Journal of Biochemistry and Molecular Biology, Vol. 37, No. 1, January 2004, pp. 28-34

JOURNAL OF
Biochemistry and Molecular Biology

© KSBMB & Springer-Verlag 2004

Review

Structure-based Functional Discovery of Proteins: Structural Proteomics

Jin-Won Jung and Weontae Lee*
Department of Biochemistry and Protein Network Research Center, College of Science, Yonsei University, Seoul 120-749, Korea

Received 11 December 2003

The discovery of biochemical and cellular functions of unannotated gene products begins with a database search of proteins with structure/sequence homologues based on known genes. Very recently, a number of frontier groups in structural biology proposed a new paradigm to predict biological functions of an unknown protein on the basis of its three-dimensional structure on a genomic scale. Structural proteomics (genomics), a research area for structure-based functional discovery, aims to complete the protein-folding universe of all gene products in a cell. It would lead us to a complete understanding of a living organism from protein structure. Two major complementary experimental techniques, X-ray crystallography and NMR spectroscopy, combined with recently developed high throughput methods have played a central role in structural proteomics research; however, an integration of these methodologies together with comparative modeling and electron microscopy would speed up the goal for completing a full dictionary of protein folding space in the near future.

Keywords: High-throughput structure determination, NMR, Protein-folding space, Structural proteomics, X-ray crystallography

Introduction

The success of genome projects has given us vast amounts of sequence information of whole genes for some species; however, the biological functions of most gene products are still unknown. In order to gain a functional understanding of unannotated genes, a number of frontiers in structural biology proposed a new paradigm for the structural determination of proteins (Kim, 1998; Zarembinski *et al.*, 1998; Hwang *et al.*, 1999; Yee *et al.*, 2002; Schmid, 2002). Structural biologists believe that the whole characters of living cells possess their own molecular and cellular functions, which could be tightly regulated by three-dimensional structures of proteins encoded in a cell (Fig. 1). For this reason, the structural genomics effort, which aims for a structure-based functional discovery of all unannotated gene products, has become a challenging and ambitious research topic. Based on the statistics of Targetdb (http://targetdb.rcsb.org), a total of 635 protein structures were currently determined from structural genomics consortium during the last three years. Among them, 500 structures were determined from X-ray crystallography and 135 from NMR spectroscopy. The Protein Data Bank (PDB; http://www.rcsb.org/pdb) currently contains 332 proteins that are deposited with a structural genomics keyword. Since there is no doubt that the three-dimensional structure of a protein could provide an essential clue for the fundamental questions about its biological functions, structural proteomics projects could fill up the universe of protein folding space provided an accurate prediction of the structure-function relationship. For structural proteomics research, two major experimental techniques (X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy) with newly developed high throughput techniques (Staunton *et al.*, 2003; Terwilliger, 2003; Yee *et al.*, 2003) have played a central role in determining the protein structures on a genomic scale. Other approaches, homology modeling (Sanchez *et al.*, 2000) and electron microscopy (Henderson, 1995) will also become important tools in the near future since these methods could be complementary to NMR and X-ray crystallography in structural proteomics research. This review will summarize recent progress and perspectives in structural proteomics research.

*To whom correspondence should be addressed.
Tel: 82-2-2123-2706; Fax: 82-2-363-2706
E-mail: wlee@spin.yonsei.ac.kr

Function Discovery from Protein Structure

The database search to predict biochemical and cellular function of unannotated gene products begins with structural homologues based on sequence information. Even though it is still not straightforward to derive biological functions of a hypothetical protein directly from its spatial arrangement of

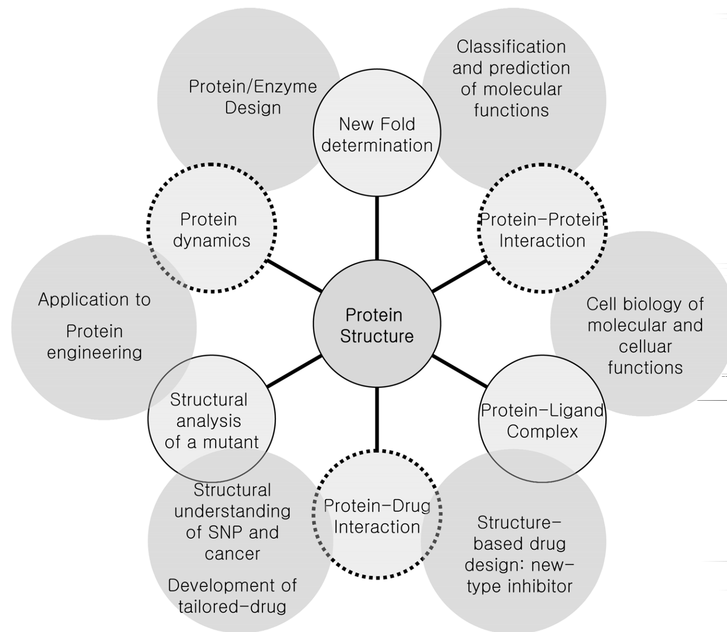

Fig. 1. The overall scope of the protein structure-function relationship. Small circles represent a basic research area from protein structure; big circles are applied areas based on structural information. The big circle is often overlapped with small circles.

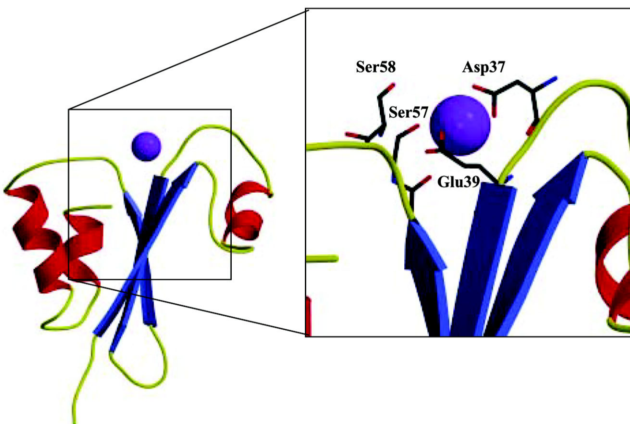

Fig. 2. Structure-based function discovery of a hypothetical protein, MTH1880. NMR structure of MTH1880 is displayed with calcium ion. Ligand binding site is located on the terminal region of β-sheet.

atoms, several pilot projects from pioneer groups in structural biology demonstrated successfully that the molecular functions of hypothetical proteins could be predicted from their three-dimensional structures (Boggon *et al.*, 1999; Hwang *et al.*, 1999; Lee *et al.*, 2004). For instance, a hypothetical protein, MTH1880 from *Methanobacterium thermoautotrophicum* (MTH), shows no sequence homology with other known proteins; however, the solution structure suggests that it is a calcium buffering protein (Lee *et al.*, 2004). The molecular surface of MTH1880 comprises a highly acidic pocket that is anchored by two loops, which is indicative that the protein would have a cation-binding site. As a result, MTH1880 is determined as a novel fold with a unique structural motif for calcium-specific binding and calcium-buffering function (Fig. 2). Another excellent example is a MJ0882, a hypothetical protein from *Methanococcus jannaschii* (MJ) that was determined by X-ray crystallography (Huang *et al.*, 2003). Even though MJ0882 has no sequence similarity with any of the known methyltransferases, the methyltransferase activity of MJ0882 was predicted from a three-dimensional structure. It was confirmed by a biochemical assay. One of major objectives in structural proteomics is to provide new functional insight for proteins with unknown functions. It is essential to categorize the determined structures into several groups together with annotated proteins for a structure-based functional prediction. Recently, Yee *et al.* (2003) demonstrated a total of 27 hypothetical proteins from MTH that could be categorized to deduce their functional annotation from a structure that was determined by NMR spectroscopy or X-ray crystallography.

NMR Spectroscopy and Structural Proteomics

A pilot project for the *Methanobacterium thermoautotrophicum* genome analysis revealed that the proteins under 30 kDa molecular weight occupy 65% of the whole genome. Also, 45% (424 proteins) out of randomly collected 900 proteins were under 20 kDa, except the membrane proteins and proteins with a known structure (Christendat *et al.*, 2000). The Montelione group reported that 25% of the yeast genome is feasible for NMR structure determination (Montelione *et al.*,

2000). The data from a gene sequence database suggested that proteins greater than 30 kDa usually consist of modules that have an independent folding module of about 175 amino acids. This implies that a NMR structure determination is useful for practical NMR-based structural proteomics research. For large proteins, a newly developed TROSY-based NMR technique (Pervushin *et al.*, 1997) could be used to determine the structures of up to 100 kDa molecular weight. TROSY experiments promise NMR spectroscopy the possibility to determine the structure of membrane proteins with a protein-micelle complex as well as multimeric proteins. In addition, the NMR-based structural proteomics would be successful combined with "high-throughput" techniques that are based on newly developed novel NMR methods.

Using conventional NMR equipments, it will take about 6 months for a NMR measurement to determine a high-resolution structure of a protein with 25 kDa (Garrett *et al.*, 1999). This is due to the intrinsic low sensitivity of NMR spectroscopy. To overcome the low sensitivity problem, a protein sample of relatively high concentration (about 1 mM) is frequently used; however, it often creates protein aggregation. Recently, the NMR magnets with ultra-high field strength (>600 MHz) were developed, providing sensitivity as well as high-resolution. For example, the signal sensitivity of a 900 MHz magnet is expected to be 4~5 times higher than that of a 500 MHz magnet. The resolving power is also twice as high. Therefore, shortening the experimental time and increasing the resolution are very useful in the NMR-based structural proteomics study. The current state-of-the-art, cryo-probe, (chilly probe) is designed to decrease heat resistance of the probe circuit, hence it dramatically improves the signal sensitivity by a factor of 4-5 over that of conventional ones (Service, 1998). This technology is especially powerful in screening a massive number of protein samples, which is prerequisite in structural proteomics for both NMR and X-ray crystallography (Yee *et al.*, 2002).

Conventional NMR techniques could not be readily applied to proteins with molecular weights over 25 kDa, due to the transverse relaxation mechanism that occurred by a slow molecular tumbling motion. It results in a significant loss of NMR signal and line broadening (Wagner, 1993). Recently, the TROSY (Transverse Relaxation Optimization Spectroscopy) technique developed by the Wüthrich group reduces transverse nuclear spin relaxation during chemical shift evolution, which provides much better sensitivity and narrow line width for large proteins (Pervushin *et al.*, 1997). The application of the TROSY method could be particularly useful in determining the structure of membrane proteins in membrane mimicking environments (Fernandez *et al.*, 2001), because the micelle condition would create a very slow tumbling motion of the protein. Even though it needed to be optimized in many different systems, it offered a great opportunity in NMR-based structural genomics by overcoming the size limitation of the protein.

In many cases, proteins with multi-domains experience a conformational change in response to biological stimuli. Previously, it was nearly impossible to determine the inter-domain orientation for multi-domain protein by NMR techniques. A recent method has demonstrated that a very weak magnetic alignment, achieved with either phage or bicelle media, could be used to measure the residual dipolar couplings (RDC) of the protein (Tjandra *et al.*, 1997). These couplings can be analyzed to determine molecular alignment tensors for individual protein domains, which provide the relative domain orientations of the protein by an examination of the principal axes of the tensors. In addition, RDC information could serve as conformational constraints to refine the low-resolution structure for rapid determination of protein folds, which is very useful in structural proteomics (Fowler *et al.*, 2000).

One of the major bottlenecks would be an extraction of because it is the most time-consuming job. An integrated structural restraints from NOESY spectrum platform for automatic data analysis would be important to overcoming this problem. Recently, the Autoassign (Moseley and Montelione, 1999), GARANT (O'Connell *et al.*, 1999), MAPPER (Güntert *et al.,* 2000) programs were developed for automated resonance assignment. They successfully demonstrated about 80~90% of the backbone atoms with a set of high quality triple-resonance spectrum. However, NOESY spectrum normally showed greater ambiguity to assign, therefore, it takes extra time to resolve ambiguity by the iterative structure calculation. The ARIA (Nilges *et al.*, 1997) and (Herrmann *et al.*, 2002) programs are designed to solve CYANA this problem by combining the structure calculation and assignment procedure. Recently, the Montelione group developed a program AutoStructure that is a combined package of Autoassign and DYANA (Greenfield *et al.*, 2001). As a pilot project, our laboratory successfully demonstrated a semi-automatic structure calculation based on the automation platform for the BCCP domain of HP0371 that is derived from *Helicobactor pyroli* (Jung and Lee, *unpublished result*) (Fig. 3). Therefore, the development of the automation platform could contribute to the rapid determination of protein folds by NMR.

The large-scale production of the protein normally requires the optimization of expression condition, solubilization and stability after protein expression. A newly developed cell-free expression system enables one to overcome many problems that are associated with the cell-based expression system, which are low-solubility, aggregation and toxicity of the gene product (Kigawa *et al.*, 1999). This method could be directly used for isotope labeling and Se-Met labeling for both NMR and X-ray crystallography in structural proteomics.

The Roles of X-ray Crystallography

X-ray crystallography has been considered the most powerful technique for the determination of the protein structure as well

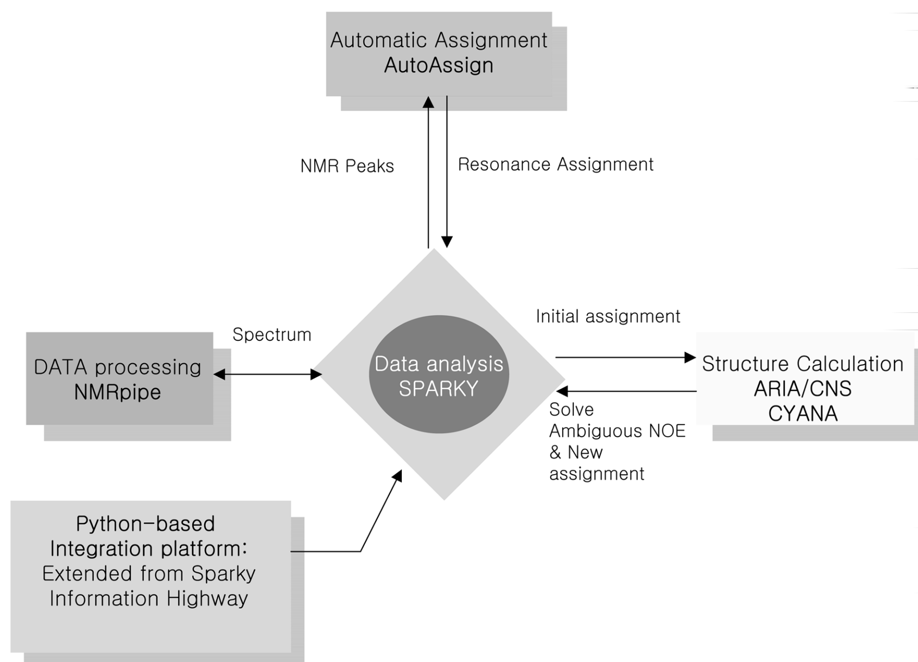

Fig. 3. A schematic drawing of automation platform to calculate solution structure from NMR data. During the structure calculation and NMR data analysis, there is no human-interpreted structural information used. All of the structure calculations will be done with program CYANA (Herrmann *et al.*, 2002).

as protein complexes during the last decade. In addition, the molecular sizes of proteins determined by X-ray crystallography with modern synchrotron facility range from a small protein up to the large 70S ribosomal complex (~2,500 kDa), which could cover a whole protein universe in a cell (Sali *et al.*, 2003). X-ray crystallography is particularly useful in determining the protein structure with post-translational modifications, which is frequently found in a living organism. However, even though X-ray crystallography is a powerful method for structural genomics study, it still has a number of hurdles to overcome. For example, the crystallization procedure of the protein solution, which is an essential step for structure determination, is not straightforward and remains a bottleneck. There are many unpredictable variables and multiple conditions that should be tried to get an optimum crystal for each protein.

A major concern of the high-throughput structure determination is robotics for the automation of the crystallization step. This procedure begins with obtaining milligram quantities of soluble proteins with high purity criteria (>95%). There are many commercially available crystallization-screening kits for high-quality protein crystals (Hui and Edwards, 2003). These kits are comprised of at least 480 conditions (5 plates of 96 well format kits) and dispensing of the protein solution to each of the 480 wells will be routinely very slow. Another major technical achievement of robotics for crystallization is nano-screening by handling a small solution quantity (Abola *et al.*, 2000). The dispensing robotics can handle a few nano-liters of solution, which enables one to greatly reduce the amount of protein that is necessary for screening a wide range of conditions with the same amount of proteins (Rupp, 2003). The newly designed robots perform nano screening of about 10,000-100,000 trials a day for various crystallization conditions. Especially, the nano-screening technique provides great benefit to proteins and protein complexes with low expression. Usually, conventional crystallization of protein required relative high protein concentrations in solution (about 10 mg/ml). Therefore, the ability to handle a small amount of solution such as the nano-screening technique, greatly speeds up the crystallization procedure. Once the crystallization step is completed, the crystals are prepared for the data collection. Since the multi-wave length anomalous dispersion (MAD) data collection induces a radiation decay to create a serious impact to protein, freezing the protein crystals at the cryogenic temperature needs to be performed before the data collection.

Recently, the data collection system, including the crystal-mounting robot, has improved tremendously for high-throughput and high performance (Karain *et al.*, 2002). Especially, the proper handling of the micro-crystals was achieved by micro-crystal diffractometers, allowing for diffraction from localized areas of twinned crystals. After the beam lines are automated, the primary data could be obtained within one to several hours. By a combination of advanced instrumentation and integration software for data collection and analysis, the structure determination step is nearly automated from crystal mounting to structure refinement (Abola *et al.*, 2000; Terwilliger, 2003), yielding a preliminary structure within

hours after crystallization of the protein. Therefore, the protein structure could be solved within a day for a good quality of crystal because the data analysis pipelines became fully automated at high-level control.

Homology Modeling and Electron Microscopy in Structural Proteomics

Homology modeling, which builds a three-dimensional structure of protein with unknown functions from template structures, has been considered as an important tool to predict information about the structure and function of a large number of unannotated gene products (Goldsmith-Fischman and Honig, 2003). A recent development in modeling methods enables one to generate a reasonable protein structure from a template structure with about a 30% sequence identity (Sanchez *et al.*, 2000). By combining the biochemical data, the cutoff range of the sequence identity could be even lower (Jung *et al.*, 2000). After the pilot projects of structural genomics initiatives provide a large number of new structures as well as new folds, the structural database could be directly used in generating the structure-function relationship of non-homologous protein with unknown functions by structural bioinformatics tool.

About 30% of the proteins that are embedded in the membrane are found insoluble. These structures are barely determined by NMR and X-ray crystallography. Electron microscopy uses images of single molecules of 2D crystals for protein structure determination (Henderson *et al.,* 1990; Nogales *et al*., 1998). This works well for small amount of insoluble membrane proteins with less purity than that of X-ray crystallography. Recently, the single-particle EM technique was successfully applied to determine the structure of large complexes of *Drosophila melanogaster* tripeptidyl peptidase II from 3D construction of asymmetric assembly of the proteins (Rockel *et al.*, 2002). However, even though electron microscopy is powerful for determining the structure of large biomolecular complexes, the resolving power is relatively low (~1 nm). In addition, since the procedure to obtain the structure still remains slow, it needs technical development for high-throughput structure determination (Zhang *et al.*, 2001).

Concluding Remarks

In the post-structural genomics era, a functional analysis of the unknown gene products would be one of most important issues in understanding the basic mechanism of living cells. The structural proteomics is a new research area to provide us a complete understanding of organisms based on structural information of a single protein or the assembly of protein complexes. From a practical viewpoint, the structural information derived from structural genomics will be directly used for drug development in the pharmaceutical industry. Especially, NMR would play an important role because it is a very sensitive method to detect the molecular interaction between the protein and other associated molecules. Therefore, NMR could be used for characterizing the molecular interaction between proteins and the protein-drug in order to gain insight about molecular functions. However, although the advantages of the individual methodology (NMR spectroscopy, X-ray crystallography, bioinformatics and electron microscopy) are very useful in structural proteomics research, a combination or integration of these methods would be mandatory to fill up a complete dictionary of protein folding space on a genomic scale. In conclusion, structural proteomics will give us a comprehensive picture of structural mechanism to govern basic cellular and biochemical functions of proteins in a living cell.

Acknowledgments This study was supported by the Ministry of Science and Technology of the Korea/the Korea Science and Engineering Foundation through the NRL program of MOST NRDP (M1-0203-00-0020) and in part by the Brain Korea 21 Project.

References

Abola, E., Kuhn, P., Earnest, T. and Stevens, R. C. (2000) Automation of X-ray crystallography. *Nat. Struct. Biol.* **7**, Suppl, 973-977.

Boggon, T. J., Shan, W. S., Santagata, S., Myers, S. C. and Shapiro, L. (1999) Implication of tubby proteins as transcription factors by structure-based functional analysis. *Science* **286**, 2119-2125.

Christendat, D., Yee, A., Dharamsi, A., Kluger, Y., Savchenko1, A., Cort, J. R., Booth, V., Mackereth, C. D., Saridakis, V., Ekiel, I. *et al.* (2000) Structural proteomics of an archaeon. *Nat. Struc. Biol.* **7**, 903-909.

Fernandez, C., Adeishvili, K. and Wüthrich, K. (2001) Transverse relaxation-optimized NMR spectroscopy with the outer membrane protein OmpX in dihexanoyl phosphatidylcholine micelles *Proc. Natl. Acad. Sci. USA* **98**, 2358-2363.

Fowler, C. A., Tian, F., Al-Hashimi, H. M. and Prestegard, J. H. (2000) Rapid determination of protein folds using residual dipolar couplings*. J. Mol. Biol.* **304**, 447-460.

Garrett, D. S., Seok, Y. J., Peterkofsky, A., Gronenborn, A. M. and Clore, G. M. (1999) Solution structure of the 40,000 Mr phosphoryl transfer complex between the N-terminal domain of enzyme I and HPr. *Nat. Struct. Biol*. **6**, 166-173.

Goldsmith-Fischman, S. and Honig, B. (2003) Structural genomics: computational methods for structure analysis. *Protein Sci.* **12**, 1813-1821.

Goto, N. K. and Kay, L. E. (2000) New development in isotope labeling strategies for protein solution NMR spectroscopy. *Curr. Opin. Struct. Biol.* **10**, 585-592.

Greenfield, N. J., Huang, Y. J., Palm, T., Swapna, G. V., Monleon, D., Montelione, G. T. and Hitchcock-DeGregori, S. E. (2001) Solution NMR structure and folding dynamics of the N terminus of a rat non-muscle alpha-tropomyosin in an

engineered chimeric protein. *J. Mol. Biol.* **312**, 833-847.

Güntert, P., Salzmann, M., Braun, D. and Wüthrich, K. (2000) Sequence-specific NMR assignment of proteins by global fragment mapping with the program MAPPER. *J. Biomol. NMR* **18**, 129-137.

Henderson, R. (1995) The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unconstrained biological molecules. *Q. Rev. Biophys*. 28, 171-193.

Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckmann, E. and Downing, K. H. (1990) Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. Mol. Biol.* **213**, 899-929.

Herrmann, T., Güntert, P. and Wüthrich, K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol*. **319**, 209-227.

Huang, L., Hung, L., Odell, M., Yokota, H., Kim, R. and Kim, S-. H. (2002) Structure-based experimental confirmation of biochemical function to a methyltransferase, MJ0882, from hyperthermophile *Methanococcus jannaschii*. *J. Struct. Funct. Genomics* **2**, 121-127.

Hui, R. and Edwards, A. (2003) High-throughput protein crystallization *J. Struct. Biol.* **142**, 154-161.

Hwang, K. Y., Chung, J. H., Kim, S. H., Han, Y. S. and Cho, Y. (1999) Structure-based identification of a novel NTPase from *Methanococcus jannaschii*. *Nat. Struct. Biol*. **6**, 691-696.

Jung, J. -W., An, J. -H., Na, K. -B., Kim, Y. S. and Lee, W. (2000) Active site and substrate binding mode of malonyl-coa synthetase determined by trandsfered nuclear overhauser effect spectroscopy, site directed mutagenesis and comparative modeling studies. *Protein Sci.* **9**, 1294-1303.

Karain, W. I., Bourenkov, G. P., Blume, H. and Bartunik, H. D. (2002) Automated mounting, centering and screening of crystals for high-throughput protein crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **58**, 1519-1522.

Kigawa, T., Muto, Y. and Yokoyama, S. (1995) Cell-free synthesis and amino acid-selective stable isotope labeling of proteins for NMR analysis. *J. Biomol. NMR* **6**, 129-134.

Kigawa, T., Yabuki, T., Yoshida, Y., Tsutsui, M., Ito, Y., Shibata, T. and Yokoyama, S. (1999) Cell-free production and stable-isotope labeling of milligram-quantities of proteins. *FEBS Lett.* **442**, 15-19.

Kim, S.-H. (1998) Shining light on structural genomics. *Nat. Struc. Biol.* **5**, 643-645.

Lee, C. -H., Jung, J. -W., Yee, A., Arrowsmith, C. H. and Lee, W. (2004) Solution Structure of a novel calcium binding protein, mth1880 from *methanobacterium thermoautotrophicum. Protein Sci.* in press.

Medek, A., Olejniczak, E. T., Meadows, R. P. and Fesik, S. W. (2000) An approach for high-throughput structure determination of proteins by NMR spectroscopy. *J. Biomol. NMR* **8**, 229-238.

Montelione, G. T., Zheng, D., Huang, Y. J., Gunsalus, K. C. and Szyperski, T. (2000) Protein NMR spectroscopy in structural genomics. *Nat. Struct. Biol*. **7** Supl, 982-985.

Moseley, H. N. B. and Montelione, G. T. (1999) Automated analysis of NMR assignments and structures for proteins. *Curr. Opin. Struct. biol*. **9**, 635-642.

Mumenthaler, C., Güntert, P., Braun, W. and Wüthrich, K. (1997) Automated combined assignment of NOESY spectra and three-dimensional protein structure determination. *J. Biomol. NMR* **10**, 351-362.

Nilges, M., Macias, M. C., ODonoghue, S. I. and Oschkinat, H. (1997) Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from alpha-spectrin. *J. Mol. Biol.* **269**, 408-422.

Nogales, E., Wolf, S. G. and Downing, K. H. (1998) Structure of the α,β tubulin dimer by electron crystallography. *Nature* **391**, 199-203.

O'Connell, J. F., Pryor, K. D., Grant, S. K. and Leiting, B. (1999) A high quality nuclear magnetic resonance solution structure of peptide deformylase from *Escherichia coli*: application of an automated assignment strategy using GARANT. *J. Biomol. NMR* **13**, 311-324.

Pervushin. K., Riek, R., Wider., G. and Wüthrich, K. (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an venue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. USA* **94**, 12366-12371.

Rockel, B., Peters, J., Kuhlmorgen, B., Glaeser, R. M. and Baumeister, W. (2002) A giant protease with a twist: the TPP II complex from Drosophila studied by electron microscopy. *EMBO J.* **21**, 5979-5984.

Rosen, M. K., Gardner, K. H., Willis, R. C., Parris, W. E., Pawson, T. and Kay, L. E. (1996) Selective methyl group protonation of perdeuterated proteins. *J. Mol. Biol.* **263**, 627-636.

Rupp, B. (2003) High-throughput crystallography at an affordable cost: The TB structural genomics consortium crystallization facility, *Acc. Chem. Res.* **36**, 173-181

Sali, A., Glaeser, R., Earnest, T. and Baumeister, W. (2003) From words to literature in structural proteomics *Nature* **422**, 216-225.

Sanchez, R., Pieper, U., Melo, F., Eswar, N., Marti-Renom, M. A., Madhusudhan, M. S., Mirkovic, N. and Sali, A. (2000) Protein structure modeling for structural genomics. *Nat. Struct. Biol.* **7** Suppl, 986-990.

Schmid, M. B. (2002) Structural proteomics: the potential of high-throughput structure determination. *Trends Microbiol*. **10**, S27-31.

Service, R. F. (1998) NMR researchers look to the next generation of machines. *Science* **279**, 1127-1128.

Staunton, D., Owen, J. and Campbell, I. D. (2003) NMR and structural genomics. *Acc. Chem. Res*. **36**, 207-214.

Terwilliger, T. C. (2003) Automated structure solution, density modification and model building. *Acta Crystallogr. D. Biol. Crystallogr.* **58**, 1937-1940.

Tjandra, N., Omichinski, J. G., Gronenborn, A. M., Clore, G. M. and Bax, A. (1997) Use of dipolar 1H-15N and 1H-13C couplings in the structure determination of magnetically oriented macromolecules in solution. *Nat. Struct. Biol.* **4**, 732-738.

Wagner, G. (1993) Prospects for NMR of large proteins. *J. Biomol. NMR* **3**, 375-385.

Yamazaki, T., Otomo, T., Oda, N., Kyogoku, Y., Uegaki, K., Ito, N., Ishino, Y. and Nakamura, H. (1998) Segmental isotope labeling for protein NMR using peptide splicing. *J. Am. Chem. Soc.* **120**, 5591-5592.

Yee, A., Chang, X., Pineda-Lucena, A., Wu, B., Semesi, A., Le B, Ramelot, T., Lee, G. M., Bhattacharyya, S., Gutierrez, P. *et al.* (2002) An NMR approach to structural proteomics *Proc. Natl. Acad. Sci. USA* **99**, 1825-1830.

Yee, A., Pardee, K., Christendat, D., Savchenko, A., Edwards, A. M. and Arrowsmith, C. H. (2003) Structural proteomics: toward high-throughput structural biology as a tool in functional genomics. *Acc. Chem. Res.* **36**, 183-189.

Zarembinski, T. I., Hung, L. W., Mueller-Dieckmann, H. J., Kim, K. K., Yokota, H., Kim, R. and Kim, S. H. (1998) Structure-based assignment of the biochemical function of a hypothetical protein: a test case of structural genomics. *Proc. Natl. Acad. Sci. USA* **95**, 15189-15193.

Zhang, P., Beatty, A., Milne, J. L. and Subramaniam, S. (2001) Automated data collection with a Tecnai 12 electron microscope: applications for molecular imaging by cryomicroscopy. *J. Struct. Biol*. **135**, 251-261.