CHAPTER FOUR

Flexible Backbone Sampling Methods to Model and Design Protein Alternative Conformations

Noah Ollikainen[*], Colin A. Smith[*,1], James S. Fraser[†,‡,2], Tanja Kortemme[*,‡,§,2]

[*]Graduate Program in Bioinformatics, University of California San Francisco, San Francisco, California, USA
[†]Department of Cellular and Molecular Pharmacology, University of California San Francisco, San Francisco, California, USA
[‡]California Institute for Quantitative Biosciences (QB3), University of California San Francisco, San Francisco, California, USA
[§]Department of Bioengineering and Therapeutic Science, University of California San Francisco, San Francisco, California, USA
[1]Present address: Max Planck Institute for Biophysical Chemistry, Göttingen, Germany
[2]Corresponding authors: e-mail address: james.fraser@ucsf.edu; kortemme@cgl.ucsf.edu

Contents

Abstract

Sampling alternative conformations is key to understanding how proteins work and engineering them for new functions. However, accurately characterizing and modeling protein conformational ensembles remain experimentally and computationally challenging. These challenges must be met before protein conformational heterogeneity can be exploited in protein engineering and design. Here, as a stepping stone, we describe methods to detect alternative conformations in proteins and strategies to model these near-native conformational changes based on backrub-type Monte Carlo moves in

Methods in Enzymology, Volume 523
ISSN 0076-6879
http://dx.doi.org/10.1016/B978-0-12-394292-0.00004-7

Rosetta. We illustrate how Rosetta simulations that apply backrub moves improve modeling of point mutant side-chain conformations, native side-chain conformational heterogeneity, functional conformational changes, tolerated sequence space, protein interaction specificity, and amino acid covariation across protein–protein interfaces. We include relevant Rosetta command lines and RosettaScripts to encourage the application of these types of simulations to other systems. Our work highlights that critical scoring and sampling improvements will be necessary to approximate conformational landscapes. Challenges for the future development of these methods include modeling conformational changes that propagate away from designed mutation sites and modulating backbone flexibility to predictively design functionally important conformational heterogeneity.

1. INTRODUCTION

Proteins are constantly fluctuating between alternative conformations (Frauenfelder, Sligar, & Wolynes, 1991). Processes including folding (Korzhnev, Religa, Banachewicz, Fersht, & Kay, 2010), ligand binding (Boehr, Nussinov, & Wright, 2009), and enzymatic catalytic cycles (Nagel & Klinman, 2009) depend on the movement across the energy landscape. While protein folding is generally driven by a large energy gap between the "native" state and the unfolded ensemble, functionally essential conformations within the "native" state are often separated by smaller energy differences (Fleishman & Baker, 2012).

Computational modeling of the "native" state can result in either a representative single structure or a limited ensemble of conformations. Several straightforward global and local metrics have been developed to compare computational predictions of a representative single structure to an experimentally derived X-ray structure (MacCallum et al., 2011). In contrast, modeling conformational heterogeneity within the "native" state presents significant complications. For example, conformational heterogeneity present in NMR structural ensembles can result from a lack of restraints, limitations in sampling methods, or genuine heterogeneity (Rieping, Habeck, & Nilges, 2005; Schneider, Brunger, & Nilges, 1999). Additionally, comparisons to simulations are often necessary to distinguish between multiple motional models suggested by NMR dynamics observables including residual dipolar couplings (Meiler, Prompers, Peti, Griesinger, & Bruschweiler, 2001), CPMG relaxation dispersion (Bouvignies et al., 2011), and side-chain order parameters (S^2) (Li, Raychaudhuri, & Wand, 1996). X-ray crystallography, which is traditionally interpreted in terms of a single static structure, can also contain information about protein conformational heterogeneity (Best, Lindorff-Larsen, DePristo, & Vendruscolo, 2006; Furnham, Blundell, DePristo, &

Terwilliger, 2006; Lang et al., 2010; Levin, Kondrashov, Wesenberg, & Phillips, 2007). All of these experimental data types can be integrated to improve the computational modeling of protein conformational ensembles.

Ultimately, to connect protein conformational dynamics to function, simulations must be leveraged to provide structural mechanisms consistent with the experimental data. Molecular dynamics simulations present the most obvious solution to identify the structural mechanisms of conformational heterogeneity (Maragakis et al., 2008). However, other than in exceptional cases (Kelley, Vishal, Krafft, & Pande, 2008; Shaw et al., 2010), the timescales accessible to molecular dynamics often preclude sampling functional motions. The computational requirements of molecular dynamics simulations also make it prohibitive to simultaneously sample sequence space for protein design.

Monte Carlo simulations, for example, as used in Rosetta (Leaver-Fay et al., 2011), can also be used to sample protein conformations but rely on having moves that result in energetically accessible conformations. Fixed backbone Monte Carlo simulations, where side-chain conformations are sampled based on a rotamer library, can provide some indications of local flexibility (DuBay & Geissler, 2009). However, it is clear that both side-chain and backbone flexibility are necessary to describe and design protein conformational heterogeneity (Friedland, Linares, Smith, & Kortemme, 2008; Mandell & Kortemme, 2009). Backbone conformations are less easily discretized compared to side-chain rotamers, leading to problems in both creating and validating Monte Carlo backbone moves. Many strategies to efficiently search through backbone space have been implemented in Rosetta including fragment insertion (Simons, Kooperberg, Huang, & Baker, 1997), loop closure with cyclic coordinate decent (Canutescu & Dunbrack, 2003), local torsion sampling with kinematic loop closure (Mandell, Coutsias, & Kortemme, 2009), and backrub (Davis, Arendall, Richardson, & Richardson, 2006; Smith & Kortemme, 2008). Additionally, these moves can be combined and iterated with sequence design to enrich for proteins with desired conformational and functional properties that could not be explored without backbone flexibility.

Here, we describe the application and validation of the backrub sampling move, which was initially inspired by observations using high-resolution X-ray crystallography (Davis et al., 2006), in Rosetta (Smith & Kortemme, 2008; Fig. 4.1A–C). To test whether these moves accurately represent protein conformational heterogeneity, we provide example command lines and scripts that can be run using Rosetta version 3.5. These commands and scripts examine how backrub sampling affects predictions of mutant structures, alternative conformations observed by X-ray crystallography, peptide-ligand binding specificities, and evolutionary properties (Fig. 4.1D). The continued development of

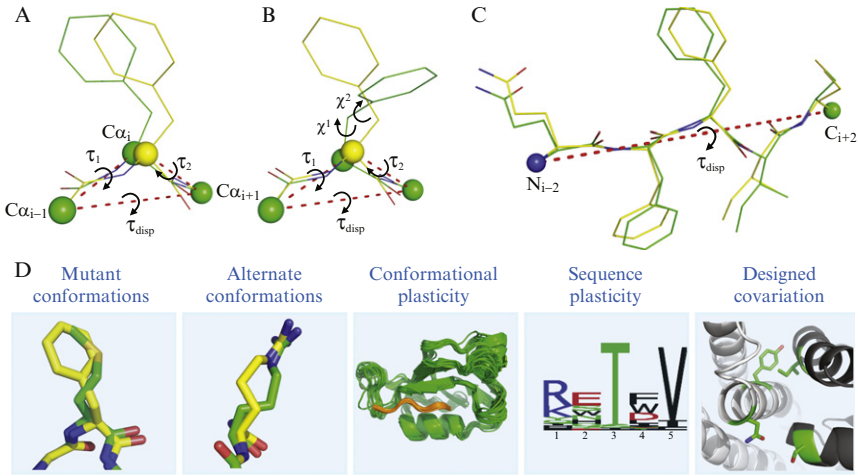


Figure 4.1 The backrub move and its applications in Rosetta. (A) The Richardson group originally described the "backrub" move as a rotation around the $C\alpha_{i-1}$ and $C\alpha_{i+1}$ axis by τ_{disp}, along with simultaneous peptide plane rotations (τ_1 and τ_2), without disturbing other surrounding atom coordinates. (B) By changing the position of the $C\alpha_i$–$C\beta_i$ bond vector, this move can couple side-chain rotameric changes with small local backbone adjustments. (C) In Rosetta, the generalized backrub move is a single rotation that can also include longer intervals and other backbone atom types as pivots for the rotations. (D) Implementing backrubs as a Monte Carlo move in Rosetta enables a variety of flexible backbone prediction and design applications that are described in this chapter: predicting mutant conformations (Fig. 4.2), modeling alternative conformations (Figs. 4.3 and 4.4), coupling conformational and sequence plasticity (Fig. 4.5), and designing amino acid covariation at protein interfaces (Fig. 4.6).

flexible backbone sampling methods that agree with diverse experimental and evolutionary data will improve our ability to design and engineer new protein functions that depend on and exploit conformational heterogeneity.

2. ROSETTA MOVES TO MODEL ALTERNATIVE CONFORMATIONS IN X-RAY DENSITY

2.1. Modeling the Richardson backrub in Rosetta

Unlike other methods for flexible backbone sampling implemented in Rosetta, which are based on fragment insertion or geometric constraints, the backrub move derives its motional model from conformational variation observed in high-resolution X-ray data (Davis et al., 2006). The Richardson group observed electron density consistent with a concerted backbone reorientation that moves a central side-chain perpendicular to the main-chain

direction for 3% of total residues in a dataset of ultra-high-resolution crystal structures. They noted that this move changed the accessible side-chain conformations while leaving flanking structure undisturbed.

In Rosetta, the backrub consists of a rotation about an axis defined by the flanking backbone atoms that changes six internal backbone degrees of freedom in the protein, namely the ϕ, ψ, and the N–Cα–C bond (α) angles at both pivots (Smith & Kortemme, 2008). In the Richardson formulation, the pivots were Cα atoms surrounding a single residue (Fig. 4.1A), but the move can be performed over any backbone atom type over varying length scales (Fig. 4.1C). Bond angle, rotational angle, and Cβ/Hα placement constraints are included to eliminate the need for costly minimization steps after every move. In addition, the backrub move in Rosetta can be adapted so that it obeys detailed balance (Smith & Kortemme, 2008). One notable aspect of the backrub move is that it makes certain side-chain conformations, which would not be accessible in the starting backbone conformation, accessible in the newly accepted backbone conformation (Fig. 4.1B). Such moves alter the potential to accommodate new side-chain conformations and mutations at the "backrubbed" position and its local neighbors.

2.2. Modeling the response to mutations

Subtle backbone adjustments are often necessary to accommodate differences between the wild-type and the mutant side-chain. An initial test of the backrub move in Rosetta was to compare its performance with fixed backbone sampling in predicting the conformation of mutated side-chains (Fig. 4.2). Based on a template of the wild-type structure, a successful prediction of a mutant structure would generate both the conformation of the mutant side-chain and any changes that propagate away from the mutated residue. In general, backrub moves decrease the RMSD between the prediction and conformation observed in the mutant crystal structure (Fig. 4.2). Particularly, dramatic successes are achieved when there would be a clash to neighboring atoms that is relieved by a small backbone adjustment or when a local backbone move changes the probability of accessing a new conformation from a backbone dependent rotamer library.

Due to the broad utility of this application for predicting the results of single or multiple point mutations, we have created a Web server that automates this task: https://kortemmelab.ucsf.edu/backrub/ (Lauck, Smith, Friedland, Humphris, & Kortemme, 2010). On the server, the user must enter the desired PDB, the site of mutation, and the new amino acid identity. By default, 10 independent simulations are performed but this can be

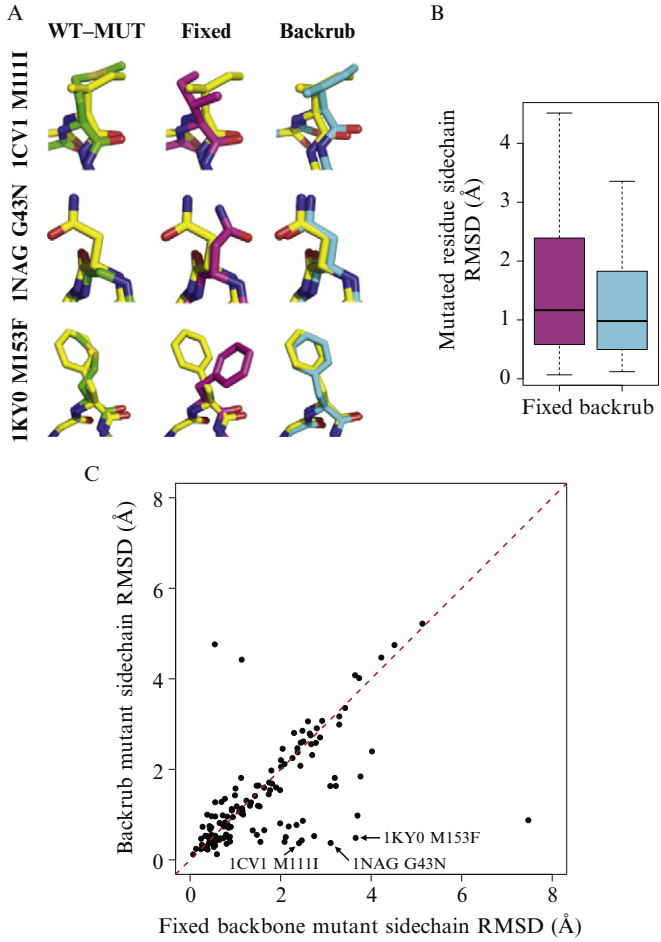


Figure 4.2 Backrub sampling improves the prediction of mutant side-chain conformation compared to fixed backbone simulations. (A) Example predictions of mutant conformations given the wild-type structure, with the indicated PDB codes and point mutations. The mutant crystal structure is shown in yellow compared to the wild-type crystal structure (green, left), prediction based on fixed backbone simulations (magenta, center), or prediction based on backrub flexible backbone sampling (cyan, right). (B) The overall quantification of the results of fixed backbone and backrub predictions over a set of 136 buried (SASA $< 5\%$) side-chains with conformations differing by more than 0.2 Å between mutant and wild type. The median RMSD decreases from 1.17 to 0.98 Å. Shown are box plots with the median as a black line and the 25–75th percentiles in the shaded box with outlier-corrected extreme values as dashed lines. (C) A scatter plot representation of the data in (B) shows that for many mutant structure predictions backrub leads to large improvements compared to fixed backbone simulations.

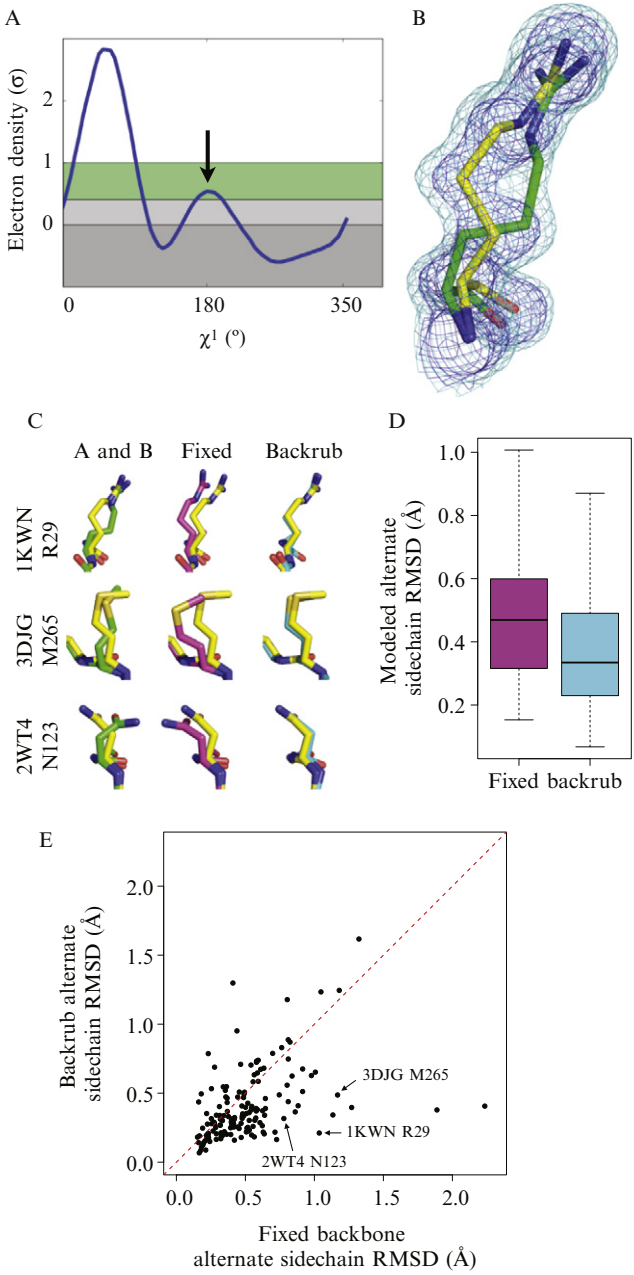
adjusted to 2–50 simulations. For each simulation, the server will attempt 10,000 moves that include backrub rotations of various lengths and angles centered on the mutated residue, and side-chain moves in a 6 Å shell around the mutated residue. The resulting conformations are scored with the Rosetta scoring function and accepted or rejected according to the Metropolis criterion using a kT of 0.6. The lowest scoring conformation out of all the simulations is returned to the user as the best prediction. Advanced users can exert greater control over these parameters by using the "backrub" Rosetta command line program or RosettaScripts (Fleishman et al., 2011). An example command line for point mutation prediction is as follows:

```
~/rosetta/rosetta_source/bin/backrub.linuxgccrelease -database
~/rosetta/rosetta_database/ -s 1CV1.pdb -ex1 -ex2 -extrachi_cutoff
0 -use_input_sc -backrub:ntrials 10000 -nstruct 10 -resfile
1CV1_M111I.resfile -pivot_residues 84 99 102 103 106 107 108 109 110
111 112 113 114 115 118
```

where the "resfile" sets positions to be mutated and repacked (allowing rotamer changes), and the pivot residues denote pivots allowed for backrub moves (necessary files to run the command line with Rosetta version 3.5 are included as example S1 at http://kortemmelab.ucsf.edu/resources/MIE_Supplement.tar.gz; results shown in Fig. 4.2 were obtained with Rosetta revision 18013). For details about command line flags and the resfile syntax, see the Rosetta manual at http://www.rosettacommons.org/.

2.3. Discovering and modeling alternative conformations from X-ray data

The backrub move was inspired by manual examination of ultra-high (sub 1 Å) resolution electron density maps (Davis et al., 2006) suggesting that conformational heterogeneity in X-ray data can be used to develop and validate new sampling methods. Subsequently, Alber and colleagues developed a method, Ringer (Lang et al., 2010), to automate the discovery of alternative side-chain conformations in high (sub 2 Å) resolution electron density maps by sampling around side-chain dihedral angles (Fig. 4.3A). Despite the limitation that Ringer uses a fixed backbone to define the sampling radius, they showed that 18% of side-chains have evidence for unmodeled alternative conformations at electron density levels of 0.3–1σ. Concurrently, van den Bedem, Dhanik, Latombe, and Deacon (2009) developed a complementary method, qFit, which includes local backbone and side-chain flexibility to compute an optimal fit to the electron density for each residue. The


A
Electron density (σ)
2
1
0
0
180
350
χ1 (°)
B
C
A and B
Fixed
Backrub
1KWN R29
3DJG M265
2WT4 N123
D
Modeled alternate sidechain RMSD (Å)
1.0
0.8
0.6
0.4
0.2
Fixed backrub
E
Backrub alternate sidechain RMSD (Å)
2.0
1.5
1.0
0.5
0.0
3DJG M265
1KWN R29
2WT4 N123
0.0
0.5
1.0
1.5
2.0
Fixed backbone alternate sidechain RMSD (Å)

resulting 1–3 backbone and side-chain conformations per residue are merged together in a multiconformer qFit model that improves R/Rfree and maintains excellent geometry statistics. Remarkably, despite an entirely different search procedure from the original observation by the Richardson group, many alternative conformations identified by qFit can be related by backrub-like moves (Fig. 4.3B). Collectively, these studies suggest that electron density maps can provide a more informative representation of the "native" state than traditionally offered by static X-ray structures.

By examining 30 pairs of matched room temperature and cryogenic X-ray datasets, Fraser et al. used Ringer and qFit to show that room temperature X-ray data increase the evidence for alternative conformations compared to data collected at conventional cryogenic temperatures (Fraser et al., 2011). We reasoned that sampling between these experimentally visualized alternative conformations would assess the ability of fixed backbone or backrub simulations to access a representative set of conformations that are significantly populated in the "native" state. To test this idea, we considered the A and B alternative conformations (Fig. 4.3C) from 30 room temperature X-ray multiconformer models refined by qFit.

We focused our analysis on alternative conformations with Cβ deviations of 0.2 Å or greater, relative SASAs less than or equal to 30%, and different χ1 rotameric bins (152 side-chains). First, we split the multiconformer model into two separate PDB files, containing all residues without alternative

Figure 4.3 Backrub sampling improves the prediction of alternative side-chain conformations observed in protein crystal structures. (A) Electron density sampling by Ringer around the χ1 of R29 from PDB 1KWN reveals high electron density for the primary conformation 60° and a secondary peak (indicated by the black arrow), above the 0.3σ threshold that enriches for alternative conformations (shaded green area), near the 180° rotameric bin. (B) 2mFo-DFc electron density surrounding R29 from PDB 1KWN contoured at 1σ (blue mesh) and 0.3σ (cyan mesh). The original PDB model is shown in yellow, with an alternative conformation identified by Ringer and modeled with qFit at 25% occupancy shown in green. (C) Example predictions with Rosetta, with the indicated PDB codes and residues. Sampling of side-chain conformations (yellow) starting from alternative conformations (green, right) is improved by flexible backbone backrub moves (cyan, right) compared to fixed backbone side-chain only sampling (magenta, center). (D) The overall quantification of the results, showing that backrub sampling increases identification of discrete side-chain local minima modeled as alternative conformations by qFit compared to fixed backbone models over a set of 152 side-chains with solvent accessibility less than 30%. The median RMSD decreases from 0.47 to 0.33. Box plots are shown as in Fig. 4.2C. (E) A scatter plot representation of the data in (D) shows that backrub leads to large improvements compared to fixed backbone for many alternative conformation predictions.

conformations and either the "A" conformations or "B" conformations (e.g., 1kwn_A.pdb and 1kwn_B.pdb). Next, we ran a RosettaScripts protocol that moves between the starting conformation specified by the flag –s (in this example, 1kwn_A.pdb) and the target alternative conformation specified by the flag –in:file:native (in this example, 1kwn_B.pdb). At the beginning of the protocol, the χ angles of a central side-chain are switched from the starting conformation to the target conformation. This script tests whether changes in the surrounding side-chains (fixed backbone) or both the backbone and surrounding side-chains (backrub) better accommodate the new χ angles and find a side-chain conformation close in RMSD to the target conformation. We tested this protocol in both directions between the A and B conformations. In the simple example below, the variables χ1 through χ4 provide the target χ angles (here, the χ angles of conformation B), and piv1 through piv3 define the positions around the central residue allowed to be pivots for the backrub move.

The command to run the protocol is

```
~/rosetta/rosetta_source/bin/rosetta_scripts.linuxgccrelease-
database ~/rosetta/rosetta_database -s 1kwn_A.pdb-in:file:
native 1kwn_B.pdb -parser:protocol model_alternate_conform-
ation.xml -parser:script_vars pos=29 chi1=-145.522 chi2=-160.509
chi3=81.9108 chi4=175.816 piv1=28 piv2=29 piv3=30
```

The contents of model_alternate_conformation.xml are

```
<ROSETTASCRIPTS>
   <SCOREFXNS>
      Include the bond angle potential scoring term
      <score12_backrub weights=score12_full>
         <Reweight scoretype=mm_bend weight=1/>
      </score12_backrub>
   </SCOREFXNS>
   <TASKOPERATIONS>
      Define the restrictions on the sidechain moves that will occur in
      the simulation
      <ExtraRotamersGeneric name=extra_rot ex1=1 ex2=2 extrachi_
      cutoff=0/>
      <IncludeCurrent name=input_sc/>
      <DesignAround name=neighbors_only allow_design=0 design_shell
      =0 repack_shell=6.0 resnums=%%pos%%/>
      <RestrictToRepacking name=repack_only/>
      <PreventRepacking name=fix_central_residue resnum=%%pos%%/>
   </TASKOPERATIONS>
```

```
<FILTERS>
    Calculates the side-chain RMSD before and after simulation
    <SidechainRmsd name=rmsd threshold=10 include_backbone=0
    res1_res_num=%%pos%% res2_res_num=%%pos%%/>
</FILTERS>
<MOVERS>
    Set the chi angles of the residue of interest
    <SetChiMover name=setchi1 chinum=1 resnum=%%pos%% angle=%%
    chi1%%/>
    <SetChiMover name=setchi2 chinum=2 resnum=%%pos%% angle=%%
    chi2%%/>
    <SetChiMover name=setchi3 chinum=3 resnum=%%pos%% angle=%%
    chi3%%/>
    <SetChiMover name=setchi4 chinum=4 resnum=%%pos%% angle=%%
    chi4%%/>
    Set backrub moves to only occur near residue of interest
    <Backrub name=backrub pivot_residues=%%piv1%%,%%piv2%%,%%
    piv3%% min_atoms=3 min_atoms=7/>
    Set side-chain moves to only include residues within 6 angstrom shell
    <Sidechain name=sidechain task_operations=extra_rot,input_sc,
    fix_central_residue,
    neighbors_only,repack_only/>
    During Monte Carlo, alternate between backrub moves (75%) and side-
    chain moves (25%)
    <ParsedProtocol name=backrub_protocol mode=single_random>
    <Add mover_name=backrub apply_probability=0.75/>
    <Add mover_name=sidechain apply_probability=0.25/>
    </ParsedProtocol>
    Set up Monte Carlo simulation with 10,000 steps and kT=0.6
    <GenericMonteCarlo name=backrub_mc mover_name=backrub_protocol
    scorefxn_name=score12_backrub trials=10000 temperature=0.6
    preapply=0/>
</MOVERS>
<PROTOCOLS>
    Set the residue of interest to the desired chi angles
    <Add mover_name=setchi1/>
    <Add mover_name=setchi2/>
    <Add mover_name=setchi3/>
    <Add mover_name=setchi4/>
    Calculate RMSD before simulation
```

```
        <Add filter_name=rmsd/>
        Run backrub simulation
        <Add mover_name=backrub_mc/>
        Calculate RMSD after simulation
        <Add filter_name=rmsd/>
    </PROTOCOLS>
</ROSETTASCRIPTS>
```

Necessary files to run this script with Rosetta version 3.5 are included as example S2 at http://kortemmelab.ucsf.edu/resources/MIE_Supplement.tar.gz; results shown in Fig. 4.3 were obtained with Rosetta revision 48648.

Similarly to the mutation data set (Fig. 4.2), backrub moves significantly improve the predictions (Fig. 4.3D and E). Additionally, we tested the effect of including larger backrub moves or using C and N atoms as pivots in place of the normal Cα pivot. Applying these larger moves or using additional pivot atoms did not significantly affect the modeled alternate side-chain RMSD over the dataset. However, there may be certain residue types, secondary structures, or local environments that benefit from distinct move sets.

These results suggest that backrub moves help to model "native" state heterogeneity. While above, we have described the validation of this procedure on high-resolution room temperature X-ray data, similar strategies can be applied to cryogenic data, where models will likely contain fewer alternative conformations, or to low-resolution data, where the electron density maps do not reveal discrete alternative conformations. Therefore, flexible backbone sampling strategies in Rosetta may help to improve the description of the "native" state offered by conventional or low-resolution X-ray crystallography experiments (Tyka et al., 2011). Such sampling methodologies will have many applications including flexible receptor docking in drug discovery (Sherman, Day, Jacobson, Friesner, & Farid, 2006).

2.4. Sampling functional alternative conformations in Cyclophilin A

While the preceding examples utilized backbone flexibility centered on a single residue, many protein motions require movement of a neighborhood of residues that may potentially spread across multiple elements of secondary structure. These movements can create loop, rigid body domain, or side-chain rearrangements that are crucial for the biological mechanism.

The difficulty of discovering and simulating correlated motions is exemplified in the intrinsic conformational exchange of the proline isomerase Cyclophilin A (CypA; Fraser et al., 2009; Fig. 4.4A). Previous NMR studies by the Kern group identified a collective exchange process extending from

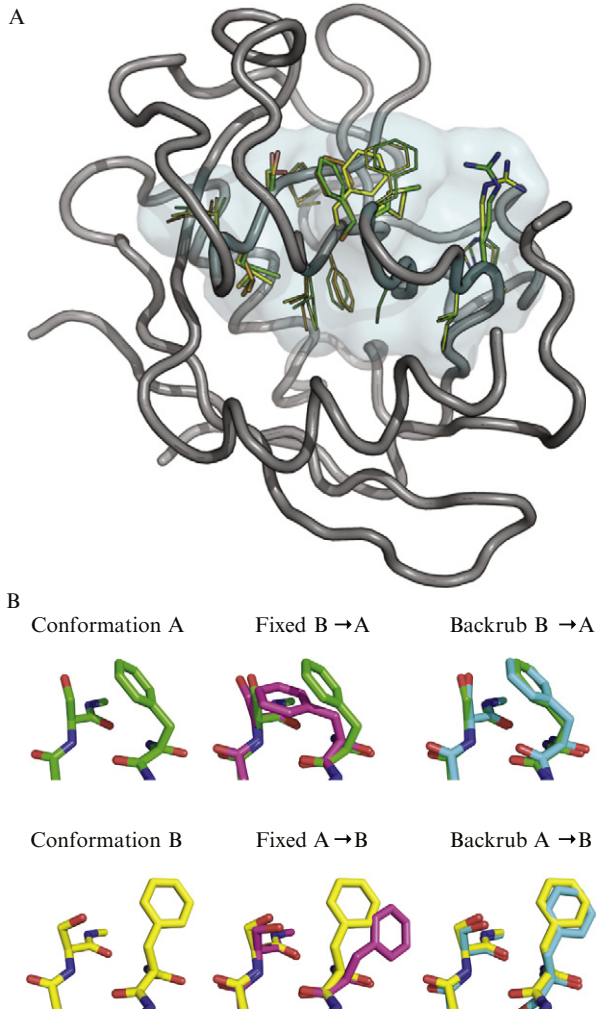


Figure 4.4 Backrub sampling can be used to model functionally relevant alternative conformations. (A) NMR relaxation experiments detect that residues in a dynamic network (cyan transparent surface) undergo a collective exchange between a major and minor conformation with and without substrate present (Eisenmesser, Bosco, Akke, & Kern, 2002; Eisenmesser et al., 2005; Fraser et al., 2009). Room temperature X-ray data collection and qFit multiconformer refinement identify a major (green) and minor (yellow) conformation providing a structural basis for the NMR observations. Additional alternative conformations are shown in orange. (B) Rosetta simulations can access the alternative conformation starting from either state (yellow/green) using backrub (right, cyan) but not fixed backbone (middle, magenta) sampling methods.

the active site into the core of the protein and established a link between the rate of conformational exchange and the catalytic cycle of the enzyme (Eisenmesser et al., 2002, 2005). Room temperature X-ray crystallography and electron density interpretation using a combination of Ringer and manual inspection were used to reveal that the exchange was due to a coupled network of alternative side-chain conformations (Fraser et al., 2009). The functional importance of the alternative conformation was tested by demonstrating a parallel reduction in dynamics and catalysis upon mutation of a residue outside the active site (Fraser et al., 2009). Intriguingly, the backbone movement of Phe113 renders its alternative conformation undetectable by Ringer, which is limited to fixed backbone sampling (Fig. 4.4B).

Due to the millisecond timescale of this correlated motion, the side-chain conformational changes cannot be sampled in conventional molecular dynamics simulations. However, recent accelerated molecular dynamics simulations that reduce torsional barriers have recapitulated several key elements of the conformational dynamics during catalysis (Doshi, McGowan, Ladani, & Hamelberg, 2012). As an initial test of the ability of Rosetta to model correlated motions between neighboring side-chains in CypA, we modified the RosettaScripts protocol used to sample alternative conformations. We defined multiple positions that are allowed to be pivots for backrub and included a call to a "resfile" that specifies the positions whose side-chains can be repacked. This protocol improves sampling of the alternative conformation over fixed backbone approaches (Fig. 4.4B).

The command to run the protocol is

```
~/rosetta/rosetta_source/bin/rosetta_scripts.linuxgccrelease -data-
base ~/rosetta/rosetta_database/ -s 3KON_A.pdb -in:file:native 3KON_B.
pdb -parser:protocol model_alternate_conformation_F113.xml -resfile
F113.resfile -parser:script_vars chi1=-53.763 chi2=-41.4727 chi3=0
chi4=0
```

The contents of model_alternate_conformation_F113.xml are:

```
<ROSETTASCRIPTS>
<SCOREFXNS>
Include the bond angle potential scoring term
<score12_backrub weights=score12_full>
<Reweight scoretype=mm_bend weight=1/>
</score12_backrub>
</SCOREFXNS>
<TASKOPERATIONS>
<ExtraRotamersGeneric name=extra_rot ex1=1 ex2=2 extrachi_
cutoff=0/>
```

```
    <ReadResfile name=read_resfile filename="F113.resfile"/>
      </TASKOPERATIONS>
      <FILTERS>
    Calculates the side-chain RMSD before and after simulation
    <SidechainRmsd name=rmsd threshold=10 include_backbone=0
res1_pdb_num=113A res2_pdb_num=113A/>
    </FILTERS>
    <MOVERS>
    Set the chi angles of the residue of interest
    <SetChiMover name=setchi1 chinum=1 resnum=113A angle=%%chi1%%/>
    <SetChiMover name=setchi2 chinum=2 resnum=113A angle=%%chi2%%/>
    <SetChiMover name=setchi3 chinum=3 resnum=113A angle=%%chi3%%/>
    <SetChiMover name=setchi4 chinum=4 resnum=113A angle=%%chi4%%/>
    Set backrub moves to only occur near residue of interest
    <Backrub name=backrub
pivot_residues=54A,55A,56A,59A,60A,61A,
62A,63A,64A,65A,90A,91A,92A,97A,98A,99A,100A,101A,102A,103A,110A,
111A,112A,113A,114A,115A,116A,118A,119A,120A,121A,122A,123A,125A,
126A,127A,128A,129A,130A min_atoms=3 min_atoms=7/>
    Set side-chain moves to only include residues within 6 angstrom shell
    <Sidechain name=sidechain task_operations=read_resfile,extra_
rot/>
    During Monte Carlo, alternate between backrub moves (75%) and side-
chain moves (25%)
    <ParsedProtocol name=backrub_protocol mode=single_random>
    <Add mover_name=backrub apply_probability=0.75/>
    <Add mover_name=sidechain apply_probability=0.25/>
    </ParsedProtocol>
    Set up Monte Carlo simulation with 10,000 steps and kT=0.6
    <GenericMonteCarlo name=backrub_mc mover_name=backrub_protocol
scorefxn_name=score12_backrub trials=10000 temperature=0.6 preapply=0/>
    </MOVERS>
    <PROTOCOLS>
    Set the residue of interest to the desired chi angles
    <Add mover_name=setchi1/>
    <Add mover_name=setchi2/>
    <Add mover_name=setchi3/>
    <Add mover_name=setchi4/>
    Calculate RMSD before simulation
    <Add filter_name=rmsd/>
```

```
Run backrub simulation
<Add mover_name=backrub_mc/>
Calculate RMSD after simulation
<Add filter_name=rmsd/>
</PROTOCOLS>
</ROSETTASCRIPTS>
```

Necessary files to run this script with Rosetta version 3.5 are included in the example S3 at http://kortemmelab.ucsf.edu/resources/MIE_Supplement.tar.gz; results shown in Fig. 4.4 were obtained with Rosetta revision 48648.

Here, we have specified neighboring residues that can undergo backrub moves. Similarly, the ability of backrub to sample functionally important loop conformations has been demonstrated for triosephosphate isomerase (TIM; Smith & Kortemme, 2008). To efficiently sample these enzymatic motions, we used prior knowledge of residues that need conformational adjustments. Therefore, these strategies present an immediate challenge: sampling large correlated motions without prior knowledge of what residues are involved in the motion. One approach is to use unbiased simulations to identify flexible regions. In the case of TIM, unbiased simulations identify that the catalytically important loop region is highly flexible, but only simulations that focus on the loop have been shown to sample the entire range of motion. Another intermediate on the road to this goal is to include constraints from NMR relaxation dispersion experiments, which specify residues that are experiencing an exchange in chemical environment but do not provide direct structural information about the exchange. Recent work using T4 Lysozyme (Bouvignies et al., 2011) suggests that Rosetta fragment insertion methods biased by experimental chemical shifts can generate structural descriptions of alternative conformations discovered by NMR. The success of backrub moves in sampling the enzyme motions of CypA and TIM indicate that "native" state sampling using backrub moves can likely be exploited in a similar fashion to link conformational dynamics discovered by NMR relaxation dispersion experiments with structural mechanisms.

3. SEQUENCE PLASTICITY AND CONFORMATIONAL PLASTICITY ARE INTERTWINED

The improvements offered by backrub moves in predicting point mutant structures (Fig. 4.2) suggest that subtle backbone rearrangements can significantly alter the prediction of tolerated mutations. It follows that conformational ensembles created through backrub moves would also

change the potential for sequences predicted to be consistent with a given protein fold. Indeed, incorporating backbone flexibility increased the overlap between sequences predicted to be consistent with the ubiquitin fold and the evolutionary record (Friedland, Lakomek, Griesinger, Meiler, & Kortemme, 2009). These results provided further evidence that the relationship between sequence and structural variability can be leveraged to develop and validate new conformational sampling methods. Both sequence alignments of orthologous proteins (natural selection) and sequences enriched in high-throughput binding experiments, such as phage display or peptide arrays (artificial selection), can be used to define the sequence variability that design methods can target.

3.1. Modeling peptide binding specificity

In addition to predicting sequences tolerated by a single protein fold, flexible backbone methods can improve the prediction of binding specificity in peptide binding domains such as PDZ, SH3, and WW domains. Phage display coupled with next-generation sequencing techniques can generate experimental position weight matrices (PWMs) based on large numbers of potential sequences (Huang & Sidhu, 2011). A challenge for interpreting these datasets is to define the structural basis for specificity in binding pockets that are quite similar. To test how well Rosetta can recapitulate the binding specificities discovered by these experiments, we sampled conformations of both the peptide and receptor protein using backrub moves. As observed previously for sampling the ubiquitin fold family sequences, the temperature parameter is key for controlling the conformational diversity sampled by the ensemble (Fig. 4.5A). For applications where there is a larger degree of backbone flexibility and corresponding sequence variability, higher temperatures can be explored. For PDZ domain–peptide interactions, a temperature of 0.6 kT was used (Smith & Kortemme, 2010, 2011). After generating an ensemble using backrub moves, design can be used to sample sequence changes of either the receptor or the peptide.

We have automated the sequence tolerance protocol for using flexible backbone ensembles and sequence design for predicting peptide binding specificity on a Web server: https://kortemmelab.ucsf.edu/backrub/ (Lauck et al., 2010). Users can also download a "protocol capture" of the sequence tolerance method, complete with example input/output and scripts, in the Supplementary Materials accompanying, http://www.elsevierdirect.com/companions/9780123942920 (Smith & Kortemme, 2011). Given a peptide-bound structure, backrub sampling methods are used

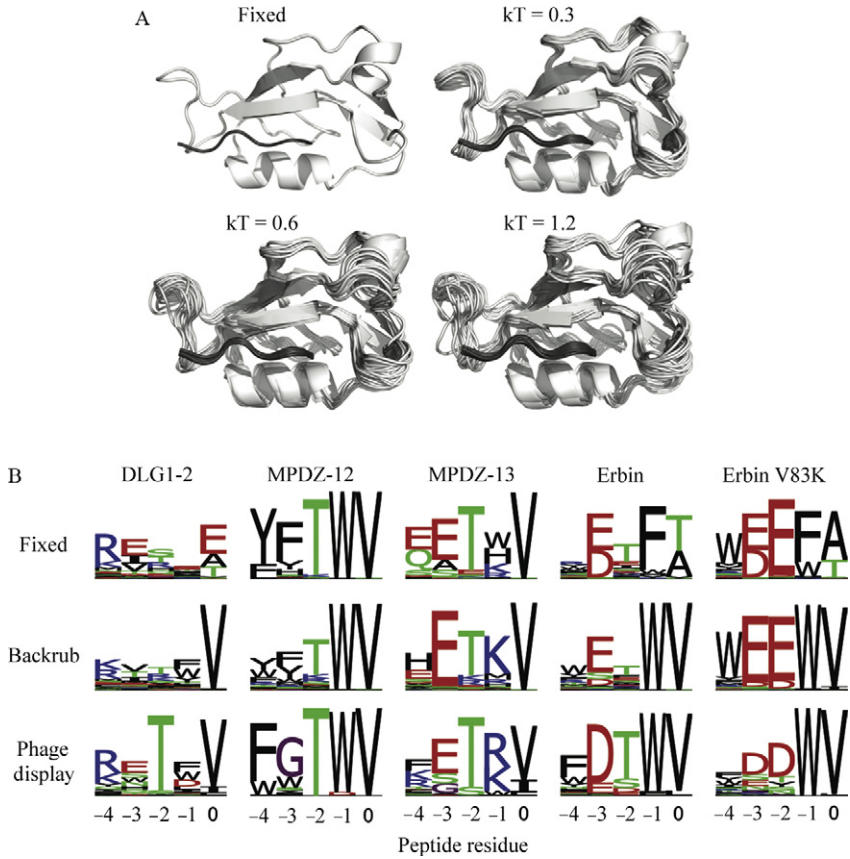


Figure 4.5 Rosetta generates near-native ensembles using backrub sampling. (A) Cα cartoons of Rosetta generated conformational ensembles using backrub sampling at different temperatures, compared to the fixed backbone (top left). Higher temperatures increase the conformational diversity and can increase agreement with experimental data. The PDZ domain structure is shown in white and peptide in gray. (B) Example results from the sequence tolerance protocol to predict peptide specificity for four PDZ domains (DLG1-2, MPDZ-12, MPDZ-13, and Erbin) and one PDZ domain point mutant (Erbin V83K); peptide positions are indicated using the standard nomenclature for PDZ domain motifs, with 0 denoting the C-terminal residue, followed by −1, −2, etc. Without backbone flexibility, Rosetta fails to predict important residue preferences observed in experimental phage display selections, such as valine at the 0 position or tryptophan at the −1 position for DLG1-2 and Erbin.

to generate an ensemble of conformations. For each conformation, a genetic algorithm is used to design sequences for high-affinity binding. In this protocol, the interface energy is given greater weight (Smith & Kortemme, 2010, 2011). To compare these predictions to experimental data, we generate a sequence logo based on the positional frequencies in the resulting designed sequences (Fig. 4.5B). Compared to fixed backbone methods, backrub sampling increases the agreement at several positions. Given the adaptable nature of many protein–protein interfaces, it is clear that flexible backbone methods will provide great insight into the structural and energetic basis for binding specificity. Additionally, as more datasets on mutant binding domains are collected, there is potential to look for covariation between the sequences tolerated between receptor and peptide positions (Ernst et al., 2010).

3.2. Covariation and interface design in two-component signaling

Testing computational protein design methods based on comparison with experimental PWMs is informative. However, it involves evaluating amino acid positions independently from each other and therefore may overlook some of the intricate details of pair-wise interactions between designed residues. In order to assess how well flexible backbone design protocols capture dependencies between designed residues, we directly compared designed residue covariation to native residue covariation. We chose to examine covariation within the bacterial two-component signaling system, since it has previously been shown that sensor histidine kinases (HKs) and their cognate response regulators (RRs) exhibit significant intermolecular covariation at their protein–protein interface (White, Szurmant, Hoch, & Hwa, 2007).

Designed sequences were obtained by generating a backrub conformational ensemble of 500 structures starting from the cocrystal structure of HK853 and RR468 from *Thermotoga maritima* (Casino, Rubio, & Marina, 2009; PDB 3DGE). Since bacterial HK and RR sequences are highly divergent, we used a temperature of 1.2 kT to produce a conformational ensemble that would yield sufficiently diverse designed sequences. We then performed sequence design using Monte Carlo simulated annealing on each structure, which resulted in 500 designed HK and RR sequences.
The command lines for this protocol are as follows:

Backrub ensemble generation

```
~/rosetta/rosetta_source/bin/backrub.linuxgccrelease -database ~/
rosetta/rosetta_database/ -s 3DGE.pdb -resfile NATAA.res -ex1 -ex2 -
extrachi_cutoff 0 -backrub:mc_kt 1.2 -backrub:ntrials 10000 -nstruct
500 -backrub:initial_pack
```

Sequence design

```
~/rosetta/rosetta_source/bin/fixbb.linuxgccrelease -database ~/
rosetta/rosetta_database/ -s 3DGE_0001_last.pdb -resfile ALLAA.res
-ex1 -ex2 -extrachi_cutoff 0 -nstruct 1 -overwrite -linmem_ig 10 -
no_his_his_pairE -minimize_sidechains
```

Necessary files to run these command lines with Rosetta version 3.5 are included as example S4 at http://kortemmelab.ucsf.edu/resources/MIE_Supplement.tar.gz; results shown in Fig. 4.6 were obtained with Rosetta revision 39284.

To compare the sequence features from interface design with those observed in naturally interacting proteins, we collected alignments of natural HK and RR sequences from Pfam (PF000512 for HK and PF00072 for RR) and concatenated all pairs of HKs and RRs that were adjacent in a particular genome (i.e., pairs with GI numbers differing by 1). To avoid bias from closely related sequences, we filtered the joint HK/RR alignment for redundancy using an 80% sequence identity cutoff. We quantified residue covariation of all intermolecular pairs of amino acid positions in designed and natural sequences using a mutual-information-based statistic (Dickson, Wahl, Fernandes, & Gloor, 2010).

We observed significant overlap between the designed and natural highly covarying intermolecular pairs within the HK/RR complex (Fig. 4.6A). Mapping the residue pairs that were highly covarying in both designed and natural sequences onto the structure of the complex revealed that all of these pairs are localized to the HK/RR interface (Fig. 4.6B). A closer examination of these pairs shows that each pair forms a physical interaction across the HK/RR interface (Fig. 4.6C), suggesting that these pairs may be important for determining specificity in bacterial two-component signaling systems. Indeed, several of these positions have previously been mutated to alter the specificity of HK–RR interactions: HK-Thr in Pair 1, HK-Tyr in Pair 9, and HK-Val in Pair 11 (Skerker et al., 2008). The remaining pairs, including those that highly covary in designed sequences but not natural sequences, represent potential opportunities for rewiring two-component signaling specificity using computational protein design.

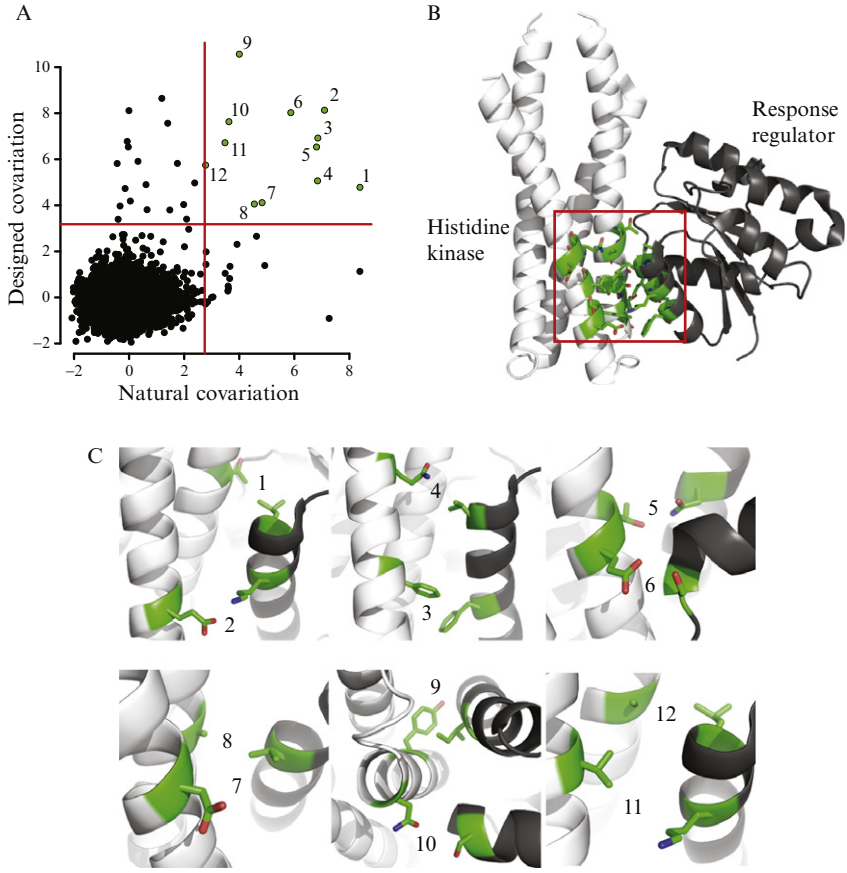


Figure 4.6 Rosetta backrub design methods capture features of evolutionary amino acid covariation. (A) Comparison between designed and natural intermolecular amino acid covariation for histidine kinases (HKs) and their cognate response regulators (RRs). Each point represents a pair of amino acid positions. Natural covariation was quantified using a mutual-information-based metric for all pairs of positions in a multiple sequence alignment of HKs concatenated to their cognate RRs. A backrub ensemble of 500 structures was generated for a HK/RR complex (PDB ID 3DGE) and RosettaDesign was used to predict one low-energy sequence for each structure in the ensemble. Designed covariation was quantified for all pairs of positions in the resulting multiple sequence alignment of 500 sequences. The red lines indicate the threshold cutoff for the top 30 designed covarying intermolecular pairs (horizontal) and the top 30 natural covarying pairs (vertical). The 12 intermolecular pairs of positions that are highly covarying in both designed and natural sequences are highlighted in green. (B) The structure of a HK/RR complex with amino acids that are involved in highly covarying intermolecular pairs in both natural and designed sequences are shown in green and stick representation. (C) Close-up of the 12 intermolecular covarying pairs. Each of these 12 pairs of amino acids forms a physical interaction across the interface of the complex.

4. FUTURE CHALLENGES

The success of flexible backbone sampling methods in predicting mutant side-chain (Fig. 4.2) and alternative (Fig. 4.3) conformations indicates the broad utility of these methods in designing sequences compatible with a target "native" structure. Previous studies have used Rosetta to provide structural mechanisms for NMR measures of protein dynamics (Friedland et al., 2008, 2009) and to design mutations that stabilize specific conformations from a dynamic ensemble (Babor & Kortemme, 2009; Bouvignies et al., 2011). Additionally, the comparisons to naturally and artificially selected sequence data suggest that flexible backbone methods can be leveraged to design libraries for generating proteins with new or improved functions (Friedland & Kortemme, 2010).

Despite these successes, exploiting backbone flexibility to design conformational heterogeneity, in contrast to design of a single target structure, remains largely unaddressed. A major challenge in the coming years will be to adapt these methods to design functionally important protein conformational dynamics. Examples of these design challenges include: designing loops to sample multiple conformations that exclude water and permit substrate flux during an enzymatic catalytic cycle, creating peptide binding domains where specificity is encoded by distinct binding modes, or generating coupled networks of side-chain conformations that respond to an allosteric binding event.

To meet these lofty challenges, scoring functions must be sensitive to the small gaps that separate these conformations on the energy landscape (Fleishman et al., 2011). In addition, to avoid having populations biased by the sampling algorithm and provide better estimates of conformational entropy, the Monte Carlo move sets must obey detailed balance (Hastings, 1970). In addition to improvements in scoring and thermodynamics, more sophisticated sampling protocols will likely be needed. Here, we have primarily focused on backrub moves around Cα. However, sampling the "native" state of some protein environments may benefit from different strategies or iterations through a combination of sampling moves. Indeed, we have recently had success at modeling conformational changes that propagate away from a designed mutation by iteratively switching between different sampling and scoring strategies during the course of a single simulation (Kapp et al., 2012). Learning from the successes and failures of these new strategies will be essential to improve both protein design and our understanding of the relationship between protein conformational dynamics and function.

ACKNOWLEDGMENTS

The Fraser lab is supported by the National Institutes of Health Early Independence Award (DP5 OD009180) and QB3. Rosetta development in the Kortemme lab is supported by awards from the National Science Foundation (NSF) to T. K. (NSF CAREER MCB-0744541; NSF EF-0849400) and the Synthetic Biology Engineering Research Center (NSF EEC-0540879; PI Keasling). N. O. was additionally supported by an NSF graduate fellowship. We thank Russell Goodman and Henry van den Bedem for helpful discussions.

REFERENCES

Babor, M., & Kortemme, T. (2009). Multi-constraint computational design suggests that native sequences of germline antibody H3 loops are nearly optimal for conformational flexibility. *Proteins*, *75*, 846–858.

Best, R. B., Lindorff-Larsen, K., DePristo, M. A., & Vendruscolo, M. (2006). Relation between native ensembles and experimental structures of proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 10901–10906.

Boehr, D. D., Nussinov, R., & Wright, P. E. (2009). The role of dynamic conformational ensembles in biomolecular recognition. *Nature Chemical Biology*, *5*, 789–796.

Bouvignies, G., Vallurupalli, P., Hansen, D. F., Correia, B. E., Lange, O., Bah, A., et al. (2011). Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature*, *477*, 111–114.

Canutescu, A. A., & Dunbrack, R. L., Jr. (2003). Cyclic coordinate descent: A robotics algorithm for protein loop closure. *Protein Science*, *12*, 963–972.

Casino, P., Rubio, V., & Marina, A. (2009). Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction. *Cell*, *139*, 325–336.

Davis, I. W., Arendall, W. B., 3rd., Richardson, D. C., & Richardson, J. S. (2006). The backrub motion: How protein backbone shrugs when a sidechain dances. *Structure*, *14*, 265–274.

Dickson, R. J., Wahl, L. M., Fernandes, A. D., & Gloor, G. B. (2010). Identifying and seeing beyond multiple sequence alignment errors using intra-molecular protein covariation. *PloS One*, *5*, e11082.

Doshi, U., McGowan, L. C., Ladani, S. T., & Hamelberg, D. (2012). Resolving the complex role of enzyme conformational dynamics in catalytic function. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 5699–5704.

DuBay, K. H., & Geissler, P. L. (2009). Calculation of proteins' total side-chain torsional entropy and its influence on protein-ligand interactions. *Journal of Molecular Biology*, *391*, 484–497.

Eisenmesser, E. Z., Bosco, D. A., Akke, M., & Kern, D. (2002). Enzyme dynamics during catalysis. *Science*, *295*, 1520–1523.

Eisenmesser, E. Z., Millet, O., Labeikovsky, W., Korzhnev, D. M., Wolf-Watz, M., Bosco, D. A., et al. (2005). Intrinsic dynamics of an enzyme underlies catalysis. *Nature*, *438*, 117–121.

Ernst, A., Gfeller, D., Kan, Z., Seshagiri, S., Kim, P. M., Bader, G. D., et al. (2010). Coevolution of PDZ domain-ligand interactions analyzed by high-throughput phage display and deep sequencing. *Molecular BioSystems*, *6*, 1782–1790.

Fleishman, S. J., & Baker, D. (2012). Role of the biomolecular energy gap in protein design, structure, and evolution. *Cell*, *149*, 262–273.

Fleishman, S. J., Leaver-Fay, A., Corn, J. E., Strauch, E. M., Khare, S. D., Koga, N., et al. (2011). RosettaScripts: A scripting language interface to the Rosetta macromolecular modeling suite. *PloS One*, *6*, e20161.

Fraser, J. S., Clarkson, M. W., Degnan, S. C., Erion, R., Kern, D., & Alber, T. (2009). Hidden alternative structures of proline isomerase essential for catalysis. *Nature*, *462*, 669–673.

Fraser, J. S., van den Bedem, H., Samelson, A. J., Lang, P. T., Holton, J. M., Echols, N., et al. (2011). Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 16247–16252.

Frauenfelder, H., Sligar, S. G., & Wolynes, P. G. (1991). The energy landscapes and motions of proteins. *Science*, *254*, 1598–1603.

Friedland, G. D., & Kortemme, T. (2010). Designing ensembles in conformational and sequence space to characterize and engineer proteins. *Current Opinion in Structural Biology*, *20*, 377–384.

Friedland, G. D., Lakomek, N. A., Griesinger, C., Meiler, J., & Kortemme, T. (2009). A correspondence between solution-state dynamics of an individual protein and the sequence and conformational diversity of its family. *PLoS Computational Biology*, *5*, e1000393.

Friedland, G. D., Linares, A. J., Smith, C. A., & Kortemme, T. (2008). A simple model of backbone flexibility improves modeling of side-chain conformational variability. *Journal of Molecular Biology*, *380*, 757–774.

Furnham, N., Blundell, T. L., DePristo, M. A., & Terwilliger, T. C. (2006). Is one solution good enough? *Nature Structural & Molecular Biology*, *13*, 184–185. Discussion 185.

Hastings, W. K. (1970). Monte-Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*, 97–109.

Huang, H., & Sidhu, S. S. (2011). Studying binding specificities of peptide recognition modules by high-throughput phage display selections. *Methods in Molecular Biology*, *781*, 87–97.

Kapp, G. T., Liu, S., Stein, A., Wong, D. T., Remenyi, A., Yeh, B. J., et al. (2012). Control of protein signaling using a computationally designed GTPase/GEF orthogonal pair. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 5277–5282.

Kelley, N. W., Vishal, V., Krafft, G. A., & Pande, V. S. (2008). Simulating oligomerization at experimental concentrations and long timescales: A Markov state model approach. *The Journal of Chemical Physics*, *129*, 214707.

Korzhnev, D. M., Religa, T. L., Banachewicz, W., Fersht, A. R., & Kay, L. E. (2010). A transient and low-populated protein-folding intermediate at atomic resolution. *Science*, *329*, 1312–1316.

Lang, P. T., Ng, H. L., Fraser, J. S., Corn, J. E., Echols, N., Sales, M., et al. (2010). Automated electron-density sampling reveals widespread conformational polymorphism in proteins. *Protein Science*, *19*, 1420–1431.

Lauck, F., Smith, C. A., Friedland, G. F., Humphris, E. L., & Kortemme, T. (2010). RosettaBackrub—A web server for flexible backbone protein structure modeling and design. *Nucleic Acids Research*, *38*, W569–W575.

Leaver-Fay, A., Tyka, M., Lewis, S. M., Lange, O. F., Thompson, J., Jacak, R., et al. (2011). ROSETTA3: An object-oriented software suite for the simulation and design of macromolecules. *Methods in Enzymology*, *487*, 545–574.

Levin, E. J., Kondrashov, D. A., Wesenberg, G. E., & Phillips, G. N., Jr. (2007). Ensemble refinement of protein crystal structures: Validation and application. *Structure*, *15*, 1040–1052.

Li, Z., Raychaudhuri, S., & Wand, A. J. (1996). Insights into the local residual entropy of proteins provided by NMR relaxation. *Protein Science*, *5*, 2647–2650.

MacCallum, J. L., Perez, A., Schnieders, M. J., Hua, L., Jacobson, M. P., & Dill, K. A. (2011). Assessment of protein structure refinement in CASP9. *Proteins*, *79*(Suppl. 10), 74–90.

Mandell, D. J., Coutsias, E. A., & Kortemme, T. (2009). Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nature Methods*, *6*, 551–552.

Mandell, D. J., & Kortemme, T. (2009). Backbone flexibility in computational protein design. *Current Opinion in Biotechnology*, *20*, 420–428.

Maragakis, P., Lindorff-Larsen, K., Eastwood, M. P., Dror, R. O., Klepeis, J. L., Arkin, I. T., et al. (2008). Microsecond molecular dynamics simulation shows effect of slow loop dynamics on backbone amide order parameters of proteins. *The Journal of Physical Chemistry. B*, *112*, 6155–6158.

Meiler, J., Prompers, J. J., Peti, W., Griesinger, C., & Bruschweiler, R. (2001). Model-free approach to the dynamic interpretation of residual dipolar couplings in globular proteins. *Journal of the American Chemical Society*, *123*, 6098–6107.

Nagel, Z. D., & Klinman, J. P. (2009). A 21st century revisionist's view at a turning point in enzymology. *Nature Chemical Biology*, *5*, 543–550.

Rieping, W., Habeck, M., & Nilges, M. (2005). Inferential structure determination. *Science*, *309*, 303–306.

Schneider, T. R., Brunger, A. T., & Nilges, M. (1999). Influence of internal dynamics on accuracy of protein NMR structures: Derivation of realistic model distance data from a long molecular dynamics trajectory. *Journal of Molecular Biology*, *285*, 727–740.

Shaw, D. E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R. O., Eastwood, M. P., et al. (2010). Atomic-level characterization of the structural dynamics of proteins. *Science*, *330*, 341–346.

Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A., & Farid, R. (2006). Novel procedure for modeling ligand/receptor induced fit effects. *Journal of Medicinal Chemistry*, *49*, 534–553.

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

Skerker, J. M., Perchuk, B. S., Siryaporn, A., Lubin, E. A., Ashenberg, O., Goulian, M., et al. (2008). Rewiring the specificity of two-component signal transduction systems. *Cell*, *133*, 1043–1054.

Smith, C. A., & Kortemme, T. (2008). Backrub-like backbone simulation recapitulates natural protein conformational variability and improves mutant side-chain prediction. *Journal of Molecular Biology*, *380*, 742–756.

Smith, C. A., & Kortemme, T. (2010). Structure-based prediction of the peptide sequence space recognized by natural and synthetic PDZ domains. *Journal of Molecular Biology*, *402*, 460–474.

Smith, C. A., & Kortemme, T. (2011). Predicting the tolerated sequences for proteins and protein interfaces using RosettaBackrub flexible backbone design. *PloS One*, *6*, e20451.

Tyka, M. D., Keedy, D. A., Andre, I., Dimaio, F., Song, Y., Richardson, D. C., et al. (2011). Alternate states of proteins revealed by detailed energy landscape mapping. *Journal of Molecular Biology*, *405*, 607–618.

van den Bedem, H., Dhanik, A., Latombe, J. C., & Deacon, A. M. (2009). Modeling discrete heterogeneity in X-ray diffraction data by fitting multi-conformers. *Acta Crystallographica Section D: Biological Crystallography*, *65*, 1107–1117.

White, R. A., Szurmant, H., Hoch, J. A., & Hwa, T. (2007). Features of protein-protein interactions in two-component signaling deduced from genomic libraries. *Methods in Enzymology*, *422*, 75–101.