Published in final edited form as:

J Mol Biol. 2008 July 18; 380(4): 742–756. doi:10.1016/j.jmb.2008.05.023.

Backrub-like backbone simulation recapitulates natural protein conformational variability and improves mutant side-chain prediction

Colin A. Smith^{1,2,3,4*} and Tanja Kortemme^{1,2,3*}

¹Graduate Program in Biological and Medical Informatics, University of California, San Francisco

²California Institute for Quantitative Biosciences, University of California, San Francisco

³Department of Biopharmaceutical Sciences, University of California, San Francisco

⁴Department of Pharmaceutical Chemistry, University of California, San Francisco

Summary

Incorporation of effective backbone sampling into protein simulation and design is an important step in increasing the accuracy of computational protein modeling. Recent analysis of high-resolution crystal structures has suggested a new model, termed backrub, to describe localized, hinge-like alternative backbone and side chain conformations observed in the crystal lattice. The model involves internal backbone rotations about axes between C α atoms. Based on this observation, we have implemented a backrub-inspired sampling method in the Rosetta structure prediction and design program. We evaluate this model of backbone flexibility using three different tests. First, we show that Rosetta backrub simulations recapitulate the correlation between backbone and side-chain conformations in the high-resolution crystal structures upon which the model was based. As a second test of backrub sampling, we show that backbone flexibility improves the accuracy of predicting point-mutant side chain conformations over fixed backbone rotameric sampling alone. Finally, we show that backrub sampling of triosephosphate isomerase loop 6 can capture the ms/ μ s oscillation between the open and closed states observed in solution. Our results suggest that backrub sampling captures a sizable fraction of localized conformational changes that occur in natural proteins. Application of this simple model of backbone motions may significantly improve both protein design and atomistic simulations of localized protein flexibility.

Keywords

flexible backbone sampling; backrub motion; point mutation; Monte Carlo; triosephosphate isomerase loop 6

Introduction

Proteins undergo conformational fluctuations in response to thermal energy, binding events, and mutation. Understanding and predicting such excursions around the native state of a protein

*corresponding authors: colin.smith@ucsf.edu, kortemme@cgl.ucsf.edu.

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.

is a key challenge in computational molecular biology. Side chain sampling¹ has been shown to be an extremely useful first-order method for predicting small-scale conformational change. Successful applications include protein-protein docking^{2,3}, total redesign of protein sequences^{4,5}, and redesign of both protein-protein⁶ and protein-DNA⁷ interfaces. However, one key approximation made by many of these applications is keeping the backbone structure fixed. In actual proteins the backbone often undergoes subtle shifts in response to binding events⁸ or sequence changes⁹. Successfully capturing such near-native shifts is thus important for many docking and design applications.

Numerous methods have been developed to take backbone flexibility into account for both the whole protein and local subsections. Molecular dynamics is currently one of the most pervasive methods. However, in the absence of a steep energy gradient, dynamics depend on random thermal velocities and a long sequence of time steps to sample motions as simple as a rotamer change. Monte Carlo minimization of backbone torsion angles^{10–12} has also been very successful, but can result in highly non-local displacements of the protein backbone and becomes increasingly less efficient with greater protein size. Insertion of peptide fragments has been used for *de novo* protein structure prediction¹³ and loop prediction¹⁴, but causes similar propagating changes. Several non-local sampling techniques have been applied to protein design including random torsion angle sampling¹⁵ and more correlated methods such as fragment insertion¹⁶, parameterized coiled-coils¹⁷, and normal mode analysis¹⁸. These methods make use of patterns commonly observed in protein structures or a harmonic approximation of intra-protein interactions to increase backbone sampling efficiency. Other methods have addressed the problem of making local perturbations using heuristics to iteratively optimize backbone torsion angles until distortions of covalent geometry are minimized^{19–21}, but those techniques sometimes leave strained chain junctions that must be relaxed with other algorithms. Another method, called wriggling²², was developed to make partially local moves in which groups of four torsion angles are changed simultaneously to minimize the displacement of distant atoms.

Deformations of protein backbones are truly local only if all consecutive atoms beyond the perturbed region remain fixed. Several local methods exist, the first being introduced by Go and Scheraga²³ with numerous subsequent refinements and adaptations^{24–27}. These methods involve making a random prerotation of one or more backbone angles, followed by solving a geometric constraint equation for six other backbone degrees of freedom to maintain the locality of the move. Several of the methods incorporated bond angle sampling, either as part of the prerotation^{24,27}, or both the prerotation and the solved constraint equation²⁶. The latter work also biased the prerotations towards less perturbed backbone conformations. The implementation of these methods is more complex than other common techniques like rotamer sampling. Another drawback is that such loop closure methods are biased towards proposing moves that satisfy bonded, geometric constraints, whose multiple free rotation axes can lead to radically different conformations, often with substantial steric clashes and unsatisfied hydrogen bonds. Those non-bonded factors are particularly relevant in highly packed protein cores and interfaces.

The work described here, instead of being motivated by geometric constraints, derives its motional model from conformational variations observed in high-resolution ($\leq 1\text{\AA}$) crystal structures²⁸. The fluctuations observed in the crystal lattice motivated Davis et al. to create a simple model, called Backrub, for subtle backbone shifts using just three residues. The core idea in this work is to use that type of motion, observed in nature, to computationally sample backbone configurations in a generalized scheme. A similar move set was recently described²⁹ in the context of a simplified energy function. Here, we investigate the utility of the backrub move to sample conformations in the context of the Rosetta all-atom force field. Rosetta has been successfully used for protein-protein docking³, protein-ligand docking³⁰,

redesign of protein cores¹⁶, design of new protein interface specificities⁶, and *de novo* prediction of small protein structures³¹. As an initial test, we recapitulate the backbone/side-chain correlations observed in the same high-resolution structures that inspired the Backrub model. We go on to show that backrub backbone flexibility improves side-chain modeling of point mutations. Finally, as a demonstration of the method's potential, we present a proof-of-concept simulation showing efficient sampling of the opening and closing of triosephosphate isomerase loop 6. Our results indicate that the backbone sampling described here captures a sizable fraction of the subtle conformational variability found in folded proteins.

Results

We implemented the backrub sampling protocol inspired by motion observed in protein structures²⁸ (see Figure 1, Figure 2, and Methods), and evaluated it using three different tests: First, we sought to determine whether the motional model, combined with an all-atom force field, could recapitulate the variation seen in occurrences of a backrub motion in high-resolution crystal structures. Secondly, we test whether backrub sampling can improve the accuracy of modeling small backbone and side chain conformational changes in response to single point mutations in a set of crystal structure pairs. Finally, we show simulations indicating that backrub sampling can capture conformational variability observed in a long time-scale loop motion.

Test 1: Simulation of 3-Residue Backrubs

Davis et al.²⁸ derived the “Backrub” model of protein backbone motion from examples of three residue segments exhibiting multiple backbone conformations in high-resolution (≤ 1.0 Å) crystal structures. To model those variations, they used the C α atoms as pivot points and enumerated the three possible rotation axes between them. By manually rotating the backbone around those axes, they were able to model the conformational transitions in a significant number of cases. They catalogued 126 such instances in the PDB that fit their model, the majority of which involved a simultaneous rotamer change. In those cases, the backbone-determined location of the C β atom significantly altered the conformation of attached side chain atoms. As an initial test of our generalized backrub sampling method, we used focused Monte Carlo simulations to determine whether we could detect distinct populations of coupled backbone/side-chain conformations centered on coordinates observed in the PDB.

Out of 161 derived starting structures (see Materials and Methods), the majority (105) came from PDB residue entries with χ_1 angles of the central side chain, i , occupying multiple rotameric bins (-60° , 60° , 180°). In our analysis, we therefore used the χ_1 angle as a one-dimensional representation of the side chain conformation. We used the C $\alpha_{i,\text{initial}}$ –C α_{i-1} –C α_{i+1} –C $\alpha_{i,\text{current}}$ pseudo-dihedral angle (τ_{disp}), to represent the backbone conformation of the 3-residue segment. (Figure 3) We wanted to determine whether the simulations showed a similar correspondence between side-chain and backbone conformation to that observed in the crystal structures. To answer that question, we calculated τ_{disp} probability distributions for each of the χ_1 bins visited during the simulations and compared those distributions to the crystallographic τ_{disp} backbone angles. An example analysis of a simulation showing good agreement with the PDB is given in Figure 4.

A simple binary metric indicating if the simulations correctly captured the side-chain/backbone bias is whether the average backbone conformations for each χ_1 bin ($\langle \tau_{\text{disp}} | \chi_1 \rangle$, circles in Figure 4B) were in the same relative orientations found in the PDB (vertical lines in Figure 4B). This is easiest to interpret for those residues with PDB side-chain conformations in exactly two χ_1 bins, as in Figure 4. There were 98 starting structures where that was the case and of those, in 76 cases the simulations did visit both χ_1 bins observed in the PDB, making the comparison possible. Out of these 76, 55 (73%) showed the correct bias, which is significantly

better (chi-square p-value $1 \cdot 10^{-4}$) than would be expected at random (50%). When only buried side chains (SASA <30%, see below) are considered, 15 out of 17 (88%) are correct.

A comparison between the mean τ_{disp} angles from the simulations and those determined from the PDB shows reasonable agreement (Figure 5). As the accuracy of rotamer prediction has been shown to be strongly dependent on the degree of residue burial^{32–35}, we show results for 24 residues with solvent accessible surface areas (SASA) of <30%, using the surface area of an extended residue flanked by glycines as the reference SASA. Deviations from the diagonal can result from both scoring/sampling problems in our modeling procedure and uncertainty in the crystallographic fitting. However, there is a reasonable positive correlation ($R = 0.64$). The correlation becomes clearer when individual simulations (connected by lines) are examined. Nearly all such lines show positive slopes, indicating that the simulations capture the direction of correlated side chain and backbone conformational changes correctly in many cases, albeit with some variation in the absolute magnitude.

A notable observation is that at certain backbone τ_{disp} angles (i.e. $< 3^\circ$ or $> 17^\circ$ in Figure 4), some side chain conformations are completely inaccessible. The intervening backbone conformations form a transitional zone where the rotameric change becomes more and more energetically favorable. In our simulations, there is little evidence for an energetic barrier between the subtle differences in backbone conformations. On the other hand, there can be significant energy barriers involved in side chain transitions, particularly in the protein core. Our data indicate that the side chain rotamers may lock the backbone into slightly different conformations, giving rise to the alternate conformations observed by Davis et al.²⁸. This mirrors another simulation study, where a side chain transition played a key role in stabilizing a relatively unconstrained backbone conformational transition³⁶.

Backbone/ χ_1 Correlation in Crystal Structures Alone—After observing the correlation of backbone conformation with the side chain χ_1 angle in our simulations, we wanted to determine whether the same biases could be observed at a global level in the Davis et al.²⁸ dataset, irrespective of the simulation results. To do so, we considered the 68 residues (out of 126) where there were at least two χ_1 bins represented in the PDB. For every alternate backbone conformation, we calculated the τ_{disp} angle, using the first conformation as the reference structure. We then normalized the τ_{disp} angles for each of the 68 residues to make the τ_{disp} weighted mean (using PDB occupancies as the weights) of each individual residue 0. The distribution of τ_{disp} for each χ_1 bin is shown in Figure 6.

Interestingly, the distribution of τ_{disp} angles for the 60° χ_1 bin was significantly different from the distributions for the $-60^\circ/180^\circ$ χ_1 bins. (Figure 6A) There was a 5.8° difference in means between the 60° and the joint $-60^\circ/180^\circ$ distributions. The structural explanation for the difference is quite clear when the orientation of the $C\beta-C/O\gamma$ bond vector is visualized on a hypothetical backbone with the central side chain pointing up and the Ca_{i-1} atom in front of the Ca_{i+1} atom. (Figure 6B) In that orientation, the 60° $C\beta-C/O\gamma$ vector points nearly perpendicular to the $Ca_{i-1}-Ca_{i+1}$ axis, pushing the backbone in a counter-clockwise direction. That rotation corresponds to a negative τ_{disp} angle. In the $-60^\circ/180^\circ$ bins, the $C\beta-C/O\gamma$ vectors point in the opposite direction but are much less perpendicular. The degree of perpendicularity helps explain the negative skew of the 60° distribution in comparison to the relative symmetry of the $-60^\circ/180^\circ$ distributions.

In principle, the dependence of backbone conformation on side-chain conformation could be used to derive coupled moves in sequence and structural optimization algorithms. For example, the differences in backbone distributions could be used to restrict sampling of backbone conformations when switching into or out of the 60° χ_1 rotameric bin.

Test 2: Point Mutant Side Chain Prediction

In addition to distinct conformations observed in the high-resolution crystal structure dataset discussed above, another context in which subtle backbone differences may be important are residue point mutation. A single-residue point mutation represents the simplest of increasingly more difficult structural modeling tasks where one is given a template and then must predict the new low energy conformation after a known perturbation. In addition, the ability to accurately predict the conformation of a side-chain upon point mutation has direct bearing on the success of protein sequence design algorithms.

We wanted to determine the extent to which generalized backrub sampling could improve the prediction of point mutant side chains, especially when using a fixed rotamer library as commonly done in computational protein design methods. Recently, Bordner and Abagyan³⁷ compiled a large benchmark set of PDB structure pairs differing by a single point mutation. We applied the generalized backrub protocol to locally refine structural models after mutation/fixed backbone rotamer optimization in Rosetta¹⁶. We found that overall, incorporation of backrub sampling improved both side chain heavy atom RMSD and χ_1/χ_2 recovery within 40°. (Figure 7) We also found that the local backbone RMSD between PDB structure pairs was correlated with prediction difficulty, in terms of both RMSD and χ_1/χ_2 recovery. The larger the backbone conformational change upon mutation, the larger was the improvement resulting from backrub sampling. In particular, the fraction of pairs with the highest starting RMSD showed the most sizeable improvement. Similar observations were made in a previous study¹⁵ which showed improvements in prediction of side chain conformations after core substitutions in T4 lysozyme when comparing flexible with fixed backbone methods. There backbone flexibility was modeled using a different mechanism employing random continuous adjustments of ± 3 degrees to each backbone angle.

In addition to the dependence on initial backbone RMSD, we also investigated how a number of other factors affected the extent of improvement, including the radius of neighboring residues allowed to change rotameric conformations (4, 5, 6, 7, and 8 Å from the mutated residue) and the degree of burial of the mutated residue (< 5%, < 30%, and $\leq 100\%$ solvent accessible surface area). Figure 7 shows results from point mutant predictions that showed the best overall improvement in prediction accuracy, considering < 5% solvent exposure and a 6 Å sampling radius. A complete enumeration of the radius of sampling and degree of burial is shown in supplementary data for both side-chain RMSD and χ_1/χ_2 recovery. Between a 4 Å and 8 Å sampling radius, the prediction of side chain RMSD does not change significantly. Backrub sampling gives better χ_1/χ_2 recovery at 6 Å than using any other radius. Considering the amount of residue burial, backrub sampling continues to improve overall RMSD prediction somewhat using a 30% SASA cutoff. When evaluating all residues including those that are largely solvent exposed, backrub sampling still improves predictions for high backbone RMSD pairs, but makes low RMSD pairs slightly worse. (see Supplementary Material)

Figure 8 shows examples of backrub sampling improving side chain prediction. The improvement can come from two sources, namely better prediction of the side-chain conformation and better prediction of the protein backbone. In some cases, the side-chain improvement comes at the cost of backbone prediction accuracy, as is shown in the last row of images. However, the worsening of backbone ($C\alpha/C\beta$) RMSD is relatively small compared with the improvement in side chain RMSD. (Table 1) The source of the error could lie in crystallographic uncertainty, inaccuracies in the scoring function, or compensation for a discretized rotameric side-chain representation.

Test 3: Triosephosphate Isomerase Loop 6 Simulation

As a third, proof-of-principle test of the backrub sampling protocol, we investigated a much larger conformational change. The hinge motion of triosephosphate isomerase (TIM) loop 6 is a well-characterized example of a protein segment undergoing significant conformational change while maintaining a relatively rigid internal conformation. The C α RMSD between an 11 residue segment (V167-T177) in the closed, (PDB [2YPI](#)38), and open (PDB [1YPI](#)39) conformations is 4.6 Å. We found that by manually rotating the closed segment 50° about the C α ₁₆₇–C α ₁₇₇ axis, the C α RMSD to the open form drops from 4.6 to 1.1 Å, indicating that a backrub simulation may capture much of the conformational variability of the TIM loop. However, such a large rotation introduces several significant steric clashes and adds sizable backbone bond angle strain at residue V167.

We wanted to determine whether generalized backrub sampling of the loop region could capture the same degree of conformational variability without producing energetically unreasonable conformations. Previous studies using molecular dynamics have had difficulty capturing the TIM loop 6 conformational transition^{40,41}. Notably, the simulations required temperatures from 1000–1200 K to see transitions from one state to the other.

We ran simulations starting from both the open conformation ([1YPI](#)) and the closed conformation ([2YPI](#)). In each simulation, we allowed backrub moves of size 2–12 on residues 165–179. Residues 128–130 showed small but potentially significant changes between the two conformations, so we allowed backrub moves of size 2–3 for those residues. In addition to all of those residues, rotamer changes were allowed for residues whose side chains were in the vicinity of the loop using a 5 Å cutoff and by visual inspection (3, 7, 95, 96, 131, 134, 139, 164, 180, 183, 208, 211, 216, 219, 220, 223, 230). Each simulation was run without the ligand, making the atomic composition identical. We ran the simulations for 1.5 million Monte Carlo moves, using a temperature of 302 K in the Metropolis criterion. Each simulation took 14 hours to complete on a single 2.0GHz Xeon processor.

To analyze the simulation trajectories, we calculated the C α RMSD of the loop from both the open ([1YPI](#)) and closed ([2YPI](#)) conformations. The sign of the difference between those RMSDs indicates whether the loop is closer to the open (positive sign) or closed (negative sign) conformation. Starting from the closed conformation with the ligand removed, the backrub simulations were able to oscillate between the open and closed forms of the loop many times during a 1.5 million step simulation. Eight example transitions are pictured in Figure 9, where the minimum RMSD for each approach to the open form ranged 1.57–2.2 Å and the RMSD values for return to the closed form ranged 1.37–2.36 Å. The loop structure (V167-T177) maintained a relatively stable internal conformation over the length of the simulation, with an average aligned C α RMSD of 1.3 Å (0.3 Å standard deviation) from the starting structure.

We found that the motion of the loop depended on the starting structure, not always showing the opening and closing behavior. In the simulation starting from the open conformation, there was a transient excursion closer to the closed form (within 2.38 Å) at the beginning of the simulation. After that, the loop stayed in a predominantly open conformation for the remainder of the simulation, in some cases migrating to a “hyper open” state up to 8.29/4.23 Å from the closed and open structures, respectively.

There are several explanations for the difference in the simulations and lack of convergence. In addition to possibly needing more sampling to equilibrate, it may be that the anchor points or other fixed regions of the different starting protein structures bias the loop motion. Another possible explanation is that the backrub motions are not sufficiently sampling the internal degrees of freedom in the loop. A likely limitation is that proline residues (at TIM residue

positions 168 and 176) are not currently allowed as pivot residues in backrub sampling, thus keeping all of their internal angles fixed. This effect may be substantial in the TIM loop case as P168 shows a ψ angle change of 40° between the two conformations. (Supplemental Figure 12)

To assess the degree to which more complete backrub sampling (not limiting sampling to just the loop 6 region) captures the flexibility of the TIM structure, we ran multiple simulations of the complete TIM dimer (with the constraint of fixing the backbone coordinates of 17 core residues in each monomer). The loop 6 region does indeed show the largest conformational variability: three of the four largest calculated B-factors from those simulations are in the tip of loop 6 (G171–G173). (Supplemental Figure 13) In addition to the high calculated B-factors for loop 6, some flexibility was observed in several other regions. These regions also showed structural differences between the open and closed crystal structures.

Discussion

We have shown that the backrub sampling method is useful for sampling small, high-resolution conformational fluctuations as well as a larger, functionally relevant conformational change. In addition to capturing the structural variability of single sequences, generalized backrub sampling also improves modeling of changes to protein structures upon point mutation. While many of the backbone movements are less than 1 Å, they can result in significant displacements of the attached side chains. In addition, the localized breathing motion that backrub sampling emphasizes can allow otherwise energetically unfavorable rotameric transitions.

This work supports the conclusion advanced by Davis et al²⁸ that protein backbones are influenced by side-chain conformations in a predictable manner, complementing the accepted notion that side-chain conformations can be backbone dependent. In backbone-dependent rotamer libraries, the side chain conformation is influenced by the ϕ and ψ angles of the residue itself³³. Our simulations and analysis support the notion of a second-order correlation between a central side-chain and the protein backbone in adjacent residues. The 3-residue simulations indicate that the energetic barriers between the relevant backbone conformations can be significantly less than those typically associated with side-chain rotamer transitions.

As a sampling method, the overall philosophy behind the move used here is somewhat different from other methods (although bearing similarities to local perturbation approaches highlighted earlier^{26,27}). First, it is a generalization of movement that is observed in nature at both small and large amplitude. Rotameric sampling was likewise inspired by observations made from crystal structures. Second, instead of treating bond angles as inviolable, it takes advantage of the small but significant flexibility in the bond angle to move a set of backbone atoms through a single, unified rotation. As indicated by the 3-residue simulations, backrub sampling helps free the protein backbone to explore an ensemble of conformations around the native state. While molecular dynamics could be used to accomplish the same goal, correlated movement of atoms can take a considerable number of time steps, unless there is already a set of forces accelerating the atoms in a concerted direction. Simultaneous, correlated rotation of many atoms is one of the strengths of rotamer sampling. Backrub sampling shares that strength.

Backrub moves are biased towards sampling hinge-like protein motion. Another type of motion sometimes seen in proteins is a shearing move, where a subsection of the protein is translated laterally in relation to the remainder of the protein. That type of motion is almost completely orthogonal to the generalized backrub move described in this work. However, the same philosophy as defined originally for the backrub move^{28,29} and described here could be applied to model shearing moves directly. This would require four Ca atoms as pivot points. One embodiment may consist of rotating Ca_2 about Ca_1 in the $\text{Ca}_1\text{-Ca}_2\text{-Ca}_3$ plane, and rotating

Ca_3 about Ca_4 in the Ca_2 - Ca_3 - Ca_4 plane, such that the Ca_2 - Ca_3 distance is preserved. Though somewhat more complex, this type of move may help model subtle shifts of alpha helices and other structural elements by small but significant distances. While development of additional move sets may prove useful in capturing the full range of protein motion, another promising avenue might involve combination of backrub and rotamer sampling with traditional molecular dynamics in a hybrid Monte Carlo approach. Towards that end, a discussion and validation of the effect of backrub sampling on detailed balance is given in supplementary information.

The combination of backrub-inspired backbone flexibility with side chain sampling and protein design protocols has a number of useful practical applications. First, we show here that employing backrub motions in a high-resolution refinement protocol improves mutant side chain predictions in two large datasets, comprising 126 backrub motions in 19 high-resolution structures and 2,023 pairs of protein point mutant structures. These results suggest that backrub sampling may enhance the applicability and accuracy of methods to estimate the change in fold stability or binding affinity of proteins upon point mutation^{15,42–45}. Second, we show in related work that incorporation of backbone flexibility using the backrub model significantly increases the agreement between modeled side chain conformational variability in folded proteins and side chain relaxation order parameters measured by NMR (Friedland et al., in press). Such simulations may provide insights into protein dynamics and mechanisms of correlated motions. Finally, the use of near-native backbone ensembles has been shown to broaden the set of sequences identified by computational methods and result in successful designs¹⁸. Similarly, we find that design simulations employing backrub-generated backbone ensembles predict protein sequence families more similar to those observed by experimental phage display selection methods than predictions using just the crystallographic backbone (Humphris & Kortemme, unpublished data). Given its relative simplicity in implementation and ability to capture relevant conformational changes inspired by observed alternative conformations in high-resolution structures, the backrub method may be generally useful for a broad spectrum of side chain sampling and protein design protocols.

Materials and Methods

Generalized Backrub Move

The backrub move (Figure 1) is applied to an internal protein segment two or more residues long and consists of a geometric rotation by a random angle, τ , about an axis defined by the flanking Ca atoms. The move simultaneously changes 6 internal backbone degrees of freedom in the protein, namely the ϕ and ψ angles at both pivot points and the N- Ca -C bond angle, α , at both pivots. (Variable names follow the conventions of Betancourt²⁹ instead of Davis²⁸, which uses τ for the N- Ca -C bond angle.)

The sampling strategy employed here is similar to the one described by Betancourt²⁹, in that three types of moves are used, namely backbone only, rotamer only, and rotamer/backbone. However, the move selection is significantly different. We were interested in selectively sampling backbone motion in specified local regions of the protein while keeping other regions fixed. Therefore, we devised a flexible scheme for specifying which parts of the protein structure were variable. At the highest level, the operator indicates for each residue whether to sample the backbone, side chain or both. Backrub moves are only allowed for segments where backbone sampling is enabled for both the beginning (i) and ending (j) residues, and all intervening residues. Because the proline side-chain rejoins to the backbone at the amide nitrogen, it has been excluded as a pivot point. In addition, the minimum and maximum segment size ($j-i+1$) can be varied. By default, the minimum segment size is 2, corresponding to a rotation of the atoms making up the peptide bond between two consecutive Ca atoms. The default maximum segment size is 12, although higher or lower values may be desired depending on the application.

Given that information, a sparse upper-triangular boolean matrix, B , is created where $B[i,j]$ indicates whether a move starting at residue i and ending at residue j is permissible. B can then be further modified to enable or disable individual residue segments. Before beginning Monte Carlo sampling, a data structure is generated from B that lists each possible segment size, along with all starting residues compatible with that particular segment size. Segment selection then becomes the simple procedure of first selecting a random segment size uniformly from all allowed segment sizes, and then selecting a random segment from all allowed segments with the selected size. As there are fewer long segments than short segments, individual long segments will be selected slightly more often than individual short segments.

Monte Carlo Sampling Protocol

During the course of an actual Monte Carlo simulation, the protocol described in Figure 2 is used to perform each move. At the beginning of a step, a decision to make a rotamer only move is made according to the adjustable probability, $P_{rotamer}$. The default value of $P_{rotamer}$ is 0.25. If a rotamer only move is chosen, a single variable side-chain is randomly selected and a rotamer is chosen from a library generated using a backbone-dependent rotamer library^{16,46}. The rotamer library is initialized using the ϕ/ψ angles from the starting structure and not updated during the simulation. If a rotamer only move is not made, then a random segment and angle is selected as described previously, and the rotation is applied. At that point, the algorithm decides whether to terminate the move (leaving it as a backbone only move) according to the second adjustable probability, $P_{backbone}$. The default value of $P_{backbone}$ is 0.75 to emphasize the more frequently accepted backbone only moves. If the move is not ended, then one or two residues (respective probabilities 0.75 and 0.25) are selected from along the length of the perturbed backbone segment and random rotamers are chosen for those residues. After all structural perturbations are complete, the move is evaluated using the Metropolis criterion and the Rosetta energy function. Constraints on the degree of angular perturbation and example acceptance probabilities are given in subsequent sections.

Rosetta Scoring Function

In a previous implementation²⁹ of the move described here, N-C α -C bond angles were not energetically scored and were constrained to being within 10° of the median bond angle observed in PDB structures. In this work we used bond angle potentials from the Amber ff94⁴⁷ and CHARMM22⁴⁸ force fields.

In addition to the added bond angle term, the Rosetta full-atom scoring function⁴⁹ uses several bonded terms including a ϕ/ψ angle term based on Ramachandran distributions and a χ angle term based on Dunbrack rotamer statistics. For evaluating non-bonded interactions, Rosetta uses a van der Waals term resembling a Lennard-Jones potential, an explicit geometry-dependent hydrogen bonding term⁴⁴, a short-range electrostatics term approximated by a residue-specific pairwise distance potential, and the Lazaridis/Karplus implicit solvation model⁵⁰.

Bond Angle Constraints

In order to reduce the amount of bond angle strain imposed, we sought to bracket the randomly chosen rotation angle such that the bond angle strain never exceeds a threshold value, α_{max} . We used a previously described method²⁹ to analytically determine the set of τ intervals satisfying that constraint. Briefly, the method involves solving for τ a trigonometric equation that relates α to τ and then plugging in $\alpha_{ideal} - \alpha_{max}$ and $\alpha_{ideal} + \alpha_{max}$ for both the starting and ending residues. The resulting values of τ establish the intervals of allowed τ angles. We term that set of intervals $I_{bond\ angle}$.

To determine how the acceptance ratio decays for increasingly strained bond angles, we performed a long (10^6 step) Rosetta Monte Carlo simulation using a PDZ domain structure (PDB [2H3L](#)51), imposing the Amber bond angle potential and limiting bond angles to within 10° of the overall bond angle minimum. Move attempts were binned by the maximum deviation (at either pivot point) from the Amber ideal bond angle and acceptance ratios were calculated (Supplemental Figure 1A). The acceptance rate remained above 20% for all moves where both bond angles remained within 6.25° degrees of ideal. At the extreme, where one of the bond angles reached a 10° deviation from ideal, the acceptance rate dropped to 6.6%. Those rates may initially seem somewhat high, given the severity of the angular strain. However, there are always two bond angles changing during any move. At equilibrium, moves may transfer bond angle strain from one residue to the other, without increasing the total amount of strain in the system.

Rotation Angle Constraints

Examining the acceptance statistics further, we made the intuitive observation that as the magnitude of angular displacement increases, the acceptance statistics drop almost exponentially (Supplemental Figure 1B). This phenomenon is best explained through sterics, where the larger the rotation, the more likely a deleterious steric clash is encountered. We therefore imposed an additional constraint upon moves that restricted the maximum angular rotation to a given threshold, τ_{max} . This was done by generating an additional interval, $I_{rotation\ angle} = [-\tau_{max}, \tau_{max}]$, and then calculating the intersection, $I = I_{bond\ angle} \cap I_{rotation\ angle}$, of that interval with the previously calculated intervals bracketing the bond angle. Importantly, this additional constraint can create an imbalance in the selection probabilities for the possible angles, as the total angular range of the intervals may be different before (I) and after (I') moves. (Supplemental Figure 2) Because the probability of selecting a given angle is inversely proportional to the number of possible values, the following acceptance criterion can be used to produce uniform selection probabilities:

$$P(\tau) = \min \left[1, \frac{l}{l'} \right]$$

Trial move τ angles are generated using this procedure:

1. Calculate the total length, l , of the set of intervals, I .
2. Choose a random threshold, t , uniformly from the interval $[0, 1]$.
3. Choose a random angle, τ , uniformly from I .
4. At angle τ , calculate the new rotational interval, $I_{rotation\ angle}' = [\tau - \tau_{max}, \tau + \tau_{max}]$.
5. Calculate $I' = I_{bond\ angle} \cap I_{rotation\ angle}'$ and the total length, l' , of I' .
6. If $l/l' \geq t$, return τ . Otherwise go back to step 3.

Because l and l' are generally quite similar, this procedure rarely iterates more than several times and is considerably less costly than other parts of the simulation.

To ameliorate the reduction in acceptance ratio for large segment sizes, the τ_{max} parameter is varied for each possible residue segment. This is distinguished from the Betancourt strategy of making equal magnitude displacements regardless of segment size. Different values of τ_{max} are stored in another sparse upper triangular matrix, T . Based on empirical observation of the acceptance statistics, we devised the following rule relating τ_{max} to segment size, s :

$$\tau_{max} = \begin{cases} 40; & s=2 \\ 23 - s; & s \geq 3 \end{cases}$$

We found peptide bonds (size 2) to be significantly more flexible than other segment sizes. The large increase in flexibility is partially due to peptide bonds lacking the steric constraints of other segment sizes. However, when one looks at the distribution of allowable τ angles, given only a 10° bond angle cutoff, it is also clear that peptide bond segments have significantly more flexibility than larger segments (Supplemental Figure 1C). In addition to Davis et al.²⁸, a similar type of motion has also been observed in unbiased computational simulations. A recent analysis of correlated ϕ/ψ motions in a large set of molecular dynamics trajectories also observed significant, localized peptide bond fluctuations⁵². Additionally, in pairs of structures of the same protein crystallized multiple times, larger “peptide flips” (involving rotations $\sim 180^\circ$) are often observed⁵³.

As a result of constraining both the N-C α -C bond angles and maximum angular displacement during a move, the acceptance statistics remain relatively high for segments sizes from 2 to 12 (Supplemental Figure 1D). For the PDZ domain test simulation, backbone only moves showed an average acceptance ratio of 29%, and rotamer only moves showed an acceptance ratio of 34% (data not shown). When combined with the much less accepted simultaneous rotamer/backbone moves, the overall acceptance ratio drops to 26% (weighted mean of all move types). Elimination of simultaneous rotamer/backbone moves would increase the overall acceptance rate to 30%.

Optimized Placement of C β and H α Atoms

An important methodological consideration in a procedure that modulates C α backbone bond angles is how the positions of the branching C β and hydrogen atoms are simultaneously varied. Placement of branching atoms means positioning of C β and H α relative to the N, C α , and C atoms. A number of angular bisecting heuristics can be applied to place those atoms in positions with acceptable geometries. In this work, to reduce bond angle strain, the branching atoms are placed in positions at the minimum of the force field bond angle potential, given the current N-C α -C bond angle. Minimization after every Monte Carlo move would be computationally expensive. Fortunately, the minimized internal coordinates of those atoms follow a predictable pattern (Supplemental Figure 4). To enable fast updates of the position of a branching atom X, quadratic functions were fit that related a series of N-C α -C backbone bond angles to the corresponding fully minimized branching atom internal coordinates, namely the C-N-C α -X torsion offset from ϕ , and the N-C α -X bond angle (Supplemental Table 2). These fits were very accurate even to highly unfavorable bond angle energies of 20 kcal/mol.

After every backrub move, the new branching atom positions are found using those quadratic fits. Subsequently, the coordinates of the side chain prior to the move are rotated about the C α atom pivot point such that the old C β atom is collinear with the new C α -C β axis. Finally, the whole side chain is rotated slightly about the C α -C β axis to restore the χ_1 angle to its original value.

Simulation of 3-Residue Backrubs (Test 1)

Davis et al.²⁸ identified 126 positions in 19 high resolution ($\leq 1.0\text{\AA}$) crystal structures where there was evidence for a localized rotation of a 3-residue segment of the protein backbone. In some cases, the conformational variability in the backbone was only implied by alternate C β atom positions in the PDB file, with a single set of C α atom coordinates representing the mean location of multiple C α atom positions. In those cases, we used a single starting structure with the C β atom optimized according to the bond angle potential. If alternate C α coordinates were present in the PDB file, we generated 2–3 starting structures, one for each variant letter (A, B, or C) in the contiguous set of atoms with alternate backbone coordinates. All other alternate atom coordinates were set to the A variants. In total, this procedure yielded 161 starting structures in the 3-residue backrub set.

The simulations for each identified backrub were as follows. Given a three residue backbone motion centered on residue i , angular perturbations were enabled for residue pairs $(i-1, i+1)$, $(i-1, i)$, and $(i, i+1)$. The side chain of residue i was also allowed to sample different rotameric states. 200,000 Monte Carlo steps were run at a temperature of 302 K. 50% of the steps consisted of a random perturbation of the $(i-1, i+1)$ angle and a simultaneous rotamer swap. The other 50% of steps consisted of a random perturbation of either the $(i-1, i)$ or $(i, i+1)$ angle and no rotameric sampling.

Point Mutant Side Chain Prediction (Test 2)

We used a benchmark set of 2,141 pairs of protein structures for which the only difference was a single point mutation, aside from extra or missing residues at the N and C termini³⁷. We removed 7 pairs from the set that had, at the mutated residue position, either missing side chain atoms or a non-canonical amino acid. We also removed 8 pairs for which the mutation was duplicated in another pair in the list. Finally we removed 103 pairs that had either missing or zero occupancy backbone atoms in the first structure in the pair. Structures with missing or zero occupancy backbone atoms in the second structure were removed during analysis (see below). That left 2,023 ordered pairs of structures.

During side-chain prediction, we sampled conformations (backbone and side-chain) for both the mutated residue and neighboring residues. Neighboring residues were selected that, prior to mutation, had any atom within a given radius of any atom in the mutated residue. Radial cutoffs of 4Å, 5Å, 6Å, 7Å, and 8Å were tested. At the beginning of each sampling run, the side-chain in the first PDB structure was mutated and then rotamer optimized along with all the neighboring side-chains using an energy-table based Monte Carlo simulated annealing protocol¹⁶. Subsequently, the backrub protocol was run for 10^4 steps at a single temperature of $kT = 0.6$, maintaining either a fixed backbone ($P_{rotamer} = 1$) or allowing backbone sampling ($P_{rotamer} = 0.25$). The lowest energy structure found during ten separate executions was used as the prediction.

To facilitate comparison of the prediction with the second PDB structure in the pair, we first superimposed the N, C α , and C atoms from a set of residues around the mutated residue. The superimposed set was defined as all residues satisfying the following condition in *both* the first and second PDB structures: a heavy atom of the residue must be within 4Å of a heavy atom in the mutated residue. All subsequent RMSD calculations used this fixed superimposition. To compare effects of the mutation on surrounding side-chains, we used a similar set of residues. The set was defined as all non-mutated residues satisfying the following condition in *either* the first or second PDB structures: a non-backbone heavy atom of the residue must be within 4Å of a non-backbone heavy atom in the mutated residue. Any RMSD calculation in which all compared atoms in the second PDB structure had zero occupancy was ignored in calculating overall statistics. All superimposition, RMSD, and chi angle calculations were done using ICM Browser 3.5-11 (Molsoft). Sequence alignments for mapping atom selections from structure to structure were created using ClustalW 1.83⁵⁴.

Code Availability

Source code for the implemented backrub model is available for download free-of-charge as part of the 2.2 release of the Rosetta molecular modeling software at <http://www.rosettacommons.org/>.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors thank Andrew Bordner for providing a text version of the point mutant benchmark. Greg Friedland gave helpful feedback about the backrub sampling code and method. Jerome Nilmeier provided useful discussions about detailed balance. Matt Jacobson gave valuable feedback as co-advisor of CAS. Christopher McClendon, Libusha Kelly, and Ian Davis provided useful comments about the manuscript. CAS was supported by NIH training grant GM067547, the Department of Defense Graduate Research Fellowship Program, and the Genentech Scholars Program. TK is an Alfred P. Sloan Scholar in Molecular Biology. This work was supported by SynBERC (NSF EEC-0540879).

References

1. Ponder JW, Richards FM. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J Mol Biol* 1987;193:775–791. [PubMed: 2441069]
2. Fernández-Recio J, Totrov M, Abagyan R. ICM-DISCO docking by global energy optimization with fully flexible side-chains. *Proteins* 2003;52:113–117. [PubMed: 12784376]
3. Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, Rohl CA, Baker D. Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J Mol Biol* 2003;331:281–299. [PubMed: 12875852]
4. Dahiyat BI, Mayo SL. De novo protein design: fully automated sequence selection. *Science* 1997;278:82–87. [PubMed: 9311930]
5. Dantas G, Kuhlman B, Callender D, Wong M, Baker D. A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins. *J Mol Biol* 2003;332:449–460. [PubMed: 12948494]
6. Kortemme T, Joachimiak LA, Bullock AN, Schuler AD, Stoddard BL, Baker D. Computational redesign of protein-protein interaction specificity. *Nat Struct Mol Biol* 2004;11:371–379. [PubMed: 15034550]
7. Ashworth J, Havranek JJ, Duarte CM, Sussman D, Monnat RJ, Stoddard BL, Baker D. Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature* 2006;441:656–659. [PubMed: 16738662]
8. Swain JF, Gierasch LM. The changing landscape of protein allostery. *Curr Opin Struct Biol* 2006;16:102–108. [PubMed: 16423525]
9. Li Y, Li H, Yang F, Smith-Gill SJ, Mariuzza RA. X-ray snapshots of the maturation of an antibody response to a protein antigen. *Nat Struct Biol* 2003;10:482–488. [PubMed: 12740607]
10. Li Z, Scheraga HA. Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *Proc Natl Acad Sci U S A* 1987;84:6611–6615. [PubMed: 3477791]
11. Abagyan R, Totrov M, Kuznetsov D. ICM - A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *J Comput Chem* 1994;15:488–506.
12. Rohl CA, Strauss CE, Misura KM, Baker D. Protein structure prediction using Rosetta. *Methods Enzymol* 2004;383:66–93. [PubMed: 15063647]
13. Simons KT, Kooperberg C, Huang E, Baker D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J Mol Biol* 1997;268:209–225. [PubMed: 9149153]
14. Rohl CA, Strauss CE, Chivian D, Baker D. Modeling structurally variable regions in homologous proteins with rosetta. *Proteins* 2004;55:656–677. [PubMed: 15103629]
15. Desjarlais JR, Handel TM. Side-chain and backbone flexibility in protein core design. *J Mol Biol* 1999;290:305–318. [PubMed: 10388574]
16. Kuhlman B, Baker D. Native protein sequences are close to optimal for their structures. *Proc Natl Acad Sci U S A* 2000;97:10383–10388. [PubMed: 10984534]
17. Harbury PB, Plecs JJ, Tidor B, Alber T, Kim PS. High-resolution protein design with backbone freedom. *Science* 1998;282:1462–1467. [PubMed: 9822371]
18. Fu X, Apgar JR, Keating AE. Modeling backbone flexibility to achieve sequence diversity: the design of novel alpha-helical ligands for Bcl-xL. *J Mol Biol* 2007;371:1099–1117. [PubMed: 17597151]

19. Wick CD, Siepmann JI. Self-adapting fixed-end-point configurational-bias Monte Carlo method for the regrowth of interior segments of chain molecules with strong intramolecular interactions. *Macromolecules* 2000;33:7207–7218.
20. Canutescu AA, Dunbrack RL. Cyclic coordinate descent: A robotics algorithm for protein loop closure. *Protein Sci* 2003;12:963–972. [PubMed: 12717019]
21. 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;55:351–367. [PubMed: 15048827]
22. Cahill M, Cahill S, Cahill K. Proteins wriggle. *Biophys J* 2002;82:2665–2670. [PubMed: 11964253]
23. Go N, Scheraga HA. Ring Closure and Local Conformational Deformations of Chain Molecules. *Macromolecules* 1970;3:178–187.
24. Brucoleri RE, Karplus M. Chain closure with bond angle variations. *Macromolecules* 1985;18:2767–2773.
25. Dinner AR. Local deformations of polymers with nonplanar rigid main-chain internal coordinates. *J Comput Chem* 2000;21:1132–1144.
26. Ulmschneider JP, Jorgensen WL. Monte Carlo backbone sampling for polypeptides with variable bond angles and dihedral angles using concerted rotations and a Gaussian bias. *J Chem Phys* 2003;118:4261–4271.
27. Coutsias EA, Seok C, Jacobson MP, Dill KA. A kinematic view of loop closure. *J Comput Chem* 2004;25:510–528. [PubMed: 14735570]
28. Davis IW, Arendall WB, Richardson DC, Richardson JS. The backrub motion: how protein backbone shrugs when a sidechain dances. *Structure* 2006;14:265–274. [PubMed: 16472746]
29. Betancourt MR. Efficient Monte Carlo trial moves for polypeptide simulations. *J Chem Phys* 2005;123:174905–174905. [PubMed: 16375567]
30. Meiler J, Baker D. ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility. *Proteins* 2006;65:538–548. [PubMed: 16972285]
31. Bradley P, Misura KM, Baker D. Toward high-resolution de novo structure prediction for small proteins. *Science* 2005;309:1868–1871. [PubMed: 16166519]
32. Holm L, Sander C. Database algorithm for generating protein backbone and side-chain co-ordinates from a C alpha trace application to model building and detection of coordinate errors. *J Mol Biol* 1991;218:183–194. [PubMed: 2002501]
33. Dunbrack RL, Karplus M. Backbone-dependent rotamer library for proteins. Application to side-chain prediction. *J Mol Biol* 1993;230:543–574. [PubMed: 8464064]
34. Shenkin PS, Farid H, Fetrow JS. Prediction and evaluation of side-chain conformations for protein backbone structures. *Proteins* 1996;26:323–352. [PubMed: 8953653]
35. De Maeyer M, Desmet J, Lasters I. All in one: a highly detailed rotamer library improves both accuracy and speed in the modelling of sidechains by dead-end elimination. *Fold Des* 1997;2:53–66. [PubMed: 9080199]
36. Formanek MS, Ma L, Cui Q. Reconciling the "old" and "new" views of protein allostery: a molecular simulation study of chemotaxis Y protein (CheY). *Proteins* 2006;63:846–867. [PubMed: 16475196]
37. Bordner AJ, Abagyan RA. Large-scale prediction of protein geometry and stability changes for arbitrary single point mutations. *Proteins* 2004;57:400–413. [PubMed: 15340927]
38. Lolis E, Petsko GA. Crystallographic analysis of the complex between triosephosphate isomerase and 2-phosphoglycolate at 2.5-A resolution: implications for catalysis. *Biochemistry* 1990;29:6619–6625. [PubMed: 2204418]
39. Lolis E, Alber T, Davenport RC, Rose D, Hartman FC, Petsko GA. Structure of yeast triosephosphate isomerase at 1.9-A resolution. *Biochemistry* 1990;29:6609–6618. [PubMed: 2204417]
40. Joseph D, Petsko GA, Karplus M. Anatomy of a conformational change: hinged "lid" motion of the triosephosphate isomerase loop. *Science* 1990;249:1425–1428. [PubMed: 2402636]
41. Derreumaux P, Schlick T. The loop opening/closing motion of the enzyme triosephosphate isomerase. *Biophys J* 1998;74:72–81. [PubMed: 9449311]
42. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol* 2002;320:369–387. [PubMed: 12079393]

43. Kortemme T, Baker D. A simple physical model for binding energy hot spots in protein-protein complexes. *Proc Natl Acad Sci U S A* 2002;99:14116–14121. [PubMed: 12381794]
44. 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;326:1239–1259. [PubMed: 12589766]
45. Yin S, Ding F, Dokholyan NV. Modeling Backbone Flexibility Improves Protein Stability Estimation. *Structure* 2007;15:1567–1576. [PubMed: 18073107]
46. Dunbrack RL, Cohen FE. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci* 1997;6:1661–1681. [PubMed: 9260279]
47. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J Am Chem Soc* 1995;117:5179–5197.
48. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J Phys Chem B* 1998;102:3586–3616.
49. 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;302:1364–1368. [PubMed: 14631033]
50. Lazaridis T, Karplus M. Effective energy function for proteins in solution. *Proteins* 1999;35:133–152. [PubMed: 10223287]
51. Appleton BA, Zhang Y, Wu P, Yin JP, Hunziker W, Skelton NJ, Sidhu SS, Wiesmann C. Comparative structural analysis of the Erbin PDZ domain and the first PDZ domain of ZO-1. Insights into determinants of PDZ domain specificity. *J Biol Chem* 2006;281:22312–22320. [PubMed: 16737969]
52. Fitzgerald JE, Jha AK, Sosnick TR, Freed KF. Polypeptide motions are dominated by peptide group oscillations resulting from dihedral angle correlations between nearest neighbors. *Biochemistry* 2007;46:669–682. [PubMed: 17223689]
53. Hayward S. Peptide-plane flipping in proteins. *Protein Sci* 2001;10:2219–2227. [PubMed: 11604529]
54. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673–4680. [PubMed: 7984417]
55. Carr DB, Littlefield RJ, Nicholson WL, Littlefield JS. Scatterplot Matrix Techniques for Large N. *Journal of the American Statistical Association* 1987;82:424–436.

Figure 1.

Schematic showing the generalized backrub move. Moves are made by first randomly selecting the polypeptide backbone segment size, typically 2–12 residues, then randomly selecting a starting residue compatible with the selected segment size. The Ca atoms of the starting and ending residues define the rotation axis. All atoms between the two Ca atoms are then rotated about that axis by a random angle up to 11–40 degrees, depending on the segment size. To minimize the bond angle penalty imposed by full atom force fields, precise placement of branching $\text{C}\beta$ and hydrogen atoms is done using quadratic equations that describe the relationship between the backbone bond angle and branching atom spherical coordinates (see Materials and Methods).

Figure 2.

Flow chart depicting the series of substeps taken during a single Monte Carlo step. The proportion of backbone, rotamer, and backbone + rotamer steps are controlled by two parameters. The first, $P_{rotamer}$, specifies the probability of only making a rotamer move. The second, $P_{backbone}$, specifies the probability that only the backbone is modified, given that a rotamer only move type was not selected.

Figure 3.

For 3-residue backrub analysis, the τ_{disp} angle was used as a one-dimensional representation of the backbone conformation. In simulations, it was defined as the $Ca_{i,initial}-Ca_{i-1}-Ca_{i+1}-Ca_{i,current}$ pseudo dihedral angle (red). In this illustration, the starting atomic coordinates are shown in gray. In some high-resolution PDB structures, alternate Ca coordinates were not provided, so the τ_{disp} angle was instead defined as the $C\beta_{i,initial}-Ca_{i-1}-Ca_{i+1}-C\beta_{i,alternate}$ pseudo dihedral angle (not shown) for all PDB analysis.

Figure 4.

Example backbone/ χ_1 populations from a 3-residue backrub simulation of PDB 1PQ7 chain A, residues 62–64, starting from the “B” alternate backbone coordinates. The central side-chain is glutamine. **A)** To monitor coupled backbone and side-chain conformational changes, we recorded both the $C\alpha_{i,\text{initial}}-C\alpha_{i-1}-C\alpha_{i+1}-C\alpha_{i,\text{current}}$ pseudo-dihedral angle (τ_{disp}) and χ_1 angle after every Monte Carlo step. Those angles are shown binned into hexagonal arrays⁵⁵. **B)** We separated the backbone pseudo-dihedral angles by χ_1 angle and generated normalized histograms for bins -60° (red), 60° (green), and 180° (blue, not shown because the overall population was $< 0.05\%$). Circles indicate the population means ($\langle \tau_{\text{disp}} | \chi_1 \rangle$). For each alternate $C\beta$ atom position found in the PDB, the $C\beta_{i,\text{initial}}-C\alpha_{i-1}-C\alpha_{i+1}-C\beta_{i,\text{alternate}}$ dihedral angle (also τ_{disp}) is indicated as a vertical line colored according to the χ_1 bin. The overall population of each bin is indicated in the upper right. The RMSD of the population means from the corresponding PDB τ_{disp} angles is shown in the upper left. Representative structures are shown from the **C)** -60° and **D)** 60° χ_1 bins. The three simulated residues are shown with a ball and stick representation. Other protein residues are shown using a surface representation. The two crystallographic alternate backbone conformations are shown using a wire representation and colored according to the χ_1 bin. Images were created with VMD.

Figure 5.

Backbone angular displacement ($\langle \tau_{disp} | \chi_1 \rangle$) is correlated between PDB structures and predicted populations from 3-residue backrub simulations. Points are colored by χ_1 bin: -60° (red), 60° (green), and 180° (blue). Points from the same simulation are connected by thin lines. Any connecting lines with positive slopes represent simulations which show the correct bias between side chain conformation and backbone conformation. Disconnected points are from simulations where only one of the χ_1 bins seen in the PDB was visited during the simulation. The thick black line shows a least squares linear fit.

Figure 6.

In high-resolution crystal structures, alternate backbone conformations are correlated with the side chain χ_1 angle, with a straightforward structural explanation. **A)** Out of 126 residues in the backrub set, 68 have χ_1 angles in multiple rotameric bins. For those residues, the calculated $C\beta_{i,initial}-C\alpha_{i-1}-C\alpha_{i+1}-C\beta_{i,alternate}$ pseudo-dihedral angles (τ_{disp}) described in Figure 4B were normalized by the average angle (weighted by PDB occupancy). Histograms of those angles are shown using 2.5° bins and colored by χ_1 bin: -60° (red), 60° (green), and 180° (blue). **B)** The clear difference between the $-60^\circ/180^\circ$ and 60° bins has a straightforward structural explanation, where side chains in the 60° bin push the backbone to the left, and the $-60^\circ/180^\circ$ side chains push the backbone to the right. Hypothetical γ atom positions are colored by χ_1 bin.

Figure 7.

Generalized backrub sampling (blue) improves **A**) the RMSD and **B**) χ^1/χ^2 recovery (within 40°) of predicted point mutant side-chains over fixed backbone sampling alone (red).

Prediction results were sorted by increasing starting backbone RMSD and divided into four equally sized groups. The improvement was most distinguished for structure pairs with a larger starting backbone RMSD. Breaks between groups are along the x-axis. Results are shown for 543 non-proline residues for which the solvent accessible surface area of the wild-type residue was $< 5\%$. Residues within a $6\text{\AA}$ radius of the wild-type residue were sampled. In the boxplots, boxes indicate the interquartile range (IQR), thick horizontal lines show the median, and dots show the mean. Whiskers extend to the most extreme datapoint within 1.5 times the IQR of the 25th or 75th percentile.

Figure 8.

Backrub backbone sampling improves side chain prediction, as shown in these examples. The fixed backbone prediction is shown in red and the backrub prediction is shown in blue. The starting PDB structure is shown in green and the target mutated PDB structure is shown in purple. Nitrogen and oxygen atoms are shown in light blue and red, respectively. Examples are sorted by the improvement in mutant residue $\text{Ca/C}\beta$ RMSD from fixed backbone to backrub protocols. The modeled mutation M153F in 1KY0 and 1LOJ is the same, but 1KY0 has a leucine at positions 118 & 121 whereas 1LOJ has a methionine at positions 118 & 121. In both cases, backrub sampling correctly shifts the backbone at residue 153 and better recovers the target side chain. In the last five examples, backrub sampling increases the $\text{Ca/C}\beta$ RMSD but improves the side chain prediction. $\text{Ca/C}\beta$ and side-chain RMSDs are listed in Table 1. Images were created using ICM Browser.

Figure 9.

Generalized backrub sampling captures triosephosphate isomerase loop 6 opening. **A**) For every 200 accepted moves in the simulations, we calculated the C α RMSD of the loop from both the open (1YPI, purple lines) and closed (2YPI, green lines) PDB structures. We defined a single reaction coordinate for the simulations as $\text{RMSD}_{\text{closed}} - \text{RMSD}_{\text{open}}$ (black lines). The green and purple lines are plotted with modified axes such that the black line is the sum of those component lines. The simulation starting in the open conformation is on top and the simulation starting in the closed conformation is on the bottom. The simulation starting from the open conformation makes an initial excursion closer to the closed conformation but then stays open for the remainder. The simulation starting in the closed conformation alternates back and forth between the open and closed conformations at least eight times during the simulation. **B**) The open structure (1YPI) is shown in purple. The closed structure (2YPI) is shown in green. For reference the substrate analogue, 2-phosphoglycolate, is shown using

space fill. (It was not present in either simulation.) The conformation at the numbered Monte Carlo step is shown in black.

Table 1
C α /C β and side-chain RMSD for examples of point mutant predictions pictured in Figure 8. Examples were selected from cases where backrub sampling improved side-chain prediction. The majority of the selected examples also showed improvement in prediction of the backbone, although this was not always the case.

PDB Chain	Mutation	Mutated Residue		Side Chain RMSD		Delta
		Fixed BB	Backrub	Fixed BB	Backrub	
1CV1:A	M111I	1.13	0.21	2.41	0.37	-2.04
1CWU:B	G138A	1.03	0.16	NA	NA	NA
1KGY:H	G1649Q	1.22	0.42	4.54	1.21	-3.33
1LVE:A	Q89L	0.64	0.09	1.14	0.30	-0.84
2HEC:A	A56F	0.70	0.16	3.70	0.98	-2.71
2MEB:A	L56M	0.57	0.08	1.18	0.59	-0.60
1THP:B	P225Y	0.75	0.31	2.24	0.38	-1.86
1LUW:B	Q30V	0.64	0.23	1.01	0.32	-0.69
1GAD:P	N313T	0.43	0.13	2.61	0.27	-2.34
5EAA:A	S191W	0.68	0.40	7.47	0.88	-6.60
1WKD:A	A102D	0.46	0.22	2.74	0.57	-2.17
1KY0:A	M153F	0.54	0.32	3.66	0.49	-3.17
1NAG:A	G43N	0.45	0.24	3.10	0.37	-2.73
2BQJ:A	A125V	0.44	0.24	2.50	0.42	-2.08
1L0J:A	M153F	0.34	0.24	3.61	0.24	-3.37
1G7L:A	S92W	0.34	0.45	3.71	0.81	-2.90
2TOD:D	A69K	0.16	0.32	4.36	1.36	-3.00
1N7X:A	E45Y	0.13	0.31	6.60	0.65	-5.95
1KX0:C	V207I	0.14	0.44	3.29	0.76	-2.53
1L82:A	L153F	0.48	0.93	3.70	0.98	-2.72