Article:
Fowler, N.J., Albalwi, M.F., Lee, S. et al. (2023) Improved methodology for protein NMR structure calculation using hydrogen bond restraints and ANSURR validation: The SH2 domain of SH2B1. Structure, 31 (8). 975-986.e3. ISSN: 0969-2126

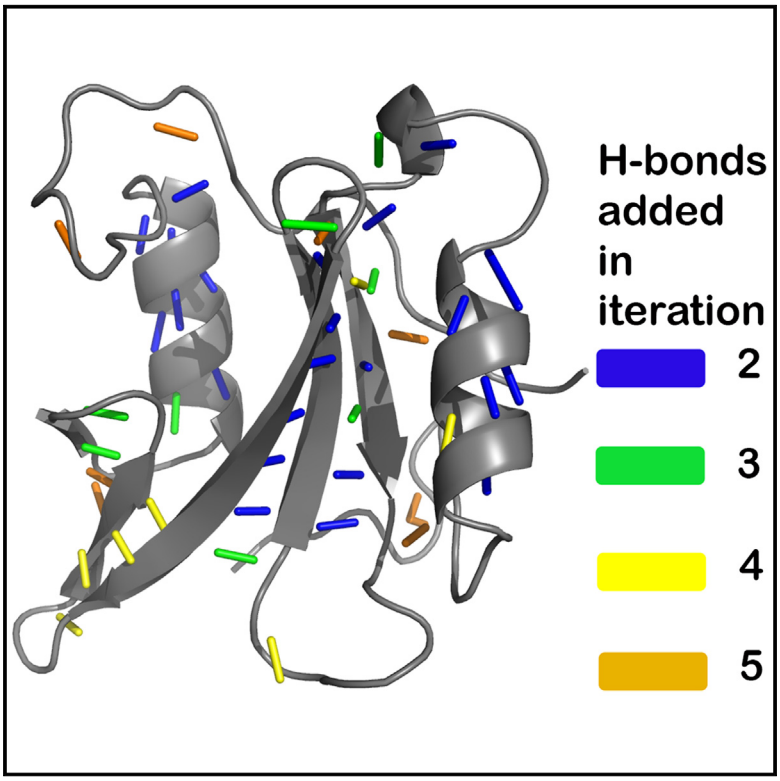
https://doi.org/10.1016/j.str.2023.05.012

Structure

Article

Improved methodology for protein NMR structure calculation using hydrogen bond restraints and ANSURR validation: The SH2 domain of SH2B1

Graphical abstract



Highlights

- We show that NMR structures typically lack hydrogen bonds
- We present a protocol for gradual inclusion of hydrogen bond restraints
- The resulting structures are better using many measures of quality
- ANSURR provides a measure for when the NMR structure calculation is completed

Authors

Nicholas J. Fowler, Marym F. Albalwi, Subin Lee, Andrea M. Hounslow, Mike P. Williamson

Correspondence

nick.j.fowler0@gmail.com (N.J.F.), m.williamson@sheffield.ac.uk (M.P.W.)

In brief

NMR structures of proteins need to be more accurate. Fowler et al. show that accuracy can be improved by inclusion of hydrogen bond restraints, and demonstrate an iterative method for inclusion of such restraints, showing a clear improvement in accuracy using a range of measures.

Fowler et al., 2023, Structure *31*, 975–986
August 3, 2023
https://doi.org/10.1016/j.str.2023.05.012

CellPress

Structure

CellPress
OPEN ACCESS

Article

Improved methodology for protein NMR structure calculation using hydrogen bond restraints and ANSURR validation: The SH2 domain of SH2B1

Nicholas J. Fowler,[1,*] Marym F. Albalwi,[1] Subin Lee,[1] Andrea M. Hounslow,[1] and Mike P. Williamson[1,2,*]
[1]School of Biosciences, University of Sheffield, S10 2TN Sheffield, UK
[2]Lead contact
*Correspondence: nick.j.fowler0@gmail.com (N.J.F.), m.williamson@sheffield.ac.uk (M.P.W.)
https://doi.org/10.1016/j.str.2023.05.012

SUMMARY

Protein structures calculated using NMR data are less accurate and less well-defined than they could be. Here we use the program ANSURR to show that this deficiency is at least in part due to a lack of hydrogen bond restraints. We describe a protocol to introduce hydrogen bond restraints into the structure calculation of the SH2 domain from SH2B1 in a systematic and transparent way and show that the structures generated are more accurate and better defined as a result. We also show that ANSURR can be used as a guide to know when the structure calculation is good enough to stop.

INTRODUCTION

We recently introduced a new method, ANSURR (Accuracy of NMR structures Using RCI and Rigidity), which permits a simple and robust indication of the accuracy of NMR structures.[1–3] We noted[2] that when estimated in this way, the accuracy of NMR protein structures improved until about 2005 but has not changed much since; and we suggested that one reason for the lack of improvement is that there has been until now no reliable way to calculate accuracy, and therefore to measure the success of technical improvements. With the introduction of ANSURR, we have a tool that increases our confidence in measuring whether new methods are effective. In this paper, we take the first step in implementing this idea, by addressing one experimental restraint that has proved difficult to implement, namely hydrogen bond restraints (HBRs).

The most widely used restraints for NMR structure calculation are nuclear Overhauser effects (NOEs). These are easily obtained from nuclear Overhauser effect spectroscopy (NOESY) spectra and have (in principle) a simple interpretation, in that under appropriate conditions, an NOE peak denotes the close distance between two protons, with an intensity proportional to r^{-6}, where r is the internuclear distance.[4] Another common restraint is to use backbone chemical shifts to produce restraints on dihedral angles, for example using TALOS-N.[5]

Hydrogen bonds are one of the few other common restraints and are very strong, because the distances are short and have a small range between upper and lower bounds. Restraint violations therefore typically become unfavorable with relatively small changes in geometry. For this reason, HBRs should be applied with caution, because they effectively force the relevant groups to form a hydrogen bond.

Hydrogen bonds are more difficult restraints to apply than NOEs because they cannot usually be obtained directly from experimental data. NOE restraints come directly from NOESY spectra and dihedral restraints from backbone chemical shifts. It is possible to observe hydrogen bonds directly from observation of ^{3}J across the hydrogen bond.[6] However, the coupling constant is small and for most proteins the signal is too weak to be detected. Therefore the presence of a hydrogen bond has to be inferred. Hydrogen bonds in regular secondary structure can be predicted based on NOE patterns in the vicinity,[7] but a more secure identification of hydrogen bonds relies on secondary factors such as slow amide exchange of the amide proton[8,9] or on the amide proton temperature coefficient.[10] Neither of these is completely reliable, meaning that any inference of a hydrogen bond is potentially wrong. Significantly, amide exchange and temperature coefficients identify the amide group but do not identify the carbonyl donor. Therefore for most hydrogen bonds, identification of the donor is based on inferences not direct observation.

For these reasons, NMR spectroscopists have generally been cautious in applying HBRs in protein structure calculation. There is no "standard protocol," but typically HBRs are only applied toward the end of the calculation, the suspicion being that it may not be scientifically justified to use them as part of the standard structure calculation. There has thus emerged a sense that HBRs are not as "respectable" as NOEs or dihedral restraints, as a result of which, the details of hydrogen bond restraints are often deeply buried in papers, and it is often very hard to work out from the literature exactly how HBRs were generated and used. This is a pity, because HBRs are such powerful restraints, and therefore very useful if used correctly. The main aim of this work is to propose a justifiable and transparent protocol. We describe a robust iterative approach to the

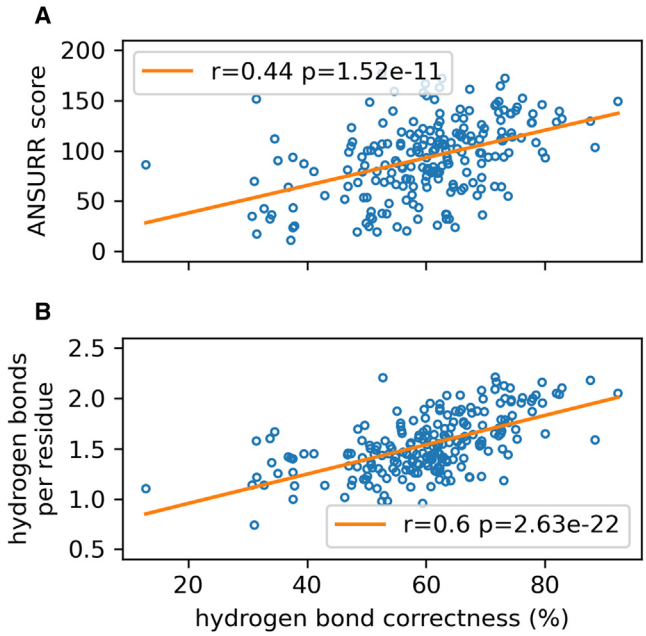


Figure 1. Correlations of structural parameters with hydrogen bond correctness

Correctness was calculated as the average percentage of hydrogen bonds in an X-ray structure that appear in the corresponding NMR ensemble.

(A) ANSURR score (B) number of hydrogen bonds per residue. Values were calculated for each structure in the ensemble and then averaged to give the value for the ensemble. The line of best fit is shown in orange, together with the Pearson correlation coefficient and two-tailed p value.

See also Table S1.

addition of HBRs, monitored using ANSURR, and demonstrate the effect of HBRs on the structure calculation of the SH2 domain from SH2B1. We show that the accuracy of the structures improves as HBRs are added.

SH2B1 is a signaling scaffold protein that acts downstream of the insulin and leptin receptors[11] and regulates the activity of JAK2 kinase. Mutations to its SH2 domain have been shown to disrupt the leptin signaling pathway[12]; deletion leads to an obese phenotype, and severely disrupts signaling downstream of the insulin receptor.[13] SH2B1 consists of a dimerization domain, a PH domain, and an SH2 domain, which together form less than half of the protein, the rest being intrinsically disordered.[11] Part of the function of SH2 is to bind to JAK2 kinase via its phosphorylated pTyr^{813} and modulate its kinase activity.[14] It is thus a protein with important biological effects. It is also an example of a well-studied domain with a crystal structure. For the purposes of this study this is useful because the crystal structure provides us with a useful guide to accuracy.

RESULTS

ANSURR suggests NMR structures lack hydrogen bonds

The aim of this study is to present an iterative approach to introducing hydrogen bond restraints (HBRs) into an NMR structure calculation, and to investigate whether this improves the accuracy of the structure. Our primary measure of accuracy is ANSURR. ANSURR calculates the rigidity of a protein structure using the algorithm Floppy Inclusions and Rigid Substructure Topography (FIRST)[15] and compares it to the rigidity derived from experimental backbone chemical shifts via the random coil index (RCI).[16] The similarity obtained by comparing these two measures provides the accuracy. The comparison uses two measures: *correlation*, which calculates the extent to which peaks and troughs in rigidity along the sequence match, and is primarily a measure of the correctness of local secondary structure; and *RMSD*, which compares rigidity along the sequence (Figure S1). Our previous studies[1,2] showed that although the correlation for NMR structures is often good, the RMSD is often poor, in almost all cases because NMR structures are too floppy in comparison to the rigidity indicated by backbone chemical shifts. We suggested that a major cause of this floppiness is that NMR structures have too few hydrogen bonds, in comparison to either crystal structures[2] or AlphaFold structures.[3] This was the rationale behind the present study. Before starting these structure calculations, we therefore undertook a more thorough study of the relationship between hydrogen bond content and accuracy.

A set of 215 pairs of published NMR and X-ray structures was assembled (Table S1). This set was used to compute the average hydrogen bond correctness of each NMR ensemble, defined as the average percentage of hydrogen bonds in an X-ray structure that appear in an NMR ensemble. This calculation relies on the assumption that the hydrogen bonds in a crystal structure are also present in solution, which is clearly not completely true, but is a good approximation.[1] We computed the average ANSURR score (the sum of correlation and RMSD scores, and thus a ranked score out of 200) for each ensemble and compared this to the average hydrogen bond correctness.

Figure 1A shows that NMR structures with higher hydrogen bond correctness tend to be more accurate as measured by ANSURR, and also indicates that the hydrogen bonds in X-ray structures generally represent those that appear in solution. Plotting the number of hydrogen bonds per residue in each ensemble against hydrogen bond correctness (Figure 1B) demonstrates that poor scoring NMR ensembles are primarily inaccurate because they lack hydrogen bonds, rather than because they contain incorrect hydrogen bonds. We conclude that the accuracy of NMR structures will be improved by incorporation of HBRs but only if they are correct. This is important, because it implies that ANSURR scores will not improve on addition of incorrect HBRs. In the subsequent text, we therefore explore how to add HBRs, monitoring the accuracy of the resultant structures using ANSURR but also comparing to other measures.

NMR assignments for the SH2 domain and preliminary structure calculation

A plasmid for expression of mouse SH2 was constructed and inserted into *Escherichia coli*. The expressed protein is 118 residues long, with an 8-residue His-tag at the N-terminus, followed by the sequence D^{9}QPLS …, with D^{9} corresponding to D^{519} of mature mouse SH2B. The protein was expressed and labeled uniformly with ^{15}N and ^{13}C. One aim of this project was to test out the automated assignment/structure calculation in FLYA/CYANA.[17] The NMR spectrum was therefore assigned automatically using CYANA, as well as manually. The NMR spectra used are listed in Table S2, and the final manually assigned HSQC is

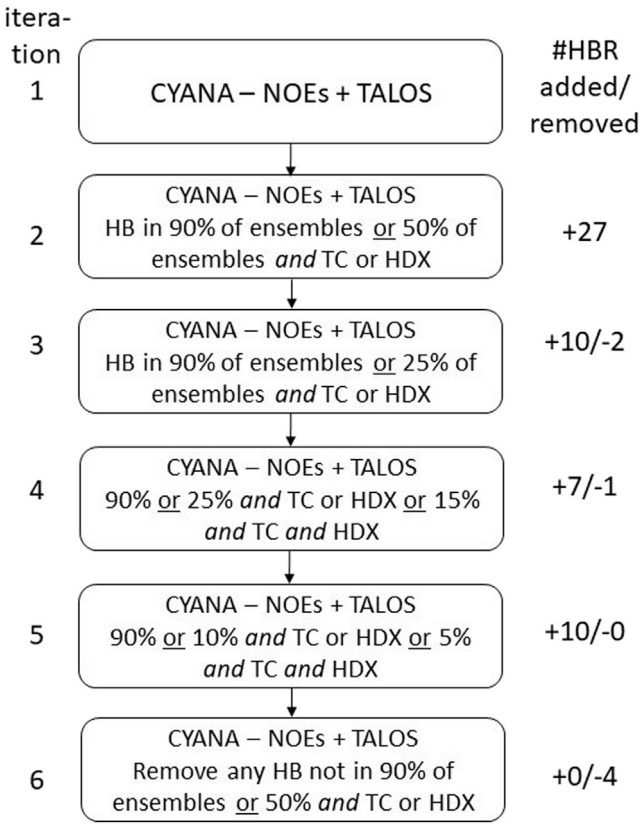


Figure 2. Protocol for addition of hydrogen bond restraints

1200 structures (60 sets of 20) were calculated at each iteration. TC: temperature coefficient greater than −4.5 ppb/K. HDX: slow amide exchange. At each iteration, HBRs were added using the criteria listed, and calculations were repeated until there were no further changes in HBRs. Figures on the right are the number of hydrogen bonds added/removed at each iteration. In iteration 2, hydrogen bonds were used as restraints if either they appeared (ie, observed in at least 6 models within the ensemble of 20) in at least 90% of the ensembles generated in iteration 1, or appeared in at least 50% of the ensembles and were supported by either temperature coefficient or slow amide exchange data. In iteration 3 we adjusted our criteria so that we could accept less frequently sampled hydrogen bonds (25% of ensembles) so long as they were supported by temperature coefficient or slow amide exchange data. In iteration 4, we further lowered our acceptance criteria to explore rarely sampled hydrogen bonds that were supported by both amide exchange and temperature coefficients, with the view that the set of restraints already applied in more restrictive previous iterations should prevent the adoption of grossly incorrect hydrogen bond restraints at this stage. The criteria were relaxed further in iteration 5. Finally, for iteration 6 we restored our initial acceptance criteria to remove any hydrogen bond restraints that failed to converge, and were therefore potentially in conflict with NOE or dihedral restraints, or with accessible protein geometry.

shown in Figure S2. A comparison of the two assignments showed that 99% of the automated backbone chemical shift assignments ($^{1}H_N$, ^{15}N, $^{1}H_\alpha$, $^{13}C_o$, $^{13}C_\alpha$, and $^{13}C_\beta$) of SH2 (9–118) agreed with the manual assignment. The only resonances that were not assigned correctly by the automated calculation are the amide chemical shifts of Trp17 and Gln61, which are two of the four resonances that had missing signals (but assigned H and N) in the manual assignment. For aliphatic methyl and methylene sidechain groups, the agreement was 97%. On the other hand, many of the NOE-based chemical shift assignments did not match the manual assignment. This includes the amide sidechains of Arg, Asn, and Gln, and many aromatics: only 53% of these were assigned correctly (Figure S3). CYANA continues to check and refine its assignments throughout the structure calculations, and it is possible that some of these incorrect assignments would have been corrected. On the other hand, the accuracy of the automated structure calculation relies heavily on the accuracy of the first iteration, which in turn depends on the accuracy of the initial set of assignments.[18] The clear implication is that assignments should be checked manually before doing automated structure calculations.[19] A manually checked set of assignments have been deposited in BioMagResBank[20] under code 51342.

Structure calculation was conducted using the standard automated routines within CYANA. We used the default CYANA calculation of 20 structures. We then repeated this calculation 60 times changing the random seed each time, in order to obtain a better "consensus bundle" of structures,[21] giving a total of 1200 calculated structures per iteration. The reason for carrying out so many calculations was to generate a reliable set of HBRs, with good discrimination between correct and incorrect HBRs. The figure of 60 was determined by calculating the ratio of true positive HBRs (that is, HBRs identified that are also present in the crystal structure) to false positive HBRs (HBRs identified that are not in the crystal structure). This ratio improved with the number of calculations, reaching a plateau at 60 (Figure S4). HBRs were identified from these structures and used as restraints for the next iteration, as described below.

Refining hydrogen bond networks using CYANA and sparse NMR data

The initial round of CYANA calculations produced an ensemble with a correct fold but low accuracy (as measured by ANSURR), and had poor convergence, a large RMSD between structures, and poor energies, all symptoms of a weak restraint set. We therefore improved the structures by iteratively and systematically including backbone hydrogen bond restraints, in such a way as to avoid incorrectly biasing the structure calculation (Figure 2). The protocol aims to generate HBRs that are present in a significant fraction of ensemble structures, supported by experimental temperature coefficients and/or amide exchange rates, and are consistent with other restraints. Iteration 6 was introduced as a way of removing any HBRs that were incompatible with the experimental restraints.

The complete process identified 47 hydrogen bond restraints, mostly in regular secondary structure. The final set of HBRs is illustrated in Figure 3. Not surprisingly, the first HBRs to be identified came from the helices and from the "core" β-strands (residues 41–69). The shorter β-sheet HBRs tended to be added later. Many of the HBRs that do not form part of regular secondary structures (for example, HBRs from the amide protons of G14, Y15, G104, and V109) were identified in the final stages. More details are available in Table S3, which also shows that most of the HBRs that were added in an early iteration and subsequently removed are present in the crystal structure.

At each stage in the calculation, the structures were analyzed to measure whether the additional HBRs were improving the quality of the ensemble. Our primary analytical tool was

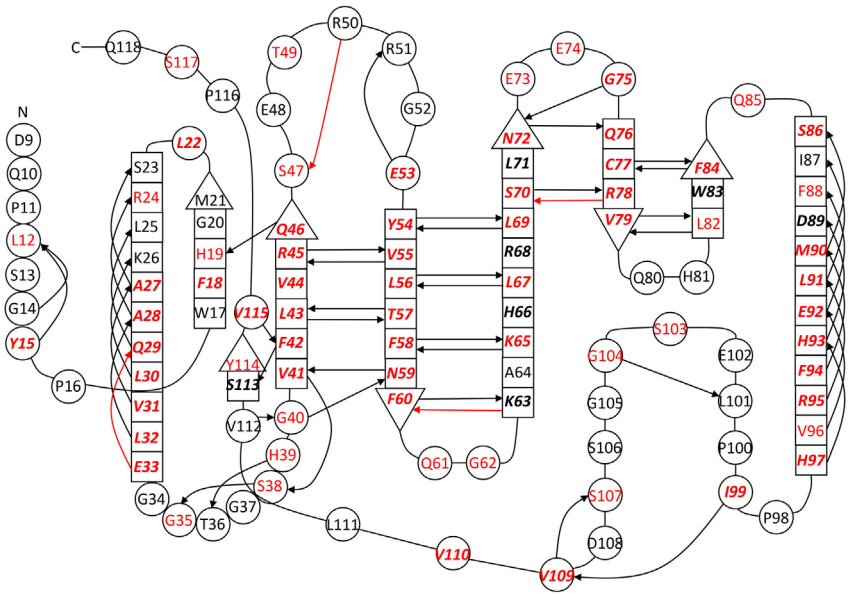


Figure 3. Secondary structure diagram for SH2

Residues in regular secondary structure are in square boxes, with an arrowhead at the C-terminal end of β-strands. Arrows indicate the hydrogen bond restraints used in the final iteration, starting at the NH and pointing toward the CO. Four restraints in red were used in iteration 5 and removed in the final iteration (Table S3). Residues are labeled in red if they have a temperature coefficient greater than −4.5 ppb/K, and in italic bold if they are slowly exchanging. We note that most residues with both slow exchange and large temperature coefficient have identified hydrogen bonding partners in our final iteration (42 out of 47).
See also Table S3.

ANSURR. Figure 4 shows how ANSURR scores steadily increase as the number of HBRs increases, suggesting that inclusion of hydrogen bond restraints does improve accuracy. The best scoring models in the final iteration 6 approach the accuracy of AlphaFold2[22] and Rosetta-Fold[23] (Figure 4A), although the average values remain worse. Note that the structures shown here come directly from CYANA and have not yet been refined in explicit solvent. The final iteration was designed to remove HBRs that do not have good supporting evidence. This iteration removed 4 out of 51 HBRs, leading to a drop in ANSURR score as well as some of the other measures of accuracy discussed below. The implication is that we probably removed correct HBRs, so this final step may have been over-cautious.

Figures 4B–4G depicts the ANSURR output for the best scoring model from each iteration. Figures 4H and 4I show ANSURR output for the X-ray structure (PDB: 5w3r) and the best scoring AlphaFold2 prediction, respectively. The crystal structure has a very good ANSURR score, but appears to be more rigid than the ''true'' solution structure around residues 70–90, judging by the rigidities shown in Figure 4H. With this exception, it would appear to be an excellent guide to the hydrogen bonds present in solution. In support, in Figure S5 we present an analysis of the HBRs at each iteration compared to those in the crystal structure, and show that they become progressively closer to those in the crystal structure. We note that the amino acid sequence of the proteins used for NMR and crystallography[24] (mouse and human, respectively) differ in 5 locations (Figure S6). It is of interest that the region for which NMR does worst is residues 46–55, which is too flexible in the NMR structures but is better in the X-ray structure. In the crystal, there is a bound phosphate here. This is where phosphotyrosine binds. The residues in contact with the phosphate in the crystal structure (residues 47–49) appear to be too rigid, suggesting that the phosphate is more tightly bound in the crystal than in solution, or perhaps that the flexibility according to chemical shifts represents an average of a bound state and unbound state. We have compared backbone chemical shifts for protein in 50 mM Tris buffer and in 50 mM phosphate buffer, pH 6 (Figure S7). It is striking that almost all of the large chemical shift changes surround the location of the phosphate ion in the crystal structure, implying that phosphate binds here in 50 mM phosphate buffer. This may serve to rigidify the region, and explain why the NMR structure, which was determined in tris buffer, has low rigidity here.

Validation of iterations against common metrics of structure quality

The previous section showed that inclusion of HBRs leads to a steady improvement in ANSURR score: i.e., a more accurate structure. Here, we report on other commonly used measures of quality, to characterize the structural improvement in more detail. The results are shown in Figure 5.

The magnitude of NOE violations increases gradually as HBRs are included, although it remains low throughout (Figure 5A). The calculation of an NMR structure is a balance between forces pulling in different directions: for example, van der Waals violations push atoms apart, while NOE violations pull them together. For this reason, it is generally found that addition of a new type of restraint leads to a slight increase in violations of existing restraints.[25] We therefore do not see this result as surprising or indicative of structural errors. In a similar way, the magnitude of dihedral angle violations has a small but insignificant increase (Figure 5B), as do van der Waals violations (data not shown). The RMSD of backbone atoms to the crystal structure improves on initial addition of HBRs, up to iteration 4 after which there is little change (Figure 5C). This can be understood by the argument that HBRs are excellent for improving local geometry, but make little difference to the positions of backbone atoms. It is widely agreed that the template modeling score (TM-score) is a more accurate measure of global similarity than RMSD,[26] and we see a more clear trend for TM-score than for RMSD (Figure 5D). TM-score slightly increases up until iteration 4. Residual dipolar couplings (RDCs) can be used either to validate the quality of NMR structures or as restraints to improve NMR structures, but not both at the same time (unless steps are taken to separate them into different groups).[27] Here we measured a set of $^{1}D_{NH}$ as a simple validation measure. Figure 5E shows the RDC Q factor,[28] which

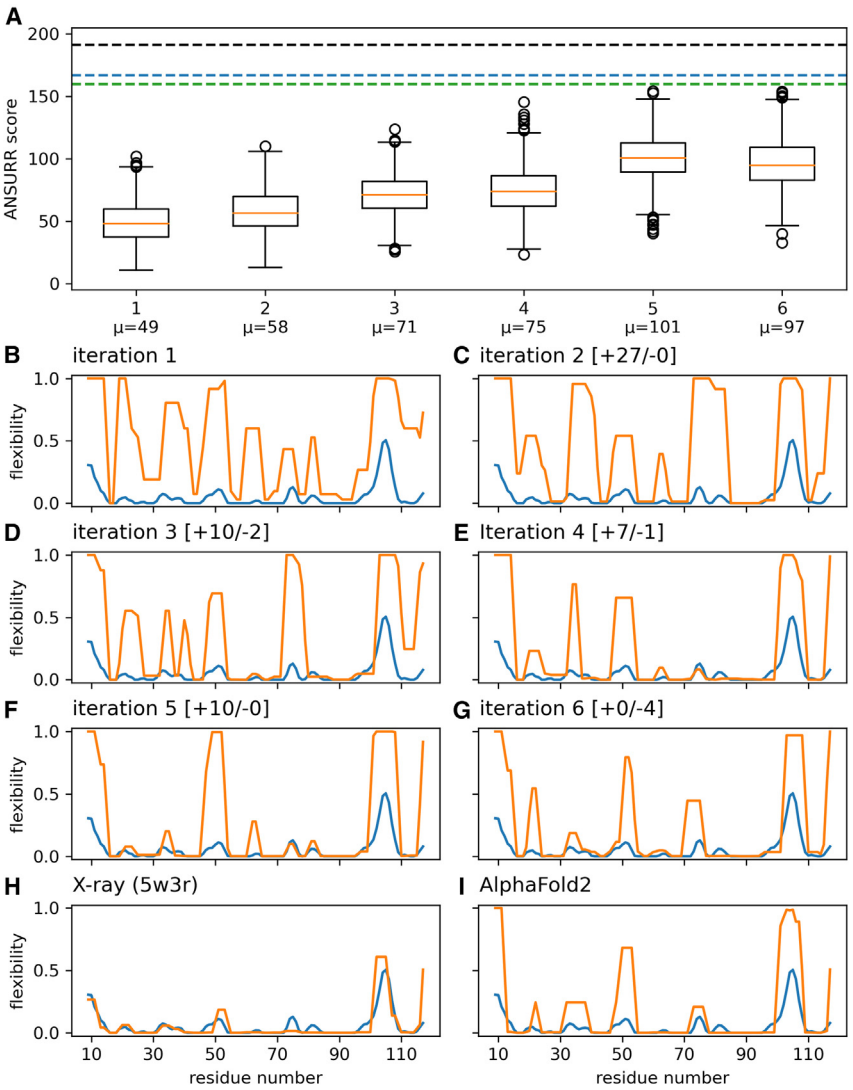


Figure 4. Improvements in ANSURR scores resulting from addition of hydrogen bond restraints, before refinement

(A) Boxplots showing the distribution of ANSURR scores for each iteration of CYANA calculations. In each panel, the orange bar is the median; box boundaries denote first and third quartiles; bars denote the range; and circles denote outliers. ANSURR scores steadily improve as the set of HBRs is refined, suggesting that inclusion of HBRs does improve accuracy. Mean ANSURR scores for each iteration are indicated. The black dashed line is the score for the X-ray structure 5w3r; the blue dashed line is the mean ANSURR score for 5 models predicted by AlphaFold2 and the green line is 5 models predicted by Rosetta-Fold.

(B–G) ANSURR comparison of flexibility according to chemical shifts (blue) and flexibility computed for the best scoring model from each iteration (orange). The number of hydrogen bond restraints added and removed between each iteration is indicated in square brackets.

(H) ANSURR output for the X-ray structure of SH2B1 (PDB: 5w3r).

(I) ANSURR output for the best scoring AlphaFold2 model.

See also Figures S5–S7.

improves up to iteration 4. The chemical shifts of amide proteins should be a sensitive measure of structural accuracy, because they are strongly affected by hydrogen bonding.[29] We therefore calculated the expected amide proton shifts using SHIFTX2[30] and compared them to the experimental shifts (Figure 5F), which improve steadily as HBRs are added. The distribution of Ramachandran angles also improves slightly up to iteration 4 (Figure 5G). Finally, we have calculated the backbone ensemble RMSD (Figure 5H). This is a measure of precision rather than accuracy, but our survey of PDB structures indicated that it does correlate with accuracy, presumably because adding more (correct) restraints serves to improve both accuracy and precision at the same time[2]; and indeed, RMSD decreases steadily as HBRs are added. Thus in summary, the gradual addition of 47 HBRs results in an improvement in accuracy, as measured by a wide range of parameters, with the improvement in accuracy slowing down as the number of HBRs increases. These metrics clearly show that there is overall no increased penalty to the structure quality from their inclusion: in other words, the HBRs are correct and do not conflict with the other structural restraints. The accuracy clearly improves until iteration 4. Iteration 5 results in an improvement on some measures but a deterioration in others, while iteration 6 (in which some HBRs were removed for lack of good evidence) represents a deterioration in almost all measures, the exception being precision (Figure 5H).

Refinement in explicit solvent

It is well established that refinement in solvent is required in order to produce good quality structures.[1,31] The 1,200 structures produced in each iteration above were therefore refined in water using CNS.[32] There is a reasonable correlation between the energies before and after refinement (Figure S8). The distance, dihedral angle, and hydrogen bond restraints used above were applied during the refinement. Interestingly, if the HBRs were removed at this stage, hydrogen bonds were lost and the structure quality got worse. This may indicate that the hydrogen bond potential in CNS is not ideal for this purpose: we note a potential recently introduced into Xplor-NIH that may handle this aspect better.[25] It is encouraging to see that the ANSURR score following refinement in solvent improves (Figure 6, compare to Figure 4), to the extent that the best scoring structures in the final iterations are competitive with the X-ray structure and AI predictions and could therefore be considered to be of good quality.

It is always difficult to know "when to stop" in an NMR structure calculation: improvements and further iterations are always possible. The ANSURR score is a ranked index, compared to all NMR structures in the PDB that we have reliable data for (e.g., at least 75% completeness of backbone chemical shifts and at

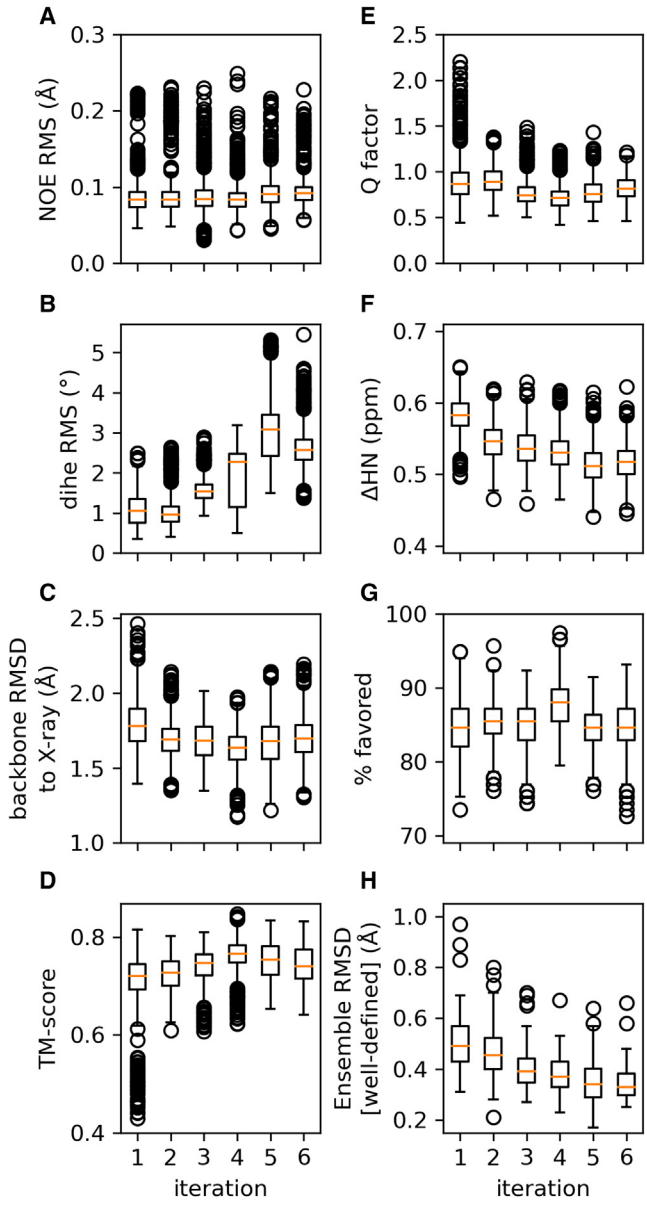


Figure 5. Change in metrics during the iterative addition of HBRs

In each panel, the orange bar is the median; box boundaries denote first and third quartiles; bars denote the range; and circles denote outliers. Data are shown for each model in the ensemble.

(A) Root-mean-square distance violation.

(B) Root-mean-square dihedral angle violation.

(C) Average backbone RMSD to the crystal structure 5w3r.

(D) Average template modeling score between NMR and crystal structures, where a value of 1 indicates a perfect match.

(E) Residual dipolar coupling *Q* factor.

(F) Mean difference between measured amide proton shifts and shifts calculated using SHIFTX2.

(G) Percentage of amino acids in the favored area of the Ramachandran plot.

(H) Average backbone RMSD to the mean structure (the ensemble precision).

least 20 residues long, which comes to 4,742 ensembles).[2] An ANSURR score of 100 is therefore by definition the PDB median; an ANSURR score of greater than 100 is "better than average" and is a good initial goal. We have produced an ensemble with an average ANSURR score of 117, which is well above 100, and with the best members at least as good as AlphaFold2, which we judge to be acceptable.

Despite this, the structure calculated using the final set of 47 HBRs is clearly not "correct" in that the flexibilities of residues 46–55 do not match well to the flexibility according to the backbone shifts (Figure 6G). As discussed above, this is probably a binding site for phosphate. It may be that addition of further restraints to a bound phosphate may improve the rigidity here, but this is not justified by the experimental NMR data.

Residues 99–112 are clearly moderately dynamic in solution. The rigidity of the crystal structure matches well to the experimental RCI data (Figure 4H), while unusually,[3] AlphaFold2 predicts this region to be more flexible than indicated by RCI (Figure 4I). HBRs have produced an overall rigidity in the NMR ensemble that matches the data reasonably well (Figure 6G). However, a comparison of the NMR-derived HBRs with hydrogen bonds in the crystal structure (Figure S9) shows that there are differences in hydrogen bonding in this region. This may indicate some structural heterogeneity in this loop in solution.

The refined structures have been analyzed (Figure 7). NOE violations (Figure 7A) change very little as HBRs are introduced, suggesting that the HBRs are consistent with the NOE restraints. Calculated amide proton shifts are slightly better in the refined structures (Figure 7F) than the unrefined structures (Figure 5F), as one might expect. The RDC *Q* factors (Figure 7E) improve by about 20% at all iterations. Similarly, Ramachandran distributions are better and more constant in the refined structures (Figure 7G), and RMSDs are rather smaller (Figure 7H). Overall, these statistics show that refinement in water consistently improves the structures at all stages. Taking all measures together, the results imply that adding HBRs in this gradual and validated manner represents a genuine improvement in accuracy.

A consensus NMR ensemble

Each of the 60 CYANA calculations carried out in the final iteration produced a different set of NOE restraints. We investigated several ways to use these to produce a consensus ensemble, and our preferred solution is based on the consensus bundle procedure[21]: the different sets of NOE restraints are combined to obtain a (smaller) consensus set of NOEs, which are then used to calculate a consensus ensemble. A big advantage of this method is that the RMSD spread (precision) of this ensemble is larger than is obtained from each of the individual calculations, and is a much more realistic number: in particular, the precision is expected to be of comparable magnitude to the accuracy, as obtained for example by comparing to a crystal structure.[21]

Our consensus NOE restraint set comprised restraints that appeared in at least 36 (60%) of the 60 restraint sets from the final iteration. We then used CYANA to produce 1000 models using only the consensus NOE restraints and TALOS-N dihedral restraints. Finally, we refined the 100 models with the best CYANA target function values in explicit solvent using CNS, adding in the 47 pairs of hydrogen bond restraints from the final iteration, and selected the 20 models with the best NOE energy as our final NMR ensemble, in the conventional structure calculation methodology (Figure 8A). The delayed addition of HBRs reduces the risk of some residues becoming trapped with incorrect

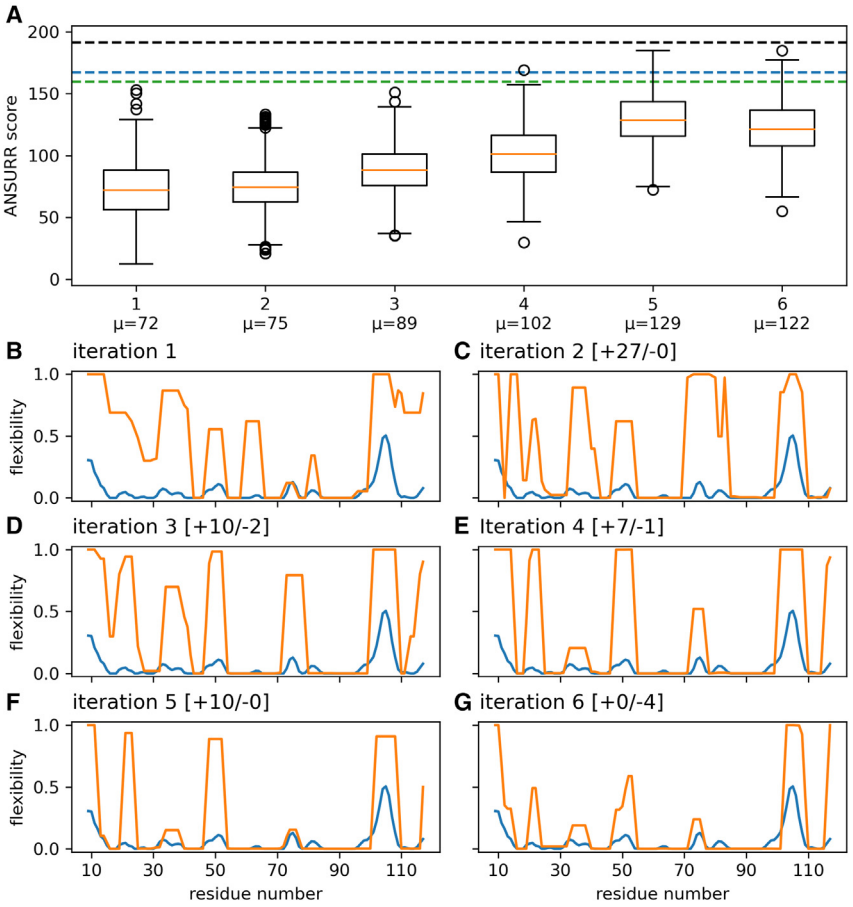


Figure 6. Improvements in ANSURR scores resulting from addition of hydrogen bond restraints, after refinement

(A) Boxplots showing the distribution of ANSURR scores for each iteration, following CNS refinement in solvent. Compare to those shown in Figure 4, which show the equivalent results before refinement. The dashed lines are as in Figure 4.

(B–G) Comparison of flexibility according to chemical shifts (blue) and flexibility computed for the best scoring model from each iteration (orange). The number of HBRs added and removed between each iteration is indicated in square brackets.

See also Figure S8.

torsion angles. It has been submitted to PDB as ID 8atk. Figure 8B shows the ANSURR scores for these structures. The distribution of ANSURR scores for the consensus ensemble is very similar to the ANSURR scores of the individual restraint sets of iteration 6, implying that the use of the consensus restraints has not produced any significant loss in accuracy. We also note that the lowest NOE energy structure (which by default would be model #1 and thus the representative ensemble structure) is one of the worst ANSURR scores in the ensemble, while the best ANSURR score is model 18, which is one of the highest NOE energies (although it is worth noting that all NOE energies are very small (Figure 7A) and therefore do not provide much discrimination). The PDB file therefore has model 18 as the first (and by default representative) model in the ensemble.

Table 1 presents a range of statistics for the ensemble. These measures serve different functions: some provide indications of the geometrical qualities of the structures, some provide indications of the accuracy of the structures, and some do both. The input restraints are fairly conventional: no RDCs and no unusual restraints, but a large number of carefully validated hydrogen bond restraints, which has been the focus of this work. The number and range of restraints encourage us to expect a well-defined but not outstandingly accurate structure: PSVS[33] calculates 2.9 restricting long-range distance restraints per restrained residue, which is average. The deviations from idealized geometry provide some reassurance that the structure calculation and energy minimization process has worked and that the structures are geometrically reasonable but provide no evidence on the accuracy of the structures. The list of restraint violations is a standard requirement and is unremarkable. Because CYANA discards NOE restraints that are repeatedly violated, such a list also provides little guidance on accuracy.

The remaining data in Table 1 provide a clearer estimate of accuracy. The RMS deviations from the mean structure are explicitly a measure of precision, not accuracy. However as seen here and elsewhere,[2,34] RMSD does correlate with measures of accuracy. These RMSD values are rather large for a good structure: however, it is important to note that the consensus calculation adopted here is deliberately aimed at producing as wide a spread of structures as possible and should ideally give a precision that is close to the true accuracy.[21]

It is widely assumed that the crystal structure is the closest one can get to the "correct" solution structure. Our work on ANSURR[1,2] has shown that the crystal structure is generally an accurate representation of the solution structure, though it tends to be too rigid. ANSURR reports that in general the crystal structure is considerably more accurate than the NMR ensemble; and a survey from the Montelione lab[35] showed that RMSD precision is usually tighter than the average distance between the NMR ensemble and the crystal structure (the "accuracy"), by a factor of 2.7 ± 1.3. We thus expect the RMS distance to the crystal structure to be small and ideally of the same magnitude as the consensus ensemble RMSD. Given the remarkable success of AlphaFold2 in predicting structures accurately,[3,36,37] we have also included an RMSD to AlphaFold, with similar expectations. The observation that these RMSD values are small, and of similar magnitude to the RMSD precision, is a good indication of the accuracy of the structures, and suggests that the precision obtained from the consensus ensemble is a better guide to the true accuracy, as claimed.[21]

In principle, the Ramachandran distribution tells nothing about the accuracy of the structure but merely reports on geometry. However, our experience with ANSURR[1,2] demonstrates that the Ramachandran score is a remarkably good guide to

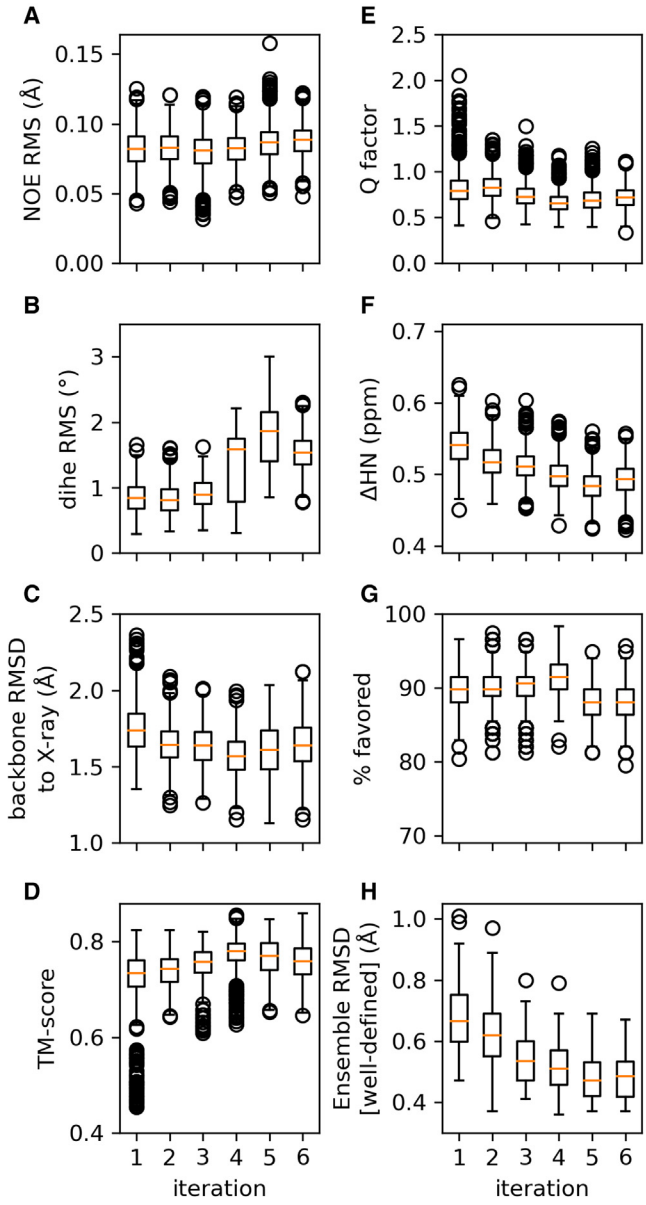


Figure 7. Change in metrics of structures refined in water during the iterative addition of HBRs

Compare to Figure 5.

(A) RMS distance violation.

(B) RMS dihedral angle violation.

(C) Average backbone RMSD to the crystal structure 5w3r.

(D) Average TM-score between NMR and crystal structures.

(E) Residual dipolar coupling *Q* factor.

(F) Mean difference between measured amide proton shifts and shifts calculated using SHIFTX2.

(G) Percentage of amino acids in the favored area of the Ramachandran plot.

(H) Average backbone RMSD to the mean (precision).

See also Figure S9.

accuracy. ANSURR itself reports that the ensemble is good but not excellent, with some members of the ensemble having high accuracy (Figure 8B). Our experience from this work and earlier studies[1,2] thus encourages us to see Ramachandran and ANSURR scores (together with the RMS to the crystal structure) as the best measures of ensemble accuracy. Table 1 also compares experimental and back-calculated amide proton shifts. This measure improved consistently during the iterations and could prove to be a simple and useful guide to accuracy.

Table 1 does not report the energies or NOE violation energies of the final ensemble. Such values are not normally reported, because they depend strongly on the details of the calculation and have little meaning. However, the rank ordering of energies is meaningful, so much so that the normal criterion for selection of the "best" structures from the ensemble is to use the structures with the lowest energy, as was done here. In other words, it is widely assumed that energy is a measure of the quality of an NMR structure. Figure 9 therefore compares ANSURR scores with energy. For total energy (Figure 9A), there is a large spread but a highly significant correlation, whereas the correlation with NOE energy (Figure 9B) is non-existent. These comparisons suggest that total energy may be a better measure of accuracy than NOE energy, and further confirm the validity of ANSURR scores as a measure of accuracy.

DISCUSSION

Protein structure determination by X-ray crystallography uses diffraction data as its input. Because there is a direct relationship between structure and diffraction pattern, the same data can also be used to check the accuracy of the resulting structure, via the *R* and R_{free} factors. NMR input data is much smaller in number, less precise and more diverse. There is also no simple relationship between the input data and output structure. The consequence is that structure calculation by NMR relies on generating an appropriate combination of different structural restraints, including a heavy reliance on knowledge-based restraints, where part of the challenge is to introduce and weight the different types of restraints appropriately. NMR has always had a problem with structure validation.[38] Here, we have taken the approach that some types of data (NOEs, dihedral restraints derived from backbone shifts, and hydrogen bonds inferred using a combination of structure, temperature coefficients and slow exchange) are used to generate restraints and/or structure selection criteria; while others (ANSURR scores based on rigidity, Ramachandran distribution, residual dipolar couplings [RDCs], back-calculated amide shifts, and [where appropriate] comparison to crystal or AlphaFold structure) are used to measure accuracy and not to produce restraints. If any of this second group is used to drive structure calculation or selection, then clearly it can no longer be used to measure accuracy. This classification is not unique: there are for example many structure calculations that have successfully used RDCs as restraints, in which case they should not be used to measure accuracy.

In this paper, we have proposed an iterative protocol for identifying HBRss based on experimental NMR data. The HBRs identified in this way were compatible with experimental restraints and significantly improved accuracy. The resulting ensemble had a lower energy, better ANSURR score, better Ramachandran distribution, better similarity to the crystal structure, better RDC *Q* factor, and lower RMSD than ensembles calculated with fewer HBRs. We suggest that this protocol could be used as the basis for a robust, transparent, and automatable system

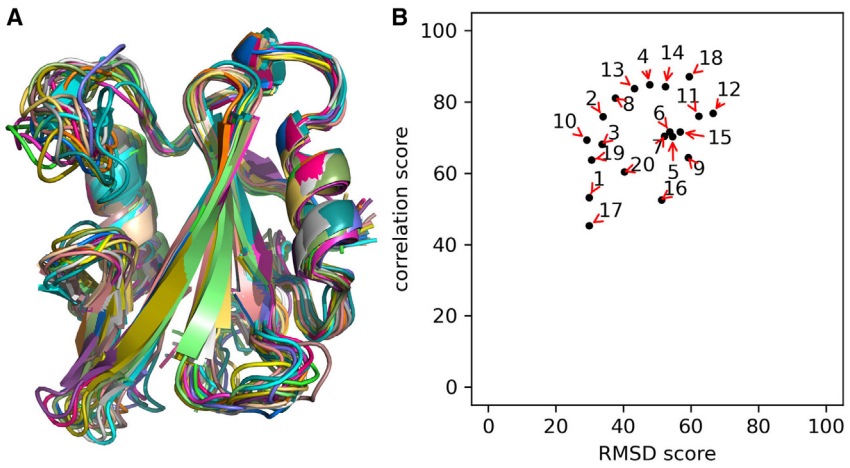


Figure 8. The ensemble formed by selecting the 20 lowest NOE energy models from the final solvent-minimized consensus ensemble

(A) Superposition of the cartoon plot of the structures. In the view shown here, the phosphate binding loop is in the bottom center, while the dynamic loop 99–112 is top left. The typical SH2 binding site for phosphotyrosine peptides has a pY-binding pocket above the phosphate loop, and a second hydrophobic pocket on the left side of the long β-strand at the front.

(B) ANSURR scores for these 20 structures. Numbering is in order of NOE energy, with 1 being the lowest.

for calculating better NMR structure ensembles. We note several key features of the protocol: HBRs were introduced gradually and iteratively, with a final step where possible incompatible HBRs were removed; the initial calculations used only NOEs and dihedral restraints; and multiple ensembles were calculated at each stage, to obtain better discrimination between correct and incorrect HBRs. The final calculation used consensus distance restraints in order to sample conformational space better and obtain a more realistic and useful RMSD.[21]

The method is here demonstrated on a single small and reasonably well behaved protein. To improve confidence in the method, it needs repeating on a range of proteins. We expect that very similar methods would be applicable to other proteins, with only minor changes to the protocol.

Only backbone-backbone HBRs were applied here. One could easily imagine adding sidechain restraints, although the experimental evidence required might be harder to obtain. For SH2, there are two pairs of sidechain hydrogen bond restraints for which we have clear evidence. (a) The sidechain amides of Asn59 are very slowly exchanging. In our structures, they are buried inside the structure and both form well-defined hydrogen bonds to the backbone carbonyls of V31 and G35, even without restraints being added. Thus the restraints could be added, but would actually not make any difference, because hydrogen bonding is already present due to the local density of restraints. (b) It is well established that α-helices often have caps at both ends, particularly at the N-terminal end where sidechains of Ser, Thr, or Asn form hydrogen bonds to (*i*+4) residues within the helix.[39] Both of the SH2 helices have serines as the N-capping residue, and slowly exchanging amides at the relevant positions. We have recalculated ensembles with these N-cap restraints in place. Structures are well formed and of low energy, but the added restraints make little difference to the overall energy or accuracy.

Many of the validation measures used show an improvement in accuracy with addition of extra HBRs (Figures 5 and 7). In the absence of a "true" structure for comparison, the best measures of accuracy across the entire refinement process appear to be ANSURR score and Ramachandran distribution. Ramachandran is not ideal because it measures how good a protein structure it is, rather than how well it matches experimental data, and therefore is liable to give misleading results if the calculation is biased toward known "good" structures. RDCs are useful for validation, though not if they are already being used as restraints. Compared to ANSURR, they lack power to differentiate between poorer quality structures.[36] We therefore propose ANSURR score as a useful, simple and rapid measure of accuracy. In particular, we noted above that by definition an ANSURR score of 100 is "average". Thus, any ensemble that achieves an ANSURR score of greater than 100 is "better than average" and is at least acceptable. If all new NMR structures in the PDB had an ensemble ANSURR score greater than 100, this would lead to a gradual overall improvement in the accuracy of NMR structures in the PDB.

During the development of the iteration protocol, we deliberately did not compare our hydrogen bonds to those observed in the crystal structure. On making such a comparison (Figure S9), there is a good agreement between the hydrogen bonds in the crystal and those identified here.

One might imagine that incorporation of explicit HBRs is not necessary, in that if a structure is almost correct, then refinement in explicit solvent should drive the formation of the correct hydrogen bonds. However, this does not in general seem to be true. There is some improvement in hydrogen bonding arising from structure refinement,[40] but it seems that the geometry has to be very close to correct in order for the forcefield to produce a hydrogen bond. We note that Schwieters et al.[25] have introduced a new potential into Xplor-NIH that may improve the situation; and that AI-based methods such as AlphaFold have good success in predicting hydrogen bonds. Nonetheless, it is currently important to maximize the number of correct HBRs in NMR structure calculation. We hope that the proposals outlined here will help.

Conclusions

The future of biological NMR as a structural tool in the post-AlphaFold era lies in its ability to detect and characterize structural and dynamic heterogeneity in solution. To carry out that task effectively it is essential that the structural details from NMR should be as accurate as possible. Here we have shown that hydrogen bond restraints can be added in a transparent and justified way, significantly improving the accuracy of the structure. We have also shown that ANSURR is a useful tool for assessing accuracy, and can be used to judge when the structure determination process is ready to complete.

Table 1. Structural statistics for the consensus SH2 structure ensemble

Parameter[a]	Value
Experimentally derived restraints	
Total unambiguous NOE restraints[b]	1288
Intraresidue	427
Sequential	401
Medium range	142
Long range ($\Delta \geq 5$)	318
NOE restraints per restrained residue	11.7
Dihedral angle restraints[c]	173
Hydrogen bond restraints	47 x 2
RMS deviations from idealised geometry	
Bond lengths (Å)	0.005
Angle geometry (°)	0.8
Deviations from experimental restraints	
RMS NOE violation (Å)	0.08
NOE violations per model >0.1 Å	8.4
RMS angle violation (°)	3.04
Angle violations per model >1°	14.9
RMS coordinate deviations[d]	
Backbone RMSD from mean structure (Å)	1.1
Heavy atom RMSD from mean structure (Å)	1.5
Heavy atom RMSD from crystal structure (Å)	1.68
Heavy atom RMSD from best AlphaFold2 (Å)	1.77
Ramachandran analysis[e]	
Most favored regions (%)	90.1
Additionally allowed regions (%)	9.9
Generously allowed regions (%)	0.1
Disallowed regions (%)	0.0
Other measures of quality	
ANSURR score	117
RDC Q[f]	0.63
RMSD of HN shifts compared to SHIFTX2 (ppm)	0.50

Values are averages over the final ensemble of 20 structures.
[a]Unless otherwise stated, the values here are calculated by PSVS 1.5.[33]
[b]Values are those used in the final consensus calculation, as discussed in the text. The consensus upper bound restraints represent a significant reduction in the number of restraints compared to those in each of the 60 runs, which had roughly 1800 in each.
[c]Derived from TALOS-N.[5] Only restraints on φ and ψ were used.
[d]For residues defined as structured by PSVS, ie residues 12–49, 53–76, 82–103 and 108–115.
[e]As calculated by Procheck.[41] using residues defined as structured.
[f]Calculated using PALES[42] using 58 NH RDCs with values varying between −22 and +27 Hz.

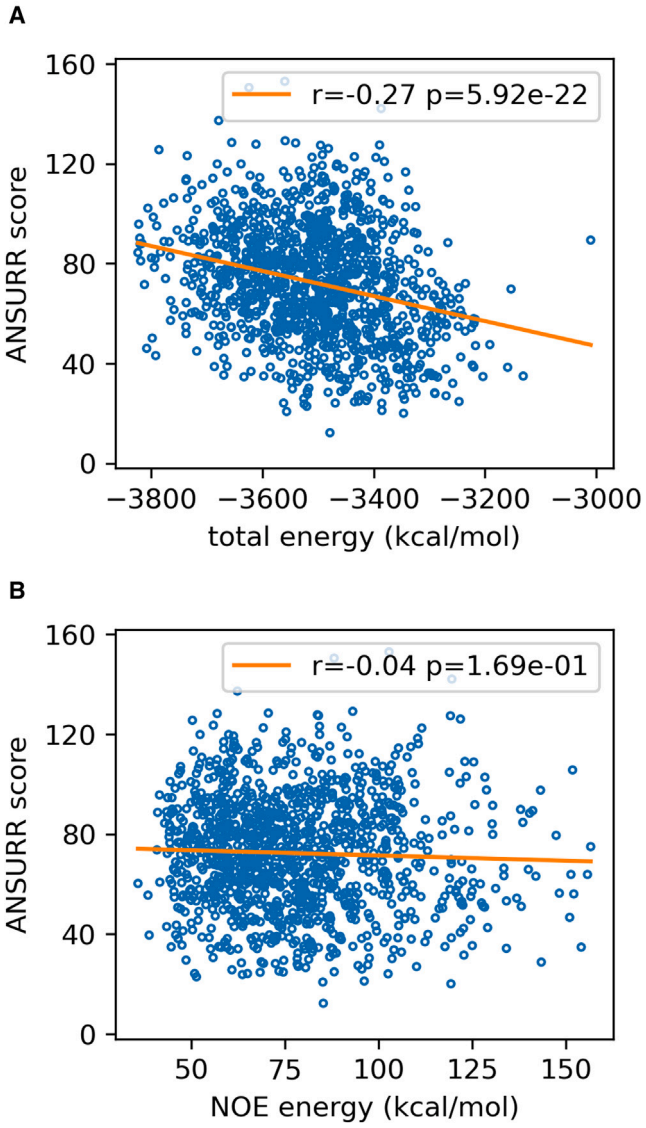


Figure 9. Correlation between ANSURR score and energy
Shown for the 1200 structures calculated in the final iteration and refined in explicit solvent. The orange line is the line of best fit, and the figures give the Pearson correlation coefficient and its two-tailed p value. (A) Total energy (B) NOE energy.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - A set of comparable NMR and X-ray structures
  - Protein expression and purification
  - NMR assignment
  - Structure restraints and calculations
  - Structure predictions using AlphaFold and RosettaFold
  - Other calculations
- QUANTIFICATION AND STATISTICAL DETAILS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.str.2023.05.012.

ACKNOWLEDGMENTS

We thank the Biotechnology and Biological Sciences Research Council for funding to N. J. F. (BB/P020038/1) and for support to upgrade the NMR spectrometer (BB/R000727/1).

AUTHOR CONTRIBUTIONS

Conceptualization, N.J.F. and M.P.W.; Investigations, N.J.F., M.F.A, S.L., and A.M.H.; Analysis, N.J.F., M.F.A., and M.P.W.; Writing, N.J.F. and M.P.W.; Funding Acquisition, M.P.W.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 10, 2023
Revised: May 2, 2023
Accepted: May 18, 2023
Published: June 12, 2023

REFERENCES

1. Fowler, N.J., Sljoka, A., and Williamson, M.P. (2020). A method for validating the accuracy of NMR protein structures. Nat. Commun. *11*, 6321. https://doi.org/10.1038/s41467-020-20177-1.
2. Fowler, N.J., Sljoka, A., and Williamson, M.P. (2021). The accuracy of NMR protein structures in the Protein Data Bank. Structure *29*, 1430–1439.e2. https://doi.org/10.1016/j.str.2021.07.001.
3. Fowler, N.J., and Williamson, M.P. (2022). The accuracy of protein structures in solution determined by AlphaFold and NMR. Structure *30*, 1–9. https://doi.org/10.1101/2022.01.18.476751.
4. Neuhaus, D., and Williamson, M.P. (2000). The Nuclear Overhauser Effect in Structural and Conformational Analysis, 2nd Edition (Wiley-VCH).
5. Shen, Y., and Bax, A. (2013). Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. J. Biomol. NMR *56*, 227–241. https://doi.org/10.1007/s10858-013-9741-y.
6. Cordier, F., and Grzesiek, S. (1999). Direct observation of hydrogen bonds in proteins by interresidue $^{3h}J_{NC'}$ scalar couplings. J. Am. Chem. Soc. *121*, 1601–1602. https://doi.org/10.1021/ja983945d.
7. Wüthrich, K., Billeter, M., and Braun, W. (1984). Polypeptide secondary structure determination by nuclear magnetic resonance observation of short proton-proton distances. J. Mol. Biol. *180*, 715–740. https://doi.org/10.1016/0022-2836(84)90034-2.
8. Wagner, G., Kumar, A., and Wüthrich, K. (1981). Systematic application of two-dimensional ^{1}H NMR techniques for studies of proteins. 2. Combined use of correlated spectroscopy and nuclear Overhauser spectroscopy for sequential ssignments of backbone resonances and elucidation of polypeptide secondary structures. Eur. J. Biochem. *114*, 375–384. https://doi.org/10.1111/j.1432-1033.1981.tb05157.x.
9. Williamson, M.P., Havel, T.F., and Wüthrich, K. (1985). Solution conformation of proteinase inhibitor IIA from bull seminal plasma by ^{1}H nuclear magnetic resonance and distance geometry. J. Mol. Biol. *182*, 295–315.
10. Baxter, N.J., and Williamson, M.P. (1997). Temperature dependence of ^{1}H chemical shifts in proteins. J. Biomol. NMR *9*, 359–369.
11. Maures, T.J., Kurzer, J.H., and Carter-Su, C. (2007). SH2B1 (SH2-B) and JAK2: a multifunctional adaptor protein and kinase made for each other. Trends Endocrinol. Metabol. *18*, 38–45.
12. Li, Z., Zhou, Y., Carter-Su, C., Myers, M.G., Jr., and Rui, L. (2007). SH2B1 enhances leptin signaling by both janus kinase 2 Tyr^{813} phosphorylation-dependent and -independent mechanisms. Mol. Endocrinol. *21*, 2270–2281. https://doi.org/10.1210/me.2007-0111.
13. Morris, D.L., Cho, K.W., Zhou, Y., and Rui, L. (2009). SH2B1 enhances insulin sensitivity by both stimulating the insulin receptor and inhibiting tyrosine dephosphorylation of insulin receptor substrate proteins. Diabetes *58*, 2039–2047. https://doi.org/10.2337/db08-1388.
14. Rui, L., Gunter, D.R., Herrington, J., and Carter-Su, C. (2000). Differential binding to and regulation of JAK2 by the SH2 domain and N-terminal region of SH2-B β. Mol. Cell Biol. *20*, 3168–3177. https://doi.org/10.1128/mcb.20.9.3168-3177.2000.
15. Jacobs, D.J., Rader, A.J., Kuhn, L.A., and Thorpe, M.F. (2001). Protein flexibility predictions using graph theory. Proteins *44*, 150–165. https://doi.org/10.1002/prot.1081.
16. Berjanskii, M.V., and Wishart, D.S. (2008). Application of the random coil index to studying protein flexibility. J. Biomol. NMR *40*, 31–48. https://doi.org/10.1007/s10858-007-9208-0.
17. López-Méndez, B., and Güntert, P. (2006). Automated protein structure determination from NMR spectra. J. Am. Chem. Soc. *128*, 13112–13122.
18. Güntert, P., and Buchner, L. (2015). Combined automated NOE assignment and structure calculation with CYANA. J. Biomol. NMR *62*, 453–471. https://doi.org/10.1007/s10858-015-9924-9.
19. Williamson, M.P., and Craven, C.J. (2009). Automated protein structure calculation from NMR data. J. Biomol. NMR *43*, 131–143.
20. Ulrich, E.L., Akutsu, H., Doreleijers, J.F., Harano, Y., Ioannidis, Y.E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., et al. (2008). BioMagResBank. Nucleic Acids Res. *36*, D402–D408.
21. Buchner, L., and Güntert, P. (2015). Increased reliability of nuclear magnetic resonance protein structures by consensus structure bundles. Structure *23*, 425–434. https://doi.org/10.1016/j.str.2014.11.014.
22. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.
23. Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., Wang, J., Cong, Q., Kinch, L.N., Schaeffer, R.D., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. Science *373*, 871–876. https://doi.org/10.1126/science.abj8754.
24. McKercher, M.A., Guan, X., Tan, Z., and Wuttke, D.S. (2018). Diversity in peptide recognition by the SH2 domain of SH2B1. Proteins *86*, 164–176. https://doi.org/10.1002/prot.25420.
25. Schwieters, C.D., Bermejo, G.A., and Clore, G.M. (2020). A three-dimensional potential of mean force to improve backbone and sidechain hydrogen bond geometry in Xplor-NIH protein structure determination. Protein Sci. *29*, 100–110. https://doi.org/10.1002/pro.3745.
26. Zhang, Y., and Skolnick, J. (2004). Scoring function for automated assessment of protein structure template quality. Proteins *57*, 702–710. https://doi.org/10.1002/prot.20264.
27. Chiliveri, S.C., Robertson, A.J., Shen, Y., Torchia, D.A., and Bax, A. (2022). Advances in NMR spectroscopy of weakly aligned biomolecular systems. Chem. Rev. *122*, 9307–9330. https://doi.org/10.1021/acs.chemrev.1c00730.
28. Cornilescu, G., Marquardt, J.L., Ottiger, M., and Bax, A. (1998). Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. J. Am. Chem. Soc. *120*, 6836–6837. https://doi.org/10.1021/ja9812610.
29. Asakura, T., Taoka, K., Demura, M., and Williamson, M.P. (1995). The relationship between amide proton chemical shifts and secondary structure in proteins. J. Biomol. NMR *6*, 227–236.
30. Han, B., Liu, Y., Ginzinger, S.W., and Wishart, D.S. (2011). SHIFTX2: significantly improved protein chemical shift prediction. J. Biomol. NMR *50*, 43–57.

31. Linge, J.P., Williams, M.A., Spronk, C.A.E.M., Bonvin, A.M.J.J., and Nilges, M. (2003). Refinement of protein structures in explicit solvent. Proteins *50*, 496–506. https://doi.org/10.1002/prot.10299.

32. Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D *54*, 905–921.

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

34. Clore, G.M., and Gronenborn, A.M. (1998). New methods of structure refinement for macromolecular structure determination by NMR. Proc. Natl. Acad. Sci. USA *95*, 5891–5898. https://doi.org/10.1073/pnas.95.11.5891.

35. Andrec, M., Snyder, D.A., Zhou, Z., Young, J., Montelione, G.T., and Levy, R.M. (2007). A large data set comparison of protein structures determined by crystallography and NMR: statistical test for structural differences and the effect of crystal packing. Proteins *69*, 449–465.

36. Huang, Y.J., Zhang, N., Bersch, B., Fidelis, K., Inouye, M., Ishida, Y., Kryshtafovych, A., Kobayashi, N., Kuroda, Y., Liu, G., et al. (2021). Assessment of prediction methods for protein structures determined by NMR in CASP14: impact of AlphaFold2. Proteins *89*, 1959–1976.

37. Robertson, A.J., Courtney, J.M., Shen, Y., Ying, J., and Bax, A. (2021). Concordance of X-ray and AlphaFold2 models of SARS-CoV-2 main protease with residual dipolar couplings measured in solution. J. Am. Chem. Soc. *143*, 19306–19310. https://doi.org/10.1021/jacs.1c10588.

38. Montelione, G.T., Nilges, M., Bax, A., Güntert, P., Herrmann, T., Richardson, J.S., Schwieters, C.D., Vranken, W.F., Vuister, G.W., Wishart, D.S., et al. (2013). Recommendations of the wwPDB NMR validation task force. Structure *21*, 1563–1570. https://doi.org/10.1016/j.str.2013.07.021.

39. Aurora, R., and Rose, G.D. (1998). Helix capping. Protein Sci. *7*, 21–38. https://doi.org/10.1002/pro.5560070103.

40. Schneider, M., Fu, X., and Keating, A.E. (2009). X-ray vs. NMR structures as templates for computational protein design. Proteins *77*, 97–110. https://doi.org/10.1002/prot.22421.

41. Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR *8*, 477–486.

42. Zweckstetter, M., and Bax, A. (2000). Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR. J. Am. Chem. Soc. *122*, 3791–3792. https://doi.org/10.1021/ja0000908.

43. Dana, J.M., Gutmanas, A., Tyagi, N., Qi, G., O'Donovan, C., Martin, M., and Velankar, S. (2019). SIFTS: updated Structure Integration with Function, Taxonomy and Sequences resource allows 40-fold increase in coverage of structure-based annotations for proteins. Nucleic Acids Res. *47*, D482–D489. https://doi.org/10.1093/nar/gky1114.

44. Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. J. Mol. Biol. *273*, 283–298.

45. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. Nat. Methods *19*, 679–682. https://doi.org/10.1038/s41592-022-01488-1.

46. Chen, V.B., Arendall, W.B., III, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. *66*, 12–21. https://doi.org/10.1107/s0907444909042073.

47. Wells, S.A. (2020). FLEXOME Software Suite.

48. Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., III, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345. https://doi.org/10.1038/nmeth.1318.

49. Reed, M.A.C., Hounslow, A.M., Sze, K.H., Barsukov, I.G., Hosszu, L.L.P., Clarke, A.R., Craven, C.J., and Waltho, J.P. (2003). Effects of domain dissection on the folding and stability of the 43 kDa protein PGK probed by NMR. J. Mol. Biol. *330*, 1189–1201.

50. Güntert, P. (2004). Automated NMR structure calculation with CYANA. Methods Mol. Biol. *278*, 353–378.

51. Bartels, C., Güntert, P., Billeter, M., and Wüthrich, K. (1997). Garant - a general algorithm for resonance assignment of multidimensional nuclear magnetic resonance spectra. J. Comput. Chem. *18*, 139–149.

52. Tomlinson, J.H., and Williamson, M.P. (2012). Amide temperature coefficients in the protein G B1 domain. J. Biomol. NMR *52*, 57–64.

53. Chou, J.J., Gaemers, S., Howder, B., Louis, J.M., and Bax, A. (2001). A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles. J. Biomol. NMR *21*, 377–382. https://doi.org/10.1023/a:1013336502594.

54. Ottiger, M., Delaglio, F., and Bax, A. (1998). Measurement of *J* and dipolar couplings from simplified two-dimensional NMR spectra. J. Magn. Reson. *131*, 373–378. https://doi.org/10.1006/jmre.1998.1361.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
BL21 (DE3) *E. coli*	Thermo Fisher	EC0114
Deposited data		
SH2 domain	This paper	PDB: 8atk
SH2 domain	This paper	BMRB: 51342
SIFTS	Dana et al., 2019[43]	https://www.ebi.ac.uk/pdbe/docs/sifts/
Recombinant DNA		
pET-28 plasmid	Merck	69864
Software and Algorithms		
ANSURR v1.2.1	Fowler et al., 2020[1]	https://zenodo.org/record/5655244
CYANA 3.98.5	Güntert et al., 1997[44]	https://www.las.jp/english/products/cyana.html
TALOS-N	Shen & Bax, 2013[5]	https://spin.niddk.nih.gov/bax/software/TALOS-N/
CNS 1.3	Brunger et al., 1998[32]	http://cns-online.org/v1.3/
PALES	Zweckstetter & Bax, 2000[42]	https://spin.niddk.nih.gov/bax/software/PALES
AlphaFold 1.2.0	Mirdita et al., 2022[45]	https://colab.research.google.com/github/sokrypton/ColabFold/blob/v1.2.0/AlphaFold2.ipynb
RosettaFold	Baek et al., 2021[23]	https://robetta.bakerlab.org/
Molprobity	Chen et al., 2010[46]	http://molprobity.biochem.duke.edu/
SHIFTX2 1.10A	Han et al., 2011[30]	http://www.shiftx2.ca/
Flexome	Wells, 2020[47]	https://doi.org/10.15125/BATH-00940
PyMol molecular graphics system	Schrödinger, LLC	https://pymol.org/2/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for information on method, dataset or computational resources should be directed to and will be fulfilled by the Lead Contact, Prof. M. P. Williamson (m.williamson@sheffield.ac.uk).

Materials availability

The plasmid for expression of SH2 is available on request.

Data and code availability

Data availability

All data reported in this paper will be shared by the lead contact upon request. Coordinates have been deposited at PDB and chemical shifts at BMRB, and are publicly available as of the date of publication. Accession codes are listed in the key resources table.

Code availability

The data reported in this study were obtained using ANSURR v1.2.1 (https://doi.org/10.5281/zenodo.5655244) with the option to re-reference chemical shifts using PANAV. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

E. coli BL21 (DE3) cells were grown in labeled M9 medium plus ampicillin.

METHOD DETAILS

A set of comparable NMR and X-ray structures

The Structure Integration with Function, Taxonomy and Sequence (SIFTS) resource[43] was used to match NMR structures to X-ray structures. NMR structures were required to have at least 80% backbone chemical shift completeness. X-ray structures were selected based on those which gave most coverage of the NMR structure, then which of those had the best resolution. Only pairs where the X-ray structure covered at least 80% of the NMR structure were used. Redundancy was reduced by randomly selecting a single structure pair for each UniProt accession number (resulting in 222 pairs) before applying a 50% sequence identity cut-off, leaving 215 structure pairs for analysis. A list of paired NMR and X-ray structures is provided in Table S1.

Protein expression and purification

The gene for SH2 was made synthetically (Integrated DNA Technologies). The gene sequence is that of mouse SH2 (Uniprot: Q91ZM2), codon optimised for expression in *E. coli* (Integrated DNA Technologies). The synthetic gene was 461 bp long and consisted of N-terminal homology region, ribosome binding site, start codon, His tag, protein sequence (118 amino acids), stop codon and C-terminal homology region. It was inserted into the pET28a plasmid, which had been linearised by cutting with Xba1 and Xho1, following standard Gibson protocol.[48] The plasmid was transformed into BL21 (DE3) *E. coli*, and grown in labelled M9 medium supplemented with 100 μg/ml ampicillin. It was incubated, induced with 0.5 mM IPTG, and grown overnight at 25°C. Cells were lysed by sonication and protein was purified using a Ni-NTA column, eluting with 300 mM imidazole, followed by polishing using a Superdex 200 column. For NMR, buffer was exchanged to 50 mM potassium phosphate pH 6, containing 1 mM trimethylsilyl propionate (TSP) as an internal standard, in 10% D_2O. The protein concentration was approximately 1 mM in a Shigemi tube.

NMR assignment

NMR data were collected on a Bruker DRX-600 equipped with a cryoprobe. Manual assignment was done using the program Astools,[49] which is based on a Monte-Carlo simulated annealing method. Three backbone amides in the HSQC spectrum have low intensity (Trp17, Gln61 and His81). The manual backbone assignment was complete except for Gly20, which appeared to be absent, and the two N-terminal residues. The final sequential assignment of SH2 backbone resonances showed that in total 95% of backbone resonances for non-proline residues 9 to 118 were assigned (96% of ^{1}H, 96% of ^{15}N, 96% of Cα, 94% of Cβ, 92% of C'), the missing signals being due to Trp17, Gly20, Gln61, and His81. Most sidechains were assigned, with the exceptions of some ends of sidechains, Phe42, Ile99, and Val115.

The automated chemical shift assignment was performed in two steps, following the recommended procedure[50]: an ensemble of chemical shift assignments was computed followed by combining the resultant raw chemical shift assignments into a single consensus resonance list. The ensemble of chemical shift assignments was obtained from 20 runs of the GARANT algorithm (a component of the FLYA/CYANA package), in which the iteration size for one generation was 100. Each independent run started from the same experimental peak lists but using a different random seed value, and optimised the match between observed peaks and expected peaks based on the knowledge of the amino acid sequence and the magnetization transfer pathways in the spectra used.[51] The matching was done with the recommended tolerance values of 0.03 and 0.4 ppm for the ^{1}H dimension and for the ^{13}C and ^{15}N dimensions, respectively.

Structure restraints and calculations

Calculations were carried out using standard procedures in CYANA 3.98.5.[44] Tolerance values for the chemical shift matching were 0.04 ppm for ^{1}H dimension, 0.03 ppm for ^{15}N or ^{13}C bound ^{1}H dimension, and 0.45 ppm for ^{15}N and ^{13}C dimensions. ^{15}N NOESY, ^{13}C NOESY for aliphatic atoms, and ^{13}C NOESY for aromatic atoms (mixing times 100 ms) were sorted in XEASY format which contains peak positions and volumes. The input to CYANA was a list of manually checked assignments plus NOE pick lists, dihedral restraints and HBRs. Dihedral angle restraints (φ and ψ) were generated according to the backbone chemical shifts in the SH2 protein using the TALOS-N program.[5] 93 pairs of TALOS-N angles that were predicted as strong or generous were converted into torsion angle restraints using a CYANA macro. No χ_1 restraints were used. Error values were given a default value of twice the standard deviation listed by TALOS-N. The predicted dihedral angles of proline residues were excluded from the CYANA torsion angles list, even for those classified as strong, namely 98, 100 and 116. The calculation started with 100 conformers generated from random torsion angle values. Each conformer is generated after 10000 torsion angle dynamics steps. For each HBR iteration, CYANA was run with different random seeds, selecting the 20 structures with the lowest target function out of a set of 100. This was duplicated a total of 60 times at each iteration starting from the same NOE peak lists, implying that the NOE restraints assigned will be different for each of these duplicate calculations. At each HBR iteration, the HBRs generated from the previous iteration were applied as starting restraints. For the final iteration of CYANA calculations, the 60 different sets of NOE restraints from the previous iteration were analysed, and restraints were adopted as consensus restraints if they were identified in 36 out of the 60 sets, using the largest upper bound distance from the set. This resulted in a reduction in the total number of NOE restraints used, from roughly 1800 to 1344.

Subsequent refinements in explicit solvent were carried out in CNS[32] using the protein-allhdg parameters. CYANA distance restraints were converted into upper distance restraints in CNS format, and TALOS-N dihedral restraints were regenerated in CNS format.

Additional structure restraints were obtained from temperature coefficients and from amide exchange. For temperature coefficients, HSQC spectra were obtained at 288, 293, 298, and 303 K. All chemical shifts were referenced to TSP at 0 ppm. Coefficients were obtained by least squares fitting of a straight line. Amides defined as potential hydrogen bond donors are those with fitted temperature coefficients of -4.5 ppb/K or greater.[10,52] Amide exchange rates were measured by lyophilising the protein from a solution at pH 5.5, redissolving in D_2O, and running a series of HSQC spectra at 288 K. Slowly exchanging amides are those visible in HSQC spectra started within 9 minutes of dissolution. A full list of amides used can be found in Table S3.

Residual dipolar coupling restraints (RDCs) were obtained using compressed polyacrylamide gels. Gels were poured and set in open-ended 5 mm NMR tubes with an inner diameter of approximately 4.2 mm, using 7% acrylamide (an 18:1 w/w mixture of acrylamide to N,N'-methylenebisacrylamide) and polymerised using 0.1% w/v ammonium persulfate and 0.1% v/v N,N,N',N'-tetramethyethylenediamine (TEMED).[53] Gels were left overnight to set, pushed out of the tube, washed extensively in distilled water, and dried in an oven at 37°C under a glass cover until they had shrunk to roughly one half of their original dimensions. They were then inserted into a Shigemi tube in the presence of 380 μM protein in 50 mM phosphate, pH 6, 10% D_2O, and the Shigemi plunger was inserted so that the swelled gel would be 45% compressed relative to its initial height. RDCs were measured using the Bruker IPAP pulse program hsqcf3gpiaphwg.2,[54] and processed to separate the two peaks into two different spectra, using zero filling in the ^{15}N dimension to achieve 0.5 Hz/pt digital resolution. The estimated accuracy of measurement of RDCs is ± 1 Hz, and RDCs were in the range of -22 to +27 Hz. RDC *Q* factors were calculated using PALES.[42]

Structure predictions using AlphaFold and RosettaFold

AlphaFold predictions were obtained using ColabFold v1.2[45] (https://colab.research.google.com/github/sokrypton/ColabFold/blob/v1.2.0/AlphaFold2.ipynb) with the following options: MSA mode: MMseqs2 (UniRef+Environmental); use templates: yes; number of recycles: 3; refine with amber: yes. RosettaFold predictions were obtained using the Robetta web server[23] (https://robetta.bakerlab.org/) with the default options.

Other calculations

Ramachandran distributions were calculated using ramalyze, part of the Molprobity suite.[46] Amide proton chemical shift calculations were made using SHIFTX2 1.10A.[30] Hydrogen bonds in crystal and NMR structures (Figure 1) were identified using FLEXOME [https://doi.org/10.15125/BATH-00940][47] using a cutoff of 0 kcal/mol for hydrogen bond energy.

QUANTIFICATION AND STATISTICAL DETAILS

Statistical details can be found in the figure legends. All statistical tests, including Pearson correlation, box and whisker plots, and two-tailed *p*-value, were carried out using standard unix routines.