, we can see how old characters develop and afterwards, how these then undergo further development, e.g., caudal fin of flatfishes: during the development first a dyphicercal tail develops and is characteristic to lungfishes, then a heterocercal tail appears which is characteristic to sturgeons and lastly a homocercal tail develops.

Body plan

We know most of our genes (more thousands) create the general body plan which makes the organism viable. These genes encoded cellular mechanisms and most of these genes do not differ between the different strains. These genes cannot tolerate mutations so good, effective mutations of these genes result only several times in association to a viable organism.

In all major groups of organisms, ("monophyletic taxa") can be recognized by a general body plan, which establishes the arrangement of the body parts and organs, regardless of the environment.

Toolkit genes

Instead of the previously described genes, we know only a few dozen genes (under 100) which influence the development, mutations of these genes change the phenotype, can create new forms and develop new functionality.

These genes specifically affect the development and influence the timing and rate of growth of certain parts. Only a few genes influence the expression of body plan encoding genes, and this creates the developmental pattern. These genes are called toolkit genes.

Following the discovery of these genes, evolution was reinterpreted meaning only a few genes can affect the development, moreover, from the same material the potential exists in the creation of a different individual organism. With this knowledge, we can easily explain the evolution, the transitions, because significantly, less change is sufficient to create big changes in the phenotype. These toolkit genes provide the base of the morphological evolution.

Toolkit genes are omnipotent and conserved, we can find them in everywhere in the multicellular animal kingdom and they can play the same or very similar role, for example pax6/eyeless gen, which induces the expression of the eye in both Drosophila and the mouse.

These genes determine and control the expression of the general body plan. For example, snakes have genes that encode the ability to grow limbs but the limb encoding genes (Distal-less genes) are not expressed or expressed in a low level. Only the chest encoding genes are expressed along the body, meaning an all-along chest organism will be created by the toolkit genes.

Non-coding sections

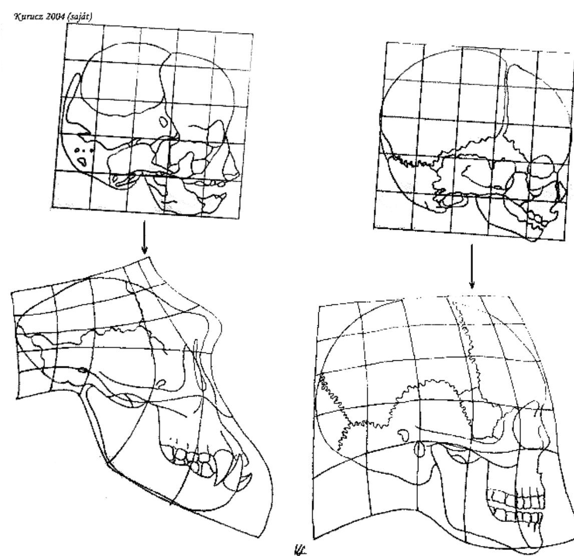
If we compare the protein coding regions of the genomes, we can see only small differences between the most distant relatives, but if we look at the non-coding, regulatory sections, that are associated to each gene, the differences become immense. Because these sections tolerate mutations more when compared to protein encoding regions, there are more differences in these sections of the genome.

These regulatory sections (for example enhancers and silencers) have a significant impact in the development to regulate the gene expression pattern evolving the final morphology. These sections can create radically different animals from the same set of genes.

Allometric growing

If the growth rate of the parts change or are altered, such as accelerate or slow down, the shape can change, and in this regard, we define this as allometric growth.

You can see on the picture how the skull of an ape and the skull of a human develop. The structures are the same, the body plan is almost the same, small changes create big differences in the resulted phenotype. Allometric growth can lead to drastic change in body proportions.



In ant colonies, allometric growth of ants determine the phenotype of the workers, the soldiers and the queen in relationship to environmental factors. Allometric growth creates most of the mouth organs found in insects.

With toolkit genes and regulatory elements, the evolution changes a basic body plan and for the changes, evolution only requires the change of a few genes and small sections only. This is a very economical way to design novel parts or entirely new functions for existing parts. The evolution tinkers, vary existing structures and toolkit genes are the primary resources in support of this.

Several examples follow:

Middle ear bones of mammals evolved from bones which former were positioned in the jaw.

The giant panda has 6 fingers, the 6th finger developed from an additional sesomoid bone.

How can these genes have such a immense effect on the body? How can so few genes control the development, the number of body parts, or expression of body parts?

They toolkit genes induce signal transduction pathways, most of them are transcription factors, but there are cell adhesion proteins, cell surface receptors or encode secreted morphogens that form the pattern and the positional information. Most of them are best defined as **master genes**.

The phrase, master genes, is best defined as the expression of these genes influencing the expression of several hundred other genes and starting whole developmental pathways, expression of whole body parts (e.g. including the regrowth of a limb after losing it, as in the case of Salamanders).

Hox genes

Hox genes belong to the group of toolkit genes.

Hox genes are also referred to as homeotic genes and homeobox genes. The name originates from a characteristic section: there is a 180 bp long, 60 amino acid encoding homeobox section in their structure. This conserved section has a role in the attachment, it contain a helix-turn-helix section.

They encode transcription factors. We refer to them as master genes due to their influence on the expression of hundreds of other genes.

We can find hox genes in every multicellular organism and they are very conserved regions.

Interestingly, the order of hox genes on the chromosome is parallel to the order of expression in the anterio-posterior axis (from the tail to the head) and this is commonly referred to as phenomenon collinearity.

There are 4 groups of hox genes in the gnathostomata grop. We highlight these genes with these capital letters of the alphabet: hoxA, hoxB, hoxC, and hoxD.

Positional information

Cells specify their position with morphogens and cell-junctions. These inform the cells where they are and this is the positional information.

Morphogenes are substances, signal molecules in the body providing positional information with their concentration or existence, so it provides information to the cells about what developmental pathway to choose. A single morphogen can help to form the dorsal-ventral and anterio-posterior direction too, such as the longitudinal axis of the body or the length of the legs. These molecules influence the gene expression.

Cellular mechanisms

Development begins with a single fertilized cell. First, cell cycle is activated, than the created cells start to specialize. In this section you can read about some important concepts and mechanisms. Every cell in the body contains the same genetic material but cells share different tasks, for this gene expression is different, unique and characteristic in each cell. The expression pattern is influenced by cell-junctions and different environmental factors, for example morphogens.

The goal of developmental biology is to understand how an organism develops from a single cell. In the embryo, there are 4 processes essential to develop a body and these are the cell proliferation, cell specialization, cell interaction and cell movement. Cells have a “memory” about the proliferations and positional information. The changes of the cells during development will inherit to the daughter cells.

During development, a cell first becomes committed to a fate (specified cells) and then becomes determined. A committed cell can reverse and transform to another developmental means, but a determined cell cannot reverse and transform to another cell type. Essentially, it means during the development and determination, cells lose their potential and are limited to provide specific functions. During this process, cells undergo special chemical and structural changes. The result is often a specialized cell that will not divide anymore.

Cell specification

There are three methods to reach the differentiated cell. These are the autonomous specification, conditional specification and syncytial specification.

Autonomous specification: only intrinsic factors influence this process and it is independent from the environment. It is caused by the concentration of intracellular determinants: the unequal dispersion of intracellular regulators (regulatory proteins, mRNA, small regulatory RNAs) that will lead to asymmetrical cell division: the daughter cells will contain different (amount of) regulators, they will differentiate and committed to different fate.

Conditional specification: extrinsic factors determine this process, extracellular factors influence the development of the cells that comes from the environment of the cell. During this process the development of (blastomer) cells will influence by signals that create the positional information (morphogens) or signals which cells exchange. Cells can interact and affect each other's fate.

Syncytial specification: in insects, it is a unique way to specialize during the development. In an early stage of the embryo we can observe a big cell with more nuclei, this is the syncytial blastoderm. The cytoplasm of this cell is not uniform, the interaction of cytoplasmic factors will decide what fate to choose.

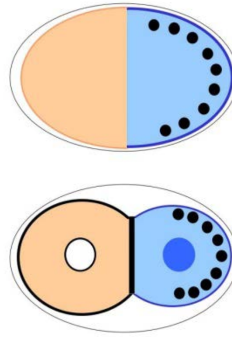
Human development

During fertilization, the egg cell (ovum) and sperm fuse and create the zygote which is a single diploid cell. Prior to fusing of the the two pronucleus (the nucleus of a sperm and egg), DNA will replicate and then fuse only following replication. The cell cycle activates, zygotes go through mitotic divisions creating a multicellular embryo in which the cells do not differentiate. During these divisions the size of the zygote does not change significantly, this proliferation process is known as cleavage. At the end of this process, there will be around 12-16 cells and we refer to this stage as the morula stage.

Maternal effect

The mother determines the descendant cell not only with its genetic material but also with other substances of the cytoplasm. Additionally, daughter cells will inherit the mitochondria of the egg. Some cytoplasmic determinants (e.g. regulatory proteins, mRNA and small regulatory RNAs) in the cytoplasm of the mother are not equally located and this will lead to asymmetrical cell division. The resulting daughter cells will contain the same genetic material, but different regulatory proteins. These regulatory proteins will create other gene expression patterns in these cells and start other developmental pathways. Because of the asymmetrical division, the cells will create poles of the zygote. Sperm provides only chromatin and centriole to the egg.

The differentiation starts with asymmetrical cell divisions. The egg contains some cytoplasmic determinants unevenly (morphogens/regulatory elements), in a mosaic form, leading to asymmetric cell division, because resulting cells will contain different regulatory proteins. This will initiate the differentiation of the cells.



Morphogens determine the differentiation in the early stage: this will begin the segregation of embryonic cells and extraembryonic cells forming the placenta, after the 8 cell stage (epiblast-hypoblast segregation).

Additionally, the activated regulatory genes will precisely separate the parts. During the future development and including pattern formation, cell junctions will have an important role, too.

Development continues with gastrulation (formation of germ layers: ectoderm, mesoderm and endoderm), then with neurulation (separation of the surface ectoderm-neuroectoderm-neural crest) and division of the tissues, formation of somites and closure of neural tube.

Morphogenes are important in the next developmetnal stages too, as Hedgehog, Wnt, Notch, TGFß, or receptor tyrosine kinases.

Morphogenes can be both inhibitors and inducers. Morphogens can move by diffusion and create the gradient in the tissues.

Homeotic mutations

The mutations of hox genes will result in surreal occurrences, in which they will sprout body parts in unusual places upon the body. For example, in Drosophila an antenna may form in the place of limbs, or eyes on limbs, or eyes upon the trunk, and throughout the entire body of the insect.

Homeotic mutations within the human body may result in the unusual number of limbs, unusual morphology of limbs, or the same with vertebrae and or ribs. If a body part or organ express and appears in a strange/abnormal place, we refer to it as being “ectopic”. Ectopic organs form by homeotic mutations, e.g., Drosophila may feature ectopic eyes if pax6 gene is expressed, even within the cells of the limbs.

Avatism

Sometimes a structure evolves and shortly afterwards, hox genes begin to regulate them. This was the circumstance with the wing formation of ancient arthropods. As it developed, it appeared on every tori segment, in which hox genes regulated where to express this structure. If and when there is a mutation, it is always featured by an earlier stage in its evolution. This appearance of the phenotype which is similar to an earlier evolutionary state is known as atavism. Normally, without regulation of hox genes we would have 8 fingers!

Summary

Most of our genes (more than thousands) create the general body plan which encode cellular mechanisms. Most of these genes do not differ between the different strains. These genes cannot tolerate mutations very well. Instead, we know only a few dozen of genes (under 100) effective in the development of which, mutations of these genes change the phenotype, can create new forms and functions. These genes are the commonly known toolkit genes (for example, hox genes). These genes provide the phenotypical evolution and the economical way to create new forms and functions from the "same material". These genes are so called master genes and affect the expression of more hundred other genes.

During the development (during the differentiation) cells shed their potential to form every cell type despite the fact every cell contains the same genetic material.

Cells specify their position according to morphogens (chemical substances) and cell junctions. In humans, the differentiation begins by asymmetrical cell divisions due to the egg containing some cytoplasmic determinants (morphogens) unevenly and following cell division, the daughter cells will contain different morphogens that will start different processes.

23.11. Gene therapy

Introduction

Several diseases are caused by the mutation of a single gene. The inheritance of these diseases follow the Mendelian rules. A few examples for monogenic diseases: phenylketonuria, Hurler disease, Duchenne muscular dystrophy, sickle cell anaemia, haemophilia.

Symptoms of these diseases are severe and there are no real cures.

Gene therapy attempts to treat these diseases by repairing the mutant DNA sequence. This can be achieved many ways:

- repairing the mutant allele
- knockout of the mutant allele
- introducing the wild type allele

Important! Prior to the start of gene therapy, the patient's genotype must be determined!

Methods

Transferring DNA into a single, isolated cell is not that difficult. However, editing the genomic material of every cells of an organism raises tons of difficulties. The two main methods:

- Indirect therapy: Cells are removed from the organism. Their DNA is modified *in vitro*, then these cells are retransplanted.
- Direct therapy: The vector, containing the modified sequence, is injected directly into the organism.

Vectors of gene therapy

Vectors are the shuttle which will transport the desired fragment of DNA into the cell. Vectors of gene therapy are divided into two groups: viral and non-viral vectors.

Non-viral vectors

- 'naked' DNA
- lipoplexes: encapsulating DNA in a lipophilic membrane improves the efficiency of transduction
- nanoparticle linked DNA: DNA molecules are physically attached to nanoparticles. These particles can be injected into cells (eg. gene gun)

Low efficiency of transduction is the biggest drawback of non-viral vectors. These vectors are mainly used for indirect therapy. Advantages: transduction by non-viral vectors is relatively easy to carry out and there is no risk of immune response.

Viral vectors

Viruses must inject their genetic material into the nucleus of the infected cell for their reproduction. The strong evolutionary pressure fine-tuned these mechanisms: viral transduction in highly efficient, most cell types can be infected. It became obvious that viruses could be used for gene therapy. (for general information on viruses, see that chapter)

The most important virus groups are the following:

Retroviruses: RNA viruses code reverse transcriptase enzyme, what transcribe RNA of virus to DNA in the host cell. This way 8-10 kilobase size DNA can be integrated into the infected cell's genome. The main drawback of using retroviruses is that they can only infect dividing cells and the genomic integration is quasi random. (see: insertional mutagenesis, insertional inactivation)

Lentiviruses: A genus of Retroviridae family. Their biggest advantage over the retroviral vectors is that they are able to infect non-dividing cells as well.

Adenoviruses: They are known for gastrointestinal and respiratory diseases. They contain double stranded DNA, which does not integrate into the host cell's genome. This extrachromosomal DNA might get lost during cellular division, but there is no risk of random insertion. Adenoviruses' main disadvantage is the immune response they trigger, since they are pathogenic. The developing immune reaction will not only reduce the effectivity of transduction but the anaphylactic shock can kill the patient. (see the lecture (Jesse Gelsinger))

Adeno associated viruses: Subgroup of adenoviruses. Small, dsDNA viruses with no known pathogeny, which means there is no immune response. They can infect dividing cells, but their cloning site is limited (4.8 kb).

Dangers of gene therapy

Like every medical intervention gene therapy also has its real risks:

- Insertional mutagenesis: The incorporating DNA fragment might inactivate a cell cycle regulator gene, which can lead to malignant neoplasia.
- Insertional inactivation: The incorporating DNA fragment might break an essential gene.
- Viral vector triggered immune response: The immune response of the host might be fatal. Typical for adenoviruses.

The risks must be minimized by choosing the suitable method and vector.

Gene therapy in practice

A lot of research groups have started to develop treatments of monogenic diseases, many therapeutic methods even reached the clinical stage 2 (details at lecture), but up to date there is only one approved gene therapy.

Alipogene tiparvovec (brand name: Glybera) is the first gene therapy approved by the EU. It treats lipoprotein lipase deficiency (LPLD). A mutation in the gene of LPL causes the loss of activity (of the protein encoded). The symptoms: chylomicronemia and chronic pancreatitis. Adeno associated viruses carry the wild type allele of lipoprotein lipase. The viruses should be injected into the femoral quadriceps. The drawbacks of this treatment are the necessary immunosuppression and the breathtakingly high price.

APPENDIX – METHODS OF MOLECULAR BIOLOGY AND GENETICS

NUCLEIC ACIDS

Isolation

DNA and RNA can be extracted from samples of almost every living tissues, cells, and even from dead tissue. Extraction can be done based on physical-chemical methods, which use different techniques. DNA and RNA extraction methods use the same techniques, with minor changes in solutions:

1. Phenol–chloroform extraction

   Aqueous samples are mixed with phenol:chloroform mixture. The mixture is centrifuged. Two distinct phases are formed: the aqueous phase is on top. Proteins will be found in the lower organic phase, the nucleic acids (as well as other contaminants such as salts, sugars, etc.) remain in the upper phase. The top phase is pipetted off avoiding pipetting organic phase. To further purify DNA, these steps can be repeated.

   If the mixture is acidic: DNA precipitates into the lower phase, RNA remains in the aqueous phase (DNA is more readily neutralized than RNA).

2. Solid phase extraction (Boom method)

   The base of this method is the fact that nucleic acids bind to the solid phase of silica under certain conditions (presence of chaotropic agents – like guanidium thiocyanate, and alkaline pH). The steps of the procedure:

   - **Sample lysis/homogenization**: Lysis buffer with detergents and degrading enzymes, physical homogenization by grinder, pestle, etc.
   - **Binding**: Chaotropic agent-containing buffer and silica beads as substance are mixed to the sample homogenate. Nucleic acids bind to the silica beads within these conditions.
   - **Washing**: Beads are washed several times to remove contaminants: beads are collected by centrifugation and the supernatant is disposed, washing buffer added, and mixed. After chaotropic agent-containing washing buffer, several alcohol-containing washing solutions are used similarly.
   - **Separating** nucleic acids from beads (elution): an elution buffer, decreasing the concentration of the chaotropic substance is added, silica bead – nucleic acid bonds are eliminated, pure nucleic acid is eluted to buffer.

3. Column purification or Spin column-based nucleic acid purification

   Quick, solid phase extraction method based on silica bead solid phase extraction method. The coloumn contains a membrane which acts as the solid phase of extraction, the other components are similar to that of Boom method. Steps:

   - **Lysis** – Breakage of the cells to reach the target (nucleic acid). → physical breakage of tissue, lysis buffer to break cell membrane.
   - **Bind**ing the nucleic acid to the silica membrane – A buffer solution is added to the sample with ethanol or isopropanol. This binding solution is transferred to a spin column which is centrifuged. → The centrifuge forces the binding solution through the silica membrane inside the spin column. If the pH and salt concentration of the binding solu-

tion are optimal, the nucleic acid will bind to the membrane as the solution passes though.

- **Wash** – The flow-through is removed and a wash buffer is added to the column. The column is put in a centrifuge again, forcing the buffer through the membrane. This washes any remaining impurities out of the membrane, leaving only the nucleic acid bound to the silica gel.
- **Elute** – The wash buffer is removed and an elution buffer (or simply water) is added to the column. The column is put in a centrifuge again, forcing the buffer through the membrane. The elution buffer removes the nucleic acid from the membrane and it is collected in a tube.

Detection

Southern Blot

Southern Blotting is a technique used for detecting specific DNA fragments based on the ability of complementary single-stranded DNA fragments to hybridize. The name Southern comes from the inventor of the technique, Edwin Southern. The procedure of the Southern blot is as follows:

1. DNA is cut by specific restriction enzymes so that fragments can enter the gel.
2. Fragments are separated by gel electrophoresis (agarose or acrylamide).
3. Denaturation in alkali to expose single-stranded DNA fragments to hybridization.
4. Transfer ssDNA fragments to a membrane.
5. Hybridization of the fragments on membrane with labeled probe (commonly radiolabeled probes are used). Labeled probes anneal only to complementary strands, which are found in a specific position (band) on the gel (therefore on the membrane).
6. Detection of the labeled band by X-ray or fluorescence.

The method can be used for determining insertion or deletion mutations in a particular locus, by monitoring the change in fragment length of a given locus.

Northern blot

This technique is similar to Southern blotting, the difference is that in this case the target sequence is RNA, and the probes can be either single stranded DNA or RNA fragments.

The steps of Northern blotting are the following:

1. RNA is isolated from cell culture cells, (animal or human) tissues, bacterial or yeast cultures or other sources.
2. RNAs are separated according to size on denaturing gel (agarose or acrylamide) – the function of denaturation is to get rid of the higher structure of RNA.
3. Separated RNA transcripts are transferred onto positively charged membrane.
4. Hybridization of RNA with labeled (fluorescently, enzymatically) probes.
5. Detection of hybridization – the presence, the size and the relative abundance of the target RNA can be seen.

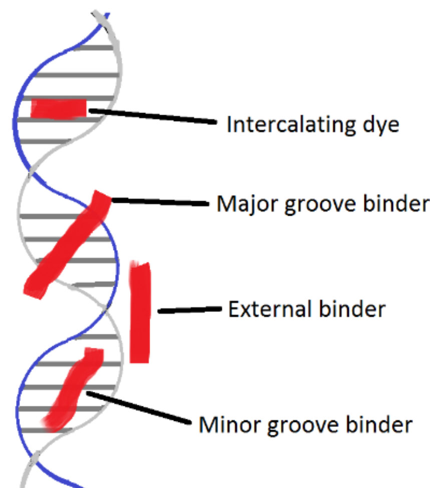
Non-specific labeling of nucleic acids

We can detect nucleic acids with UV light at the wavelength of about 260 nm. We can also use markers to which we can bind to the DNA non-specifically, thereby making the entire amount of DNA detectable or the entire amount of nucleic acids.

Nucleic acids are acidic compounds meaning they can bind basic dyes (nucleic acids are basophilic).

The three most important classes of nucleic acid strains are as follows:

1. Intercalating dyes: they can incorporate into the 2 strands of the DNA: e.g. propidium iodide (PI), ethidium bromide (EtBr), or daunomycin. We can use these intercalating dyes in gels, too.
2. Minor-groove binders: DAPI and Hoechst dyes (note diagram).
3. Other nucleic acid stains: for example, acridine orange (AO). Acridine orange has the capability to distinguish single- and double stranded nucleic acids: in double stranded DNA it has green fluorescent emission but if bound to a single-stranded nucleic acid, it has red emission.



Radioactive labeling

Radioactive isotopes

Radioactivity is the process where non-stable (=radioactive) nuclei degrade.

Types of radioactive radiation:

α: nuclei of He,
β: electron radiation,
γ: electromagnetic radiation

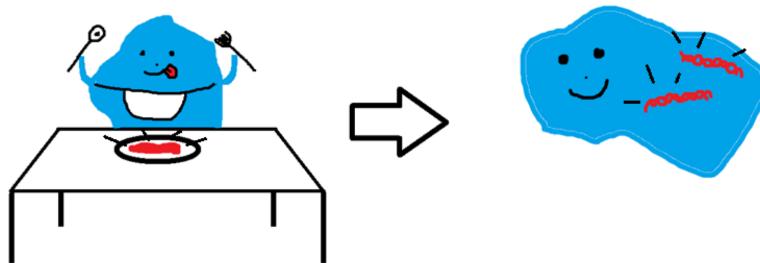
In both molecular biology and medicine, one of the most important properties of radioactive molecules is half-life. Half-life is the length of time required to one half of the given amount of radioactive atoms to decay.

The most commonly used radioactive isotopes and their half-lives

Radioactive isotope	Emitted radiation	Half-life
^{3}H	β particle	12,35 year
^{14}C	β particle	5730 year
^{35}S	β particle	87,5 day
^{32}P	β particle	14,3 day
^{131}I	γ particle	8,1 day

We detect radioactivity in molecular biology with the following:

- Specific photopapers (autoradiography films), PET, MRI and particle detectors, e.g. Geiger-Muller tube, scintillation detectors and ionization chambers.
- In the radioactive labeling of nucleic acids, we can bind probe nucleic acids containing radioactive phosphates (^{32}P) by a hybridization method, such as FISH.
- Additionally, we can use the thymidine incorporation method and incorporate/build in a radioactive base (^{3}H- or ^{14}C-thymidine) into the DNA of a living cell: we make the radioactive base available as nutrition for the cells so they will incorporate them into their new DNA strands.



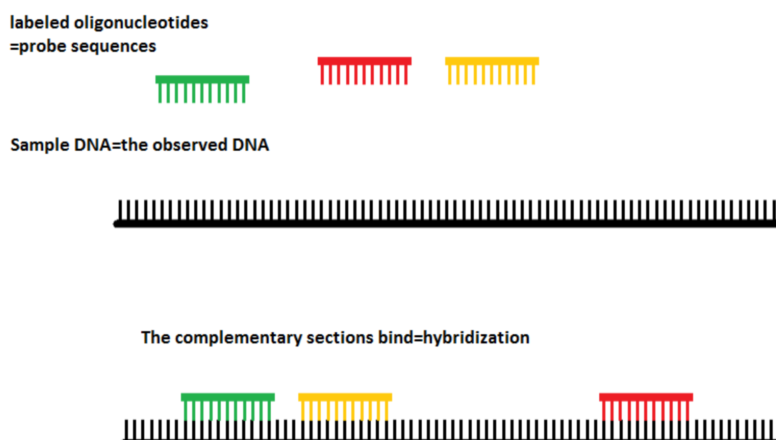
FISH: Sequence specific labeling of nucleic acids

FISH is best defined as being fluorescent in situ hybridization, a cytogenetic technique to detect or identify a specific DNA or RNA section.

We use small sequence-specific probe DNAs to locate DNA motifs and sequence specific probe RNAs to find specific RNA sections. These sequence-specific probes are specific because they are complementary to the targeted (observed) DNA or RNA section. These small complementary sections are fluorescently tagged to make them detectable. In this regard, we can use not only fluorescent molecules but also radioactive substances or biotin.

Procedure: Our objective is to effectively design a proper probe section large enough to hybridize and not so incredibly large so as not to thwart and obstruct the binding.

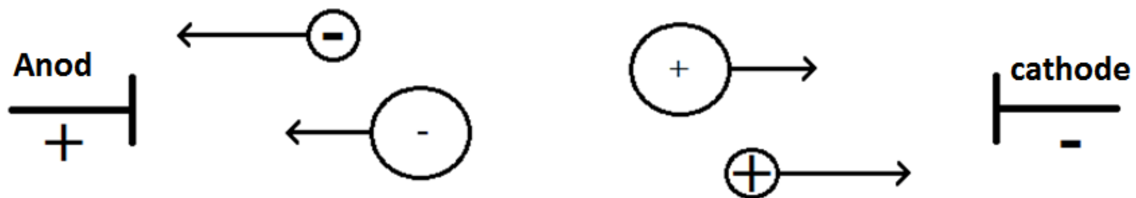
First we denaturate the observed nucleic acid (with formamid or heat) then we add the labeled small probe oligonucleotides and provide the suitable conditions in order for the probe DNAs to hybridize (bind) to the complementary sections (it takes approximately 12 hours). Next, we apply several rinse steps to wash out the non-hybridized oligonucleotides from the system and detect the bound sequences.



Electrophoresis of nucleic acids

Electrophoresis

The base of separation: the charged particles move in the homogenous electric field. The speed of the motion depends on their charge and size. Smaller and higher charged particles move faster, bigger and lower charged particles move slower.



The efficiency of the separation is improved if we use different solid matrices e.g. various gels, however, the measurement can be achieved in other media, such as buffers.

Gel electrophoresis

We typically use gels in the electrophoresis of nucleic acids.

During the electrophoretic separation in gels, the most important phenomenon occurs when smaller molecules move/pass through the gel matrix quicker than the bigger ones.

The two most commonly used gel matrices:

Agarose: aqueous suspension of polysaccharide which makes a cross-linked gel matrix following a boiling and cooling procedure. The density of the crosslink is dependent on the concentration of agar. In this case, the electrophoresis is usually accomplished in the horizontal position.

Polyacrylamide gel (=PAGE): it is a three dimensional polymer of acrylamide and bisacrylamide in which these two molecules bound covalently. Physical properties of this gel are better when compared with the agarose gel. The concentration of the monomers determines the pore size of the gel. In this case the electrophoresis is attained generally in the vertical position. Its resolution is better when compared to the resolution of agarose gels.

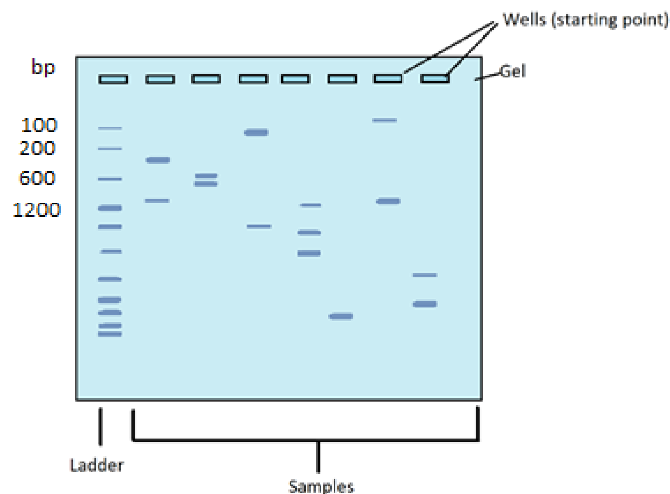
Denaturing gel electrophoresis

Gel electrophoresis can be performed in non-denaturing and denaturing conditions, too. Denaturing conditions are used to disrupt the intramolecular interactions, which is very important in case of RNAs to determine their exact molecular weight.

Denaturing gels: generally used for the separation of proteins and RNAs. By nucleic acids we use urea, and often DMSO and glyoxal, formamid to denature RNA.

Native gels: gels which are run in non-denaturing conditions.

In this analysis, we can make a separation and determine the size of the molecule with using so called ladders (mixture of molecules with a known size: we can correlate the sample molecules to the molecules of the ladder).



The distance from the start point and the size of the molecules demonstrate a logarithmic relationship.

Cloning

As a general description; a fragment of DNA is inserted into a cloning vector, and the recombinant DNA is propagated in a host organism. The process in brief:

1. **Insert DNA** isolation and purification.
2. Isolation of **vector DNA**.
3. **Digestion** of the target DNA and vector using restriction endonucleases.
4. **Ligation**: target is inserted into the vector.
5. The vector containing the insert is **transformed** into competent host bacterial cells.
6. **Screening** for recombinant molecules by PCR reaction (colony PCR), analytical restriction digestion or DNA sequencing.
   - *Colony PCR*: several colonies are checked for the presence of the plasmid: after a selective amplification of the target, the product is run on agarose gel to find out if a band corresponding to the size of the insert is present.
   - *Analytical restriction digestion* is performed using appropriate restriction endonuclease(s), then digests are analyzed by agarose gel electrophoresis for the presence of a recombinant clone.
   - *DNA sequencing* is the definitive identification of the recombinant plasmid. The technique of DNA sequencing is described later in this chapter.

Insert: The DNA fragment we want to incorporate in the vector. In other words, the target. The source for target can be genomic DNA, complementary DNA (cDNA transcribed generally from mRNA) or PCR-amplified fragment.

Vector: extrachromosomal circular DNA molecules that can be maintained in the host cell. Commonly used vectors for cloning: bacterial and yeast **plasmids**, artificial chromosomes (YAC, BAC), bacteriophage or retroviral vectors. Important structural features of plasmids:

- MCS: **Multiple Cloning Site**: contains many restriction sites – sequences where restriction endonucleases can cleave them selectively. After cleavage, ligases can attach the insert into the vector. Using this method, the researcher can insert the target sequence to a known position of the plasmid.
- **Selectable marker**: A sequence of the vector which allows the selection of transformed cells. Antibiotic resistance is often used as marker (eg. Ampicillin, Kanamycin resistance).

- A vector may contain a **reporter gene** as well: these allow screening for successfully cloned cells. For instance the gene coding for β-galactosidase, which is capable to hydrolyze the colourless X-gal molecule to a blue product. If an insert is cloned successfully within the lacZα sequence of the β-galactosidase (in the vector MCS), the production of the enzyme is prvented. If X-gal is included in the selective agar plates, transformant colonies appear to be blue in the case of a vector with no inserted DNA and white in the case of a vector containing a fragment of cloned DNA.

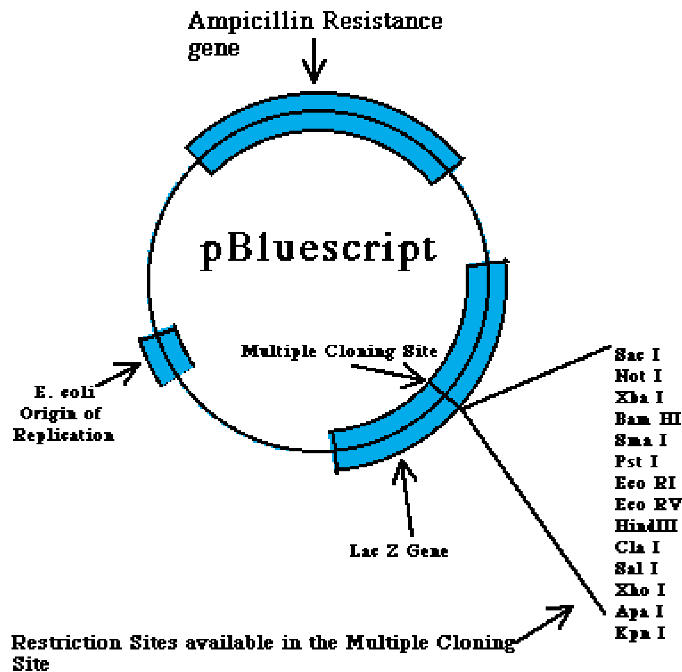


Fig. 1.:
A general description of the pBluescript plasmid.
Note the MCS, the restriction sites within it, and Amp resistance gene.

Restriction endonucleases: enzymes that cut DNA at specific sequences. The restriction sites are usually between 4 and 8 bases. Many of them are palindromic, meaning the base sequence reads the same backwards and forwards. Some restriction endonucleases produce "sticky" ends (an overhang of one strand), like EcoRI, others produces “blunt” ends, like SmaI, as seen below:

Fig. 2.:
restriction sites of EcoRI and SmaI

When target and vector DNA are cut with the same enzyme, they are easier to ligate (especially when endonucleases producing sticky ends are used).

DNA ligase: an enzyme capable of joining two DNA fragments together by forming a phosphodiester bond between 3’ and 5’ ends. Together with restriction endonucleases it is the basic tool for inserting a DNA sequence into the vector.

Plasmid transformation: uptake and incorporation of the plasmid to competent bacterial cells. Heat shock or electroporation forces the bacterial cells to uptake the vector (with the insert). Bacterial cells are then plated on agar plates containing antibiotics and incubated for colony development. Only those cells incorporating the plasmid with antibiotic resistance gene will grow and form colonies. Transformant colonies are picked up and grown in liquid media, and plasmid can be isolated from the host cell. The process of introducing DNA into eukaryotic cells is called **transfection**. Electroporation, lipofection and microinjection are techniques used for transfection. For lipofection, the vector is packed into liposomes, which fuse with the host cell's membrane, while microinjection is the injection of the DNA directly to the nucleus.

Mutagenesis: engineering of DNA to produce gene mutations. One goal of mutagenesis it to examine the functions of a gene, by mutating its control elements or product. Sometimes mutations produce mutant proteins with altered functions that may be of commercial use. Random mutagenesis happens when the organism is exposed to mutagens (UV or mutagenic chemicals). Mutants with interesting phenotypes are selected. Other techniques for random mutagenesis include error-prone PCR reaction in conditions that enhance misincorporation of nucleotides. PCR products which contain mutation are then cloned into an expression vector and the mutant proteins produced can be characterized. *Site-directed mutagenesis* was first produced using nucleotide analogs or chemicals to generate localized point mutations. Current techniques use mutagenic oligonucleotides in a primer extension reaction with DNA polymerase to generate point mutation, or deletion or insertion of small stretches of DNA to be introduced at specific sites. *Combinatorial mutagenesis* modifies a few positions on DNA by excision of that portion and replacement with many different sequences from a library of mutations. This technique is useful if we want to screen for a particular phenotype. *Insertional mutagenesis* is a useful tool to identify genes involved in carcinogenesis and to understand the biological pathways of specific cancer and to examine the function of a gene. *Homologous recombination* can be used to produce a specific mutation in an organism. Vector containing DNA sequence similar to the gene to be modified is introduced to the cell, and by a process of recombination replaces the target gene in the chromosome. This method can be used to introduce a mutation or knock out a gene.

Analysation of nucleic acids

PCR

qRT-PCR

Quantitative reverse transcription is able to judge the relative amount of a specific RNA transcript at different conditions (with and without of a treatment for example). In this case after RNA isolation reverse transcriptase enzyme is used to generate a single stranded DNA copy (cDNA) of all RNA molecules (by using oligo d(T) only mRNAs are copy, using the 3' poly (A) tail as template). The synthetized cDNA will be used in a PCR reaction with gene-specific primers. This way the relative amount of the PCR product (in the exponential phase of the reaction) is proportional to the amount of the starting RNA.

Real-time PCR

Sequencing of DNA

Nucleic acid interactions: Hybridizations, protein interactions: footprinting, EMSA, ChIP and RNA interference

EMSA- Electrophoretic mobility shift assay

Electrophoretic mobility shift assay (EMSA) is commonly referred to as gel retardation and/or band shift assay, is best defined as a rapid and sensitive method utilized in the detection of sequence-

specific DNA-binding proteins. This method is suitable for qualitative and quantitative determinations. The assay is based upon the ability of a transcription factor to bind in a sequence-specific manner, manner, next onto a radio-labeled oligonucleotide probe which serves to retard its migration through a non-denaturing polyacrylamide gel. In this regard, both crude nuclear extracts and purified factors are suitable as a source of the DNA-binding protein.

Determination of specificity of the DNA/protein interaction

Two common approaches are often used to determine the specificity of DNA/protein interaction towards the effective identification of the specific protein involved within the complex formation.

1. *Unlabeled competitions*: prior to the addition of radiolabeled probe DNA, a 50-100 fold molar excess of unlabeled competitor DNA is added to the reaction mix. Individual reactions are performed with oligonucleotides containing the target DNA sequence and oligonucleotides which have been specifically mutated within the target sequence. Specific binding is indicated by a loss of factor binding to the radiolabeled probe.
2. *Antibody supershifts*: antibodies specific to the putative DNA binding protein are incubated with in the binding reaction prior to addition of radiolabeled probe. If the antibody recognizes the target protein two distinctive results are possible. If the antibody does not inhibit binding it will create a higher molecular weight complex which will be observed as a “supershift” on the autoradiograph.

Assay

1. Nuclear extract preparation
2. Measurement of protein concentration
3. Design of the DNA oligonucleotide
4. Annealing of the DNA oligonucleotide
5. Labeling of DNA probe
6. Binding reactions
7. Non-denaturing gel electrophoresis and autoradiography

Protein extracts may be prepared from whole cells or nuclei. The amount of nuclear extract required may vary dependent on the protien concentration of the extracts and the amount and affinity of the transcription factor intended for additional study.

The type and size of the probe is dependent on the nature of the investigation. In the event of a previously identified DNA binding site intended for study, a synthetic oligonucleotide probe is ideally suitable for use. Synthetic binding sites are made by choosing two complementary single-stranded DNA oligonucleotides including the sequence of interest and annealing them one to another. These oligonucleotides are designed to possess overhanging ends at the 3'-extremity one fully annealed.

The labeling of fragment probes depends on the nature of DNA ends. In the case of fragments with 3'-overhanging ends, T7 DNA polymerase, which possesses a 5’ to 3’ polymerase activity is used for labeling. An alternative method used in the precise labelling for the oligonulceotides involves the addition of ^{32}P to the 5' end using T4 polynucleotide kinase and $\gamma^{32}P$-ATP. In this specific circumstance, the oligonucleotides are designed with blunt ends.

Poly d[I-C] is added to the binding reaction as a competitor for non -specific DNA binding protein. Sonicated salmon sperm DNA (ssDNA) may alsobe used. However, generally speaking, the very simple copolymers assure the best results. Concentration and combination of poly d[I-C] and ssDNA often needs to be determined empirically.

Competition analysis using an unlabelled DNA fragment (same sequence as for the labeled probe) can be used to test the specificity of the complex formation to the DNA sequence. Binding of the unla-

belled competitor DNA to the transcription factor of interest will result in a decrease in the amount of protein available for binding to the probe. This will lead to an attenuation or elimination of the band corresponding to the complex formed by the specific protein.

To positively identify the proteins which complex to the DNA-binding sites, antibodies against known transcription factors are added to the binding reaction. These antibodies may bind to the complex, causing an alteration in the mobility of the complex, characterized by a super-shift of the DNA-protein complex or it can completely inhibit the complex formation by binding to an essential site on the transcription factor required for DNA binding, resulting in the absence of the DNA-protein complex on the gel.

ChIP- Chromatin immunoprecipitation

Chromatin immunoprecipitation is an invaluable method for studying **interactions between specific proteins and a genomic DNA region**. It can be used to determine whether a transcription factor interacts with the target gene and it is used to monitor the presence of histone proteins with posttranslational modifications at specific genomic locations.

Assay

1. The cells are treated with formaldehyde which crosslinks DNA with DNA-binding proteins.
2. Chromatin is isolated from the cells.
3. Genomic DNA is shared by mild sonication (200-1000 bps) and precleared to reduce non-specific immunoprecipitation.
4. Preclarification is carried out with Protein A- or G-agarose resins which have been blocked with ssDNA.
5. Target protein specific antibodies bound to protein A- or protein G- agarose resins are added to the samples.
6. The samples are incubated with the resins.
7. After incubation the resins are washed to wash out nonspecific protein-DNA complexes.
8. Protein complexes are eluted from the resins.
9. Samples are heated to reverse the covalent crosslinks.
10. The DNA fragments are purified and analyzed by PCR.

Footprinting

Footprinting is a method for **determining the exact DNA sequence** to which a particular DNA-binding protein binds.

Assay

1. We clone a piece of DNA cotaining the site to which the transcrition factor binds.
2. Label one end of the DNA molecules with a radioactive molecule, e.g., radioactive ATP.
3. Digest the DNA with **DNase I** enzyme
   - DNase I cuts DNA molecules randomly (it does not have recognition sequence)
4. The result yields a mixture of radioactive fragments of varying length, with the smallest increment in length represented by a single nucleotide.
5. Next, the fragments are separated by electrophoresis.
6. Binding of the *transcription factor* to the DNA sequence prevents DNase I from attacking the binding region.
7. When the fragments are separated by electrophoresis, those representing the lengths covered by the transcription factor will be missing from the autoradiogram.
8. The resulting gap is the "footprint".

9. The same sample of DNA (without the transcription factor) is subjected to normal DNA sequencing and the resulting ladder aligned with the footprint autoradiogram.
10. The exact sequence of bases in the *DNA* can then be read directly because they represent the rungs of the ladder missing in the footprint.

RNA interference

RNA interference (RNAi) is a process in which a specific mRNA is degraded due to the presence of a small **double-stranded siRNA** whose sequence is contained within the mRNA sequence.

Assay

Determining the roles of genes at the cellular level can be accomplished by studying the effects of siRNAs on the activities of cultured cells. RNAi can be used to study gene function in mammalian cells by incubating the cells with small dsRNAs encapsulated in lipids (transfection) or by genetically engineering the cells to produce the siRNAs themselves. Inside the cell, the siRNA regulates the **degradation** of the target mRNA. Thus, the cell is unable to produce the protein encoded by the target gene. We can use **libraries** containing thousands of siRNAs, or vectors containing DNA encoding these RNAs to study of gene function in a number of model organisms and in humans. In this way one can study the effects of the elimination of a gene's expression on any cellular process.

The role of RNAi in the cells

RNAi is probably a type of genetic immune system which protects organisms from the presence of foreign or unwanted genetic material (infections). RNAi is probably evolved as a mechanism to block the replication of viruses and/or to suppress the movements of transposons within the genome. Cells can recognize dsRNAs as foreign molecules because they are not produced by the cell's normal genetic activities.

The double-stranded RNA is first cleaved into small (21–23 nucleotide), double-stranded fragments, called small interfering RNAs (siRNAs), by a particular type of ribonuclease enzyme, called Dicer. These small dsRNAs are then loaded into a complex called pre-RISC. One of the strands of the RNA duplex (passenger strand) is cleaved in two and then dissociates from the pre-RISC. The other strand of the RNA duplex (guide strand) is incorporated into a related protein complex named RISC. The RISC provides the machinery for the single-stranded siRNA to bind to a complementary RNA sequence. Once bound, the target RNA is cleaved at a specific site by the ribonuclease. Each siRNA can carry out the destruction of numerous copies of the target RNA.

PROTEINS

Protein isolation

There are specific procedures to extract and purify specific proteins from observed biological samples but there are some general steps which are common during the isolation procedures. It starts with lysis of the cell, of which, we can perform with liquid nitrogen, ultrasound, frosting and high pressure or with different chemical substances. We can choose the proper technique according to the type of the cell (bacterial, plant, yeast, or mammal cell) with accordance to the subsequent purifying steps. This step will determine the yield and the quality of proteins, too. After the lysis, we should separate the non-protein components, such as nucleic acids, lipids and cell organelles with specific precipitation and centrifugation steps. During centrifugation procedure first cell debris will separate from the lysate then we precipitate nucleic acids and also separate them using centrifugation. During the process the use of protease inhibitors is recommended for the maximum yield.

Following the separation of proteins, we can further purify them with regard to the characteristics of the targeted protein: amino acid composition, size, shape, net charge, isoelectric point, solubility, heat-stability, hydrophobicity, binding behavior, and other properties.

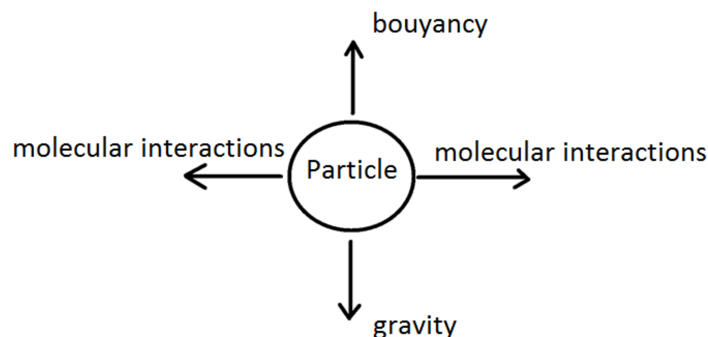
The distinctive property	Applicable technique
Solubility	Selective extraction
Different binding / sorption ability	Chromatographic methods, ion exchange methods
Molecular weight / size	Sedimentation / centrifugation, gel electrophoresis
Shape	Sedimentation / centrifugation
Density	Density gradient centrifugation
Charge	Electrophoresis, isoelectric focusing, ion exchange chromatography
Specific and non-specific binding sites	A variety of chromatographic procedures, immune techniques

Separations

Centrifugation

In this technique separation, is generated using the gravitational acceleration (g). With this technique we can separate proteins in solutions. We can use this technique with other particles, too, including cells, cell organelles, macromolecules or smaller molecules).

There are different forces acting on a particle in solution : buoyancy, gravity and various molecular interactions.



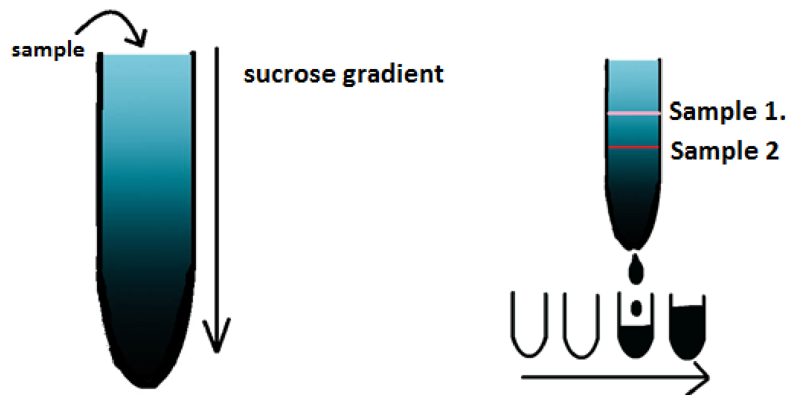
During centrifugation we increase the force of gravitational acceleration upon the sample. This will act on **the dispersed particles** and they **will sediment** in the solution **depending on their size, shape and density**.

Gradient centrifugation

Gradient centrifugation is a specific centrifugation method where we use a density gradient solution, usually a sucrose or a cesium chloride gradient. The density of these solutions increase towards to the bottom of the centrifuge tube. We add the sample (for example, a cell lysate) to the gradient and we

centrifuge the sample in this gradient. The particles will sink and create a band in the tube: the bands will appear at the density that is the same as their own density. In this technique only the density is important, the shape and size influences the time only, meaning how fast the particles reach their appropriate bands.

When the separation is complete we collect the fractions (=bands) in different tubes.



Ultracentrifugation

We call it ultracentrifugation because it is a high performance centrifugation technique, the rotor of this equipment could obtain 600.000xg or 2.000.000xg acceleration. At this speed, air resistance is an important factor and so, a vacuum is used inside the centrifuge. Theodor Svedberg (Nobel Prize winning chemist) invented and created the first ultracentrifuge in 1925. **From the measurements we can calculate the sedimentation constant, commonly referred to as the Svedberg** (taking into account of the sedimentation of the particles, the speed and the acceleration of the centrifuge). **This sedimentation constant is characteristic** to the distinct proteins. (Memo: consider how we distinguish the parts of ribosomes: we name them 40S, 60S. These numbers with the S signal means subunits has 60 and 40 Svedberg sedimentation constant.)

Chromatography

Chromatography is **one of the most important** separation techniques in pharmaceutical research. To understand chromatography, one should understand several important definitions.

Chromatographic systems consist of a **stationary phase** and a **mobile phase**. Mobil phase carries the sample through the parts of the system and through the stationary phase.

The basis of the chromatography: the sample carrier mobile phase passes through a stationary phase, and the separation is based on differential partitioning between the mobile and stationary phase. Because the particles interact in different ways, the constituents of the mixture pass through the stationary phase with different speed, causing the separation of the particles.

The stronger interactions create slower motion through the stationary phase, the particles which do not interact with the stationary phase (or the interactions are weaker) quickly pass through it. The time particles spend in the system is the “retention time”. **Chromatogram** is the visual output of the chromatography.

Mobile phase could be a liquid, a gas or a supercritical fluid.

We can distinguish many types of chromatographic methods as follows:

- Gas chromatography (GC): mobile phase is gas
- Liquid chromatography (LC): mobile phase is liquid
- Paper chromatography (PC): stationary phase is a paper
- Thin layer chromatography (TLC): stationary phase is a “planar” thin player as by PC

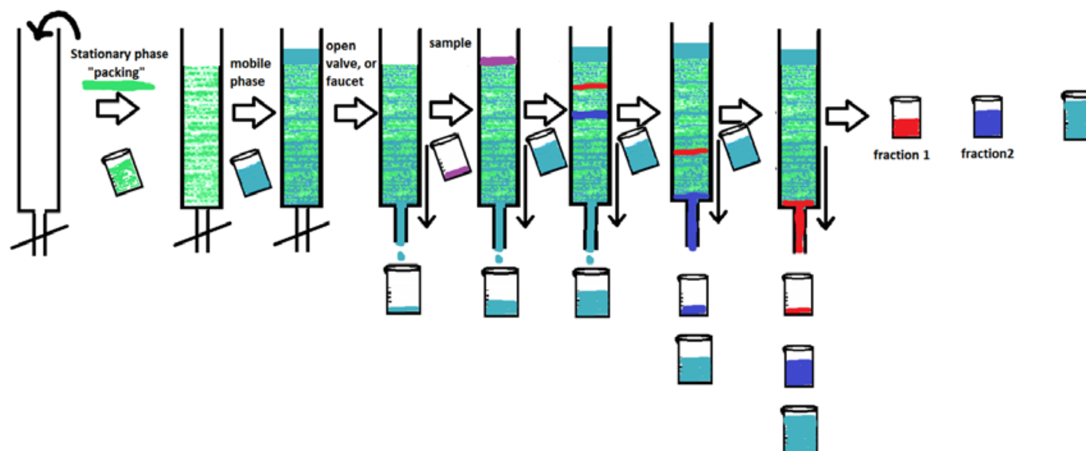
The driving force moving the mobile phase may be different: electric voltage (by electrochromatography), pressure (by most chromatographic methods), and capillarity caused movement (by planar methods: paper and thin layer chromatography).

Column chromatography

The definition „column chromatography” means the separation takes place in a “**column shape**” system! This **does not determine the type of interaction** between the stationary phase and the sample.

This is one of the most widely used methods, largely due to its use both in very small microscopic sizes and in immense sizes.

Note the example pictured: the stationary phase is packed into a column, we fill the column with the mobile phase and pipette or fill the sample on the top of the stationary phase. As we open the valve at the bottom of the column, the sample starts to flow down, through the stationary phase. The column is refilled continuously with the mobile phase (the pressure should be continuous). The samples separate from each other in the column and we collect the fractions separately.

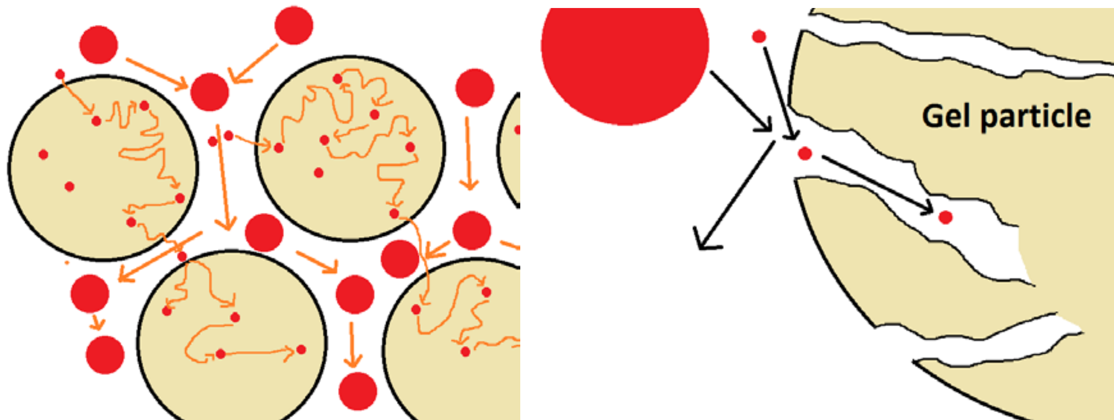


The most widely applied chromatographic methods according to the interactions

Size exclusion chromatography

The stationary phase consists of **small, porous gel particles** which can be made of e.g. agarose, polyacrylamide or dextran.

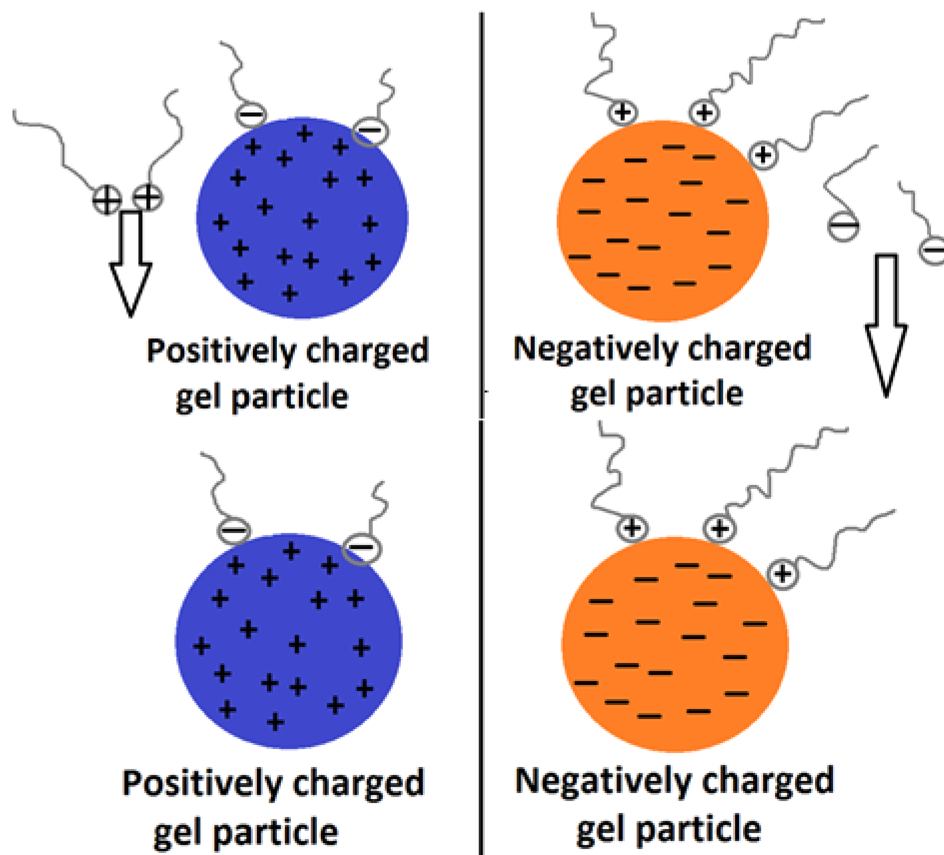
As in the mobile phase, once the sample passes through the system, the smaller particles will enter into the pores of the gel particles, but the bigger sample particles cannot go in the pores of the gel particles, therefore they travel around it (see illustration). The smaller sample molecules which can move into the gel particles “take a longer trip” in the column meaning their retention time is longer. With this method we can separate by **size**.



Ion exchange chromatography

The stationary phase of the ion exchange chromatography is particles with charged surface (these particles may be gel particles, too).

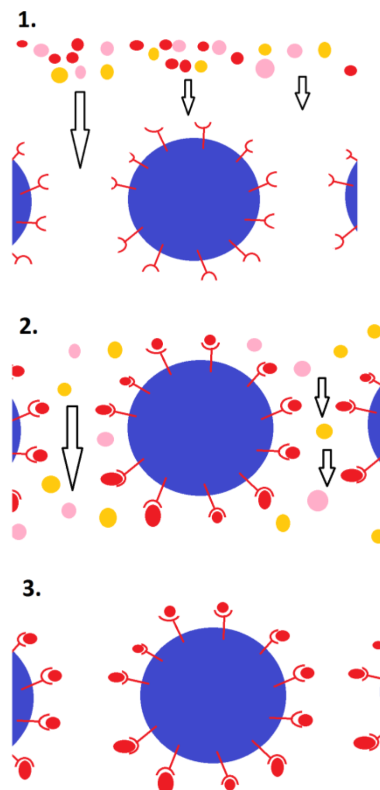
The base of separation is defined as when the opposite charged particles will be bound on the surface of the gel. The power of binding is different, it depends on the strength of the charge. The interacted particles pass slowly or do not pass through the column.



Affinity chromatography

In this method we bind specific substances to the surface of the gel particles during the preparation of the stationary phase. These bound substances can bind specifically targeted molecules of the sample.

The bound molecules cannot pass through the column, the others flow through the system. After the outflow of the non-bound particles, we disconnect the bound particles from the gel particles: we elute them with an elution buffer and collect them separately.



*HPLC = **H**igh **P**ressure (or **H**igh **P**erformance) **L**iquid **C**hromatography*

We can separate compounds with HPLC that are not volatile enough for gas chromatography and cannot derivate to volatile compound to a GC analysis (for example, proteins). This equipment uses high pressure to press the sample through the stationary phase.

The conventional HPLC instrument can use 400 bars and even more, 1200 or 1500 bars overpressure. With an increase in high pressure, there are HPLC instruments identified as UHPLC= ultra high pressure liquid chromatography. It is one of the most important pieces of equipment of analytics.

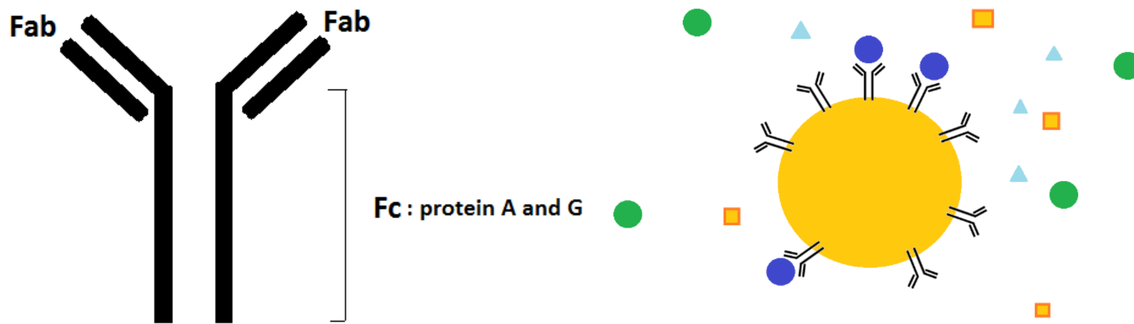
In the use of HPLC methods we can implement all of the liquid chromatography methods.

Immunoprecipitation

Immunoprecipitation is a technique based on the precipitation of a targeted protein (antigen) by the effect of an antibody. The process can purify or concentrate determined proteins from a sample with accuracy. In utilizing this technique, we can purify particular proteins from a tissue lysate. We can use this technique to make an observed protein visible, for example, during a Western blot.

Isolation of particular proteins:

We can bind antibodies to agarose gel by the Fc region of the antibody. Special agarose gels can bind the protein A or protein G region of the antibody. After the precipitation of the targeted protein on the agarose gel, we can sediment the gel particles with centrifugation. Following isolation, we can analyze them further using electrophoresis.

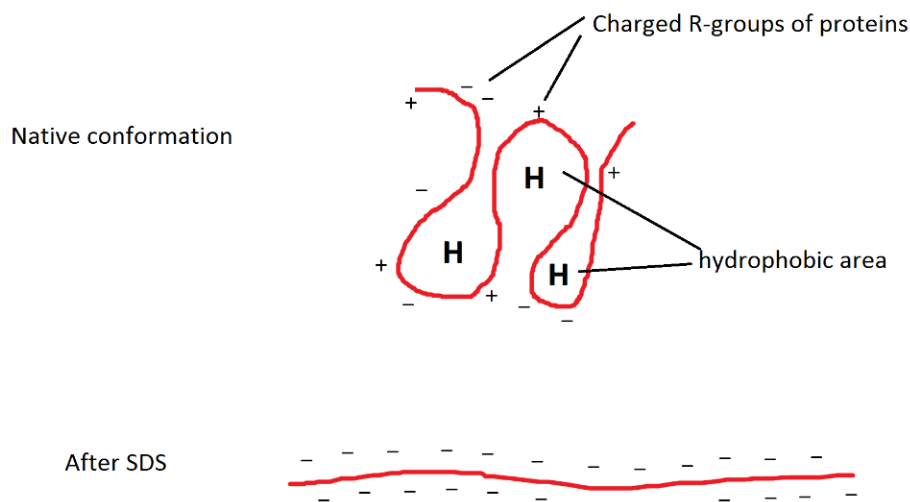


Electrophoresis of proteins

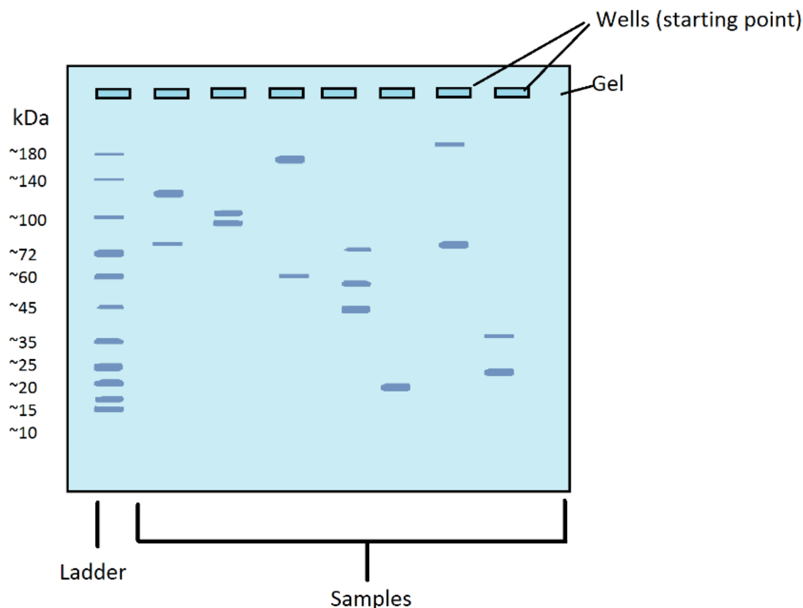
(General information about electrophoresis: see electrophoresis of DNA)

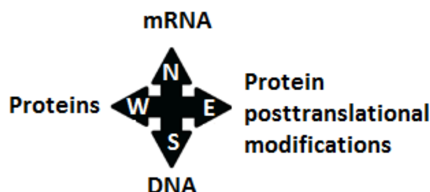
Electrophoresis of proteins can be performed in both in fluid and in gel environments. We typically use denaturing conditions to avoid disrupting the effect of the 3D structure of the protein.

Denaturing gels: generally used for the separation of proteins (and RNAs). In proteins, the denaturant is typically **SDS** (sodium dodecyl sulfate) and **mercaptoethanol**. The detergent SDS linked to the proteins' hydrophobic amino acids and provides a strong negative charge to the protein (the sulfate groups of the SDS gives the negative charge). The negative charge is formed in the entire length of the molecule and it exerts a repulsive force so the protein loses its original conformation and the non-covalent interactions are removed. Mercaptoethanol reduces the disulfide bonds within and between polypeptide chains, resulting in the formation of a negatively charged protein molecule and the strength of the charge will depend only on its length! Because the charge/size ratio is constant, the protein molecules will migrate in the gel matrix only according to the size of the molecules: the bigger molecules cannot move as quickly as the smaller ones.



In a denaturing gel, we can separate proteins according to their size. With this analysis we can make a separation and determine the size of the molecule using ladders, which are defined as a mixture of molecules with a known size, and then we can correlate the sample molecules to the molecules of the ladder.





The distance from the starting point and the size of the molecules demonstrate a logarithmic relationship.

Detection of proteins

Western blot (or immunoblot)

Western blot begins with a gel electrophoresis step (it can be a native electrophoresis that separates according to the 3D structure of the protein or a denaturing electrophoresis that separates according to the length of the protein) as in all of the blot techniques. Next, the separated proteins will be transferred to a membrane (usually nitrocellulose) where it will be visualized. Visualization is acheived using a staining procedure with antibodies (immunolabeling).

Immunolabeling

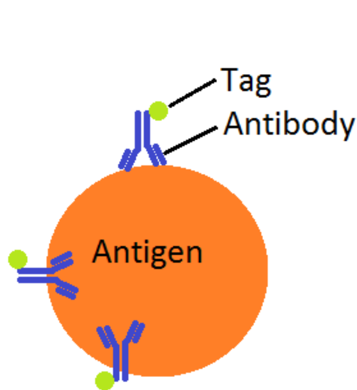
Antibodies bind with high affinity and very specifically to the surface of the recognized molecule. We can distinguish monoclonal and polyclonal antibodies according to the specificity of the antibodies. Polyclonal antibodies are secreted by different B cell lineages (they recognize the same antigen but they bind to different parts of the same molecule). Monoclonal antibodies derive from the same cell lineage and they are chemically precisely the same. In molecular biology we prefer the monoclonal antibodies that can recognize a special part of the antigen and only that particular part.

There are two types of immunolabeling: direct and indirect method.

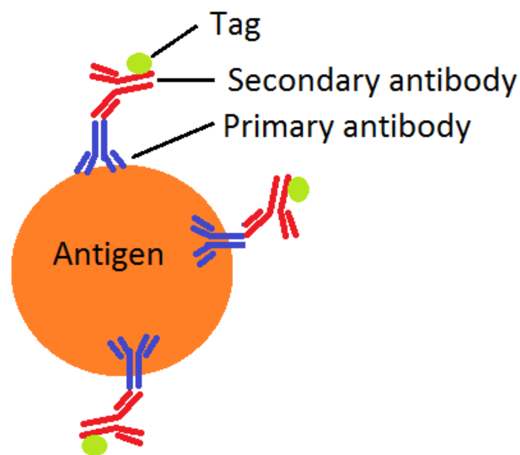
Direct method is where the antibody binding to the antigen is tagged. An indirect label occurs when we bind a specific antibody to the antigen and we make that initial bound antibody detectable with another antibody which is labeled (contains a tag). In this method we identify the first bound antibody as the primary antibody and the second as secondary antibody. This indirect immunolabeling amplifies the signal and in utilizing this technique we can mass produce the secondary tagged antibody.

In this technique we can label the antibodies with different tags: a fluorescent compound, an enzyme which produces colored compounds, or gold beads.

Direct labeling



Indirect labeling

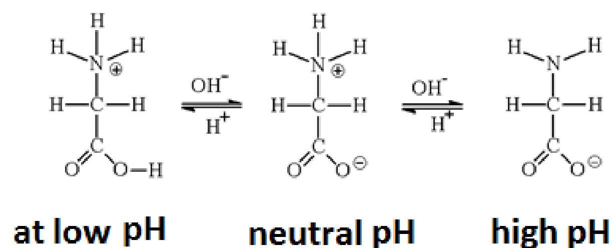


Analysis of proteins

Isoelectric focusing

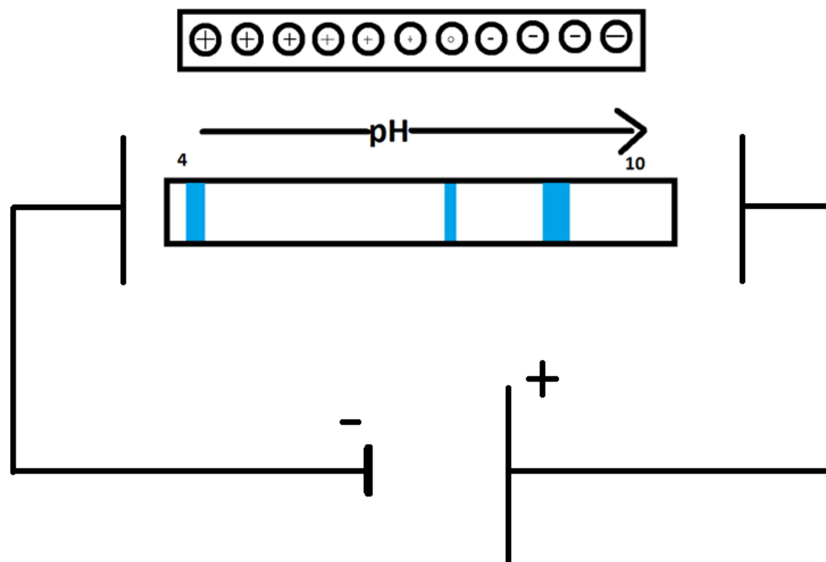
We use this technique typically in the separation of proteins.

The proteins may feature both positively and negatively charged groups. If the pH of the environment changes, it causes change in the charge, because these groups can protonate and deprotonate. In high pH, the net charge is more negative, in low pH the net charge is more positive.



Proteins have amphoteric properties, meaning they can act as a weak base in acidic solutions and as weak acids in base solutions (due to the charged groups).

During isoelectric focusing, we place the proteins into a pH gradient and connect DC power into the system forming a homogenous electric field in this pH gradient.

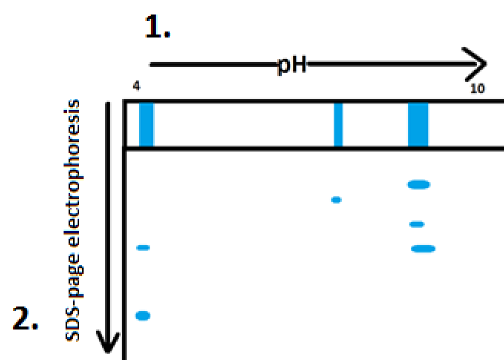


The added sample molecules will migrate/move in the electric field and in the pH gradient according to their net charge (electrophoretic phenomenon). In a point of the pH gradient, the proteins do not have a charge (their net charge is zero), in that point they do not move and produce a band. Notably, the particles move in the homogenous electric field only if they have charge!

The **isoelectric point** is the point, where the net charge of a protein is zero (0).

2D electrophoresis = 2 dimensional electrophoresis

Two dimensional techniques are separation techniques where we separate the sample particles with two separation techniques, one following the other. In order for proteins to separate all proteins from one another, we typically use a 2D technique consisting of two parts: isoelectric focusing and SDS PAGE. According to these two properties (**isoelectric point** and **weight),** the different proteins will make unique spots.



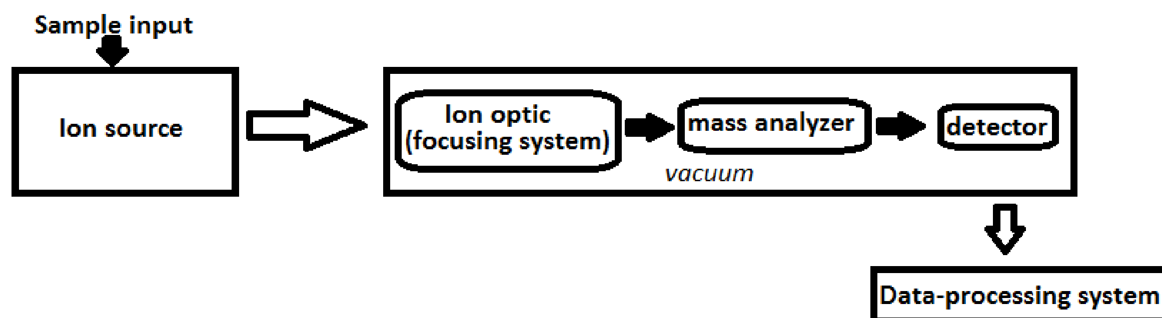
1st dimension: isoelectric focusing

2nd dimension: SDS-page electrophoresis

Mass spectrometry (MS)

Mass spectrometry is an analytical technique used to identify the amount and the type of ions of an observed sample by measuring the mass-to-charge (m/z) ratio. The sample is ionized by an **ion source,** then the ions are sorted and separated according to their mass and charge (with a **mass analyzer**). The separated ions are measured with a **detector**. The created ions are short lived and

reactive so the analysis is performed in a vacuum. Ionization (creation of analyzable ions from molecules and atoms) can be performed by kinetic energy, light, electric or chemical energy. After the ion source, there is a focusing system that ensures to gather the ions with the same kinetic energy and in one beam to the detector. This ion optics transfers the ions to the mass analyzer into the vacuum. A mass analyzer separates, then the ions according to the mass-to-charge ratio. The detector measures the intensity of the ions and creates a mass spectrum according to the ion current intensity and specific mass.



According to the above described method, a mass spectrometer consists of an ion source, a mass analyzer and a detector. A vacuum system provides the vacuum conditions and a data system processes the data from the detector and controls all the functions.

Ion sources

Ion sources of mass spectrometers can be different. There are ion sources which work with atmospheric pressure and others which operate within a vacuum.

Some important types of ion source are as follows: electrospray (ESI), atmospheric pressure chemical ionization (APCI) and atmospheric pressure photoionization (APPI).

Electrospray ionization: uses an electric field to generate charged droplets and a heated nitrogen drying gas that shrinks the droplets and carries away uncharged materials. It works in atmospheric pressure.

Atmospheric pressure chemical ionization is very similar to ESI, of which there is an atomizer that makes droplets, too, yet here a corona discharge ionizes the sample in the gas phase.

During atmospheric pressure photoionization we add additives to the sample (e.g. toluol) that significantly absorbs UV. During the sputtering, the solution is irradiated with UV light, and the sample is ionized by the additives.

The most widely used mass analyzers are the quadrupole analyzer (Q), ion trap (IT) and time of flight (TOF) method.

We can use this technique in the field of proteomics after sample preparation of liquid chromatography-based methods.

We can use MS to identify the mass of the proteins, peptide fragments, determine the protein structure, function, interactions and post-translational modifications.

Protein sequencing

The sequencing method is based on the so called Edman reaction that cleaves the N-terminal amino acid of the peptide chain.

The method is very complex and begins with the digestion of the protein followed by a 2D chromatography step (usually 2D column chromatography or RP-HPLC). In the Edman reaction, we cleave the N-terminal amino acid using phenyl isothiocyanate. The separated N-terminal amino acid will be identified using mass spectrometry one followed by the other. After the sequencing of digested parts, we repeat the operation with different enzymes having different cutting sites. After sequencing we assemble the sequences from the overlapping sections.

Nuclear magnetic resonance (NMR)

NMR active nuclei absorb and reemit electromagnetic radiation in magnetic electric field. The resonance frequency is characteristic and depends on the strength of the magnetic field and the magnetic properties of the isotope of the atoms. We use this technique especially in structural analysis, for example, the 3D structure of peptides.

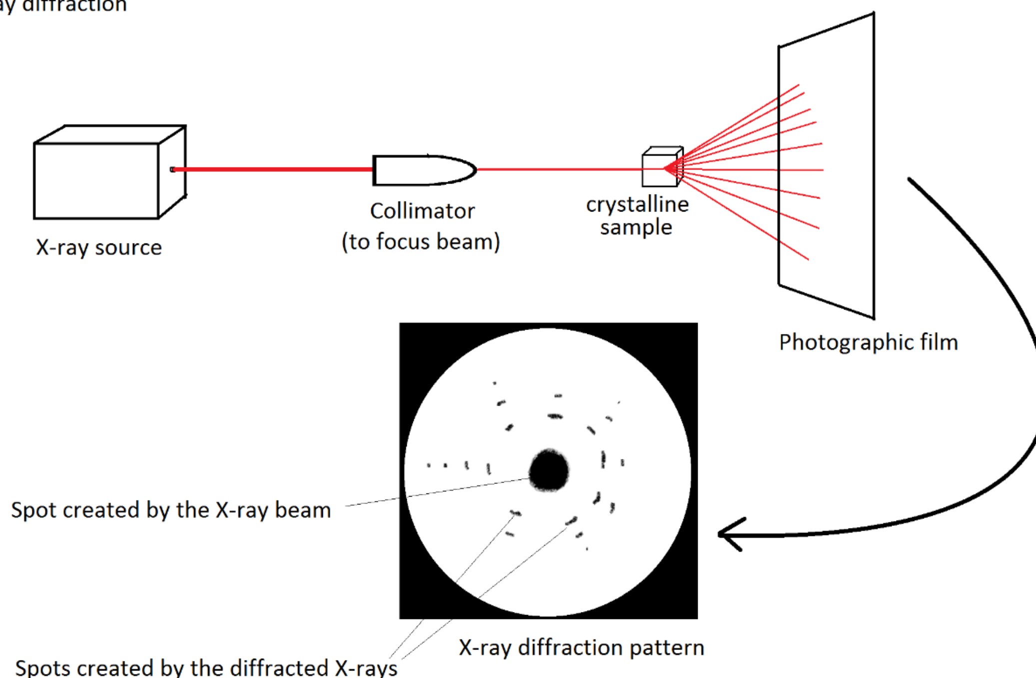
X-ray diffraction

Also known as X-ray deflection

X-rays scatter and deflect on the electron shell. Because of the interference, these interfered X-ray waves draw patterns with small patches or circles on light-sensitive materials as they travel through the crystallized sample, for example protein samples. The radiations interfere with one another which is scattered by two or more atoms! We can determine the exact location of an atom in the crystal cell using the information of the interference image.

From the size of the patches we can conclude to the quality of the atoms (the composition of the material). The method is suitable to examine large molecules, such as proteins.

This technique is important because the structure of DNA was solved with the help of X-ray diffraction images. (Rosalin Franklin)



Interactions of proteins

It is an important fact about a protein (maybe a potential drug target) what other proteins are interacting with it. The sequence data can help by revealing functional groups with known domains (specific binding activities). Here we discuss briefly two methods which will result in protein interactions clarification: pull-down methods and yeast two-hybrid assays.

Pull-down methods

These methods are in vitro, qualitative methods for protein-protein interaction detection. They can be used for proving of suspected interactions; to search for interacting partners of a bait protein; to carry out screening with cell lysate.

In general the bait protein is expressed with N- or C-terminal tag (fusion protein: protein of interest with an affinity tag). The most frequently used affinity tags are: GST (glutathione S-transferase: binding of glutathione); His_6 (6 histidines: binding divalent metal ions); biotin: binding streptavidin. The elution methods depend on the type of the tag (reduced glutathione, EDTA, or low pH). The bait protein is bound to the affinity column material via specific ligand. The material (e.g. cell lysate) carrying the potential interacting proteins, is incubated with the column and the bound bait protein. The non-bound components are washed out and the bound proteins are eluted. Eluted proteins can be analyzed by gel electrophoresis, Western blotting or mass spectrometry.

Disadvantages of pull-down assays: the expressed protein is not properly folded, so the binding activity of it is changed; tags may have effect on the structure and/or the binding sites of the bait protein; occasionally the elution of the proteins is difficult. We have to take into consideration, whether a revealed interaction is stable or just a transient one. It can be a considerable component, that a protein has to get changed (phosphorylated, GTP-bound) for certain interactions.

Yeast two-hybrid assay

In this assay two examined proteins have to bind each other within the yeast cell (in vivo). To detect this interaction we have to clone the two proteins separately into two different expression vectors, and transform yeast cells with both plasmids at the same time. One of the plasmids carries the coding sequence of one protein with fusion of one part (DNA binding domain) of a transcription factor. The other protein (or a library for screening) is expressed in another vector with fusion of the other domain (activation domain) of the transcription factor. The yeast cells have the coding sequence of a reporter gene, the activity of which can be determined by a simple (enzymatic, colorimetric) measurement. If the examined proteins bind each other in the yeast cells, the two domains of the transcription factor with get in contact, and the activity of the reporter gene will be measurable. Without the interaction of the proteins the reporter gene has no or only a minimal activity.

Förster Resonance Energy Transfer (FRET) (or Fluorescence Resonance Energy Transfer) is a biophysical phenomenon that is used to examine molecular interactions in the nanometer scale. A light-sensitive fluorophore molecule, say a *donor*, when absorbs a given wavelength light, gets into a higher energy state (*excitation*). To get back to a lower energy state, fluorescence radiation can occur (*emission*), or a non-radiative energy transfer may occur to an *acceptor* fluorophore molecule through long-range dipo-dipol interactions. See the Jablonski-diagram below:

Fig. 3.:
Jablonski-diagram including FRET

The ***efficiency*** of this energy transfer is inversely proportional to the sixth power of the ***distance*** between donor and acceptor. Therefore FRET is extremely sensitive to small changes in distance, which makes it a valuable tool for studying molecular structure and molecular interactions. One common pair of fluorophores for biological use is a cyan fluorescent protein (CFP) – yellow fluorescent protein (YFP) pair.

Expression of recombinant proteins

High level of pure proteins can be easily achieved by using the cDNA of the mRNA of the protein, cloning it into special expression vectors, generate high amount of the target protein in host cells and purify the protein with the help specific sequences or unique feature of the protein. This way therapeutic peptides and proteins can also be generated (recombinant technology).

Aims of expression of recombinant proteins: study the structure and function of target protein, use protein in therapy, to produce specific antibodies, to use protein as target of drugs. The expression vector is the tool to introduce foreign DNA into the host cell and generate large amount of the cloned protein. The vector can be plasmid or recombinant virus. The host cells can be: bacteria (cheap, easy to deal with, but not always properly folded eukaryotic proteins); yeast (for eukaryotic proteins); mammalian cells (expensive, difficult to get high quantity).

Common steps of the technique:

- gain the coding sequence of the protein
- choose the proper expression system (how to use the purified protein, purification method, origin of the protein, potential posttranslational modifications, high expression level is toxic or not)
- cloning cDNA into expression vector with restriction enzymes
- transform host cells with recombinant plasmid
- grow and induce cells for expression
- purify and identify the expressed protein.

For easier purification proteins may be **tagged** on the C- or N-termini. Tag could be added to the protein sequence with PCR, or the protein can be cloned into a vector which already carries the sequence of the tag (make sure the sequences are in frame!). When using bacterial host cells, we have to consider the differences in codon usage, posttranslational modifications, folding or toxicity between prokaryotic and eukaryotic proteins.

An **expression vector** used in prokaryotic host cell must contain the following sequences: origin of replication; antibiotic resistance gene (for selection of transformed cells); multiply cloning site (for cDNA insertion in proper orientation and reading frame); promoter sequence (constitutive or inducible); terminator sequence, ribosome binding site for translation; potentially tag (possibility for affinity purification even with cleavage site between the protein and the tag sequence).

Induced expression: transformed cells (containing the recombinant plasmid) are grown to mid-log phase, then the inducer is added to the medium (concentration of the inducer and the induction time has to be determined experimentally), after further growing cells are collected (centrifuged), washed, lysed and the protein can be purified.

MODEL ORGANISMS

“What is true for *E. coli* is true for elephant.” - Jacques Monod

Model organisms are used for studying different biological phenomena. Many biological, biochemical processes are highly conserved and similar in a wide range of organisms. It is more convenient to study these processes on model organisms.

The choice of model organism can range from viruses to apes, but they have common features:

- easy to breed and maintain in the lab
- short lifecycle
- genetically well defined (ideally full genome known)
- characterized strains

Usually model organisms are not important medically or economically per se, but the knowledge obtained from them can be used in other species. However care must be taken with the extrapolation of data.

Examples of model organisms

Bacteria

Escherichia coli: Probably the most studied organism of the world. It is a Gram negative gut bacterium with a short (20-30 min.) generation time and a single circular genome. Haploid. Its genome is sequenced and there are many tools for genetic modification. There are some pathogenic strains as well.

Yeasts

Saccharomyces cerevisiae (Baker’s yeast): Diploid single cell organism with many practical applications: baking, brewing and biofuel production. Many fundamental pathways are highly conserved and similar between S. a. and other eukaryotes. Its generation time is approximately 90 minutes under optimal conditions.

Plants

Arabidopsis thaliana: Arabidopsis is a tiny flowering plant. While it has no agricultural significance it is one of the most studied plants due to its small genome and short generation time. Its genetic modification is relatively easy using a soil bacterium capable of infecting the plant, *Agrobacterium tumefaciens.*

Animals

Caenorhabditis elegans: It is a roundworm of the size about 1 mm. Its transparent body and the interesting fact that adult worms have always the same number of cells make it a perfect model organism of developmental biology. Apoptosis and RNA interference was extensively studied.

Fruit fly (*Drosophila melanogaster*): Common fruit fly or vinegar fly is a tiny diptera which can be found around rotting fruit. It has a very short life cycle, only four chromosomes what made fruit fly one of the most important model organism in genetics. The lab of Thomas Hunt Morgan produced the first genetic linkage map on *Drosophilas.*

Zebrafish (*Danio rerio*): Zebrafish is a common pet fish. It is a favoured organism of developmental biology because the larvae are translucent and they develop externally.

Mouse (*Mus musculus*): Many biological processes should be studied in mammals. The small size and short generation time of the common lab mouse makes it an ideal mammalian model organism. The genome of mouse is known, and many well defined laboratory strains are available as a model for different genetic diseases. Furthermore, knock out and transgenic mice are also available.

Further examples of mammalian model organisms: rat, guinea pig, dog, cat, pig, chimpanzee.

Genetically modified organisms or transgenic organisms

Genetically modified organism (GMO) is an organism whose genome has been changed by modern molecular biological methods.

Strictly speaking transgenic organism is when a gene or gene fragment from a different species is transferred into the host organism's genome. It can yield several advantages in basic and applied science. For example: it is possible to study the function of human genes in model animals, important peptides can be produced in bulk quantities (e.g. insulin in yeasts) or heat tolerant crops can be cultivated.

Knock-out (KO) modification is when a gene is selectively mutated, what leads to loss of function. We can differentiate into permanent knock-out and conditional knock-out. In the latter the expression of the studied gene can be turned off with a chemical signal. It makes it possible to study KO mutations which otherwise would be lethal.

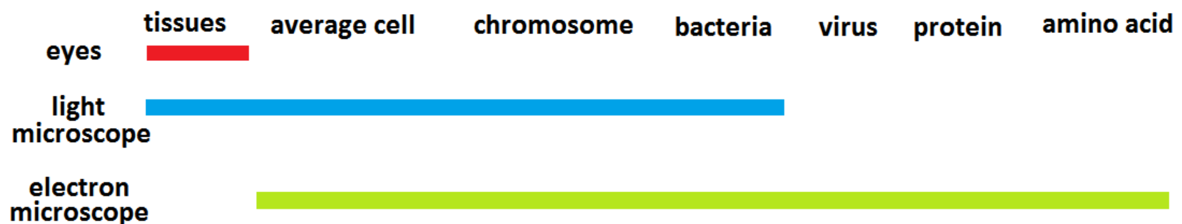
Imaging techniques

Microscope

There are some structures including cells and tissues in which we can only examine their intricate structures utilizing high magnification.

Structure	Avarage size
Tissue structure	100 µm
An avarage cell	10-50 µm
Red blood cells	7,5 µm
Chromosomes	1-5 µm
Bacteria	0,1-1 µm
Virus	10-100 nm
Proteins	2-20 nm
Amino acids	0,33 nm

We can examine these small structures with different types of microscopes. The first developed microscope was a light microscope in the 17th century. With this microscope, we were able to see and examine cells. Today, we have many types of microscopes, which belong to two large groups: light microscopes (=optical microscopes) and electron microscopes.



Light microscope

Traditional light microscope (=optical microscope)

This microscope uses visible light (~400-800nm) for the imaging. A system of lenses (optical system) enlarges the image of the sample.

Optical system

We typically use compound microscopes, meaning this microscope uses more lenses to collect the light from the sample and transfers it to our eyes.

The sample is very thick, the light travels through the sample and passes through the lenses to the eyes in this pathway. The lenses enlarge the image and rotates it several times.

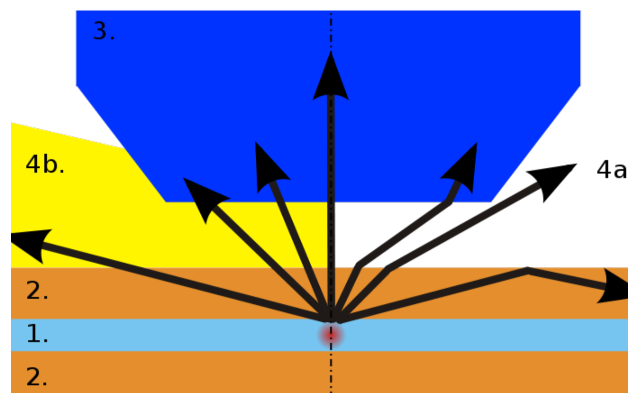
We need at least two lenses: one near the sample, which is the objective; and one near the eyes, which is the eyepiece lens. There may be more lenses in the tube, but these are the most important

parts. As the real and magnified image appears through the tube of the microscope, the eyepiece lens further enlarges the image.

The total magnification is the multiplication of the magnification of the objective lens and the eyepiece lens. It is labeled on the tube of these lenses. If we use a 10x objective lens and a 20x eyepiece lens, the total magnification totals 200x.

Objective: The lens or a lens system is the most important part of the entire lens system. This determines the quality and the resolution of the image.

Immersion lenses: special type of objective lenses. The air (the refractive index of the air) in the distance between the objective and the sample determine the amount of maximum magnification. We can optimize it if we use other materials between the sample and the objectives. In this regard, we can use specific objectives, such as immersion lenses, featuring higher magnification (60-100x), and in the use of these immersion lenses, we use oil between the objective lens and the sample. The oil has higher refractive index, resulting in collecting more light from the sample and ultimately examining a larger part of the image.



1. Sample
2. Slide
3. Objective
4. a. traditional use, air between the sample and the objective
   b. oil between the sample and the objective: immersion oil

Illumination system

Light source: a lamp or a mirror which guides light through the sample.

Between the lamp and the stage there is an important part: the **condenser, condenser diaphragm** or **iris diaphragm**. This is an adjustable part to alter the amount of light entering into the system, and image contrast (if we close them, the contrast will be higher, but the image will be darker).

Resolution: the shortest distance between two separate points in a microscope's field of view that can still be distinguished as distinct entities, meaning the minimum distance between two points which we can see as two different points. This is one of the most important properties of the microscope.

Stereo microscope (=stereoscopic microscope, dissecting microscope)

This type of microscope has a lower magnification. This microscope uses incident light illumination, meaning the light does not go through the sample, the light source is positioned above it.

There are two separate optical paths for the two objectives and two eyepieces: and in this method, we acquire two different views from different angles to the left and right eyes. The result is a three-dimensional image. In using this type of microscope we can examine surfaces in much greater detail, we can use these microscopes in micro surgery, as a jeweler microscope and dental microscopes.

The conventional, traditional light microscopes also feature two eyepieces, but there are not two optical paths, only one and the microscope shares this one image to the two eyes! However, the image in such a microscope is not different from that obtained with a single monocular eyepiece. The binocular eyepieces only provide a greater viewing comfort.

Staining

During examinations, it could be necessary to detect cell compounds specifically or just make a better contrast of the sample. We can use different contrast enhancers and indicator substances. We can use fluorescent dyes, if we have fluorescent illumination and excitation (e.g. with a mercury arc lamp).

Electron microscopes

Electron microscopes have a much better resolution than light microscopes, because it uses beam of electrons to create image and the wavelength of an electron can be up to 100.000 times shorter than the wavelength of visible light photons.

There are more types of electron microscopes. In the transmission electron microscope (TEM), a beam of electrons passes through the sample and will be detected in front of the electron source. In a reflection electron microscope (REM), reflected electrons are detected. It provides information about the surface. In a scanning electron microscope (SEM), a narrow focused electron beam moves across the sample and scans the sample from part to part.

CULTURED CELLS

Cells in cultures are “kept” within controlled conditions out of the organism, in cell culture flasks or tubes. Generally, when we talk about cell cultures, we think of cultured cells originating from Eukaryotic multicellular organisms. As every technique has limitations, the data achieved from cell line experiments can not be extrapolated in whole to the human body.

Primary cultures

Primary cultures are made from cells, ‘freshly’ isolated from tissue. These cells’ functions and physiological characteristics are closer to cells within the body and therefore act as best models for biology experiments. However, technically they are harder to get and keep in culture. One goal of primary cell isolation is to get a pure culture containing only one cell type. Primary cells are capable of limited number of cell divisions only (Hayflick limit). Primary cell cultures mainly are isolated from animal tissues. Steps of primary cell culture production:

- Isolation of the organ/tissue
- digestion of extracellular matrix (trypsin, collagenase)
- insertion into cell culture medium

Secondary cell cultures

Secondary cell cultures can be produced by subculturing (passage) a primary culture.

Cell strains

Cell strains are generated by a number of subculturing primary cells. These cultures have finite lifespan (40-50 passages).

Cell lines

Cell lines comprise of immortalized, proliferating cells with genetic homogeneity. These originate from tumorous tissues, hybridomas, adenoviruses. Due to telomerase activity of the cell, these lines are capable of infinite number of divisions. These lines are easy to maintain in culture, and also more standardized in “quality”, therefore these provide lower standard deviations between experiments. Some examples:

- HeLa: the first immortalized cell line, isolated from cervix carcinoma (originates from a woman called Henrietta Lacks), generally used for many applications (Salk’s polio vaccine)
- Jurkat: immortalized T-lymphocytes used mainly for immunological experiments. These cells are grown in suspension.
- CHO: Chinese Hamster Ovary: non-human origin, low chromosomal number (2n=22) and intensive protein expression characterizes this cell line.

Cell lines are used for many applications including drug screening, toxicology screening, protein production, virus breeding, gene expression experiments.

Cell banks

Cell banks store cell strains, cell lines in special cryoprotective media in liquid nitrogen.

The **advantages** of using cell cultures for experiments:

- specific, homogene cell population
- precise and reproducible experiment design
- controllable environment
- cell functions can be monitored (note not to extrapolate in whole to living organism!)
- replace in vivo animal experiments (partly)

Transfection of cell cultures

Transfection of cell cultures is the incorporation of a DNA fragment to the cell, with the aim to investigate gene function and control, and protein function. At **stable transfection** the foreign genetic material integrates into the host genome, and is expressed constantaneously. **Transient transfection** means the foreign gene does not integrate into the host genome, causing limited time gene expression.

Methods for transfection:

- biological (transduction): virus-mediated process using eukaryotic expression vectors (plasmids)
- chemical: complexes of negatively charged nucleic acids and positively charged molecules are able to connect the negatively charged membrane, and get into the cell
- physical: microinjection, electroporation, laser-based transfection

Transfection in steps using eukaryotic expression vectors

- Target gene (the foreign sequence) is cloned into an expression vector
- Vectors are amplified in competent bacteria
- Bacteria having vectors are selected based on a selection marker
- Plasmid DNA is isolated from selected clones
- Transfection of host cells with plasmid
- Detection of gene expression/protein synthesis changes

siRNAs

Small interfering RNAs (siRNAs) are short, 20-25 base pair double-stranded RNA molecules playing role in gene silencing. siRNAs specifically bind to complementary mRNA and induce its degradation. This way it interferes with the expression of those genes. RNA interference is a hot spot of today's science, as this also provides a potential for synthetic siRNAs to be used in therapy.

SPECIAL METHODS IN GENETICS

Chromosome number and structure, labeling techniques

Chromosomal mutations are changes on the chromosomes, which form rearrangement of chromo–somes (structural changes), or abnormalities in number of chromosomes and parts of chromosomes (numerical changes). We can identify mutations with the tools of cytogenetic: genetically and microscopically.

Most chromosome abnormalities occur as an accident in the egg or sperm, therefore, the abnormality is present in every cell of the adult body. However some abnormalities can be formed after conception, resulting mosaicism, where only a part of the cells will carry abnormal genetic material.

Chromosome abnormalities can be inherited from a parent (such as a translocation) or be "de novo" (newly formed). This is why chromosome studies are often performed on parents when a child is found to have an abnormality.

Cytogenetics

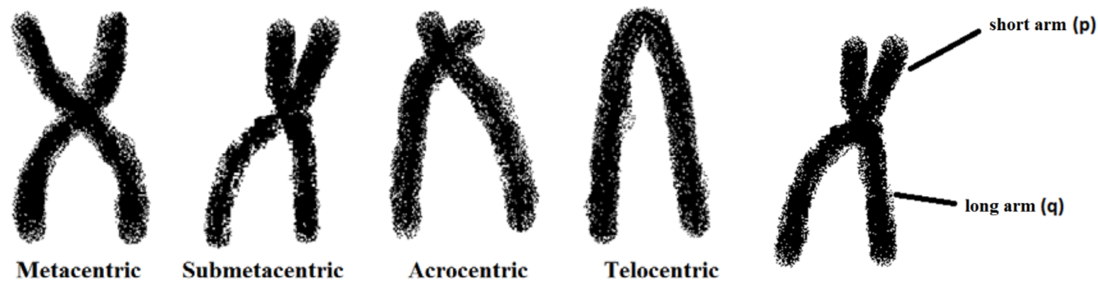
Cytogenetics is a discipline in biology linking the study of **cytology** with **genetics**. This branch of biology deals with analysis of chromosomes, especially detection of inheritable diseases. Cytogenetics analyze normal and pathological chromosomes: their **number**, **structure**, **phenotype** and the relationship between these, cytogenetics describe chromosomal disorders and their causes.

For this we usually use banding techniques and in situ hybridization techniques to analyze the chromosome set of a cell and draw conclusions.

We usually examine condensed chromosomes and because of this we need cells in metaphase of mitosis. (Remember: at metaphase of mitosis DNA has been doubled and it is highly condensed, sister chromatids are together.) We can use for the analyzis spontaneously proliferating cells (bone marrow, tumor cells), or we can induce the proliferation of other cells with mitogenic substances (e.g.: we can induce periferal white blood cells). After it we can obstruct specific mechanism of the cell division e.g. we can obstruct the formation of mitotic spindle with colcemid or colchicine so it stop the celll division on metaphase.

Chromosome morphology

We can classify chromosomes into morphological groups according to the relative position of centromeres. Because of the rate of the condensing the lenght of the chromosomes can differ but the proportion of the parts is always the same.



Banding techniques

The morphology of different chromosomes, the lenght of the q and p arms can be very similar, so the identification of the native chromosomes can be very difficult. Because of this we use the so called „banding" techniques. We can stain chromosomes more ways and form bands on it. We number bands from the centromeres by Arabic numbers.

The number, the shape and the banding of the chromosomes are characteristic and species-specific. Of course, there is chromosome-heteromorphism (there are small differences), but the differences between the healthy karyotypes in the same species are very small.

Karyotype (chromosome set): is the description of individual chromosomes, that describe the complete set of chromosomes in the individual cells.

By a karyotype analysis, we fix cells in metaphase of mitosis (as described above). After we stop the cell division in metaphase, we use hypotonic conditions, the cells swell (become bigger), so the individual cells (and the chromosomes in the cell) will be further from each other that will make the examination easier.

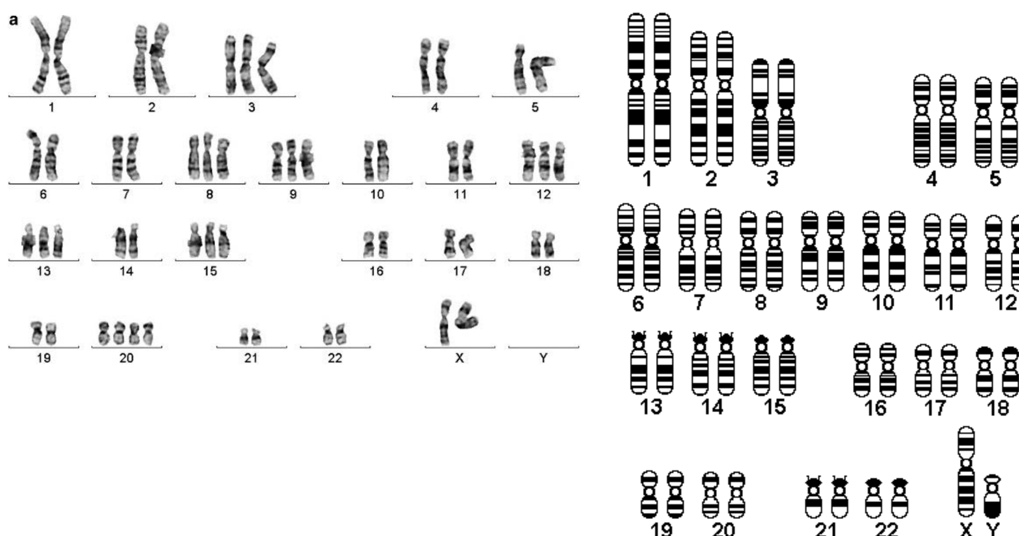
Some banding techniques:

- Q-banding: staining with fluorescent dyes, e.g.: quinacrine (quinacrine hidrochlorid)
- G-banding: staining with giemsa after digestion with trypsin
- C-banding (=reversed Giemsa): treating the chromosomes with acid and alkali solution and staining with Giemsa.

Karyogram vs ideogram

Karyogram: a photo about a cell's chromosome set: where the chromosomes are arranged according to the size, shape, and banding.

Idiogram: a schematic picture (drawing/graphic) about a cell's chromosome set: an ideal picture, how chromosomes should look like.



Molecular cytogenetics

In situ hybridization

In this technique we add known, marked oligonucleotides which can bind to the complementary sections on the sample DNA. If these marked oligonucleotides can bind, this verify the existance of the complementary sections on the sample. Benefit of this technique is that we can use this technique with not condensed chromosomes too. If we use this technique with condensed chromosomes we can see where the complementer sections are.

This technique is much more sensitive than the banding technique. (But much more expensive too.) With banding technique we can see changes which are bigger than 2 Mbp. With in situ hybridization we can see a few bp rearrangement too. At the beginning there was isotopic labeling, nowdays we use mainly fluorescent dyes to mark the oligonucleotides. (*See FISH*)

SPECIAL METHODS IN PHARMACEUTICAL RESEARCH

Structure based drug design

The **basic steps** of structure based drug design are the following: to choose, clone and purify the target molecule; to determine the structure of the target; to find the proper region of the target for interactions; with computer programs find compounds which bind the selected part of the target molecule; synthesis of the lead compounds; biological assays of the interacting compound; structure of the target-lead complex and possible optimization of the lead molecule.

For **choosing a drug target** molecule, we have to rely on biological and biochemical knowledge. To use structure based drug design the target molecule has to have connection with a human disease and it should bind a small molecule for activity. This small molecule is binding to the target in a specific binding site so other, mimicking molecules can compete for this binding. Good targets for structure based drug design are G protein coupled receptors, ion channels, proteases, kinases and hormone receptors, to mention the most frequent ones. In case of human disease the drug should modulate the activity of the target molecule and possibly leave the function of the normal cells intact. A drug against a pathogen should complete inhibit the pathogen organism, so their targets are essential, pathogen specific molecules.

The **structure of the target** is determined usually by using the crystal structure of the molecule. If this cannot be used for some reason, a homology model can substitute. To locate a ligand binding site (an active site of an enzyme or a site for communication for example), a pocket with potential hydrogen bond donors/acceptors or hydrophobic interactions may be found. There are **two basic methods** for creating molecules with target binding activity: experimental or computer based. The latter one has **three categories**: inspection, virtual screening and de novo generation. Inspection means that the known binding partner of the target molecule is modified to generate inhibiting activity. Virtual screening is a process, where databases of small molecules interact with the target in silico. Or small fragments can be generated in silico and find the best interacting ones with computer programs. The potential lead compounds have to be synthesized and tested. Before this step the candidate molecules have to be evaluated: affinity to the target, oral bioavailability, stability, process of synthesis.

Combinatorial chemistry

The aim of the method is to generate large chemical libraries in order to find lead compounds for drug discovery. These libraries are collections of different but related molecules which can be generated synthetically and screened in one time for biological activity. The method can be used to find new lead for a defined target, or to find a better modification of an original lead. The essence of the method is to connect covalently different building blocks to create a large variation of the structure. These reactions may occur in solution or on a solid support. According to the mechanism it can be parallel synthesis or split and mix synthesis. In the parallel synthesis method each material is reacted with every small group, then the reaction is split into a number, which is the number of the next reacting building block. In the split and mix method after every reaction step the products are pooled, mixed, then separated into reactions with the next reactant – these steps are repeated more times resulting in large and complex libraries.

High-throughput screening (HTS)

The number of drug targets and the number of potential drug compounds increased so largely that the everyday screening and assay methods had to change lately. This change was helped by the miniaturization, and automation techniques. Testing 10 000 – 100 000 compounds in an assay system per day is the definition of HTS. Before this time the use of the 96 well plates provided the fastest results. Nowadays 384 and rather 1536 well plates are used for screening. This has the advantage of applying small reaction volumes as well. The regular target families (enzymes, ion channels, GPCRs) are complemented with new ones: transporters, receptors, signaling pathways, protein-nucleic acid interactions, DNA/RNA interactions, which are not only biochemical but also cell based targets.

Because of the increase in library sizes there are new processes developed in drug discovery. New tools are introduced in the screening methods (recombinant cell lines for example). There are different types of screening: counter-screening is the procedure when artifacts are selected, while selectivity screens are carried out among the target and other targets. These days in drug discovery the understanding of the mechanism of action is more important than before. More energy is paid to the lead optimization section of the work and to gain biological active compound from the screening. New trend is that not the whole library is screened always but smaller groups of the best ranked compounds (“focused libraries”) are used which are examined to minimize the number of artifacts in a solution.