

ICM Protein Modeling

Updated on 5/29/2014.

Contents

Internal Coordinate Mechanics (ICM).....	1
Global Energy Optimization	2
Electrostatics.....	3
ICM Force-Fields	3
Homology Modeling.....	3
GPCR Modeling	4
GPCR Modeling Using Experimental Restraints.....	4
Loop Modeling	4
Modeling Protein Flexibility	5
Predicting the Effect of Mutation	6
Protein-Protein Docking.....	7
Prediction of Protein-Protein Interface Sites.....	7
Case Study: Structure-Guided Design of HIV-1 Immunogens.....	8
Case Study: Modeling Protein-Protein Interaction Sites.	8
Case Study: Modeling of GPCR Agonist Conformation.....	9
References	9

Internal Coordinate Mechanics (ICM)

One of the advantages of the ICM-Pro desktop modeling package is related to the coordinate method and optimization procedure implemented in the software. We use Internal Coordinate Mechanics (ICM) whereby four types of internal variable are considered: bond lengths (b), bond angles (ω), torsion angles (φ) and phase dihedral angles (ϕ) (Figure 1a). In normal conditions the bond lengths and planar angles are rigid and constant and only the torsion angle changes. This reduces the number of variables defining the conformation of the system compared to cartesians methods which have three variables (x,y, and z) per atom. Fewer variables improves the simulation convergence time at least 1000 fold without sacrificing accuracy¹. A tree-like graph is imposed on all atoms as well as some virtual atoms which link to other molecules in other branches of the tree (Figure 1b). Further reduction of free variable space and system complexity in ICM can be achieved by effectively freezing IC variables in more rigid or less important parts of the model. When carefully applied and validated, such complexity reductions reduce

unnecessary noise in the modeling system and enable faster and more reproducible energy optimizations.

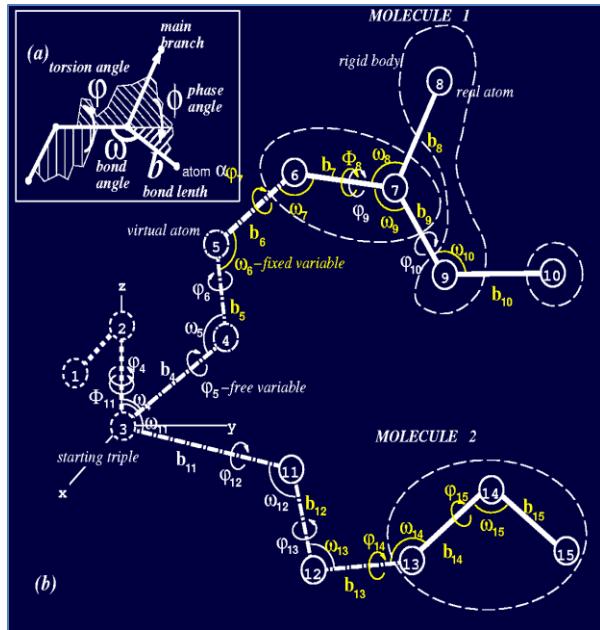


Figure 1 (a) The four types of internal variables in ICM. (b) The ICM tree depicting the geometry of two molecules using containing atoms, bonds and virtual atoms (dotted bonds). The image is adapted from ¹.

Global Energy Optimization

Simulations are supported by an accurate internal coordinate force field and a very efficient conformational state sampling algorithm called Biased Probability Monte Carlo (BPMC)². Any subset of internal variables can undergo BPMC. The method consists of the following steps:

1. One or several system variables are changed randomly. This can include a positional Pseudo-Brownian random move or internal torsion modification.
2. A local energy gradient minimization is then undertaken.
3. The energy of the system is evaluated. If the energy E_{new} is lower than E_{old} then the new conformation is always accepted and used in the next iteration. If not, the new conformation is accepted with a probability given by:
 $P_{\text{acc}} = \exp[-(E_{\text{new}} - E_{\text{old}})/kT]$ Where k is Boltzman's constant and T is the effective temperature of the simulation.
4. Return to Step 1.

The length of the simulation is based on adaptive heuristics which are based on the number of atoms in the molecules. Each random step is based on energy or statistical properties of conformational subspaces (e.g. ϕ - φ zones or torsion angles), the positions, sizes and preferences for high-probability zones were calculated for different residue types from representative structures from the PDB.

Electrostatics

Solvation is an important effect to consider when undertaking protein energy simulations. Many solvation methods are too computationally expensive to be used efficiently for protein simulation. MolSoft has developed a fast and accurate electrostatics method called REBEL (**R**apid **E**xact-**B**oundary **E**lectrostatics)³. The method solves the Poisson equation for a molecule without a grid and with exact positions of electric charges and is a powerful implementation of the boundary element method with analytical molecular surface as dielectric boundary. The energy calculated by this method consists of the intramolecular Coulomb energy and the solvation energy which can be analyzed separately.

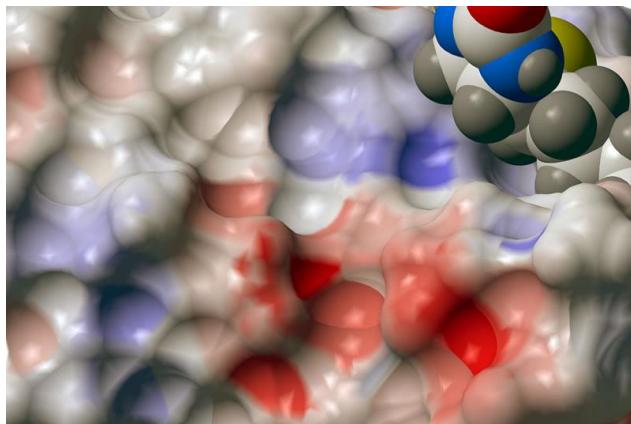


Figure 2 The REBEL feature calculates the electrostatic potential away from the surface of the graphics object and colors surface elements according to this potential from red to blue. The potential is calculated either by the REBEL boundary element solution of the Poisson equation, or, if option fast is specified, by a simple Coulomb formula. Used in conjunction with the electrostatic solvation component, REBEL allows you to determine the specific energy for binding. This includes determination of the electrostatic component of drug-receptor binding activity.

ICM Force-Fields

The ICMFF internal mechanics force field developed by MolSoft⁴ is used for protein modeling inside ICM other IC force fields such as ECEPP⁵ can also be used. The Merck Molecular Force Field is used for all small molecule chemical modeling^{6,7}.

ICMFF was parameterized using small molecule crystal data and quantum mechanics calculations and differs to the original ECEPP force field. The main differences include that the dielectric constant is parameterized to the condensed state rather than a vacuum, and there is an improved description of hydrogen bonds and backbone covalent geometry. ICMFF compares very well with other force-fields and it was tested on a challenging set of loop conformations (see Loop Modeling section).

Homology Modeling

ICM homology modeling requires an initial placement of the aligned polypeptide chain onto the template structure, the side-chain torsion angles are predicted by simultaneous global optimization of the energy for all non-identical residues. Methodology for conformational modeling of protein side chains and loops (see Loop Modeling section), relies on internal coordinate definition of the molecular object¹ combined with computationally efficient Biased Probability Monte Carlo (BPMC) optimization².

An extended force field includes surface terms, electrostatics with the boundary element solution of the Poisson equation³, side chain entropy terms, and a fast algorithm for calculating molecular surfaces⁸. The quality of the resulting 3D model is assessed by a specialized ICM procedure, which also predicts possible backbone deviations between the homologues⁹. The modeling method has demonstrated excellent accuracy in blind predictions at the CASP2 competition⁸ and in several protein engineering applications^{9,10}.

GPCR Modeling

In 2008, when the structure of the Adenosine receptor A2a G-Protein Coupled Receptor (GPCR) was solved in a complex with a drug, the crystallographers who solved the structure announced a challenge for all modelers to predict the binding interactions of the drug with the receptor. Two teams from Molsoft (using ICM-Pro + VLS), Katrich-Abagyan and Lam-Abagyan built the models that had the largest number of correct ligand-receptor interatomic contacts, 45 and 34 out of 70, respectively out of all participants. Moreover, both models were ranked number 1 in the set of submitted complexes^{13,14}. Ruben Abagyan's group at UCSD has organized subsequent GPCR modeling competition using ICM to evaluate the accuracy of the submitted models^{15,16}. The modeling method has been successful in finding new ligands for Melanin Concentrating Hormone¹⁷, adenosine receptor agonists¹⁸ and antagonists¹⁹, beta-2 adrenergic receptor^{20,21}.

GPCR Modeling Using Experimental Restraints

Although in recent years more atomic detail GPCRs has been published they still remain a challenging modeling target. This is particularly the case with Family B GPCRs where limited modeling templates are available. ICM has been used to build high quality models of Family B GPCRs using cysteine trapping²², photoaffinity labeling²³⁻²⁵ data. This experimental data can be used to incorporate distance restraints and tethers during the protein modeling procedure^{24,26-28}.

Loop Modeling

Accurate loop modeling is important for high quality predictions of protein-protein and protein-drug interactions. ICM has a good track record in predicting the conformation of loop regions. As an example, ICM loop modeling was used to design two new 7 residue loops in a monomeric variant of the dimeric trypanosomal enzyme triose phosphate isomerase. In both cases the designs were successful. The predicted conformations turned out to be correct (accuracy of 0.5A RMSD) after the crystallographic structures of the designed proteins were determined in Rik Wierenga's lab²⁹.

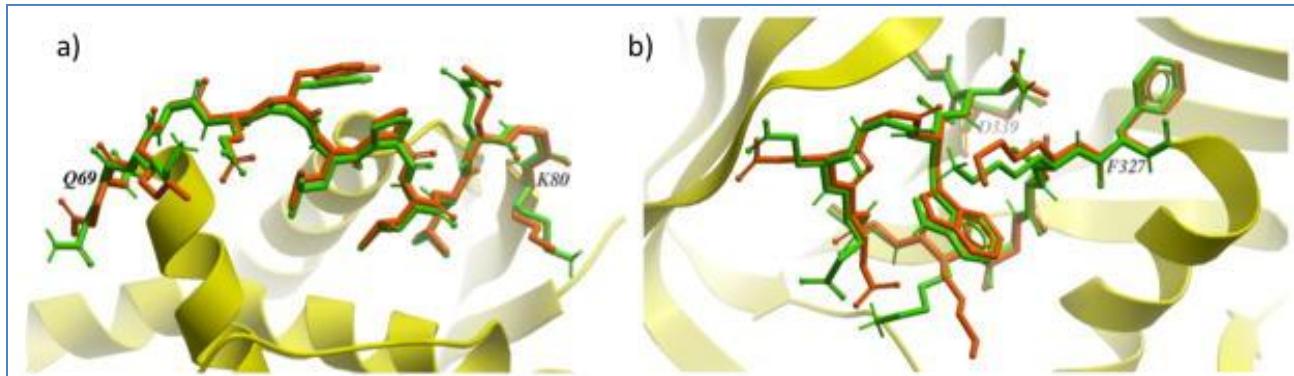


Figure 3 Comparison of modeled (green) loop with crystal structure (yellow) for loops with (a) 12 and (b) 13 residues. Image adapted from Arnautova et al 2011.

Recently, MolSoft developed a new physics-based internal coordinate mechanics force field (ICMFF) that was evaluated on a set of loops⁴ ranging from 4 to 13 residues in length. To the best of our knowledge the modeling results are the best reported for the benchmark used. The new force-field contains new parametrization for the dielectric constant, an improved hydrogen bond determination method, and implementation of novel backbone atom torsional potentials, which include bond angles of the carbon (alpha) atoms into the internal variable set. ICMFF and a solvent-accessible surface area solvation model was used to sample the loop using ICM BPMC³⁰. Average/median backbone root-mean-square deviations of the lowest energy conformations from the native structures were 0.25/0.21 Å for 4 residues loops, 0.84/0.46 Å for 8 residue loops, and 1.16/0.73 Å for 12 residue loops (Figure 3).

Modeling Protein Flexibility

Another modeling aspect that ICM addresses is the importance of taking into account ligand induced fit. MolSoft has been the leader in developing ligand guided modeling methods such as Alibero³¹ and the development of resources such as the Pocketome database^{32,33}. MolSoft has developed a suite of methods to tackle various protein flexibility problems:

- **Multiple Receptor Conformations** – this approach uses an ensemble of predicted (generated using ICM Biased Probability Monte Carlo) and/or experimental structures to represent the flexible nature of the binding pocket^{32,34,35}.
- **Fumigation** – this technique can generate a more “druggable” pocket by sampling the side-chains of the receptor in the presence of a repulsive density representing a generic ligand. As an example, this technique was used to discover small molecules disrupting the subunit interaction of the protein kinase CK2^{33,36,37}.
- **Normal Modes** – this method provides a spring-like representation of the pocket backbone atoms allowing a wide conformational space to be sampled^{38,39}. An all heavy atom ICM Elastic Network NM modeling approach was successfully used in the 2008 “blind” G-Protein Coupled Receptor (GPCR) modeling competition. The method yielded the best model in terms of ligand-receptor contacts for the Adenosine A2a receptor^{13,14}.

- **Ligand Guided Modeling** - this model uses a fully flexible seed ligand, which is known to bind to the receptor, docked to the protein and the pocket side-chain, and in some cases backbone atoms are sampled and optimized. This approach generates an ensemble of structures, which can be clustered and filtered down to a few selected conformations. The ability of the model to be able to discriminate binders from non-binders is then tested by screening a database of decoy ligands mixed with known binders^{17,40,41}.

Predicting the Effect of Mutation

MolSoft has developed methods to predict the effect of mutation on binding and stability (*publication in preparation*).

The binding free energy change, ΔG_{bind} , is computed as a difference between the free energy of mutant and wild type:

$$\Delta \Delta G_{bind} = \Delta G_{bind}^{mut} - \Delta G_{bind}^{wt}, \text{ where}$$

$$\Delta G_{bind} = (E_{intra}^{comp} - E_{intra}^{parts}) + (\Delta G_{solv}^{comp} - \Delta G_{solv}^{parts})$$

The energy is calculated for fixed backbone and all the side chains except those in the vicinity of the mutable residue. BPMC simulations are carried out to relieve possible atomic clashes created as a result of mutations to larger amino acid residues. "Scan Protein Interface" option allows to mutate all residues (one by one) of the Interacting Part in close contact with the second part of the complex.

The free energy change in protein stability is computed as follows:

$$\Delta \Delta G = \Delta G^{mutant} - \Delta G^{wt}$$

$$\Delta G = \Delta G_{folded} - \Delta G_{unfolded}$$

The free energy of the unfolded and misfolded states is approximated by a sum of the residue-specific energies. The residue-specific energies were derived empirically using a large set of experimental data. Mutation of a given residue is followed by BPMC simulations with flexible side chains for the mutated residue and its neighboring residues. The rest of the protein structure is considered rigid.

Protein-Protein Docking

The ICM protein-protein docking methodology uses a pseudo-Brownian method¹ with a biased probability global optimization (BPMC³⁰) procedure in the ICM or ECEPP⁵ force field. In 1994, the first application of the method was reported resulting in a detailed *ab initio* prediction of lysozyme antibody to within 1.6 Å of the crystal structure⁴². The top energy docked pose had considerably lower energy (by 20 kcal mol⁻¹) than any other solution. Following this success MolSoft competed in a blind protein-protein docking prediction contest with similar good results although the speed of the calculations were becoming a hindrance to larger systems⁴³. To speed up the calculations grid potential maps were introduced for the receptor (or larger binding partner), the resulting complexes were then refined and scored⁴⁴. In 2003 and 2005 ICM performed very well in the CAPRI Protein-Protein docking competitions⁴⁵, a new scoring method incorporating Accessible Surface Area solvation parameters improved the ranking of the docked complexes (Figure 4).

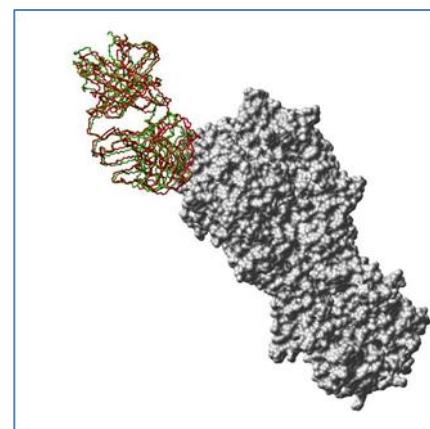


Figure 3 Docked complex of hemagglutinin / Fab antibody submitted in the 2003 CAPRI protein-protein docking completion with <2Å deviation from the crystal structure.

Prediction of Protein-Protein Interface Sites

To help focus the protein-protein docking to a specific patch on the surface of the receptor and/or ligand a method called Optimal Docking Areas (ODA) has been developed (Figure 4). The method identifies optimal surface patches with the lowest docking desolvation energy values as calculated by atomic solvation parameters (ASP) derived from octanol/water transfer experiments and adjusted for protein-protein docking. The predictor has been benchmarked on 66 non-homologous unbound structures, and the identified interaction points (top 10 ODA hot-spots) are correctly located in 70% of the cases (80% if we disregard NMR structures).

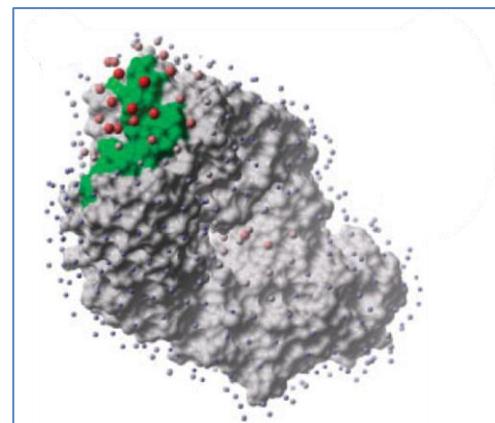


Figure 4 Accurate ODA prediction of Fab-lysozyme interface. Red are the lowest energy values and larger spheres indicate stronger prediction.

The ODA method correctly predicted Fab-lysozyme interface (PDB 1MLC). Figure 4 shows the surface points around the unbound antibody colored according to the energy values of their respective ODAs and the size of the ODA points is proportional to their energy values. The surface of the unbound molecule is represented in white and known interface residues (i.e. residues within 5 Å from any non-hydrogen atom of a protein partner in a known complex) are shown in green.

Case Study: Structure-Guided Design of HIV-1 Immunogens.

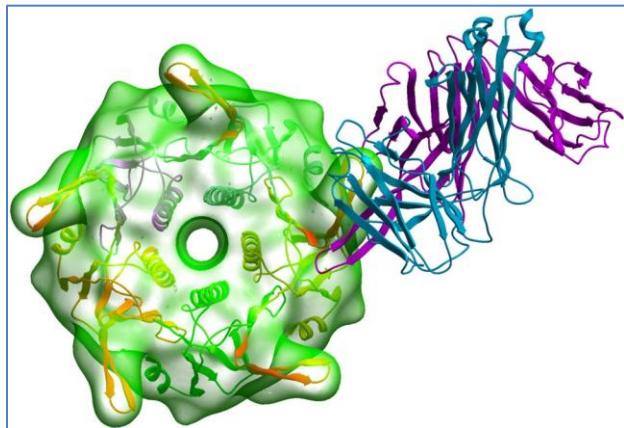


Figure 5 Model of the short V3-CTB construct (ribbon and transparent surface) in complex with the Fab fragment of mAb 447-52D (magenta and blue ribbons for heavy and light chains, respectively).

sera that potently neutralize multiple HIV virus strains.

Case Study: Modeling Protein-Protein Interaction Sites.

Here we describe results from collaboration between Prof. Ruben Abagyan (UCSD, MolSoft co-founder) and Prof. Stephen Tomlinson (Medical University, South Carolina).

CD59 is a glycoprotein that inhibits the formation of the Membrane Attack Complex (MAC) of complement host cells. Using a combination of ICM modeling and mutagenesis experiments, the active site of CD59 was mapped. The site was found to be located next to a hydrophobic groove on the face of the molecule. The model was used to design mutations that can improve CD59 inhibitory activity⁵¹ and identify the functionally important residues that helped in understanding the reasons for species selectivity⁵². The residues responsible for species selective function were mapped to a region between residues number 40 and 66.

To identify the individual residues involved in human CD59 species selective function, a model of rat CD59 was built and the location and distribution of non-conserved surface patches within the region 40-66 were analyzed. The key residues were identified to be 47, 51, and 55, which are located on a short single helix of CD59⁵³. This patch was further refined using NMR models and optimized in ICM⁵⁴. A short peptide spanning this region was shown to inhibit binding of human of CD59. Further modeling of the structure revealed a potential small molecule binding pocket, which was favorable for the development of small molecule antagonists of CD59-mediated complement inhibition.

Self-assembly of complement proteins C5b, C6, C7, C8, and from 1 to 18 molecules of C9 forms MAC. CD59 binds to C8, which subsequently binds to C9. ICM was used to predict the interactions between C9

MolSoft was funded by the Gates Foundation to work on modeling the V3 Loop which is a conserved structure of gp120 that can induce broadly neutralizing antibodies against HIV-1⁴⁶⁻⁵⁰.

Two (short and full length) V3-scaffold immunogen constructs (V3-CTB) were designed using 3D structures of cholera toxin B subunit (CTB), complete V3 in the gp120 context, and V3 bound to a monoclonal antibody (mAb) (Figure 5). Both short and full length were recognized by a large majority of 24 mAbs in the screening panel. The immunogens were evaluated in rabbits using DNA-prime/protein-boost protocol. Boosting with the full-length V3-CTB induced high anti-V3 titers in

and CD59. ICM modeling identified a “closed” and “open” pocket, and ICM-Docking was used to predict the interactions⁵⁵.

Case Study: Modeling of GPCR Agonist Conformation.

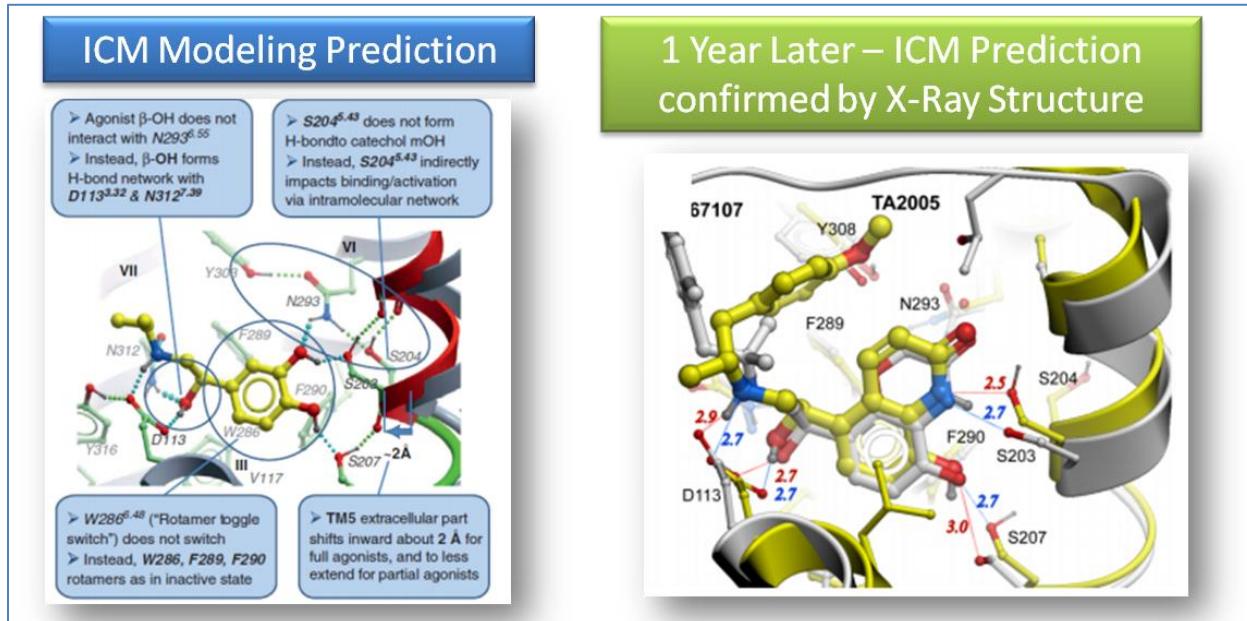


Figure 6 An ICM model accurately predicted the key structural changes in the agonist GPCR ligand binding pocket, this was confirmed one year later when the crystal structure was published.

GPCR agonist molecules can induce structural changes in the transmembrane region of the receptor. Modeling these changes is important for agonist design and understanding GPCR biology and drug action. In 2009, the agonist models of Beta 2 Adrenergic receptor and Adenosine A2A generated with ICM were published^{20,21} and in 2010 the crystal structure was published⁵⁶. A comparison between the ICM models and the atomic crystal structure showed that the binding pose of the agonist differed by only 0.8 Å (adrenergic) and 1.7 Å (adenosine) (Figure 6). The models predicted key side-chain rotations and transmembrane helical shifts in the pocket which determined full, partial or inverse agonism.

References

1. Abagyan, R., Totrov, M. & Kuznetsov, D. ICM - a new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* **15**, 488–506 (1994).
2. Abagyan, R. & Totrov, M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J. Mol. Biol.* **235**, 983–1002 (1994).
3. Totrov, M. & Abagyan, R. Rapid boundary element solvation electrostatics calculations in folding simulations: successful folding of a 23-residue peptide. *Biopolymers* **60**, 124–133 (2001).
4. Arnautova, Y. A., Abagyan, R. A. & Totrov, M. Development of a new physics-based internal coordinate mechanics force field and its application to protein loop modeling. *Proteins* **79**, 477–498 (2011).

5. Arnautova, Y. A., Jagielska, A. & Scheraga, H. A. A new force field (ECEPP-05) for peptides, proteins, and organic molecules. *J. Phys. Chem. B* **110**, 5025–5044 (2006).
6. Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **17**, 490–519 (1996).
7. Halgren, T. A. & Nachbar, R. B. Merck molecular force field. IV. Conformational energies and geometries for MMFF94. *J. Comput. Chem.* **17**, 587–615 (1996).
8. Abagyan, R. *et al.* Homology modeling with internal coordinate mechanics: deformation zone mapping and improvements of models via conformational search. *Proteins Suppl* **1**, 29–37 (1997).
9. Maiorov, V. & Abagyan, R. Energy strain in three-dimensional protein structures. *Fold. Des.* **3**, 259–269 (1998).
10. Borchert, T. V., Abagyan, R., Jaenicke, R. & Wierenga, R. K. Design, creation, and characterization of a stable, monomeric triosephosphate isomerase. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 1515–1518 (1994).
11. Borchert, T. V. *et al.* Three new crystal structures of point mutation variants of monoTIM: conformational flexibility of loop-1, loop-4 and loop-8. *Struct. Lond. Engl.* **1993** **3**, 669–679 (1995).
12. Borchert, T. V., Abagyan, R., Kishan, K. V., Zeelen, J. P. & Wierenga, R. K. The crystal structure of an engineered monomeric triosephosphate isomerase, monoTIM: the correct modelling of an eight-residue loop. *Struct. Lond. Engl.* **1993** **1**, 205–213 (1993).
13. Katritch, V., Rueda, M., Lam, P. C.-H., Yeager, M. & Abagyan, R. GPCR 3D homology models for ligand screening: lessons learned from blind predictions of adenosine A2a receptor complex. *Proteins* **78**, 197–211 (2010).
14. Michino, M. *et al.* Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008. *Nat. Rev. Drug Discov.* **8**, 455–463 (2009).
15. Kufareva, I., Rueda, M., Katritch, V., Stevens, R. C. & Abagyan, R. Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. *Struct. Lond. Engl.* **1993** **19**, 1108–1126 (2011).
16. Stevens, R. C. *et al.* The GPCR Network: a large-scale collaboration to determine human GPCR structure and function. *Nat. Rev. Drug Discov.* **12**, 25–34 (2012).
17. Cavasotto, C. N. *et al.* Discovery of novel chemotypes to a G-protein-coupled receptor through ligand-steered homology modeling and structure-based virtual screening. *J. Med. Chem.* **51**, 581–588 (2008).
18. Tosh, D. K. *et al.* Optimization of adenosine 5'-carboxamide derivatives as adenosine receptor agonists using structure-based ligand design and fragment screening. *J. Med. Chem.* **55**, 4297–4308 (2012).
19. Katritch, V. *et al.* Structure-based discovery of novel chemotypes for adenosine A(2A) receptor antagonists. *J. Med. Chem.* **53**, 1799–1809 (2010).
20. Katritch, V. *et al.* Analysis of full and partial agonists binding to beta2-adrenergic receptor suggests a role of transmembrane helix V in agonist-specific conformational changes. *J. Mol. Recognit. JMR* **22**, 307–318 (2009).
21. Reynolds, K. A., Katritch, V. & Abagyan, R. Identifying conformational changes of the beta(2) adrenoceptor that enable accurate prediction of ligand/receptor interactions and screening for GPCR modulators. *J. Comput. Aided Mol. Des.* **23**, 273–288 (2009).
22. Dong, M. *et al.* Mapping spatial approximations between the amino terminus of secretin and each of the extracellular loops of its receptor using cysteine trapping. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **26**, 5092–5105 (2012).
23. Cawston, E. E. *et al.* Molecular basis for binding and subtype selectivity of 1,4-benzodiazepine antagonist ligands of the cholecystokinin receptor. *J. Biol. Chem.* **287**, 18618–18635 (2012).

- 24.Dong, M. *et al.* Molecular approximations between residues 21 and 23 of secretin and its receptor: development of a model for peptide docking with the amino terminus of the secretin receptor. *Mol. Pharmacol.* **72**, 280–290 (2007).
- 25.Dong, M. *et al.* Secretin occupies a single protomer of the homodimeric secretin receptor complex: insights from photoaffinity labeling studies using dual sites of covalent attachment. *J. Biol. Chem.* **285**, 9919–9931 (2010).
- 26.Dong, M. *et al.* Spatial approximation between secretin residue five and the third extracellular loop of its receptor provides new insight into the molecular basis of natural agonist binding. *Mol. Pharmacol.* **74**, 413–422 (2008).
- 27.Dong, M. *et al.* Role of lysine187 within the second extracellular loop of the type A cholecystokinin receptor in agonist-induced activation. Use of complementary charge-reversal mutagenesis to define a functionally important interdomain interaction. *Biochemistry (Mosc.)* **46**, 4522–4531 (2007).
- 28.Harikumar, K. G. *et al.* Fluorescence resonance energy transfer analysis of secretin docking to its receptor: mapping distances between residues distributed throughout the ligand pharmacophore and distinct receptor residues. *J. Biol. Chem.* **282**, 32834–32843 (2007).
- 29.Thanki, N. *et al.* Protein engineering with monomeric triosephosphate isomerase (monoTIM): the modelling and structure verification of a seven-residue loop. *Protein Eng.* **10**, 159–167 (1997).
- 30.Abagyan, R. & Totrov, M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J. Mol. Biol.* **235**, 983–1002 (1994).
- 31.Rueda, M., Totrov, M. & Abagyan, R. ALiBERO: Evolving a Team of Complementary Pocket Conformations Rather than a Single Leader. *J. Chem. Inf. Model.* **52**, 2705–2714 (2012).
- 32.Kufareva, I., Ilatovskiy, A. V. & Abagyan, R. Pocketome: an encyclopedia of small-molecule binding sites in 4D. *Nucleic Acids Res.* **40**, D535–540 (2012).
- 33.Kufareva, I. & Abagyan, R. Strategies to overcome the induced fit effects in molecular docking. *Hansmann UH E Meinke JH Mohanty Nadler W Zimmermann Oeds Comput. Biophys. Syst. Biol. CBSB08 H John Von Neumann Inst. Comput. NIC JüL*. 1–6 (2008).
- 34.Bottegoni, G., Rocchia, W., Rueda, M., Abagyan, R. & Cavalli, A. Systematic exploitation of multiple receptor conformations for virtual ligand screening. *PLoS One* **6**, e18845 (2011).
- 35.Totrov, M. & Abagyan, R. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr. Opin. Struct. Biol.* **18**, 178–184 (2008).
- 36.Cozza, G., Bortolato, A. & Moro, S. How druggable is protein kinase CK2? *Med. Res. Rev.* **30**, 419–462 (2010).
- 37.Villoutreix, B. O., Eudes, R. & Miteva, M. A. Structure-based virtual ligand screening: recent success stories. *Comb. Chem. 38 High Throughput Screen.* **12**, 1000–1016 (2009).
- 38.Rueda, M., Bottegoni, G. & Abagyan, R. Consistent improvement of cross-docking results using binding site ensembles generated with elastic network normal modes. *J. Chem. Inf. Model.* **49**, 716–725 (2009).
- 39.Cavasotto, C. N., Kovacs, J. A. & Abagyan, R. A. Representing receptor flexibility in ligand docking through relevant normal modes. *J. Am. Chem. Soc.* **127**, 9632–9640 (2005).
- 40.Katritch, V., Rueda, M. & Abagyan, R. Ligand-guided receptor optimization. *Methods Mol. Biol. Clifton NJ* **857**, 189–205 (2012).
- 41.Bisson, W. H. *et al.* Discovery of antiandrogen activity of nonsteroidal scaffolds of marketed drugs. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 11927–11932 (2007).
- 42.Totrov, M. & Abagyan, R. Detailed ab initio prediction of lysozyme-antibody complex with 1.6 Å accuracy. *Nat. Struct. Biol.* **1**, 259–263 (1994).

- 43.Strynadka, N. C. *et al.* Molecular docking programs successfully predict the binding of a beta-lactamase inhibitory protein to TEM-1 beta-lactamase. *Nat. Struct. Biol.* **3**, 233–239 (1996).
- 44.Fernández-Recio, J., Totrov, M. & Abagyan, R. ICM-DISCO docking by global energy optimization with fully flexible side-chains. *Proteins* **52**, 113–117 (2003).
- 45.Fernández-Recio, J., Abagyan, R. & Totrov, M. Improving CAPRI predictions: optimized desolvation for rigid-body docking. *Proteins* **60**, 308–313 (2005).
- 46.Gorny, M. K. *et al.* Human anti-V3 HIV-1 monoclonal antibodies encoded by the VH5-51/VL lambda genes define a conserved antigenic structure. *PloS One* **6**, e27780 (2011).
- 47.Jiang, X. *et al.* Conserved structural elements in the V3 crown of HIV-1 gp120. *Nat. Struct. Mol. Biol.* (2010). doi:10.1038/nsmb.1861
- 48.Spurrier, B. *et al.* Structural analysis of human and macaque mAbs 2909 and 2.5B: implications for the configuration of the quaternary neutralizing epitope of HIV-1 gp120. *Struct. Lond. Engl.* **1993** **19**, 691–699 (2011).
- 49.Totrov, M. *et al.* Structure-guided design and immunological characterization of immunogens presenting the HIV-1 gp120 V3 loop on a CTB scaffold. *Virology* **405**, 513–523 (2010).
- 50.Zolla-Pazner, S. *et al.* Cross-clade HIV-1 neutralizing antibodies induced with V3-scaffold protein immunogens following priming with gp120 DNA. *J. Virol.* **85**, 9887–9898 (2011).
- 51.Yu, J. *et al.* Mapping the Active Site of CD59. *J. Exp. Med.* **185**, 745–754 (1997).
- 52.Yu, J. *et al.* Mapping the regions of the complement inhibitor CD59 responsible for its species selective activity. *Biochemistry (Mosc.)* **36**, 9423–9428 (1997).
- 53.Zhang, H.-F. *et al.* Identification of the Individual Residues That Determine Human CD59 Species Selective Activity. *J. Biol. Chem.* **274**, 10969–10974 (1999).
- 54.Huang, Y. *et al.* Insights into the Human CD59 Complement Binding Interface Toward Engineering New Therapeutics. *J. Biol. Chem.* **280**, 34073–34079 (2005).
- 55.Huang, Y., Qiao, F., Abagyan, R., Hazard, S. & Tomlinson, S. Defining the CD59-C9 Binding Interaction. *J. Biol. Chem.* **281**, 27398–27404 (2006).
- 56.Wacker, D. *et al.* Conserved binding mode of human beta2 adrenergic receptor inverse agonists and antagonist revealed by X-ray crystallography. *J. Am. Chem. Soc.* **132**, 11443–11445 (2010).