

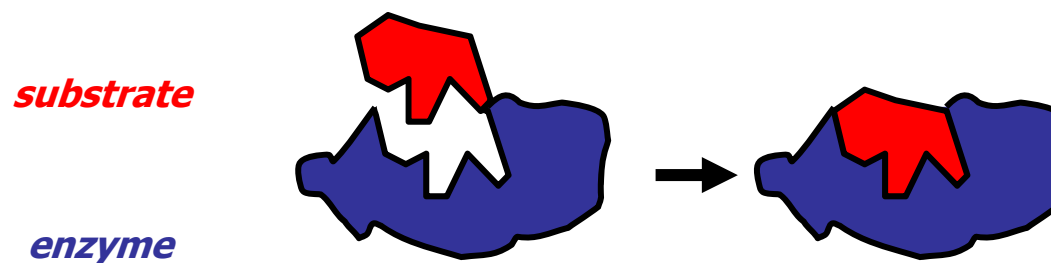
# Modeling the 3D structure of the ligand – protein complexes

E. Kellenberger

Structural Chemogenomics laboratory  
Medalis Drug Discovery Center  
Faculté de Pharmacie, Illkirch

# The lock & key principle

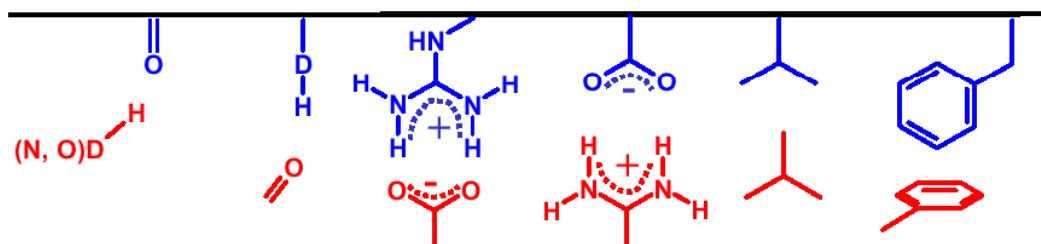
## Geometrical Complementarity



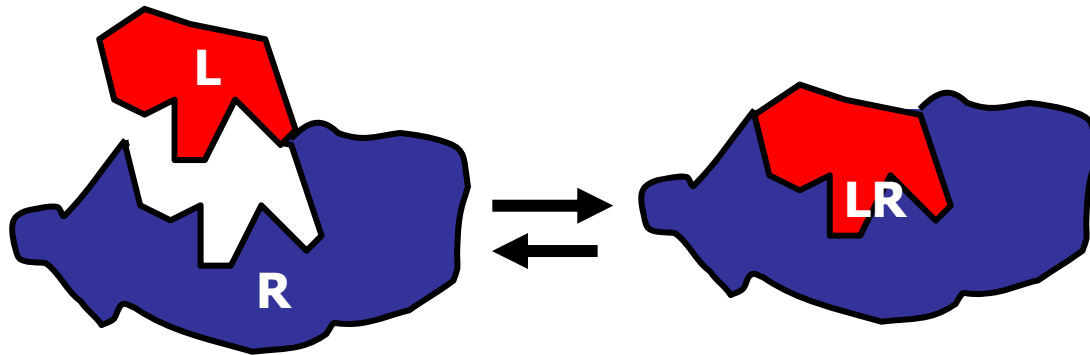
## Non-covalent inter-molecular interactions

*protein*

*ligand*



# Thermodynamics of the association



**Association constant (equilibrium)**

$$K_A = \frac{[LR]}{[L][R]}$$

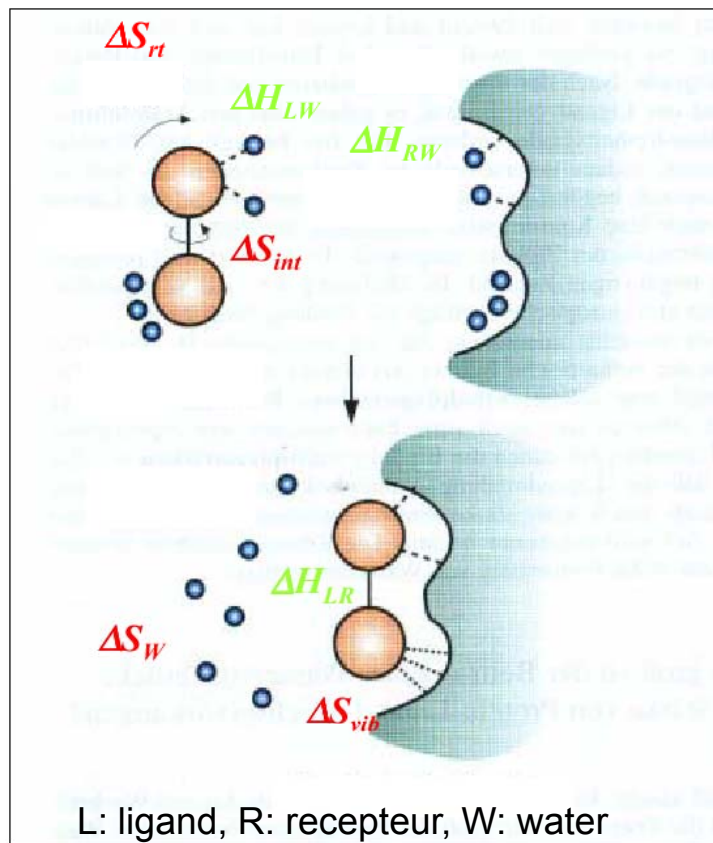
no unit

$$K_A = \exp - \frac{\Delta G^0}{RT}$$

$\Delta G^0$ ; RT in J/mole

$K_D$ (M)	$\Delta G^0$ (kcal/mol) at 298K
$10^{-15}$	-20.4
$10^{-12}$	-16.4
$10^{-9}$	-12.3
$10^{-6}$	-8.2
$10^{-3}$	-4.1

# Thermodynamics of association



**Free energy**

$$\Delta G = \Delta H - T\Delta S$$

**Enthalpy H**

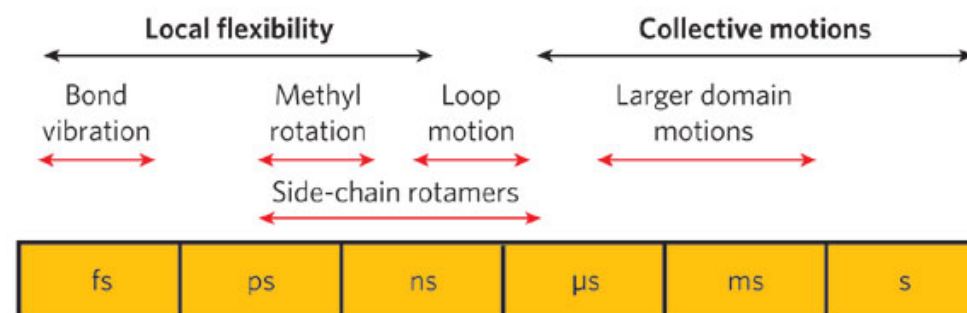
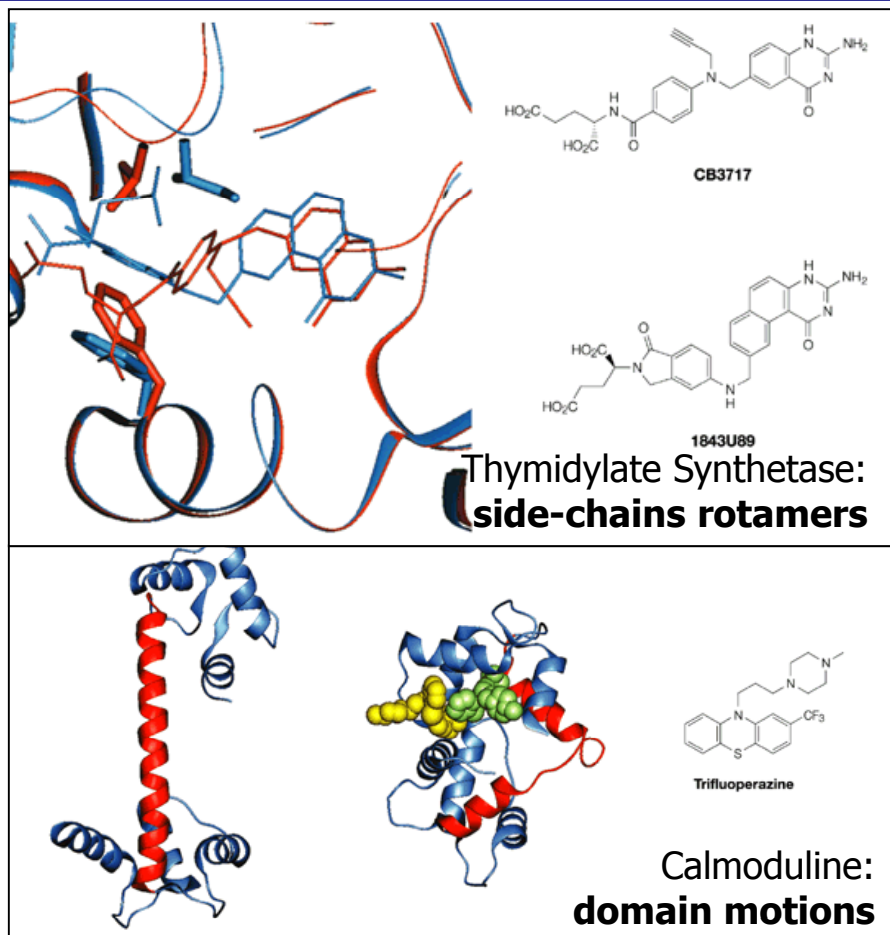
~ sum of interactions

**Entropy S**

~ order

***What is gained/lost upon binding?***

# Molecular flexibility: Lock & key or induced fit



*Nature 450, 964-972 ( 2007)*

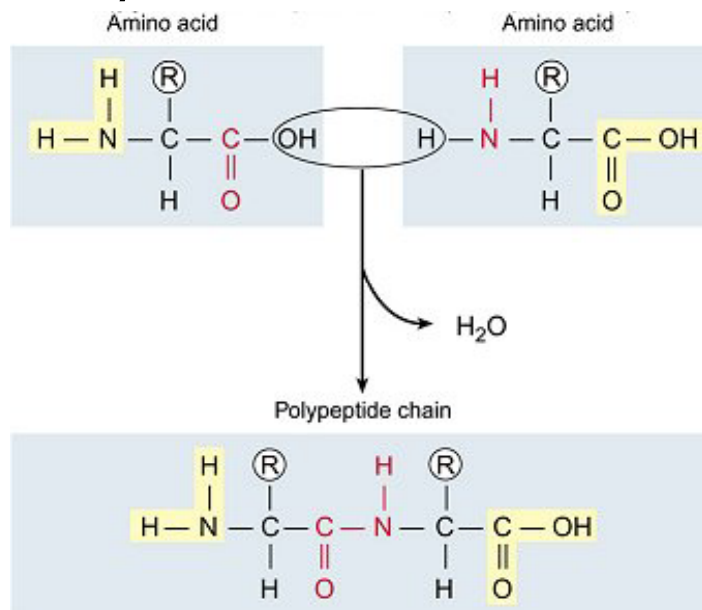
# Chapter 1:

## Proteins are folded

## biopolymers

# primary, secondary, tertiary & quaternary structures

## Polymerisation reaction

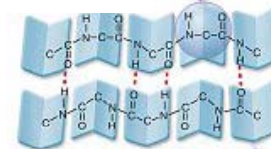


20 monomers differ by their R group

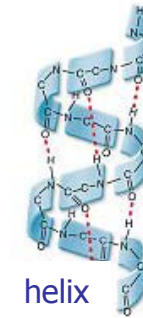
primary structure amino acids



secondary structure

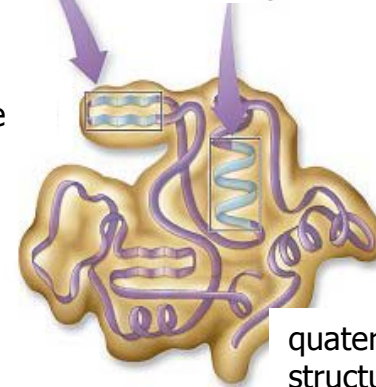


sheet

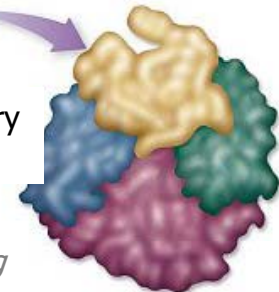


helix

tertiary structure



quaternary structure



<http://www.yellowtang.org>

# Three classes of proteins for a huge variety of functions

Human collagen

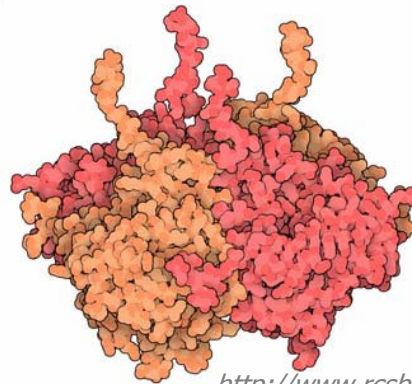


<http://www.rcsb.org>

## Fibers

- auto assembly  
→ supramolecules
- **collagen** (cartilage, bone, teeth), **keratin** (hair, nail), **fibrin** (blood clots)

Influenza neuraminidase  
(H1N1 surface protein)

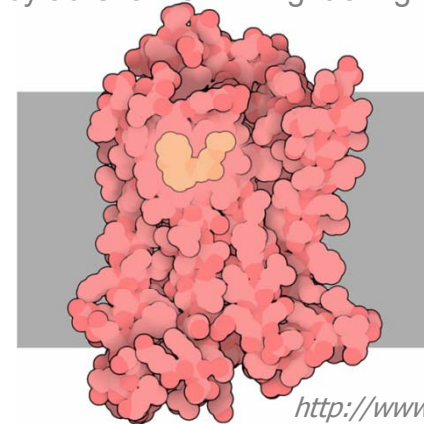


<http://www.rcsb.org>

## globular proteins

- hydrosoluble (cytoplasm, blood)
- **enzyme catalysis**, **defense** (toxin, immunoglobulin), **transporter** (O<sub>2</sub>, electrons), **motion** (actin, myosin), **regulation** (osmotic protein, gene regulators, hormone) **ion storage** (ferritin, calmodulin)

$\beta$ 2 adrenergic receptor stimulated  
by adrenaline → “flight or fight”

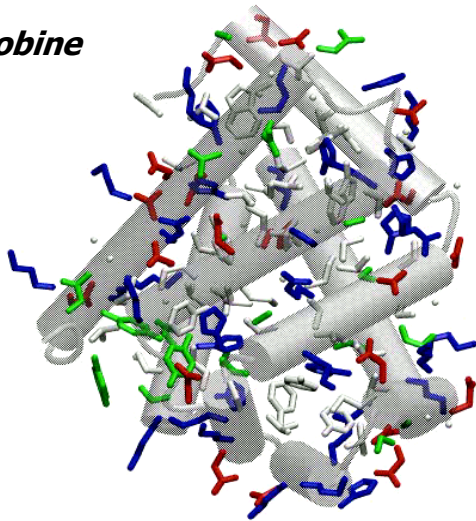
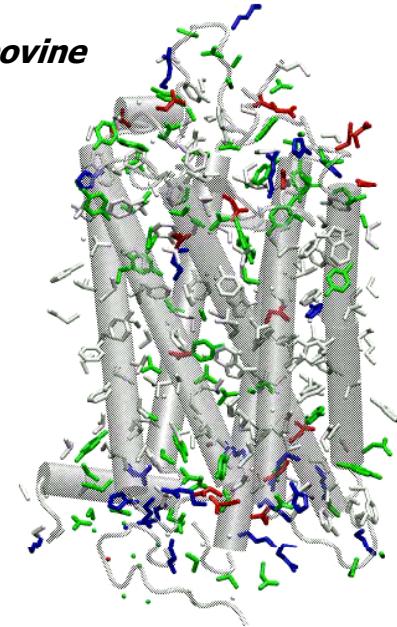


<http://www.rcsb.org>

## membrane proteins

- ~25-30% of total genes in many eukaryotes
- **Signal transducer**, **transporter** (Na<sup>+</sup>, proton, glucose), **channels**
- ~10,000 proteins encoded by human genome
- =targets of 1/2 FDA approved drugs (ionic channels, GPCRs)

# A limited number of structural organisations globular vs membrane proteins

*myoglobine**Rhodopsine bovine*

hydrophobe

polar

basic

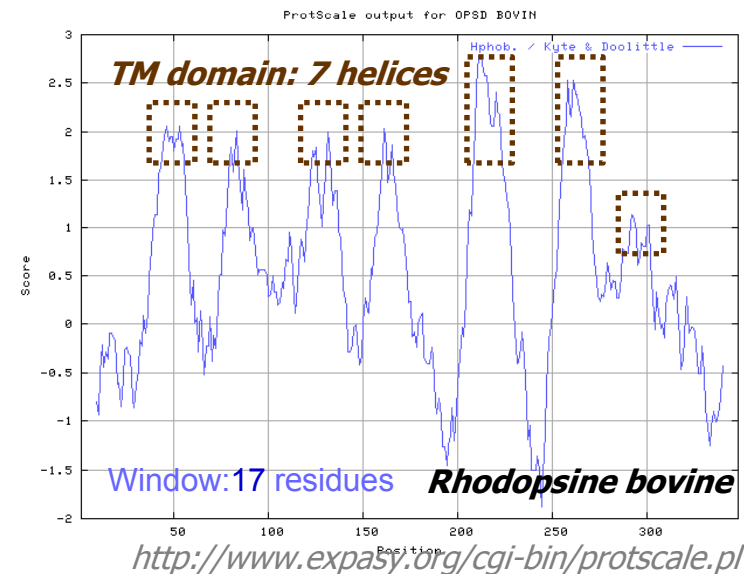
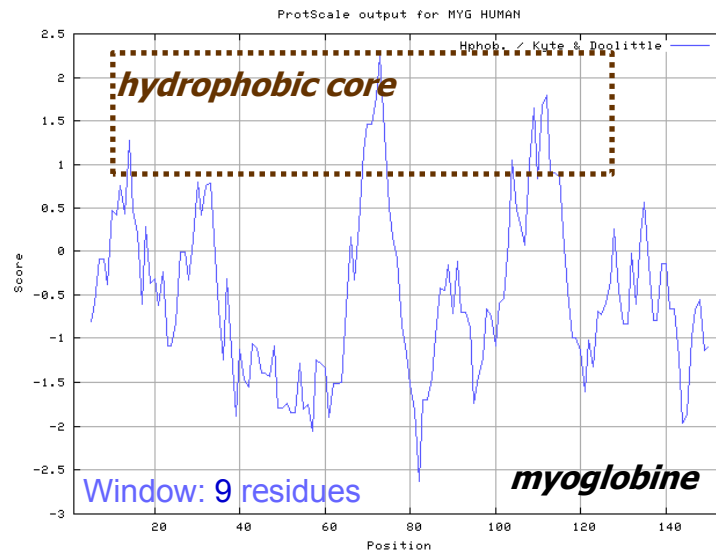
acidic

- generally soluble in water → the most studied by biophysical approaches
- a diverse but finite number folds (3D-arrangement of alpha-helices and beta-sheets)
- **Hydrophobic** residues usually **conserved** and buried to form the **3D core**
- **Hydrophylic** residues at the **surface**

- simple organization: **anti-parallel assembly of alpha helices** (helix axis perpendicular to the membrane plane)
- **Hydrophobic** residues **predominant** in the **TM** domain
- **Hydrophylic** residues in the **TM** domain generally conserved for **function**

# Hydrophobicity profiles

## globular vs membrane proteins



Many tables of amino acid hydrophobicity, based on experimental  $\Delta G$  measurements. All agree on extreme values. In a sequence, the residue hydrophoby is corrected depending on local sequence.

$H_i$  from *Kyte et Doolittle* (1982)

Ile	4,5	Val	4,2	Leu	3,8
Phe	2,8	Cys	2,5	Met	1,9
Ala	1,8	Gly	-0,4	Thr	-0,7
Ser	-0,8	Trp	-0,9	Tyr	-1,3
Pro	-1,6	His	-3,2	Asx	-3,5
Glx	-3,5	Lys	-3,9	Arg	-4,5

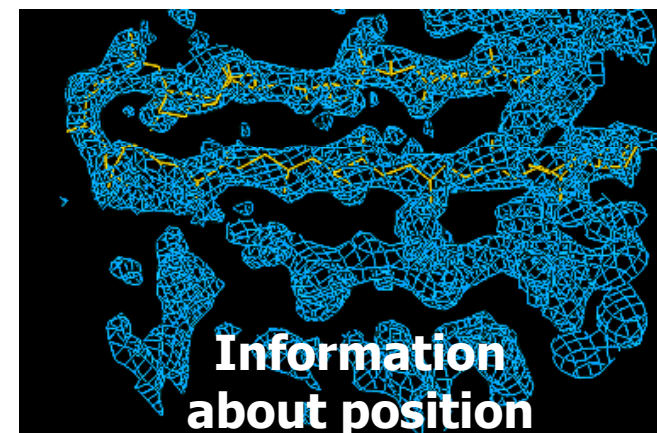
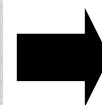
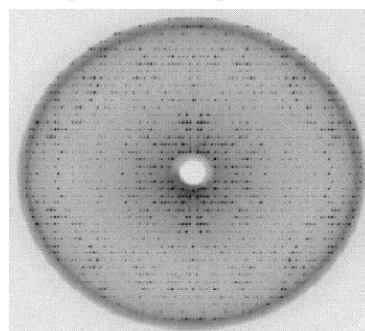
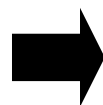
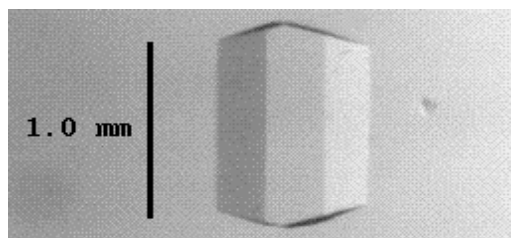
$$\bar{H}(i) = \frac{1}{2i+1} \sum_{j=i-i}^{i+i} H(i+j)$$

$$H(E) = \dots - N - V - E - D - S - G - \dots$$

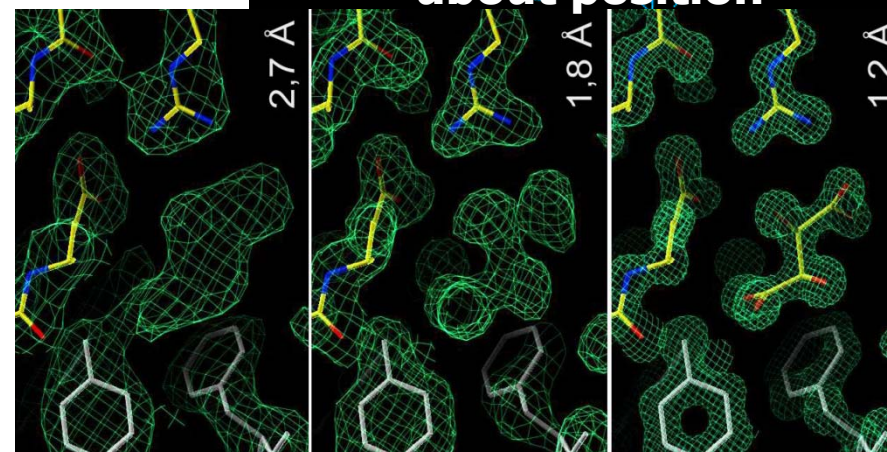
$$(-3,5 + 4,2 - 3,5 - 3,5 - 0,8) / 5 = -1,42$$

# Experimental determination of protein 3D structure

## X-ray structure of the crystal protein



Quality check:  
the resolution ( $\text{\AA}$ )



# Experimental determination of protein 3D structure

## X-ray structure of the crystal protein

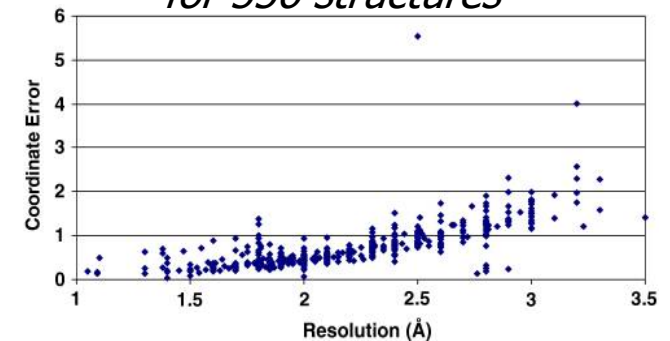
### the diffraction-component precision index (DPI)

- estimation the uncertainty of the position of atoms in a structure
- alternative to using the average B-factor (which may be over-fitted)
- average precision for the atomic coordinates in a protein structure

$$\sigma(r, B_{\text{avg}}) = 2.2 N_{\text{atoms}}^{1/2} V_a^{1/2} n_{\text{obs}}^{-5/6} R_{\text{free}}$$

$\sigma(r)$	standard error of position
$N_{\text{atoms}}$	number of atoms in the unit cell
$V_a$	volume of the unit cell
$n_{\text{obs}}$	number of crystallographic observations
$R_{\text{free}}$	agreement between model and data (not used for model creation)

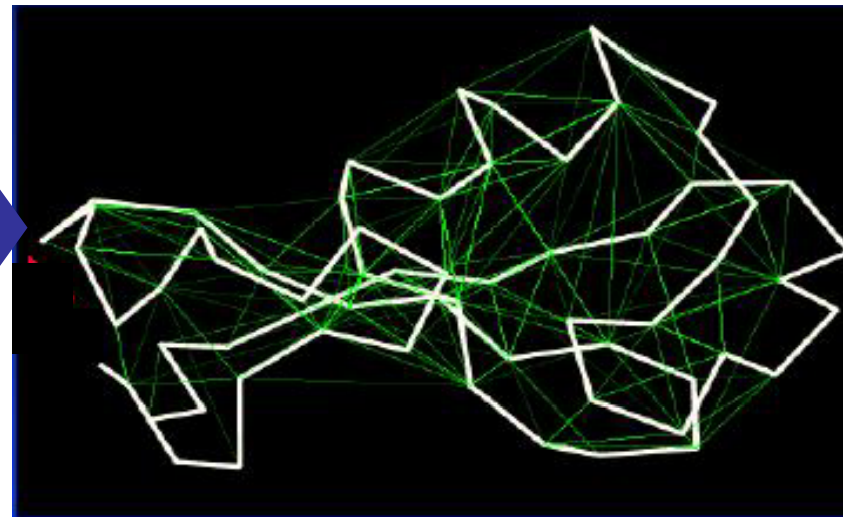
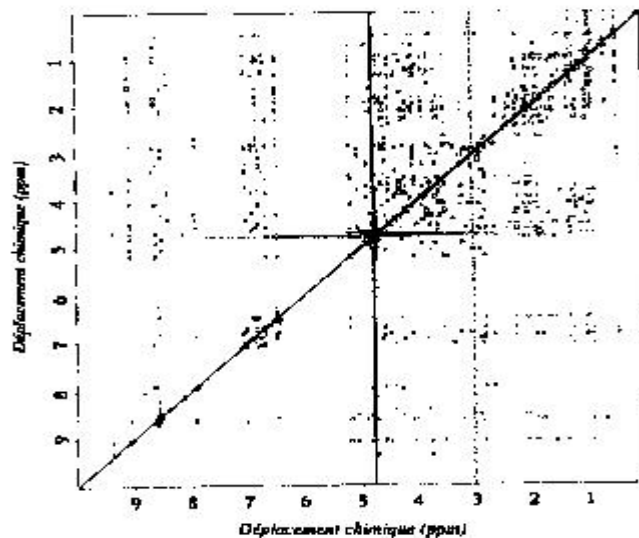
*coordinate error  
for 556 structures*



Blow DM (2002) *Acta Cryst D* 58:792  
Hawkins P et al. (2008) *Comput Aided Mol Des*  
22: 179–190

# Experimental determination of protein 3D structure

## NMR structure of the protein in solution

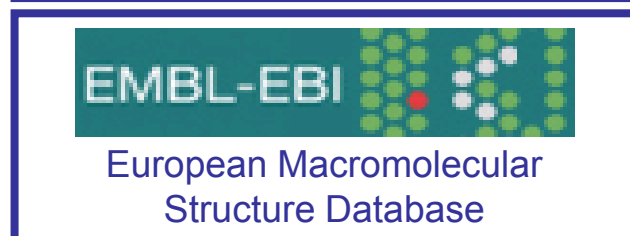


Quality check:  
quantity/quality/distribution of information

Information about

- DISTANCE (NOE)
- ANGLE (coupling constant)
- RELATIVE POSITION (residual dipolar coupling)

# Since 2003: a single archive of public 3D structures of biomolecules



**Weekly updated  
data synchronization on mirror sites**



**deposition, analysis & distribution**

**X-ray: data and structures, NMR: structures**

# PDB statistics: June 2014

Exp.Method	Proteins	Nucleic Acids	Protein/NA Complexes	Other	Total
X-RAY	83452	1530	4347	4	89333
NMR	9185	1083	210	7	10485
ELECTRON MICROSCOPY	550	56	176	0	782
HYBRID	59	3	2	1	65
other	155	4	6	13	178
Total	93401	2676	4741	25	100843

**78887** structures in the PDB have a structure factor file.

**7799** structures in the PDB have an NMR restraint file.

**21 171 non redundant sequence (>30 identity)**

**2 626 different CATH superfamilies, since 2010**

# Errors left after PDB remediation: structural errors

## The easily detectable:

- erroneous bond lengths and bond angles
- steric clashes
- missing atoms

## These errors are very frequent but can be fixed!

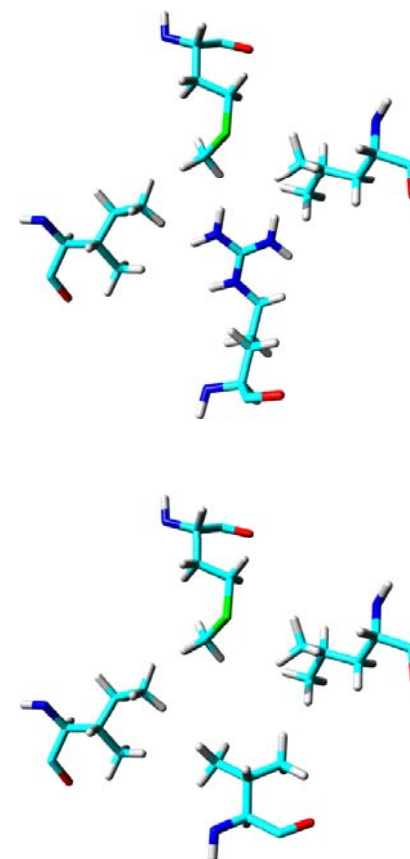
- number of atomic clashes in the PDB in 2010 ~13 million
- fixed by re-computing coordinates from structure factors or NMR restraints using a **proper force field**
- **PDBREPORT** registers, for each PDB entry, all structural anomalies in biopolymers
- **PDB\_REDO** holds re-refined copies of the PDB structures solved by X-ray

Joosten *et al.* (2011) *Nucleic Acids Research*, **39**, D411-9.

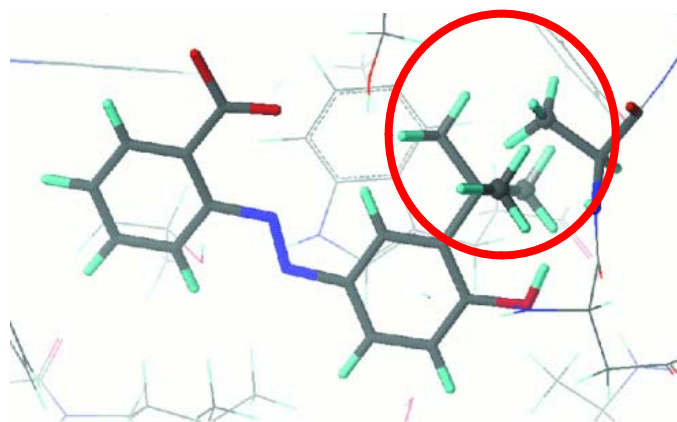
# Errors left after PDB remediation: structural errors

## The difficultly detectable, but common:

- wrong side chain packing
  - polar atoms buried in a hydrophobic environment must be engaged in an electrostatic interaction
  - opposite charged residues close in space are expected to form salt bridges
- Most common yet undisclosed structural ambiguities concern the ionisation and the tautomerisation of biopolymers and ligands (3 protonation states for HIS, 2 positions of N and O of amide groups in ASN, GLN)



## example of Ligand – protein clashes: a low resolution does not guarantee structure accuracy

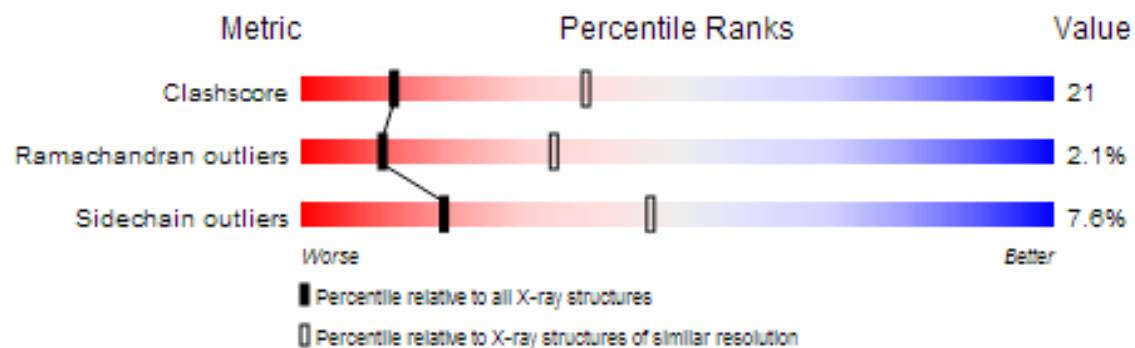


PDB entry 1srf, dC-C=2.2Å  
resolution= 2 Å

*Nissink et al (2002) Proteins: Structure, Function, and  
Bioinformatics 49: 457 - 471*

# Useful tools:

## Worldwide PDB Validation Reports



PDB ID : 1F88  
Title : CRYSTAL STRUCTURE OF BOVINE RHODOPSIN  
Authors : Okada, T.; Palczewski, K.; Stenkamp, R.E.; Miyano, M.  
Deposited on : 2000-06-29  
Resolution : 2.80 Å (reported)

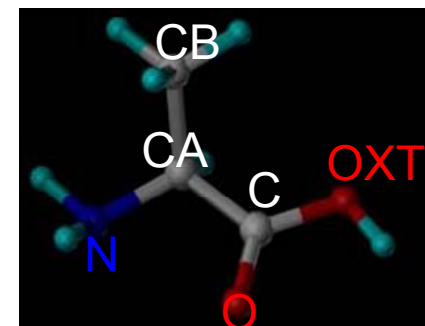
# PDB entry: format and content

Format: Flat file organized into tagged fields

- Header: information
- Connexion table: atom section only (element: C, O, N, S, H)  
no explicit bonds for standard amino acids

ATOM	1	N	ALA	1	-0.970	1.066	0.185
ATOM	2	CA	ALA	1	-0.489	-0.306	0.185
ATOM	3	C	ALA	1	1.041	-0.306	0.185
ATOM	4	O	ALA	1	1.663	0.755	0.185
ATOM	5	CB	ALA	1	-0.918	-1.001	-1.109
ATOM	6	HA	ALA	1	-0.865	-0.845	1.067
ATOM	7	OXT	ALA	1	1.638	-1.498	0.185
ATOM	8	HOCA	ALA	1	2.509	-1.408	0.552
ATOM	9	HB1	ALA	1	-0.557	-2.040	-1.113
ATOM	10	HB2	ALA	1	-0.495	-0.474	-1.976
ATOM	11	HB3	ALA	1	-2.016	-0.982	-1.179
ATOM	12	H	ALA	1	-0.304	1.812	0.185
ATOM	13	HNCA	ALA	1	-1.965	1.290	0.185

e.g., alanine



Columns

2, 3: atom number and name

4,5: residue name and number

6, 7, 8: x, y and z coordinates

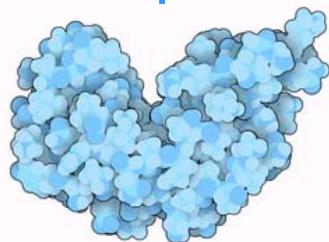
# PDB entry: format and content

Incomplete description of protein structure

METHOD	X-ray	NMR
hydrogen atoms	✗	✓
Def. of imidazole, amide	✗	✓
water molecules	✓	✗
Metal ions, cofactors	✓	✓

# Homology modeling

template

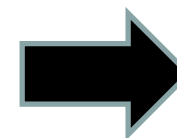


target

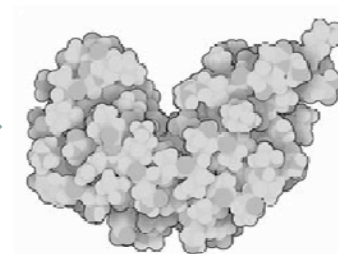
```
MNPLILITFVAAALAAPFDDDDKIVGGYNC  
EENSVPYQVSLNSGYHFCGSLINEQWVVS  
AGHCYKSRIQVRLGEHNIEVLEGNEQFINA  
GADYPDELQCLDAPVLSQAKCEASYPGKIT  
SNMFCVGFLEGGKDSQGDSGGPVVCNGQL  
QGVVSWGDCGAQKNKPGVYTKVYNYVKWIK  
NTIAAN
```



Sequence  
alignment



Coordinates  
generation

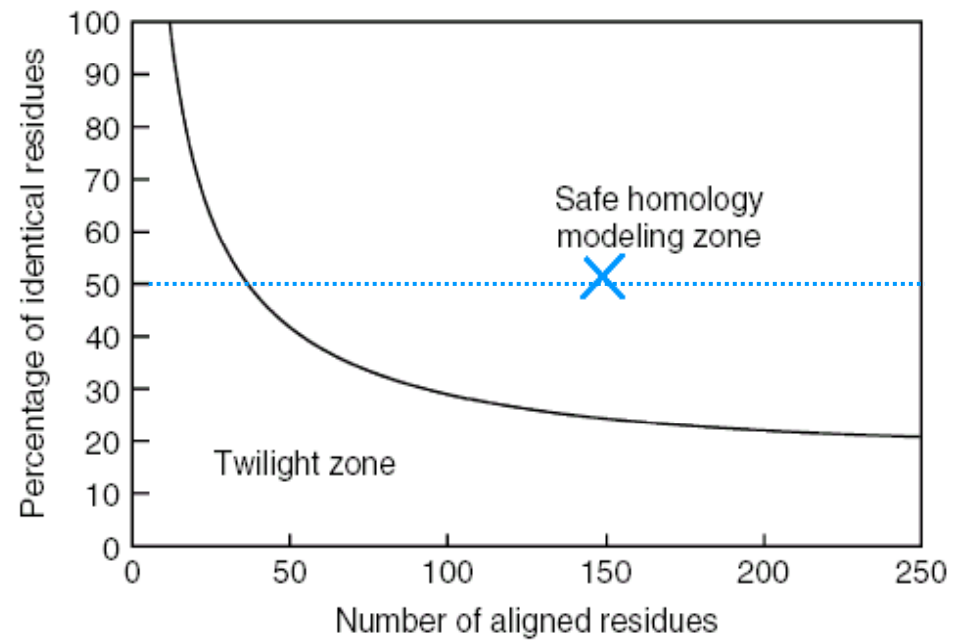


- 1- modification in the template structure
- 2- folding of target using restraints extracted from the template structure

# Homology modeling

## What is a good template?

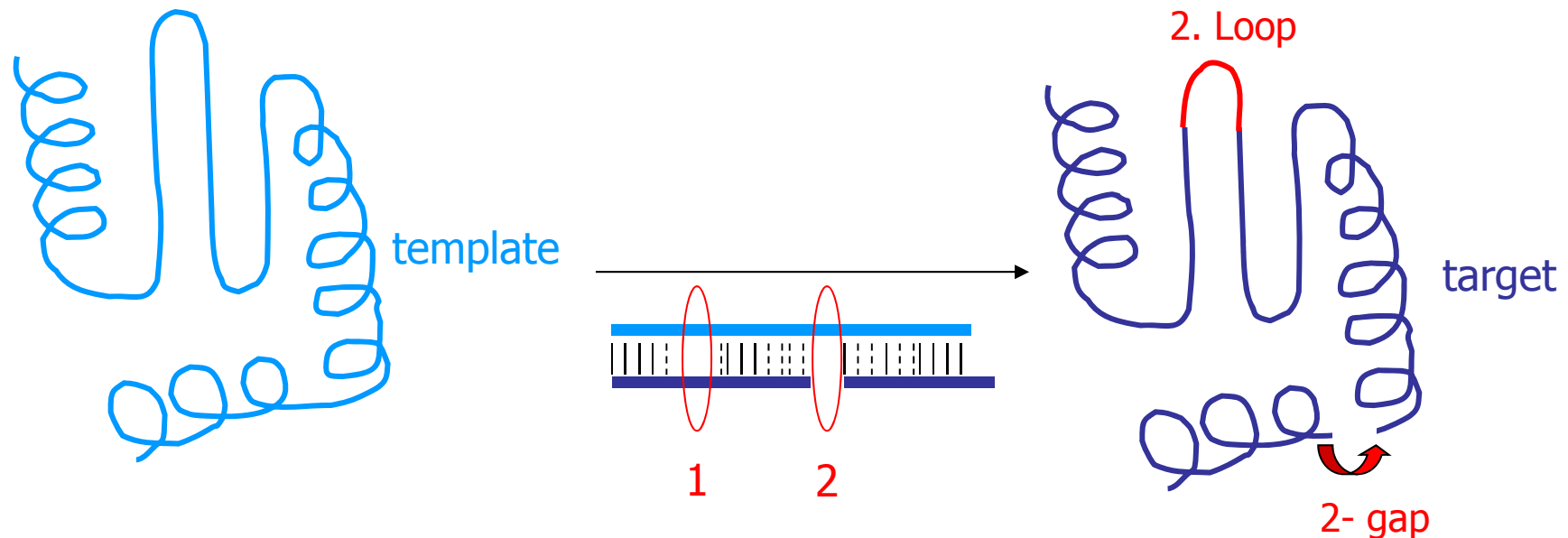
- close homologous protein
- high quality structure



# Homology modeling

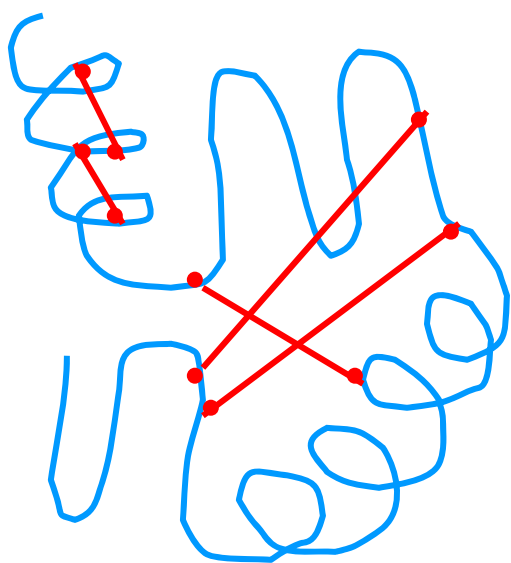
## Modification of template structure

- creation of backbone: mutation (substitution), gap filling (deletion) and loop modeling (insertion)
- side chain modeling: sampling of rotameric states

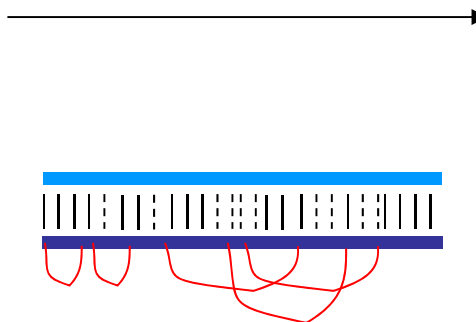


# Homology modeling

**folding of target using restraints extracted from the template structure**



template



target

# Homology modeling

## Next steps:

- Refinement (molecular mechanics)
- Validation by experimental approach (mutation, binding of ligand)

**The model cannot be more precise or more accurate than the template structure**

# Chapter 2a:

# Docking chemicals

# into proteins

# Docking into a protein pocket / cavity

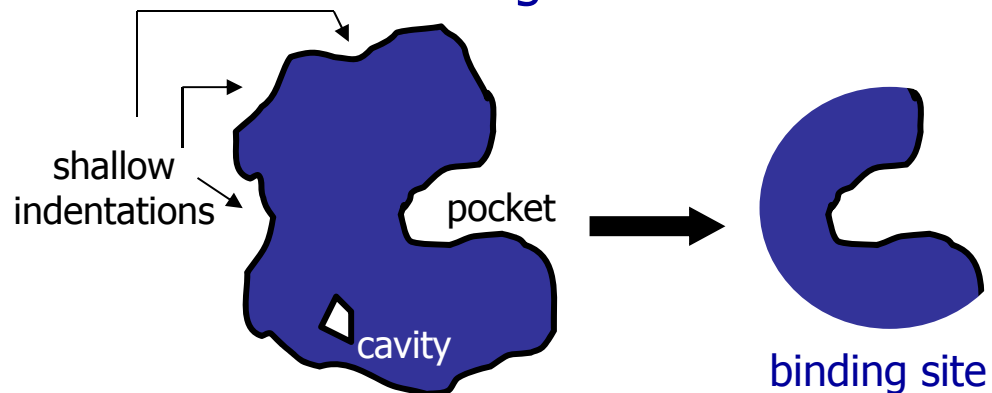
Scan the entire protein surface?

- not applicable to most docking projects (computational cost)
- not relevant: docking only efficient for “druggable” sites

What is a protein binding site suitable for docking?

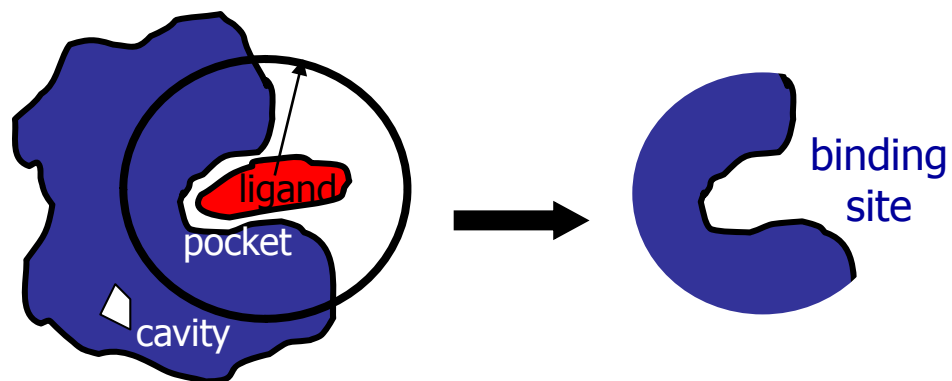
important features are

- depth/enclosure
- volume
- lipophilic surface



# The identification of pockets and cavities

in apo-protein      using geometrical methods  
in holo-protein      based on the ligand



9283 « druggable » sites in PDB in 2014

<http://bioinfo-pharma.u-strasbg.fr/scPDB/>

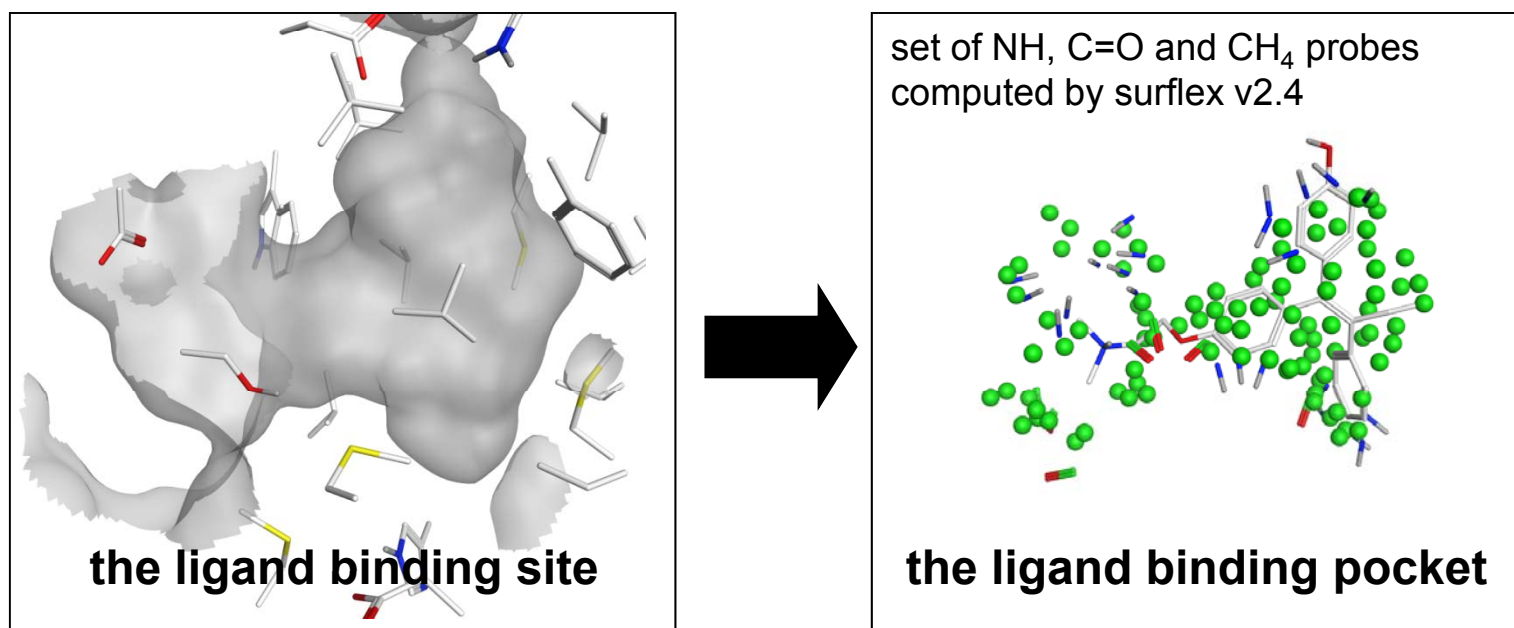
# The representation of pockets and cavities

The size and the complexity of the protein prevent the direct use of its atomic coordinates. In docking, the **structural information** is **reduced**.

The available space for the ligand is characterized by **geometric descriptors** combined with **physicochemical descriptors**.

- grid points
- surface points or vectors
- pharmacophoric points or vectors

In docking, the protein representation is the **negative image** of the binding site, “**colored**” with the properties of the residues flanking the pocket (or cavity).

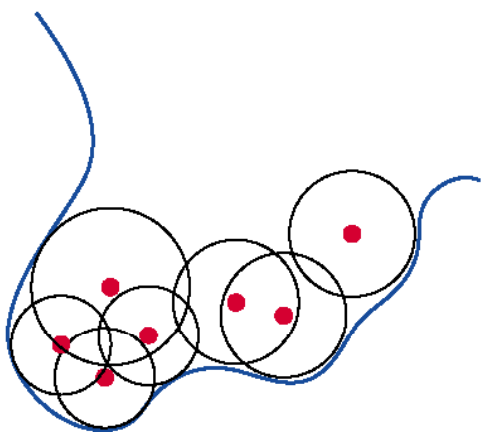


Example of the human estrogen receptor  $\alpha$  in the inactivated state (PDB code: 3ert)

## other examples of representation

### DOCK

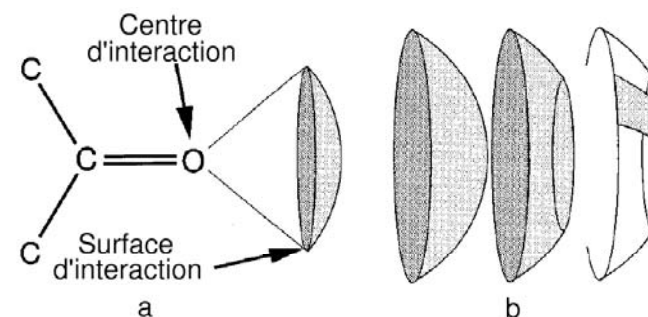
*Kuntz, I.D et al. (1982) J. Mol. Biol*



- Protein cavity filled with overlapping spheres (variable radius).
- Feature points: sphere center colored according to physico-chemical properties

### Flexx

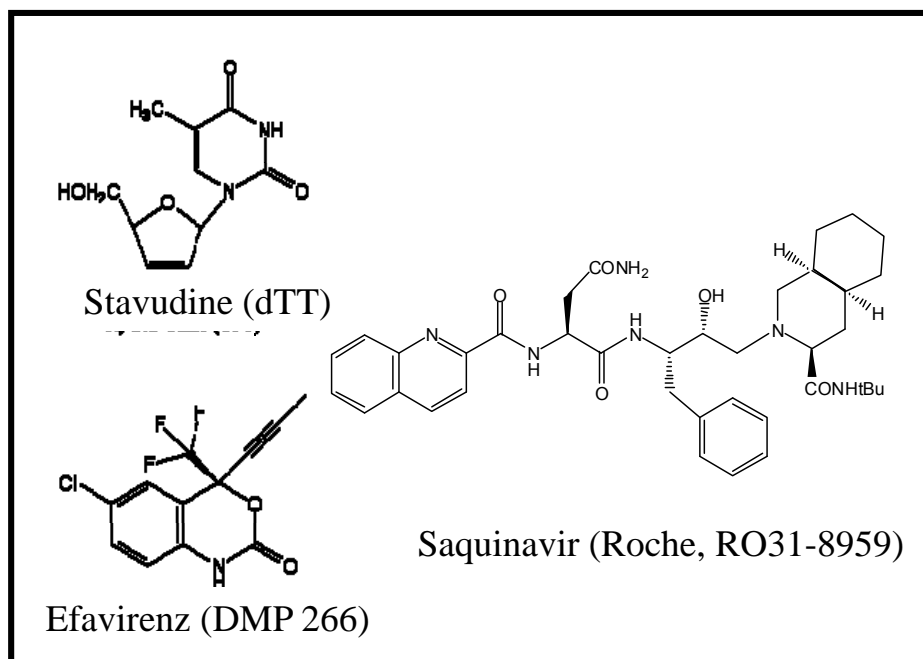
*M. Rarey et al. (1996) J. Comp.-Aid. Mol. Design*



Interaction centers and interaction surfaces identified on both receptor (a) and ligand (b)

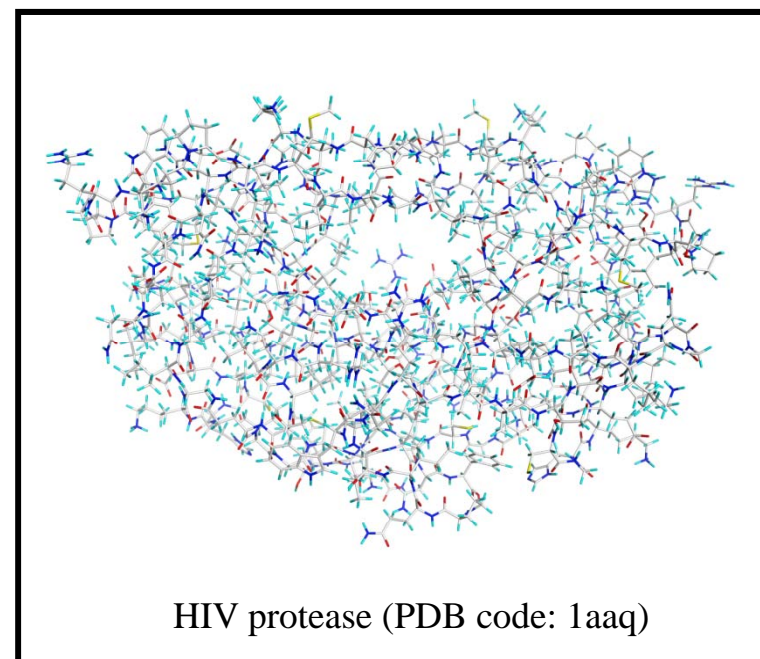
- H bond
- Salt bridges
- Aromatics
- methyl-aromatics
- amide-aromatics

# Docking a flexible ligand into a rigid protein



< 25-30  
→ exhaustive  
seconds to hours

number of rotatable bonds  
conformational search  
cpu time of conf. search



≥ 3 per amino acids  
partial  
huge!

# The search for the ligand best pose

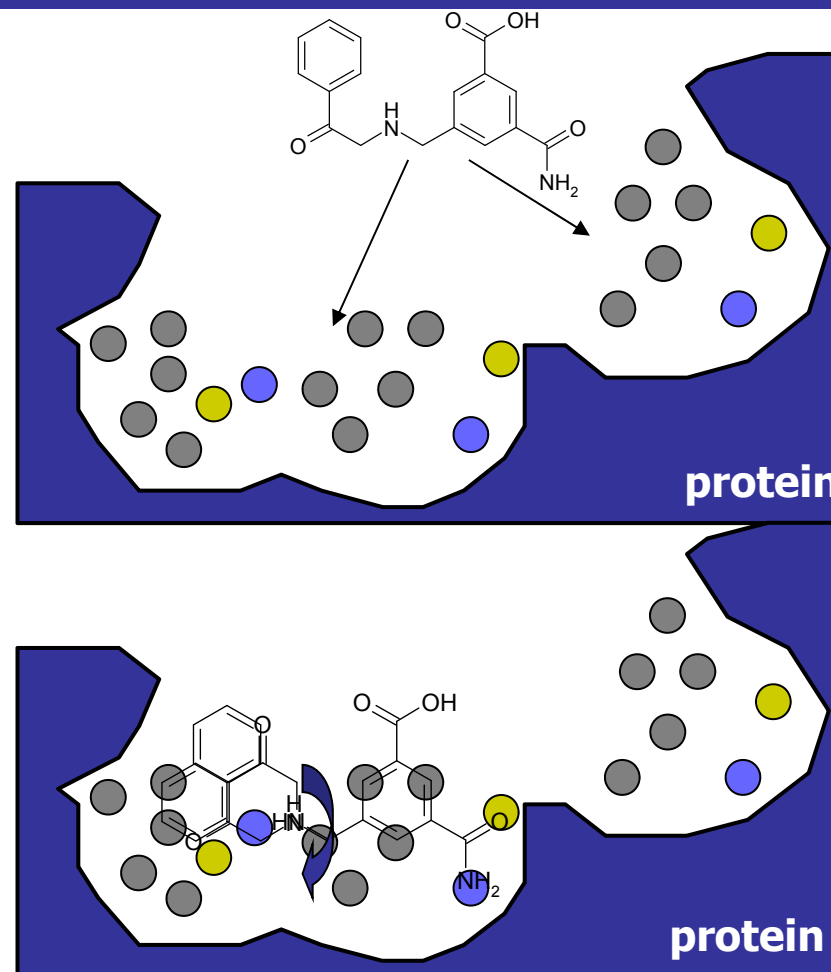
Pose = Orientation & conformation

## Orientation

- Position of the ligand in the protein
- Rigid body motions
- Translations + rotations of the whole molecule

## Conformation

- 3D structure of the molecule
- Molecular flexibility



# Handling of ligand flexibility: determinist approaches

Search: translations + rotations of rigid **ligand conformers** or rigid **ligand fragments**

Systematic sampling of torsion angles:

## 1. Library of conformers

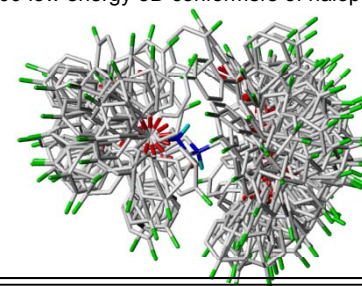
- rigid docking of all conformers, individually

## 2. Incremental construction

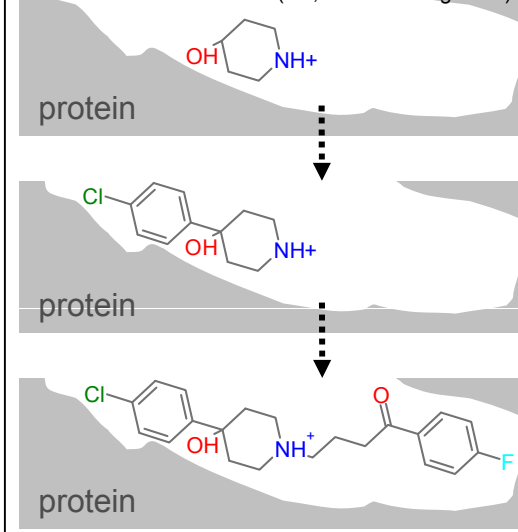
- The molecule is fragmented
- First placement of a base fragment
- Molecule is built incrementally
- Outcome depends on the selection and placement of the base fragment

**1**

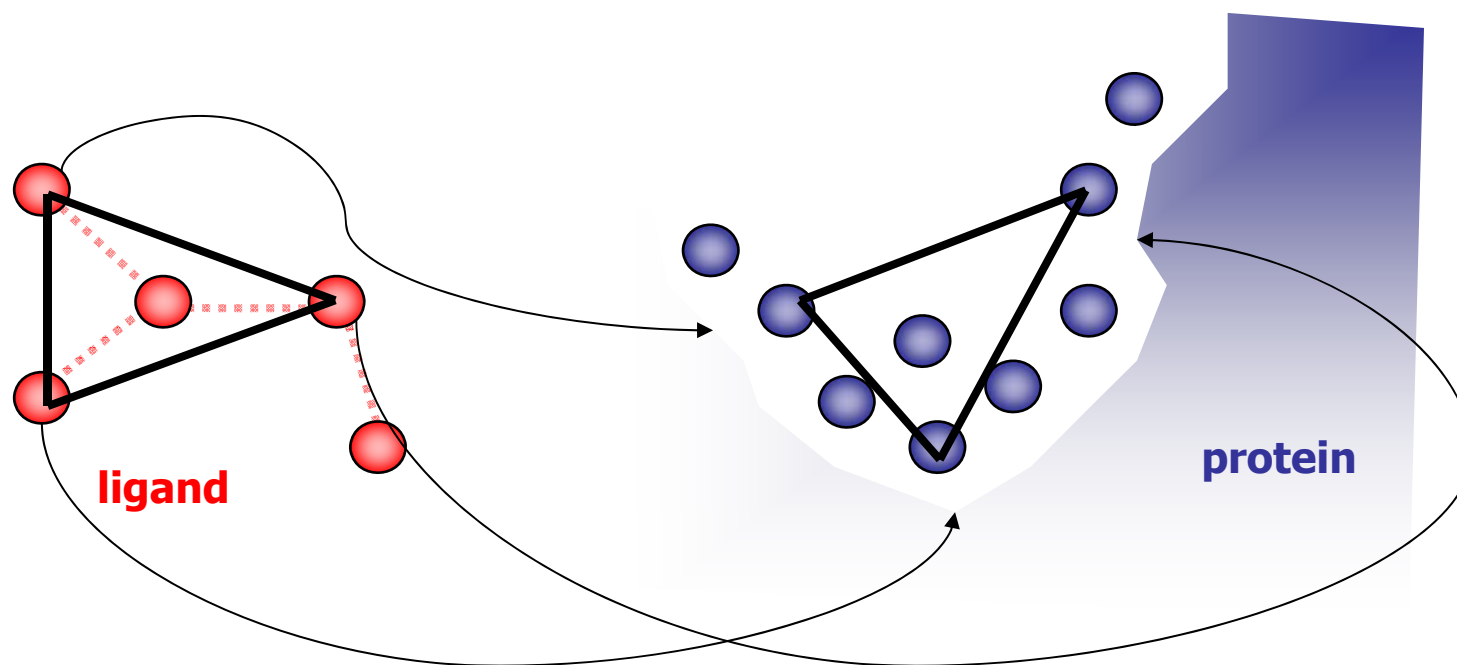
The 200 low energy 3D conformers of haloperidol

**2**

Incremental construction (i.e., anchor-and-growth)



## Algorithmic: Geometric matching of triangles



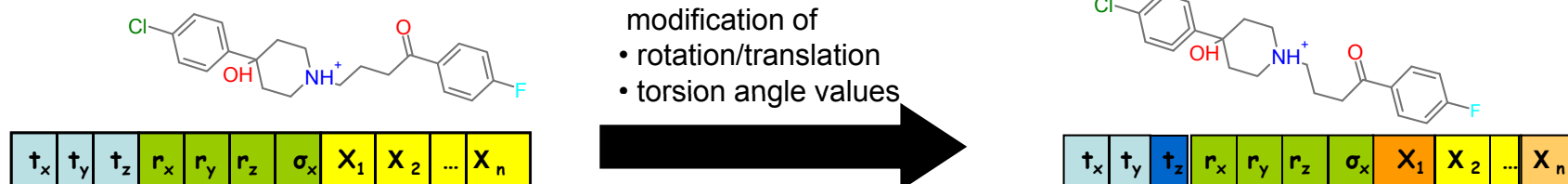
Other common algorithms: geometric hashing, clique detection

# Handling of ligand flexibility: stochastic approaches

Search: **iterative** and **random** modifications of ligand position and conformation  
the best solution is selected at each stage using a scoring function

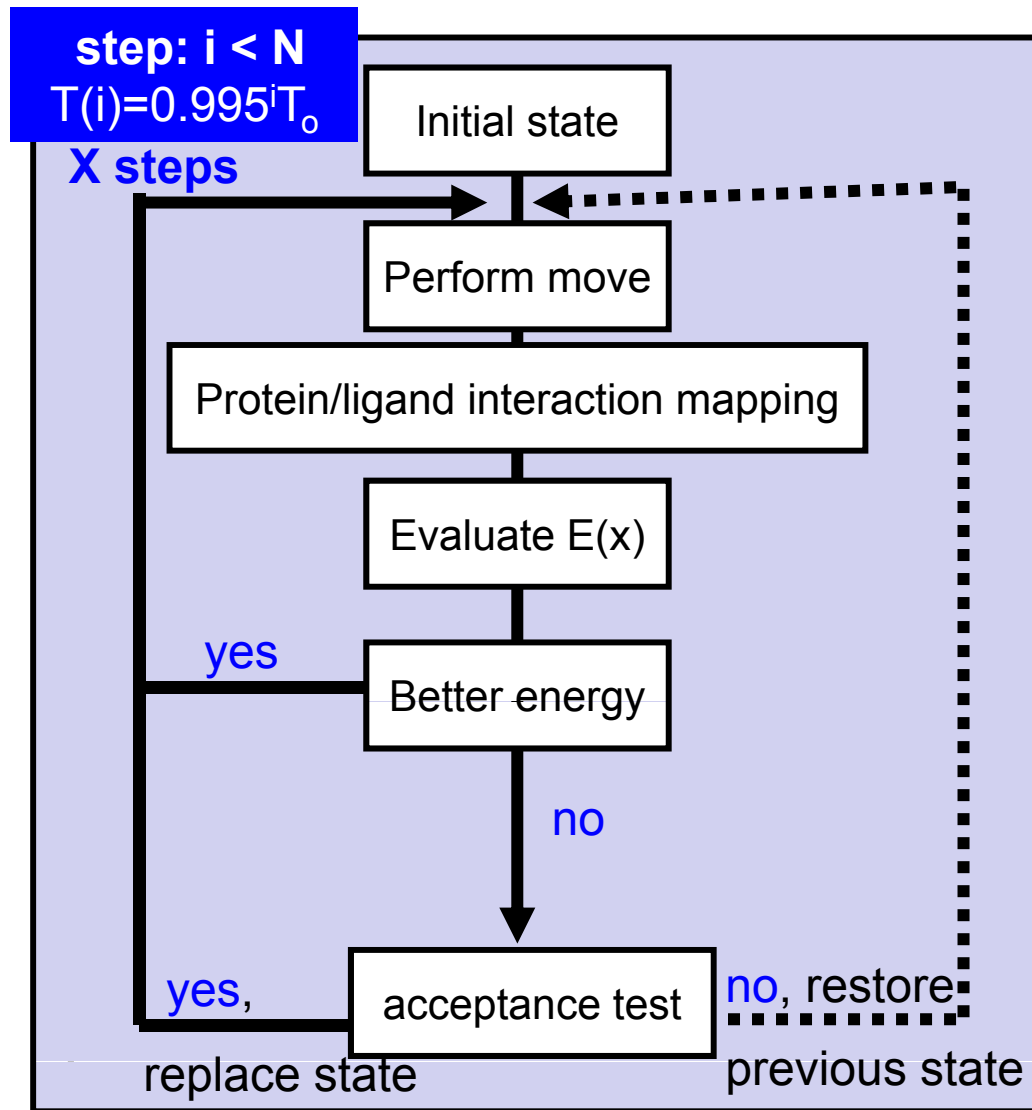
Docking: an **optimization procedure**, where the “energy” function describes molecular interactions

Example of ligand numerical representation for conformational sampling



The position of the ligand is encoded by three translations ( $t_x$ ,  $t_y$ , and  $t_z$ ) and three rotations ( $r_x$ ,  $r_y$  and  $r_z$ ,  $\sigma_x$  defining the rotation axis). The conformation of the ligand is encoded by the torsion angles of its rotatable bonds

# the Monte-Carlo/simulated annealing algorithm



## Acceptance Test

if  $E_f < E_i$  new pose is accepted  
 if  $E_f > E_i$  calculate **probability P**

$$P = \exp \left( -\frac{E_f - E_i}{kT} \right)$$

k: boltzman constant  
 T: temperature

Compare P with **random number h**

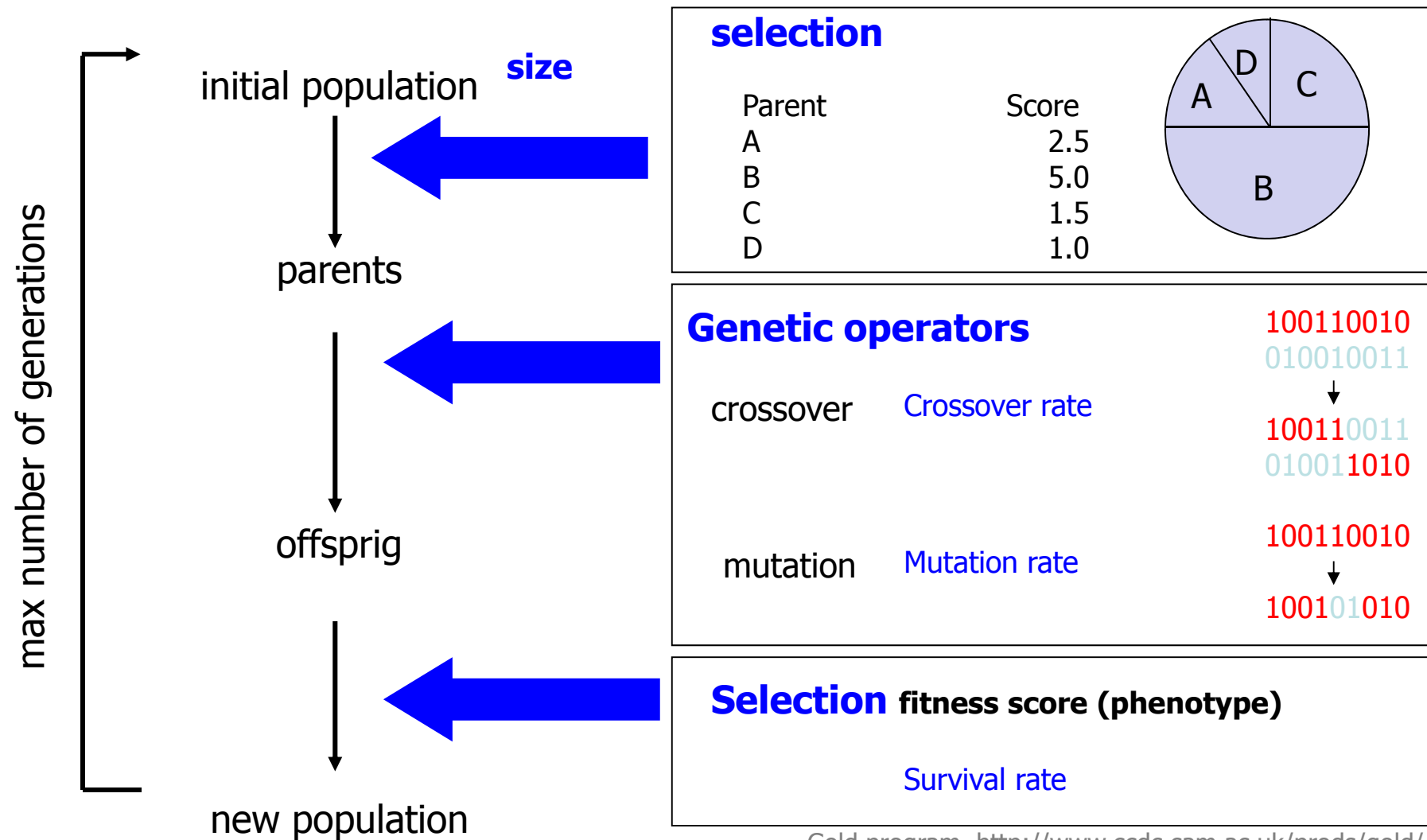
if  $h < P$  new pose accepted  
 if  $h > P$  restart based on last accepted pose

Temperature T decreased in defined number of cooling cycles:

T high broad region of space is sampled  
 T low space is explored locally

**Multiple independent runs are required for convergence (10-50)**

# the genetic algorithm (GA)



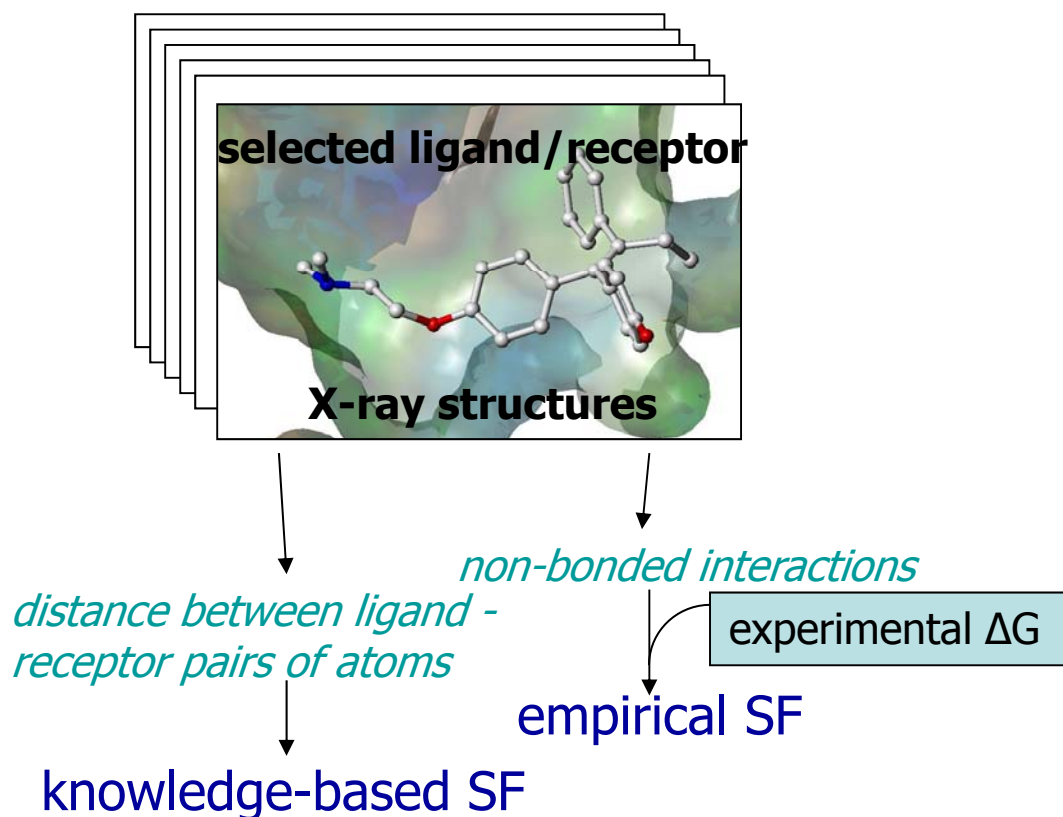
Gold program, <http://www.ccdc.cam.ac.uk/prods/gold/>

# Chapter 2b:

## Scoring ligand/receptor interaction

# empirical scoring functions vs force field

## Empirical SF

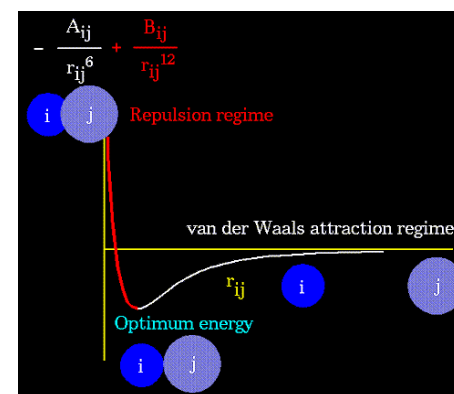


## Force Field SF

$$S = E_{\text{ligand}} + E_{\text{complex}}$$

$E_{\text{ligand}}$  = internal energy

$$E_{\text{complex}} = \sum_i^{\text{lig}} \sum_j^{\text{rec}} \left[ \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + 332.0 \frac{q_i q_j}{\epsilon r_{ij}} \right]$$



# Prediction of free energy by Empirical SF

$$\text{Score} \approx \Delta G = \sum k_i * F_i$$

$F_i$  : function which describes protein – ligand interaction

- H-bond, salt bridge
- Van der Waals interactions ...

computed using geometry predicted by docking

$k_i$  : constant adjusted using a training set composed of ligand-protein complexes

- 3D structure is known
- experimental value of free energy  $\Delta G$  is available

Böhm J. Comp. Aid Mol. Des. 8 243 (1994)

$$\Delta G = \Delta G_o + \Delta G_{hb} \sum_{\text{h-bonds}} f(\Delta R, \Delta \alpha) + \Delta G_{\text{ionic}} \sum_{\text{ionic int.}} f(\Delta R, \Delta \alpha) + \Delta G_{\text{lipo}} A_{\text{lipo}} + \Delta G_{\text{rot}} N_{\text{rot}}$$

$\Delta G_o$ :	constant term $\Leftrightarrow$ entropy lost
$\Delta G_{hb}$ :	contribution of one perfect H-bond
$f(\Delta R, \Delta \alpha)$ :	penalty for bad geometry
$\Delta G_{\text{ionic}}$ :	ionic contribution
$\Delta G_{\text{lipo}}$ :	lipophilic contribution
$A_{\text{lipo}}$ :	lipophilic contact surface
$\Delta G_{\text{rot}}$ :	lost of free energy due to internal rotation
$N_{\text{rot}}$ :	number of rotatable bonds frozen upon receptor binding

# Empirical statistical functions or « knowledge-based » functions

$g_{xy}^2$  : probability to find 2 atoms distant from  $r$  Å (wo unit)  
obtained by statistical count using a training set of X-ray structures of PDB  
complexes (distribution of distances between type x atoms and type y atoms)

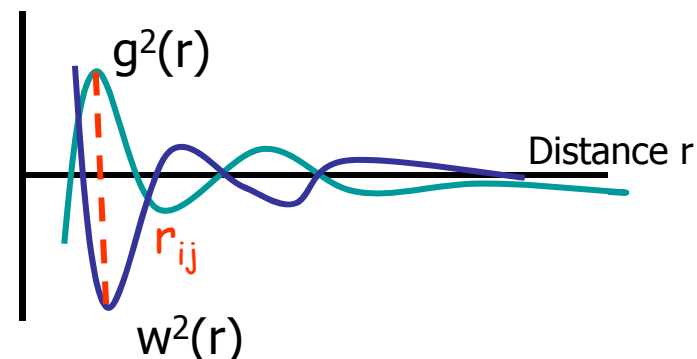
$w^2$  **mean-force Potential**, logarithmic function of  $g_{xy}^2$

**score** sum of  $w^2(r_{ij})$  values for each our i (ligand) j (protein) atoms pair  
based on the 3D structure of the complexe:

$$\text{Score} = \sum_i \sum_j w^2(r_{ij})$$

$$w^2(r) = -RT \ln g_{xy}^2(r)$$

R : rare gaz constant , T : temperature



Grzybowski et al Acc. Chem. Res. 2002 35 261

# Strength and weakness of SF

## Empirical SF



- fast calculation
- adaptable to custom target



- training set (incompleteness, inaccuracy of data)
- binding mode (if few polar interactions, underestimation of score)
- missing penalty (steric clashes, polar/apolar match, internal ligand energy, lost of entropy, geometry of directional interaction, local environment of interaction)

## Force Field SF



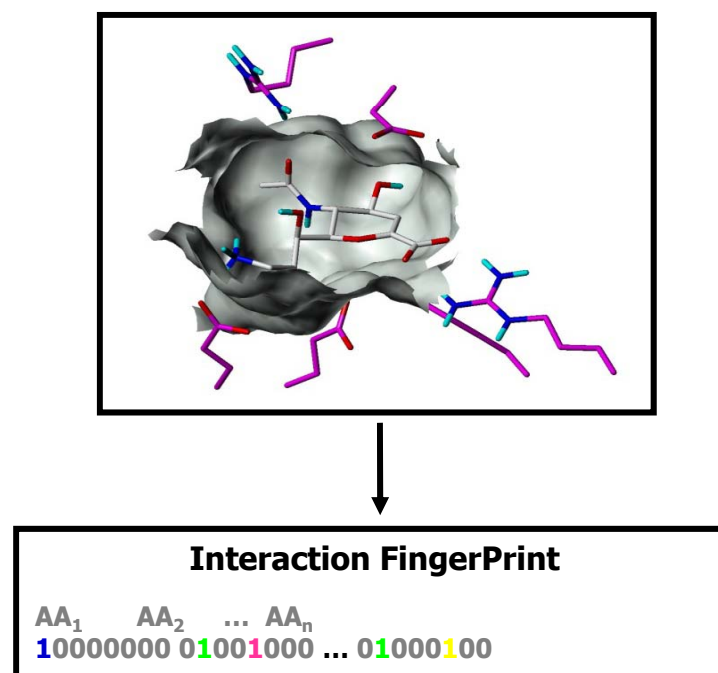
- independant of training set



- strongly depends on ligand size
- force field accuracy, difficulty to set parameters
- no account of entropy
- includes empirical terms !

# post-processing output: pose selection by *Interaction Fingerprint* analysis

- **Intermolecular interactions between ligand and protein encoded in a binary string**
  - Fixed number of bit / site residue
  - Each bit describes an interaction (hydrophobic, aromatic, H-bond, ionic bond, bond to metal)
- **Similarity of binding to a given protein evaluated by Tanimoto coefficient**
- **Successful applications in pose prediction, compound ranking and scaffold hopping**
- **Requires a reference ligand!**



Marcou, G et al (2007) J Chem Inf Model

# In 2013, 50 reported docking programs (Swiss Institute of Bioinformatics)

- **the representation of the protein cavity**

Regular (grid) or irregular (set of probes) ensemble of geometrical objects

Object: Point, vector or more complex geometrical object

Properties assigned to object (pharmacophoric feature, atom type, potential..)

- **handling of ligand flexibility**

Determinist approach: systematic conformational exploration, and rigid-body motion

Stochastic approach: random modification of ligand structure to optimize of a function

- **the algorithmic**

Determinist approach: geometric matching (triplet matching, geometric hashing, clique detection)

Stochastic approach: Monte-Carlo/Simulated annealing, evolutionary algorithms, swarm intelligence methods

- **scoring function:** various empirical scoring functions, force field

# The state of the art

- **docking accuracy**

- many programs able to reproduce x-ray conformation
- but performance is highly dependend on the studied protein

- **cpu time**

- few seconds to several minutes to dock one compound

- **hit rate** (true positives in hit list) in screening by high throughput docking

- ~ 50 % retrospective studies
- from 10% to 30% in prospective studies using X-ray structures
- lower rates for docking using homology models
- docking hits are usually not very potent (micromolar affinity)

# The limitations

## Pre-processing of the protein

which ionisation state for his, arg, lys, asp, glu?

addition of missing hydrogens, optimization of hydrogen bonds network

selection of bound water molecules?

## Pre-processing of the ligands

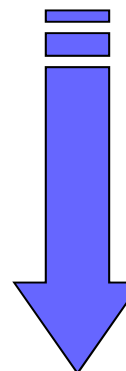
which protonation state?

which tautomeric state?

## Flexibility of the ligand , covalent ligand

## Flexibility of the protein /binding site

## Fuzzy scoring functions



*no serious problem*

*a difficult problem*

*the biggest problem*

Moitessier N, Englebienne P, Lee D, Lawandi J, Corbeil CR. *British Journal of Pharmacology* (2008), 153, S7-26.

# Chapter 3:

# Pharmacophore

# What is a pharmacophore?

*" a pharmacophore is the **ensemble of steric and electronic features** that is necessary to ensure the **optimal supra-molecular interactions** with a **specific biological target** and to trigger (or block) its biological response."*

*C-G Wermuth et al., Pure Appl. Chem. 1998, 70, p1129-1143*



# Feature-based pharmacophore models

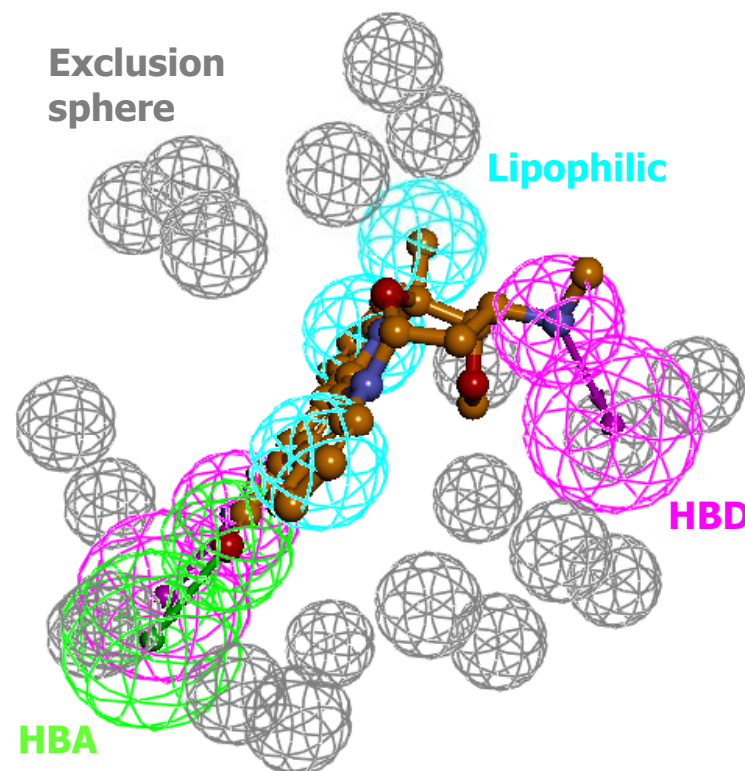
## Pharmacophore

### features

- H bond donors/acceptors
- + / - charged groups
- Lipophilic groups
- Aromatic groups
- Exclusion volumes

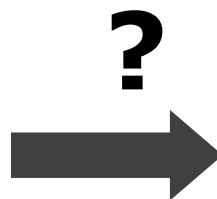
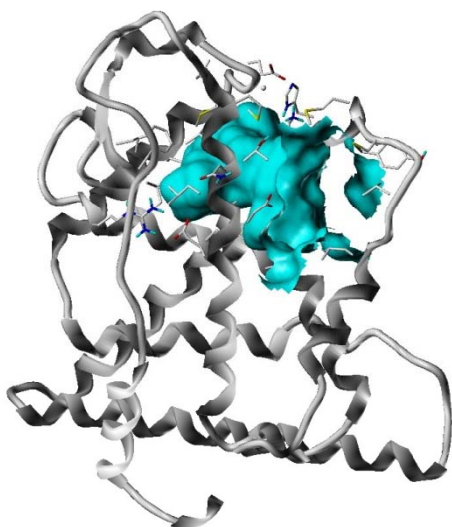
### and **relative spatial orientation**

- 2D: number of bonds
- 3D: distance



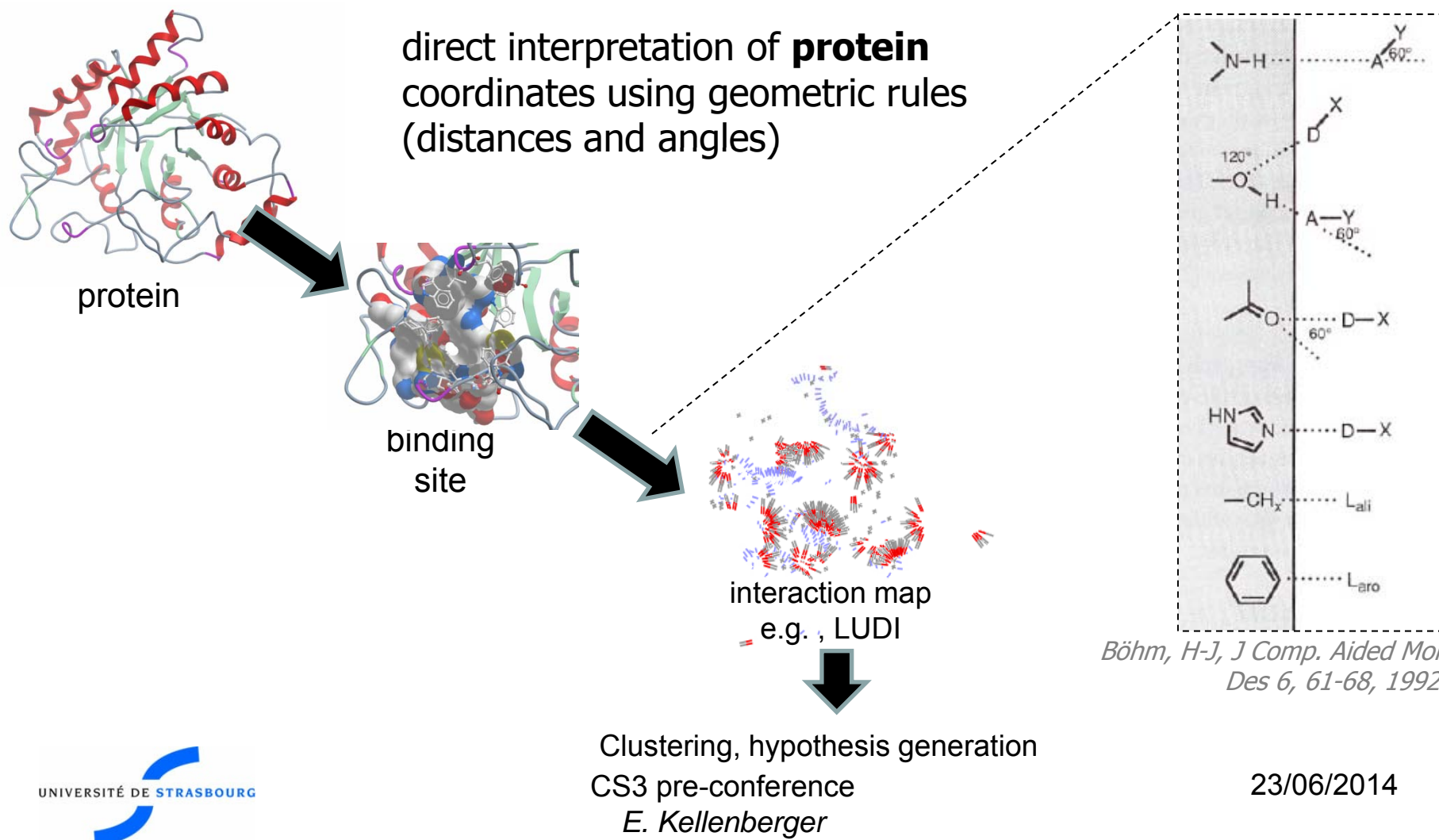
# Hot spots at protein-ligand binding sites

Hot spots are chemical groups which contribute a large part of the binding energy. It is commonly accepted that hot spots in protein are the amino acid atoms involved in non covalent interactions with a ligand.



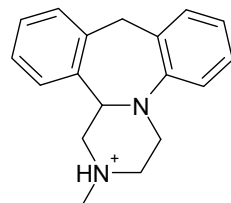
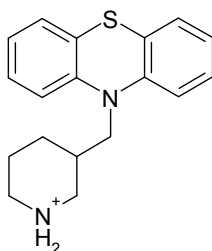
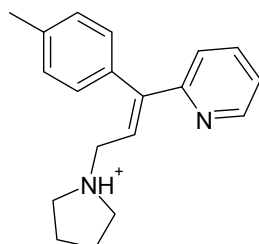
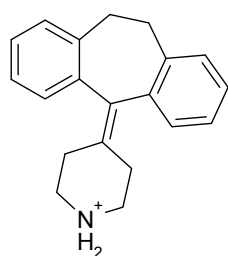
Protein-based  
pharmacophore

# Protein-based pharmacophore

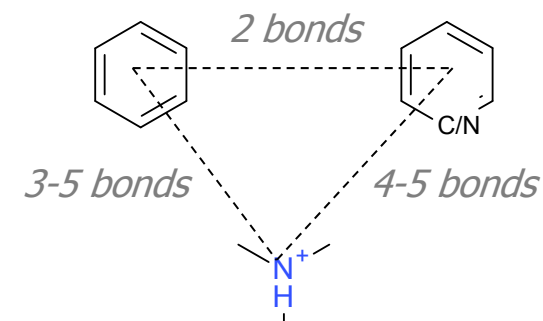


# Ligand-based pharmacophore

requires **several known actives**



**selection of  
common features**

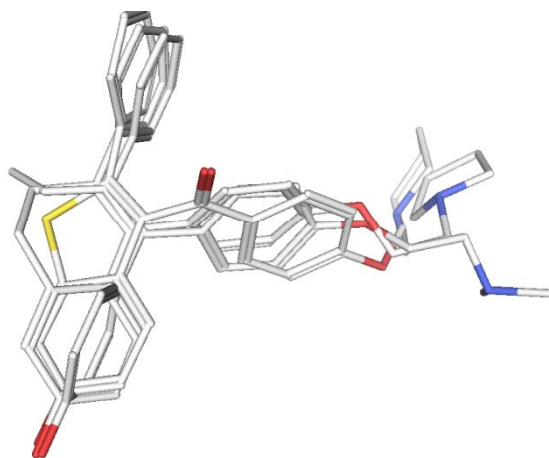


pharmacophore

e.g., histamine receptor H1 antagonists

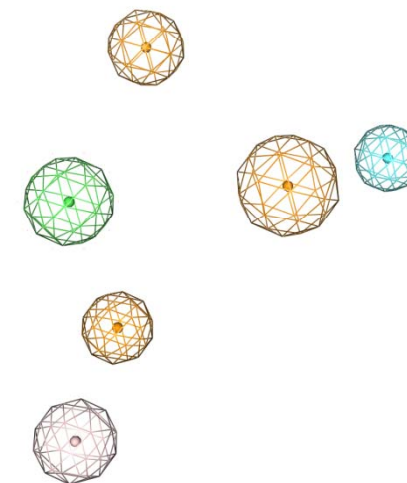
# Ligand-based pharmacophore

requires **several known actives**, and 3D **overlap** of structures



3D-aligned ligands

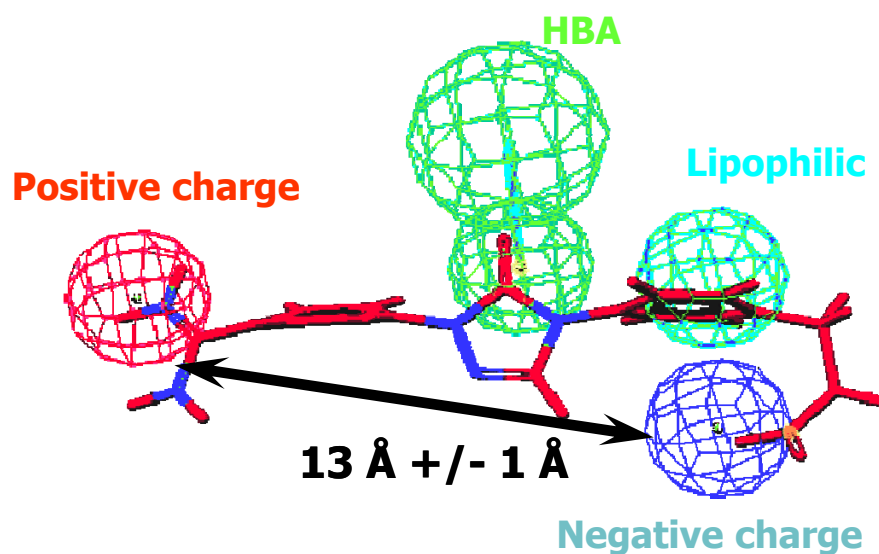
selection of  
common features



3D pharmacophore

# Tuning the precision of pharmacophore

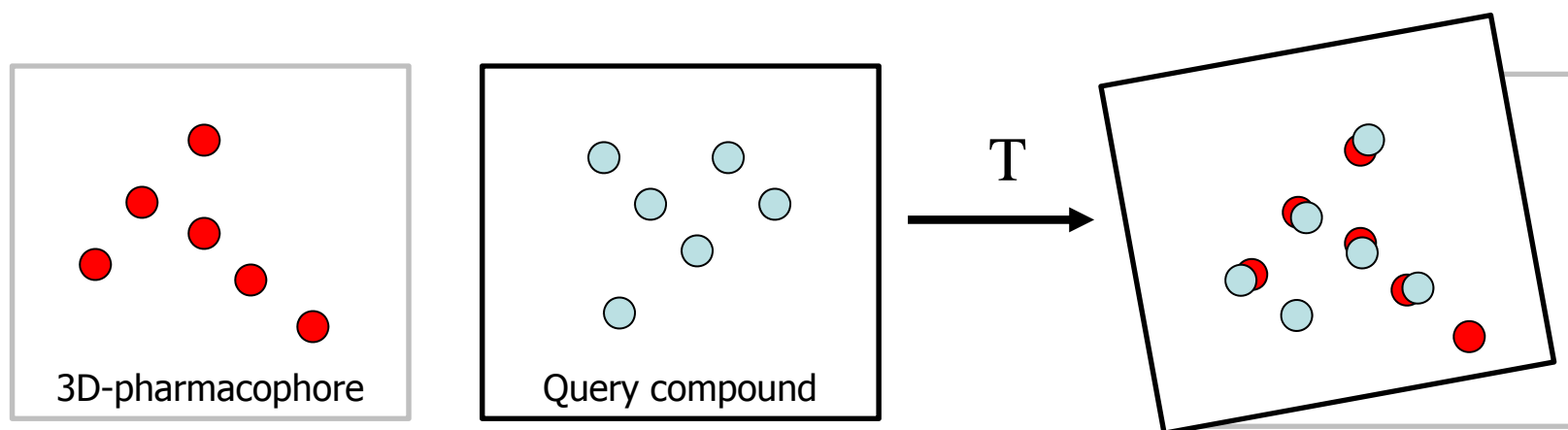
- distance **tolerance**
- **number** of **features**
- **features** weight / selection / combination
- **features** representation: point, sphere, vector



# Pharmacophore search in a query compound

**... is a 3D-alignment problem, *i.e.* a geometrical problem**

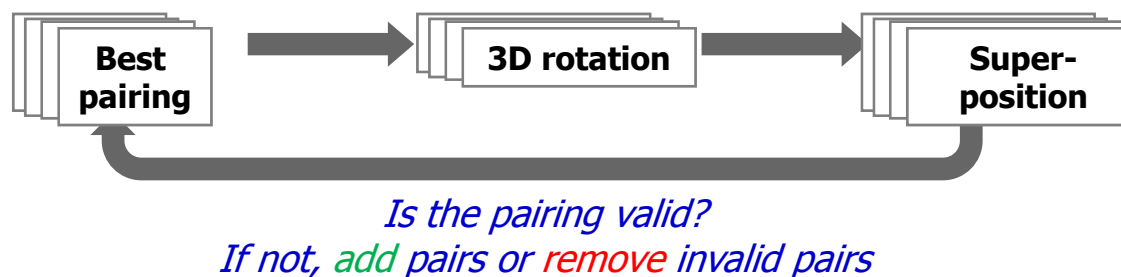
Given 2 ensembles of points in the tridimensional space,



What is the transformation  $T$  (rotation and translation) which yield the best superposition of the two sets?

# Example of protocol for 3D-alignment

## Iterative protocol



e.g., “hiphop” algorithm in Catalyst (Accelrys) starts by finding all two-features models and expands the model until no more configurations can be found

# Scoring the pharmacophore match

## Geometric fit values

- number of matched features
- quality of the match (RMSD)
- no considerations about energetical cost of solvation and flexibility

Root mean Square deviation RMSD:

computed from the distances  $\delta$  between  
the N pairs of equivalent points

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^{i=N} \delta_i^2}$$

# What about ligand flexibility?

If **distances** in the pharmacophore are **real**

(by contrast with topological distances):

- generation of conformers for the searched ligand
- individual testing of each conformers
- ranking of the conformers by the fit values

# why use pharmacophore?

- **Universal**

- pharmacophore models represent chemical functions
- Two chemically dissimilar molecules can match the same pharmacophore
- The approach well suits scaffold hopping and design of peptidomimetics

- **Computationally Efficient**

- fast method
- suitable for virtual screening
- excellent filter tool for rapid pre-screening of large databases ( $> 10^4$  compounds)

- **Comprehensive & Editable**

- Selectivity-tuning by adding or omitting chemical features
- It is often worth testing multiple pharmacophores and then fusing data.

# Structure-based approaches to drug discovery

## Virtual screening

Docking or pharmacophore  
(depends on prior knowledge on  
the studied system)

## Hit/lead optimization

Docking of a focalized library

