



## REVIEW

# Recent Advances in Computational Methods for Nuclear Magnetic Resonance Data Processing

Xin Gao \*

Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia

Received 22 October 2012; revised 12 December 2012; accepted 28 December 2012

Available online 11 January 2013

## KEYWORDS

Nuclear magnetic resonance;  
Protein structure;  
Computational methods;  
Bioinformatics

**Abstract** Although three-dimensional protein structure determination using nuclear magnetic resonance (NMR) spectroscopy is a computationally costly and tedious process that would benefit from advanced computational techniques, it has not garnered much research attention from specialists in bioinformatics and computational biology. In this paper, we review recent advances in computational methods for NMR protein structure determination. We summarize the advantages of and bottlenecks in the existing methods and outline some open problems in the field. We also discuss current trends in NMR technology development and suggest directions for research on future computational methods for NMR.

## Introduction

Nuclear magnetic resonance (NMR) spectroscopy is one of the main methods for determining three-dimensional (3D) structures of proteins [1]. The underlying idea for NMR protein structure determination is that if a large number of distance constraints are known between atom pairs of a target protein, the conformational space of possible protein structures will be restricted to a few structures [2]. The physical principle of NMR structure determination is that when a certain isotope (e.g.,  $^1\text{H}$ ,  $^{13}\text{C}$  or  $^{15}\text{N}$ ) is placed in a strong magnetic field, the nucleus will absorb electromagnetic radiation at a frequency that is characteristic of the isotope. Depending on different

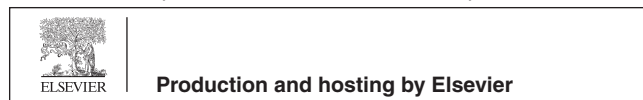
local chemical and geometric environments, different nuclei resonate at different frequencies. Since frequency is a magnetic field-dependent measure, it is often converted into a relative frequency with respect to a reference frequency. Such relative frequencies are referred to as chemical shifts. The resonances of nuclei that are close in Euclidean space couple, either through covalent bonds or through space. NMR experiments capture such coupling.

The outputs from NMR experiments are NMR spectra, which are, mathematically speaking, multi-dimensional matrices. The indices for each dimension are the discrete chemical shift values of a certain nucleus, and the entries of the matrices are the intensity values of the coupling. For instance,  $^{15}\text{N}$ -HSQC is one of the most commonly-used NMR spectra. It captures the coupling between the backbone nitrogen (N) and the hydrogen (H) that is attached to this nitrogen. For a protein with  $n$  amino acids, there are  $(n-p)$  expected peaks in the  $^{15}\text{N}$ -HSQC spectrum, where  $p$  is the number of proline (Pro) in the protein. However, the amine groups in the side chains of some amino acids are also visible in the  $^{15}\text{N}$ -HSQC spectrum, such as arginine (Arg), asparagine (Asn) and glutamine (Gln). To eliminate the peaks of these side chains,

\* Corresponding author.

E-mail: [xin.gao@kaust.edu.sa](mailto:xin.gao@kaust.edu.sa) (Gao X).

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.



information from different spectra needs to be combined. There are additional sources of error in NMR spectra, including missing signals, chemical shift degeneracy, sample impurity, water bands, artifacts and experimental errors [2]. All of these sources of error need to be taken into account.

Another important NMR spectrum is the nuclear Overhauser enhancement (NOE) spectrum, which is a through-space experiment that captures certain atoms that are close to each other in the Euclidean space. Here, ‘close’ often refers to a distance smaller than 6 Å. Thus, the NOE spectrum is a through-space spectrum and each peak in the NOE spectrum provides a distance constraint that can reduce the conformational space of possible protein structures.

In contrast to NOE that provides short-range interactions ( $<6$  Å), there are experiments that can provide long-range information. One example is residual dipolar couplings (RDCs), which provides long-range orientational information relative to an external alignment tensor [3–5]. Another example is paramagnetic relaxation enhancement (PRE) [6,7]. PRE effect can be detected in large magnetic moment of protons and unpaired electron up to 35 Å.

Traditionally, determination of NMR protein structure mainly follows the four-step process described by Wüthrich [1]. After the spectra are collected, the four steps involve peak picking, resonance assignment, NOE assignment and structure calculation. The peak picking step takes the through-bond and through-space NMR spectra as inputs and identifies peaks in these spectra. The peaks of certain through-bond spectra are then used to assign the chemical shift values to the corresponding atoms of the protein, which is the so-called resonance assignment step. After resonance assignment, mapping between the chemical shift values and the indices of the atoms is built. Such mapping is applied to interpret the NOE peaks and extract distance constraints. Since the chemical shift values of all the atoms of the protein are distributed within a small range, overlaps in chemical shift values are expected. Thus, the interpretation of the NOE peaks can be ambiguous. The structure calculation step takes the distance constraints (both ambiguous and unambiguous) to determine the final structure(s) of the protein.

Most NMR labs process NMR data either manually or semi-automatically with the help of visualization tools. The entire process is computationally costly and time-consuming. Recently, attention has been paid to developing computational methods that can significantly accelerate the NMR data processing and reduce the errors introduced by manual processing. However, NMR is still a new field to the computational community. Even in the field of bioinformatics and computational biology, computational problems in NMR structure determination have not been well studied. Here, we review some recent advances in computational methods for NMR protein structure determination.

## Peak picking

The goal of the peak picking step is to identify peaks, *i.e.*, the chemical shift coordinates of the coupling nuclei, in any given spectrum. This is the key step in the entire NMR protein structure determination process because the following steps are all built upon this step [8,9]. The automated peak picking problem was first studied two decades ago [10]. Expected properties of peak shapes, such as the symmetry property, were used to

identify peaks. Since then, a variety of computational methods have been utilized, including peak-property-based methods [11,12], machine learning methods [13–16], and spectra-decomposition-based methods [17–19].

Recently, image processing techniques have been applied to the peak picking problem and they have demonstrated promising performance [20,21]. Alipanahi et al. proposed a multi-stage method, PICKY, to automatically identify peaks from a given set of N–H-rooted NMR spectra [20]. PICKY considers an NMR spectrum as an image and estimates the noise level by estimating the variance in local neighborhoods, which is based on the assumption that the noise is white Gaussian noise. All the ‘pixels’ of the image, *i.e.*, data points of the spectrum, that have intensity values lower than the estimated noise level are believed to contain no signal and are thus removed. The disconnected components of the remaining spectrum are identified, some of which may contain a number of peaks due to peak overlapping or inaccuracy in the estimation of the noise level. The components are further decomposed to smaller ones by checking the levels of overlapping of adjacent local maxima. Rank-one singular value decomposition (SVD) is applied to each small component to identify peaks, which can eliminate false local maxima in the component. Finally, cross-referenced information between spectra that share common nuclei, such as  $^{15}\text{N}$  and  $^1\text{H}$ , is used to refine the peak lists. Another contribution of [20] is to propose a benchmark set that contains 32 2D and 3D spectra extracted from eight proteins. This is the most comprehensive data set to date for the peak picking problem.

Although PICKY demonstrated significantly better performance than previous peak picking methods, it has two bottlenecks. PICKY is not sensitive enough to replace manual peak picking in the sense that weak peaks may be eliminated in the denoising step of PICKY if they have intensity values lower than the estimated noise level. On the other hand, the number of false positives is high in PICKY peak lists due to the fact that PICKY ranks peaks by intensity values, which can be badly biased. WaVPeak was developed to overcome these two bottlenecks [21]. Like PICKY, WaVPeak is also based on image processing techniques. Specifically, WaVPeak uses wavelets. Wavelets are mathematical functions that cut data into different frequency components. Each component is then studied with a resolution matched to its scale. WaVPeak applies multi-dimensional wavelets to the NMR spectra to smooth the spectra. In contrast to PICKY, WaVPeak aims to eliminate noise from the data points instead of eliminating noisy data points. This can preserve the shapes of the peaks, including the weak ones. Furthermore, WaVPeak ranks the peaks by their estimated volumes. On PICKY’s benchmark set, WaVPeak showed significantly higher sensitivity and included a smaller number of false positives than did PICKY. To be more specific, WaVPeak achieved an average of 88% recall value and 74% precision value.

One remaining problem in automatic peak picking is how to select true peaks from a large number of predicted peaks [9]. If a set of spectra is available for a target protein, the peak lists for these spectra can be used as cross-checks for each other [20,22]. For instance, the chemical shifts of  $^{15}\text{N}$  and  $^1\text{H}$  in a true peak in a CBCA(CO)NH spectrum are expected to be visible in the  $^{15}\text{N}$ -HSQC spectrum of the same protein and they can be cross-checked. It is also possible to select the true peaks of a single spectrum. To do so, Abbas et al. cast the peak selection problem as a multiple testing problem in statistics [22]. They first converted the peak ranking criterion,

such as intensity or volume, into a  $P$ -value. They then applied a Benjamini–Hochberg algorithm to control the false discovery rate (FDR) and select the true peaks. Their method can be potentially applied to different bioinformatic problems in which true predictions must be differentiated from a large number of false ones, such as protein function annotation [23]. However, the Benjamini–Hochberg algorithm only selects a ‘cutting point’ in the ranked peak list. Its performance therefore depends on the quality of the ranking criteria. Designing a ranking measure that is better than volume or symmetry still remains an open problem in peak picking.

## Resonance assignment

After the peaks are identified, the peak lists from the through-bond spectra are first combined to assign the chemical shift values to the corresponding atoms of the protein. For resonance assignment, the peaks that share common nuclei,  $^{15}\text{N}$  and  $^1\text{H}$ , are first grouped into spin systems. The spin systems are then assigned to the residues of the protein using both inter-residue and intra-residue information contained in the spin systems. Ideally, there are  $n$  spin systems to be assigned to  $n$  residues. However, due to incomplete peak picking, there are often missing spin systems, missing chemical shifts in spin systems and false spin systems, which make the resonance assignment problem practically difficult. A variety of computational methods have been explored to solve the resonance assignment problem, including search algorithms [24–27], maximum independent set algorithms [28], sequential algorithms [29,30], logic algorithms [31], fragment-based algorithms [32,33] and optimization algorithms [34–37].

Many target proteins of NMR experiments have closely homologous structures that are stored in the protein data bank (PDB) [38]. Depending on whether the homologous structures are utilized to assist the assignment process, resonance assignment methods can be classified as either *ab initio* or structure-based assignments. To make an assignment method practically useful, the method has to be error-tolerant because the input peak lists or spin systems could contain missing or false information. Another major difficulty is caused by chemical shift degeneracy, that is, the same nucleus may have slightly different chemical shift values in different spectra. This introduces ambiguities in the assignment process, especially for large proteins and proteins containing residues with similar chemical shift values, such as all- $\alpha$  proteins, which is a class of structural domains in which the secondary structure is composed entirely of  $\alpha$ -helices.

IPASS was developed as an error-tolerant assignment method that automatically takes picked peaks as inputs [34]. IPASS is built based on the optimization techniques. The peaks from different spectra are first grouped into spin systems by a two-round algorithm that can eliminate the effects of chemical shift degeneracy. The spin systems are then evaluated by a probabilistic model to calculate the probability of being assigned to different residues. After that, the problem becomes one of finding the mapping between the spin system set and the residue set. Finding the optimal mapping, however, is NP-hard in the worst case. IPASS formulates the problem as an integer linear programming (ILP) formulation. For most of the cases, the probabilistic model can reduce the search space to a reasonable size in which state-of-the-art ILP solvers can find the optimal solutions. Tycko and Hu, on the other hand,

solved the resonance assignment problem in a completely probabilistic manner [30]. They formulated the assignment problem as a local search problem and developed a Monte Carlo simulated annealing algorithm to explore the assignment search space. In this way, they could handle chemical shift degeneracy and missing/false chemical shifts in spin systems.

When close homology to the target protein can be found in PDB, the problem becomes more tractable. Jang et al. proposed the structure-based assignment problem and developed a general integer linear programming framework to solve the problem [35,36]. Their method simultaneously assigns backbone chemical shifts and interprets NOE peaks. The underlying idea is that given the homologous structure, a contact graph can be built in which each node is a residue and each edge denotes a pair of residues that are closer than 6 Å in Euclidean space. A similar graph can also be built based on spin systems and the NOE peaks that are associated with such spin systems. In this graph, each node is a spin system and each edge represents two spin systems that are associated by an NOE peak. The goal is to find the common edge matching between the two graphs that maximizes the matching scores. Their method was highly accurate, even when automatically picked peaks were used as the inputs.

The performance of all the aforementioned methods, however, largely depends on the accuracy of amino acid typing and secondary structure prediction of spin systems. Probabilistic models have been built based on statistics from the Biological Magnetic Resonance Bank (BMRB) [39], to predict amino acid and secondary structure types of spin systems to reduce the search space [34,35,40]. However, the accuracy of such models remains modest, which leaves room for improvement.

## NOE assignment and structure calculation

NOE assignment and structure calculation are often combined together to calculate final structures [34,41–44]. A widely used method is the CYANA package [43]. CYANA is based on local search techniques, *i.e.*, simulated annealing by molecular dynamics simulations in the torsion angle space. However, CYANA requires manually processed assignments and NOE peaks to accurately determine the final structures. To make the structure calculation more error-tolerant, Gao et al. developed AMR (automated NMR protocol) [2,34]. AMR is an end-to-end computational pipeline that consists of the peak picking module, PICKY, the resonance assignment module, IPASS, and the NOE assignment and structure calculation module, FALCON-NMR [45]. Given a target protein and its resonance assignment, FALCON-NMR first searches for homologs of the protein in PDB. If homologs are found, it refines the structure by encoding chemical shift information. Otherwise, it makes an *ab initio* prediction of the structure of the protein. The chemical shifts are used to search for fragments of the target protein, from which the backbone angle distributions are extracted. An order-nine hidden Markov model (HMM) is built to sample the conformational space. It has been shown recently that little information is worthwhile beyond the residues that are more than nine residues apart [46]. The sampled structures are thus ranked by the ambiguous NOE constraints and the top ones are selected to generate fragments for the next iteration. FALCON-NMR works in an iterative manner until convergence.

The main bottleneck to *ab initio* protein structure calculation methods is that the size of the search space is intractable. Although the aforementioned methods use chemical shift information to significantly reduce the search space, they do not work well on large proteins. Besides, NMR information has mainly been used in the scoring function and the fragment selection parts of such methods. A method that can encode the chemical shift information to direct the search procedure may give better scalability.

### Automated structure determination from spectra

The ultimate goal for all the aforementioned efforts is to greatly accelerate, and even fully automate, the currently time-consuming NMR protein structure determination process, *i.e.*, from the set of NMR spectra to the final 3D structure of the protein. Despite the large number of computational methods developed for different steps of the NMR data processing procedure, a crucial question is that whether the “isolated” methods can be combined into a pipeline to work together. In fact, this is one of the most important questions for the general bioinformatics field. In bioinformatics, a complex problem is often decomposed into smaller ones or consecutive steps. Computational efforts can usually solve the smaller problems relatively well. However, such methods are developed independently of each other and often have different assumptions, inputs and outputs, and error tolerant levels. From a user point of view, it is very difficult to make a correct combination of the methods to solve the big problem.

As mentioned in the previous section, Gao et al. developed a fully automated pipeline, AMR, as a proof-of-concept [2]. PICKY was applied to identify peaks from a set of six spectra, including  $^{15}\text{N}$ -HSQC, HNCOC or HNCA, CBCA(CO)NH, HNCACB, HCONH-TOCSY and N-NOESY [20]. The six peak lists were then used to cross check to remove false positives. The refined peak lists were fed into IPASS for resonance assignment [34]. IPASS was specifically developed to deal with highly noisy and incomplete peak lists generated by automatic peak picking methods. The resonance assignment was then applied to assign NOE peaks. FALCON-NMR was used to calculate the final 3D structure by using both chemical shift information and distance constraints [34]. AMR was applied on the spectrum sets of four proteins and generated final structures within 1.5 Å to the experimentally determined ones. Another successful attempt is FLYA [47,44], which uses AUTOPSY as the peak picking tool [17], GARANT as the chemical shift assignment tool [48], ARIA as the NOE assignment tool [49] and CYANA as the structure calculation tool [43].

### Outlook

Despite of some progress in developing computational methods for NMR data processing, the main bottlenecks to analysis of NMR spectroscopy data remain, *i.e.*, solving structures of large proteins and solving loop structures. If the target protein is a large protein, the number of atoms will be higher and the spectra will become more crowded. On the other hand, if the target protein contains flexible loops, their peaks tend to have weak intensities and sometimes overlap with each other. To overcome these bottlenecks, efforts have been extended in three directions. First, NMR spectrometers with stronger mag-

netic fields, such as 950 MHz, have been developed and utilized in labs. Such machines can generate spectra with much higher resolutions and their peaks are more concentrated. Second, higher-dimensional NMR experiments have been developed and used. Up to now, 6D spectra have been used in practice [50]. Far fewer overlapping peaks are expected in higher-dimensional spectra. Third,  $^{13}\text{C}$ -labeled spectra can be used to replace traditional  $^1\text{H}$ -labeled proteins to reduce the number of peaks significantly and thus reduce ambiguities. Any of these directions will require computational efforts to extend the current methods or develop novel methods to deal with new types of data, especially for the peak picking step and the structure calculation step.

### Conclusion

Here, we have briefly reviewed recent advances in computational methods for NMR protein structure determination, which is a relatively new field of inquiry for bioinformaticians and computational biologists. We have provided a summary of the advantages to and bottlenecks in existing methods and outlined some open questions. We have also discussed current trends in the development of NMR technologies and have pointed out directions for the development of future computational methods.

### Competing interests

None declared.

### Acknowledgements

We are grateful to Ahmed Abbas, Babak Alipanahi, Cheryl Arrowsmith, Vladimir B. Bajic, Frank Balbach, Dongbo Bu, Meghana Chitale, Logan Donaldson, Jianhua Huang, Richard Jang, Bing-Yi Jing, Emre Karakoc, Daisuke Kihara, Xinbing Kong, Ming Li, Shuai Cheng Li, Zhi Liu, Mehdi Maadooliat, Mario Messih and Jinbo Xu for their contributions to the projects discussed in this review. We thank Virginia Unkefer for editorial work on the manuscript. This work was supported by the GRP-CF award (Grant No. GRP-CF-2011-19-P-Gao-Huang) and a GMSV-OCRF award from King Abdullah University of Science and Technology (KAUST).

### References

- [1] Wüthrich K. NMR of proteins and nucleic acids. New York: John Wiley and Sons; 1986.
- [2] Gao X. Towards automating protein structure determination from NMR data. PhD dissertation. University of Waterloo; 2009.
- [3] Tjandra N, Omichinski JG, Gronenborn AM, Clore GM, Bax A. Use of dipolar  $^1\text{H}$ - $^{15}\text{N}$  and  $^1\text{H}$ - $^{13}\text{C}$  couplings in the structure determination of magnetically oriented macromolecules in solution. *Nat Struct Biol* 1997;4:732–8.
- [4] Clore GM. Accurate and rapid docking of protein–protein complexes on the basis of intermolecular nuclear overhauser enhancement data and dipolar couplings by rigid body minimization. *Proc Natl Acad Sci U S A* 2000;97:9021–5.
- [5] Bax A, Kontaxis G, Tjandra N. Dipolar couplings in macromolecular structure determination. *Methods Enzymol* 2001;339:127–74.

- [6] Solomon I. Relaxation processes in a system of two spins. *Phys Rev* 1955;99:559.
- [7] Clore GM, Iwahara J. Theory, practice and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem Rev* 2009;109:4108–39.
- [8] Li M. Can we determine a protein structure quickly? *J Comput Sci Technol* 2010;25:95–106.
- [9] Gao X. Mathematical approaches to the NMR peak-picking problem. *J Appl Comput Math* 2012;1:1.
- [10] Kleywegt GJ, Boelens R, Kaptein R. A versatile approach toward the partially automatic recognition of cross peaks in 2D  $^1\text{H}$  NMR spectra. *J Magn Reson* 1990;135:288–97.
- [11] Garrett DS, Powers R, Gronenborn AM, Clore GM. A common sense approach to peak picking in two-, three-, and four-dimensional spectra using automatic computer analysis of contour diagrams 1991. *J Magn Reson* 2011;213:357–63.
- [12] Johnson BA, Blevins RA. NMR view: a computer program for the visualization and analysis of NMR data. *J Biomol NMR* 1994;4:603–14.
- [13] Corne SA, Jorgenson AP, Fisher J. An artificial neural network for classifying cross peaks in two dimensional NMR spectra. *J Magn Reson* 1992;100:256–66.
- [14] Carrara EA, Pagliari F, Nicolini C. Neural networks for the peak-picking of nuclear magnetic resonance spectra. *Neural Netw* 1993;7:1023–32.
- [15] Rouh A, Louis-Joseph A, Lallemand J. Bayesian signal extraction from noisy FT NMR spectra. *J Biomol NMR* 1994;4:505–18.
- [16] Antz C, Neidig KP, Kalbitzer HR. A general Bayesian method for an automated signal class recognition in 2D NMR spectra combined with a multivariate discriminant analysis. *J Biomol NMR* 1995;5:287–96.
- [17] Koradi R, Billeter M, Engeli M, Güntert P, Wüthrich K. Automated peak picking and peak integration in macromolecular NMR spectra using AUTOPSY. *J Magn Reson* 1998;135:288–97.
- [18] Orekhov VY, Ibraghimov IV, Billeter M. MUNIN: a new approach to multi-dimensional NMR spectra interpretation. *J Biomol NMR* 2001;20:49–60.
- [19] Korzhneva DM, Ibraghimov IV, Billeter M, Orekhov VY. MUNIN: application of three-way decomposition to the analysis of heteronuclear NMR relaxation data. *J Biomol NMR* 2001;21:263–8.
- [20] Alipanahi B, Gao X, Karakoc E, Donaldson L, Li M. PICKY: a novel SVD-based NMR spectra peak picking method. *Bioinformatics* 2009;25:i268–75.
- [21] Liu Z, Abbas A, Jing B, Gao X. WaVPeak: picking NMR peaks through wavelet-based smoothing and volume-based filtering. *Bioinformatics* 2012;28:914–20.
- [22] Abbas A, Kong X, Liu Z, Jing B, Gao X. Automatic peak selection by a Benjamini–Hochberg-based algorithm. *PLoS One* 2013;8:e53112.
- [23] Messih MA, Chitale M, Bajic VB, Kihara D, Gao X. Protein domain recurrence and order can enhance prediction of protein functions. *Bioinformatics* 2012;28:i444–50.
- [24] Zimmerman DE, Kulikowski CA, Huang Y, Feng W, Tashiro M, Shimotakahara S, et al. Automated analysis of protein NMR assignments using methods from artificial intelligence. *J Mol Biol* 1997;269:592–610.
- [25] Coggins BE, Zhou P. PACES: protein sequential assignment by computer-assisted exhaustive search. *J Biomol NMR* 2003;26:93–111.
- [26] Volk J, Herrmann T, Wüthrich K. Automated sequence-specific protein NMR assignment using the memetic algorithm MATCH. *J Biomol NMR* 2008;41:127–38.
- [27] Lemak A, Steren CA, Arrowsmith CH, Llinás M. Sequence specific resonance assignment via Multicanonical Monte Carlo search using an ABACUS approach. *J Biomol NMR* 2008;41:29–41.
- [28] Wu KP, Chang JM, Chen JB, Chang CF, Wu WJ, Huang TH, et al. RIBRA – an error-tolerant algorithm for the NMR backbone assignment problem. *J Comput Biol* 2006;13:229–44.
- [29] Wan X, Lin G. CISA: combined NMR resonance connectivity information determination and sequential assignment. *IEEE/ACM Trans Comput Biol Bioinform* 2007;4:336–48.
- [30] Tycko R, Hu KN. A Monte Carlo/simulated annealing algorithm for sequential resonance assignment in solid state NMR of uniformly labeled proteins with magic-angle spinning. *J Magn Reson* 2010;205:304–14.
- [31] Masse JE, Keller R. Autolink: automated sequential resonance assignment of biopolymers from NMR data by relative-hypothesis-prioritization-based simulated logic. *J Magn Reson* 2005;174:133–51.
- [32] Güntert P, Salzmann M, Braun D, Wüthrich K. Sequence-specific NMR assignment of proteins by global fragment mapping with the program MAPPER. *J Biomol NMR* 2000;18:129–37.
- [33] Jung YS, Zweckstetter M. Mars–robust automatic backbone assignment of proteins. *J Biomol NMR* 2004;30:11–23.
- [34] Alipanahi B, Gao X, Karakoc E, Li SC, Balbach F, Donaldson L, et al. Error tolerant NMR backbone resonance assignment and automated structure generation. *J Bioinform Comput Biol* 2011;9:15–41.
- [35] Jang R, Gao X, Li M. Towards automated structure-based NMR resonance assignment. *Lect Notes Comput Sci* 2010;6044:189–207.
- [36] Jang R, Gao X, Li M. Towards fully automated structure-based NMR resonance assignment of  $^{15}\text{N}$ -labeled proteins from automatically picked peaks. *J Comput Biol* 2011;18:347–63.
- [37] Jang R, Gao X, Li M. Combining automated peak tracking in SAR by NMR with structure-based backbone assignment from  $^{15}\text{N}$ -NOESY. *BMC Bioinformatics* 2012;13:S4.
- [38] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. *Nucleic Acids Res* 2000;28:235–42.
- [39] Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, et al. BioMagResBank. *Nucleic Acids Res* 2008;36:D402–8.
- [40] Pons JL, Delsuc MA. RESCUE: an artificial neural network tool for the NMR spectral assignment of proteins. *J Biomol NMR* 1999;15:15–26.
- [41] Herrmann T, Güntert P, Wüthrich K. Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J Biomol NMR* 2002;24:171–89.
- [42] Gronwald W, Kalbitzer HR. Automated structure determination of proteins by NMR spectroscopy. *Prog Nucl Magn Reson Spectrosc* 2003;44:33–96.
- [43] Güntert P. Automated NMR structure calculation with CYANA. *Methods Mol Biol* 2004;278:353–78.
- [44] Güntert P. Automated structure determination from NMR spectra. *Eur Biophys J* 2009;38:129–43.
- [45] Li SC, Bu D, Xu J, Li M. Fragment-HMM: a new approach to protein structure prediction. *Protein Sci* 2008;17:1925–34.
- [46] Maadooliat M, Gao X, Huang JZ. Assessing protein conformational sampling methods based on bivariate lag-distributions of backbone angles. *Brief Bioinform* 2012. <http://dx.doi.org/10.1093/bib/bbs052>.
- [47] López-Méndez B, Güntert P. Automated protein structure determination from NMR spectra. *J Am Chem Soc* 2006;128:13112–22.
- [48] Bartels C, Billeter M, Güntert P, Wüthrich K. Automated sequence-specific NMR assignment of homologous proteins using the program GARANT. *J Biomol NMR* 1996;7:207–13.
- [49] Nilges M, Macias MJ, O'Donoghue SI, Oschkinat H. Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin. *J Mol Biol* 1997;269:408–22.
- [50] Fiorito F, Hiller S, Wider G, Wüthrich K. Automated resonance assignment of proteins: 6D APSY-NMR. *J Biomol NMR* 2006;35:27–37.