Using NMR Chemical Shifts as Structural Restraints in Molecular Dynamics Simulations of Proteins

Paul Robustelli,[1] Kai Kohlhoff,[1] Andrea Cavalli,[1] and Michele Vendruscolo[1,*]
[1]Department of Chemistry, University of Cambridge Lensfield Road, Cambridge CB2 1EW, UK
*Correspondence: mv245@cam.ac.uk

SUMMARY

We introduce a procedure to determine the structures of proteins by incorporating NMR chemical shifts as structural restraints in molecular dynamics simulations. In this approach, the chemical shifts are expressed as differentiable functions of the atomic coordinates and used to compute forces to generate trajectories that lead to the reduction of the differences between experimental and calculated chemical shifts. We show that this strategy enables the folding of a set of proteins with representative topologies starting from partially denatured initial conformations without the use of additional experimental information. This method also enables the straightforward combination of chemical shifts with other standard NMR restraints, including those derived from NOE, J-coupling, and residual dipolar coupling measurements. We illustrate this aspect by calculating the structure of a transiently populated excited state conformation from chemical shift and residual dipolar coupling data measured by relaxation dispersion NMR experiments.

INTRODUCTION

There has recently been significant progress in the development of computational techniques for the determination of protein structures from NMR chemical shifts (Berjanskii et al., 2009; Cavalli et al., 2007; Das et al., 2009; Montalvao et al., 2008; Robustelli et al., 2008, 2009; Shen et al., 2008, 2009; Wishart et al., 2008), which are the most readily and accurately measured NMR observables. It has been shown that the structures of proteins (Berjanskii et al., 2009; Cavalli et al., 2007; Robustelli et al., 2008, 2009; Shen et al., 2008, 2009; Wishart et al., 2008) and protein complexes (Das et al., 2009; Montalvao et al., 2008) can be determined with a molecular fragment approach (Delaglio et al., 2000; Simons et al., 1997) that utilizes sequence homology information together with chemical shift restraints. In these methods, structural motifs are selected from existing protein structures on the basis of chemical shift and sequence homology and assembled to generate a set of candidate structures; these structures are then refined using state-of-the-art force fields and chemical shift information.

These studies have demonstrated that the information contained in backbone chemical shifts, when used in combination with structural databases and molecular mechanics force fields, can be sufficient to determine the structures of proteins and may allow for a significant reduction in the amount of data acquisition required for the determination of the native structure of proteins. The use of the structural information obtained from chemical shifts with molecular fragment replacement approaches is not, however, easily combined with other data used in standard NMR structure calculations, such as NOEs, J-couplings, and residual dipolar couplings (Brunger, 2007; Schwieters et al., 2006), or readily used to calculate conformational changes from existing structural models, such as those required, for example, to define the effects ligand binding events, changes in solution conditions, chemical modifications, or amino acid substitutions. One way to achieve these goals is to develop an approach in which chemical shifts are incorporated as restraints in a manner analogous to that adopted in standard NMR structure calculations, where a penalty function is used to bias a conformational search toward structures consistent with the available experimental restraints. This strategy has been previously used to perform a structural refinement of previously determined structures (Clore and Gronenborn, 1998; Kuszewski et al., 1995; Schwieters et al., 2006). We have also recently shown that it is possible to determine the structures of small proteins from extended conformations by directly incorporating chemical shifts as structural restraints in a Monte Carlo search of the conformational space (Robustelli et al., 2009); in this approach a tunable soft-square harmonic well was used to convert differences between experimental and calculated backbone chemical shifts into a penalty function to direct the conformational search toward regions of conformational space consistent with the measured chemical shifts. We have demonstrated that by adaptively tuning the parameters of a chemical shift penalty function (in particular the size of the flat-bottom tolerance, which designates the magnitude of the differences between experimental and calculated chemical shifts that produce energetic penalties), it is possible to overcome, at least for short polypeptide chains, the rugged nature of the restrained energy landscape and achieve the large structural adjustments required to satisfy the chemical shift restraints (Robustelli et al., 2009).

The extension of this type of approach to establish more general structure calculation protocols has been challenging because of the complexity of the mapping between chemical shifts and atomic coordinates (Wishart and Case, 2001; Xu and Case, 2002), which makes it very difficult to explore

Table 1. Summary of the Calculations Presented in This Work

Reference Structure (PDB Code)	Protein Length	rmsd of the Starting Structure	rmsd of the Sampled Structure of Lowest Energy		Sampled Structure of Lowest rmsd	
			CS-MD	Control MD	CS-MD	Control MD
2oed	56	3.48	0.84	5.63	0.62	3.01
2jvw	61	6.40	1.11	3.10	0.93	1.79
2jtv	64	3.23	1.30	2.35	0.97	1.38
2jvm	55	4.63	1.51	4.02	1.11	2.43
2jxt	78	3.32	1.59	7.90	1.31	2.30
2jt1	71	6.76	1.67	4.40	1.56	4.01
2jva	108	5.25	1.88	6.23	1.18	2.75
1d3z	76	3.57	1.92	4.30	1.76	3.49
1faf	68	6.35	2.02	7.41	1.73	4.05
1mjc	69	7.02	2.08	7.58	1.26	4.84
1icb	74	4.87	2.15	10.45	1.11	2.60
2jv8	73	3.63	6.82 (*2.48*)	6.35	2.23 (*1.37*)	1.74

Results of chemical shift restrained molecular dynamics (CS-MD) simulations of partially unfolded structures of proteins compared to unrestrained molecular dynamics controls. All results are reported for the lowest temperature replica of replica exchange simulations. rmsd is reported in Å for backbone atoms of the ordered regions in reference structures, which were determined with standard methods. Italicized values for 2jv8 are for CS-MD simulations where chemical shift restraints where not included for the disordered loop regions in the structure.

conformational space under the guidance of chemical shift restraints. Thus, the incorporation of chemical shift restraints in Monte Carlo conformational searches, at least in the initial implementation that we presented (Robustelli et al., 2009), is subject to several limitations. In predicting chemical shifts from protein structures using currently available methods (Lehtivarjo et al., 2009; Meiler, 2003; Neal et al., 2003; Shen and Bax, 2007; Xu and Case, 2002), small changes in protein conformations can often produce large changes in the predicted values of chemical shifts, resulting in large energetic penalties. These energetic barriers can frustrate a conformational search and make it difficult to achieve large structural adjustments that are required to minimize chemical shift penalties if the starting structure is not close to the target structure. Another complication is that the SHIFTX method (Neal et al., 2003), which we used to calculate the chemical shifts from the structures (Robustelli et al., 2009), predicts the chemical shifts using discontinuous functions of the atomic coordinates; therefore, it was necessary to adopt a Monte Carlo sampling, rather than molecular dynamics simulations, resulting in a large percentage (up to 90%) of moves rejected due to a generation of large penalties. This effect was particularly pronounced in the case of near-native structures, where structural adjustments that disrupt the packing of structure elements or aromatic rings produce dramatic changes in the predicted chemical shifts. Additionally, because the SHIFTX method utilizes look-up tables to predict chemical shifts from protein structures, extra calculations were needed at each step in the conformational search in addition to those required to calculate the molecular mechanics force field energies. The additional calculations were shown to significantly increase the computational cost of the restrained simulations compared with unrestrained control simulations (Robustelli et al., 2009). Since the chemical shift calculations from SHIFTX are not differentiable with respect to the positions of the atoms of the protein, the chemical shift restraints cannot be implemented in standard NMR structure calculation protocols that utilize molecular dynamics simulations, and therefore cannot be easily combined with the many existing tools for protein structure calculations utilizing the wealth of structural information contained in other NMR observables.

In the investigation presented here, we demonstrate that it is possible to incorporate chemical shift restraints in conformational searches carried out by molecular dynamics simulations by using the CamShift method (Kohlhoff et al., 2009), a tool recently introduced for the rapid prediction of NMR chemical shifts from protein structures based on an approximation of the chemical shifts as polynomial functions of interatomic distances. This incorporation is possible as the CamShift functions used to compute chemical shifts are fast to compute and readily differentiable with respect to the Cartesian coordinates of the atoms, in contrast to previously developed methods for the semiempirical calculation of protein backbone chemical shifts (Lehtivarjo et al., 2009; Meiler, 2003; Neal et al., 2003; Shen and Bax, 2007; Xu and Case, 2002).

To assess the capability of molecular dynamics simulations of proteins with chemical shift restraints (CS-MD) to efficiently sample the conformational space and correctly identify native states, we carried out a test in which CS-MD simulations were performed for 12 proteins of representative topologies starting from partially unfolded initial conformations. We found that in 11 of the 12 cases, CS-MD simulations successfully folded the proteins and identified structures with backbone rmsd between 0.8 and 2.2 Å from the corresponding reference structures determined by X-ray crystallography or standard NMR spectroscopy methods; for comparison, control simulations without chemical shift restraints only successfully folded just one of the 12 proteins to a near-native structure (Table 1). In the case where CS-MD simulations failed to correctly fold the protein, we show that the removal of chemical shift restraints for the disordered loop regions of the protein improves sampling and leads to the correct identification of the native fold.
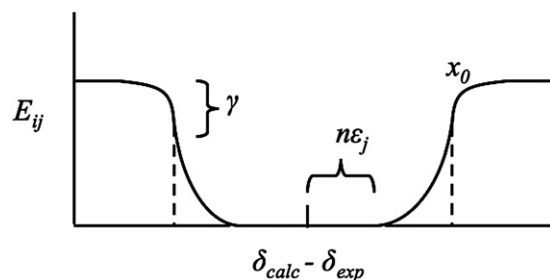
Figure 1. Schematic Illustration of the Chemical Shift Penalty Function Used in the Restrained Molecular Dynamics Procedure Described in This Work

The term E_{ij} gives the contribution of the chemical shift of atom j of residue i to total penalty E_{CS} from the difference between the calculated (δ_{calc}) and experimental shift (δ_{exp}). The function E_{ij} has a flat bottom with a width determined by the term $n\varepsilon_j$, where n is a tolerance parameter and ε_j is the accuracy of the predictions for the chemical shifts of type j. The penalty is harmonic until the deviation reaches a cutoff value x_0; deviations in excess of x_0 contribute to the penalty according to a hyperbolic tangent function that is selected to maintain a continuous derivative at the point x_0. The parameter γ determines how large the penalty can grow for chemical shift deviations beyond x_0.

One additional advantage of the methodology described here over molecular fragment replacement techniques for the determination of protein structures from chemical shifts is that in enables the straightforward combination of chemical shift restraints with NMR restraints that are routinely used in molecular dynamics simulations. In order to demonstrate the utility of this approach, we used chemical shift and residual dipolar coupling data measured with relaxation dispersion NMR to calculate the structure of a transiently populated excited state of a SH3 domain protein.

RESULTS

In this work, we present a molecular dynamics procedure that enables the determination of the structures of proteins by using chemical shift restraints when an initial low-resolution structural model is present. In this method, a penalty function based on the differences between the experimentally measured and calculated chemical shifts is defined by a soft-square harmonic well, E_{CS} (Figure 1), and the forces required for the integration of the equations of motion are generated between atom pairs based on the derivative of E_{CS} with respect to the coordinates of the atom pairs (see Experimental Procedures). The program CamShift (Kohlhoff et al., 2009) is used to predict the backbone chemical shifts ($^1H_\alpha$, $^{13}C_\alpha$, $^{13}C_\beta$, $^{13}C'$, ^1H_N, ^{15}N) at each time step in the simulations (see Experimental Procedures). To provide an initial test of the method, replica exchange chemical shift restrained molecular dynamics simulations (CS-MD) starting from partially denatured protein conformations were run for 12 proteins using all available backbone chemical shifts deposited in the BMRB. Eleven of the 12 proteins contained assignments for all backbone atom types ($^1H_\alpha$, $^{13}C_\alpha$, $^{13}C_\beta$, $^{13}C'$, ^1H_N, and ^{15}N) and one protein (PDB code 2jxt) did not contain $^{13}C'$ assignments.

The results of the CS-MD and control simulations are presented for each protein in Table 1. The backbone rmsd from

the reference structures is reported for both the lowest energy and the lowest rmsd structures sampled by the lowest temperature replica of the replica exchange simulations. Backbone rmsd was calculated from C_α, C', and N backbone atoms for the structured regions of the corresponding reference structures (Table 1). The energies of the structures from the lowest temperature replica of the CS-MD simulations were recalculated using a E_{CS} weight $\alpha = 5$, which was found to be optimal for decreasing the energies of the native states of proteins with respect to other structures for all proteins examined here. Molecular dynamics simulations were run with $\alpha = 1$ because the forces resultant from the use of $\alpha = 5$ were too large and generated instabilities.

For 11 of the 12 proteins examined here, the CS-MD simulations successfully identified structures with backbone rmsd between 0.8 and 2.2 Å from the corresponding reference structures (Table 1); control simulations run without chemical shift restraints only successfully identified a near-native structure, with a backbone rmsd of 2.4 Å, as the lowest in energy for 1 of the 12 proteins (Table 1). In these 11 cases, structures with backbone rmsd less then 1.8 Å from the reference structures were sampled, and in 6 cases, structures with backbone rmsd less then 1.2 Å were sampled. In the case of protein NE1242, for which the lowest energy CS-MD structure had a backbone rmsd of 6.8 Å from the reference structure (PDB code 2jv8) (Table 1), the lowest rmsd structure sampled had a backbone rmsd of 2.2 Å. In the control simulations structures with backbone rmsd of less then 2 Å from the reference structures were only sampled in three cases.

The energy landscapes of the structures sampled by the lowest temperature replicas in the CS-MD and control simulations as a function of the rmsd from the reference structures are illustrated for all proteins in Figure 2, along with an overlay of the lowest energy CS-MD structure and the reference structure and the partially unfolded starting structure. For 10 of the 12 CS-MD simulations considered here, the cluster of structures containing the lowest energy conformation was the most extensively sampled by the lowest temperature replica, and most of the sampling of alternative conformations was limited to higher temperature replicas. The two exceptions were NE1242 (PDB code 2jv8) and ubiquitin (reference PDB code 1d3z). In the case of NE1242, the most extensively sampled cluster of conformations was the most native-like and included structures at about 2.9 Å backbone rmsd from the reference structure. In the case of ubiquitin, the lowest temperature replica was initially trapped in a local minimum at about 3.4 Å backbone rmsd from the reference structure; after escaping from this metastable state, structures with less than 2.0 Å backbone rmsd from the reference structure were the most extensively sampled.

The results presented here suggest that for a number of topologies of globular proteins (Table 2), a native structure can be defined using CamShift backbone chemical shifts as the only experimental restraints in molecular dynamics simulations carried out by using a standard molecular mechanics force field (see Experimental Procedures). Although only one force field (AMBER03) was tested here, the funneled nature of the chemical shift penalty function for each protein (see, e.g., Figures 2 and 3) suggests that the method should be transferable to most molecular mechanics force field for which the native state of a given protein is energetically accessible. The optimal weighting of
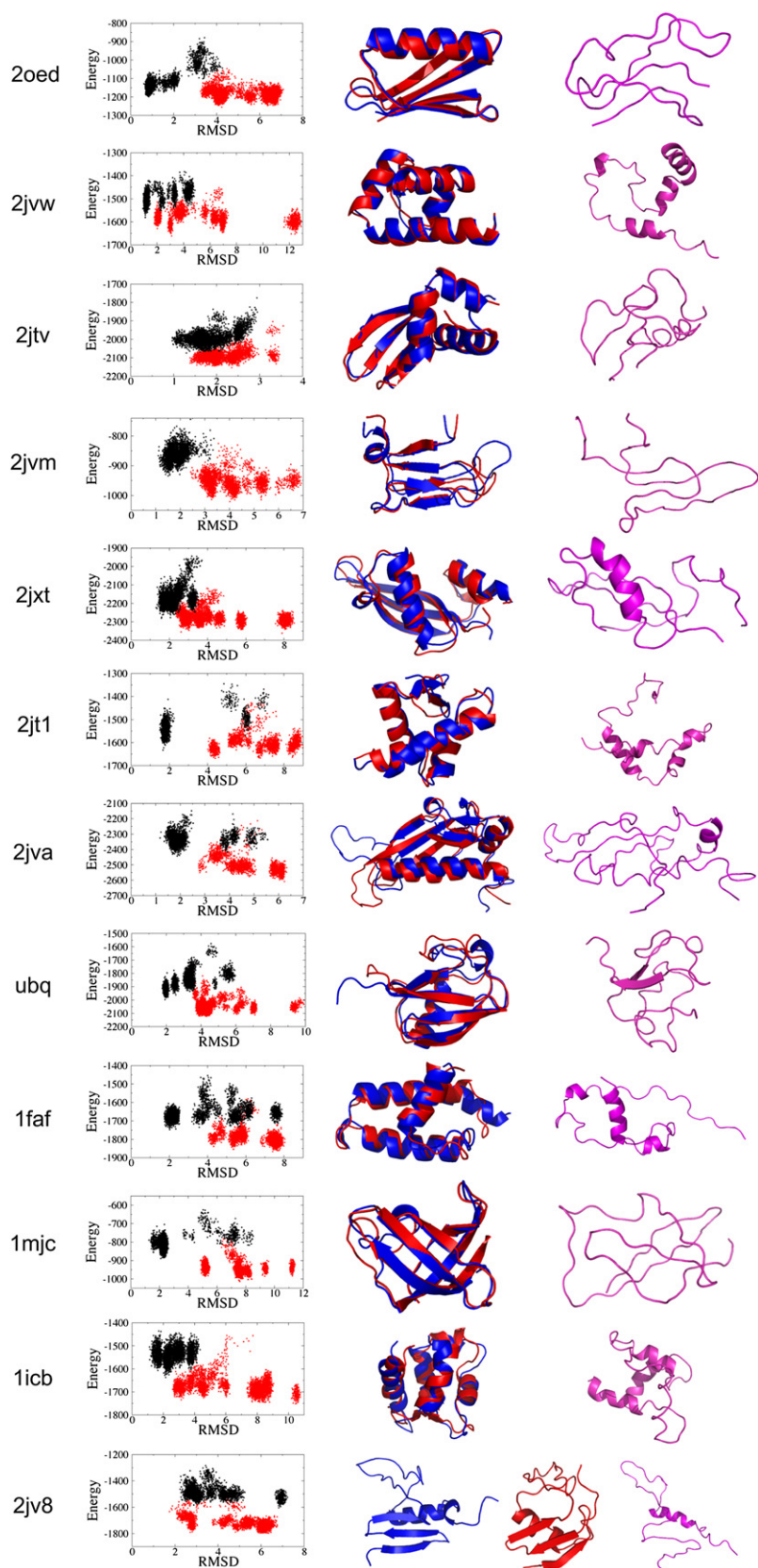
Figure 2. Comparison of the Energy Landscapes of the Structures Generated by CS-MD Simulations and Unrestrained Control Simulations Started from Partially Unfolded Proteins as a Function of the Backbone rmsd from the Reference Structures, Indicated by the PDB Code

The total energy ($E_{CS} + E_{FF}$) for the CS-MD simulations is shown in black and the energies of the unrestrained control simulations are shown in red. Energy landscapes are shown for the lowest temperature replica of replica exchange simulations. For each protein, the overlay of the lowest energy CS-MD structure (red) with the previously determined structure (blue) is also compared with the partially unfolded starting structure (magenta).

Aee Table 1.

Table 2. Secondary Structure Content of the Proteins Discussed in This Work

PDB	SCOP Class	Secondary Structure
2oed	α+β	25% α, 42% β
2jvw	NA (α)	48% α
2jtv	NA (α+β)	50% α, 23% β
2jvm	NA (β)	3% α, 23% β
2jxt	α+β	33% α, 36% β
2jt1	NA (α)	42% α, 12% β
2jva	NA (α+β)	29% α, 20% β
1d3z	α+β	23% α, 32% β
1faf	α	62% α
1mjc	β	4% α, 49% β
1icb	α	52% α, 2% β
2jv8	NA (α+β)	19% α. 24% β

For some PDB entries, there is not yet a SCOP classification (v1.75). These entries are marked as NA.

the chemical shift penalty relative to the force field energy may, however, require empirical adjustments between different force fields.

The results also demonstrate that, by applying chemical shift restraints with a flat-bottom soft-square harmonic well in replica exchange simulations, it is possible to overcome the ruggedness of chemical shift restrained conformational space and obtain large conformational fluctuations to satisfy chemical shift restraints with molecular dynamics, whereas previous implementations of direct refinements from chemical shift restraints have been limited to smaller adjustments of the native fold (Clore and Gronenborn, 1998; Kuszewski et al., 1995; Schwieters et al., 2006).

The protein structures examined here have previously been determined from chemical shift information using molecular fragment based structure calculations (Berjanskii et al., 2009; Cavalli et al., 2007; Das et al., 2009; Montalvao et al., 2008; Robustelli et al., 2008, 2009; Shen et al., 2008, 2009; Wishart et al., 2008). In most cases, the rmsd from the reference structures of the structures calculated with the method that we introduce in this work are only slightly higher than those obtained with Cheshire (Cavalli et al., 2007), CS-Rosetta (Das et al., 2009; Shen et al., 2008, 2009), and CS23D (Berjanskii et al., 2009; Wishart et al., 2008). These results are likely to be due to the use of

more sophisticated knowledge-based force fields in the later methods, which contain additional terms that have been parameterized to favor native structures of proteins in structural databases, and are superior to standard force fields in the selection of protein models (Das and Baker, 2008). It is possible to add similar additional knowledge based terms to the potentials used in this investigation; the results presented here demonstrate the baseline level of structural information contained in the chemical shift restraints implemented in this fashion, without the introduction of further bias from structural databases.

In the case of protein NE1242, CS-MD simulations failed to identify the reference structure (PDB code 2jv8) (Table 1) as the lowest in energy, even though near-native structures were sampled. The CS-MD approach failed in this case because of the presence of a large partially disordered loop region in the native state of this protein. CamShift predictions for loop regions are significantly poorer than for secondary structure elements (Kohlhoff et al., 2009), which is likely due to the inherent difficulty of determining the parameters for the CamShift predictions using loop regions extracted from crystal structures, given that chemical shifts are measured in solution where loop regions often exist as an ensemble of multiple conformations. Indeed, flexible loop regions in native states exhibit chemical shifts that are very close to those in the random coil state (De Simone et al., 2009). It is likely that by attempting to apply chemical shift restraints to determine an average structure in loop regions, the initial conformational search method that we have presented here has difficulty in finding structures that satisfy the chemical shift restraints even if the global fold is correct. In proteins with smaller loop regions and more extended secondary structure elements, the energetic stabilization of the well-defined tertiary interactions in the protein core can outweigh deviations in the loop regions, and the energy of native structure remains the minimum, but, in the case of protein NE1242, it is likely that there is not a large enough proportion of stable tertiary packing in the structure to stabilize the native state.

To test this hypothesis, the CS-MD calculations for NE1242 were repeated with the chemical shift restraints of the disordered loop regions, residues 28–58 and 68–73, removed. The energy landscape of the structures generated in the CS-MD simulation as a function of the rmsd from the reference structures along with a comparison of the lowest energy CS-MD structure and the reference structure are shown in Figure 4. With the shift restraints of the disordered loop region removed, the lowest energy structure sampled has a backbone rmsd of 2.48 Å from
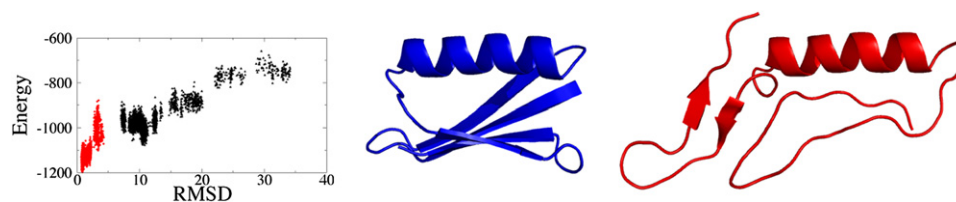

Figure 3. Energy Landscape of the Structures Generated By CS-MD Simulation of GB3 (PDB Code 2oed) Beginning from an Extended Structure as a Function of the Backbone rmsd from the Reference Structure

The total energy ($E_{CS} + E_{FF}$) for the CS-MD simulations is shown in black. For comparison the energy landscape of simulation beginning from the partially denatured structure of 2oed is shown in red. Energy landscapes are shown for the lowest temperature replica of replica exchange simulations. The lowest energy CS-MD structure obtained from the extended structure (red) is compared with the previously determined structure (blue).

Figure 4. Energy Landscape of the Structures Generated By CS-MD Simulation of NE1242 (PDB Code 2jv8) with Chemical Shift Restraints of the Disordered Loop Regions Removed as a Function of the Backbone rmsd from the Reference Structures
The total energy ($E_{CS} + E_{FF}$) for the CS-MD simulations is shown in black. The energy landscape is shown for the lowest temperature replica of a replica exchange simulation. The lowest energy CS-MD structure (red) is compared with the previously determined structure (blue).

the reference structure, and the lowest rmsd structure sampled has a backbone rmsd of 1.37 Å from the reference structure. These results represent a marked improvement over simulations containing chemical shift restraints for the disordered loop regions, suggesting that these restraints do indeed frustrate the conformational search, preventing the sampling of native-like structures, and increase the energy of native state. It is interesting to note, however, that without chemical shift restraints in the disordered loop region, two secondary structure elements, an additional β strand added to the core β sheet and a small α-helical segment, are incorrectly formed. Additionally, structures with the lowest rmsd from the reference structure are not identified as the lowest energy, demonstrating the inability of the force field to correctly model the disordered region as well. For cases where the target structure is not known, it is possible to identify disordered residues for which chemical shift restraints should not be enforced from chemical shift and sequence information alone, by comparing experimental shifts to predicted random coil shift values (De Simone et al., 2009; Schwarzinger et al., 2001; Wishart et al., 1991), in a manner analogous to use of the Random Coil Index (Berjanskii and Wishart, 2007) in CS-Rosetta (Shen et al., 2008).

In all 12 proteins examined here, significantly more accurate structures, with backbone rmsd as low as 0.62 Å from the reference structures, were sampled then those identified as the lowest in energy by the chemical shift restrained energy function. The average lowest backbone rmsd structure sampled was 1.24 Å. These results demonstrate that more sophisticated methods for model selection will result in more accurate structures. Knowledge based potentials are one possibility, but a more rigorous option for overcoming deficiencies in the chemical shift restrained energy function, when possible, would be the use of an independent set of unrestrained NMR data, such as RDCs, to select conformations from those sampled.

Comparison with Monte Carlo Simulations with Chemical Shift Restraints

A replica exchange chemical shift restrained Monte Carlo simulation (Robustelli et al., 2009) was run for the protein GB3 (PDB code 2oed). To allow for direct comparison to the CS-MD protocol, all simulation parameters were left identical (see Experimental Procedures), and 15,000 Monte Carlo steps were carried out at each temperature before attempting replica exchange swaps. The simulation was run for equivalent CPU time compared with the CS-MD simulation. Consistent with previous observations (Robustelli et al., 2009), using our simple moveset

(see Experimental Procedures), the chemical shift restrained Monte Carlo sampling was inefficient for exploring conformational space of compact structures, with >90% of proposed moves rejected. In the timescale examined here, these simulations failed to sample any significant conformational changes from the starting partially denatured structure. The energy landscape of the structures sampled by the lowest temperature replica as a function of the rmsd from the reference structure and an overlay of the starting structure and lowest energy structure sampled are shown Figure 5.

Molecular Dynamics Simulations with Chemical Shift Restraints from Fully Extended Conformations

Replica exchange CS-MD simulations were run for GB3 (PDB code 2oed) starting from a fully extended conformation. The simulation was run for 2.0 ns, with the same parameters as CS-MD simulations begun from partially denatured conformations. The energy landscape of the structures sampled by the lowest temperature replica as a function of the rmsd from the reference structures is shown in Figure 3, along with a comparison of lowest energy CS-MD structure and the reference structure. The lowest energy structure sampled has a backbone rmsd of 10.82 Å from the reference structure. The helical portion of the structure has a backbone rmsd of 0.39 Å from the reference structure; however, nonnative hydrophobic packing and erroneous tertiary contacts between β strands have resulted in an incorrect global fold. Comparison to the energy landscape of lowest temperature replica from the CS-MD simulation from the partially unfolded starting structure demonstrates that the
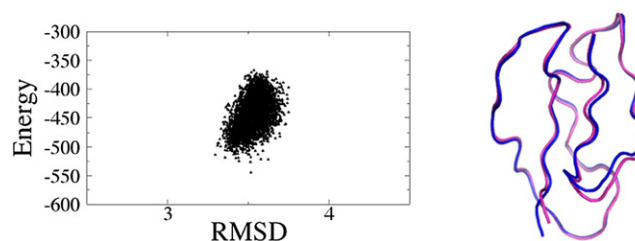

Figure 5. Energy Landscape of the Structures Generated By a Chemical Shift Restrained Monte Carlo Simulations of GB3 (PDB Code 2oed) as a Function of the Backbone rmsd from the Reference Structure
The total energy ($E_{CS} + E_{FF}$) for the lowest temperature replica of the replica exchange simulation is shown in black. The lowest energy structure (blue) is compared to the starting structure (magenta).
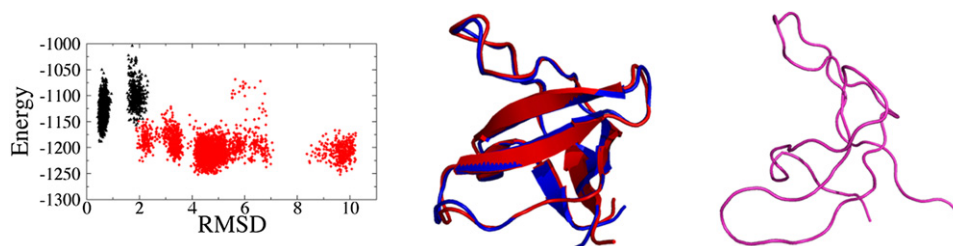
Figure 6. Comparison of the Energy Landscape of the Structures Generated By a CS-MD Simulation with $^{13}C_\alpha$, $^{13}C'$, ^1H_N, and ^{15}N Chemical Shifts Measured By Relaxation Dispersion NMR and an Unrestrained Control Simulation of the Peptide Bound Form of the Abp1p SH3 Domain as a Function of the Backbone rmsd from the Previously Determined NMR Structure (PDB code 2k3b)

The total energy ($E_{CS} + E_{FF}$) for the CS-MD simulations is shown in black and the energy of the unrestrained control simulations is shown in red. Energy landscapes are shown for the lowest temperature replica of replica exchange simulations. The lowest energy CS-MD structure (red) is compared with the previously determined structure (blue), and the partially unfolded starting structure is shown in magenta.

simulation from the extended conformation has become frustrated in a local minimum, and that the native structure is significantly lower in energy. An additional 1.5 ns replica exchange CS-MD simulation (data not shown), begun using the 21 lowest energy structures sampled starting from the extended conformations as seeds, failed to sample new conformations in the lowest temperature replica. These results suggest that current replica exchange temperature distribution, force field, solvation model, and chemical shift penalty parameters used here are not ideally suited for escaping local folding minima with incorrect hydrophobic cores on these short timescales. It is likely that folding proteins from extended conformations using a CS-MD approach will require significantly longer simulations and, as previously demonstrated for Monte Carlo simulations (Robustelli et al., 2009), would benefit from an optimization of simulation parameters suited for the efficient sampling of the necessary conformational transitions, which involve different interactions and energetics than the smaller conformational fluctuations observed when starting from near topologically correct partially unfolded structures.

Combining Chemical Shift and Residual Dipolar Coupling Restraints

A benefit of implementing chemical shift restraints in molecular dynamics simulations is that they can be directly combined with other NMR restraints, including NOEs, J-couplings, and residual dipolar couplings (RDCs), that have previously been implemented in structure calculation packages utilizing restrained molecular dynamics simulations (Brunger, 2007; Schwieters et al., 2006). One area where combining chemical shift restraints with additional restraints is likely to be of crucial importance is in the determination of the structures of transiently populated excited state conformations studied with relaxation dispersion NMR techniques (Baldwin and Kay, 2009). It has recently been demonstrated that NMR relaxation dispersion can be used to measure the magnitudes and signs of chemical shift differences between excited state and ground state populations for all protein backbone atom types ($^1H_\alpha$, $^{13}C_\alpha$, $^{13}C_\beta$, $^{13}C'$, ^1H_N, and ^{15}N) and to measure excited state RDCs in small amounts of alignment media (Vallurupalli et al., 2008). These advances have resulted in the first high-resolution structure calculation of a protein structure using restraints measured from relaxation dispersion NMR (Vallurupalli et al., 2008). In that investigation,

an exchanging system was created by adding a small mole fraction of a peptide ligand, a 17-residue Ark1p peptide, to a solution of the Abp1p SH3 domain, such that the peptide bound form of the protein is populated at about 5% the mole fraction of the apo form. Relaxation dispersion measurements were used to measure $^{13}C_\alpha$, $^{13}C'$, ^1H_N, and ^{15}N backbone chemical shifts and ^1H_N-^{15}N, $^1H_\alpha$-$^{13}C_\alpha$, and ^1H_N-$^{13}C'$ RDCs of the excited state peptide bound form of the Abp1p SH3 domain. We have used these data to demonstrate that RDC restraints can be added to the chemical shift restrained molecular dynamics methodology presented to produce structures of excited state conformations.

The structure calculation was conducted in two phases; calculations were first run with chemical shifts alone, and in a second phase RDC restraints were added. The crystal structure of the apo form (PDB code 1jo8) (Fazi et al., 2002) was used to generate a partially denatured starting conformation as previously described (see Experimental Procedures). A partially denatured starting conformation was created to increase sampling and avoid potential structural bias that might result from being trapped in local energetic minima near the starting conformation. A 1.5 ns replica exchange CS-MD simulation with all available $^{13}C_\alpha$, $^{13}C'$, ^1H_N, and ^{15}N backbone chemical shifts used as restraints and an unrestrained 1.5 ns replica exchange control simulation were conducted as described previously (see Experimental Procedures). The energy landscapes of the structures generated in the CS-MD and control simulation as a function of the rmsd from the previously determined reference structure of the ligand bound state (PDB code 2k3b) are shown in Figure 6, along with an overlay of the lowest energy CS-MD structure and the reference structure and the partially unfolded starting structure. The lowest energy CS-MD structure had a backbone rmsd of 0.50 Å from the reference structure.

The 21 lowest energy structures of the CS-MD simulation were used as starting points for a replica exchange CS-MD simulation with added RDC restraints (see Experimental Procedures). All RDCs measured in Pf1 phage particles were used as restraints (24 ^1H_N-^{15}N, 16 $^1H_\alpha$-$^{13}C_\alpha$, and 21 ^1H_N-$^{13}C'$ RDCs), and 26 ^1H_N-^{15}N RDCs measured in PEG/hexanol were left unrestrained for use as an independent method of cross-validation. The initial conformations had an average Q-factor of 0.44 for Pf1 phage RDCs and 0.42 for PEG/hexanol RDCs. The CS-MD simulation with added RDC restraints was run for a total of 2.0 ns. The five lowest energy conformations are displayed in Figure 7.
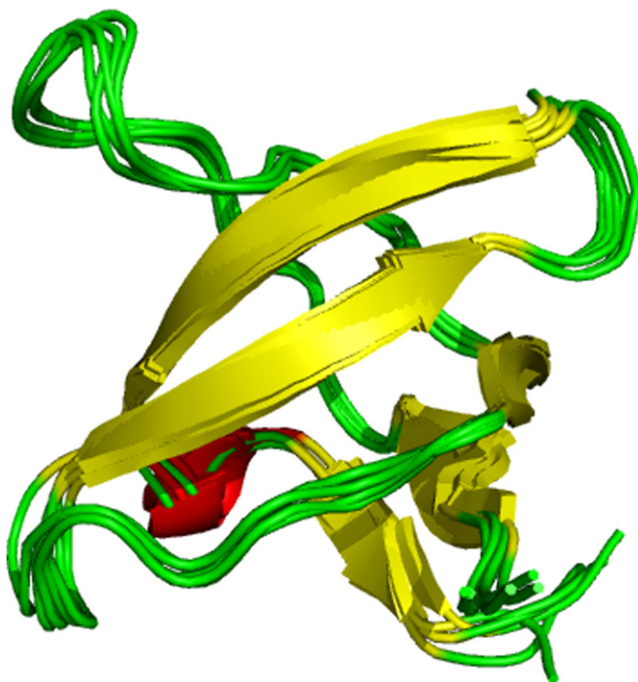
Figure 7. The Five Lowest Energy Structures of the Peptide Bound Form of the Abp1p SH3 Domain Determined from a CS-MD Simulation with Added RDC Restraints

$^{13}C_{\alpha}$, $^{13}C'$, $^{1}H_N$, and ^{15}N chemical shifts and $^{1}H_N$-^{15}N, $^{1}H_{\alpha}$-$^{13}C_{\alpha}$, and $^{1}H_N$-$^{13}C'$ RDCs were measured by relaxation dispersion NMR.

The backbone rmsd of the five lowest energy structures is 0.51 Å, similar to the variation among the previously determined NMR ensemble (Vallurupalli et al., 2008). The lowest energy conformation has a Q-factor of 0.31 for the restrained Pf1 phage RDCs and a Q-factor of 0.34 for the unrestrained PEG/hexanol RDCs, and the five lowest energy structures have average Q-factors of 0.34 for the restrained Pf1 phage RDCs and 0.35 for the unrestrained PEG/hexanol RDCs. Of the five lowest energy conformations, the smallest Q-factor observed for the unrestrained PEG/hexanol RDCs was 0.29.

The generation of structures with low Q-factors for unrestrained RDCs demonstrates the utility of combining chemical shift restrained molecular dynamics simulations with additional restraints and shows that is possible to produce accurate structures of conformations of proteins using data measured by relaxation dispersion NMR, which continue much larger errors than data measured from ground state conformations, with the methodology presented here. Additionally, the structures obtained from CS-MD simulations prior to refinement with RDC restraints using only $^{13}C_{\alpha}$, $^{13}C'$, $^{1}H_N$, and ^{15}N chemical shifts measured by relaxation dispersion produced reasonable Q-factors and were substantially more accurate then the quality of structures obtained using only φ/ψ restraints obtained from TALOS+ (Cornilescu et al., 1999), as described previously (Vallurupalli et al., 2008), which showed significant conformational heterogeneity. This suggests that CS-MD simulations may be of use for the determination of conformations of proteins studied by relaxation dispersion NMR even when RDC restraints are not available.

DISCUSSION

We have demonstrated that the structural information provided by protein backbone chemical shifts can be directly incorporated as restraints in molecular dynamics simulations. In order to achieve this result, the chemical shifts are expressed as differentiable functions of the atomic coordinates and a square well harmonic penalty function is implemented to convert the differences between experimental and calculated chemical shifts into forces that are then used to integrate the equations of motions.

We have tested the method on a set of 12 proteins representative of the major structural classes (Table 2) in the Protein Data Bank, showing that this approach is capable of determining the correct native conformations starting from initial conformations at various degrees of unfolding. The method that we have presented is alternative to existing ones, which enable protein structures to be determined with similar accuracy from chemical shifts using molecular fragment replacement approaches, where chemical shift and sequence homology are used to select fragments from structural databases, and these fragments are assembled to generate new structures. These approaches have been shown to be effective for proteins up to 130 residues in size (Berjanskii et al., 2009; Cavalli et al., 2007; Robustelli et al., 2008; Shen et al., 2008, 2009; Wishart et al., 2008). It has also been previously demonstrated that it is possible to fold small proteins from completely unstructured initial conformations utilizing a Monte Carlo protocol analogous to the molecular dynamics protocol described here, where chemical shifts are directly incorporated as restraints in conformational searches without the use of molecular fragment replacement or sequence and structural homology (Robustelli et al., 2009). As we have shown here, the use of chemical shifts as restraints in molecular dynamics instead of in the Monte Carlo methods enables the structures of longer proteins to be determined, at least when an initial partially folded conformation is available. An initial test to fold the protein GB3 from a completely extended conformation failed to fold the protein on the nanosecond timescales examined here. This result suggests that CS-MD parameters and timescales used in this investigation are most reliably suited for cases where large topological changes are not required, and it is not necessary to escape local minima generated from incorrect topologies. In order to fold proteins from extended conformations with a CS-MD protocol, much longer simulations are expected to be required, and the efficiency of these simulations is likely to drastically improved by the optimization of the simulation parameters for this task, as was done in the previous investigation (Robustelli et al., 2009).

As it is well known that in most cases all-atom molecular simulations are not the most efficient means of folding proteins from extended conformations, we expect the protocol presented here to be most useful when applied in combination with molecular fragment replacement techniques, such as those implemented in Cheshire (Cavalli et al., 2007), CS-Rosetta (Shen et al., 2008), and CS23D (Wishart et al., 2008), to address challenging cases, including proteins larger than 130 amino acids or with complex or novel topologies. In these cases, molecular fragment replacement methods can be used to provide an initial set of candidate structures with incompletely assembled folds, which

can then be used as starting point for structural refinement by chemical shift driven molecular dynamics simulations.

The method introduced here has also the advantage that it can be incorporated in existing structure calculation packages and combined in standard molecular dynamics simulations with other types of structural restraints, including those derived from NOEs, J-couplings, and residual dipolar couplings, allowing the information contained in chemical shift restraints to be fully integrated in standard NMR structure calculation protocols and allowing the calculation of protein conformations from sparse NMR data where initial structural models are available. One such application, the calculation of excited state conformations from data measured by relaxation dispersion NMR, which utilizes the ground state structure for the generation of a starting model, is illustrated here. Additional applications where chemical shift restrained molecular dynamics simulations are likely to be of use include defining conformational changes that occur upon ligand binding, changes in solution conditions, chemical modifications, or amino acid substitutions, or in the refinement of models produced by structural homology. In these cases, if an initial model is available, molecular dynamics with chemical shift restraints are not subject to the current size limit of molecular fragment replacement techniques. The application of this approach should therefore broaden the scope of chemical shift based protein structure determination in structural biology.

EXPERIMENTAL PROCEDURES

CamShift Chemical Shift Predictions

Chemical shift restraints were implemented in the molecular dynamics simulation software *almost* (http://www.open-almost.org) with a modified version of the program CamShift (Kohlhoff et al., 2009) (http://www-vendruscolo.ch.cam. ac.uk/camshift/camshift.php). CamShift is a tool for the rapid prediction of protein backbone chemical shifts ($^1H_\alpha$, $^{13}C_\alpha$, $^{13}C_\beta$, $^{13}C'$, ^1H_N, ^{15}N), based on an approximation of the chemical shifts as polynomial functions of interatomic distances. The CamShift equations for the prediction of a chemical shift, $_{Calc}$, consist of the following components (Kohlhoff et al., 2009)

$$\delta_{Calc} = \delta_{Coil} + \delta_{Backbone} + \delta_{Side-Chains}$$
$$+ \delta_{Dihedrals} + \delta_{Through-Space} + \delta_{Ring} + \delta_{Hbonds}. \quad (1)$$

In this equation, δ_{Coil} is a residue-specific constant that represents the random coil (i.e., dependent on the primary structure of the proteins but not on their secondary and tertiary structures) value for a given atom type, $\delta_{Backbone}$ describes the effects of the local configuration of the protein backbone, $\delta_{Side-Chains}$ captures the effects of the configuration of the side chain atoms for a given residue, $\delta_{Dihedrals}$ takes into account the effects of the ϕ, ψ, and χ_1 dihedral angles, $\delta_{Through-Space}$ defines the through-space effects of nonbonded atoms within a 5 Å sphere of the backbone atom of interest, δ_{Ring} accounts for the influence of ring currents, and δ_{Hbonds} defines the contributions due to the presence and orientation of hydrogen bonds. A version of CamShift was used in the molecular dynamics simulations presented here in which the δ_{Hbonds} term was excluded to avoid discontinuities in forces arising from changes of hydrogen bonding partners during simulations. The $\delta_{Backbone}$, $\delta_{Side-Chains}$, and $\delta_{Through-Space}$ terms are of the form (Kohlhoff et al., 2009)

$$\delta_X = \sum_{j,k} \alpha_{jk} d_{jk}^{\beta_{jk}}. \quad (2)$$

In this equation $_X$ is the contribution of a given term to a predicted chemical shift for a specified query atom i, d_{jk} is the distance between atoms j and k, and α_{ij} and β_{jk} are parameters that were fit on the RefDB database of structures for which backbone chemical shifts were assigned (Zhang et al., 2003). The $\delta_{Backbone}$ term contains the sum over a selected set of distances between

the query atom and backbone atoms in the neighboring residues in addition to a set of extra distances between pairs of backbone atoms that do not include the query atom which better capture ϕ, ψ, and χ_1 torsion angles (Kohlhoff et al., 2009). The $\delta_{Side-Chains}$ term contains the sum over a set of distances between the query atom and atoms in the side chain of that residue. The $\delta_{Through-Space}$ term sums over the set all of distances between the query atom and all atoms contained in a 5 Å sphere which are not backbone atoms of the residue of the query atoms or the directly neighboring residues. For the $\delta_{Backbone}$ and $\delta_{Side-Chains}$ terms, the value of the parameter β_{jk} is equal to 1. For the $\delta_{Through-Space}$ term two separate terms are included for each distance with β_{jk} parameters set to 1 and −3, respectively. The values of α_{ij} were fit for each type of distance using the RefDB database.

The contribution of the term δ_{Ring} is defined using the classical point-dipole method (Pople, 1958)

$$\delta_{ring} = \sum_i \alpha_{ring,i} \sum_{R(i)} \left[\frac{1 - 3\cos^2(\theta)}{r^3} \right], \quad (3)$$

where θ is the angle between the normal vector to the ring plane and the vector connecting the ring center and the query atom and r is the distance between that target atom and the ring center; i runs over the five different ring types (Phe, Tyr, His, Trp_1, and Trp_2) with the latter two denoting the two different rings of tryptophan), and R(i) is the set of all rings in a protein of type i. As in the case of the values of α_{ij} in Equation 2, the values of $\alpha_{ring,i}$ were fit for all pairs of backbone atoms and ring types i on the same database of high-resolution crystal structures and backbone chemical shifts.

The Dihedrals term is defined as (Kohlhoff et al., 2009)

$$\delta_{Dihedrals}(\theta) = p_1 \cos[3(\theta + p_4)] + p_2 \cos(\theta + p_5) + p_3, \quad (4)$$

where θ indicates in turn the three backbone dihedral angles ϕ, ψ, and χ_1 and p_1, p_2, p_3, p_4, and p_5 are fit on the structural database for each atom type with each dihedral angle.

Chemical Shift Penalty Function

At each time step in the molecular dynamics simulations, all backbone chemical shifts for which experimental values are available are computed with CamShift as described above. The differences between the calculated shift and the experimental shift for each atom are converted into a penalty

$$E_{CS} = \alpha \sum_{i=1}^N \sum_j E_{ij}, \quad (5)$$

where N is the number of assigned residues in the protein and j is the chemical shift type where $j = {}^1H_\alpha$, $^{13}C_\alpha$, $^{13}C_\beta$, $^{13}C'$, ^1H_N, ^{15}N. The parameter α defines the weight of the chemical shift term, E_{CS}, relative to the force field term, E_{FF}, and the total energy of the system is defined as $E = E_{FF} + E_{CS}$. In this investigation we set $\alpha = 1$ in all CS-MD simulations, but energy landscapes were recalculated for all trajectories using $\alpha = 5$. It would have been optimal to run CS-MD simulations using a weight $\alpha = 5$, as the larger weight of the chemical shift restraints respective to the force field energies was found to significantly lower the energies of the native states of proteins with respect to the energies of other structures sampled and produce more funneled energy landscapes; however, CS-MD simulations had to be run with $\alpha = 1$, and the energies of the structures sampled recalculated $\alpha = 5$, because the forces resultant from the use of larger values of α were too large and generated instabilities. The penalty E_{ij} calculated for each chemical shift (Figure 1) is defined by

$$E_{ij} = \begin{cases} 0 & \text{if } \left| \delta_{calc}^{ij} - \delta_{exp}^{ij} \right| < n\varepsilon_j \\ \left(\frac{|\delta_{calc}^{ij} - \delta_{exp}^{ij}| - n\varepsilon_j}{\beta_j} \right)^2 & \text{if } n\varepsilon_j < \left| \delta_{calc}^{ij} - \delta_{exp}^{ij} \right| < x_0 \\ \left(\frac{(x_0 - n\varepsilon_j)}{\beta_j} \right)^2 + \gamma \cdot \tanh\left(\frac{2(x_0 - n\varepsilon_j)\left(|\delta_{calc}^{ij} - \delta_{exp}^{ij}| - x_0 \right)}{\gamma \beta_j^2} \right) & \text{for } x_0 \leq \left| \delta_{calc}^{ij} - \delta_{exp}^{ij} \right| \end{cases}$$

$$(6)$$

where δ_{exp} and δ_{calc} are the experimental and calculated chemical shifts, respectively. The function E_{ij} has a flat bottom (Figure 1) so that chemical shifts calculated to within a given accuracy of the experimental value do not produce a penalty. The width of the flat region of the potential is determined by the term

$n\varepsilon_j$, where n is a tolerance parameter and ε_j is the accuracy of the CamShift predictions used for the chemical shifts of type j (Kohlhoff et al., 2009). The penalty is harmonic until the deviation reaches a cutoff value x_0, at which point the penalty grows according to a hyperbolic tangent function defined to maintain a continuous derivative at the point x_0. The magnitude of the penalty is scaled for each chemical shift type j by the variable β_j, which is a function of the variability of that chemical shift in folded proteins reported in the BMRB database (Ulrich et al., 2008). The scaling factor β_j is used to obtain relative contributions of comparable magnitude of each chemical shift type to E_{CS}. The parameter γ determines how large the penalty can grow for deviations beyond x_0. In this investigation all simulations were run with n = 1 for all chemical shifts. The harmonic truncation point x_0 was set to 4.0 ppm for $^1H_\alpha$ and ^1H_N and 20.0 ppm for $^{13}C_\alpha$, $^{13}C_\beta$, $^{13}C'$, and ^{15}N. The penalty truncation factor γ was set to 20 for all chemical shifts. These values of x_0 and γ result in an essentially harmonic penalty for most chemical shifts, with penalties only reaching the hyperbolic tangent region of the penalty function in the case of very large outliers.

Calculation of Forces from CamShift Restraints

CamShift is used to incorporate chemical shift restraints at each step in molecular dynamics simulations by computing the chemical shifts for all experimentally assigned atoms using Equation 2, calculating the energetic penalty of each chemical shift with Equation 6, and generating forces between all atoms that are included in the calculations. Force vectors are generated between two atoms a and b in the x, y, and z directions by computing the derivative of the Equation 5, with respect to the positions of the atoms in each direction

$$f_{(x,y,z)}(a,b) = -\frac{\partial E_{CS}}{\partial_{(x,y,z)}} . \qquad (7)$$

The size of the forces generated is proportional to the slope of the chemical shift penalty E_{ij}. If an atom has an E_{ij} value that falls within the flat-bottom portion of the penalty, no forces are generated. As the value of E_{ij} increases toward x_0 within the harmonic portion of the potential the magnitude of the forces generated increase linearly. As E_{ij} exceeds x_0 the forces generated decrease until they asymptotically approach zero. The rate of this decrease is controlled by the variable γ, with larger values of γ resulting in a slower decay of forces and a larger maximum penalty.

The functional form of E_{ij} has the effect that portions of a protein that are in agreement with chemical shifts do not experience any disrupting forces, portions of a protein that are near-agreement with the chemical shifts experience strong forces to alter the structure, and portions of a protein with very poor chemical shifts have smaller correcting forces. This set up ensures that outliers with very poor chemical shifts do not dominate the force calculations and distort the integration of the equations of motion.

Molecular Dynamics Simulations

All molecular dynamics simulations were run with the software *almost* (http://www.open-almost.org). All software and example scripts are available for download, or upon request from the authors. Simulations were run using the Amber03 force field (Duan et al., 2003) with a generalized Born implicit solvation model (Bashford and Case, 2000), an integration time step of 2 fs with the SHAKE algorithm (Ryckaert et al., 1977) applied to all bond lengths and bond angles. It was necessary to apply the SHAKE algorithm too all bond lengths and angles because of the nature of the chemical shift restraints. CamShift predicts backbone chemical shifts as a function of pairs of interatomic distances between the query atom and atoms of neighboring residues in the protein backbone, between the query atom and atoms within a 5 Å sphere of the query atom, and between a selected set of extra distance pairs in the backbone atoms of the neighboring residues used to capture the effects of local ϕ, ψ, and χ_1 torsion angles. The forces generated from E_{CS} are manifested as attractions or repulsions between each of the atom pairs used in the CamShift calculations. If the local structure of a protein is not held rigid by applying the SHAKE algorithm to all bond lengths and bond angles large deviations between experimental and predicted chemical shift can result in the generation of strong forces between atoms that can exceed the force field bond stretch and bond angle force terms which restrict bond lengths and bond angles to ideal values. This can result in unphysical distortions of local

protein geometries, which in turn generate even stronger restoring forces that can destabilize the protein structure if the time step is too large.

Replica exchange (RepEx) simulations (Sugita and Okamoto, 1999) were run with 21 replicas spanning temperatures from 270 to 515K. Temperature spacing was empirically adjusted to obtain swap acceptances rates between 30% and 50% for all temperatures. Structures were saved every 0.6 ps, and swaps were attempted between adjacent replicas every 6 ps. RepEx simulations were run for lengths between 1.1 and 3.2 ns, based on available machine time. All analyses were conducted on the structures sampled by the lowest temperature replica of the RepEx simulations.

Generation of the Starting Conformations

Partially unfolded conformations were generated as starting points for the CS-MD simulations for 12 proteins whose native states have been determined using standard X-ray crystallography and NMR spectroscopy methods (Table 1). Torsion angle molecular dynamics simulations were run using only ϕ, ψ, and χ dihedral angles as degrees of freedom to ensure that the denatured structures did not contain distorted ω bond angles. Short unfolding simulations were run for 100 ps with a time step of 20 fs at a temperature of 10,000K (Chen et al., 2005; Schwieters and Clore, 2001). The denatured starting structures for the folding simulations were randomly selected from structures with backbone rmsd between 3.5 and 7.0 Å from the original structures, considering all elements of regularly ordered structure. The backbone rmsd of the selected denatured conformation from the starting structure of each protein is reported in Table 1.

Chemical Shift Restrained Monte Carlo Simulations

RepEx chemical shift restrained monte carlo (MC) simulations were run using the Amber03 force field (Duan et al., 2003) with a generalized Born implicit solvation model (Bashford and Case, 2000). Chemical shift restraints were enforced using the same parameters of E_{CS} used in the CS-MD simulations, and RepEx simulations were run with the same parameters as the MD simulations, with 21 replicas spanning temperatures from 270 to 515K. 15,000 MC steps were carried out at each temperature before attempting swaps, and structures were saved every 1500 moves. The MC moveset used was based on the most conservative move set employed in our previous investigation of chemical shift restrained MC simulations. The step size of each move was drawn from a Gaussian distribution with a mean of 0 and a designated standard deviation. The move set contained a side-chain move, which rotates a randomly selected χ angle, with a step size distribution with a standard deviation of 14.3 degrees, two single residue backbone moves, which simultaneously rotate the ϕ and ψ angles of a randomly selected residue, with step size distributions with standard deviations of 5.7 and 2.9 degrees, and two double residue backbone moves, which simultaneously rotate the ϕ and ψ angles of two randomly selected residues, with step size distributions with standard deviations of 2.9 and 1.4 degrees.

Residual Dipolar Coupling Restraints

Residual dipolar coupling (RDC) restraints were imposed by using a flat-bottom harmonic potential with a width determined by the magnitude of the experimental errors (Schwieters et al., 2006). The parameters characterizing the alignment tensor, D_a and R, were initially computed from starting models and allowed to float during the calculations. Simulations were performed with variable relative weights of the RDC restraints and the weight of 0.020 kcal·mol^{-1}Hz^{-2} was found to produce structures with the best agreement to unrestrained RDCs.

REFERENCES

Baldwin, A.J., and Kay, L.E. (2009). NMR spectroscopy brings invisible protein states into focus. Nat. Chem. Biol. *5*, 808–814.

Bashford, D., and Case, D.A. (2000). Generalized born models of macromolecular solvation effects. Annu. Rev. Phys. Chem. *51*, 129–152.

Berjanskii, M.V., and Wishart, D.S. (2007). The RCI server: Rapid and accurate calculation of protein flexibility using chemical shifts. Nucleic Acids Res. 35, W531–W537.

Berjanskii, M., Tang, P., Liang, J., Cruz, J.A., Zhou, J.J., Zhou, Y., Bassett, E., MacDonell, C., Lu, P., Lin, G.H., and Wishart, D.S. (2009). GeNMR: a web server for rapid nmr-based protein structure determination. Nucleic Acids Res. 37, W670–W677.

Brunger, A.T. (2007). Version 1.2 of the crystallography and NMR system. Nat. Protoc. 2, 2728–2733.

Cavalli, A., Salvatella, X., Dobson, C.M., and Vendruscolo, M. (2007). Protein structure determination from NMR chemical shifts. Proc. Natl. Acad. Sci. USA 104, 9615–9620.

Chen, J.H., Im, W., and Brooks, C.L. (2005). Application of torsion angle molecular dynamics for efficient sampling of protein conformations. J. Comput. Chem. 26, 1565–1578.

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.

Cornilescu, G., Delaglio, F., and Bax, A. (1999). Backbone angle restraints from searching a database for chemical shift and sequence homology. J. Biomol. NMR 13, 289–302.

Das, R., and Baker, D. (2008). Macromolecular modeling with Rosetta. Annu. Rev. Biochem. 77, 363–382.

Das, R., Andre, I., Shen, Y., Wu, Y.B., Lemak, A., Bansal, S., Arrowsmith, C.H., Szyperski, T., and Baker, D. (2009). Simultaneous prediction of protein folding and docking at high resolution. Proc. Natl. Acad. Sci. USA 106, 18978–18983.

De Simone, A., Cavalli, A., Hsu, S.T.D., Vranken, W., and Vendruscolo, M. (2009). Accurate random coil chemical shifts from an analysis of loop regions in native states of proteins. J. Am. Chem. Soc. 131, 16332–16333.

Delaglio, F., Kontaxis, G., and Bax, A. (2000). Protein structure determination using molecular fragment replacement and NMR dipolar couplings. J. Am. Chem. Soc. 122, 2142–2143.

Duan, Y., Wu, C., Chowdhury, S., Lee, M.C., Xiong, G.M., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., et al. (2003). A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J. Comput. Chem. 24, 1999–2012.

Fazi, B., Cope, M., Douangamath, A., Ferracuti, S., Schirwitz, K., Zucconi, A., Drubin, D.G., Wilmanns, M., Cesareni, G., and Castagnoli, L. (2002). Unusual binding properties of the SH3 domain of the yeast actin-binding protein ABP1—structural and functional analysis. J. Biol. Chem. 277, 5290–5298.

Kohlhoff, K.J., Robustelli, P., Cavalli, A., Salvatella, X., and Vendruscolo, M. (2009). Fast and accurate predictions of protein NMR chemical shifts from interatomic distances. J. Am. Chem. Soc. 131, 13894–13895.

Kuszewski, J., Gronenborn, A.M., and Clore, G.M. (1995). The impact of direct refinement against proton chemical-shifts on protein-structure determination by NMR. J. Magn. Reson. B. 107, 293–297.

Lehtivarjo, J., Hassinen, T., Korhonen, S.P., Perakyla, M., and Laatikainen, R. (2009). 4D prediction of protein H-1 chemical shifts. J. Biomol. NMR 45, 413–426.

Meiler, J. (2003). Proshift: protein chemical shift prediction using artificial neural networks. J. Biomol. NMR 26, 25–37.

Montalvao, R.W., Cavalli, A., Salvatella, X., Blundell, T.L., and Vendruscolo, M. (2008). Structure determination of protein–protein complexes using NMR chemical shifts: case of an endonuclease colicin–immunity protein complex. J. Am. Chem. Soc. 130, 15990–15996.

Neal, S., Nip, A.M., Zhang, H.Y., and Wishart, D.S. (2003). Rapid and accurate calculation of protein H-1, C-13 and N-15 chemical shifts. J. Biomol. NMR 26, 215–240.

Pople, J.A. (1958). Molecular orbital theory of aromatic ring currents. Mol. Physiol. 1, 175–180.

Robustelli, P., Cavalli, A., and Vendruscolo, M. (2008). Determination of protein structures from solid-state NMR chemical shifts. Structure 16, 1764–1769.

Robustelli, P., Cavalli, A., Dobson, C.M., Vendruscolo, M., and Salvatella, X. (2009). Folding of small proteins by Monte Carlo simulations with chemical shift restraints without the use of molecular fragment replacement or structural homology. J. Phys. Chem. B 113, 7890–7896.

Ryckaert, J.P., Ciccotti, G., and Berendsen, H.J.C. (1977). Numerical-integration of cartesian equations of motion of a system with constraints—molecular-dynamics of n-alkanes. J. Comp. Physiol. 23, 327–341.

Schwarzinger, S., Kroon, G.J.A., Foss, T.R., Chung, J., Wright, P.E., and Dyson, H.J. (2001). Sequence-dependent correction of random coil NMR chemical shifts. J. Am. Chem. Soc. 123, 2970–2978.

Schwieters, C.D., and Clore, G.M. (2001). Internal coordinates for molecular dynamics and minimization in structure determination and refinement. J. Magn. Reson. 152, 288–302.

Schwieters, C.D., Kuszewski, J.J., and Clore, G.M. (2006). Using Xplor-NIH for NMR molecular structure determination. Prog. Nucleic Mag. Res 48, 47–62.

Shen, Y., and Bax, A. (2007). Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. J. Biomol. NMR 38, 289–302.

Shen, Y., Lange, O., Delaglio, F., Rossi, P., Aramini, J.M., Liu, G.H., Eletsky, A., Wu, Y.B., Singarapu, K.K., Lemak, A., et al. (2008). Consistent blind protein structure generation from NMR chemical shift data. Proc. Natl. Acad. Sci. USA 105, 4685–4690.

Shen, Y., Vernon, R., Baker, D., and Bax, A. (2009). De novo protein structure generation from incomplete chemical shift assignments. J. Biomol. NMR 43, 63–78.

Simons, K.T., Kooperberg, C., Huang, E., and Baker, D. (1997). Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. J. Mol. Biol. 268, 209–225.

Sugita, Y., and Okamoto, Y. (1999). Replica-exchange molecular dynamics method for protein folding. Chem. Phys. Lett. 314, 141–151.

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.

Vallurupalli, P., Hansen, D.F., and Kay, L.E. (2008). Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy. Proc. Natl. Acad. Sci. USA 105, 11766–11771.

Wishart, D.S., and Case, D.A. (2001). Use of chemical shifts in macromolecular structure determination. Methods Enzymol. 338, 3–34.

Wishart, D.S., Sykes, B.D., and Richards, F.M. (1991). Relationship between nuclear-magnetic-resonance chemical-shift and protein secondary structure. J. Mol. Biol. 222, 311–333.

Wishart, D.S., Arndt, D., Berjanskii, M., Tang, P., Zhou, J., and Lin, G. (2008). Cs23d: a web server for rapid protein structure generation using NMR chemical shifts and sequence data. Nucleic Acids Res. 36, W496–W502.

Xu, X.P., and Case, D.A. (2002). Probing multiple effects on N-15, C-13 alpha, C-13 beta, and C-13 ′ chemical shifts in peptides using density functional theory. Biopolymers 65, 408–423.

Zhang, H., Neal, S., and Wishart, D.S. (2003). RefDB: a database of uniformly referenced protein chemical shifts. J. Biomol. NMR 25, 173–195.