



Published in final edited form as:

Structure. 2009 February 13; 17(2): 151–159. doi:10.1016/j.str.2008.12.014.

## Outcome of a Workshop on Applications of Protein Models in Biomedical Research

Torsten Schwede<sup>1</sup>, Andrej Sali<sup>2</sup>, Barry Honig<sup>3</sup>, Michael Levitt<sup>4</sup>, Helen M. Berman<sup>5</sup>, David Jones<sup>6</sup>, Steven E. Brenner<sup>7</sup>, Stephen K. Burley<sup>8</sup>, Rhiju Das<sup>9</sup>, Nikolay V. Dokholyan<sup>10</sup>, Roland L. Dunbrack Jr<sup>11</sup>, Krzysztof Fidelis<sup>12</sup>, Andras Fiser<sup>13</sup>, Adam Godzik<sup>14</sup>, Yuanpeng Janet Huang<sup>21</sup>, Christine Humblet<sup>15</sup>, Matthew P. Jacobson<sup>16</sup>, Andrzej Joachimiak<sup>17</sup>, Stanley R. Krystek Jr.<sup>18</sup>, Tanja Kortemme<sup>19</sup>, Andriy Kryshchuk<sup>20</sup>, Gaetano T. Montelione<sup>21</sup>, John Moulton<sup>22</sup>, Diana Murray<sup>23</sup>, Roberto Sanchez<sup>24</sup>, Tobin R. Sosnick<sup>25</sup>, Daron M. Standley<sup>26</sup>, Terry Stouch<sup>27</sup>, Sandor Vajda<sup>28</sup>, Max Vasquez<sup>29</sup>, John D. Westbrook<sup>5</sup>, and Ian A. Wilson<sup>30</sup>

<sup>1</sup> Swiss Institute of Bioinformatics, Biozentrum, University of Basel, Klingelbergstrasse 50-70, CH-4056 Basel, Switzerland <sup>2</sup> Department of Biopharmaceutical Sciences, Department of Pharmaceutical Chemistry, California Institute for Quantitative Biosciences, Byers Hall Room 503B, University of California at San Francisco, 1700 4th Street, San Francisco, CA 94158-2330 <sup>3</sup> Department of Biochemistry and Molecular Biophysics, HHMI, Center for Computational Biology and Bioinformatics, Columbia University, 1130 Saint Nicholas Ave, New York, NY 10032 <sup>4</sup> Department of Structural Biology, Stanford University School of Medicine, Fairchild Building, Stanford, CA 94305 <sup>5</sup> The Research Collaboratory for Structural Bioinformatics, Protein Data Bank, Rutgers, The State University of New Jersey, Piscataway, NJ 08854 <sup>6</sup> Department of Computer Science, University College London, Gower Street, London WC1E 6BT, UK <sup>7</sup> Department of Plant & Microbial Biology, 461A Koshland Hall #3102, University of California at Berkeley, Berkeley, CA 94720-3102 <sup>8</sup> SGX Pharmaceuticals, Inc., San Diego, CA 92121 <sup>9</sup> Department of Biochemistry, J Wing, Health Sciences Building Box 357350, University of Washington, Seattle, WA 98195 <sup>10</sup> Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 <sup>11</sup> Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, PA 19111 <sup>12</sup> Genome and Biomedical Sciences Facility, University of California at Davis, Davis, CA 95616 <sup>13</sup> Systems and Computational Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave, Bronx, NY 10461 <sup>14</sup> Bioinformatics and Systems Biology Program, The Burnham Institute for Medical Research, La Jolla, CA 92037 <sup>15</sup> Computational Chemistry/Cheminformatics Chemical & Screening Sciences, Wyeth Research, 865 Ridge Road, Monmouth Junction, NJ 08852 <sup>16</sup> Department of Pharmaceutical Chemistry, Genentech Hall Room N472C, University of California at San Francisco, 600 16th Street, San Francisco, CA 94158-2517 <sup>17</sup> Midwest Center for Structural Genomics and Structural Biology Center, Biosciences Division, Argonne National Laboratory, Argonne, IL 60439 <sup>18</sup> Bristol-Myers Squibb, Research & Discovery, P.O. Box 5400, Princeton, NJ 08543 <sup>19</sup> Department of Biopharmaceutical Sciences, University of California at San Francisco, 1700 4th Street, San Francisco, CA 94143-2540 <sup>20</sup> Protein Structure Prediction Center, University of California at Davis, 451 Health Sciences Drive, Davis, CA 95616 <sup>21</sup> Department of Molecular Biology and Biochemistry, Center for Advanced Research in Biotechnology, Rutgers, The State University of New Jersey, Piscataway, NJ 08854 <sup>22</sup> Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute, Rockville, MD 20850 <sup>23</sup> Center for Computational Biology and Bioinformatics, Columbia University, 622 West 168th Street, New York, NY 10032 <sup>24</sup> Department of Structural and Chemical Biology, Mount Sinai School of Medicine, 1425 Madison

All attendees of the Workshop, except students and NIH officers, are listed as authors.

Avenue, New York, NY 10029 <sup>25</sup> Biochemistry and Molecular Biology, The University of Chicago, 929 E. 57th Street, Chicago, IL 60637 <sup>26</sup> Immunology Frontier Research Center, Systems Immunology Laboratory, Osaka University, Osaka 565-0871, Japan <sup>27</sup> 6211 Kaitlyn Court, West Windsor, NJ 08550 <sup>28</sup> Biomedical Engineering Structural Bioinformatics Laboratory, Boston University, 44 Cummington Street, Boston, MA 02215 <sup>29</sup> Bocaraca, 2443 Ash St, Palo Alto, CA 94306 <sup>30</sup> Joint Center for Structural Genomics, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, CA 92037

## Summary

We describe the proceedings and conclusions from a “Workshop on Applications of Protein Models in Biomedical Research” that was held at University of California at San Francisco on 11 and 12 July, 2008. At the workshop, international scientists involved with structure modeling explored (i) how models are currently used in biomedical research, (ii) what the requirements and challenges for different applications are, and (iii) how the interaction between the computational and experimental research communities could be strengthened to advance the field.

## 1. Introduction: Background and Goals of the Workshop

### Molecular Modeling is Well-Established

Three-dimensional modeling of biological molecules and their interactions has a long history and is now established as a cornerstone of modern structural biology. Classic examples include the molecular model of the DNA double helix that was built by James Watson and Francis Crick in 1953 (Watson and Crick, 1953), models for the polypeptide  $\alpha$ -helix and  $\beta$ -sheet proposed by Linus Pauling some two years earlier (Pauling *et al.*, 1951), and the first homology model of a protein, built by David Phillips and coworkers for  $\alpha$ -lactalbumin based on hen egg white lysozyme (Browne *et al.*, 1969). While not every model can have the same impact as these early landmark examples, the potential of molecular modeling to produce new biological insights has never been greater than it is today, thanks to the recent explosion of sequence and structural data, advances in modeling methods and vastly more powerful computers.

### Types of Molecular Models

Protein structure prediction methods differ in terms of the needed input information and the aspects of protein structure that can be computed. The secondary structure, trans-membrane segments, and disordered regions can be predicted from a protein sequence (Bryson *et al.*, 2005; Rost, 2003); an atomic model of a domain can be obtained from the sequence alone by *ab initio* or *de novo* prediction methods (Das and Baker, 2008); fold assignment and sequence-structure alignment can be achieved by threading against a library of known folds (Godzik, 2003); atomic models of a protein can be calculated on the basis of known template structures by homology modeling (Marti-Renom *et al.*, 2000; Petrey and Honig, 2005; Schwede *et al.*, 2003); and atomic and reduced representation models of protein complexes with small ligands and other macromolecules, such as nucleic acids, can be derived with various docking methods (Lensink *et al.*, 2007). Increasingly, integrative or hybrid methods rely on more than one type of information, especially for the structural characterization of protein assemblies (Alber *et al.*, 2008).

### New Context for Modeling Provided by PSI

A stimulating catalyst for molecular modeling is the Protein Structure Initiative (PSI) that aims to determine representative atomic structures of most major protein families by X-ray crystallography and NMR spectroscopy, so that the remainder of protein sequences can be

characterized by homology modeling (<http://www.nigms.nih.gov/Initiatives/PSI/>) (Chandonia and Brenner, 2006; Liu et al., 2007). In the PSI, experimental structure determination and molecular modeling are especially mutually reinforcing. On the one hand, the experiments provide essential template structures for homology modeling specific sequences and the expanded dataset of protein structures provides opportunities for developing better modeling methods. On the other hand, modeling greatly leverages experimentally determined structures. By judicious selection of target proteins determined by experiment, each experimental structure enables the modeling of many protein sequences that could not be modeled well before (Liu et al., 2007). Molecular modeling can also add value to both experimentally determined structures and models; for example, docking of small molecules to proteins can be used for functional annotation (Hermann et al., 2007a) and docking of proteins can be used for characterization of large macromolecular machines (Lensink *et al.*, 2007). Finally, integrative methods have actually begun to improve the process of experimental structure determination itself (Alber *et al.*, 2007a; Qian *et al.*, 2007).

### PSI Knowledgebase and Protein Model Portal

To make the fruits of PSI available as widely as possible, the PSI Structural Genomics Knowledgebase was launched in February 2008 (<http://kb.psi-structuralgenomics.org>) (Berman, 2008). The Knowledgebase is designed to provide a “marketplace of ideas” that connects protein sequence information to experimentally determined structures and computationally predicted models, enhances functional annotation, and facilitate access to new experimental protocols and materials. The initial version of the Knowledgebase is a web portal to a series of modules, including the Experimental Tracking, Material Repository, Models, Annotation, and Technology portals. The Protein Model Portal in particular provides access to models calculated by SWISS-MODEL (Kopp and Schwede, 2004), ModBase (Pieper *et al.*, 2006), as well as models produced by the four PSI large-scale production centers (<http://www.proteinmodelportal.org/>). Its design and implementation are based on the recommendations proposed at the Workshop on Biological Macromolecular Structure Models in 2005 (Berman *et al.*, 2006). The Model Portal aims to foster effective usage of molecular models in biomedical research by providing convenient and comprehensive access to the models and their annotations. An associated annual workshop will be a forum for developers and users of modeling methods on best practices, including methods for estimating model accuracy, guidelines for publishing theoretical models, and educational resources on using models for different biological applications. Thus, the Model Portal is a major opportunity to increase the impact of molecular modeling on biology and medicine.

### Workshop Aims

Sixty-four participants from 30 academic, industry, and government institutions worldwide, including 9 from non-US locations, attended a workshop at the University of California at San Francisco (<http://www.proteinmodelportal.org/workshop/>). The participants discussed state-of-the-art applications of molecular modeling to biomedical problems, the requirements and challenges for various applications, as well as ways to strengthen the collaboration between the modeling and experimental communities. While the workshop was concerned primarily with applications of homology modeling as a cornerstone of the PSI, other relevant molecular modeling areas were also covered, including application of modeling to improving experimental structure determination (*eg*, molecular replacement in X-ray crystallography) and the use of homology models in conjunction with other methods (*eg*, docking of small molecules and proteins). The participants’ consensus was formulated as specific recommendations, aimed to increase the impact of molecular modeling in biology and medicine.

## Workshop Program

On the first day, 16 presentations were given on topics that ranged from coverage of protein sequence-structure space (Section 2) to the uses of modeling in biology and medicine (Section 3). On the second day, four independent discussion groups were asked to address the same set of specific questions covering the topics of the workshop, report on their findings, and make recommendations for the future (Section 5). Thus, each set of participants approached the issues in their own way; the resulting redundancy provided a rich source of ideas revealing both a commonality and a diversity of opinions that are incorporated in this document.

## 2. Coverage of Sequence Space by Homology Modeling

The utility of molecular modeling hinges on its coverage and accuracy. In other words, modeling needs to be applicable to many proteins and the models need to be sufficiently accurate for biological applications. The coverage issue was addressed in a recent comprehensive analysis of the current sampling of the protein universe (Levitt, 2008). The protein universe is the set of protein sequences and structures in all organisms. It was explored in terms of sequence families that have single or multi-domain architectures, with or without known structures. The domains were defined based on the CDART resource at NCBI (Geer *et al.*, 2002), which contains almost 30,000 domain families. Growth of single domain families has now saturated: almost all current growth comes from multi-domain architectures that are combinations of single domains. Structures are known for a quarter of the single-domain families and half of all known sequences can be partially modeled due to their membership in these families; 20% of the structures for such modeling come from the structural genomics effort, in particular, from the PSI. Multi-domain architecture families continue to grow rapidly and at the same rate as deposited sequences; almost all novelty, therefore, arises from the arrangement of known single domains within a chain, particularly for eukaryotes. A quarter of the sequences do not appear to match any domain pattern and constitute the dark matter of the protein universe.

These empirical observations demonstrate the relatively high degree of applicability of homology modeling and the important role that structural genomics plays in increasing this coverage. Moreover, the generation of novel proteins through combining individual domains increases the importance of molecular docking as a means to characterize the structures of the multi-domain proteins.

## 3. Applications of Modeling for Biology and Medicine

Modeling is not only widely applicable (Section 2), but often sufficiently accurate to make an impact on biology and medicine. To demonstrate this point, we do not discuss here the purely technical measures of geometrical accuracy of a model; instead, we focus on the bottom line corresponding to the numerous published studies where models have helped provide important biological insights. In most examples presented at the Workshop, the models have been combined with experimental efforts to produce results of significant biomedical impact. Therefore, despite its remaining limitations, modeling can certainly add substantial value to experimentally determined protein structures.

### 3.1 Drug Discovery

Homology modeling is widely applied in the pharmaceutical industry and is integrated into most stages of pharmaceutical research (Tramontano, 2006). For example, it is used to design protein constructs and to enhance protein production, solubility, and crystallization. Once a protein is established as a viable pharmaceutical target, homology modeling is used in assay development, compound screening, identification of biologically active small molecules, and further optimization of the potency of those compounds.

Homology models are used in “structure-based ligand discovery”, facilitating investigation of ligand-protein interactions in an effort to find ligands and improve their potency (Rester, 2008). One technique, “virtual screening”, computationally screens large libraries of organic molecules for those that complement the structure of a protein binding site (Huey *et al.*, 2007). Success rates for identifying compounds with biological activity range typically from 1% to 15% of those molecules that are predicted to bind (Babaoglu *et al.*, 2008; Doman *et al.*, 2002). This relatively high false positive rate reflects the remaining challenges with accurate prediction of affinity. Nevertheless, virtual screening was found to be as useful as experimental “high-throughput screening” in side-by-side prospective studies (Babaoglu *et al.*, 2008; Doman *et al.*, 2002). Homology models accelerate the virtual screening process and can help make helpful suggestions before crystal structures are available or experimental high-throughput screening begins (Oshiro *et al.*, 2004).

Other applications of structural models involve “optimization” of hits from virtual screening or high throughput screening by detailed examination of the ligand-protein interactions and the exploitation of new contacts with the protein *via* ligand modification (Noble *et al.*, 2004). The discovery and development of neuraminidase inhibitors is an important case where structure-based methods were used to guide the design of the first anti-influenza drug Relenza (zanamivir), brought to market by GlaxoSmithKline (von Itzstein *et al.*, 1993). Coupled with informed molecular biology efforts, even crude homology models based on remotely related structures have been successful in facilitating drug discovery (de Paulis, 2007). Modeling is especially robust and informative when used in a target class mode; for example, homology modeling of kinases has been applied to ligand discovery, as well as optimization of binding potency and selectivity (Buckley *et al.*, 2008; Diller and Li, 2003; Rockey and Elcock, 2006). Long before experimental structures of GPCRs were determined, models helped the selection and introduction of GPCR ligands into the clinic (Engel *et al.*, 2008; Webb and Krystek; Webb *et al.*, 1996). Clearly, the recent GPCR structures (Cherezov *et al.*, 2007; Rasmussen *et al.*, 2007; Warne *et al.*, 2008) will further aid modeling of this important class of biological targets.

### 3.2 Biotherapeutics, Biologicals, and Industrial Enzymes

Several biotherapeutics have been developed with the aid of homology modeling. Antibody construct design and humanization is a mature field (Lippow *et al.*, 2007). Of the 21 antibodies on the market as of 2007, it is estimated that 11 were the result of computational design of humanized constructs *via* homology modeling. Three examples are Zenapax (humanized anti-Tac or daclizumab), Herceptin (humanized anti-HER2 or trastuzumab), and Avastin (humanized anti-VEGF or bevacizumab) (Carter *et al.*, 1992; Presta *et al.*, 1997; Queen *et al.*, 1989). Many more have reached clinical trials. Similar techniques have been used to engineer smaller antibody fragments with improved specificity, affinity, and half-life (Hinton *et al.*, 2004; Lazar *et al.*, 2006; Lippow *et al.*, 2007).

Enzymes and other biologicals are widely used in biotechnology and industrial processes; they are key components of detergents and animal feed, and are used in the production of bread, wine and fruit juice, as well as in the treatment of textiles, paper, and leather. Enzymes frequently replace traditional chemicals or additives and help to save water and energy in a variety of production processes. Molecular modeling often provides the basis for understanding and engineering their biophysical properties, such as stability at high temperature and oxidation, activity at low temperatures, and substrate specificity (Alquati *et al.*, 2002; Hult and Berglund, 2003).

### 3.3 Protein-Protein Interactions

Most proteins act in the cell through interactions with other proteins. Therefore, the impact of individual models, as well as experimentally determined atomic structures, can be increased



by computational docking methods that produce models of protein complexes. The need for computational docking is emphasized by the difficulty of experimental structure determination for complexes, especially the more transient ones. Despite remaining challenges, the results of the CAPRI effort (Critical Assessment of Predicted Interactions) (Janin *et al.*, 2003) demonstrate that substantial progress in docking methods has been made during the last few years (Lensink *et al.*, 2007). The ClusPro docking server, which returns best-scoring models of a complex between two input atomic structures or models, is a case in point (Comeau *et al.*, 2004). The main applications of the server have included modeling multi-domain proteins and oligomers, frequently in combination with additional data from experimental or other computational techniques.

For example, the configuration of the histone domain relative to the Dbl-homology, pleckstrin-homology and catalytic domains in the Ras-specific nucleotide exchange factor son of sevenless (SOS) was determined by filtering top scoring docking models by small-angle X-ray scattering, mutagenesis, and calorimetry data (Sondermann *et al.*, 2005); the orientation and position of the histone domain implicated it as a potential mediator of membrane-dependent activation signals. Similarly, the high-resolution solution structure of the 15.4 kDa homodimer CylR2, the regulator of cytolysin production from *Enterococcus faecalis*, was solved by combining paramagnetic relaxation enhancement data with docking (Rumpel *et al.*, 2008). Further, the binding of cofilin to monomeric actin (Kamal *et al.*, 2007) was characterized by a combination of docking with mass spectrometry data (Kamal and Chance, 2008). Additional examples of docking include a model of the human p53-controlled ribonucleic reductase (p53R2) homodimer, which was used to explain mutations that cause mitochondrial DNA depletion (Bourdon *et al.*, 2007); and an L-type Ca<sup>2+</sup> channel, which was used for the characterization of binding interactions with 1,4-dihydropyridines (Cosconati *et al.*, 2007).

### 3.4 Membrane Binding Specificity

The recognition by peripheral membrane proteins of different biological membranes and distinct phospholipids underlies a variety of signaling processes. What is the molecular basis of these recognition mechanisms? In close collaboration with experimental groups, modelers studied this problem by first building homology models of proteins, both within functional families and across genomes, and then predicting the sub-cellular localization of proteins based on the calculated electrostatic properties of those models. For example, a computational study of structures and models for all retroviral matrix domains, such as those from HIV-1, revealed that matrix domains contain a characteristic basic surface patch and, thus, exploit electrostatic interactions to bind membrane surfaces (Dalton *et al.*, 2005; Murray *et al.*, 2005). This discovery provides insight into the mechanism used by matrix domains to localize to the plasma membrane of infected cells.

The construction of models of the membrane binding domains from different families (Ananthanarayanan *et al.*, 2002; Blatner *et al.*, 2004; Stahelin *et al.*, 2004; Yu *et al.*, 2004) also illustrates how homology modeling allows the identification of functional properties of proteins that are different than a family member whose structure has been determined by experiment. Specifically, calculations with a homology model for the PX domain from phospholipase D-1 showed that this domain binds membranes containing the cellular growth-inducing PI, PI (3,4,5)P<sub>3</sub>, primarily through electrostatic interactions, although the model was built using the structure of a PX domain that binds to PI(3,4)P<sub>2</sub>-containing membranes with significant hydrophobic penetration (Stahelin *et al.*, 2004).

### 3.5 Ligand Specificity of Receptors

Members of the NSS transporter family are responsible for uptake of neurotransmitters (such as glycine,  $\gamma$ -amino butyric acid, serotonin, dopamine, and norepinephrine) from the synaptic

cleft; mutations in NSS transporters have been implicated in psychological and digestive disorders including schizophrenia. Furthermore, several NSS transporters have been shown to be targets for psychoactive compounds such as cocaine. Thus, an understanding of the molecular mechanisms underlying transport by these proteins is of considerable interest. It has been extremely difficult to crystallize mammalian members of this family, but bacterial substitutes have been more tractable. These structures can then be used as templates to construct homology models of mammalian homologs, which in turn can be used to deduce function. In a specific example, the chloride binding site of the serotonin transporter, SerT, was identified from a homology model built from the previously published structure of a bacterial amino acid transporter, LeuT, which does not bind chloride (Forrest *et al.*, 2007). The prediction was confirmed experimentally. The work was highlighted in an Editor's Choice in *Science*, emphasizing the importance of homology modeling to this class of problems (Chin and Yeston, 2007).

### 3.6 Substrate Specificity of Enzymes

Many enzymes encoded by sequenced genomes and metagenomes have unknown functions. One promising approach to leverage structures for functional annotation is to dock libraries of possible substrates or chemical intermediates against the enzyme active site (Hermann *et al.*, 2006; Hermann *et al.*, 2007b; Kalyanaraman *et al.*, 2005). Homology models can extend the utility of this approach to the many uncharacterized enzymes lacking experimental structures, and enable prediction of substrate specificity among related enzymes in protein families.

In a joint computational and experimental effort, homology models were created for approximately 100 homologs of an Ala-Glu epimerase enzyme for which a crystallographic structure was available (Kalyanaraman *et al.*, 2008). Docking possible substrates against the models suggested that many had different substrate specificities and, hence, biological functions. Subsequent experimental screening confirmed several novel functions, including N-succinyl-Arg racemase (Song *et al.*, 2007) and Ala-Phe epimerase (Kalyanaraman *et al.*, 2008), and crystal structures confirmed the predicted binding modes. Because enzyme specificity is related to fine details of the binding site, such as precise orientations of side chains, one promising approach is to treat the binding site of homology models as flexible during docking, reducing the sensitivity of the results to small errors in the model (Hamblin *et al.*, 2008; Song *et al.*, 2007).

### 3.7 Analysis of Mutations

The onrush of personal genetic data adds new urgency for more effective computational analysis of the structural and functional impact of mutations, such as non-synonymous, single DNA base variants (*ie*, those that change the encoded amino acid residue type) (Karchin *et al.*, 2007). Exon sequencing is already providing single base somatic mutation information in individual cancer cell lines. Many more data of this type are expected shortly (Di Bernardo *et al.*, 2008; Sjoblom *et al.*, 2006; Stacey *et al.*, 2008). It is impossible to characterize functional consequences of all mutations by experiment, because there are too many of them. Therefore, computational approaches are required that are based on general principles of protein evolution, structure, and function. Full utilization of the mass of mutation data will require knowledge of the structure of human proteins, and that knowledge will come primarily from models.

With a particular machine learning method, homology models based on experimental templates down to 40% sequence identity provide as accurate a prediction of functional impact of a DNA base variant as do experimental structures (Yue *et al.*, 2005). Use of these models doubles the number of human common base variants that can be fully analyzed for likely impact, compared with using experimentally determined structures alone. Further improvements in modeling methods enabling the use of models based on sequence identity down to 20% would add a

further 50% to the number of analyzable single point mutations. Recent progress measured in the CASP experiments (Kopp *et al.*, 2007; Kryshafovich *et al.*, 2007) suggests this coverage is not an unreasonable expectation. A particularly successful example is provided in the next section.

### 3.8 Cancer Biology

Homology modeling and other computational tools have also been used to study structure-function relationships of proteins involved in DNA repair, cell cycle progression, chromatin formation, apoptosis, and other cellular processes associated with cancer development. Recent examples include explaining mutant phenotypes in a complex of yeast cyclin C and its cyclin-dependent kinase, cdk8p (Krasley *et al.*, 2006), analysis of patient-derived mutants of c-kit in gastrointestinal stromal tumors (Tarn *et al.*, 2005), and a prediction of the docking structure of BAK with p53 in apoptosis that relied on structure-based design of mutants (Pietsch *et al.*, 2008).

One of the most useful applications of molecular modeling in cancer biology is to dissect the roles of multiple interacting proteins in various pathways associated with cancer (Huang *et al.*, 2008). As an example, collaboration between experimental biologists and molecular modelers at the Fox Chase Cancer Center was aimed at understanding different phenotypes of overexpression of the chromatin remodeling protein ASF1a in humans (Tang *et al.*, 2006; Zhang *et al.*, 2005). Overexpression of this protein causes two different phenotypes: an increase in the formation of senescence-associated heterochromatin foci (SAHF) and G2-cell-cycle arrest. A homology model of the human ASF1a protein was constructed based on an experimentally determined yeast protein structure. It was found that mutations affecting SAHF formation were clustered together at one end, whereas mutations that did not affect SAHF formation were scattered in other regions of the structure (Tang *et al.*, 2006). To investigate the cell-cycle arrest phenotype, modelers searched for a cluster of surface residues elsewhere in the model that were conserved within ASF1a, but different from ASF1b (which does not exhibit the cell-cycle arrest phenotype). Mutations of residues that were predicted to affect cell-cycle arrest, but not the SAHF phenotype, were subsequently verified experimentally.

## 4. Integrative or Hybrid Structure Determination Methods

Molecular modeling plays an increasing role in experimental structure determination. In point of fact, the experimentally or theoretically derived information about a structure being determined must always be converted to an explicit structural model through computation. The “integrative” or “hybrid” approaches explicitly combine diverse experimental and theoretical information, with the aim to increase the accuracy, precision, coverage, and efficiency of structure determination (Alber *et al.*, 2008; Robinson *et al.*, 2007). Input information may vary greatly in terms of resolution (*i.e.*, precision), accuracy, and quantity. To be precise, all structure determination methods are integrative, but there is a difference in degree. At one end of the spectrum, even atomic structure determinations by X-ray crystallography and NMR spectroscopy rely on a molecular mechanics force field as well as on the “raw” X-ray and NMR data, respectively. An archetypal hybrid method is flexible docking of comparative models for component proteins into an electron density map of their assembly determined by cryo-electron microscopy (Rossmann *et al.*, 2005; Topf *et al.*, 2008). Such hybrid methods begin to blur the distinction between models based primarily on theoretical considerations and those based primarily on experimental data about the characterized system.

### 4.1 Atomic Structure Determination

Modelers have begun to contribute directly to atomic structure determination of proteins. In crystallography, *de novo* protein structure prediction can sometimes solve the phase problem,



via molecular replacement models for proteins of distant homology or even no detectable homology to previously solved structures (Qian *et al.*, 2007). In structure determination by satisfaction of NMR-derived restraints, high-resolution physics-based refinement can now consistently improve the accuracy of NMR model ensembles (Bhattacharya *et al.*, 2008; Qian *et al.*, 2007). Perhaps most promising are methods that can dramatically accelerate NMR-based structural inference, by bringing together limited chemical shift data with modeling techniques to achieve structures with near-atomic resolution (Cavalli *et al.*, 2007; Shen *et al.*, 2008).

#### 4.2 Structural Characterization of Large Assemblies at Low Resolution

Even low-resolution biophysical and biochemical data can provide a rich source of structural information that can be integrated into realistic representations of macromolecular assemblies, as shown by determining the positions of the 456 constituent proteins in the yeast nuclear pore complex (NPC) (Alber *et al.*, 2007a; Alber *et al.*, 2007b). The structure was determined at approximately 5 nm resolution by satisfying spatial restraints that encoded protein and nuclear envelope excluded volumes (from the protein sequences and ultracentrifugation), protein positions (from immunoelectron microscopy), protein contacts (from affinity purification), and the eight-fold and two-fold symmetries of the NPC (from electron microscopy). Although each individual restraint may contain little structural information, the concurrent satisfaction of all restraints derived from independent experiments drastically reduced the degeneracy of the structural solutions. The resulting low-resolution map was combined with atomic structures and homology models of constituent proteins, resulting in insights about the evolution and function of the NPC. This study illustrates how structural genomics and the PSI can make a major impact even on the most challenging structural biology problems, through providing atomic structures and homology models of the individual proteins that are then assembled into models of large macromolecular machines and processes.

### 5. Recommendations

We now summarize the recommendations reached by consensus among the four independent workshop discussion groups. The recommendations are concerned with (i) coverage of the sequence space by homology modeling; (ii) publication and archiving of models; (iii) standards for data formats; (iv) estimating model accuracy; (v) communication between modelers and experimentalists; and (vi) development and role of the Protein Model Portal.

#### 5.1 Coverage of Sequence Space by Homology Modeling Needs to be Quantified

As discussed above, modeling can significantly expand the structural coverage of the protein universe. It remains unclear how best to integrate the experimental structure determination and computational modeling to maximize the impact of structural genomics on biology. The present focus of the PSI on large families that have no structural representatives and on very large families with limited structural coverage is a promising approach to achieve this goal.

**Recommendation**—We recommend that the modeling and structural genomics communities interact closely to formulate how maximizing the structural coverage can be most efficiently achieved. Suitable metrics for measuring structural coverage must be developed by the modeling community. Once these metrics are adopted, the PSI Knowledgebase will continually update and report them.

#### 5.2 Standards for the Publication of Models Must be Established

**Journal Requirements**—At the present time, models are published with different amounts of information about how these models were derived. A set of guidelines for what should be included in a modeling paper needs to be established. For homology modeling, these guidelines may include decisions leading to choice of the template structure(s), details of sequence

alignment, methods used to derive the model, indication of the expected accuracy of the model, and how the model may be accessed publicly. These guidelines should be shared with journal editors and reviewers.

**Model Access and Archiving**—Models that have been peer reviewed and referred to in published literature should be publicly available. Without access to the model coordinates and sufficient annotation of the model, it is impossible for the reader to interpret the results and to assess the validity of the interpretations. In the past, some of the models were archived in the Protein Data Bank (PDB). Since 2006, only structures that have been determined experimentally are allowed to be deposited in the PDB (Berman *et al.*, 2006).

**Recommendation**—We recommend that a Model Working Group be established to set standards for journal publication, to define minimum annotation standards, and to establish the scope and requirements of a public archive of *in silico* models. Membership of this group should consist of a representative of the wwPDB (Berman *et al.*, 2003), the Protein Model Portal, as well as members of the modeling and user communities.

### 5.3 Standards for Data Formats Must be Established to Facilitate Data and Software Exchange

While the experimental structural biology community has essentially reached a consensus on the definition of common data formats that enable the seamless exchange of data and algorithms (Westbrook and Fitzgerald, 2003; Winn, 2003), most software tools for protein structure modeling use proprietary data formats for input data, parameters, and results. Although data formats from experimental structures can be applied to the protein model coordinates, data types specific to computational modeling, such as target-template alignments, error estimates, force field parameters, and specific details of the individual modeling algorithms, frequently vary between different applications. This incompatibility is a serious impediment for the exchange of tools and algorithms; it hinders both method development and the widespread use of tools outside of the developer groups themselves.

**Recommendation**—We recommend that the Model Working Group initiates a community-wide mechanism for reaching an agreement on a common open data format for information related to molecular modeling, with the aim of facilitating the exchange of algorithms and data. Once these standards are established, the services offered by the Protein Model Portal should be based exclusively on these common formats.

### 5.4 Standards for the Assessment of Models must be Established

**Model Accuracy Criteria**—As with structures determined by X-ray crystallography and other methods, accuracy can be estimated globally, akin to the crystallographic R-value, or locally, akin to residue-specific, real-space correlation coefficients and R-values. Applications of models strongly depend on their accuracy, with different applications having varied requirements on accuracy and precision. Even if the overall accuracy of the model is high, the accuracy of specific regions (binding sites, loops, pockets, surface features, and overall fold) may vary. Criteria based on the global correctness of C $\alpha$  coordinates are often insufficient to decide whether a model is suitable for a specific application, such as modeling ligand binding (Kopp *et al.*, 2007). Accuracy measures that convey the suitability of models for specific applications need to be established.

**Estimating Model Accuracy**—Methods for estimating model accuracy are being actively studied. No accurate or dominant method has yet emerged. In one type of approach, global and local model properties are compared against expected values from statistical analyses of experimentally determined structures, such as main-chain dihedral angle distributions, rotamer

probabilities, and solvation properties (Benkert et al., 2008; Bhattacharya et al., 2008; Pettitt et al., 2005; Shen and Sali, 2006; Sippl, 1993; Wallner and Elofsson, 2003). However, it is still possible for an inaccurate model to pass these checks. In cases where a number of independent models are available for a given target, consensus-based approaches can be applied (Ginalski et al., 2003; Wallner et al., 2003).

**Recommendation**—We recommend that the Model Working Group establishes guidelines for estimating model accuracy, with special emphasis on identifying criteria reflecting the suitability of models for specific biological applications. For this purpose, the Group should work most closely with members of the experimental research community representing specific model application requirements. The Protein Model Portal should provide a technical platform to make validated tools for estimating model accuracy available to the users of the models; it should also establish a mechanism for a continuous evaluation and improvement of these tools.

### 5.5 The Scientific Community Needs to be Aware of the Strengths and Limitations of Models

At present, many members of the scientific community are unaware of the advances in molecular modeling, its limitations and its applications. It is primarily the responsibility of modelers to educate the community about their area of research (*eg*, in the form of scientific publications, presentations, collaborative projects, and web resources). However, the Workshop participants felt that molecular modeling is often not used to its full potential in biomedical research, and that the impact of structural biology in general could be increased by better education about the optimal use of existing modeling methods.

**Recommendation**—We recommend that the PSI Knowledgebase and its Protein Model Portal proactively solicit educational contributions from the modeling community in the form of reviews, tutorials or even open workshops, aimed at demonstrating applications and limitations of computational modeling methods.

### 5.6 Protein Model Portal can Play a Key Role

The discussion at the Workshop explored how to maximize the impact of the Protein Model Portal (<http://www.proteinmodelportal.org/>) on the application of molecular models in biomedical research.

**Recommendation**—We recommend that the Portal provide unified access to molecular models and their annotations, and support the development of data standards to facilitate exchange of information and algorithms. The Portal should play an active role in facilitating discussions between developers of computational methods and their users, provide access to tools for estimating model accuracy, and promote their further development. Its user interface should allow a broad range of queries to the participating model databases as well as links to experimental data. Tools for estimating model errors and selecting the likely best model among the available models should be included. An interface to interactive services for modeling should be established. Mechanisms to notify users when a particular sequence is modeled (or experimental data becomes available) should be implemented. The Portal should work closely with the Knowledgebase to establish a series of online documents with community feedback to explain the value and limitations of protein structure models. Finally, the Portal should be as inclusive of all method developers and prediction methods as technically feasible.

## Acknowledgments

The workshop on Applications of Protein Models in Biomedical Research and the PSI Knowledgebase Protein Model Portal were supported by the National Institutes of Health (P20 GM076222-02S1, Roland Dunbrack, PI; and U54 GM074958-04S2, Helen Berman, PI).

## References

- Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, Devos D, Suprpto A, Karni-Schmidt O, Williams R, Chait BT, et al. Determining the architectures of macromolecular assemblies. *Nature* 2007a;683–694. 621–622.
- Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, Devos D, Suprpto A, Karni-Schmidt O, Williams R, Chait BT, et al. The molecular architecture of the nuclear pore complex. *Nature* 2007b; 450:695–701. [PubMed: 18046406]
- Alber F, Forster F, Korkin D, Topf M, Sali A. Integrating diverse data for structure determination of macromolecular assemblies. *Annu Rev Biochem* 2008;77:443–477. [PubMed: 18318657]
- Alquati C, De Gioia L, Santarossa G, Alberghina L, Fantucci P, Lotti M. The cold-active lipase of *Pseudomonas fragi*. Heterologous expression, biochemical characterization and molecular modeling. *Eur J Biochem* 2002;269:3321–3328. [PubMed: 12084074]
- Ananthanarayanan B, Das S, Rhee SG, Murray D, Cho W. Membrane targeting of C2 domains of phospholipase C-delta isoforms. *J Biol Chem* 2002;277:3568–3575. [PubMed: 11706040]
- Babaoglu K, Simeonov A, Irwin JJ, Nelson ME, Feng B, Thomas CJ, Cancian L, Costi MP, Maltby DA, Jadhav A, et al. Comprehensive mechanistic analysis of hits from high-throughput and docking screens against beta-lactamase. *J Med Chem* 2008;51:2502–2511. [PubMed: 18333608]
- Benkert P, Tosatto SC, Schomburg D. QMEAN: A comprehensive scoring function for model quality assessment. *Proteins* 2008;71:261–277. [PubMed: 17932912]
- Berman H, Henrick K, Nakamura H. Announcing the worldwide Protein Data Bank. *Nat Struct Biol* 2003;10:980. [PubMed: 14634627]
- Berman HM. Harnessing knowledge from structural genomics. *Structure* 2008;16:16–18. [PubMed: 18184578]
- Berman HM, Burley SK, Chiu W, Sali A, Adzhubei A, Bourne PE, Bryant SH, Dunbrack RL Jr, Fidelis K, Frank J, et al. Outcome of a workshop on archiving structural models of biological macromolecules. *Structure* 2006;14:1211–1217. [PubMed: 16955948]
- Bhattacharya A, Wunderlich Z, Monleon D, Tejero R, Montelione GT. Assessing model accuracy using the homology modeling automatically software. *Proteins* 2008;70:105–118. [PubMed: 17640066]
- Blatner NR, Stahelin RV, Diraviyam K, Hawkins PT, Hong W, Murray D, Cho W. The molecular basis of the differential subcellular localization of FYVE domains. *J Biol Chem* 2004;279:53818–53827. [PubMed: 15452113]
- Bourdon A, Minai L, Serre V, Jais JP, Sarzi E, Aubert S, Chretien D, de Lonlay P, Paquis-Flucklinger V, Arakawa H, et al. Mutation of RRM2B, encoding p53-controlled ribonucleotide reductase (p53R2), causes severe mitochondrial DNA depletion. *Nat Genet* 2007;39:776–780. [PubMed: 17486094]
- Browne WJ, North AC, Phillips DC, Brew K, Vanaman TC, Hill RL. A possible three-dimensional structure of bovine alpha-lactalbumin based on that of hen's egg-white lysozyme. *J Mol Biol* 1969;42:65–86. [PubMed: 5817651]
- Bryson K, McGuffin LJ, Marsden RL, Ward JJ, Sodhi JS, Jones DT. Protein structure prediction servers at University College London. *Nucleic Acids Res* 2005;33:W36–38. [PubMed: 15980489]
- Buckley GM, Ceska TA, Fraser JL, Gowers L, Groom CR, Higuero AP, Jenkins K, Mack SR, Morgan T, Parry DM, et al. IRAK-4 inhibitors. Part II: a structure-based assessment of imidazo[1,2-a]pyridine binding. *Bioorg Med Chem Lett* 2008;18:3291–3295. [PubMed: 18482836]
- Carter P, Presta L, Gorman CM, Ridgway JB, Henner D, Wong WL, Rowland AM, Kotts C, Carver ME, Shepard HM. Humanization of an anti-p185HER2 antibody for human cancer therapy. *Proc Natl Acad Sci U S A* 1992;89:4285–4289. [PubMed: 1350088]
- Cavalli A, Salvatella X, Dobson CM, Vendruscolo M. Protein structure determination from NMR chemical shifts. *Proc Natl Acad Sci U S A* 2007;104:9615–9620. [PubMed: 17535901]
- Chandonia JM, Brenner SE. The impact of structural genomics: expectations and outcomes. *Science* 2006;311:347–351. [PubMed: 16424331]
- Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK, Stevens RC. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* 2007;318:1258–1265. [PubMed: 17962520]

- Chin G, Yeston J. Studying Ions in Depth. *Science* 2007;317:873.
- Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics* 2004;20:45–50. [PubMed: 14693807]
- Cosconati S, Marinelli L, Lavecchia A, Novellino E. Characterizing the 1,4-dihydropyridines binding interactions in the L-type Ca<sup>2+</sup> channel: model construction and docking calculations. *J Med Chem* 2007;50:1504–1513. [PubMed: 17335186]
- Dalton AK, Murray PS, Murray D, Vogt VM. Biochemical characterization of rous sarcoma virus MA protein interaction with membranes. *J Virol* 2005;79:6227–6238. [PubMed: 15858007]
- Das R, Baker D. Macromolecular modeling with rosetta. *Annu Rev Biochem* 2008;77:363–382. [PubMed: 18410248]
- de Paulis T. Drug evaluation: PRX-00023, a selective 5-HT<sub>1A</sub> receptor agonist for depression. *Curr Opin Investig Drugs* 2007;8:78–86.
- Di Bernardo MC, Crowther-Swanepoel D, Broderick P, Webb E, Sellick G, Wild R, Sullivan K, Vijayakrishnan J, Wang Y, Pittman AM, et al. A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia. *Nat Genet* 2008;40:1204–1210. [PubMed: 18758461]
- Diller DJ, Li R. Kinases, homology models, and high throughput docking. *J Med Chem* 2003;46:4638–4647. [PubMed: 14561083]
- Doman TN, McGovern SL, Witherbee BJ, Kasten TP, Kurumbail R, Stallings WC, Connolly DT, Shoichet BK. Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J Med Chem* 2002;45:2213–2221. [PubMed: 12014959]
- Engel S, Skoumbourdis AP, Childress J, Neumann S, Deschamps JR, Thomas CJ, Colson AO, Costanzi S, Gershengorn MC. A virtual screen for diverse ligands: discovery of selective G protein-coupled receptor antagonists. *J Am Chem Soc* 2008;130:5115–5123. [PubMed: 18357984]
- Forrest LR, Tavoulari S, Zhang YW, Rudnick G, Honig B. Identification of a chloride ion binding site in Na<sup>+</sup>/Cl<sup>-</sup>-dependent transporters. *Proc Natl Acad Sci U S A* 2007;104:12761–12766. [PubMed: 17652169]
- Geer LY, Domrachev M, Lipman DJ, Bryant SH. CDART: protein homology by domain architecture. *Genome Res* 2002;12:1619–1623. [PubMed: 12368255]
- Ginalski K, Elofsson A, Fischer D, Rychlewski L. 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics* 2003;19:1015–1018. [PubMed: 12761065]
- Godzik A. Fold recognition methods. *Methods Biochem Anal* 2003;44:525–546. [PubMed: 12647403]
- Hamblin K, Standley DM, Rogers MB, Stechmann A, Roger AJ, Maytum R, van der Giezen M. Localization and nucleotide specificity of Blastocystis succinyl-CoA synthetase. *Mol Microbiol* 2008;68:1395–1405. [PubMed: 18452512]
- Hermann JC, Ghanem E, Li Y, Raushel FM, Irwin JJ, Shoichet BK. Predicting substrates by docking high-energy intermediates to enzyme structures. *J Am Chem Soc* 2006;128:15882–15891. [PubMed: 17147401]
- Hermann JC, Marti-Arbona R, Fedorov AA, Fedorov E, Almo SC, Shoichet BK, Raushel FM. Structure-based activity prediction for an enzyme of unknown function. *Nature* 2007a;448:775–779. [PubMed: 17603473]
- Hermann JC, Marti-Arbona R, Fedorov AA, Fedorov E, Almo SC, Shoichet BK, Raushel FM. Structure-based activity prediction for an enzyme of unknown function. *Nature* 2007b;448:775–779. [PubMed: 17603473]
- Hinton PR, Johlfs MG, Xiong JM, Hanestad K, Ong KC, Bullock C, Keller S, Tang MT, Tso JY, Vasquez M, Tsurushita N. Engineered human IgG antibodies with longer serum half-lives in primates. *J Biol Chem* 2004;279:6213–6216. [PubMed: 14699147]
- Huang YJ, Hang D, Lu LJ, Tong L, Gerstein MB, Montelione GT. Targeting the human cancer pathway protein interaction network by structural genomics. *Mol Cell Proteomics* 2008;7:2048–2060. [PubMed: 18487680]
- Huey R, Morris GM, Olson AJ, Goodsell DS. A semiempirical free energy force field with charge-based desolvation. *J Comput Chem* 2007;28:1145–1152. [PubMed: 17274016]



- Hult K, Berglund P. Engineered enzymes for improved organic synthesis. *Curr Opin Biotechnol* 2003;14:395–400. [PubMed: 12943848]
- Janin J, Henrick K, Moult J, Eyck LT, Sternberg MJ, Vajda S, Vakser I, Wodak SJ. CAPRI: a Critical Assessment of PRedicted Interactions. *Proteins* 2003;52:2–9. [PubMed: 12784359]
- Kalyanaraman C, Bernacki K, Jacobson MP. Virtual screening against highly charged active sites: Identifying substrates of alpha-beta barrel enzymes. *Biochem* 2005;44:2059–2071. [PubMed: 15697231]
- Kalyanaraman C, Imker HJ, Federov AA, Federov EV, Glasner ME, Babbitt PC, Almo SC, Gerlt JA, Jacobson MP. Discovery of a new dipeptide epimerase enzymatic function guided by homology modeling and virtual screening. *Structure*. 2008In Press
- Kamal JK, Benchaar SA, Takamoto K, Reisler E, Chance MR. Three-dimensional structure of cofilin bound to monomeric actin derived by structural mass spectrometry data. *Proc Natl Acad Sci U S A* 2007;104:7910–7915. [PubMed: 17470807]
- Kamal JK, Chance MR. Modeling of protein binary complexes using structural mass spectrometry data. *Protein Sci* 2008;17:79–94. [PubMed: 18042684]
- Karchin R, Monteiro AN, Tavtigian SV, Carvalho MA, Sali A. Functional impact of missense variants in BRCA1 predicted by supervised learning. *PLoS Comput Biol* 2007;3:e26. [PubMed: 17305420]
- Kopp J, Bordoli L, Battey JN, Kiefer F, Schwede T. Assessment of CASP7 predictions for template-based modeling targets. *Proteins* 2007;69(Suppl 8):38–56. [PubMed: 17894352]
- Kopp J, Schwede T. The SWISS-MODEL Repository of annotated three-dimensional protein structure homology models. *Nucleic Acids Res* 2004;32:D230–234. [PubMed: 14681401]
- Krasley E, Cooper KF, Mallory MJ, Dunbrack R, Strich R. Regulation of the oxidative stress response through Slf2p-dependent destruction of cyclin C in *Saccharomyces cerevisiae*. *Genetics* 2006;172:1477–1486. [PubMed: 16387872]
- Kryshafovich A, Fidelis K, Moult J. Progress from CASP6 to CASP7. *Proteins* 2007;69(Suppl 8):194–207. [PubMed: 17918728]
- Lazar GA, Dang W, Karki S, Vafa O, Peng JS, Hyun L, Chan C, Chung HS, Eivazi A, Yoder SC, et al. Engineered antibody Fc variants with enhanced effector function. *Proc Natl Acad Sci U S A* 2006;103:4005–4010. [PubMed: 16537476]
- Lensink MF, Méndez R, Wodak SJ. Docking and scoring protein complexes: CAPRI 3rd Edition. *Proteins* 2007;69:704–718. [PubMed: 17918726]
- Levitt M. *The Nature of the Protein Universe*. 2008
- Lippow SM, Wittrup KD, Tidor B. Computational design of antibody-affinity improvement beyond in vivo maturation. *Nat Biotechnol* 2007;25:1171–1176. [PubMed: 17891135]
- Liu J, Montelione GT, Rost B. Novel leverage of structural genomics. *Nat Biotechnol* 2007;25:849–851. [PubMed: 17687356]
- Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* 2000;29:291–325. [PubMed: 10940251]
- Murray PS, Li Z, Wang J, Tang CL, Honig B, Murray D. Retroviral matrix domains share electrostatic homology: models for membrane binding function throughout the viral life cycle. *Structure* 2005;13:1521–1531. [PubMed: 16216583]
- Noble ME, Endicott JA, Johnson LN. Protein kinase inhibitors: insights into drug design from structure. *Science* 2004;303:1800–1805. [PubMed: 15031492]
- Oshiro C, Bradley EK, Eksterowicz J, Evensen E, Lamb ML, Lanctot JK, Putta S, Stanton R, Grootenhuys PD. Performance of 3D-database molecular docking studies into homology models. *J Med Chem* 2004;47:764–767. [PubMed: 14736258]
- Pauling L, Corey RB, Branson HR. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci U S A* 1951;37:205–211. [PubMed: 14816373]
- Petrey D, Honig B. Protein structure prediction: inroads to biology. *Mol Cell* 2005;20:811–819. [PubMed: 16364908]
- Pettitt CS, McGuffin LJ, Jones DT. Improving sequence-based fold recognition by using 3D model quality assessment. *Bioinformatics* 2005;21:3509–3515. [PubMed: 15955780]

- Pieper U, Eswar N, Davis FP, Braberg H, Madhusudhan MS, Rossi A, Marti-Renom M, Karchin R, Webb BM, Eramian D, et al. MODBASE: a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res* 2006;34:D291–295. [PubMed: 16381869]
- Pietsch EC, Perchiniak E, Canutescu AA, Wang G, Dunbrack RL, Murphy ME. Oligomerization of BAK by p53 utilizes conserved residues of the p53 DNA binding domain. *J Biol Chem* 2008;283:21294–21304. [PubMed: 18524770]
- Presta LG, Chen H, O'Connor SJ, Chisholm V, Meng YG, Krummen L, Winkler M, Ferrara N. Humanization of an anti-vascular endothelial growth factor monoclonal antibody for the therapy of solid tumors and other disorders. *Cancer Res* 1997;57:4593–4599. [PubMed: 9377574]
- Qian B, Raman S, Das R, Bradley P, McCoy AJ, Read RJ, Baker D. High-resolution structure prediction and the crystallographic phase problem. *Nature* 2007;450:259–264. [PubMed: 17934447]
- Queen C, Schneider WP, Selick HE, Payne PW, Landolfi NF, Duncan JF, Avdalovic NM, Levitt M, Junghans RP, Waldmann TA. A humanized antibody that binds to the interleukin 2 receptor. *Proc Natl Acad Sci U S A* 1989;86:10029–10033. [PubMed: 2513570]
- Rasmussen SG, Choi HJ, Rosenbaum DM, Kobilka TS, Thian FS, Edwards PC, Burghammer M, Ratnala VR, Sanishvili R, Fischetti RF, et al. Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature* 2007;450:383–387. [PubMed: 17952055]
- Rester U. From virtuality to reality - Virtual screening in lead discovery and lead optimization: a medicinal chemistry perspective. *Curr Opin Drug Discov Devel* 2008;11:559–568.
- Robinson CV, Sali A, Baumeister W. The molecular sociology of the cell. *Nature* 2007;450:973–982. [PubMed: 18075576]
- Rockey WM, Elcock AH. Structure selection for protein kinase docking and virtual screening: homology models or crystal structures? *Curr Protein Pept Sci* 2006;7:437–457. [PubMed: 17073695]
- Rossmann MG, Morais MC, Leiman PG, Zhang W. Combining X-ray crystallography and electron microscopy. *Structure* 2005;13:355–362. [PubMed: 15766536]
- Rost B. Prediction in 1D: secondary structure, membrane helices, and accessibility. *Methods Biochem Anal* 2003;44:559–587. [PubMed: 12647405]
- Rumpel S, Becker S, Zweckstetter M. High-resolution structure determination of the CylR2 homodimer using paramagnetic relaxation enhancement and structure-based prediction of molecular alignment. *J Biomol NMR* 2008;40:1–13. [PubMed: 18026911]
- Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res* 2003;31:3381–3385. [PubMed: 12824332]
- Shen MY, Sali A. Statistical potential for assessment and prediction of protein structures. *Protein Sci* 2006;15:2507–2524. [PubMed: 17075131]
- Shen Y, Lange O, Delaglio F, Rossi P, Aramini JM, Liu G, Eletsky A, Wu Y, Singarapu KK, Lemak A, et al. Consistent blind protein structure generation from NMR chemical shift data. *Proc Natl Acad Sci U S A* 2008;105:4685–4690. [PubMed: 18326625]
- Sippl MJ. Recognition of errors in three-dimensional structures of proteins. *Proteins* 1993;17:355–362. [PubMed: 8108378]
- Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, et al. The consensus coding sequences of human breast and colorectal cancers. *Science* 2006;314:268–274. [PubMed: 16959974]
- Sondermann H, Nagar B, Bar-Sagi D, Kuriyan J. Computational docking and solution x-ray scattering predict a membrane-interacting role for the histone domain of the Ras activator son of sevenless. *Proc Natl Acad Sci U S A* 2005;102:16632–16637. [PubMed: 16267129]
- Song L, Kalyanaraman C, Fedorov AA, Fedorov EV, Glasner ME, Brown S, Babbitt PC, Almo SC, Jacobson MP, AGJ. Assignment and Prediction of Function in the Enolase Superfamily: A Divergent N Succinyl Amino Acid Racemase from *Bacillus cereus*. *Nature Chemical Biology* 2007;3:486–491.
- Stacey SN, Manolescu A, Sulem P, Thorlacius S, Gudjonsson SA, Jonsson GF, Jakobsdottir M, Bergthorsson JT, Gudmundsson J, Aben KK, et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* 2008;40:703–706. [PubMed: 18438407]

- Stahelin RV, Ananthanarayanan B, Blatner NR, Singh S, Bruzik KS, Murray D, Cho W. Mechanism of membrane binding of the phospholipase D1 PX domain. *J Biol Chem* 2004;279:54918–54926. [PubMed: 15475361]
- Tang Y, Poustovoitov MV, Zhao K, Garfinkel M, Canutescu A, Dunbrack R, Adams PD, Marmorstein R. Structure of a human ASF1a-HIRA complex and insights into specificity of histone chaperone complex assembly. *Nat Struct Mol Biol* 2006;13:921–929. [PubMed: 16980972]
- Tarn C, Merkel E, Canutescu AA, Shen W, Skorobogatko Y, Heslin MJ, Eisenberg B, Birbe R, Patchefsky A, Dunbrack R, et al. Analysis of KIT mutations in sporadic and familial gastrointestinal stromal tumors: therapeutic implications through protein modeling. *Clin Cancer Res* 2005;11:3668–3677. [PubMed: 15897563]
- Topf M, Lasker K, Webb B, Wolfson H, Chiu W, Sali A. Protein structure fitting and refinement guided by cryo-EM density. *Structure* 2008;16:295–307. [PubMed: 18275820]
- Tramontano A. The role of molecular modelling in biomedical research. *FEBS Lett* 2006;580:2928–2934. [PubMed: 16647064]
- von Itzstein M, Wu WY, Kok GB, Pegg MS, Dyason JC, Jin B, Phan VT, Smythe ML, White HF, Oliver WS, et al. Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* 1993;363:418–423. [PubMed: 8502295]
- Wallner B, Elofsson A. Can correct protein models be identified? *Protein Sci* 2003;12:1073–1086. [PubMed: 12717029]
- Wallner B, Fang H, Elofsson A. Automatic consensus-based fold recognition using Pcons, ProQ, and Pmodeller. *Proteins* 53 Suppl 2003;6:534–541.
- Warne T, Serrano-Vega MJ, Baker JG, Moukhametzianov R, Edwards PC, Henderson R, Leslie AG, Tate CG, Schertler GF. Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature* 2008;454:486–491. [PubMed: 18594507]
- Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 1953;171:737–738. [PubMed: 13054692]
- Webb ML, Krystek SR, Jr. Molecular and Structural Biology of Endothelin Receptors. In: Pollock, DM.; Highsmith, RF., editors. *Endothelin Receptors and Signaling Mechanisms*. Austin, TX: Landes Bioscience; 1998.
- Webb ML, Patel PS, Rose PM, Liu EC, Stein PD, Barrish J, Lach DA, Stouch T, Fisher SM, Hadjilambris O, et al. Mutational analysis of the endothelin type A receptor (ETA): interactions and model of selective ETA antagonist BMS-182874 with putative ETA receptor binding cavity. *Biochemistry* 1996;35:2548–2556. [PubMed: 8611558]
- Westbrook JD, Fitzgerald PM. The PDB format, mmCIF, and other data formats. *Methods Biochem Anal* 2003;44:161–179. [PubMed: 12647386]
- Winn MD. An overview of the CCP4 project in protein crystallography: an example of a collaborative project. *J Synchrotron Radiat* 2003;10:23–25. [PubMed: 12511787]
- Yu JW, Mendrola JM, Audhya A, Singh S, Keleti D, DeWald DB, Murray D, Emr SD, Lemmon MA. Genome-wide analysis of membrane targeting by *S. cerevisiae* pleckstrin homology domains. *Mol Cell* 2004;13:677–688. [PubMed: 15023338]
- Yue P, Li Z, Moulton J. Loss of protein structure stability as a major causative factor in monogenic disease. *J Mol Biol* 2005;353:459–473. [PubMed: 16169011]
- Zhang R, Poustovoitov MV, Ye X, Santos HA, Chen W, Daganzo SM, Erzberger JP, Serebriiskii IG, Canutescu AA, Dunbrack RL, et al. Formation of MacroH2A-containing senescence-associated heterochromatin foci and senescence driven by ASF1a and HIRA. *Dev Cell* 2005;8:19–30. [PubMed: 15621527]