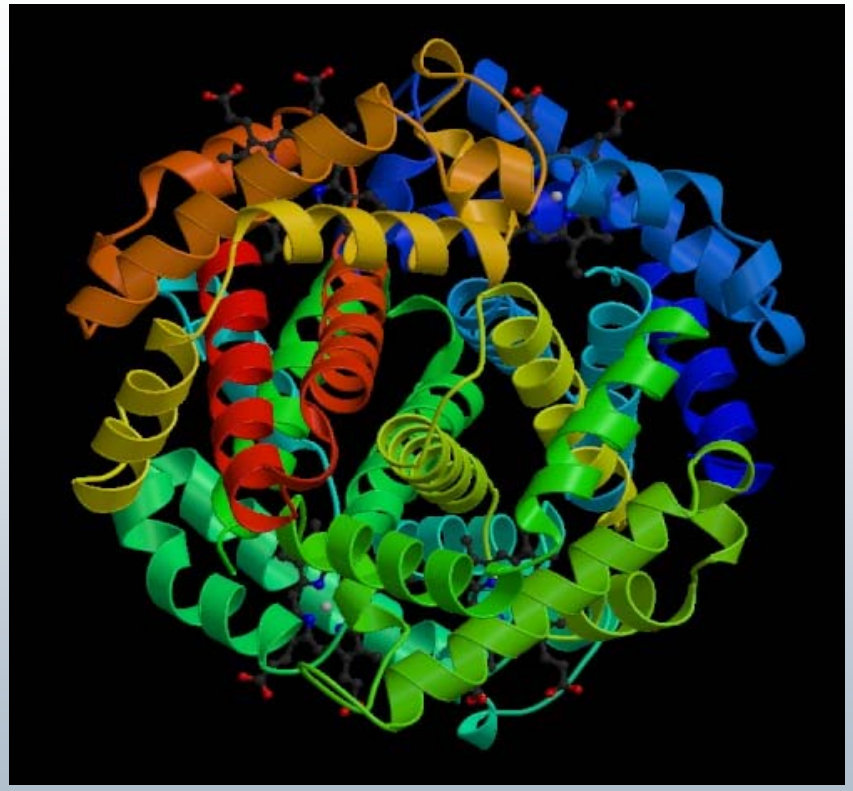
Biomolekulare Strukturvorhersage mit stochastischen Optimierungsverfahren: von der Sequenz zum Medikament



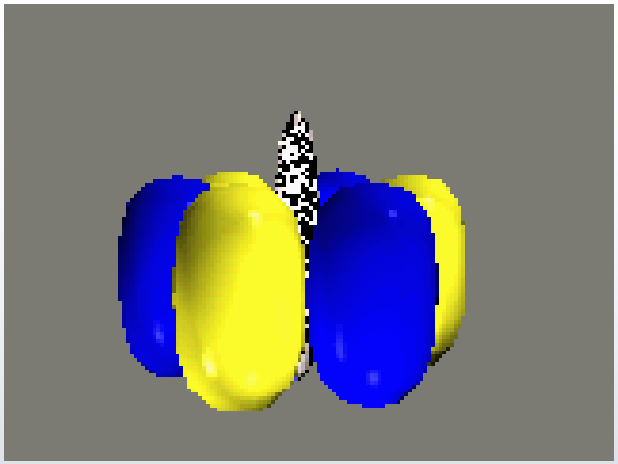
Wolfgang Wenzel
Forschungszentrum Karlsruhe
Institut für Nanotechnologie

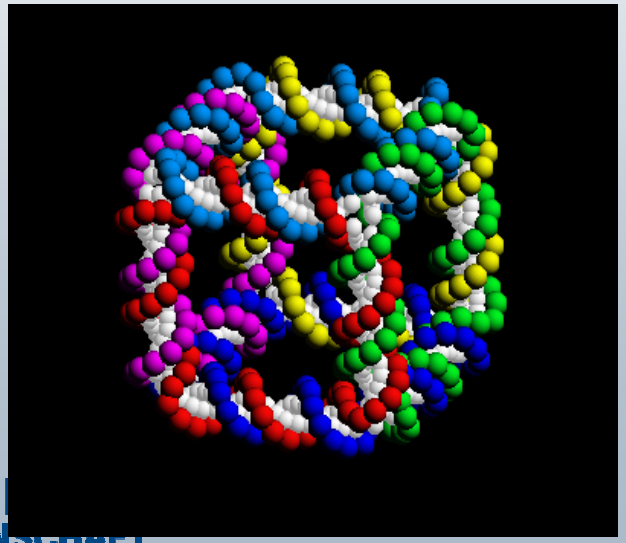
email: wenzel@int.fzk.de

http://www.fzk.de/biostruct

HELMHOLTZ GEMEINSCHAFT

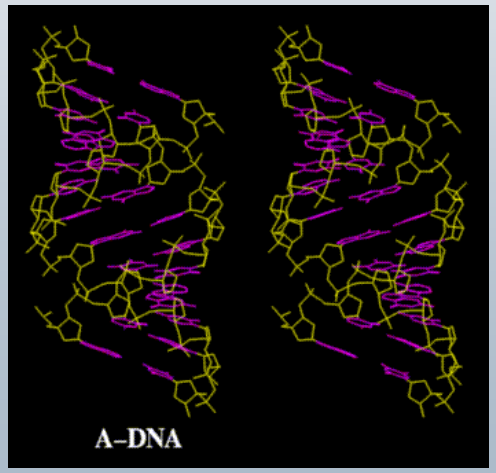
Nanobiotechnology





Biological macromolecules play an increasing role as functional units in nano-devices

Urgent need to understand and predict their structural properties and stability.



Computational Nanophysics

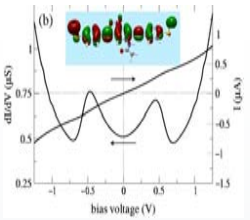
DFT, MRCI, DMRG

Landauer Theory, Rate Equations

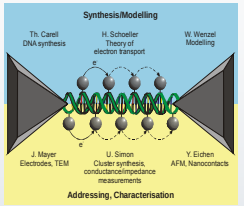
Kinetic Theory, Molec. Dynamics

Stochastic Optimization

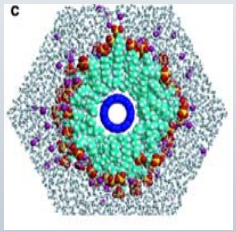
Single Molecule Transport
Hettler, Weber, Schoeller (Aachen), Cuevas (KA)
DFG



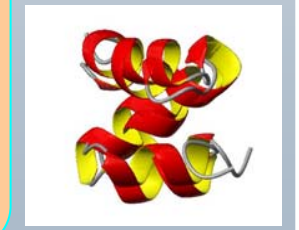
DNA Based Molecular Wires
Carell(München), Simon (Aachen)
VW Stiftung



Materials Modelling of Nanotubes and Nanofilms
Krupke, Balaban, Wahlheim, Carrell(München)
VW Stiftung, KIST



Protein Folding & Drug Development
Scheraga (Cornell), Lee (KIST), 4SC AG (München)
DFG, KIST

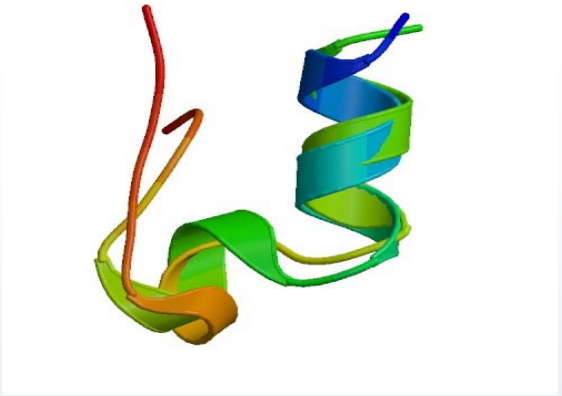


HELMHOLTZ GEMEINSCHAFT

Biomolecular Structure Prediction

Protein Folding: from sequence to structure

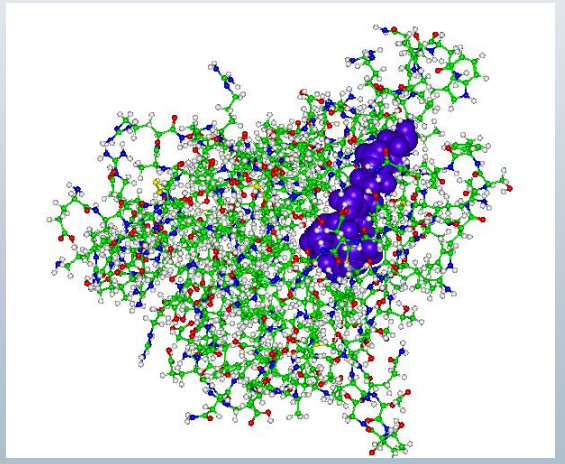
- *All-atom free-energy forcefield* that can fold a family of helical proteins
- *Stochastic Optimization Methods* to reproducibly fold proteins with up to 60 amino acids



Schug et.al., Phys. Rev. Lett 91,159102(2003)

Drug Development: from structure to drug

- *FlexScreen* for in-silico high-thoughput screening with flexible protein receptors and ligands for up to 250,000 compounds
- *IntelliScore:* adaptive scoring functions



Herges et.al., Nanotechnology 14, 1 (2003)

HELMHOLTZ GEMEINSCHAFT

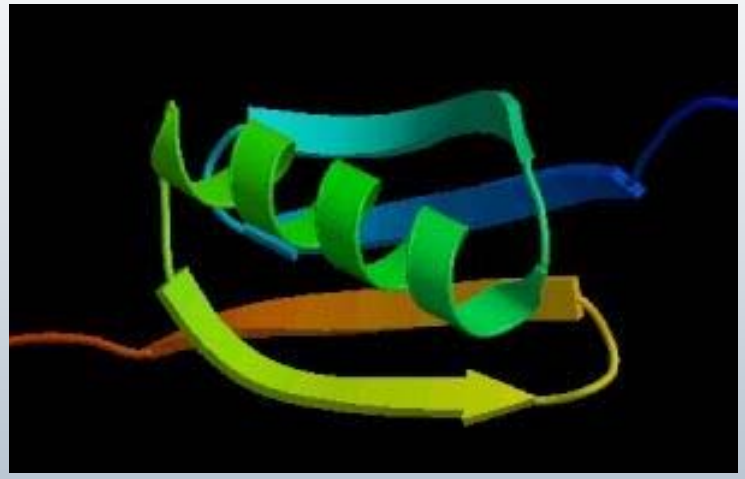
Protein Structure Prediction

- Proteins are the building blocks and machinery of life
- sequential molecules assembled from 20 aminoacid building blocks
- efficient methods to determine the sequence are available
- but, the knowledge of the sequence is insufficient to understand the biological functions
- structure determinaton is much more expensive than sequencing

from sequence:

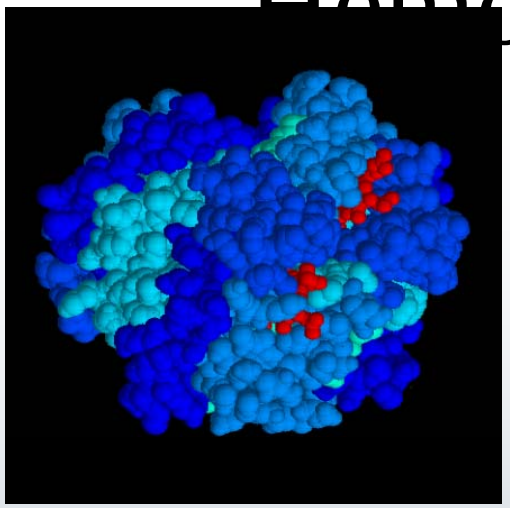
VAL LEU SER PRO ALA ASP
LYS THR ASN VAL GLY .....

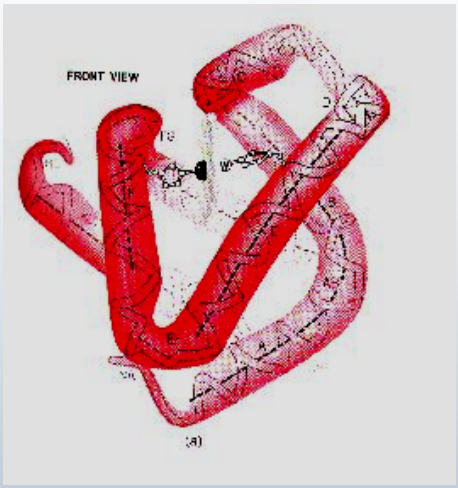
to structure:

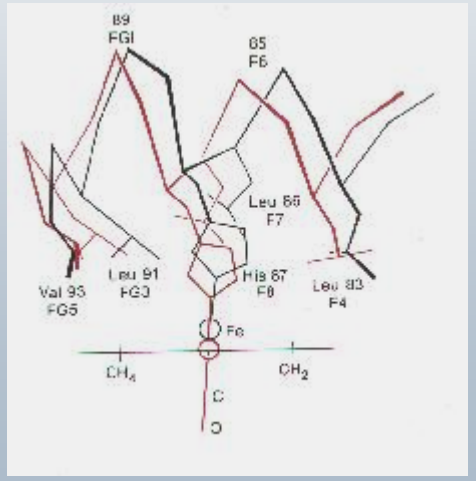


HELMHOLTZ GEMEINSCHAFT

Representation of the Hemoglobin Protein



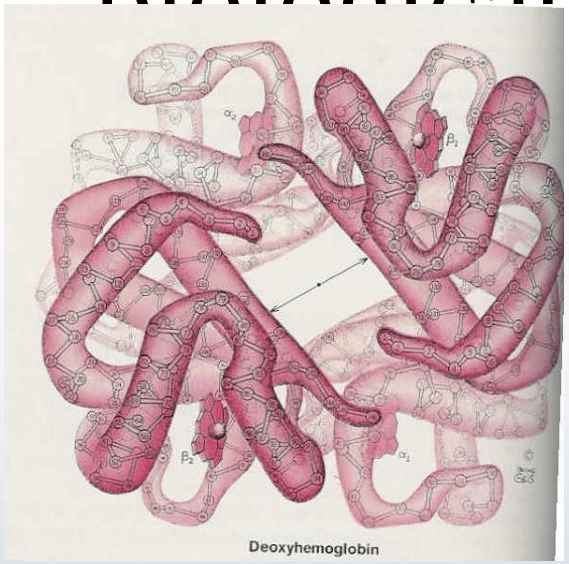


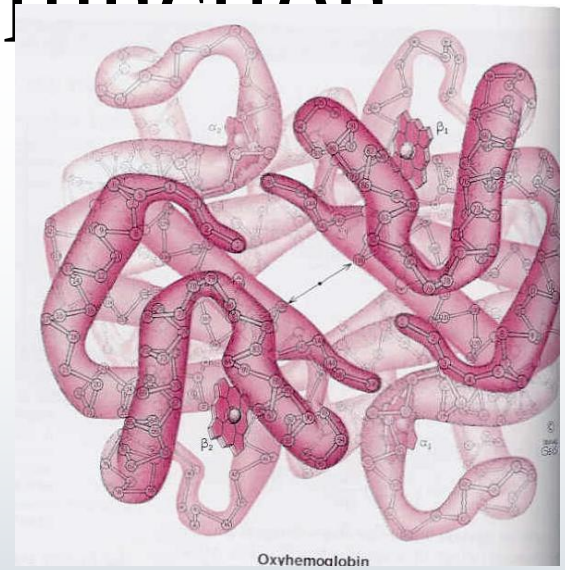


Structure resolution permits the analysis of biological function

HELMHOLTZ GEMEINSCHAFT

Hemoglobin: Control of biological function

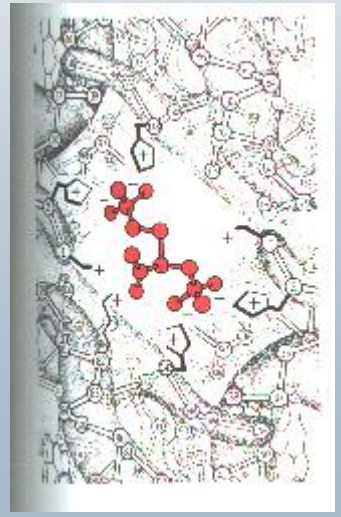




Structure analysis permits control of biological function.

There are 10,000,000 sequences available, but only 25,000 structures.

Movies instead of snapshots, design of inhibitors or enzymes



HELMHOLTZ GEMEINSCHAFT

Prediction Methods

Homology Models

Transfer structural information from databases of resolved proteins on the basis of partial sequence similarity.

Advantage: Fast, wins present day prediction competitions (CASP)

Problem: can reproduce only what is in the database, requires large degree of similarity for successful prediction, need to rank different propositions

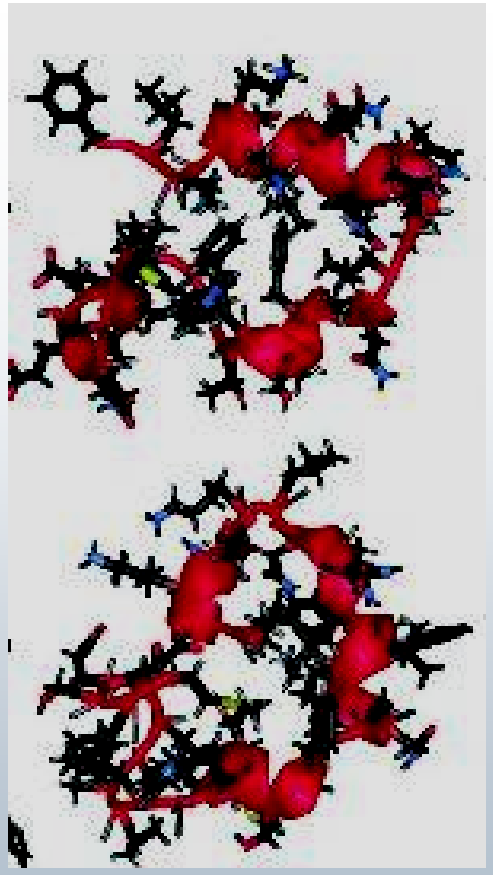
Prediction by Folding

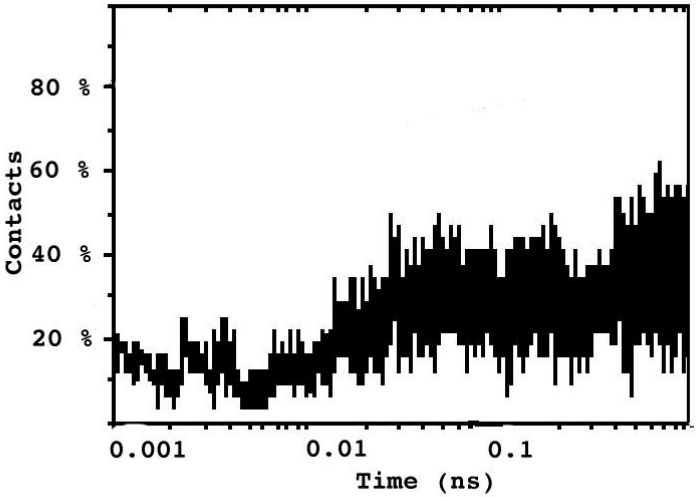
Solve protein equations of motion: Protein folding occurs on the milisecond time-scale, while molecular dynamics time steps are on the femtosecond timescale

$$m\,\ddot{x}_i = -\frac{\partial V(x)}{\partial x_i}$$

HELMHOLTZ GEMEINSCHAFT

Folding Pathway with Molecular Dynamics





256 nodes CRAY T3E
= **85 CPU Years**

Reproducible folding / unfolding has been observed for peptides in helices/bends/beta-sheets for up to 20 amino acids in direct simulation

For larger proteins, such as the villin headpiece (36 amino acids), one has to rely on rare events (Folding@Home)

HELMHOLTZ GEMEINSCHAFT

Protein Structure Prediction by Free Energy Optimization

- Thermodynamics hypothesis (Anfinsen, 1972):

  Proteins are in thermodynamic equilibrium with their environment !

- Native conformation is the global optimum of the free energy
- replace internal energy in the simulation by *effective free energy*
- simulation problem is replaced by *structure optimization* problem
- structure optimum can be found

  without recourse to the folding dynamics

- Enormous gain in efficiency, because optimization methods can visit unphysical intermediates

HELMHOLTZ GEMEINSCHAFT

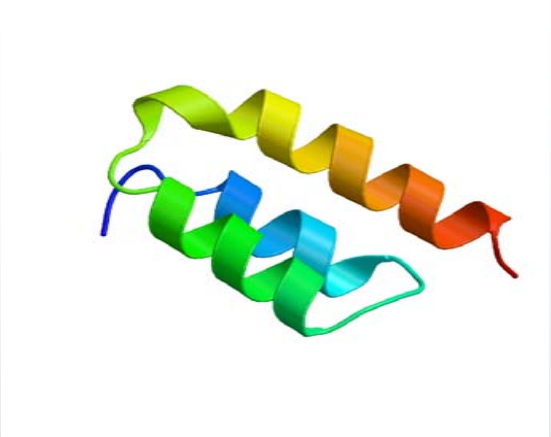
Protein Forcefield PFF01/PFF02

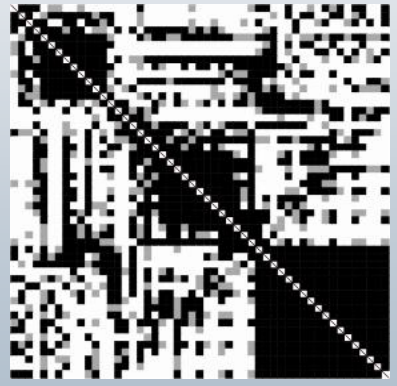
- All atom forcefield (except CH_n)
- Bond distances and angles are **fixed**
- Dihedral angles of backbone and sidechains are **free**
- Lennard Jones
  parameterized to experimental structures of 137 proteins
- Electrostatic interaction
  group specfic dielectric constants (Avbelj, Moult 1992)
  correction for main-chain dipole-dipole interaction
- Solvent Model
  SASA model based on Eisenberg/McLachlan parameters
- Hydrogen Bonding
  parameterized to a set of helical fragments and bends
- *Torsional Potential for Backbone dihedral angles*

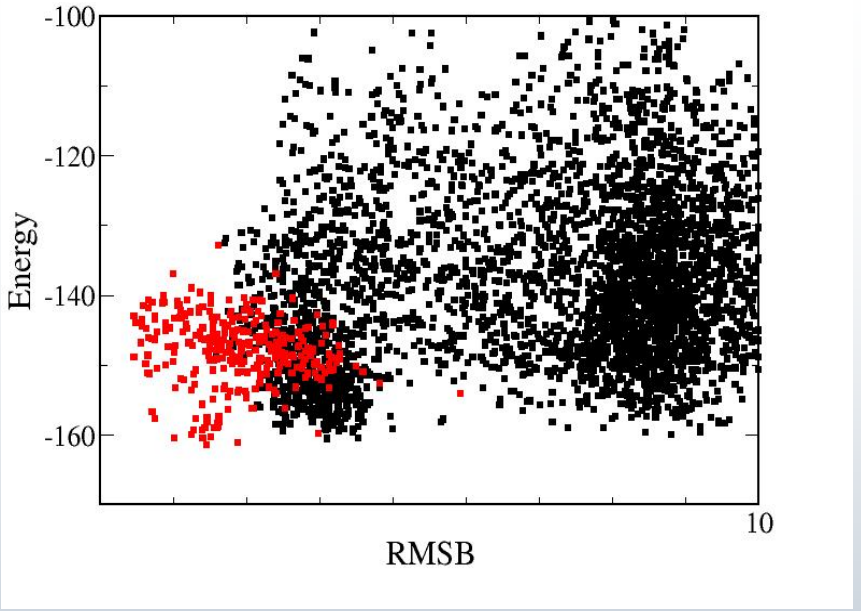
Herges, et.al. *Biophysical Journal* (2004)
Verma, et.al. (in prep)

HELMHOLTZ GEMEINSCHAFT

Decoy-Generation for Protein A







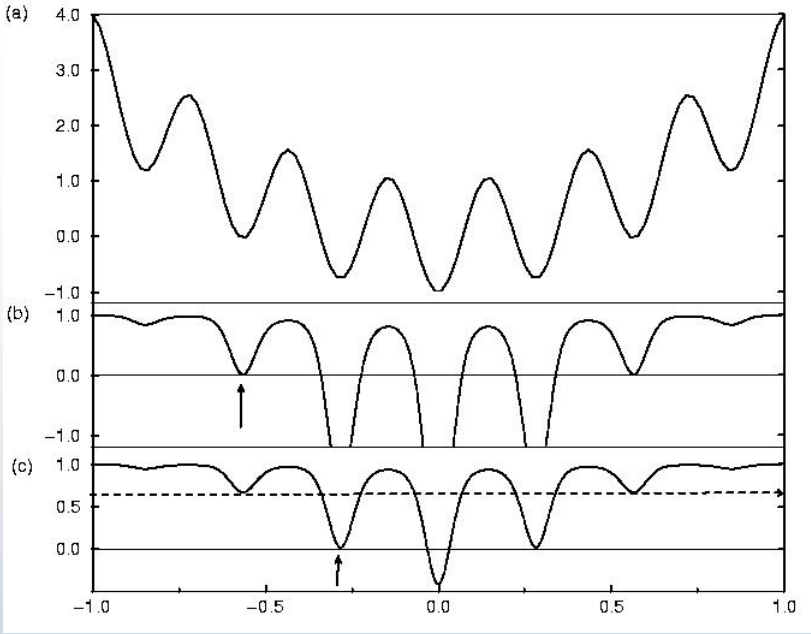
Generate 10,000 decoys from random and NMR starting configurations, improve the best through repeated optimization (cost 2 CPU years).

Herges, Biophysical Journal (2004)

HELMHOLTZ GEMEINSCHAFT

Optimization by Stochastic Tunneling

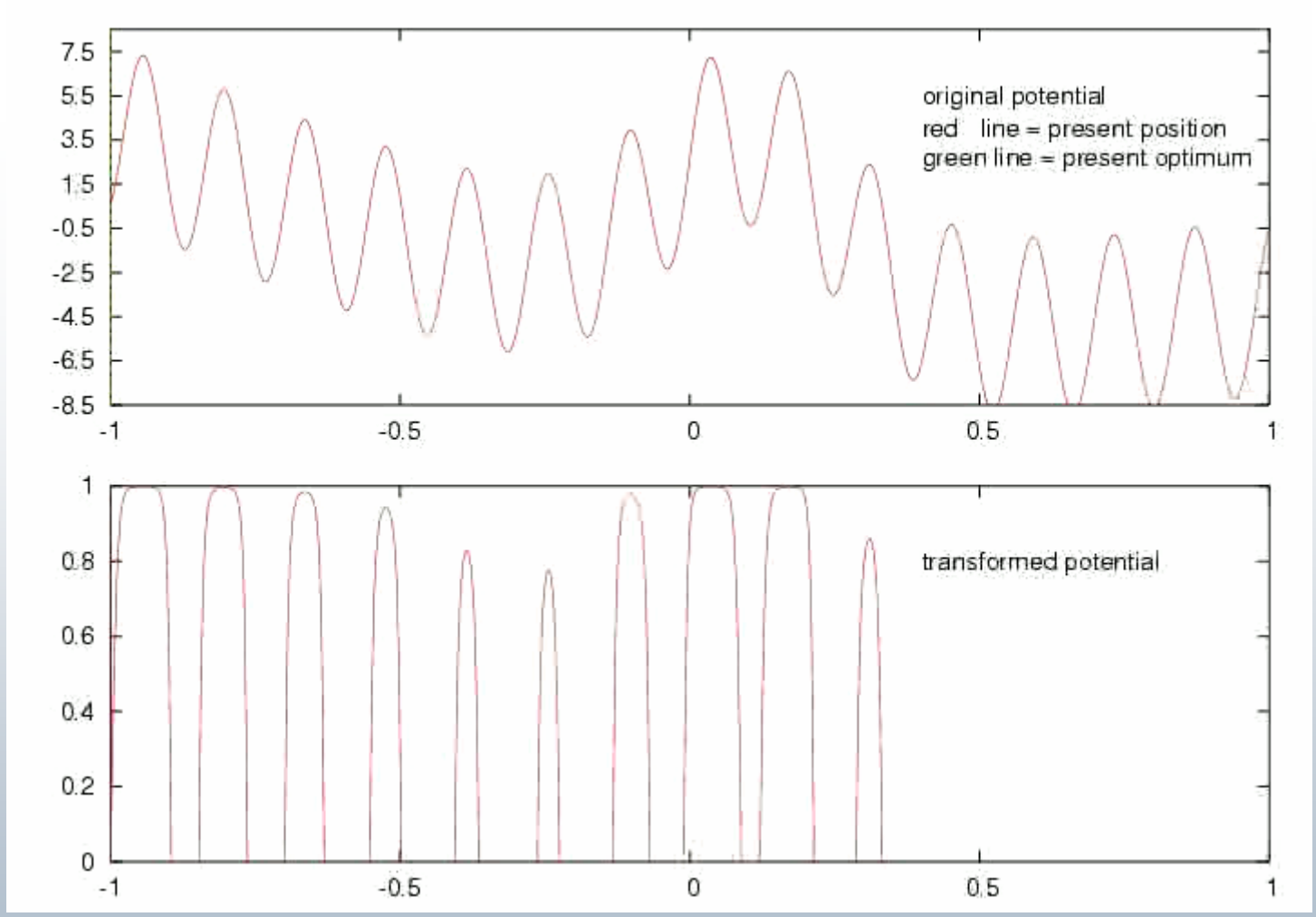
- at any point in the simulation, the detailed structure of the potential above the present best energy E(R) is irrelevant, while the details of the potential below the best energy found are very important
- compress the potential above E(R) to a fixed interval and stretch the potential below
- preserve the location and relative order of the minima



$$f_{eff}(x) = 1 - e^{-\gamma \otimes [f(x) - f(x_0)]}$$

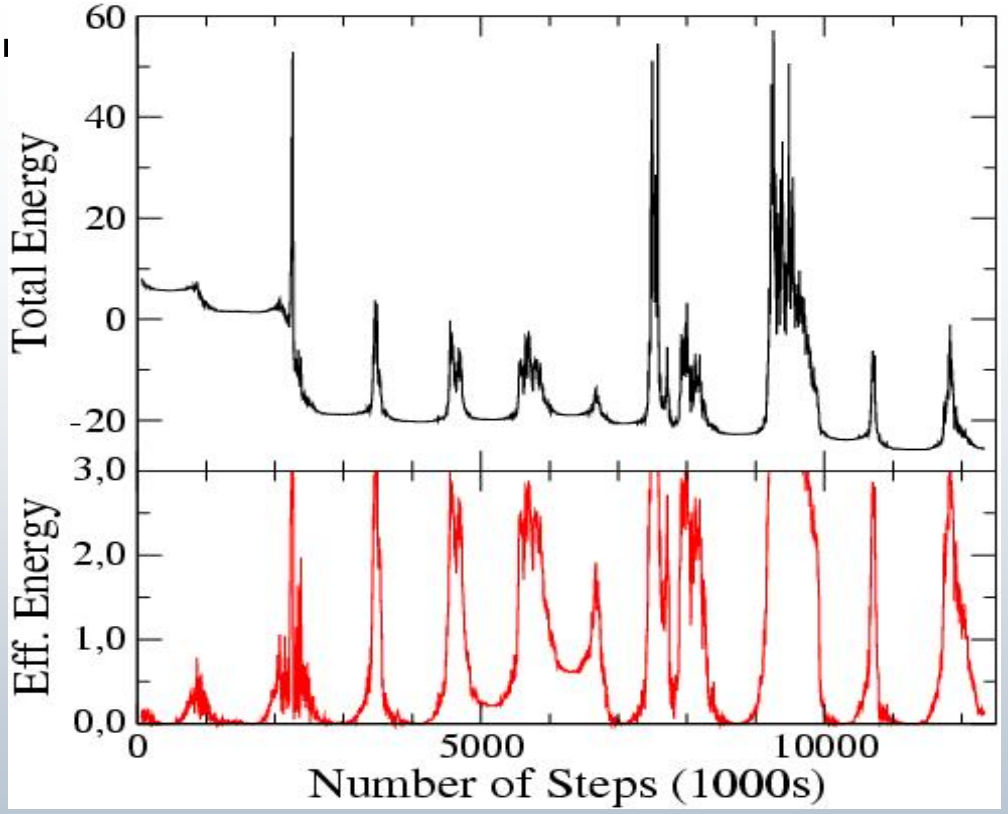
HELMHOLTZ GEMEINSCHAFT

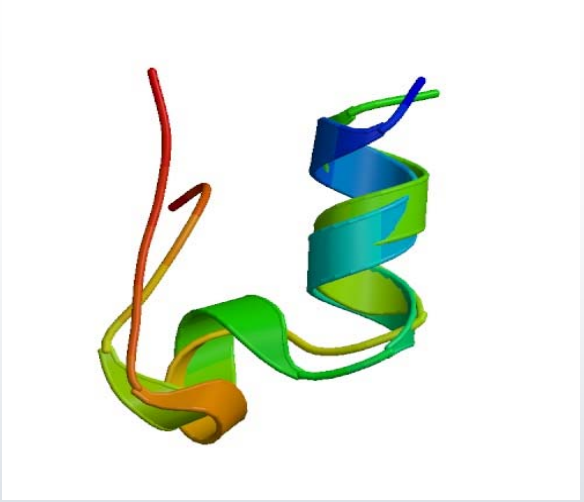
Wenzel, et.al. *Phys. Rev. Lett.* (1999)



HELMHOLTZ GEMEINSCHAFT

Reproducible and Predictive Folding of the Trp-Protein with the Stochastic





Schug; et. al PRL 2003

The energetically lowest 8 of 25 simulations converged to structures within 1kcal/mol and 2-3 A RMSB to the native conformation.

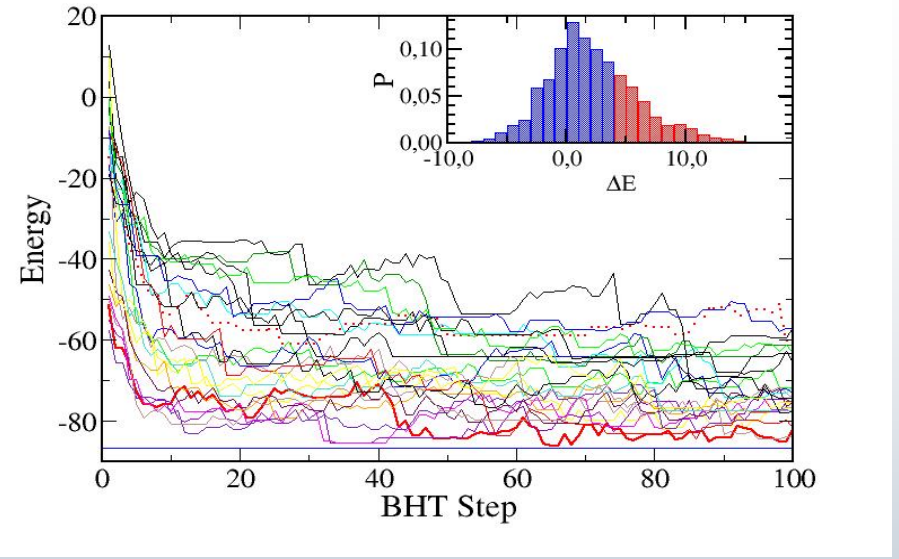
HELMHOLTZ GEMEINSCHAFT

Basin Hopping Technique

Map the original potential energy surface to a simplified potential by associating each conformation with the conformation of an associated local minimum,

optimize on this potential.

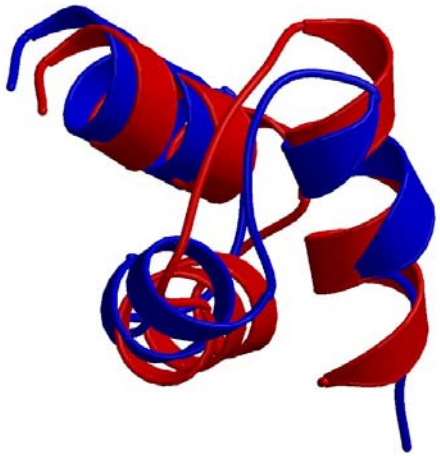
For proteins: local minimization by simulated annealing

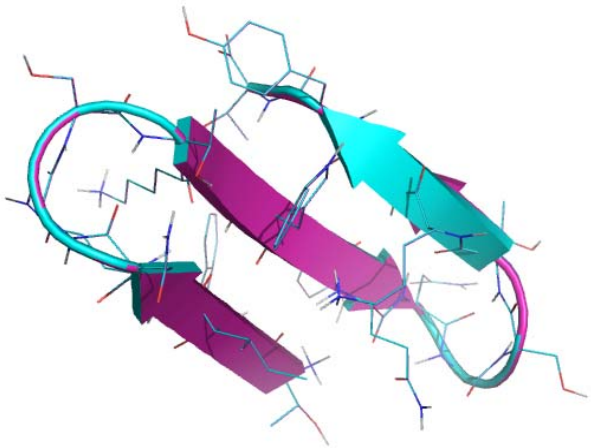


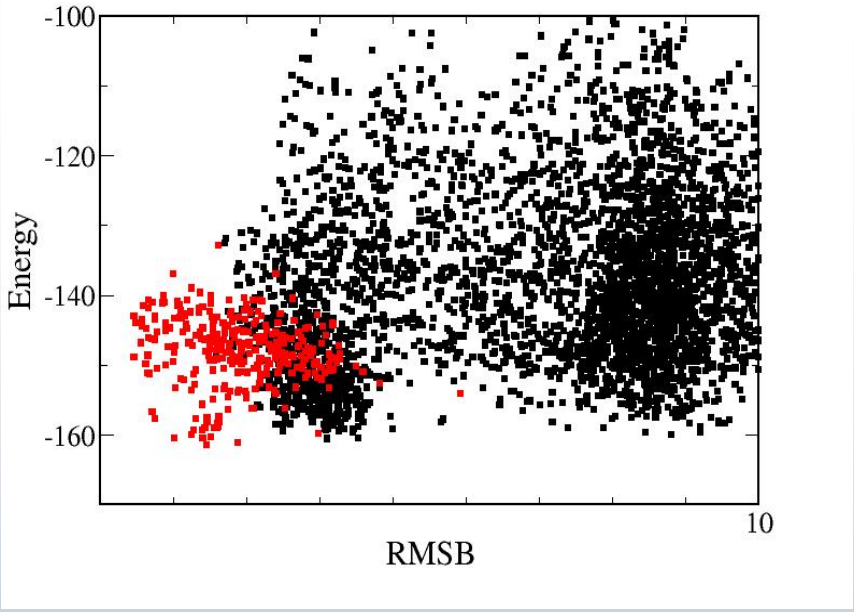
Herges, Wenzel, PRL (2005)
Schug, Verma, Wenzel: ChemPhysChem (in press), J. Chem. Phys (in press)

HELMHOLTZ GEMEINSCHAFT

Proteins Folded with Basin Hopping

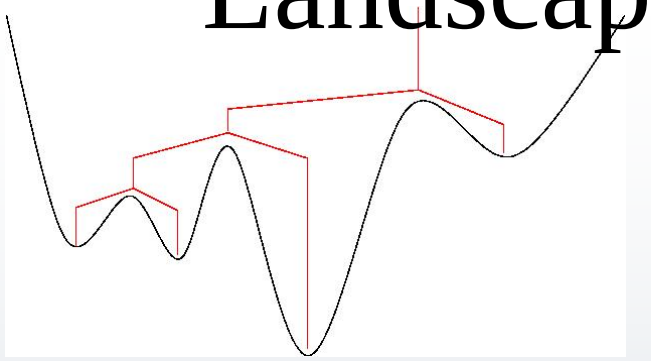






The energetically lowest six of 20 independent simulations converged to 2-3 A RMSB to the native conformation.

Visualization of the Folding Landscape



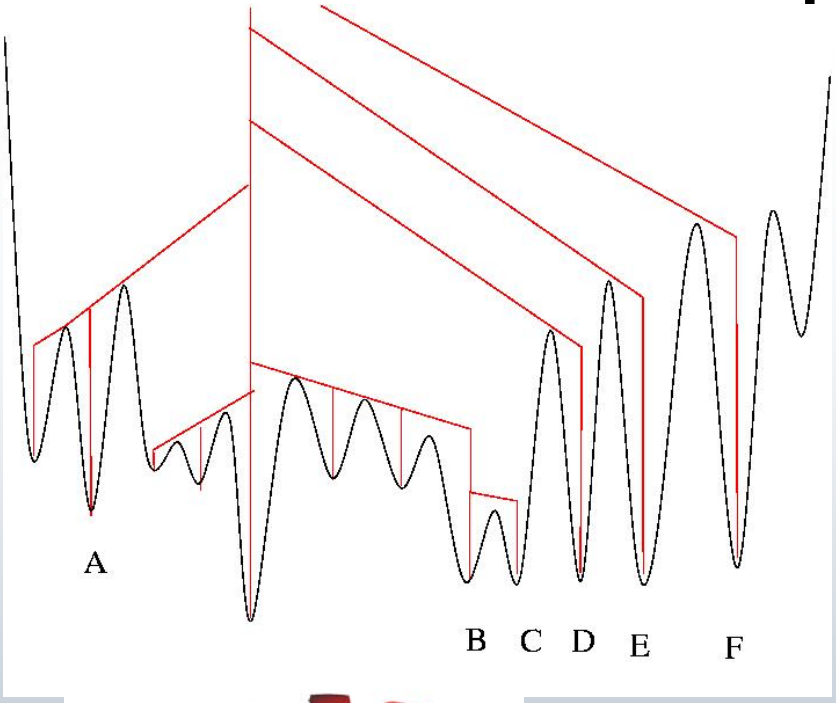
Complete topological characterization of the low energy part of the free energy surface

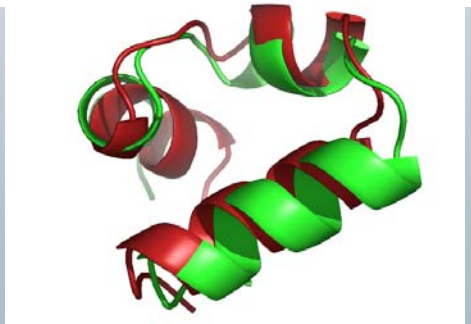
- generate decoys that explore the entire low energy surface
- start with the lowest energy decoy
- Associate all decoys in the next higher energy window with existing families, when they are structurally similar, otherwise create a new family
- Family membership is associative: if A is in the same family as B, and B in the same family as C, A and C are also in the same family.
- As the energy increases, families unite
- Generates inverted tree-structure

Herges, et. al. Structure (2005)

HELMHOLTZ GEMEINSCHAFT

Energy Landscape of the Villin Headpiece





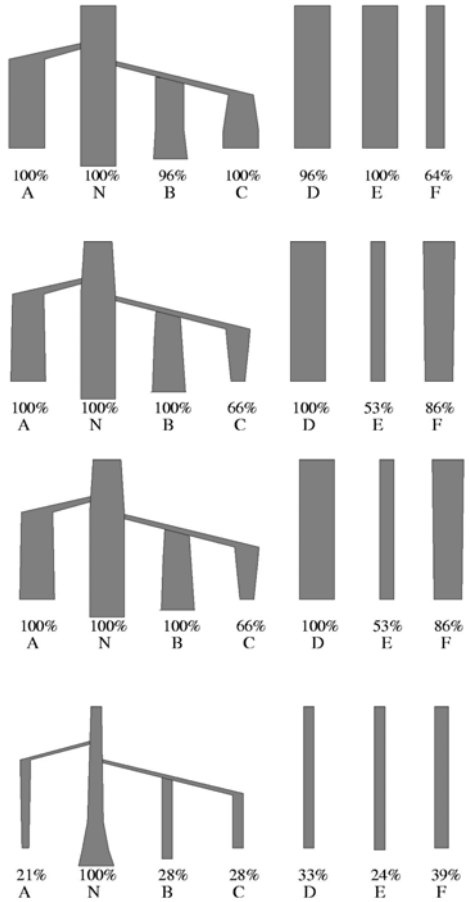
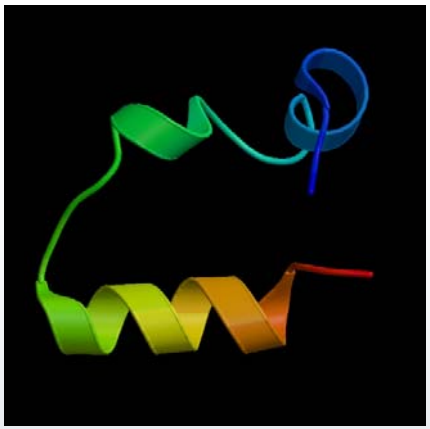


Figure 4

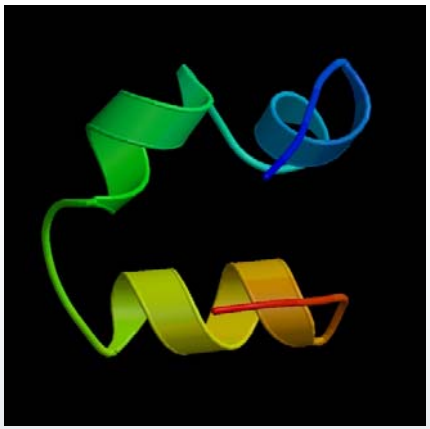
HELMHOLTZ GEMEINSCHAFT

Herges, Structure (2005)

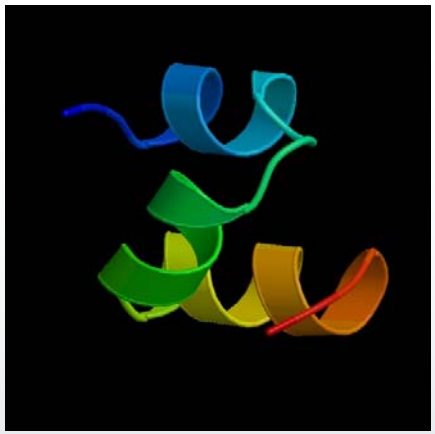
1VII Decoys



NMR

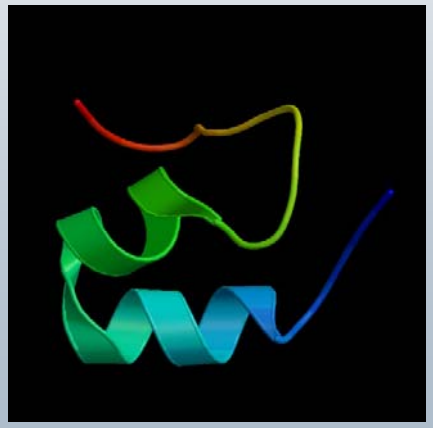


N

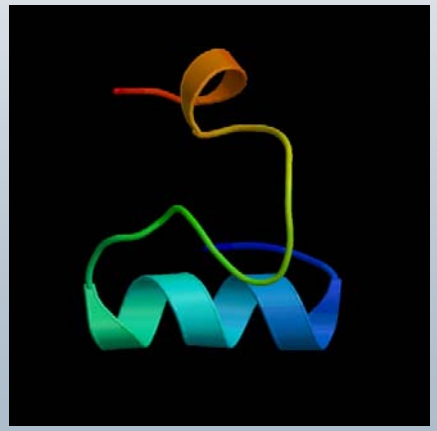


M

A



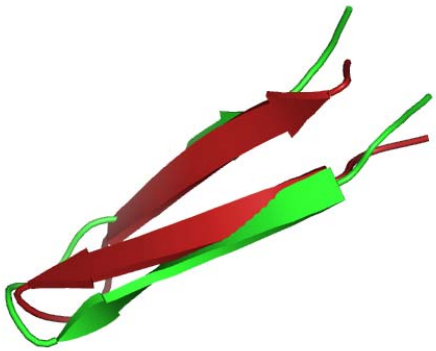
B



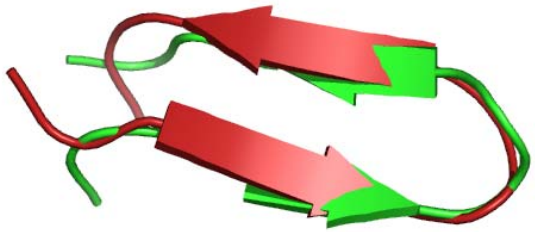
C

HELMHOLTZ GEMEINSCHAFT

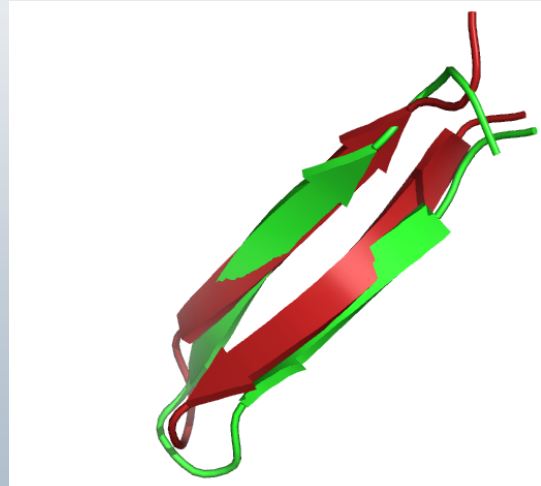
Beta Peptide



1E0Q, 17AA, 2.62 Å



1K43, 14AA, 2.67 Å

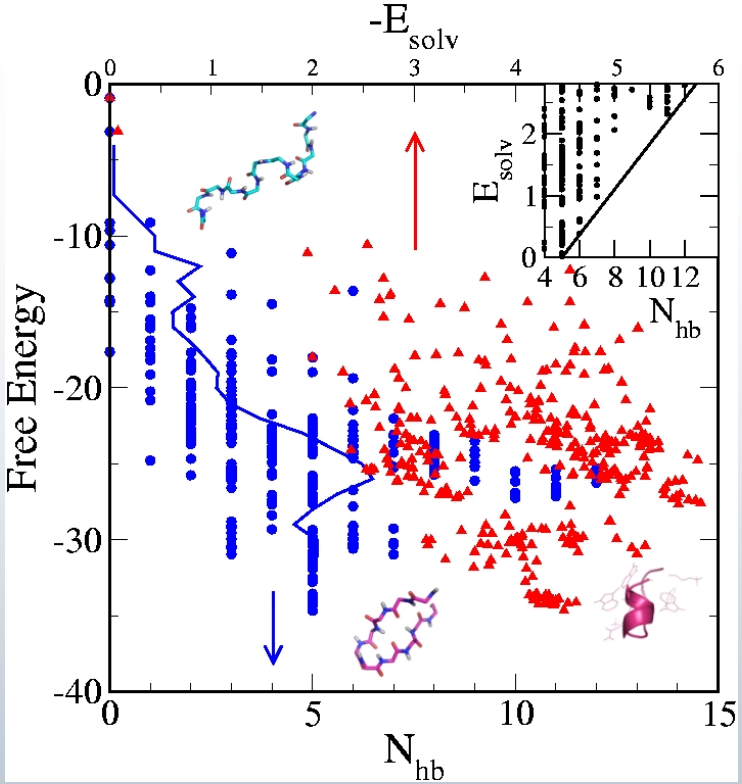


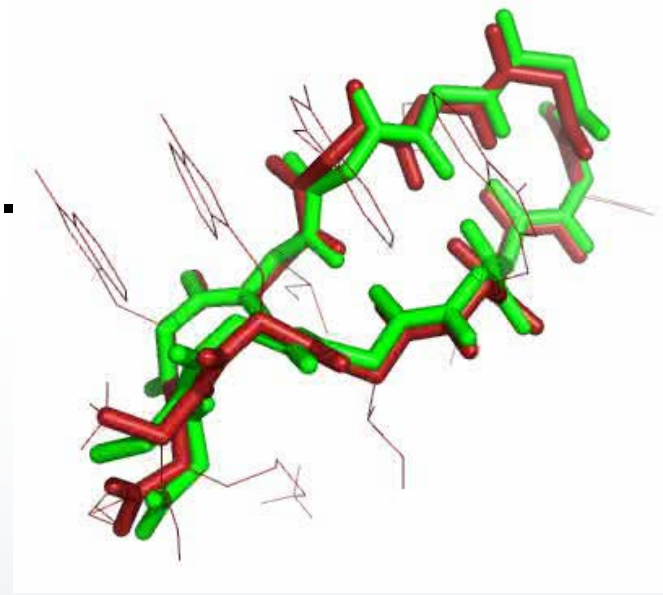
1A2P (85-102), 17AA, 2.53 Å

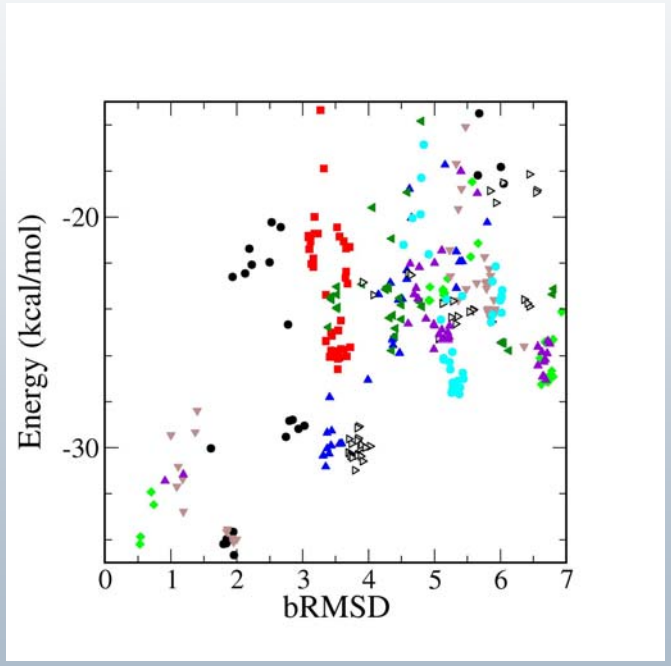
PFF02 stabilizes small beta peptides, reproducible folding Up to 24 amino acids, no mixed systems to date. Decoy studies show that *the helical proteins are not destabilized* in the new forcefield !

HELMHOLTZ GEMEINSCHAFT

Folding the trp-





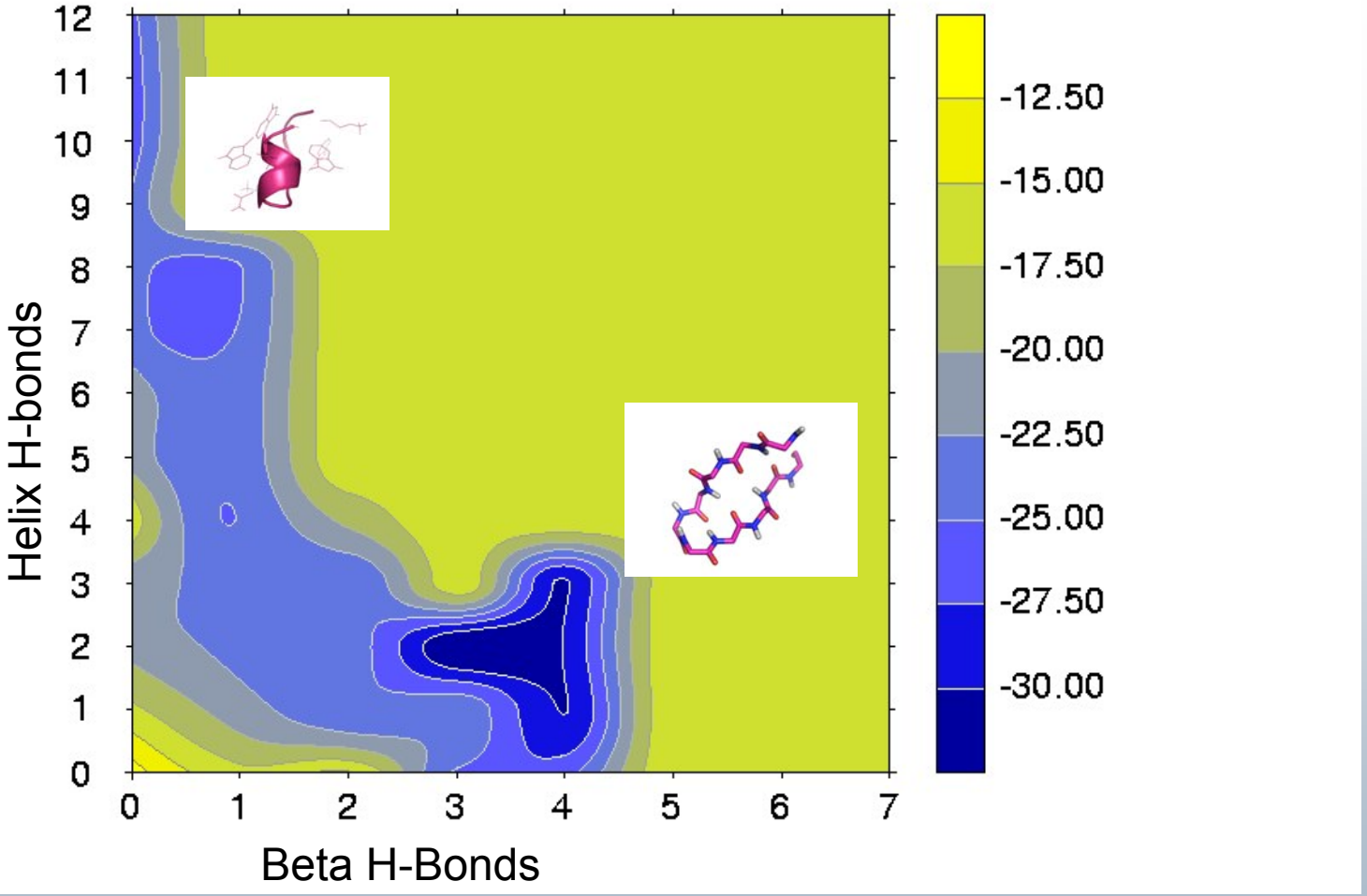


30% of the simulations converge to the native conformation within exp. Resolution, speedup: 10^5

HELMHOLTZ GEMEINSCHAFT

Cochran, PNAS (02), Snow, PNAS (04), Yang, JMB (04)

Internal Free Energy Surface



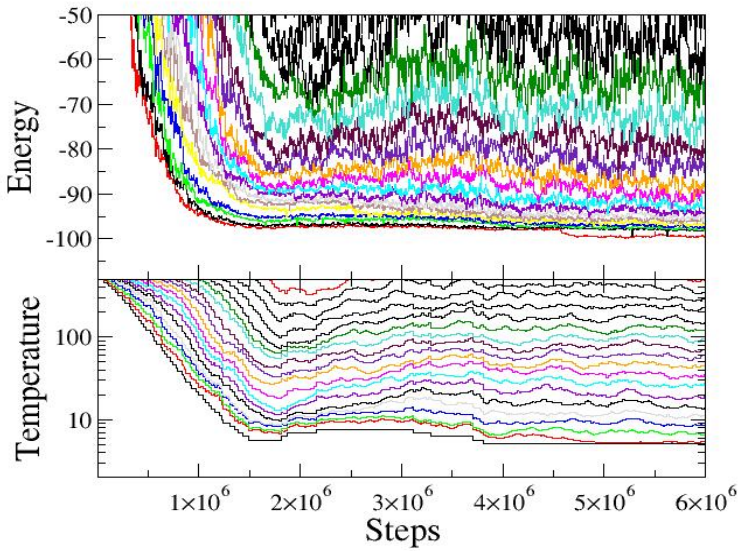
HELMHOLTZ GEMEINSCHAFT

Adaptive Parallel Tempering

Run a number of parallel simulations at different temperatures and exchange their conformations according to:

$$p = \max\left(1, e^{-\Delta\beta\Delta E}\right)$$

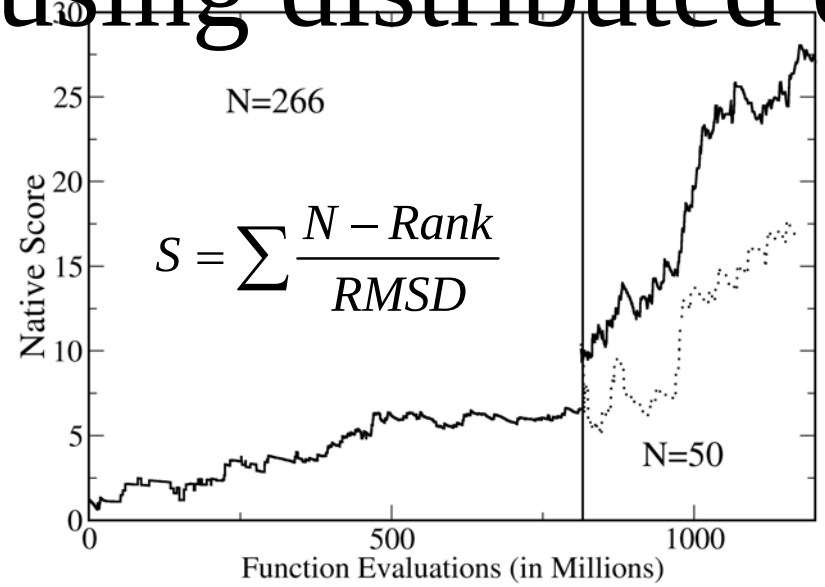
(preserve thermodynamic equilibribum)
Better: adjust temperatures to control exchange rates
Even better: duplicate the best conformation to highest temperature

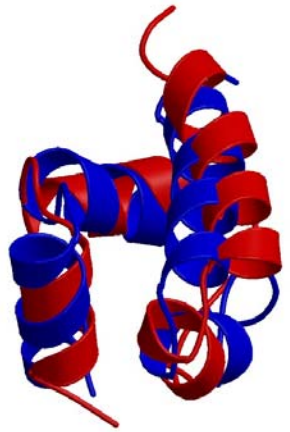


Schug, Herges, Wenzel, *Eur. Phys. Lett.* (2004),
Herges, Schug, Wenzel, *Proteins* (2004)

HELMHOLTZ GEMEINSCHAFT

Reproducible Folding of the Bacterial Ribosomal Protein L20 using distributed optimization





In a population of 2000/200/50 structures in a distributed optimization approach the native state occupies the three lowest conformations and occurs 4 additional times.

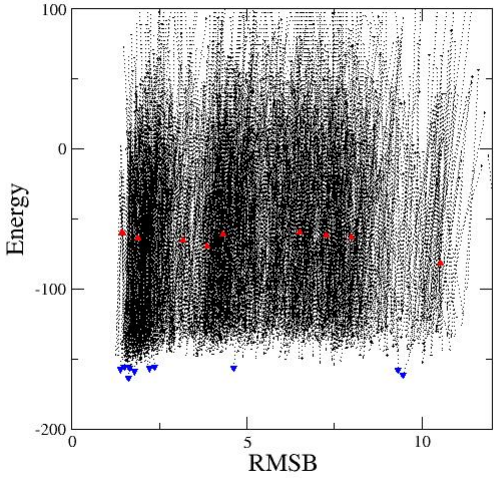
HELMHOLTZ GEMEINSCHAFT

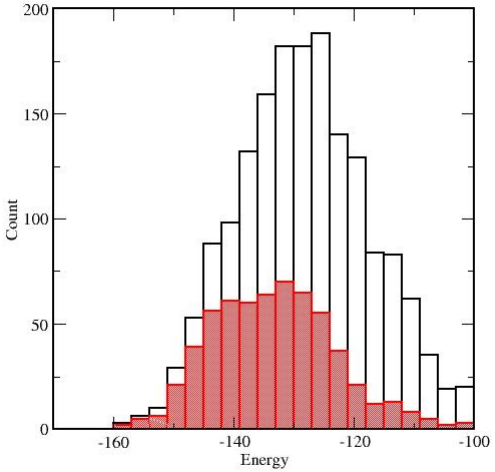
A. Schug, W. Wenzel, *J. Am. Chem. Soc*. (2004)

Protein Structure Prediction with Homology Based Methods and Forc[illegible]

Refinement

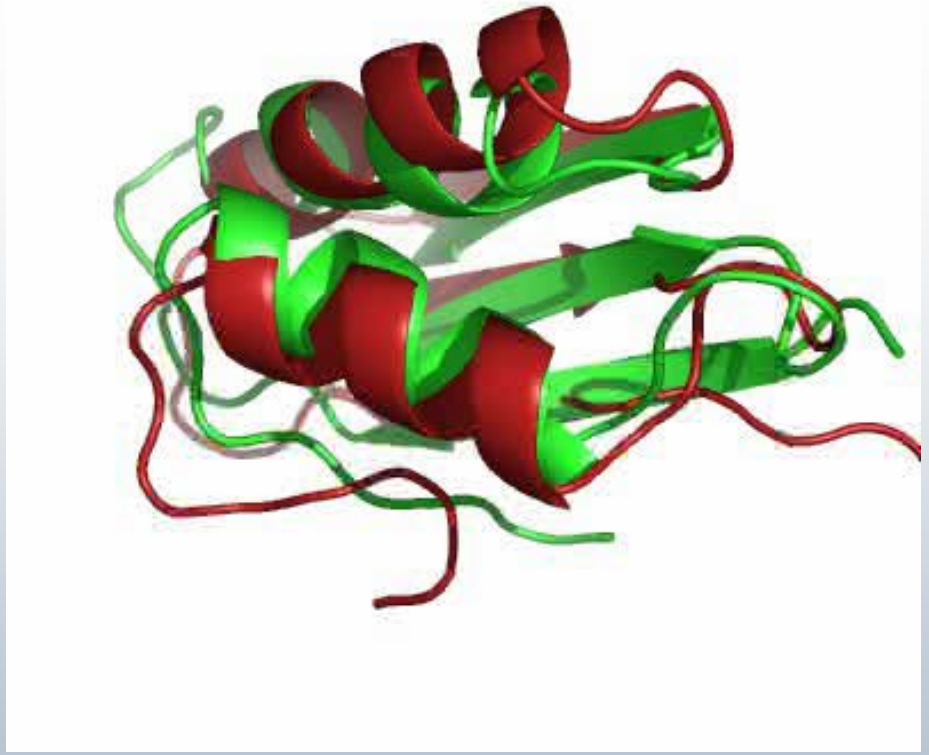
- Decoy set from Rosetta, 43 Proteins, (Tsai et al. Proteins 2003), over 1800 decoys for each protein
- PFF01 stabilizes all helical proteins except one.
- For the helical proteins, where near-native decoys are in the set, PFF01 selects a near native decoy in 9 of 21 cases, but always in the top ten.
- For the one exceptional case, the experimental structure has since been replaced in the PDB
- Significant enrichment even for mixed and beta-sheet systems, but no predictive selection
- average Z-score < -3





HELMHOLTZ GEMEINSCHAFT

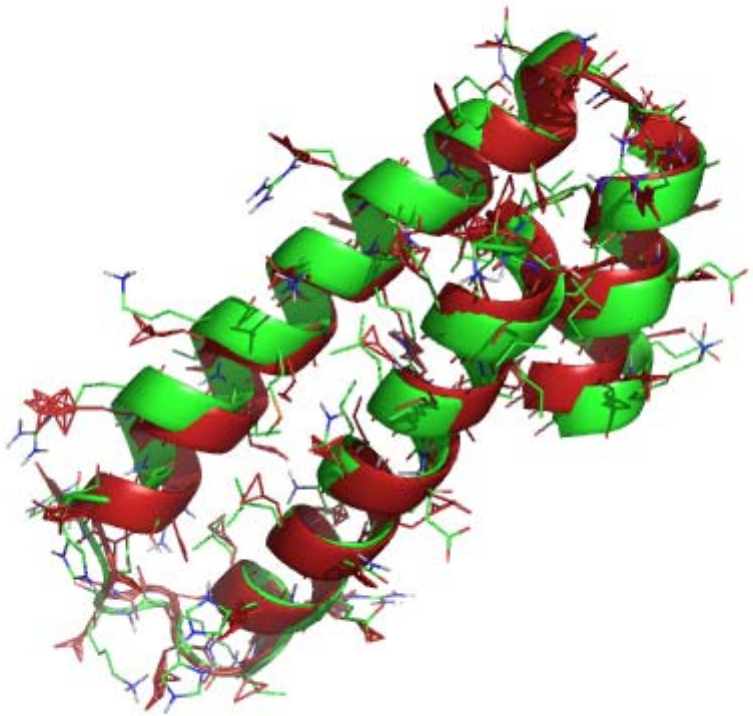
Protein structure prediction by de novo refinement



Pdb: 1afi, 72 amino acids, 2.2 A bRMSD

HELMHOLTZ GEMEINSCHAFT

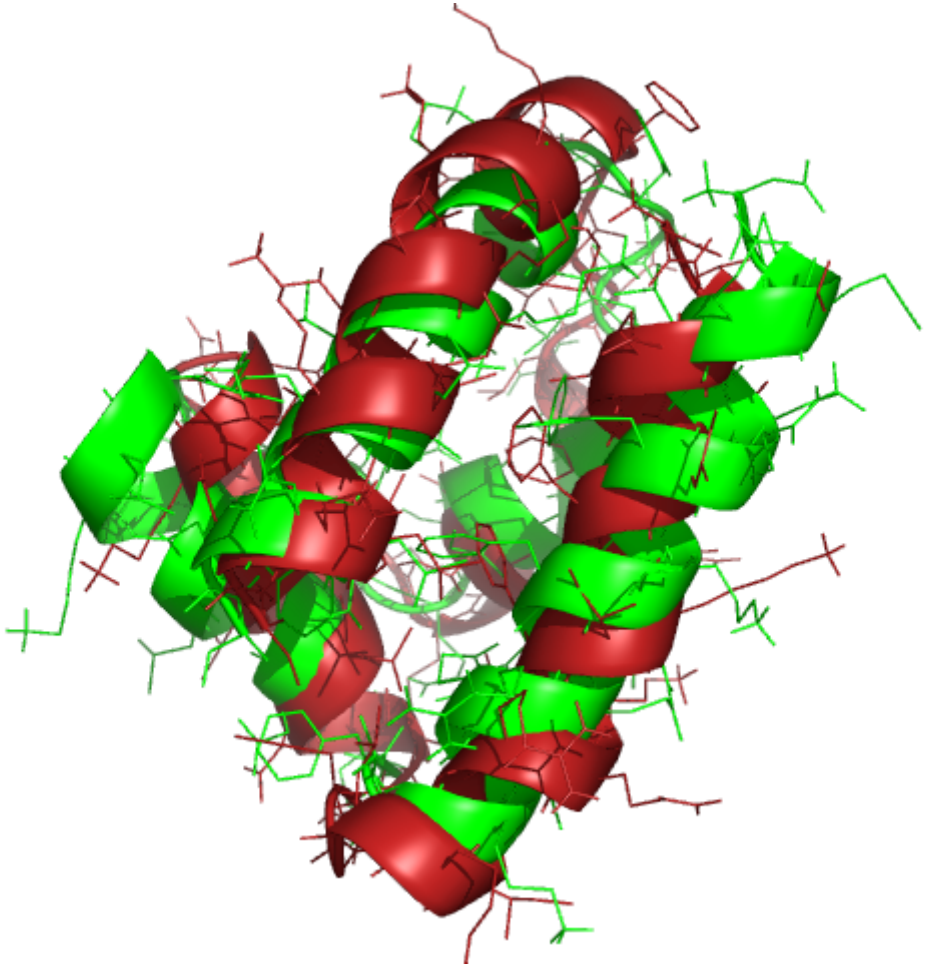
Protein Structure Prediction



1A32, 65 AA, 1.01 Å bRMSD

HELMHOLTZ GEMEINSCHAFT

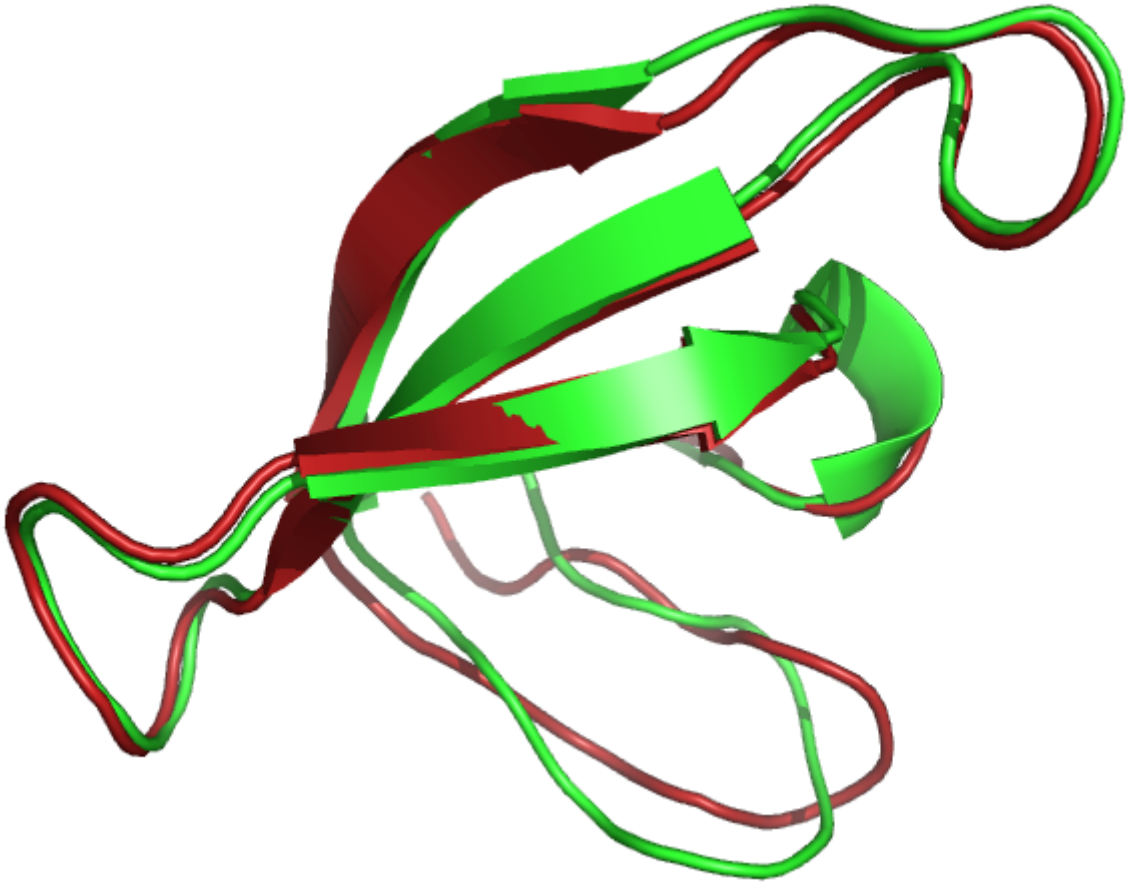
Protein Structure Prediction



1POU, 70 AA, 2.71 Å bRMSD

HELMHOLTZ GEMEINSCHAFT

Protein Structure Prediction



1VIF, 48 AA, 1.45 Å bRMSD

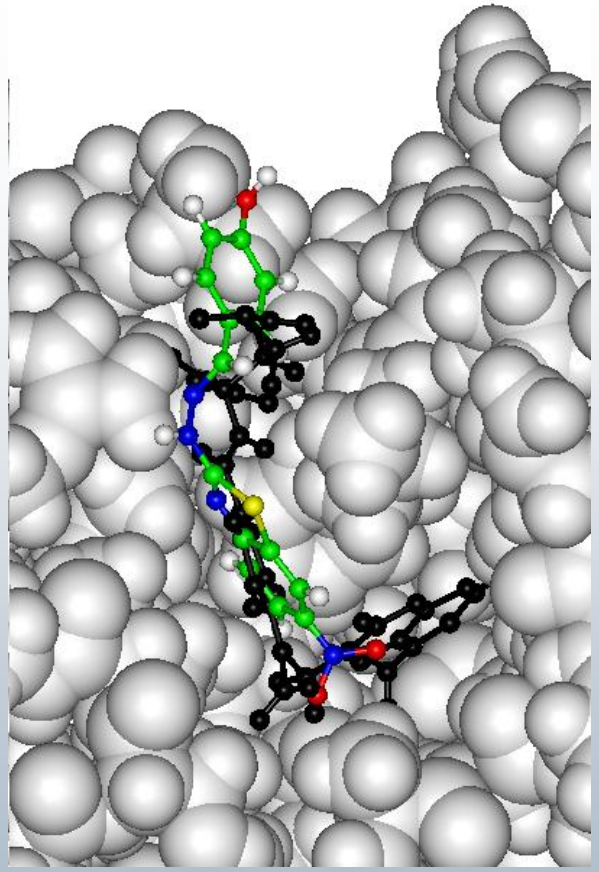
HELMHOLTZ GEMEINSCHAFT

Conclusions

- We have developed and validated all-atom free-energy forcefields that stabilize the native conformation of many proteins as their global optimum
- We have developed and adapted efficient optimization methods that find the global optimum of the protein free-energy surface
- Based on the thermodynamic hypothesis we have predictively folded several proteins with both alpha-helix and beta-sheet secondary structure
- We can characterize the low-energy structure of the protein free energy surface (and possibly reconstruct the folding dynamics)
- Using decoy sets generated from heuristic methods we can predict the structure of proteins from many distinct structure classes

HELMHOLTZ GEMEINSCHAFT

Computational Drug Discovery



Selection of ligands as molecular switches for structurally characterized protein receptors.

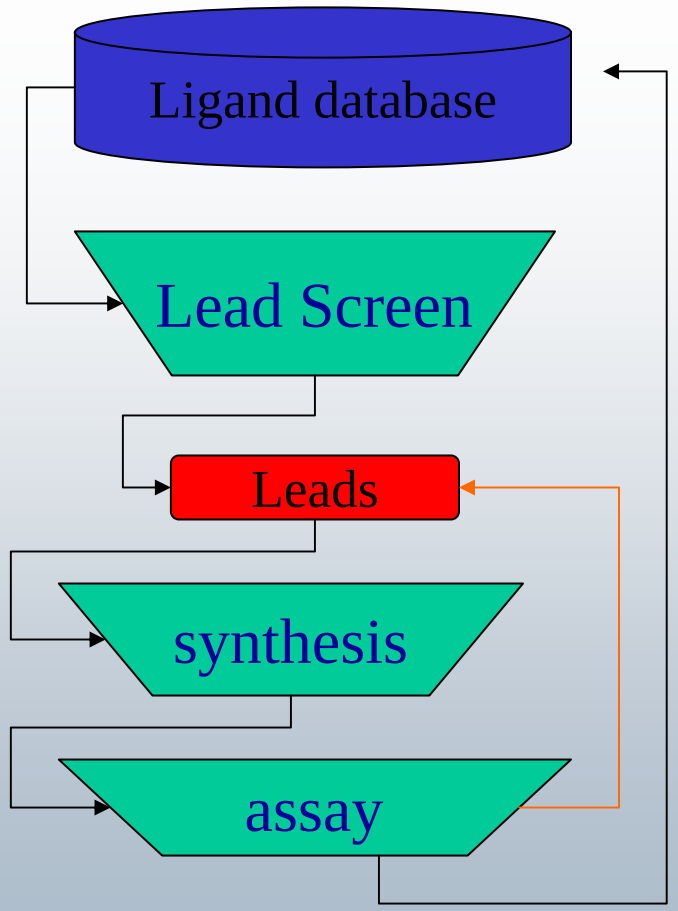
Old approach: QSAR, fast but unspecific

New approach: Atomistic simulation of the docking process

In 2002: 18 drugs in clinical trials worldwide

HELMHOLTZ GEMEINSCHAFT

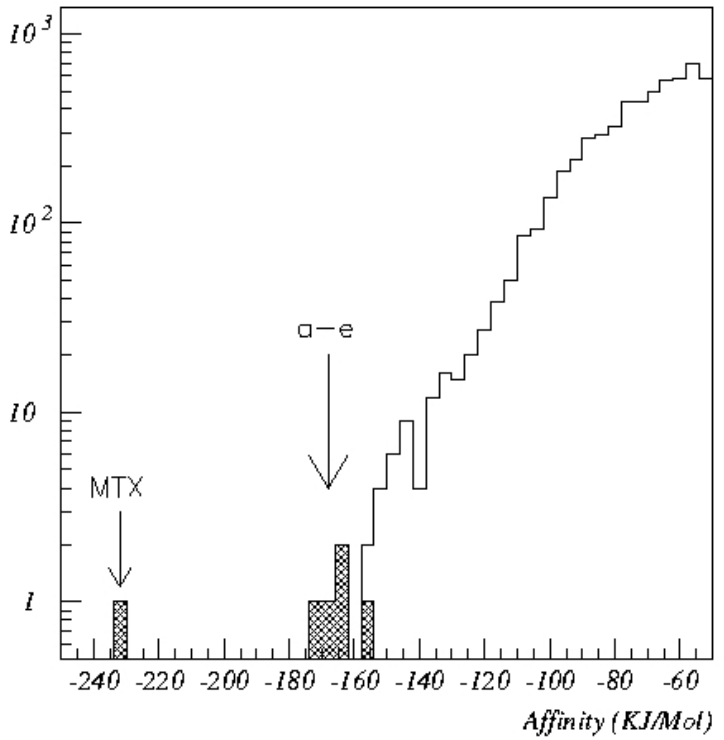
In silico Lead Screening



- Choice of possible ligands from the database
- Synthesis and test of the selected ligands (expensive !!!)
- Improvement through combinatorial chemistry and high throughput screening
- Data base size:10,000,000, i.e. approx 50 ms / molecule
- High specificity of the receptor-ligand pair (key-lock principle) requires atomistic simulations
- Affinity depends on intermolecular interactions

Screening of dihydrofolate reductase

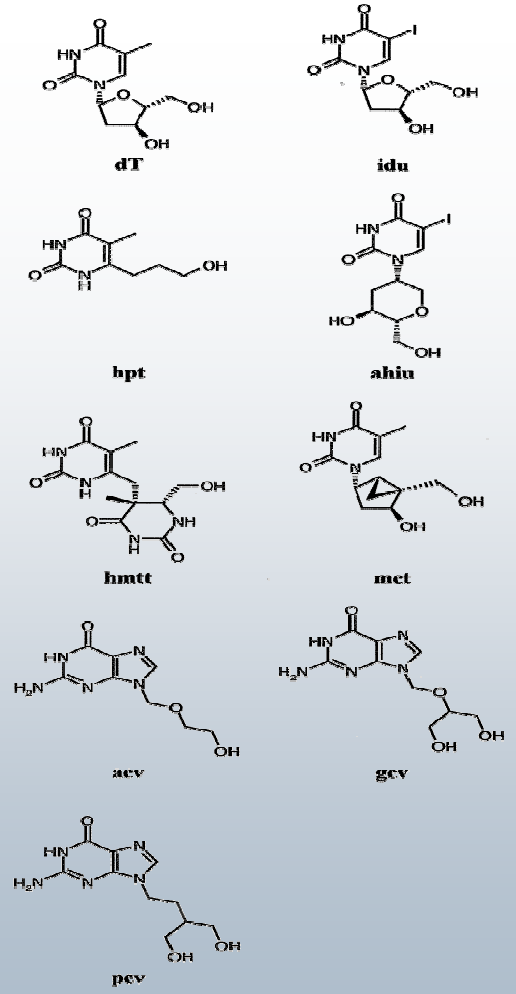
- Receptor for methotrexate (MTX, pdb-entry 4dfr)
- 10000 chemical compounds from nciopen3D database
- MTX was scoring best
- Other top ranking leads display specific binding pattern

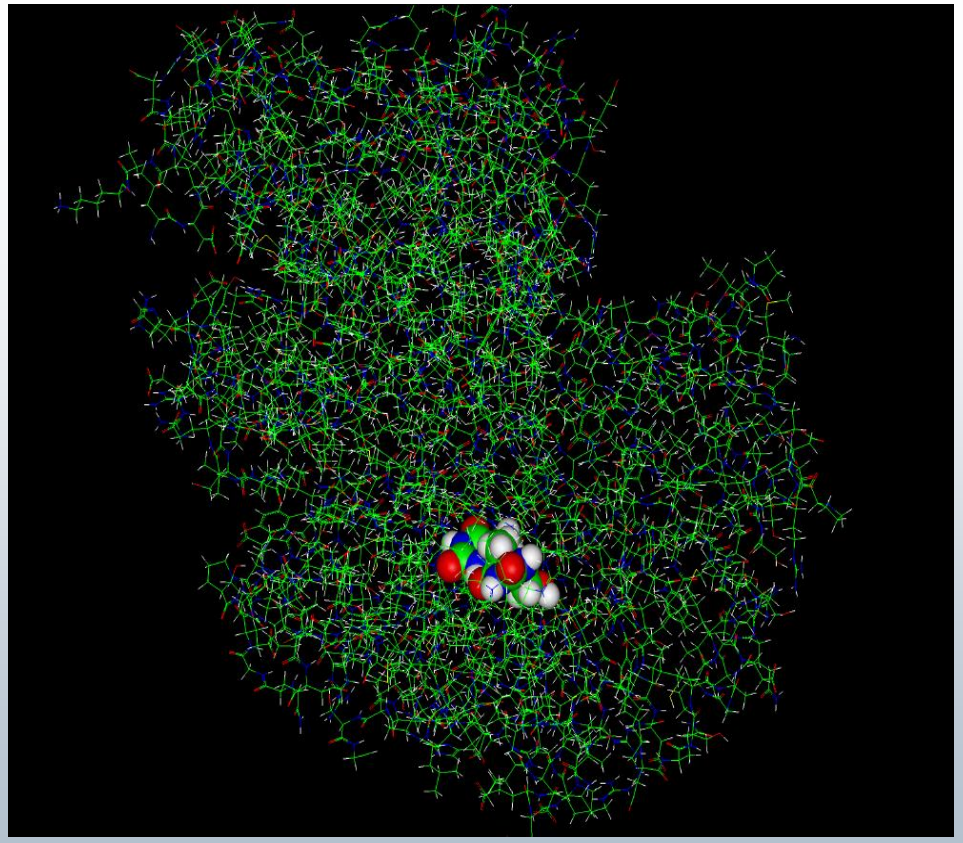


H. Merlitz et al., Chem. Phys. Lett. 370, 68 (2003)

HELMHOLTZ GEMEINSCHAFT

Docking to thymidine kinase





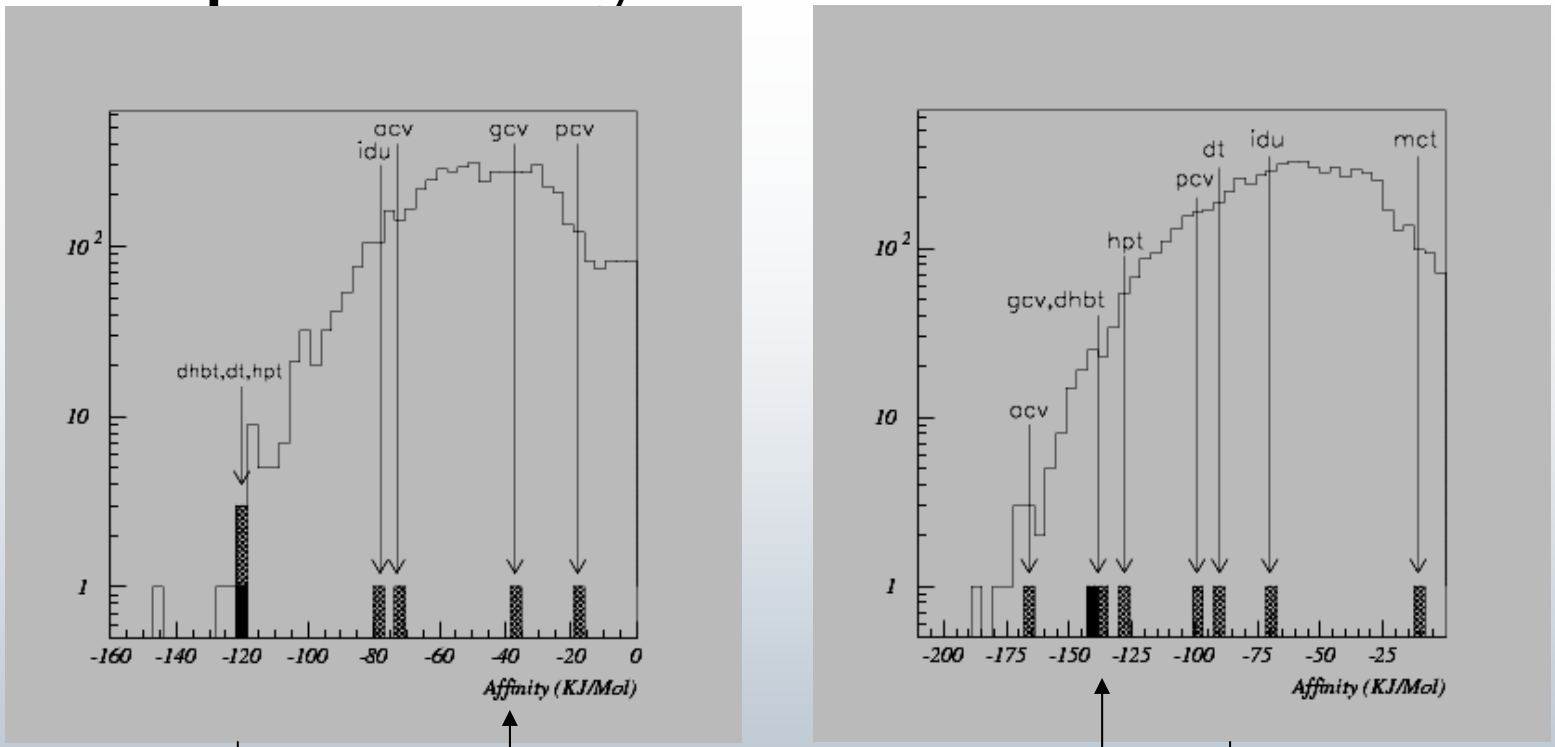
C. Bissantz et al., J. Med. Chem 43, 4759 (2000)

HELMHOLTZ GEMEINSCHAFT

Ranking of 10 substrates against 10000 database ligands for different receptor X-ray conformations

TK in complex with dt

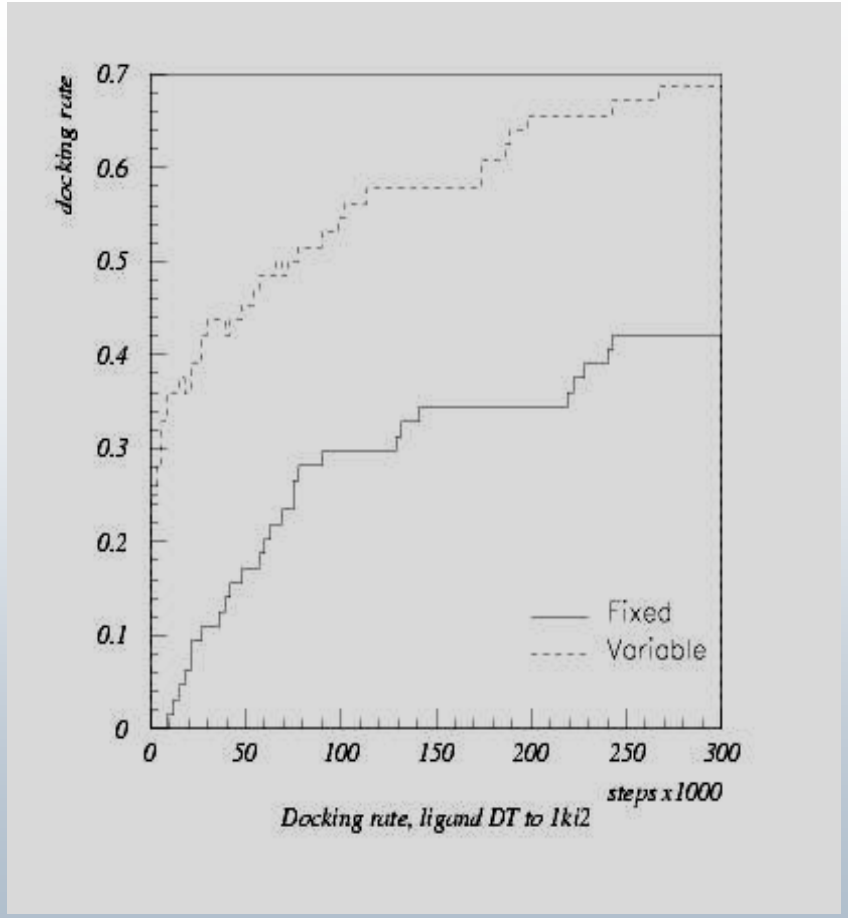
TK in complex with gcv



HELMHOLTZ GEMEINSCHAFT

Merlitz, et.al. : J. Comput. Chem. (2004)

FlexScreen: Receptor Flexibility



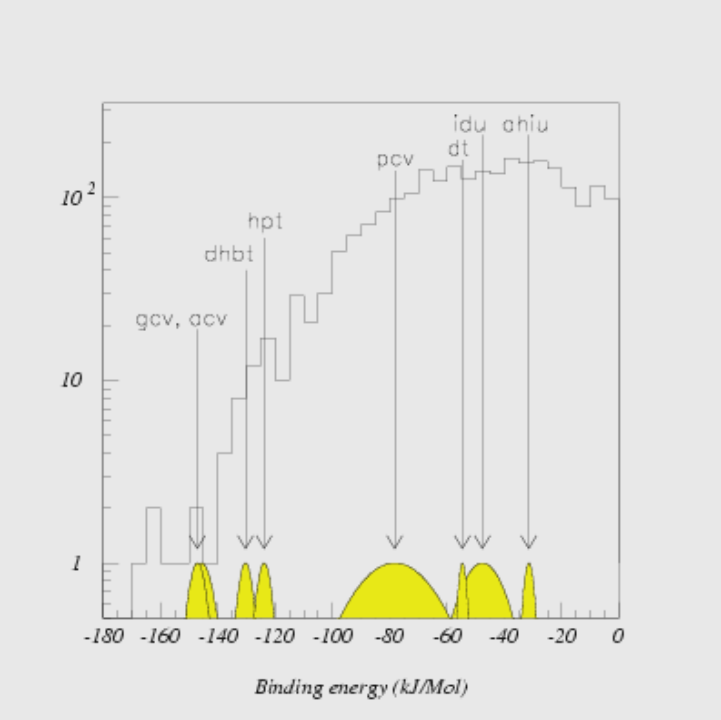
- The consideration of side-chain mobility is a signifcant improvement in model
- The price is a dramatic increase in the number of variables in the optimimization problem
- While energy evaluations are more expensive, the optimization method is unaffected

Merlitz, et.al. : J. Comp. Chem. (submitted)

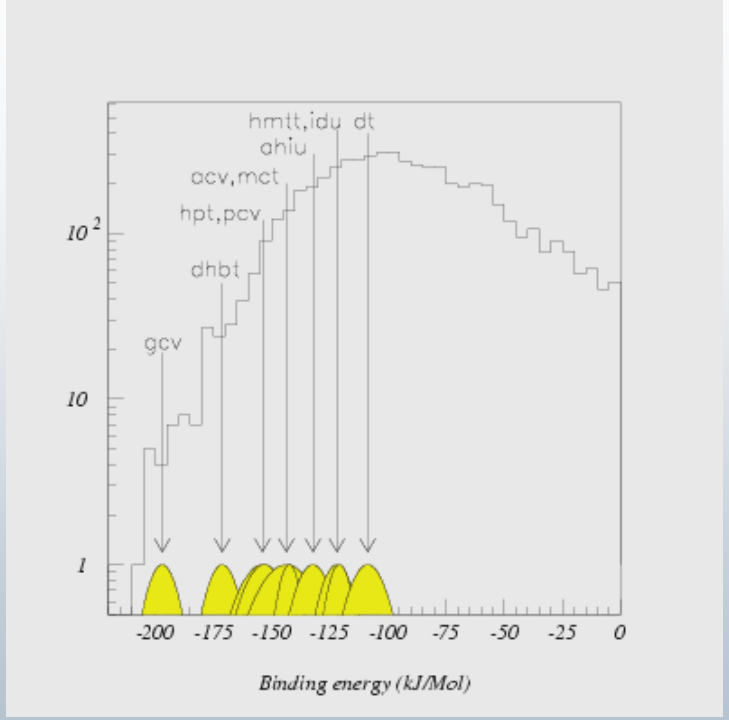
HELMHOLTZ GEMEINSCHAFT

Database screen with receptor flexibility

Rigid receptor



Flexible receptor



Left: Screen to rigid receptor conformation (1ki2, gcv). Docked: 8 of 10 substrates.
Right: 15 flexible bonds enabled. Docked: All 10 of 10 substrates.

HELMHOLTZ GEMEINSCHAFT

IntelliScore: Adaptable Scoring Functions

Rational development of scoring functions for particular receptors and databases

(1) Perform a screen using FlexScreen to obtain ranking of ligands
(2) Synthesis and Affinity measurement
(3) Rationally adjust the Parameterization of the Scoring Function to improve the correlation between the measured and predicted affinities

Repeat steps (2)-(4) until a suitable ligand has been found

HELMHOLTZ GEMEINSCHAFT

FlexScreen / IntelliScore

- The stochastic tunneling method provides an efficient docking algorithm for flexible ligand / flexible receptor screens in *FlexScreen*
- *FlexScreen* screens the NCIopen database (ca. 250,000 ligands) in about 1 week turnaround time
- FlexScreen is able to identify known ligands in the top of the database using an atomistic representation of receptor and ligand (industrial test with 4SC AG, München).
- The *IntelliScore* approach permits a rational evolution of exisiting all-atom scoring functions for specific recpeptors and databases.

HELMHOLTZ GEMEINSCHAFT

Group Members, Collaborators and Funding

Protein Folding:

- Dr. T. Herges
- A. Schug
- A. Verma
- S. Murthy

Drug Development:

- Dr. H. Merlitz
- B. Fischer
- S. Basili

Computational Materials:

- Dr. E. Starikov
- Dr. S. Mujamder
- A. Quintilla

Collaborations:

- J. Moult (Maryland)
- S. Gregurick (NIST)
- K.-Y. Lee (KIST)
- H. Scheraga (Cornell)
- U. Hansmann (Jülich)
- M. Seifert (4 SC AG)
- S. Tanaka (Kobe)
- B. Loeffler, NEC Life Science
- H. Schoeller, U. Simon (RWTH)

Funding:

DFG, BMWF, Bode Foundation, Volkswagen Foundation, KIST

http://www.fzk.de/biostruct

HELMHOLTZ GEMEINSCHAFT