Protein Science (1997), 6:1167–1178. Cambridge University Press. Printed in the USA.

De novo design of the hydrophobic core of ubiquitin

GREG A. LAZAR, JOHN R. DESJARLAIS, AND TRACY M. HANDEL
Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, California 94720

(RECEIVED December 26, 1997; ACCEPTED March 6, 1997)

Abstract

We have previously reported the development and evaluation of a computational program to assist in the design of hydrophobic cores of proteins. In an effort to investigate the role of core packing in protein structure, we have used this program, referred to as Repacking of Cores (ROC), to design several variants of the protein ubiquitin. Nine ubiquitin variants containing from three to eight hydrophobic core mutations were constructed, purified, and characterized in terms of their stability and their ability to adopt a uniquely folded native-like conformation. In general, designed ubiquitin variants are more stable than control variants in which the hydrophobic core was chosen randomly. However, in contrast to previous results with 434 cro, all designs are destabilized relative to the wild-type (WT) protein. This raises the possibility that β-sheet structures have more stringent packing requirements than α-helical proteins. A more striking observation is that all variants, including random controls, adopt fairly well-defined conformations, regardless of their stability. This result supports conclusions from the cro studies that non-core residues contribute significantly to the conformational uniqueness of these proteins while core packing largely affects protein stability and has less impact on the nature or uniqueness of the fold.

Concurrent with the above work, we used stability data on the nine ubiquitin variants to evaluate and improve the predictive ability of our core packing algorithm. Additional versions of the program were generated that differ in potential function parameters and sampling of side chain conformers. Reasonable correlations between experimental and predicted stabilities suggest the program will be useful in future studies to design variants with stabilities closer to that of the native protein. Taken together, the present study provides further clarification of the role of specific packing interactions in protein structure and stability, and demonstrates the benefit of using systematic computational methods to predict core packing arrangements for the design of proteins.

Keywords: computational; genetic algorithm; hydrophobic core packing; protein design; ubiquitin

Protein design is an extremely powerful method for critically testing the relationship between the primary sequence and three-dimensional structure of proteins. One goal of protein design is to reconstruct known three-dimensional folds from completely novel amino acid sequences. The concepts used to design the amino acid sequences constitute a hypothesis of what interactions are necessary for a protein to adopt a given three-dimensional structure with the properties of natural proteins, and at what level of detail these interactions must be considered.

The hypotheses of early protein design efforts were based on the assumption that simple empirical rules would be sufficient. These designs included, foremost, patterns of hydrophobic and hydrophilic residues, and to a lesser extent, secondary structure propensities and potential electrostatic interactions. The common result of such designs has been a protein that contains significant amounts of the desired secondary structure and folds into an approximately correct topology (Hecht et al., 1990; Kamtekar et al., 1993; Quinn et al., 1994; Yan & Erickson, 1994), sometimes with considerable stability (Regan & DeGrado, 1988). This result strongly supports the notion that burial of hydrophobic surface area is largely sufficient to direct folding of a polypeptide chain to the correct global fold (Dill, 1990; Handel et al., 1993; Kamtekar et al., 1993). However, a protein that possesses all of the characteristics of a natural protein has yet to be designed. The major obstacle has been the inability to design proteins with a well-ordered (i.e., unique) tertiary structure. Consequently, an important goal in current design strategies is to determine the key interactions that encode the well-ordered properties of natural proteins, and to develop methods to adequately model them.

One explanation for the non-native behavior of designed proteins is that they lack the specific packing interactions observed in the cores of natural proteins (Richards, 1977; Lim & Sauer, 1991; Richards & Lim, 1993). This suggestion has prompted several groups to consider packing interactions more seriously in their design strategies (Raleigh & DeGrado, 1992; Quinn et al., 1994; Tanaka et al., 1994). The role of core packing in protein design has been investigated both in a de novo system (Regan & DeGrado, 1988) and in the redesign of natural proteins (Harbury et al., 1993;

Reprint requests to: Tracy M. Handel, Department of Molecular and Cell Biology, 229 Stanley Hall, University of California at Berkeley, Berkeley, California 94720, e-mail: handel@paradise1.berkeley.edu.

Munson et al., 1994, 1996). In these studies, qualitative approaches to core packing were used to design helical bundle proteins. Although successful, these strategies are limited in that one can only consider a few alternative core sequences among the enormous number of potential sequences. A method that can efficiently predict a large number of alternative core sequences using more objective and quantitative measures would clearly be desirable for protein design. Success in the prediction of side chain structure, core sequence, and relative stabilities in natural proteins (Ponder & Richards, 1987; Summers & Karplus, 1989; Holm & Sander, 1991; Lee & Levitt, 1991; Lee & Subbiah, 1991; Tuffery et al., 1991; Hellinga & Richards, 1994; Kono & Doi, 1994) has recently encouraged the use of computational approaches for the design of hydrophobic cores (Hurley et al., 1992; Desjarlais & Handel, 1995; Dahiyat & Mayo, 1996; P.B. Harbury, J.J. Plecs, B. Tidor, T. Alber, P.S. Kim, in prep.). In these studies various optimization procedures were used to evaluate the "fitness" of a significant number of alternative core sequences. However, despite the success of these studies, a fully predictive understanding of hydrophobic core packing in proteins has not been realized, and the de novo design of stable, unique proteins remains a challenging problem.

This study is an extension of previous work in which we reported the development of Repacking of Cores (ROC), a computational program that attempts to find novel core sequences given the backbone structure of a protein of interest (Desjarlais & Handel, 1995). The program uses a genetic algorithm (Holland, 1992) to optimize the search for alternative core structures, which are scored on the basis of a van der Waals potential energy function. The effectiveness of ROC was tested by characterizing several hydrophobic core variants of the protein 434 cro. Results showed that variants predicted to be energetically favorable were comparable in stability to wild-type (WT) and much more stable than control sequences in which the hydrophobic core was chosen randomly, validating the methodology.

Because 434 cro is an all α-helical protein, we became interested in the predictive ability of ROC to redesign cores of proteins with high β-sheet content. Given that the hydrogen bonding that stabilizes an α-helix is predominantly local, whereas that stabilizing a β-sheet is nonlocal, it has been suggested that α-helical folds are intrinsically easier to design than β-sheet folds (Hecht, 1994). This is supported by the fact that the most successful attempts at de novo protein design have used helical bundles as model systems (Betz et al., 1993). We therefore chose the $\alpha + \beta$ protein ubiquitin, a single-domain eukaryotic protein involved in proteolytic degradation. Several other aspects of ubiquitin besides its β-sheet content make it a good model system for this study. First, it is extremely stable. 434 cro is only moderately stable, and therefore variants with randomized hydrophobic cores, made as controls, resulted in unfolded proteins. As a consequence, expression, purification, and characterization of these proteins was difficult. Ubiquitin's high stability allows enough "folding room" so that even highly destabilized variants are structured and can be easily expressed and characterized. Secondly, ubiquitin is small (76 residues), extremely soluble, and gives well-dispersed NMR spectra, all of which should allow detailed structural characterization of variants by NMR. Third, there is a significant amount of structural (Di Stefano & Wand, 1987; Vijay-Kumar et al., 1987; Weber et al., 1987), dynamic (Schneider et al., 1992; Wand et al., 1996), thermodynamic (Wintrode et al., 1994), and kinetic folding data (Briggs & Roder, 1992; Khorasanizadeh et al., 1993; Khorasanizadeh et al., 1996) available for the WT protein, which will be useful for comparison with our designed proteins. In the present study, we designed and expressed several core variants of ubiquitin, and investigated how structure, stability, and conformational uniqueness is affected by alternative core packing arrangements.

Results

Description of program and parameters

Details of ROC have been described previously (Desjarlais & Handel, 1995). Briefly, ROC uses a genetic algorithm (GA) to optimize a search for low energy core structures for a given protein. The search is conducted using a library of side-chain rotamers "customized" for a particular protein backbone which remains rigid throughout the search. The potential energy function is based predominantly on a Lennard-Jones potential function, with the inclusion of a side-chain torsional potential for some versions of the program. The starting population for a GA run consists of model structures whose core sequences and structures are chosen randomly. Each round of evolution begins by calculating the energies for each model structure in the population. These structures are then recombined, with the more energetically favorable sequences in the population recombining more frequently. The round ends with the introduction of random mutations of either side-chain identity or rotamer conformation. The program proceeds through several hundred iterations of a cycle consisting of energy calculation, recombination, and mutation. At the end of the search the program outputs a list of energetically favorable core sequences with their predicted side-chain orientations.

Concurrent with our use of ROC to design ubiquitin variants, we attempted to improve the program by optimizing the atomic radii and well depths parameters of the Lennard-Jones potential function. The parameters were modified twice, totaling three sets. The first set of parameters is that used with the initial version of ROC (Desjarlais & Handel, 1995). The motivation for improving these parameters arose from experimental results of variants designed with the first parameter set, which were predicted to be more stable than WT but found to be less stable.

The program with each respective parameter set is designated ROC1, ROC2, and ROC3 (see Materials and methods for details).

Redesign of ubiquitin's hydrophobic core

We have used ROC to design nine hydrophobic core variants of ubiquitin using the backbone coordinates of the X-ray structure (accession code 1UBI) (Vijay-Kumar et al., 1987). Fourteen core positions were chosen by visual inspection of the structure, and are displayed in the ribbon diagram of WT ubiquitin (Fig. 1). These residues make up a single continuous hydrophobic core at the interface between the α-helix and the β-sheet. In the designed proteins, the core residues were allowed to mutate to Val, Leu, Ile, and Phe. Other hydrophobic amino acids such as Met, Trp, and Tyr have partial polar character and were therefore not included in the search.

Sequences were evaluated on the basis of several quantitative and qualitative criteria. Design candidates were identified by comparing the output energies reported by the program. We then considered the frequency with which a sequence recurred during the evolution; a frequently predicted core sequence indicates that it is less sensitive to subtle differences in packing arrangements. Preference was also given to sequences with core volume and second-

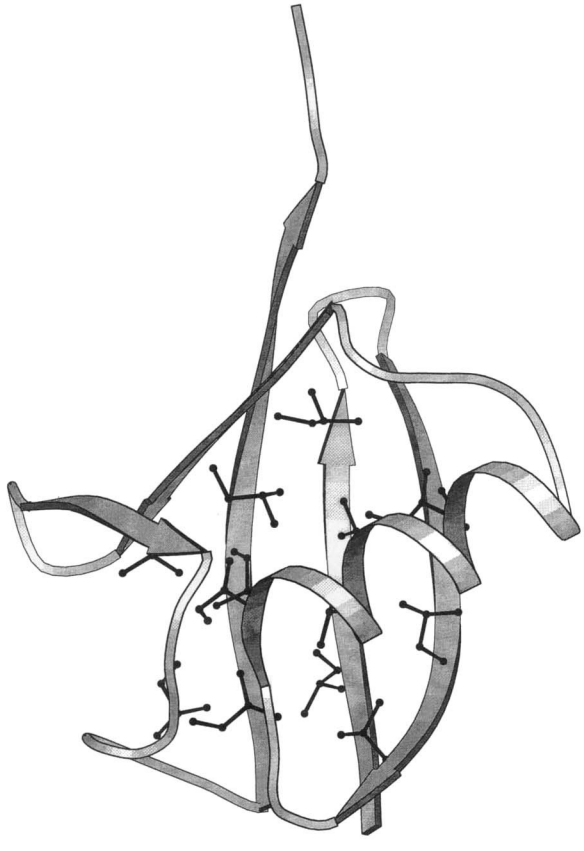

Fig. 1. Molscript diagram (Kraulis, 1991) of human WT ubiquitin backbone with the 14 chosen core side chains. The structure is based on the X-ray coordinates (Vijay-Kumar et al., 1987).

ary structure propensity comparable to the WT protein so that experimental results could be interpreted primarily in terms of packing. Models of selected designs were also visually inspected with the program INSIGHT II (Biosym Technologies, San Diego, California) as further confirmation of a "well-packed" core.

In addition to the above criteria, it is important to emphasize that the dominant criterion used to select design candidates was the number of mutations relative to the WT sequence. Studies of high resolution structures of hydrophobic core mutants have shown that natural proteins typically relax to accommodate a small number of mutations, resulting in a slightly perturbed WT hydrophobic core (Eriksson et al., 1992, 1993; Baldwin et al., 1993; Lim et al., 1994). Our aim was to design entirely novel packing arrangements in the core, and therefore we selected sequences with a large (six to eight) number of changes from WT. Two variants have fewer than six mutations; these were chosen because the output energies indicated the superiority of these core sequences over other sequences.

Characterization of the solubilities, circular dichroism (CD) spectra, and stabilities of the proteins distinguished two types of variants, which we refer to as class I and class II (see below). The core composition and calculated energies of the designs are shown in Figure 2 along with the volume difference from the WT core and class to which each belongs.

Solubility of ubiquitin variants

An initial indication of a successful protein design, particularly one involving β-sheet structure, can be solubility. The six class I variants were all expressed in soluble form and purified in the same way as the WT protein (see Materials and methods). The three class II variants, however, were purified from inclusion bodies. The tendency of class II variants to aggregate was corroborated during the preparation of NMR samples. At pH 5, 1 mM NMR samples of all class I proteins were prepared with no apparent aggregation, but it was not possible to achieve this concentration with the class II proteins. Among the class I variants there is some variability in solubility as well. Attempts at preparing 1 mM samples at higher pH resulted in aggregation of 3D6, and to a small

Protein	Energy	ΔV	Class	Residue													
				3 s	5 s	13 s	15 s	17 s	23 h	26 h	30 h	43 s	50 s	56 c	61 c	67 s	69 s
WT	-117.9	-	-	I	V	I	L	V	I	V	I	L	L	L	I	L	L
1D8	-96.5	-3	I	**L**	V	**L**	**V**	**L**	**V**	**L**	I	L	L	L	**L**	**I**	L
1D7	-113.1	+8	I	**V**	**L**	**V**	**I**	V	**V**	**F**	I	L	L	L	I	**I**	L
2D6	-107.2	+25	I	**L**	**I**	**V**	L	V	I	**L**	**L**	L	L	L	I	L	**I**
2D7	-103.8	+60	II	**L**	**L**	**V**	L	V	I	**I**	**L**	**F**	L	L	I	L	**I**
3D6	-102.7	0	I	**L**	V	I	**I**	**I**	**V**	V	I	L	L	L	**L**	**I**	L
3D4	-105.5	-29	I	**L**	V	I	L	**L**	**V**	V	I	L	L	L	**V**	L	L
3D3	-107.0	-2	I	**L**	V	I	L	**L**	**V**	V	I	L	L	L	I	L	L
R7	-66.7	0	II	I	**I**	**L**	**V**	V	I	V	**L**	L	**I**	**I**	**L**	L	L
R6	-72.5	+30	II	I	V	I	**I**	**I**	I	V	**L**	**I**	**I**	**I**	I	L	L

Fig. 2. Designed and control variants of ubiquitin. For variant names, the first number designates the parameter set used to design the protein, the letter that follows designates whether the protein is a design (D) or a random control (R), and the final number designates the number of changes from the WT core sequence. The letters under the residue numbers indicate the secondary structure of that position: s indicates sheet, h indicates helix, and c indicates coil. Boxed residues indicate mutations from the WT protein. The output energies from the program are presented, but because different parameter sets were used for the designs these energies are those predicted by our final version of the program ROC* (see below). Also listed is the volume difference from WT, ΔV in Å^3, and the class to which each variant is assigned.

extent 2D6. These results indicate that packing interactions in the hydrophobic core of ubiquitin affect its solubility.

CD spectra

To verify that the designs have retained the overall fold of the WT protein we examined the CD spectra of all 10 proteins (Fig. 3). Five of the six class I variants, 1D8, 2D6, 3D3, 3D4, and 3D6, have spectra that are within error of the WT protein. The other class I variant, 1D7, has a spectrum that has an amplitude comparable to the WT signal, but also a larger positive band at approximately 218 nm. This difference is perhaps due to the presence of an extra phenylalanine, which can make significant contributions to features in both the near and far UV CD spectrum (Manning & Woody, 1989). In contrast, the three class II variants (R6, R7, and 2D7) have CD spectra which differ significantly from the WT protein in shape and/or amplitude. These results can be attributed to loss of stability, subtle differences in native structure (e.g., twist of the sheet), or a combination of both. Stability data (see next section) taken under identical conditions to the CD spectra indicate that for 2D7 and R6 the ratio of folded to unfolded protein is on the order of 100:1. Such a small population of unfolded protein cannot account for the amplitude loss in these spectra. Furthermore, although R6 is more stable than R7, it has a smaller amplitude. Finally, the spectra of all three class II proteins were unaffected by the addition of 40% glycerol, a protein renaturant (data not shown). Together with the NMR and ANS data described below, these results suggest that the anomalies in the class II CD spectra are due to subtle differences in the local structure of these proteins rather than stability differences or gross changes of the fold.

Stability and cooperativity

The stabilities of designed proteins are typically characterized by thermal denaturation because the cooperativity of this transition can be a sensitive probe of structural uniqueness. However, at physiological pH, WT ubiquitin is stable to temperatures in excess of 100 °C. The more stable variants showed a similar resistance to thermal denaturation, whereas the less stable variants aggregated before an unfolding transition could be observed. For these reasons, the stability of each protein was investigated by monitoring the change in CD signal at 222 nm as a function of guanidine-hydrochloride (GuHCl) concentration. These data are presented in Figure 4 as the apparent fraction of unfolded protein along with the thermodynamic parameters describing the folding transitions: ΔG_{H_2O}, the free energy for unfolding in the absence of denaturant, m, the cosolvation free energy change per mole of denaturant, and C_m, the concentration of denaturant at the midpoint of the transition. These results for WT ubiquitin at pH 7.0 (ΔG_{H_2O} = 7.2 kcal mol^{-1}, m = 1.9 kcal mol^{-1} M^{-1}, and C_m = 3.8 M) are comparable to those of Roder and colleagues who report ΔG_{H_2O} = 6.7 kcal mol^{-1}, m_{den} = 1.8 kcal mol^{-1} M^{-1}, and C_m = 3.9 M at pH 5.0 (Khorasanizadeh et al., 1993).

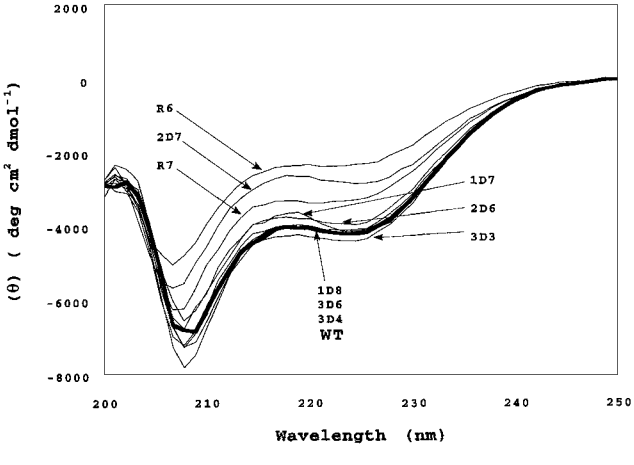

Fig. 3. Far UV CD spectra of the ubiquitin proteins. The spectrum of the WT protein is shown in bold.

Three distinct stability ranges are observed among the 10 proteins, which is the primary reason for separating them into two classes: WT is extremely stable, class I variants are moderately destabilized relative to WT, and class II variants are extremely destabilized relative to WT. This, in conjunction with the fact that ROC1 predicted the WT core sequence to be far from the lowest in energy, is what motivated us to reparameterize the program. Nonetheless, the results are encouraging because all designs except for 2D7 are considerably more stable than the two control proteins, i.e., only one of the seven designs is in class II. The wide range of stabilities of the 10 ubiquitin variants emphasizes the fact that packing interactions in the hydrophobic core play a crucial role in determining the stability of a protein.

In addition to stability differences, there are also slight differences in the m values of the proteins. Although these differences may be close to the error of the measurement, there are clear trends. All of the designs, including 2D7, have m values 0.1–0.5 greater than the two control proteins. m values are thought to reflect the amount of hydrophobic surface area that becomes exposed to solvent as a protein folds or unfolds (Schellman, 1978; Shortle et al., 1990). Therefore, one interpretation of the m values is that the designed proteins have more efficiently packed hydrophobic cores than the random control proteins; not only does better packing make the designs more stable, it also allows them to better sequester their hydrophobic surface area from solvent. Alternatively, m values can be affected by a loss of cooperativity in the GuHCl induced unfolding transition. Further experiments are required to distinguish between these possibilities.

An unexpected result is that all of the designs have m values 0.1–0.4 greater than WT. Although the above explanations might also apply to this observation, neither of these explanations seems likely. If surface area burial were the cause of this trend, the differences in m values between variants would correlate to some extent with the differences in hydrophobic core residue volume among the variants, but they do not. An alternative explanation involves destabilization of any residual structure in the unfolded state (Shortle et al., 1990). Consistent with this interpretation, we have observed a significant amount of CD signal for unfolded WT ubiquitin at high temperature but not at low temperature, which we cautiously interpret as cold denaturation of residual structure in the unfolded state (unpublished data). These results are supported by the observation of a cold denaturing early folding intermediate of ubiquitin by Roder and coworkers (Khorasanizadeh et al., 1993). If the mutations have destabilized such residual structure in the denatured state of the variants, this could also account for the larger m values. Given the highly conservative nature of these mutations, such an explanation requires that residual structure in the unfolded state is determined by the specific identity and not just the hydrophobic pattern of the residues in the hydrophobic core of ubiquitin.

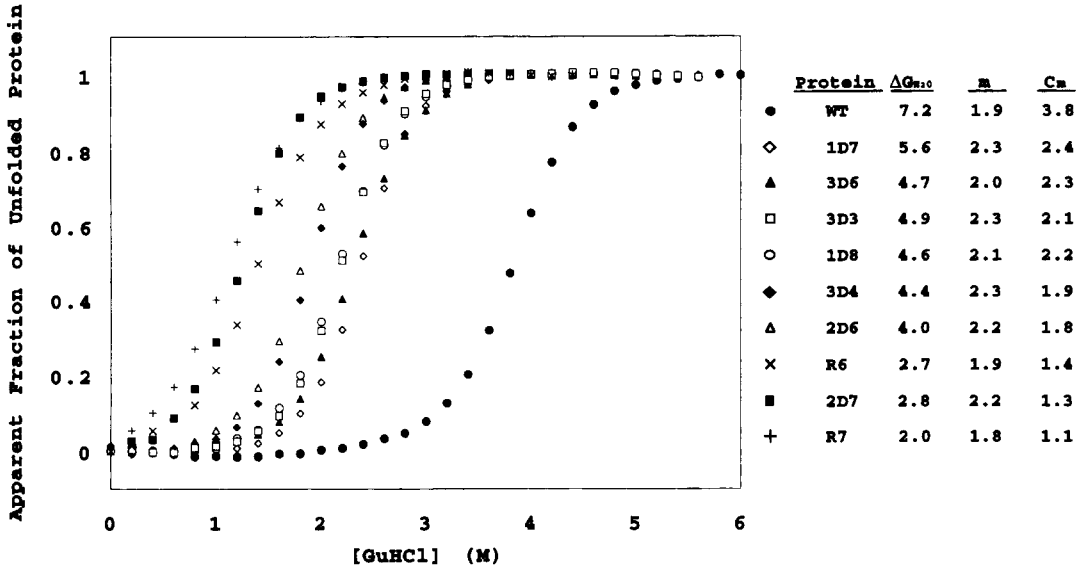

Fig. 4. Stability curves for the 10 ubiquitin proteins as measured by GuHCl-induced denaturation. Data were acquired by monitoring the change in CD ellipticity at 222 nm. Shown is the apparent fraction of unfolded protein, obtained after curve fitting of the original data. The adjacent table presents the thermodynamic parameters obtained from the fits: the Gibbs free energy for unfolding in the absence of denaturant, ΔG_{H_2O}, in kcal mol^{-1}, the slope of the curve in the folding transition, m, in kcal mol^{-1} M^{-1}, and the concentration of denaturant at the midpoint of the transition, C_m, in M.

ANS binding

A common feature of de novo designed proteins is that they lack well-packed tertiary structures and in this respect resemble the molten globule acid states of natural proteins (Kuwajima, 1989). A qualitative measure for good packing is the accessibility of the core to the hydrophobic dye 8-anilino-1-naphthalenesulfonic acid (ANS) (Mulqueen & Kronman, 1982; Semisotnov et al., 1991). These data are shown in Figure 5 along with the fluorescence spectrum of ANS in the presence of the acid state form of ribonuclease H (Dabora & Marqusee, 1994), used here as a positive control. In contrast to ribonuclease H, all ubiquitin variants do not bind ANS, suggesting they are more well-packed than many previously designed proteins. This is particularly significant in the case of the random control proteins which do not bind ANS despite their marginal stability.

1D NMR spectra

A qualitative probe of conformational uniqueness is the level of dispersion in NMR spectra. De novo designed proteins generally have poorly dispersed spectra (Betz et al., 1993), presumably because increased structural disorder and dynamics leads to averaging of chemical shifts close to random coil values. We therefore collected one-dimensional ^{1}H NMR spectra for all nine variants and WT ubiquitin. These spectra are shown in Figure 6, along with the WT protein in 6 M GuHCl as an example of dispersion in the unfolded state. The chemical shift dispersion of all ubiquitin variants, including the class II proteins, is comparable to that of the WT protein. This is most evident upon comparison of the spectra with that of the unfolded WT protein, and indicates that all of the ubiquitin variants, regardless of stability, have significant structural order. Several of the variants, particularly the two control proteins, have slightly broadened peaks in some regions of their spectra. This result may reflect conformational exchange, possibly indicating subtle differences in the degree of order/dynamics among the proteins. However, given the solubility problems observed in preparing these NMR samples (see Materials and methods), this broadening could also be due to nonspecific protein aggregation.

The dispersive nature of the spectra is emphasized by the presence of both downfield-shifted and upfield-shifted resonances. Several spectra, including 1D7, 1D8, 3D3, 3D4, and 3D6, have resonances which are further downfield than those in the WT spectrum. This shift is particularly dramatic in the 1D7 spectrum. Figure 6 also shows the presence of upfield-shifted resonances in all of the spectra. Such dramatic upfield shifts are typically attributed to ring currents from nearby aromatic residues. These results fur-

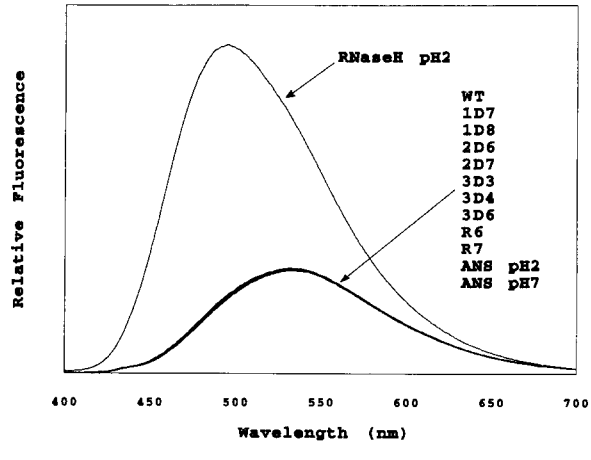

Fig. 5. Fluorescence emission spectra of the ubiquitin proteins in the presence of ANS. The lower intensity spectrum represents all nine ubiquitin variants and WT in the presence of ANS at pH 7.0, as well as ANS alone at pH 7.0 and pH 2.0. The higher intensity spectrum represents the fluorescence of ANS in the presence of *E. coli* ribonuclease H in its acid state at pH 2.0 (Dabora & Marqusee, 1994).

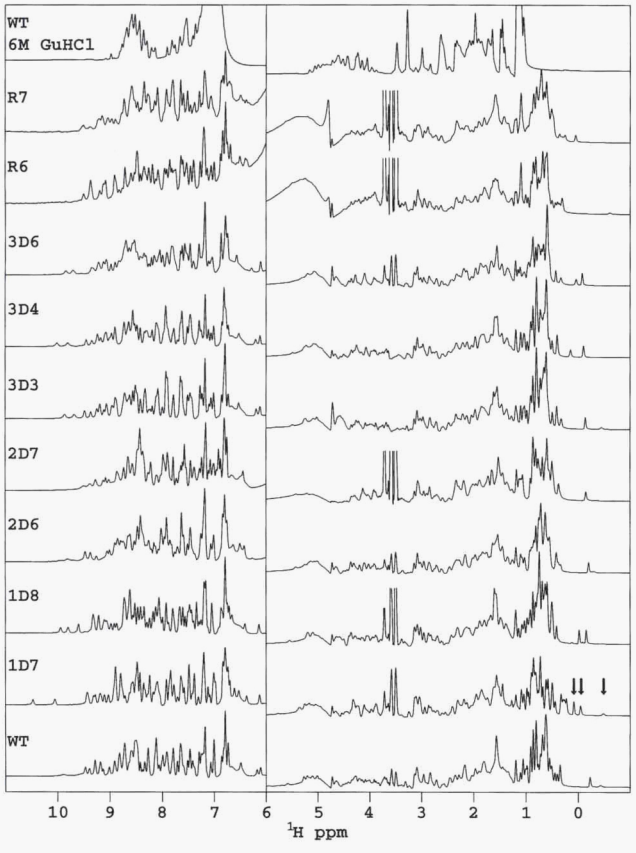

Fig. 6. NMR spectra of the all 10 ubiquitin proteins under native conditions, and the WT protein under unfolded conditions in 6 M GuHCl. All spectra were scaled equivalently, and the region downfield of 6.0 ppm was scaled to 3.7 times that of the upfield region. The broad resonance at approximately 7 ppm in the unfolded WT spectrum is due to GuHCl, and the three starred peaks, present in several of the spectra, are due to a small glycerol contaminant. The fact that the class II samples were less concentrated resulted in the broad, distorted water peak after convolution, as well as more prominant contaminant peaks, which were truncated for clarity. The arrows indicate the three upfield-shifted 1D7 peaks referred to in the text.

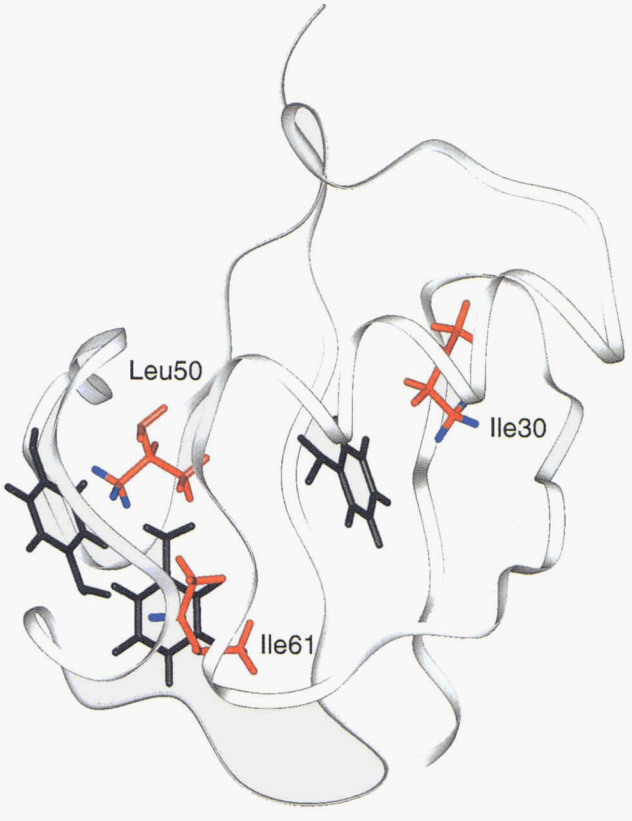

Fig. 7. Predicted core structure for the 1D7 ubiquitin design showing the Ile30-Phe26 interaction, and the two WT-like Leu50–Tyr59 and Ile61–Phe45 interactions. Aliphatics are shown in red and aromatics are shown in black. The upfield-shifted protons, namely the δ-CH_3 protons of Ile30, the δ-CH_3 protons of Leu50, and the γ-CH proton of Ile61, are shown in blue. The predicted structure is from ROC*. The figure is displayed using the program INSIGHT II (Biosym Technologies, San Diego, California).

ther support the presence of relatively well-ordered core packing in all of the ubiquitin variants including the random controls.

The two upfield peaks in the WT spectrum at -0.22 ppm and -0.43 ppm, respectively, are assigned to the δ-CH_3 of Leu50, which is packed against the aromatic ring of Tyr59, and one of the two γ-CHs of Ile61, which is packed against the ring of Phe45 (Di Stefano & Wand, 1987). We are in the process of solving the structure of 1D7 by NMR and currently have this protein fully assigned (E.C. Johnson, G.A. Lazar, J.R. Desjarlais, T.M. Handel, in prep.). The three upfield-shifted protons in the 1D7 spectrum at 0.09 ppm, -0.05 ppm, and -0.48 ppm, respectively, are assigned to the δ-CH_3 of Leu50, the δ-CH_3 of Ile30, and one of the two γ-CHs of Ile61. Figure 7 shows selected side chains in the predicted 1D7 model that rationalize these upfield-shifted peaks. In agreement with the NMR data, the model preserves the two WT-like aliphatic–aromatic interactions: Leu50 is packed against Tyr59, and Ile61 against Phe45. In this model Ile30 is packed against the mutant residue, Phe26, providing an explanation for the Ile30 shift from 0.65 ppm in the WT protein (Di Stefano & Wand, 1987) to -0.05 ppm in the 1D7 protein. High resolution structural information will ultimately verify if these predictions are correct.

Predictive ability of ROC

Having stability data for 10 hydrophobic core sequences of ubiquitin provides an excellent opportunity to evaluate and improve the predictive ability of ROC. Our strategy involved maximizing the correlation between experimentally determined stabilities and predicted energies for all 10 proteins, and maximizing the number of correctly predicted WT rotamers. The two main variables that were modified to improve the program were the potential function parameters and the library of side-chain conformers. For the potential function, we evaluated the predictions of ROC using three different parameter sets (see Materials and methods). For the rotamer library, we explored the use of a customized rotamer library for ubiquitin versus a standard rotamer library obtained from the PDB (Tuffery et al., 1991). We have also developed a new version of ROC which incorporates the non-bonded parameters from the AMBER/OPLS potential (Weiner et al., 1984; Jorgensen & Tirado-Rives, 1988) or the AMBER95 potential (Cornell et al., 1995), and includes side-chain torsional potentials taken from those force fields. For the AMBER versions of ROC, a nearly continuous set of rotamers was used instead of a library of rotamers, allowing a much finer search of side-chain dihedral angles. The results of comparing the different strategies are shown in Table 1. The best results are obtained when ROC is used with the AMBER95 potential and a continuous rotamer search. We therefore chose this set

Table 1. *Comparison of ROC with various parameter sets and rotamer searches*[a]

	Rotamers	
Parameters	Custom library	Standard library
1	0.69 (1,1)	0.72 (1,2)
2	0.79 (1,2)	0.55 (1,1)
3	0.81 (2,2)	0.26 (0,3)
	Continuous search	
AMBER/OPLS	0.63 (1,1)	
AMBER95	0.82 (1,1)	

[a]The first number represents the correlation between predicted and experimental energies. The two numbers in parentheses represent the number of χ_1 and χ_2 rotamers, respectively, which are incorrectly predicted for the WT protein.

of conditions for the final version of ROC, and refer to this program as ROC*.

The final correlation for ROC* is shown in Figure 8. The program predicts the WT core to be the most stable, and clearly distinguishes the control proteins from the designs. However, 2D7, which has the third highest experimental energy (-2.8 kcal mol^{-1}), is incorrectly predicted to be quite stable by ROC* and all other versions. It is an obvious outlier, and when excluded the correlation improves to 0.91. We offer two possible explanations for this problem. First, 2D7 has two extra methylene groups relative to WT, more volume than any other variant. Additional interactions provided by these methylenes should result in a more favorable predicted energy but this may be offset by strain or other steric costs due to adding more volume to the core. Apparently the potentials do not accurately represent the balance between these opposing interactions. Alternatively, the phenylalanine in 2D7 may be the cause of the overestimation in stability. The two variants that contain an extra phenylalanine (1D7 and 2D7) are predicted to be significantly more stable than they are. In other studies using AMBER it has been observed that packing interactions of phenylalanines are predicted to be more energetically favorable than those of aliphatics (Hurley et al., 1992), and our results may reflect the same problem.

Figure 9 shows the core side chains of the predicted WT structure compared with those from the X-ray crystal structure. 13 out of 14 χ_1 and 10 out of 11 χ_2 WT rotamers are predicted correctly. χ_1 of Ile13, one of the two incorrectly predicted conformers, is in a non-standard rotamer. However, in no case is this rotamer predicted when customized libraries are used. The other incorrectly predicted rotamer, χ_2 of Leu69, is at the C-terminus of the protein and may simply be insufficiently constrained for ROC to correctly predict it.

Discussion

The purpose of this study has been to investigate the role of hydrophobic core packing on the structure, stability, and uniqueness of proteins. Our approach involves the development of systematic computational methods to predict core packing arrangements, coupled with experimental investigation of these predictions. The most significant results of this study are the wide range of protein stabilities that result from alternative hydrophobic cores of ubiquitin, and the fact that, regardless of stability, all proteins adopt a relatively unique folded state.

It is important to point out those factors which are not responsible for the greater stability of the WT protein relative to the designs, and the greater stability of the designs relative to the

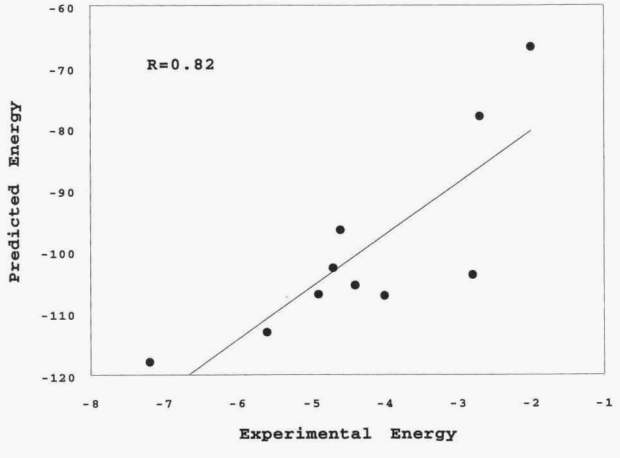

Fig. 8. Plot of predicted versus experimental stabilities of the ten ubiquitin proteins using ROC*. The line represents the best linear fit to the data using Kaleidagraph, and the correlation value, R, for the data is presented. Experimental energies are the Gibbs free energy of folding, ΔG_{H_2O}, in kcal mol^{-1}, obtained from the GuHCl denaturation data, and calculated energies are in the arbitrary units of the program and have no comparative meaning.

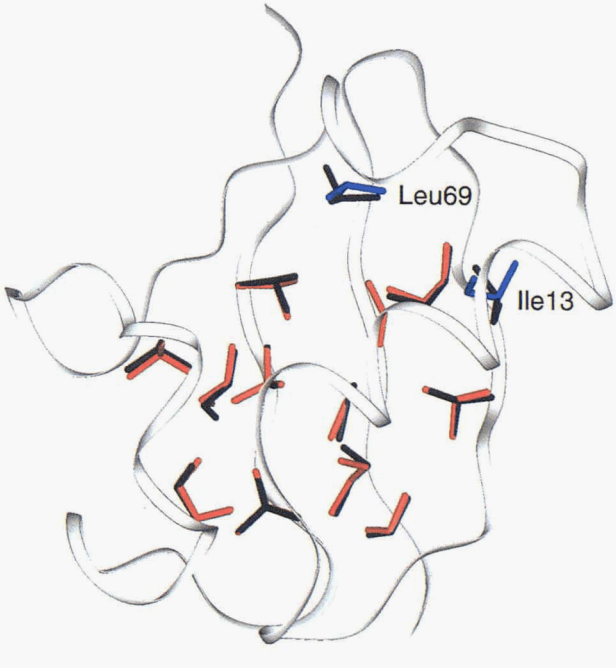

Fig. 9. Structural comparison of the WT ubiquitin core predicted by ROC* to that from the X-ray structure (Vijay-Kumar et al., 1987). X-ray structure core side chains are in black and predicted core side chains are in red. Side chains which have incorrectly predicted rotamers in the predicted structure are shown in blue. Hydrogens are not shown for clarity. The figure is displayed using the program INSIGHT II (Biosym Technologies, San Diego, California).

random controls. First of all, the number of substitutions relative to the WT sequence does not correlate with stability. For example, R6 and R7 have comparable numbers of mutations as the designs. Secondly, core volume does not appear responsible for differences in stability. R7 has essentially the same volume as several of the designs as well as the WT protein. Additionally, if volume of the core were solely responsible for stability, then 2D7 would be the most stable. Finally, stability cannot be attributed to secondary structure propensity because R6 has by far the highest secondary structure propensity score of any of the proteins (Chou & Fasman, 1978), and yet is one of the least stable. In the absence of a major influence from these factors, we believe this study makes an undeniable argument for the importance of core packing to the stability of a protein. A stark example is made by comparing WT, 3D6, and R7. Figure 2 reveals that these three proteins contain not only the same volume, but exactly the same composition. The only difference among them is the packing of their core residues, and yet these three proteins have stabilities spanning a range of greater than 5 kcal mol^{-1}. Such a result, independent of a quest for uniqueness, justifies the consideration of specific packing interactions in de novo protein design.

We conclude that variation in stability of these 10 ubiquitin proteins is due predominantly to differences in packing and/or conformational strain. This interpretation is consistent with the clear correlation found between the energies calculated with ROC* and experimentally determined stabilities. In a simplified view of these effects on the energetics, the conformational strain lies in the use of different side-chain rotamers to achieve optimal packing of the core. The lowest energy structures are able to efficiently pack while using a set of near optimal rotamers, while those which are destabilized have to rely on the use of rotamers which are non-optimal. In reality, this rotamer strain may translate to conformational strain which involves adjustment of the backbone geometry. However, backbone adjustment is not allowed in our current version of ROC.

Contrary to the stability differences among the 10 ubiquitin proteins, all of them, even the randomized core controls, appear to possess a fairly high degree of conformational specificity. All proteins have well-dispersed NMR spectra, and none bind ANS. Even the randomly designed variants R6 and R7 have very well-dispersed NMR spectra, despite the fact that they are destabilized by approximately 5 kcal mol^{-1}. A related example of this effect has been observed in the case of a T4 lysozyme variant which is active and cooperatively folded although a significant portion of the core has been replaced by methionine residues (Gassner et al., 1992). These results are quite striking, and are at odds with the view that hydrophobic packing interactions play a dominant role in determining native state specificity and uniqueness. In an extreme interpretation, these results would seem to suggest that energetically favorable packing interactions in the hydrophobic core do not contribute to structural specificity. However, we prefer the alternative explanation that the uniqueness of natural proteins is overdetermined by a combination of interactions involving core and noncore residues. This view does not necessarily extend to de novo designed proteins, which typically lack the many specific tertiary interactions of native proteins. In fact, specific packing in the hydrophobic core may be necessary for designing conformational uniqueness into de novo proteins. Whether specific packing is sufficient for uniqueness is as yet an unanswered question.

Does core packing affect the fold? There are some differences in the structures of the proteins, indicated by the anomalous class II CD spectra and the variability in solubility among the 10 ubiquitin proteins. An example of such structural differences is shown by 3D6 and 3D3. Both have comparable volume, and even comparable stability, yet 3D3 appears to be as soluble as the WT protein while 3D6 readily aggregates at physiological pH. Another example is given by R6 and 2D6, two proteins with essentially the same core volume, yet different CD spectra. The fact that the stability of R6 is -2.7 kcal mol^{-1}, and that addition of glycerol has absolutely no effect on its CD spectrum indicate that, although R6 is destabilized relative to 2D6, it is still completely folded under the conditions of these experiments. These structural differences among the proteins are unlikely due to major changes in their folds. If that were the case, we would not expect the alternative conformations to be as well-structured as the NMR and ANS results indicate. Instead, these differences are most likely the result of minor adjustments of the protein backbone due to differences in core packing, a well-recognized response to mutation (see below). These results support the view that packing interactions in the hydrophobic core are important determinants of local structure in proteins but do not determine a protein's fold. Exceptions to this have been published, and involve multimeric proteins which change their oligomeric state depending on the nature of core residues (Harbury et al., 1993; Munson et al., 1996). It is also possible that packing interactions may play a greater role in determining the global fold in de novo designed proteins, where non-core residues may be less optimally chosen.

It is important to note that, in an effort to design novel packing arrangements in the core of ubiquitin, we chose to study variants with a significant number of core mutations from WT. This, along with the extreme stability of the WT protein, may be the reason we observe such a large stability gap between our designs and the WT protein. In fact, the four lowest energy output sequences from ROC* have only one to four changes from the WT protein. The inability to find disparate hydrophobic cores of ubiquitin as stable as the WT is in contrast to our previous experiment with 434 cro, where variants designed with five to eight substitutions had stabilities comparable to that of the WT protein (Desjarlais & Handel, 1995). We propose two possible reasons for this. The first is the predominance of β-sheet structure in ubiquitin. Alternative core packing arrangements may be more difficult to accommodate in β-sheet structures than in α-helical ones due to the non-local nature of the stabilizing interactions in β-sheets. Secondly, ubiquitin is unusually stable for its size. Its stability may have evolved to resist degradation in the continual presence of the proteolytic machinery. Regardless of the reason, it appears that evolution may have designed by far the best hydrophobic core for the ubiquitin fold.

In regard to ROC's effectiveness for the design of hydrophobic cores, the most encouraging result is that, with one exception, all designs are significantly more stable than randomized controls. Furthermore, all but one of the designs appear to be more structurally similar to the WT protein than randomized controls. Although in the present study we have characterized only two random core variants, other reports have shown that random core design typically leads to a significant loss of stability and/or activity. For example, Fersht and colleagues find that 77% of barnase variants with randomized hydrophobic cores do not retain enzymatic activity (Axe et al., 1996). We conclude from these observations that ROC, or similar packing algorithms, will generally be superior to approaches that utilize random sequence selection for the design of hydrophobic cores.

The large decrease in stability of all of the designs relative to the WT protein indicates that a complete understanding of core packing, at least in terms of predictive ability, has not been fully realized. Our examination of different parameter sets and different methods of sampling side-chain orientations identified the AMBER95/continuous rotamer search as the best set of conditions for ROC, and we therefore chose this version, referred to as ROC*, as the final version of the program. ROC* does identify the WT sequence as the most stable protein of the current test set. Overall, the correlation between calculated versus experimental stabilities using this version is quite good ($R = 0.82$). ROC* also correctly predicts 13 out of 14 χ_1 and 10 out of 11 χ_2 rotamers for the WT protein, of which the incorrect χ_1 is a nonstandard rotamer and the incorrect χ_2 is poorly constrained.

There have been a number of studies that use existing experimental data to evaluate the predictive ability of computational methods. The emphasis in our study on trying to design proteins with stabilities comparable to or greater than WT provides a particularly stringent test of our methods. The difficulty of improving on a highly evolved core sequence almost guarantees that these designed variants will be less stable than the WT, leading to repeated reassessment of the computational methods. Indeed, ROC2 and ROC3 initially resulted in near perfect correlations of calculated versus experimental stabilities until they were used to generate more test sequences. At this point, the ROC* algorithm, using the AMBER95 potential, results in the best overall correlation. While this is quite promising, until several variants are designed using this version its true potential for the prediction of relative stabilities remains speculative.

The lack of complete success with our core designs, as well as the imperfect correlation with our final program, may be attributed to a number of terms which are not included in the predictions such as side-chain entropy effects, surface area burial, and backbone flexibility. Because of the strong correlation observed without including these effects, and the possibility that these effects may not be independent, we believe that these terms should only be included and optimized when a much larger data set is available. Of these terms, we recognize that a significant limitation of the current work is the assumption of a fixed backbone. It is now well established that protein backbones can and do shift to accommodate mutations in the hydrophobic core (Eriksson et al., 1992, 1993; Baldwin et al., 1993; Lim et al., 1994). We know of one other protein design program which takes into account protein backbone movement (P.B. Harbury, J.J. Plecs, B. Tidor, T. Alber, P.S. Kim, in prep.). Only when such backbone adjustment is explicitly treated can we hope to achieve a better predictive understanding of the structure and energetics of hydrophobic core variants. Considering such backbone shifts will also be important for the application of the current methods to de novo protein design. We have modified ROC* to account for backbone relaxation, and are currently carrying out a similar investigation of its use for the prediction of hydrophobic core packing (J.R. Desjarlais & T.M. Handel, in prep.).

Materials and methods

Program description

Details of the core evolution program ROC have been described previously (Desjarlais & Handel, 1995). The energy function used for ROC is based on a Lennard-Jones potential, defined as follows: For two atoms i and j:

$$E = \left[\left(\frac{R}{d}\right)^{12} - 2\left(\frac{R}{d}\right)^{6}\right]\sqrt{\epsilon_i \cdot \epsilon_j} \quad (1)$$

R_i and R_j are the van der Waals radii which are used in the following combining rule:

$$R = 2\sqrt{R_i \cdot R_j} \quad (2)$$

except for the AMBER95 parameter set, which uses the combining rule:

$$R = R_i + R_j \quad (3)$$

ϵ_i and ϵ_j are the well depths for atoms i and j, and d is the distance between the two atoms. We attempted to improve the program by changing the atomic van der Waals radii and the well depths of this potential energy function. The parameters were modified twice, totaling three sets. The first set of parameters is that used with the initial version of ROC (Desjarlais & Handel, 1995) derived from Hagler (Dauber & Hagler, 1980). The second and third parameter sets were derived in two stages, and for both sets the parameters were divided into three classes: intra-side-chain interactions, side chain-backbone interactions, and side-chain–side-chain interactions. For generation of the second set, intra-side-chain and side-chain–backbone parameters were derived by computationally searching for individual radii and well depths that resulted in an optimal match between calculated side-chain rotamer preferences and probabilities obtained from the PDB (McGregor et al., 1987). A Monte Carlo search procedure using moves that modified individual atomic parameters was performed to minimize the difference between calculated side-chain rotamer distributions (in an ideal alpha-helix) and the distributions derived from the PDB. Side-chain–side-chain parameters were derived by searching for radii and well depths which resulted in prediction of the crystal structure core to be superior to many alternative decoy core structures that were generated with ROC. This was done simultaneously for a set of six different proteins. The final atomic parameters were those which resulted in the lowest squared sum of native core rankings over the six proteins and corresponding decoy sets. The proteins and pdb files used were ubiquitin (1UBI), 434 cro (2CRO), major cold shock protein (1MJC), lambda repressor (1LMB), basic fibroblast growth factor (2FGF), and thioredoxin (2TRX). Derivation of the third parameter set was similar in nature, except instead of modifying individual atomic parameters, a common base set of parameters was globally scaled depending on the interaction type. Scaling factors for each interaction type were derived using the same criteria as for the second set, using monte carlo optimization of the global scaling factors. Van der Waals parameters for the first and second parameter sets are shown in Table 2. The third parameter set uses the parameters from the first as a basis set with global radii scaling factors of 1.0 for intra-side-chain interactions, 0.84 for side-chain-backbone interactions, and 0.96 for side-chain–side-chain interactions.

Independent of the derived parameter sets described above, we also created versions of ROC which use the AMBER/OPLS united atom parameters (Weiner et al., 1984; Jorgensen & Tirado-Rives, 1988) or the AMBER95 potential (Cornell et al., 1995). The AMBER95 potential was modified slightly so that methyl group hydrogens have zero radii and methyl carbons were assigned the

Table 2. *Atomic parameters*[a]

	Parameter set					
	1		2			
			sc-bb		sc-sc	
	R	ϵ	R	ϵ	R	ϵ
Hydrogen	1.38	0.038	1.61	0.246	1.49	0.004
Methyl hydrogen	0.00	0.000	0.00	0.000	0.00	0.000
Carbon	2.03	0.148	2.16	0.233	2.72	0.048
Carbonyl carbon	2.18	0.039	2.06	0.093	—	—
Methyl carbon	2.19	0.160	2.12	0.049	2.26	0.160
Aromatic carbon	2.00	0.110	2.29	0.117	2.24	0.043
Nitrogen	1.97	0.167	1.50	0.003	—	—
Oxygen	1.61	0.228	0.52	0.000	—	—
Sulfur	2.19	0.160	2.19	0.160	2.40	0.216

[a]Radii, R, are in Å, and well depths, ϵ, are in arbitrary units. Dashes indicate that those parameters do not apply for the residues involved in these interactions.

united atom values from the AMBER/OPLS potential. Partial charges from the AMBER95 potential were not included. For these versions of ROC, a side-chain torsional potential is also included and the one to four nonbonded interactions are scaled by a factor of eight for the AMBER/OPLS potential and a factor of five for the AMBER95 potential.

Design of ubiquitin variants using ROC

Three versions of the program, ROC1–ROC3, were used to design the variants described in the present paper. The number in the name of the designs designates which version of the program was used for that particular variant. The rotamer library used for all designs was a customized library based on the ubiquitin backbone (Desjarlais & Handel, 1995). Each design trial used a custom rotamer library generated with its respective parameter set, and each library was additionally supplemented with a statistically derived standard rotamer set (Tuffery et al., 1991). ROC was typically run using 100 supercycles of 500 rounds of genetic algorithm.

Evaluating the predictability of ROC

Calculation of energies for each ubiquitin protein under various sets of conditions was carried out by allowing each core residue to mutate only its rotamers. The output from such a run consists of a list of energetically favorable cores that are identical to the input sequence in residue identity but different in their rotamers and calculated energies. In cases where a custom rotamer library was used, each library was generated with its respective parameter set and supplemented with the standard rotamer set. In cases where a standard library was used, the χ_1 and χ_2 angles of each standard rotamer were incremented at 10° over a deviation of ±20° from the standard rotamer set. Finally, in the case where a "continuous" rotamer search was used, all dihedral angle values within 50° of each standard rotamer dihedral value were allowed, at 5° increments. Under all conditions, ROC was run using 10 supercycles of 300 rounds of genetic algorithm for each protein. For situations in which rotamer libraries were constructed at 10° increments, each output was subsequently refined by rerunning the core sequence through ROC using a new rotamer library consisting of a spread of rotamers around those of the lowest energy output structure (as done with the standard library). This library was residue specific for each sequence, and was generated by varying both the χ_1 and χ_2 angles at 3° increments over a deviation of ±15° from the rotamer angles for that residue in the best output structure. Because it searches an approximately continuous set of rotamers, ROC* did not require such refinement.

Correlations were obtained by plotting the energy of the best output core for each set of conditions versus the experimental free energy. These data were fit to a linear equation using the program Kaleidagraph (Synergy Software, Reading, Pennsylvania), and the correlation in Table 1 represents the goodness of fit for this line. The number of correctly predicted WT rotamers was also determined by comparing the rotamers of the best WT output core after refinement for each set of conditions with those of the X-ray crystal structure (Vijay-Kumar et al., 1987). A rotamer was considered to be correctly predicted if it was within ±40° of the dihedral angle given by the X-ray structure. If a χ_1 rotamer was incorrect, an incorrect χ_2 rotamer for the same residue was not counted.

Construction of ubiquitin variants

The plasmid pNMHUB, containing a synthesized WT human ubiquitin gene (Ecker et al., 1987), was a generous gift from Dr. Tauseef R. Butt. The ubiquitin gene was subcloned into a p'AED vector, a pUC-based plasmid that utilizes a T7 expression system (Doerring, 1992). In addition, because of an Arg to Lys mistranslation problem (Calderone et al., 1996), the four codons for arginine were mutated to those preferred in *E. coli*. Ubiquitin mutants were constructed using site-directed mutagenesis, using the removal of unique restriction sites to select and screen for desired mutations. The entire ubiquitin gene for each mutant was sequenced to confirm the presence of the mutations and the fidelity of the sequence.

Expression, purification, and sample preparation of proteins

Ubiquitin proteins were expressed in a BL21/plysS strain (Studier et al., 1990) of *E. coli* by inducing cultures at mid-log growth phase with 0.5 mM IPTG. Proteins were purified by one of two procedures, depending on whether the protein partitioned into the lysis supernatant or lysis pellet. All class I proteins were purified by the supernatant procedure. Briefly, cells were collected by centrifugation and lysed, and the supernatant was dialyzed into 50 mM sodium acetate, pH 5.0, 5 mM EDTA. The sample was centrifuged, and the supernatant was loaded onto a Fast Flow SP Sepharose column (Pharmacia) and eluted with a linear salt gradient from 0 to 1 M NaCl in the same buffer. Fractions containing ubiquitin were pooled and concentrated. The sample was then run over a Superdex 75 sizing column (Pharmacia) using an FPLC in 50 mM potassium phosphate, pH 7.0, 0.5 M NaCl, 5 mM EDTA. Fractions containing ubiquitin were pooled and further purified on a Shimadzu HPLC using a reversed-phase C8 semipreparative column (Vydac). The ubiquitin sample was then lyophilized.

Class II proteins were purified from the lysate pellet. Cells were centrifuged and lysed, and the pellet was resuspended in 50 mM sodium acetate, pH 5.0, 6 M urea. The sample was run over the Fast Flow SP Sepharose column as indicated above except that all

buffers contained 6 M urea. Fractions were pooled and dialyzed into water. The precipitated sample was centrifuged, and the pellet containing ubiquitin was resuspended in 6 M GuHCl and purified over the HPLC reversed-phase column as above.

It is not possible to solubilize lyophilized ubiquitin directly into aqueous solution. All experimental samples were therefore prepared by dissolving lyophilized protein in 6 M GuHCl, and then refolding by dialysis into several volumes of buffer or doubly-distilled water.

All ubiquitin protein samples were determined to be greater than 95% pure as judged by the presence of a single band on a coomassie stained polyacrylamide gel, and the mass of each protein was confirmed by mass spectrometry. Concentrations of all ubiquitin samples were determined by measuring the absorbance at 280 nm using a molar extinction coefficient of 1280 M^{-1} cm^{-1} (Gill & von Hippel, 1989).

Purified *E. coli* ribonuclease H was a generous gift from Dr. Susan Marqusee.

CD spectroscopy and denaturation experiments

All CD experiments were performed on an Aviv 62DS CD spectrometer. CD spectra were recorded at 25 °C in 10 mM potassium phosphate, pH 7.0. Protein concentration was 10 μM, and the cell path length was 1 cm. The signal was scanned every 0.5 nm in the range of 200 nm to 300 nm, with an averaging time of 5 s per nm.

GuHCl-induced denaturation experiments were carried out using protein concentrations of 10 μM in 10 mM potassium phosphate, pH 7.0 in a 1 cm path length cell. The CD signal at 222 nm was monitored in kinetics mode with an averaging time of 1 s, where each data point is the average of 200 s. Data points were taken at 0.2 M increments of GuHCl concentration.

Assuming a two-state folding transition and a linear dependence of the free energy on denaturant concentration (Schellman, 1978), the data were fit to the following equation using the program Kaleidagraph (Synergy Software, Reading, Pennsylvania):

$$\theta_{obsd} = \frac{(S_N[\text{GuHCl}] + I_N) - (S_U[\text{GuHCl}] + I_U)}{\left[1 + \exp\left(\frac{(-\Delta G_{H_2O} + m[\text{GuHCl}])}{RT}\right)\right]} + (S_U[\text{GuHCl}] + I_U) \quad (4)$$

where θ_{obsd} is the observed ellipticity, S_N, I_N, S_U, and I_U are the slopes and y intercepts of the native and unfolded baselines, respectively, [GuHCl] is the molar concentration of guanidine-hydrochloride, R is the gas constant 1.98 cal mol^{-1} K^{-1}, T is the temperature in Kelvin, ΔG_{H_2O} is the Gibbs free energy for unfolding in the absence of denaturant, and m is the slope of the curve in the unfolding transition. Data were converted to the apparent fraction of unfolded protein (F_{app}) using the equation:

$$F_{app} = \frac{(\theta_{obsd} - \theta_N)}{(\theta_U - \theta_N)} \quad (5)$$

where θ_N and θ_U are the ellipticities of the native and unfolded forms, respectively, obtained from the above-fitted baselines.

ANS binding fluorescence

Fluorescence data were collected on a Perkin Elmer MPF-44B Fluorescence Spectrophotometer. Samples contained 250 μM ANS and 1 μM protein in 10 mM potassium phosphate, pH 7.0. Data were collected at 25 °C by exciting at 380 nm and monitoring emission every 1 nm from 400 nm to 700 nm with an averaging time of 1 s. Excitation and emission bandwidths were both 6 nm.

NMR

NMR spectra were collected on a Bruker DMX 600 spectrometer at 25 °C and processed using the program FELIX 1.1 (Hare Research). Except for the denatured WT sample, all samples were prepared in 10 mM deuterated sodium acetate, pH 5.0, 90% H_2O–10% D_2O. The denatured WT protein sample was prepared in 25 mM deuterated sodium acetate, pH 5.0, 90% H_2O–10% D_2O, 6 M GuHCl. The concentration of all class I proteins was 1.0 mM. At this concentration, class II proteins aggregated; however, adequate protein concentrations for one-dimensional spectra could be obtained after centrifugation. Final protein concentrations for these samples were 0.2 mM for both R6 and R7, and 0.7 mM for 2D7. The unfolded WT protein sample was 3.7 mM protein. One-dimensional ^1H spectra were collected with 1,024 complex points using low power presaturation to suppress the water signal. Each free induction decay was convolved with a sine function to remove residual water, apodized with a 75°-shifted sinebell, and zero-filled to 2 K points prior to Fourier transformation. Chemical shifts were referenced to sodium 2,2-dimethyl-2-silapentane-5-sulfonate at 0 ppm and 25 °C (Wishart et al., 1995). These data were taken under different conditions than those of the published assignments (Di Stefano & Wand, 1987; Weber et al., 1987), and therefore chemical shifts are slightly different.

Acknowledgments

T.M.H. is an NSF Young Investigator and J.R.D. is a fellow of the Jane Coffin Childs Memorial Fund for Medical Research, fellowship number 61-948. This material is based upon work supported under an NSF Graduate Research Fellowship awarded to G.A.L., and is supported in part by an NSF grant, MCB9458201. We thank Eric Johnson for assisting with molecular biology, Tauseef Butt for providing the ubquitin clone, Dave King for mass spectrometry, Alex Glazer for the use of his fluorimeter, and Susan Marqusee both for the use of her CD spectrometer and for allowing G.A.L. to work in her lab during the initial stages of his graduate study.

References

Axe DD, Foster NW, Fersht AR. 1996. Active barnase variants with completely random hydrophobic cores. *Proc Natl Acad Sci USA 93*:5590-5594.

Baldwin EP, Hajiseyedjavadi O, Baase WA, Matthews BW. 1993. The role of backbone flexibility in the accommodation of variants that repack the core of T4 lysozyme. *Science 262*:1715–1718.

Betz SF, Raleigh DP, DeGrado WF. 1993. De novo protein design: From molten globules to native-like states. *Current Opinion in Structural Biology 3*:601–610.

Briggs MS, Roder H. 1992. Early hydrogen-bonding events in the folding reaction of ubiquitin. *Proc Natl Acad Sci USA 89*:2017–2021.

Calderone TL, Stevens RD, Oas TG. 1996. High-level misincorporation of lysine for arginine at AGA codons in a fusion protein expressed in Escherichia coli. *J Mol Biol 262*:407–412.

Chou PY, Fasman GD. 1978. Empirical predictions of protein conformation. *Ann Rev Biochem 47*:251-276.

Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. 1995. A second-generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc 117*:5179–5197.

Dabora JM, Marqusee S. 1994. Equilibrium unfolding of *Escherichia coli* ribonuclease H: Characterization of a partially folded state. *Protein Sci 3*:1401–1408.

Dahiyat BI, Mayo SL. 1996. Protein design automation. *Protein Sci 5*:895–903.

Dauber P, Hagler AT. 1980. Crystal packing, hydrogen bonding, and the effect of crystal forces on molecular conformation. *Accounts of Chemical Research 13*:105–112.

Desjarlais JR, Handel TM. 1995. De novo design of the hydrophobic cores of proteins. *Protein Sci 4*:2006–2018.

Di Stefano DL, Wand AJ. 1987. Two-dimensional ^{1}H NMR study of human ubiquitin: A main chain-directed assignment and structure analysis. *Biochemistry 26*:7272–7281.

Dill KA. 1990. Dominant forces in protein folding. *Biochemistry 29*:7133–7155.

Doerring DS. 1992. *Functional and structural studies of a small f-actin binding domain*. Cambridge, Massachusetts: MIT Press.

Ecker DJ, Butt TR, Marsh J, Sternberg EJ, Margolis N, Monia BP, Jonnalagadda S, Khan MI, Wever PL, Mueller L, Crooke ST. 1987. Gene synthesis, expression, structures, and functional activities of site-specific mutants of ubiquitin. *J Biol Chem 262*:14213–14221.

Eriksson AE, Baase WA, Matthews BW. 1993. Similar hydrophobic replacements of Leu99 and Phe153 within the core of T4 lysozyme have different structural and thermodynamic consequences. *J Mol Biol 229*:747–769.

Eriksson AE, Baase WA, Zhang XJ, Heinz DW, Blaber M, Baldwin EP, Matthews BW. 1992. Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science 255*:178–183.

Gassner NC, Baase WA, Zhang XJ, Heinz DW, Blaber M, Baldwin EP, Matthews BW. 1992. A test of the "jigsaw puzzle" model for protein folding by multiple methionine substitutions within the core of T4 lysozyme. *Proc Natl Acad Sci USA 93*:12155–12158.

Gill SC, von Hippel PH. 1989. Calculation of protein extinction coefficients from amino acid sequence data. *Analytical Biochemistry 182*:319–326.

Handel TM, Williams SA, DeGrado WF. 1993. Metal ion-dependent modulation of the dynamics of a designed protein. *Science 261*:879–885.

Harbury PB, Zhang T, Kim PS, Alber T. 1993. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science 262*:1401–1407.

Hecht MH, Richardson JS, Richardson DC, Ogden RC. 1990. De novo design, expression, and characterization of Felix: A four-helix bundle protein of native-like sequence. *Science 249*:884–891. [published erratum appears in *Science 249*:4972–4973]

Hecht MH. 1994. De novo design of β-sheet proteins. *Proc Natl Acad Sci USA 91*:8729–8730.

Hellinga HW, Richards FM. 1994. Optimal sequence selection in proteins of known structure by simulated evolution. *Proc Natl Acad Sci USA 91*:5803–5807.

Holland JH. 1992. *Adaptation in natural and artificial systems*. Cambridge, Massachusetts: MIT Press.

Holm L, Sander C. 1991. Database algorithm for generating protein backbone and side-chain co-ordinates from a Cα trace. Application to model building and detection of co-ordinate errors. *J Mol Biol 218*:183–194.

Hurley JH, Baase WA, Matthews BW. 1992. Design and structural analysis of alternative hydrophobic core packing arrangements in bacteriophage T4 lysozyme. *J Mol Biol 224*:1143–1159.

Jorgensen WL, Tirado-Rives J. 1988. The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *J Am Chem Soc 110*:1657–1666.

Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH. 1993. Protein design by binary patterning of polar and nonpolar amino acids. *Science 262*:1680–1685.

Khorasanizadeh S, Peters ID, Butt TR, Roder H. 1993. Folding and stability of a tryptophan-containing mutant of ubiquitin. *Biochemistry 32*:7054–7063.

Khorasanizadeh S, Peters ID, Roder H. 1996. Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. *Nature Struct Biol 3*:193–205.

Kono H, Doi J. 1994. Energy minimization method using automata network for sequence and side-chain conformation prediction from given backbone geometry. *Proteins 19*:244–255.

Kuwajima K. 1989. The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins Struct Funct Genet 6*:87–103.

Lee C, Levitt M. 1991. Accurate prediction of the stability and activity effects of site-directed mutagenesis on a protein core. *Nature 352*:448–451.

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

Lim WA, Hodel A, Sauer RT, Richards FM. 1994. The crystal structure of a mutant protein with altered but improved hydrophobic core packing. *Proc Natl Acad Sci USA 91*:423–427.

Lim WA, Sauer RT. 1991. The role of internal packing interactions in determining the structure and stability of a protein. *J Mol Biol 219*:359–376.

Manning MC, Woody RW. 1989. Theoretical study of the contribution of aromatic side chains to the circular dichroism of basic bovine pancreatic trypsin inhibitor. *Biochemistry 28*:8609–8613.

McGregor MJ, Islam SA, Sternberg MJ. 1987. Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. *J Mol Biol 198*:295–310.

Mulqueen PM, Kronman MJ. 1982. Binding of naphthalene dyes to the N and A conformers of bovine α-lactalbumin. *Arch Biochem Biophys 215*:28–39.

Munson M, Balasubramanian S, Fleming K, Nagi AD, O'Brien R, Sturtevant JM, Regan L. 1996. What makes a protein a protein? Hydrophobic core designs that specify stability and structural properties. *Protein Sci 5*:1584–1593.

Munson M, O'Brien R, Sturtevant JM, Regan LR. 1994. Redesigning the hydrophobic core of a four-helix-bundle protein. *Protein Sci 3*:2015–2022.

Ponder JW, Richards FM. 1987. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J Mol Biol 193*:775–791.

Quinn TP, Tweedy NB, Williams RW, Richardson JS, Richardson DC. 1994. Betadoublet: De novo design, synthesis, and characterization of a beta-sandwich protein. *Proc Natl Acad Sci USA 91*:8747–8751.

Raleigh DP, DeGrado WR. 1992. A de novo designed protein shows a thermally induced transition from a native to a molten globule-like state. *J Am Chem Soc 114*:10079–10081.

Regan L, DeGrado WF. 1988. Characterization of a helical protein designed from first principles. *Science 241*:976–978.

Richards FM. 1977. Area, volumes, packing, and protein structure. *Ann Rev Biophys Bioeng 6*:151–176.

Richards FM, Lim WA. 1993. An analysis of packing in the protein folding problem. *Q Rev Biophys 26*:423–498.

Schellman JA. 1978. Solvent denaturation. *Biopolymers 17*:1305–1322.

Schneider DM, Dellwo MJ, Wand AJ. 1992. Fast internal main-chain dynamics of human ubiquitin. *Biochemistry 31*:3645–3652.

Semisotnov GV, Rodionova NA, Razgulyaev OI, Uversky VN, Gripas AF, Gilmanshin RI. 1991. Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers 31*:119–128.

Shortle D, Stites WE, Meeker AK. 1990. Contributions of the large hydrophobic amino acids to the stability of staphylococcal nuclease. *Biochemistry 29*:8033–8041.

Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol 185*:60–89.

Summers NL, Karplus M. 1989. Construction of side-chains in homology modelling. Application to the C-terminal lobe of rhizopuspepsin. *J Mol Biol 210*:785–811.

Tanaka T, Hayashi M, Kimura H, Oobatake M, Nakamura H. 1994. De novo design and creation of a stable artificial protein. *Biophys Chem 50*:47–61.

Tuffery P, Etchebest C, Hazout S, Lavery R. 1991. A new approach to the rapid determination of protein side-chain conformations. *J Biomol Struct Dyn 8*:1267–1289.

Vijay-Kumar S, Bugg CE, Cook WJ. 1987. Structure of ubiquitin refined at 1.8 Å resolution. *J Mol Biol 194*:531–544.

Wand JA, Urbauer JL, McEvoy RP, Bieber RJ. 1996. Internal dynamics of human ubiquitin revealed by ^{13}C-relaxation studies of randomly fractionally labeled protein. *Biochemistry 35*:6116–6125.

Weber PL, Brown SC, Mueller L. 1987. Sequential ^{1}H NMR assignments and secondary structure identification of human ubiquitin. *Biochemistry 26*:7282–7290.

Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S Jr, Weiner P. 1984. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J Amer Chem Soc 106*:765–784.

Wintrode PL, Makhatadze GI, Privalov PL. 1994. Thermodynamics of ubiquitin unfolding. *Proteins Struct Funct Genet 18*:246–253.

Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, Markley JL, Sykes BD. 1995. ^{1}H, ^{13}C, and ^{15}N chemical shift referencing in biomolecular NMR. *J Bio NMR 6*:135–140.

Yan Y, Erickson B. 1994. Engineering of betabellin 14D: Disulfide-induced folding of a β-sheet protein. *Protein Sci 3*:1069–1073.