Computational Structural Proteomics

Lars Brive and Ruben Abagyan

Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA

1 Introduction

1.1 New sequences from proteomics and genomics

Genomics and proteomics efforts generate new results at a remarkable pace. Over 30 organisms have been sequenced and contribute evolutionary information and new proteins with yet unknown function that are important from a basic research point of view and for biomedical research. We now need ways to analyze the information and make something useful of it, which may turn out to be a major obstacle. From a protein structure point of view, the DNA sequences have to be analyzed for coding regions, protein constructs shall be expressed correctly at good yields, functions and properties shall be determined, and the structure needs to be solved. For example, of the more than 600 predicted G-protein coupled receptors (GPCRs) in the human genome, 250-300 have been annotated in swissprot, 100 have been experimentally characterized and only 2 crystal structures have been reported.

1.2 Structural proteomics

A complete understanding of protein function requires the structures of the individual components and their complexes to be determined. The huge amount of data from genomics and proteomics requires high-throughput and highly automated procedures for structure determination. There are currently twelve public structure proteomics centers (see www.rcsb.org/strucgen.html) that aim to solve long term scientific goals and generate data available for the public. The private efforts are more geared towards proteins that are easier to solve and are drug targets. The major companies in the U.S.A are Structural Genomics, Syrrx and Plexicon. The majority of structures are solved by crystallographic methods. Determination of structures that are new, or difficult to crystallize well, often requires the use of synchrotron resources due to the availability of high-energy source and MAD phasing. In cases where a similar structure exists, e.g. co-crystals of small molecules used in drug discovery processes, molecular replacement methods make it possible to utilize smaller, in-house instruments with higher availability and lower cost. The experimental techniques that allow crystallization of membrane proteins are progressing rapidly and will soon bring lots of data, see for example Chang & Roth (Chang and Roth 2001). *Nuclear magnetic resonance* (NMR) is currently used to

determine structures for proteins with molecular weights less than 25 kDa. Recent developments in hardware and pulse sequences, including TROSY sequences, may extend the practical range to about 50 kDa. NMR is indispensable for structure determination of proteins that cannot crystallize or that yield low-quality crystals. Automated structure determination methods have been developed but are still not capable of routine analysis. NMR also provides useful tools for screening for folded structures (line shape analysis will quickly reveal how well-folded a structure is and may be used to select constructs for structure determination with NMR or crystallography) and for determination of physical properties (e.g. dynamics and pK_as of ionizable groups).

Despite the recent developments in structure proteomics, the resources are still too limited. To date, the protein databank (www.rcsb.org) contains 16245 entries as of October 9, 2001 (Berman *et al.* 2000), with almost 3000 structures a year growth rate. Curiously, the structural genomics initiative does not change this number dramatically. Only about 300 to 1000 structures are expected to be added per year. Even if the number of solved structures would double, it's still an order of magnitude smaller than the known protein sequences resulting from the genomics and proteomics initiatives, and a majority of important targets will be missed. In addition, the cost for each solved structure is substantial (typically $100-300 k). The total budget of the seven public structure proteomics centers in the U.S. is roughly 35 million dollars per year. Determination of biomolecule interactions is of major importance, and determination of complex structures is a greater challenge compared to the individual parts. And how can effects of single nucleotide polymorphisms and mutants be evaluated promptly and cost effectively? Computational biology addresses these questions and may play a significant role to decrease the gap between available sequences and structures.

Overall, to cover the difference between sequences and structures we need to be able to at least build models by homology (short of being able to accurately predict structure from scratch), and both predicted and experimental structures can then be further studied to predict their function and to design specific ligands if necessary.

2 Computational tasks

We will now list the current problems in structural biology where computational methods can contribute. We will outline the problems, describe a solution, focus on some points that are of special importance and give examples of case studies. The framework for the discussion will be the ICM program that has been designed to approach all of the following problems, and we will initially explain the internal coordinate system that is the basis for the following discussion.

2.1 Internal coordinate mechanics, ICM

Predicting structure means finding a global minimum of an ill-behaved energy function of hundreds of variables and special approaches are needed to be developed to deal with the problem. The most commonly used molecular mechanics force fields operate in Cartesian space and provides solutions to many of the tasks that were mentioned above. However, the large number of degrees of freedom of even a small size protein makes it impossible to efficiently search the global conformational space, e.g. in protein folding. By switching to an internal coordinate system operating in torsion space and fixing the high frequency variables (bond lengths, bond angles), the number of degrees of freedom is reduced seven-fold. If planar or tetrahedral geometry is assumed for every atom, the decrease is roughly ten-fold, and provides a faster and still accurate means of energy calculations. Force fields working in torsion space were first described 25 years ago (Momany *et al.* 1975) and are currently being developed in programs such as ICM (Abagyan 2000; Abagyan *et al.* 1994), ECEPP (Nemethy *et al.* 1992) and DYANA (Guntert and Wuthrich 2001). Description the molecular system in torsion space also enables efficient sampling of the conformational space. A Monte Carlo minimization step in torsion space changes a randomly selected group of coupled variables according to their local probability distribution and performs a local energy minimization (Figure 1). (Abagyan and Argos 1992; Abagyan R *et al.* 1994; Li and Scheraga 1987) This step is repeated and coupled with mechanisms to limit sampling of uninteresting areas and encourage sampling of new areas until the procedure converges. New conformations are selected if the energy difference between the new and old conformation ($\Delta E = E_{n+1} - E_n$) is less than 0, or if a random number between 0 and 1 is smaller than $e^{-\Delta E/RT}$, where R is 8.314 kJ/(mol K) and T is the simulation temperature (Metropolis selection). The

Metropolis selection allows the procedure to traverse local energy barriers, and the size of the barriers that can be crossed is dependent on T. A stack of the lowest energy conformations is continuously updated, and also contains the number of times a particular conformation has been visited. If a certain number of steps have yielded the same conformation, the simulation temperature is doubled to allow escape from local minima.

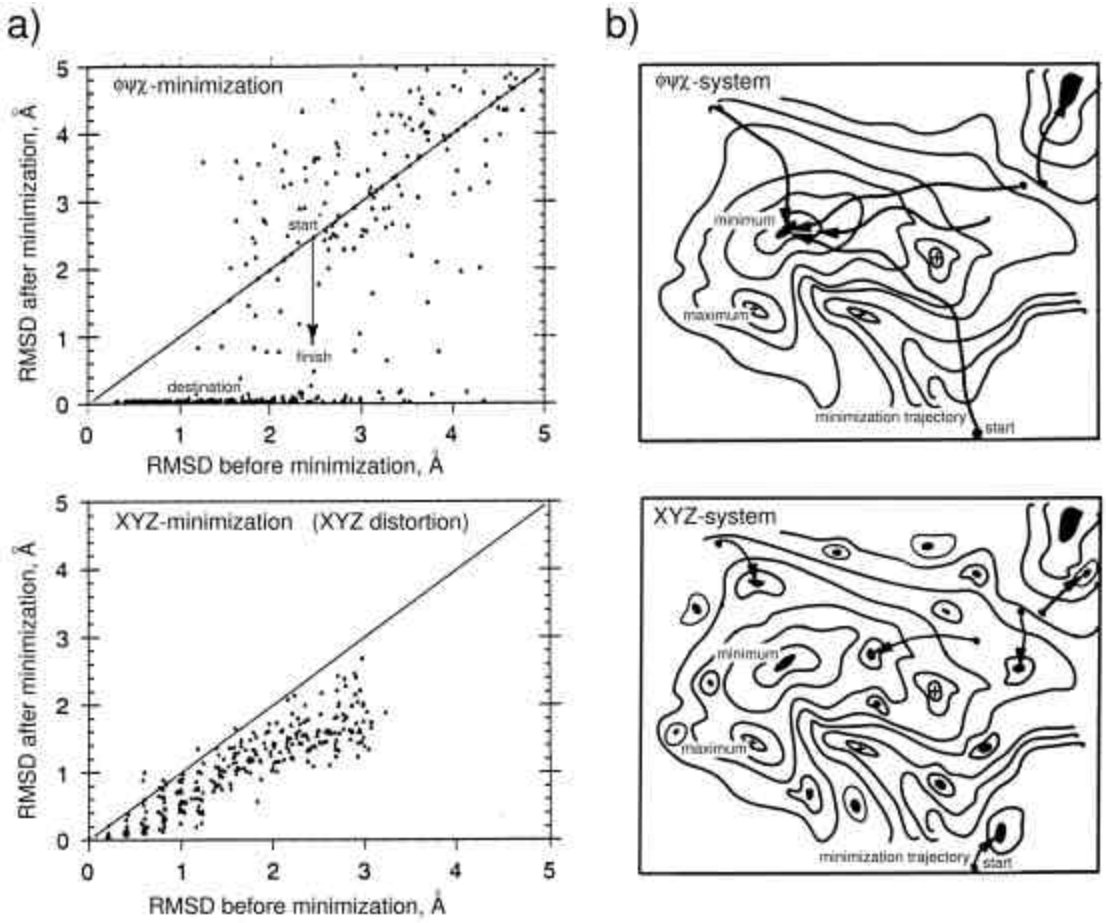

Figure 1. Conformation minimization in torsion space is more efficient for finding the global minimum conformation of a polypeptide. (Abagyan R *et al*. 1994) a) Monte Carlo minimization for the folded structure in torsion space using the ICM protocol outlined in the text (top) and in Cartesian space using molecular mechanics minimization (bottom). Random deviations from the folded structures were generated in torsion and Cartesian space, respectively. The RMS deviation between the starting and target conformations is shown on the horizontal axes, the RMS between the minimized and target structure is shown on the vertical axes. Monte Carlo minimization in torsion space can find the folded structure even if the distortions are more than three Å. Note that a full conformational search employs additional steps and can find the global minimum from *any* given starting structure. b) Minimization trajectories drawn on the energy landscape projected into two dimensions. The torsion space landscape with fixed covalent geometry (top) contains less local minima than the corresponding landscape for Cartesian space with relaxed covalent geometry, and provides a faster route to the global minimum.

Side chain and backbone torsion angle changes are partly biased to make the search procedure more efficient and still allow rare conformations (that do exist in protein structures) to be sampled. This so called optimal biased Monte Carlo minimization (OBMCM, a.k.a. BPMC) protocol improves the sampling efficiency by an order of magnitude compared to Monte Carlo minimization and has been shown to predict structures ab initio with high efficiency.(Abagyan and Totrov 1999; Totrov and Abagyan 2001) A critical issue for all molecular mechanics calculations is the accuracy of the potentials. ICM uses parameters that have been optimized for proteins based on the ECEPP3 force field (Nemethy G *et al.* 1992) and also includes entropy and implicit solvent energy terms. (Abagyan 1997) The accuracy of the energy function is critical and we have recently developed a new Internal Coordinate Force Field (ICFF) that includes *implicit* flexibility of bond length, bond angle and "1-4" van der Waals interactions by projection of a Cartesian force field onto the internal coordinate molecular model with fixed bond geometry. (Katritch V., Totrov M., and Abagyan R., submitted) The main object with ICFF is to improve the force field description of both protein and non-protein molecules and to enhance ligand docking accuracy.

2.2 Homology modeling.

Roughly 10% of the available protein sequences resulting from the genomics efforts can be expected to be solved despite the major progress in structural proteomics. The remaining, unsolved structures will contain a wealth of information of interest for the understanding of how cells work and how we can treat disease. One important role of computational biology is to predict three-dimensional structures of proteins by homology modeling to cover this gap. Simple comparative modeling tools have been used for three decades and are well established.

The easy step in modeling homology is to copy aligned parts of the backbone from a homologue. In general, good models can be generated if the template structure has more than 30% sequence identity, and the predicted structures often have a C_α RMSD less than 1.5 compared with the crystal structure. Although the accuracy of the predicted structures is limited, the models are still valuable for a range of tasks including prediction of function to ligand design.

The first step in homology modeling is to find template structures and produce a (multiple) sequence alignment. PSI-BLAST (Altschul *et al.* 1997) and hidden Markov models are commonly used tools for both tasks. Alignment quality can be further improved by inclusion of structural information, as in the secondary-structure-and-residue-accessibility-enhanced ZEGA algorithm (Abagyan R 2000; Abagyan and Batalov 1997), and by manual inspection and editing. The quality of the alignment is crucial for the result since an error will propagate throughout the procedure. The query polypeptide chain is then threaded onto the template structure and the backbone and conserved residues are overlaid on the template structure. In the next step, side chains conformations for non-conserved residues are optimized. Finally, the inserted or deleted loops are placed. Loop modeling is currently one of the most challenging parts of homology modeling and should be treated with caution in the subsequent analysis. One solution is to search for peptide fragments with similar sequence and end point distances in a structure database, and score all hits on the model. Loop modeling is more successful if the end point coordinates are known, such as for redesign of available proteins. (Borchert *et al.* 1994; Borchert *et al.* 1993)

2.3 Ab initio folding

Three-dimensional structures of proteins still cannot be predicted from first principles without knowledge about homologous structures. In the future, it will hopefully be possible by efficient and accurate global optimization of the free energy function. Ab initio folding requires a fast global search algorithm and a complete and correct energy function. The solvation energy is an important term and was, until recently, a difficult problem to address. The structure of a well-folded 23-residue peptide with $\beta\beta\alpha$ topology was recently predicted using a biased probability Monte Carlo search including implicit solvent electrostatics. (Totrov M and Abagyan R 2001)

2.4 Annotation

Once a new structure (model) has been determined, it needs to be annotated with respect to properties and function. This information includes the character of the surface, functional residues, hinge points, local rigidity, potential binding pockets, sequence similarity to related proteins, energetic strain and local accuracy of the model. In

addition, information from biological and structural databases can be projected on the new structure, such as mutation data, SNP sites, known binding and catalytic sites, dimer interfaces, dynamics parameters and so forth.

2.5 Target selection.

The number of potential, or hypothetical, biomedical targets is far larger than the number that can currently be exploited. It is therefore necessary to select, from a list of structures, those that have a potential for further development. Are there pockets with the right size and character for drug design (see below)? What are the chances that a particular peptide binding groove that is important for function can be blocked with a small molecule or a peptido-mimetic? These kinds of annotation results are critical for target selection, and although some steps in the procedures can be automated, interactive analysis and inspection are still required for optimal results.

2.6 Detecting drugable pockets

Structure-based design of ligands can, in principle, be targeted towards a whole protein molecule or complex, but it is more efficient to find and focus on specific sites. Given a suitable receptor structure, one of the first steps is to identify sites that are “drugable”. A simple and efficient method is to map out cavities in the receptor structure that is not occupied with ligands/inhibitors or solvent molecules. More sophisticated methods calculate potential maps for pocket size, charge distribution and similar properties, and finds potential pockets by contouring the sum of the maps. A benefit from the latter method is that binding sites on the surface that are solvent exposed will also be identified (Figure 2). These methods are highly reliable for finding and characterizing binding pockets in structures where a ligand has been co-crystallized with the receptor. Additional potential binding sites are often found that may represent pockets with trapped solvent molecules or well-defined groves on the protein surface. To our knowledge, there are still no examples of rational ligand design based on a non-natural binding pocket, i.e. a pocket that is not known to bind any ligands, but there are already successful designs of ligands targeting protein-protein interaction interfaces.

2.7 Small molecule docking and virtual ligand screening

A protein that can affect disease progression have the potential to be a drug target. The most desirable approach to regulate a target's activity is by design of small, orally bioavailable, ligands. (Peptides and proteins are more difficult for this task in humans for a number of reasons, including delivery issues and immune response.) Screening of tens of thousands molecules is standard practice in industry, but it is costly. It is therefore desirable to reduce the number of compounds to be tested.

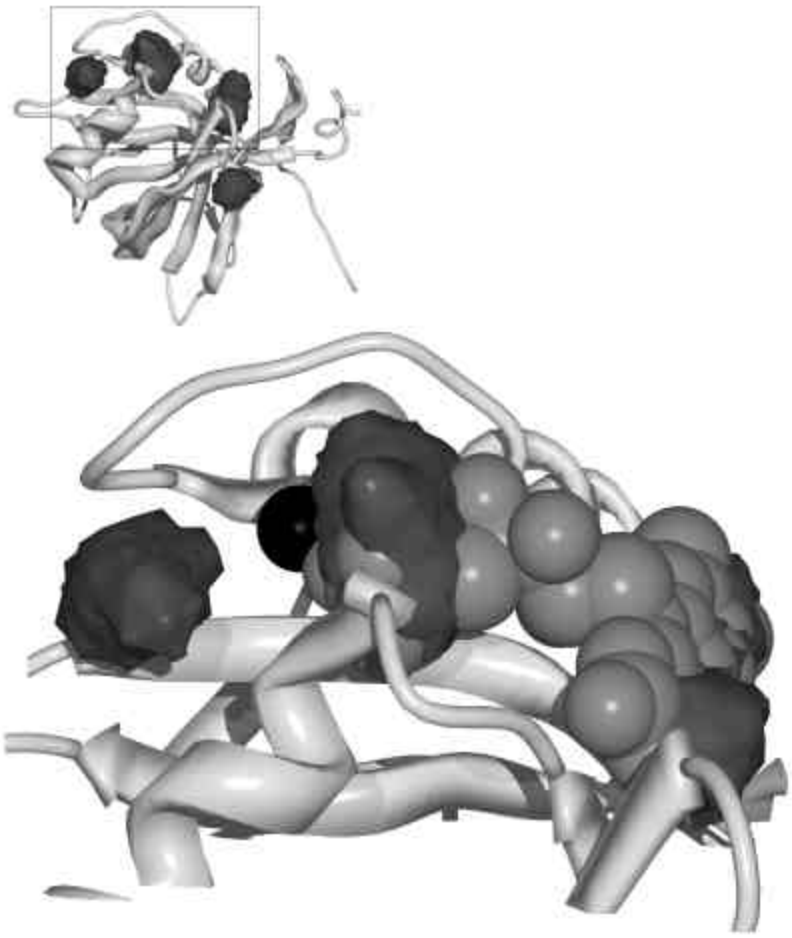

Figure 2. Four potential drug binding pockets of the MTH152 flavin mononucleotide binding protein. (Christendat *et al*. 2000) The top panel shows the overall structure with the pockets shown as dark gray blobs. The lower panel shows a close-up view of the top part of the structure, along with the FMN molecule as a gray cpk model and a magnesium ion as a black sphere. Note that the pockets are all partly surface exposed and that the FMN unit covers two pockets. The two remaining pockets are "empty" in the pdb file and could be targets for discovery of new ligands, conceivably in combination with an adjacent site. The structure was solved in a structural proteomics initiative that comprised 424 proteins from *Methanobacterium thermoautotrophicum.*

Virtual ligand screening (VLS) drastically reduces the number of compounds that need to be screened *in vitro* and *in vivo* in the search for active ligands. (See Abagyan and Totrov 2001 and references therein.) Pharmacophore modeling has traditionally been used for this purpose, but the method is hampered by the low diversity of the resulting libraries and the bias towards the template ligands. If instead a large number of compounds are docked to a binding site without bias of known ligands, a better diversity, selectivity and

toxicity properties may be achieved. The starting point of the procedure is to calculate potential maps of the target site of a receptor model (from X-ray, NMR or homology modeling) in the desired structural state (e.g. open or closed state, as for nuclear receptor ligand binding domains). The maps may cover two vicinal pockets to provide additional binding energy and specificity (see for example Shuker *et al.* 1996). A library with compounds to be screened shall be generated. The compounds are virtual and are not restricted to available, synthesized molecules and provides one clear advantage of VLS versus high throughput screening (HTS). They should be selected based on chemical feasibility an on drug-like properties, such as absorption, metabolic properties and low toxicity. The efficiency of automatic docking of flexible ligands improves if potential maps are used instead of full atom representation of the receptor. These maps describe the shape, charge distribution and hydrogen bonding properties of the fixed receptor. Each ligand is positioned in a defined site and the structure is optimized (Figure 3). The ligands can be targeted to a specific binding site if it is known (which is not an option for HTS), or it can be the whole receptor. Due to limitations in the free energy calculation of the bound and unbound ligand, the ranking of the generated structures is based on a scoring function rather than of the calculated energy. Manual inspection of the top hits has also been a successful strategy for the final selection. Special properties of the ligands can easily be included in the processing of the hit list, e.g. the presence of a functional group in space to provide selectivity for a particular protein isoform. Two recent examples of true structure based ligand design involve the ligand binding domain (LBD) of the nuclear receptor retinoic acid receptor (RAR). In the first case, a model of the RARγ antagonist-bound structure was constructed based on the inactive RARα structure and the antagonist bound estrogen receptor LBD. (Schapira *et al.* 2000) Over 150 000 compounds were screened *in silico*, of which 30 were selected manually from the top 300 hits selected by the scoring algorithm. This significantly reduced set of ligands was tested for activity in vitro. Two new antagonists and one agonist were found. This example demonstrates the validity of the *in silico* screening procedure, and also provides an example where a modeled receptor was successfully used. In a similar study, the agonist-bound structure of RARα LBD was constructed from the RARγ LBD by replacing three residues in the binding pocket. (Schapira *et al.* 2001) Again, 30 structures were selected out of 150 000 initial compounds for *in vitro* binding. Two agonists were active at 50 nM and one has an entirely novel structure as compared to known RAR ligands.

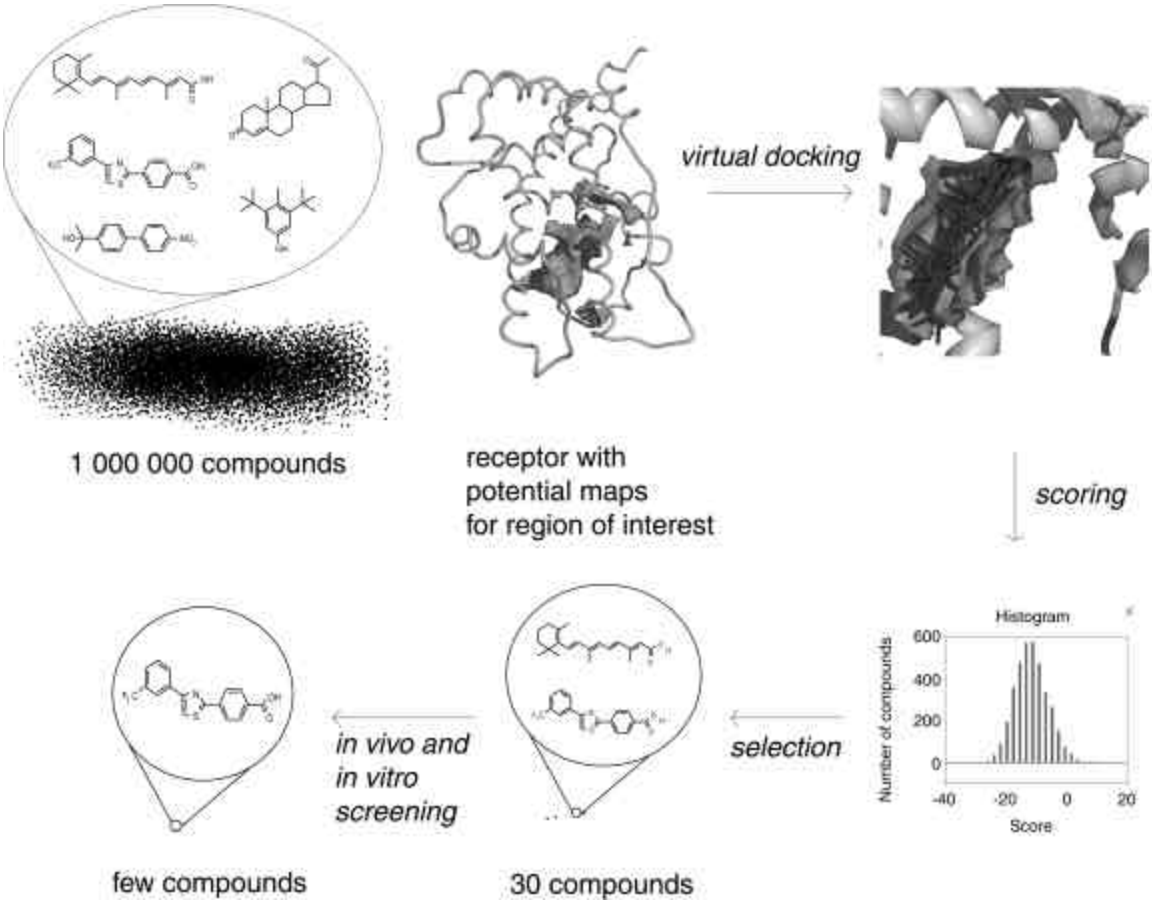

Figure 3. The general procedure for virtual ligand screening starts from a receptor model and a compound library and results in a limited list of potential ligands.

2.8 Peptide-protein docking

Docking of flexible peptides to a protein represents a more difficult problem than small molecule docking due to the dramatic increase of number of freedom for the ligand, and also because the binding surface of the receptor is often flat and more permissive to uncorrect solutions. The accurate calculation of solvent electrostatics and entropy becomes more important for surface docking experiments. A recent example is the docking of phosphotyrosine-containing peptides to SH2 and PTB domains. (Zhou and Abagyan 1998) It was shown that three flexible penta- or hexa-peptides were correctly docked to an SH2 domain model. Docking of longer peptides (around 11 residues) to SH2 or PTB domains was more challenging and gave correct results only when the phosphotyrosine moiety was restrained to its binding pocket in the protein. It is probably a general rule that longer peptides are more difficult to dock and that biological information will aid in the prediction. A less complex task is to dock flexible peptides to

well-defined grooves on the protein surface. Recent results from the docking of peptides to HLA (human MHC) proteins are encouraging, and will be useful in structural immunology research.

2.9 Protein-protein docking.

The association of two biomolecules is a fundamental process in the cell. It is a major challenge for computational biology to predict the structure of a complex given the two, separately solved structures. Formation of a biomolecule complex is accompanied by conformational changes of both structures. It would therefore be intuitive to dock fully flexible molecules, but this is far out of reach with the current methods and resources. The solution is to represent the receptor with potential grids (as described in virtual ligand screening above) and to ignore the flexibility of the ligand until a late stage of the procedure. Several groups have successfully recreated complex structures from the separate parts by rigid body docking (reviewed in (Sternberg *et al.* 1998)), but using the individually solved structures as starting points is the realistic exercise and will be a challenge for the future when many, single structures are expected to be solved.

The initial step in protein-protein docking is to find relative positions of the molecules. This can be done by defining a grid around the receptor, placing the ligand in all grid points and evaluating the interaction energy. This rigorous method is computationally expensive and impractical. A more efficient sampling strategy selects starting points over the receptor surface and performs stochastic pseudo-Brownian moves followed by Monte Carlo minimization of the rigid structures (Figure 4). However, the binding of two biomolecules is always accompanied by some degree of structural changes, everywhere from side chain rearrangements to domain motions. Soft docking methods, where van der Waals repulsion terms are downscaled, can partly compensate for the mobility (Fernández-Recio J, Totrov M and Abagyan R, in press) but suffer from bad discrimination of false positives. Local minimization of the interface of the candidate complexes would provide more robust predictions. Optimal biased Monte Carlo minimization of the initial complexes reproduced the structure of a lysozyme-antibody complex with 1.6 Å accuracy (Totrov and Abagyan 1994). An improved version of this procedure was recently tested on 24 protein-protein complexes. Seven proteinase-inhibitor complexes out of eleven were correctly predicted as the highest rank solution

and mark a clear improvement over previous attempts. (Fernández-Recio J, Totrov M and Abagyan R, in press) However, other classes of complexes, such as antibody-antigen complexes, proved to be more challenging. Overall, the success rate for protein-protein docking is about 30%. Although the precision is still too low for general predictions, *ab initio* dockings provide useful hypotheses that can be confirmed with experimental data such as mutagenesis analysis or NMR chemical shift perturbations (see for example Morelli *et al.* 2000).

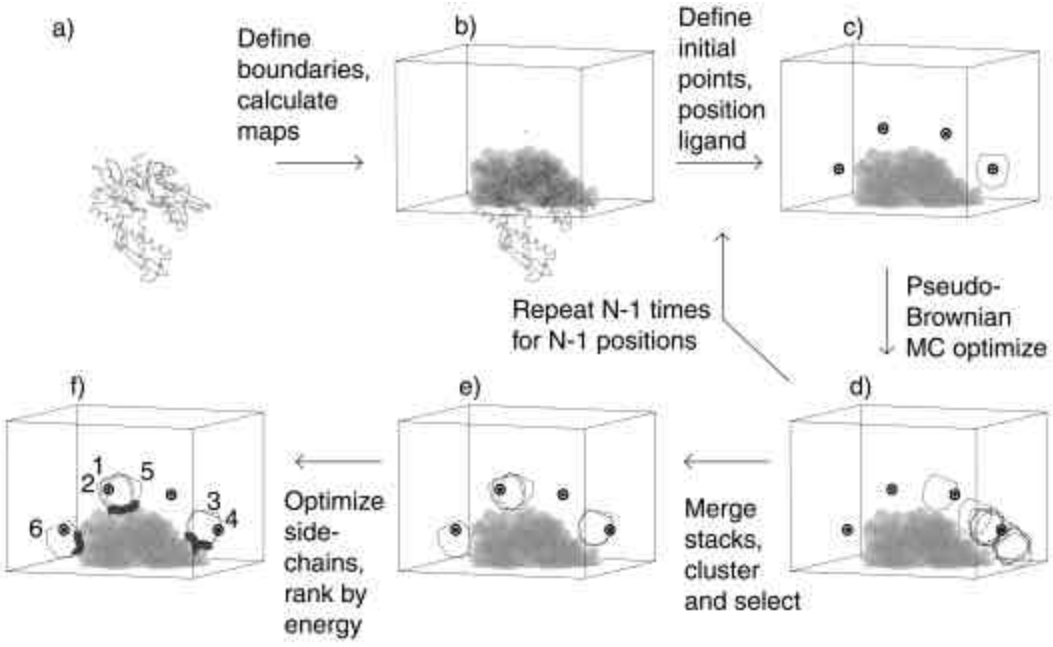

Figure 4. Protein-protein docking using pseudo-Brownian Monte Carlo procedure followed by optimal biased Monte Carlo Minimization. A weighted potential map (gray, diffuse cloud) is calculated from five potential maps for the receptor (black tangle), and N starting positions for the ligand are distributed around the map. Each ligand is docked and minimized to yield a set of M x N optimized rigid body docking complexes, where M is the number of conformations in the stack for each starting position. After clustering and removal of redundant complexes, the lowest energy complexes are selected for OBMCM optimization of the interface side-chains of the ligand. The resulting structures are ranked according to energy (numbers 1-6 in panel f) and in the optimal case, the correct complex structure is the one with lowest energy.

2.10 Protein health

In the process of determining structures by crystallography, by NMR and by modeling it is necessary to check for errors in the intermediate and final solutions. This is becoming

even more important as structure generation is being increasingly automated and the large number of solved structures (and publishing pressure) provides less time for critical human inspection, thus increasing the risk for errors. Structural fitness is commonly measured by comparison with the experimental data (if available), by local geometry criteria and by checking for steric clashes and provides a fast way to identify local abnormalities.(Laskowski *et al.* 1993; Laskowski *et al.* 1996) More rigorous methods to assess structure quality have been suggested. One approach is to compare the energy of residues in a target structure with a table of calculated energies for residues in a database of X-ray crystal structures with resolution better than 2.0 Å, grouped by residue type. (Maiorov and Abagyan 1998) In addition to the geometry check, this method probes the non-bonded interactions (including electrostatics and hydrogen bonding) more carefully than the simple bump-check and provides a complementary test of structural healthiness.

2.11 Understand SNPs and mutations

Naturally occurring single nucleotide polymorphisms (SNPs) may modify both structure and function of a protein. This variation in the genetic information is one reason why the response to drugs varies between individuals. The functional consequences of SNPs are tested *in vivo* and *in vitro*, and it is also of interest to understand what the structural effects are. Similarly, mutations of genes are responsible for a large number of disease states, and an understanding on the molecular level will aid in treatments. Only a fraction of all known proteins are being structure determined by X-ray and NMR, so how can we cope with all information from SNPs and mutants? Can we predict the effect of a mutation to focus experimental characterization efforts to the important targets? On the most coarse level, surface exposed residues are expected to influence the folding of a protein to lesser extent whereas it can modify the ability to interact with other molecules. Increased accuracy in the prediction of mutation effects is challenging and requires more sophisticated analysis. (Maiorov V and Abagyan R 1998; Rashin *et al.* 1997; Wright and Lim 2001)

3 Acknowledgments

We thank Vsevolod Katritch and Juan Fernández-Recio for helpful discussions and comments on the manuscript. L.B. is supported by a post-doctoral fellowship from the Human Frontier Science Program.

4 References

Abagyan R (1997) Protein structure prediction by global energy optimization. In: van Gunsteren WF, Weiner PK and Wilkinson AJ (eds) Computer simulation of biomolecular systems: Theoretical and experimental applications. (3rd edition) Kluwer Academic Publishers, Dordrecht. pp 363-394

Abagyan R (2000) ICM 2.8: Users manual. Molsoft L.L.C. La Jolla CA

Abagyan R and Argos P (1992) Optimal Protocol and Trajectory Visualization For Conformational Searches of Peptides and Proteins. J. Mol. Biol. 225:519-532

Abagyan R and Totrov M (2001) High-throughput docking for lead generation. Curr. Opin. Chem. Biol. 5:375-382

Abagyan R, Totrov M and Kuznetsov D (1994) ICM - A new method for protein modeling and design - Applications to docking and structure prediction from the distorted native conformation. J. Comp. Chem. 15:488-506

Abagyan RA and Batalov S (1997) Do aligned sequences share the same fold? J. Mol. Biol. 273:355-368

Abagyan RA and Totrov M (1999) Ab initio folding of peptides by the optimal-bias Monte Carlo minimization procedure. J. Comp. Phys. 151:402-421

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W and Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389-3402

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN and Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res. 28:235-242

Borchert TV, Abagyan R, Jaenicke R and Wierenga RK (1994) Design, Creation, and Characterization of a Stable, Monomeric Triosephosphate Isomerase. Proc. Natl. Acad. Sci. USA 91:1515-1518

Borchert TV, Abagyan R, Kishan KVR, Zeelen JP and Wierenga RK (1993) The Crystal-Structure of an Engineered Monomeric Triosephosphate Isomerase, Monotim - the Correct Modeling of an 8-Residue Loop. Structure 1:205-213

Chang G and Roth CB (2001) Structure of MsbA from E-coli: A homolog of the multidrug resistance ATP binding cassette (ABC) transporters. Science 293:1793-1800

Christendat D, Yee A, Dharamsi A, Kluger Y, Savchenko A, Cort JR, Booth V, Mackereth CD, Saridakis V, Ekiel I, Kozlov G, Maxwell KL, Wu N, McIntosh LP, Gehring K, Kennedy MA, Davidson AR, Pai EF, Gerstein M, Edwards AM and Arrowsmith CH (2000) Structural proteomics of an archaeon. Nature Struct. Biol. 7:903-909

Guntert P and Wuthrich K (2001) Sampling of conformation space in torsion angle dynamics calculations. Comp. Phys. Comm. 138:155-169

Laskowski RA, Macarthur MW, Moss DS and Thornton JM (1993) Procheck - a Program to Check the Stereochemical Quality of Protein Structures. J. Appl. Cryst. 26:283-291

Laskowski RA, Rullmann JAC, MacArthur MW, Kaptein R and Thornton JM (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR 8:477-486

Li Z and Scheraga H (1987) Monte Carlo minimization approach to the multiple-minima problem in protein folding. Proc. Nat. Acad. Sci. 84:6611-6615

Maiorov V and Abagyan R (1998) Energy strain in three-dimensional protein structures. Fold. & Des. 3:259-269

Momany R, McGuire R, Burgess A and Scheraga H (1975) Energy parameters in polypeptides. VII. Gemoetric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions and intrinsic torsional potentials for the naturally occurring amino acids. J. Phys. Chem. 79:2361

Morelli X, Dolla A, Czjzek M, Palma PN, Blasco F, Krippahl L, Moura JJG and Guerlesquin F (2000) Heteronuclear NMR and soft docking: An experimental approach for a structural model of the cytochrome c(553)-ferredoxin complex. Biochemistry 39:2530-2537

Nemethy G, Gibson KD, Palmer KA, Yoon CN, Paterlini G, Zagari A, Rumsey S and Scheraga HA (1992) Energy Parameters in Polypeptides .10. Improved Geometrical Parameters and Nonbonded Interactions For Use in the Ecepp/3 Algorithm, With Application to Proline-Containing Peptides. J. Phys. Chem. 96:6472-6484

Rashin AA, Rashin BH, Rashin A and Abagyan R (1997) Evaluating the energetics of empty cavities and internal mutations in proteins. Protein Sci. 6:2143-2158

Schapira M, Raaka BM, Samuels HH and Abagyan R (2000) Rational discovery of novel nuclear hormone receptor antagonists. Proc. Natl. Acad. Sci. USA 97:1008-1013

Schapira M, Raaka BM, Samuels HH and Abagyan R (2001) In silico discovery of novel retinoic acid receptor agonist structures. BMC Struct. Biol. 1:1-7

Shuker SB, Hajduk PJ, Meadows RP and Fesik SW (1996) Discovering high-affinity ligands for proteins: SAR by NMR. Science 274:1531-1534

Sternberg MJE, Gabb HA and Jackson RM (1998) Predictive docking of protein-protein and protein-DNA complexes. Curr. Opin. Struct. Biol. 8:250-256

Totrov M and Abagyan R (1994) Detailed Ab-Initio Prediction of Lysozyme-Antibody Complex With 1.6-Angstrom Accuracy. Nature Struct. Biol. 1:259-263

Totrov M and Abagyan R (2001) Rapid boundary element solvation electrostatics calculations in folding simulations: Successful folding of a 23-residue peptide. Biopolymers 60:124-133

Wright JD and Lim C (2001) A fast method for predicting amino acid mutations that lead to unfolding. Protein Eng. 14:479-486

Zhou YY and Abagyan R (1998) How and why phosphotyrosine-containing peptides bind to the SH2 and PTB domains. Fold. & Des. 3:513-522