Published in final edited form as:

J Mol Biol. 2008 July 18; 380(4): 757–774. doi:10.1016/j.jmb.2008.05.006.

A simple model of backbone flexibility improves modeling of side-chain conformational variability

Gregory D. Friedland^{1,5}, Anthony J. Linares^{2,3,5}, Colin A. Smith^{4,5}, and Tanja Kortemme^{1,4,5,*}

¹Graduate Group in Biophysics, University of California at San Francisco, 1700 4th St, UCSF MC 2540, San Francisco, CA 94158-2330, USA

²Pomona College, University of California at San Francisco, 1700 4th St, UCSF MC 2540, San Francisco, CA 94158-2330, USA

³UCSF Summer Research Training Program, University of California at San Francisco, 1700 4th St, UCSF MC 2540, San Francisco, CA 94158-2330, USA

⁴Graduate Group in Biological and Medical Informatics, University of California at San Francisco, 1700 4th St, UCSF MC 2540, San Francisco, CA 94158-2330, USA

⁵Department of Biopharmaceutical Sciences & California Institute for Quantitative Biosciences, University of California at San Francisco, 1700 4th St, UCSF MC 2540, San Francisco, CA 94158-2330, USA

Abstract

The considerable flexibility of side-chains in folded proteins is important for protein stability and function, and may play a role in mediating pathways of energetic connectivity between allosteric sites. While sampling side-chain degrees of freedom has been an integral part of several successful computational protein design methods, the predictions of these approaches have not been directly compared to experimental measurements of side-chain motional amplitudes. In addition, protein design methods generally keep the backbone fixed, an approximation that may substantially limit the ability to accurately model side-chain flexibility. Here we describe a Monte Carlo approach to modeling side-chain conformational variability and validate our method against a large dataset of methyl relaxation order parameters derived from Nuclear Magnetic Resonance experiments (17 proteins and a total of 530 data points). We also evaluate a model of backbone flexibility based on Backrub motions, a type of conformational change frequently observed in ultra-high resolution X-ray structures that accounts for correlated side-chain backbone movements. The fixed-backbone model performs reasonably well with an overall rmsd between computed and predicted side-chain order parameters of 0.26. Notably, including backbone flexibility leads to significant

© 2008 Elsevier Ltd. All rights reserved.

*Corresponding author: Kortemme@cgl.ucsf.edu.

AUTHOR CONTRIBUTIONS

GDF and TK conceived and designed the experiments. GDF and AJL performed the experiments. GDF, TK, and AJL analyzed the data and wrote the paper. CAS contributed reagents/materials/analysis tools.

SUPPORTING INFORMATION AVAILABLE

One figure with population distributions of ubiquitin I13 and L15 chi 1 and chi 2. One table with detailed results from Model 1; one table with results from Models 2 and 3 for nonpolar methyl groups only; and one table with detailed parameter sensitivity results from Model 2.

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

improvements in modeling side-chain order parameters for 10 of the 17 proteins in the set. Higher accuracy of the flexible backbone model results from both increases and decreases in side-chain flexibility relative to the fixed-backbone model. This simple flexible-backbone model should be useful for a variety of protein design applications, including improved modeling of protein-protein interactions, design of proteins with desired flexibility or rigidity, and prediction of energetic pathways within proteins.

Keywords

protein dynamics; side-chain dynamics; NMR order parameters; protein design; flexible backbone

INTRODUCTION

As suggested by Frauenfelder many years ago it is becoming increasingly recognized that representing the “native” state of a protein using a single conformation, while useful for the analysis of many protein properties, is a substantial simplification¹. A more realistic but also more complex description views proteins as conformational ensembles in both the unfolded^{2–4} and folded states⁵. In particular, the ability of side-chains to adopt several conformations in non-surface positions in folded proteins has received recent attention^{5–7} and it has long been known that aromatic residues are mobile in protein cores⁸. As a consequence of the development of new experimental techniques to characterize side-chain conformational variability, such as nuclear magnetic resonance (NMR) methyl spin relaxation experiments, considerable amounts of data are now available for different types of methyl-group containing side-chains⁷.

Interpretation of side-chain methyl relaxation experiments has led to the suggestion that changes in side-chain conformational entropy can contribute substantially to the free energy of binding^{9; 10}. More accurate modeling of side-chain conformational flexibility may also be important for structure-based drug design, when a target protein changes its binding site in response to binding different small molecules¹¹. Work by Ranganathan and others^{12–18} has provided intriguing evidence for the existence of “communication pathways” in proteins to facilitate the transmission of signals between allosteric and active sites. NMR experiments on several systems suggest that side-chains may play an important role in mediating this conformational coupling¹⁹.

Given the importance of side-chain conformational variability in binding and allostery, modeling this flexibility may lead to considerable improvements in the characterization and design of functional proteins and protein interactions. Several representations of side-chain flexibility incorporating multiple higher-energy conformations have been used in prediction and design^{20–23}. However, the predictions resulting from models commonly used in protein design simulations have not been directly compared to experimental data on the amplitude of side-chain motion. Moreover, even when side-chain flexibility has been explicitly considered, the protein backbone is generally held static in the crystal structure conformation, an approximation that likely leads to inaccuracies modeling side-chain conformational freedom.

In contrast, molecular dynamics (MD) simulations model backbone conformational changes and have been compared to side-chain dynamics data, yielding reasonable agreement with measured side-chain order parameters for several proteins^{21; 24–26}. However, it was noted that estimating order parameters from MD trajectories is difficult for side-chains that make few rotameric transitions during the simulation, although sampling was improved using replica exchange MD²⁴. While these MD-based methods achieve considerable accuracy,

they are generally computationally prohibitive for use during protein design, which seeks to simultaneously search in sequence and structure space for low energy amino acid combinations and conformations.

In this paper, we describe a Monte Carlo-based approach to model side-chain conformational variability in protein design simulations and validate our method on a dataset of 17 proteins with 530 methyl relaxation order parameters. We find that motions within the native rotamer well are not sufficient to explain the range of experimentally observed side-chain relaxation order parameters. While a multiple-rotamer Monte Carlo model of side-chain conformational variability performs reasonably well for some proteins (with correlation coefficients between calculated and experimental order parameters above 0.6 for 8 out of the 17 proteins in the dataset), prediction accuracy is limited by the fixed backbone approximation and the use of an implicit solvent model. Using the same dataset, we also evaluate a simple representation of backbone flexibility that allows correlated backbone and side-chain motions inspired by conformational changes observed in ultra-high resolution crystal structures²⁷. Notably, we find that this simple model of correlated backbone-side-chain (“Backrub”²⁷) motions leads to significant overall improvements in modeling side-chain order parameters in our dataset. Incorporating backbone flexibility using Backrub simulations lowers the rmsd between computed and predicted side-chain order parameters for 10 of the 17 proteins, has no significant effect on the rmsd for 5 of the other proteins, and increases the rmsd for 2 proteins. Our results suggest that this flexible backbone protocol is a useful method to sample near-native conformational space. This approach could have substantial impact on many applications, including the computational design of proteins with new functions requiring flexibility.

RESULTS AND DISCUSSION

Rationale and Computational Strategy

We aimed to develop and assess a simple model for fast timescale protein side-chain flexibility that can be applied efficiently in protein structure prediction and design simulations. The major sources of experimental data for model evaluation were methyl relaxation side-chain order parameters. These measurements capture the amplitude of side-chain motions on the picosecond to nanosecond timescale and range from 0 (flexible) to 1 (rigid). Our dataset of 17 proteins containing 530 experimentally characterized methyl groups is shown in Table 1. We tested three different models with increasing complexity (Figure 1a): (1) The first model evaluates the extent to which simulations only allowing side-chain motions within the native rotamer well recapitulate experimentally measured side-chain order parameters. (2) The second model allows more extensive side-chain flexibility by using Metropolis Monte Carlo simulations to sample side-chain conformations from multiple rotameric states. (3) The third model tests the effect of including backbone flexibility by calculating side-chain order parameters from side-chain Monte Carlo simulations performed over an ensemble of backbone conformations. These ensembles were generated using backbone motions inspired by conformational variability observed in ultra-high resolution protein structures (Figure 1b and c)²⁷. All 3 models sample side-chains by varying chi dihedral angles while side-chain bond lengths and bond angles are unchanged (see Methods).

Side-chain Monte Carlo simulations allowing motions only around the native rotamer (Model 1)

We first compared experimental (S^2_{exp}) and computed (S^2_{calc}) side-chain order parameters using the simplest approximation of side-chain conformational variability, where motions are only allowed within the rotamer well of the side-chain conformation observed in the

crystal structure (Model 1, see Figure 1a for a schematic and Methods for details). Figure 2 shows that Model 1 fails to recapitulate the range of order parameters observed experimentally. (See Table S1 for statistics on all proteins in the data set.) The methyl groups mostly have high values for S^2_{calc} , whereas S^2_{exp} values cover a larger range with both high and low values. The finding that native rotamer side-chain motions do not represent the range of experimentally observed side-chain motions also holds when sampling on multiple backbones but still restricting side-chain motions to be within one rotameric state (referred to as “Model 1*” in Figure 2; white boxes), sampling conformers in wider rotamer wells (3 standard deviations around the Dunbrack mean chi 1 and 2 angles) or increasing the leniency towards steric clashes by ‘softening’ the Lennard-Jones repulsive term in the energy function (see Methods; data not shown). This result agrees with previous work calculating side-chain order parameters from MD methods^{5; 21; 24} and from a toy model of side-chain motion²¹. Thus, rotamer transitions, included in Models 2 and 3 below, are necessary to explain measured side-chain order parameters, even for many buried residues.

Side-chain Monte Carlo simulations on a fixed-backbone allowing rotamer transitions (Model 2)

Table 2a summarizes the results of side-chain Monte Carlo simulations employing Model 2. As in Model 1, the simulations were carried out on a fixed backbone, but side-chains were allowed to change rotameric conformations during the simulations. The results for each protein are evaluated by the correlation coefficient (r) between experimental (S^2_{exp}) and calculated (S^2_{calc}) side-chain order parameters, and the root mean squared deviation (rmsd) between them. Additionally, we measured how often we correctly model the qualitative rigidity or flexibility of a side-chain dihedral angle. The results from Model 1 (Figure 2) and the work of others^{5; 21; 24} provide a useful distinction between “rigid” and “flexible” side-chain dihedrals, as they indicate that methyl groups with order parameters above 0.7–0.8 are likely to sample a single rotameric well, and methyl groups with order parameters below this threshold are likely to switch between multiple rotameric states. When sampling within the native rotamer well on a fixed backbone, 95% of methyl groups have S^2_{calc} values above 0.9 and 0.8 for chi 1 and chi 2, respectively (Figure 2). When using an ensemble with backbone variations and no rotamer transitions, the respective S^2_{calc} values are 0.85 and 0.7 for chi 1 and chi 2. Thus, for the purposes of this study, we chose a cutoff value of $S^2=0.75$ to separate “rigid” and “flexible” methyl groups.

We evaluated several parameters that determine which “conformers” (defined here as side-chain conformational microstates, with rotamers being macrostates containing many conformers within a dihedral energy basin, see Methods) are included in the side-chain Monte Carlo simulations: the number of base rotamers to use for each residue (the size of the base rotamer library), the largest chi angle distance between a conformer and its base rotamer (the rotamer well width), and the chi angle degree increment between adjacent conformers (the conformer resolution). The results described in Table 2a use the “large” base rotamer library, a rotamer well width of 1.5 times the standard deviation taken from the Dunbrack rotamer library (see Methods) for that base rotamer, and a conformer resolution of 10 degrees for chi 1 and chi 2.

A standard test for the accuracy of side-chain sampling is whether a given method correctly predicts the side-chain conformations (usually within 40 degrees) observed in the crystal structure. To evaluate our conformer library, we used it in repacking simulations of all side-chains in a published test set of 65 high-resolution crystal structures²⁸ (see Methods for details.) Out of the residues in this set, 87% had correctly assigned chi 1 rotamers and 76% had correctly assigned chi 1 and chi 2 rotamers. These values are similar to repacking results

on the same dataset from several other studies^{29; 30} and somewhat lower than a study using a library of nearly 50,000 conformers²⁸.

Results of Model 2

Introducing rotamer transitions into the simulations results in order parameters spanning the range of the experimental values. For all 17 proteins (Table 2a), allowing rotamer flips substantially improves the correlation coefficient (r) and rmsd between S^2_{exp} and S^2_{calc} (compare Table 2a and Table S1). (Results for the 439 nonpolar residues, excluding threonines, are given in parentheses; reasons for analyzing threonine residues separately are discussed below; see Table S2a.) Out of the 17 proteins in the set, 5(6) have $r \geq 0.7$ and 8(11) have $r \geq 0.6$, indicating that Model 2 is a reasonable model of fast-timescale side-chain motion in some proteins. The rmsd over the whole dataset is 0.26, 69% of the 215 rigid methyl groups were correctly modeled as rigid, and 70% of the 315 flexible methyl groups were correctly modeled as flexible. Thus Model 2 had similar success modeling rigid and flexible methyl groups.

Sensitivity analysis

We next tested how the model performance was affected by changes in the strength of the Lennard-Jones repulsive term, the base rotamer library size, the rotamer well width, and the conformer resolution (see Methods for details).

Reduced Lennard-Jones repulsive term—A number of flexible methyl groups were incorrectly modeled as rigid in our Model 2 simulations. Since Model 2 does not allow motion along the backbone degrees of freedom, we tested whether reducing the Lennard-Jones (LJ) repulsive term would help in the cases where a small backbone shift would allow the side-chain to become flexible. We evaluated two different ways of reducing the strength of the LJ repulsive term: in the first we scaled the LJ radii down by 0.95 (the “small radii” LJ repulsive term), and in the second we decreased the slope of the LJ repulsive term (the “soft repulsive” LJ term; see Methods for details on both).

The results in Table 3a illustrate that there is a clear tradeoff between correctly modeled rigid and correct modeled flexible residues. The original LJ repulsive term gives mostly equivalent modeled fractions of rigid and flexible residues. With the “smaller radii” and the “soft repulsive” LJ terms, the balance shifts in favor of flexible residues, with more methyl groups correctly modeled as flexible and many rigid methyl groups incorrectly modeled as flexible. Thus, these changes to the LJ repulsive term seem to modulate the flexibility or rigidity across all residues in the set, but do not have the environmental specificity needed to increase accuracy.

Base rotamer selection—The size of the base rotamer library determines the number of rotameric states accessible to the protein’s side-chains. One strategy is to choose a library size that is small and hence allows fast sampling. This approach is useful especially for applications such as *ab initio* structure prediction where large numbers of conformations are generated. Since we are sampling high-energy states rather than trying to find the lowest energy conformation, our strategy here was instead to generate a larger rotamer library including many rotamers that have low but nonnegligible probability. As shown in Table 3b, using the large rotamer set has no effect on the rmsd over the data set, but balances prediction of flexible and rigid methyl groups, whereas the default rotamer set modeled rigid residues better than flexible residues. For the flexible *de novo* designed protein $\alpha 3D$, this improvement in rmsd is substantial (0.35 to 0.22; Table S3), as it is for gb1 (0.36 to 0.26). Therefore, the increase in rotamer library size appears useful for modeling some proteins.

Width of rotamer wells—Increasing the width of the rotamer wells has a strong effect on the modeling accuracy (Table 3c). Sampling in wells with widths of 0.5 or 1.0 standard deviations (sds) around the Dunbrack rotamer chi angles shows improvements in rmsd and prediction of rigidity/flexibility. This trend continues, although weaker, up to 1.5 and 2 sds. There is a very large drop in rmsd for α 3D from 1 to 1.5 sds, again highlighting the flexibility of this protein (Table S3).

Conformer resolution—There is not much difference in the performance at 10- and 15-degree conformer resolutions (Table 3d); however, when using the 20- and 25-degrees conformer resolutions the rmsd increases and the fraction of correct flexible methyl groups drops. As expected, excluding all conformers except for the base rotamers performs poorly overall. The single exception is fyn-sh3, which performs well even with no added conformers per base rotamer (Table S3).

Examples of good and bad predictions using Model 2

Figure 3 depicts several examples of Model 2 results. Two of the proteins in our dataset, albp and fyn-sh3 (Figure 3a and 3b), are modeled very well with r of 0.79 and 0.89, and rmsds below 0.2 (Table 2a). We correctly classify the flexibility/rigidity of 10 of the 12 methyl groups in fyn-sh3; of the 28 methyl groups in albp, we correctly classify 77% of its rigid and 73% of its flexible residues. Flavodoxin-holo and ubiquitin (Figure 3c and 3d) are also modeled well, at least qualitatively: their correlation coefficients are lower (0.63 and 0.6), but 83% and 75% of the methyl groups are correctly classified as rigid/flexible for flavodoxin-holo and ubiquitin, respectively.

For ubiquitin, this leaves 5 nonpolar methyl group outliers: 3 are modeled to be too rigid (I61 C δ , L50 C δ , and V70 C γ) and 2 are modeled to be too flexible (I3 C δ and I44 C γ). I61 and I50 are both located on the loop between strands β 3 and β 4, the longest loop in the protein (13 residues), and V70 is located near the C-terminus. The errors modeling these residues could be the result of backbone flexibility that is not taken into account in Model 2.

Examples of proteins that did not perform well with Model 2 are also shown in Figure 3. For cytochrome c2 (Figure 3e), the low fraction of correctly modeled flexible and rigid methyl groups (0.67 and 0.5, respectively) may be related to keeping the large buried heme prosthetic group rigid during the simulations. For example, V107 C γ , I57 C δ and L100 C δ are all less than 4.5Å away from the heme group (Figure 3e) and have S^2_{calc} (S^2_{exp}) values of 0.28(0.9), 0.27(0.7), and 0.58(0.83). In addition, V114 C γ , V115 C γ , I27 C δ and I20 C γ are modeled as too rigid. These residues are located near the end of a beta-hairpin or at the C-terminus, which may be flexible in solution, as suggested by the comparatively low backbone amide order parameters of these residues and their neighbors: 0.75 for E26 (which is the residue in register with I20 on the adjacent strand), 0.72 for I27, 0.78 for S113 and 0.81 for V115. In addition, cytochrome c2 has the lowest resolution crystal structure of any protein in the set at 2.5Å.

As illustrated in the examples above, inaccuracies are likely due to a number of simplifications in our model, including the use of a fixed backbone (as discussed above for ubiquitin and cytochrome c) and the approximation of ligand rigidity. Other likely sources of error include the lack of timescale information from the Monte Carlo simulations (flexibility in our model may occur on timescales longer than those reflected in the experimental measurements) and the use of an implicit solvation model, which does not capture effects related to the defined size and properties of water molecules. We expected this latter effect to be most dramatic for solvent-exposed threonine residues, as discussed below.

Slow transitions / Solvent model inaccuracies

Excluding threonines from the correlation and rmsd calculations improved the results significantly for two proteins: ubiquitin and protein L. For protein L (Figure 3f), excluding the threonine methyl groups (8 of 20 total data points), increased the correlation coefficient from -0.07 to 0.68 and decreased the rmsd from 0.35 to 0.25 . These large differences are caused by several surface-exposed threonine side-chains (T5, T19, T25, T39, and T48), which are modeled as flexible but have high S^2_{exp} values (Figure 3f). A study by Millet et al.³¹ observed that T19, T25, T39, and T48 have one or two ^3J scalar coupling values inconsistent with a singly populated rotamer (^3J couplings could not be measured for T5 due to signal overlap). Thus, our Monte Carlo simulations may in fact correctly capture the flexibility of these threonines on timescales longer than the picosecond to nanosecond motions reflected in the relaxation order parameters. Millet et al. suggest that these slow rotamer transitions occur because particular backbone conformations change the height of the energy barrier between rotamers, and that these barriers can be altered by relatively modest backbone conformational changes in response to mutation³¹.

Alternatively, the slow threonine transitions may be the result of hydrogen bonds to water molecules in the first solvation shell. Inspecting the 5 crystal forms of protein L reveals many potential hydrogen bonds between the side-chain hydroxyl groups of these threonines and nearby water molecules with low temperature factors. As mentioned above, the implicit solvent model used in our simulations does not capture effects due to specific water-mediated hydrogen bonds. Water-mediated hydrogen bonds could restrict the rate of transition between rotameric states for the 5 above-mentioned solvent-exposed threonines in protein L as well as in other proteins, such as ubiquitin (Figure 3c), which also has several threonine residues near ordered water molecules in the X-ray structure. The idea that missed water interactions could be responsible for modeling inaccuracies is supported by the facts that: (a) predictions for threonine $\text{C}\gamma$ had the highest rmsd of any methyl type between S^2_{exp} and S^2_{calc} (0.3), and (b) a lower percentage of threonine $\text{C}\gamma$ methyl groups were correctly modeled as rigid (58%) than either the valine $\text{C}\gamma$ s (75%) or the isoleucine $\text{C}\gamma$ (91% ; data not shown). If the problem modeling threonines is indeed related to the implicit solvent model, it may be ameliorated in future studies by using a “solvated rotamer” approach³².

Side-chain Monte Carlo simulations on backbone ensembles (Model 3)

We show above that Model 2 is a reasonable approximation capturing the flexibility of side-chains with low methyl relaxation order parameters in some proteins. However, a substantial simplification in Model 2 not present in MD simulations is that the backbone is held fixed. We next asked whether a model of backbone flexibility that is simple enough to be computationally feasible in the context of protein design simulations would improve modeling of side-chain flexibility. For each protein in our set, an ensemble of ten near-native backbone structures was used to represent small backbone variations. The backbones were generated using Backrub Monte Carlo simulations with the Rosetta all-atom scoring function (see Figure 1 and Methods for details), and the resulting structures had Ca rmsds to the crystal structure ranging from $0.01\text{\AA}$ to $0.37\text{\AA}$ (see Table 2b for averaged pair-wise Ca rmsds of the ensembles to the crystal structure). This protocol was repeated over ten different ensembles for each protein to estimate the sensitivity to variation in the composition of the ensembles.

We first tested whether inclusion of small backbone variability in this way led to predictions of side-chain order parameter values that were in the experimentally observed range. Figure 2 shows that this is the case, but only when rotameric transitions are allowed in the simulations (compare Model 1*, white boxes, and Model 3, grey boxes, to the experimental data, black boxes). Table 2b summarizes the results of Model 3 for all methyl groups (results

for nonpolar methyl groups only are in parentheses; see Table S2b). 4(8) out of the 17 proteins have correlation coefficients between S^2_{exp} and $S^2_{\text{calc}} \geq 0.7$, and 11(14) proteins have correlation coefficients ≥ 0.6 . Although the ability of Model 3 to correctly classify rigid methyl groups is affected (reduced by 7%) by the increased conformational degrees of freedom introduced with the backbone perturbations, Model 3 shows clear improvements over Model 2 with respect to rmsd values.

The boxplots in Figure 4 illustrate the rmsd values resulting from the different backbone ensembles generated for each protein. Including backbone flexibility results in noticeable improvement in the order parameter predictions for 10 out of 17 proteins, with at least 75% of the rmsd values from Model 3 (purple boxes) below the rmsd value for Model 2 (yellow line). Of the 7 cases where Model 3 does not improve the rmsd relative to Model 2, five proteins have essentially identical results (albp, calmodulin, eglin c, flavodoxin-holo, and NTnC). Of the two cases that worsen substantially (hiv-protease-holo and fyn-sh3), one (fyn-sh3) already performed very well under Model 2, with a correlation coefficient of 0.89 and an rmsd of 0.15.

To assess whether the improvements with Model 3 described above are significant, we performed several statistical tests (see Methods). Tests for the 17 individual proteins confirmed that the results depicted in Figure 4 are statistically significant for 12 proteins, leading to improved rmsds with Model 3 over Model 2 in 10 cases (p-values < 0.005 from the Student's t-test and < 0.01 from the Wilcoxon signed-rank test; asterisks in Figure 4) and a decrease in agreement with experimental data for 2 proteins (same criteria as above, pound signs in Figure 4). We also evaluated whether Model 3 errors (defined as the magnitude of the difference between S^2_{calc} and S^2_{exp} for each methyl group) were significantly less than the Model 2 errors when considering results for all 530 methyl groups together. Using the paired Wilcoxon signed-rank test and the paired Student's t-test, we found that the Model 3 errors were indeed smaller than the Model 2 errors with p-values of 8×10^{-6} and 0.003, respectively. Thus, the overall improvement of Model 3 over the dataset suggests that the ensembles contain relevant conformations that may be populated in the solution experiments at the timescale of interest.

An interesting property of the experimental order parameter measurements is that they suggest the existence of many flexible side-chains in buried positions. We were curious whether our models would be able to capture this core flexibility. On a fixed backbone, buried methyl groups were correctly modeled as flexible in 63% of the cases. With a flexible-backbone this value increased slightly to 66%.

A specific example of the improvements seen for Model 3 over Model 2 is shown in Figure 5 for the B1 domain of protein G (gb1). Fixed-backbone simulations (Model 2) of gb1 yield a correlation coefficient of 0.49 and an rmsd of 0.26. In contrast, incorporating backbone flexibility (Model 3) increases the correlation coefficient to 0.68 and decreases the rmsd (calculated over combined simulations using all 10 ensembles of size 10) to 0.22. As illustrated in Figure 5, improvements from Model 3 result both from methyl groups becoming more flexible when incorporating backbone flexibility (L7 C δ and V21 C γ show approximate S^2_{calc} decreases of -0.25 and -0.1 , respectively) and from methyl groups becoming more rigid (residues V29 C γ and V54 C γ have S^2_{calc} increases of 0.4 and 0.1, respectively).

To rationalize the observed changes in side-chain dynamics with Model 3, we structurally analyzed the lowest rmsd ensemble (#6 out of 10) of gb1. Intuitively, backbone moves are expected to increase the freedom of protein regions to sample conformational space and hence lead to lower order parameters. Examples consistent with this behavior are V21 C γ

and L7 C δ . V21 is located in a turn region between a beta sheet and an alpha helix. The averaged C α rmsd in the ensemble (relative to the crystal structure 1PGB) at this position is 0.61Å (with a standard deviation of 0.26Å) and the largest C α rmsd to the crystal structure for an individual backbone conformation is 1.23Å. L7 is located on an extended beta strand, and in the ensemble its C α and C β atoms move up to 0.35Å and 0.49Å from their crystal structure positions, respectively. However, the slight rotation about the backbone resulting from Backrub moves causes larger Cartesian coordinate changes at the C δ atoms (Figure 6c) and allows L7 to more extensively sample the chi 2 dihedral degree of freedom. As a result, the chi 2=180 degrees rotamer that was infrequently visited in the fixed-backbone simulations (Figure 6a) has a much higher population in the flexible-backbone simulations (Figure 6b), resulting in a lower order parameter that is closer to the experimental value (Figure 5).

Notably, backbone and side-chain conformations simulated in the gb1 backrub ensembles also serve to increase some modeled side-chain order parameters. For example, V29 is located in the center of the helix and is facing into solvent with a ~50% solvent-accessible surface-area. Its flexible-backbone S^2_{calc} is substantially higher than the value from Model 2, and is closer to the experimental value (Figure 5). On the fixed backbone, V29 populates all three rotamers (Figure 6d) but in the flexible-backbone ensemble only the chi 1=180 degrees rotamer is significantly populated (Figure 6e). Figure 6f illustrates a possible mechanism for this increase in modeled rigidity. We observe a hinge motion at residue V29 in a representative backbone in the ensemble, resulting from Backrub rotations that cause a 0.4Å movement of the V29 C α atom away from its position in the X-ray structure and a 0.76Å movement of the C β atom (Figure 6f). This puts the C γ atoms in a different environment where they form closer packing interactions with the alpha helix. Thus, small backbone variations may have considerable effects on the rotamer populations of surface-exposed residues (and likely buried residues as well), highlighting the importance of using a flexible-backbone model for prediction and design.

Another interesting case is residues 13 and 15 of ubiquitin. These side-chains, despite being buried in the protein core, were identified as highly flexible in a key study determining ensembles of protein conformations that represent simultaneously the native structure and its associated dynamics⁵. Our dataset contains measured order parameters for the C γ and C δ methyl groups of I13 and the C δ methyl groups of L15. In all three cases, both Models 2 and 3 predict order parameters below 0.6. Thus, in agreement with reference 5, both side-chains are modeled to populate multiple rotameric states. As described above for V29 in gb1, inclusion of backbone flexibility in ubiquitin increases the S^2_{calc} of two of these methyl groups (I13 C δ and L15 C δ). For I13 C δ , S^2_{calc} rises substantially from 0.1 (Model 2) to 0.49 (Model 3), which is closer to the S^2_{exp} value of 0.55. To rationalize this observation, we analyzed the population distributions of the I13 chi 1 and chi 2 angles predicted from Models 2 and 3. Supplemental Figure S1a shows that I13 chi 2 is relatively unaffected by the inclusion of backbone flexibility; however, the chi 1 distribution shifts from one major rotamer with a moderate and a minor rotamer (Model 2) to a different major rotamer with two more equally-populated moderate rotamers. This change in the population distribution for I13 chi 1 will affect the positions of the C δ atoms and is likely responsible for the modeled increase in order parameters for the I13 C δ methyl group. These observations indicate that relatively subtle redistributions in rotamer populations may have substantial effects propagated through the chi dihedral angles to the ends of the side-chain. While this increased order parameter is closer to the experimental value, the order parameter of the I13 C γ methyl group, which is correctly identified as flexible, is under-predicted by both Models 2 and 3 (with an S^2_{calc} of 0.22 and 0.21, respectively, compared to an S^2_{exp} of 0.6). For L15 C δ , S^2_{calc} changes slightly from 0.32 (Model 2) to 0.4 (Model 3), closer to the S^2_{exp}

value of 0.6, with relatively minimal change in the predicted chi angle population distributions between Models 2 and 3 (Figure S1d).

Comparison to other methods

Side-chain order parameters are generally difficult to predict using simple models based only on packing density (which work well for backbone amide order parameters)³³ or solvent accessibility³⁴. A different method³⁵ reached correlation coefficients between modeled and observed order parameters of $r > 0.6$ for 4 out of 7 proteins. This model described the number of contacts around each methyl carbon and their distances from the backbone with 4 parameters, which were determined by fitting optimal values to 5 proteins in the dataset³⁵.

Another approach for modeling side-chain order parameters used the structural variation present in ensembles of crystal structures of the same protein³⁶. These ensembles with $>98\%$ sequence-identity contain small but significant conformational differences due to different crystallization conditions or small numbers of mutations. From the data available, this method seems to perform similarly to ours in overall average correlation coefficient and rmsd (Table 4). An advantage of our method is that it only requires a single structure of the protein in question.

Several other studies performed MD and calculated order parameters from the structures in the trajectory. A thorough comparison with MD studies is difficult because MD statistics with correlation coefficients and/or rmsds have been published for only a few proteins. In one study, Best et al. performed MD in implicit solvent for 5ns on each of 18 proteins and found correlation coefficients between 0.4–0.7; however, the results were not reported per protein²⁵. Three proteins simulated with explicit solvent MD are shown in Table 4; of these, eglin c (run for 80ns) and TNfn3 (run for 3ns) agree better with experimental data than do our simulations, and FNfn10 (run for 3ns) performs similarly^{21; 24}. A drawback observed in the MD studies is the difficulty sampling rotamer transitions for some residues on the timescale of the simulation, while our models, which do not directly consider a timescale, do not have this problem. In fact, this difference may contribute to the improved performance of MD in modeling side-chain motions on the relatively short picosecond to nanosecond timescale of the side-chain relaxation measurements, while the rotamer transitions modeled by our method may occur on longer timescales (e.g. 5 surface-exposed threonines of protein L appear to make slow transitions; see above).

The study by Lindorff-Larsen et al [5] mentioned above provides another useful point of comparison. This work derived an ensemble of protein conformations that is simultaneously consistent with both experimentally determined order parameters for the native state of ubiquitin as well as distance information from nuclear Overhauser effect (NOE) data. By incorporating the experimental data as restraints in molecular dynamics simulations, the authors conclude that ubiquitin displays significant conformational heterogeneity in solution, with several side chains populating multiple rotameric conformations. The chi angle probability distributions from our simulations (Figure S2) can be compared with those depicted in Figure 3 of reference [5]. The corresponding distributions show substantial agreement, except for the chi 2 angle distribution for Leu 67 that we model to be closer to a previously determined NMR ensemble³⁷.

Our method has several useful advantages, especially in the absence of extensive experimental data on the structure and dynamics of the protein in question. Our method uses a single crystal structure rather than requiring at least ten crystal structures with high sequence identity. It performs similarly to implicit solvent MD, but not as well as explicit solvent MD in two of the published cases. Nevertheless, our methods model motions (such

as slow rotamer transitions) that are difficult to sample with MD. It also runs quickly, taking about 4 minutes to run the fixed-backbone protocol and about 13 hours to run the flexible-backbone protocol (numbers are for modeling ubiquitin on a single AMD Opteron 240 processor). (These time lengths are for the protocol used here in which computational efficiency was not a consideration; we expect that these methods can be sped up significantly.) This lower computational cost provides a significant advantage over MD in the case of protein design methods where both sequence and structure space need to be searched.

CONCLUSIONS

We have compared 3 different models of side-chain flexibility to experimental relaxation side-chain order parameter measurements. While native-rotamer motions (Model 1) do not reproduce the range of S^2_{exp} values, consistent with other studies^{5; 21; 24}, our fixed-backbone Monte Carlo method to sample rotameric transitions (Model 2) gives reasonable agreement between S^2_{calc} and S^2_{exp} values. Importantly, we expand upon this Monte Carlo model of side-chain motion by incorporating near-native ensembles of backbone structures into our simulations (Model 3). These backbone motions were inspired by Backrub conformational changes observed in ultra-high resolution X-ray structures, and allow correlated backbone-side-chain movements. Using this flexible-backbone model (Model 3), we find statistically significant overall improvements in rmsd (over Model 2), which are consistent over the majority of proteins in our dataset. This result demonstrates that Backrub simulations are a useful method for sampling near-native conformational space. Both Models 2 and 3 achieve these results despite inherent simplifications: (i) ligands are treated as rigid, (ii) the experiments are limited to motions on the picosecond-nanosecond timescale while our simulations are timescale-independent, and (iii) we do not model water molecules explicitly.

Conceptually similar to the dipeptide Backrub move in Model 3 is the 1-D Gaussian Axial Fluctuation (GAF) model. The similarity is worthy of note because this model was shown to explain motions present in Residual Dipolar Coupling experiments, which measure events that can take up to milliseconds to occur³⁸. Our model extends the 1-D GAF model by adding both tripeptide moves and rotamer changes, allowing significant anisotropy and structural deviations from the native structure.

The treatment of backbone flexibility here is simple, and leaves room for improvement from more sophisticated models. The dataset used in this study provides a useful benchmark to evaluate such models. Notably, the new degrees of freedom introduced by backbone motions, while producing the expected increase in flexibility of some residues, can also cause increased side-chain rigidity, as was observed in gb1 and ubiquitin.

The models of side-chain flexibility presented here have many uses in prediction or design applications. First, incorporating information about side-chain flexibility changes resulting from macromolecular interactions should improve prediction and design of binding. Specifically, having an improved picture of the high-energy states of a protein will help in prediction and design of binding by conformational selection. Second, we show that modeling backbone flexibility leads to significant differences in sampled side-chain conformations, an effect that is likely to increase the diversity of sequences sampled during protein design. Sequences sampled during fixed-backbone design are strongly biased towards the particular backbone conformation used. Thus, removing the restraint of a fixed backbone (even with the small amount of conformational variability used here) will allow a greater diversity of amino acids at designed positions and may enable the computational design of sequence libraries matched to specific engineering tasks (E. Humphris and T.K.,

unpublished results). Third, our method can be used to design for flexibility or rigidity of a protein. This can be accomplished by adapting the flexibility prediction technique described here into either a post-processing filter or a score term calculated on the fly during the design protocol. In the latter method, short side-chain Monte Carlo simulations could be used to evaluate whether an amino acid substitution results in the desired flexibility profile. Considering flexibility explicitly may also prove useful in the design of enzymes³⁹, an idea supported by recent NMR data highlighting the importance of conformational dynamics in the rate of catalytic turnover^{40–42}. Finally, these and similar simulation methods might be useful for investigating energetically connected pathways of residues in proteins, which could lead to the design of proteins with new modes of allosteric regulation.

MATERIALS AND METHODS

Dataset of experimental protein structures and relaxation measurements

The proteins used in this study are: ubiquitin (PDB id: 1UBQ)⁴³, the third fibronectin type III domain from human tenascin (TNfn3; PDB id: 1TEN)⁴⁴, the tenth fibronectin type III domain from human fibronectin (FNfn10; PDB id: 1FNA)⁴⁵, the B1 domain from protein G (gb1; PDB id: 1PGB)⁴⁶, the SH3 domain in human Fyn (fyn-sh3; PDB id: 1SHF)³⁴, the *de novo* designed three-helix bundle α 3D (a3D; PDB id: 2A3D)⁶, eglin c (PDB id: 1CSE)^{7; 47}, an SH2 domain of phospholipase C-gamma1 (plcc-sh2; PDB id: 2PLD)⁷, the B1 domain of protein L (proteinL; PDB id: 1HZ6)³¹, HIV-1 protease homodimer bound to the inhibitor DMP323, (hiv-protease-holo; PDB id: 1QBS), flavodoxin bound to flavin mononucleotide (flavodoxin-holo; PDB id: 1OBO)⁴⁸, cytochrome c2 bound to its heme prosthetic group (cytochrome-c2-holo; PDB id: 1C2R)⁴⁹, the N-terminal domain of chicken skeletal troponin C (NTnC; PDB id: 1AVS)⁵⁰, Cdc42Hs⁵¹, adipocyte lipid-binding protein (albp; PDB id: 1LIB)⁵², muscle fatty acid-binding protein (mfabp; PDB id: 1HMT)⁵², and Ca²⁺-loaded calmodulin (PDB id: 1AHR)⁵³.

For NMR structures (i.e. 2PLD and 2A3D), the first submitted conformation was used after minimizing all atoms with the Protein Local Optimization Program (which uses a variant of the Truncated Newton method with the OPLS force field and Generalized Born solvation)^{54; 55}. The highest resolution structure was used for proteins with multiple crystal structures, and the first chain was chosen for structures with multiple chains (except for the homodimer HIV protease). If a protein had relaxation measurements in both apo and holo states, we chose to model the apo state; however if no apo structures were available, the ligands were removed from the structure. If measurements were only available for a ligand-bound protein, we included the ligand atoms held fixed at their crystal structure coordinates.

There are 9 types of methyl groups with relaxation side-chain order parameter (S^2) measurements: alanine β , valine γ_1 & γ_2 , threonine γ , isoleucine γ & δ , leucine δ_1 & δ_2 , and methionine. We did not analyze alanine methyl group dynamics as alanine side-chains lack non-hydrogen torsional degrees of freedom. Our models are based on idealized bond geometry and thus treat the symmetric methyl groups of valine and leucine identically; for these methyl groups we compare the computed S^2 to the average of the two experimental S^2 values (if both were available). If backbone and side chain motions are anisotropic, then the two methyl groups of leucine and valine residues are not equivalent and will have slightly different order parameters. However, averaging the values can be justified as they are quite close, with differences less than 0.1 for 53 out of the 60 pairs of Leucine or Valine methyl groups in our dataset and less than 0.05 for 37 of these methyl group pairs.

Fixed-backbone Monte Carlo side-chain simulations (Models 1 and 2)

All simulations use the Rosetta program for protein structure modeling and design⁵⁶. Three different models were used to simulate side-chain flexibility. Models 1 and 2 (Model 3 is described in a later section) used Monte Carlo simulations consisting of side-chain conformer changes on a fixed polypeptide backbone evaluated with the Metropolis criterion. This method is similar to side-chain repacking in Rosetta, with the difference here that the temperature is fixed at $kT=1$ after the initial annealing procedure (see below). A Monte Carlo move in this simulation consisted of randomly choosing a residue in the protein and changing its side-chain conformation to a “conformer” with side-chain χ dihedral angles chosen according to PDB statistics (described in the next paragraph) using idealized bond geometry. (To avoid confusion we use the term “rotamer” to describe the conformational macrostate including all nearby microstates within the same dihedral energy basin. The term “base rotamer” is used to describe the side-chain conformational microstate at the center of this basin, and the term “conformer” is used to describe any side-chain conformational microstate.) An initial simulated annealing procedure was used to equilibrate the protein to the force field; this consisted of starting the Monte Carlo simulations at high temperature ($kT=100$) and exponentially decreasing the temperature in stages to $kT=1$. The number of Monte Carlo moves performed in this initial annealing process was 200 times the number of conformers included in the conformer library of a protein. The number of moves performed at fixed temperature was 800 times the number of conformers. Each such simulation on a given protein was performed 10 times (unless otherwise noted) with different seeds for the random number generator. The simulations used the Rosetta all-atom scoring function, which is dominated by Lennard-Jones packing interactions, an orientation-dependent hydrogen bonding potential⁵⁷ and an implicit solvation model⁵⁸, as described in detail in⁵⁹. The results were somewhat dependent on the simulation temperature, but $kT=1$ was found optimal overall in the context of the Rosetta all-atom scoring function (data not shown).

Side-chain conformer libraries

The “conformer” library of possible side-chain conformations for a given residue was created by first selecting the base rotamers using Dunbrack’s backbone-dependent rotamer library^{60; 61} and then adding conformers around each base rotamer. The library was defined by several attributes: (a) the number of base rotamers to include, (b) the conformer resolution, or χ angle separation between adjacent conformers for χ_1 and χ_2 ; and (c) the rotamer well width, or maximum χ_1 and χ_2 angle distance of conformers from the base rotamer (expressed as the number of standard deviations tabulated in the Dunbrack library). For Model 1, one base rotamer was chosen per residue by finding the base rotamer with the lowest heavy-atom rmsd to the crystal side-chain conformation (with conformers added around it as described). For Model 2, the base rotamer library was either: (i) the “default” library: consisting of 95% or 98% accumulated probability of occurrences in the PDB but restricted to at most 24 or 30 conformers for surface or buried base rotamers, respectively, or (ii) the “large” library: consisting of 99% accumulated probability with at most 45 conformers for a given surface or buried base rotamer. Only the base rotamers were used for χ_3 and χ_4 , without adding neighboring conformers.

Side-chain repacking test

To test side-chain repacking accuracy, we used the same dataset of 65 X-ray structures described in²⁸. For each protein, all residues were repacked simultaneously using Rosetta with the large base rotamer library, a rotamer well width of 1.5 times the Dunbrack standard deviation for that base rotamer, and a conformer resolution of 10 degrees. A repacked residue was classified as having a correctly assigned χ_1 or χ_{1+2} conformation if the

modeled chi values deviated by less than 40 degrees from the corresponding X-ray structure values.

Modified Lennard-Jones terms

Rosetta models the Lennard-Jones (LJ) term using the classical 6–12 potential for attractive contributions and some repulsive contributions; however, at inter-atomic distances in the repulsive regime less than a “switchover” distance, the repulsive potential is modeled as a line with the same slope as the 6–12 potential at this distance⁵⁶. The default (“hard repulsive”) value of this switchover distance is $d_{ij}/r_{ij} = 0.6$, which gives a well depth-independent slope of ~ -9000 (where d_{ij} and r_{ij} are the inter-atomic distance and summed van der Waals radii, respectively, for atoms i and j). The LJ radii are derived from fitting atom distances in protein X-ray structures to the 6–12 LJ potential using CHARMM well depths⁶². Two variants of a “reduced” LJ repulsive term were used. The first “small radii” modification used LJ radii values scaled by 0.95⁶³. The second “soft repulsive” modification reduced the linear slope of the LJ repulsive term. The adjusted value of the switchover distance for this “soft repulsive” modification is $d_{ij}/r_{ij} = 0.91$, which gives a well depth-independent slope of ~ -18 .

Generation of conformational ensembles using Backrub Monte Carlo simulations

To generate protein conformational ensembles with varying backbone conformations, we ran Metropolis Monte Carlo simulations using two types of moves with equal probability: a) a side-chain conformer change, or b) a backbone and side-chain change resulting from a “Backrub” move. The Backrub move was motivated by a type of conformational variability frequently observed in alternate conformations of the same chain of ultra-high resolution crystal structures²⁷. In our implementation, the Backrub move consisted of: (i) choosing a random peptide segment of 2 or 3 successive C α atoms with endpoint residues a and b , (ii) performing a rigid body rotation of main-chain and side-chain atoms between C α_a and C α_b about the axis connecting C α_a and C α_b , (Figure 1b and c) and (iii) optimizing the bond angles extending from C α_a and C α_b using the CHARMM22⁶⁴ bond angle potential. (C.A.S and T.K., unpublished results) This Backrub Monte Carlo simulation was run for 10,000 steps at $kT=0.6$. Side-chain conformers were taken from the Dunbrack library⁶⁰ with conformations around each base rotamer added for chi 1 and chi 2 as described⁶⁵.

Simulations including backbone flexibility (Model 3)

Backbone flexibility was included into the side-chain simulations by running fixed-backbone Monte Carlo simulations on an ensemble of backbone conformations generated using Backrub Monte Carlo simulations. Each backbone was generated by selecting the lowest energy structure from a Backrub Monte Carlo simulation (as described above). For each protein, ten ensembles were used. For each ensemble, 100 structures were generated and then pruned down to the ten with the lowest energy. One fixed-backbone side-chain Monte Carlo simulation was then run on each backbone in the 10-member ensemble.

Calculation of order parameters

For each fixed-backbone side-chain Monte Carlo step that a residue was in a particular conformer, the count for that conformer was incremented. At the end of a simulation on a particular backbone, the population of each conformer was calculated as the sum of the conformer counts, divided by the total number of non-annealing steps. For multiple independent simulations (on the same or different backbones) all conformers in the simulations were accumulated and their probabilities were renormalized to a sum of 1. The order parameters were calculated from these conformer populations and the coordinates (x ,

y, z) of the conformer's relevant methyl carbon (using C γ 1 for valines and C δ 1 for leucines)⁴⁷:

$$S^2 = \frac{3}{2} \left[\langle x^2 \rangle^2 + \langle y^2 \rangle^2 + \langle z^2 \rangle^2 + 2\langle xy \rangle^2 + 2\langle xz \rangle^2 + 2\langle yz \rangle^2 \right] - \frac{1}{2} \quad \text{Equation 1}$$

Analysis of goodness-of-fit

The level of agreement between the experimental and simulated order parameters was calculated in three ways: (a) the linear correlation coefficient between the two sets of order parameters, (b) the root mean squared deviation (rmsd) between these two sets, and (c) the percentage of methyl groups that were correctly modeled as “rigid” (defined as $S^2 \geq 0.75$) or “flexible” (defined as $S^2 < 0.75$). The cutoff value for the order parameters of 0.75 is an approximation of the threshold that was observed in multiple studies^{21; 25; 60} (including this one) to distinguish qualitatively between side-chains populating one or multiple rotameric states.

Analysis of statistical significance

For each of the 17 proteins in the dataset, the performance of Models 2 and 3 were analyzed by applying the Student's t-test to compare the Model 2 rmsd to the Model 3 ensemble rmsds. The difference between the rmsds of two models was judged significant when the one-tailed p-value was less than 0.005. The Wilcoxon signed-rank test was performed on the same data and judged significant when the p-value was less than 0.01.

The error between the calculated and experimental order parameters for Models 2 and 3 were also compared across each of the 530 methyl groups in the dataset. The errors were calculated as the unsigned distances between S^2_{calc} and S^2_{exp} . (The order parameters for Model 3 were averaged over the 10 ensembles). The difference in the errors between Models 2 and 3 were evaluated with the paired Student's t-test and the paired Wilcoxon signed-rank test against the null hypothesis that there is no difference in the magnitude of the errors).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Lewis Kay, Anthony Mittermaier, Martin Stone, Robert Oswald, and Adrienne Loh for providing data on side-chain methyl relaxation order parameters and Dan Mandel, Chris McClendon and Ashley Conrad-Saydah for critical reading of the manuscript. Kristian Kaufman provided scripts to model ligands. G.D.F. is the recipient of an NSF Graduate Research Fellowship. A.J.L. was supported by the UCSF Summer Research Training Program and Genentech. C.A.S. received support from an NIH Training Grant GM067547, a DOD NDSEG fellowship and the Genentech Scholars Program. T.K. is an Alfred P. Sloan Foundation Fellow in Molecular Biology. This work was additionally supported by Sandler start-up funding to T.K.

ABBREVIATIONS

r	correlation coefficient
sd	standard deviation
rmsd	root mean squared deviation
S²	order parameter
S²_{calc}	calculated order parameters

S²_{exp}	experimental relaxation order parameters
MD	Molecular Dynamics
GAF	Gaussian Axial Fluctuation
PDB	Protein Data Bank
LJ	Lennard-Jones

References

1. Hartmann H, Parak F, Steigemann W, Petsko GA, Ponzi DR, Frauenfelder H. Conformational substates in a protein: structure and dynamics of metmyoglobin at 80 K. *Proc Natl Acad Sci U S A*. 1982; 79:4967–71. [PubMed: 6956905]
2. Dill KA, Shortle D. Denatured states of proteins. *Annu Rev Biochem*. 1991; 60:795–825. [PubMed: 1883209]
3. Kortemme T, Kelly MJ, Kay LE, Forman-Kay J, Serrano L. Similarities between the spectrin SH3 domain denatured state and its folding transition state. *J Mol Biol*. 2000; 297:1217–29. [PubMed: 10764585]
4. Choy WY, Forman-Kay JD. Calculation of ensembles of structures representing the unfolded state of an SH3 domain. *J Mol Biol*. 2001; 308:1011–32. [PubMed: 11352588]
5. Lindorff-Larsen K, Best RB, Depristo MA, Dobson CM, Vendruscolo M. Simultaneous determination of protein structure and dynamics. *Nature*. 2005; 433:128–32. [PubMed: 15650731]
6. Walsh ST, Lee AL, DeGrado WF, Wand AJ. Dynamics of a de novo designed three-helix bundle protein studied by ¹⁵N, ¹³C, and ²H NMR relaxation methods. *Biochemistry*. 2001; 40:9560–9. [PubMed: 11583155]
7. Kay LE, Muhandiram DR, Farrow NA, Aubin Y, Forman-Kay JD. Correlation between dynamics and high affinity binding in an SH2 domain interaction. *Biochemistry*. 1996; 35:361–8. [PubMed: 8555205]
8. Wagner G, DeMarco A, Wuthrich K. Dynamics of the aromatic amino acid residues in the globular conformation of the basic pancreatic trypsin inhibitor (BPTI). I. ¹H NMR studies. *Biophys Struct Mech*. 1976; 2:139–58. [PubMed: 9165]
9. Frederick KK, Kranz JK, Wand AJ. Characterization of the backbone and side chain dynamics of the CaM-CaMKII complex reveals microscopic contributions to protein conformational entropy. *Biochemistry*. 2006; 45:9841–8. [PubMed: 16893184]
10. Kay LE, Muhandiram DR, Wolf G, Shoelson SE, Forman-Kay JD. Correlation between binding and dynamics at SH2 domain interfaces. *Nat Struct Biol*. 1998; 5:156–63. [PubMed: 9461082]
11. Thanos CD, DeLano WL, Wells JA. Hot-spot mimicry of a cytokine receptor by a small molecule. *Proc Natl Acad Sci U S A*. 2006; 103:15422–7. [PubMed: 17032757]
12. Lockless SW, Ranganathan R. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science*. 1999; 286:295–9. [PubMed: 10514373]
13. Ota N, Agard DA. Intramolecular signaling pathways revealed by modeling anisotropic thermal diffusion. *J Mol Biol*. 2005; 351:345–54. [PubMed: 16005893]
14. Gunasekaran K, Ma B, Nussinov R. Is allostery an intrinsic property of all dynamic proteins? *Proteins*. 2004; 57:433–43. [PubMed: 15382234]
15. Clarkson MW, Lee AL. Long-range dynamic effects of point mutations propagate through side chains in the serine protease inhibitor eglin c. *Biochemistry*. 2004; 43:12448–58. [PubMed: 15449934]
16. Igumenova TI, Lee AL, Wand AJ. Backbone and side chain dynamics of mutant calmodulin-peptide complexes. *Biochemistry*. 2005; 44:12627–39. [PubMed: 16171378]
17. Fuentes EJ, Gilmore SA, Mauldin RV, Lee AL. Evaluation of energetic and dynamic coupling networks in a PDZ domain protein. *J Mol Biol*. 2006; 364:337–51. [PubMed: 17011581]

18. Scheer JM, Romanowski MJ, Wells JA. A common allosteric site and mechanism in caspases. *Proc Natl Acad Sci U S A*. 2006; 103:7595–600. [PubMed: 16682620]
19. Kern D, Zuiderweg ER. The role of dynamics in allosteric regulation. *Curr Opin Struct Biol*. 2003; 13:748–57. [PubMed: 14675554]
20. Mendes J, Baptista AM, Carrondo MA, Soares CM. Improved modeling of side-chains in proteins with rotamer-based methods: a flexible rotamer model. *Proteins*. 1999; 37:530–43. [PubMed: 10651269]
21. Hu H, Hermans J, Lee AL. Relating side-chain mobility in proteins to rotameric transitions: insights from molecular dynamics simulations and NMR. *J Biomol NMR*. 2005; 32:151–62. [PubMed: 16034666]
22. Koehl P, Delarue M. Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy. *J Mol Biol*. 1994; 239:249–75. [PubMed: 8196057]
23. Zhang J, Liu JS. On side-chain conformational entropy of proteins. *PLoS Comput Biol*. 2006; 2:e168. [PubMed: 17154716]
24. Best RB, Clarke J, Karplus M. What contributions to protein side-chain dynamics are probed by NMR experiments? A molecular dynamics simulation analysis. *J Mol Biol*. 2005; 349:185–203. [PubMed: 15876377]
25. Best RB, Clarke J, Karplus M. The origin of protein sidechain order parameter distributions. *J Am Chem Soc*. 2004; 126:7734–5. [PubMed: 15212494]
26. Prabhu NV, Lee AL, Wand AJ, Sharp KA. Dynamics and entropy of a calmodulin-peptide complex studied by NMR and molecular dynamics. *Biochemistry*. 2003; 42:562–70. [PubMed: 12525185]
27. Davis IW, Arendall WB 3rd, Richardson DC, Richardson JS. The backrub motion: how protein backbone shrugs when a sidechain dances. *Structure*. 2006; 14:265–74. [PubMed: 16472746]
28. Peterson RW, Dutton PL, Wand AJ. Improved side-chain prediction accuracy using an ab initio potential energy function and a very large rotamer library. *Protein Sci*. 2004; 13:735–51. [PubMed: 14978310]
29. Xiang Z, Honig B. Extending the accuracy limits of prediction for side-chain conformations. *J Mol Biol*. 2001; 311:421–30. [PubMed: 11478870]
30. Liang S, Grishin NV. Side-chain modeling with an optimized scoring function. *Protein Sci*. 2002; 11:322–31. [PubMed: 11790842]
31. Millet O, Mittermaier A, Baker D, Kay LE. The effects of mutations on motions of side-chains in protein L studied by 2H NMR dynamics and scalar couplings. *J Mol Biol*. 2003; 329:551–63. [PubMed: 12767834]
32. Jiang L, Kuhlman B, Kortemme T, Baker D. A “solvated rotamer” approach to modeling water-mediated hydrogen bonds at protein-protein interfaces. *Proteins*. 2005; 58:893–904. [PubMed: 15651050]
33. Mittermaier A, Kay LE, Forman-Kay JD. Analysis of deuterium relaxation-derived methyl axis order parameters and correlation with local structure. *Journal of Biomolecular NMR*. 1999; 13:181–185. [PubMed: 20700817]
34. Mittermaier A, Davidson AR, Kay LE. Correlation between 2H NMR side-chain order parameters and sequence conservation in globular proteins. *J Am Chem Soc*. 2003; 125:9004–5. [PubMed: 15369343]
35. Ming D, Bruschweiler R. Prediction of methyl-side chain dynamics in proteins. *J Biomol NMR*. 2004; 29:363–8. [PubMed: 15213434]
36. Best RB, Lindorff-Larsen K, DePristo MA, Vendruscolo M. Relation between native ensembles and experimental structures of proteins. *Proc Natl Acad Sci U S A*. 2006; 103:10901–6. [PubMed: 16829580]
37. Cornilescu GMJ, Ottiger M, Bax A. Validation of Protein Structure from Anisotropic Carbonyl Chemical Shifts in a Dilute Liquid Crystalline Phase. *J Am Chem Soc*. 1998; 120:6836–6837.
38. Bernado P, Blackledge M. Anisotropic Small Amplitude Peptide Plane Dynamics in Proteins from Residual Dipolar Couplings. *J Am Chem Soc*. 2004; 126:4907–4920. [PubMed: 15080696]

39. Bolon DN, Mayo SL. Enzyme-like proteins by computational design. *Proc Natl Acad Sci U S A*. 2001; 98:14274–9. [PubMed: 11724958]
40. Eisenmesser EZ, Millet O, Labeikovsky W, Korzhnev DM, Wolf-Watz M, Bosco DA, et al. Intrinsic dynamics of an enzyme underlies catalysis. *Nature*. 2005; 438:117–21. [PubMed: 16267559]
41. Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, Kern D. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature*. 2007; 450:913–6. [PubMed: 18026087]
42. Wolf-Watz M, Thai V, Henzler-Wildman K, Hadjipavlou G, Eisenmesser EZ, Kern D. Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. *Nat Struct Mol Biol*. 2004; 11:945–9. [PubMed: 15334070]
43. Lee AL, Flynn PF, Wand AJ. Comparison of ²H and ¹³C NMR Relaxation Techniques for the Study of Protein Methyl Group Dynamics in Solution. *J Am Chem Soc*. 1999; 121:2891–2902.
44. Best RB, Rutherford TJ, Freund SM, Clarke J. Hydrophobic core fluidity of homologous protein domains: relation of side-chain dynamics to core composition and packing. *Biochemistry*. 2004; 43:1145–55. [PubMed: 14756550]
45. Geierhaas CD, Best RB, Paci E, Vendruscolo M, Clarke J. Structural comparison of the two alternative transition states for folding of TI I27. *Biophys J*. 2006; 91:263–75. [PubMed: 16603501]
46. Goehrlert VA, Krupinska E, Regan L, Stone MJ. Analysis of side chain mobility among protein G B1 domain mutants with widely varying stabilities. *Protein Sci*. 2004; 13:3322–30. [PubMed: 15537756]
47. Hu H, Clarkson MW, Hermans J, Lee AL. Increased rigidity of eglin c at acidic pH: evidence from NMR spin relaxation and MD simulations. *Biochemistry*. 2003; 42:13856–68. [PubMed: 14636053]
48. Liu W, Flynn PF, Fuentes EJ, Kranz JK, McCormick M, Wand AJ. Main chain and side chain dynamics of oxidized flavodoxin from *Cyanobacterium anabaena*. *Biochemistry*. 2001; 40:14744–53. [PubMed: 11732893]
49. Flynn PF, Bieber Urbauer RJ, Zhang H, Lee AL, Wand AJ. Main chain and side chain dynamics of a heme protein: 15N and 2H NMR relaxation studies of *R. capsulatus* ferrocytochrome c2. *Biochemistry*. 2001; 40:6559–69. [PubMed: 11380250]
50. Gagne SM, Tsuda S, Spyropoulos L, Kay LE, Sykes BD. Backbone and methyl dynamics of the regulatory domain of troponin C: anisotropic rotational diffusion and contribution of conformational entropy to calcium affinity. *J Mol Biol*. 1998; 278:667–86. [PubMed: 9600847]
51. Loh AP, Pawley N, Nicholson LK, Oswald RE. An increase in side chain entropy facilitates effector binding: NMR characterization of the side chain methyl group dynamics in Cdc42Hs. *Biochemistry*. 2001; 40:4590–600. [PubMed: 11294626]
52. Constantine KL, Friedrichs MS, Wittekind M, Jamil H, Chu CH, Parker RA, et al. Backbone and side chain dynamics of uncomplexed human adipocyte and muscle fatty acid-binding proteins. *Biochemistry*. 1998; 37:7965–80. [PubMed: 9609689]
53. Lee AL, Kinnear SA, Wand AJ. Redistribution and loss of side chain entropy upon formation of a calmodulin-peptide complex. *Nat Struct Biol*. 2000; 7:72–7. [PubMed: 10625431]
54. Jacobson MP, Pincus DL, Rapp CS, Day TJ, Honig B, Shaw DE, Friesner RA. A hierarchical approach to all-atom protein loop prediction. *Proteins*. 2004; 55:351–67. [PubMed: 15048827]
55. Zhu K, Shirts MR, Friesner RA, Jacobson MP. Multiscale Optimization of a Truncated Newton Minimization Algorithm and Application to Proteins and Protein-Ligand Complexes. *J Chem Theory Comput*. 2007; 3:640–648.
56. Rohl CA, Strauss CE, Misura KM, Baker D. Protein structure prediction using Rosetta. *Methods Enzymol*. 2004; 383:66–93. [PubMed: 15063647]
57. Kortemme T, Morozov AV, Baker D. An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *J Mol Biol*. 2003; 326:1239–59. [PubMed: 12589766]
58. Lazaridis T, Karplus M. Effective energy function for proteins in solution. *Proteins*. 1999; 35:133–52. [PubMed: 10223287]

59. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. Design of a novel globular protein fold with atomic-level accuracy. *Science*. 2003; 302:1364–8. [PubMed: 14631033]
60. Dunbrack RL Jr, Cohen FE. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci*. 1997; 6:1661–81. [PubMed: 9260279]
61. Dunbrack, RL, Jr. The Backbone-Dependent Rotamer Library (May 2002). 2003.
62. Neria E, Fischer S, Karplus M. Simulation of activation free energies in molecular systems. *J Chem Phys*. 1996; 105:1902–1921.
63. Dahiyat BI, Mayo SL. Probing the role of packing specificity in protein design. *Proc Natl Acad Sci U S A*. 1997; 94:10172–7. [PubMed: 9294182]
64. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B*. 1998; 102:3586–3616.
65. Dahiyat BI, Mayo SL. Protein design automation. *Protein Sci*. 1996; 5:895–903. [PubMed: 8732761]

Figure 1. Computational strategy and motional models. (a) Flowchart of the methods used for the 3 models of motion. Schematic of (b) dipeptide and (c) tripeptide “Backrub” conformational changes used to model backbone changes in Model 3. The Backrub motion consists of a rigid body rotation of all atoms between two Ca atoms, about the axis connecting the Ca atoms. This rotation is followed by optimization of bond angles involving the endpoint Ca atoms (see Methods).

Figure 2.

Side-chain motions within the native rotamer well do not sample the conformational flexibility observed in methyl relaxation experiments. Shown are boxplots representing the distributions of order parameters for C γ and C δ methyl groups from different models and from the experimental measurements. White boxes: native rotamer motions on a fixed backbone (Model 1) or on an ensemble of backbones (generated using Backrub Monte Carlo simulations that kept the side-chains in their native rotamer well, Model 1*). Grey boxes: results from Model 3 simulations, using an ensemble of backbone conformations and allowing multiple rotameric states. Black boxes: experimental relaxation measurements. The boxes represent the middle 25–75% of the values; the horizontal bar inside the box is the median value; the “whiskers” extending out of the box cover the ~ 1.5 times the range of the box (or up to the furthest data point); and dashes outside of the whiskers represent outliers.

Figure 3.

Side-chain motions allowing rotameric transitions on a fixed backbone (Model 2). Plots of S^2_{calc} from Model 2 vs. S^2_{exp} for several proteins that were modeled well—(a) albp, (b) fyn-sh3, (c) flavodoxin-holo, and (d) ubiquitin—and for several proteins that were modeled poorly—(e) cytochrome-c2-holo, and (f) protein-L. Blue circles: nonpolar methyl groups (valine, leucine, isoleucine and methionine); Orange circles: threonine methyl groups; Open circles: cytochrome-c2-holo residues pointing towards the heme group and within $4.5\text{\AA}$. Dashed lines are drawn at $S^2_{\text{calc}} = 0.75$ and $S^2_{\text{exp}} = 0.75$ to reflect the threshold used to classify rigid and flexible side-chain methyl groups.

Figure 4.

A simple model of backbone conformational variability (Model 3) improves modeling of side-chain motions. Rmsd between experimental order parameters (S^2_{exp}) and S^2_{calc} from Model 2 (yellow lines) and Model 3 (purple boxes). *: the 10 proteins for which the Model 3 ensemble rmsds are significantly lower than the Model 2 rmsd (Student's t-test p-value < 0.005 and Wilcoxon-signed rank test p-value < 0.01). #: the 2 proteins for which the Model 2 rmsd is lower than the Model 3 rmsds (by the same measure). See Figure 2 for an explanation of boxplots. Proteins are depicted in order of increasing rmsd between experimental and computed order parameters from Model 2.

Figure 5.

Improvement of S^2_{calc} values for Model 3 over Model 2 showing both increased and decreased modeled flexibility of residues in the protein gb1. S^2_{exp} vs. S^2_{calc} from Model 2 (yellow) and Model 3 (purple). Error bars for Model 3 are the standard deviations of S^2_{calc} over the 10 ensembles.

Figure 6.

Backbone flexibility (Model 3) in gb1 results in the side-chain of L7 becoming more flexible and V29 becoming more rigid. Probability distributions of L7 chi 2 from (a) Model 2, and (b) ensemble 6 of Model 3. (c) Structures of all conformers for L7 from Model 2 (yellow) and Model 3 (purple). The C α -C β and C β -C γ bonds are drawn with lines while the C δ 1 atoms are drawn as disconnected spheres for clarity. Probability distributions of V29 chi 1 from (d) Model 2, and (e) ensemble 6 of Model 3. (f) Structures of all V29 conformers from Model 2 (yellow) and a selected backbone from ensemble 6 of Model 3 (purple). This backbone is representative of those that keep V29 predominantly in 1 rotamer.

Table 1

The dataset of proteins modeled in this paper.

	# methyls ^a	PDB	Chain	Minimized	Ligand	Class
a3D ^a	15	2A3D	–	y	n	Alpha
alpb ^b	28	1LJB	A	n	n	Beta
calmodulin ^c	36	1AHR	A	n	y	Alpha
Cdc42hs	47	1AN0	A	n	y	Alpha/Beta
cytochrome-c2-holo ^d	31	1C2R	A	n	y	Alpha
Egline	17	1CSE	I	n	n	Alpha/Beta
flavodoxin-holo ^e	56	1OBO	A	n	y	Alpha/Beta
FNfn10 ^f	35	1FNA	–	n	n	Beta
fyn-sh3 ^g	12	1SHF	A	n	n	Beta
gb1 ^h	10	1PGB	–	n	n	Alpha/Beta
hiv-protease-holo ⁱ	56	1QBS	dimer	n	y	Beta
mfabp ^j	42	1HMT	–	n	n	Beta
NTnC ^k	26	1AVS	A	n	n	Alpha
plcc-sh2-free ^l	27	2PLD	A	y	n	Alpha/Beta
proteinL ^m	20	1HZ6	A	n	n	Alpha/Beta
TNfn3 ⁿ	40	1TEN	A	n	n	Beta
ubiquitin	32	1UBQ	A	n	n	Alpha/Beta
Totals	530					

^a a3D: the *de novo* designed three-helix bundle α3D

^b alpb, adipocyte lipid-binding protein

^c calmodulin: Ca²⁺-loaded calmodulin

^d cytochrome-c2-holo: cytochrome c2 bound to its heme prosthetic group

^e flavodoxin-holo: flavodoxin bound to flavin mononucleotide

^f FNfn10: the tenth fibronectin type III domain from human fibronectin

$g_{\text{fyn-sh3}}$: the SH3 domain in human Fyn
 h_{gb1} : the B1 domain from protein G
 $i_{\text{hiv-protease-holo}}$: HIV-1 protease homodimer bound to the inhibitor DMP323
 j_{infabp} : muscle fatty acid-binding protein
 k_{NTnC} : the N-terminal domain of chicken skeletal troponin C
 $l_{\text{plcc-sh2}}$: an SH2 domain of phospholipase C-gamma1
 m_{proteinL} : the B1 domain of protein L
 n_{TNfn3} : the third fibronectin type III domain from human tenascin
 o : values for the methyl groups of valine and leucine were averaged over the two experimental values (see Methods)

Table 2

Summary of (a) Model 2 and (b) Model 3 results for all proteins. The values are: the correlation coefficient and slope for the linear fit between S^2_{exp} and S^2_{calc} ; the rmsd between S^2_{exp} and S^2_{calc} ; the number of rigid ($S^2_{\text{exp}} > 0.75$) and flexible ($S^2_{\text{exp}} \leq 0.75$) methyl groups based on the experimental order parameters; the fraction of rigid and flexible methyl groups correctly identified as rigid or flexible from the simulations.

(a) Model 2														(b) Model 3			
	# of methyls	# Rigid (exp)	# Flexible (exp)				Fraction Correct Rigid (calc)		Fraction Correct Flexible (calc)					Mean Ca. RMSD of backbone ensembles ^a			
				r	slope	rmsd				r	slope	rmsd					
a3D	15	0	15	0.64	1.24	0.22	N/A	0.73	0.74	1.35	0.17	N/A	0.86	0.30			
albp	28	13	15	0.79	0.98	0.19	0.77	0.73	0.76	0.93	0.19	0.66	0.79	0.19			
calmodulin	36	4	32	0.81	1.22	0.21	0.75	0.69	0.78	1.14	0.21	0.78	0.69	0.27			
cdc42hs	47	23	24	0.31	0.33	0.33	0.61	0.58	0.41	0.41	0.31	0.46	0.80	0.24			
cytochrome-c2-holo	31	21	10	0.37	0.35	0.24	0.67	0.50	0.52	0.51	0.22	0.56	0.83	0.22			
egline	17	6	11	0.55	0.76	0.25	0.67	0.82	0.59	0.82	0.25	0.83	0.78	0.22			
flavodoxin-holo	56	36	20	0.63	0.74	0.21	0.83	0.85	0.62	0.74	0.21	0.79	0.85	0.06			
FNfn10	35	12	23	0.40	0.57	0.29	0.67	0.65	0.42	0.49	0.24	0.55	0.66	0.19			
fyn-sh3	12	4	8	0.89	1.45	0.15	0.75	0.88	0.62	0.88	0.20	0.55	0.99	0.22			
gb1	10	0	10	0.49	0.64	0.26	N/A	0.80	0.68	0.92	0.22	N/A	0.77	0.24			
hiv-protease-holo	56	36	20	0.73	0.86	0.22	0.69	0.95	0.62	0.68	0.23	0.57	0.92	0.19			
mfabp	42	20	22	0.59	0.69	0.27	0.55	0.82	0.60	0.67	0.26	0.56	0.85	0.15			
NTnC	26	2	24	0.77	1.13	0.23	1.00	0.58	0.73	1.05	0.22	1.00	0.59	0.24			
plec-sh2-free	27	3	24	0.57	0.78	0.26	0.33	0.67	0.60	0.74	0.24	0.36	0.84	0.26			
protl	20	8	12	-0.07	-0.10	0.35	0.62	0.50	-0.06	-0.09	0.33	0.56	0.50	0.11			
TNfn3	40	11	29	0.34	0.40	0.32	0.64	0.55	0.38	0.45	0.31	0.62	0.57	0.12			
ubiquitin	32	16	16	0.60	0.82	0.27	0.75	0.75	0.64	0.81	0.24	0.74	0.83	0.22			
Total	530	215	315	.	.	0.26 ^b	0.69	0.70	.	.	0.25 ^b	0.62	0.76				

^a All backbone ensembles had standard deviations of these values less than 0.05.

^b Total rmsd calculated by considering all 530 data points together.

Table 3

Effect of parameter values of Model 2 on the rmsd and fraction of correctly modeled rigid or flexible methyl groups (see Table 2 caption for more details). **(a)** Effect of using different types of Lennard-Jones repulsive terms. **(b)** Different base rotamer library. **(c)** Increasing rotamer well width. **(d)** Conformer resolution.

	rmsd	Fraction Correct Rigid (calc)	Fraction Correct Flexible (calc)	Fraction Correct Total (calc)
(a) LJ Repulsive				
hard repulsive ^a	0.26	0.69	0.70	0.70
small radii	0.25	0.60	0.80	0.72
soft repulsive	0.28	0.36	0.94	0.70
(b) Base Rotamer Library				
default	0.26	0.73	0.64	0.68
large ^a	0.26	0.69	0.70	0.70
(c) Rotamer Well Width				
0.5 sd ^b	0.30	0.73	0.58	0.64
1.0 sd ^b	0.27	0.73	0.64	0.67
1.5 sd ^{a c}	0.26	0.69	0.70	0.70
2.0 sd ^c	0.26	0.69	0.71	0.70
3.0 sd ^c	0.26	0.69	0.70	0.70
(d) Conformer Resolution				
10 degrees ^a	0.26	0.69	0.70	0.70
15 degrees	0.26	0.68	0.71	0.70
20 degrees	0.27	0.67	0.68	0.67
25 degrees	0.27	0.69	0.64	0.66
1 rotamer	0.33	0.80	0.47	0.60

^aIndicates parameter values used for the results in Table 2

^bSampled with 5 degree conformer resolution

^cSampled with 10 degree conformer resolution

Table 4

Comparison of the results of Models 2 and 3 to results from other methods as indicated.

	Ensembles of X-ray structures ^a				Explicit Solvent MD ^b		Model 2		Model 3	
	# of structs	r	rmsd	r	rmsd	r	rmsd	r ^c	rmsd	
a3D						0.64	0.22	0.74	0.17	
albp	14	0.73	0.19			0.79	0.19	0.76	0.19	
calmodulin	28	0.72	0.20			0.81	0.21	0.78	0.21	
Cdc42hs	13	0.53	0.30			0.31	0.33	0.41	0.31	
cytochrome c2						0.37	0.24	0.52	0.22	
eglin c	10	0.37	0.30	0.84		0.55	0.25	0.59	0.25	
flavodoxin						0.63	0.21	0.62	0.21	
FNfn10				0.51	0.23	0.40	0.29	0.42	0.24	
fyn-sh3	12	0.74	0.21			0.89	0.15	0.62	0.20	
gb1						0.49	0.26	0.68	0.22	
hiv1 protease	330	0.74	0.17			0.73	0.22	0.62	0.23	
mfabp						0.59	0.27	0.60	0.26	
NTnC	13	0.69	0.19			0.77	0.23	0.73	0.22	
plec sh2						0.57	0.26	0.60	0.24	
protein L						−0.07 ^d	0.35 ^d	−0.06 ^d	0.33 ^d	
TNfn3				0.62	0.23	0.34	0.32	0.38	0.31	
ubiquitin	13	0.76	0.18			0.60 ^d	0.27 ^d	0.64 ^d	0.24 ^d	
<hr/>										
Average (Proteins from ^a)		0.66	0.22			0.68 [*]	0.23 [*]	0.65 [*]	0.23 [*]	
Average (All proteins)						0.55 [*]	0.25 [*]	0.57 [*]	0.24 [*]	

^aData from

^bEglin c data from ²¹; FNfn10 and TNfn3 data from ²⁴

^cCalculated over all S²_{calc} values for the 10 backbone ensembles

^dThese proteins exhibited problems modeling threonines (see Results for details)

^{*}Sum of the values for each protein divided by the number of proteins (i.e. different that the metric at the bottom of Table 2); provided for comparison with values from ³⁶