J Biomol NMR (2008) 42:155–158
DOI 10.1007/s10858-008-9277-8

PERSPECTIVE

Solution NMR structure determination of proteins revisited

Martin Billeter · Gerhard Wagner · Kurt Wüthrich

Received: 11 September 2008 / Accepted: 12 September 2008 / Published online: 1 October 2008

Abstract This 'Perspective' bears on the present state of protein structure determination by NMR in solution. The focus is on a comparison of the infrastructure available for NMR structure determination when compared to protein crystal structure determination by X-ray diffraction. The main conclusion emerges that the unique potential of NMR to generate high resolution data also on dynamics, interactions and conformational equilibria has contributed to a lack of standard procedures for structure determination which would be readily amenable to improved efficiency by automation. To spark renewed discussion on the topic of NMR structure determination of proteins, procedural steps with high potential for improvement are identified.

Keywords NMR structure determination of proteins · Protein biochemistry · Automation · Structural biology · Structural genomics

M. Billeter (✉)
Department of Chemistry, University of Gothenburg, Box 462, 40530 Gothenburg, Sweden
e-mail: martin.billeter@chem.gu.se

G. Wagner
Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02467, USA

K. Wüthrich
Department of Molecular Biology and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA, USA

K. Wüthrich
Institute of Molecular Biology and Biophysics, ETH Zürich, CH-8093 Zurich, Switzerland

During the past decades, NMR spectroscopy of biological macromolecules in solution has demonstrated its potential with a wide range of applications for studies of structure and function (Wüthrich 2003). Solution NMR has also firmly established its role for de novo structure determination of protein-, RNA- and DNA-constructs with molecular weights up to about 30 kDa. However, although the Protein Data Bank (PDB; Berman et al. 2000) contains over 6000 NMR structures, no standard procedures have evolved for NMR structure determination of proteins, and different laboratories use a variety of different approaches. Quite generally, this presents barriers for efficient NMR structure determination, and start-up NMR groups or biochemists and biologists with structural biology projects cannot readily find guidance for the use of NMR techniques without collaboration with "experts". In spite of this rather special situation, the NMR field has flourished and generated a wealth of important new insights in the molecular biological sciences, which extend the characterization of proteins from atomic resolution three-dimensional structure models, obtained by NMR or by X-ray diffraction, to dynamic and binding properties. Because the NMR-derived information is often unique and could not be obtained by other available methods, high efficiency of NMR data collection is usually not a priority in "hypothesis-driven" projects. An obvious illustration is the situation where a protein cannot be crystallized, and NMR is then the only technique capable of yielding a three-dimensional structure at atomic resolution.

Starting in the year 2000, the advent of "structural genomics" (SG) resulted in a new challenge, since high-throughput protein structure determination is an important element in the attempt to reach the stated Protein Structure Initiative (PSI) goal of a representative sampling of protein fold space (Chandonia and Brenner 2006; Levitt 2007;

Stuart et al. 2006). In recent reviews, the response of X-ray crystallography to this challenge has been lauded as being highly successful, with technical advances that now more and more also benefit discovery- and hypothesis-driven research in the entire crystallography community. In contrast, the response of NMR-based structure determination to the SG-challenge has been judged to be less impressive.[1]

In part this is due to the fact that NMR structure determination of proteins has to a lesser degree been supported by robotics and other automation than protein crystal structure determination. In an upcoming 'Perspective' in this Journal, Michael Williamson and Jeremy Craven will investigate the present state and likely future developments of the automation of NMR structure determination.

The standard data processing in NMR structure determination of proteins consists of a sequence of successive steps, starting with Fourier transformation of the time-domain data to obtain the frequency-domain spectra, identifying the NMR signals in the frequency-domain spectra, assigning the NMR signals to individual atoms in the polypeptide chain, collecting three-dimensional structure information in the form of restraints on interatomic distances, torsion angles about chemical bonds and the relative directions of chemical bonds linking different atom pairs of the polypeptide chain, and using the experimental data thus obtained as input for the structure calculation. Automation of this procedure will foreseeably include going through part or all of these steps in multiple cycles, in order to enable feedback mechanisms such as the use of preliminary protein structures obtained in early cycles to supplement the initial resonance assignments and the initial collection of conformational constraints.

An overview of the published NMR structure determinations of proteins reveals that nearly all of these results were obtained using computer-supported interactive procedures rather than automation of at least some of the steps involved. The limited extent of automation is primarily due to the fact that NMR spectroscopy with biological macromolecules is typically at the limit of a workable signal-to-noise ratio. As a consequence, automated distinction between NMR signals, artifacts and noise peaks in the experimental data sets is a challenge, as will be discussed in more detail in the 'Perspective' by Williamson and Craven. Here we want to focus on factors other than automation of the NMR data handling that may affect protein NMR structure determination in a "high-throughput environment". Considering the impressive achievements of protein crystallography (Chandonia and Brenner 2006; Levitt 2007), we start with a glance at the crystal structure determination approach.

For the purpose of this presentation, we divide protein crystallography into two sub-projects, which are in typical SG-consortia performed by different groups of scientists. The first sub-project covers all the steps involved in obtaining diffracting protein crystals for structure determination, including the annotation of complete genomes, the selection of target proteins, the cloning of multiple constructs for a given target, the expression, purification and physico-chemical characterization of the protein preparations, crystallization trials with a wide range of solution conditions, and finally the screening of the resulting crystals for high resolution diffraction. The role of automation and robotics in this subproject of protein crystallography and the implications for NMR structure determination of proteins are the subject of the remainder of the present text. The second subproject is the actual crystal structure determination, starting from crystals that yield high-resolution diffraction for the native protein and, in as far as needed, suitably derivatized forms of the protein. Software for automated processing of the diffraction data for routine protein structure determinations is highly developed, and once a structure has been automatically "solved", standardized procedures are available for interactive structure refinement, which also includes quality control mechanisms for checking on the intermediate and final results.

Important factors for the success of protein crystal structure determination in "high-throughput" environments include: (i) As the dominant player in protein structure determination, crystallography has succeeded to attract massive investments. This enabled the installation of efficient, method-focused target selection, heavily robot-supported protein preparation laboratories that ensure a steady protein supply, and ample access to high-intensity X-ray beamlines. (ii) The foundations of the target selection by informatics specialists include a wealth of statistical data on both the crystal structures deposited in the PDB and on structure determination attempts that failed when using the standard protocols of individual PSI centers (for example, Slabinski et al. 2007a, b). This significantly contributes to the probability that the chosen targets are indeed amenable to X-ray structure determination. (iii) The protein production pipelines are designed to generate extensive optimization of protein constructs, and they routinely screen the protein preparations for their suitability for crystallization trials, using chemical and

[1] Quotation from the "Report of the Protein Structure Initiative Assessment Panel" (http://www.nigms.nih.gov/News/Reports/PSIAssessmentPanel2007.htm). "*A.3.2C NMR*. NMR technology from the PSI is not seen as having the general impact described above for crystallography. Many of the NMR advances (G-matrix Fourier-transform NMR, automated data analysis and structure determination) are not readily applicable to complex systems with higher molecular weight. Thus, these have not changed the practices of the NMR community in the way PSI-supported technologies have changed the crystallographic community."

physico-chemical techniques. (iv) Highly developed dedicated robotics are available for exhaustive automated crystallization trials, using nanoliter volumes of protein solution for each assay. (v) The screening of thousands of crystals for their suitability for structure determination is fully automated in dedicated synchrotron beamlines. (vi) Access to synchrotrons is streamlined and user-friendly, and ever smaller crystals can be used for structure determination. (vii) Quality assessment is well-established at all stages of the high-throughput process.

When turning to NMR, one notes in most "high-throughput" environments that, as a consequence of the impressive performance of X-ray crystallography, a trend has evolved to consider NMR an option for structure determination only when repeated crystallization attempts have failed. With this philosophy, unique information is again requested from NMR, and high efficiency is not of overriding importance. Given this situation, it is not surprising that, in contrast to X-ray crystallography (Slabinski et al. 2007a, b), there are no commonly known and accepted criteria for the use of bioinformatics tools to evaluate target protein sequences for the probability of success in NMR structure determination. Regarding protein production, the historical development of protein X-ray crystallography confirms in impressive ways that a highly productive protein expression laboratory is the key to efficient protein structural biology. In the NMR field, efficient protein production should ensure that optimization of protein constructs, as well as solution conditions, is not a bottleneck for NMR structure determination (Acton et al. 2005; Yee et al. 2002). Finally, there is a need for a quantifiable quality assessment of the protein solutions used for NMR structure determination, which would be the equivalent to protein crystal screening for high-resolution diffraction. So far, no generally accepted systematic procedure for such evaluation of protein solutions is in place, in spite of early work on this key step in NMR structure determination (Bagby et al. 1997, 2001). The absence of generally accepted, streamlined procedures extends further from the sample preparation to the actual NMR methodology, where the use of a wide array of different combinations of experiments and different pulse sequences (Cavanagh et al. 2007) makes it difficult to compare achievements in different laboratories. Overall, notwithstanding the structural genomics centers that had from the start a special focus on solution NMR (Liu et al. 2005; Pan et al. 2007; Yee et al. 2002; Yokoyama et al. 2002), success stories about biomolecular NMR structure determinations include as many accounts of heroics by graduate students and postdoctoral fellows combined with well-deserved good luck, as achievements based on rational sample optimization.

The limited role of NMR structure determination in high-throughput environments has well-recognizable fundamental roots. First, as mentioned in the introduction to the present text, de novo protein structure determination is only one of many uses of solution NMR spectroscopy in structural and functional biology, and there is an upper size limit of about 25 kDa for efficient use of the method. NMR investigation of molecular dynamics, conformational equilibria, ligand binding, characterization of "unstructured functional polypeptides" and their structural transitions in active complexes are just a few examples of highly attractive, alternative NMR applications. Second, a wide arsenal of different NMR experiments is available for different applications (Cavanagh et al. 2007). The outstanding flexibility of the method is then quite naturally reflected in variations of the protein structure determination protocols used in different laboratories, and also in extensive investment of resources for intellectually stimulating attempts to use minimal sets of NMR data for "protein structure predictions". Third, few NMR spectroscopists have been under pressure to pursue protein structure determination in a high-throughput fashion.

In view of all that has been mentioned so far, one might question the wisdom of working toward protocols for "high-throughput" NMR determination of high-resolution protein structures in solution. With this "Perspective", we want to strongly advocate to resume and intensify discussions of this intellectual and organizational challenge (Bhattacharya et al 2007; Yee et al. 2002; Yokoyama et al. 2002). Major progress toward NMR structure determination in a high-throughput context will depend on new methods development, which will at the very least lead to higher-output technology and thus benefit the entire biomolecular NMR community. X-ray crystallography has made commendable use of this opportunity and, as a consequence, the entire field of crystal structure determination has evolved in important ways during the last 5 years. NMR runs the risk of being left out of this ongoing evolution of structural biology. If an effort is made, then the standards for NMR structure determination will have to be in line with the high standards established by the use of X-ray crystallography in structural genomics (Bhattacharya et al. 2007; Brown and Ramaswamy 2007). In addition to progress required for the actual structure determination protocol (see the forthcoming 'Perspective' by Williamson and Craven), optimization with robotics and automation of the procedures used to select, prepare and evaluate the protein constructs for structure determination is needed also in the NMR field.

Acknowledgements We thank Dr. Ian A. Wilson for a critical reading of the manuscript. M.B. acknowledges support by the EU (LSHG-CT-2005-018988), G.W. by the NIH grant GM 47467, and

K.W. by the Joint Center for Structural Genomics in La Jolla, CA, USA (JCSG, NIH #GM074898) and the NCCR Structural Biology (Swiss National Science Foundation and ETH Zürich, Switzerland).

References

Acton TB, Gunsalus KC, Xiao R, Ma LC, Aramini J, Baran MC, Chiang Y-W, Climent T, Cooper B, Denissova NG, Douglas SM, Everett JK, Ho CK, Macapagal D, Rajan PK, Shastry R, Shih L-Y, Swapna GV, Wilson M, Wu M, Gerstein M, Inouye M, Hunt JF, Montelione GT (2005) Robotic cloning and protein production platform of the northeast structural genomics consortium. Methods Enzymol 394:210–243

Bagby S, Tong KI, Liu D, Alattia JR, Ikura M (1997) The button test: a small scale method using microdialysis cells for assessing protein solubility at concentrations suitable for NMR. J Biomol NMR 10:279–282

Bagby S, Tong KI, Ikura M (2001) Optimization of protein solubility and stability for protein nuclear magnetic resonance. Methods Enzymol 339:20–41

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN (2000) The protein data bank. Nucleic Acids Res 28:235–242

Bhattacharya A, Tejero R, Montelione G (2007) Evaluating protein structures determined by structural genomics consortia. Proteins 66:778–795

Brown EN, Ramaswamy S (2007) Quality of protein crystal structures. Acta Crystallogr D 63:941–950

Cavanagh J, Fairbrother WJ, Palmer AG, Rance M, Skelton NJ (2007) Protein NMR spectroscopy. Academic Press, New York

Chandonia J-M, Brenner SE (2006) The impact of structural genomics: expectations and outcomes. Science 311:347–351

Levitt M (2007) Growth of novel protein structural data. Proc Natl Acad Sci USA 104:3183–3188

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 (2005) NMR data collection and analysis protocol for high-throughput protein structure determination. Proc Natl Acad Sci USA 102:10487–10492

Pan X, Wesenberg GE, Markley JL, Fox BG, Phillips GN Jr, Bingman CA (2007) A graphical approach to tracking and reporting target status in structural genomics. J Struct Funct Genomics 8:209–216

Slabinski L, Jaroszewski L, Rychlewski L, Wilson IA, Lesley SA, Godzik A (2007a) XtalPred: a web server for prediction of protein crystallizability. Bioinformatics 23:3403–3405

Slabinski L, Jaroszewski L, Rodrigues APC, Rychlewski L, Wilson IA, Lesley SA, Godzik A (2007b) The challenge of protein structure determination—lessons from structural genomics. Protein Sci 16:2472–2482

Stuart DI, Jones EY, Wilson KS, Daenke S (2006) Structural proteomics in Europe—the best of both worlds. Acta Crystallogr D. doi:10.1107/S0907444906035347

Wüthrich K (2003) NMR studies of structure and function of biological macromolecules. J Biomol NMR 27:13–39

Yee A, Chang X, Pineda-Lucena A, Wu B, Semsei A, Le B, Ramelot T, Lee GM, Bhattacharyya S, Gutierrez P, Denisov A, Lee C-H, Cort JR, Kozlov G, Liao J, Finak G, Chen L, Wishart D, Lee W, Mcintosh LP, Gerhring K, Kennedy MA, Edwards AM, Arrowsmith CH (2002) An NMR approach to structural proteomics. Proc Natl Acad Sci USA 99:1825–1830

Yokoyama S, Hirota H, Kigawa T, Yabuki T, Shirouzu M, Terada T, Ito Y, Matsuo Y, Kuroda Y, Nishimura Y, Kyogoku Y, Miki K, Masui R, Kuramitsu S (2002) Structural genomics projects in Japan. Nat Struct Biol 7:943–945