

Achievements and challenges in structural bioinformatics and computational biophysics

Ilan Samish^{1,2}, Philip E. Bourne³ and Rafael J. Najmanovich^{4,*}

¹Department of Plant Sciences, Weizmann Institute of Science, Rehovot, 76100, Israel, ²Ort Braude College, Karmiel, 2161002, Israel, ³Office of the Director, National Institutes of Health, Bethesda, MD 20814, USA and ⁴Department of Biochemistry, University of Sherbrooke, Sherbrooke, J1H 5N4, Canada

ABSTRACT

Motivation: The field of structural bioinformatics and computational biophysics has undergone a revolution in the last 10 years. Developments that are captured annually through the 3DSIG meeting, upon which this article reflects.

Results: An increase in the accessible data, computational resources and methodology has resulted in an increase in the size and resolution of studied systems and the complexity of the questions amenable to research. Concomitantly, the parameterization and efficiency of the methods have markedly improved along with their cross-validation with other computational and experimental results.

Conclusion: The field exhibits an ever-increasing integration with biochemistry, biophysics and other disciplines. In this article, we discuss recent achievements along with current challenges within the field.

Contact: Rafael.Najmanovich@USherbrooke.ca

1 INTRODUCTION

Structural bioinformatics, originally known as structural computational biology, predates other forms of bioinformatics. It can be argued that the seminal 1953 article by Watson and Crick (Watson and Crick, 1953) is in fact a modeling paper and arguably the first structural bioinformatics paper. Thus, the 2014 Nobel prize for ‘multiscale modeling’ to Martin Karplus, Arie Warshel and Michael Levitt marks an important hallmark acknowledging the impact of structural bioinformatics on science. In his account of the birth of the field, Levitt (2001) describes how computation was required to accurately refine the tRNA model predicted by Crick in building an actual model that was taller than himself. Thus, computation has been an integral part of structural biology from its early days and has had an ever-increasing role in biochemistry and molecular biology with the passing of years. Indeed, from the first simulations of small systems and a few picoseconds acknowledged by the Nobel committee, we are now at a stage where millisecond simulations (Beauchamp *et al.*, 2011) or massive searches of sequence and structure space as required for, e.g. computational protein design (Gu and Bourne, 2011; Kiss *et al.*, 2013; Samish *et al.*, 2011) are achievable.

Structural bioinformatics or structural computational biology, broadly defined, is a field at the intersection between computer science, physics, chemistry and molecular biology. Historically, the term ‘structural bioinformatics’ describes data-driven statistical, knowledge-based research of representative non-redundant ensembles of structures to understand the statistical behavior of

the system under investigation. Alternatively, ‘computational biophysics’ describes a hypothesis-driven physics-based treatment of biological molecular systems. The ergodic hypothesis guarantees that conclusions from the two types of approaches converge over large non-redundant samples or long simulations (Samish, 2009). Currently, numerous methodologies employ ideas from both approaches. Consequently, hereafter we will refer to both as structural bioinformatics.

Biologically, structural bioinformatics aims to understand the factors that influence and determine the function of biological macromolecules, the interplay between evolution, kinetics and thermodynamics, the determinants of specificity and selectivity in molecular interactions, the dynamic aspects of macromolecular structures and their effect on function and stability and, finally, the ability to use all these for engineering, design and biotechnology. In fact, a complete understanding of biological processes must inescapably pass through an understanding of the factors influencing such processes at the atomic and sometimes even subatomic levels. In this article, we discuss some of the most notable achievements in structural bioinformatics over the past 10 years and discuss existing challenges in the field. Without a doubt, the topics and the specific articles mentioned here are biased by the opinion of the authors.

2 ACHIEVEMENTS

Some of the numerous achievements in the field in the last 10 years include the following (Fig. 1).

2.1 Data coverage and community resources

The human genome sequencing revolution marks the availability of sequence data at scale. The realization that sequence alone is not enough to understand/predict function led to the establishment of large structural genomics initiatives where the structure of proteins with low sequence similarity to proteins with known structures were targeted to increase the coverage of fold space and permit more accurate fold recognition, threading and homology modeling (Baker and Sali, 2001). In the last decade, the number of structures deposited in the protein data bank (PDB) (Berman *et al.*, 2000) grew 4-fold to over 100,000 structures including an unprecedented number of membrane protein structures. The PDB—one of the first biological databases—and derived resources such as CATH (Sillitoe *et al.*, 2013), SCOP (Andreeva *et al.*, 2008) and PFAM (Finn *et al.*, 2014) are fundamental enabling tools for the entire field and their long-term maintenance is of great importance to the field.

*To whom correspondence should be addressed.

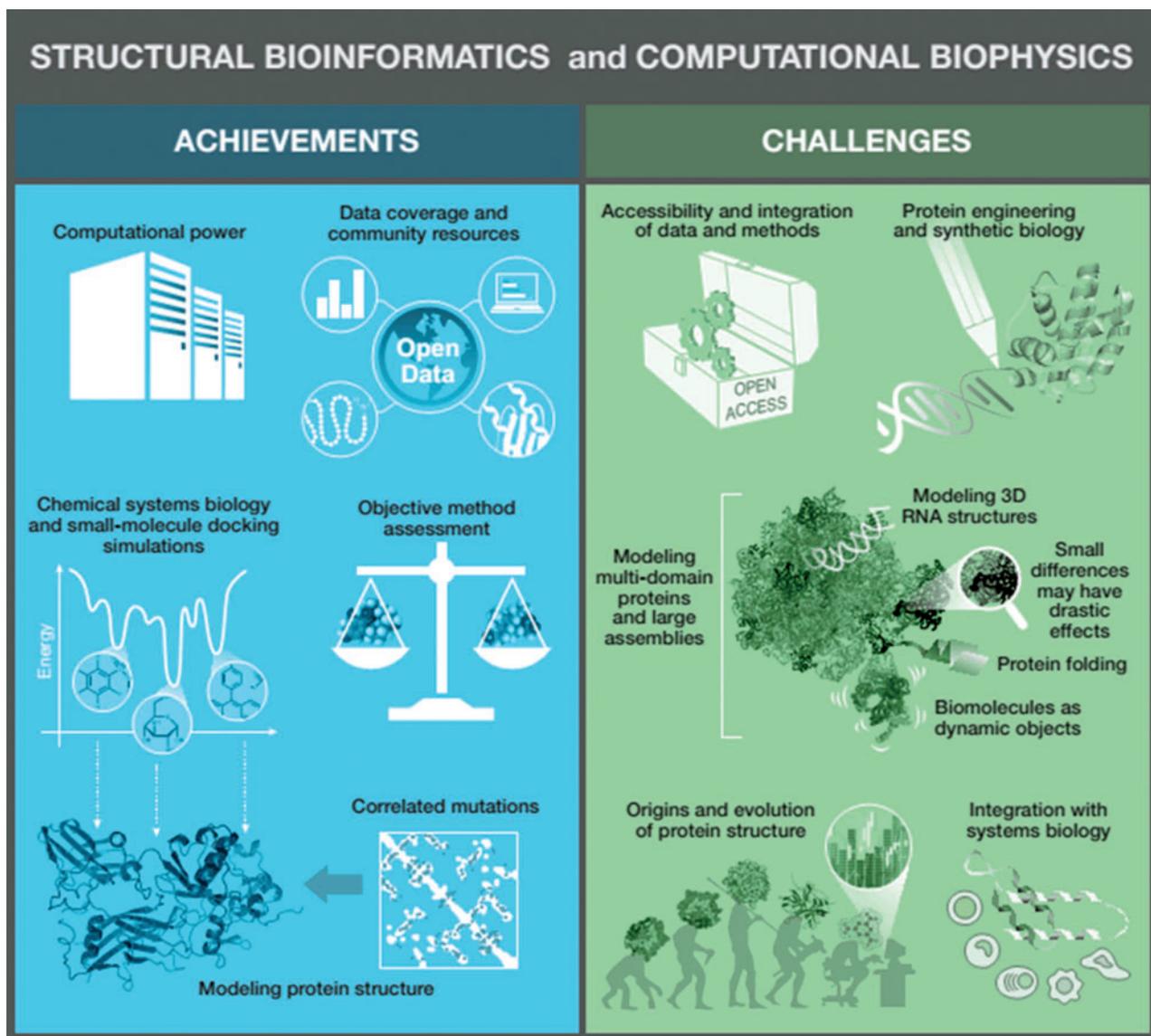


Fig. 1. Schematic representation of the main achievements and challenges in the field of structural bioinformatics and computational biophysics as discussed in the text

2.2 Computational power

The increased availability of computer power has enabled applications that were beyond imagining just a few years ago. Dedicated hardware [e.g. ANTON (Shaw *et al.*, 2009) at DE Shaw Research designed uniquely for molecular dynamics simulations], the use of graphical processing units (Friedrichs *et al.*, 2009) or distributed crowd sourcing with the BOINC interface (Anderson, 2004) [e.g. Folding@home (Shirts and Pande, 2000)] vastly increased the reach of current methods. An exciting development is the involvement of interested laypeople as active problem solvers such as in the Foldit game (Cooper *et al.*, 2010).

2.3 Objective method assessment

The CASP critical assessment of protein structure prediction began a new era in the field providing an objective double

blind test for structure prediction methods (Moult *et al.*, 2014) with other areas following suit: protein interactions (Janin *et al.*, 2003), function prediction (Radivojac *et al.*, 2013), membrane protein docking (Kufareva *et al.*, 2014) or automated structure prediction (Bourne, 2003). This also led to meta-prediction methods (Ginalski *et al.*, 2003) and community collaborations, e.g. WeFold (Khoury *et al.*, 2014).

2.4 Correlated mutations and modeling protein structure

Correlated mutation data have enabled the generation of pairwise amino acid contact maps from sequence data (Marks *et al.*, 2011; Nugent and Jones, 2012). Such contact maps are then used as spatial constraints to generate models of globular and membrane proteins. In the cases where enough sequence data allow the use of this methodology, a high level of accuracy has been achieved.

2.5 Chemical systems biology

Also known as systems pharmacology, the integration of the vast amounts of 'omics' data with accessible structural methods such as the detection of binding site similarities (Kurbatova *et al.*, 2013; Xie *et al.*, 2011a) can be used for drug repositioning and discovery (Chang *et al.*, 2010; Kinnings *et al.*, 2009; Xie *et al.*, 2011b, 2014).

2.6 Small-molecule docking simulations

Docking simulations (Brooijmans and Kuntz, 2003; Halperin *et al.*, 2002) are widely used (Kitchen *et al.*, 2004; Warren *et al.*, 2006). Predicted ligand protein complex structures are then used to generate hypotheses about binding and for virtual screening in the early stages of drug design.

3 CHALLENGES

Although considerable progress has been made, advances are still needed in the following areas (Fig. 1).

3.1 Modeling large or multi-domain proteins and assemblies

Most proteins are large and multi-domain and take part in complex assemblies requiring fine-tuned recognition (Chothia and Janin, 1975; Jones and Thornton, 1996; Levy *et al.*, 2008). Examples include large complexes such as the ribosome (Yonath, 2010) and the proteasome (Adams, 2003) or supercomplexes such as the 60-subunit pyruvate dehydrogenase complex (Patel *et al.*, 2014). Most of cell biology requires understanding the interplay of such large multi-domain proteins and complexes and lacks the level of detail that comes from atomic-level structural information. Targeting such complexes and assemblies experimentally and looking at them in the context of the complete cell is an emergent challenge.

3.2 Biomolecules as dynamic objects

The accurate modeling of large conformational changes due to ligand binding, allosteric effects, post-translational modifications or as the result of protein-protein interactions is essential. Techniques such as molecular dynamics (Adcock and McCammon, 2006) and elastic network models (Bahar and Rader, 2005; Frappier and Najmanovich, 2014) provide two techniques at different granularities. Descriptions of biomolecules representing the conformational ensembles in which they are found under physiological conditions will expand our understanding of the relationship between structure and function. The use of such conformational ensembles as opposed to single static structures is challenging for complex structures.

3.3 Modeling 3D RNA structures

The prediction of RNA structures is still in its infancy (Laing and Schlick, 2010; Rother *et al.*, 2011) and often requires processing low-resolution data (Parisien and Major, 2012). With simplified rules for molecular interactions in RNA when compared with proteins, we should be able to model larger and more complex RNA molecules than currently possible.

3.4 Small differences may have drastic effects

Although protein structure is resilient to mutation (Baase *et al.*, 2010), function is not necessarily as resilient (Najmanovich *et al.*, 2008; Rajamani *et al.*, 2004). This combination of structural robustness and functional plasticity is at the core of evolutionary change. It is essential that we learn to recognize and accordingly weigh such functional determinants to predict the outcome of natural or engineered perturbations at the molecular level. In particular, the prediction of function is based on the detection of similarities (Najmanovich *et al.*, 2005) where small differences are ignored. One needs to only look at the existence of vast protein families performing the 'same' function to realize that we do not fully understand the effect of small differences. Some of these differences are likely required to modulate selectivity rather than specificity within the different cellular (or temporal) contexts where the same function is required (Najmanovich *et al.*, 2008). This level of understanding will require the integration between structural and systems biology.

3.5 Integration with systems biology

Beyond the detection of cross-reactivity targets to a given drug and the impact on the complete system, a full understanding of specificity and selectivity at the molecular level requires studying all macromolecules that are sharing the crowded (Ellis, 2001; McGuffee and Elcock, 2010; Minton, 1993) cellular milieu. Computationally understanding a full cell from a structural point of view is a major challenge. A full structural understanding of a living cell at all scales (Stein *et al.*, 2007), integrating macromolecules at atomic resolution all the way to phenotypes is still missing and once achieved will define a new era in biology (Hallock *et al.*, 2014; Peterson *et al.*, 2014; Winter *et al.*, 2012).

3.6 Protein engineering and synthetic biology

Protein engineering of single proteins (Kiss *et al.*, 2013) and protein-protein interfaces (Potapov *et al.*, 2008) is an advancing field, in particular with the success of Rosetta (Rohl *et al.*, 2004) and the advent of a whole generation of 'Rosetta Engineers' (The term was heard at the Protein Engineering Canada 2014 conference but the author is unknown.). However, computational protein design (Samish *et al.*, 2011) and enzyme redesign (Gerlt and Babbitt, 2009) are still restricted by the complexity of the sequence and structure search space. The approximations that are needed to tackle this complexity often require a choice of simplified methods that all but ignore atomic structure. Specifically, side-chain placement, iterative homology modeling (Q. Wang *et al.*, 2008) and flexible backbone sampling (Ollikainen *et al.*, 2013) remain major challenges. Lastly, synthetic biology (Way *et al.*, 2014) promises to extend biological systems well beyond what is naturally observed and its integration with computational drug design (Marchisio and Stelling, 2009; Tew *et al.*, 2010) and structural bioinformatics will open new and exciting possibilities.

3.7 Origins and evolution of protein structure

Is fold space discrete or continuous? If discrete, why are some folds so much more common than others? Have we identified nearly all possible folds? How do new folds appear? All these

questions have been partially addressed but definitive answers still remain illusive.

3.8 Protein folding

The biggest challenge in structural bioinformatics remains unsolved. That is, the ability to consistently predict the structure of a protein based solely on its amino acid sequence. The Levinthal paradox inspires us to continue. Thus, although conformational space is so immense as to be intractable, proteins do fold, leading to a variety of advances based on different ways of addressing the paradox (Dill and Chan, 1997; Karplus, 1997; Wolynes et al., 1995). However, we are far from solving the protein folding problem and arguably progress is based more on existing structures than first principles.

3.9 Accessibility and integration of data and methods

Data, methods and publications must be open (Masum et al., 2013) and reproducible (Sandve et al., 2013). One success of the field has been the number of mature open source software. Widely used software includes CCP4 for macromolecular structure determination and refinement (Potterton et al., 2004), CHARMM (Hynninen and Crowley, 2014), NAMD (Y. Wang et al., 2011) and GROMACS (Pronk et al., 2013) for molecular dynamics, Modeller (Esvar et al., 2007) for homology modeling and Rosetta (Rohl et al., 2004) for structure prediction and design. One challenge for the future is how to make the plethora of existing methods accessible to newcomers in the field and to the scientific community at large. Just as one cannot publish a structure or a sequence without submitting the data to a public repository, methods and data must be stored in repositories that guarantee their accessibility to the community immediately at the time of publication. The availability of data and methods will help ensure reproducibility and cross-validation.

4 3DSIG

3DSIG is a special interest group within the International Society for Computational Biology. 3DSIG holds an annual meeting preceding the annual ISMB meeting. The insights shared in this article come as a synthesis of the trends observed at 3DSIG since its inception 10 years ago.

ACKNOWLEDGEMENTS

The authors would like to thank John Moult for his work as chair of the scientific committee until 2011 and the support of Marvin Edelman in creating 3DSIG. The success of 3DSIG would not be possible without the support of its attendees over the years as well as the hard work of Matthieu Chartier as a member of the organizing committee. R.J.N. is part of the CR-CHUS, Institute of Pharmacology of Sherbrooke, PROTEO (the Québec network for research on protein function, structure and engineering) and GRASP (Groupe de Recherche Axé sur la Structure des Protéines).

Conflict of Interest: none declared.

REFERENCES

- Adams, J. (2003) The proteasome: structure, function, and role in the cell. *Cancer Treat Rev.*, **29** (Suppl. 1), 3–9.
- Acock, S.A. and McCammon, J.A. (2006) Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem. Rev.*, **106**, 1589–1615.
- Anderson, D.P. (2004) BOINC: a system for public-resource computing and storage. In: *Fifth IEEE/ACM International Workshop on Grid Computing*. IEEE, pp. 4–10, Nov. 8, 2004. IEEE. doi: <http://10.1109/GRID.2004.14>.
- Andreeva, A. et al. (2008) Data growth and its impact on the SCOP database: new developments. *Nucleic Acids Res.*, **36**, D419–D25.
- Baase, W.A. et al. (2010) Lessons from the lysozyme of phage T4. *Protein Sci.*, **19**, 631–641.
- Bahar, I. and Rader, A.J. (2005) Coarse-grained normal mode analysis in structural biology. *Curr. Opin. Struct. Biol.*, **15**, 586–592.
- Baker, D. and Sali, A. (2001) Protein structure prediction and structural genomics. *Science*, **294**, 93–96.
- Beauchamp, K.A. et al. (2011) MSMBuild2: modeling conformational dynamics at the picosecond to millisecond scale. *J. Chem. Theory Comput.*, **7**, 3412–3419.
- Berman, H.M. et al. (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235–242.
- Bourne, P.E. (2003) CASP and CAFASP experiments and their findings. *Methods Biochem. Anal.*, **44**, 501–507.
- Brooijmans, N. and Kuntz, I.D. (2003) Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.*, **32**, 335–373.
- Chang, R.L. et al. (2010) Drug off-target effects predicted using structural analysis in the context of a metabolic network model. *PLoS Comput. Biol.*, **6**, e1000938.
- Chothia, C. and Janin, J. (1975) Principles of protein-protein recognition. *Nature*, **256**, 705–708.
- Cooper, S. et al. (2010) Predicting protein structures with a multiplayer online game. *Nature*, **466**, 756–760.
- Dill, K.A. and Chan, H.S. (1997) From Levinthal to pathways to funnels. *Nat. Struct. Biol.*, **4**, 10–19.
- Ellis, R. (2001) Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Curr. Opin. Struct. Biol.*, **11**, 114–119.
- Esvar, N. et al. (2007) Protein structure modeling with MODELLER. *Methods Mol. Biol.*, **426**, 145–159.
- Finn, R.D. et al. (2014) Pfam: the protein families database. *Nucleic Acids Res.*, **42**, D222–D230.
- Frapplier, V. and Najmanovich, R.J. (2014) A coarse-grained elastic network atom contact model and its use in the simulation of protein dynamics and the prediction of the effect of mutations. *PLoS Comput. Biol.*, **10**, e1003569.
- Friedrichs, M.S. et al. (2009) Accelerating molecular dynamic simulation on graphics processing units. *J. Comput. Chem.*, **30**, 864–872.
- Gerlt, J.A. and Babbitt, P.C. (2009) Enzyme (re)design: lessons from natural evolution and computation. *Curr. Opin. Chem. Biol.*, **13**, 10–18.
- Ginalski, K. et al. (2003) 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics*, **19**, 1015–1018.
- Gu, J. and Bourne, P.E. (2011) Structural bioinformatics. Wiley-Blackwell.
- Hallock, M.J. et al. (2014) Simulation of reaction diffusion processes over biologically relevant size and time scales using multi-GPU workstations. *Parallel Comput.*, **40**, 86–99.
- Halperin, I. et al. (2002) Principles of docking: an overview of search algorithms and a guide to scoring functions. *Proteins*, **47**, 409–443.
- Hynninen, A.-P. and Crowley, M.F. (2014) New faster CHARMM molecular dynamics engine. *J. Comput. Chem.*, **35**, 406–413.
- Janin, J. et al. (2003) CAPRI: a critical assessment of predicted interactions. *Proteins*, **52**, 2–9.
- Jones, S. and Thornton, J.M. (1996) Principles of protein-protein interactions. *Proc. Natl Acad. Sci. USA*, **93**, 13–20.
- Karplus, M. (1997) The Levinthal paradox: yesterday and today. *Fold Des.*, **2**, S69–S75.
- Khoury, G.A. et al. (2014) WeFold: a competition for protein structure prediction. *Proteins*, **82**, 1850–1868.
- Kinnings, S.L. et al. (2009) Drug discovery using chemical systems biology: repositioning the safe medicine Comtan to treat multi-drug and extensively drug resistant tuberculosis. *PLoS Comput. Biol.*, **5**, e1000423.
- Kiss, G. et al. (2013) Computational enzyme design. *Angew. Chem. Int. Ed. Engl.*, **52**, 5700–5725.
- Kitchen, D.B. et al. (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug. Discov.*, **3**, 935–949.

- Kufareva, I. *et al.* (2014) Advances in GPCR modeling evaluated by the GPCR dock 2013 assessment: meeting new challenges. *Structure*, **22**, 1120–1139.
- Kurbatova, N. *et al.* (2013) IsoCleft finder—a web-based tool for the detection and analysis of protein binding-site geometric and chemical similarities. *Fl1000Res*, **2**, 117.
- Laing, C. and Schlick, T. (2010) Computational approaches to 3D modeling of RNA. *J. Phys. Condens. Matter*, **22**, 283101.
- Levitt, M. (2001) The birth of computational structural biology. *Nat. Struct. Mol. Biol.*, **8**, 392–393.
- Levy, E.D. *et al.* (2008) Assembly reflects evolution of protein complexes. *Nature*, **453**, 1262–1265.
- Marchisio, M.A. and Stelling, J. (2009) Computational design tools for synthetic biology. *Curr. Opin. Biotechnol.*, **20**, 479–485.
- Marks, D.S. *et al.* (2011) Protein 3D structure computed from evolutionary sequence variation. *PLoS One*, **6**, e28766.
- Masum, H. *et al.* (2013) Ten simple rules for cultivating open science and collaborative R&D. *PLoS Comput. Biol.*, **9**, e1003244.
- McGuffee, S.R. and Elcock, A.H. (2010) Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Comput. Biol.*, **6**, e1000694.
- Minton, A. (1993) Macromolecular crowding and molecular recognition. *J. Mol. Recognit.*, **6**, 211–214.
- Moult, J. *et al.* (2014) Critical assessment of methods of protein structure prediction (CASP)—round x. *Proteins*, **82** (Suppl. 2), 1–6.
- Najmanovich, R. *et al.* (2008) Detection of 3D atomic similarities and their use in the discrimination of small molecule protein-binding sites. *Bioinformatics*, **24**, i105–i111.
- Najmanovich, R.J. *et al.* (2005) Prediction of protein function from structure: insights from methods for the detection of local structural similarities. *BioTechniques*, **38**, 847, 849, 851.
- Nugent, T. and Jones, D.T. (2012) Accurate de novo structure prediction of large transmembrane protein domains using fragment-assembly and correlated mutation analysis. *Proc. Natl Acad. Sci. USA*, **109**, E1540–E1547.
- Ollikainen, N. *et al.* (2013) Flexible backbone sampling methods to model and design protein alternative conformations. *Methods Enzymol.*, **523**, 61–85.
- Parisien, M. and Major, F. (2012) Determining RNA three-dimensional structures using low-resolution data. *J. Struct. Biol.*, **179**, 252–260.
- Patel, M.S. *et al.* (2014) The pyruvate dehydrogenase complexes: structure-based function and regulation. *J. Biol. Chem.*, **289**, 16615–16623.
- Peterson, J.R. *et al.* (2014) Towards a computational model of a methane producing archaeum. *Archaea*, **2014**, 898453.
- Potapov, V. *et al.* (2008) Computational redesign of a protein-protein interface for high affinity and binding specificity using modular architecture and naturally occurring template fragments. *J. Mol. Biol.*, **384**, 109–119.
- Potterton, L. *et al.* (2004) Developments in the CCP4 molecular-graphics project. *Acta Crystallogr. D*, **60**, 2288–2294.
- Pronk, S. *et al.* (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, **29**, 845–854.
- Radivojac, P. *et al.* (2013) A large-scale evaluation of computational protein function prediction. *Nat. Methods*, **10**, 221–227.
- Rajamani, D. *et al.* (2004) Anchor residues in protein-protein interactions. *Proc. Natl Acad. Sci. USA*, **101**, 11287–11292.
- Rohl, C.A. *et al.* (2004) Protein structure prediction using Rosetta. *Methods Enzymol.*, **383**, 66–93.
- Rother, M. *et al.* (2011) ModeRNA: a tool for comparative modeling of RNA 3D structure. *Nucleic Acids Res.*, **39**, 4007–4022.
- Samish, I. (2009) Search and sampling in structural bioinformatics. In: Bourne, P.E. and Gu, J. (eds) *Structural Bioinformatics*. Wiley-Blackwell, New York, pp. 207–236.
- Samish, I. *et al.* (2011) Theoretical and computational protein design. *Annu. Rev. Phys. Chem.*, **62**, 129–149.
- Sandve, G.K. *et al.* (2013) Ten simple rules for reproducible computational research. *PLoS Comput Biol.*, **9**, e1003285.
- Shaw, D.E. *et al.* (2009) Millisecond-scale molecular dynamics simulations on Anton. SC '09: *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*. Portland, OR. 14–20 Nov. 2009. doi: 10.1145/1654059.1654126.
- Shirts, M. and Pande, V.S. (2000) COMPUTING: screen savers of the world unite! *Science*, **290**, 1903–1904.
- Sillitoe, I. *et al.* (2013) New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures. *Nucleic Acids Res.*, **41**, D490–D498.
- Stein, M. *et al.* (2007) Bridging from molecular simulation to biochemical networks. *Curr. Opin. Struct. Biol.*, **17**, 166–172.
- Tew, G.N. *et al.* (2010) De novo design of antimicrobial polymers, foldamers, and small molecules: from discovery to practical applications. *Acc. Chem. Res.*, **43**, 30–39.
- Wang, Q. *et al.* (2008) SCWRL and MolIDE: computer programs for side-chain conformation prediction and homology modeling. *Nat. Protoc.*, **3**, 1832–1847.
- Wang, Y. *et al.* (2011) Implementation of accelerated molecular dynamics in NAMD. *Comput. Sci. Discov.*, **4**, 015002.
- Warren, G.L. *et al.* (2006) A critical assessment of docking programs and scoring functions. *J. Med. Chem.*, **49**, 5912–5931.
- Watson, J.D. and Crick, F.H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**, 737–738.
- Way, J.C. *et al.* (2014) Integrating biological redesign: where synthetic biology came from and where it needs to go. *Cell*, **157**, 151–161.
- Winter, C. *et al.* (2012) Protein interactions in 3D: from interface evolution to drug discovery. *J. Struct. Biol.*, **179**, 347–358.
- Wolynes, P.G. *et al.* (1995) Navigating the folding routes. *Science*, **267**, 1619–1620.
- Xie, L. *et al.* (2011a) Structure-based systems biology for analyzing off-target binding. *Curr. Opin. Struct. Biol.*, **21**, 189–199.
- Xie, L. *et al.* (2011b) Drug discovery using chemical systems biology: weak inhibition of multiple kinases may contribute to the anti-cancer effect of nelfinavir. *PLoS Comput. Biol.*, **7**, e1002037.
- Xie, L. *et al.* (2014) Towards structural systems pharmacology to study complex diseases and personalized medicine. *PLoS Comput. Biol.*, **10**, e1003554.
- Yonath, A. (2010) *Polar Bears, Antibiotics, and the Evolving Ribosome (Nobel Lecture)*. Wiley-VCH Verlag.