J|A|C|S

ARTICLES

Published on Web 05/18/2006

Direct Use of Unassigned Resonances in NMR Structure Calculations with Proxy Residues

Eiso AB,[†,§] David J. R. Pugh,[‡] Robert Kaptein,[†] Rolf Boelens,[†] and
Alexandre M. J. J. Bonvin*,[†]

Contribution from the NMR Research Group, Faculty of Science, Utrecht University, Padualaan 8, NL-3584 CH Utrecht, The Netherlands, and University of the Western Cape, Private Bag X17, Bellville 7535, South Africa

Received January 5, 2006; E-mail: a.m.j.j.bonvin@chem.uu.nl

Abstract: We present a method that significantly enhances the robustness of (automated) NMR structure determination by allowing the NOE data corresponding to unassigned NMR resonances to be used directly in the calculations. The unassigned resonances are represented by additional atoms or groups of atoms that have no interaction with the regular protein atoms except through distance restraints. These so-called "proxy" residues can be used to generate NOE-based distance restraints in a similar fashion as for the assigned part of the protein. If sufficient NOE information is available, the restraints are expected to place the proxies at positions close to the correct atoms for the unassigned resonance, which can facilitate subsequent assignment. Convergence can be further improved by supplying additional information about the possible identities of the unassigned resonances. We have implemented this approach in the widely used automated assignment and structure calculation protocols ARIA and CANDID. We find that it significantly increases the robustness of structure calculations with regard to missing assignments and yields structures of higher quality. Our approach is still able to find correctly folded structures with up to 30% randomly missing resonance assignments, and even when only backbone and β resonances are present! This should be of significant value to NMR-based structural proteomics initiatives.

Introduction

Obtaining the last 10−15% of missing resonance assignments for NMR structure determination is a process that often requires a considerable amount of effort. Typical examples of difficult assignments are the aromatic side-chain resonances and long-chain aliphatic methylene resonances. Often those are obtained only by manual analysis of NOE spectra once an initial model of the structure is available.

Widely used protocols that perform combined iterative NOE assignment and structure calculation protocols, like ARIA[1] and CANDID,[2] are defined in terms of a complete structural model of the protein with chemical shift assignments for (some of) its nuclei. These protocols are quite sensitive to missing assignments: for example, missing aromatic resonances might very well prevent finding the correct fold in the first iteration, or at least have a negative effect on the reliability of the initial fold. Some NOEs will remain unassigned because of missing resonance assignments, which will decrease the precision of the structure. Even worse, some NOEs will have wrong assignments, leading to incorrect restraints, which will have a negative impact on the accuracy of the structures.

A possible circumvention of this problem is offered by so-called assignment-free methods, where structure calculation precedes resonance assignment. There have been several earlier efforts at developing methods for structure determination based only on NOE data, such as the ANSRS method[3] and, more recently, the CLOUDS protocol.[4,5] These methods cast the structure calculation problem completely in terms of free atoms, which have to be assigned later. Although these methods have been shown to work in a number of cases, the requirements of very high quality NMR data and particularly a lack of resonance overlap seem to limit their applicability, thus far, to proteins below 10 kDa.

Here, we propose a method that does not aim at replacing the conventional methods of (semi)automatic NOE assignment and structure calculation, but rather tries to improve these procedures by introducing a free atom approach to treat the portion of the NOE data that is normally ignored or misinterpreted due to missing resonance assignments.

In our approach, the unassigned resonances are represented by additional atoms or groups of atoms that interact with the regular protein atoms exclusively through distance restraints. These so-called proxy atoms or proxy residues can be used to generate NOE-based distance restraints in a similar fashion as

[†] Utrecht University.

[‡] University of the Western Cape.

[§] Current address: Gorlaeus Laboratories, Leiden University, Einsteinweg 55, 2333 CC Leiden, The Netherlands.

(1) Linge, J. P.; Habeck, M.; Rieping, W.; Nilges, M. *Bioinformatics* **2003**, *19*, 315−316.
(2) Herrmann, T.; Güntert, P.; Wüthrich, K. *J. Mol. Biol.* **2002**, *319*, 209−227.
(3) Kraulis, P. J. *J. Mol. Biol.* **1994**, *243*, 696−718.
(4) Grishaev, A.; Llinas, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6713−6718.
(5) Grishaev, A.; Llinas, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6707−6712.

10.1021/ja058504q CCC: $33.50

for the assigned part of the protein. If sufficient information is available, the NOE-based distance restraints are expected to place the proxy residues at positions close to the correct atoms for the unassigned resonance. This can facilitate subsequent assignment. Additional information about the nature of the unassigned resonances can be conveniently provided in the form of identity restraints (IDRs): ambiguous distance restraints which restrain the position of the proxy atoms to be close to at least one of the atoms corresponding to possible assignments for the unassigned resonance.

We demonstrate the feasibility of this approach and show that it leads to improved structural quality and suggests possible assignments at the same time. Our method makes structure determination protocols such as ARIA and CANDID significantly more robust to missing resonances. It even works if only backbone and β resonances, but no other side-chain resonances, are present.

Theory

Our computational model includes a complete description of the protein under investigation, as well as a number of free-floating dummy atoms or groups of dummy atoms. Because these act as temporary replacements for the missing correct assignments, we refer to these as *proxy residues*. To allow the proxy residues to occupy the same position in space as the (yet unknown) atoms they represent, they are defined to have no nonbonded interactions with the protein atoms, but to interact with them exclusively via NOE-based distance restraints, and via identity restraints: ambiguous distance restraints that carry information about the possible assignments (see below).

If the NOE data contain enough information, the proxy atoms are expected to find a position in the calculated structures in the proximity of the position of the atoms corresponding to the unassigned resonance. Analysis of the positions of the proxies should, in many cases, give useful suggestions for the correct assignment of the resonance, or at least narrow-down the number of possibilities.

Figure 1 contrasts the conventional approach (Figure 1A) with our proposed method (Figure 1B). The use of proxy atoms allows us to benefit from the information contained in the NOEs involving unassigned resonances, even about long-range interactions. In the conventional approach, this information is lost because the NOEs cannot lead to “correct” distance restraints. Even if the atoms corresponding to the target assignment do not have a correct conformation due to lack of restraints, the restraints to the proxy atoms can still carry valid long-range information (Figure 1C). In such a case, the proxy would not end up in the proximity of the correct atom.

Additional information about possible assignments can often be inferred from, e.g., chemical shifts, constant-time ^{13}C HSQC, or other spectra. Usually the type of chemical moiety (e.g., methine, methylene, methyl, atomatic ring, amide) is known, even if the exact assignment is not. IDRs can be used to pull the proxy atom close to atoms or groups of atoms corresponding to possible assignments (see Figure 2). The requirement that distinct resonances should not have the same assignment, and thus distinct proxy residues should not occupy the same position, is easily enforced by applying a repulsive potential between the atoms of different proxy residues.

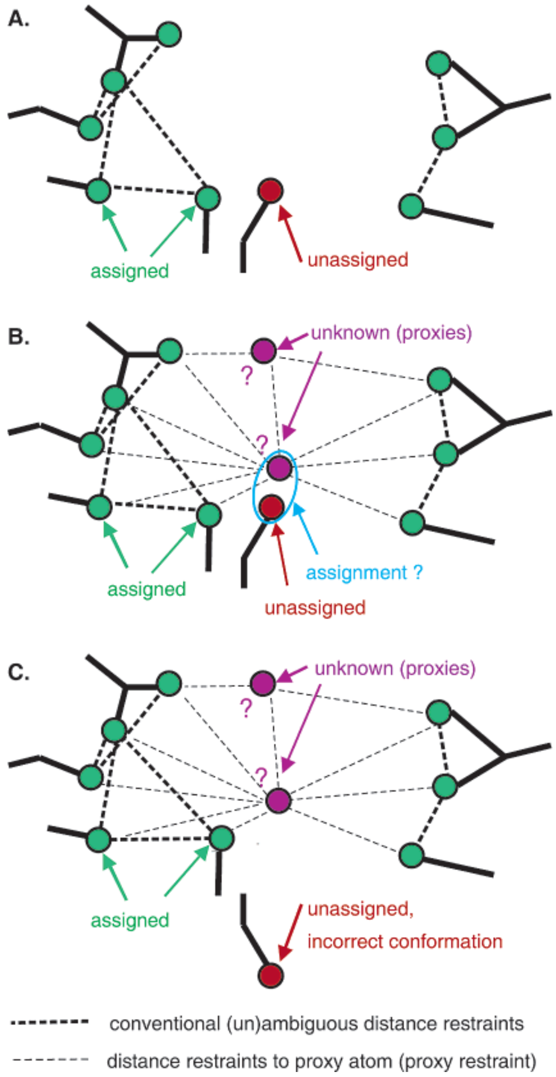

Figure 1. Concept of proxy restraints. Panel A shows the conventional situation where NOE-based distance restraints (thick dashed lines) are defined only between protein atoms with known resonance assignments (green atoms). Panel B shows that additional distance restraints (narrow dashed lines) corresponding to the unassigned resonances can be introduced if proxy atoms (purple) are used, and that the resulting restraint network can carry long-range information by connecting the assigned atoms on the left and the right. This can be the case even if the atoms corresponding to the correct position do not have the correct conformation in the current model (panel C).

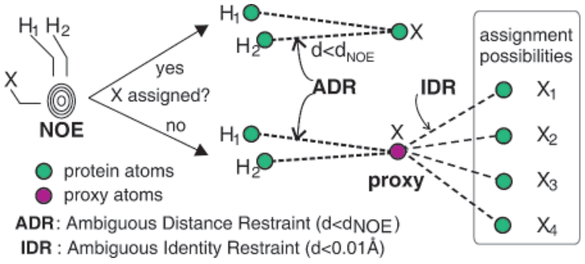

Figure 2. Generation of NOE-based distance restraints in case of assigned resonances (upper arrow) or unassigned resonances using proxies (lower arrow). Information about possible assignment candidates (X_1, X_2, X_3, X_4) can be provided by defining identity restraints between the proxy and all assignment candidates.

Results and Discussion

Our initial implementation of the proxy method in ARIA1.2 was used during the refinement of the DWNN protein. At a

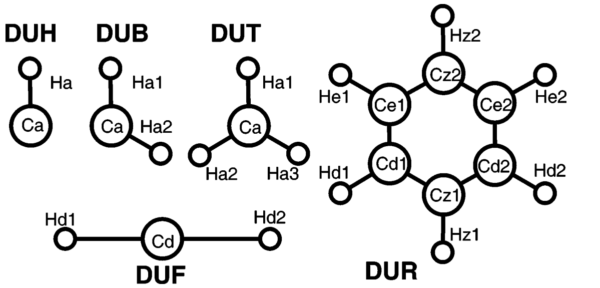

Figure 3. Proxy residues used in this paper.

Table 1. Quality of DWNN Structures: Comparison of an ARIA Run with and without Proxy Residues[a]

	without proxies	with proxies	ref
Ramachandran:			
most favored	62.6	73.2	92.6
additionally allowed	29.9	20.5	6.58
generously allowed	6.1	4.6	0.1
disallowed	1.4	1.7	0.68
PROCHECK *G*-factors:			
dihedrals	−0.82	−0.75	−0.17
covalent	0.44	0.48	0.15
overall	−0.33	−0.27	−0.07
rmsd to ref bb:8−78	4.25	3.10	0.58

[a] Ramachandran scores and *G*-factors were calculated using PROCHECK.[6] The rmsd is defined with respect to the final set of reference structures, determined with virtually complete assignments.

later stage, we implemented the proxy method in CYANA/CANDID, mainly to benefit from the network-anchoring implementation which is, in our experience, very important for reliable structure determination. The performance of the proxy method in CANDID was tested with two proteins: DWNN (PDB code 2c7h) and enzyme IIB^{Chb} (PDB code 1h9c). The calculations were performed using the proxy residues shown in Figure 3.

The Use of Proxies To Guide Manual Assignment. During the structure calculation and structural refinement of DWNN, we used proxy residues in ARIA 1.2 to find the assignments of several thus far unassigned resonances in the ^{13}C NOESY-HSQC. A total of 38 proxy residues were defined, of which 24 converged to an average displacement of less than 4 Å after superposition on the backbone atoms of residues 8−78. No IDRs were used at this stage. Visual inspection of the proxy residue positions led to proposed assignments for 10 resonances that were later confirmed using H[C]CH COSY and HC[C]H TOCSY spectra. Notably, a number of these assignments involved protein nuclei that were already, but incorrectly, assigned to other resonances. Our proxy residues method thus allowed us to "correct" mistaken assignments. Although only a small fraction of the proxy atoms yielded new assignments, the inclusion of proxy residues resulted in a better restraint list. As a consequence, the quality of the structures significantly improved, as judged, for example, by the Ramachandran quality (Table 1). Note that these are not the final structural statistics, since further completion of the assignment and structural refinement were performed.

Test Cases. (1) Assignment of Aromatic Rings. The assignment of aromatic ring side-chain resonances for proteins is notoriously problematic. In many cases, these are assigned only after analysis of NOE spectra, on the basis of the expected intra-residual NOE correlations with the β-protons. However, if multiple aromatic rings are in close proximity, this method can easily lead to errors.

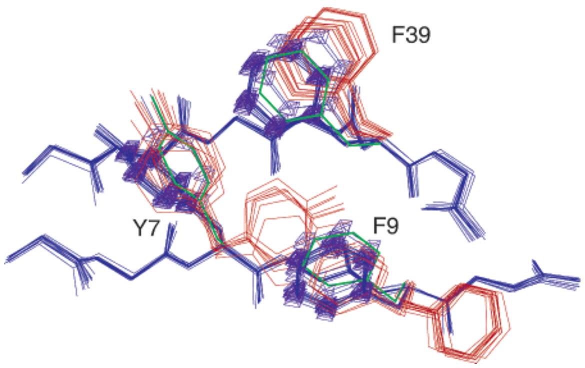

Figure 4. Structures of the IIB^{Chb} aromatic cluster consisting of side chains of Y7, F9, and F39, obtained with proxy residues representing the aromatic rings are shown in blue (the heavy atoms of the proxy residues are shown as tetrahedral shapes). For comparison, the side chains of the aromatic residues calculated with the same incomplete assignments but without proxy residues are shown in red. A representative of the structures calculated with full assignments (and without proxy residues) is shown in green.

The use of proxy residues should simplify the assignments of aromatic residues, for a number of reasons. In the first place, aromatic rings often show a lot of NOE correlations, due to their size and because they are often located in the core of the protein. So the position of these structurally important residues should, in general, be well determined by the NOE restraints. Second, the number of possible assignments is usually well defined and limited (i.e., only the aromatic side chains). Often the type of side chain can be determined on the basis of the chemical shifts. So the number of possible solutions can be effectively reduced with IDRs. Finally, all the available NOE information can be used before an initial fold is known, which is likely to make the initial fold more reliable. Because information on which aromatic resonances are part of the same spin system is relatively easy to obtain (e.g., from H[C]CH TOCSY or filtered 2D ^{1}H TOCSY spectra), we have used assignments to complete aromatic ring proxies to take advantage of this.

We tested this approach on NOESY data for enzyme IIB^{Chb}, which contains a cluster of aromatic residues (Y7, F9, F39). The aromatic side-chain resonances of these residues were difficult to assign manually because the presence of several *inter*-residual Hβ-to-aromatic ring NOEs complicated the assignments based on *intra*-residual correlations. We removed the assignments for resonances of the three rings and reassigned them to the DUR proxy ring constructs shown in Figure 3.

As can be seen in Figure 4, treating these resonances as proxy rings easily resolves this problem. All three proxy rings are closely superimposed on their corresponding protein rings and have correct H$^{\delta}$/H$^{\epsilon}$/H$^{\zeta}$ orientation (not shown). The correct chemical shift assignments can easily be deduced from these structures. When the intra-ring connections are left undefined by assigning each single aromatic resonance to a DUF-type proxy, designed to deal with single aromatic resonances (Figure 3), still 7 of the 8 proxies (representing 2 × 3 phenylalanine resonances + 2 tyrosine resonances) end up closest to the atoms corresponding to the correct assignments.

(2) Random Missing Assignments. As a general test of the performance of the proxy residue method, we randomly deleted various fractions of the assignment of proton−heteroatom pairs

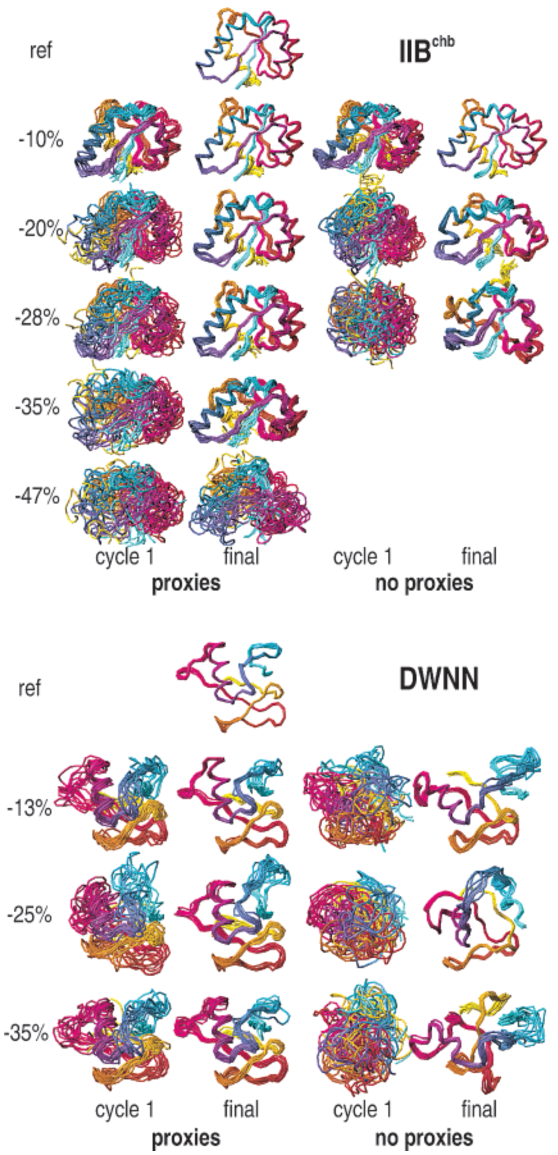

Figure 5. Comparison of structures calculated with and without proxies for IIB^{Chb} (top) and DWNN (bottom). The reference structures (labeled "ref") are the final water-refined structures, calculated with complete assignments, without the use of proxies. The rows of structures show the 10 lowest energy structures of the first and the final cycles of structures calculated with and without the use of proxies, for various amounts of missing assignments (percentage indicated on the left). Assignments were randomly replaced by proxy residue assignments, except in the bottom row for DWNN (−35%): for those runs all side-chain (except β) resonance assignments were replaced by proxy assignments. All structures are shown in the orientations obtained after superposition onto the reference structure.

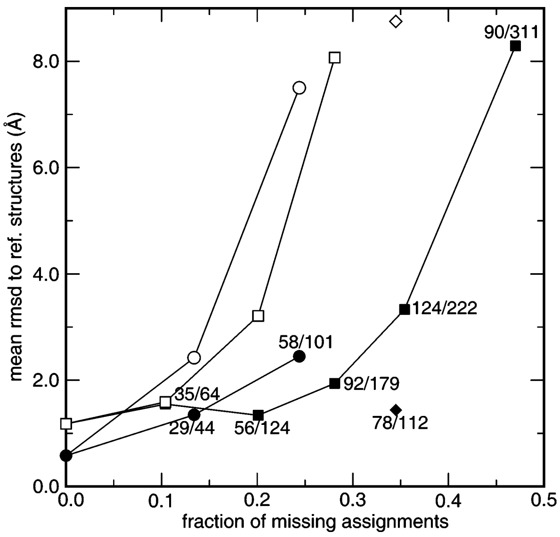

Figure 6. Comparison of the accuracy and convergence of automated assignment and structure calculations in CANDID with (filled symbols) and without (open symbols) proxy residues for various amounts of deleted assignments. Connected symbols show results for randomly deleted assignments from DWNN (circles) and IIB^{Chb} (squares). Diamonds show results for DWNN using only backbone and β resonance assignments. The rmsd values are avarge pairwise rmsds between the cluster of calculated structures and the ensemble of reference structures. The numbers next to the filled symbols indicate the convergence ratios (number of converged proxies/total number of proxies). Proxy residues were considered converged if their heteroatoms have a mean displacement less than 4 Å after superposition of the structures on the backbone atoms of residues 8−80 for the DWNN protein and residues 3−103 for enzyme IIB^{Chb}.

for IIB^{Chb} and DWNN. For both proteins, the use of proxy residues leads to structures that are much closer to the reference structure compared to that obtained when no proxy atoms are used (Figure 5). When CANDID is used with proxy residues, it still gives structures that are reasonably close to the reference structure when 25−30% of the resonances are missing. Without proxies, this is only the case at maximally 10−15% missing resonances. This is in accordance with earlier reports that the assignment completeness needs to approach 90% for CANDID to perform reliably.[7] Figure 5 shows the clusters of structures for DWNN produced in the first and final CANDID cycles. Up to 35% missing assignments, CANDID including proxy residues produced a correct fold in the first cycle, which is a prerequisite for reliability of the algorithm. Without proxy residues, CANDID failed to converge in the first round, which prevented it from subsequently finding the correct fold.

(3) Missing Side Chain Assignments. To mimic a situation typical for a protein after the backbone assignment stage, we performed a CANDID run with the DWNN protein, in which all side-chain resonance assignments were replaced by proxy assignments, except for the β resonances. This corresponds to 34.5% missing assignments.

When proxies are used, the correct fold is reproduced quite closely: the average pairwise backbone rmsd from the reference structures is 3.25 Å for the cluster of structures in the first CANDID cycle and 1.37 Å (2.43 Å for all heavy atoms) for the final structures (filled diamonds in Figure 6). The average pairwise rmsd within the cluster of final structures is only slightly smaller (1.27 Å), indicating that the cluster of reference structures falls largely within the structures obtained in the final run. It is interesting to note that the average pairwise rmsd to the mean for the first cycle is 2.39 Å, and the average target function is 86 Å.[2] Both values are well within the criteria indicative of proper convergence, determined by Jee and Güntert.[7] In the absence of proxies, the CANDID run fails to converge and results in structures almost 9 Å away from the reference.

Convergence. In the successful calculations, on average about 50−70% of the proxy residues converged to a rather well-defined location, with a mean displacement after superposition of less than 4 Å, as indicated in Figure 6. To obtain the correct fold, CANDID selects in the initial structure calculations (cycle

[6] Laskowski, R. A.; Rullmannn, J. A.; MacArthur, M. W.; Kaptein, R.; Thornton, J. M. *J. Biomol. NMR* **1996**, *8*, 477−486.

[7] Jee, J.; Güntert, P. *J. Struct. Funct. Genomics* **2003**, *4*, 179−189.

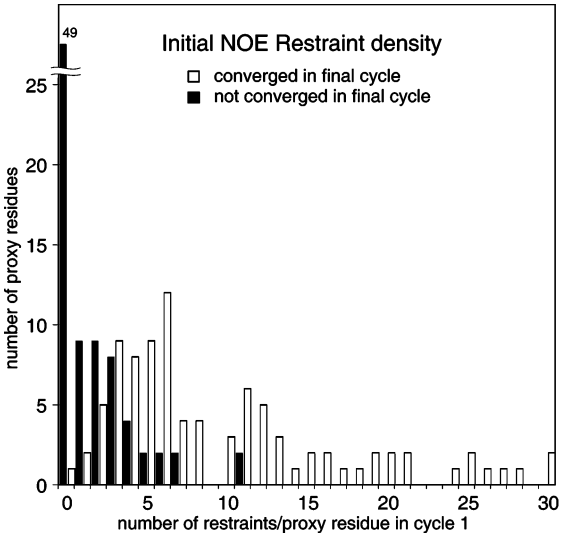

Figure 7. Comparison of the initial restraint density for proxy residues that have (white bars) and have not (black bars) converged to a mean heavy-atom displacement of less than 4 Å in the final CANDID structures. The data were taken from the runs for IIB^{Chb} with 28.1 missing assignments. Note that proxy residues can have non-integer number of restraints due to the use of ambiguous restraints. For example, there are 49 proxies that did not converge in the final round with less than one restraint in the first round.

1) only restraints that are sufficiently supported by other covalent or NOE contacts. In subsequent cycles, this network anchoring becomes less important and structure-based filters take over. So if restraints for a proxy atom show only low network anchoring support, no restraints for that particular proxy might be accepted in the first cycle. As a result, that proxy will not converge to a stable position in the structure.

We analyzed how many restraints are needed for a proxy in the first cycle in order to show convergence in the final cycle. For the data shown in Figure 7, 87% of the proxy residues with four or more NOE restraints in the first CANDID cycle converged. Proxies with less than four restraints generally did not converge in the first cycle. Although about 30% of the proxy residues did not converge (in this case corresponding to 14% of missing resonances), the impact of these missing resonances is limited because they will usually represent the resonances for which only little NOE data is available.

It is interesting to see whether the generated structures can be used to deduce additional resonance assignments. For the DWNN run with missing side chain assignments, we examined whether we could find criteria that allow for reasonably safe assignment of the converged proxy residues. Because the distances between proxy and protein atoms that are used in the IDRs carry information about possible assignments, we used those to evaluate putative assignments. We consider the possible assignment that is closest to the proxy as the best assignment for a given structure.

We found that an assignment can be made with reasonable confidence if at least half of the proxy atom−protein atom distances, as defined in the IDR, are smaller than 0.3 Å. This ensures that (1) the most likely assignment is found in at least half of the structures and (2) that assignment will not be chosen for another proxy, because of the repulsive potential between the proxy residues. Of the 74 proxy residues we analyzed using this criterion, 50 yielded an assignment, of which only one was different from the manual assignments. Of the remaining 24 that were not assigned, in 20 cases the atoms corresponding to the correct assignment were, on, average within 4 Å, and in 15 cases the most likely assignment was still equal to the manual assignment. Although this initial analysis shows promising results, we do not want, at this stage, to claim any success rate since more work will be needed to define robust assignment criteria and automate the proxies assignment procedure.

Computational Aspects. In general, the computational burden of adding a proxy residue is limited, because their number will usually be small. In CYANA's torsion angle dynamics (TAD) implementation, this is slightly worse than would be the case in Cartesian space because of the linkers that need to be added between proxy residues. This is, however, not an intrinsic requirement of TAD, but only of the current implementation of TAD in CYANA, where the protein is described as a tree with only one root. In principle, one could define the proxy residues to have separate roots; no linkers would then be needed, which would be computationally more efficient.

The use of IDRs and repulsive restraints results in a considerable number of interactions that have to be evaluated during each step of the simulated annealing protocol. For example, in the DWNN run with only backbone and β assignments, the 107 IDRs that were added had an average multiplicity of 80; additionally, 5671 lower-bound restraints between the heavy atoms of different proxy residues were used. Nevertheless, the use of these extra restraints led to an increase of only 18% in the computation time needed for the first CANDID cycle.

To prevent possible problems with sampling and convergence of the proxy residues, we increased both the number of structures calculated in the first cycles from 100 to 200, and the number of simulated annealing steps from 10 000 to 25 000. As this still yielded a total run time of somewhat less than 3 days on a modern dual-CPU PC, we have not tried, at this time, to optimize the procedure for minimal run time.

Comparison with Other Approaches. Our proxy atom method differs from other methods that use anonymous atoms, such as ANSRS and CLOUDS, in the fact that it uses a combination of anonymous atoms for the unassigned resonances and a conventional model of the protein under study at the same time. It thus circumvents the problem that the assignments have to be deduced from the coordinates of the anonymous atoms: during the entire procedure, a complete model of the protein is present, whose coordinates should already reflect both the chemical knowledge and the information that is present in the NOE and identity restraints.

One important advantage of representing unassigned resonances by proxy residues is that more complete NOE information can be used from the first iterations of structure calculations. This reduces the risk of finding an incorrect fold and of possible subsequent incorrect resonance assignments. Because of the sensitivity of the conventional ARIA and CANDID protocols to missing resonance assignments, one is often tempted to make premature assignments. With proxy residues, one has the choice to leave the resonance unassigned, treat it with a proxy residue, and express the assignment ambiguity as an IDR. In this way, one can prevent incorrect assignments and still be confident that as much as possible of the available information is used.

Conclusions

With the introduction of proxy residues, we have presented a simple modification of widely used structure calculation protocols that significantly improves their robustness with regard to missing resonance assignments. By assigning the unassigned resonances to proxy residues, the information contained in NOEs corresponding to unassigned resonances is cast in a form that can be used readily by structure calculation protocols like ARIA and CANDID. The extra structural information that is gained by using proxy residues obviously helps to find the correct three-dimensional fold of a protein. Furthermore, because it is possible to have an assignment for every observed resonance, either to a protein atom if the resonance is assigned or to a proxy atom if it is not assigned, the probability of generating erroneous restraints because of missing assignments is reduced.

Our results show that the use of proxy residues makes it possible to find missing as well as incorrect assignments. The perfomance depends on the amount of available NOE cross-peaks and on the IDRs that drive the proxy residues close to atoms that are likely to represent correct assignments. We have achieved already quite satisfactory results by creating IDRs for relatively general classes such as amide, aromatic, and methyl resonances. In principle, the IDRs can be carefully tuned to reflect the assignments that are possible for a given resonance, using information from chemical shifts, the number of neighboring carbon nuclei, and the presence or absence of assignments and their reliability.

Our calculations with various amounts of missing assignments show that the use of proxy residues can make the CANDID protocol much less sensitive to missing assignments: while the conventional protocol becomes unreliable when the number of missing assignments exceeds 10%, we are still able to generate structures that are very close to the (correct) reference structures in cases where between 20 and 35% of the resonances are missing, as well as when only backbone and β resonance assignments are present. This opens the possibility to start structure calculations right after assignment of the backbone and β resonances and should significantly impact NMR-based structural genomics initiatives by both reducing the effective structure determination time and increasing the reliability of the structures generated in the initial stages.

Materials and Methods

We have constructed CNS and CYANA library definitions for the proxy residues shown in Figure 3. These include proxy residues for CH, CH_2, CH_3, NH, and NH_2 resonances, for complete aromatic ring systems, and for single aromatic resonances that allow a proper treatment of degenerate resonances due to ring flipping. The only information that has to be supplied to the algorithm is a sequence that contains the proxy residues and a chemical shift list that contains the assignments for the proxy spin systems. From the point of view of the host algorithm (ARIA or CANDID), there is no difference in treatment of the proxy residues compared to the regular protein residues. Protocols for calculations with proxy residues were implemented in ARIA version 1.2 and CYANA version 2. In the latter, the CANDID implementation as found in the standard macro ‘noeassign.cya’ was used.

Using proxy residues in torsion angle dynamics (TAD), as currently implemented in CYANA, is slightly more complicated than in Cartesian space: the proxy residues should contain anchors to attach them via linker residues to the protein and to each other. Although this is not a fundamental requirement of TAD, this is a limitation of the current TAD implementation in CYANA, where the molecular topology is described in terms of a tree with a single root node. In this work, we have used relatively short linkers of five LL5 flanked by two LL2 linker residues from the standard CYANA library.

The force field parameters are defined in such a way that the proxies do not have any nonbonded (van der Waals + electrostatic) interactions with the protein, as they should be able to occupy the same positions as the protein atoms they represent. Distinct proxy atoms, however, should not have the same assignment and thus should not take the same position. This requirement was implemented by defining lower-bound restraints of 2.0 Å between the heteroatoms of different proxy residues.

For the generation of NOE-based distance restraints, the proxy atoms are treated in the same way as the assigned protein atoms with regard to network anchoring, restraint combination, selection and rejection of assignment possibilities, or whatever methods are provided by the host algorithm, in this case ARIA/CNS[1,8,9] and CANDID/CYANA.[10,2] No parameters were adjusted except for some that influence the sampling, like the number of simulated annealing steps and the number of calculated and/or selected structures.

Additional information about the identity of a proxy residue that might be available, for example, from chemical shifts was cast in the form of IDRs, ambiguous distance restraints with a very short upper bound (0.01 Å), from one proxy atom to all atoms that are potential candidates for the associated unassigned resonance. To limit the amount of restraints that have to be evaluated during structure calculations, IDRs were defined only between heteronuclei of the proxy and protein residues.

The robustness of the CANDID protocol with and without proxy residues with regard to missing assignments was tested on two proteins for which almost complete assignments are available: the 86-residue DWNN protein[11] (99.0% completeness) and the 106-residue enzyme IIB^{Chb} [12] (99.5% completeness). Chemical shift lists with varying degrees of randomly missing assignments were produced by replacing the assignments of proton−heteroatom pairs by proxy residue assignments with a replacement probability of $P = 0.1$, 0.2, 0.3, 0.4, and 0.5, respectively. IDRs were defined to restrict assignment to broad classes like aromatic, aliphatic, methylene, amide, and side-chain amide NH_2 groups. Peaks from the following NOE spectra were used: for the DWNN protein, 2D NOESY (2514 peaks), 2D NOESY in D_2O (1510 peaks), ^{15}N NOESY-HSQC (871 peaks), and ^{13}C NOESY-HSQC (2487 peaks); for enzyme IIB^{Chb}, ^{15}N HSQC-NOESY (1279 peaks), ^{13}C HSQC-NOESY (3452 peaks), and 2D NOESY (774 peaks). No pre-existing NOE assignments were used during the calculations.

As an additional test case, we replaced all side chain assignments except the β resonances by proxy residue assignments, restricting the IDRs to unassigned atoms.

Software Availability

The PROXIES libraries for use in CNS and CYANA, together with example scripts and demo data, can be found on the Internet at http://www.nmr.chem.uu.nl/proxies.

Acknowledgment. This project was supported by two fifth framework EU grants: FIND-Structure, contract QLG2-CT2002-00988, and SPINE, contract QLG2-CT-1999-01003.

JA058504Q

(8) Linge, J. P.; O’Donoghue, S. I.; Nilges, M. *Methods Enzymol.* **2001**, *339*, 71−90.
(9) Brunger, A. T.; et al. *Acta Crystallogr. D: Biol. Crystallogr*. **1998**, *54*, 905−921.
(10) Güntert, P.; Mumenthaler, C.; Wüthrich, K. *J. Mol. Biol.* **1997**, *273*, 283−298.
(11) Pugh, D. J. R.; AB, E.; Faro, A.; Lutya, P. T.; Hoffmann, E.; Rees, D. J. G. *BMC Struct. Biol.* **2006**, *6*, 1−12.
(12) AB, E.; Schuurman-Wolters, G. K.; Nijlant, D.; Dijkstra, K.; Saier, M. H.; Robillard, G. T.; Scheek, R. M. *J. Mol. Biol.* **2001**, *308*, 993−1009.
(13) Doreleijers, J. F.; Mading, S.; Maziuk, D.; Sojourner, K.; Yin, L.; Zhu, J.; Markley, J. L.; Ulrich, E. L. *J. Biomol. NMR* **2003**, *26*, 139−146.