Oslo universitetssykehus

UNIVERSITY OF OSLO

Structural biology and drug design

Bjørn Dalhus

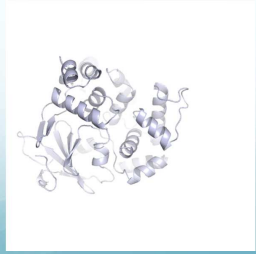
Research Scientist, Department for Medical Biochemistry & Microbiology,
Oslo University Hospital, Rikshospitalet

E-mail: bjornda@medisin.uio.no

Telephone: 23070908

Group: Magnar Bjørås (http://www.rr-research.no/bjoras)

Main research area: DNA repair & Structural Biology



Outline of lecture

Part 1

1. Group presentation
   1. DNA repair & methods
   2. Core facility
2. Drug design
   1. From idea to market
   2. Strategies
   3. Target selection and validation
3. Role of structural biology in drug design
   1. 3D structure determination (X-ray and NMR)
   2. Experimental challenges
   3. 3D models – strengths and weaknesses (demo)

Part 2

4. Structure-based drug design
   1. Ligand based methods
   2. Receptor-based methods
5. Docking
   1. Ligands; databases, ligand preparation
   2. Target preparation
   3. Running a docking job
   4. Scoring and validation

Outline of lecture

Part 1

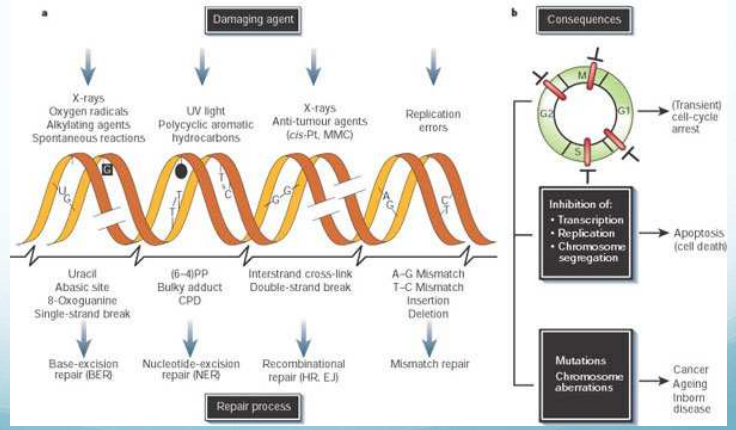
1. **Group presentation**
   1. **DNA repair & methods**
   2. **Core facility**
2. Drug design
   1. From idea to market
   2. Strategies
   3. Target selection and validation
3. Role of structural biology in drug design
   1. 3D structure determination (X-ray and NMR)
   2. Experimental challenges
   3. 3D models – strengths and weaknesses (demo)

Part 2

4. Structure-based drug design
   1. Ligand based methods
   2. Receptor-based methods
5. Docking
   1. Ligands; databases, ligand preparation
   2. Target preparation
   3. Running a docking job
   4. Scoring and validation

DNA damage

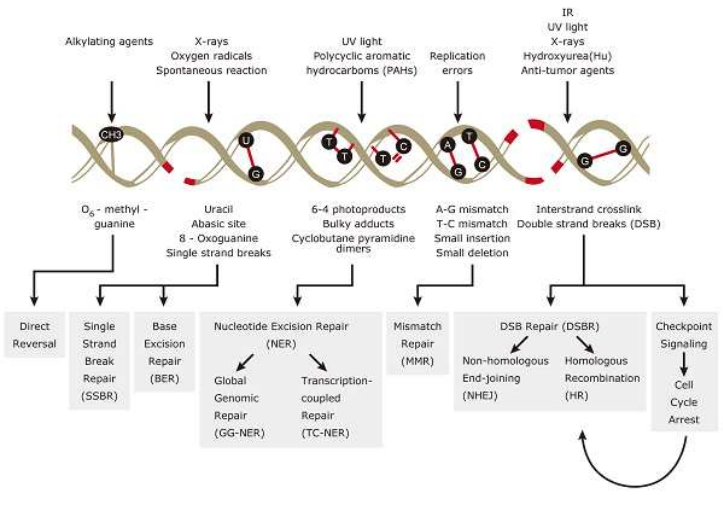
- Since DNA carries all genetic information, one would think that DNA is a chemically stable molecule
- However, there are betweeen 10.000 – 100.000 damages per genom per cell per day in humans
- Ideally, all these damages must be detected and/or removed/repaired



Hoeijmakers, J. H. J. Genome maintenance mechanisms for preventing cancer. *Nature* **411,** 366–374 (2001)

DNA repair

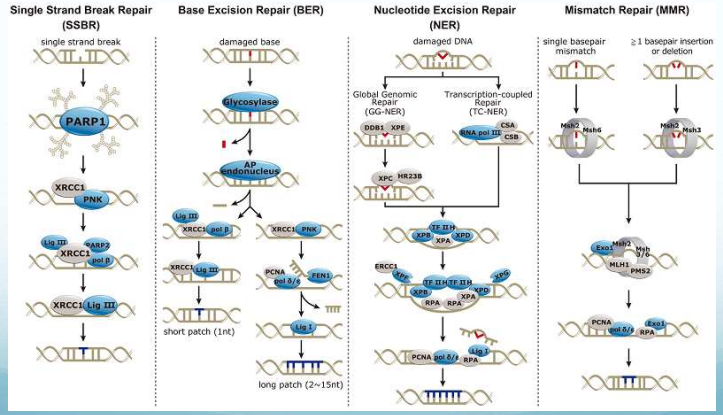
- At least 8 different DNA repair mechanisms have evolved



www.genetex.com

DNA repair pathways

- Repair pathways involve many proteins/enzymes, sometimes also with backup systems
- Failure in any essential step may lead to cancer
- But DNA repair proteins/enzymes are also targets for cancer treatment ….
- … because DNA repair is a <u>resistance factor</u> in current therapies



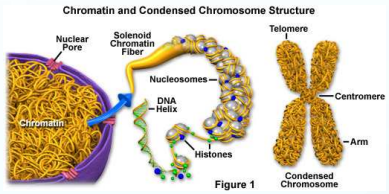
Our research

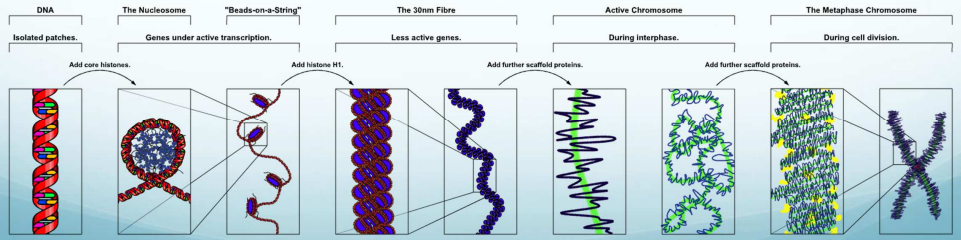
- **Aims and methods**
  - Look at DNA repair in model organisms (*E. coli*, yeast, mouse), particularly by making knock-out mutants where DNA repair genes are removed
  - Determine and identify proteins involved in DNA repair or other processes to maintain genomic stability
  - Biochemical charactization of repair proteins/enzymes
  - Structure determination of protein/DNA complexes

- **Questions we want to answer**
  - How can enzymes detect damages in DNA?
  - How does these proteins work at the atomic/molecular level?
  - What is the biological/biochemical role of partner molecules?
  - What is the effect of mutations in these genes on the function?

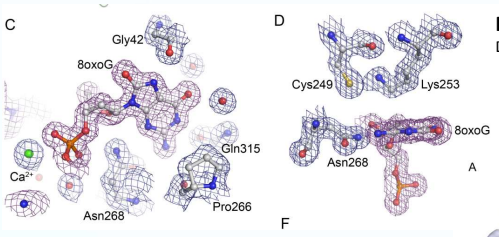
Chromatin – the haystack

- **Chromatin is a protein-DNA mixture**
  - Euchromatin = ”loose” complex with active genes
  - Heterochromatin = ”dense” complex with silent genes
  - Condensed chromatin = chromosome structure

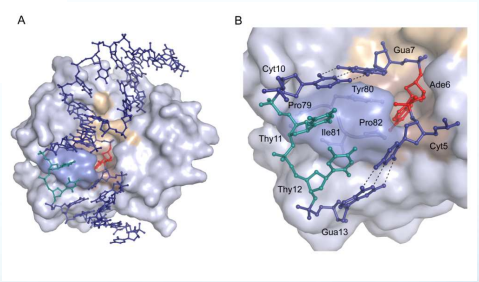




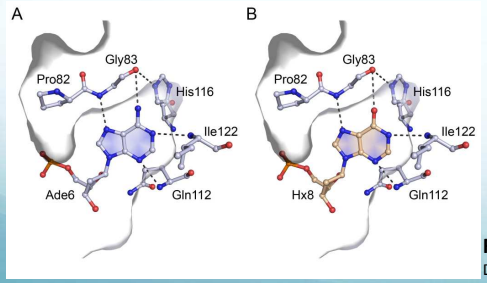
Examples



Recognition of 8-oxoG in OGG1
Dalhus et al, Structure, 2011



Recognition of loop by EndoV
Rosnes et al, Structure, 2013



Recognition of deaminated Ade by EndoV
Dalhus et al, Nature Struct Mol Biol, 2009

Core facility for structural biology and bioinformatics

- We are responsible for a joint MLS^{UiO} & HSØ Regional Technology Platform
  - Financed by HSØ (2012-2014) and MLS^{UiO} (2012-?)
  - Regional service, but mainly within OUS and UiO
  - **Personnel**: Alex Rowe (Optical tweezers), Bjørn Dalhus (PX, SAXS, docking, modeling), Jon K Lærdahl (Bioinformatics, modeling), Paul H Backe (PX, molecular biology), Pernille Strøm-Andersen (Protein purification), Rune J Forstrøm (SPR/Biacore) & Torbjørn Rognes (Bioinformatics, sequence analysis)
- Two types of interaction modes
  - Access to instruments (fees)
  - Collaborative projects with co-authorship (free of charge)

Services/Methods

- Experimental methods
  - Structure determination by crystallography (PX) and SAXS
  - Interaction studies by SPR (Biacore)
  - Single particle imaging/manipulation (Optical tweezers)
  - Protein expression, purification
  - General molecular biology (e.g. cloning, mutagenesis)
- Analysis and modeling/computational methods
  - Sequence analysis
  - Structural modeling
  - Interpretation of clinical data with respect to structural models
  - Protein-protein and protein-ligand docking

Outline of lecture

Part 1

1. Group presentation
   1. DNA repair & methods
   2. Core facility
2. **Drug design**
   1. **From idea to market**
   2. **Strategies**
   3. **Target selection and validation**
3. Role of structural biology in drug design
   1. 3D structure determination (X-ray and NMR)
   2. Experimental challenges
   3. 3D models – strengths and weaknesses (demo)

Part 2

4. Structure-based drug design
   1. Ligand based methods
   2. Receptor-based methods
5. Docking
   1. Ligands; databases, ligand preparation
   2. Target preparation
   3. Running a docking job
   4. Scoring and validation

Drug design – from idea to market

- **What is a drug?**
  - Bioactive compound used to treat, cure, diagnose or prevent a disease
  - Most drugs are small organic compounds, but there are drugs that are also larger molecules such as proteins (insulin) or protein fragments (vaccines)

- **What are the targets of small-molecule drugs?**
  - Most small-molecule drugs bind to proteins, but there are examples of drugs that bind to e.g. DNA (cancer therapy) or RNA (ribosomes; antibiotics)
  - Different classes of proteins are suitable drug targets

    Enzymes (inhibit, or someetimes also enhance, the activity)

    Membrane receptors (Signal blocking)

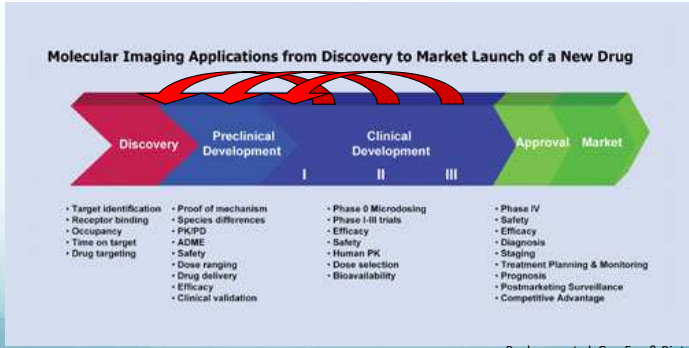
    Non-enzymatic proteins (stablize protein/complexes)
  - Ca 1/3 of current drugs target membrane proteins, GPCRs in particular
  - Proteins involved in signal transduction pathways are attractive targets since these processes are key elements in the pathology of cancer, inflamation, cardiovascular, metabolic and neuropsyciatric diseases: **GPCRs**, **protein kinases** and **nuclear receptors**.

Drug design – from idea to market

- **Selection of diseases and targets**
  - Development of new drugs are very time consuming (10+ years) and expensive (1-2 billion $) [Paul et al, Nature Reviews, 2010]
  - Companies have to earn back this kind of money in a short period after the drug has been approved and the patent expires – the time window is short
  - The best way to make money is to design a drug that many patiens needs to use at a regular basis, or treats serious life-threathening conditions in a rich population
  - For these reasons, many rare diseases, or social/patient groups, are not interesting for the industry
  - Companies are quite conservative when selecting targets, and the process is focused on elliminating ”dead ends” as soon as possible before costs start running high
  - Failure in the late stages of a project are particularly expensive
  - Companies often work with many targets in parallell for the same disease

Drug design – from idea to market

- **Major steps in the process**
  - Target selection & **hit discovery** (biochemistry)
  - **Hit to lead development** (chemistry)
  - In vivo testing (ADME-Tox) (pharmacology)
  - Clinical phases (I, II and III)
  - Notably, the process becomes cyclic for each set-back in the project



Buchanan et al, Gen Eng & Biotech News, 2007

The entry point

- **Where to start?**
  1. Start from scratch with a broad screen of a library (industry standard)
  2. Start with a known drug/lead and develop it further (better effects; less side effects; increased specificity; improved kinetics etc)
  3. Observations of biological effects of natural products

- **HTS – High-throughput screening**
  - Pharmaceutical companies have large libraries of molecules (millions of compounds) they can use in automated screens to discover **hits** for a new target
  - Academic groups have also access to smaller libraries for semi-HTS (typically 50.000 compounds)

- **Virutal Screening (docking)**
  - Alternative computational route to discover hits for a given target
  - More often used by industry as part of the modeling process in lead development/optimization rather than hit discovery
  - **Drawback**: requires knowlede of the 3D structure of the target protein

Outline of lecture

Part 1

1. Group presentation
   1. DNA repair & methods
   2. Core facility
2. Drug design
   1. From idea to market
   2. Strategies
   3. Target selection and validation
3. **Role of structural biology in drug design**
   1. **3D structure determination (X-ray and NMR)**
   2. **Experimental challenges**
   3. **3D models – strengths and weaknesses (demo)**

Part 2

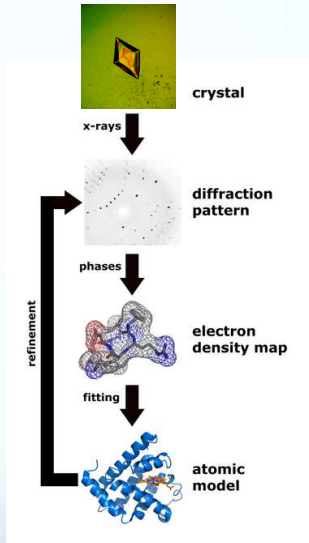
4. Structure-based drug design
   1. Ligand based methods
   2. Receptor-based methods
5. Docking
   1. Ligands; databases, ligand preparation
   2. Target preparation
   3. Running a docking job
   4. Scoring and validation

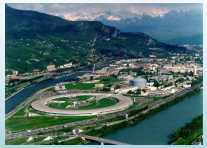
Protein structure determination

- **Virtual screening requires a known 3D structure of the target protein**
  - The aim of virtual HTS screening (docking) is to evaluate the fit between
  - a set of small molecules in a library and the drug target
  - To calculate interaction energies, surface contacts and steric volume complementarity the positions of all relevant atoms in the protein and ligand are an absolute requirement
- **Two methods to determine 3D structures of proteins at high resolution**
  - Protein crystallography
  - Protein NMR
  - Each method has advantages and drawbacks, but crystal structures are normally preferred over NMR structures for docking if availible
  - Both methods gives a 3-dimensional model of the protein with coordinates of the atoms that form the protein

Protein crystallography - basics

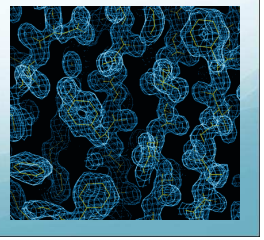
- **Solving a protein structure by X-ray crystallography involves several steps**
  - Make protein crystals
  - Expose crystals to X-rays (typically at a synchrotron facility)
  - Collect diffraction images
  - Process images to calccuate the intensity of the spots
  - Calculate an electron desity map from the spot intensities (Fourier transform)
  - Build atoms and amino acids into the density map
  - Analyse the model with molecular graphics software





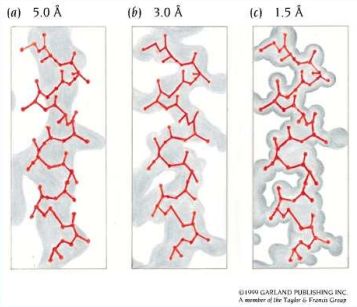
Protein crystallography

- **Requirement: Protein crystals**
  - The method is absolutely depending on formation of protein crystals
  - Many proteins form crystals under optimal conditions (which are normally not known in advance, hence crystal screening is nessesary)
  - Thousands of combinations of diffrent protein forms and crystallization conditions may have to be tested
- **Diffraction of X-rays by crystals**
  - X-ray radiation will scatter from electrons in molecules, and with a symmetrical arrangement of molecules in a crystal, the scattering forms a characteristic diffration pattern
  - Analysis of this pattern makes it possible to calculate the **electron density** inside the crystal – hence the position of the atoms can be determined

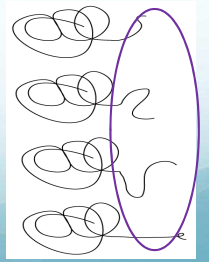


Protein crystallography

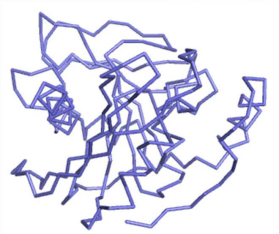
- **Resolution**
  - The structural models have different levels of accuracy and detail – known as resolution
  - For docking purposes, the resolution must be high enough to determine the accurate positions of all relevant side chains in the "active site"
  - Resolution is measured in Ångstrøm, where low numbers means high resolution



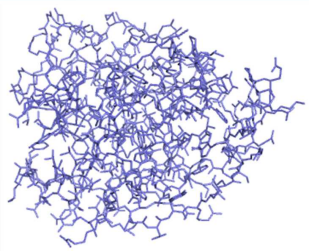
- **Limitations**
  - There is no size limitation to the method as long as crystals of the particles can be formed – e.g. ribosomes and viruses can also be crystallized
  - Only odered parts of the molecules can be modeled



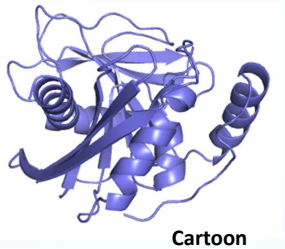
Representation of protein structures



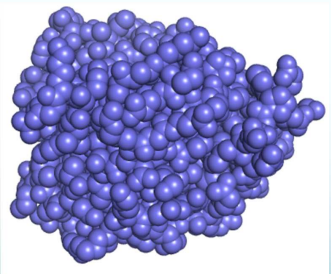
Cα-trace



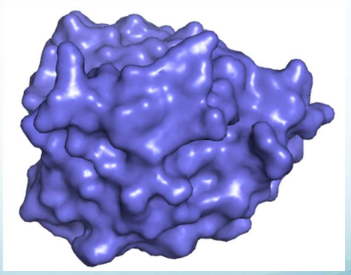
Lines/wire-frame



Cartoon



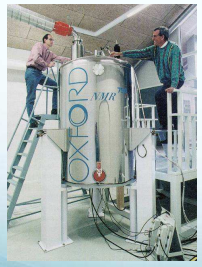
Space-filling spheres (CPK)

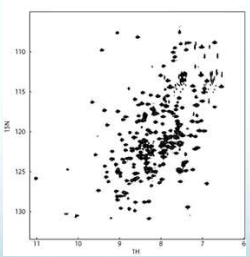


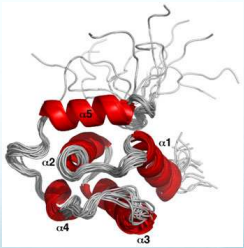
Surface

Protein NMR

- The method is based on signals from atoms (^{13}C or ^{15}N) in a strong magnetic field
- The method determines distances between atoms
- Several models can be built that will satisfy the distance matrix
  → which model is correct?
- The method has an upper-size limit (typically < 200 residues)
- The method gives information about dynamics/flexibility
- Experiments are performed in solution – no crystals needed



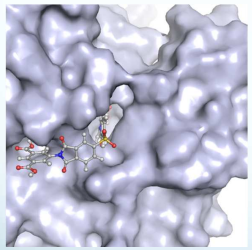


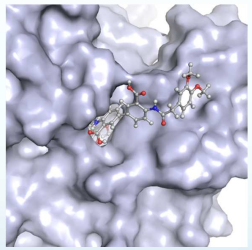


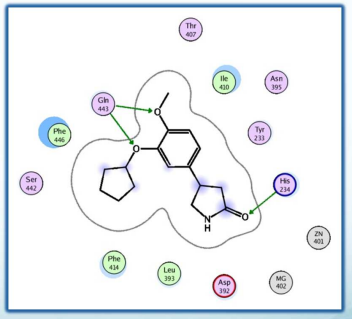
M. Tuberculosis Rv0543c
PDB code: 2KVC

Amino acids – a reminder

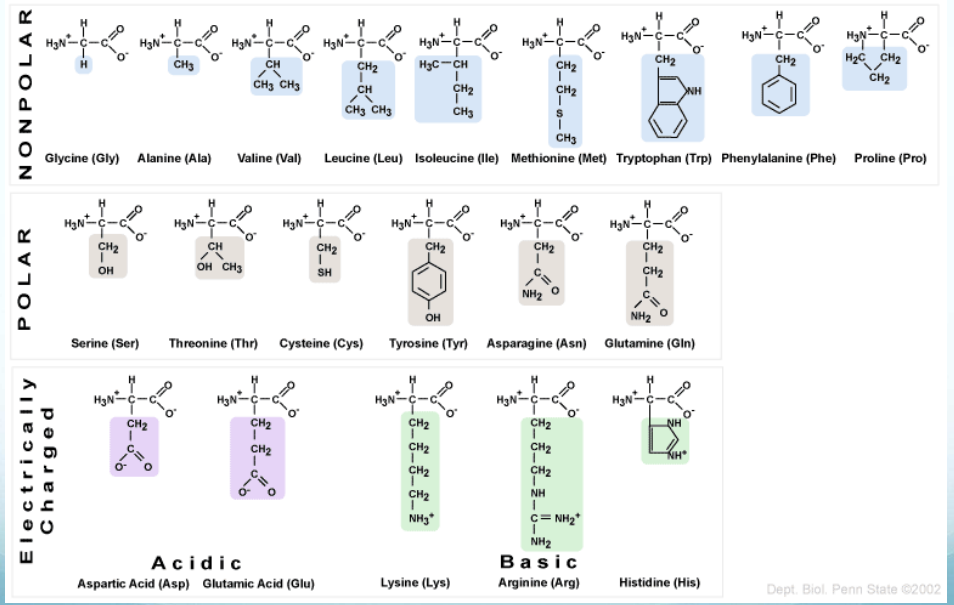
- Protein surfaces are mainly formed by amino acid side chains
- The chemical properties of the amino acid side chains are essential for intercations between the protein and the drug molecule
- Detailed knowledge of preferred interaction partners makes drug design an excercise in "match-making" – who's the best partner ?







Amino acids – a reminder



Mark Troyan, Penn State Univ.

Outline of lecture

Part 1

1. Group presentation
   1. DNA repair & methods
   2. Core facility
2. Drug design
   1. From idea to market
   2. Strategies
   3. Target selection and validation
3. Role of structural biology in drug design
   1. 3D structure determination (X-ray and NMR)
   2. Experimental challenges
   3. 3D models – strengths and weaknesses (demo)

Part 2

4. **Structure-based drug design**
   1. **Ligand based methods**
   2. **Receptor-based methods**
5. Docking
   1. Ligands; databases, ligand preparation
   2. Target preparation
   3. Running a docking job
   4. Scoring and validation

What does a typical drug look like?

- **Lipinski's rule-of-five for drug-like molecules**
  - Looking at small-molecule drugs, there are some physical/chemical properties that are common to most drugs on the market
  - These properties makes it possible to judge if a given molecule is "druglike"
  - These properties relates to the pharmacokinetics / ADME properties of a molecule
  - Lipinski's rule-of-five states that an <u>orally active drug does not violate more than one</u> of the following criteria
    1. Not more than 5 hydrogen bond donors
    2. Not more than 10 hydrogen bond acceptors
    3. A molecular mass below 500 Dalton (g/mol)
    4. An octanol-water partition-coefficient logP not greater than 5
  - These rules ensures that the compound is not too soluble and polar, so that it can cross biological membranes, like cell walls, and also gets decomposed in the kidneys

Structure-based drug design (SBDD)

- Also known as structure-based ligand design (SBLD)
- Set of <u>methods/strategies</u> with <u>associated tools</u> used to find/design ligands that binds to a given protein (called the target/receptor) where structural information is used to guide the process
- Typically divided into two sub categories
  - Ligand-based design
  - Receptor-based design

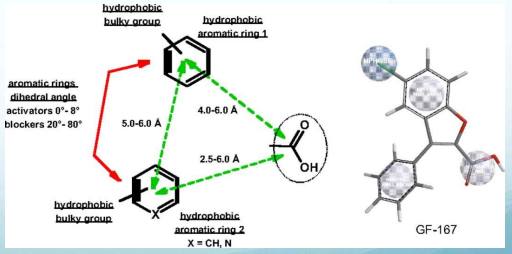
  depending on the availability of structural information

- **Ligand-based design**
  - Used when several inhibitors/ligands are already known
  - Tools/methods: similarity search, pharmacophore design, QSAR

- **Receptor-based design**
  - Used when the atomic structure of the receptor is availible
  - Can be used without prior knowledge of any inhibitors

Ligand-based methods

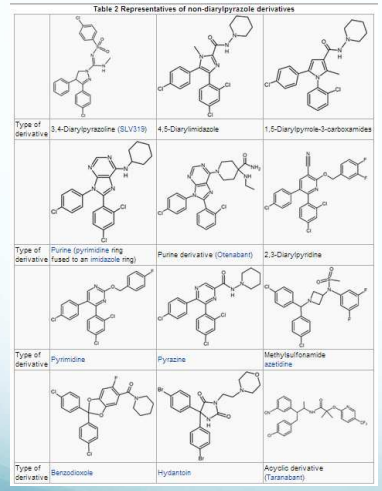
- Ligand based methods are used when several inhibitors are known.
- The structures of the ligands are used to derive a **pharmacophore model**
- If biological data is availible (e.g. ”activity”, binding affinity) a **QSAR model** can be designed
- **Pharmacophore**
  - 3D description of chemical properties of a set of ligands/inhibitors
  - The model should reflect the ”least common” steric and electronic features expected to be important for binding
  - The pharmacophore model can be used to <u>search databases for molecules with similar properties</u>



Pharmacophore model and ”master” ligand GF-167 for K Cl channel; Liantonio A et al. PNAS 2008;105:1369-1373

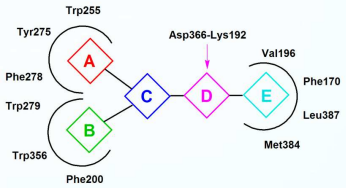
Pharmacophore - example

- Rimonabant – an inhibitor of CB_1 receptor for treatment of obesity and overweight
- Acts on the cannabinoid receptor

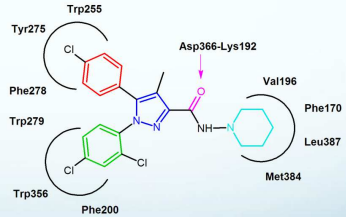


Wikipedia

Pharmacophore model



Rimonabant

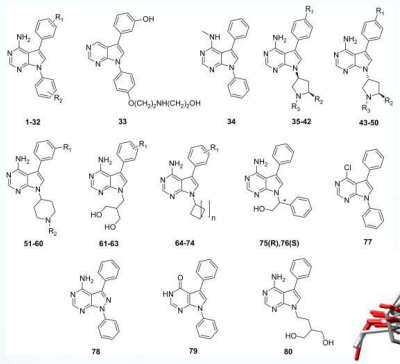


QSAR – Qantitative Structure-Activity Relationship

- More advanced ligand-based drug design strategy/method
- In addition to a set of ligands, biochemical activity data must be availible
- Uses statistical methods (multi-variable regression) to relate a set of "predictor" variables (X) to the potency of the response variable (Y)
- In QSAR, <u>X is the “molecular structure”</u> of the ligands, and <u>Y is the biological response</u>
- For a good QSAR model, the biological response for new molecules can be estimated from the “molecular structure” without testing

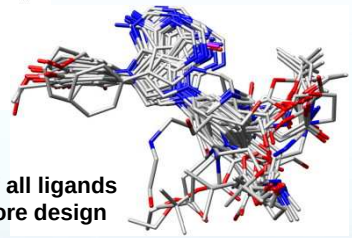
- The difficult part is to quantify the “molecular structure” in a meaningful way which describes the molecules
- Examples of terms in the formula are
  - Items in the Lipinski’s rule-of-five list
  - pKa-values of acidic groups, solubility, distance between pharmacophore groups
  - Electronegativity, polarizability, etc
- Known ligands are divided into a training and a validation set. Only molecules in the training set are used in the regression, while the validation set is used to check the model
- When the model is validated, it can be used to predict activity for untested compounds

QSAR - example

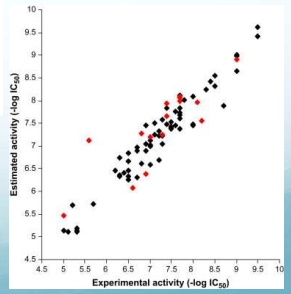


Test set molecules with quantified activity

Tintori et al, Eur J Med Chem, 2009, 990-1000



Superposition of all ligands for pharmacophore design



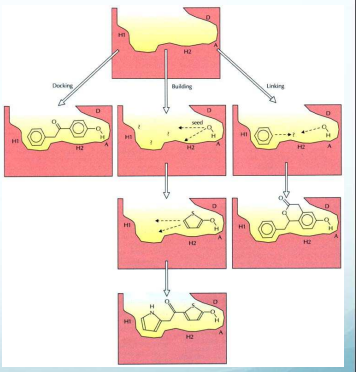
Final QSAR model predicion (black = training set; red = cross-validation set)

Receptor-based methods

- Receptor-based drug design uses structural information of the target/receptor to predict which molecules that will bind to the protein
- Here, we will discuss two different approaches
  - Fragment-based drug design
  - Docking (also known as high throughput virtual screening, HTVS)

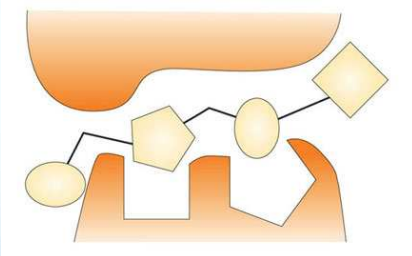
- **Fragment based drug design**
  - Experimental technique (X-ray, NMR, ITC, SPR)
  - Uses small fragments typical for drug-like molecules

- **Docking (HTVS, high-throughput virtual screening)**
  - Computational technique
  - Uses databases with drug-like "whole" molecules
  - Results depend on algorithm and scoring function

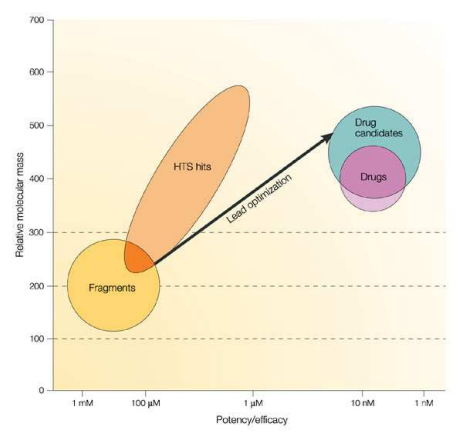


Fragment-based drug design

- One problem with docking is that the average size of hits are in the upper range according to the Lipinski's rule-of-five
- This means, that in order to develop the molecule further, some parts have to be removed before new parts will be added
- Fragment-based drug-design tries to circumvent this problem by using smaller fragments of drug like molecules



Typical hit from HTVS – no really good fit, all regions must be optimized (one by one)

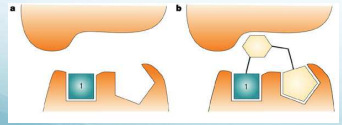


Rees et al, Nature Reviews Drug Design (2004) 660.

Fragment-based drug design

- **Advantages**
  - Reduced number of molecules to be tested (500 fragments = 125.000.000 compounds with three groups)
  - Easier to extend and gain affinity than to reduce and keep affinity
  - Larger inhibitors can be designed by fragment expansion or fragment connection
- **Disadvantages**
  - Can't dock fragments, their affinity must be determined experimentally
  - Their position(s) must be determined using X-ray or NMR structure

Fragment expansion



Fragment connection

Rees et al, Nature Reviews Drug Design (2004) 660.

Outline of lecture

Part 1

1. Group presentation
   1. DNA repair & methods
   2. Core facility
2. Drug design
   1. From idea to market
   2. Strategies
   3. Target selection and validation
3. Role of structural biology in drug design
   1. 3D structure determination (X-ray and NMR)
   2. Experimental challenges
   3. 3D models – strengths and weaknesses (demo)

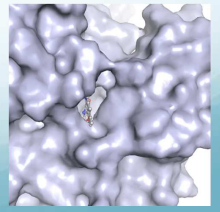
Part 2

4. Structure-based drug design
   1. Ligand based methods
   2. Receptor-based methods
5. **Docking**
   1. **Ligands; databases, ligand preparation**
   2. **Target preparation**
   3. **Running a docking job**
   4. **Scoring and validation**

Docking – HTVS (High-throughput virtual screening)

• Docking is used to search for possible small molecules that can bind to active sites/pockets on protein surfaces
• The method is an alternative approach to experimental high-throughput screening normally used in the pharmaceutical industry to find lead compounds that can be modified to make new drugs
• The method is based on energy calculations (scoring function) that ranks the small molecules according to the predicted interaction energy

• A library of 100.000 – several million compounds can be screened in days (parallel computation systems)
• Each moelcule in the library is docked onto the protein surface one-by-one
• The best candidates can be selected for experimental testing in enzyme assays or cell based studies
• Determination of the structures of these complexes may give clues about further improvements



Docking – the method(s)

- **Ligand**
  - Each ligand can exist in several conformations
  - Normally, databases must be converted from 2D to 3D structures
  - A ligand with, say 5 rotable bonds, may form as many as 243 conformations (3^5)
  - With 500.000 compounds, each with 200 conformations, makes > 100.000.000 ligands; with 10 sec per docking, a screening would take >200.000 CPU hours ….
  - Typically, the ligand database is processed once, and only a few (or one) low-energy conformation is stored for docking (rigid ligand)
  - Some programs splits the ligand into fragments, places the core first, and then builds the ligand by fragment extension (simulating flexible ligand)

- **Receptor**
  - The receptor must be prepared for docking
  - The docking site must be defined
  - A grid is calculated within a box surrounding the docking site
  - The grid stores information about receptor shape/volume, charge and hydrophobicity
  - Some programs store several grids for different side chain conformations (simulating protein flexibility)
  - Each ligand is matched with the properties in the grid
  - A score is calculated for each ligand and listed in a scoring table

Docking – what's the catch

- **Scoring function**
  - All docking programs uses a scoring function that tries to calculate the binding affinity of the ligand
  - Scoring functions are only approximate (there's not time for a full quantum-mechanical calculation of the energy)
  - Entropy and (de)solvatisations are very difficult to estimate, yet these are very important factors for the binding affinity
  - For instance, entropy may be approximated by the number of rotable bonds and solvatisation by the area of exposed hydrophobic surfaces
- **Ligand**
  - Charges must be assigned
  - The correct low energy conformation must be found
  - Correct protonation states of titrable groups must be set

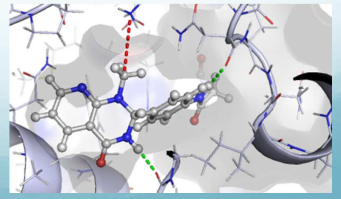
- **Receptor**
  - The correct (or best) docking site must be defined
  - The correct side-chain conformation must be determined
  - Should different forms of the protein be used? Induced fit? Flexibility?

Docking - evaluation

- **After docking – what then?**
  - Any docking program/algorithm will produce a scoring list with some top-ranking compounds
  - The challenge is to know what's "rubbish" and what's relevant

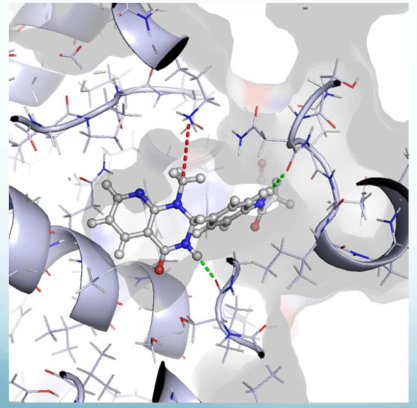
- **Hit rate**
  - Expect a hit rate as low as 1-5 % (i.e. only 3 of 100 top scores are true inhibitors)
  - Each solution must be inspected manually, and judged (Lipinsky & gut feeling/experience)
  - Select diverse molecules from the hit list for testing

- **Docking is like a funnel**
  - Docking can remove a lot of unproductive compounds from the database
  - More difficult to find the best ligands

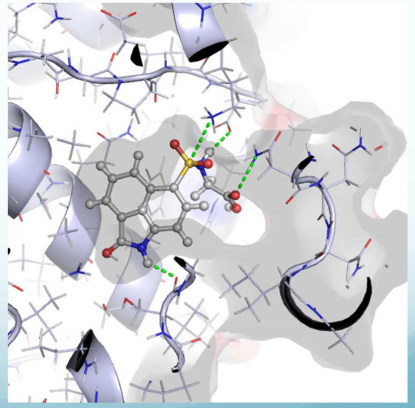


Docking – a final example of ”hit picking”

Probably bad



Hopefully good



Docking programs

- **Dock 6.5**
  - http://dock.compbio.ucsf.edu/DOCK_6/index.htm
  - Free
  - Flexible receptor and molecular dynamics
- **Autodock**
  - http://autodock.scripps.edu/
  - Free
  - Flexible receptor (side chain rotamers)
- **DockBlaster & Zinc database**
  - http://blaster.docking.org/
  - Free, online, version of Dock 3.6
  - Rigid receptor only
- **Commercial docking programs**
  - Glide from Schrödinger
  - Gold from CCDC
  - FlexX from BioSolveIt

Thanks for your attention!