Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Magnetic Resonance 173 (2005) 310–316
JMR Journal of Magnetic Resonance
www.elsevier.com/locate/jmr

Communication

The fumarate sensor DcuS: progress in rapid protein fold elucidation by combining protein structure prediction methods with NMR spectroscopy

Jens Meiler*, David Baker*

Department of Biochemistry, University of Washington, BOX 357350, Seattle, WA 98195, USA

Received 29 September 2004; revised 24 November 2004
Available online 22 January 2005

Abstract

We illustrate how moderate resolution protein structures can be rapidly obtained by interlinking computational prediction methodologies with un- or partially assigned NMR data. To facilitate the application of our recently described method of ranking and subsequent refining alternative structural models using unassigned NMR data [Proc. Natl. Acad. Sci. USA 100 (2003) 15404] for such "structural genomics"-type experiments it is combined with protein models from several prediction techniques, enhanced to utilize partial assignments, and applied on a protein with an unknown structure and fold. From the original NMR spectra obtained for the 140 residue fumarate sensor DcuS, 1100 ^{1}H, ^{13}C, and ^{15}N chemical shift signals, 3000 ^{1}H–^{1}H NOESY cross peak intensities, and 209 backbone residual dipolar couplings were extracted and used to rank models produced by de novo structure prediction and comparative modeling methods. The ranking proceeds in two steps: first, an optimal assignment of the NMR peaks to atoms is found for each model independently, and second, the models are ranked based on the consistency between the NMR data and the model assuming these optimal assignments. The low-resolution model selected using this ranking procedure had the correct overall fold and a global backbone RMSD of 6.0Å, and was subsequently refined to 3.7 Å RMSD. With the incorporation of a small number of NOE and residual dipolar coupling constraints available very early in the traditional spectral assignment process, a model with an RMSD of 2.8 Å could rapidly be built. The ability to generate moderate resolution models within days of NMR data collection should facilitate large scale NMR structure determination efforts.

Keywords: De novo fold prediction; ROSETTA; Monte Carlo optimization; Assignment; Histidine kinase; Fumarate sensor; DcuS

1. Introduction

Knowledge of the three-dimensional structures of proteins is critical for many biological questions but the time-consuming processes of structure elucidation via X-ray crystallography or NMR spectroscopy cannot keep up with the rapidly growing number of sequenced genes and genomes [1]. Efforts are underway to accelerate protein structure elucidation by predicting structures computationally [2–6]. The lack of an experimental validation for the predicted models has been a major drawback of these methods so far.

In the recent years several algorithms were proposed that use either unassigned dipolar couplings [7,8] or unassigned NOESY intensities [9,10] to utilize unassigned NMR spectra in combination with modeling for structure determination and/or simultaneous resonance assignment. We developed a novel algorithm that utilizes NOESY intensities, dipolar couplings, chemical shifts, as well as sequential assignments for these purposes [11] and Jung et al. [12] published recently another approach to the same problem. Such algorithms have

* Corresponding authors. Fax: +1 206 6851792 (J. Meiler).
E-mail addresses: jens@jens-meiler.de (J. Meiler), dabaker@u.washington.edu (D. Baker).

1090-7807/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2004.11.031

the potential not only to increase the usability of structural models predicted from protein sequence alone, but also to greatly speed up structure elucidation from NMR spectroscopy.

DcuS is a histidine kinase that is composed of a sensory domain for stimulus perception, and a kinase domain catalyzing autophosphorylation of the protein at a conserved His residue [13–15]. The phosphoryl residue is then transferred to an aspartate residue of the response regulator which converts the regulator to the active state and is the second protein in the two component system [16,17]. The stimulus is generally a C_4–dicarboxylate such as fumarate, succinate, or malate [13,16–18]. DucS has a transmembrane topology with the sensory, or signal perceiving, domain in the periplasm, and the responding kinase domain in the cytoplasm. The structure of DcuS has been recently solved independently by NMR spectroscopy [19] and X-Ray crystallography [20].

In this communication, we report the application of structure prediction techniques in combination with unassigned and partially assigned NMR spectra to the DcuS example. To evaluate the potential application of such a protocol in "structural genomics" the setup is highly automated and maximum tolerance levels for inaccuracies of the predicted structural models as well as the NMR data, such as noise, incompleteness, or measurement artifacts were assumed (see below). In this respect, we enhance the protocol described in our previous PNAS publication [11] by utilizing a wider range of models generated from various protein structure prediction techniques and include the usage of automatically generated partial assignments as would be desired to maximize the theoretical and experimental input for the protocol. Whereas our previous work focused on the development of the computational algorithm, here we illustrate with a realistic test case how the protocol profits from the usage of several different structure prediction techniques or partially sequential assignments.

2. Results

Parallel to the traditional signal assignment and structure elucidation process (and therefore prior to the completion of any high-resolution structure [19]) we applied a variety of protein structure prediction techniques to generate possible low-resolution models for DcuS (compare Fig. 1). These models were afterwards filtered utilizing the original unassigned NMR spectra as input for our algorithm [11]. The validated model was then refined to improve its resolution, first with only unassigned NMR spectra [11] later using partially assigned spectra [11,21,22]. This procedure not only provides the best possible model at the highest possible confidence level at each point in time but can also give valuable feedback for the high-resolution structure elucidation process.

2.1. Fold recognition

The consensus fold recognition server 3D-Jury [23] was used in a first experiment to detect potential homologous sequences with known structure. For the sequence of DcuS the GAF domain 1f5mA [24] (FSSP 278.2.2, SCOP d.110.2) has the highest score (38) and a primary sequence identity of 12.1%. The next three hits are 1f5mA in a slightly different alignment, the homologue GAF domain 1f5mB or 1mc0A, and have scores between 21 and 33 depending on the protein and the alignment. However, since only scores above 50 indicate a high confidence prediction (>90% correct) [23], the GAF domain can only be assumed as one of several possible folds. Nevertheless, it was used as a template for comparative modeling.

2.2. Building comparative models

A first comparative model (comparative model 1, Fig. 1) was built with ROSETTA [25] restricting the backbone of all secondary structure elements to their respective positions in 1f5mA. Since the sequence similarity is very low, DcuS and 1f5mA are likely to have significantly different structures, even if the folds are the same. To allow more significant structural changes, a second model was built by trimming the alignment to exclude the helical regions at the beginning and the end of the sequence (AA 41–95, 170–180). Hence, only the topology of the β-sheet is conserved from the structure of 1f5mA (compare Fig. S1) and helices as well as loops are built subsequently de novo (comparative model 2, Fig. 1) using the Rosetta protocol [25].

2.3. Building de novo models

To sample the possible structural space in an unbiased fashion 10,000 de novo models are built using the ROSETTA fragment replacement algorithm [26,27] and the CASP5 protocol described elsewhere [5]. The sequence of DcuS is cut into overlapping fragments of three and nine amino acids. The PDB is subsequently screened for fragments that have a high primary sequence homology and a secondary structure that matches the predicted secondary structure of the query sequence. These fragments, which sample possible local conformations for the protein backbone, are combined using a Monte Carlo algorithm. Six representative models were selected using a clustering procedure (de novo models 1–6, Fig. 1).

The six cluster centers and the two comparative models were generated without the use of experimental information. The overall folds of comparative model 2 and

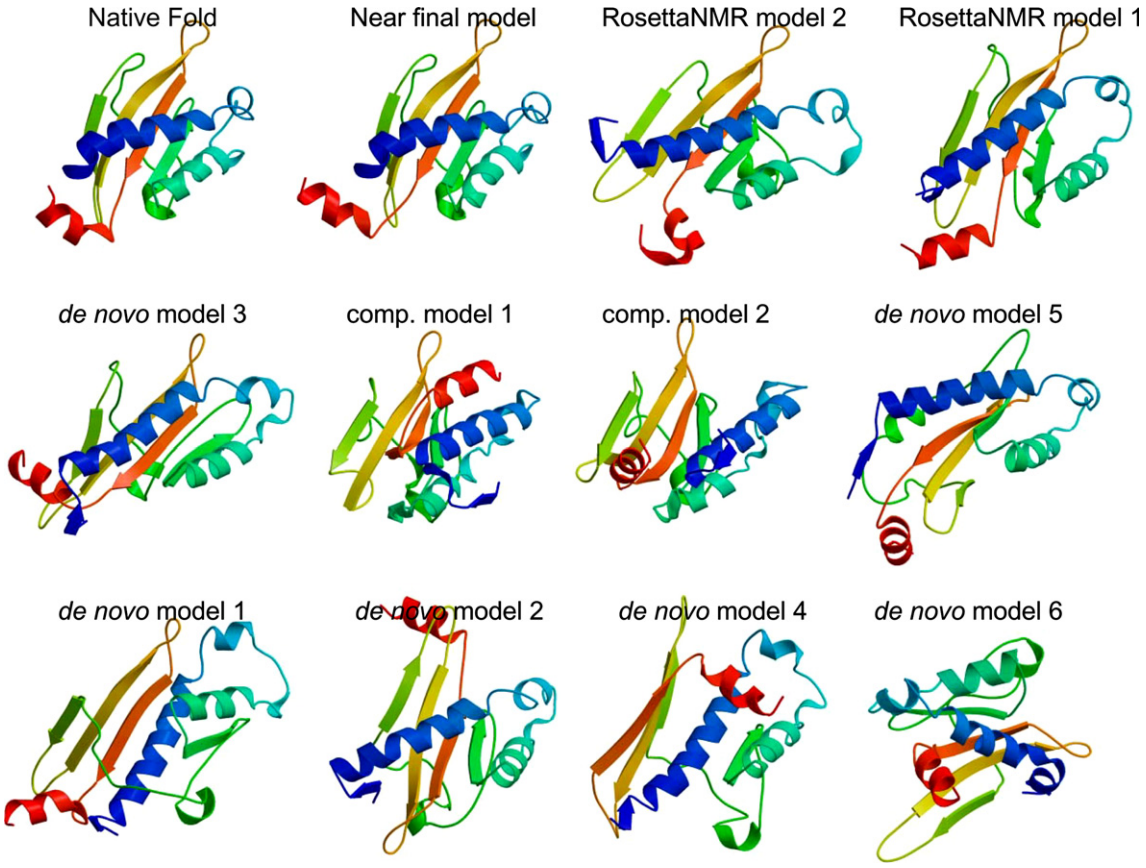

Fig. 1. Ribbon diagrams of the DcuS models. The amino acids 41 (blue)–180 (red) are shown in a rainbow color scheme. The ROSETTANMR model 1 and 2, de novo model 3, and comparative models 2 resemble the native fold. Comparative model 1 and de novo model 5 have the correct β-sheet topology but the helix packing is significantly different from the native fold. De novo models 1, and 4 form two local β-sheets, models 2 and 6 form a parallel instead of an anti-parallel contact between the C-terminal three stranded sheet and the fourth strand. The two models within each of these three groups of incorrect topologies can further be distinguished by differences in their helix packing.

de novo model 3 were found to be similar (89 amino acids superimposable at 6 Å). This increased our confidence that these models represent the correct fold, however without the incorporation of experimental information both models cannot be validated and the other possibilities cannot be excluded.

2.4. NMR data generation

The NMR experiments are described in detail in Pappalardo et al. [19]. Distance restraints were obtained from the intensities of ^{1}H–^{1}H NOESY cross peak intensities (NOEs) extracted from ^{15}N-edited three-dimensional NOESY-HSQC, and ^{13}C-edited three-dimensional NOESY-HSQC spectra. A set of $^{1}D_{NH}$, $^{1}D_{NC\alpha}$, $^{1}D_{C\alpha C'}$, and $^{1}D_{C\alpha H\alpha}$ backbone residual dipolar couplings (RDCs) of DcuS were calculated from the difference in the corresponding J splitting measured in a protein sample containing 10 mg/ml Pf1 filamentous phage [28] and in a protein sample in the absence of phage [29]. $^{1}D_{NH}$ and $^{1}D_{NC\alpha}$ residual dipolar couplings were measured simultaneously using a modified interleaved three-dimensional TROSY-HNCO experiment [30] and $^{1}D_{C\alpha H\alpha}$ and $^{1}D_{C\alpha C'}$ residual dipolar couplings using a modified interleaved three-dimensional CBCA-CONH experiment [31]. One crucial point for the use of any computer program based on frequency and intensity lists of different NMR spectra is the tolerance level for CS changes between spectra (because of differences in temperature, spectrometer, or sample). To account for a potentially decreased quality of spectra in larger scale application structural genomics the tolerance levels were set rather generously to be 0.25 ppm for ^{1}H CSs, 1 ppm for ^{13}C CSs, and 2.5 ppm for ^{15}N CSs. If no assignment within these ranges was possible, the tolerance was increased up to 1 ppm for ^{1}H CSs, 4 ppm for ^{13}C CSs, and 10 ppm for ^{15}N CSs.

A combined list of the total of 1100 ^{1}H, ^{13}C, and ^{15}N chemical shift signals (CSs) was compiled from RDCs and NOEs. The almost 3000 NOEs were obtained from the XEASY-integrated spectra and linked to the respective CSs values as was done for the 209 RDCs. In addition a list of 315 connectivity restraints was compiled. This list contains pairs of CSs where the corresponding atoms are known to be bonded from their RDC (e.g., Cα–C′) or from the $^{15}N/^{13}C$-edited three-dimensional

NOESY-HSQC spectra (N–H and Cα–Hα). 25% randomly selected sequential assignments were input as 111 additional connectivity restraints for the corresponding $C\alpha^{i}$–C'^{i}, C'^{i}–N^{i+1}, N^{i+1}–$C\alpha^{i+1}$ bonds in some experiments. However, no initial actual assignment of any atom to any CS was input in any of the experiments.

2.5. Model selection using unassigned/partially assigned NMR data

To assess these models in a short period of time, a recently developed algorithm that compares protein models with unassigned NMR [11] data was applied. In contrast to X-ray crystallography, where the growing of suitable crystals is often the time-consuming step, the collection of data can be done rapidly by NMR spectroscopy, once the protein is expressed. Comparing the eight models with unassigned/partially assigned NMR data is one of the fastest validations conceivable yielding structural information at atomic resolution. NMR spectra assignment is comprised of three general steps: (1) sequential backbone resonance assignment, (2) side chain resonance assignment, and (3) NOE cross peak assignment. While the two latter steps are often time-expensive and error-prone, the first step is generally at least in part straightforward and fast.

In a first experiment for each of the eight models, five optimized assignments were obtained from completely unassigned spectra (compare reference [11]). Their average consistency score S was translated into a Z score with respect to the complete set of 500 optimized assignments (50 for each of the 10 models) to prove the statistical relevance of the rather small changes in the overall consistency score. The Z score is defined as $Z_i = (S_i - \bar{S})/\sigma_S$ with $\bar{S}$ as the average consistency score and σ_S as the standard deviation over all S_i.

In a second experiment, the protocol was repeated including sequential assignments from triple resonance spectra to test its ability to utilize the assignment information which can be generated fast and straightforward. Randomly selected 25% of these sequential assignments were assumed to be known and included as boundary conditions in the calculation. This rather low level was again chosen to account for expected lower accuracies and higher ambiguities of an automated protocol. Note that only the sequential assignment but not the direct assignment of backbone CSs to amino acids was used. The selected subset of known sequential assignments spanned the complete sequence. According to their increased Z score (Fig. 2) de novo model 3 and comparative model 2 were selected as most consistent with the NMR data. The restraints for refinement were obtained from the consistently assigned signals for de novo model 3 [11].

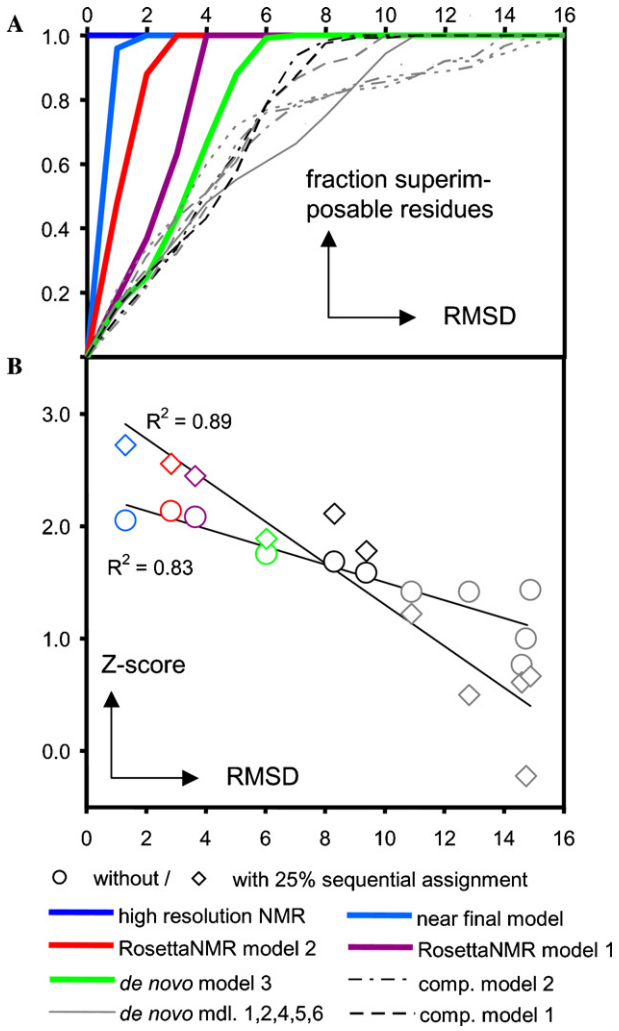

Fig. 2. Model quality (A) and correlation with NMR consistency score (B). (A) Fraction of all amino acids that can be superimposed below a certain RMSD threshold; the superior quality of the ROSETTANMR models (red and magenta), the de novo model 3 (green), and the comparative model 2 (black) is evident from the steeper increase of the corresponding curves. (B) Models with higher NMR consistency Z scores have higher accuracy. Assignments built with no sequential constraints are shown as a dashed line with circles. Assignments built with sequential constraints for 25% of all amino acids are shown as solid line with squares. The correlation as well as discrimination becomes better as sequential assignment information is added.

2.6. Refining the model with sparse assigned NMR data

In the final step de novo model 3 (the best scoring model from the above evaluation) was refined with ROSETTANMR using the ROBETTA server (http://robetta.bakerlab.org/) [32]. The first model was built using 33 assigned RDCs and 19 low resolution backbone–backbone restraints as obtained from the partial assignments in the prior protocol [11]. The second calculation was undertaken to demonstrate the potential to use these initial models as supportive information in the traditional path of protein structure elucidation

from NMR data. It utilized 92 assigned N–H^N RDCs and 130 low-resolution backbone–backbone distance restraints to refine de novo model 3 from partially assigned NOESY spectra which became available at a later stage of the traditional structure elucidation [19] but well before the availability of a high-resolution structure (compare Table 1). Out of 100 models built the one in best agreement with the assigned experimental constraints was chosen in both cases (ROSETTANMR models 1 and 2, Fig. 1).

2.7. The high-resolution structure

The structure of the periplasmic domain of this protein was recently solved by NMR spectroscopy in the absence of fumarate [19]. The structure of the perimplasmic domain of the citrate sensor cocrystallized with citrate has also been solved by X-ray crystallography [20]. The 140 amino acid periplasmic domain of the protein contains about equal amounts of α-helix and β-strand. The numbering of the amino acids used in the following starts with 41 and goes up to 180 to be consistent with the common numbering scheme [19].

In a late stage of the project but prior to the completion of the NMR solution structure, a near final model was utilized for evaluating the quality of the applied methods [11]. When completed, the NMR solution structure determined using ~1000 assigned NOEs and ~300 RDCs (native fold, PDB code 1OJG, Fig. 1) [19] was compared with the nine models introduced in the previous paragraphs. All RMSD values discussed in the following refer to the backbone of the region between amino acids 45–169. To facilitate comparison, the evaluation of the consistency of the high-resolution NMR structure, the near final model, as well as the two ROSETTANMR models with the unassigned NMR data was carried out as done earlier for the eight models [11].

3. Discussion

The two comparative models have aspects of the correct overall topology, which suggests the potential remote homology detected by the 3D-Jury server is real. The more conservative model 1 has a slightly higher RMSD (Table 1), mainly because the C-terminal α-helix packs against the N-terminal helix causing an unnatural bend in the last strand. A tight packing of the C-terminal α-helix is unlikely if the increased dynamics obtained in the NMR experiments is considered. In comparative model 2 both issues are resolved by repacking all α-helices.

However, significantly better in its RMSD is de novo model 3 which can be completely superimposed with the native fold at 6 Å—a surprisingly low value for a protein of this length [33]. De novo models 1, 2, 4, 5, and 6 are significantly worse in their quality. As ensured by the clustering protocol, the cluster centers must be significantly different from model 3 and therefore their lower similarity with the native fold is expected. The only other model with the correct β-sheet topology (de novo model 5) has the helices packed on the opposite sides of the sheet but has the lowest RMSD among the remaining five cluster centers.

Table 1
Quality of the generated models and results of the consistency analysis with unassigned NMR data

Model description	RMSD in Å/correct topology	Number AA superimposable below			Estimated[a] time limits with/without NMR valid	NMR cons. score S[c]		Z score[c] $Z_i = (S_i - \bar{S})/\sigma_S$		Fraction correctly assigned signals	
		4 Å	6 Å	8 Å			psa.[b]		psa.[b]		psa.[b]
High-resolution model	0.00/☺	**125**	**125**	**125**	180d/—	—	—	—	—	—	—
Near final model	1.31/☺	**125**	**125**	**125**	90d/—	0.798	0.792	2.05	2.72	2.5%	8.2%
ROSETTANMR model 2	2.83/☺	**125**	**125**	**125**	40d/—	0.799	0.789	2.13	2.56	2.8%	11.2%
ROSETTANMR model 1	3.67/☺	**125**	**125**	**125**	20d/—	0.798	0.787	2.08	2.44	2.5%	9.8%
De novo model 3	6.03/☺	83	**124**	**125**	10d/2d	0.794	0.777	1.75	1.89	2.1%	8.3%
Comparative model 2	8.31/☺	64	98	**124**	10d/2d	0.793	0.781	1.68	2.11	1.8%	6.8%
De novo model 1	12.83/☹	60	76	94	10d/2d	0.790	0.752	1.42	0.50	0.7%	5.8%
De novo model 2	14.73/☹	75	96	102	10d/2d	0.785	0.739	1.00	−0.22	0.7%	3.5%
De novo model 4	14.59/☹	58	95	101	10d/2d	0.782	0.754	0.77	0.61	0.9%	4.9%
De novo model 5	10.90/😐	64	98	114	10d/2d	0.790	0.765	1.42	1.22	0.8%	7.6%
De novo model 6	14.89/☹	64	89	101	10d/2d	0.790	0.755	1.43	0.67	1.0%	3.7%
Comparative model 1	9.38/😐	54	98	122	10d/2d	0.792	0.775	1.58	1.78	0.9%	7.3%
Average consistency score $\bar{S}$[d]						0.773	0.743				
Standard deviation σ_S[d]						0.012	0.018	0.02	0.02		
Correlation coeff. R [2] to RMSD	**0.71**	**0.78**	**0.77**			**0.83**	**0.89**	**0.83**	**0.89**	**0.86**	**0.81**

[a] Time limits given with NMR validation are estimates assuming an optimal coordination of the NMR experiments with the calculations which was not always achievable due to the long-distance collaboration of the two research groups.
[b] Randomly chosen 25% partial sequential assignments were used.
[c] These values are computed over the five best scoring assignments out of 50 optimized assignments for every of the 10 models.
[d] These values are computed over 500 optimized assignments (50 for each of the 10 models) as used for computing the Z scores.

The results of the comparison of the eight models with unassigned NMR data are summarized in Table 1 and Fig. 2B. De novo model 3 and comparative model 2 score best among the structures predicted from sequence alone. Since both models have the same overall topology, it is safe to assume that the correct fold for the unknown protein is found. The advantage of this approach is the very rapid experimental validation of the fold prediction result. The inclusion of sequential assignments into the calculation increases the information content and thus leads to a better discrimination of correct from incorrect topologies as can be seen from the increased correlation coefficients and the increased slope of the regression line (Fig. 2). Also, the average percentage of correctly assigned signals increases from 2.1 to 8.3% which allows the detection of consistently assigned signals and in turn the refinement of the structure [11] to medium resolution (ROSETTANMR model 1).

Unassigned NOEs describe a three-dimensional cloud of hydrogen atoms that fits the three-dimensional structure of the protein. The rather low overall percentages of correctly assigned signals reflects the fact that after translating NOEs into generous allowed regions and considering the incorrectness of even the best models available at this stage many assignments will score well because of the general agreement of the shape of the cloud and the shape of the molecule. In most of these assignments only a small number of hydrogen atoms will be correctly assigned since the interchange of spatially close hydrogen atoms in the assignment has little influence on the score. Similar conclusions hold for RDCs, where the interchange of assignments within a secondary structure element (e.g., amino acid i with amino acids $i+3$, $i+4$, or $i+7$ in an α-helix or amino acid i with amino acids $i+2$ in a β-sheet) is frequently possible without affecting the score significantly, in particular after large tolerance ranges were added to the dipolar couplings to account for the worse quality of predicted protein models compared to an experimental X-ray or NMR structure.

Not surprisingly, the best model is generated by the ROSETTANMR (model 2) due to the larger number of experimental constraints used. During the assignment of NMR spectra certain experimental values such as backbone–backbone NOEs and N–H^N residual dipolar couplings can be more easily and faster assigned than others. In this particular example, a set of 92 N–H^N dipolar couplings and 130 NOEs was manually assigned first and used as input for ROSETTANMR. This information, although later available than unassigned spectra, is still much earlier available than completely assigned NOESY spectra. Fewer signals from mostly backbone atoms in this region of the NOESY spectra make assignment easier and mis-assignment less likely. With less than two experimental constraints per amino acid an overall RMSD below 3 Å is achieved. The structure of the N-terminal helix and the β-sheet in particular profited from including these experimental data.

4. Conclusion

Applying the protein structure prediction protocol developed for the CASP5 experiment [5] it was possible to predict the correct fold of the 140 amino acid protein DcuS in less than 2 days of computation. An experimental validation for the suggested models was achieved in less than 10 days by comparing unassigned NMR spectra with the structural proposals and scoring their agreement [11]. A first refinement of the best de novo model from 6 to 4 Å was achieved in 20 days by using partial sequential assignments, a further refinement to 3 Å was achieved in 30 days with less than 200 assigned NOEs and RDCs that were particularly readily accessible [21,22]. Hence, the time required for solving the fold of DcuS is drastically reduced compared to the complete structure elucidation via NMR spectroscopy or X-ray crystallography. The predicted models were used as starting point for refinement using only sequential or sparse assignments. The approach has potential for structural genomics approaches, where the folds of sequences in a medium-sized sequence databases could be solved much faster. The models generated can guide the further structure elucidation process and can help to identify and focus on so far unknown folds or folds of special interest for high-resolution structure elucidation.

Acknowledgments

We thank Dylan Chivian for building the comparative model 1. Further we thank Christian Griesinger, Boris Coligaev, Markus Zweckstetter, Vinesh Vijayan, and Jochen Junker for providing the NMR spectra, peak lists, and partial assignments prior publication. J.M. thanks the Human Frontier Science Program (HFSP) for financial support. This work was also supported by the HHMI.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmr.2004.11.031.

References

[1] A. Sali, 100,000 protein structures for the biologist, Nat. Struct. Biol. 5 (12) (1998) 1029–1032.

[2] R. Bonneau, J. Tsai, I. Ruczinski, D. Chivian, C. Rohl, C.E.M. Strauss, D. Baker, Rosetta in CASP4: progress in ab initio protein structure prediction, Proteins 45 (Suppl. 5) (2001) 119–126.
[3] D. Fischer, L. Rychlewski, R.L. Dunbrack Jr., A.R. Ortiz, A.. Elofsson, CAFASP3: the third critical assessment of fully automated structure prediction methods, Proteins 53 (Suppl. 6) (2003) 503–516.
[4] D. Fischer, A. Elofsson, L. Rychlewski, F. Pazos, A. Valencia, B. Rost, A.R. Ortiz, R.L. Dunbrack Jr., CAFASP2: the second critical assessment of fully automated structure prediction methods, Proteins (Suppl. 5) (2001) 171–183.
[5] P. Bradley, D. Chivian, J. Meiler, K. Misura, W. Wedemeyer, C. Rohl, B. Schief, D. Baker, Rosetta in CASP5: progress in ab initio protein structure prediction, Proteins Struct. Funct. Genet. 53 (Suppl. 6) (2003) 457–468.
[6] R.L. Dunbrack Jr., Comparative modeling of CASP3 targets using PSI-BLAST and SCWRL, Proteins (Suppl. 3) (1999) 81–87.
[7] F. Tian, H. Valafar, J.H. Prestegard, A dipolar coupling based strategy for simultaneous resonance assignment and structure determination of protein backbones, J. Am. Chem. Soc. 123 (47) (2001) 11791–11796.
[8] J.-C. Hus, J. Prompers, R. Brueschweiler, Assignment strategy for proteins with known structure, J. Magn. Res. 157 (2002) 119–123.
[9] A. Grishaev, M. Llinas, CLOUDS, a protocol for deriving a molecular proton density via NMR, Proc. Natl. Acad. Sci. USA 99 (10) (2002) 6707–6712.
[10] A. Grishaev, M. Llinas, Protein structure elucidation from NMR proton densities, Proc. Natl. Acad. Sci. USA 99 (10) (2002) 6713–6718.
[11] J. Meiler, D. Baker, Rapid protein structure elucidation utilizing unassigned NMR data, Proc. Natl. Acad. Sci. USA 100 (26) (2003) 15404–15409.
[12] Y.S. Jung, M. Sharma, M. Zweckstetter, Simultaneous assignment and structure determination of protein backbones by using NMR dipolar couplings, Angew. Chem. Int. Ed. Engl. 43 (26) (2004) 3479–3481.
[13] I.G. Janausch, I. Garcia-Moreno, G. Unden, Function of DcuS from *Escherichia coli* as a fumarate-stimulated histidine protein kinase in vitro, J. Biol. Chem. 277 (42) (2002) 39809–39814.
[14] I.G. Janausch, E. Zientz, Q.H. Tran, A. Kroger, G. Unden, C4-dicarboxylate carriers and sensors in bacteria, Biochim. Biophys. Acta 1553 (1–2) (2002) 39–56.
[15] J.S. Parkinson, E.C. Kofoid, Communication modules in bacterial signaling proteins, Annu. Rev. Genet. 26 (1992) 71–112.
[16] E. Zientz, J. Bongaerts, G. Unden, Fumarate regulation of gene expression in *Escherichia coli* by the DcuSR (dcuSR genes) two-component regulatory system, J. Bacteriol. 180 (20) (1998) 5421–5425.
[17] P. Golby, S. Davies, D.J. Kelly, J.R. Guest, S.C. Andrews, Identification and characterization of a two-component sensor-kinase and response-regulator system (DcuS–DcuR) controlling gene expression in response to C4-dicarboxylates in *Escherichia coli*, J. Bacteriol. 181 (4) (1999) 1238–1248.
[18] S. Kaspar, R. Perozzo, S. Reinelt, M. Meyer, K. Pfister, L. Scapozza, M. Bott, The periplasmic domain of the histidine autokinase CitA functions as a highly specific citrate receptor, Mol. Microbiol. 33 (4) (1999) 858–872.
[19] L. Pappalardo, I.G. Janausch, V. Vijayan, E. Zientz, J. Junker, W. Peti, M. Zweckstetter, G. Unden, C. Griesinger, The NMR structure of the sensory domain of the membraneous two-component fumarate sensor (histidine protein kinase) DcuS of *Escherichia coli*, J. Biol. Chem. 278 (40) (2003) 39185–39188.
[20] S. Reinelt, E. Hofmann, T. Gerharz, M. Bott, D.R. Madden, J. Biol. Chem. 278 (2003) 39189–39196.
[21] P.M. Bowers, C.E.M. Strauss, D. Baker, Denovo protein structure determination using sparse NMR data, J. Biomol. NMR 18 (2000) 311–318.
[22] C. Rohl, D. Baker, De Novo determination of protein backbone structure from residual dipolar couplings using rosetta, J. Am. Chem. Soc. 124 (11) (2002) 2723–2729.
[23] K. Ginalski, A. Elofsson, D. Fischer, L. Rychelewski, 3D-Jury: a simple approach to improve protein structure predictions, Bioinformatics 19 (8) (2003) 1015–1018.
[24] Y.S. Ho, L.M. Burden, J.H. Hurley, Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor, EMBO J. 19 (20) (2000) 5288–5299.
[25] C.A. Rohl, C.E. Strauss, D. Chivian, D. Baker, Modeling structurally variable regions in homologous proteins with rosetta, Proteins 55 (3) (2004) 656–677.
[26] K.T. Simons, C. Kooperberg, E. Huang, D. Baker, Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions, J. Mol. Biol. 268 (1997) 209–225.
[27] R. Bonneau, C.E.M. Strauss, D. Baker, Improving the performance of rosetta using multiple sequence alignment information and global measures of hydrophobic core formation, Proteins Struct. Funct. Genet. 43 (2001) 1–11.
[28] M. Zweckstetter, A. Bax, Characterization of molecular alignment in aqueous suspensions of Pf1 bacteriophage, J. Biomol. NMR 20 (2001) 365–377.
[29] M.R. Hansen, M. Rance, A. Pardi, Observation of long-range 1H–1H distances in solution by dipolar coupling interactions, J. Am. Chem. Soc. 120 (43) (1998) 11210–11211.
[30] J.J. Chou, F. Delaglio, A. Bax, Measurement of one-bond 15N–13C′ dipolar couplings in medium sized proteins, J. Biomol. NMR 18 (2000) 101–105.
[31] J.J. Chou, A. Bax, Protein side-chain rotamers from dipolar couplings in a liquid crystalline phase, J. Am. Chem. Soc. 123 (2001) 3844–3845.
[32] D.E. Kim, D. Chivian, D. Baker, Protein structure prediction and analysis using the Robetta server, Nucleic Acids Res. 32 (Web Server issue) (2004) W526–W531.
[33] R. Bonneau, C.E.M. Strauss, C. Rohl, D. Chivian, P. Bradley, L. Malmström, T. Robertson, D. Baker, De novo prediction of three-dimensional structures for major protein families, J. Mol. Biol. 322 (1) (2002) 65–78.