Accuracy of Side-Chain Prediction Upon Near-Native Protein Backbones Generated by Ab Initio Folding Methods

Enoch S. Huang,[1] Patrice Koehl,[2] Michael Levitt,[2] Rohit V. Pappu,[1] and Jay W. Ponder[1]*
[1]*Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri*
[2]*Department of Structural Biology, Stanford University School of Medicine, Stanford, California*

ABSTRACT The ab initio folding problem can be divided into two sequential tasks of approximately equal computational complexity: the generation of native-like backbone folds and the positioning of side chains upon these backbones. The prediction of side-chain conformation in this context is challenging, because at best only the near-native global fold of the protein is known. To test the effect of displacements in the protein backbones on side-chain prediction for folds generated ab initio, sets of near-native backbones ($\leq$ 4 Å Cα RMS error) for four small proteins were generated by two methods. The steric environment surrounding each residue was probed by placing the side chains in the native conformation on each of these decoys, followed by torsion-space optimization to remove steric clashes on a rigid backbone. We observe that on average 40% of the $\chi1$ angles were displaced by 40° or more, effectively setting the limits in accuracy for side-chain modeling under these conditions. Three different algorithms were subsequently used for prediction of side-chain conformation. The average prediction accuracy for the three methods was remarkably similar: 49% to 51% of the $\chi1$ angles were predicted correctly overall (33% to 36% of the $\chi1+2$ angles). Interestingly, when the inter-side-chain interactions were disregarded, the mean accuracy increased. A consensus approach is described, in which side-chain conformations are defined based on the most frequently predicted χ angles for a given method upon each set of near-native backbones. We find that consensus modeling, which de facto includes backbone flexibility, improves side-chain prediction: $\chi1$ accuracy improved to 51–54% (36–42% of $\chi1+2$). Implications of a consensus method for ab initio protein structure prediction are discussed. Proteins 33:204–217, 1998. © 1998 Wiley-Liss, Inc.

Key words: rotamer libraries; energy minimization; self consistent mean-field theory; torsion space; modeling

INTRODUCTION

The ab initio protein folding problem, when solved in its entirety, considers as input only the sequence information, and provides as output an all-atom model of the protein. Because an average-sized protein comprises thousands of atoms, the conformational space one must search to find its native conformation is astronomically large, making this task computationally impossible.[1] As one approach to circumvent this problem, investigators have divided the protein folding problem into two separate and more tractable tasks of roughly equal complexity,[2] namely defining the backbone conformation and positioning the side chains on this backbone. A stepwise solution is desirable since the combinatorial complexity of the entire problem is merely additive for the two steps, rather than multiplicative. For the first problem, i.e. definition of the overall fold of the protein, so-called ab initio folding methods normally start with a random or extended chain conformation and search for a native-like backbone trace. The search space associated with the backbone is reduced by constraining atoms to a lattice[3–8] or through discretization in torsion space.[9–19] Furthermore, the degrees of freedom associated with a side chain are often neglected by defining a "center of interaction" positioned at the Cα atom or at a virtual centroid location.[20] When side chains are used explicitly in the folding simulation, they mainly serve to drive residue-specific or chemical group specific energy functions, in contrast to the all-atom force fields used in molecular simulation.[21,22,23] Side-chains aid ab initio folding methods by enforcing excluded volume constraints on the backbone, but their final conformations are not considered explicitly, and only the accuracy of the

backbone with respect to that of the target structure is measured. Homology modeling is another approach to the problem of generating the backbone of a protein. It entails that a related structure be available as a starting point for the all-atom modeling of the target. In that case, the backbone of the homologue defines the fold of the unknown protein, with possible gaps that are usually filled using a library of fragments.[24,25] Upon completion of this first task in the folding problem, one now has a backbone upon which the side chains may seek an energetically favorable (ideally the native) combination of conformations. This modeling of side-chain conformations corresponds to the second task in the folding problem. If the generated backbone is sufficiently close to the native structure, then side-chain placement algorithms (recently reviewed in References 2 and 26) will presumably position the side-chain atoms accurately. These search procedures are often made more efficient by the use of side-chain rotamer libraries.[27,28,29]

Current progress in ab initio folding algorithms has allowed the generation of near native folds. Levitt and co-workers[30,31] describe a discrete-state model restricted to four points in Ramachandran space that can reproduce the native backbone to 2–4 Å root-mean-square (RMS) accuracy, even if these near-native structures cannot be identified with the exclusion of all the non-native folds.[31,32] Others have used simplified energy functions to drive the folding of a polypeptide chain through torsion space in the hopes of yielding near-native structures. Most attempts have enforced native secondary structures.[11,15,17] Accuracy within 4 Å of the native structure were achieved by Mumenthaler and Braun,[15] who managed to fold helices into clusters of native-like tertiary arrangements in 8 out of 10 test cases. Likewise, Sun et al.[17] reported similarly impressive results in four out of ten simulations. Recently, Simons et al.,[19] using a method that assembles fragments in the database, were able to fold small proteins to 3–4 Å RMS error without prior knowledge of the secondary structure, though complete convergence towards a native-like fold was not achieved.

How accurate must the backbones be in order for side-chain packing methods to be effective? Chung and Subbiah[33,34] systematically explored the relationship between backbone accuracy and the effectiveness of a side-chain prediction algorithm.[35,36] Specifically, they noted that when the backbone of a homologous protein departs 2 Å RMS from the experimentally-determined structure, the accuracy of their algorithm approached that of a random prediction. These baseline expectations are 22% for $\chi 1$ angles and 29% for $\chi 2$ angles, using the criterion that a given side-chain dihedral angle be predicted within 40° of its corresponding native angle[34] In a later study, Tuffery et al.[37] explored the same relation-

ship between side-chain prediction accuracy using their method[28,38] upon near-native backbones generated by Monte Carlo perturbation of the experimentally-determined structure. For the buried residues, they report ~70% accuracy in $\chi 1$ prediction and ~65% in the $\chi 2$ prediction, even when the mean backbone RMS deviation is 2 Å. The extent to which the different methods of generating these near-native backbones affects the prediction accuracy is unclear. Nevertheless, the outer limit of these studies (2 Å RMS from the native structure) is still outside the reach of even the best ab initio methods.[2]

Not surprisingly, side-chain prediction was not explicitly attempted in any of the ab initio studies described above. It would therefore be useful to evaluate side-chain prediction accuracy upon the near-native folds generated by ab initio methods. There are many reasons to be pessimistic that the prediction would be any better than random. Ab initio backbones present a stiff challenge for side-chain modeling techniques beyond their relatively high RMS deviation from the ideal structure. Consider two backbones at 2 Å RMS from the native fold, one generated ab initio and the other taken from a homologue. The backbone from the homologue will have a certain fraction of the residues identical to that of the target structure;[39] accordingly, the local steric environment surrounding each side-chain to be modeled will also be similar to that in the target. Ab initio folds, on the other hand, do not have the benefit of "sequence memory" to mold native-like steric environments, since the excluded volume surrounding each side chain is often implicit.[19,30] Moreover, backbone subtleties that better accommodate native side-chain packing are removed by idealized secondary structures and/or restricted backbone torsion angles. Buried residues are normally predicted correctly at a higher rate than the surface[2] when the backbone is ideal; this is a result of the more restrictive steric environment surrounding each residue in the core as compared to the surface, where residues are more labile. This tendency is also seen in homology modeling experiments.[33,34] It is unclear whether ab initio backbones, because of the limitations described above, will afford the same kind of benefit to core prediction.

Here we take as starting structures the nearest-native structures (< 4 Å RMS deviation) of four small proteins gathered from two different ab initio protocols.[19,30] First, we assess the theoretical limits of side-chain prediction methods in general by explicitly probing the steric environment surrounding the side chains as constrained by fixed near-native backbone scaffolds. This is accomplished by first building side chains in their native conformation upon each near-native backbone, followed by energy minimization of the side chains in torsion space. The deviation from the starting conformation indicates how permissive the steric environments of the ab initio folds are.

Next, side chains were positioned using three independent, established methods. The first is SCWRL, which minimizes the steric clash of backbone-dependent rotamers.[40] The second is SCMF, which uses self-consistent mean-field theory to position rotamers according to a van der Waals (Vdw) potential.[41] The third algorithm, SegMod,[25] pastes in side-chain conformers directly from a structural database in random order. Each approach is described in more detail in Materials and Methods.

The average prediction accuracy for each method towards each backbone set is evaluated and compared against the expectation that one would be correct by random prediction, defined in two ways. The first, which we designate the "random chi" method, is the measure suggested by Chung and Subbiah.[34] A value between −180° and 180° is randomly chosen for each χ angle in an unbiased and independent fashion. If one allows a tolerance of ±40°, the accuracy of this method would be 22% for $\chi 1$ and $0.22 \times 0.29 = 6\%$ for $\chi 1+2$. The random expectation for $\chi 2$ (29%) is somewhat higher than that of $\chi 1$ because the distal parts of some side chains are symmetric about $\chi 2$. However, the examination of rotamer libraries shows a tendency of χ angles to assume standard values and the interdependence of the various χ values. Hence, we suggest a second measure of random prediction, which we call the "random rotamer" method. Assuming that the rotamer approximation is sufficient for all side chains to be predicted, the unbiased choice of a $\chi 1$ value from the three standard angles of $g+$, $g-$, and t [27] produces a baseline expectation of 33% for $\chi 1$ for 17 residues. For Pro, there are essentially only two possibilities for $\chi 1$. This results in a weighted average of 34% over all $\chi 1$ angles, assuming a uniform distribution of the amino acids in a sequence. For $\chi 1+2$, the random expectation is equal to $1/n$, where n is the number of rotamers with at least two discrete χ angles. The weighted average over the Ponder and Richards library[27] yields a random prediction accuracy of 22% for $\chi 1+2$.

Recent data from Tuffery et al.[37] suggest that when side-chain prediction is performed upon a collection of near-native backbone conformations, increased prediction accuracy results from taking the most frequently predicted rotamer at each residue position. Similarly, Shenkin et al.[42] demonstrated that consensus side-chain predictions on a single fixed backbone also improves accuracy. We discuss the implications of taking a consensus-based prediction on side-chain prediction in the context of ab initio folding.

MATERIALS AND METHODS
Generation of Near-Native Folds

Two ab initio folding protocols were used to generate full polypeptide backbones. The method by Park and Levitt[30,31] builds all backbones using only four discrete points in Ramachandran space. The native secondary structure is idealized and held fixed while designated loop residues are permitted to explore all four ϕ,ψ possibilities in a combinatorial fashion. The approach by Simons et al.[19] starts with an extended polypeptide chain and uses Monte Carlo torsion moves with simulated annealing to generate compact folds. The energy function is a Bayesian scoring function. The move set for each residue is restricted to those in a library of fragments of unrelated protein structures with similar local sequences.[19,48] For each protein, 500 low-energy, compact folds were generated. For the purposes of side-chain prediction, only structures with RMS deviation in the Cα positions of ≤ 4 Å were saved. Models of 434 repressor (1r69) and ubiquitin (1ubq) were built by the method of Park and Levitt[30,31] and models of Protein A (1fc2:C) and the homeodomain protein (1hdd:C) by the method of Simons et al.[19] Full backbones were constructed from the Park and Levitt folds (which only had Cα atoms) by custom software developed in the David Baker lab at the University of Washington.

Side-chain atoms were not present in the construction of any model save the Cβ atom for the structures generated by Simons et al.[19] Thus these ab initio backbones have no knowledge of the explicit volumes or shapes of the side chains.

Optimization of Native Side-Chain Conformations on Non Native Backbones

The native side-chain conformations were built upon each ab initio backbone using custom software developed in the Ponder lab. The side chains were constructed using the χ angles from the experimentally-determined structures and the standard bond lengths and angles used by the TINKER software package (http://dasher.wustl.edu/tinker/). Energy minimization was performed in torsion space via the OPTIROT module of the TINKER software package to a gradient of 0.1 (kcal/mole/radian) using the CHARMM22 force field[21] (and A. McKerell, personal communication). OPTIROT employs the variable metric optimization described by Davidon.[49] Parameters involving charge interactions were neglected in the minimization, and a distance cutoff of 12.0 Å was enforced. All backbone torsion angles were fixed during the minimization. Since the two methods for generating the ab initio backbones did not account for side-chain volume exclusion, building side chains may result in severe clashes that can only be removed by energy minimization with a severe displacement cost. The importance of the minimizer itself on defining these displacements was tested by performing a second round of minimization, in which "soft" VdW potentials were used: the Lennard-Jones term was truncated at 10 kcal/mol. "Soft" VdW potentials are common in side-chain prediction methods (see below).

Prediction of Side-Chain Conformation

A battery of different methods were applied to the four sets of backbones in order to predict side-chain conformations. Prediction algorithms vary in how they discretize side-chain conformations, how they model potential energy, and how they combine these two considerations in a search strategy. Two of the three restrict side-chain conformations to a rotamer library.[40,41] Of these two, the method of Bower et al.[40] is the only one that uses a backbone-dependent rotamer library. The third method borrows side-chain conformations directly from a database.[25] The search strategies include combinatorial repacking[40] and self-consistent mean-field theory.[41] The method by Levitt[25] randomly selects residues to model and uses a Boltzmann-weighted probability to choose the side-chain conformation in the context of the backbone and the side chains already positioned. All methods use truncated van der Waals energy to soften the steric penalty associated with the imperfect modeling of the target side-chain conformation. However, whereas two of the three use Lennard-Jones potentials to model van der Waals forces, one chooses a simpler functional form to model steric energy.[40] None of the energy functions model electrostatic interactions, hydrogen bonds, or interactions with solvent. All three methods assume a fixed backbone at the time of side-chain placement. The method of Levitt,[25] which utilizes energy refinement of an average of ten independent initial structures, permits restrained backbone motion in the final modeling step.

For the homeodomain protein, Arg 18 was changed to an alanine since this side-chain conformation was not determined by X-ray crystallography (protein data bank (PDB)[53] entry 1hdd).

Evaluation of Side-Chain Accuracy

A predicted χ angle was considered correct if it was within $\pm 40°$ of its experimentally-determined value. Some side chains have two identical symmetry-related conformations about the $\chi 2$ angle (Phe, Tyr, Asp); only the lower of the two possible $\chi 2$ values was considered in the evaluation.

Consensus Modeling of Side-Chain Conformation

For each set of model structures, $\chi 1$ and $\chi 2$ dihedral angles were computed. Each $\chi 1$ angle was counted as $g+$, $g-$, or t, whichever was closest. Those side chains that had the consensus $\chi 1$ value were then collected to determine the consensus $\chi 2$ angle. For Leu, Ile, Met, Glu, Gln, Lys, and Arg, the procedure for $\chi 1$ was repeated for $\chi 2$. For residues with sp^2 hybridization at the gamma atom, a different counting system was applied. For Trp, Asn, and His, the consensus $\chi 2$ angles were 0°, 90°, 180°, and −90°, based on values suggested by the litera-

ture.[27,50,51] For Asp, Phe, and Tyr the consensus values were 0° and 90° because of symmetry considerations.

Designation of Buried and Exposed Residues

Computation of solvent accessibility for each side chain was performed by the software NACCESS.[52] Buried side chains were those that had relative solvent exposure of 20% or less in the native structure only.

Assessing Prediction Variability: Computation of Shannon Entropy

To evaluate the effect of prediction variability on consensus modeling, the Shannon entropy was calculated for all residues with χ angles.[42] For a given χ angle of a given residue, the entropy of prediction S is computed as:

$$S = -\sum_j f_j \ln f_j$$

where f_j is the frequency of occurrence of a standard χ angle j.

RESULTS AND DISCUSSION
Generation of Native-Like Backbones

Details for structure generation protocols are discussed in Materials and Methods. Three of the four are alpha-helical proteins; their PDB identifiers are enclosed in parentheses: 434 repressor (1r69, 97 structures), the homeodomain protein (1hdd:C, 32 structures), and Protein A (1fc2:C, 177 structures). The fourth is ubiquitin, which has both alpha helices and beta strands (1ubq, 19 structures). Table I lists the method of generation, the radius of gyration (R_G) and the mean RMS error for each of the four proteins.

Torsion-Space Minimization of Native Side Chains on Fixed Backbones

In general, side-chain prediction methods rely heavily upon the resolution of steric clash for their prediction of side-chain conformations, and this principle is true for the three methods evaluated in this study. It is therefore useful to probe explicitly the steric environment created by the fixed backbone template and evaluate its fitness for native side-chain conformations. After energy minimization in torsion space using the standard CHARM22 force field without electrostatics, we observe the frequency with which χ angles depart from their "native" conformation by more than $\pm 40°$. Thus an estimate of the collective steric tolerance of the ab initio backbone is obtained. Table I lists the average fraction of $\chi 1$ angles and $\chi 1+2$ angles that remained in their native conformations after resolution of steric clashes.

TABLE I. Torsion-Space Optimization of Native Side-Chain Conformations Upon Ab Initio Backbones, Based on the CHARMM VdW Potential[†]

Protein	nStr	$\langle$RMS$\rangle$	$\langle$RG$\rangle$	$\langle\chi^1\rangle$	$\langle\chi^{1,2}\rangle$	B$\langle\chi^2\rangle$	B$\langle\chi^{1,2}\rangle$	Generation
1fc2:C	177	3.72	9.27	0.60	0.42	0.50	0.33	Simons et al.[19]
1hdd:C	32	3.51	11.33	0.63	0.48	0.71	0.48	Simons et al.
1r69	97	3.35	10.66	0.58	0.39	0.53	0.36	Park and Levitt[31]
1ubq	19	3.19	11.76	0.58	0.41	0.49	0.30	Park and Levitt
Average				0.60	0.43	0.56	0.37	

[†]Columns marked with a "B" refer to prediction accuracy for buried residues, defined in Materials and Methods. nStr = Number of structures; $\langle$RMS$\rangle$ = average Cα RMS deviation (Å); $\langle$RG$\rangle$ = average radius of gyration (Å).

Overall, 60% of the side chains remained within 40° of their native $\chi1$, and 43% preserved their $\chi1+2$ dihedral angles. The corresponding values for the buried residues (see Materials and Methods) was somewhat lower (56% and 37%, respectively). The minimized side chains did not settle into standard conformations. χ dihedral angles cluster around three preferred conformations, $g+$, $g-$, and t, and any move from one of these conformation to another would result in a change of χ values by 120°; however, the distribution of angular displacement for all residues did not show a peak at 120° (Fig. 1). In comparison, steric minimization of side chains upon native backbones did not result in a large departure from the starting native side-chain conformations (Table II). After energy minimization of the side chains upon native scaffolds, 94% of the $\chi1$ values remained within 40° of the native starting conformations, and 81% of the native $\chi1+2$ dihedral angles were preserved. The homeodomain protein (1hdd:C) exhibited somewhat higher displacements from the native side-chain conformations. For the buried residues, one sees a higher conservation in the side-chain dihedral angles (95% and 93% for $\chi1$ and $\chi1+2$, respectively).

Building side chains on the ab initio backbones resulted in severe clashes, which were removed by energy minimization as described above, but at a severe displacement cost. It is important to know, however, what fraction of this displacement was due to the potential energy function of the minimizer itself. This was tested by modifying the van der Waals component of the CHARMM force field to mimic the energy functions employed by side-chain prediction methods. The classic Lennard-Jones 6–12 potential was truncated at 10 kcal/mol and the torsion-space optimization repeated. Table III shows that when a softened potential was used, the side-chains tended to stay closer to their original conformations: an average of 77% of $\chi1$ angles and 62% of $\chi1+2$ angle remained within 40°. We therefore conclude that the softer potential allows side-chain prediction methods to place approximately-correct rotamers in an environment that would otherwise be inhospitable.

The results indicate that a protein scaffold generated without explicit side chains, that crudely mod-
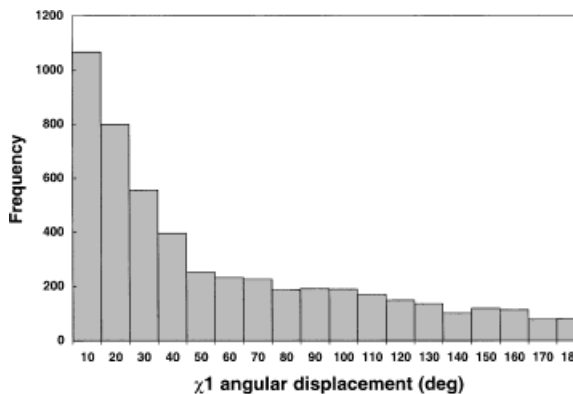

Fig. 1. Histogram of $\chi1$ angular displacement. Histogram of absolute angular difference between the $\chi1$ angle of the native structure and the predicted structures is plotted for 97 structures of 1r69 after torsion-space minimization. Electrostatic interactions were not considered in the optimization.

TABLE II. Torsion-Space Side-Chain Optimization Upon Native Backbones[†]

Protein	$\langle\chi^1\rangle$	$\langle\chi^{1,2}\rangle$	B$\langle\chi^1\rangle$	B$\langle\chi^{1,2}\rangle$
1fc2	1.00	0.83	1.00	1.00
1hdd:C	0.84	0.77	0.92	0.91
1r69	0.96	0.85	0.94	0.93
1ubq	0.94	0.80	0.94	0.86
Average	0.94	0.81	0.95	0.93

[†]Columns marked with a "B" refer to prediction accuracy for buried residues, defined in Materials and Methods.

els the correct number of interactions (Simons et al.[19]) and are quantitatively (< 4 Å RMS error) and qualitatively similar to the native structure,[31] cannot faithfully recreate the correct steric environments critical for side-chain prediction. Rather, placing side chains in their native conformations yields extremely high steric energies, and they were often driven from their starting conformations by more than 40° by van der Waals repulsion in a typical all-atom force field. The buried residues, which typically are predicted with higher accuracy, apparently face even a more demanding steric environment. Because the non-bonded terms in the force field involve only van der Waals interactions, the exposed side chains, which presumably face less

TABLE III. Modified Torsion-Space Optimization of Native Side-Chain Conformations Upon Ab Initio Backbones: Application of a Soft Lennard Jones Potential*

Protein	$\langle\chi^1\rangle$	$\langle\chi^{1+2}\rangle$	$B\langle\chi^1\rangle$	$B\langle\chi^{1+2}\rangle$
1fc2:C	0.78	0.63	0.70	0.62
1hdd:C	0.76	0.66	0.78	0.62
1r69	0.81	0.62	0.74	0.61
1ubq	0.72	0.55	0.63	0.52
Average	0.77	0.62	0.71	0.59

*Columns marked with a "B" refer to prediction accuracy for buried residues, defined in Materials and Methods.

steric stress, are less inclined to change conformation. The homeodomain protein (1hdd:C) has an anomalously high deviation in its energy-minimized side-chain positions. This result can be explained by its relatively expanded structure: its amino-terminal tail and carboxy-terminal helices have fewer than expected tertiary contacts in the X-ray structure. Consequently, its native radius of gyration of 11.3 Å is greater than that of 434 repressor (1r69; 10.1 Å), even though the latter protein has a longer sequence (56 amino acids for 1hdd:C; 63 for 1r69).

The data presented above provide an upper limit for expected side-chain prediction accuracies: on average, a maximum of 78% of $\chi 1$ can be correctly predicted overall, if a "soft" Lennard-Jones function is used to mimic VdW interactions. While this idealized upper-boundary does not strictly apply to side-chain prediction protocols (see below), they do place the prediction accuracy in its proper context.

Side-Chain Prediction on Ab Initio Backbones

Table IV and Figure 2 list the results for the three side-chain prediction methods applied on our four test proteins. Performance is evaluated in terms of the percentage of $\chi 1$ and $\chi 1+2$ angles correctly predicted within 40° of the native dihedral angles. Buried side-chains (defined in Materials and Methods) are tabulated separately. Although the three methods performed similarly in each test set, SegMod had the highest overall $\chi 1$ accuracy (50%) and SCWRL had the highest overall $\chi 1+2$ accuracy (36%). For buried residues, SegMod had the highest overall prediction accuracy (54% and 44%, respectively). The accuracy for the buried residues was higher for SCWRL and SegMod, but was not necessarily true on an individual basis (e.g. for 1r69). SCMF performed less well in the core. The standard deviations for each set of predictions are also listed in Table IV. A typical standard deviation in the $\chi 1$ prediction is 5% (6% for $\chi 1+2$). For reference, the results for self-modeling the four proteins by the three algorithms is shown in Table V.

In light of the data showing that side chains in their native conformation must change position to accommodate their placement upon non-native back-

bones, it would seem that algorithms designed to place side chains in their native conformation when the backbone is ideal should fail. This expectation is consistent with the degradation of prediction accuracy for backbones within ∼2 Å RMS from the ideal documented by earlier studies.[34,37] Neither study attempted side-chain placement upon backbones between 3–4 Å RMS from the ideal, as they did not utilize backbones gathered from ab initio folding experiments. Our idealized test case, in which we start with the correct side-chain conformation and relieve steric clashes, would indicate a best-case scenario of 71% in the $\chi 1$ predictions and 59% in $\chi 1+2$, given a perfect search method in effectively continuous torsion space (for the buried residues). Finally, the self-modeling experiments for the same four proteins indicate that the intrinsic accuracy of the methods ranges from roughly 65 to 75% for $\chi 1$ (60 to 65% for $\chi 1+2$) for the buried residues. The factors that affect the intrinsic accuracy include imperfections in the scoring functions, limitation in the discretization of side-chain conformations (e.g. rotamers), limitation in the quality of the native structure itself, and search strategies that are not guaranteed to find the global energy minimum. In light of these results, we were pleasantly surprised that the results from the SegMod prediction resulted in a mean accuracy of 54% for $\chi 1$ and 44% for $\chi 1+2$ (for buried residues).

Moreover, this accuracy appears to be general: SCMF, SegMod, and SCWRL, three completely different approaches, performed very similarly. All methods were well above the $\chi 1$ prediction accuracy of the "random chi" model (22%) and that of the more stringent "random rotamer" model (34%). For $\chi 1+2$, these results exceeded the "random chi" expectation of 6% and the "random rotamer" expectation of 22%.

There are at least two reasons why these prediction methods are well suited to the problem of side-chain placement upon non-optimal backbones. First, the methods do not use a molecular force field akin to CHARMM. Detailed all-atom force fields are appropriate for molecular dynamics simulation because they model the many types of inter-atomic interactions accurately. A hallmark of these potentials is inter-atomic repulsion modeled by a classic Lennard-Jones 6–12 function, which increases without limit as the inter-atomic distance approaches zero. Side-chain repacking algorithms have softened this repulsive term, thereby alleviating the problems associated with a rotamer representation of side-chain conformations and an unyielding backbone. In the three methods we tested, similar measures have been taken to make these repulsive terms more forgiving. SCWRL specifies atomic radii that are smaller than those found in standard force fields and a simplified steric energy function that is truncated

TABLE IV. Mean Side-Chain Prediction Accuracy for Three Algorithms[†]

Protein	Method	χ^1	sd	χ^{1+2}	sd	B-χ^1	sd	B-χ^{1+2}	sd
1fc2:C	SCWRL	0.51	0.06	0.31	0.06	0.55	0.15	0.55	0.19
	SCMF	0.52	0.04	0.35	0.04	0.42	0.17	0.30	0.19
	SegMod	0.49	0.06	0.29	0.06	0.53	0.13	0.56	0.20
1hdd:C	SCWRL	0.53	0.05	0.40	0.04	0.61	0.10	0.41	0.09
	SCMF	0.53	0.05	0.40	0.05	0.60	0.13	0.28	0.11
	SegMod	0.53	0.06	0.39	0.05	0.65	0.11	0.40	0.09
1r69	SCWRL	0.49	0.06	0.38	0.10	0.42	0.09	0.29	0.14
	SCMF	0.50	0.04	0.37	0.05	0.37	0.09	0.21	0.09
	SegMod	0.49	0.06	0.37	0.07	0.48	0.09	0.43	0.10
1ubq	SCWRL	0.44	0.04	0.34	0.05	0.45	0.09	0.36	0.09
	SCMF	0.42	0.05	0.28	0.03	0.36	0.10	0.26	0.07
	SegMod	0.49	0.04	0.27	0.06	0.49	0.08	0.38	0.13
Average	SCWRL	0.49	0.05	0.36	0.06	0.51	0.11	0.40	0.13
	SCMF	0.49	0.05	0.35	0.04	0.44	0.12	0.26	0.12
	SegMod	0.50	0.06	0.33	0.06	0.54	0.10	0.44	0.13

[†]Standard deviations for each prediction set is indicated by "sd." Columns marked with a "B" refer to prediction accuracy for buried residues, defined in Materials and Methods.

such that it cannot exceed a certain value as two atoms approach other.[40] Similarly, SCMF and Seg-Mod employ truncated Lennard-Jones potentials.[25,41]

In addition to tailoring the energy function to the task of side-chain placement on a fixed backbone, prediction methods typically discretize the space to be searched by means of a rotamer library. SCWRL chooses from a sophisticated backbone-dependent library, whereas SCMF is based on a more conventional extension of the Ponder and Richards[27] library. SegMod selects conformers already existing in its database of protein structures. In any case, side chains are restricted to discrete states, and the final structure is a low-energy combination (if not the global minimum) of these rotamers. Since the ab initio backbones have departed from the ideal structure such that 40% of the side chains must rotate 40° in the $\chi 1$ angle, a rotamer approximation may in fact regularize the side-chain conformation in many instances. In other words, if side-chain dihedral angles were only permitted to assume idealized values (e.g. $g+$, $g-$, and t in the $\chi 1$ angle) during energy minimization, it is likely that a greater fraction of residues would have remained in their most native-like rotamer conformations. Since the $\chi 1$ angle has only three ideal values, an energy function would merely need to report that the $\chi 1$ associated with the native structure resulted in *less* steric clash than the other two $\chi 1$ possibilities, irrespective of whether it itself clashes or not. Support for this hypothesis is seen in comparing results of a rotamer-based method[37] and a non-rotamer method[35,36] in the challenge of side-chain prediction upon backbones at 2 Å RMS from the native. For buried residues, the rotamer method yielded ~70% accuracy for the $\chi 1$ angle (~65% for $\chi 1+2$ angle). The other algorithm, which explores torsion space in fine 10° increments, only predicted ~50% of the $\chi 1$ angles correctly and ~30% of the

$\chi 1+2$ angles correctly,[34] though the search methods and test sets were different in the two studies.

Our data, which showed ~55% accuracy for $\chi 1$ prediction of buried residues using backbones 3–4 Å RMS from the ideal, is consistent with the trends reported by the earlier rotamer-based study.[37] Moreover, it approaches the idealized 60% $\chi 1$ accuracy of our energy minimization probe while remaining well over the random success rate. Taken all together, our data suggest that a softened energy function in conjunction with a discrete-state representation can be effective even when the global energy minimum is not guaranteed because of limitations in the search strategy.

Modeling Side Chains on Non-Native Backbones: Which Interactions Should Be Considered?

Positioning side chains on ab initio backbones that did not include side-chain information upon construction is a challenging task. We have seen that the native side-chain conformations do not fit in these backbones since they induce large steric clashes. Interestingly, removal of these clashes by energy minimization resulted in greater angular displacement for core residues than for exposed residues (see Tables I and III). For ab initio side-chain prediction on native backbones, core residues are usually better predicted, a result of their making more contacts with other side chains than exposed residues. The situation is not as clear in the case of side-chain prediction on non-native backbones: while both SCWRL and SegMod perform better on core residues, SCMF performs poorly on the same residues (see Table IV and Figure 2). To further investigate these issues, side-chain prediction was performed using only interactions between side chains and the backbone for two different methods showing differ-

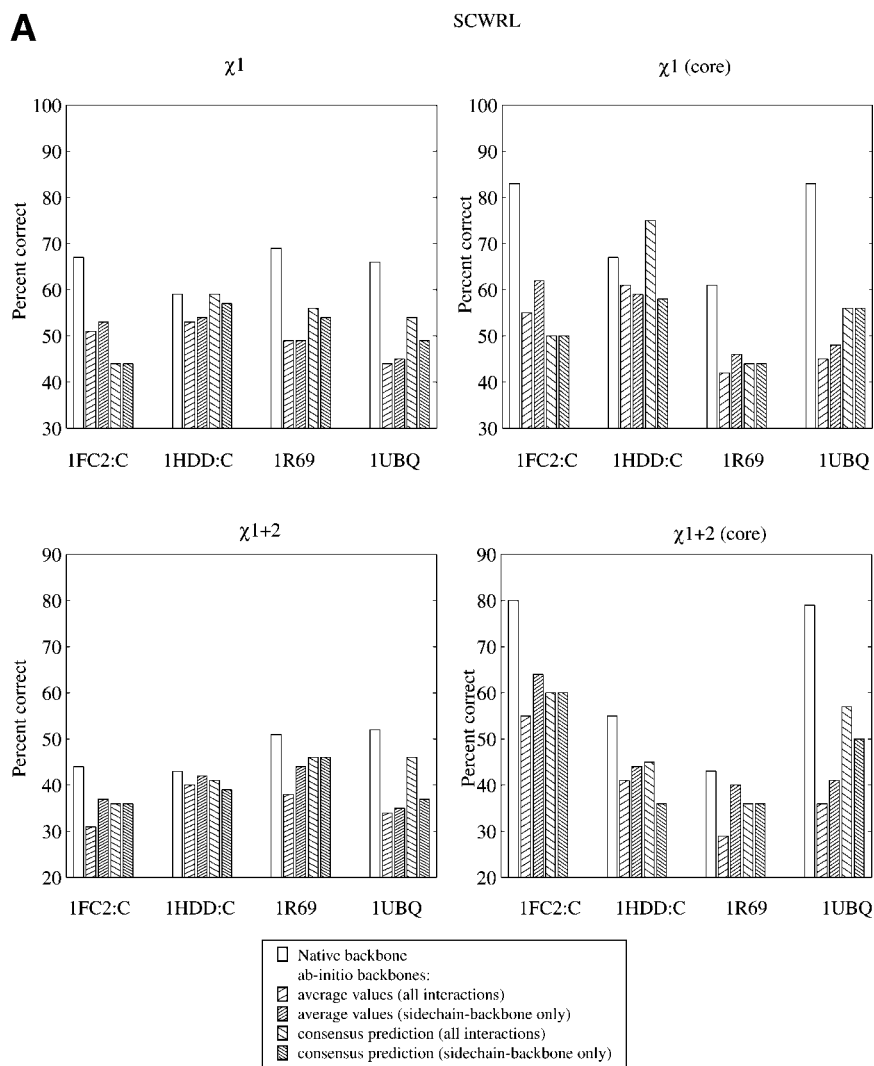**A**   SCWRL

Fig. 2.   Side-chain prediction accuracy on native and non-native backbones of four small proteins. Prediction of side-chain conformation for 1fc2:C, 1hdd:C, 1r69, and 1ubq based on three different methods: SCWRL **(A)**,[40] SCMF **(B)**,[41] and SegMod **(C)**.[25] Results are provided for both $\chi 1$ and $\chi 1+2$, for all residues and for core residues only. A predicted torsion angle χ is considered accurate if it lies within 40° of its value in the native conformation. A residue is considered buried if it has less than 20% relative accessible surface area. For SCWRL and SCMF, the fraction of side-chain torsion angles predicted correctly for five different prediction schemes are shown: prediction on the native backbone, the mean accuracy observed for a collection of decoy structures when considering interactions with the backbone and other side chains, the mean accuracy observed when considering only backbone interactions, and the two consensus accuracies corresponding to the alternative sets of interactions considered. Side-chain prediction based on backbone interaction alone was not available for the SegMod procedure.

ent trends, SCMF and SCWRL. Results are shown in Table VI and Figure 2.

The accuracy with respect to the native structure was usually higher when the inter-side-chain interactions were ignored rather than included. This tendency is true for both exposed and buried residues for both methods. Hence SCWRL and SCMF only differ in the extent to which non-native inter-side-chain interactions displace side chains from their conformations obtained from interactions with the backbone only. The difference in the VdW potentials used in both methods is a good explanation for this discrepancy: while both methods flatten the steric potential at 10 kcal/mol, the potential in SCMF retains the sharpness characteristic of the Lennard-Jones 6–12 function[41] while SCWRL uses a softer potential.[40]

Consensus Modeling of Side-Chain Conformations

The key to consensus modeling is that ab initio methods can generate many folds clustered near the native fold but slightly different from each other.[15] Thus, each individual fold models most of the backbone such that accurate side-chain placement can occur, while other parts are less well-modeled. In a
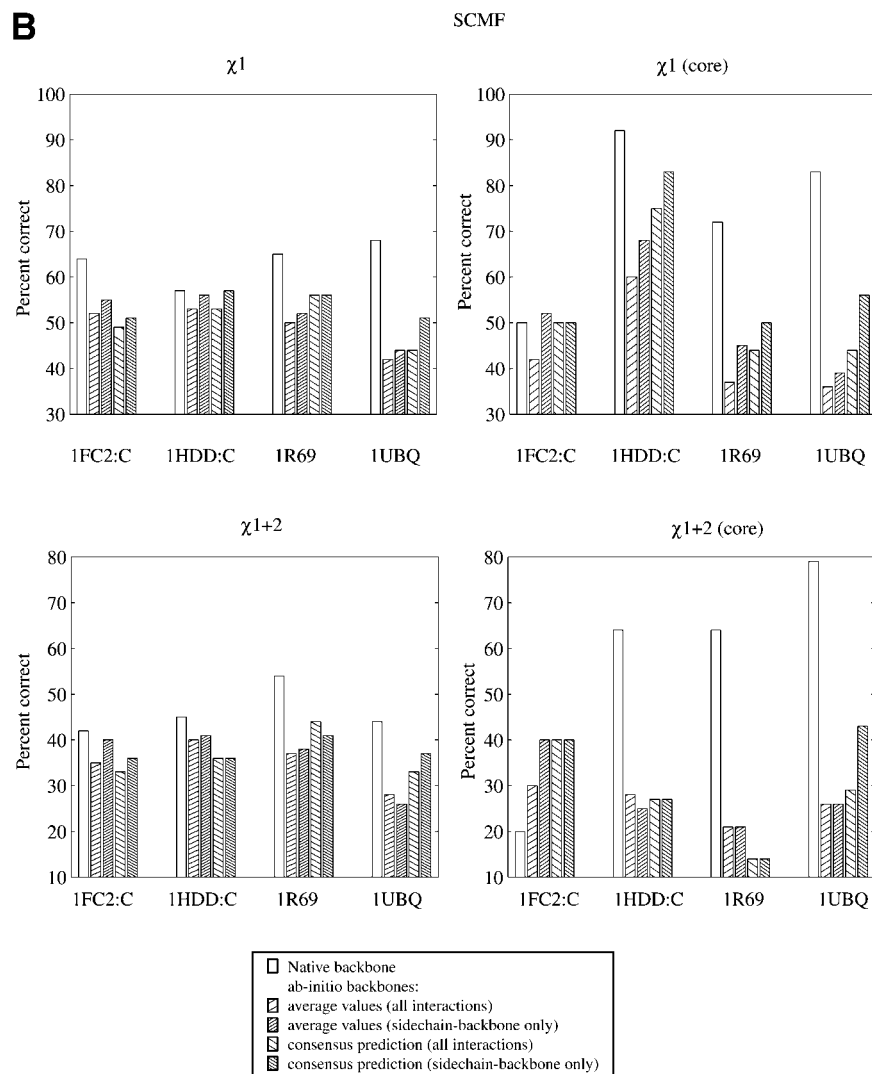
Figure 2. (Continued.)

collection of near-native folds, it may be that the errors in the predicted side-chain conformations are distributed differently on each backbone. When this happens, taking the consensus then masks the backbone-specific errors, thereby improving the overall prediction. However, the consensus approach breaks down if each individual backbone in the collection is so dissimilar from the others, and from the ideal structure, that each correctly predicted subset of residues is different.

For each near-native set of backbones, the most frequently predicted $\chi 1$ angle is taken as the consensus prediction. Likewise, the consensus $\chi 2$ is the most frequently predicted $\chi 2$ angle in the set of side chains with the consensus $\chi 1$ angle. The accuracy of each consensus prediction is shown in Table VII and Figure 2.

Using SCWRL, the mean consensus accuracy over all four proteins was 53% for $\chi 1$ and 42% for $\chi 1+2$,

slightly higher than the corresponding values for both SegMod and SCMF. Consensus prediction failed to maintain the average prediction accuracy for only one protein (1fc2, for all three methods). For the other three proteins, SCWRL and SegMod improved upon the mean prediction accuracy by taking the consensus. The consensus SCMF prediction did not show consistent improvement relative to the mean. Consensus modeling tended to perform better for the buried residues. The consensus prediction by SCWRL had the highest $\chi 1$ accuracy (56%), though all three methods were over 50%. The best $\chi 1+2$ consensus predictions was by SegMod at 51%.

Consensus modeling was similarly performed for the models obtained by applying only the interactions between side chains and the backbone. This resulted in an improvement in the accuracy of the prediction compared to the average over the set of decoys, but this improvement was less than that of
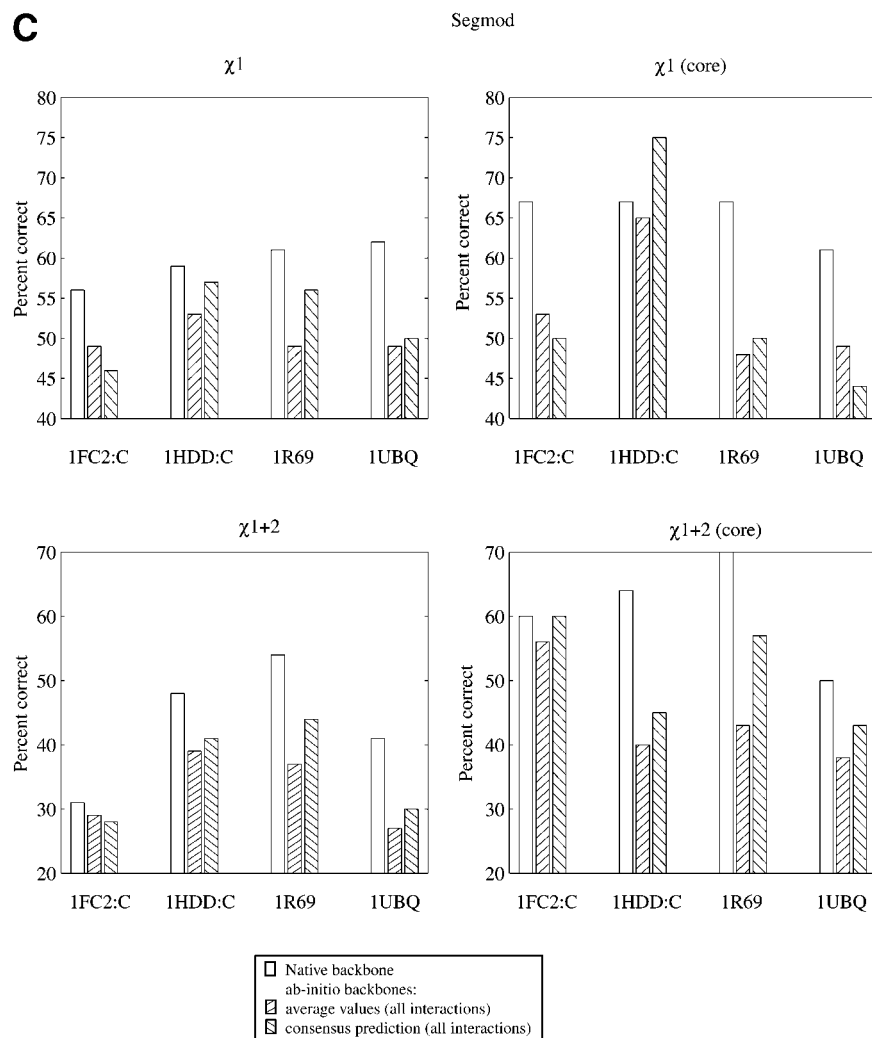
C

Segmod

Figure 2. (Continued.)

the consensus modeling with all interactions considered. Even though the improvement was somewhat less, the predictions themselves were comparable to the best consensus predictions after taking into account all interactions. Consensus SCMF results from backbone-only predictions yielded 54% $\chi1$ accuracy (60% in the core) on average, both of which are the highest of any method. This result is undoubtedly a function of the accurately positioned side chains already present in the models prior to taking a consensus.

Because modeling with only backbone interactions involves placing each side chain independently of the other side chains, the variability in a given side-chain prediction is less than the corresponding case when inter-side-chain interactions apply. The relationship between prediction variability and the potential benefit of taking a consensus is seen in the limiting case where every protein predicts an identical set of side-chain conformations, precluding the

TABLE V. Side-Chain Placement Upon Native Backbones[†]

Protein	Method	$\langle\chi^1\rangle$	$\langle\chi^{1,2}\rangle$	$\langle B\chi^1\rangle$	$\langle B\chi^{1,2}\rangle$
1fc2:C	SCWRL	0.67	0.44	0.83	0.80
	SCMF	0.64	0.42	0.50	0.20
	SegMod	0.56	0.31	0.67	0.60
1hdd:C	SCWRL	0.59	0.43	0.67	0.55
	SCMF	0.57	0.45	0.92	0.64
	SegMod	0.59	0.48	0.67	0.64
1r69	SCWRL	0.69	0.51	0.61	0.43
	SCMF	0.65	0.54	0.72	0.64
	SegMod	0.61	0.54	0.67	0.71
1ubq	SCWRL	0.69	0.52	0.83	0.79
	SCMF	0.68	0.44	0.83	0.79
	SegMod	0.62	0.41	0.61	0.50
Average	SCWRL	0.66	0.48	0.74	0.64
	SCMF	0.64	0.46	0.74	0.57
	SegMod	0.60	0.44	0.66	0.61

[†]Columns marked with a "B" refer to prediction accuracy for buried residues, defined in Materials and Methods.

TABLE VI. Side-Chain Predictions on Non-Native Backbones, Using Interactions Between Side Chains and the Backbone Only[†]

Protein	Method	$\langle\chi^1\rangle$	$\langle\chi^{1+2}\rangle$	$B\langle\chi^1\rangle$	$B\langle\chi^{1+2}\rangle$
1fc2:C	SCWRL	0.53	0.37	0.62	0.64
	SCMF	0.55	0.40	0.52	0.40
1hdd:C	SCWRL	0.54	0.42	0.59	0.44
	SCMF	0.56	0.41	0.68	0.25
1r69	SCWRL	0.49	0.44	0.46	0.40
	SCMF	0.52	0.38	0.45	0.21
1ubq	SCWRL	0.45	0.35	0.48	0.41
	SCMF	0.44	0.26	0.39	0.26
Average	SCWRL	0.50	0.40	0.54	0.47
	SCMF	0.52	0.36	0.51	0.28

[†]Columns marked with a "B" refer to prediction accuracy for buried residues, defined in Materials and Methods.

TABLE VII. Consensus Predictions[†]

Protein	Method	$\langle\chi^1\rangle$	$\langle\chi^{1+2}\rangle$	$B\langle\chi^1\rangle$	$B\langle\chi^{1+2}\rangle$
1fc2:C	SCWRL	0.44	0.36	0.50	0.60
	SCWRL-bb	0.44	0.36	0.50	0.60
	SCMF	0.49	0.33	0.50	0.40
	SCMF-bb	0.51	0.36	0.50	0.40
	SegMod	0.46	0.28	0.50	0.60
1hdd:C	SCWRL	0.59	0.41	0.75	0.45
	SCWRL-bb	0.57	0.39	0.58	0.36
	SCMF	0.53	0.36	0.75	0.27
	SCMF-bb	0.57	0.36	0.83	0.27
	SegMod	0.57	0.41	0.75	0.45
1r69	SCWRL	0.56	0.46	0.44	0.36
	SCWRL-bb	0.54	0.46	0.44	0.36
	SCMF	0.56	0.44	0.44	0.14
	SCMF-bb	0.56	0.41	0.50	0.14
	SegMod	0.56	0.44	0.50	0.57
1ubq	SCWRL	0.54	0.46	0.56	0.57
	SCWRL-bb	0.49	0.37	0.56	0.50
	SCMF	0.44	0.33	0.44	0.29
	SCMF-bb	0.51	0.37	0.56	0.43
	SegMod	0.50	0.30	0.44	0.43
Average	SCWRL	0.53	0.42	0.56	0.50
	SCWRL-bb	0.51	0.40	0.52	0.46
	SCMF	0.51	0.37	0.53	0.28
	SCMF-bb	0.54	0.38	0.60	0.31
	SegMod	0.52	0.36	0.55	0.51

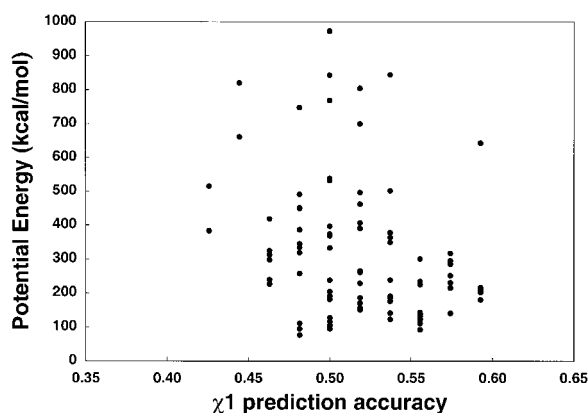[†]SCWRL-bb and SCMF-bb indicate that only interactions between side-chains and the backbone were taken into account for SCWRL and SCMF methods, respectively. Columns marked with a "B" refer to prediction accuracy for buried residues, defined in Materials and Methods.

possibility of improvement. In contrast, the addition of inter-side-chain interactions forces an increase in the sampling of side-chain conformations compatible with a rough tertiary fold, thereby allowing the consensus approach to improve upon the average prediction. Table VIII lists the variability for all predicted $\chi1$ and $\chi2$ angles in a set of folds. Each value in the table is the average Shannon entropy for a given χ angle over all residues in that protein that have the particular χ angle associated with them. Clearly, the backbone-only predictions are less variable overall than the predictions involving full interactions. Moreover, the somewhat lower variability seen in both full and backbone-only interactions for SCMF (Table VIII) explains why consensus modeling does little to aid this particular prediction method.

In our test set of four proteins, three out of the four consensus predictions were generally higher than the corresponding average prediction accuracy. Only the consensus modeling of Protein A (1fc2:C) failed to improve upon the average accuracy with any consistency. However, because of the high mean RMS error of 1fc2 backbones (3.72 Å) this result is not surprising because of reasons discussed above. In fact, if one only considers the nearest-native subset of the 177 backbones (i.e., those < 3.5 Å RMS), then the SCWRL consensus prediction rises above the average prediction accuracy for the subset (data not shown). The overall consensus accuracy of SCWRL, a consistently good performer, was 53% for all $\chi1$ angles and 42% for $\chi1+2$. In the core, the accuracy increased to 56% and 50%, respectively. While we were impressed that consensus modeling can apparently position half of the buried side chains correctly, we note that the statistics are subject to large fluctuation due to small numbers. For instance, although 1r69 and 1ubq had 18 buried residues with a $\chi1$, 1hdd:C had 12, and 1fc2:C only 6. The result is that prediction of buried subset of residues is more variable than side chains overall (Table IV).

Selecting a Model Protein Based on Side-Chain Prediction

Besides the improvement in the average side-chain prediction, consensus modeling also provides a single set of side chains with which it is possible to complete the all-atom model of the protein. Without a consensus set, it would be difficult to justify a single choice amongst the dozens of possible side-chain conformations provided by side-chain modeling upon each backbone. The data in Table IV indicate that for a given family of near-native ab initio folds, the side-chain prediction can be quite variable.

One could instead use the potential energy reported by the side-chain prediction methods to select a best model. However, we find that the potentials used to position the side-chains do not correlate well with the actual accuracy in side-chain prediction. For example, Figure 3 shows that the final potential energy reported by SCMF is essentially uncorrelated with the actual $\chi1$ accuracy. Indeed, the lowest energy model has a prediction accuracy (48%) below the mean (52%). This result is typical of the potentials used to position side chains on our backbone sets.

One might alternatively consider an energy-based approach for selecting a best candidate in a single

TABLE VIII. Prediction Variability Expressed as Shannon Entropies[†]

Protein	Method	$S(\chi^1)$	$S(\chi^2)$
1fc2:C	SCWRL	0.48	0.17
	SCWRL-bb	0.39	0.05
	SCMF	0.29	0.14
	SCMF-bb	0.23	0.06
	SegMod	0.52	0.44
1hdd:C	SCWRL	0.52	0.14
	SCWRL-bb	0.50	0.04
	SCMF	0.26	0.14
	SCMF-bb	0.18	0.08
	SegMod	0.56	0.38
1r69	SCWRL	0.61	0.35
	SCWRL-bb	0.56	0.09
	SCMF	0.48	0.36
	SCMF-bb	0.38	0.19
	SegMod	0.70	0.55
1ubq	SCWRL	0.68	0.32
	SCWRL-bb	0.58	0.09
	SCMF	0.64	0.27
	SCMF-bb	0.53	0.19
	SegMod	0.69	0.53

[†]SCWRL-bb and SCMF-bb indicate that only interactions between side-chains and the backbone were considered for the SCWRL and SCMF methods, respectively.

Fig. 3. Plot of potential energy versus $\chi 1$ prediction accuracy. The truncated Lennard-Jones 6–12 potential described by Koehl and Delarue[41] is plotted as a function of the prediction accuracy for the 97 structures of 1r69.

Fig. 4. Plot of $\chi 1$ accuracy versus Cα RMS error. The correlation between $\chi 1$ prediction accuracy and backbone quality is weak for 1r69 backbones 3–4 Å RMS from the native structure.

that soft potentials coupled with rotameric states make the predictions as robust as possible. Figure 4 depicts the $\chi 1$ accuracy of SCWRL as a function of the Cα RMS deviation. Even though the average prediction accuracy decreases slightly as the Cα RMS error increases, the relationship is clearly not monotonic and is of little predictive value, especially at 3–4 Å. Since structures with relatively high Cα RMS deviation and/or high potential energies are often as good as those with lower RMS deviation and/or low potential energies, it is useful to include all the information in the set. Taking a consensus prediction is apparently a reasonable and stable method of incorporating the information in aggregate.

Although consensus prediction allows selection of side-chain conformations derived from different backbones, the method described here will still consider one of the full backbones of the decoy set as a candidate for the backbone of the final model built ab initio. Ultimately, we would like to extend the consensus protocol to pick up fragments of backbone from different decoy structures, such as generation of a hybrid backbone closer to the native backbone.

CONCLUSION

Predicting side-chain conformations to complete an all-atom model in the context of ab initio folding is a true challenge, the difficulties of which we have enumerated above. We have shown that despite the apparently high RMS deviation of the backbone conformations, conventional homology modeling techniques do serve to predict the correct $\chi 1$ angle ~50% of the time on average. In the core, essentially the entire side chain is predicted more than 50% of the time on average. If the structures cluster sufficiently close to the ideal structure, this accuracy is assured (if not enhanced) by consensus modeling. Certainly, the final accuracy of a given model is inextricably linked to the prowess of the ab initio folding method used to generate the backbones. The structures tested by this study are 3–4 Å RMS from the ideal

modeling attempt. Assuming a best-case scenario of finding the backbone in the set with the lowest RMS deviation from the native structure, there is still no guarantee that the backbone would produce the model of highest accuracy. Given the exquisite sensitivity of side-chain prediction methods to steric clashes and limitations of the search heuristic, the overall quality of a backbone with respect to side-chain modeling is too subtle to be assessed by the relatively crude metric of RMS deviation when the structures lie between 3–4 Å from the ideal. The weak relationship between Cα RMS error and $\chi 1$ accuracy lends further support to our hypothesis
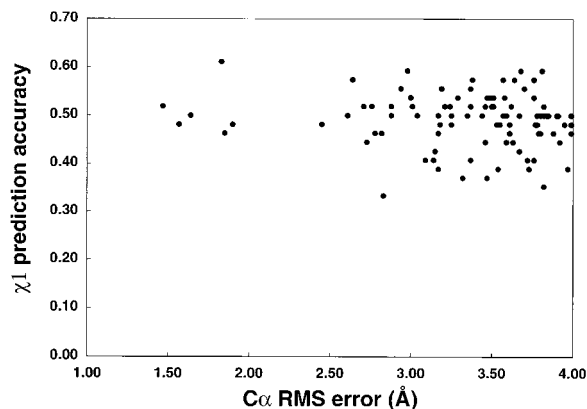
structure, but simplified energy functions cannot reliably identify them as near-native. The emergence of powerful all-atom knowledge-based potentials of mean force may lead to more accurate fold generation in the future.[43–47] The desired goal is the collection of several backbones tightly clustered about the native structure. An accurate backbone set increases the likelihood that the pool of possible side chains is highly enriched with native conformers and provides leverage to the consensus method by the cancellation of individual errors. Efforts are currently underway to determine a suitable method to build a single structure consistent with the information in the multiple backbones and the consensus set of side-chain conformations.

ACKNOWLEDGMENTS

REFERENCES

1. Levinthal, C. Are there pathways for protein folding? J. Chim. Phys. 65:44–45, 1968.
2. Levitt, M., Gerstein, M., Huang, E., Subbiah, S., Tsai, J. Protein folding: The endgame. Annu. Rev. Biochem. 66:549–579, 1997.
3. Covell, D.G. Folding protein α-carbon chains into compact forms by Monte Carlo methods. Proteins 14:409–420, 1992.
4. Covell, D.G. Lattice model simulations of polypeptide chain folding. J. Mol. Biol. 235:1032–1043, 1994.
5. Hinds, D.A., Levitt, M. A lattice model for protein structure prediction at low resolution. Proc. Natl. Acad. Sci. USA 89:2536–2540, 1992.
6. Hinds, D.A., Levitt, M. Exploring conformational space with a simple lattice model for protein structure. J. Mol. Biol. 243:668–682, 1994.
7. Kolinski, A., Skolnick, J. Monte Carlo simulations of protein folding. I. Lattice model and interaction scheme. Proteins 18:338–352, 1994.
8. Vieth, M., Kolinski, A., Brooks, C.L. 3rd, Skolnick, J. Prediction of the folding pathways and structure of the GCN4 leucine zipper. J. Mol. Biol. 237:361–367, 1994.
9. Wilson, C., Doniach, S. A computer model to dynamically simulate protein folding: Studies with crambin. Proteins 6:193–209, 1989.
10. Sun, S. Reduced representation model of protein structure prediction: statistical potential and genetic algorithms. Protein Sci. 2:762–785, 1993.
11. Bowie, J.U., Eisenberg, D. An evolutionary approach to folding small α-helical proteins that uses sequence information and an empirical guiding fitness function. Proc. Natl. Acad. Sci. USA 91:4436–4440, 1994.
12. Dandekar, T., Argos, P. Folding the main chain of small proteins with the genetic algorithm. J. Mol. Biol. 236:844–861, 1994.
13. Dandekar, T., Argos, P. Identifying the tertiary fold of small proteins with different topologies from sequence and secondary structure using the genetic algorithm and extended criteria specific for strand regions. J. Mol. Biol. 256:645–660, 1996.
14. Monge, A., Lathrop, E.J.P., Gunn, J.R., Shenkin, P.S., Friesner, R.A. Computer modeling of protein folding: Conformational and energy analysis of reduced and detailed protein models. J. Mol. Biol. 247:995–1012, 1995.
15. Mumenthaler, C., Braun, W. Predicting the helix packing of globular proteins by self-correcting distance geometry. Protein Sci. 4:863–871, 1995.
16. Srinivasan, R., Rose, G.D. LINUS: A hierarchic procedure to predict the fold of a protein. Proteins 22:81–99, 1995.
17. Sun, S., Thomas, P.D., Dill, K.A. A simple protein folding algorithm using a binary code and secondary structure constraints. Protein Eng. 8:769–778, 1995.
18. Yue, K., Dill, K.A. Folding proteins with a simple energy function and extensive conformational searching. Protein Sci. 5:254–261, 1996.
19. Simons, K.T., 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. 268:209–225, 1997.
20. Miyazawa, S., Jernigan, R.L. Estimation of effective inter-residue contact energies from protein crystal structures: Quasi-chemical approximation. Macromolecules 18:534–552, 1985.
21. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem. 4:187–217, 1983.
22. Cornell, W.D., Cieplak, P., Bayly, C.I., et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J. Am. Chem. Soc. 117:5179–5197, 1995.
23. Levitt, M., Hirshberg, M., Sharon, R., Daggett, V. Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution. Comp. Phys. Comm. 91:215–231, 1995.
24. Jones, T. A., Thirup, S. Using known substructures in protein model building and crystallography. EMBO J. 5:819–822, 1986.
25. Levitt, M. Accurate modeling of protein conformation by automatic segment matching. J. Mol. Biol. 226:507–533, 1992.
26. Vasquez, M. Modeling side-chain conformation. Curr. Opin. Struct. Biol. 6:217–221, 1996.
27. Ponder, J.W., Richards, F.M. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. J. Mol. Biol. 193:775–791, 1987.
28. Túffery, P., Etchebest, C., Hazout, S., Lavery, R. A new approach to the rapid determination of protein side-chain conformations. J. Biomol. Struct. Dyn. 8:1267–1289, 1991.
29. Dunbrack, R. L., Jr, Karplus, M. Backbone-dependent rotamer library for proteins: Application to side-chain prediction. J. Mol. Biol. 230:543–574, 1993.
30. Park, B., Levitt, M. The complexity and accuracy of discrete state models of protein structure. J. Mol. Biol. 249:493–507, 1995.
31. Park, B., Levitt, M. Energy functions that discriminate X-ray and near-native folds from well-constructed decoys. J. Mol. Biol. 258:267–392, 1996.
32. Park, B.H., Huang, E.S., Levitt, M. Factors affecting the ability of energy functions to discriminate correct from incorrect folds. J. Mol. Biol. 266:831–846, 1997.
33. Chung, S.Y., Subbiah, S. The use of side-chain packing methods in modeling bacteriophage repressor and cro proteins. Protein Sci. 4:2300–2309, 1995.
34. Chung, S.Y., Subbiah, S. How similar must a template protein be for homology modeling by side-chain packing methods? In: "Proceedings of the Pacific Symposium on Biocomputing, Hawaii, USA." Hunter, L., Klein, T.E., (eds.). New Jersey: World Scientific, 1996: 126–141.

35. Lee, C., Subbiah, S. Prediction of protein side-chain conformation by packing optimization. J. Mol. Biol. 217:373–388, 1991.

36. Lee, C. Predicting protein mutant energetics by self-consistent ensemble optimization. J. Mol. Biol. 236:918–939, 1994.

37. Tuffery, P., Etchebest, C., Hazout, S. Prediction of protein side chain conformations: A study on the influence of backbone accuracy on conformation stability in the rotamer space. Protein Eng. 10:361–372, 1997.

38. Tuffery, P., Etchebest, C., Hazout, S., Lavery, R. A critical comparison of search algorithms applied to the optimization of protein side-chain conformations. J. Comput. Chem. 14:790–798, 1993.

39. Summers, N. L., Carlson, W. D., Karplus, M. Analysis of side-chain orientations in homologous proteins. J. Mol. Biol. 196:175–198, 1987.

40. Bower, M.J., Cohen, F.E., Dunbrack, R.L., Jr. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: A new homology modeling tool. J. Mol. Biol. 267:1268–1282, 1997.

41. Koehl, P., Delarue, M. Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy. J. Mol. Biol. 239:249–275, 1994.

42. Shenkin, P.S., Farid, H., Fetrow, J.S. Prediction and evaluation of side-chain conformations for protein backbone structures. Proteins 26:323–352, 1996.

43. Delarue, M., Koehl, P. Atomic environment energies in proteins defined from statistics of accessible and contact surface areas. J. Mol. Biol. 249:675–690, 1995.

44. Jernigan, R.L., Bahar, I. Structure-derived potentials and protein simulations. Curr. Opin. Struct. Biol. 6:195–209, 1996.

45. Sippl, M.J., Ortner, M., Jaritz, M., Lackner, P., Flöckner, H. Helmholtz free energies of atom pair interactions in proteins. Folding Des. 1:289–298, 1996.

46. Melo, F., Feytmans, E. Novel knowledge-based mean force potential at atomic level. J. Mol. Biol. 267:207–222, 1997.

47. Samudrala, R., Moult, J. An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction. J. Mol. Biol. 275: 895–916, 1998.

48. Bystroff, C., Han, K.F., Simons, K.T., Baker, D. Local sequence-structure correlations in proteins. Curr. Opin. Struct. Biol. 7:417–421, 1996.

49. Davidon, W.C. Optimally conditioned optimization algorithms without line searches. Math. Programming 9:1–30, 1975.

50. Janin, J., Wodak, S., Levitt, M., Maigret, B. Conformation of amino acid side-chains in proteins. J. Mol. Biol. 125:357–386, 1978.

51. Dunbrack, R.L., Jr , Cohen, F.E. Bayesian statistical analysis of protein side-chain rotamer preferences. Protein Sci. 6:1661–1681, 1997.

52. Hubbard, S.J., Thornton, J.M. 'NACCESS', Computer Program. London: University College, Department of Biochemistry and Molecular Biology, 1993.

53. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., et al. Protein Data Bank: A computer-based archival file for macromolecular structures. J. Mol. Biol. 112:535–542, 1977.