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

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

[1]Swiss Institute of Bioinformatics, Biozentrum, University of Basel, Klingelbergstrasse 50-70, CH-4056 Basel, Switzerland
[2]Department of Biopharmaceutical Sciences, Department of Pharmaceutical Chemistry, California Institute for Quantitative Biosciences, Byers Hall, Room 503B, University of California, San Francisco, 1700 4th Street, San Francisco, CA 94158, USA
[3]Department of Biochemistry and Molecular Biophysics, Howard Hughes Medical Institute,
Center for Computational Biology and Bioinformatics, Columbia University, 1130 Saint Nicholas Avenue, New York, NY 10032, USA
[4]Department of Structural Biology, Stanford University School of Medicine, Fairchild Building, Stanford, CA 94305, USA
[5]The Research Collaboratory for Structural Bioinformatics, Protein Data Bank, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA
[6]Department of Computer Science, University College London, Gower Street, London WC1E 6BT, UK
[7]Department of Plant and Microbial Biology, 461A Koshland Hall, #3102, University of California, Berkeley, Berkeley, CA 94720
[8]SGX Pharmaceuticals, Inc., San Diego, CA 92121, USA
[9]Department of Biochemistry, J Wing, Health Sciences Building, Box 357350, University of Washington, Seattle, WA 98195, USA
[10]Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
[11]Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, PA 19111, USA
[12]Genome and Biomedical Sciences Facility, University of California, Davis, Davis, CA 95616, USA
[13]Systems and Computational Biology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA
[14]Bioinformatics and Systems Biology Program, The Burnham Institute for Medical Research, La Jolla, CA 92037, USA
[15]Department of Molecular Biology and Biochemistry, Center for Advanced Research in Biotechnology, Rutgers,
The State University of New Jersey, Piscataway, NJ 08854, USA
[16]Computational Chemistry/Cheminformatics Chemical and Screening Sciences, Wyeth Research, 865 Ridge Road,
Monmouth Junction, NJ 08852, USA
[17]Department of Pharmaceutical Chemistry, Genentech Hall, Room N472C, University of California, San Francisco, 600 16th Street, San Francisco, CA 94158, USA
[18]Midwest Center for Structural Genomics and Structural Biology Center, Biosciences Division, Argonne National Laboratory, Argonne, IL 60439, USA
[19]Bristol-Myers Squibb, Research and Discovery, P.O. Box 5400, Princeton, NJ 08543, USA
[20]Department of Biopharmaceutical Sciences, University of California, San Francisco, 1700 4th Street, San Francisco, CA 94143, USA
[21]Protein Structure Prediction Center, University of California, Davis, 451 Health Sciences Drive, Davis, CA 95616, USA
[22]Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute, Rockville, MD 20850, USA
[23]Center for Computational Biology and Bioinformatics, Columbia University, 622 West 168th Street, New York, NY 10032, USA
[24]Department of Structural and Chemical Biology, Mount Sinai School of Medicine, 1425 Madison Avenue, New York, NY 10029, USA
[25]Biochemistry and Molecular Biology, The University of Chicago, 929 East 57th Street, Chicago, IL 60637, USA
[26]Immunology Frontier Research Center, Systems Immunology Laboratory, Osaka University, Osaka 565-0871, Japan
[27]6211 Kaitlyn Court, West Windsor, NJ 08550, USA
[28]Biomedical Engineering Structural BioInformatics Laboratory, Boston University, 44 Cummington Street, Boston, MA 02215, USA
[29]Bocaraca, 2443 Ash Street, Palo Alto, CA 94306, USA
[30]Joint Center for Structural Genomics, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA
*Correspondence: torsten.schwede@unibas.ch (T.S.), sali@salilab.org (A.S.)
All attendees of the Workshop, except for students and National Institutes of Health officers, are listed as authors.

We describe the proceedings and conclusions from the "Workshop on Applications of Protein Models in Biomedical Research" (the Workshop) that was held at the University of California, 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) the requirements and challenges for different applications, and (iii) how the interaction between the computational and experimental research communities could be strengthened to advance the field.

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 α helix and β sheet proposed by Linus Pauling some 2 years earlier (Pauling et al., 1951); and the first homology model of a protein (α-lactalbumin),

built by David Phillips and coworkers, based on hen egg white lysozyme (Browne et al., 1969). Although 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, transmembrane 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 the Protein Structure Initiative

A stimulating catalyst for molecular modeling is the Protein Structure Initiative (PSI), which aims to determine representative atomic structures of most major protein families by X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy, so that most of the remaining 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 of specific sequences, and the expanded data set 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., 2007), and docking of proteins can be used for the 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).

Protein Structure Initiative Knowledgebase and Protein Model Portal

To make the fruits of PSI available as widely as possible, the PSI Structural Genomics Knowledgebase (the Knowledgebase) was launched in February 2008 (http://kb.psi-structuralgenomics.org) (Berman et al., 2009). 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 facilitates 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 (Arnold et al., 2008) in particular provides access to models calculated by SWISS-MODEL (Kiefer et al., 2009) and ModBase (Pieper et al., 2009), 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 Protein 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 to discuss 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 Protein Model Portal presents 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, 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. Although 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 (e.g., molecular replacement in X-ray crystallography) and the use of homology models in conjunction with other methods (e.g., docking of small molecules and proteins). The participants' consensus was formulated as specific recommendations, aimed to increase the impact of molecular modeling on biology and medicine.

Workshop Program

On the first day, 16 presentations were given on topics that ranged from the coverage of protein sequence-structure space to the uses of modeling in biology and medicine. 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. 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.

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 (M.L., unpublished data). 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 multidomain architectures, with or without known structures. The domains were defined based on the Conserved Domain Architecture Retrieval Tool resource at the National Center for Biotechnology Information (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 multidomain 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. Multidomain 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 multidomain proteins.

Applications of Modeling for Biology and Medicine
Modeling is not only widely applicable, but it is 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 the geometrical accuracy of a model; instead, we focus on the bottom line corresponding to the numerous published studies in which 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.

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 the 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 typically range from 1% to 15% of those molecules that are predicted to bind (Babaoglu et al., 2008; Doman et al., 2002). The 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 the "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 in which 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 to the optimization of binding potency and selectivity (Buckley et al., 2008; Diller and Li, 2003; Rockey and Elcock, 2006). Long before experimental structures of G-protein-coupled receptor (GPCRs) were determined, models helped the selection and introduction of GPCR ligands to the clinic (Engel et al., 2008; Webb and Krystek, 1998; 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.

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 the 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 they 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).

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 Critical Assessment of Predicted Interactions (CAPRI) effort (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 multidomain 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). Furthermore, 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^{2+} channel, which was used for the characterization of binding interactions with 1,4-dihydropyridines (Cosconati et al., 2007).

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 subcellular 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 for 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, $PI(3,4,5)P_3$, primarily through electrostatic interactions, although the model was built by using the structure of a PX domain that binds to $PI(3,4)P_2$-containing membranes with significant hydrophobic penetration (Stahelin et al., 2004).

Ligand Specificity of Receptors

Members of the neurotransmitter/sodium symporter (NSS) transporter family are responsible for the uptake of neurotransmitters (such as glycine, γ-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).

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, 2007; Kalyanaraman et al., 2005). Homology models can extend the utility of this approach to the many uncharacterized enzymes lacking experimental structures, and they enable prediction of substrate specificity among related enzymes in protein families.

In a joint computational and experimental effort, homology models were created for ~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).

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 nonsynonymous, single DNA base variants (i.e., 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 the 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 the functional impact of a DNA base variant as 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 an additional 50% to the number of analyzable single-point mutations. Recent progress measured in the CASP experiments (Kopp et al., 2007; Kryshtafovych et al., 2007) suggests that this coverage is not an unreasonable expectation. A particularly successful example is provided in the next section.

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 the 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.

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 a 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.

Atomic Structure Determination

Modelers have begun to contribute directly to the 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).

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, 2007b). The structure was determined at ~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.

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 the role of the Protein Model Portal.

The Coverage of the 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.

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 the 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 interpreta-

tions. 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.

Standards for Data Formats Must Be Established to Facilitate Data and Software Exchange

Although 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 for 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.

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 for 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_α coordinates are often insufficient to determine 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 in which 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.

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 (e.g., 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 on 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 the applications and limitations of computational modeling methods.

The 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 Protein Model 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 Protein Model 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 for 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 become available) should be implemented. The Protein Model 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 Protein Model Portal should be as inclusive of all method developers and prediction methods as technically feasible.

REFERENCES

Alber, F., Dokudovskaya, S., Veenhoff, L.M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., Karni-Schmidt, O., Williams, R., Chait, B.T., et al. (2007a). Nature *450*, 683–694.

Alber, F., Dokudovskaya, S., Veenhoff, L.M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., Karni-Schmidt, O., Williams, R., Chait, B.T., et al. (2007b). Nature *450*, 695–701.

Alber, F., Forster, F., Korkin, D., Topf, M., and Sali, A. (2008). Annu. Rev. Biochem. *77*, 443–477.

Alquati, C., De Gioia, L., Santarossa, G., Alberghina, L., Fantucci, P., and Lotti, M. (2002). Eur. J. Biochem. *269*, 3321–3328.

Ananthanarayanan, B., Das, S., Rhee, S.G., Murray, D., and Cho, W. (2002). J. Biol. Chem. *277*, 3568–3575.

Arnold, K., Kiefer, K., Kopp, J., Battey, J.N.D., Podvinec, M., Westbrook, J., Berman, H., Bordoli, L., and Schwede, T. (2008). J. Struct. Funct. Genomics, in press. Published online November 27, 2008. 10.1007/s10969-008-9048-5.

Babaoglu, K., Simeonov, A., Irwin, J.J., Nelson, M.E., Feng, B., Thomas, C.J., Cancian, L., Costi, M.P., Maltby, D.A., Jadhav, A., et al. (2008). J. Med. Chem. *51*, 2502–2511.

Benkert, P., Tosatto, S.C., and Schomburg, D. (2008). Proteins *71*, 261–277.

Berman, H., Henrick, K., and Nakamura, H. (2003). Nat. Struct. Biol. *10*, 980.

Berman, H.M., Burley, S.K., Chiu, W., Sali, A., Adzhubei, A., Bourne, P.E., Bryant, S.H., Dunbrack, R.L., Jr., Fidelis, K., Frank, J., et al. (2006). Structure *14*, 1211–1217.

Berman, H.M., Westbrook, J.D., Gabanyi, M.J., Tao, Y., Shah, R., Kouranov, A., Schwede, T., Arnold, K., Kiefer, F., Bordoli, L., et al. (2009). Nucleic Acids Res. *37*, D365–D368.

Bhattacharya, A., Wunderlich, Z., Monleon, D., Tejero, R., and Montelione, G.T. (2008). Proteins *70*, 105–118.

Blatner, N.R., Stahelin, R.V., Diraviyam, K., Hawkins, P.T., Hong, W., Murray, D., and Cho, W. (2004). J. Biol. Chem. *279*, 53818–53827.

Bourdon, A., Minai, L., Serre, V., Jais, J.P., Sarzi, E., Aubert, S., Chretien, D., de Lonlay, P., Paquis-Flucklinger, V., Arakawa, H., et al. (2007). Nat. Genet. *39*, 776–780.

Browne, W.J., North, A.C., Phillips, D.C., Brew, K., Vanaman, T.C., and Hill, R.L. (1969). J. Mol. Biol. *42*, 65–86.

Bryson, K., McGuffin, L.J., Marsden, R.L., Ward, J.J., Sodhi, J.S., and Jones, D.T. (2005). Nucleic Acids Res. *33*, W36–W38.

Buckley, G.M., Ceska, T.A., Fraser, J.L., Gowers, L., Groom, C.R., Higueruelo, A.P., Jenkins, K., Mack, S.R., Morgan, T., Parry, D.M., et al. (2008). Bioorg. Med. Chem. Lett. *18*, 3291–3295.

Carter, P., Presta, L., Gorman, C.M., Ridgway, J.B., Henner, D., Wong, W.L., Rowland, A.M., Kotts, C., Carver, M.E., and Shepard, H.M. (1992). Proc. Natl. Acad. Sci. USA *89*, 4285–4289.

Cavalli, A., Salvatella, X., Dobson, C.M., and Vendruscolo, M. (2007). Proc. Natl. Acad. Sci. USA *104*, 9615–9620.

Chandonia, J.M., and Brenner, S.E. (2006). Science *311*, 347–351.

Cherezov, V., Rosenbaum, D.M., Hanson, M.A., Rasmussen, S.G., Thian, F.S., Kobilka, T.S., Choi, H.J., Kuhn, P., Weis, W.I., Kobilka, B.K., and Stevens, R.C. (2007). Science *318*, 1258–1265.

Chin, G., and Yeston, J. (2007). Science *317*, 873.

Comeau, S.R., Gatchell, D.W., Vajda, S., and Camacho, C.J. (2004). Bioinformatics *20*, 45–50.

Cosconati, S., Marinelli, L., Lavecchia, A., and Novellino, E. (2007). J. Med. Chem. *50*, 1504–1513.

Dalton, A.K., Murray, P.S., Murray, D., and Vogt, V.M. (2005). J. Virol. *79*, 6227–6238.

Das, R., and Baker, D. (2008). Annu. Rev. Biochem. *77*, 363–382.

de Paulis, T. (2007). Curr. Opin. Investig. Drugs *8*, 78–86.

Di Bernardo, M.C., Crowther-Swanepoel, D., Broderick, P., Webb, E., Sellick, G., Wild, R., Sullivan, K., Vijayakrishnan, J., Wang, Y., Pittman, A.M., et al. (2008). Nat. Genet. *40*, 1204–1210.

Diller, D.J., and Li, R. (2003). J. Med. Chem. *46*, 4638–4647.

Doman, T.N., McGovern, S.L., Witherbee, B.J., Kasten, T.P., Kurumbail, R., Stallings, W.C., Connolly, D.T., and Shoichet, B.K. (2002). J. Med. Chem. *45*, 2213–2221.

Engel, S., Skoumbourdis, A.P., Childress, J., Neumann, S., Deschamps, J.R., Thomas, C.J., Colson, A.O., Costanzi, S., and Gershengorn, M.C. (2008). J. Am. Chem. Soc. *130*, 5115–5123.

Forrest, L.R., Tavoulari, S., Zhang, Y.W., Rudnick, G., and Honig, B. (2007). Proc. Natl. Acad. Sci. USA *104*, 12761–12766.

Geer, L.Y., Domrachev, M., Lipman, D.J., and Bryant, S.H. (2002). Genome Res. *12*, 1619–1623.

Ginalski, K., Elofsson, A., Fischer, D., and Rychlewski, L. (2003). Bioinformatics *19*, 1015–1018.

Godzik, A. (2003). Methods Biochem. Anal. *44*, 525–546.

Hamblin, K., Standley, D.M., Rogers, M.B., Stechmann, A., Roger, A.J., Maytum, R., and van der Giezen, M. (2008). Mol. Microbiol. *68*, 1395–1405.

Hermann, J.C., Ghanem, E., Li, Y., Raushel, F.M., Irwin, J.J., and Shoichet, B.K. (2006). J. Am. Chem. Soc. *128*, 15882–15891.

Hermann, J.C., Marti-Arbona, R., Fedorov, A.A., Fedorov, E., Almo, S.C., Shoichet, B.K., and Raushel, F.M. (2007). Nature *448*, 775–779.

Hinton, P.R., Johlfs, M.G., Xiong, J.M., Hanestad, K., Ong, K.C., Bullock, C., Keller, S., Tang, M.T., Tso, J.Y., Vasquez, M., and Tsurushita, N. (2004). J. Biol. Chem. *279*, 6213–6216.

Huang, Y.J., Hang, D., Lu, L.J., Tong, L., Gerstein, M.B., and Montelione, G.T. (2008). Mol. Cell. Proteomics *7*, 2048–2060.

Huey, R., Morris, G.M., Olson, A.J., and Goodsell, D.S. (2007). J. Comput. Chem. *28*, 1145–1152.

Hult, K., and Berglund, P. (2003). Curr. Opin. Biotechnol. *14*, 395–400.

Janin, J., Henrick, K., Moult, J., Eyck, L.T., Sternberg, M.J., Vajda, S., Vakser, I., and Wodak, S.J. (2003). Proteins *52*, 2–9.

Kalyanaraman, C., Bernacki, K., and Jacobson, M.P. (2005). Biochemistry *44*, 2059–2071.

Kalyanaraman, C., Imker, H.J., Federov, A.A., Federov, E.V., Glasner, M.E., Babbitt, P.C., Almo, S.C., Gerlt, J.A., and Jacobson, M.P. (2008). Structure *16*, 1668–1677.

Kamal, J.K., Benchaar, S.A., Takamoto, K., Reisler, E., and Chance, M.R. (2007). Proc. Natl. Acad. Sci. USA *104*, 7910–7915.

Karchin, R., Monteiro, A.N., Tavtigian, S.V., Carvalho, M.A., and Sali, A. (2007). PLoS Comput. Biol. *3*, e26.

Kamal, J.K., and Chance, M.R. (2008). Protein Sci. *17*, 79–94.

Kopp, J., Bordoli, L., Battey, J.N., Kiefer, F., and Schwede, T. (2007). Proteins *69* (*Suppl 8*), 38–56.

Kiefer, F., Arnold, K., Künzli, M., Bordoli, L., and Schwede, T. (2009). Nucleic Acids Res. *37*, D387–D392.

Krasley, E., Cooper, K.F., Mallory, M.J., Dunbrack, R., and Strich, R. (2006). Genetics *172*, 1477–1486.

Kryshtafovych, A., Fidelis, K., and Moult, J. (2007). Proteins *69* (*Suppl 8*), 194–207.

Lazar, G.A., Dang, W., Karki, S., Vafa, O., Peng, J.S., Hyun, L., Chan, C., Chung, H.S., Eivazi, A., Yoder, S.C., et al. (2006). Proc. Natl. Acad. Sci. USA *103*, 4005–4010.

Lensink, M.F., Méndez, R., and Wodak, S.J. (2007). Proteins *69*, 704–718.

Lippow, S.M., Wittrup, K.D., and Tidor, B. (2007). Nat. Biotechnol. *25*, 1171–1176.

Liu, J., Montelione, G.T., and Rost, B. (2007). Nat. Biotechnol. *25*, 849–851.

Marti-Renom, M.A., Stuart, A.C., Fiser, A., Sanchez, R., Melo, F., and Sali, A. (2000). Annu. Rev. Biophys. Biomol. Struct. *29*, 291–325.

Murray, P.S., Li, Z., Wang, J., Tang, C.L., Honig, B., and Murray, D. (2005). Structure *13*, 1521–1531.

Noble, M.E., Endicott, J.A., and Johnson, L.N. (2004). Science *303*, 1800–1805.

Oshiro, C., Bradley, E.K., Eksterowicz, J., Evensen, E., Lamb, M.L., Lanctot, J.K., Putta, S., Stanton, R., and Grootenhuis, P.D. (2004). J. Med. Chem. *47*, 764–767.

Pauling, L., Corey, R.B., and Branson, H.R. (1951). Proc. Natl. Acad. Sci. USA *37*, 205–211.

Petrey, D., and Honig, B. (2005). Mol. Cell *20*, 811–819.

Pettitt, C.S., McGuffin, L.J., and Jones, D.T. (2005). Bioinformatics *21*, 3509–3515.

Pieper, U., Eswar, N., Webb, B.M., Eramian, D., Kelly, L., Barkan, D.T., Carter, H., Mankoo, P., Karchin, R., Marti-Renom, M.A., et al. (2009). Nucleic Acids Res. *37*, D347–D354.

Pietsch, E.C., Perchiniak, E., Canutescu, A.A., Wang, G., Dunbrack, R.L., and Murphy, M.E. (2008). J. Biol. Chem. *283*, 21294–21304.

Presta, L.G., Chen, H., O'Connor, S.J., Chisholm, V., Meng, Y.G., Krummen, L., Winkler, M., and Ferrara, N. (1997). Cancer Res. *57*, 4593–4599.

Qian, B., Raman, S., Das, R., Bradley, P., McCoy, A.J., Read, R.J., and Baker, D. (2007). Nature *450*, 259–264.

Queen, C., Schneider, W.P., Selick, H.E., Payne, P.W., Landolfi, N.F., Duncan, J.F., Avdalovic, N.M., Levitt, M., Junghans, R.P., and Waldmann, T.A. (1989). Proc. Natl. Acad. Sci. USA *86*, 10029–10033.

Rasmussen, S.G., Choi, H.J., Rosenbaum, D.M., Kobilka, T.S., Thian, F.S., Edwards, P.C., Burghammer, M., Ratnala, V.R., Sanishvili, R., Fischetti, R.F., et al. (2007). Nature *450*, 383–387.

Rester, U. (2008). Curr. Opin. Drug Discov. Devel. *11*, 559–568.

Robinson, C.V., Sali, A., and Baumeister, W. (2007). Nature *450*, 973–982.

Rockey, W.M., and Elcock, A.H. (2006). Curr. Protein Pept. Sci. *7*, 437–457.

Rossmann, M.G., Morais, M.C., Leiman, P.G., and Zhang, W. (2005). Structure *13*, 355–362.

Rost, B. (2003). Methods Biochem. Anal. *44*, 559–587.

Rumpel, S., Becker, S., and Zweckstetter, M. (2008). J. Biomol. NMR *40*, 1–13.

Schwede, T., Kopp, J., Guex, N., and Peitsch, M.C. (2003). Nucleic Acids Res. *31*, 3381–3385.

Shen, M.Y., and Sali, A. (2006). Protein Sci. *15*, 2507–2524.

Shen, Y., Lange, O., Delaglio, F., Rossi, P., Aramini, J.M., Liu, G., Eletsky, A., Wu, Y., Singarapu, K.K., Lemak, A., et al. (2008). Proc. Natl. Acad. Sci. USA *105*, 4685–4690.

Sippl, M.J. (1993). Proteins *17*, 355–362.

Sjoblom, T., Jones, S., Wood, L.D., Parsons, D.W., Lin, J., Barber, T.D., Mandelker, D., Leary, R.J., Ptak, J., Silliman, N., et al. (2006). Science *314*, 268–274.

Sondermann, H., Nagar, B., Bar-Sagi, D., and Kuriyan, J. (2005). Proc. Natl. Acad. Sci. USA *102*, 16632–16637.

Song, L., Kalyanaraman, C., Fedorov, A.A., Fedorov, E.V., Glasner, M.E., Brown, S., Babbitt, P.C., Almo, S.C., Jacobson, M.P., and Gerlt, J.A. (2007). Nat. Chem. Biol. *3*, 486–491.

Stacey, S.N., Manolescu, A., Sulem, P., Thorlacius, S., Gudjonsson, S.A., Jonsson, G.F., Jakobsdottir, M., Bergthorsson, J.T., Gudmundsson, J., Aben, K.K., et al. (2008). Nat. Genet. *40*, 703–706.

Stahelin, R.V., Ananthanarayanan, B., Blatner, N.R., Singh, S., Bruzik, K.S., Murray, D., and Cho, W. (2004). J. Biol. Chem. *279*, 54918–54926.

Tang, Y., Poustovoitov, M.V., Zhao, K., Garfinkel, M., Canutescu, A., Dunbrack, R., Adams, P.D., and Marmorstein, R. (2006). Nat. Struct. Mol. Biol. *13*, 921–929.

Tarn, C., Merkel, E., Canutescu, A.A., Shen, W., Skorobogatko, Y., Heslin, M.J., Eisenberg, B., Birbe, R., Patchefsky, A., Dunbrack, R., et al. (2005). Clin. Cancer Res. *11*, 3668–3677.

Topf, M., Lasker, K., Webb, B., Wolfson, H., Chiu, W., and Sali, A. (2008). Structure *16*, 295–307.

Tramontano, A. (2006). FEBS Lett. *580*, 2928–2934.

von Itzstein, M., Wu, W.-Y., Kok, G.B., Pegg, M.S., Dyason, J.C., Jin, B., Phan, V.T., Smythe, M.L., White, H.F., Oliver, W.S., et al. (1993). Nature *363*, 418–423.

Wallner, B., and Elofsson, A. (2003). Protein Sci. *12*, 1073–1086.

Wallner, B., Fang, H., and Elofsson, A. (2003). Proteins *53* (*Suppl 6*), 534–541.

Warne, T., Serrano-Vega, M.J., Baker, J.G., Moukhametzianov, R., Edwards, P.C., Henderson, R., Leslie, A.G., Tate, C.G., and Schertler, G.F. (2008). Nature *454*, 486–491.

Watson, J.D., and Crick, F.H. (1953). Nature *171*, 737–738.

Webb, M.L., and Krystek, S.R., Jr. (1998). Endothelin Receptors and Signaling Mechanisms, D.M. Pollock and R.F. Highsmith, eds. (Austin, TX: Landes Bioscience).

Webb, M.L., Patel, P.S., Rose, P.M., Liu, E.C., Stein, P.D., Barrish, J., Lach, D.A., Stouch, T., Fisher, S.M., Hadjilambris, O., et al. (1996). Biochemistry *35*, 2548–2556.

Westbrook, J.D., and Fitzgerald, P.M. (2003). Methods Biochem. Anal. *44*, 161–179.

Winn, M.D. (2003). J. Synchrotron Radiat. *10*, 23–25.

Yu, J.W., Mendrola, J.M., Audhya, A., Singh, S., Keleti, D., DeWald, D.B., Murray, D., Emr, S.D., and Lemmon, M.A. (2004). Mol. Cell *13*, 677–688.

Yue, P., Li, Z., and Moult, J. (2005). J. Mol. Biol. *353*, 459–473.

Zhang, R., Poustovoitov, M.V., Ye, X., Santos, H.A., Chen, W., Daganzo, S.M., Erzberger, J.P., Serebriiskii, I.G., Canutescu, A.A., Dunbrack, R.L., et al. (2005). Dev. Cell *8*, 19–30.