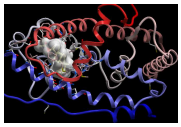
Ruben Abagyan,
The Scripps Research Institute, La Jolla, California

Computational Structural Proteomics and Therapeutic Discovery

Brno, 2006

- Predicting Protein Function and Flexibility
- Improved virtual ligand screening. Induced fit. Case studies
- Predicting Protein Structure and Association Geometry

The Receptor Modeling Challenges

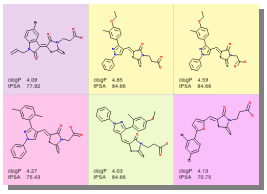


Over 38,000 structure files in PDB (~ 25 /1 new structures a w.day).

A need to extend crystallography with structure prediction:

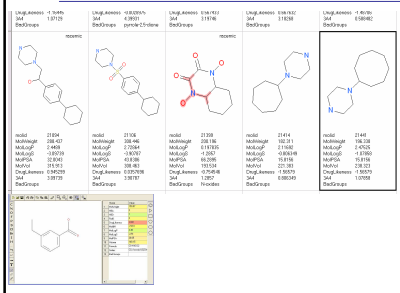
- low *gene* and *domain* coverage of the proteome : "the dark matter" of the structural proteome (predict by homology & ab initio)
- imperfect interpretations of incomplete electron density: predict H, Asn, Gln, His, protonation, errors, missing side chains, loops
- predict conf. of alternative functional states, *induced fit*
- predict *protein or domain association*
- predict *mutations,* SNPs, post-translational modifications
- predict ligand docking for virtual *screening and* ***design***

Revolution in Virtual Chemistry and Pharmacology



- Millions of easily available vendor compounds
- 10 NIH centers: public bio-screening data
- PubChem
- Predicting compound properties: LogP,LogS, CNS, hERG, PgP, CYPs: 3A4,2D6,2C9..

Predicting Molecular Properties



E-ADMET.

LogP, LogS, PSA,
Ubiquitous binding,
Drug-Likeness,
hERG, PgP, CYPs:
3A4,2D6,2C9
Half-life
Metabolites

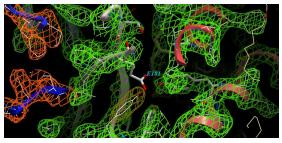
A critical bottle-neck for e-ADMET predictions:

Available e-ADMET data (prediction algorithms are easier)

Goal for public e-ADMET initiatives: generate e-ADMET databases

Preparing receptor coordinates

- PDB coordinates: imperfect interpretation of incomplete electron density.
- Build a complete model (missing side-chains, loops etc.)
- Predict correct Asn, Gln, His orientations, protons, detect errors.

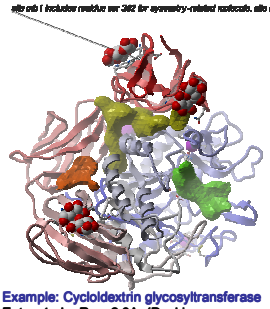


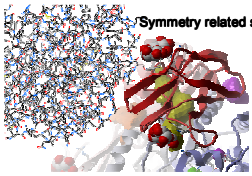
Preparing a pdb-structure for docking

A. Search for a pdb with the closest sequence to your protein of interest
B. Choose the most suitable entry (or several entries)
C. Find, build and edit the pocket composition and geometry.

- **X-ray with up to 2.5-2.8A resolution is preferable over NMR**
- **NMR or homology models are only dockable by skillful operators**
- **Forget electron microscopy**
- **X-ray Resolution < 2.2 A is preferable.**
  (Structures with resolution > 2.3 A may have up to 30% peptide flips, the maps are not self-refinable)
- **Analyze symmetry if the pocket might be at the interface**
- **Analyze relative b-factors. B > 100. are not credible**
- **Pay attention to occupancies (in many cases pocket geometries of ligand conformations/presence are pure fantasies of the authors!).**
- **Analyze alternative positions**
- **Check orientations of His, Asn, Gln**
- **Check protonation states of Glu, Asp, His**
- **Analyze stongly bound water molecules, ions and co-factors .**

Preparations: symmetry





Symmetry related subunit

Example: Cycloldextrin glycosyltransferase
Entry: 1cdg, Res. 2.0A (Docking Rmsd without symmetry: 9.76)
More examples: transthyretin 1f41 (thyroid hormone binds at the dimer interface)

Problem: the true pocket is formed by chains which are not explicitly present in a pdb entry.

Goal: Find all molecules/subunits or chains involved in the interaction with the ligand.

Warning signs: ICM pocket finder does not show pocket density; Binding site is obviously exposed

Recovery: generate symmetry related subunits (View/Cryst.Cell)

Preparations: occupancies, b-factors and alternatives

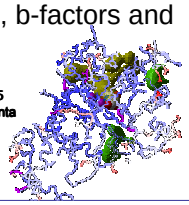
Glossary:

B-factor (or temperature factor):
mean-square displacement of atom from its position in the model. Bi = 79*<u^2> (B of 80 means 1A dev.)
Normal range: 5. – 50. A^2.

Occupancy:
A fraction of atomic density at a given center. It there are two equally occupied conformers, both will have occupancies of 0.5
Normal value: 1. Range: 0.-1.

Alternatives:
If two or more alternative conformations for the same atom or group are discernable in the density, several alternative sets of coordinates are deposited.

Occupancies <= 0.5 are shown in magenta High b-factors are colored red



Problem: sometimes, when electron density is poor and/or ambiguous, crystallographers make things up (or just deposit an arbitrary conformation from a refinement program)

Goal: Identify fantasy atoms/groups

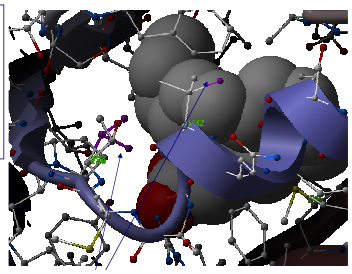
Warning signs: occupancies less than 0.5, b-factors larger than 60-80 A^2.
Tool: Color/label pocket atoms by occupancies/b-factors.

Recovery: Choose another entry, or refine with a ligand, or perform restrained minimization. Choose one of alternatives, or create alternative models

Preparations: occupancies, b-factors and alternatives. Example.

This is a very high resolution structure. For some key residues two alternative conformations are provided.

Recovery:
Choose one alternative or generate several separate docking models

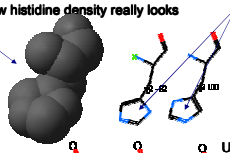


Alternative positions for Thr and Val32

Entry: 1hmt.
Res. 1.4
Fatty Acid Binding Protein with stearic acid

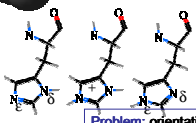
Preparations: fixing histidines

That is how histidine density really looks



Orientation at the heavy atom level
We need to discriminate between These two conformations

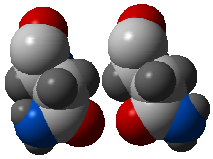
Often the xi2 angle needs to be Corrected by 180 degrees.



Uncertainly at the protonation level
You need to decide which of the three conformations is correct for each important location. The charged conformation is rare.

Problem: orientations and protonation states of histidines are frequently wrong on pdb entries and need to be fixed to ensure correct docking results.
Placement principle: maximization of hydrogen bonds and other interactions with the rest of the protein and/or with the ligand.
Recovery: ICM procedure optimizeHisProAsnGln finds the best orientation and protonation state

Preparations: determining orientations of Gln, Asn, side chains



Orientation at the heavy atom level
The two conformations shown give similar electron density.
We need to discriminate between these two conformations of the **Asn** side chains. The same ambiguity needs to be resolved for the xi3 angle of **Gln**

Background: xi2 in asparagines and xi3 in glutamines are frequently wrong or undefined and need to be corrected ensure correct docking.

Placement principle: maximization of hydrogen bonds and other interactions with the rest of the protein and/or with the ligand.

Recovery: ICM optimizeHisProAsnGln procedure.

Preparations: do I need to uncharge Asp, Glu, Lys and Arg?

Definitions: DERK is Asp (D), Glu (E), Arg ® or Lys (K)
Facts: pKs: His 6.0, Cys 8.3 Glu 4.2, Asp 3.9

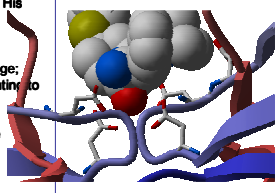
General recommendation: keep the DERK residues charged.

Problem: while in most cases DERKs are charged, in some special cases ED need to be uncharged or His needs to be charged.

Warning signs:
a DERK is buried and NOT involved in a salt bridge;
Several DERKs of the same kind/charge are pointing to the same space.

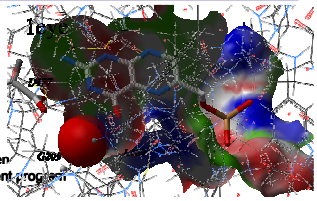
Example: HIV protease. 1ida. Asp 25 and 25' are protonated.

Recovery: Modify them to the uncharged forms.



Preparations: which waters to keep?

Example: 1eye dihydropteroate synthase, anti-mycobacterial/TB target. It binds to the buried Asp177 and improves electrostatic desolvation by ~10 units.



Definition: crystallographic water: an oxygen placed by a crystallographer or a refinement program to a blob of electron density.

General recommendation: get rid of all water molecules,
Keep only water molecules with three or four hydrogen bonds with the protein or ligand atoms.

Reason: keeping inappropriate water(s) will prevect correct docking, while dropping good waters is usually tolerated.
However some tightly bound water molecules help docking and scoring and prevet from erroneous placement of H-bond-rich ligand groups in water sites.

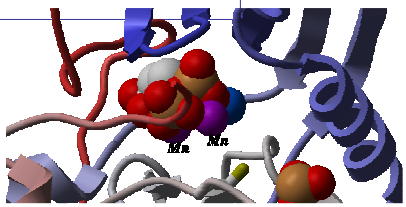
Recovery: Find interface waters with 3 or more protein/ligand neighbors and include them into your model.

Preparations: cofactors and metals?

Problem: metals may be required to dock a charged native ligand (e.g. ATP is charged and requires 2 Mn++ ions.)

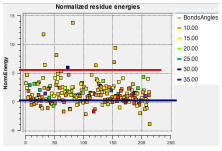
However, to the metals are not necessary for docking of neutral drugs.

Example: a kinase domain. 1atp



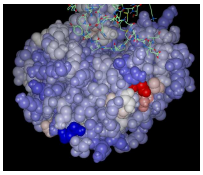
Local quality: Energy Strain

- More sensitive that geometrical and clash criteria
- Based on fast ICM calculation of residue energies (recently accelerated 1000 fold)
- Energy = Vacuum Energy + Solvation + Entropy



Deriving energy distributions for each amino-acid type

- All high resolution PDB structures (<1.5A) collected
- Distributions of residue energies calculated
- Energy Distribution for each amino acid derived
- Normalized energies derived

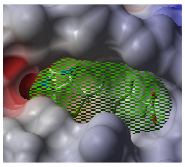


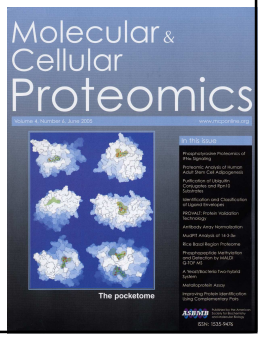
Calculating normalized residue energies for a model

- Calculate Z-score (normalized energy) for each residue
- Residues with E_{norm} > 5 are probably wrong

Maiorov, Abagyan, 1998 Proteins, The 1000-times faster version: 2004

The Pocketome





An, Totrov, Abagyan. (2005) Mol. Cell Proteomics,

***Pocketome**: Comprehensive Identification & Classification of Ligand Binding Envelopes*

Detecting Small Molecule Pockets from Structure

The problem

- We do not know the nature of the ligand
- Find the location and the extent / envelope of a pocket
- The Lennart Jones potential is short-range and **does not** predict the location of the small molecule site.

A Physical Idea:

- The CUMULATIVE potential integrated over a typical size of a ligand may predict the site location and extent

Detecting Binding Pockets

Challenge: Predicting Ligand Binding Sites *Without Knowing the Ligand*

Method:

1.Calculate this potential

$$P^0(\vec{r}) = \sum_a \frac{A_{aC}}{r_{ag}^{12}} - \frac{B_{aC}}{r_{ag}^{6}}$$

$$P(\vec{r}) = \int e^{-((\vec{\rho}-\vec{r})/\lambda)^2} P^0(\vec{\rho})$$

QuickTime™ and a YUV420 codec decompressor are needed to see this picture.

$$\lambda = 2.6\,\text{A}$$

2. Contour the potential
3. Filter out small blobs

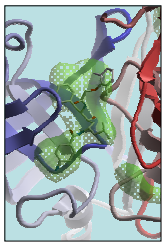
Pocket Database

17,000 pockets

Example: Biotin-binding protein (2izi)

Benchmarking the Pocket Prediction Algorithm

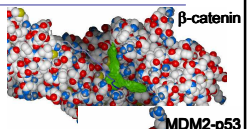
- 95% of 11535 pockets in apo structures overlap >50% with a predicted pocket. (96.8% out of 5656 complexed entries)
- In 82.3% of apo-cases the predicted pocket covers > 80% of the ligand contact atoms!

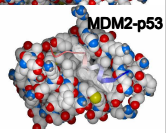


Binding Site Prediction: Conclusions

Pockets can be used to :

- Identify allosteric sites and alternative druggable pockets



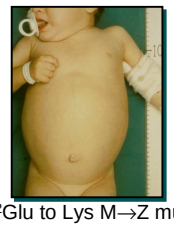


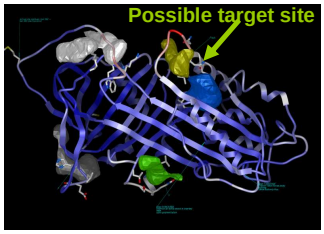
- De-orphanize (pre-docking): Identification of ligand binding potential and site location for orphan receptors

- Evaluate druggability of protein-protein interaction inhibition by applying the icmPocketFinder to separated protein subunits and evaluating the "pocket" strength

α_1-Antitrypsin deficiency and pathological aggregation

Collaboration with David Lomas, Cambridge



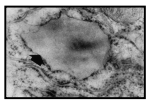


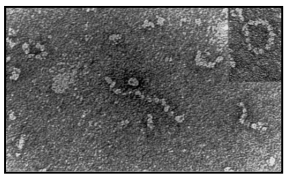
- ^{342}Glu to Lys M→Z mutant
- 1:1700 of North European Caucasians
- Risk of death from liver disease during childhood is 2-3%
- Low plasma α_1-antitrypsin level (10-15, 85% retained in liver %), emphysema and higher risk of lung cancer

Z α_1-antitrypsin: finding a polymerization inhibitor

Collaboration with David Lomas, Cambridge

α_1-antitrypsin is retained in ER and forms polymers *in vivo*





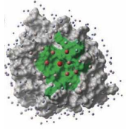
Lomas *et al*, *Nature* 1992; 357: 605-607

Lomas *et al, J.Biol.Chem.* 1993; 268: 15333-15335

Predicting

Protein Interfaces

interface location
oligomeric state
orphan interfaces
membrane interface



The problem of predicting transient interfaces

- Proteins do not have open hydrophobic surfaces
- Previous efforts that looked are residue frequences were not sufficiently predictive
- We do not know the partner to look for complementarity

A physical idea: Desolvation & entropy

- The transient interaction patch may have lower desolvation energy and lower entropy loss upon association.
- Both terms can evaluated via atomic surface areas
  (Eisenberg&McLachlan, 1986, Abagyan, Totrov, 1994)

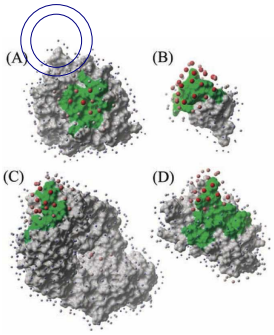
Optimal Docking Area: A New Method for Predicting Protein–Protein Interaction Sites

Juan Fernandez-Recio,[1] Max Totrov,[2] Constantin Skorodumov,[2] and Ruben Abagyan[1]* **Proteins, 2004**

- ODA identifies contiguous patches
- Atomic desolvation energy

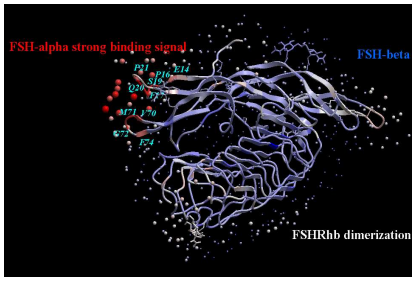
$$E_{desolv} = -\sum_i \sigma_i ASA_i$$

- Optimized on 66 complexes using protein docking results to include other physical components, like entropy
- Located 80% of the interfaces
- Larger study with 1568 complexes



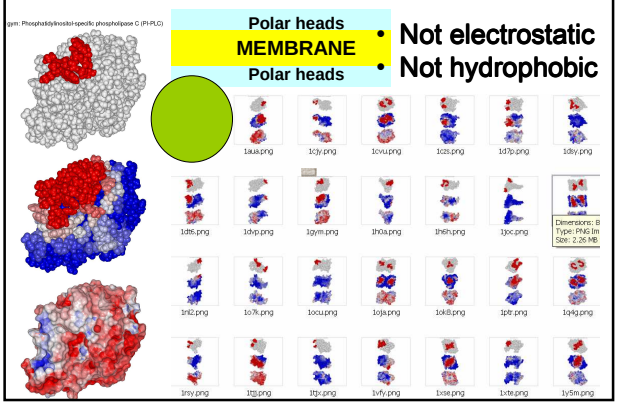
Bordner, Abagyan, Proteins (2005) 60, 353

Predicting Interaction Propensity for FSHR-FSH comples



Predicting transient protein-membrane interfaces

Irina Kufareva, Collaboration with the Overduin Group

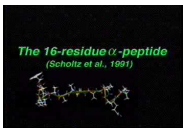


- Not electrostatic
- Not hydrophobic

Transient Protein Interactors Like the Membrane

Structure Prediction

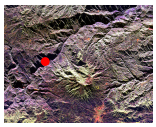
Conformational Searching for the Global Free Energy Minimum in Internal Coordinates (ICM)



Energy optimization.
ICM : Internal Coordinate Mechanics

Goal: Find the global minimum of

$\Delta G_{free} = \Delta E_{vacuum} + \Delta G_{solv,S}$



Representation, Energy, Derivatives, Eqs. of Motion

Charmm,CFF,..	$\{x,y,z\}$	*1970 s, 1981*
ECEPP	$\{\phi\}$,	*1975, 1983*
ICM	$\{\phi,\Phi,\alpha,b\}$	*1985, 1989, 1994*

Vacuum Energy

$$\Delta E_{vacuum} = \sum_{vw,el}\left(\frac{A_{ij}}{d_{ij}^{12}} - \frac{cB_{ij}}{d_{ij}^{6}} + \frac{332q_iq_j}{\varepsilon d_{ij}}\right) + \sum_{h-bonds}\left(\frac{A'_{ij}}{d_{ij}^{12}} - \frac{D_{ij}}{d_{ij}^{10}}\right) + \sum_{torsions} U_m \cos(n\phi - \phi_0)$$

Electrostatic Solvation and Entropy

$$E_{solv} = \sum_i \sigma_i a_j \quad \text{or} \quad \Delta E^{electrostatic}$$ **Poisson(/Boltzmann)**

$$E_{entropy} = -RT \sum_{residue\,\rho} \Delta S_\rho^{max} a_\rho / a_\rho^{max}$$

Main ICM References:

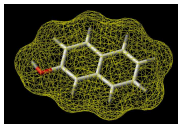
Abagyan, Mazur (1989) JBSD

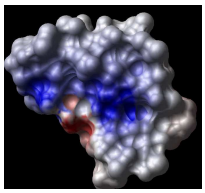
Abagyan et al. (1994) "ICM - a new method for protein modeling.." J. Comp. Chem. 15, 488-506

Abagyan, and Totrov, (1994)."Biased Probability Monte Carlo searches …" J. Mol. Biol. 235, 983-1002

Implicit Solvation Models in ICM

1. Atomic Surface model (*Eisenberg McLachlan, 1986*)
2. MIMEL: mirror image approximation for proteins *(JMB, 1994)*
3. Generalized Born approximation
4. REBEL: Boundary Element Algorithm
Poisson Equation:





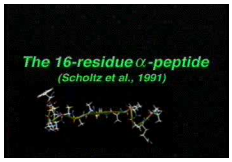
Use fast contour building and triangulation algorithm with grouping of surface triangles into atomic patches.
(Totrov, Abagyan, Biopolymers, 2002 - REBEL)
Totrov, Abagyan, J.Str.Bio. 1996 - Contour Build-up Algorithm for the analytical Connolly surface construction

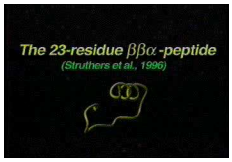
ICM Stochastic Global Optimization

- **Full atom, selected internal coordinates for the area of interest**
- **Gradient local minimization after random moves**
- **Optimally biased, designed, continuous group moves:**

- **Double energy scheme** $P(\theta^{group}) \sim \sqrt{F^{obs}(\theta)}$
- **Reactive history mechanism, stack**

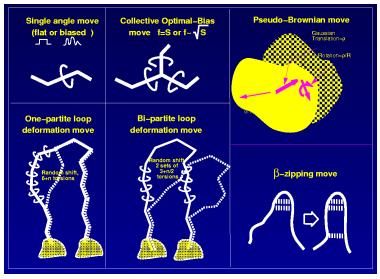
Not simulated annealing (T=const), *Not* Monte Carlo **(RHM, no local balance)**





Abagyan, Totrov, JMB 1994, JCP, 1999
Zhou, Abagyan, JCP, 1999

Collective moves for ligand optimization, protein structure prediction and docking in ICM



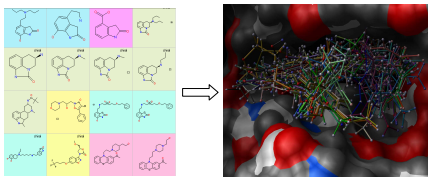
Automated Homology Modeling for Docking with ICM

- Find close template(s)
- Align sequence to the template
- Copy the aligned backbone
- Predict side chains
- Predict loops
  - Best Db fragment
  - Explicit ICM-local SGO
  - Grid simulation
- Refine by ICM SGO
- Predict local reliability (B_i)
- Validate by docking a known ligand if possible

The only input for ICM-homology builder: a sequence and a template structure

Side-Chain prediction: JMB 1993
Proteins, 1995, 1997
Marsden, Abagyan, submitted, 2003

Compound Docking

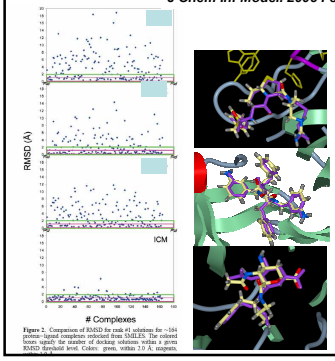


- **Fast docking: atomic ligand to the grid potentials of the receptor**
- **Method: stochastic global optimization in internal coordinates**
- **ICM performance from multiple benchmarks: 60-90% poses are correct**
- **Speed: takes 20 seconds per compound per processor.**

On Evaluating Molecular-Docking Methods for Pose Prediction and Enrichment Factors

Hongming Chen,*,† Paul D. Lyne,‡ Fabrizio Giordanetto,§ Timothy Lovell,*,†,§ and Jin Li†
GDECS Computational Chemistry, AstraZeneca R&D, Mölndal, Sweden, Cancer Discovery, AstraZeneca R&D, Boston, Massachusetts, and Medicinal Chemistry, AstraZeneca R&D, Mölndal, Sweden

J Chem Inf Model. 2006 Feb;46(1):401-15.



- 1bhx – alpha thrombin
- Rmsd 1.18 (0.71), rank 1
- 1dmp – HIV protease
  Rmsd 0.90, rank 1
- 1bji – neuraminidase
  Rmsd 0.89, rank 1

ICM Binding Score

A COMPROMISE between physics and errors

Coordinate errors due to *induced fit,* charge errors, docking errors, etc.

$$S_{binding} = \Delta E_{VW\,\mathrm{int}} + \Delta E_{ligStrain} + T\Delta S_{tor} + \alpha_1 \Delta \mathrm{E}_{\mathrm{HBond}} + \alpha_2 \Delta \mathrm{E}_{\mathrm{HBDesol}} + \alpha_3 \Delta \mathrm{E}_{\mathrm{SolEl}} + \alpha_4 \Delta \mathrm{E}_{\mathrm{HPhob}} + \alpha_5 \mathrm{Q}_{\mathrm{Size}}$$

α_{1-5} were optimized on a benchmark

- Van der Waals truncated at 4kcal/mole
- Hbonds calculation is based on lone pairs
- Penalty for desolvated hydrogen bonding donors/acceptors
- Electrostatics by Poisson equation (boundary element)

Preparing pdb compounds for docking

Problem1: compounds/ligands in PDB are not suitable for automated conversion. They lack bond types, formal charges and chirality flags.

Problem 2: compound databases contain only 2D drawings. They need to be converted to 3D.

To fix a PDB ligand follow these steps:

- Assign correct bond order manually
- Assign correct formal charges manually
- Assign chirality if necessary (less validated)
- Save is as a mol file or Run the conversion tool

The conversion tool performs these steps:

- Adds hydrogen according for elements, bond orders and formal charges
- Runs ICM MMFF atom type assignment routine
- Assigns partial electrostatic charges
- Assigns rotatable torsions
- Creates a 3D model by full MMFF94 energy optimization

Preparing compound database for screening

Background:
Preparation of the compound database depends on software used. Some software requires rigid conformations pre-generated. Some will generate 3D structures of ligands and sample them on the fly.

Typically, some kind of index is required to speed up access to the compounds in a very large compound file.

ICM just needs a mol/sdf file with correct drawings

Each molecule from a database will be converted on the fly and flexibly docked into a pocket. If the score is lower than a predefined threshold, it will be retained in the "answers" file.

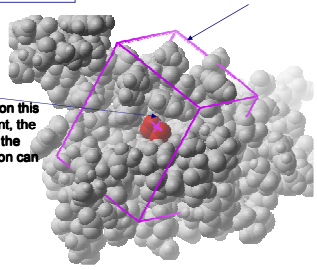
Things to decide:

1) To keep (or not) the carboxyls neutral
2) To charge or not the amino/imidazole groups
3) Filters (rotatable bonds, donors, acceptors, mass, etc.)

Choosing a grid box and a probe for intial placement

Background:
The docking procedure needs the force field precalculated as grid potentials.
Also, one need to define the initial placement of the ligand.
Both decisions can be semi-automated.

Energy maps (or grid potentials) will be calculated inside this box.



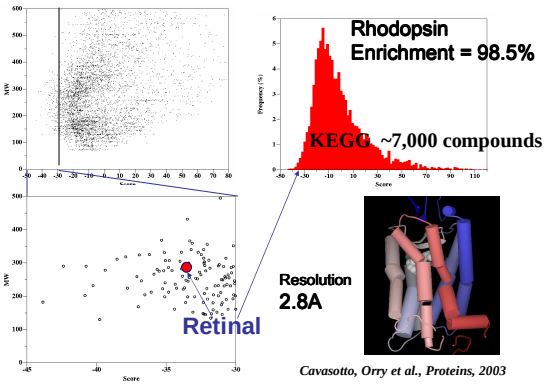
Initial orientations will be based on this probe. If sampling is sufficient, the answer does not depend on the initial position. A good position can make the search shorter.

De-orphanization

in silico with the 'native' coordinates

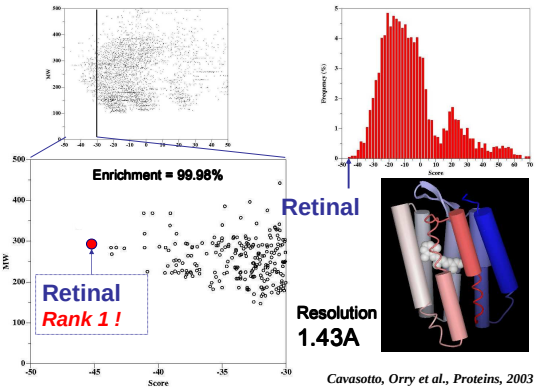
Q: *given an empty pocket and the metabolome, can we identify the native substrate in-silico?*

De-orphanizing a GPCR by docking



Cavasotto, Orry et al., Proteins, 2003

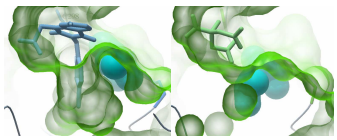
Bacteriorhodopsin fdVLS and ranking



Cavasotto, Orry et al., Proteins, 2003

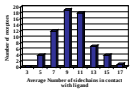
Receptor Flexibility

upon ligand binding



Receptor flexibility statistics

- 1132 PDB complexes of **65** receptors with > 5 different ligands each analyzed



Sidechains

- A ligand contacts with ~ 10 side chains
- ~75% ligand contact atoms are s.c. (vs 50% in protein core)
- 3 s.c. in 85% of receptors will move by > 1.5A
- But only 14% severe clashes with 1s.c. and 3% with > 1 s.c.

Backbone

- ~ 30% receptors had substantial backbone movements: >1A backbone deviations leading to ligand clashes
- 8 elastic deformations, 8 loop, 1 secondary structure

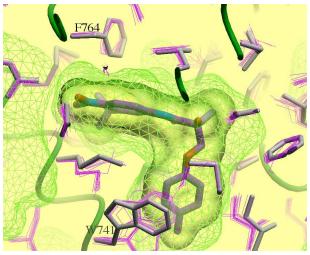
Totrov, Barcelona 2006

Evaluating side-chain flexibility

- Identify the side-chains of interest
- Perform an ICM simulation (~15min)
- Cluster and space-filter (retain best Ei)
- Evaluate Boltzmann-weighted RMSD for each sidechain atom

$\langle D^2_i * \exp(-\Delta E_i/kT) \rangle$

- ICM Flexibility tool



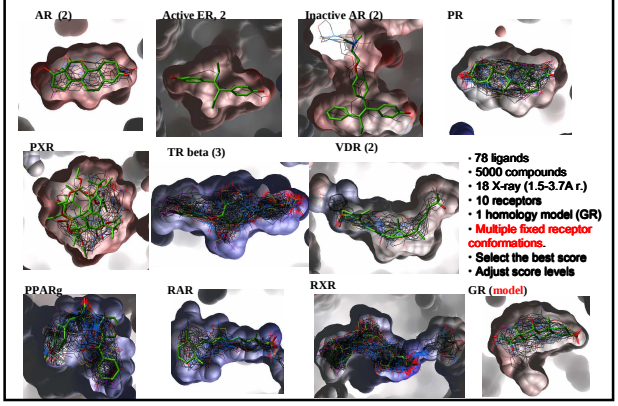
Androgen receptor

Pocket Conformations

Representing receptor by multiple static conformations

Nuclear Receptors: Predicting Specificity

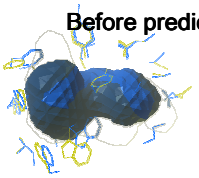
Schapira, Abagyan, Totrov. J. Med. Chem. 2003

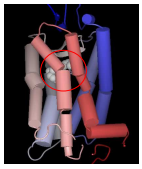


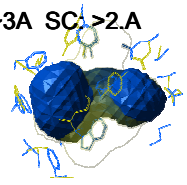
- 78 ligands
- 5000 compounds
- 18 X-ray (1.5-3.7A r.)
- 10 receptors
- 1 homology model (GR)
- Multiple fixed receptor conformations.
- Select the best score
- Adjust score levels

Rhodopsin: pocket-flexible docking

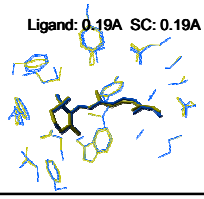
Before prediction: Ligand: ~3A SC: >2.A



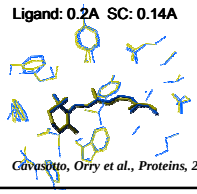




Ligand: 0.19A SC: 0.19A

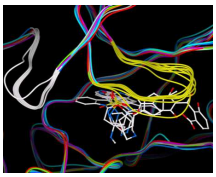


Ligand: 0.2A SC: 0.14A



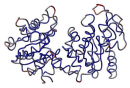
Cavasotto, Orry et al., Proteins, 2003

Predicting Larger Backbone Deformations: Normal Modes and Hinge areas



Cα-Cα spring strengths:

$$C_{ij} = \left(\frac{r_0}{r_{ij}}\right)^6 + a\, s_{ij}$$



Derive

$\mathbf{u}^n$ = normal modes

ω_n = vibrational frequencies

Deformability and hinge regions:

$$S_{\mathbf{u}} = \mathrm{sym}(\nabla \mathbf{u}) - \tfrac{1}{3}\,\mathrm{div}\,\mathbf{u}\, I$$

$$d_{\mathbf{u}} = \|S_{\mathbf{u}}\|$$

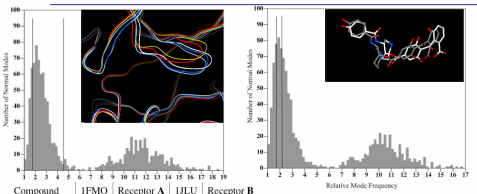
$$d_M^2 = \sum_{n=7}^{3N} \left(\frac{d_{\mathbf{u}^n}}{\omega_n}\right)^2$$

Idea & Protocol

- Soft, low-res and smooth harmonic model of residue interactions (atomic model does not work)
- Find normal modes U
- Derive deformability

Kovacs, Chacon, Abagyan, Proteins, 2004

Choosing Relevant Normal Modes



Compound	1FMO	Receptor A	1JLU	Receptor B
Adenosine	0.4	0.6	0.7	4.0
Balanol	8.5	1.1	1.6	4.0
Staurosporine	9.8	7.0	10.1	0.8
H7	0.8	0.8	0.5	3.4
H8	0.8	1.1	0.8	3.5
H89	9.2	1.1	10.2	1.5

- **Very few normal modes are needed for docking (< 10)**
- **These modes are NOT the lowest frequency modes !**
- **The small number of relevant modes can be combined**

Cavasotto, Kovacs, Abagyan, JPC, 2006

Mutants and Mutations

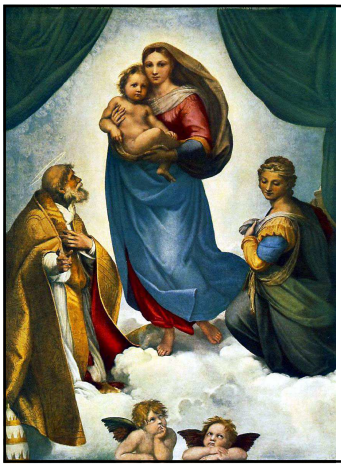
"Portrait of a Girl Covered in Hair"

By Lavinia Fontana (1552-1614)

We are all different at 0.1% level (almost every protein has one amino acid different)

8% of liveborns will suffer from a genetically based disorder by age 25

Spontaneous mutations occur continuously (smoking, tanning, eating, age)



Geometry, stability and functional effects of single point mutations

Growing volume of **SNP** and Pharmacogenetics data

Predicting the effect on

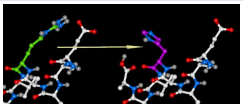
- geometry and dynamics
- stability changes
- bio-function and binding
- drug binding

"The Sistine Madonna"

by Rafael (1513)

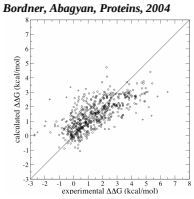
Look at Pope Sixtus IV

Predicting energy and geometry of mutants



▪ The largest database of 2141 ordered pairs of structures with a mutation compiled
▪ A filtered training set of 1816/2 mutants including "*small-to-big*"with Protherm ΔΔGs.
▪ Cross-Prediction accuracy on the second half ~ 1.1 kcal/mole (correlation=0.66)
▪ Regression from a subset of 317 mutants gives the same prediction accuracy.
▪ 20 unfolded state energies derived for each residue
▪ Terms: van der Waals, electrostatics, hydrogen bonds, solvation, entropy, residue type constant.

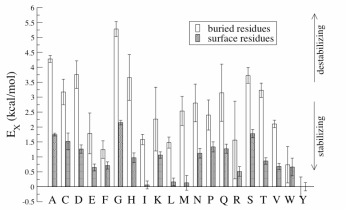
Bordner, Abagyan, Proteins, 2004



$$\Delta E_{X \to Y} = \sum_{terms} w_t \Delta E^t{}_{X \to Y} + (E^0{}_X - E^0{}_Y)$$

Stability prediction without structure

- Fit simple energy function $\Delta\Delta G=E_{X'}-E_X$ for the mutation X-X' to the entire data set without outliers (1768 values).
- Buried residues: r=0.71 (std=1.21 kc/m); surface res.: r=0.55 (std=1.14 kc/m);
- Only includes residue energies: useful when no structure is available.
- Residues with small side chains (*glycine, serine, and alanine*) most destabilizing
- Most stabilizing residues are *tyrosine, isoleucine and leucine*. Agrees with their high occurrence frequency in β sheets.
- Also separately fit parameters for buried and surface residues
- Mutation from *Lys* to *Arg* stabilize protein by 0.5-1 kcal/mole

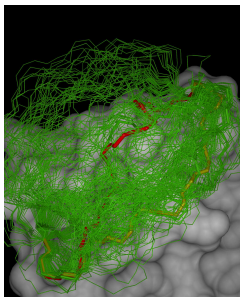


Loop Prediction

QuickTime™ and a
YUV420 codec decompressor
are needed to see this picture.

Predicting and redesigning the 15 residues of the triosephosphate isomerase backbone to 8-res. loop
Collaboration with the Wierenga group
Structure, PNAS, Prot. Eng. 1993-2002

Predicting Short Loops (benchmark)



Loop Benchmarks: Fiser & Sali, 2003 for 4,8 and 12-residue loops. Friesner et al.

• Challenge: prediction methods break down at 8-12 residue loops.

• 10 years of CASP did not result in X-ray quality loop prediction (NIH)

• Idea: loop is a separate chain with *loose closure conditions*.

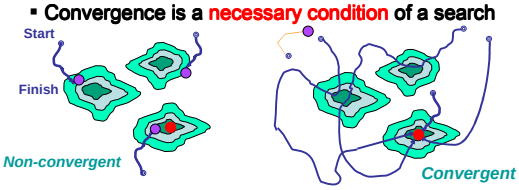
• Randomized starting conformation

• Run from 5 starts to ***convergence***

• A homology loop benchmark was also compiled and tested

Convergence and Freedom

- Convergence is a necessary condition of a search

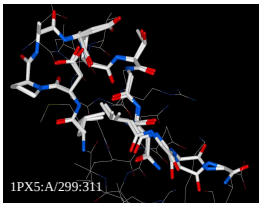


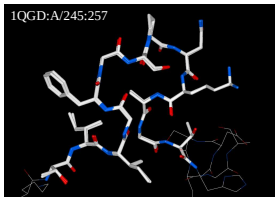
- Set them free.. Departing from a strict loop closure search

12-residue loops predicted by the ICM optimization after convergence

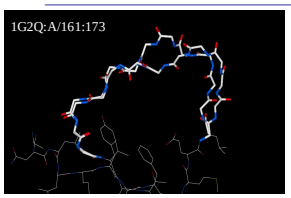
In most cases the prediction is virtually identical to the crystal structure!

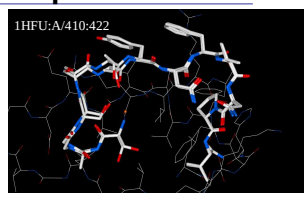
An, Totrov, Abagyan, 2006

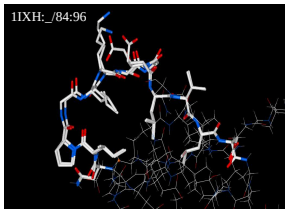


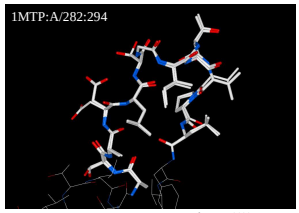


More 12 residue Loops Predictions

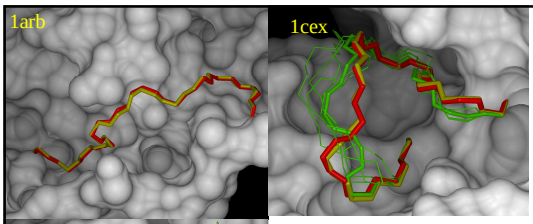








An, Totrov, Abagyan, 2006



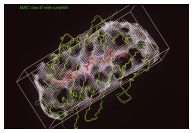
States within 10 kc/m from the lowest E

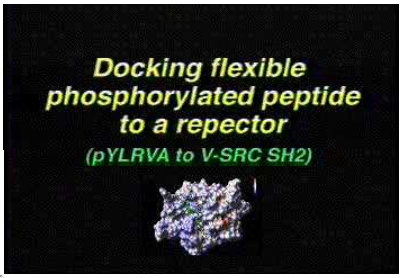
How unique is the crystallographic state?

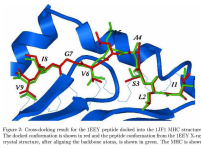
We observed a large diversity of types of low-energy conformational ensembles.

Red:Native Yellow: The Prediction Others:within 10kcal/mol

Predicting peptide protein association







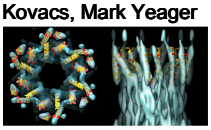
End-guided docking of 27 peptides (8-9) to the HLA receptors, including homology models.
Bordner, Abagyan, 2006

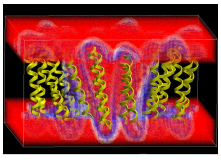
Ab initio docking to a receptor of 24 peptides to SH2 and PTB domains
Zhou et al. 1998, Folding&Design, 3, 513

EM-guided Atomic Models

Julio Kovacs, Mark Yeager

- Full atom global energy + densityFit optimization. Flexible backbones
- Sampling strategy combines systematic grid and overlapping stochastic searches
- Solvation models with specific geometry built through solvation maps.
- Benchmark reconstitutions for KcsA tetramer and MscL pentamer show about 1 to 2A RMSD for the contact residues.

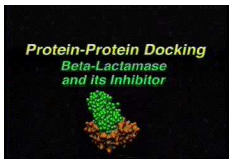




Protein Docking

Procedure:

- **Both receptor and ligand are presented by atomic models**
- **Convergent Multistart ICM Stochastic Energy optimization with pseudo-Brownian moves (JMB, JCC, 1994) and side-chain minization**
- **Explicit simulaneous global optimization side-chain and 6 positional variables of candidate solutions**



Benchmarks

GCN4 ab initio helix docking (*JCC, 1994*)
Lysozyme-Antibody (*Nature SB, 1994*)

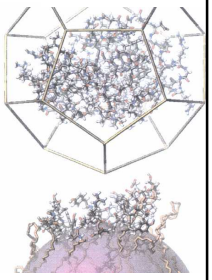
Competitions. **Docking challenge** (*Nature SB 1995,96*)
CAPRI Rounds 1:5

Local Minimization

Nature, SB, 1994

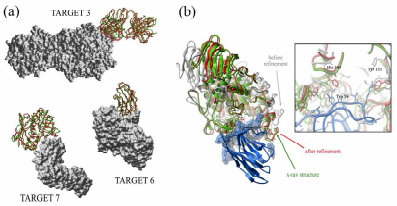
Detailed *ab initio* prediction of lysozyme-antibody complex with 1.6 Å accuracy

Maxim Totrov and Ruben Abagyan



The fundamental event in biological assembly is association of two biologic l macromolecules. Here we present a successful, accurate *ab initio* predictio of binding of uncomplexed lysozyme to the HyHel5 antibody. The predictio combines pseudo Brownian Monte Carlo minimization with a biased-probabilit global side-chain placemer+ procedure. It was effected in an all-atom representation, with ECEPP 2 potentials complemented with the surface energy side-chain entropy and electrostatic polarization free energy. The near-native solution found was surprisingly close to the crystallographic structure (root-mean-square deviation of 1.57 Å for all backbone atoms of lysozyme) and had considerably lower energy (by 20 kcal mol^{-1}) than any other solution.

CAPRI Round 1-2 results



- **3 out of 7 targets predicted correctly**
- **Refinement of Target 6 Dramatically improves the near-native solution**

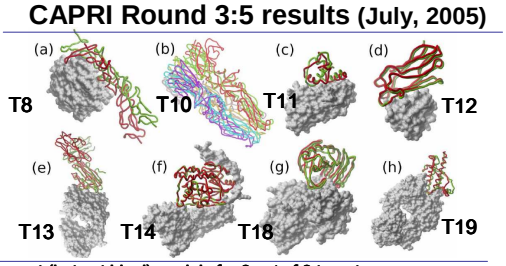
64 R. MENDEZ ET AL.

TABLE IV. Summary of Docking Predictions

Predictor group	T51	T52	T53	T54	T55	T56	T57	Predictor summary
Scripps US (Abagyan)	0	0	**	0	0	***	**	3/2**/1***
Boston U. US (Camacho)	*	0	0	0	0	***	***	3/2***
Weizmann Inst. IL (Eisenstein)	*	*	0	0	0	0	***	3/1***
Imperial Coll. UK (Sternberg)	0	*	0	0	0	***	*	3/1***
UCSD, US (Torda/Lyck)	*	*	0	0	0	**	0	3/1**
Sheffield U. UK (Gerstein)	*	*	—	—	—	—	—	2
U. Washington US (Gray/Baker)	0	0	0	0	0	**	***	2/1**/1***
U. Lisbon,Pt (Palma/Krippahl)	—	0	—	0	0	**	*	2/1**
Aberdeen U. UK, UK (Ritchie)	0	0	**	0	0	***	0	2/1**/1***
Boston U. US (Wong)	0	**	0	0	0	0	**	2/2**
TAU IL-NIH US	*	0	0	0	0	0	***	2/1***

Proteins, Mendez, ..Wodak, 52, *51-53, 2003*

CAPRI Round 3:5 results (July, 2005)



- Decent (but not ideal) models for 8 out of 9 targets
- 64-71% of native contacts
- 0.4-1A interface RMSD for the best cases
- For T14, Rmsd 0.6A, Rank 1 by energy
- Successfully used new scoring function for T14, T18 & T19
- T19: antibody - prion. Used no CDR bias + NMR model for prion.

Fernandez-Recio, Totrov (2005) Proteins , 60, 308

- One failure: T9 with large hinge-bending movements,

PROTEINS: Structure, Function, and Bioinformatics 60:150–169 (2005)

Assessment of CAPRI Predictions in Rounds 3–5 Shows Progress in Docking Procedures

Raúl Méndez,[1,2] Raphaël Leplae,[1] Marc F. Lensink,[1] and Shoshana J. Wodak[1,3]*

TABLE II. Summary of Docking Predictions

Predictor group	T08	T09	T10	T11	T12	T13	T14	T18	T19	Predictor summary
Abagyan	**	0	*	**	***	*	***	**	**	8/4***/2***
Wolfson	**	*	*	*	*	0	**	**	*	8/3**
Weng	**	0	0	*	***	***	***	**	**	7/3***/3***
Bates	*	0	*	**	*	0	**	**	*	7/3**
Baker	—	0	0	**	***	**	***	0	***	6/2**/4***
Camacho	**	0	0	0	***	***	**	**	*	6/0**/2***
Gray	***	—	—	**	***	0	0	0	**	5/2**/3***
Bonvin	—	—	**	**	0	***	***	0	0	5/3**/2***
ClusPro	**	0	0	0	***	*	0	0	*	5/2**/1***
Sternberg	**	0	0	*	*	0	**	0	*	5/2**
Eisenstein	***	0	0	*	***	0	**	0	0	4/1**/2***
Ritchie	0	0	0	**	***	*	*	0	0	4/1**/1***
Zhou	—	—	0	—	***	**	*	*	0	4/1**/1***
Ten Eyck	0	0	0	0	***	***	**	0	0	3/1**/2***
Zacharias	**	0	—	—	—	—	***	0	**	3/2**/1***
Valencia	*	0	0	*	*	—	0	0	—	3
Vakser	—	—	0	—	—	—	**	**	0	2/2**
Umeyama	0	0	0	**	*	0	0	0	0	2/1**
Kamesnis	—	—	0	0	***	0	0	0	0	1/1***
Fass	—	—	0	*	0	0	0	0	0	1
Gottschalk	—	—	—	*	—	—	—	—	—	1
Palma	0	0	0	—	0	0	0	0	0	1
Poupon	—	—	—	—	0	*	0	0	0	1
Wang	0	0	0	*	0	0	0	0	0	1
Target summary	11/7**/2***	1	4/1**	15/7**	16/11***	10/2**/4***	14/7**/5***	2/0**	10/4**/1***	

This table summarizes the results obtained by all the groups that submitted one or more predictions of acceptable quality or better for at least one

Summary

- Accurate cross-docking to receptors represented by ‘static’ grid potentials works in most cases.
- Receptor flexibility can be predicted in advance
- A combination of ligand based methods with receptor structure methods can help to de-orphanize receptors.
- Stochastic global optimization in internal coordinates is a powerful and general method for modeling membrane proteins.

Acknowledgements

Scripps Group Members

- **Julio Kovacs (normal modes,membrane proteins)**
- **Irina Kufareva (orphan protein interfaces)**
- **Adrian Saldanha (antitrypsin)**
- **William Bisson, Anton Cheltsov (AR inhibitors)**
- **Giovanni Bottegoni (receptor flexible docking)**

Molsoft (www.molsoft.com)

- **Maxim Totrov (ICM,Ligand Docking)**
- **Andrew Bordner (peptide docking, Mutations)**
- **Claudio Cavasotto (kinase docking, flexibility)**
- **Andrew Orry (GPCRs)**

Former Group Members

- **Juan Fernandez-Recio (Barcelona, protein docking)**
- **Matthieu Schapira (Lyon, TR, NR)**
- **Jianghong An (Vancouver, pockets, loops)**

Collaborators

AAT: **David Lomas, Meera Mallya and the team, Cambridge**

EM: **Mark Yeager, Scripps**

AR: **Patrick Sexton, Melbourne, Xiaokun Zhang, Burnham**

Membrane: **Michael Overduin Group**

Funding

NIH grant on protein docking