The Protein Structure Initiative: achievements and visions for the future

Gaetano T. Montelione

Address: Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, Rutgers University, Department of Biochemistry and Molecular Biology, Robert Wood Johnson Medical School, Northeast Structural Genomics Consortium, Piscataway, NJ 08854, USA

Email: guy@cabm.rutgers.edu

Abstract

The Protein Structure Initiative (PSI) was established in 2000 by the National Institutes of General Medical Sciences with the long-term goal of providing 3D (three-dimensional) structural information for most proteins in nature. As advances in genomic sequencing, bioinformatics, homology modelling, and methods for rapid determination of 3D structures of proteins by X-ray crystallography and nuclear magnetic resonance (NMR) converged, it was proposed that our understanding of the biology of protein structure and evolution could be greatly enabled by 'genomic-scale' protein structure determination. Over the past 12 years, the PSI has evolved from a testing bed for new methods of sample and structure production to a core component of a wide range of biology programs.

Introduction

The vision of the PSI is to make 3D protein structure information an integral part of biology research. Structural Genomics has the potential to transform biomedical research, creating a powerful new infrastructure capable of addressing some of the most challenging molecular problems of modern biology. Large-scale genome sequencing efforts have provided new insights into the richness and diversity of life, and the genomic bases of evolution and function. However, natural selection is largely driven by the physical properties of the 3D protein structure. A complete understanding of protein function and evolution, thus, requires information about both protein sequence *and* 3D structure [1,2]. The PSI has developed over three phases: the first phase (from 2000-2001) was a pilot phase to test the feasibility and develop the methodology; the second phase was to solve large numbers of structures using insights from the first phase; and the third phase, PSI:Biology, aims to expand the role of the 3D structure in biological research using advances from the first two phases. In this commentary, I summarize some of the achievements of the PSI and the vision for expanding these in the new

National Institute of General Medical Science PSI: Biology program.

Major recent advances

Several recent reviews have outlined the progress and achievements of the PSI program [3-9]. Ultimately, the success of the initiative will be determined by the scientific impact of the new technologies, reagents, and 3D structures provided into the public domain, and the knowledge that is gained from these data. One operational metric of the program is a count of 3D structures of 'distinct' proteins (or domains), referred to as 'Distinct Structures', deposited into the Protein Data Bank (PDB). Two protein sequences are 'distinct' if they share < 98% sequence identity over the full-length of the shortest sequence of the pair, e.g. though each provides uniquely valuable information, two crystal structures of the same protein bound to different ligands count as a single Distinct Structure. "Novel" structures are defined as those which have < 30% sequence identity with any structure in the PDB at the time of deposition. In the second phase of the PSI program (called PSI2, 2005-2010), investigators achieved their goal set in 2005 of depositing more than

3,000 Distinct Structures into the PDB. Most of these were also Novel Structures, greatly expanding our knowledge of the relationship between protein sequences and 3D structure. Over the full ten years of the PSI program, investigators completed and deposited into the public domain more than 5,000 3D protein structures, including protein-ligand complexes and pairs of X-ray and NMR structures which, though not counted as "Distinct", have important scientific value (see for example [10]).

Many of the structures determined in first and second phases of the PSI program were, at the time of deposition into the PDB, the first representatives from extensive protein domain families [11]. A 'protein domain family' is a set of homologous protein domains likely to have similar structures and possibly similar biochemical functions. These included both domains with known biochemical functions and domains of unknown function, known as DUFs. These structures are being used as templates for modelling tens of thousands of homologous proteins [12-17] and provide a database of protein sequences, structures, and biophysical properties (e.g. chemical shifts) that also inform the fields of protein structure prediction, design, and engineering. By focusing the choice of targets on proteins that have minimal sequence similarity with known structures, PSI structures have greatly increased the size of the non-redundant protein structure knowledge base that is being used to develop improved structure prediction algorithms, including fragment-based search algorithms and knowledge-based atomic potentials. In some cases, these structural data are accompanied by extensive chemical shift, nuclear overhauser effect (NOE), and other NMR data that are being used in hybrid structure determination methods [18].

The PSI has also become the primary contributor of structural data that can be used for testing new methods for protein structure prediction and automated data analysis, including data used in such projects as the Critical Assessment of Structure Prediction [19,20] and the Critical Assessment of Automated Structure Determination of Proteins from NMR data [21]. Structural Genomics projects are unique as they can provide their data for community-wide tests of computational methods without concern regarding how it impacts their priorities for publishing a particular structure. PSI Centers are also involved in collaborative projects aimed at accelerating the field of protein NMR structure analysis [22-28] and computational protein design [29-31].

As part of the community-outreach goals of the PSI program, the National Institute of General Medical Science has created the PSI Structural Biology Knowledge Base (PSI-SBKB) [32], for organizing and disseminating the entire repertoire of scientific information generated by the PSI program, and the PSI Materials Repository (PSI-MR) [33], designed to provide easy, rapid, and broad access to the biochemical reagents produced by PSI Centers, particularly the protein expression systems. These resources serve as a platform for PSI-funded investigators to provide information on protein samples and 3D structures to the broad biological community in an "open source" fashion, in which intermediate results, protein expression systems and protocols, protein structures, and new technologies are made available to the community as soon as the data and/or methods are deemed to be reliable. The PSI-SBKB also provides access to 3D protein models generated using various comparative modeling methods [14], together with coordinates of structures solved by the PSI program.

The PSI program has also instituted a Community Nomination Target (CNT) program, through which scientists can nominate targets for study by PSI centers and collaborate on functional follow on studies (http://sbkb.org/cnt/). This program provides a unique method of connecting PSI investigators with important biological problems and top-tier biological investigators and provides access to PSI Centers by a wide range of collaborators who are not directly funded by the PSI. Several of these CNT projects have yielded important and challenging structures enabling the research programs of individual investigators across the globe (see for example [34-42]).

Future directions: a vision for Structural Genomics in PSI:Biology

The successful demonstration of the feasibility of 'high-throughput structure production' opens doors to a wide range of new opportunities for biological research that could not be considered without such infrastructure. The third phase of the PSI program, PSI:Biology, aims to expand the role of 3D structure in biological research by supporting several "high-throughput-enabled biology partnerships", designed to leverage the protein sample and structure production horsepower of the PSI High-Throughput Production Centers in applications involving broad and/or challenging biological questions. Examples of project areas that are emphasized in the PSI:Biology program include the following:

(a) Metagenomics and microbiomes. Providing one or a few representative 3D structures from the largest protein domain families from ecosystems of organisms (e.g. the human gut microbiome) that are being characterized by large-scale metagenomic sequencing projects.

(b) <u>Individual organisms or organelles</u>. Structural Genomics platforms are providing extensive structural coverage of proteins from a specific organisms, such as *Thermatoga maritima* [43] and *Mycobacterium tuberculosis* [44], or proteins from organelles, such as mitochondria [45] or nuclear pore complexes [46].

(c) <u>Systems biology: networks, pathways, and complexes</u>. Structural Genomics platforms are being directed to provide hundreds of experimental structures and high-quality homology models of proteins and protein complexes involved in co-functioning networks, metabolic systems, such as networks of interacting proteins involved in cancer biology [47], and specific biochemical pathways, such as biosynthetic pathways [48-50].

(d) <u>Individual protein domain families</u>. Structural Genomics platforms are providing (i) extensive, fine-grained structural coverage of particular protein domain families in order to understand the atomic details of structure-function relationships, (ii) studies of ligand binding affinities aimed at understanding the rules of molecular recognition, and (iii) studies of the relationships between structure, function, and internal dynamics across an extensive enzyme family.

(e) <u>Protein engineering and design</u>. Structural Genomics platforms are being applied to validate *de novo* protein designs, to discover new principles for protein engineering, and to illuminate mechanisms for altering enzyme specificity and chemistry.

These exciting applications provide a vision of the broad impact Structural Genomics platforms and technologies will have on biological and biomedical research. Examples of key areas that are being explored in the PSI:Biology program are outlined in the following three sections.

The role of Structural Genomics in functional genomics and systems biology

Structural Genomics provides 3D atomic-resolution structural information of large numbers of gene products (so far, primarily proteins) and, thus, lays the foundation required for systems biology. For example, having protein samples, affinity capture reagents (e.g. phage display antibodies), and complete 3D structural descriptions of the enzymes and protein-protein complexes associated with a specific biological process, such as epigenetically-regulated gene expression or protein translation, will open new avenues to model and understand such complex biological systems. Such a comprehensive view would also allow improved diagnosis and treatment of diseases.

Structural Genomics of protein complexes

Most proteins function by forming complexes. Indeed, some proteins are simply not folded in the absence of their macromolecular and/or small molecule partners. Genomes encode large numbers of natively disordered proteins or protein regions that are functionally important for protein-protein interactions, modulating binding affinities, and regulating signaling pathways [51-55]. This has been born out by biophysical studies on thousands of proteins expressed and purified in the PSI program, demonstrating that a large portion of the eukaryotic proteome codes for intrinsically disordered proteins and/or protein regions. In recent years, research groups of the PSI have begun to address such disordered regions of proteins (or entire intrinsically-disordered protein families), particularly those that become ordered upon complex formation [56]. Important technological goals include development of high-throughput methods for protein co-expression, crystallization-enhancing chaperones generated by phage display methods [57,58], and various technologies for identifying, co-expressing, and forming complexes between proteins, including those that involve disorder-order transitions.

Structural Genomics of membrane proteins

Membrane proteins remain a major challenge to structural biology. However, our understanding of biology will not be complete without extensive structural information on integral membrane proteins. Structural genomics pipelines, involving coordinated teams of scientists working together with shared resources and infrastructure (e.g. [59-61]), have the potential to make major breakthroughs in creating new technologies and protocols for determining 3D structures and dynamics of integral membrane proteins. Some important integral membrane protein structures, including several G protein-coupled receptors, known as GPCRs, and their complexes with ligands, have recently been determined in PSI-funded projects using both X-ray crystallography [62-71] and NMR spectroscopy [72,73]. In the PSI:Biology program, expanded support is provided for the technology development needed for membrane protein sample production, coordinated, project-wide structural analysis of human integral membrane proteins, and community-nominated studies of membrane protein structure and function.

Cooperation between PSI centers

One of the unique features of the PSI program is the extensive cooperation and synergy between potentially competing National Institutes of Health (NIH)-funded centers. For example, in the second phase of the PSI

program, target selection of proteins for structure determination was coordinated by a BioInformatics Group committee that included members from each of the computational biology teams associated with the four Large Scale Centers [11]. The successful collaboration of the different teams resulted not only in more rational and comprehensive target selection but also minimized duplications of effort by providing coordination among hundreds of scientists and synergies that would not be possible in smaller individual laboratory research projects.

Challenges for PSI:Biology

The third phase of the PSI program, PSI:Biology, is an experiment in large-scale biological research. Preliminary progress using large-scale 3D protein structure production to enable biological research partnerships is very exciting. The program currently consists of four Centers for High-Throughput Structure Determination, nine Centers for Membrane Protein Structure Determination, 12 High-Throughput Enabled Biology Partners, and two Resource Centers (the PSI-SBKB and PSI-MR). In addition, several of these centers host extensive CNT projects, involving many research groups that are not directly funded by the PSI:Biology program. These CNT projects have the potential to create new and unexpected uses of protein structure, and to enable, with 3D protein structures, a wide range of biological and biochemical studies.

A challenge faced by the PSI:Biology program is to create strong cooperation and synergies among the 27 PSI:Biology centers, each of which is itself a multi-investigator team, as well between the PSI:Biology centers and individual investigators associated with the CNT programs. Such an integrated program will be a unique engine for biological discovery. Obviously, issues may arise within such a research network that require innovative thinking and creative management. A model for this integration is the cooperation and synergy achieved between the Large Scale Centers. Despite challenges that may present themselves in the early phases of the program, the power of such a unique, integrated infrastructure for biological research has the potential to rapidly advance biology and biomedical science.

Conclusions

The PSI program provides a novel paradigm for biological science discovery. Rather than determining 3D structures as a means for testing specific hypotheses, the Structural Genomics approach aims to discover new science by analyzing the information provided by 3D structures, sometimes even before the biological significance of the protein is recognized. It has been a powerful and successful driving force for a wide range of method

developments that have been realized only by collecting homogeneous, fully-documented data across large numbers of protein samples and structures (see for example [10,21,26,74-81]). The PSI has also provided a test bed for "network biological science", enabling discovery through cooperative interactions across a network of collaborating scientists. The unique multi-laboratory structure of the PSI "centers" also provides a model of how the internet can be used to integrate research activities across a network of real time collaborations. This paradigm is making significant contributions to biology, utilizing the high-throughput platforms developed in the PSI program to enable biology with 3D structural information. Indeed, the concerted effort of the more than 500 scientists participating in the PSI:Biology program has the potential to revolutionize the utility and impact of protein 3D structure information for the broad biological community.

Abbreviations

CNT, community nominated target; NIH, National Institutes of Health; NMR, nuclear magnetic resonance; NOE, nuclear overhauser effect; PDB, Protein Data Bank; PSI, Protein Structure Initiative; PSI-SBKB, PSI Structural Biology Knowledge Base; 3D, three-dimensional.

Competing interests

The author is Principal Investigator of the Northeast Structural Genomics Consortium, one of several NIH National Institute of General Medical Science funded projects of the PSI.

Acknowledgements

References

1. Burley SK, Almo SC, Bonanno JB, Capel M, Chance MR, Gaasterland T, Lin D, Sali A, Studier FW, Swaminathan S: **Structural genomics: beyond the human genome project.** *Nat. Genet* 1999, **23**:151-7.

2. Montelione GT, Anderson S: **Structural genomics: keystone for a Human Proteome Project.** *Nat. Struct. Biol* 1999, **6**:11-2.

3. Burley SK, Joachimiak A, Montelione GT, Wilson IA: **Contributions to the NIH-NIGMS Protein Structure Initiative from the PSI Production Centers.** *Structure* 2008, **16**:5-11.

4. Chandonia J, Brenner SE: **The impact of structural genomics: expectations and outcomes.** *Science* 2006, **311**:347-51.

   > F1000 Factor 8
   > Evaluated by Christine Orengo 03 Apr 2006

5. Levitt M: **Nature of the protein universe.** *Proc. Natl. Acad. Sci. U.S.A* 2009, **106**:11079-84.

   > F1000 Factor 6
   > Evaluated by Tobin Sosnick 01 Dec 2009

6.   Montelione GT, Szyperski T: **Advances in protein NMR provided by the NIGMS Protein Structure Initiative: impact on drug discovery.** *Curr Opin Drug Discov Devel* 2010, **13**:335-49.

7.   Montelione GT, Arrowsmith C, Girvin ME, Kennedy MA, Markley JL, Powers R, Prestegard JH, Szyperski T: **Unique opportunities for NMR methods in structural genomics.** *J. Struct. Funct. Genomics* 2009, **10**:101-6.

8.   Yee A, Gutmanas A, Arrowsmith CH: **Solution NMR in structural genomics.** *Curr. Opin. Struct. Biol* 2006, **16**:611-7.

9.   Terwilliger TC: **The success of structural genomics.** *J. Struct. Funct. Genomics* 2011, **12**:43-4.

10. Mao B, Guan R, Montelione GT: **Improved technologies now routinely provide protein NMR structures useful for molecular replacement.** *Structure* 2011, **19**:757-66.

11. Dessailly BH, Nair R, Jaroszewski L, Fajardo JE, Kouranov A, Lee D, Fiser A, Godzik A, Rost B, Orengo C: **PSI-2: structural genomics to cover protein domain family space.** *Structure* 2009, **17**:869-81.

12. Liu J, Montelione GT, Rost B: **Novel leverage of structural genomics.** *Nat. Biotechnol* 2007, **25**:849-51.

13. Nair R, Liu J, Soong T, Acton TB, Everett JK, Kouranov A, Fiser A, Godzik A, Jaroszewski L, Orengo C, Montelione GT, Rost B: **Structural genomics is the largest contributor of novel structural leverage.** *J. Struct. Funct. Genomics* 2009, **10**:181-91.

     F1000 Factor 8
     Evaluated by Thomas Szyperski 25 Feb 2009

14. Arnold K, Kiefer F, Kopp J, Battey JN, Podvinec M, Westbrook JD, Berman HM, Bordoli L, Schwede T: **The Protein Model Portal.** *J Struct Funct Genomics* 2009, **10**:1-8.

15. Förster F, Lasker K, Nickell S, Sali A, Baumeister W: **Toward an integrated structural model of the 26S proteasome.** *Mol. Cell Proteomics* 2010, **9**:1666-77.

16. Mirkovic N, Li Z, Parnassa A, Murray D: **Strategies for high-throughput comparative modeling: applications to leverage analysis in structural genomics and protein family organization.** *Proteins* 2007, **66**:766-77.

17. Pieper U, Eswar N, Webb BM, Eramian D, Kelly L, Barkan DT, Carter H, Mankoo P, Karchin R, Marti-Renom MA, Davis FP, Sali A: **MODBASE, a database of annotated comparative protein structure models and associated resources.** *Nucleic Acids Res* 2009, **37**:D347-54.

18. Grant TD, Luft JR, Wolfley JR, Tsuruta H, Martel A, Montelione GT, Snell EH: **Small angle X-ray scattering as a complementary tool for high-throughput structural studies.** *Biopolymers* 2011, **95**:517-30.

19. Moult J: **A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction.** *Curr. Opin. Struct. Biol* 2005, **15**:285-9.

20. Moult J, Fidelis K, Kryshtafovych A, Rost B, Tramontano A: **Critical assessment of methods of protein structure prediction - Round VIII.** *Proteins* 2009, **77**(Suppl 9):1-4.

21. Rosato A, Bagaria A, Baker D, Bardiaux B, Cavalli A, Doreleijers JF, Giachetti A, Guerry P, Güntert P, Herrmann T, Huang YJ, Jonker HRA, Mao B, Malliavin TE, Montelione GT, Nilges M, Raman S, van der Schot G, Vranken WF, Vuister GW, Bonvin AMJJ: **CASD-NMR: critical assessment of automated structure determination by NMR.** *Nat. Methods* 2009, **6**:625-6.

     F1000 Factor 8
     Evaluated by Gottfried Otting 11 Mar 2010

22. Raman S, Lange OF, Rossi P, Tyka M, Wang X, Aramini J, Liu G, Ramelot TA, Eletsky A, Szyperski T, Kennedy MA, Prestegard J, Montelione GT, Baker D: **NMR structure determination for larger proteins using backbone-only data.** *Science* 2010, **327**:1014-8.

     F1000 Factor 8
     Evaluated by H Jane Dyson and Gira Bhabha 03 Mar 2010, Antonio Rosato 03 Mar 2010, Gabriele Varani and Yu Chen 06 Apr 2010

23. Tang Y, Schneider WM, Shen Y, Raman S, Inouye M, Baker D, Roth MJ, Montelione GT: **Fully automated high-quality NMR structure determination of small (2)H-enriched proteins.** *J. Struct. Funct. Genomics* 2010, **11**:223-32.

24. Sgourakis NG, Lange OF, DiMaio F, André I, Fitzkee NC, Rossi P, Montelione GT, Bax A, Baker D: **Determination of the structures of symmetric protein oligomers from NMR chemical shifts and residual dipolar couplings.** *J. Am. Chem. Soc* 2011, **133**:6288-98.

25. Raman S, Huang YJ, Mao B, Rossi P, Aramini JM, Liu G, Montelione GT, Baker D: **Accurate automated protein NMR structure determination using unassigned NOESY data.** *J. Am. Chem. Soc* 2010, **132**:202-7.

     F1000 Factor 6
     Evaluated by Thomas Szyperski 06 Jul 2010

26. Shen Y, Lange O, Delaglio F, Rossi P, Aramini JM, Liu G, Eletsky A, Wu Y, Singarapu KK, Lemak A, Ignatchenko A, Arrowsmith CH, Szyperski T, Montelione GT, Baker D, Bax A: **Consistent blind protein structure generation from NMR chemical shift data.** *Proc. Natl. Acad. Sci. U.S.A* 2008, **105**:4685-90.

     F1000 Factor 17
     Evaluated by Antonio Rosato 26 Mar 2008, Gottfried Otting 04 Apr 2008, Deyou Zheng 21 Apr 2008, Torleif Härd 21 May 2008, Mikael Akke 27 May 2008, Dorothee Kern 18 Aug 2008

27. Atreya HS, Szyperski T: **Rapid NMR data collection.** *Meth. Enzymol* 2005, **394**:78-108.

28. Liu G, Shen Y, Atreya HS, Parish D, Shao Y, Sukumaran DK, Xiao R, Yee A, Lemak A, Bhattacharya A, Acton TA, Arrowsmith CH, Montelione GT, Szyperski T: **NMR data collection and analysis protocol for high-throughput protein structure determination.** *Proc. Natl. Acad. Sci. U.S.A* 2005, **102**:10487-92.

     F1000 Factor 7
     Evaluated by Gottfried Otting 25 Oct 2005, Deyou Zheng 03 Nov 2005

29. Jha RK, Wu YI, Zawistowski JS, MacNevin C, Hahn KM, Kuhlman B: **Redesign of the PAK1 autoinhibitory domain for enhanced stability and affinity in biosensor applications.** *J. Mol. Biol* 2011, **413**:513-22.

30. Karanicolas J, Corn JE, Chen I, Joachimiak LA, Dym O, Peck SH, Albeck S, Unger T, Hu W, Liu G, Delbecq S, Montelione GT, Spiegel CP, Liu DR, Baker D: **A de novo protein binding pair by computational design and directed evolution.** *Mol. Cell* 2011, **42**:250-60.

31. Der BS, Machius M, Miley MJ, Mills JL, Szyperski T, Kuhlman B: **Metal-mediated affinity and orientation specificity in a computationally designed protein homodimer.** *J. Am. Chem. Soc* 2012, **134**:375-85.

32. Gabanyi MJ, Adams PD, Arnold K, Bordoli L, Carter LG, Flippen-Andersen J, Gifford L, Haas J, Kouranov A, McLaughlin WA, *et al.*: **The Structural Biology Knowledgebase: a portal to protein structures, sequences, functions, and methods.** *J Struct Funct Genomics* 2011, **12**:45-54.

33. Cormier CY, Park JG, Fiacco M, Steel J, Hunter P, Kramer J, Singla R, LaBaer J: **PSI:Biology-materials repository: a biologist's resource for protein expression plasmids.** *J. Struct. Funct. Genomics* 2011, **12**:55-62.

34. Forouhar F, Yang Y, Kumar D, Chen Y, Fridman E, Park SW, Chiang Y, Acton TB, Montelione GT, Pichersky E, Klessig DF, Tong L: **Structural and biochemical studies identify tobacco SABP2 as a methyl salicylate esterase and implicate it in plant innate immunity.** *Proc. Natl. Acad. Sci. U.S.A* 2005, **102**:1773-8.

     F1000 Factor 6
     Evaluated by Eric Lam 08 Mar 2005

35. Bayro MJ, Mukhopadhyay J, Swapna GVT, Huang JY, Ma L, Sineva E, Dawson PE, Montelione GT, Ebright RH: **Structure of antibacterial peptide microcin J25: a 21-residue lariat protoknot.** *J. Am. Chem. Soc* 2003, **125**:12382-3.

36. Zhang H, Constantine R, Vorobiev S, Chen Y, Seetharaman J, Huang YJ, Xiao R, Montelione GT, Gerstner CD, Davis MW, Inana G, Whitby FG, Jorgensen EM, Hill CP, Tong L, Baehr W: **UNC119 is required for G protein trafficking in sensory neurons.** *Nat. Neurosci* 2011, **14**:874-80.

37. Forouhar F, Anderson JLR, Mowat CG, Vorobiev SM, Hussain A, Abashidze M, Bruckmann C, Thackray SJ, Seetharaman J, Tucker T, Xiao R, Ma L, Zhao L, Acton TB, Montelione GT, Chapman SK, Tong L: **Molecular insights into substrate recognition and catalysis by tryptophan 2,3-dioxygenase.** *Proc. Natl. Acad. Sci. U.S.A* 2007, **104**:473-8.

38. Forouhar F, Kuzin A, Seetharaman J, Lee I, Zhou W, Abashidze M, Chen Y, Yong W, Janjua H, Fang Y, Wang D, Cunningham K, Xiao R, Acton TB, Pichersky E, Klessig DF, Porter CW, Montelione GT, Tong L: **Functional insights from structural genomics.** *J. Struct. Funct. Genomics* 2007, **8**:37-44.

39. Aramini JM, Ma L, Zhou L, Schauder CM, Hamilton K, Amer BR, Mack TR, Lee H, Ciccosanti CT, Zhao L, Xiao R, Krug RM, Montelione GT: **Dimer interface of the effector domain of non-structural protein 1 from influenza A virus: an interface with multiple functions.** *J. Biol. Chem* 2011, **286**:26050-60.

40. Yin C, Khan JA, Swapna GVT, Ertekin A, Krug RM, Tong L, Montelione GT: **Conserved surface features form the double-stranded RNA binding site of non-structural protein 1 (NS1) from influenza A and B viruses.** *J. Biol. Chem* 2007, **282**:20584-92.

41. Das K, Ma L, Xiao R, Radvansky B, Aramini J, Zhao L, Marklund J, Kuo R, Twu KY, Arnold E, Krug RM, Montelione GT: **Structural basis for suppression of a host antiviral response by influenza A virus.** *Proc. Natl. Acad. Sci. U.S.A* 2008, **105**:13093-8.

42. Guan R, Ma L, Leonard PG, Amer BR, Sridharan H, Zhao C, Krug RM, Montelione GT: **Structural basis for the sequence-specific recognition of human ISG15 by the NS1 protein of influenza B virus.** *Proc. Natl. Acad. Sci. U.S.A* 2011, **108**:13468-73.

43. Zhang Y, Thiele I, Weekes D, Li Z, Jaroszewski L, Ginalski K, Deacon AM, Wooley J, Lesley SA, Wilson IA, Palsson B, Osterman A, Godzik A: **Three-dimensional structural view of the central metabolic network of Thermotoga maritima.** *Science* 2009, **325**:1544-9.

> F1000 Factor 12
> Evaluated by Chantal Abergel 12 Oct 2009, Torsten Schwede 14 Oct 2009

44. Chim N, Habel JE, Johnston JM, Krieger I, Miallau L, Sankaranarayanan R, Morse RP, Bruning J, Swanson S, Kim H, Kim C, Li H, Bulloch EM, Payne RJ, Manos-Turvey A, Hung L, Baker EN, Lott JS, James MNG, Terwilliger TC, Eisenberg DS, Sacchettini JC, Goulding CW: **The TB Structural Genomics Consortium: a decade of progress.** *Tuberculosis (Edinb)* 2011, **91**:155-72.

45. Pagliarini DJ, Calvo SE, Chang B, Sheth SA, Vafai SB, Ong S, Walford GA, Sugiana C, Boneh A, Chen WK, Hill DE, Vidal M, Evans JG, Thorburn DR, Carr SA, Mootha VK: **A mitochondrial protein compendium elucidates complex I disease biology.** *Cell* 2008, **134**:112-23.

> F1000 Factor 8
> Evaluated by Massimo Zeviani and Erika Fernandez-Vizarra 24 Jul 2008

46. Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, Devos D, Suprapto A, Karni-Schmidt O, Williams R, Chait BT, Sali A, Rout MP: **The molecular architecture of the nuclear pore complex.** *Nature* 2007, **450**:695-701.

> F1000 Factor 15
> Evaluated by Petra Fromme 18 Dec 2007, Tom Misteli 18 Dec 2007, Paul Clarke 21 Dec 2007

47. 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**:2048-60.

48. Burley SK, Bonanno JB: **Structural genomics of proteins from conserved biochemical pathways and processes.** *Curr. Opin. Struct. Biol* 2002, **12**:383-91.

49. Bonanno JB, Edo C, Eswar N, Pieper U, Romanowski MJ, Ilyin V, Gerchman SE, Kycia H, Studier FW, Sali A, Burley SK: **Structural genomics of enzymes involved in sterol/isoprenoid biosynthesis.** *Proc. Natl. Acad. Sci. U.S.A* 2001, **98**:12896-901.

50. Lu LJ, Sboner A, Huang YJ, Lu HX, Gianoulis TA, Yip KY, Kim PM, Montelione GT, Gerstein MB: **Comparing classical pathways and modern networks: towards the development of an edge ontology.** *Trends Biochem. Sci* 2007, **32**:320-31.

51. Uversky VN, Oldfield CJ, Dunker AK: **Intrinsically disordered proteins in human diseases: introducing the D2 concept.** *Annu Rev Biophys* 2008, **37**:215-46.

52. Oldfield CJ, Meng J, Yang JY, Yang MQ, Uversky VN, Dunker AK: **Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners.** *BMC Genomics* 2008, **9**(Suppl 1):S1.

53. Romero PR, Zaidi S, Fang YY, Uversky VN, Radivojac P, Oldfield CJ, Cortese MS, Sickmeier M, LeGall T, Obradovic Z, Dunker AK: **Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms.** *Proc. Natl. Acad. Sci. U.S.A* 2006, **103**:8390-5.

54. Wright PE, Dyson HJ: **Linking folding and binding.** *Curr. Opin. Struct. Biol* 2009, **19**:31-8.

55. Huang YJ, Montelione GT: **Structural biology: proteins flex to function.** *Nature* 2005, **438**:36-7.

56. Oldfield CJ, Ulrich EL, Cheng Y, Dunker AK, Markley JL: **Addressing the intrinsic disorder bottleneck in structural proteomics.** *Proteins* 2005, **59**:444-53.

57. Uysal S, Vásquez V, Tereshko V, Esaki K, Fellouse FA, Sidhu SS, Koide S, Perozo E, Kossiakoff A: **Crystal structure of full-length KcsA in its closed conformation.** *Proc. Natl. Acad. Sci. U.S.A* 2009, **106**:6644-9.

58. Tereshko V, Uysal S, Koide A, Margalef K, Koide S, Kossiakoff AA: **Toward chaperone-assisted crystallography: protein engineering enhancement of crystal packing and X-ray phasing capabilities of a camelid single-domain antibody (VHH) scaffold.** *Protein Sci* 2008, **17**:1175-87.

> F1000 Factor 8
> Evaluated by David Waugh 01 Jul 2008

59. Love J, Mancia F, Shapiro L, Punta M, Rost B, Girvin M, Wang D, Zhou M, Hunt JF, Szyperski T, Gouaux E, MacKinnon R, McDermott A, Honig B, Inouye M, Montelione G, Hendrickson WA: **The New York Consortium on Membrane Protein Structure (NYCOMPS): a high-throughput platform for structural genomics of integral membrane proteins.** *J. Struct. Funct. Genomics* 2010, **11**:191-9.

60. Lee JK, Stroud RM: **Unlocking the eukaryotic membrane protein structural proteome.** *Curr. Opin. Struct. Biol* 2010, **20**:464-70.

61. Stroud RM, Choe S, Holton J, Kaback HR, Kwiatkowski W, Minor DL, Riek R, Sali A, Stahlberg H, Harries W: **2007 annual progress report synopsis of the Center for Structures of Membrane Proteins.** *J. Struct. Funct. Genomics* 2009, **10**:193-208.

62. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SGF, Thian FS, Kobilka TS, Choi H, 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**:1258-65.

> F1000 Factor 30
> Evaluated by Xiayang Qiu 31 Oct 2007, Mark von Zastrow 06 Nov 2007, Peter Artymiuk 27 Nov 2007, Patricia C Weber 03 Dec 2007, Ram Samudrala 14 Dec 2007, Petra Fromme 18 Dec 2007, Marc Caron 10 Jan 2008, Susan Steinberg 30 Jan 2009

63. Rosenbaum DM, Cherezov V, Hanson MA, Rasmussen SGF, Thian FS, Kobilka TS, Choi H, Yao X, Weis WI, Stevens RC, Kobilka BK: **GPCR**

engineering yields high-resolution structural insights into beta2-adrenergic receptor function. *Science* 2007, **318:**1266-73.

F1000 Factor 19
Evaluated by Xiayang Qiu 31 Oct 2007, Senyon Choe 19 Nov 2007, Marc Caron 10 Jan 2008, Thomas Baranski 26 Feb 2008

64. Jaakola V, Griffith MT, Hanson MA, Cherezov V, Chien EYT, Lane JR, Ijzerman AP, Stevens RC: **The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist.** *Science* 2008, **322:**1211-7.

F1000 Factor 10
Evaluated by John Auchampach 24 Oct 2008, John Saunders 12 Jan 2009, Hao Wu and Yu-Chih Lo 26 Jan 2009

65. Xu F, Wu H, Katritch V, Han GW, Jacobson KA, Gao Z, Cherezov V, Stevens RC: **Structure of an agonist-bound human A2A adenosine receptor.** *Science* 2011, **332:**322-7.

F1000 Factor 13
Evaluated by Manoj Desai and Britton Corkey 19 Apr 2011, John Overington and Felix Alexander Krüger 06 May 2011, Nikolay Dokholyan and Feng Ding 13 May 2011

66. Shimamura T, Shiroishi M, Weyand S, Tsujimoto H, Winter G, Katritch V, Abagyan R, Cherezov V, Liu W, Han GW, Kobayashi T, Stevens RC, Iwata S: **Structure of the human histamine H1 receptor complex with doxepin.** *Nature* 2011, **475:**65-70.

F1000 Factor 12
Evaluated by John Kyongwon Lee 30 Jun 2011, John Overington and Felix Alexander Krüger 04 Jul 2011

67. Wu B, Chien EYT, Mol CD, Fenalti G, Liu W, Katritch V, Abagyan R, Brooun A, Wells P, Bi FC, Hamel DJ, Kuhn P, Handel TM, Cherezov V, Stevens RC: **Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists.** *Science* 2010, **330:**1066-71.

68. Chien EYT, Liu W, Zhao Q, Katritch V, Han GW, Hanson MA, Shi L, Newman AH, Javitch JA, Cherezov V, Stevens RC: **Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist.** *Science* 2010, **330:**1091-5.

F1000 Factor 10
Evaluated by Bryan Roth 26 Nov 2010

69. Chen Y, Hu L, Punta M, Bruni R, Hillerich B, Kloss B, Rost B, Love J, Siegelbaum SA, Hendrickson WA: **Homologue structure of the SLAC1 anion channel for closing stomata in leaves.** *Nature* 2010, **467:**1074-80.

F1000 Factor 13
Evaluated by Enrico Martinoia 15 Nov 2010, Julian Schroeder and Aaron Stephan 16 Nov 2010, Christophe Maurel and Lionel Verdoucq 07 Dec 2010

70. Savage DF, O'Connell JD, Miercke LJW, Finer-Moore J, Stroud RM: **Structural context shapes the aquaporin selectivity filter.** *Proc. Natl. Acad. Sci. U.S.A* 2010, **107:**17164-9.

71. Gruswitz F, Chaudhary S, Ho JD, Schlessinger A, Pezeshki B, Ho C, Sali A, Westhoff CM, Stroud RM: **Function of human Rh based on structure of RhCG at 2.1 A.** *Proc. Natl. Acad. Sci. U.S.A* 2010, **107:**9638-43.

F1000 Factor 6
Evaluated by Jeff Kraut 24 May 2010

72. Reckel S, Gottstein D, Stehle J, Löhr F, Verhoefen M, Takeda M, Silvers R, Kainosho M, Glaubitz C, Wachtveitl J, Bernhard F,

Schwalbe H, Güntert P, Dötsch V: **Solution NMR structure of proteorhodopsin.** *Angew. Chem. Int. Ed. Engl* 2011, **50:**11942-6.

73. Berardi MJ, Shih WM, Harrison SC, Chou JJ: **Mitochondrial uncoupling protein 2 structure determined by NMR molecular fragment searching.** *Nature* 2011, **476:**109-13.

74. Gräslund S, Nordlund P, Weigelt J, Hallberg BM, Bray J, Gileadi O, Knapp S, Oppermann U, Arrowsmith C, Hui R, Ming J, dhe-Paganon S, Park H, Savchenko A, Yee A, Edwards A, Vincentelli R, Cambillau C, Kim R, Kim S, Rao Z, Shi Y, Terwilliger TC, Kim C, Hung L, Waldo GS, Peleg Y, Albeck S, Unger T, Dym O, et al.: **Protein production and purification.** *Nat. Methods* 2008, **5:**135-46.

F1000 Factor 6
Evaluated by Gottfried Otting 20 Aug 2008

75. Xiao R, Anderson S, Aramini J, Belote R, Buchwald WA, Ciccosanti C, Conover K, Everett JK, Hamilton K, Huang YJ, Janjua H, Jiang M, Kornhaber GJ, Lee DY, Locke JY, Ma L, Maglaqui M, Mao L, Mitra S, Patel D, Rossi P, Sahdev S, Sharma S, Shastry R, Swapna GVT, Tong SN, Wang D, Wang H, Zhao L, Montelione GT, et al.: **The high-throughput protein sample production platform of the Northeast Structural Genomics Consortium.** *J. Struct. Biol* 2010, **172:**21-33.

76. Pantazatos D, Kim JS, Klock HE, Stevens RC, Wilson IA, Lesley SA, Woods VL: **Rapid refinement of crystallographic protein construct definition employing enhanced hydrogen/deuterium exchange MS.** *Proc. Natl. Acad. Sci. U.S.A* 2004, **101:**751-6.

F1000 Factor 13
Evaluated by Patricia C Weber 23 Jan 2004, Celerino Abad-Zapatero 30 Jan 2004, Song Tan 10 Feb 2004, David Waugh 26 Feb 2004

77. Sharma S, Zheng H, Huang YJ, Ertekin A, Hamuro Y, Rossi P, Tejero R, Acton TB, Xiao R, Jiang M, Zhao L, Ma L, Swapna GVT, Aramini JM, Montelione GT: **Construct optimization for protein NMR structure analysis using amide hydrogen/deuterium exchange mass spectrometry.** *Proteins* 2009, **76:**882-94.

78. Vinarov DA, Lytle BL, Peterson FC, Tyler EM, Volkman BF, Markley JL: **Cell-free protein production and labeling protocol for NMR-based structural proteomics.** *Nat. Methods* 2004, **1:**149-53.

79. Yee AA, Savchenko A, Ignachenko A, Lukin J, Xu X, Skarina T, Evdokimova E, Liu CS, Semesi A, Guido V, Edwards AM, Arrowsmith CH: **NMR and X-ray crystallography, complementary tools in structural proteomics of small proteins.** *J. Am. Chem. Soc* 2005, **127:**16512-7.

F1000 Factor 7
Evaluated by Deyou Zheng 06 Dec 2005, Thomas Szyperski 20 Jan 2006

80. Snyder DA, Chen Y, Denissova NG, Acton T, Aramini JM, Ciano M, Karlin R, Liu J, Manor P, Rajan PA, Rossi P, Swapna GVT, Xiao R, Rost B, Hunt J, Montelione GT: **Comparisons of NMR spectral quality and success in crystallization demonstrate that NMR and X-ray crystallography are complementary methods for small protein structure determination.** *J. Am. Chem. Soc* 2005, **127:**16505-11.

F1000 Factor 7
Evaluated by Deyou Zheng 06 Dec 2005, Thomas Szyperski 20 Jan 2006

81. Bhattacharya A, Tejero R, Montelione GT: **Evaluating protein structures determined by structural genomics consortia.** *Proteins* 2007, **66:**778-95.

F1000 Factor 6
Evaluated by Linda Ball and Victoria Ann Higman 31 Dec 2010