

# Modeling conformational changes during docking

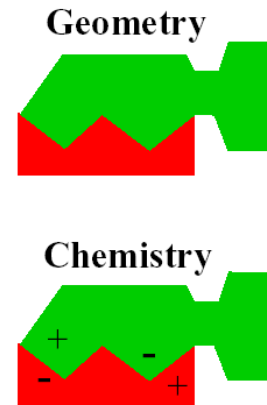
**Martin Zacharias**  
**Computational Biophysics**  
**Jacobs University Bremen**

# Outline

- **Conformational changes in proteins upon association**
- **Methods to model conformational changes**
- **Strategies to account for conformational changes**
- **Explicit flexibility during docking**
- **Attract docking approach**

# Lock-and-key and induced fit binding

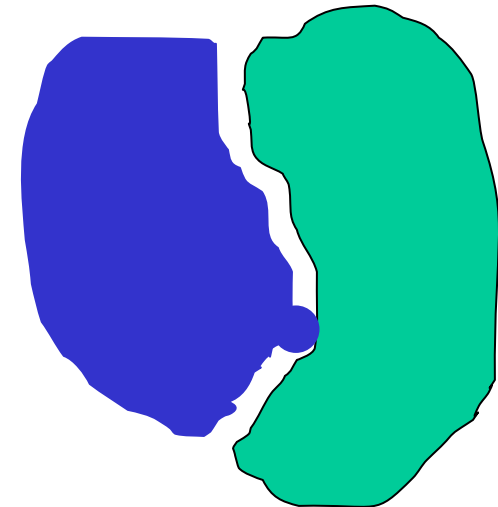
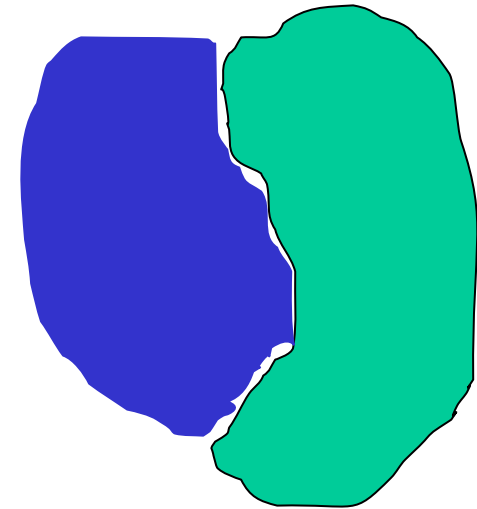
Emil Fischer 1894: *“To use an image, I would say that enzyme and glycoside have to fit into each other like a lock and a key, in order to exert a chemical effect on each other.”*



- **Comparison of protein conformations in the bound and unbound states indicates:**
  - A variety of conformational changes can accompany protein association.
  - Ranging from local adjustments of side chains involving atom displacements of  $< 1 \text{ \AA}$  to folding/refolding of protein segments
- „true induced-fit“ vs. conformational selection of near bound conformations from an ensemble of unbound conformations.

# Docking with bound protein structures

- Docking with „bound“ protein structures is easier than using „unbound“ conformations
  - Algorithms that are based purely on surface complementarity can often detect near-native docking solutions as top ranking (using bound structures)
- Even local conformational changes at an interface can significantly perturb surface complementarity.



# Types of conformational changes in proteins

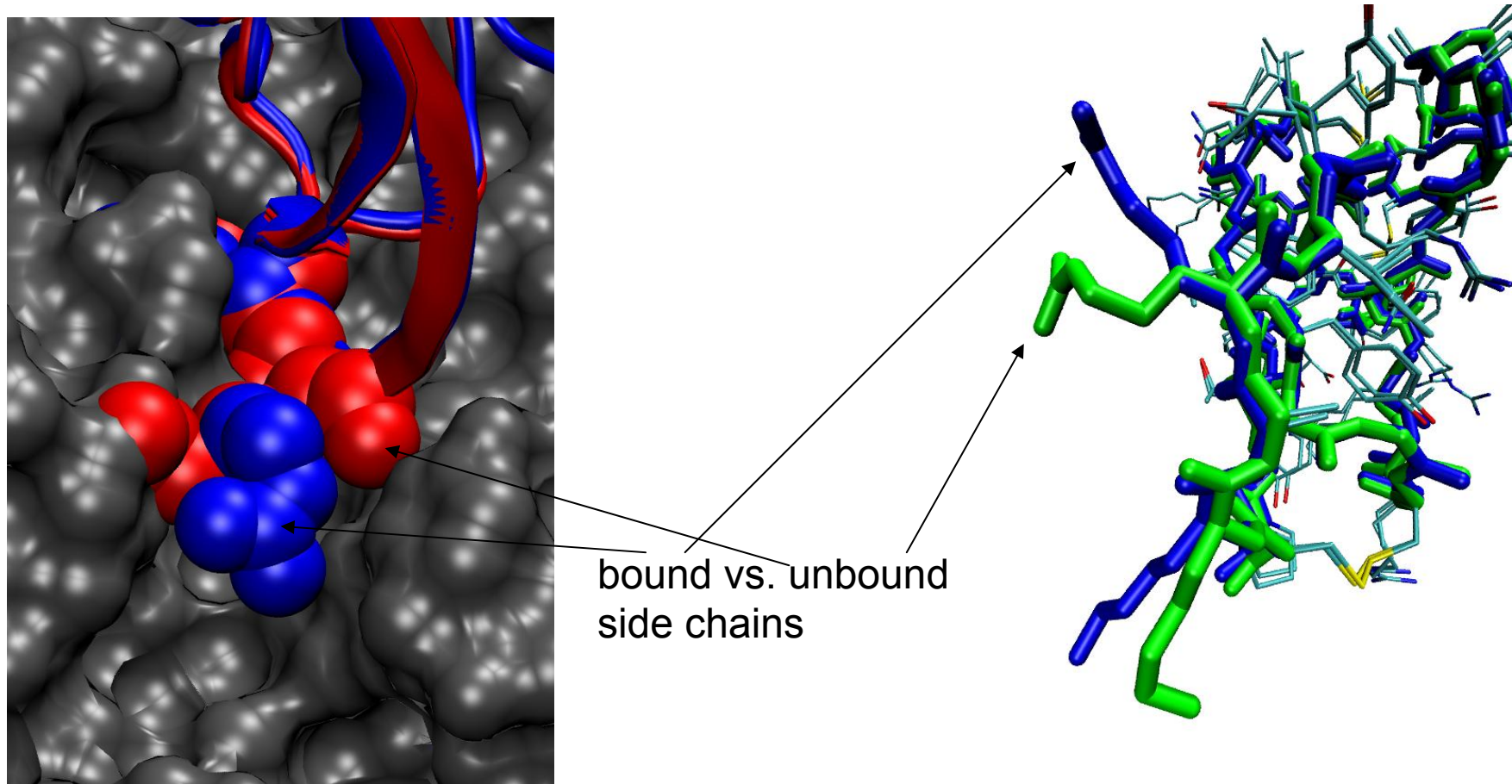
- **Large scale motions:**

Type of motion	Time Scale	Amplitude
Side chain motions (protein surface)	0.1 ps- 0.1 ns	1-5 Å
Backbone motions in protein loop regions :	several ns	1-10 Å
Motions of the N- or C-terminus of a protein:	several ns	1-5 Å
Rigid body motions of secondary structures :	0.05 – 1 µs	1-5 Å
Protein domain motions :	1 µs – 1 ms	5-10 Å
(for example hinge bending motions)		
Allosteric transitions:	1 µs – 100 ms	5-10 Å
(correlated motion of several subunits)		
Local folding and unfolding transitions	0.1 µs – 10 ms	~5 Å
(helix-coil transitions, loop folding)		

(from McCammon & Harvey, Dynamics of proteins and nucleic acids, Cambridge University Press)

# Types of conformational changes upon complex formation

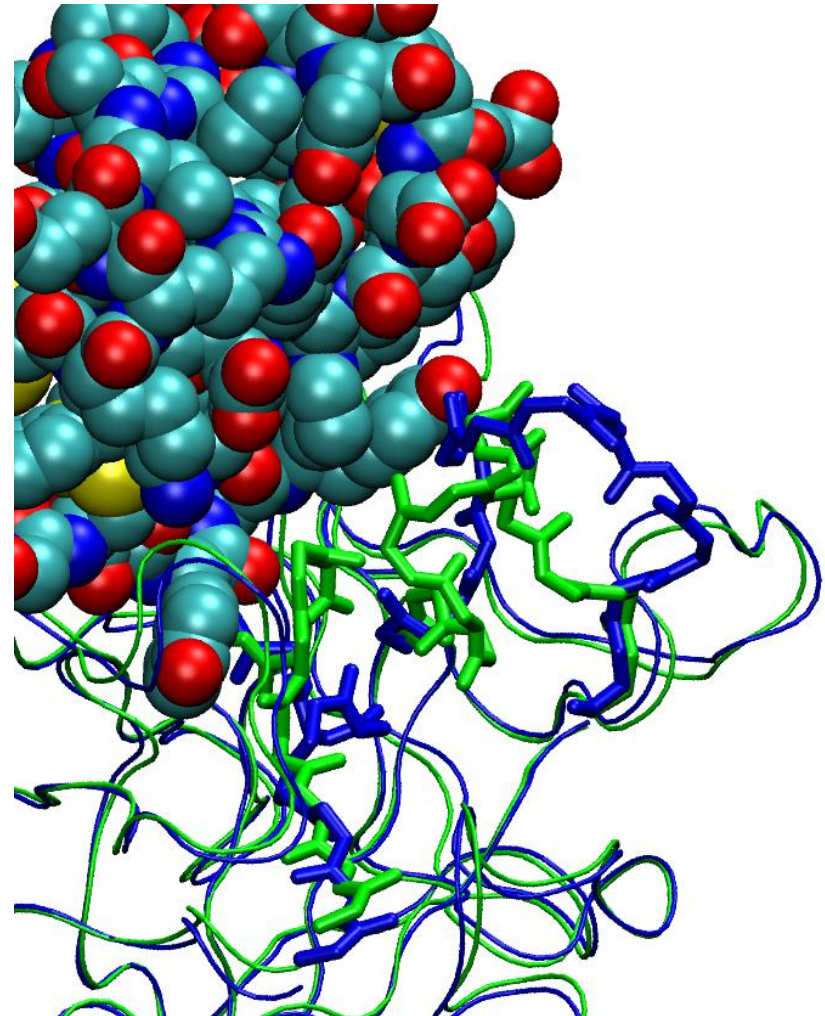
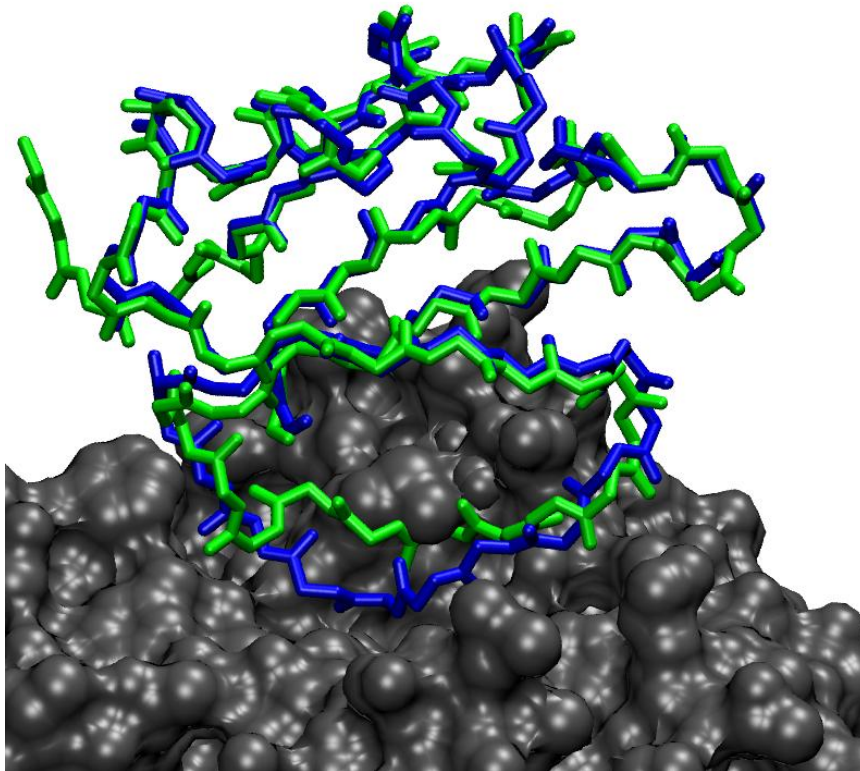
- Side chain conformations in bound and unbound structures may differ.
  - Often seen for side chains such as Lys and Arg with long flexible aliphatic tail.
- Can result in sterical overlap in case of rigid docking.





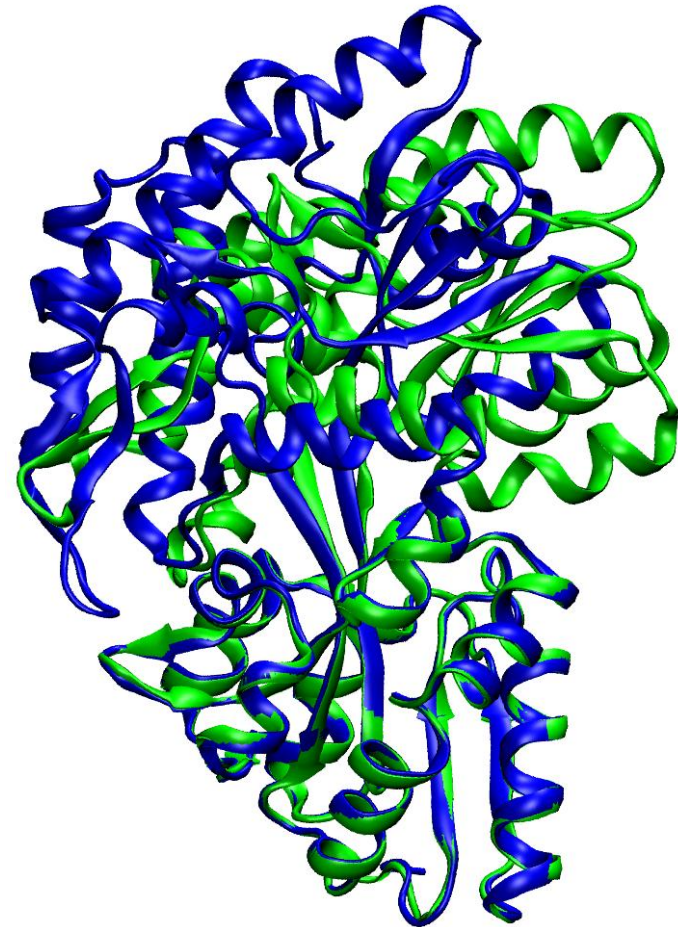
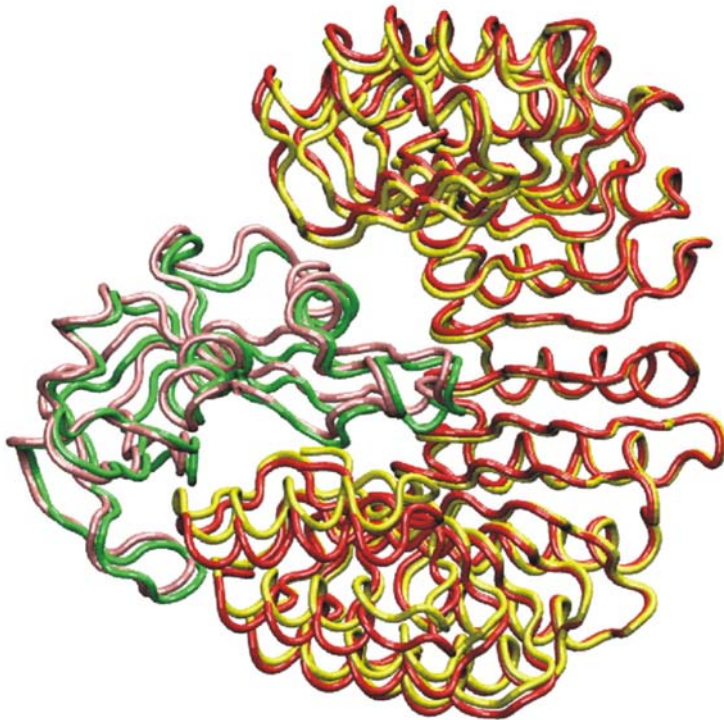
# Localized backbone changes upon association

- Frequently, not only side chains but also local backbone segments (loops) undergo conformational changes during complex formation.
- Sterical overlap; strong deviation of docked complex from native complex structure



## Global backbone changes upon association

- **Global changes**
  - may involve domain-domain rearrangement
  - collective adjustment of large protein segments

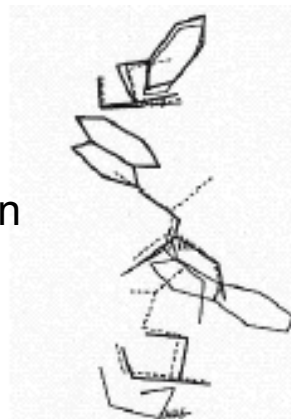




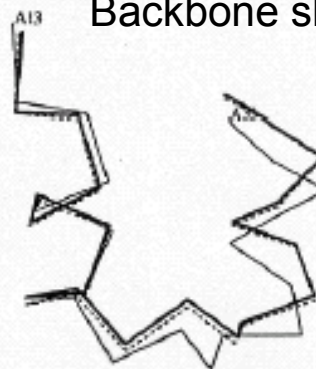
# Docking using protein model structures

- Frequently protein-protein docking requires to use homology modeled structures.
  - Quality of model structures depends on sequence similarity to template structure and on the modeling procedure.
    - Possible errors in target-template alignment
    - Structural inaccuracies in segments with low sequence similarity
    - Possible errors in modeled surface loops and side chains

Incorrect side chain placement



Backbone shift



Incorrect loop



## **Docking using protein model structures**

- **Docking of model structures is typically more difficult than docking using experimental structures**
  - **Most difficult CAPRI-targets involved homology models**
  - **Docking procedure must either tolerate large errors in protein conformation**
  - **or allow explicitly for significant conformational changes at the interface during docking that “reverse” the modeling errors**

# Outline

- Conformational changes in proteins upon association
- **Methods to model conformational changes**
- Strategies to account for conformational changes
- Explicit flexibility during docking
- Own docking approach

# Computational methods to model protein conformations

- **Systematic conformational generator approaches**

- based on peptide backbone segments
- based on systematic dihedral angle sampling
- based on stable side chain rotamer states

Example: CONGEN (Brucoleri& Karplus 1987. Biopolymers 26, 127)

- **Molecular dynamics simulations**

- **Monte Carlo simulations**

- **Normal mode calculations**

- **Distance geometry methods**

- Method generates possible structures compatible with a set of distances between atoms

Examples: CONCOORD (de Groot et al. 1997. Proteins 29, 240)

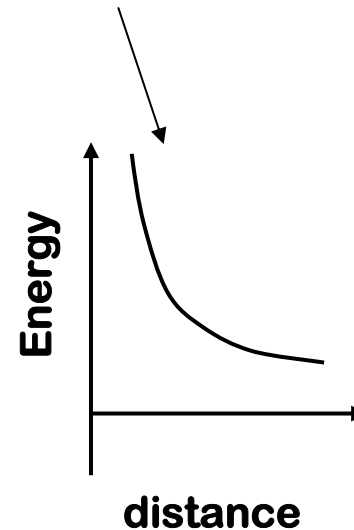
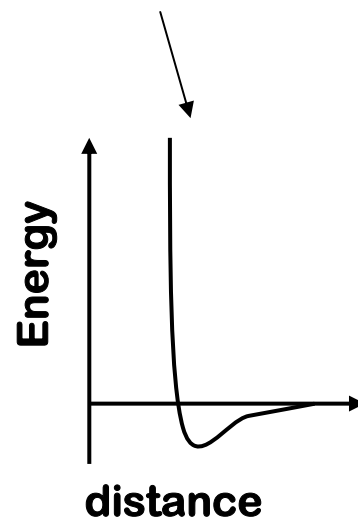
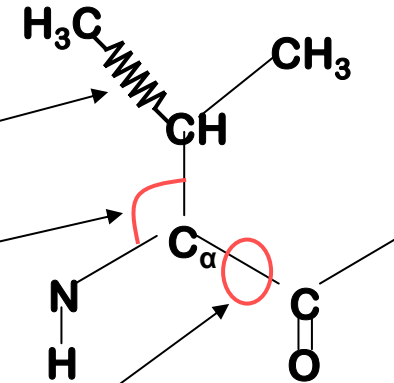
- **Basis of most methods is a molecular mechanics force field**

# Molecular mechanics force field for a protein

Force field energy of a molecule:

$$V(r_1, r_2, \dots, r_n) =$$

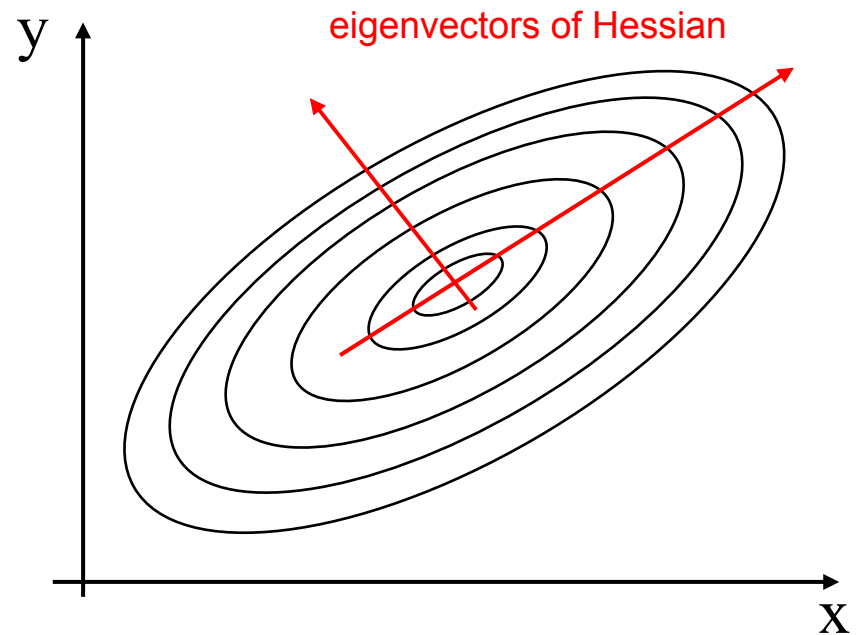
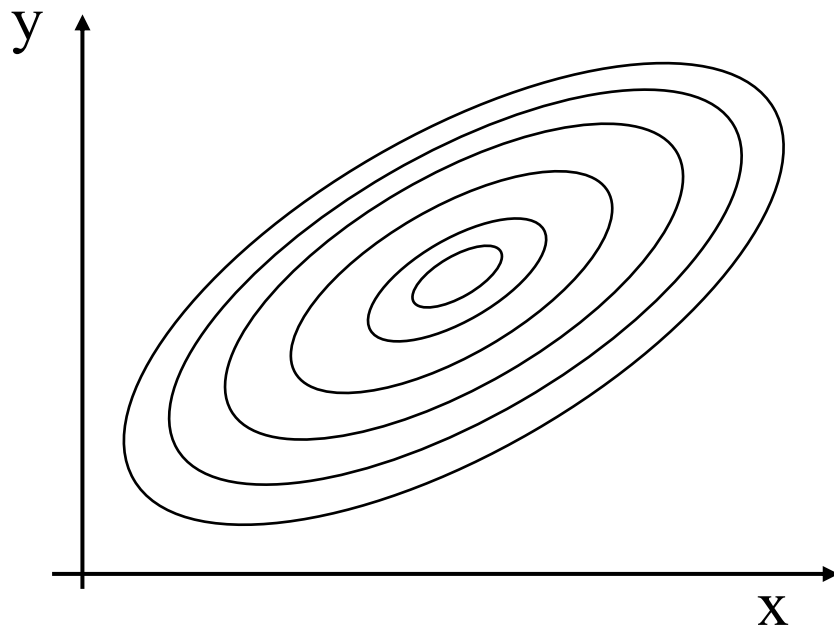
$$\begin{aligned} & \sum_{\text{Nbonds}} \frac{1}{2} k_{bi} (b_i - b_{i,0})^2 \\ & + \sum_{\text{Nangles}} \frac{1}{2} k_{\theta i} (\theta_i - \theta_{i,0})^2 \\ & + \sum_{\text{Ntorsions}} \sum_{n=1..N_i} k \tau_{ni} (1 + \cos [n_i \tau_i - \delta_i]) \\ & + \sum_{\text{nbpairs}} \epsilon_{ij} [(\sigma_{ij}/d_{ij})^{12} - (\sigma_{ij}/d_{ij})^6] + q_i q_j / (4\pi\epsilon_o d_{ij}) \end{aligned}$$





# Normal mode analysis

- Taylor expansion of the energy function at energy minimum
  - First derivative of energy function is zero.
  - Curvature locally determined by second derivative (Hessian) of the energy function
  - Diagonalization of the Hessian yields eigenvectors that correspond to collective (orthogonal) degrees of freedom.
  - Eigenvectors can be ordered according to eigenvalues (corresponding to force constants (or frequencies) for deformations along corresponding eigenvectors)



# Approximate normal mode calculations based on elastic network models

- Elastic networks describe the interaction between atoms in a protein by harmonic springs.
- Model by Hinsen (Proteins 1998, 33, 417.):

$$E(R_1, \dots, R_N) = \sum_{\text{C}\alpha\text{-pairs}} E_{ij}(R_i - R_j)$$

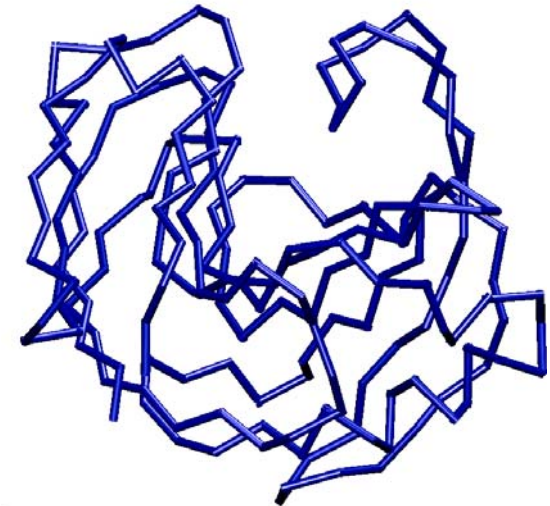
$$E_{ij}(r) = k(R_{ij}^0) (|r| - R_{ij}^0)^2$$

$$k(r) = c \text{Exp}[-|r|^2 / r_o^2]$$

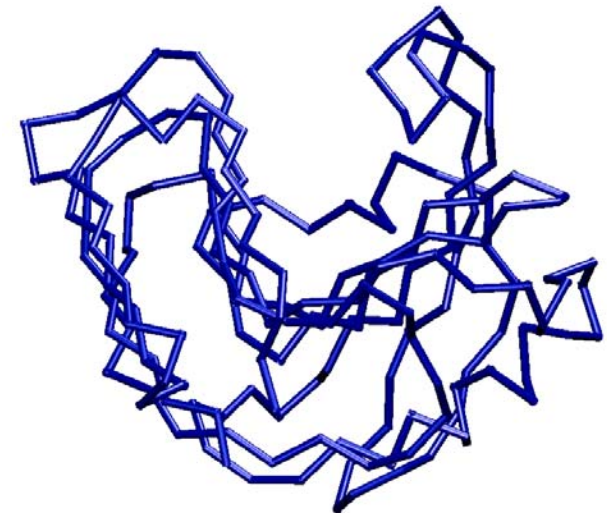
- Spring force constant decreases with distance (other methods use a cutoff)
- Results in global collective modes that are similar to normal modes calculated at atomic resolution.

Tirion, Phys Rev Lett 1996;77:1905-1908.  
Bahar et al. Folding Design 1997;2:173-181.  
Hinsen K. Proteins. 1998;33:417-429.

Backbone of Xylanase



Mode 1

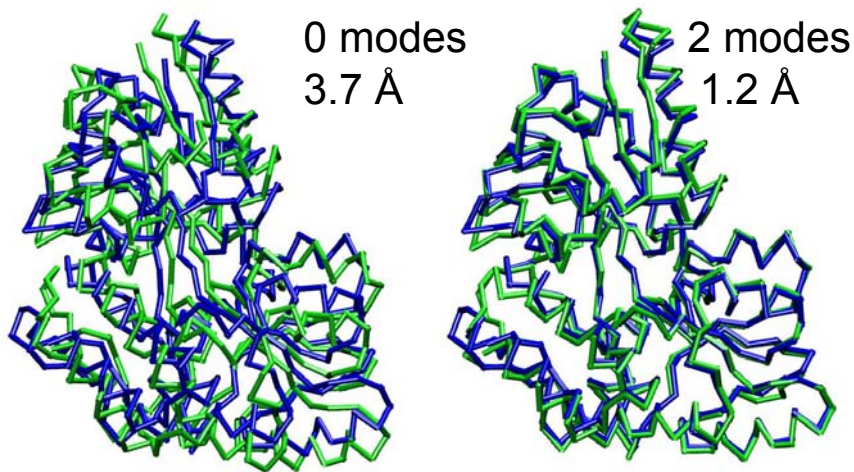
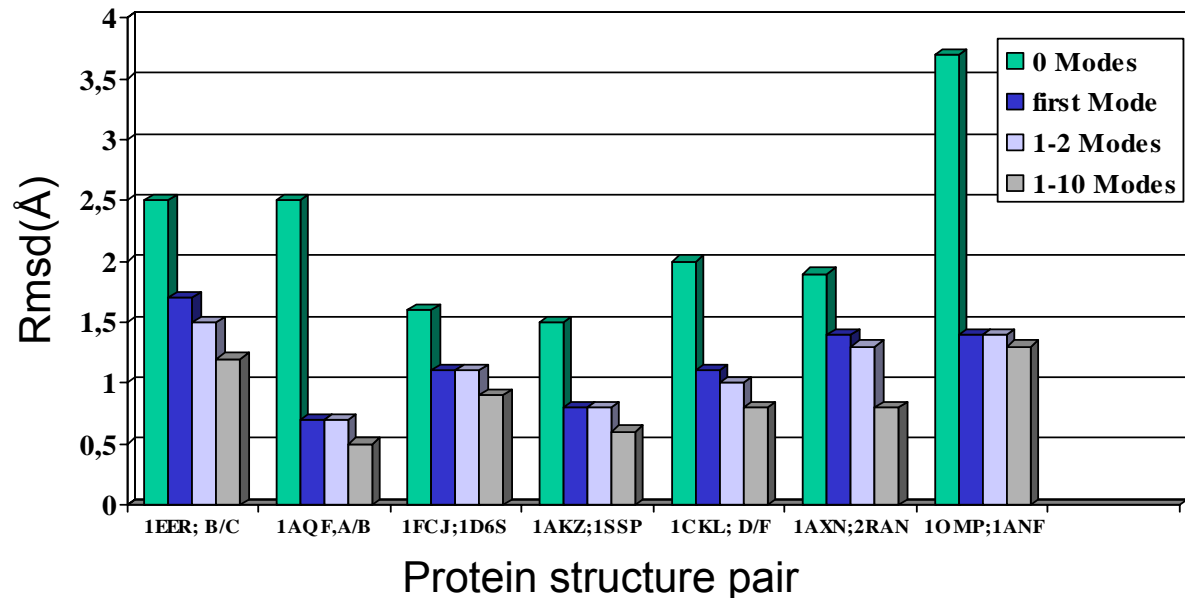


Mode 2

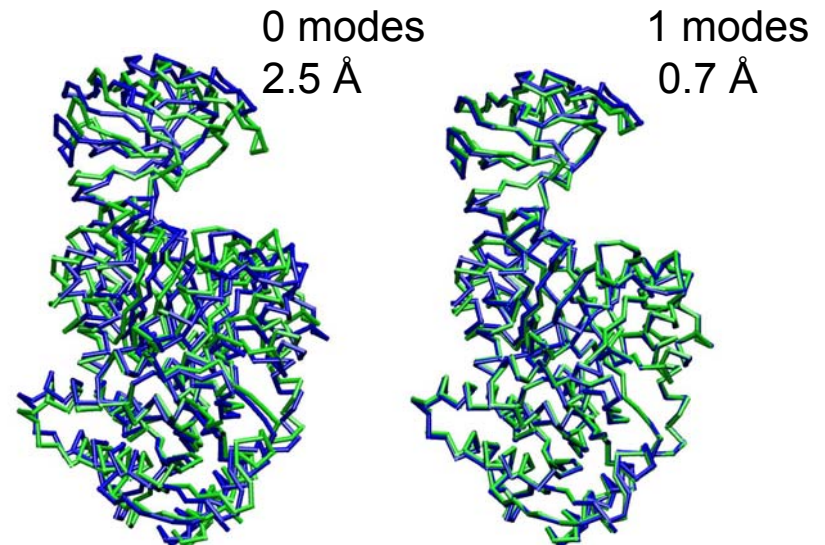
# Observed global motions vs. approximate harmonic modes

- Can experimentally observed global changes be approximated by pre-calculated soft modes?

Maltose-binding protein (**bound** vs. **unbound** (1anf vs 1omp)



Pyruvate kinase (1aqf; chain A/B)



Investigated by:

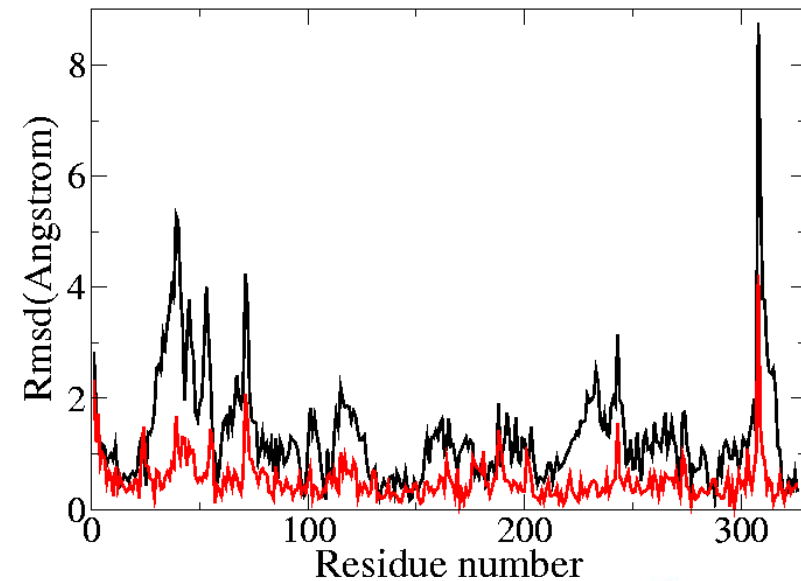
Tama & Sanejouand 2001. Protein Eng. 14, 1.

Lindahl & Delarue 2005, NAR 33, 4496.

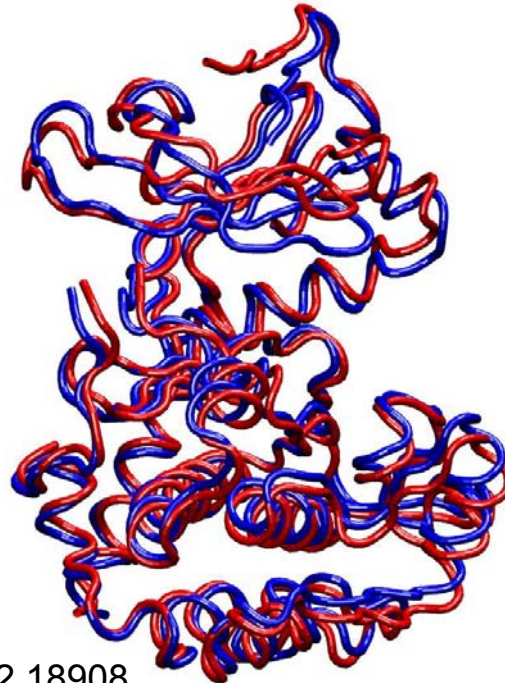
Dobbins et al. 2008, PNAS 105, 10390.

# Proteinkinase A (apo vs. bound structure)

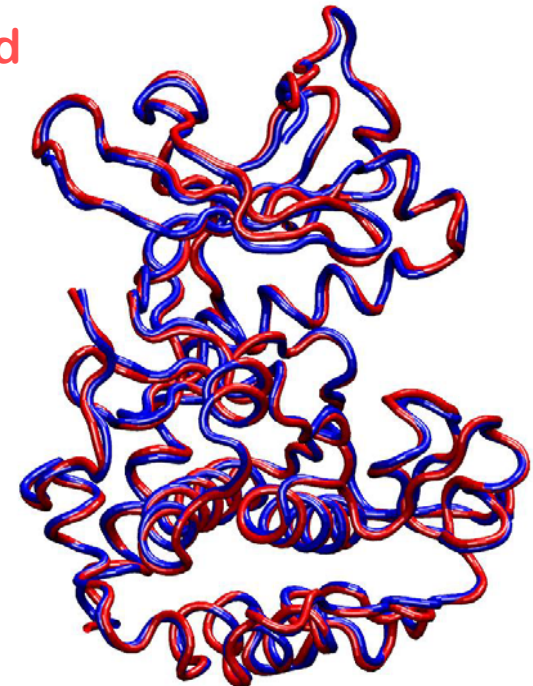
- cAMP-dependent protein kinase (PKA) undergoes global conformational changes upon ligand binding
  - Apo form: pdb1j3h
  - Balanol bound form: pdb1bx6
- 10 modes (Apo-form) can reduce backbone RMSD from 1.65 Å to 0.65 Å
- First mode alone: 0.93 Å



Apo vs.  
bound PKA

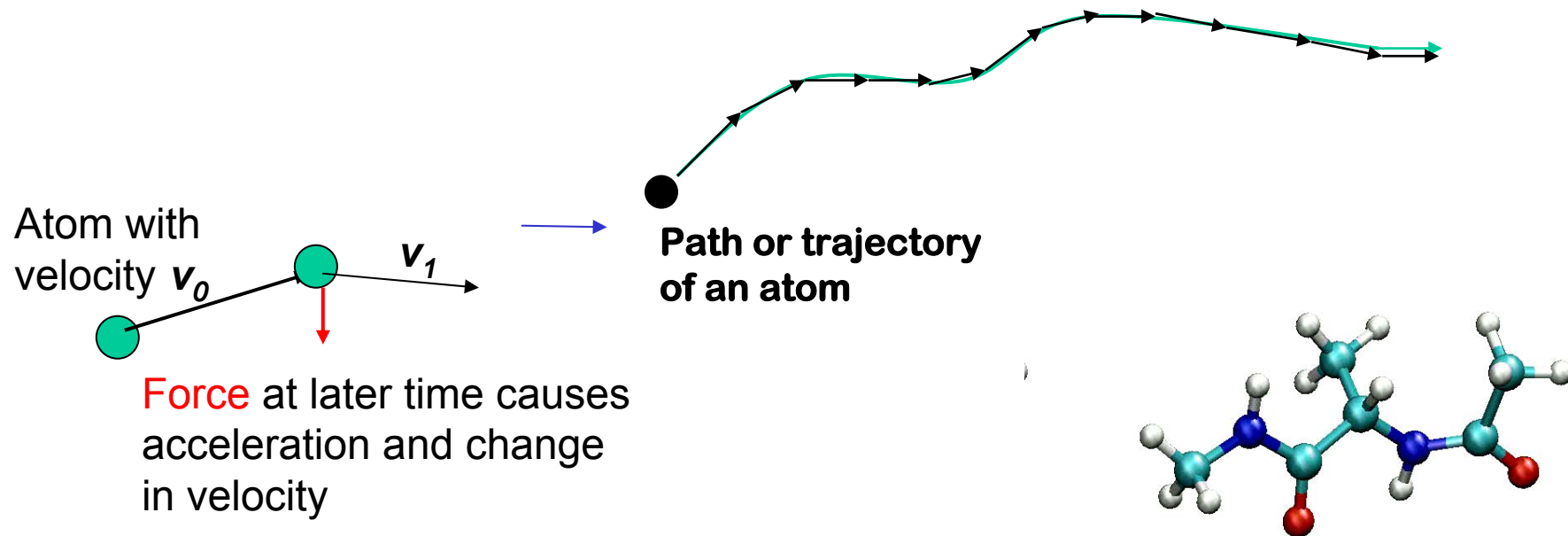


Mode deformed  
vs. bound PKA



# Molecular dynamics simulations

- The equations of motion for a system of interacting particles can be integrated numerically in small time steps.
- The resulting set of (discrete) coordinates (trajectory) for each atom (particle) is an approximation to the “real” path the atom takes in time:





# Replica-exchange molecular dynamics

- **Multi-temperature replica exchange MD:**
  - Replicas of the system are run at  $N$  temperatures ( $T_1, \dots, T_i, T_j, \dots, T_N$ )
  - Exchange between replicas  $i, j$  (at neighboring  $T$ ), accepted according to:

$$w(x_i \rightarrow x_j) = 1 \quad \text{for } \Delta \leq 0;$$

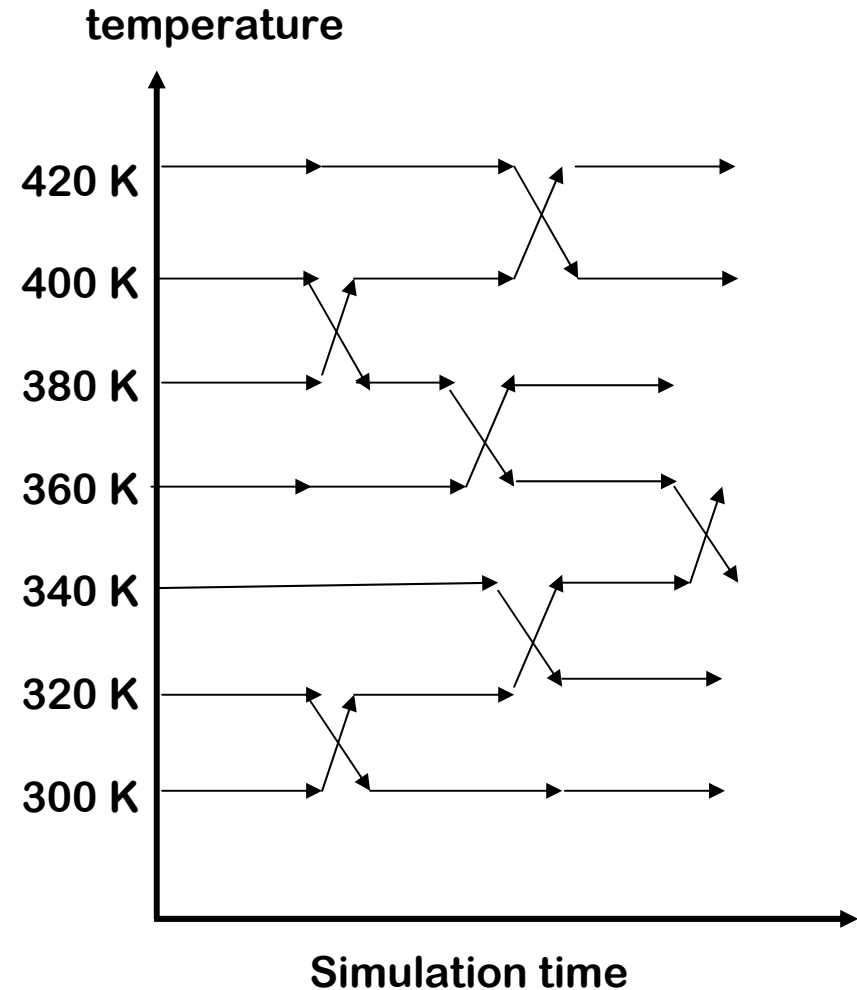
$$w(x_i \rightarrow x_j) = \exp(-\Delta) \quad \text{for } \Delta > 0$$

where

$$\Delta = (\beta_i - \beta_j) [E(r_j) - E(r_i)]$$

Momenta are adjusted according to:

$$p[i] = \sqrt{T(i)/T(j)} p[j]$$



# **Molecular dynamics simulations can be used to study local and global motions of a protein**

- **Side chain and loop motion on the nanosecond time scale**
- **Can be used to select alternative side chain and loop structures**
  - **Camacho et al. (2004, 2005) used MD simulations to predict near native side chain structures for anchor residues in unbound protein structures.**
- **Global motions can be extracted by principle component analysis of the positional covariance matrix (essential dynamics, Amadei et al., 1993)**
  - **Smith et al. (2005) have used MD simulations to analyse global conformational fluctuations in proteins and the relation to conformational changes upon association.**

Rajamani et al. 2004. PNAS 101, 11287.

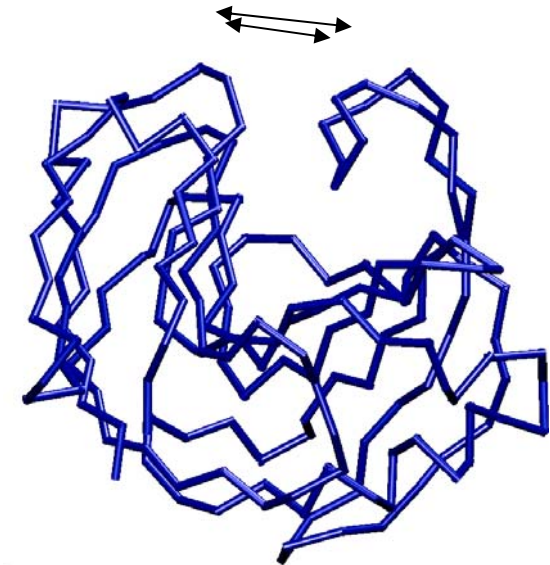
Camacho, 2005. Proteins, 60, 245.

Amadei et al. 1993. Proteins 17, 412.

Smith et al. 2005. JMB 347, 1077.

# Combining elastic network calculations and molecular dynamics simulations

- ENM calculations can help to rapidly identify soft flexible degrees of freedom of a protein.
  - Low resolution view of a structure
- Distance fluctuations compatible with the ENM model can be calculated by excitation in each mode
- The distance fluctuations indicate the range of sterically allowed deformations.

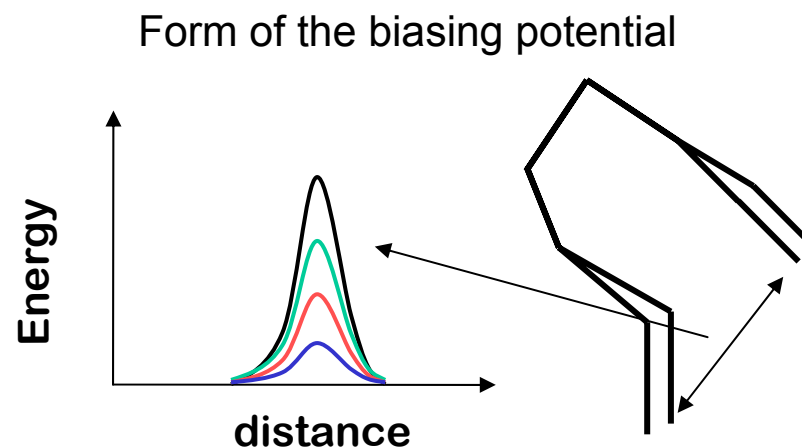


# How to combine ENM analysis and MD simulation?

- Add a biasing (flooding) potential for distance fluctuations derived from ENM analysis for each replica.
- Biasing potential for C $\alpha$ -C $\alpha$  distances or heavy atom distances

$$E(d_{ij}) = k \left( \left[ d_{ij} - d_{ij0} \right]^2 - \Delta d_{ij}^2 \right)^2, \quad \text{if } |d_{ij} - d_{ij0}| \leq \Delta d_{ij}$$

$$E(d_{ij}) = 0, \text{ otherwise}$$



- Use Hamiltonian replica exchange with different levels of the biasing potential

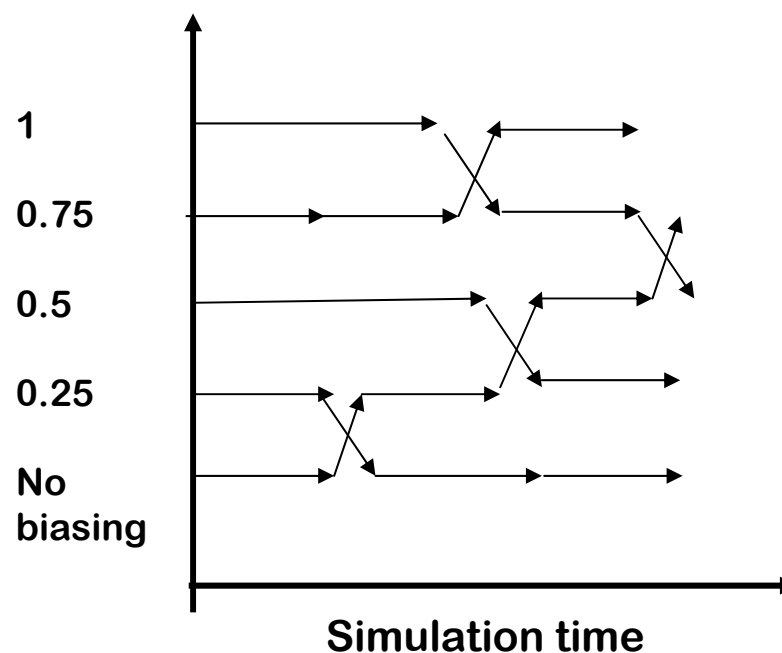
$$w(x_i \rightarrow x_j) = 1 \quad \text{for } \Delta \leq 0;$$

$$w(x_i \rightarrow x_j) = \exp(-\Delta) \quad \text{for } \Delta > 0$$

where

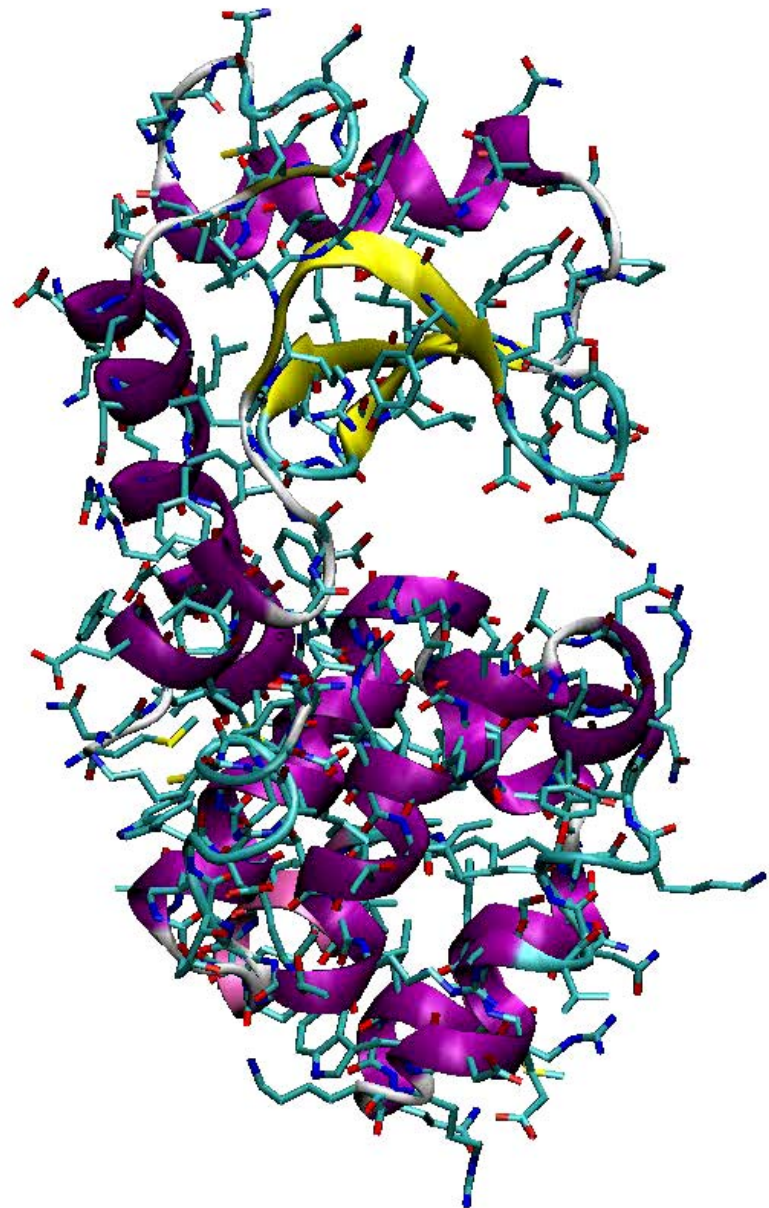
$$\Delta = \beta \left[ \left( E^j(r_j) - E^j(r_i) \right) - \left( E^i(r_j) - E^i(r_i) \right) \right]$$

Biasing level



# Application to T4 lysozyme

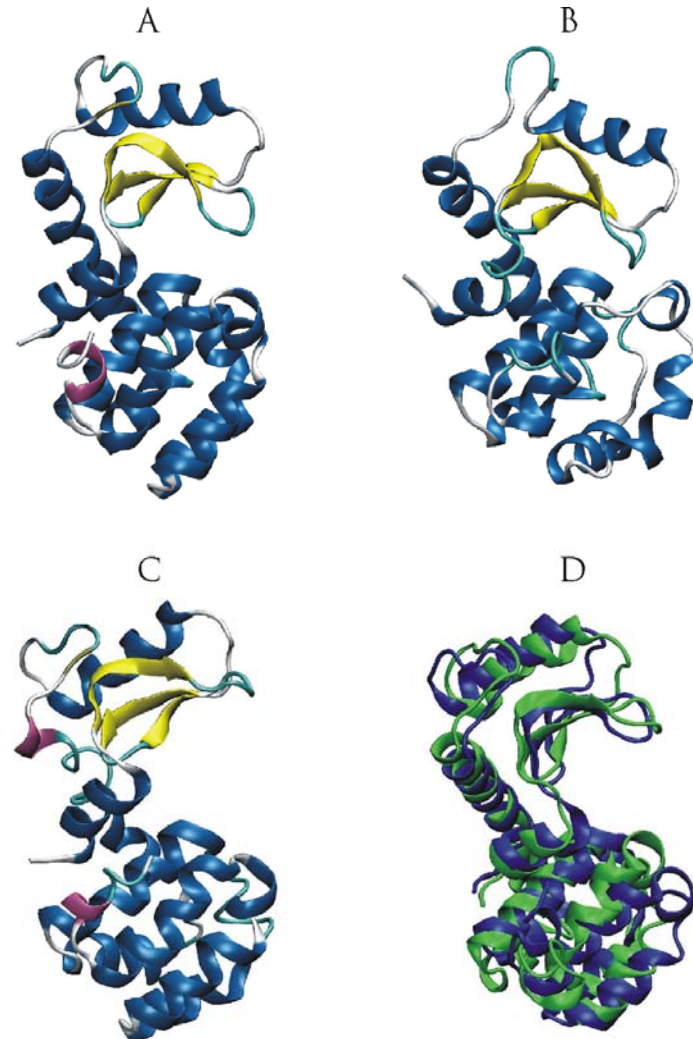
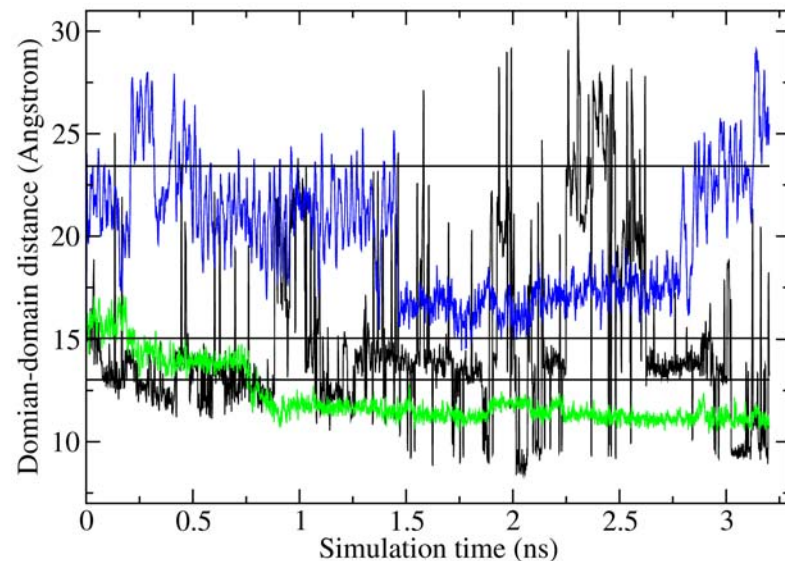
- More than 200 structures of T4L in the data base
- Can adopt open and closed structures
  - Simulations using Amber parm03 force field at 310 K, GB model
  - 2LZM start (a closed form)
  - 5 biasing levels (including the original force field)
  - ENM calculation for CA atoms every 20 ps.
- Total simulation time: 3.2 ns





# Application to T4 lysozyme

- T4L flips between open and closed states many times
- Comparison with conventional MD simulation starting from closed and from an open form
  - No open-closed transition during conventional MD on the 3.2 ns time scale



# Outline

- Conformational changes in proteins upon association
- Methods to model conformational changes
- **Strategies to account for conformational changes**
- Explicit flexibility during docking
- Attract docking approach

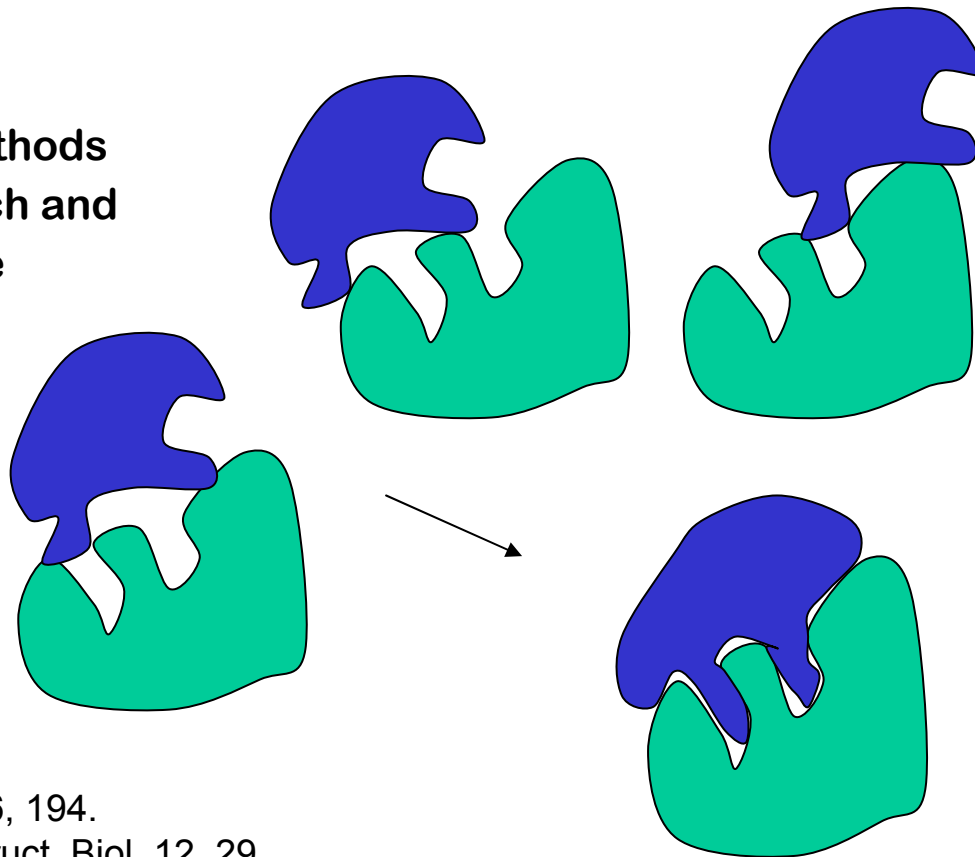
# Strategies to account for conformational changes during docking

## Two possibilities:

Inclusion of conformational changes during entire docking search

Rigid docking followed by allowing conformational changes in a second step

- The majority of docking methods follows the second approach and may include several flexible refinement steps



Reviewed in:

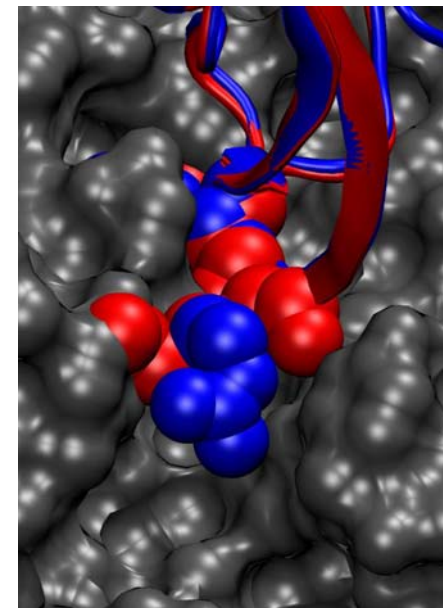
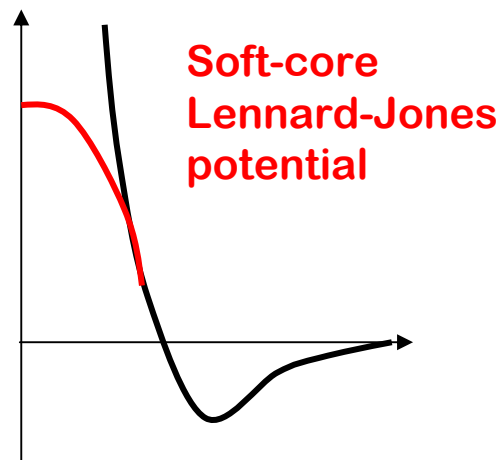
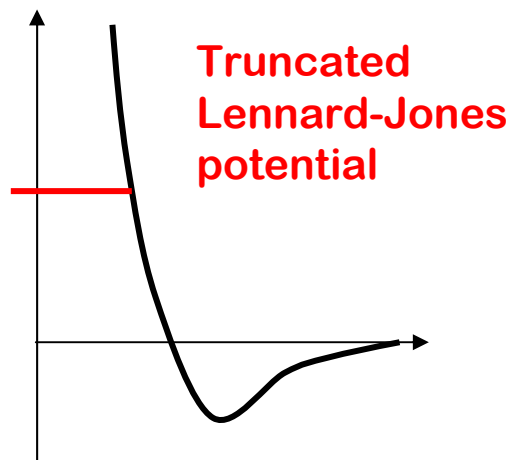
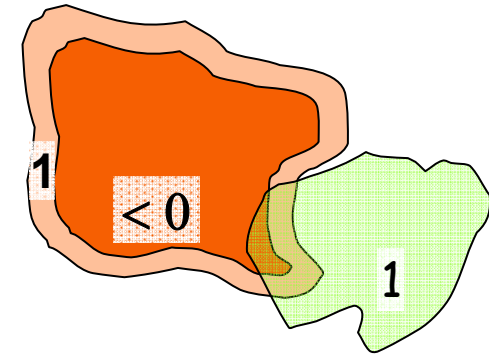
Andrusier et al. 2008. Proteins 73,271.

Bonvin, 2006. Curr. Opin. Struct. Biol. 16, 194.

Smith & Sternberg, 2002. Curr. Opin. Struct. Biol. 12, 29.

# Soft docking: Accounting implicitly for small conformational changes

- Rigid docking with a soft protein boundary
  - Correlation methods:
    - Smoothing/softening the protein surface boundary
    - Increasing the tolerance for receptor-ligand overlap
- Rigid docking with soft or truncated non-bonded potentials
- Pruning (removing) of side chains during docking



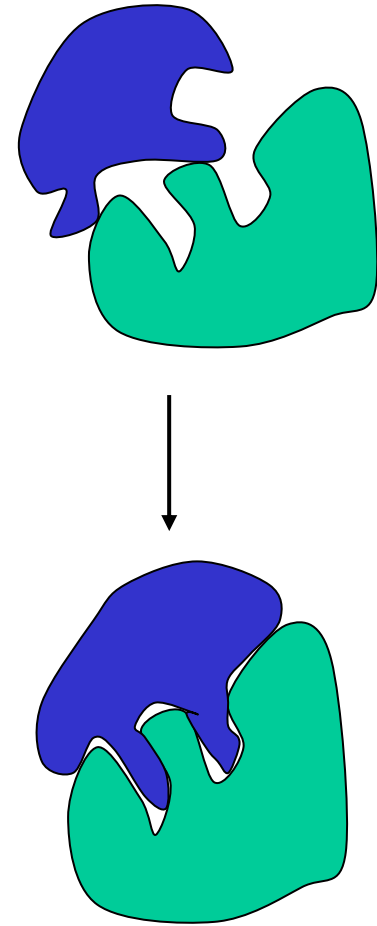
# Accounting for conformational changes on a subset of docking solutions

- The first rigid docking phase results in a large set of structures.
- It is hoped that the pool of solutions contains complex geometries sufficiently close to the native complex.
  - Experimental information, application of different scoring schemes can help to limit the number of docking solutions.



# Accounting for conformational changes on a subset of docking solutions

- In principle, changes of both backbone and side chain structure need to be allowed.
- Procedure must be sufficiently fast to deal with several hundred or even thousands of complexes.
- Ideally, docking refinement should improve complex geometry and ranking.



# Modeling side chain conformational changes

- Side chain refinement by:
  - Systematic methods
  - All systematic methods assume rigid backbone
  - Reduction of search space by considering only discrete side chain conformations (rotamers)
    - Side chain rotamer structures have been derived from analysis of known structures
    - Backbone dependent and independent rotamer libraries
  - Global optimization problem to minimize sterical overlap between side chains

Energy-score of a side chain structure:

$$E_{\text{rotamer combination}} = \sum_i^{\text{Nresidue}} E_i(\text{rotamer } r) + \sum_{i,j} E_{i,j}(\text{i} \rightarrow \text{rotamer } r, \text{j} \rightarrow \text{rotamer } s)$$

# Modeling side chain conformational changes

- **Systematic exploration of all possible combinations**
  - Possible for a small set of side chains
  - Efficient if the side chains show little overlap (independent search for each side chain)
- **Self-consistent mean field optimization**
  - **Algorithm:**
    - 1.Stores a weight for each side chain rotamer
    - 2.Calculates the interactions of each side chain rotamer with all other residues (multiplied with the weight)
    - 3.Update of weights (Boltzmann Probability based on Interactions)
    - 4. go to 1 or terminate if weights do not change.
  - **Used in 3D-DOCK (Jackson et al. 1998), Mc2 (Bastard et al. 2003) and Attract (Bastard et al. 2006)**

Jackson et al. 1998. JMB 276, 265.

Bastard et al. 2003. JCC 24, 1910.

Bastard et al. 2006. Proteins 62, 956.

# Modeling side chain conformational changes

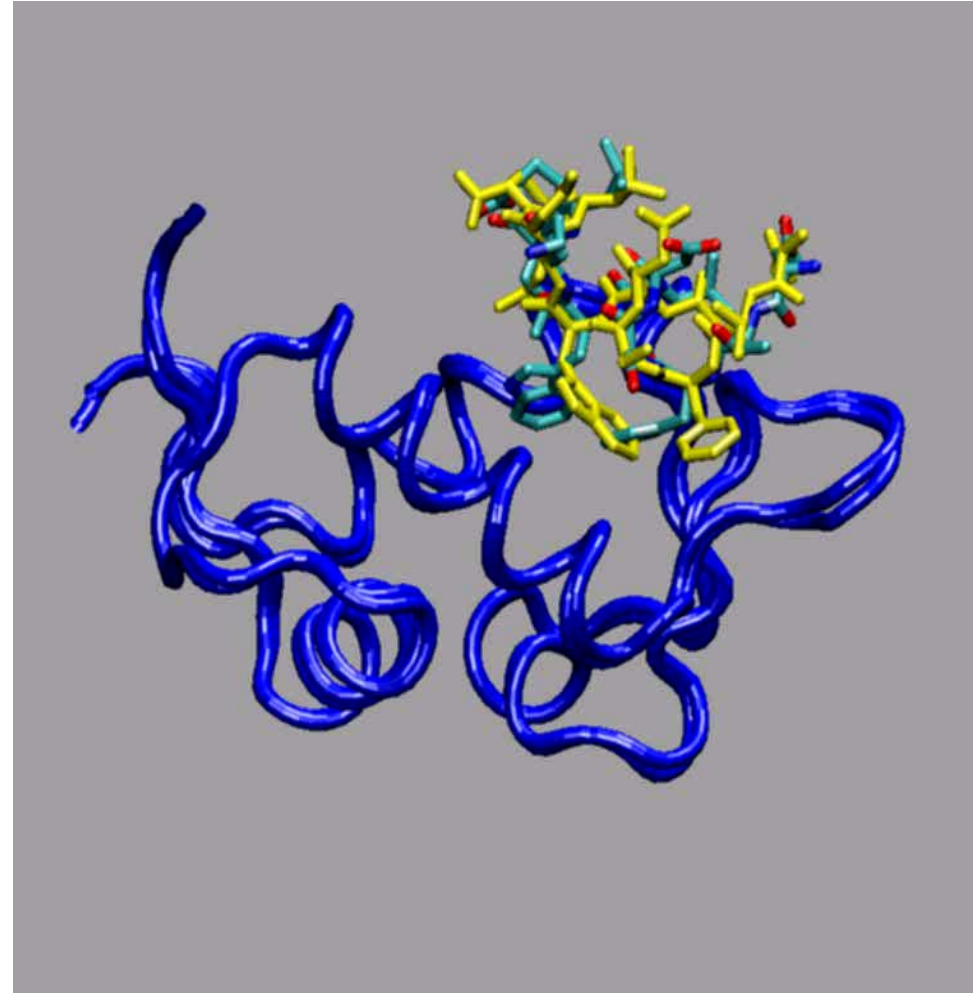
- **Dead-end-elimination methods**
  - A method to systematically eliminate side chain rotamers that cannot be part of the global minimum
  - A rotamer is removed if another rotamer has a lower energy for every rotamer combination of all other residues.
  - Variants of DEE are implemented for example in SCWRL (Canutescu et al., 2003) and FireDock (Andrusier et al., 2007)

# Molecular dynamics simulations of docked complexes

- Conformational adjustments by molecular dynamics (MD) simulations:
  - Allows for larger conformational changes (by crossing energy barriers) compared to EM.
  - Backbone and side chain motions can be included
  - Solvent molecules can be included.
  - Coupling with advanced sampling methods (simulated annealing, replica-exchange)
  - Quality of final results depends on force field conditions and experimentally derived restraints

# Refinement of docking interfaces

- Low resolution protein-protein docking models may require refinement at atomic resolution
- Development of an efficient molecular dynamics/potential scaling method for side chain refinement at interfaces
- Test system:
  - **MDM2 domain** docked to human p53 peptide domain



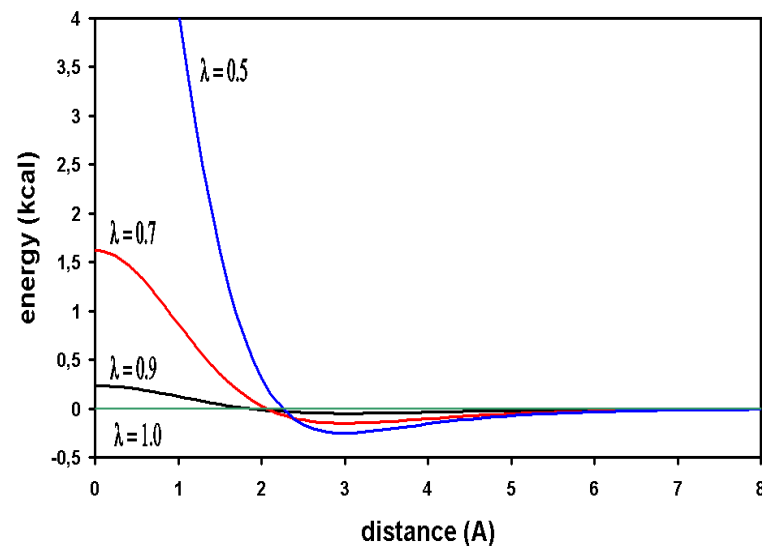
# Interface refinement using potential scaling

## Method:

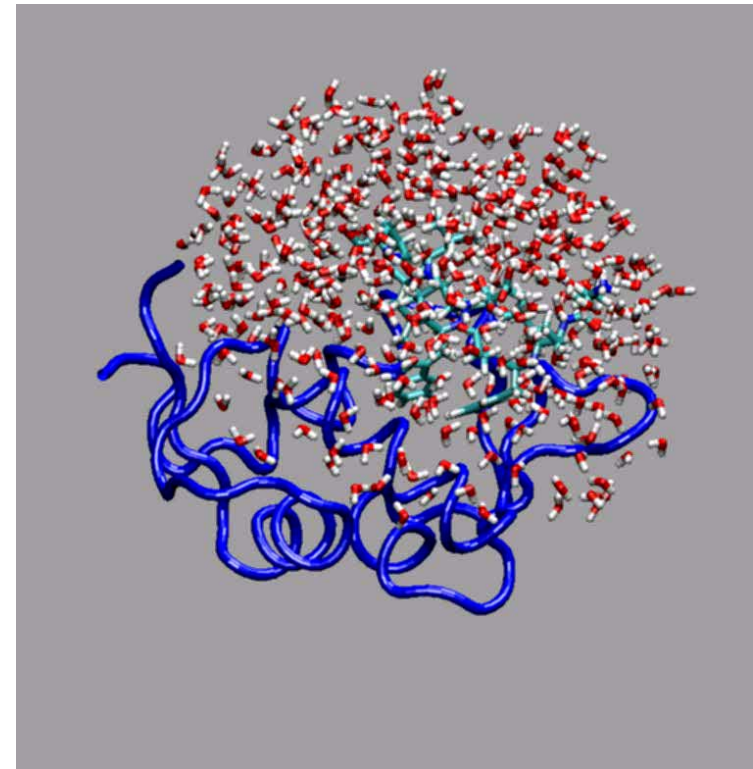
- select residues at the protein-protein interface
- set *van der Waals* and *electrostatic* interactions including these residues to zero at MD simulation start.
- Smooth re-scaling of interactions using a scaling factor  $\lambda$  during MD

## Advantage:

- inclusion of *adjustable flexibility* ( $\pm 2\text{\AA}$ ) of all (non-interface) atoms
- applicable in explicit water

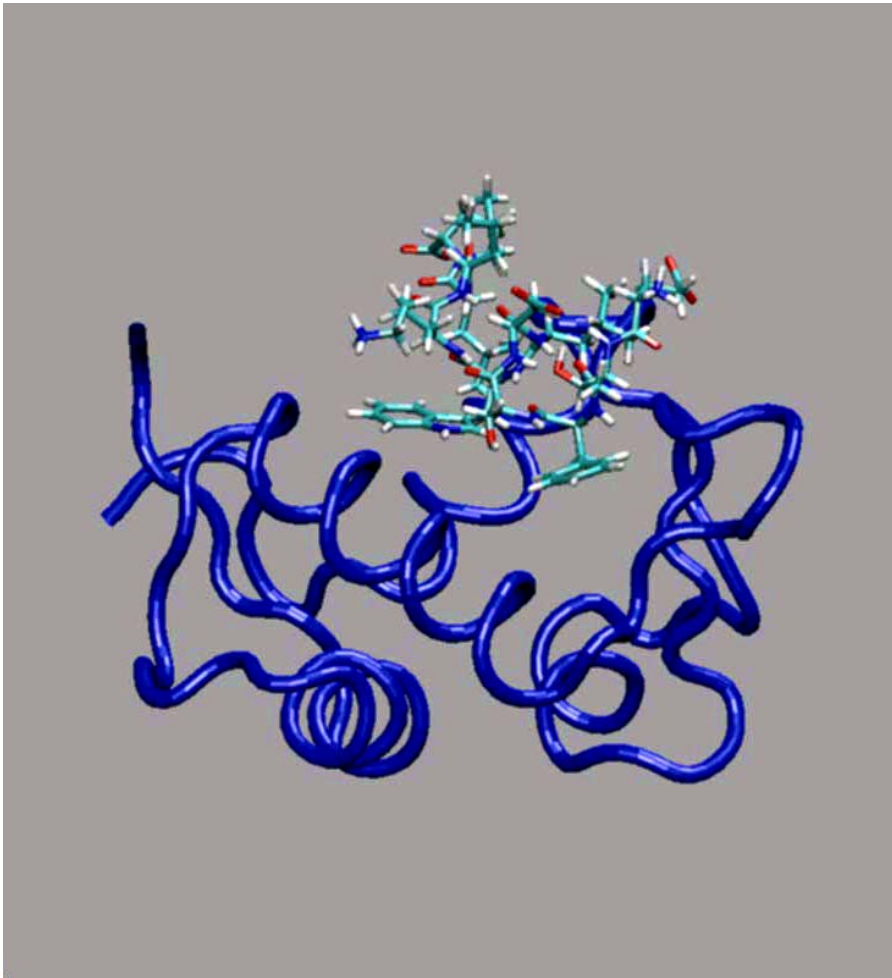


Softcore scaling according to: Zacharias, Straatsma, McCammon, J. Chem. Phys., 100, 9025 (1994).

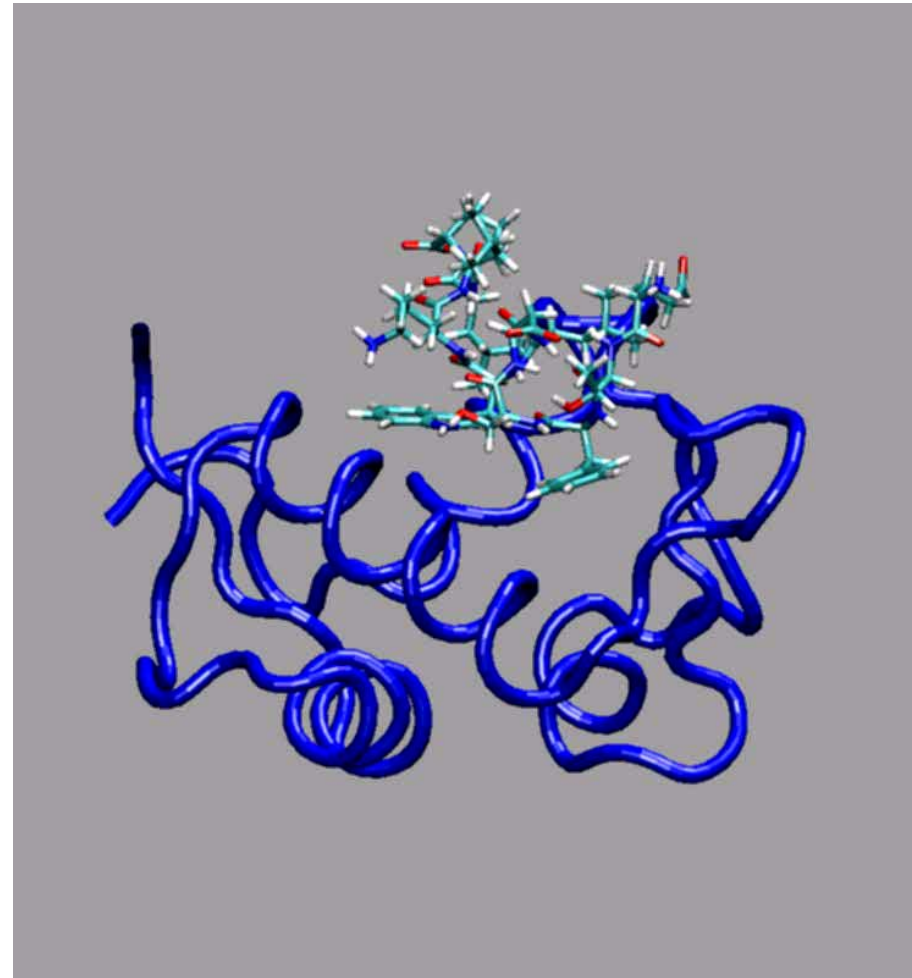




# Application to a docking interface

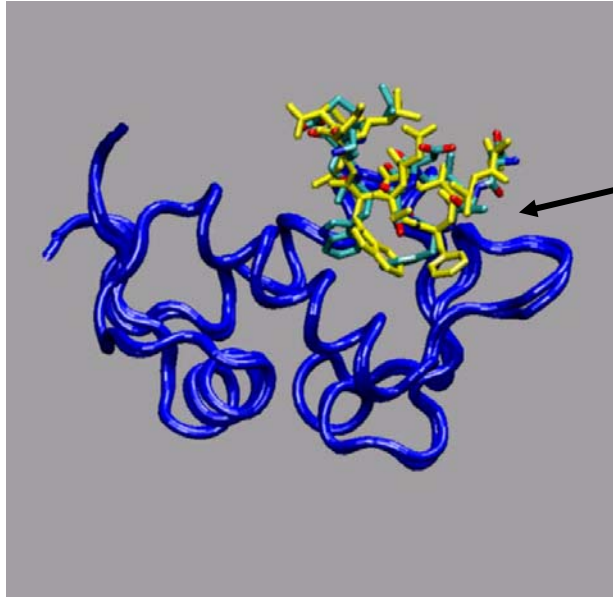


**Standard MD simulation of the protein-peptide complex**

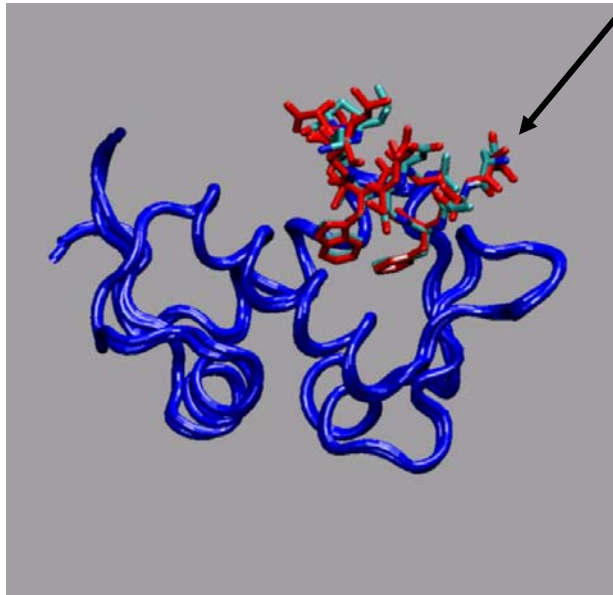


**Increased mobility of interface residues during potential scaling**

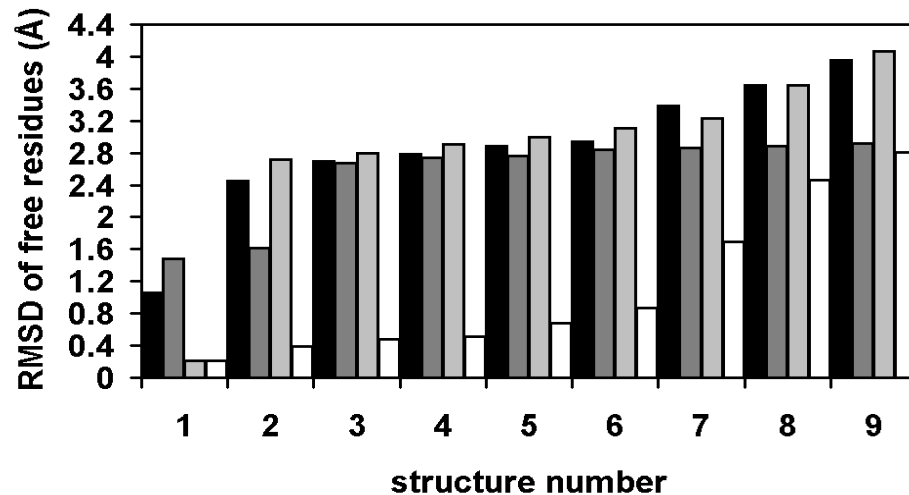
# Side chain prediction at docking interface



- Refinement of 9 start structures with „incorrect“ side chains and backbone deviation of up to 1.5 Å
- PS-MD and standard MD in the presence of a water cap
- PS-MD results in realistic predictions within ~0.3 ns simulation time.



**Black bar:** start side chain structure  
**Dark grey:** SCWRL3.0 (Canutesu et al. 2003)  
**Light grey:** standard MD (315 ps)  
**White:** potential scaling MD (315 ps)



# Monte Carlo methods

- Heuristic method (similar to MD no guarantee for finding best possible solution)
- Use of simulated annealing to overcome energy barriers
  - Fast because only interactions close to mobile side chains need to be calculated
  - Various (non-differentiable) energy functions can be used
  - Step size can be adapted, e.g. switching between rotamer states (larger conformational changes per step than in MD simulations)
  - Possibility to combine it with (limited) backbone motion

# Approaches that employ Monte Carlo simulations

- **RosettaDock** (Gray et al., 2003; Wang et al.2005)
  - Uses MC steps in side chain rotamers + gradient based EM of dihedral angles; MC steps in backbone dihedrals can also be included.
- **Biased probability MC methods** (Fernandez-Recio et al., 2002;2007)
  - Uses random changes in backbone and side chain dihedrals and subsequent EM.
- **Replica-Exchange MC simulations** (Lorenzen & Zhang, 2007)
  - T-RexMC simulation on side chain dihedrals and rotational + translational degrees of freedom of the partners

Wang et al. 2005. Protein Sci 14, 1328.

Jackson et al. 1998. J Mol Biol 276, 265.

Gray et al. 2003. J Mol Biol 331, 281.

Fernandez-Recio et al. 2002 Prot. Sci. 11,280; 2007, Proteins 52, 113.

Lorenzen & Zhang 2007. Prot. Sci. 16, 2716.

# Outline

- Conformational changes in proteins upon association
- Methods to model conformational changes
- Strategies to account for conformational changes
- **Explicit flexibility during docking**
- Attract docking approach

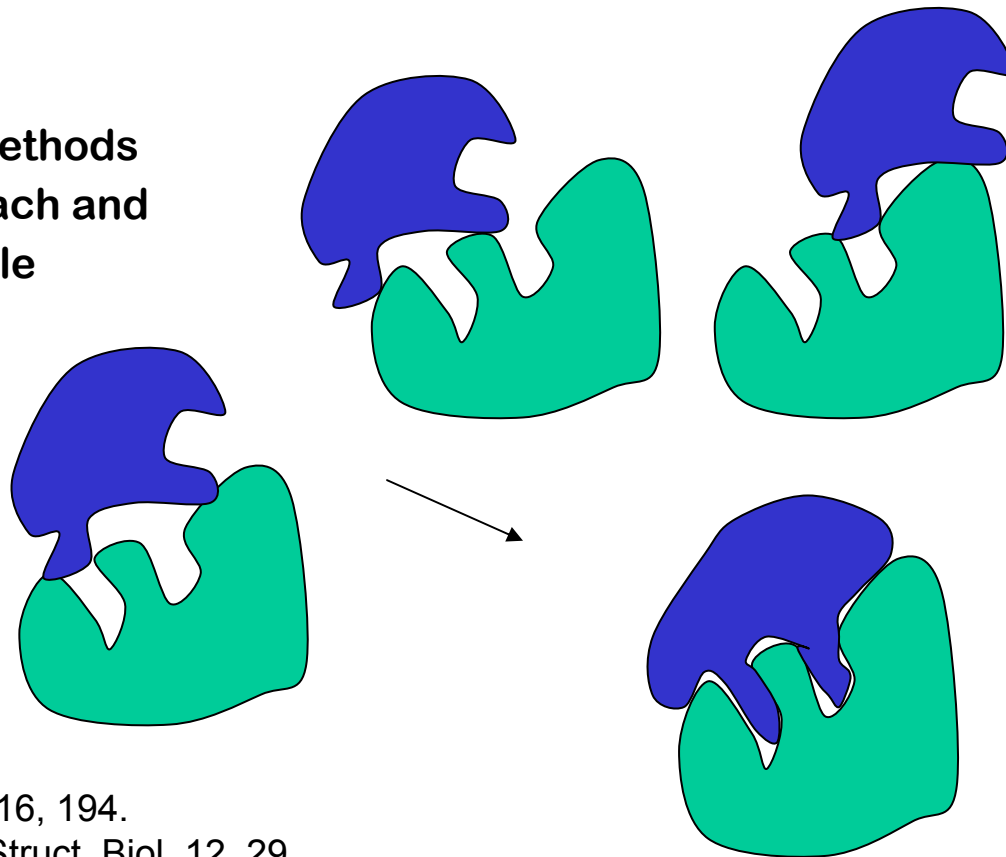
# Strategies to account for conformational changes during docking

## Two possibilities:

**Inclusion of conformational changes during entire docking search**

**Rigid docking followed by allowing conformational changes in a second step**

- The majority of docking methods follows the second approach and may include several flexible refinement steps.



Reviewed in:

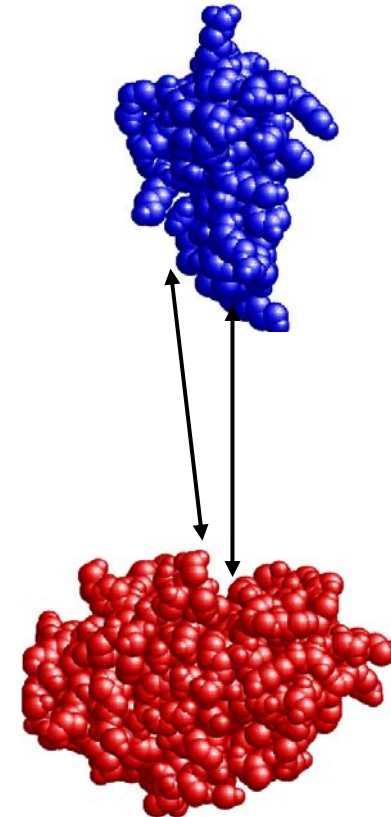
Andrusier et al. 2008. Proteins 73,271.

Bonvin, 2006. Curr. Opin. Struct. Biol. 16, 194.

Smith & Sternberg, 2002. Curr. Opin. Struct. Biol. 12, 29.

# Inclusion of conformational changes during docking

- **Cross-docking to members of an ensemble of structures (Krol et al., 2007)**
  - Can handle both changes in backbone as well as side chains
  - No modification to existing methods necessary
  - Linear increase of computational demand and also docking solutions
- **Docking using MD simulations including experimental restraints**
  - Implemented in HADDOCK (Dominguez et al., 2003)
  - Involves different MD phases (rigid, inclusion of dihedral degrees of freedom, Cartesian coordinates)
  - Very successful if sufficient experimental restraints are available



Krol et al. 2007. Proteins 69, 750.

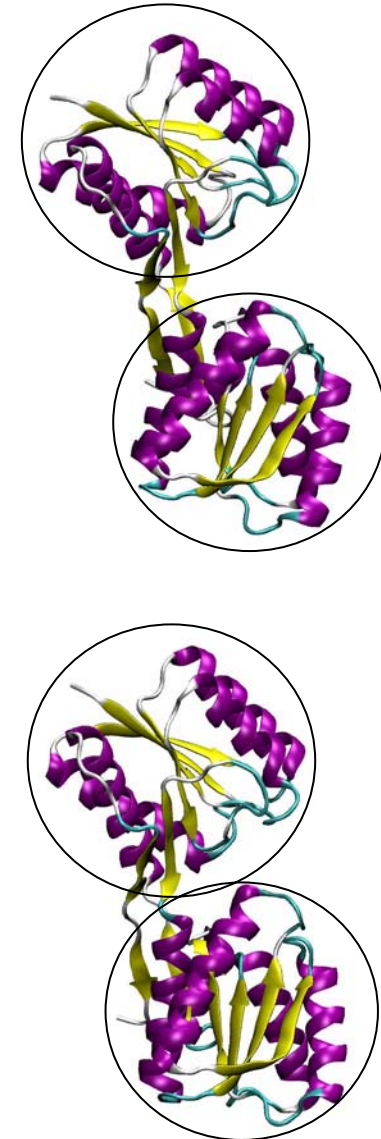
Dominguez et al. 2003. JACS 125, 1731.



# Inclusion of backbone conformational changes during docking

- **Identification of flexible hinge regions in proteins**
  - Several methods available to detect flexible backbone hinge regions:
    - ENM/GNM analysis (e.g. HingeProt; Emekli et al. 2008)
    - Comparison of experimental structures (DynDom; Hayward & Berendsen, 1998), HingeFind; Wriggers & Schulten, 1997; FlexProt; Emekli et al., 2008)
- **Separate docking of rigid domains after hinge detection (Schneidman-Duhovny et al. 2007)**
  - Retain only those solutions that allow appropriate domain connectivity

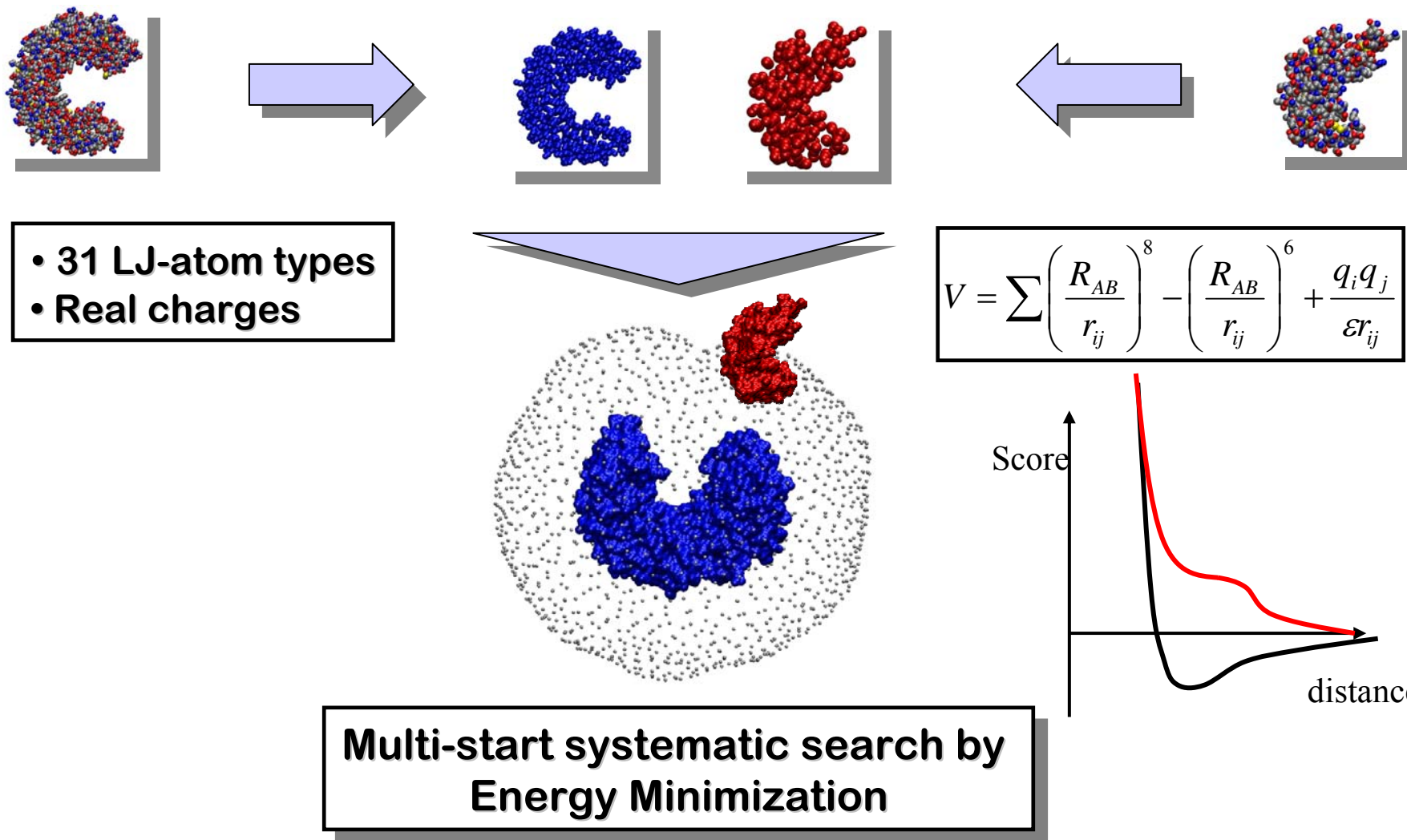
Hayward & Berendsen, 1998. Proteins 30, 144.  
Wriggers & Schulten, 1997. Proteins 29, 1.  
Shatsky et al. 2004. J.Comp.Biol. 11, 83.  
Emekli et al. 2008. Proteins 70, 1219.  
Schneidman-Duhovny et al. 2007. Proteins 69, 764.



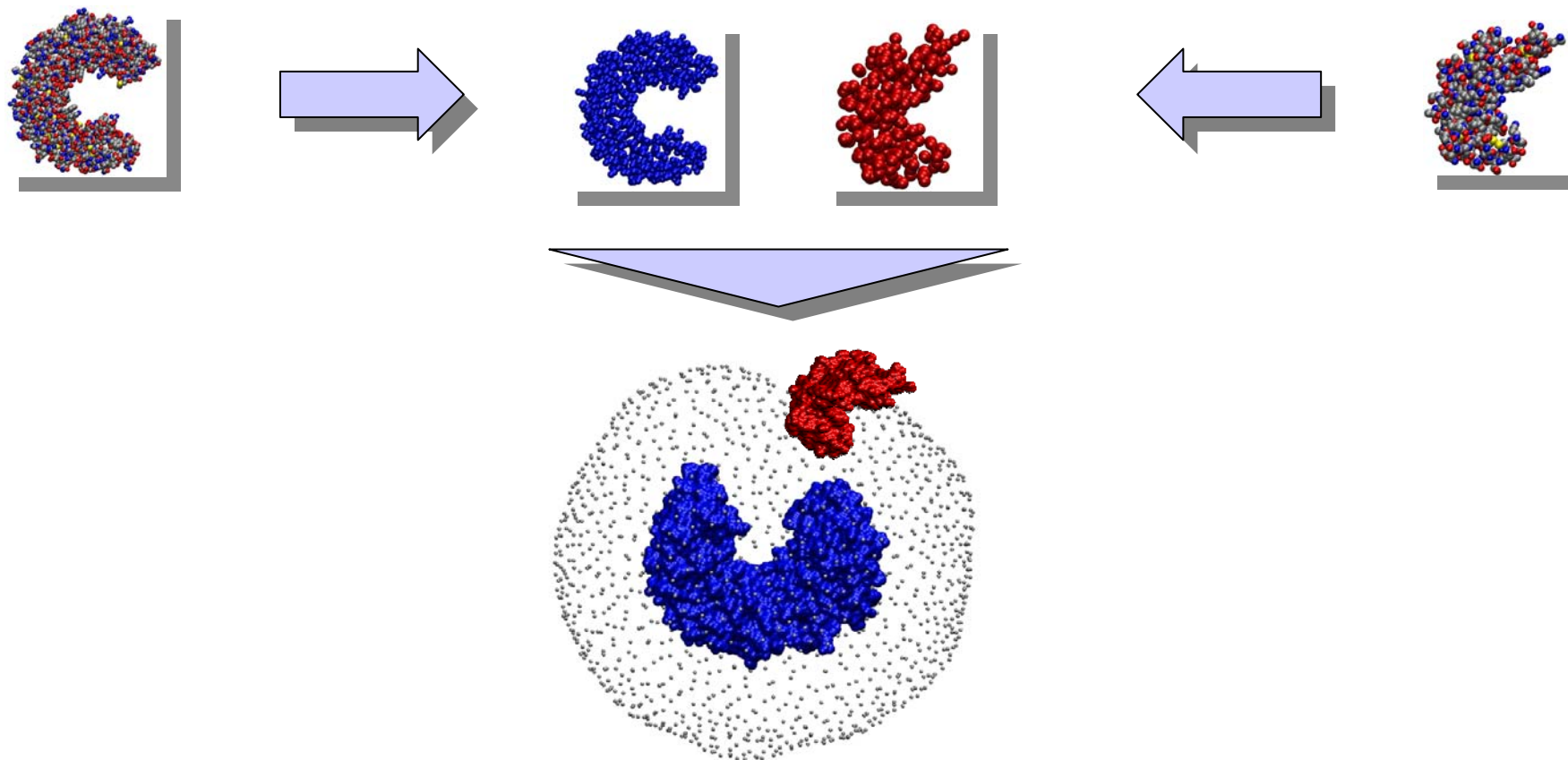
# Outline

- Conformational changes in proteins upon association
- Methods to model conformational changes
- Strategies to account for conformational changes
- Explicit flexibility during docking
- **Attract docking approach**

# The ATTRACT approach

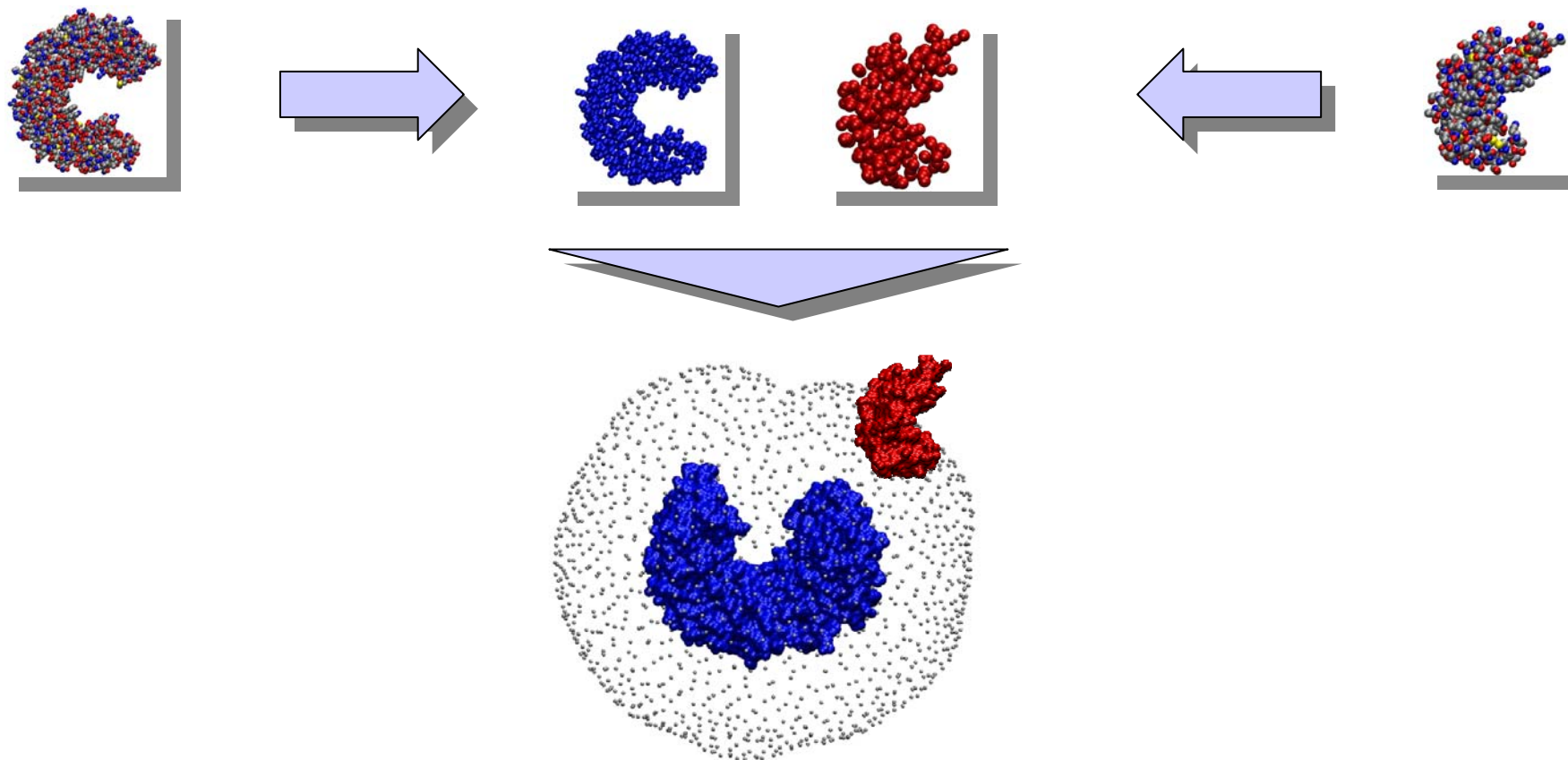


# The ATTRACT approach



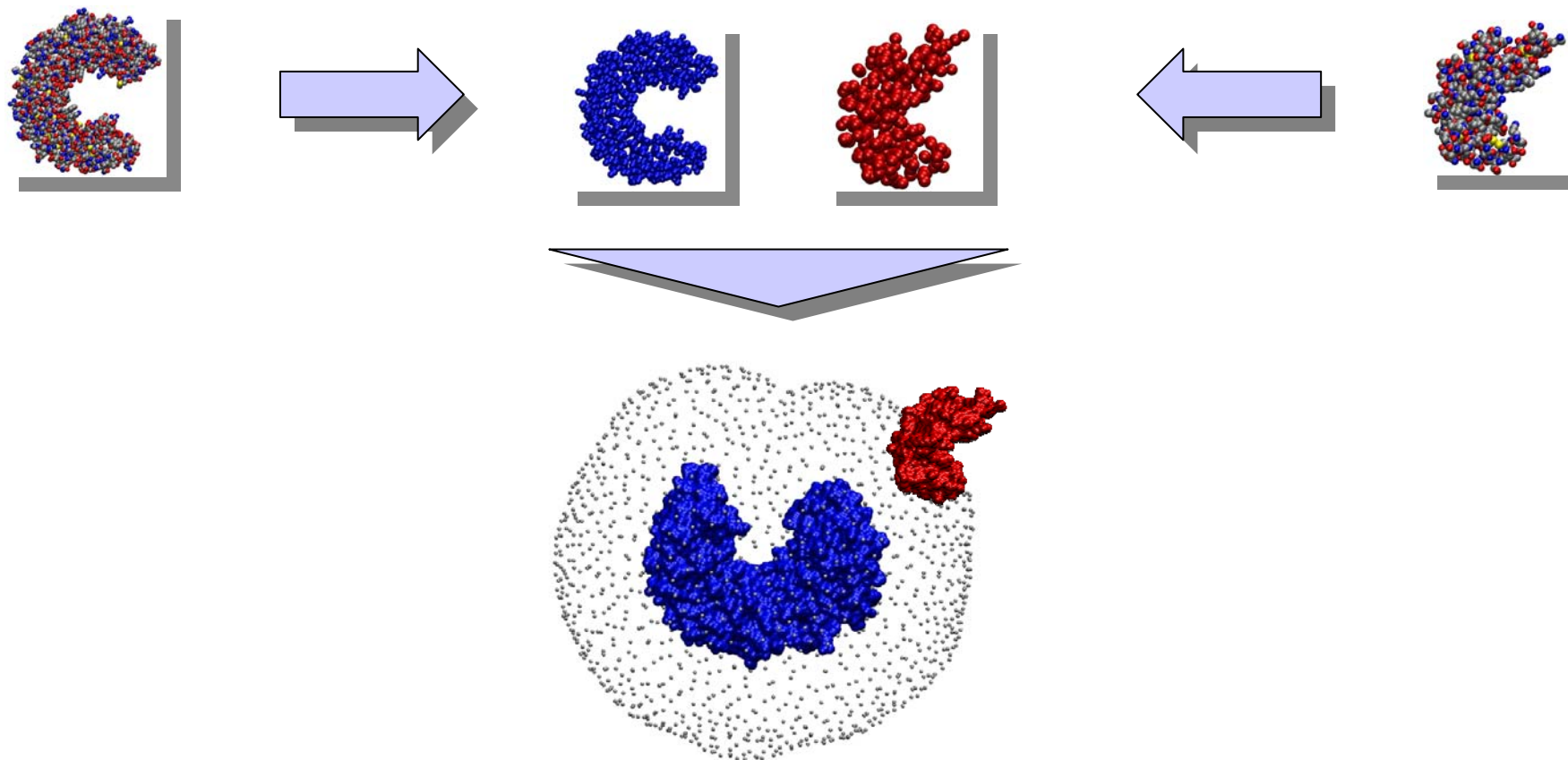
**Multi-start systematic search by  
Energy Minimization**

# The ATTRACT approach



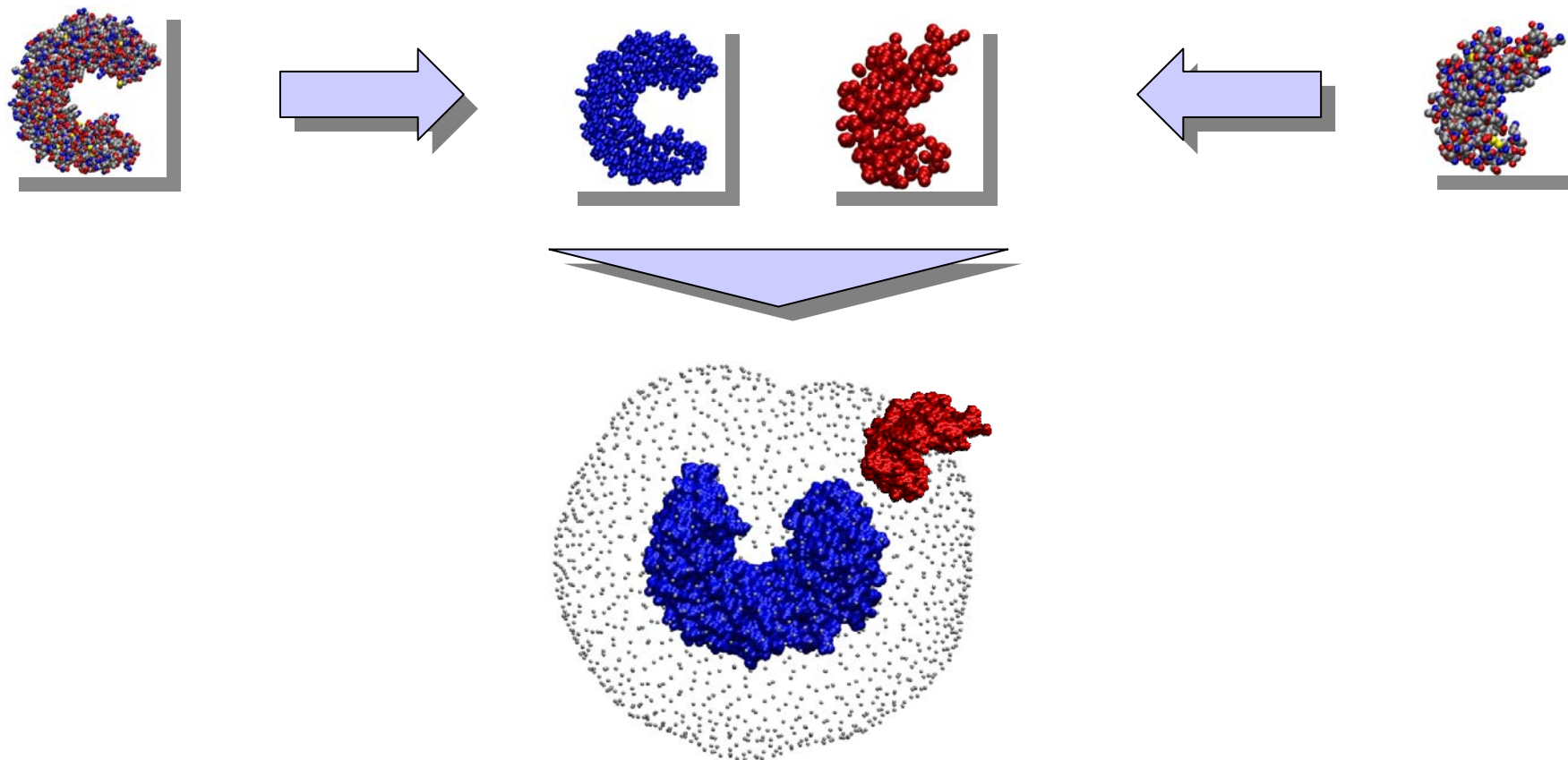
**Multi-start systematic search by  
Energy Minimization**

# The ATTRACT approach



**Multi-start systematic search by  
Energy Minimization**

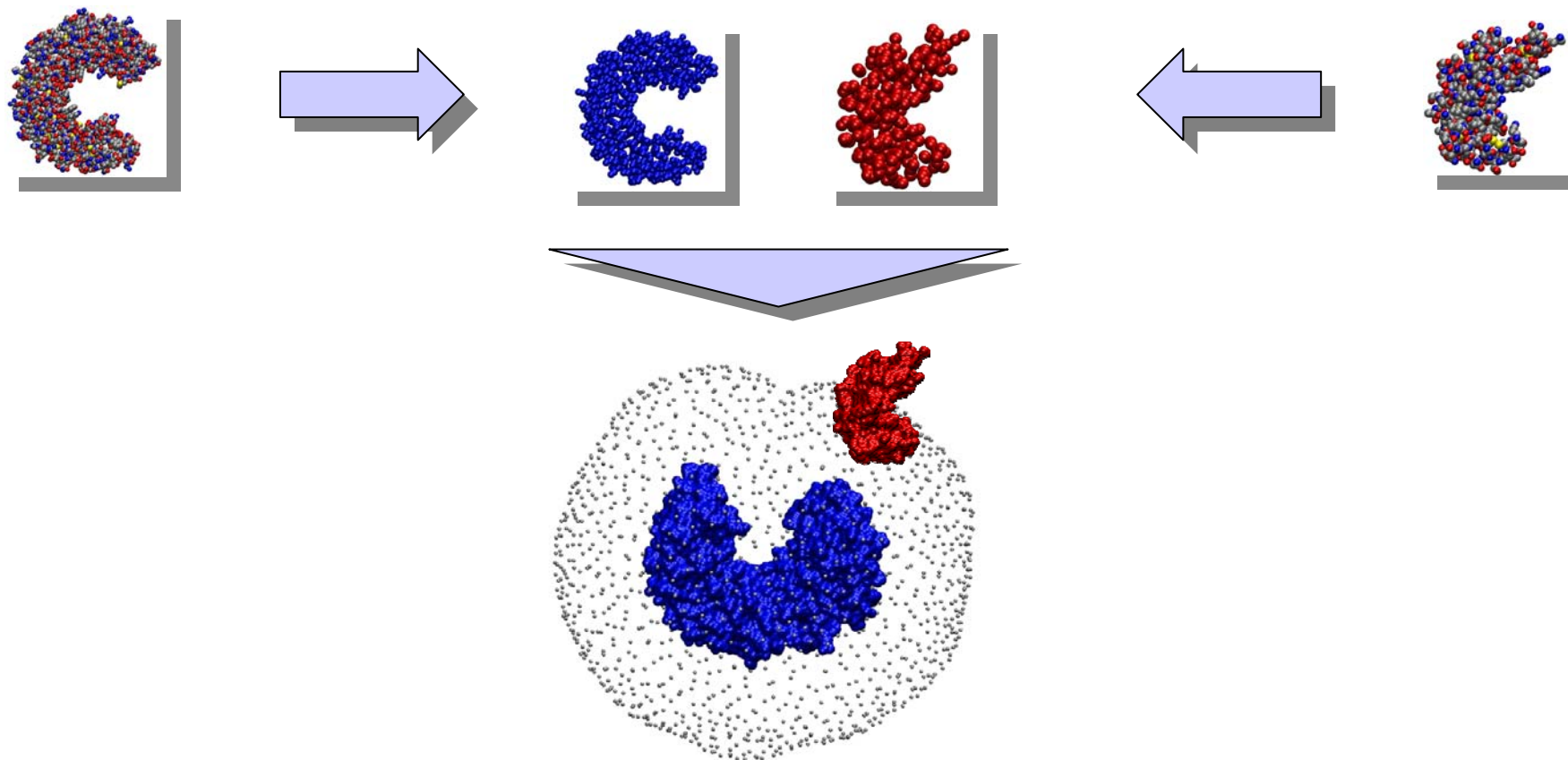
# The ATTRACT approach



**Multi-start systematic search by  
Energy Minimization**



# The ATTRACT approach



**Multi-start systematic search by  
Energy Minimization**

# Reduced vs. atomic resolution representation

## Pros

Fewer pairwise interactions compared to atomic resolution

Fewer local minima compared to atomic resolution

Limited implicit flexibility by soft interaction potentials

## Cons

Structures must be transferred back to atomic resolution

Scoring performance to be improved

# Systematic improvement of the scoring function

## Aim

Scoring optimization of **near-native** vs. **alternative docking minima** for a large set of training complexes

## Target function

Top ranking of native solution (large gap to incorrect solutions)

## Step 1

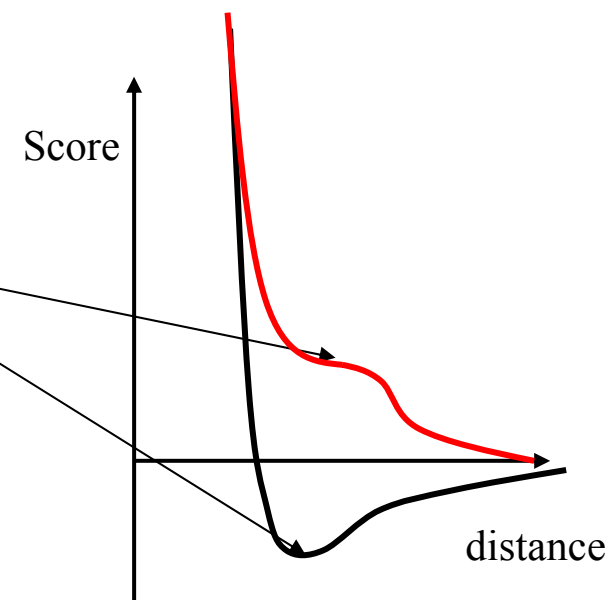
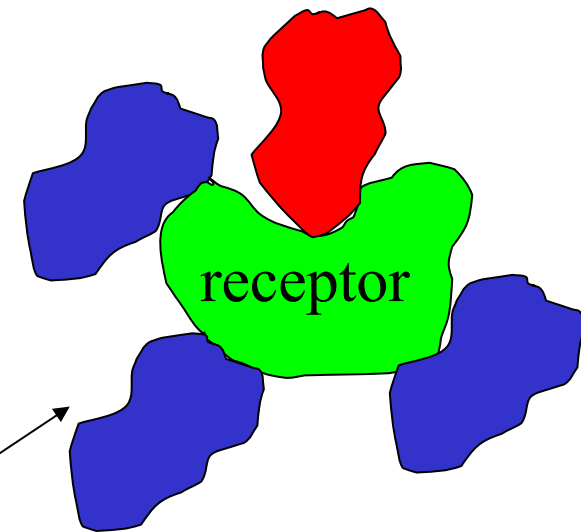
Generation of „high-ranked“ incorrect solutions

## Step 2

Optimization of pairwise interactions with respect to target function

## Step 3

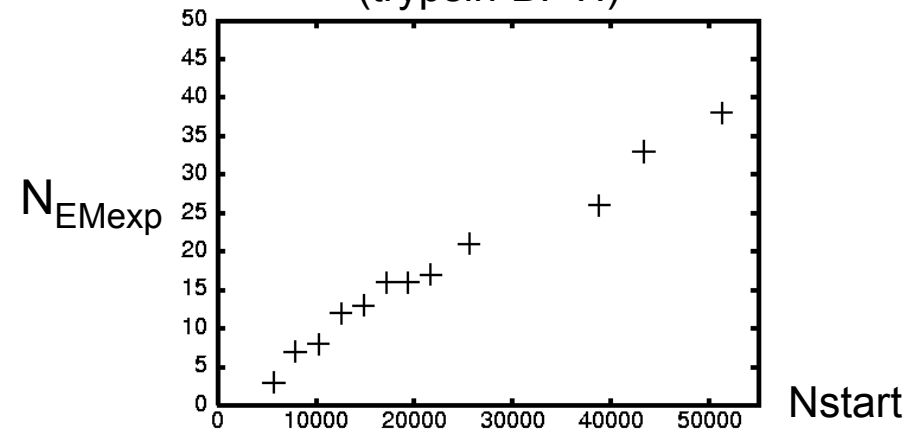
Test of scoring on separate set of test complexes



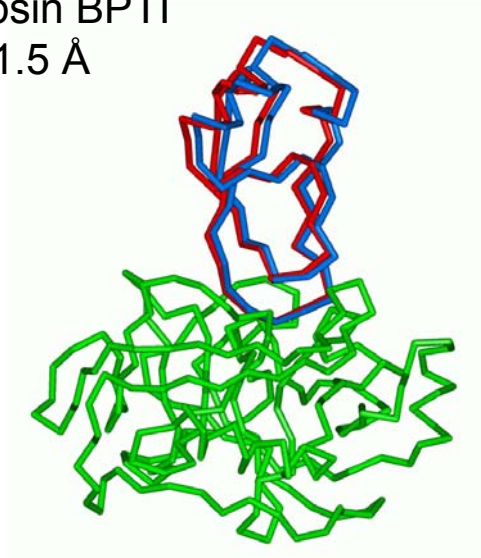
# Systematic docking of „bound“ structures

Complex	Rank <sub>EMexp</sub>	N <sub>EMexp</sub>
Trypsin-BPTI	2	25
Subtilisin-Inh.	2	22
Kallikrein-BPTI	1	31
Chym-OvoM	2	26
Chym-EglinC	3	17
U-Glycosidase	1	5
hGrowthh.Rec.	12	3
Barnase-Barstar	1	5

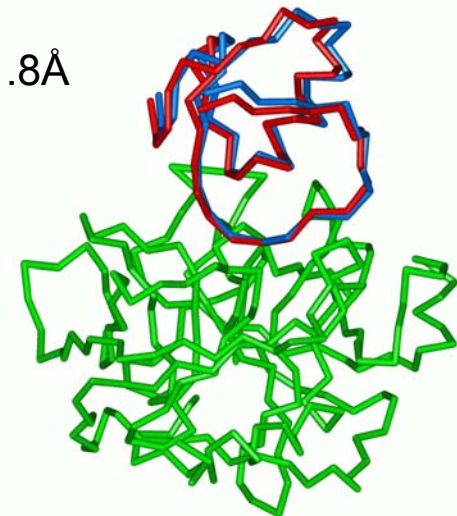
Number of EMexp complexes vs. Nstart  
(trypsin-BPTI)



Docking of trypsin BPTI  
Rmsd(ligand):1.5 Å



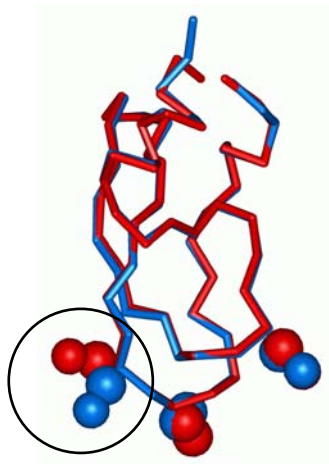
Docking of chymotrypsin-  
ovomucoid, Rmsd(ligand):1.8Å



# Docking with „unbound“ protein structures

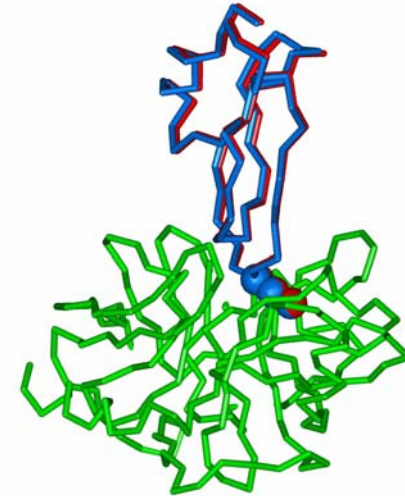
Docking with experimental structures of isolated protein partners:

Complex	Rank	Rmsd(Å)
Trypsin-BPTI	>120	7.1/2.9
	>760	3.5/1.4
Kallikrein-BPTI	>160	3.4/1.5
Chymo-OvoM	9	2.8/1.2
Chymo-BPTI	>150	3.4/1.7
U-Glycosidase	1	3.2/1.3
SubtiN/ChymInh	>1100	3.8/1.6

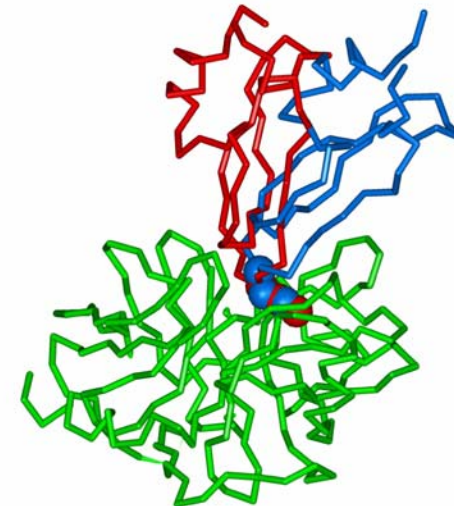


Backbone of BPTI in „bound“ (red) and „unbound“ (blue) conformation (+ side chains 15, 17, 39)

EM of „bound“ BPTI (blue)



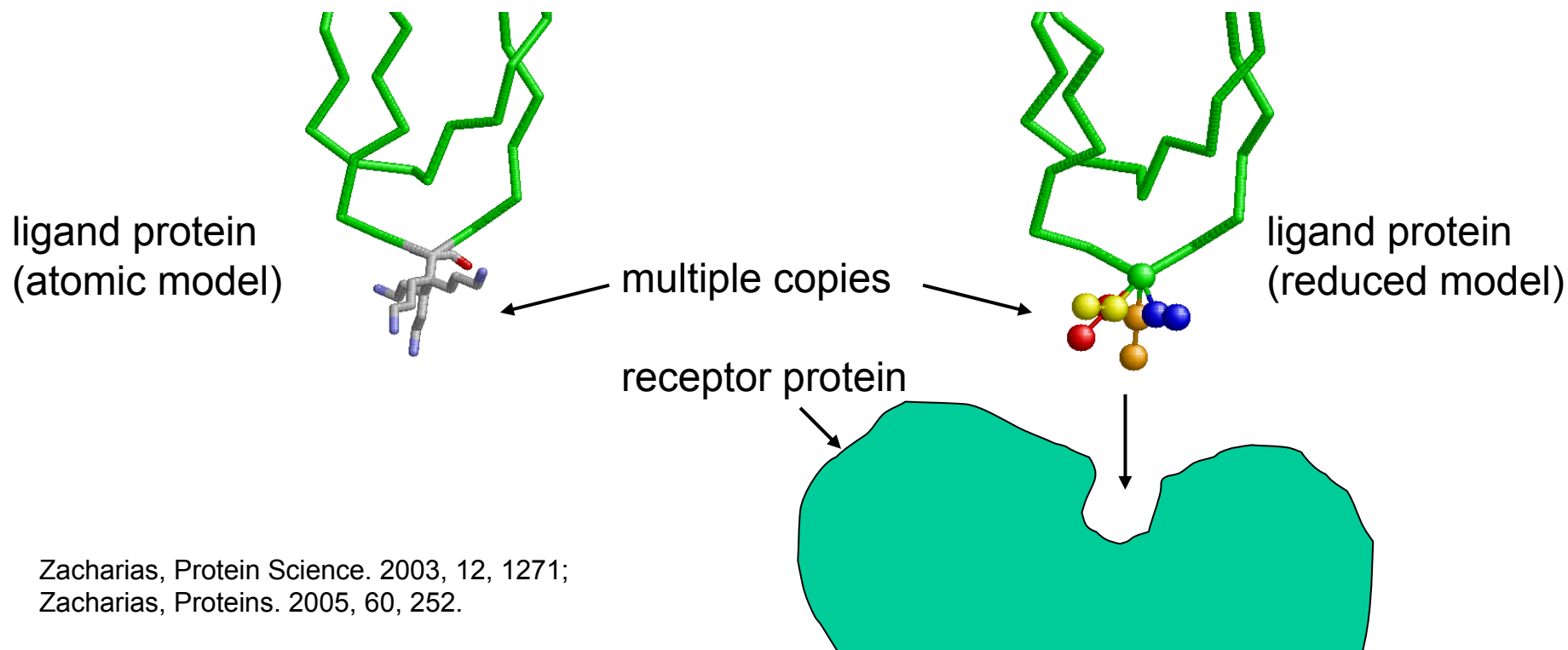
EM of „unbound“ BPTI (blue)



# Docking with multiple side chain copies

- Surface side chains are represented by several sterically allowed rotamer copies.
- Selection of most favorable copies during docking using a meanfield or switching approach.

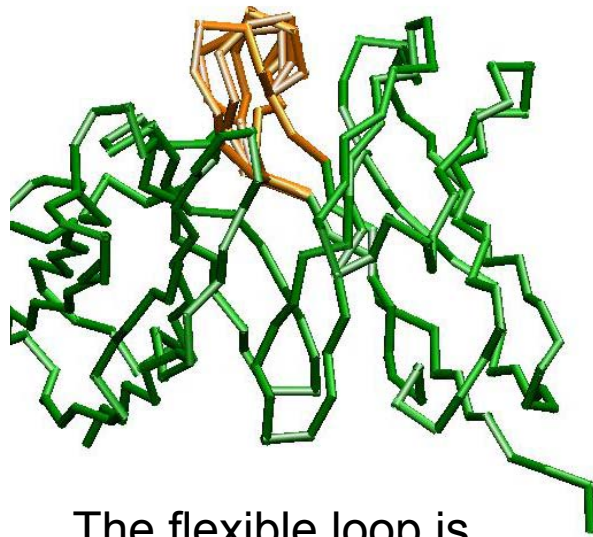
Complex	- copies		+ copies	
	Rank	Lrmsd <sub>Å</sub>	Rank	Lrmsd <sub>Å</sub>
Trypsin-BPTI	>120	7.1	11	1.4
Kallikrein-BPTI	>160	3.4	31	2.4
Chymo-BPTI	>150	3.4	29	3.2
SubtiN/Chy-Inh	>1100	3.8	59	3.0





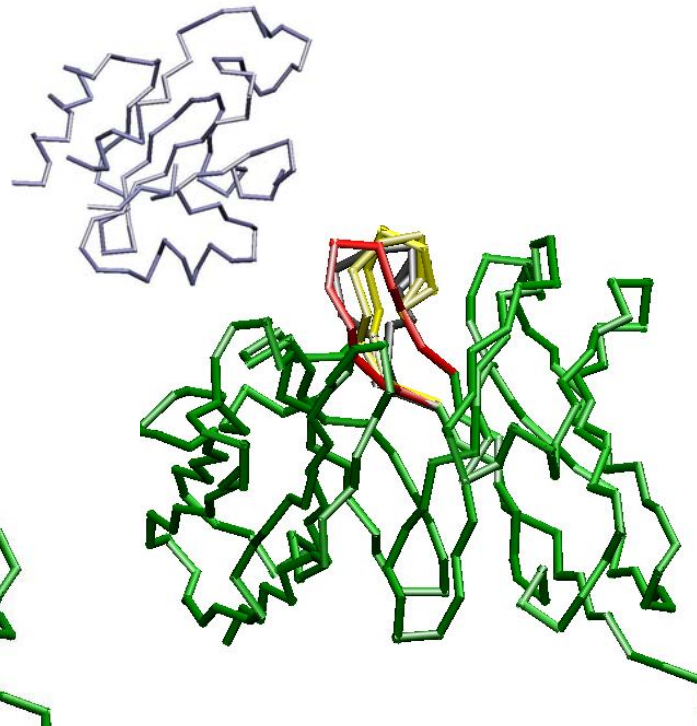
# Loop flexibility with multicopy mean-field approach

Before docking



The flexible loop is represented by an ensemble of copies (multi-copy approach)

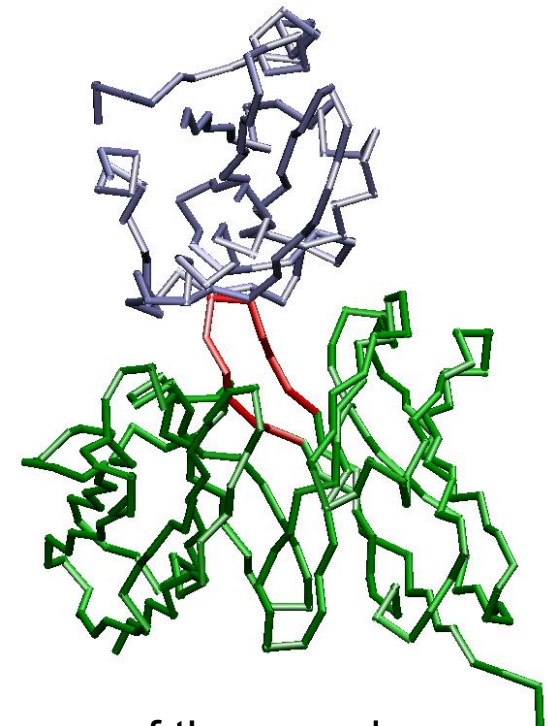
For each starting position



A weight is attributed to each copy  $k$  as a function of interaction energy with the ligand:

$$W_k = \exp(-E_k/RT)/(\sum W)$$

Minimization



Energy of the complex:

$$E = E_{\text{simple}} + \sum (W_k E_k)$$

Copies with highest weight drive the minimization



# Loop flexibility with a multicopy mean-field approach

- Four test cases in which significant changes in loop regions occur upon complex formation
  1. 1OAZ/1OAQ (IGE-Thioredoxin complex; loop: 100-107)
  2. 1A0O/1CHN (CheA-CheY complex; loop: 89-101)
  3. 1CGI/1CGH (Chymotryp.-Inh. complex; loop: 143-155)
  4. 1BTH/2HNT (Thrombin-BPTI; loop: 48-55; 77-86)

Case	unbound receptor		+10 loop copies(b)		+10 loop copies (u)	
	score	Lrmsd	score	Lrmsd	score	Lrmsd
1OAZ/1OAQ	-12.4(44)	4.5Å	-19.0(2)	0.8Å	-15.3(11)	5.0Å
1A0O/1CHN	-17.7(7)	3.6Å	-20.6(1)	2.6Å	-16.9(3)	2.9Å
1CGI/1CGH	-18.3(44)	1.2Å	-23.9(1)	1.7Å	-20.1(1)	1.1Å
1BTH/2HNT	-4.1(>6000)	5.5Å	-21.9(1)	2.4Å	-19.3(1)	4.5Å

# Docking challenge CAPRI

- CAPRI**

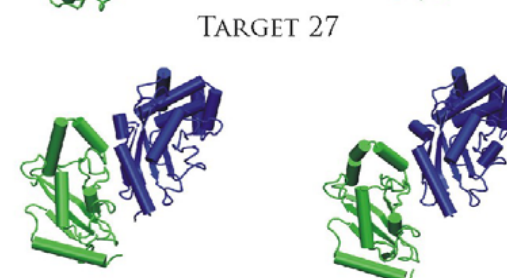
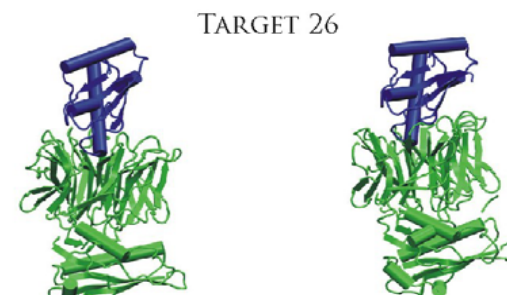
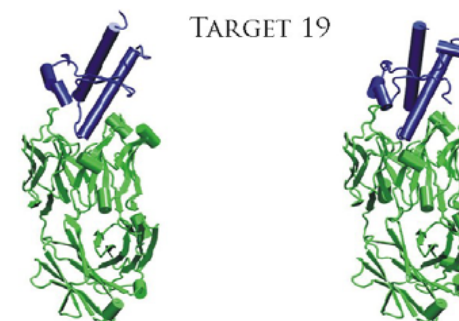
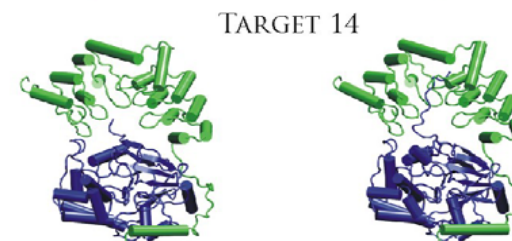
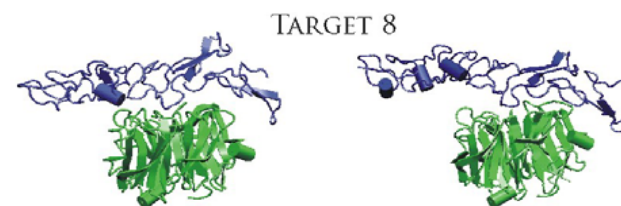
(Critical Assessment of Predicted Interactions)

<http://capri.ebi.ac.uk/>

– Binding geometry predictions before experimental complex structures are available:

Target	% native contacts	Interface-Rmsd(Å)
8	40	0.9
9	18	9.5
14	60	0.6
18	0	22.5
19	65	1.8
20	26	9.8
21	34	5.1
25	21	4.4
26	45	2.1
27	39	3.6
28	7	7.2
29	2	11.5
30	45	2.5 (best prediction)

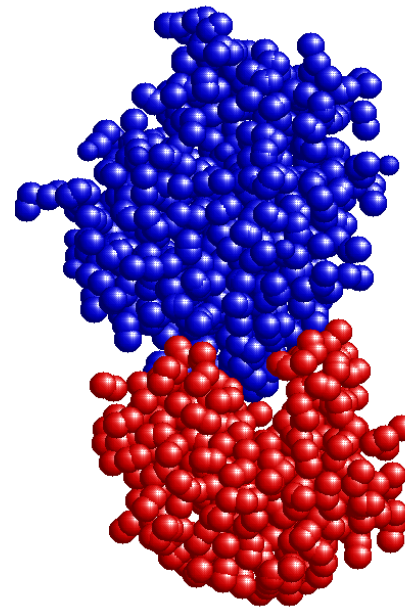
PREDICTED STRUCTURE      EXPERIMENTAL STRUCTURE



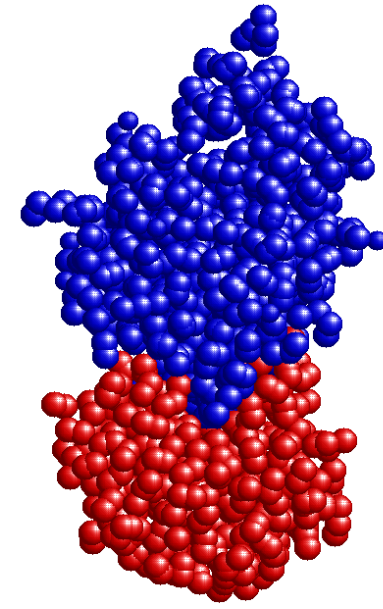
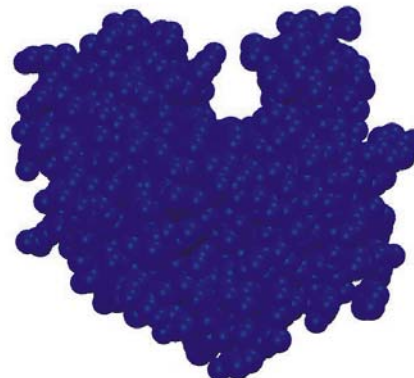
# Global changes can affect docking performance

- A.niger xylanase-taxi-inhibitor complex
- Rigid docking:
  - Solution closest to native structure deviates by  $> 6 \text{ \AA}$ .
- Xylanase undergoes global opening motion upon complex formation

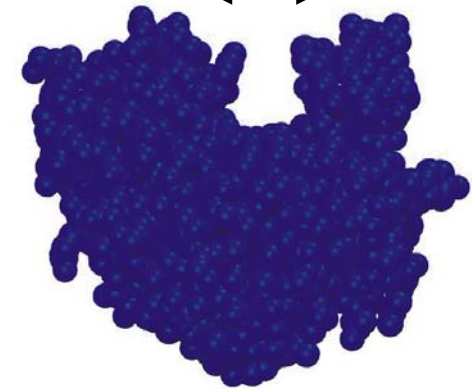
closest prediction / exp. complex



Xylanase:  
unbound



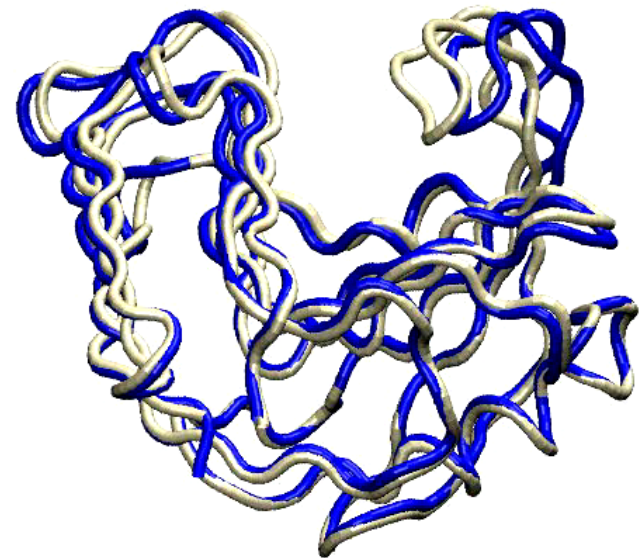
Xylanase:  
bound



# Accounting for global deformations during protein-protein docking

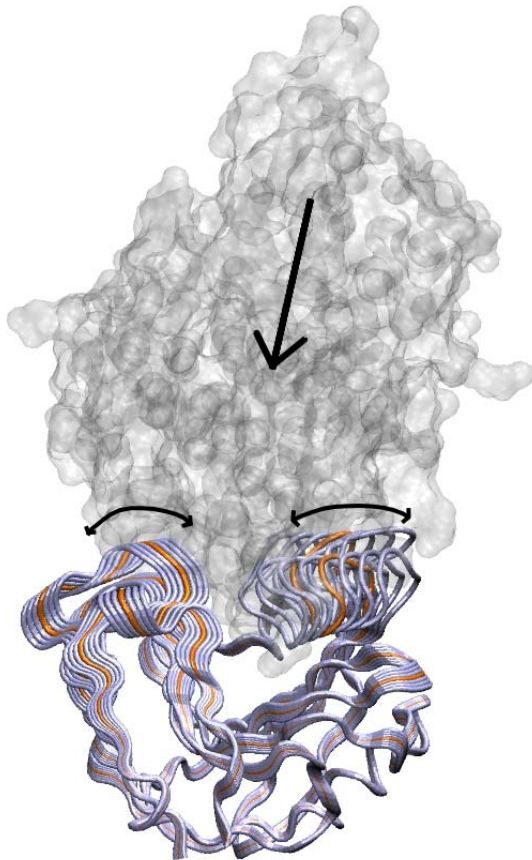
- Protein-protein association can involve global deformations.
- The deformation directions may overlap with „soft collective degrees of freedom“ of the protein partners.
- Approach:
  - Pre-calculation of soft modes of protein partners
    - Principal component analysis of a MD simulation
    - Approximate normal mode calculation
  - Structure relaxation in soft degrees of freedom during docking by energy minimization

Softest ENM-mode of Xylanase overlaps very well with observed conformational change



# A slightly modified optimization problem

6 rigid body degrees of freedom  
+ one additional for every soft mode  $m$



$$V = V_{\text{intermolecular}} + V_{\text{intramolecular}}(m)$$

$$V_{\text{intramolecular}}(m) = \sum \text{eig}_m^2 (R_m^0 - R_m)^4$$

$m$ : number of soft modes

$\text{eig}_m$ : corresponding eigenvalue of mode  $m$

$R_m^0$ : equilibrium coordinate set of mode  $m$

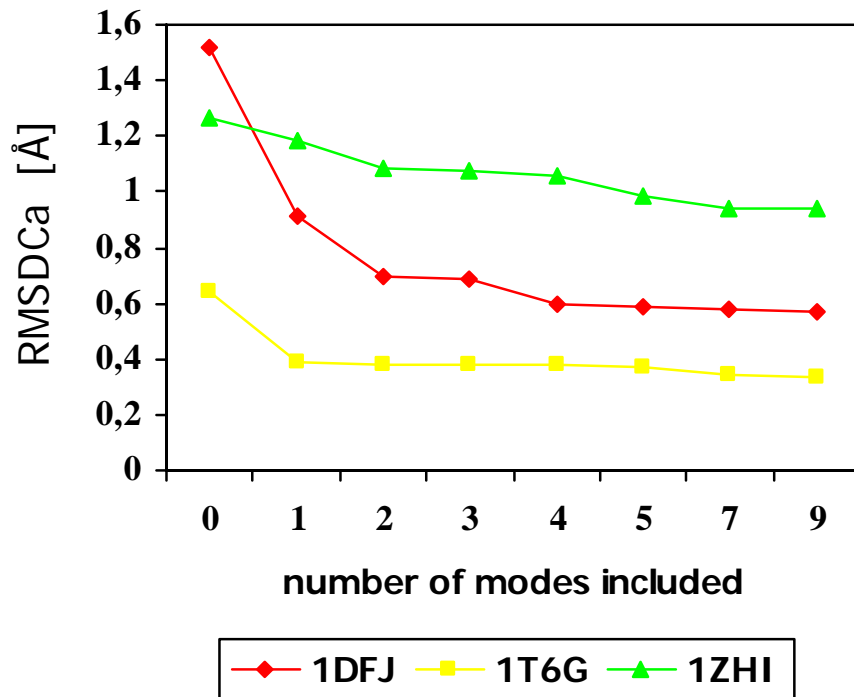
$R_m$ : coordinate set after deflection of mode  $m$

$R_m^0 - R_m$ : amplitude of mode  $m$

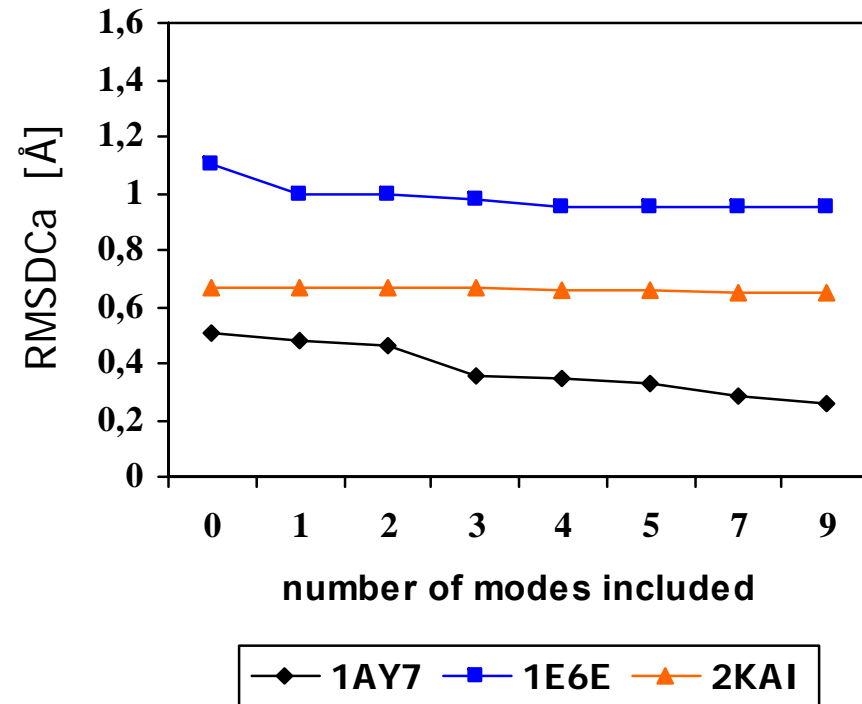
Any deformation in the bonded geometry  
can be removed by a Shake-algorithm

# Overlap between normal mode directions and experimentally observed conformational difference

Maximum achievable refinement for systems showing overlap



Maximum achievable refinement for systems showing little or no overlap





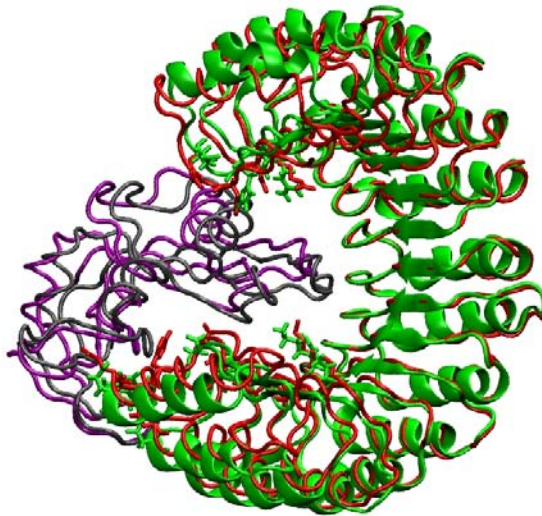
# Inclusion of global flexibility through normal mode minimization

- First test:
  - Normal modes only for the flexible partner (receptor)
  - Apo-receptor (with near native side chains)
  - Bound ligand protein
- EM of ~60000 start configurations including NM-minimization at every stage (no clustering)

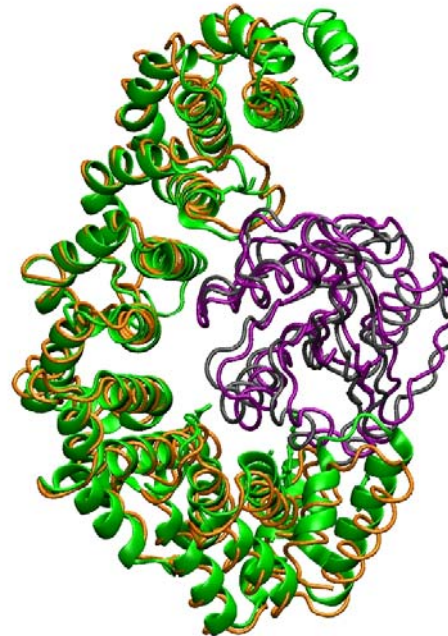
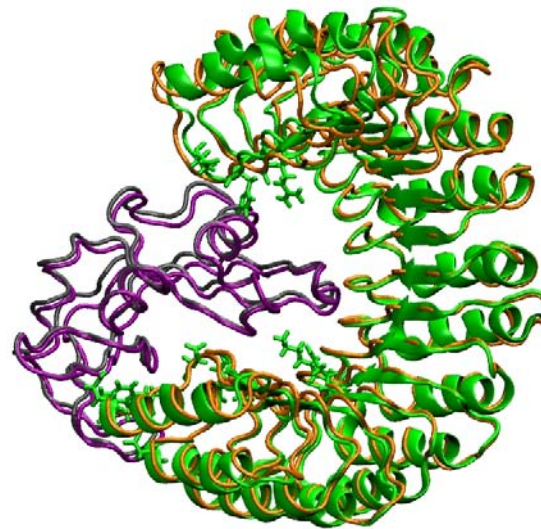
Complex	Lrmsd (Å)			Rank		
	Rigid	1-mode	5-modes	Rigid	1-mode	5-modes
<b>1DFJ</b>	<b>3.3(1.5)</b>	<b>1.5(0.9)</b>	<b>1.1(1.4)</b>	<b>3</b>	<b>1</b>	<b>1</b>
<b>1T6G</b>	<b>9.4(0.65)</b>	<b>1.6(0.4)</b>	<b>1.5(0.6)</b>	<b>25</b>	<b>1</b>	<b>1</b>
<b>1ZHI</b>	<b>2.0(1.35)</b>	<b>2.0(1.1)</b>	<b>0.9(1.3)</b>	<b>5</b>	<b>3</b>	<b>6</b>
<b>1IBR</b>	<b>7.5(2.9)</b>	<b>6.0(2.7)</b>	<b>2.2(2.0)</b>	<b>12</b>	<b>1</b>	<b>1</b>
<b>1AY7</b>	<b>0.8(0.5)</b>	<b>0.9(0.5)</b>	<b>0.9(0.6)</b>	<b>21</b>	<b>21</b>	<b>26</b>
<b>1E6E</b>	<b>2.0(1.0)</b>	<b>2.2(0.9)</b>	<b>2.0(1.1)</b>	<b>9</b>	<b>13</b>	<b>13</b>
<b>2KAI</b>	<b>3.9(0.9)</b>	<b>3.9(0.9)</b>	<b>4.0(0.9)</b>	<b>35</b>	<b>33</b>	<b>38</b>

# RNAse / Inhibitor (1DFJ) and Importin/RanGTPase

rigid



flexible (5 modes)

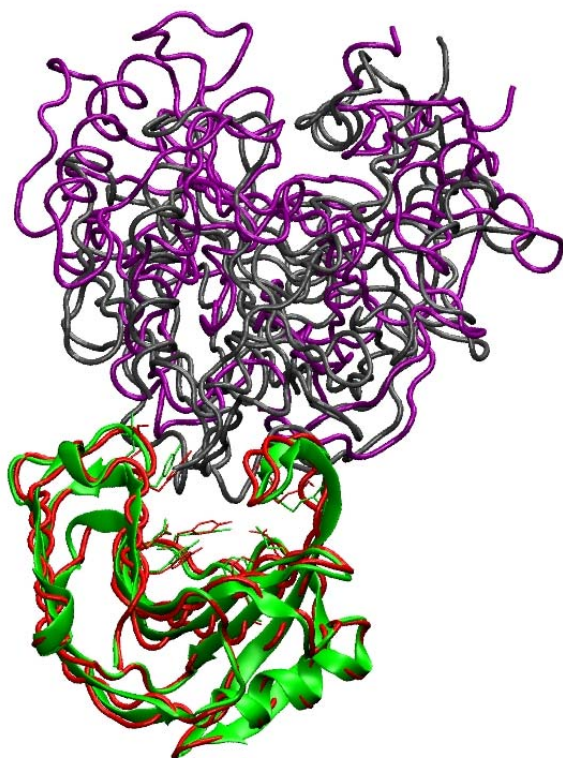


Apo rec., holo rec., rec. after flexible docking, exp. ligand position, docked ligand

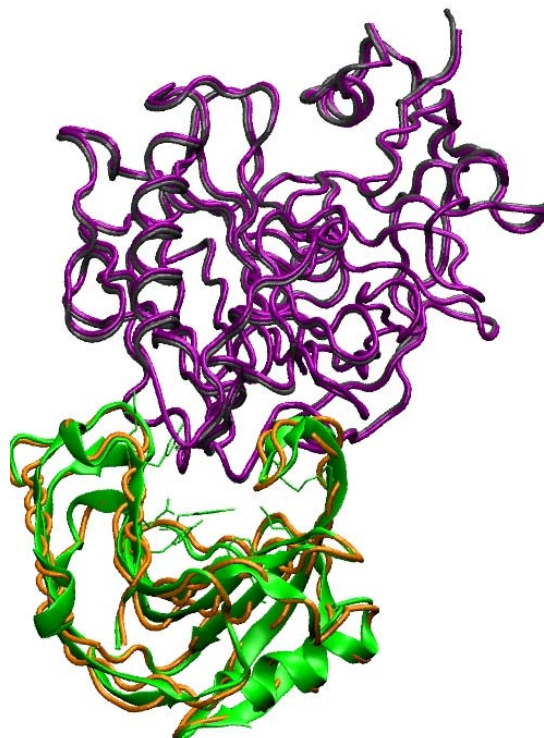


# Most accurate solution for Xylanase / TAXI Inhibitor (1T6G) system

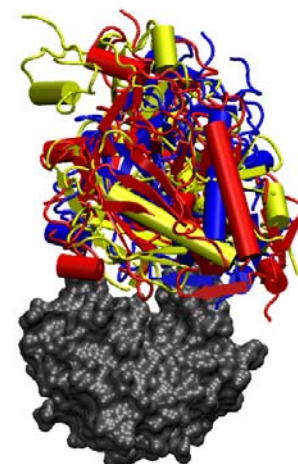
rigid



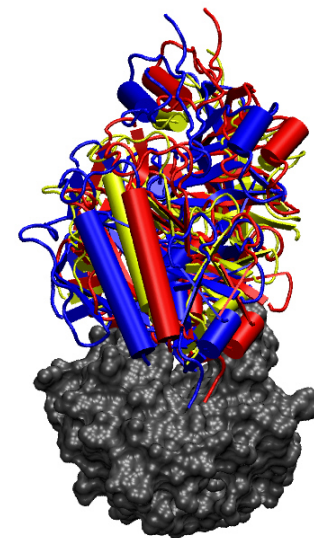
flexible (5 modes)



Enrichment effect:  
3 solutions closest to native  
complex after  
rigid docking



flexible docking



Apo rec., holo rec., rec. after flexible docking,  
exp. ligand position, docked ligand

# Normal mode relaxation on both partners

- Second test:
  - Normal modes minimization for both partners (receptor and ligand)
  - Unbound receptor and ligand (may also have incorrect side chain conformations and other backbone problems)
- EM of ~60000 start configurations including NM-minimization at every stage (no clustering)

Complex	Lrmsd (Å)			Rank		
	Rigid	Rflex	Allflex	Rigid	Rflex	Allflex
<b>1DFJ</b>	<b>4.8</b>	<b>4.8</b>	<b>4.4</b>	<b>2</b>	<b>2</b>	<b>13</b>
<b>1T6G</b>	<b>9.7</b>	<b>5.0</b>	<b>4.6</b>	<b>14</b>	<b>1</b>	<b>2</b>
<b>1ZHI</b>	<b>6.3</b>	<b>7.7</b>	<b>7.2</b>	<b>36</b>	<b>4</b>	<b>11</b>
<b>1IBR</b>	<b>9.5</b>	<b>9.8</b>	<b>10.5</b>	<b>42</b>	<b>19</b>	<b>24</b>
<b>1AY7</b>	<b>4.3</b>	<b>3.8</b>	<b>3.9</b>	<b>111</b>	<b>105</b>	<b>164</b>
<b>1E6E</b>	<b>2.0</b>	<b>2.7</b>	<b>3.2</b>	<b>19</b>	<b>17</b>	<b>16</b>
<b>2KAI</b>	<b>8.7</b>	<b>7.7</b>	<b>5.5</b>	<b>64</b>	<b>114</b>	<b>97</b>

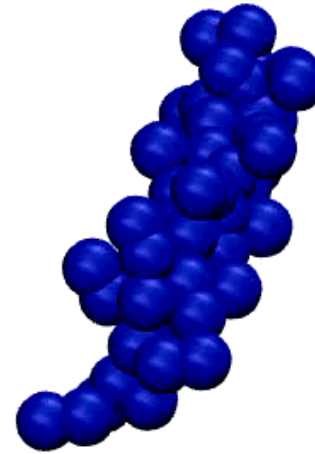
# **Impact of Normal mode minimization during docking**

- **Significant improvement of ligand placement when eliminating the side-chain problem**
- **Modest improvement of receptor RMSD**
- **Improvement of ranking in case of unbound partners**
- **Increase in computational burden very modest (factor ~3)**
- **Lacking overlap between ENMs and conformational change results in no or only small deterioration**
- **Enrichment of solution space with near-native solutions: detection of binding funnels is facilitated**

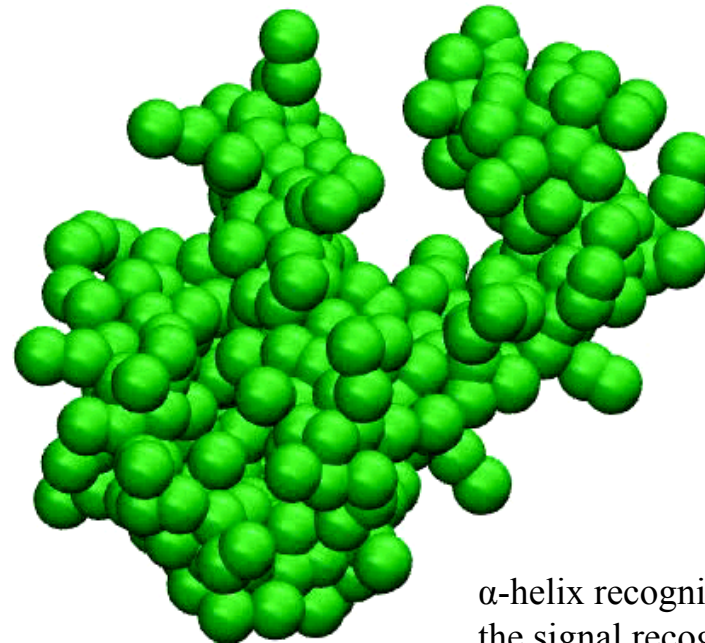
# Future work on protein-protein docking

- Study of multiple protein interactions and protein assemblies
- Extension to DNA-protein and RNA-protein docking
- Inclusion of low-resolution structural data (electron-density; CryoEM)
- Combination with Brownian dynamics:
  - Influence of flexibility on association
  - Analysis of intermediate docked states

Simulation of the complex formation of two proteins



Approaching  $\alpha$ -helical peptide



$\alpha$ -helix recognition domain of the signal recognition particle

# Conclusions

- Accounting (efficiently!) for conformational changes during docking remains a challenge
- Longterm goal: docking model structures
  - Docking procedure must tolerate or correct errors in the model
  - Better protein models
- Characterization of transient interactions and encounter complexes

## Reviews:

Andrusier , Mashiac, Nussinov & Wolfson 2008. Principles of flexible protein-protein docking. Proteins 73,271.

May, Sieker & Zacharias (2008) How to efficiently include receptor flexibility during computational docking. Curr. Comput. Aided Drug Des., 4, 143-153.

Gray, 2006. High-resolution protein-protein docking. Curr. Opin. Struct. Biol. 16, 1.

Bonvin, 2006. Flexible protein-protein docking. Curr. Opin. Struct. Biol. 16, 194.

May & Zacharias, 2005. Accounting for global protein deformability during protein-protein and protein-ligand docking. BBA 1754, 225.

Smith & Sternberg, 2002. Prediction of protein-protein interactions by docking methods. Curr. Opin. Struct. Biol. 12, 29.

# Acknowledgements



## Group members:

Ranjit Bahadur

Jeremy Curuksu

Sebastien Fiorucci

Ragav Kannan

Zdravko Kochovski

Andreas May

Florian Sieker

## Collaboration on protein docking

Chantal Prevost (IBPC, Paris)

Adrien Saladin (University 7, Paris)

-> designed a Python/C++ version of  
Attract (-> docking with arbitrary  
number of partners)

## Current Funding:

DFG (Protein-Protein-Docking)

DFG (RNA-Protein-Interaktion)

VW-Stiftung (Modellierung kompl. Sys.)

EU-STREP (Bacterial. Vaccine-Epitopes)

PNNL-Comput. Grand Challenge:

gc2002:DNA-damage;

gc2006:petascale-Modeling