Published in final edited form as:

J Magn Reson. 2012 August ; 221: 76–84. doi:10.1016/j.jmr.2012.05.020.

Parameterization of Solvent-Protein Interaction and Its Use on NMR Protein Structure Determination

Yu Wang¹, Charles D. Schwieters², and Nico Tjandra^{1,*}

¹Laboratory of Molecular Biophysics, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892, U. S. A

²Division of Computational Bioscience, Center for Information Technology, National Institutes of Health, Bethesda, MD 20892, U. S. A

Abstract

NMR structure determination is frequently hindered by an insufficient amount of distance information for determining the correct fold of the protein in its early stages. In response we introduce a simple and general structure-based metric that can be used to incorporate NMR-based restraints on protein surface accessibility. This metric is inversely proportional to the sum of the inverse square distances to neighboring heavy atoms. We demonstrate the use of this restraint using a dataset from the water to protein magnetization transfer experiment on the protein Bax and the solvent paramagnetic relaxation enhancement experiment on the protein ubiquitin and Qual homodimer. The calculated solvent accessibility values using the new empirical function are well correlated with the experimental data. By incorporating an associated energy term into Xplor-NIH, we show that structure calculation with a limited number of additional experimental restraints, improves both the precision and accuracy of the resulting structures. This new empirical energy term will have general applicability to other types of solvent accessibility data.

Keywords

protein structure; structure refinement; solvent accessibility; solvent PRE

Introduction

Protein structure determination by high resolution Nuclear Magnetic Resonance (NMR) has evolved greatly in the recent years. This NMR procedure has traditionally relied on NOE based distance restraints and dihedral restraints based on scalar or J-coupling to determine a protein structure. These restraints contain only local structural information. Newer classes of structural restraints have emerged that provide longer range or even distance independent information. For instance, residual dipolar coupling [1; 2; 3; 4], residual chemical shift anisotropy [5; 6; 7; 8], ratios of NMR relaxation times [9; 10; 11; 12] and cross-correlated relaxation rates [13; 14] can define bond vector orientation in a fixed common frame, or can describe the relative angles between different bond vectors. In addition to these new classes of restraints, pseudo contact shifts [15; 16] and paramagnetic relaxation enhancement (PRE) [17; 18; 19] also offer long-range orientation and distance restraints relative to the

*Corresponding author: Nico Tjandra, tjandra@nhlbi.nih.gov, Phone: (301)-402-3029, Fax: (301)-402-3405.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

paramagnetic center. Inclusion of all or some of these restraints in a single domain structure calculation has been shown to improve the precision and, more importantly, the accuracy of the resulting structures. They also aid in determining the relative orientation of protein domains or subunits in a complex.

The NMR structure determination process generally is initiated by first obtaining a low-resolution structure of the protein with the correct fold using the minimum NMR data sets. This initial structure is further iteratively refined by adding more NMR restraints into the calculation. Arriving at the first correctly folded structure can be a bottleneck in the process due to the difficulty of direct interpretation of NMR data and the nature of the energy landscape defined by various experimental and knowledge-based terms. The above-mentioned newer restraints can help overcome this problem to a certain degree and improve the convergence of the initial fold-determining calculation. Another recurring problem for NMR structure determination is the lack of overall compactness of the structure. One proposal to address this was to use a radius [20] or volume [21] of gyration potential term during the structure calculation to keep the structure compact. An alternative is to use small angle X-ray scattering data that can provide structure information in place of the radius of gyration [22]. Ryabov *et al.* has also shown that implementing the rotational diffusion anisotropy information into the shape of the protein can improve the quality of the refined structures [23; 24]. In addition, Nilges *et al.* also has proposed carrying out the final structure refinement in the presence of water molecules to regularize the protein structure and improve its overall packing [25]. However, easily obtained experiment-based restraints are sorely needed for initial NMR structure determination.

NMR has also been a useful tool to study protein solvent interactions. For instance, hydrogen-deuterium exchange studies by NMR provide a measure of protection from the solvent that signifies stability at various sites in the protein. In solution, the PRE by paramagnetic solvent probe has been shown to depend on the water accessibility of residues and provides proton-solvent distances [26; 27; 28; 29]. Confinement of a protein within a reverse micelle also provides the residue-specific information of protein-solvent interactions [30]. In addition, water-protein NOE or ROE measurements specify water lifetime and accessibility at various residues in a protein [31; 32; 33]. All of these experiments can easily identify which residues are solvent exposed, and thus are on the surface of the protein, and which are buried in the core of the protein. Interestingly this information has not been commonly used as structural restraints. In this work we propose a method to take advantage of this information using an empirical relationship between a simple measure of protein surface accessibility and observed solvent protein interaction. Specifically, we evaluated data sets from an experiment where water magnetization was transferred to the protein as well as data sets from paramagnetic solvent PRE experiment. Both experiments are easy to conduct and it is straightforward to analyze the data. We show that this approach improves the convergence and accuracy of structure calculations of the protein Bax, ubiquitin as well as the Qual homodimer using a reduced data set simulating an early stage of structure determination and cases where large numbers of NOEs are difficult to obtain. In addition this term should be generally applicable to various other measures of solvent protein interactions.

Materials & Methods

Sample preparation

The ^{15}N -labeled protein Bax was expressed and purified as described previously [34]. While the ^{15}N , ^2H -labeled ubiquitin was prepared as described by Lazar *et al.* [35]. The ^{15}N -labeled protein Bid was expressed in *Escherichia coli* BL21 (DE3) cells transformed with the pET15b-Bid plasmid and purified using Ni^{2+} affinity column followed by ion-exchange

(Q) column. Samples for NMR experiments contained 0.5 mM Bax and 0.8 mM Bid in 50 mM phosphate buffer at pH 6.7, with 2mM DTT, 0.02% NaN₃, while the ubiquitin NMR sample contained 0.5 mM ²H, ¹⁵N labeled ubiquitin in 20 mM Na-acetate buffer at pH 5.2 and 0.02% NaN₃.

Measurements of the water and protein interaction

The water to protein magnetization transfer data was acquired by using the enhanced protein hydration observed by gradient spectroscopy (e-PHOGSY) with NOE transfer step pulse sequence with no modification [36; 37]. The selective water inversion pulse used was a 50 ms sinc shaped pulse. The ¹H, ¹⁵N-HSQC spectrum was acquired as a reference. The water accessibility value reported is the intensity ratio between the water to protein magnetization transfer and the reference HSQC spectra. Both spectra were acquired with 1024 and 192 complex data in the ¹H and ¹⁵N dimension, respectively. The ¹H carrier frequency was set at the water frequency, while the ¹⁵N carrier was at 115 ppm. The reference HSQC spectrum for the protein Bax was acquired with 32 scans for each t₁ increment, while the water to protein magnetization transfer spectrum was collected with 64 scans. The transfer mixing time used was 70 ms. Both experiments were acquired at 32 °C on a Bruker DMX 600 MHz spectrometer equipped with cryogenic probe and Z-pulsed field gradient. For the protein Bid, the same experiments were carried out with 32 and 64 scans for the reference HSQC and water to protein magnetization transfer spectra at 25 °C, respectively. For the deuterated ¹⁵N-ubiquitin, 16 and 64 scans were acquired for the reference HSQC and water to protein magnetization transfer spectra at 25 °C, respectively. NMR data were processed using NMRPipe [38] and analyzed with NMRView [39] and PIPP [40].

Results

The measurement of water and protein interaction

The e-PHOGSY detection pulse sequence was applied to detect the interaction between water and the protein amide-proton [36; 37]. The uniformly ¹⁵N labeled apoptotic protein Bax was used to carry out the measurement. Meanwhile, a 2D ¹H-¹⁵N HSQC spectrum was acquired and used as a reference. An overlay of the spectra of Bax from these two experiments is shown in Figure 1A. The variation in the strengths of the interaction between water and the protein amide protons could clearly be observed as modulation in their signal intensities compared to the HSQC spectrum. Residues with overlapping resonances were taken out of the analysis. A total of 156 out of 184 non-proline residues could be analyzed. In order to eliminate the discrepancy from the rapidly exchanging hydroxyl protons, residues (Ser, Thr and Tyr) possessing the hydroxyl group were omitted. The ratios in intensities of 138 residues from these two different spectra are plotted as a function of residue number as shown in Figure 1B. Qualitatively, this water-protein interaction data differentiates the exposed and buried residues. For instance the termini and long exposed loops of the protein show high intensity ratios (> 0.07), while buried residues show low ratios (<0.05). Some residues in the same helix can also show differential exposure to solvent (for instance, helices α₂ & α₃).

Finding an appropriate empirical function to describe solvent-protein interaction

The test for an empirical function was carried out using the previously published solution structure of Bax (PDB: 1F16). Exposed residues that are close to the protein surface tend to have fewer heavy atoms within a certain distance range, while buried residues will have more (Supplemental Figure 1). Following this observation, we chose this simple metric as a stand-in for solvent accessibility S_{Acc}^c (Å²):

$$S_{Acc}^c = \left(\sum_{i=1}^N \frac{1}{r_i^2} \right)^{-1} \quad (1)$$

where r_i is the distance (Å) between an amide proton and heavy atom i . The sum is over all heavy atoms within a cutoff distance R_c . Heavy atoms of the residue to which the amide-proton belongs to are excluded from this calculation. This is done to avoid residue type dependence. For instance a residue on the surface with bulky, but mobile, side chain may artificially show significant solvent protection compared to non-bulky residue, since a static structure is used in the calculation. This effect, however, is not large. In the case of the protein Bax, excluding the intra-residue heavy atoms improves the correlation by 1% (0.80 to 0.81). The calculated (using the lowest energy Bax structure) values for Eq. (1) are plotted against the experimental intensity ratios, corresponding to the different strength of water-protein interaction, and show a good linear correlation with a correlation coefficient $R=0.80$ (Figure 2). Using the 10 lowest energy NMR structures of Bax, the calculated values correlated with the experimental data with an average correlation coefficient R of 0.83 and standard deviation of 0.02. Other choices for power of distance in Eq. (1) result in worse agreement.

The linear regression fitting of the empirical function of the 10 lowest energy Bax structures (PDB ID: 1F16) to the water-protein interaction data on all the identified Bax residues resulted in the empirical conversion of intensity ratios (I_{ratio}) to an effective surface area S_{Acc}^m :

$$S_{Acc}^m = 2.06 I_{ratio} + 0.06 \quad (2)$$

The fitting for the 10 lowest energy Bax structures showed the slope and offset values in the range of 1.83 Å² to 2.30 Å² and 0.04 Å² to 0.07 Å², respectively. The distance from each amide proton to all heavy atoms was calculated using a cutoff radius $R_c = 20$ Å. It is important to note that a cutoff range exceeding 20 Å showed very small improved agreement to the experimental values because long distances (r_i) contribute much less to the final values of the empirical term. For the case of Bax, the correlation coefficient R improved from 0.805 to 0.809 when $R_c = 80$ Å was used. The choice of the distance cutoff R_c is a compromise between computational cost and structural improvement. Using Eq. (2), the experimental water-protein interaction's intensity ratios could be converted into restraints and used for the structure calculations.

Bax structure refinement

In order to test the effect of the new term on structure calculation, the experimental water-protein interaction data were used in a simulated annealing calculation to determine the structure of Bax. The solvent accessibility potential was implemented as a new energy term in Xplor-NIH (<http://nmr.cit.nih.gov/xplor-nih/>) [41], given by:

$$E_{Acc} = k(S_{Acc}^c - S_{Acc}^m)^2, \quad (3)$$

where S_{Acc}^c is the calculated solvent accessibility which is given by Eq. (1). During the calculation, the force constant k was increased from 0.2 kcal/Å² to varying final values to test for convergence. This was done in concert with cooling the system from an initial temperature of 3500 °K to a final value of 25 °K. All structure calculations were done using extended random coil starting structures. All original dihedral and hydrogen bond restraints

along with different sets of NOE distance restraints were used throughout the structure calculations.

The solvent accessibility potential was developed to distinguish buried and exposed residues. Therefore inclusion of the new potential is not expected to significantly improve the overall quality of the refined structures when a large number of other experimental NMR data are available. However this new potential should assist in determining the overall fold of the protein during the initial stage of the structure determination when relatively few restraints are available. To test the convergence efficiency using this potential to determine the overall fold of the protein structure starting from a random coil conformation, a total of 500 non-intraresidue $^1\text{H}_\text{C}$ - $^1\text{H}_\text{C}$ NOE restraints were randomly chosen from the 4689 total NOEs [34]. Among the 500 randomly chosen NOEs, there are 197 long-range (greater than 5 residues apart), 135 medium-range (2 to 4 residues apart), 21 sequential and 147 intraresidual NOEs. Dihedral and hydrogen bond restraints were included, but the RDCs were omitted. The calculations with and without the new solvent accessibility potential were performed. A total of 50 structures were calculated, from which the 5 lowest energy structures were chosen as a representative ensemble. To check for accuracy the resulting structures were compared to the lowest energy structure of Bax (PDB ID: 1F16). The optimal force constant for the new potential used was $70 \text{ kcal}/\text{\AA}^2$, based on the structure precision (convergence), accuracy and the energy contribution from the new potential. As would be the case in a *de novo* structure calculation we assumed the optimal slope and offset values were unknown. Therefore, a coarse grid search of the slope was carried out by varying its value from $1.4 \text{ \AA}^2$ to $2.4 \text{ \AA}^2$ while keeping the offset fixed. This was followed by a search for offset by varying their values from $0.0 \text{ \AA}^2$ to $0.12 \text{ \AA}^2$ while keeping the slope fixed at its optimal value. The slope and offset for structures with minimum energies were shown to be at $1.6 \text{ \AA}^2$ and $0.06 \text{ \AA}^2$, respectively (Figure 3A&B). To further optimize them, structures are calculated using these slope and offset values. These structures were then used to fit the experimental intensity ratios to give a new set of slope and offset. This iterative operation was performed three times to arrive at the final optimized parameters (slope of $1.69 \text{ \AA}^2$ and offset of $0.09 \text{ \AA}^2$) with the lowest energy structures. The optimal slope and offset values found in this manner are close to those reported above using the original Bax structure. The convergence and accuracy statistics of the calculated structures with or without the new potential (E_{Acc}) are summarized in Table 1. An obvious improvement in structure convergence was observed for the 5 lowest energy structures when the new potential was used. The correlation between calculated and measured water-protein interaction intensity ratios improved from a correlation coefficient R of 0.79 to 0.86 (Figure 3C&D). This improvement is accompanied by an increase of precision, as measured by the backbone RMSD (calculated for residues 15–189) of the 5 lowest energy structures, of the calculated structures from $3.12 \pm 0.97 \text{ \AA}$ to $1.59 \pm 0.21 \text{ \AA}$ when the new potential is used. The structure comparison was illustrated in Figure 3E&F. The backbone of calculated structures using the new potential is $2.77 \pm 0.04 \text{ \AA}$ away from the published Bax structures (PDB ID: 1F16), while without using the potential it is $3.93 \pm 1.40 \text{ \AA}$ away.

To test for compatibility of the new energy term with the other terms, calculations were carried out with full restraints, including all distance restraints and residual dipolar couplings as previously described [34]. Similar to the calculation described above, using force constant of $70 \text{ kcal}/\text{\AA}^2$ the calculated structures showed some improvement in precision with the addition of E_{Acc} (Table 1). For instance, the backbone RMSD for the 5 lowest energy structures was $1.27 \pm 0.16 \text{ \AA}$ and $1.16 \pm 0.17 \text{ \AA}$ for structures calculated without and with the solvent accessibility potential, respectively. The correlation between calculated and measured water-protein interaction intensity ratios values improved from a correlation coefficient R of 0.82 to 0.85. To check whether the new potential is pulling the structure to a different energy minimum, the calculated structures were compared to the original lowest

energy structure. The difference between the accuracy of the two calculations is not significant. The backbone structures calculated with and without the new energy term are 1.92 ± 0.23 Å and 1.95 ± 0.25 Å away from the original structure, respectively. This result demonstrates that inclusion of this new potential term in the context of full experimental restraints did not result in structures that were completely different from the ones calculated without the new potential, but can still improve the precision of the structures.

To show that this method is generally applicable we evaluated the new empirical solvent accessibility term against experimental data from other proteins. For instance water to protein magnetization transfer data obtained from the protein Bid [42] (PDB ID: 2BID) could be correlated to the surface accessibility values and yield a linear correlation coefficient of 0.79 (Supplemental Figure 2A). The linear regression of the correlation plot for Bid resulted in a slope for Eq. (2) of 1.72 Å² with the offset value of 0.03 Å². In addition we also measured water to protein magnetization transfer data for ¹⁵N, ²H-labeled ubiquitin. The measured intensity ratios are correlated linearly with the calculated water accessibility, using the lowest ubiquitin structure (PDB ID: 1D3Z) [6], with a linear correlation coefficient of 0.88 (supplemental Figure 2B). The linear regression of the correlation plot for ubiquitin resulted in a relationship of $S_{Acc}^m = 1.62 I_{ratio} + 0.07$.

Ubiquitin structure refinement

In order to test the applicability of the new term with another type of water accessibility data, previously published solvent PRE data of ubiquitin [28] were tested. Values calculated using Eq. (1) for the lowest energy NMR structures of ubiquitin (PDB ID: 1D3Z) correlated well with the experimental solvent PRE values (Figure 4A) with correlation coefficient R of 0.86 (0.84 ± 0.03 for 10 lowest energy ubiquitin structures). Furthermore, the experimental solvent PRE (Γ_{PRE}) can be fit to a linear empirical function similar to Eq. (2). Using the 10 lowest energy ubiquitin structures with the a cutoff radius R_c of 20 Å, the following linear relationship to the effective surface area S_{Acc}^m was found:

$$S_{Acc}^m = 0.353 \Gamma_{PRE} + 0.128 \quad (4)$$

Based on Eq. (4), the experimental solvent PREs ($s^{-1} \cdot mM^{-1}$) ($0 < \Gamma_{PRE} < 1$) could be converted into restraints and used for the structure calculations.

By applying a calculation protocol similar to that used for Bax, we tested the convergence efficiency using the new solvent accessibility potential for ubiquitin [28]. Starting from a random coil conformation, a total of 202 NOE restraints were randomly chosen from the 2727 total NOEs [6]. Dihedral and hydrogen bond restraints were included, but the other restraints were omitted. The calculations of protein ubiquitin with and without the solvent PRE restraints were performed. Using the same optimization protocol as in the calculation of Bax, the optimal force constant found was 300 kcal/Å² and the optimal slope and offset were 0.396 Å²·s·mM and 0.096 Å², respectively. The convergence and accuracy statistics of the calculated structures are summarized in Supplemental Table 1. An obvious improvement in structure convergence was shown for the 5 lowest energy structures when the new potential was used. The correlation between calculated and measured intensity ratios values improved from a correlation coefficient R of 0.79 to 0.90 without and with the new potential, respectively (Supplemental Figure 3). This improvement is accompanied by an increase of precision, as measured by the backbone RMSD (calculated for residues 2–71) of the 5 lowest energy structures, of the calculated structures from 1.76 ± 0.31 Å to 1.24 ± 0.20 Å when the new potential is used. The backbone accuracy of calculated structures using the new potential is 2.01 ± 0.14 Å away from the published ubiquitin structures (PDB ID: 1D3Z),

compared to $2.38 \pm 0.40 \text{ \AA}$ without using the potential. In comparison, calculations using the water-protein interaction data also improved the ubiquitin structures to $1.10 \pm 0.08 \text{ \AA}$ in precision and $2.03 \pm 0.14 \text{ \AA}$ in accuracy (Supplemental Table 1).

Qua1 homodimer structure refinement

The new solvent accessibility term is expected to be useful in determining structures of protein complexes. In order to demonstrate this, we tested a structure calculation for the protein Sam68 Qua1 homo-dimerization domain. We applied the new solvent accessibility potential using the previously published solvent PRE data of the Qua1 homodimer [43]. The correlation of values calculated using Eq. (1) for the lowest energy NMR structures of Qua1 homodimer (PDB ID: 2XA6) correlated well with the experimental solvent PRE values (Figure 4B) with correlation coefficient R of 0.82 (0.84 ± 0.01 for 10 lowest energy Qua1 homodimer structures) and the following linear relationship to the effective surface area S_{Acc}^m was found:

$$S_{Acc}^m = 0.123\Gamma_{PRE} + 0.132. \quad (5)$$

Experimental solvent PREs ($s^{-1} \cdot mM^{-1}$) ($0 < \Gamma_{PRE} < 1$) were used as restraints in structure calculations. High PREs ($\Gamma_{PRE} > 1$) of surface protons were intentionally excluded due to the contamination of chemical exchange with water protons [43].

The calculation protocol used was similar to the one used for Bax and ubiquitin. Starting from a random coil conformation, a total of 150 out of 883 intra-molecular and 40 out of 125 inter-molecular NOE restraints were randomly chosen for each monomer subunit [43]. Dihedral restraints were included, but the other restraints were omitted. The optimal force constant used was $300 \text{ kcal/\AA}^2$. Again we assumed that the actual slope and offset described in Eq. (5) were not available. A grid search for the slope (between $0.1 \text{ \AA}^2 \cdot s \cdot mM$ and $0.4 \text{ \AA}^2 \cdot s \cdot mM$) and offset (between $0.05 \text{ \AA}^2$ and $0.18 \text{ \AA}^2$) was conducted (Figure 5A&B). These values were further optimized by the iterative method to arrive at the final slope of $0.137 \text{ \AA}^2 \cdot s \cdot mM$ and offset of $0.151 \text{ \AA}^2$. The resulting 5 lowest energy structures calculated with the new solvent accessibility potential showed an improvement compared to structures calculated without the new potential (E_{Acc}) (Table 2). The correlation between experimental and calculated solvent PRE values improved from a correlation coefficient R of 0.77 to 0.94 when the new solvent accessibility potential was used (Figure 5C&D). In addition, the precision of the calculated structures increased from $2.59 \pm 0.34 \text{ \AA}$ to $1.70 \pm 0.29 \text{ \AA}$ evaluated using the backbone atoms of the 5 lowest energy structures (calculated residues 99–134 for both subunit A and B) when the new potential was applied (Figure 5E&F). For each subunit of the Qua1 homodimer, the precision of the calculated structure increased from $2.34 \pm 0.48 \text{ \AA}$ to $1.49 \pm 0.34 \text{ \AA}$ and $1.99 \pm 0.70 \text{ \AA}$ to $1.41 \pm 0.40 \text{ \AA}$ for subunit A and B, respectively. The backbone of calculated structures using the new potential is $3.01 \pm 0.20 \text{ \AA}$ away from the published Qua1 homodimer structures (PDB ID: 2XA6), while without using the potential it is $3.44 \pm 0.37 \text{ \AA}$ away. The accuracy of the each subunit of the Qua1 homodimer increased from $3.17 \pm 0.45 \text{ \AA}$ to $2.77 \pm 0.35 \text{ \AA}$ and $2.98 \pm 0.67 \text{ \AA}$ to $2.83 \pm 0.30 \text{ \AA}$ for subunit A and B, respectively.

Conclusion

We demonstrate the utility of a new empirical function to represent the degree of solvent exposure of an individual residue in a protein. This new term was designed based on the location of neighboring atoms or “crowding” around the residue, which is directly related to the solvent accessibility (Supplemental Figure 1). The ability to distinguish “inside” and “outside” residues in principle should assist in determining a global fold of the protein. We

illustrated the feasibility of this approach using both experimentally acquired water to protein magnetization transfer data and solvent PRE data. These data correlated well with the values calculated from the known protein structures using the new empirical term (Figure 2&4). The addition of this new potential to a limited number of NMR-based restraints can significantly improve structure convergence to the correct conformation as evidenced for protein Bax (Table 1) and ubiquitin (Supplemental Table 1). Furthermore, these types of solvent accessibility restraints were used in determining the protein complex structures as illustrated with the Qua1 homodimer (Table 2). These restraints are very useful in this case since the interaction surfaces of the complex have different solvent accessibility profiles than the free proteins. During the calculation the proper slope and offset in Eq. 2, 4, or 5 can be searched by a coarse grid method and further optimized using an iterative operation by fitting to the experimental data to the calculated structures. From the calculations tested using proteins Bax, ubiquitin and Qua1 homodimer, we demonstrated that access to this type of solvent accessibility information could substantially expedite the initial stage of the NMR structure determination.

The water to protein magnetization transfer experiment used to illustrate our method probes very complicated interactions between the protein and the water molecules. This measurement alone will not differentiate the water to a protein-proton NOE rate from the solvent exchange rate. In experimental conditions used here, the chemical exchange rates between bulk water and exchangeable NH protons are generally faster than the inefficient NOE rates between the two due to the short residence time [33; 44]. For those amide hydrogen's protected by hydrogen-bonds, most of the amide hydrogen-water NOE correlations seen in bulk solution are also likely dominated by exchange-relayed NOE interactions with labile side chain hydrogen, not by direct protein-water NOEs [30; 45; 46]. Nevertheless, it is important to point out that an increase in both NOE and solvent exchange rates still indicates that the residue is solvent exposed. In addition to the above issue, long lived bound water or buried water molecule, C_{α} protons that resonate close to water frequency, as well as rapidly exchanging hydroxyl protons can result in measured intensity ratios that will not follow the relationship described here. Some of these contributions can be distinguished by carrying out a complementary ROE experiment as previously suggested [36; 47]. Interestingly, the intensity ratios for Bax and Bid that we measured were not screened for any of these possible effects and still resulted in an acceptable outcome. We did, however, use a deuterated ubiquitin sample in our water to protein magnetization transfer experiment to try to overcome the problem due to some H_{α} resonances that are close to the frequency of water. Alternatively, any intensity ratios that do not follow the empirical relationship can be identified during the structure refinement process and treated accordingly, since they most likely will not agree with the distance restraints of the protein. This is similar to the PRE refinement when there is contribution from minor, previously unknown, conformation that would result in a deviation from the predicted PRE values [48].

The results presented here provide some evidence that the S_{Acc} metric can be of general use in quantitatively interpreting various observations that primarily depend on solvent accessibility. The inclusion of this type of structural information that differentiates surface exposed from buried residues can significantly improve convergence and accuracy in structure calculations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to Motoshi Suzuki for the help on experimental setup and Yi He for fermentation preparation of isotopically labeled Bax. We thank Tobias Madl and Michael Sattler for providing us with the solvent PRE data. This work was supported by the NIH Intramural Research Programs of CIT (C.D.S) and NHLBI (N.T.).

References

1. Prestegard JH, al-Hashimi HM, Tolman JR. NMR structures of biomolecules using field oriented media and residual dipolar couplings. *Q Rev Biophys.* 2000; 33:371–424. [PubMed: 11233409]
2. Tjandra N, Bax A. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science.* 1997; 278:1111–4. [PubMed: 9353189]
3. Tjandra N, Omichinski JG, Gronenborn AM, Clore GM, Bax A. Use of dipolar ^1H - ^{15}N and ^1H - ^{13}C couplings in the structure determination of magnetically oriented macromolecules in solution. *Nat Struct Biol.* 1997; 4:732–8. [PubMed: 9303001]
4. Tolman JR, Flanagan JM, Kennedy MA, Prestegard JH. Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. *Proc Natl Acad Sci U S A.* 1995; 92:9279–83. [PubMed: 7568117]
5. Choy WY, Tollinger M, Mueller GA, Kay LE. Direct structure refinement of high molecular weight proteins against residual dipolar couplings and carbonyl chemical shift changes upon alignment: an application to maltose binding protein. *J Biomol NMR.* 2001; 21:31–40. [PubMed: 11693566]
6. Cornilescu G, Marquardt JL, Ottiger M, Bax A. Validation of Protein Structure from Anisotropic Carbonyl Chemical Shifts in a Dilute Liquid Crystalline Phase. *J Am Chem Soc.* 1998; 120:6836–6837.
7. Hansen AL, Al-Hashimi HM. Insight into the CSA tensors of nucleobase carbons in RNA polynucleotides from solution measurements of residual CSA: towards new long-range orientational constraints. *J Magn Reson.* 2006; 179:299–307. [PubMed: 16431143]
8. Lipsitz RS, Tjandra N. Carbonyl CSA restraints from solution NMR for protein structure refinement. *J Am Chem Soc.* 2001; 123:11065–6. [PubMed: 11686713]
9. Fushman D, Tjandra N, Cowburn D. Direct Measurement of ^{15}N Chemical Shift Anisotropy in Solution. *J Am Chem Soc.* 1998; 120:10947–10952.
10. Ryabov Y, Clore GM, Schwieters CD. Direct use of ^{15}N relaxation rates as experimental restraints on molecular shape and orientation for docking of protein-protein complexes. *J Am Chem Soc.* 2010; 132:5987–9. [PubMed: 20392103]
11. Tjandra N, Garrett DS, Gronenborn AM, Bax A, Clore GM. Defining long range order in NMR structure determination from the dependence of heteronuclear relaxation times on rotational diffusion anisotropy. *Nat Struct Biol.* 1997; 4:443–9. [PubMed: 9187651]
12. Tjandra N, Wingfield P, Stahl S, Bax A. Anisotropic rotational diffusion of perdeuterated HIV protease from ^{15}N NMR relaxation measurements at two magnetic fields. *J Biomol NMR.* 1996; 8:273–84. [PubMed: 8953218]
13. Reif B, Hennig M, Griesinger C. Direct measurement of angles between bond vectors in high-resolution NMR. *Science.* 1997; 276:1230–3. [PubMed: 9157875]
14. Schwalbe H, Carlomagno T, Hennig M, Junker J, Reif B, Richter C, Griesinger C. Cross-correlated relaxation for measurement of angles between tensorial interactions. *Methods Enzymol.* 2001; 338:35–81. [PubMed: 11460558]
15. Banci L, Bertini I, Savellini GG, Romagnoli A, Turano P, Cremonini MA, Luchinat C, Gray HB. Pseudocontact shifts as constraints for energy minimization and molecular dynamics calculations on solution structures of paramagnetic metalloproteins. *Proteins.* 1997; 29:68–76. [PubMed: 9294867]
16. Gaponenko V, Sarma SP, Altieri AS, Horita DA, Li J, Byrd RA. Improving the accuracy of NMR structures of large proteins using pseudocontact shifts as long-range restraints. *J Biomol NMR.* 2004; 28:205–12. [PubMed: 14752254]
17. Battiste JL, Wagner G. Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry.* 2000; 39:5355–65. [PubMed: 10820006]

18. Iwahara J, Schwieters CD, Clore GM. Ensemble approach for NMR structure refinement against (1)H paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule. *J Am Chem Soc.* 2004; 126:5879–96. [PubMed: 15125681]
19. Tang C, Iwahara J, Clore GM. Visualization of transient encounter complexes in protein-protein association. *Nature.* 2006; 444:383–6. [PubMed: 17051159]
20. Kuszewski J, Gronenborn AM, Clore GM. Improving the Packing and Accuracy of NMR Structures with a Pseudopotential for the Radius of Gyration. *J Am Chem Soc.* 1999; 121:2337–2338.
21. Schwieters CD, Clore GM. A pseudopotential for improving the packing of ellipsoidal protein structures determined from NMR data. *J Phys Chem B.* 2008; 112:6070–3. [PubMed: 18088109]
22. Svergun DI, Koch MH. Advances in structure analysis using small-angle scattering in solution. *Curr Opin Struct Biol.* 2002; 12:654–60. [PubMed: 12464319]
23. Ryabov YE, Geraghty C, Varshney A, Fushman D. An efficient computational method for predicting rotational diffusion tensors of globular proteins using an ellipsoid representation. *J Am Chem Soc.* 2006; 128:15432–44. [PubMed: 17132010]
24. Ryabov Y, Suh JY, Grishaev A, Clore GM, Schwieters CD. Using the experimentally determined components of the overall rotational diffusion tensor to restrain molecular shape and size in NMR structure determination of globular proteins and protein-protein complexes. *J Am Chem Soc.* 2009; 131:9522–31. [PubMed: 19537713]
25. Linge JP, Williams MA, Spronk CA, Bonvin AM, Nilges M. Refinement of protein structures in explicit solvent. *Proteins.* 2003; 50:496–506. [PubMed: 12557191]
26. Otting G. Protein NMR using paramagnetic ions. *Annu Rev Biophys.* 2010; 39:387–405. [PubMed: 20462377]
27. Fesik SW, Gemmecker G, Olejniczak ET, Petros AM. Identification of solvent-exposed regions of enzyme-bound ligands by nuclear magnetic resonance. *J Am Chem Soc.* 1991; 113:7080–7081.
28. Madl T, Bermel W, Zangger K. Use of relaxation enhancements in a paramagnetic environment for the structure determination of proteins using NMR spectroscopy. *Angew Chem Int Ed Engl.* 2009; 48:8259–62. [PubMed: 19774576]
29. Madl T, Guttler T, Gorlich D, Sattler M. Structural analysis of large protein complexes using solvent paramagnetic relaxation enhancements. *Angew Chem Int Ed Engl.* 2011; 50:3993–7. [PubMed: 21442693]
30. Nucci NV, Pometun MS, Wand AJ. Site-resolved measurement of water-protein interactions by solution NMR. *Nat Struct Mol Biol.* 2011; 18:245–9. [PubMed: 21196937]
31. Gruschus JM, Ferretti JA. Quantitative measurement of water diffusion lifetimes at a protein/DNA interface by NMR. *J Biomol NMR.* 2001; 20:111–26. [PubMed: 11495243]
32. Wang YX, Freedberg DI, Grzesiek S, Torchia DA, Wingfield PT, Kaufman JD, Stahl SJ, Chang CH, Hodge CN. Mapping hydration water molecules in the HIV-1 protease/DMP323 complex in solution by NMR spectroscopy. *Biochemistry.* 1996; 35:12694–704. [PubMed: 8841113]
33. Otting G, Liepinsh E, Wuthrich K. Protein hydration in aqueous solution. *Science.* 1991; 254:974–80. [PubMed: 1948083]
34. Suzuki M, Youle RJ, Tjandra N. Structure of Bax: coregulation of dimer formation and intracellular localization. *Cell.* 2000; 103:645–54. [PubMed: 11106734]
35. Lazar GA, Desjarlais JR, Handel TM. De novo design of the hydrophobic core of ubiquitin. *Protein Sci.* 1997; 6:1167–78. [PubMed: 9194177]
36. Dalvit C, Hommel U. Sensitivity-Improved Detection of Protein Hydration and Its Extension to the Assignment of Fast-Exchanging Resonances. *Journal of Magnetic Resonance, Series B.* 1995; 109:334–338.
37. Dalvit C, Hommel U. New pulsed field gradient NMR experiments for the detection of bound water in proteins. *JOURNAL OF BIOMOLECULAR NMR.* 1995; 5:306–310.
38. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR.* 1995; 6:277–93. [PubMed: 8520220]
39. Johnson BA, Blevins RA. NMR View: A computer program for the visualization and analysis of NMR data. *Journal of Biomolecular NMR.* 1994; 4:603–614.

40. Garrett DS, Powers R, Gronenborn AM, Clore GM. A common sense approach to peak picking two-, three- and four-dimensional spectra using automatic computer analysis of contour diagrams. *J Magn Reson.* 1991; 95:214–220.
41. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM. The Xplor-NIH NMR molecular structure determination package. *J Magn Reson.* 2003; 160:65–73. [PubMed: 12565051]
42. Chou JJ, Li H, Salvesen GS, Yuan J, Wagner G. Solution structure of BID, an intracellular amplifier of apoptotic signaling. *Cell.* 1999; 96:615–24. [PubMed: 10089877]
43. Meyer NH, Tripsianes K, Vincendeau M, Madl T, Kateb F, Brack-Werner R, Sattler M. Structural basis for homodimerization of the Src-associated during mitosis, 68-kDa protein (Sam68) Qual domain. *J Biol Chem.* 2010; 285:28893–901. [PubMed: 20610388]
44. Liepinsh E, Otting G. Proton exchange rates from amino acid side chains--implications for image contrast. *Magn Reson Med.* 1996; 35:30–42. [PubMed: 8771020]
45. Fitzkee NC, Torchia DA, Bax A. Measuring rapid hydrogen exchange in the homodimeric 36 kDa HIV-1 integrase catalytic core domain. *Protein Sci.* 2011; 20:500–12. [PubMed: 21213249]
46. Halle B. Protein hydration dynamics in solution: a critical survey. *Philos Trans R Soc Lond B Biol Sci.* 2004; 359:1207–23. discussion 1223–4, 1323–8. [PubMed: 15306377]
47. Dalvit C. Homonuclear 1D and 2D NMR Experiments for the Observation of Solvent-Solute Interactions. *J Magn Reson B.* 1996; 112:282–8. [PubMed: 8812917]
48. Tang C, Schwieters CD, Clore GM. Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature.* 2007; 449:1078–82. [PubMed: 17960247]

Highlights

- We formulated a new structural metric for solvent-protein interaction
- We showed that this new metric correlated well with experimental NMR data
- We developed a protocol to use this information in structure calculations
- Addition of this new information improved precision and accuracy of the structures

Figure 1.

(A) Overlay of the 2D water-protein interaction spectrum (red) with the reference HSQC spectrum (black). The inset shows an expanded region of the HSQC spectrum to illustrate the variations of the water-protein interaction peak intensities. The decrease of peak intensity indicates less water-protein interaction and thus less water accessibility. (B) Plot of the intensity ratio between the water-protein interaction and the reference HSQC spectra for the protein Bax. The missing values are due to proline residues or residues whose resonances overlap in the spectrum. The secondary structure composition of Bax are shown as helices at the top of the panel.

Figure 2.

Correlation plot of calculated water accessibility values of Bax based on Eq. (1) over the experimental relative water-protein interaction intensity ratios. The calculation was carried out using the lowest energy Bax structure (PDB ID: 1F16), resulting in a linear correlation coefficient value of 0.80. The red line is a result of a linear regression fit of the data.

Figure 3.

The dependence of total energy for Bax structure calculations on slope (while offset was fixed at $0.06 \text{ \AA}^2$) (A) and offset (with slope fixed at $2.06 \text{ \AA}^2$) (B) from Eq. (2). The correlation plot of the calculated solvent accessibility values based on Eq. (1) for the mean structure of 5 lowest energy structures calculated without (C) and with (D) the new water accessibility potential to the experimental relative water-protein interaction intensity ratio. An improvement of the correlation coefficient (R) was shown from 0.79 to 0.86 with the incorporation of the new water accessibility term. Overlay of the 5 lowest energy structures calculated with only 500 randomly chosen non-intra residue 1H_C - 1H_C NOE, dihedral and hydrogen bond restraints in the absence (E) and presence (F) of the new water accessibility potential term [34]. An obvious increase of structure convergence was observed and the protein precision was improved from $3.12 \text{ \AA}$ to $1.59 \text{ \AA}$.

Figure 4.

Correlation plot of calculated solvent accessibility values over the experimental solvent PRE values Γ_{PRE} for ubiquitin (**A**) [28] and Qua1 homodimer (**B**) [43]. The calculation was carried out based on Eq. (1) using the lowest energy ubiquitin structure (PDB ID: 1D3Z) [6] and Qua1 homodimer structure (PDB ID: 2XA6) [35], resulting in a linear correlation coefficient value of 0.86 and 0.82, respectively. The red line is a result of a linear regression fit of the data.

Figure 5.

The total energy as a function of slope, while offset it set to $0.0 \text{ \AA}^2$ (A) and offset (B) with slope fixed at $0.2 \text{ \AA}^2 \cdot \text{s} \cdot \text{mM}$ for the Qua1 homodimer structure calculations. The correlation plot of the calculated solvent accessibility values based on Eq. (1) for the mean structure of the 5 lowest energy structures calculated without (C) and with (D) the new water accessibility potential to the experimental solvent PRE values. An improvement of the correlation coefficient (R) from 0.77 to 0.94 was observed. Overlay of the 5 lowest energy structures calculated with only 150 out of 883 intra-molecular and 40 out of 125 inter-molecular randomly chosen NOE restraints and dihedral restraints in the absence (E) and presence (F) of the new water accessibility potential term [43]. An obvious increase of structure convergence was seen and the overall protein precision was improved from $2.59 \text{ \AA}$ to $1.70 \text{ \AA}$.

Table 1

Structure statistics of calculations performed with full or partial sets of restraints in the absence and presence of the new water accessibility restraints for the protein Bax.

Structure calculation	Restraint Set ^a	Backbone RMSD (Å) ^b	Backbone pair- wise RMS to 1F16 (Å) ^c	Correlation coefficient (R) ^d
without E_{Acc}	500 NOEs	3.12±0.97	3.93±1.40	0.79±0.06
with E_{Acc}	500 NOEs	1.59±0.21	2.77±0.04	0.86±0.002
without E_{Acc}	Full	1.27±0.16	1.92±0.23	0.82±0.04
with E_{Acc}	Full	1.16±0.17	1.95±0.25	0.85±0.02

The optimal final force constants used for solvent accessibility term were 70 kcal/Å² for both the full and partial restraint calculations.

^aThe “500 NOEs” restraint set corresponds to 500 randomly chosen non-intra residue ¹H_C-¹H_C NOEs, dihedral, and hydrogen bond restraints; while “Full” indicates all NOE, dihedral, hydrogen bond, and RDC restraints as previously reported [34].

^bThe backbone RMSD was calculated using backbone atoms of residues 15–189 in the 5 lowest energy structures to the mean structure.

^cThe backbone pair-wise RMS was calculated from the 5 lowest energy structures against the lowest energy structure of Bax (PDB ID: 1F16).

^dThe correlation coefficient R is the linear correlation between the calculated solvent accessibility using Eq. (1) and the measured water NOE values.

Table 2

Structure statistics of calculations performed with limited number of restraints in the absence and presence of the new solvent accessibility restraints for the protein Qua1 homodimer.

	Calculation Condition ^a	Backbone RMSD (Å) ^b	Backbone pair-wise RMS to 2XA6 (Å) ^c	Correlation coefficient (R) ^d
Overall	without E_{Acc}	2.59±0.34	3.44±0.37	0.77±0.07
	with E_{Acc}	1.70±0.29	3.01 ± 0.20	0.94±0.02
Subunit A	without E_{Acc}	2.34 ± 0.48	3.17 ± 0.45	0.77±0.07
	with E_{Acc}	1.49 ± 0.34	2.77 ± 0.35	0.94±0.02
Subunit B	without E_{Acc}	1.99 ± 0.70	2.98 ± 0.67	0.76±0.07
	with E_{Acc}	1.41 ± 0.40	2.83 ± 0.30	0.94±0.01

The optimal final force constants used for solvent accessibility term were 300 kcal/Å² for calculation using partial NOE restraints of Qua1 homodimer.

^aThe partial NOE restraints correspond to 150 out of 883 intra-molecular and 40 out of 125 inter-molecular NOE restraints that were randomly chosen for each monomer subunit [43].

^bThe backbone RMSD was calculated using backbone atoms of residues 99–134 for each subunit in the 5 lowest energy structures to the mean structure.

^cThe backbone pair-wise RMS was calculated from the 5 lowest energy structures against the lowest energy structure of Qua1 homodimer (PDB ID: 2XA6).

^dThe correlation coefficient R is the linear correlation of the calculated solvent accessibility using Eq. (1) and the Qua1 homodimer solvent PRE data [43].