Selection of Optimal Variants of Gō-Like Models of Proteins through Studies of Stretching

Joanna I. Sułkowska and Marek Cieplak
Institute of Physics, Polish Academy of Sciences, Warsaw, Poland

ABSTRACT The Gō-like models of proteins are constructed based on the knowledge of the native conformation. However, there are many possible choices of a Hamiltonian for which the ground state coincides with the native state. Here, we propose to use experimental data on protein stretching to determine what choices are most adequate physically. This criterion is motivated by the fact that stretching processes usually start with the native structure, in the vicinity of which the Gō-like models should work the best. Our selection procedure is applied to 62 different versions of the Gō model and is based on 28 proteins. We consider different potentials, contact maps, local stiffness energies, and energy scales—uniform and nonuniform. In the latter case, the strength of the nonuniformity was governed either by specificity or by properties related to positioning of the side groups. Among them is the simplest variant: uniform couplings with no *i*, *i* + 2 contacts. This choice also leads to good folding properties in most cases. We elucidate relationship between the local stiffness described by a potential which involves local chirality and the one which involves dihedral and bond angles. The latter stiffness improves folding but there is little difference between them when it comes to stretching.

INTRODUCTION

All-atom simulations have been established as a standard approach to interpret behavior of biomolecules on timescales up to ~100 ns. However, studies of large conformational changes, such as those occurring during folding or stretching at experimentally realistic rates, require access to considerably longer timescales. Coarse-grained molecular dynamics models offer tools to provide this access, in an approximate way, by reducing the number of degrees of freedom, e.g., by making the solvent implicit, by dealing only with the C^{α} atoms, and by introducing effective interactions that pertain to the larger scale level of description. The coarse-grained models gain further advantages when considering biomolecular complexes, such as multiple linkages of proteins and ribosomes (1–3), and when comparing properties of a large number of proteins as in the literature (4,5). The procedure of adopting a larger scale of description together with its new set of relevant couplings embodies the spirit of the renormalization group approach in field theory and phase transitions. One should be able to iterate it further when considering larger and larger systems.

Among the coarse-grained models of proteins, the Gō-like systems are currently most widely used (6–17) since they are easy to implement and yet are specific to a protein. The general idea is to devise a model which is consistent with the experimentally established structure of the native state (18,19). Clearly, there is no unique prescription for how to do it, so it seems worthwhile to ponder whether some choices lead to a greater consistency with nonstructural properties of proteins than others. One would expect that the Gō-like models would be more reliable when used primarily in the vicinity of the native state than far away from it. Therefore, considering experiments on protein stretching should provide more reliable benchmarks than those on protein folding. When a protein is stretched at a constant speed, $v_{\rm p}$, it resists the pull and the force of resistance depends on the extension in a nonmonotonic way. The largest peak force, $F_{\max}$, gives a characteristic scale for the resistance and it has been established experimentally for at least 28 proteins. The collected data used in this article are listed in Table 1 (see also (20–42)). If data at various pulling speeds are available, we select those corresponding to the speed of 600 nm/s, as this is the value most commonly used. In this article, we use these data to assess performance of variants of the Gō model.

We first consider the simplest reduction of the degrees of freedom—the one in which a protein is represented by its C^{α} atoms. Later on, we also discuss models with the C^{α} and C^{β} atoms. In the C^{α} case, the potential energy can be written as a sum of four terms:

$$E_{\rm p}(\{\boldsymbol{r}_{\rm i}\}) = V^{\rm BB} + V^{\rm S} + V^{\rm NON} + V^{\rm NAT}. \qquad (1)$$

The first term is responsible for tethering of the consecutive C^{α} beads into a chain. The simplest choice is to represent it by harmonic potentials with minima at 3.8 Å. The second term is responsible for the local backbone stiffness. We discuss two choices for $V^{\rm S}$: one involving bond and dihedral angles (7,15) and another, faster numerically, as given by the chirality potential (14,43). The last two terms represent the remaining interactions between the C^{α} values and they depend on the construction of the contact map. We consider seven ways of choosing the contact map and demonstrate sensitivity of the results to the choice. $V^{\rm NON}$ generates excluded volume (we take it at a distance of 4 Å) for beads which may come to proximity in nonnative conformations. $V^{\rm NAT}$ corresponds to

Submitted December 13, 2007, and accepted for publication May 29, 2008.

Address reprint requests to Joanna I. Sułkowska, E-mail: kwiatek@ifpan.edu.pl.

Editor: Kathleen B. Hall.

0006-3495/08/10/3174/18 $2.00

doi: 10.1529/biophysj.107.127233

TABLE 1 Comparison between experimentally measured values $F_{\rm max}$ with theoretical predictions in $\{$6-12, C, M3, $E_{\rm o}\}$ Gō-like model

PDB	N	$F^{\rm e}_{\rm max}$ [pN]	$v_{\rm p}$ [nm/s]	$F^{\rm t}_{\rm max}[\varepsilon/\text{Å}]$		References
1tit	89	204 ± 30	600	2.15	I27*8	(20,21)
1nct	98	210 ± 10	500	2.4 ± 0.2	I54-I59	(22,23)
1g1c	97	127 ± 10	600	2.3 ± 0.2	I5 titin	(24)
1b6i	164	64 ± 30	1000	1.2	T4 lysozyme(21–141)	(25)
1aj3	106	68 ± 20	3000	1.23	Spectrin R16	(26)
1qjo	80	15 ± 10	600	1.2	eE2lip3(N-C)	(27)
1qjo	40	177 ± 10	600	2.0	E2lip3(N-41)	(27)
1dqv	127	60 ± 15	600	1.5	Calcium binding C2A	(28)
1rsy	127	60 ± 15	600	1.7 ± 0.2	Calcium binding C2A	(28)
1byn	127	60 ± 15	600	1.4	Calcium binding C2A	(28)
1cfc	148	<20	600	0.55	Calmodulin	(28)
1n11	33	37 ± 9	0.2	0.4	Ankyrin*1	(29,30)
1bni	108	70 ± 15	300	1.4, 1.7	Barnase/i27	(31)
1bnr	108	70 ± 15	300	1.05	Barnase/i27	(31)
1bny	108	70 ± 15	300	1.1, 1.3	Barnase/i27	(31)
1hz6	67	152 ± 10	700	3.5	Protein L	(32)
1hz5	67	152 ± 10	700	2.8	Protein L	(32)
2ptl	67	152 ± 10	700	2.2 ± 0.2	Protein L	
1ksr	100	45 ± 20	350	2.0 ± 0.3	DdFLN-4	(33,34)
2rn2	155	19 ± 10	700	1.8 ± 0.2	Ribonuclease H	(35)
1ubq	76	230 ± 34	1000	2.32	Ubiquitin	(36)
1ubq	76	203 ± 35	410	2.32	Ubiquitin(N-C)*9	(36,37)
1ubq	28	85 ± 20	300	0.9	Ubiquitin(K48-C)*(2–7)	(36,37)
1emb	129	350 ± 30	3600	5.15 ± 0.4	GFP(3–132)	(39)
1emb	219	130 ± 30	3600	2.3, 4.3	GFP(3–212)	(39)
1emb	80	120 ± 30	3600	2.2 ± 0.2	GFP(132–212)	(39)
1emb	235	104 ± 40	3600	2.3 ± 0.2	GFP(N-C)	(38)
1fnf	94	75 ± 20	3000	1.6, 1.8	Fniii-10	(40,41)
1ttf	94	75 ± 20	600	0.7, 1.2	Fniii-10	(42)
1ttg	94	75 ± 20	600	0.7, 1	Fniii-10	(42)
1fnh	92	124 ± 18	600	1.8	Fniii-12	(41)
1fnh	89	89 ± 18	600	1.4, 1.7	Fniii-13	(41)
1oww	93	220 ± 31	600	2.1 ± 0.2	FNiii-1	(41)
1ten	90	135 ± 40	500	1.7	TNFNiii-3	(41,67)
1pga	56	190 ± 20	400	2.4, ± 0.2	Protein G	(47)
1gb1	56	190 ± 20	400	1.65 ± 0.2	Protein G	(47)

The theoretical results are averaged over 10 trajectories to account for several pathways if any. The symbol after the asterisk indicates the number of domains.

attractive interactions that are judged to be operational in the native state. In typical folding (14) and stretching (16) simulations, $V^{\rm NAT}$ has a well-defined minimum, such as the commonly used Lennard-Jones potential

$$V^{6-12} = 4E_{\rm ij}\left[\left(\frac{\sigma_{\rm ij}}{r_{\rm ij}}\right)^{12} - \left(\frac{\sigma_{\rm ij}}{r_{\rm ij}}\right)^{6}\right], \tag{2}$$

where $r_{\rm ij}$ is the distance between the C^{α} values in amino acids i and j, in which $\sigma_{\rm ij}$ is determined pair-by-pair so that the minimum is located at the experimentally established native distance $r^{\rm n}_{\rm ij}$, i.e., $\sigma_{\rm ij} = r^{\rm n}_{\rm ij}/\sqrt[6]{2}$. However, the energy parameters in the pair potentials, such as the $E_{\rm ij}$ in Eq. 2, determine the depths of the potential wells and turn out to influence properties of the system more substantially. In this article, we consider six choices of the energy scales, including the one with uniform values: $E_{\rm ij} = \varepsilon$.

Overall, we consider 62 variants for the C^{α}-based Gō models and estimate the square of the Pearson correlation coefficient, R^2, with the set of the experimental values of $F_{\rm max}$ for each of them. We find that there is a considerable spread in the values of R^2, and several models come out the best. Among these is the Lennard-Jones model with uniform energy parameters, recently used in the theoretical survey of stretching of nearly 8000 proteins (4,5) with two different contact maps. Surprisingly then, despite the complexity of features encountered in proteins, such as the differing properties of the constituent amino acids, the simplest uniform choice of the couplings is found to be optimal for modeling of the dynamics.

The article is organized as follows. In the next section, we outline the methods of the molecular dynamics procedure and of the statistical assessment of the data. In Models, we define the variants of the C^{α}-based models. In Models with Side Groups, we define models which also involve the C^{β} atoms. In The Selection of Temperature for Stretching Simulations, we discuss issues related to the selection of the temperature at which the simulations are performed. In Results, we compare performances of various models when confronted with the experimental data on stretching. In Thermodynamics and

Folding and the section following, Conclusion, we discuss thermodynamic stabilities and folding properties of selected models respectively. Models with the local stiffness described through the bond and dihedral angles are found to yield better folding properties than those incorporating values of the chirality. In stretching, the differences between the two are minor.

METHODS

The molecular dynamics procedure

The time evolution of folding or unfolding is simulated through methods of molecular dynamics as described in detail in the literature (14,16,44). The beads representing the amino acids are coupled to the Langevin noise and damping terms to mimic the effect of surrounding solvent and provide thermostatting at a temperature T. The equations of motion for each bead are

$$m\ddot{\mathbf{r}}_{\mathrm{i}} = -\gamma\dot{\mathbf{r}}_{\mathrm{i}} + \mathbf{F}_{\mathrm{c,i}} + \Gamma_{\mathrm{i}}, \tag{3}$$

where m is the mass of an amino acids represented by each bead, considered to be uniform and equal to the average amino-acid mass. $\mathbf{F}_{\mathrm{c,i}}$ is the net force due to molecular potentials and external forces that act on the i^{th} bead located at $\mathbf{r}_{\mathrm{i}}$, γ is the damping coefficient, and Γ_{i} is a Gaussian noise term with the dispersion $\sqrt{2\gamma k_{\mathrm{B}}T}$, where T is the temperature. The dynamics of any bead are overdamped and thus the natural single-bead timescale in the problem, τ, is not related to the period of oscillations in the potential well but to diffusion. It is of ~1 ns (7,45,46).

The stretching simulation were accomplished by attaching both ends of the protein in its native state to harmonic springs of elastic constant $k = 0.06$ ε/Å^2, which is close to the values corresponding to the elasticity of experimental cantilevers. One spring is anchored at one end and the second spring is pulled at its head with the velocity v_{p} of 0.005 Å/τ, which exceeds the experimental speeds by approximately two orders of magnitude.

To study folding, we start from at least 300 unfolded conformations and determine the folding time, t_{fold}, as the median first passage time, i.e., the time needed to arrive at the native conformation. The native state is declared to be reached if all of its native contacts are established for the first time. For the Lennard-Jones potential, a native contact is established if r_{ij} does not exceed $1.5\sigma_{\mathrm{ij}}$ and similar criteria apply for other potentials. The value t_{fold} depends on T typically in an U-shape fashion and the center of the U defines an optimal temperature for folding, T_{min}. Throughout the article, we shall use dimensionless temperatures denoted by $\tilde{T}$. For the uniform Lennard-Jones potential, $\tilde{T} = k_{\mathrm{B}}T/\varepsilon$. For other potentials, ε is replaced by the average strength of the native contact in a given protein.

Thermodynamic stability of a protein can be characterized by providing the folding temperature $\tilde{T}_{\mathrm{f}}$ at which half of the native bonds are established on average in an equilibrium run (based on at least five long trajectories that start in the native state).

In all of our studies, contacts involving disulphide bonds have their energy parameters enhanced by an order of magnitude to prevent their rupture.

Statistical measures of correlation with the experimental data

F_{max} has been measured experimentally for a set of $D = 28$ proteins. The experimental values of F_{max} are denoted by F^{e}_{λ}, where $\lambda = 1,\ldots,D$. The corresponding theoretical values are denoted by F^{t}_{λ}.

The level of correlation between the set of F^{t}_{λ} and F^{e}_{λ} values can be assessed through the square of the Pearson correlation coefficient

$$R^2 = 1 - \sqrt{\frac{1}{D}\sum_{\lambda=1}^{\mathrm{D}}\left(\frac{F^{\mathrm{t}}_{\lambda} - F^{\mathrm{e}}_{\lambda}}{F^{\mathrm{e}}_{\lambda}}\right)^2}. \tag{4}$$

The closer the R^2 is to unity, the more perfect the correlation becomes, whereas a zero value of R^2 signifies absence of any correlation. The actual optimal form of the linear relationship

$$F^{\mathrm{t}}_{\lambda} = a^{*}F^{\mathrm{e}}_{\lambda} \tag{5}$$

is obtained through the least-square method. The method is constrained to involve only the slope, but not a constant shift, on physical grounds: on dissolving the coupling constants to zero, both the experimental and theoretical systems should generate a zero force.

The quantity R^2 may overemphasize rare large deviations in an otherwise well-behaving model. Therefore, we also consider a complementary statistical measure known as Theil's U coefficient. This coefficient incorporates consecutive changes as one goes from one protein to its neighbor in an ordered set as though the correlational trend corresponded to a passage of time. We order the set of proteins from the smallest to the largest value in F^{e}_{λ} and define

$$W_{\lambda+1} = \frac{F^{\mathrm{t}}_{\lambda+1} - F^{\mathrm{e}}_{\lambda}}{F^{\mathrm{e}}_{\lambda}}, \quad w_{\lambda+1} = \frac{F^{\mathrm{e}}_{\lambda+1} - F^{\mathrm{e}}_{\lambda}}{F^{\mathrm{e}}_{\lambda}}. \tag{6}$$

Here, $W_{\lambda+1}$ and $w_{\lambda+1}$ denote the predicted and actual relative single-step changes, respectively. The U coefficient is then given by

$$U = \sqrt{\frac{\sum_{\lambda=1}^{\mathrm{D}-1}(W_{\lambda+1} - w_{\lambda+1})^2}{\sum_{\lambda=1}^{\mathrm{D}-1}(w_{\lambda+1})^2}}. \tag{7}$$

A perfect prediction corresponds to $U = 0$. Values >1 correspond to theoretical predictions that are opposite to the actual ones, and values equal to 1 ($W_{\lambda+1} = 0$) signify prediction of no change. Thus, a good theoretical model should yield small values of the U coefficient.

Proteins used for analysis

The experimental results on F_{max} that we used are listed in Table 1 together with the theoretical results based on the Lennard-Jones model with the uniform energy parameters (5). This list is essentially identical to the one considered in the literature (4,5) to justify the model used in the PDB-wide protein survey except that now we have also included the data on protein G (47).

The proteins are usually linked in tandem with the same or other repeat units and the linkages typically involve the terminal amino acids. If the linkage involves other amino acids, like the sites 21 and 141 in the case of lysozyme, then this feature is indicated in brackets next to the common name of the protein in the last column. The first column lists the PDB code of the protein. The tandem nature of the biomolecules used in the stretching experiments is often interpreted as leading to a serial character of unwinding which should allow for extraction of data for the domain of interest. However, such an interpretation need not necessarily be always correct. Nevertheless, we followed the assignment of the force to a domain as decided by the experimentalists. If a protein has several PDB structures associated with it, and therefore several somewhat differing contact maps, we average the theoretical results over these structures. We also average over several trajectories.

In our tests of the folding properties within a set of models, we used crambin (1crn), ubiquitin (1ubq), and the 27th domain of titin (1tit).

MODELS

The models have several basic attributes that can be listed as

$$model = \{V^{\mathrm{NAT}}, S, \mathcal{M}, E\}, \tag{8}$$

where the subsequent entries mean making a decision about the functional form of the potential in the native contact, the description of the local backbone stiffness, the contact map, and the set of values of the energy parameters consecutively. The choices considered are summarized in Table 2 and will be described in the following. The total number of possibilities encompassed by Table 2 is equal to 504, but we investigate dynamically ~12% of them by making judgments on their physical relevance.

The C^{α}-C^{α} potentials for the native contacts

We consider six variants of the interactions V^{NAT} in the native contacts. In each case, the relevant length parameters are chosen so that the minimum in the potential coincides with the C^{α}-C^{α} contact distance in the native state. The amplitudes of these potentials, E_{ij}, will be discussed later on. The native contact distances are distinct from the distances, $l_{\mathrm{e,ij}}$, which set the threshold for establishing or rupturing a contact during molecular dynamics simulations: if $r_{\mathrm{ij}} < l_{\mathrm{e,ij}}$, then the contact is considered to be present dynamically. The functional forms of the potentials considered here are shown in Fig. 1.

The 6-12 potential

The first variant is an ordinary Lennard-Jones potential defined in Eq. 2 for which $l_{\mathrm{e,ij}} = 1.5\sigma_{\mathrm{ij}}$.

The 6-12 potential with constant shape

The second variant is the Lennard-Jones potential, whose width does not depend on the actual distance between the C^{α} atoms (only the location of the minimum does). The shape of potential has been proposed in Wojciechowski and Cieplak (48) and is given by

$$V^{6-12}_{\mathrm{const}} = 4E_{\mathrm{ij}}\left[\left(\frac{s}{r_{\mathrm{ij}} - \sqrt[6]{2}(r^{\mathrm{n}}_{\mathrm{ij}} - s)}\right)^{12} - \left(\frac{s}{r_{\mathrm{ij}} - \sqrt[6]{2}(r^{\mathrm{n}}_{\mathrm{ij}} - s)}\right)^{6}\right], \qquad (9)$$

where s is a constant parameter which determines the width of the potential. Studies of 500 proteins (48) suggest that $s \sim 4$ Å on average. However, there is a possibility, though not explored here, to distinguish between the hydrogen bonds and all other bonds for which one has $s \sim 2.4$ Å and $s \sim 5.6$ Å, respectively. We take $l_{\mathrm{e,ij}} = 1.5 \cdot s + \sqrt[6]{2}(r^{\mathrm{n}}_{\mathrm{ij}} - s)$.

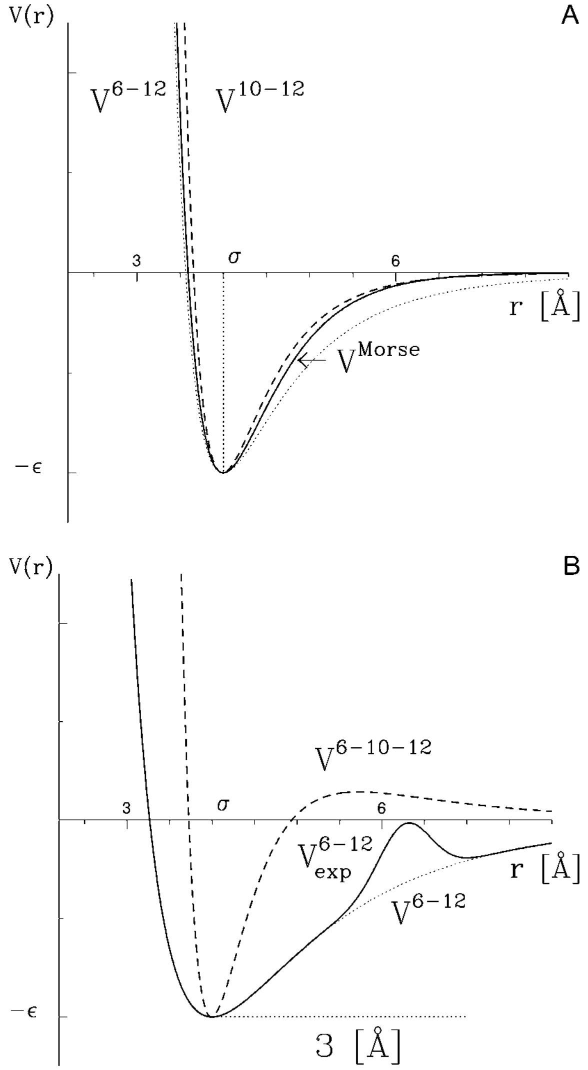

FIGURE 1 Examples of the functional forms of contact potential corresponding to a native distance of $\sigma = 4$ Å. (*A* and *B*) Lennard-Jones potential (or $V^{6\text{-}12}$) is represented by the dotted line. The remaining potentials are denoted as follows: (*A*) $V^{10\text{-}12}$, dashed line and V^{Morse}, solid line; and (*B*) $V^{6\text{-}10\text{-}12}$, dashed line and V^{6-12}_{exp}, solid line. The minima of the potentials 6-12, 6-10-12 are set at $r = 4$ Å.

TABLE 2 Notation used to describe different models

Notation	
V^{NAT}	$\in \{$6-12, 6-12$_{\mathrm{const}}$, 10-12, 6-10-12, 6-12$_{\mathrm{exp}}$, Morse$\}$
S	$\in \{C, A\}$
M	$\in \{$M2, M3, M4, M2c75, M3c7, M3c75, CSU$\}$
E	$\in \{E^{\mathrm{o}}, E^{\mathrm{L1,\ L2}}, E^{\mathrm{G1,\ G2,\ G3}}, E^{\nu1,\ \nu2}, E^{\mathrm{HB\text{-}MJ}}, E^{\mathrm{HB\text{-}HH}}\}$

V^{NAT} denotes possible choices of the functional form of the potentials that describe native contacts. *S* stands for the potentials describing the local backbone stiffness: chirality-based (*C*) and bending-angle- and dihedral-angle-based (*A*). *M* stands for the choices in the contact map, as described in the text. *E* denotes choices for energy scales, or amplitudes, that multiply the contact potentials.

The 6-10-12 potential

The third variant of the potential corresponds to the interaction that is mediated by water molecules which is understood as giving rise to a second minimum and, in consequence, an additional energy barrier compared to the standard Lennard-Jones potential. This potential has been used in the literature (49–53) and is described by

$$V^{6-10-12} = 4E_{\mathrm{ij}}\left[13\left(\frac{\sigma_{\mathrm{ij}}}{r_{\mathrm{ij}}}\right)^{12} - 18\left(\frac{\sigma_{\mathrm{ij}}}{r_{\mathrm{ij}}}\right)^{10} + 4\left(\frac{\sigma_{\mathrm{ij}}}{r_{\mathrm{ij}}}\right)^{6}\right]. \qquad (10)$$

Note that the location of the barrier in the 6-10-12 potential depends on the location of the minimum. This barrier generates the energy penalty needed to be paid to establish a contact. However, once a contact is established, its stability is enhanced. In this case, we consider contact to be present if r_{ij} is smaller than the position of the energetic barrier and thus $l_{e,ij} = 1.4\sigma_{ij}$.

The 6-12 potential with a second minimum

The fourth variant is similar to the third one, but we construct it in a way that introduces an intermediate minimum in the contact mediated exactly by one water molecule, so that it takes the form

$$V^{6-12}_{\exp} = V^{6-12}_{\mathrm{const}} + \mathcal{A}\exp(-(r_{ij} - \sigma_{ij} + \mathcal{C}\sqrt[6]{2} - 9)^2/\mathcal{C}^2\mathcal{B}), \tag{11}$$

where $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ control the shape of the potential. We take 0.3, 0.005, 6 for $\mathcal{A}$–$\mathcal{C}$, respectively, which describe interaction mediated by one water molecule as most accurate. The Gaussian term represents an energy barrier with the maximum of 0.01 ε that is followed by the second energy minimum at a distance $r'_{ij} = r_{ij}+3$ Å corresponding to amino acids separated by one water molecule. The separation of 3 Å is indicated in Fig. 1 *B* by the dotted horizontal line at the lowest minimum. Similar form of the potential was used in Wojciechowski and Cieplak (48). Again, we take $l_{e,ij}$ to agree with the position of the energetic barrier and thus $l_{e,ij} = \sigma_{ij}+2.32$ Å.

The 10-12 potential

The fifth variant is the 10-12 potential, which takes the form

$$V^{10-12} = E_{ij}\left[5\left(\frac{r^n_{ij}}{r_{ij}}\right)^{12} - 6\left(\frac{r^n_{ij}}{r_{ij}}\right)\right]^{10}. \tag{12}$$

This potential has frequently been used to describe hydrogen bonds (15). Following Clementi et al. (15), we take $l_{e,ij} = 1.2r^n_{ij}$. It has been shown (54) that a precise definition of $l_{e,ij}$ is not essential in studies of folding in this case.

The Morse potential

The sixth variant is the Morse potential, which takes the form

$$V^M = E_{ij}[(1 - \exp(-\alpha(r_{ij} - r^n_{ij}))^2 - 1]. \tag{13}$$

We chose the parameter α to be equal to $1.7/r^n_{ij}$. For this choice, the shape of the Morse potential is similar to the shape of the ordinary Lennard-Jones potential at the most typical distance of ~4 Å. We take $l_{e,ij} = 3.4\, r^n_{ij}$, which corresponds to the inflection point in the Lennard-Jones potential. The variants considered here are certainly not exhaustive. For instance, another class of possible potentials that is relevant physically may involve Coulombic terms.

Potentials for the backbone stiffness

We have considered two choices, denoted as *A* and *C*, for describing the local conformation of the backbone where the symbols correspond to the angular and chiral methods. The more widely known angular method makes use of a potential V^A, which depends on the bond (θ_i) and dihedral (ϕ_i) angles and favors their native values (θ^n_i and ϕ^n_i) (7,15,55). It is given by

$$V^A = \Upsilon\left[\sum_{i=1}^{N-2} K_\theta(\theta_i - \theta^n_i)^2 + \sum_{i=1}^{N-3}\left(K^1_\phi(1 - cos(\phi_i - \phi^n_i)) + K^3_\phi(1 - cos3(\phi_i - \phi^n_i))\right)\right]. \tag{14}$$

Following Clementi et al. (15) we take 20ε, ε, and 0.5ε for the force constants K_θ, K^1_ϕ, and K^3_ϕ, respectively. The quantity Υ is an overall control parameter of the potential strength such that when $\Upsilon = 1$, the customarily used strength is obtained (15). The bond and dihedral angle are determined by three and four subsequent residues, respectively. Customarily, when using V^A, one discards contacts which may arise in pairs *i*, *i* + 2 and *i*, *i* + 3 because they may contradict the action of V^A.

A simpler way to represent the local stiffness is through the chirality potential, V^C, which favors the native sense of the local chirality (14,43). The specific choice of V^C employed here is given by

$$V^C = \sum_{i=2}^{N-2}\frac{1}{2}\kappa(C_i - C^n_i)^2, C_i = \frac{(\mathbf{w}_{i-1}\times\mathbf{w}_i)\cdot\mathbf{w}_{i+1}}{d^3_0}, \tag{15}$$

where C^n_i is the chirality of residue *i* in the native conformation and $d_0 = |\mathbf{w}_i|$ is the distance between subsequent C^α atoms. Here, $\mathbf{w}_i = \mathbf{r}_{i+1} - \mathbf{r}_i$. We take κ equal to 1 as discussed in Kwiecinska and Cieplak (43). The action of V^C is similar to V^A, but it is weaker and less conformation-specific. Therefore, when using V^C one should keep the shorter-ranged contacts like *i*, *i* + 3 contacts, if any.

We now demonstrate that V^A acts approximately like V^C but it also incorporates an extra term which makes it stiffer. Notice that these potentials are expressed in terms of three consecutive unit vectors $\mathbf{w}_i$, which join four consecutive C^α atoms. Let ϕ denote a dihedral angle between the plane spanned by the vectors $\mathbf{w}_{i-1}$ and $\mathbf{w}_i$, and the plane spanned by the vectors $\mathbf{w}_i$ and $\mathbf{w}_{i+1}$. We denote a deviation from its native value ϕ^n by $\delta\phi = \phi - \phi^n$. The dihedral angle can be determined from the relation

$$tg\,\phi = \frac{C_i}{D_i}, \tag{16}$$

where

$$D_i = (\mathbf{w}_{i-1}\times\mathbf{w}_i)\cdot(\mathbf{w}_i\times\mathbf{w}_{i+1})/d^4_0. \tag{17}$$

Let us now consider the potential V^A given in Eq. 14. For small values of $\delta\phi$, one can expand the cosine functions in the dihedral angle part of V^A to get (for the i^{th} residuum)

$$V_{\mathrm{i}}^{\mathrm{dih}} \simeq \frac{2\Upsilon(K_{\phi}^{1} + 9K_{\phi}^{3})}{2}\delta\phi^{2} + \mathcal{O}((\delta\phi)^{3}). \qquad (18)$$

V^{dih} can be expressed in terms of vectors $\mathbf{w}_{\mathrm{i}}$ by noticing that for small $\delta\phi$

$$\delta\phi \simeq tg\,\delta\phi = \frac{tg\,\phi - tg\,\phi^{\mathrm{n}}}{1 + tg\,\phi tg\,\phi^{\mathrm{n}}} \simeq \frac{C_{\mathrm{i}} - C_{\mathrm{i}}^{\mathrm{n}}}{D_{\mathrm{i}}^{\mathrm{n}} + C_{\mathrm{i}}C_{\mathrm{i}}^{\mathrm{n}}/D_{\mathrm{i}}} + \ldots, \qquad (19)$$

where $D_{\mathrm{i}}^{\mathrm{n}}$ denotes the native value of D_{i} and the dots indicate other terms arising from the expansion of D_{i} around its native value. Substituting this form of $\delta\phi$ into Abe and Gō (18), we find that $V_{\mathrm{i}}^{\mathrm{dih}}$ indeed contains a part which, in the first approximation, can be identified with the rescaled chiral potential (15)

$$V_{\mathrm{i}}^{\mathrm{dih}} \simeq \frac{2\Upsilon(K_{\phi}^{1} + 9K_{\phi}^{3})}{2(D_{\mathrm{i}}^{\mathrm{n}} + (C_{\mathrm{i}}^{\mathrm{n}})^{2}/D_{\mathrm{i}}^{\mathrm{n}})^{2}}(C_{\mathrm{i}} - C_{\mathrm{i}}^{\mathrm{n}})^{2} + \ldots.$$

This shows that V_{C} is responsible essentially for the dihedral angle part of the V^{A} potential. However, potential V^{A} also contains the bond-angle terms

$$V^{\mathrm{A}} \simeq V^{\mathrm{C}} + \Upsilon K_{\theta}\sum_{\mathrm{i}}(\delta\theta_{\mathrm{i}})^{2}. \qquad (20)$$

Thus, V^{A} leads to stronger local stiffness energies. The argument above suggests that one could replace the potential V^{A} by its more convenient numerically approximation provided by Eq. 20.

The contact maps

We consider three basic techniques to determine the native contact map of a protein for a C^{α}-based Gō model. The simplest of them is to introduce a cutoff distance, R_{c}, for r_{ij} between the C^{α} atoms, which are not sequential neighbors. The usual choices are R_{c} of 7 Å or 7.5 Å, as used previously in, e.g., the literature (13,56). The corresponding contact maps will be denoted by Mc7 and Mc75, respectively.

The cutoff-based approach often misses many important couplings at larger distances, since the contact lengths may extend up to ~12 Å (14). A simple yet more sophisticated approach has been proposed by Tsai et al. (57). It involves reading in native positions of all nonhydrogen atoms in an amino acid and assigning spheres to them. In this way, an amino acid is represented by a cluster of grapes. The radii of the spheres are equal to the van der Waals radii multiplied by 1.24 to account for attraction. If two such clusters of grapes overlap, one declares existence of a native contact between the corresponding amino acids. The contact map determined by this overlap technique will be denoted by M. It has been applied to Gō-like models in the literature (16,44,58).

A third technique is chemistry-based, and it was used in Onuchic et al. (54). It takes into account specific geometrical properties that correspond to various types of bonds as considered at the electronic level. There is a commonly used software, known as CSU, which has been developed by Sobolev et al. (59). This software determines which type of contacts (e.g., hydrogen bonds, hydrophobic-hydrophobic, aromatic-aromatic, aromatic-polar, etc.) contributes to stabilization of the native structure the most. In particular, it works by analyzing the interface surface between two amino acids.

The CSU-based contact map cannot be used for large-scale computations since determination of the contact map this way becomes a tedious task in itself. However, we have used it in conjunction with the overlap technique in the following way. We have considered proteins 1tit, 1aj3, 1ubq, and 1crn and found that the CSU-based and overlap-based contacts maps are very similar with the exception of the i, $i + 2$ contacts. These short-ranged contacts are often found by the overlap criterion, but they usually turn out to correspond to the van der Waals polarizational couplings which are an order-of-magnitude weaker than the hydrogen bonds. For this reason, we consider three kinds of the overlap-derived contact maps: M2, M3, and M4. In M2, all overlap-based contacts are taken into account, whereas in M3, the i, $i + 2$ are discarded. Finally, in M4 both i, $i + 2$ and i, $i + 3$ are discarded; M4 is used when considering the angular way to describe the local stiffness. The distinction between these maps is more meaningful when dealing with α-proteins but it is not very relevant when dealing with β-proteins. Sulkowska and Cieplak used the M3 (4) and M2 (5) contact maps.

The energy parameters

Uniform energy parameters E^{o}

In the simplest case we assume a uniform energy parameter, so that all amino acids interact with the same strength and $E_{\mathrm{ij}} = E_{\mathrm{ij}}^{\mathrm{o}} = \varepsilon$. This parameter should be of ~1–2 kcal/mol.

A quasiuniform variant of this approach is obtained when one uses the 10-12 potential for the hydrogen bonds, but the 6-12 potential (or the 6-12 with constant shape) for all other contacts while keeping the energy scale for all potentials the same. In this variant, the hydrogen-bond contacts get enhanced at shorter distances. The corresponding models will be denoted as (6-12, 10-12), E^{o} or (6-12$_{\mathrm{const}}$, 10-12), E^{o}.

Nonuniform energy parameters

We now introduce various nonuniform energy scales, by defining more general energy parameters for various subsets of native contacts. In what follows it is assumed that native contacts for which new energy parameters are not explicitly specified, interact with the uniform energy ε.

The $E_{\mathrm{ij}}^{\mathrm{L}}$ energy parameters

This model is motivated by Srinivasan and Rose (60); it aims at reducing the effect of contacts that correspond to large native C^{α}-C^{α} native distances. This goal is achieved by introducing a cutoff distance m. Below this distance, the strength of the contact is given by $E^{\mathrm{L}}_{\mathrm{ij}} = \varepsilon$. Between m and 12 Å, it is modulated as

$$E^{\mathrm{L}}_{\mathrm{ij}} = \varepsilon\left[1 - \frac{r^2_{\mathrm{ij}} - m^2}{(m+n)^2 - m^2}\right], \qquad (21)$$

where n is another parameter. We consider two cases, $E^{\mathrm{L1}}_{\mathrm{ij}}$ and $E^{\mathrm{L2}}_{\mathrm{ij}}$, for which $m = 8.5$ Å, $n = 3.5$ Å and $m = 10.6$ Å, $n = 1.4$ Å, respectively. The latter choice has been used in Srinivasan and Rose (60). For native distances ≥12 Å, $E^{\mathrm{L}}_{\mathrm{ij}}$ is set equal to zero.

The $E^{\mathrm{G}}_{\mathrm{ij}}$ energy parameters

This model is based on geometrical properties of amino acids and it takes into account the specific nature of the native atomic overlaps. We define three situations denoted as *aa*, *ab*, and *bb*, consecutively. In the first situation, usually corresponding to the backbone-backbone hydrogen bonds, the atomic overlaps in the contact involve at least one pair of atoms that both belong to the backbone. If such an overlap does not exist, but there is an overlap between a backbone piece of one amino acid and a side group of another, then this case corresponds to the second situation. The third situation arises when the atomic overlaps develop only between the atoms of the side groups. Such contacts are usually long-ranged and hydrophobic.

We consider three variants of the energy scales $E^{\mathrm{G1}}_{\mathrm{ij}}$, $E^{\mathrm{G2}}_{\mathrm{ij}}$, and $E^{\mathrm{G3}}_{\mathrm{ij}}$, which take into account the three situations described above in a different way:

$$E^{\mathrm{G1,G2,G3}}_{\mathrm{ij}} = \begin{cases} \varepsilon & \varepsilon & \varepsilon & \text{for aa} \\ \varepsilon & \varepsilon & \varepsilon \times 0.75 & \text{for ab, ba}\,. \\ \varepsilon \times 0.75 & \varepsilon \times 0.9 & \varepsilon \times 0.9 & \text{for bb} \end{cases} \qquad (22)$$

Here, each column corresponds to a different model and, for each model, the couplings are between the C^{α} atoms. These models reduce the role of the side groups relative to the hydrogen bonds that involve the backbone. Moreover, we simultaneously consider energy scales E^{G1}, or E^{G2}, or E^{G3} with E^{L}.

The E^{ν}_{ij} energy parameters

Another way of introducing a nonuniform set of the energy parameters is to take into account the actual number of the overlapping atomic pairs in a contact. In the overlap-determined contact map there is no reference to the number of the existing overlaps so one may consider making up for this by enhancing contacts in proportion to the number, ν_{ij}, of overlapping pairs in a contact between the amino acids i and j. The distributions of ν_{ij} for three proteins are shown in Fig. 2. The issue here is, however, how to normalize this proportion and there are several reasonable propositions of

$$E^{\nu}_{\mathrm{ij}} = \varepsilon\nu_{\mathrm{ij}}/\Omega_{\mathrm{k}}, \qquad (23)$$

where Ω_{k} denotes different normalizations. We consider three ways ($k = 1, 2, 3$) to implement the normalization by:

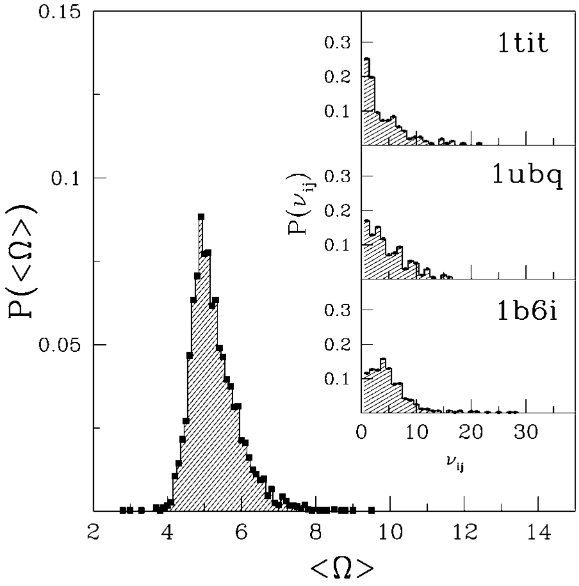

FIGURE 2 Distribution of the average number of atomic contacts normalized to the number of interacting pairs i, j in the native state of a protein, as determined based on 3500 proteins. The insets show distributions of the numbers of atomic contacts between i, j in proteins 1tit, 1ubq, and 1b6i, top to bottom, respectively.

the number, Ω_1, of all atomic pairs in the given protein; the most typical number, $\Omega_2 = 4.9$, of interresidue interactions between two amino acids based on a set of 3500 proteins (see Fig. 2); and the maximal number, Ω_3, of the atomic overlap in the protein under consideration. Still another way to normalize was used in Cecconi et al. (61), in which one first uses a cutoff distance to tell a native contact, and only then do the studies overlap.

The $E^{\mathrm{HH}}_{\mathrm{ij}}$ and $E^{\mathrm{HB}}_{\mathrm{ij}}$ energy parameters

So far, the energy parameters were varied based on the native state geometry. An alternative approach is to focus on the chemical properties and take into account the hydrophobicity scale or/and the presence of the hydrogen bonds. The simplest approach to include the hydrophobic properties ($E^{\mathrm{HH}}_{\mathrm{ij}}$) is to follow Srinivasan and Rose (60) and set E_{ij} equal to 2ε if both interacting amino acids are hydrophobic, equal to ε if one is hydrophobic and other amphipathic, and equal to zero when neither i nor j is hydrophobic. The hydrophobic residues we consider are Cys, Ile, Leu, Met, Phe, Trp, and Val. The amphipathic residues are Ala, His, Thr, and Tyr.

We now discuss ways to incorporate the energies in the hydrogen bonds ($E^{\mathrm{HB}}_{\mathrm{ij}}$) between the N and C atoms (60). We first identify all hydrogen bonds by the method of Kabsch and Sander (62). This method estimates the hydrogen-bond energy based on geometry and electrostatics, which are a function of both the hydrogen-bond distance and the alignment of the N and C atoms. A hydrogen bond is present when this energy lies below some threshold value. However, positions of protons are not provided in all PDB files, and therefore we simplified this procedure by only checking for overlaps of the enlarged van der Waals spheres of the N and C atoms. Consequently, we used this simplification to determine contacts for all proteins. In this model, we assume that each hydrogen contributes an ε to the amplitude of the potential. In particular, if two hydrogen bonds contribute to the same contact, the energy parameter is 2ε. However, in such a case, or if there is a compound contact (e.g., a backbone-backbone hydrogen-bond overlap and a side-chain atomic overlap), we assign the energy of $(1/4)\varepsilon$ to the surrounding couplings $(i, j-1)$, $(i, j+1)$, $(i-1, j)$, $(i+1, j)$ if they are not yet connected by other native contact. A similar idea of amplification was previously introduced in the literature (49,63). It has to be noted that we also combined energy scale E^{HB} with the E^{HH} discussed above.

The $E_{\mathrm{ij}}^{\mathrm{HB,MJ}}$ energy parameters

Another way to deal with the variety of chemical properties of the 20 amino acids is to introduce a table of amino-acid-dependent interactions. The first example of such a table was provided by Miyazawa and Jernigan (64). We denote this table by $\varepsilon_{\mathrm{ij}}^{\mathrm{MJ}}$. It comprises 210 different entries and it reflects an uneven frequency of occurrence of different amino-acid pairs in contacts. Energies in such a table should be normalized with respect to the energy of hydrogen bonds in the remaining contacts (the E^{HB} energy scale). We follow the implementation proposed by Karanicolas and Brooks (49). The corresponding energy scale is defined as

$$E_{\mathrm{ij}}^{\mathrm{HB,MJ}} = \begin{cases} E^{\mathrm{HB}} & \text{for hydrogen bonds} \\ \varepsilon \dfrac{\varepsilon_{\mathrm{ij}}^{\mathrm{MJ}}}{\langle \varepsilon_{\mathrm{ij}}^{\mathrm{MJ}} \rangle} & \text{for non-HB side chain-side contacts}, \\ \varepsilon & \text{for other contacts} \end{cases} \tag{24}$$

where the average is over the 210 $\varepsilon_{\mathrm{ij}}^{\mathrm{MJ}}$ parameters. This energy function distinguishes between three types of native contacts: contacts arising through the side-chain-to-side-chain interaction, hydrogen bonds described by E^{HB}, and all remaining contacts of strength ε. $E_{\mathrm{ij}}^{\mathrm{HB,MJ}}$ is sometimes denoted as $E_{\mathrm{ij}}^{\mathrm{MJ}}$ for short.

Table 3 shows native energies for 1tit, 1ubq, and 1crn as calculated based on all various energy scales considered here and for the M3 contact map. We find that most of them are very close (the variations are <8%) to the value calculated with the uniform energy parameters. Significantly above and below are $E^{\mathrm{MJ,HB}}$ and $E^{\mathrm{L1,G3}}$, respectively. This is an important result because it provides a justification for using the same reduced temperature for all models in the studies of stretching.

MODELS WITH SIDE GROUPS

The C^{α}-based models can acquire a finer structure by introducing side groups. The simplest implementation involves representing an amino acid by two beads: one at the location of C^{α} and another at that of C^{β}. In the native state, the locations of the C^{β} atoms are given in the PDB file. However, in a dynamical simulation one needs to know how to keep tethering the C^{β} values to the backbone at an angle. We follow the procedure of Sułkowska and Cieplak (4) and introduce a tethering potential that has a minimum when the C^{β} atom is at a distance of $l = 1.5$ Å from the C^{α} in the direction $\vec{r}_{\mathrm{i}}^{C^{\beta}}$ $(l = |\vec{r}_{\mathrm{i}}^{C^{\beta}}|)$. This direction is calculated based on the placement, $\vec{r}_{\mathrm{i}}^{C^{\alpha}}$, of the corresponding C^{α} atom and of its sequential neighbors along the chain. The directional characteristics are described by the equation

$$\vec{r}_{\mathrm{i}}^{C^{\beta}} = l(\hat{a}\cos\theta + \hat{b}\sin\theta), \tag{25}$$

which was deduced from studies of the peptide geometry (51,65). Here, the angle θ is chosen optimally to be equal to 37.6° and

$$\hat{a} = \frac{\hat{s}_{\mathrm{i,i-1}} + \hat{s}_{\mathrm{i,i+1}}}{|\hat{s}_{\mathrm{i,i-1}} + \hat{s}_{\mathrm{i,i+1}}|} \quad \hat{b} = \frac{\hat{s}_{\mathrm{i,i-1}} \times \hat{s}_{\mathrm{i,i+1}}}{|\hat{s}_{\mathrm{i,i-1}} \times \hat{s}_{\mathrm{i,i+1}}|}, \tag{26}$$

where $\hat{s}_{\mathrm{i,j}}$ is a unit vector defined by

$$\hat{s}_{\mathrm{ij}} = \frac{\vec{r}_{\mathrm{i}}^{C^{\alpha}} - \vec{r}_{\mathrm{j}}^{C^{\alpha}}}{|\vec{r}_{\mathrm{i}}^{C^{\alpha}} - \vec{r}_{\mathrm{j}}^{C^{\alpha}}|}. \tag{27}$$

This model with the side groups is denoted by $C^{\alpha-\beta}$.

The potential energy of a protein where each amino acid is represented by two beads is given by the expression

$$E_{\mathrm{p}}^{\alpha-\beta}(\{\mathbf{r}_{\mathrm{i}}\}) = V^{\mathrm{BB}} + V^{\mathrm{BS}} + V_{\alpha-\beta}^{\mathrm{NAT}} + V^{\mathrm{NON}} + V^{\mathrm{S}}, \tag{28}$$

where the terms V^{BB}, V^{NON}, and V^{S} are the same as for the C^{α}-based model. The V^{BS} potential links C^{β} atom to the C^{α} for the same amino acid by a harmonic tethering term with a minimum at a location $\vec{r}_{\mathrm{i}}^{C^{\beta}}$. The $V_{\alpha-\beta}^{\mathrm{NAT}}$ term represents interaction between four types of native contact pairs. The contact map is determined by the all heavy-atoms overlaps, as this procedure is the most efficient in this case. Now, however, we differentiate between four types of overlap and thus between four types of contacts:

1. $\mathrm{C}_{\mathrm{i}}^{\beta}$ and $\mathrm{C}_{\mathrm{j}}^{\beta}$ (if the side-group-effective atoms overlap).
2. $\mathrm{C}_{\mathrm{i}}^{\alpha}$ and $\mathrm{C}_{\mathrm{j}}^{\beta}$ (if the side group on j overlaps with a C^{α} on contact-type 1).
3. $\mathrm{C}_{\mathrm{i}}^{\beta}$ and $\mathrm{C}_{\mathrm{j}}^{\alpha}$.
4. $\mathrm{C}_{\mathrm{i}}^{\alpha}$ and $\mathrm{C}_{\mathrm{j}}^{\alpha}$ (which arise primarily within secondary structures).

TABLE 3 Value of the energy in the native state

Protein	E^{o}	E^{L1}	E^{L2}	E^{G1}	E^{G2}	E^{G3}	$E^{\nu 1}$	$E^{\nu 2}$	$E^{\mathrm{HB,MJ}}$
1tit	167	166.55	162.26	157.25	163.1	151.65	167	153.26	236.21
1crn	100	100	99.37	96.75	98.70	92.95	100	97.75	124.67
1ubq	171	170.23	164.02	159.25	166.3	154.95	171	167.95	243.30

Calculated using different energy scales for the experimental structures of three proteins, based on contact map M3.

It should be noted that, in the $C^{\alpha-\beta}$ model, the contact may involve between one and four attractive potentials with a minimum. On average, there are ~1.5 contact potentials involved in a pair of amino acids which suggests doing stretching simulations at the correspondingly increased reduced temperature relative to the C^{α}-based model. We consider the usual Lennard-Jones potential and the $V_{\rm exp}^{6\text{-}12}$ potential (Eq. 11) for interaction between native contacts. Each of four possible existing contacts between each pair of amino acids is represented by different $\sigma_{\rm ij}$ with a minimum located at the native distance between the interacting entities (e.g., between $C_{\rm i}^{\beta}$ and $C_{\rm j}^{\beta}$).

The considered energy parameters are either uniform or nonuniform. In the former case, all four possible interactions between the amino acids have the same strengths $E_{\rm ij} = E^{\rm o} = \varepsilon$. In the nonuniform case, we use the energy scale $E_{\rm ij}^{\nu}$, which takes into account the number of the atomic overlaps between residues i and j in the native state, denoted previously as $\nu_{\rm ij}$. The $\nu_{\rm ij}$ numbers are now split into $\nu_{\rm ij}^{\alpha\alpha}$, which denotes the number of backbone-to-backbone overlaps; $\nu_{\rm ij}^{\alpha\beta}$ is the number of backbone-to-side chain overlaps; and the number of side-chain-to-side-chain overlaps is $\nu_{\rm ij}^{\beta\beta}$. The effective energy parameter $E_{\rm ij}$ is then given by

$$E_{\rm ij}^{\nu} = \begin{cases} \varepsilon\nu_{\rm ij}^{\alpha\alpha}/\Omega_{\alpha\alpha} & \text{for } C^{\alpha-\alpha} \\ \varepsilon\nu_{\rm ij}^{\alpha\beta}/\Omega_{\alpha\beta} & \text{for } C^{\alpha-\beta} \\ \varepsilon\nu_{\rm ij}^{\beta\alpha}/\Omega_{\beta\alpha} & \text{for } C^{\beta-\alpha}, \\ \varepsilon\nu_{\rm ij}^{\beta\beta}/\Omega_{\beta\beta} & \text{for } C^{\beta-\beta} \end{cases} \tag{29}$$

where, e.g., $\Omega_{\beta\beta}$ denotes normalization related to the average number of atomic interactions between all $\beta\beta$ contacts in the protein under study.

THE SELECTION OF TEMPERATURE FOR STRETCHING SIMULATIONS

There are several characteristic temperatures that can be associated with a protein. Among them, we distinguish $\tilde{T}_{\rm min}$, $\tilde{T}_{\rm f}$, and $\tilde{T}_{\rm max}$. The first of them characterizes the kinetics, and it corresponds to the temperature of the fastest folding (or the temperature of the least frustration) as determined by the first-passage-time criterion; the second is the folding temperature at which the equilibrium probability of having all native contacts established crosses 1/2; and the third denotes the location of the maximum in the specific heat. In an ideal model, the three temperatures should be near one another. However, for most models considered here, there is a certain shift between their values, and one has to make decisions as to what temperature to pick for stretching studies. Furthermore, $\tilde{T}_{\rm min}$ and, especially, $\tilde{T}_{\rm f}$ depend on specific criteria that test establishment of a native contact. $\tilde{T}_{\rm f}$ and $\tilde{T}_{\rm max}$ also depend on the duration of simulations and statistics. We have opted for performing stretching at or near $\tilde{T}_{\rm min}$, since this temperature appears to correspond to a kinetically optimal state of the model protein, i.e., when the model is most proteinlike. In many models, including the simplest uniform-energy-scale Lennard-Jones model, this choice is close to the effective room temperature (4,5), whereas the model $\tilde{T}_{\rm max}$ often exceeds it substantially.

Fig. 3 shows the dependence of the folding time on temperature for crambin based on models with the uniform energy scale for a sample of the potentials. The value $t_{\rm fold}$ has a U-shaped dependence on T that is centered on $\tilde{T}_{\rm min}$. There is a clear distinction between the models that use $V^{\rm C}$ to account for the local stiffness and $V^{\rm A}$. The former have $\tilde{T}_{\rm min}$ at ~0.3 whereas the latter correspond to a $\tilde{T}_{\rm min}$ which is higher. For crambin, $\tilde{T}_{\rm min}$ is twice as high. Also, the angular stiffness typically leads to a broader region of temperatures where folding is optimal.

Fig. 4 summarizes results on $\tilde{T}_{\rm min}$ and the approximate temperature range of the optimal folding for all models considered here and for three proteins: crambin, titin, and ubiquitin. It shows that the values of $\tilde{T}_{\rm min}$ for various models cluster around 0.3 if $V^{\rm C}$ is used. If $V^{\rm A}$ is used instead, then the resulting $\tilde{T}_{\rm min}$ clusters around three values: 0.8 (for $V^{6\text{-}10\text{-}12}$ with $E^{\rm MJ}$), 0.7 (for $V^{6\text{-}12}$ or $V_{\rm exp}^{-12}$ with $E^{\rm MJ}$), and 0.6 (for remaining models). Thus, whenever we use $V^{\rm C}$, we simulate stretching at $\tilde{T} = 0.3$. Otherwise, the simulations are performed at $\tilde{T}$ equal to 0.6, 0.7, and 0.8, and the use of these higher temperatures is indicated by the symbols *, †, and ‡, respectively.

The upward shift in $\tilde{T}_{\rm min}$ on replacing $V^{\rm C}$ by $V^{\rm A}$ is due to the fact that the latter incorporates an extra stability term involving the bond angles. The right panel of Fig. 4 shows that the enhanced stiffness associated with the potential $V^{\rm A}$

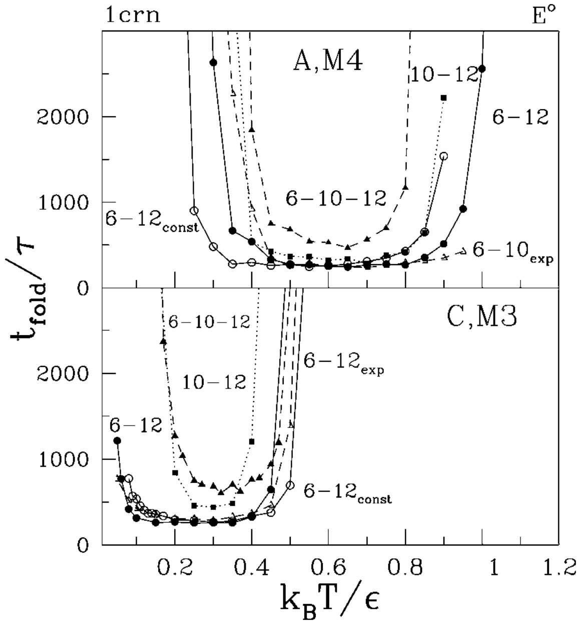

FIGURE 3 The dependence of the folding time on temperature for various models for crambin. The top panel is for $V^{\rm C}$ whereas the bottom panel is for $V^{\rm A}$. The solid line with circles $V^{6\text{-}12}$, solid line with open circles $V_{\rm const}^{6\text{-}12}$, solid (*dashed*) with solid square $V^{10\text{-}12}$, dotted with solid triangle $V^{6\text{-}10\text{-}12}$ and dotted with open triangle $V_{\rm exp}^{6\text{-}12}$. The same notation is used for arrows which indicate position of the folding temperature $\tilde{T}_{\rm f}$.

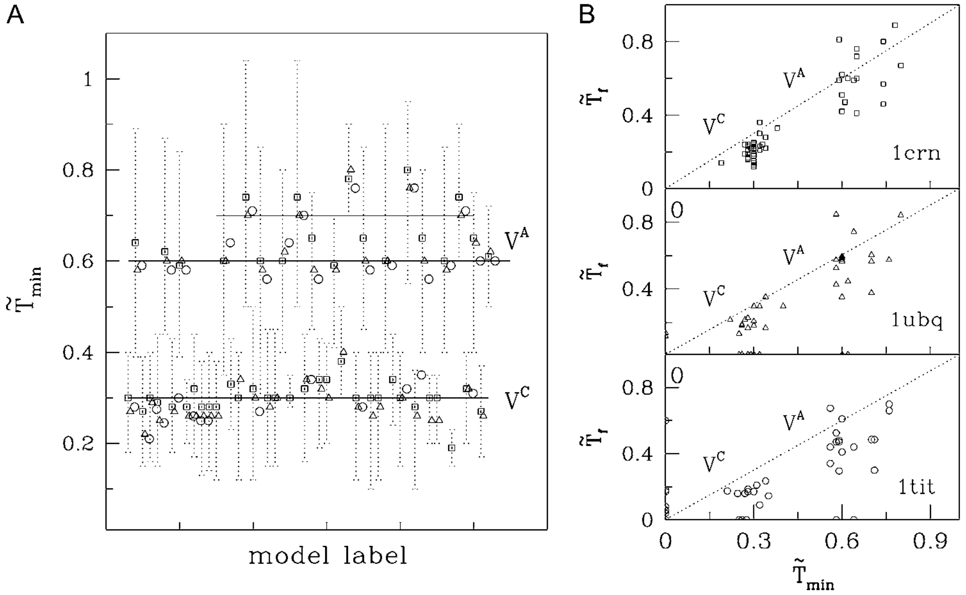

FIGURE 4 (*Left panel*) the values of the $\tilde{T}_{\rm min}$ for proteins 1crn (*square*), 1ubq (*triangle*), and 1tit (*circle*) based on all models considered here. The dotted error bars show the approximate temperature range of the optimal folding kinetics. Solid lines shows the main trend in the temperature which corresponds to the $\tilde{T}_{\rm min} = 0.3$, 0.6, 0.7. (*Right panel*) Correlation between $\tilde{T}_{\rm min}$ and $\tilde{T}_{\rm f}$ for proteins 1crn, 1ubq, and 1tit, top to bottom, respectively.

leads both to a larger thermodynamic stability (a larger value of $\tilde{T}_{\rm f}$) and to an enhanced value of $\tilde{T}_{\rm min}$. The zero values of $\tilde{T}_{\rm min}$ in this figure indicate lack of folding. The zero values of $\tilde{T}_{\rm f}$ indicate values which are smaller than 0.05.

The degree to which $F_{\rm max}$ and the *F-d* patterns are sensitive to the choice of the temperature is model-dependent. In the case of the $V^{\rm A}$-based local stiffness, the temperature range of the optimal folding is usually broader than with $V^{\rm C}$ and, within this range, the sensitivity is especially acute for the $V^{6\text{–}10\text{–}12}$ and $V^{10\text{–}12}$ potentials with $E^{\rm HB,\ MJ}$ as illustrated in Fig. 5. For the first of these potentials, stretching should be done at $\tilde{T}_{\rm f}{\sim}0.75$, which is above $\tilde{T}_{\rm min}$, to match the experimental position of the major peak and the type of the *F-d* pattern. For the second potential, there is no difference between stretching at $\tilde{T}_{\rm min}$ and $\tilde{T}_{\rm f}$, but they both differ from stretching at $\tilde{T} = 0.3$ significantly. However, if we take proteins from another class, like the helical proteins, we can also see high sensitivity to the choice of temperature. In such cases, it is better to do unfolding at a temperature which is lower than $\tilde{T}_{\rm min}$. The choice of the temperature for stretching can change R^2 substantially in these models and then the criterion for the selection of T is provided by selection of the largest value of R^2. The resulting choices of the T are indicated by asterisks in the tables.

Using $V^{\rm C}$ is physically well motivated and has advantages compared to $V^{\rm A}$, because the resulting relevant temperature ranges are narrower and quite similar for various proteins, and $T_{\rm f}$ is closer to $T_{\rm min}$.

RESULTS

Correlations between the theoretical and experimental values of $F_{\rm max}$

Table 4 summarize the statistical assessment of the performance of various C^{α}-based models as grouped by the choice of the functional form of the contact potentials. The last row of Table 4 shows a similar assessment for the $C^{\alpha-\beta}$-based models. The assessment is quantified by the parameters R^2 and U and the best slope a. The coefficient a serves as translation factor between the theoretical (e.g., ε/Å) and the experimental (pN) force units. The first entry is for the {6-12, C, M3, $E^{\rm o}$} model and its coefficient a is equal to 0.0140. This value means that ε/Å corresponds to $1/a = 71$ pN, i.e., ε is of ~1 kcal/mol which is equivalent to ~500 K. In our previous article (5), we have reported the value of 67 pN for

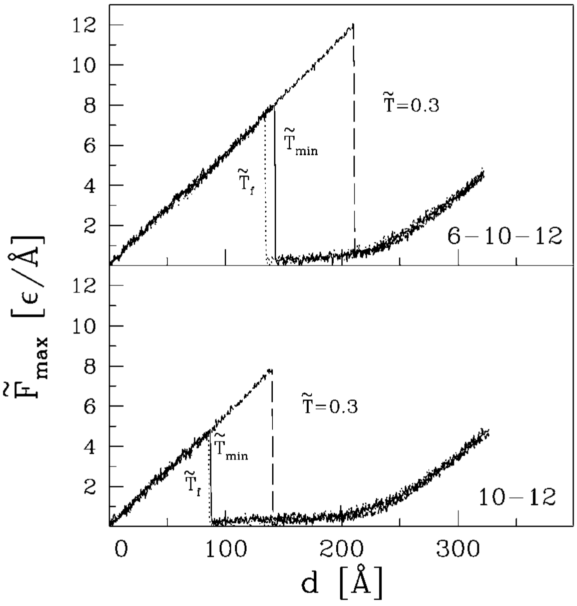

FIGURE 5 Dependence of *F-d* pattern on the temperature of stretching when the $V^{\rm A}$ backbone stiffness is used. The dashed, solid, and dotted lines are for stability temperature $\tilde{T}_{\rm f}$, optimal folding temperature $\tilde{T}_{\rm min}$, and $\tilde{T} =$ 0.3 which is outside of the range of good folding. The top and bottom panels are for the 6-10-12 and 10-12 potentials, respectively.

TABLE 4 Results of comparison between experimentally measured F^{e}_{max} and theoretical predictions based on 6-12 potential for different models

Model	a	Δa	R^2	U	Model	a	Δa	R^2	U
6-12					6-12$_{exp}$				
$\{C, \mathrm{M3}, E^{o}\}$	0.0140	0.0047	0.89	0.22	$\{C, \mathrm{M3}, E^{o}\}$	0.0153	0.0051	0.85	0.25
$\{C, \mathrm{M2}c75, E^{o}\}$	0.0160	0.0053	0.65	0.39	$\{A, \mathrm{M3}, E^{o}\}*$	0.0179	0.0024	0.79	0.29
$\{C, \mathrm{M3}c7, E^{o}\}$	0.0144	0.0051	0.73	0.36	$\{C, \mathrm{M3}c75, E^{o}\}$	0.0185	0.0068	0.77	0.33
$\{C, \mathrm{M3}c75, E^{o}\}$	0.0164	0.0042	0.77	0.33	$\{C, \mathrm{M3}, E^{\nu 2}\}$	0.0164	0.0055	0.82	0.28
$\{C, \mathrm{M2}, E^{o}\}$	0.0122	0.0046	0.85	0.23	$\{C, \mathrm{M3}, E^{\mathrm{HB,MJ}}\}$	0.0223	0.0076	0.77	0.30
$\{C_{0.5}, \mathrm{M3}, E^{o}\}$	0.0137	0.0047	0.88	0.23	$\{A, \mathrm{M3}, E^{\mathrm{HB,MJ}}\}^{\dagger}$	0.0241	0.0083	0.83	0.26
$\{A, \mathrm{M4}, E^{o}\}$	0.0256	0.0058	0.79	0.29	10-12				
$\{A, \mathrm{M4}, E^{o}\}*$	0.0162	0.0055	0.69	0.35	$\{C, \mathrm{M2}, E^{o}\}$	0.0167	0.0058	0.75	0.32
$\{C, CSU, E^{o}\}$	0.0150	0.0049	0.86	0.21	$\{C, \mathrm{M3}, E^{o}\}$	0.0144	0.0050	0.84	0.26
$\{C, \mathrm{M3}, E^{\mathrm{L1}}\}$	0.0142	0.0050	0.77	0.31	$\{C, \mathrm{M3}c75, E^{o}\}$	0.0169	0.0065	0.79	0.33
$\{C, \mathrm{M3}, E^{\mathrm{L2}}\}$	0.0146	0.0018	0.82	0.27	$\{A, \mathrm{M4}, E^{o}\}$	0.0289	0.0094	0.85	0.25
$\{C, \mathrm{M3}, E^{\mathrm{G2}}\}$	0.0132	0.0016	0.83	0.28	$\{A, \mathrm{M4}, E^{o}\}*$	0.0174	0.058	0.73	0.33
$\{A, \mathrm{M4}, E^{\mathrm{G2}}\}$	0.0250	0.0080	0.76	0.31	$\{C, \mathrm{M3}, E^{\nu 1}\}$	0.0136	0.0044	0.71	0.34
$\{C, \mathrm{M3}, E^{\nu 1}\}$	0.0142	0.0044	0.79	0.23	$\{C, \mathrm{M3}, E^{\mathrm{HB,MJ}}\}$	0.0213	0.0074	0.67	0.36
$\{C, \mathrm{M3}, E^{\nu 2}\}$	0.0147	0.0049	0.67	0.36	$\{A, \mathrm{M4}, E^{\mathrm{HB,MJ}}\}$	0.0388	0.0125	0.87	0.24
$\{A, \mathrm{M4}, E^{\nu 1}\}*$	0.0156	0.0053	0.66	0.36	$\{A, \mathrm{M4}, E^{\mathrm{HB,MJ}}\}^{\dagger}$	0.0242	0.0083	0.82	0.28
$\{C, \mathrm{M3}, E^{\mathrm{HB,MJ}}\}$	0.0209	0.0069	0.87	0.23	6-10-12				
$\{A, \mathrm{M4}, E^{\mathrm{HB,MJ}}\}$	0.0332	0.0076	0.83	0.27	$\{C, \mathrm{M2}, E^{o}\}$	0.0167	0.0081	0.75	0.32
$\{A, \mathrm{M4}, E^{\mathrm{HB,MJ}}\}*$	0.0223	0.0042	0.84	0.26	$\{C, \mathrm{M3}c75, E^{o}\}$	0.0209	0.0077	0.83	0.27
6-12$_{const}$					$\{C, \mathrm{M3}, E^{o}\}$	0.0154	0.0051	0.81	0.28
$\{C, \mathrm{M2}, E^{o}\}$	0.014	0.0059	0.81	0.28	$\{A, \mathrm{M4}, E^{o}\}*$	0.0208	0.0073	0.81	0.28
$\{C, \mathrm{M}c75, E^{o}\}$	0.016	0.0062	0.78	0.33	$\{C, \mathrm{M3}, E^{\nu 1}\}$	0.0190	0.0060	0.66	0.36
$\{C, \mathrm{M3}, E^{o}\}$	0.014	0.0050	0.77	0.31	$\{C, \mathrm{M3}, E^{\mathrm{HB,MJ}}\}$	0.0268	0.0089	0.76	0.32
$\{A, \mathrm{M3}, E^{o}\}^{\ddagger}$	0.020	0.0067	0.76	0.31	$\{A, \mathrm{M4}, E^{\mathrm{HB,MJ}}\}$	0.0493	0.0176	0.87	0.24
$\{C, \mathrm{M3}, E^{\nu 1}\}$	0.015	0.0051	0.79	0.29	$\{A, \mathrm{M4}, E^{\mathrm{HB,MJ}}\}^{\dagger}$	0.0296	0.0108	0.83	0.29
$\{C, \mathrm{M3}, E^{\mathrm{HB,MJ}}\}$	0.021	0.0074	0.80	0.29	Morse				
Mixed 10-12 with					$\{C, \mathrm{M3}, E^{o}\}$	0.015	0.0051	0.83	0.26
$\{\text{6-12}, C, \mathrm{M3}, E^{o}\}$	0.014	0.0047	0.87	0.23	$\{A, \mathrm{M3}, E^{o}\}$	0.029	0.0092	0.73	0.32
$\{\text{6-12}, A, \mathrm{M4}, E^{o}\}*$	0.016	0.0053	0.74	0.32	$\{A, \mathrm{M3}, E^{o}\}*$	0.018	0.0088	0.78	0.30
$\{\text{6-12}_{const}, A, \mathrm{M4}, E^{o}\}*$	0.018	0.0054	0.80	0.29	$\mathrm{C}^{\alpha-\beta}$				
$\{\text{6-12}_{const}, C, \mathrm{M3}, E^{o}\}$	0.015	0.0052	0.79	0.32	$\{\text{6-12}, C, \mathrm{M2}, E^{o}\}$	0.0220	0.0082	0.79	0.31
					$\{\text{6-12}, C, \mathrm{M3}, E^{o}\}$	0.0220	0.0074	0.81	0.30
					$\{\text{6-12}, C, \mathrm{M2}, E^{\nu 2}\}$	0.0278	0.0097	0.79	0.34
					$\{\text{10-12}, C, \mathrm{M2}, E^{o}\}$	0.0301	0.0086	0.81	0.32

The data are fitted to a line $F^{t}_{max} = a^{*}F^{e}_{max}$. Δa denotes deviation of particular F^{t}_{max} from the fit. R^2 denotes correlation coefficient and U denotes Theil's U-Statistic, both are described in the text. The symbol $C_{0.5}$ indicates reduction in the amplitude in the chirality term by two compared to the standard value.

*Indicates calculations performed at $\tilde{T} = 0.6$.

†Indicates calculations performed at $\tilde{T} = 0.7$.

‡Indicates calculations performed at $\tilde{T} = 0.8$.

the same model. The current result incorporates one more protein (protein G) and is based on bigger trajectory statistics. In most cases, we considered 10 trajectories and picked the most common behavior in case several pathways were possible, since some proteins show more than one pathway to fold.

The standard deviations of the theoretical results away from the best slope are characterized by the parameter Δa. For the $\{$6-12, C, M3, $E^{o}\}$ model, it is equal to 0.0047 indicating that for most proteins, the effective value of ε/Å ranges between 58 and 108 pN. Ubiquitin and titin and several other proteins are better described by the larger value which corresponds to ε of ~1.5 kcal/mol, i.e., 750 K. In this case, the room temperature is close to 0.35 ε/k_B. For the general trend, however, 0.55 ε/k_B would seem more appropriate. On the other hand, the temperature of optimal folding (0.3 ε/k_B) appears to be nearly common for the proteins studied, suggesting that it also could play the role of the effective room temperature.

It should be noted that the experimental results on stretching have been obtained not necessarily at the same pulling speed (even though we tended to pick 600 nm/s whenever possible) whereas all theoretical results correspond to one speed. This fact introduces additional uncertainties in the assessment and in the value of Δa.

A graphical representation of all of these results is shown in Fig. 6 where each model is located on the R^2–U plane. The best models are those for which R^2 is large and U is small. However, a correct model should also display a reasonable folding behavior.

Examples of models that fail the folding test are $\{$6-12, C, M3, $\mathrm{E}^{\mathrm{HB,\ MJ}}\}$ and $\{$(6-12, 10-12),C, M3, $E^{o}\}$. The four best

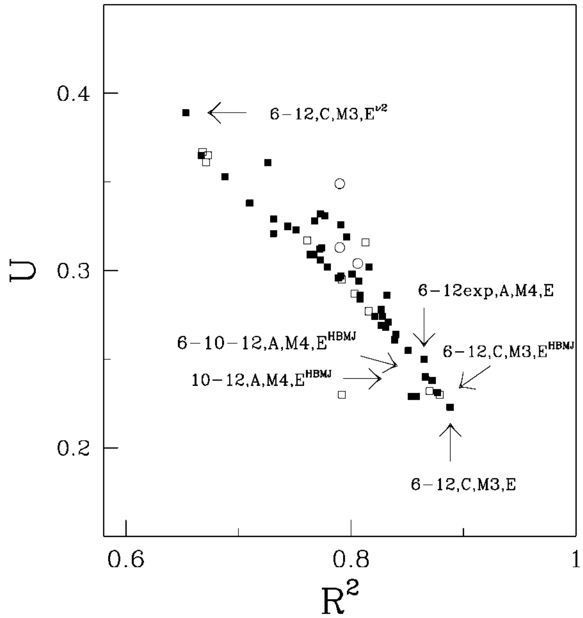

FIGURE 6 Correlation between R^2 and U coefficients. The squares correspond to the C^{α}-based models and circles to the $C^{\alpha\text{-}\beta}$-based models. The solid symbols indicate good folding properties and open symbols, poor folding properties. The overall appearance of the plot suggests existence of a relationship between R^2 and U, indicating that the two quantities are not independent of each other.

stretching models that employ V^C (and also lead to folding) correspond to: {6-12, C, M3, E^o}, i.e., the uniform Lennard-Jones potential with the chirality-described stiffness and no i, $i+2$ contacts; {6-12$_{\text{exp}}$, C, M3, E^o} where, instead of 6-12, we used 6-12 with second minimum at 3 Å; {10-12, C, M3, E^o}; and {6-10-12, C, M3, E^o}, which correspond to R^2 of 0.89, 0.87, 0.85, and 0.84, respectively. The correlation plots for the first two models, together with the one that is the worst, are shown in Fig. 7. Even though the model {(6-12, 10-12), C, M3, E^o} has the second-ranked R^2 (of 0.87), it has to be rejected because it does not lead to folding for 1tit. The remaining good models should perform quite similarly in practice.

When we split proteins into structural classes, we find that the uniform energy scales work for the α-β class proteins with the better ($R^2 \sim 0.82$) and the best energy scales E^{G1} or E^{G2} ($R^2 \sim 89$), although E^{MJ} seems to be more adequate for the β-class proteins ($R^2 \sim 83$) (note that the F-d patterns for the α-proteins with the E^{MJ} scale are poor when compared to the experimental data ($R^2 \sim 64$)). Another observation is that the M3 contact map is better than the M2 map.

For the models with the 6-12 potential, good correlations with the experimental data are also found for models with the energy scales $E^{L1,\ L2}$, $E^{G1,\ G2,\ G3}$ and their combinations, where the strength of the interaction between side groups is lower than for others native contacts (data not shown). These models work for the α-class proteins pretty well.

The best four models that employ V^A correspond to {6-12, A, M4, $E^{HB,MJ}$}, {6-10-12, A, M4, $E^{HB,MJ}$}(49), {10-12, A, M4, E^o}, and {10-12, A, M4, $E^{HB,MJ}$}. The commonly used model of Clementi et al. (56) has a somewhat lower correlation coefficient of $R^2 \sim 0.81$. The data points for the model of Karanicolas and Brooks (49) are shown in Fig. 7. For the models with $V^{A,}$ the R^2 coefficient is rather sensitive to the choice of the temperature and can be in the range from 0.60 to 0.87 for the same model while still being within the temperatures which are optimal for folding. For instance, in the case of $V^{6\text{-}10\text{-}12}$, $R^2 = 0.82$ for stretching at $\tilde{T} = 0.8$ and 0.86 for stretching at $\tilde{T} = 0.5$, where the first choice corresponds to $\tilde{T}_{\min} = 0.8$ for 1tit and the second to $\tilde{T}_{\min} = 0.5$ for 1aj3.

Poor correlations with the experimental data on $F_{\max}$ come with models incorporating $E^{\nu 2}$, but not $E^{\nu 1}$, for almost all V^{NAT} with V^C and models contained parameters 10-12, C, $E^{MJ,HB}$. These models also lead to bad folding, at least for 1tit.

The assessment of the models has been accomplished by using one pulling speed of 0.005 Å/τ. Most of the experimental data we used corresponded to pulling speeds that cluster around 600 nm. The dependence of $F^e_{\max}$ on v_p is often found to be logarithmically weak. Thus, small variations in the experimental v_p are expected to be of no consequence for the assessment. However, six entries in Table 1 correspond either to a much larger value of v_p (1aj3 and four cases of 1emb) or to a much lower (1n11). To determine the degree of robustness of our results, we reevaluate R^2 and U based on two new ensembles: one obtained by removing the six proteins from the set and another in which all proteins are kept but the theoretical pulling speeds for the six outliers are either scaled up or down in proportion to the relation of the experimental pulling speed to 600 nm/s. All this is done only for the best 15 models selected based on Table 4. The resulting values of R^2 and U are compared to the same-speed-results in Table 5.

Generally, the results are found to be robust. The removal of the six proteins makes R^2 somewhat smaller and still selects {6-12, C, M3, E^o} as the best model. However, when taking the speed variations into account we pick three nearly equivalent models as winners: {6-12, C, M3, E^o}, {10-12, A, M4, $E^{HB,\ MJ}$}, and {6-10-12, A, M4, $E^{HB,\ MJ}$} that have R^2 close to 0.84 and U close to 0.26 with {6-12, A, M4, $E^{HB.MJ}$} and {6-12$_{\text{exp}}$, A, M4, $E^{HB.MJ}$} coming in the close second tier.

Models with the side groups

The C^{α}-C^{β}-based models are studied only with the V^C potential. The $\tilde{T}_{\min}$ is ~0.4 and this is the temperature used in the simulations (see Table 7). Table 4 suggests that the best model corresponds to the $V^{10\text{-}12}$ potential with the uniform energy scale. However, its R^2 is still lower than in the case of the C^{α}-based Gō model. We also find that that $E^{\nu 2}$ is very good for the α–β proteins, but it is poor for the α-proteins.

The form of the *F-d* patterns

$F_{\max}$ sets the characteristic scale for the force but it does not relate to the appearance of the whole F-d pattern. Various

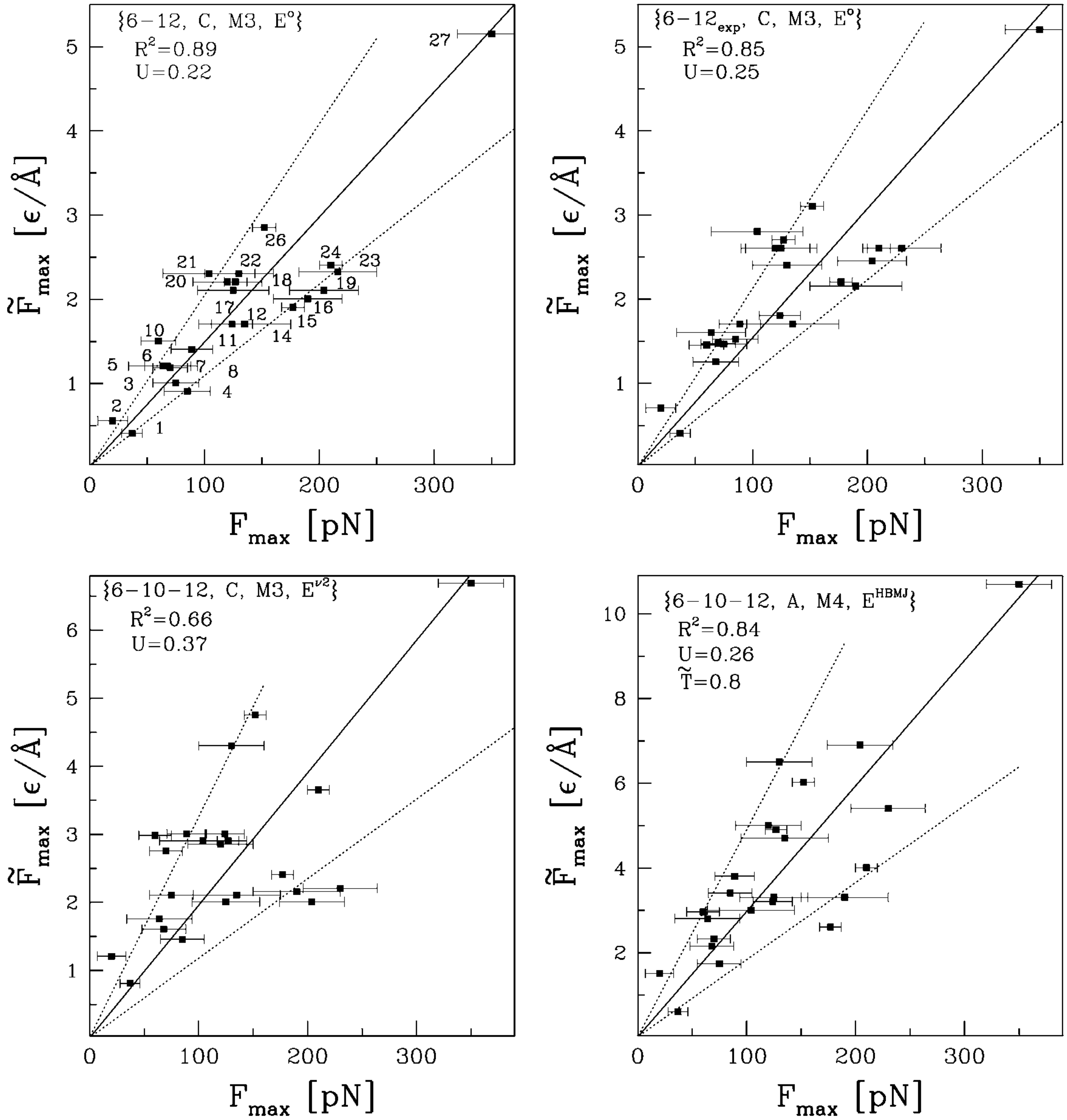

FIGURE 7 Correlation between the experimental F^{e}_{max} and theoretical F^{t}_{max} for four models. The top two panels are the best working models with the V^{C} potential: {6-12, *C*, M3, *E*°}, and {6-12$_{exp}$, *C*, M3, *E*°}. The left bottom panel illustrates a poorly working model, {6-10-12, *C*, M3, $E^{\nu 2}$}, corresponding to R^2 of 0.66. The right bottom panel is for one of the best performing models with the V^{A} stiffness: {6-10-12, *A*, M4, $E^{HB,MJ}$}. The numbers in the top left panel indicate particular proteins. These are: 1(1n11), 2(1cfc), 3(1hci), 4(^{10}*FNII*), 5(1u4q), 6(1aj3), 7(B), 8(1ubq$_{48\text{–}N}$), 9(1b6i), 10(1rsy), 11(^{13}FNII), 12(^{12}FNIII), 14(TNFNIII), 15(1qjo$_{N\text{–}41}$), 16(G), 17(^{1}FNII), 18(I1), 19(I27), 20(1emb), 21(1emb$_{132\text{–}212}$), 22(1emb$_{3\text{–}212}$), 23(1ubq), 24(1nct), 25(1glc), 26(L), 27(1emb$_{3\text{–}132}$), 28(1vsc). *B*, *L*, and *G* denote barnase, protein L, and protein G, respectively.

models may, in principle, lead to different patterns. However, if we use the uniform energy scale and V^{C} then any potential studied here leads to similar force patterns. For instance, for a single domain of 1tit a major maximum is followed by a minor second peak. It should be noted that in tandem linkages of several domains the minor peaks are usually shed in the pattern, except in the initial segment (16). (Here, we study only single domain situations—these yield robust values of F_{max} though not necessarily robust *F-d* patterns when used in multiple linkages.) When one uses V^{A} then more variety appears. For instance, the $V^{6\text{-}10\text{-}12}$ potential with the uniform energy scale does not lead to the emergence of the minor peak for one domain of 1tit.

Nonuniform energy scales lead to a greater variety in the resulting force patterns. For instance, the $E^{\nu 2}$ and $E^{HB,MJ}$ energy scales have respectively the worst and the best influence on the *F-d* pattern for any V^{NAT} when confronted with the experiment. The corresponding patterns for 1tit are shown in Fig. 8. The $E^{\nu 2}$ scale yields a pattern which bears no resemblance to the experimental *F-d* curves for 1tit, 1ubq,

TABLE 5 Similar to Table 4 but for selected 15 best models

Model	a	Δa	R^2	U	Model	a	Δa	R^2	U
6-12					6-10-12				
$\{C$, M3, $E^{\mathrm{o}}\}$	0.0140	0.0047	0.89	0.22	$\{C$, M3, $E^{\mathrm{o}}\}$	0.0154	0.0051	0.81	0.28
	0.0129	0.0059	0.79	0.26		0.0147	0.0051	0.72	0.32
	0.0141	0.0051	0.84	0.26		0.0174	0.0051	0.79	0.30
$\{A$, M4, $E^{\mathrm{o}}\}$	0.0256	0.0058	0.79	0.29	$\{A$, M4, $E^{\mathrm{o}}\}$*	0.0208	0.0073	0.81	0.28
	0.0255	0.0064	0.72	0.27		0.0200	0.0086	0.71	0.28
	0.0201	0.0078	0.73	0.41		0.0212	0.0072	0.76	0.32
$\{A$, M4, $E^{\mathrm{o}}\}$*	0.0162	0.0055	0.69	0.35		0.0212	0.0072	0.76	0.32
	0.0158	0.0085	0.59	0.33	$\{C$, M3, $E^{\mathrm{HB,MJ}}\}$	0.0268	0.0089	0.76	0.32
	0.0173	0.0059	0.72	0.32		0.0244	0.0095	0.72	0.27
$\{C$, M3, $E^{\mathrm{HB,MJ}}\}$	0.0209	0.0069	0.87	0.23		0.0264	0.0091	0.75	0.31
	0.0196	0.0073	0.75	0.26	$\{A$, M4, $E^{\mathrm{HB,MJ}}\}$	0.0493	0.0176	0.87	0.24
	0.0218	0.0078	0.72	0.32		0.0483	0.0196	0.76	0.27
$\{A$, M4, $E^{\mathrm{HB,MJ}}\}$	0.0332	0.0076	0.83	0.27		0.0501	0.0180	0.83	0.26
	0.0332	0.0089	0.69	0.28	$\{A$, M4, $E^{\mathrm{HB,MJ}}\}$†	0.0296	0.0108	0.83	0.29
	0.0332	0.0081	0.82	0.26		0.0278	0.0137	0.69	0.32
$\{A$, M4, $E^{\mathrm{HB,MJ}}\}$*	0.0223	0.0042	0.84	0.26		0.0301	0.0126	0.79	0.28
	0.0217	0.0053	0.71	0.30	10-12				
	0.0219	0.0044	0.80	0.29	$\{C$, M3, $E^{\mathrm{o}}\}$	0.0144	0.0050	0.84	0.26
6-12$_{\mathrm{exp}}$						0.0130	0.0073	0.71	0.29
$\{C$, M3, $E^{\mathrm{o}}\}$	0.0153	0.0051	0.85	0.25		0.0149	0.0064	0.79	0.29
	0.0146	0.0060	0.74	0.26	$\{A$, M4, $E^{\mathrm{HB,MJ}}\}$	0.0388	0.0125	0.87	0.24
	0.0164	0.0058	0.78	0.28		0.0323	0.0149	0.76	0.29
$\{A$, M3, $E^{\mathrm{HB,MJ}}\}$†	0.0241	0.0083	0.83	0.26		0.0353	0.0131	0.84	0.25
	0.0228	0.0091	0.72	0.27					
	0.0246	0.0086	0.80	0.28					

Each model comes with three lines. The first line is a repeat of the corresponding entry in Table 4. The second line is for the situations in which the six outliers in terms of the pulling speed are not taken into consideration. The third line corresponds to the situation in which all proteins are included but the pulling speeds are either increased to 0.03 Å/τ for 1aj3 and 1emb or decreased to 0.0001 Å/τ for 1n11.

*Indicates calculations performed at $\tilde{T} = 0.6$.

†Indicates calculations performed at $\tilde{T} = 0.7$.

and all FNIII proteins, especially in combination with $V^{6\text{-}12}$ and $V^{6\text{-}10\text{-}12}$ with V^{C}.

Energy scale $E^{\mathrm{HB,MJ}}$ with V^{C} generally makes the F-d pattern look rougher (sometimes additional small force peaks are observed). We have found that for this energy scale taken with the potential $V^{6\text{-}10\text{-}12}$ or $V^{6-12}_{\mathrm{const}}$ one obtains a small shoulder peak on the rising side of the major peak in 1tit, which agrees with the experimental result (66). Thus, the additional roughness may actually have a physical meaning.

All models with the uniform energy parameters yield a major peak followed by an after-peak in the F-d pattern for one domain of 1tit. (For tandem linkages of 1tit, the after-peak remains only in the first period of unwinding.) The exception is the $\{6\text{-}10\text{-}12, A, \mathrm{M4}, E^{\mathrm{o}}\}$ model which yields no after-peak. However, when we replace the uniform energy parameters by $E^{\mathrm{HB,MJ}}$ then all V^{NAT} potentials, but $V^{6\text{-}12}_{\mathrm{exp}}$, do not generate the after-peak for 1tit.

It has to be noted that even though the calculated R^2 coefficients for models with the energy scales like E^{G1}, E^{G2} and E^{L1}, E^{L2} are low, the corresponding F-d patterns for helical proteins appear to be closer to the experimental plots. This is the reason why we observe good correlation for those proteins in one of the worst models, as shown in one the bottom panel of Fig. 7.

To summarize, even though the correlation levels between several models are similar, analysis of the F-d pattern shows that the model with the uniform energy scale and the $V^{6\text{-}12}$ with V^{C} potential usually reproduces the experimental shapes of the F-d patterns very well. These patterns are not improved by adjustments in the strength of V^{C} and by considering the CSU contact map. The corresponding values of $F_{\max}$ are listed in Table 1.

THERMODYNAMICS AND FOLDING

We now discuss performance of the models in nonstretching situations. Gō models are not reliable far away from the native state so we do not use the experimental folding temperatures and experimental folding times as benchmarks. However, it is interesting to check the behavior of the models that have bean selected by the mechanical benchmarks.

The theoretical values of $\tilde{T}_{\mathrm{f}}$, $\tilde{T}_{\min}$, and t_{fold} for 1crn, 1ubq, and 1tit for the 62 various models are shown in Tables 6 and 7. We have found that only 1crn folds to the native state in all models. Other proteins fold only in a subset of the models.

In an earlier study (14) it has been shown, based on the 51 different proteins, that $\tilde{T}_{\min}$ in the case of the $\{6\text{-}12, \mathrm{M2}, C, E^{\mathrm{o}}\}$ model depends on the class of proteins and the length N

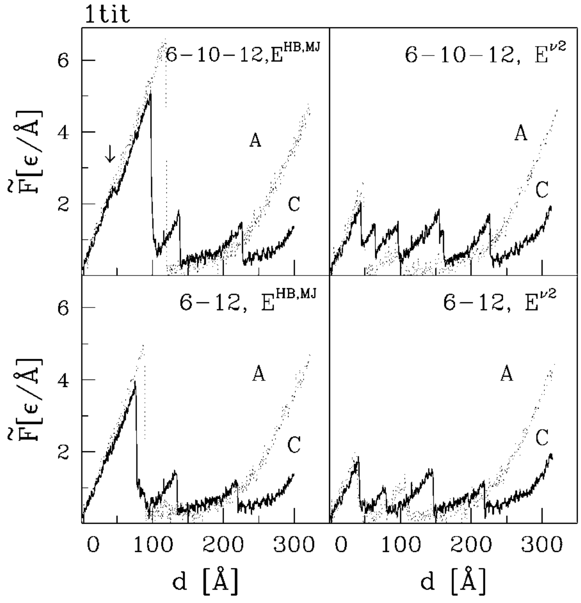

FIGURE 8 The dependence of the *F-d* patterns on the choice of the backbone stiffness potentials for the nonuniform energy scales. The dotted and solid lines are for V^A and V^C, respectively. The left and right panels are for the energy scales $E^{HB,MJ}$ and $E^{\nu 2}$, respectively. The top and bottom panels are for the 6-10-12 and 6-12 potentials, respectively.

TABLE 6 Stability temperature $\tilde{T}_f$, optimal folding temperature $\tilde{T}_{min}$, and the folding time t_{fold}/τ at $\tilde{T}_{min}$ for 1crn, 1ubq, 1tit for the models listed

		1crn		1ubq		1tit	
Model	$\tilde{T}_f$	$\tilde{T}_{min}$	t_{fold}/τ	$\tilde{T}_{min}$	t_{fold}/τ	$\tilde{T}_{min}$	t_{fold}/τ
6-12							
$\{C, \mathrm{M3}, E^{o}\}$	0.24	0.3	256	0.27	460	0.25	3926
$\{A, \mathrm{M4}, E^{o}\}$	0.59	0.64	243	0.58	459	0.59	1055
$\{C, \mathrm{M3}, E^{L1}\}$	0.24	0.28	328	0.22	773	0.21	4100
$\{C, \mathrm{M3}, E^{L2}\}$	0.25	0.30	337	0.29	824	0.27	2817
$\{C, \mathrm{M3}, E^{G1}\}$	0.22	0.29	318	0.25	747	0.24	6300
$\{A, \mathrm{M4}, E^{G1}\}$	0.60	0.62	216	0.60	419	0.58	1283
$\{C, \mathrm{M3}, E^{G2}\}$	0.23	0.28	318	0.27	696	0.30	5038
$\{A, \mathrm{M4}, E^{G2}\}$	0.59	0.59	225	0.60	460	0.58	1039
$\{C, \mathrm{M3}, E^{G3}\}$	0.21	0.28	351	0.26	766	0.26	4500
$\{C, \mathrm{M3}, E^{G1,\ L1}\}$	0.23	0.32	317	0.26	755	0.25	4947
$\{C, \mathrm{M3}, E^{G2,\ L1}\}$	0.24	0.28	320	0.26	778	0.25	3800
$\{C, \mathrm{M3}, E^{\nu 1}\}$	0.17	0.28	336	0.26	771	—	—
$\{C, \mathrm{M3}, E^{\nu 2}\}$	0.16	0.28	322	0.26	751	—	—
$\{A, \mathrm{M4}, E^{\nu 2}\}$	0.62	0.60	209	0.60	443	0.64	1561
$\{C, \mathrm{M3}, E^{HB,MJ}\}$	0.22	0.30	300	0.34	1060	—	—
$\{A, \mathrm{M4}, E^{HB,MJ}\}$	0.80	0.74	229	0.70	348	0.72	1337
$6\text{-}12_{const}$							
$\{C, \mathrm{M3}, E^{o}\}$	0.21	0.32	265	0.3	563	0.27	6647
$\{A, \mathrm{M3}, E^{o}\}$	0.51	0.60	247	0.58	544	0.56	1296
$\{C, \mathrm{M3}, E^{\nu 1}\}$	0.14	0.3	388	0.28	908	—	—
$\{C, \mathrm{M3}, E^{\nu 2}\}$	0.15	0.3	375	0.3	915	—	—
$\{A, \mathrm{M4}, E^{\nu 2}\}$	0.51	0.6	214	0.62	563	0.64	2680
$\{C, \mathrm{M3}, E^{HB,MJ}\}$	0.19	0.3	664	—	—	—	—
$\{A, \mathrm{M4}, E^{HB,MJ}\}$	0.57	0.74	338	0.74	905	0.70	1500
6-10-12							
$\{C, \mathrm{M3}, E^{o}\}$	0.36	0.32	603	0.34	1053	0.34	7500
$\{A, \mathrm{M4}, E^{o}\}$	0.76	0.65	519	0.58	851	0.56	2500
$\{C, \mathrm{M3}, E^{\nu 1}\}$	0.28	0.34	839	0.32	1880	—	—
$\{C, \mathrm{M3}, E^{\nu 2}\}$	0.28	0.34	802	0.30	1832	—	—
$\{A, \mathrm{M4}, E^{\nu 2}\}$	0.81	0.59	485	0.58	910	—	—
$\{C, \mathrm{M3}, E^{HB,MJ}\}$	0.33	0.38	875	0.4	8034	—	—
$\{A, \mathrm{M4}, E^{HB,MJ}\}$	0.89	0.78	601	0.80	1183	0.76	3259

The long-dash symbol indicates lack of folding in this variant of the model.

but most proteins have optimal folding temperatures of $\sim\tilde{T} = 0.3$. In case of the $\{$10-12, M2, C, $E^{o}\}$ model, the values of $\tilde{T}_{min}$ cluster at ~0.225 and do not depend on N. In both cases, the values of $\tilde{T}_f$ are slightly below $\tilde{T}_{min}$, however, for the 10-12 potential the separation is bigger. We now make a similar comparison between various models based only on 1ubq, 1crn, and 1tit.

Fig. 4 shows correlation between $\tilde{T}_{min}$ and $\tilde{T}_f$ separately for the three proteins. For the models that employ V^C, $\tilde{T}_f$ is found to be generally comparable to the $\tilde{T}_{min}$ (with the biggest deviations occurring for potentials 10-12, 6-10-12, and the energy scale E^{ν}). However, when V^A is used, the relationship between $\tilde{T}_{min}$ and $\tilde{T}_f$ is less clearly defined. Furthermore, in a few cases the points are above the diagonal in Fig. 4, which divides models into those for which $\tilde{T}_f$ is either bigger or smaller than $\tilde{T}_{min}$. In this second group, even though $\tilde{T}_f$ is found to be smaller than $\tilde{T}_{min}$, the folding times at $\tilde{T}_f$ are comparable to those at $\tilde{T}_{min}$. This indicates that these models are unfrustrated and lead to folding. This statement also applies to models with the nonuniform energy scales which usually perform poorly in the stretching simulations. Only in a few cases, like for V^{6-12}_{const} with E^{MJ} (when V^C or V^A is used) and V^{10-12} with E^{MJ} or $E^{\nu 1,\nu 2}$, the folding times at $\tilde{T}_f$ are too long to be determined in our simulations, especially for 1tit and 1ubq (such cases correspond to $\tilde{T}_{min} = 0$ in Fig. 4). These models are also endowed with small values of $\tilde{T}_f$.

For several models with the V^A stiffness, $\tilde{T}_f$ and $\tilde{T}_{min}$ are closer to T_{max}. The best situation in which $\tilde{T}_f$ is very close to $\tilde{T}_{max}$ is found in models based on $V^{6-10-12}$ and with the energy scales E^{o}, E^{MJ}, and $E^{\nu 2}$. This closeness is combined with reasonable kinetics: $\tilde{T}_f$ and T_{max} are located close to the upper border of the optimal kinetic behavior.

As mentioned before, only small and simple proteins, such as 1crn, lead to folding in all variants of the models and V^A generates broader region of optimality than V^C and has a higher T_{min}. In most cases, folding at $\tilde{T}_{min}$ is comparable to that at $\tilde{T}_f$. However, the V^{10-12} and V^C potentials with the E^{MJ} or $E^{\nu 1,\nu 2}$ and all variants of the $V^{6-10-12}$ potential come with long values of t_{fold}. Generally we found that the narrowest U-curves correspond to the 6-10-12 potential. The 10-12 potential yields a bit broader range of good folding conditions. The range gets still broader for 6-12 and especially $6\text{-}12_{exp}$. It becomes very broad for $6\text{-}12_{const}$.

To find out which models are reasonable in studies of folding, we focused on two proteins, 1ubq and 1tit, which are harder to fold. Only some of the models lead to folding in these two proteins. Among them are all uniform-energy-parameter models with the M3 contact map. The $\{6\text{-}12_{exp}$,C,M3,$E^{o}\}$

TABLE 7 The same as Table 6, but for other potentials

Model	$\tilde{T}_f$	1crn		1ubq		1tit	
		$\tilde{T}_{min}$	t_{fold}/τ	$\tilde{T}_{min}$	t_{fold}/τ	$\tilde{T}_{min}$	t_{fold}/τ
6-12$_{exp}$							
$\{C, \mathrm{M3}, E^{o}\}$	0.24	0.3	285	0.28	702	0.28	2171
$\{A, \mathrm{M3}, E^{o}\}$	0.6	0.65	242	0.6	482	0.58	997
$\{C, \mathrm{M3}, E^{\nu 1}\}$	0.18	0.3	239	0.26	532	—	—
$\{C, \mathrm{M3}, E^{\nu 2}\}$	0.18	0.3	332	0.28	715	—	—
$\{A, \mathrm{M4}, E^{\nu 2}\}$	0.62	0.60	205	0.60	450	0.59	1201
$\{C, \mathrm{M3}, E^{\mathrm{HB,MJ}}\}$	0.23	0.3	261	0.30	670	0.32	9980
$\{A, \mathrm{M4}, E^{\mathrm{HB,MJ}}\}$	0.67	0.78	244	0.76	411	0.75	1163
10-12							
$\{C, \mathrm{M3}, E^{o}\}$	0.19	0.28	440	0.26	638	0.35	6300
$\{A, \mathrm{M3}, E^{o}\}$	0.41	0.65	339	0.6	529	0.56	1500
$\{C, \mathrm{M3}, E^{\nu 1}\}$	0.13	0.3	972	0.25	1100	0.29	5130
$\{C, \mathrm{M3}, E^{\nu 2}\}$	0.12	0.3	928	0.25	1083	0.28	5280
$\{A, \mathrm{M4}, E^{\nu 2}\}$	0.42	0.6	285	0.58	537	0.59	2320
$\{C, \mathrm{M3}, E^{\mathrm{HB,MJ}}\}$	0.14	0.19	2469	—	—	—	—
$\{A, \mathrm{M4}, E^{\mathrm{HB,MJ}}\}$	0.46	0.74	423	0.70	1914	0.71	2909
Morse							
$\{C, \mathrm{M3}, E^{o}\}$	0.3	0.32	358	0.32	677	0.31	3200
$\{A, \mathrm{M3}, E^{o}\}$	0.72	0.65	344	0.64	614	0.60	1688
Mixed							
$\{(10\text{-}12, 6\text{-}12), C, \mathrm{M3}, E^{o}\}$	0.19	0.27	253	0.26	485	—	—
$\{(10\text{-}12, 6\text{-}12), A, \mathrm{M3}, E^{o}\}$	0.47	0.61	217	0.62	428	0.60	1213
$\{(10\text{-}12, 6\text{-}12_{const}), C, \mathrm{M3}, E^{o}\}$	0.23	0.30	223	0.30	451	0.33	2940
$\{(10\text{-}12, 6\text{-}12_{const}), A, \mathrm{M3}, E^{o}\}$	0.58	0.61	209	0.60	397	0.60	1107
$\mathrm{C}^{\alpha-\beta}$							
$\{6\text{-}12, C, \mathrm{M2}, E^{o}\}$	0.29	0.39	543	0.42	1280	0.40	3289
$\{6\text{-}12, C, \mathrm{M3}, E^{o}\}$	0.25	0.39	490	0.40	1010	0.40	2520
$\{6\text{-}12, C, \mathrm{M3}, E^{\nu 2}\}$	0.31	0.38	591	0.40	1183	0.42	3629
$\{10\text{-}12, C, \mathrm{M3}, E^{o}\}$	0.22	0.36	641	0.38	1132	—	—

model has been found to be an exceptionally good folder: 90% of trajectories fold (with the small RMSD) and t_{fold} is short.

The models with the energy scale $E^{L1,L2}$ are interesting in the context of folding. Their native energies nearly coincide with those of the uniform energy model, but the amplitude of interaction gets screened depending on the distance between the pair of native amino acids. We have found that this feature smoothes the folding landscape out especially for E^{L2} with $V^{6\text{-}12}$.

When we discuss models with nonuniform energy scales we find a clear correlation between the shape of the temperature dependence of the folding time and the choice of the energy scale. We find that at least for potential 6-12 the narrowest U-curves correspond to the energy scale $E^{\nu 1,\mathrm{nu}2}$ and then they broaden increasingly in the order E^{L1}, E^{L2}, (6-12, 10-12), E^{o}, E^{o}, and $E^{\mathrm{HB,MJ}}$.

However, generally across all choices of the energy scale, we can say that all variants of E^{G3}, E^{L2}, and $E^{\mathrm{HB,MJ}}$ with V^{A} always produce folding to the native state and the temperature dependence of t_{fold} is U-shaped. The fastest folding arises when one combines $E^{\nu 2}$ with $V^{6\text{-}12}$ and V^{A}, or $V^{6\text{-}12}_{exp}$ and $E^{\mathrm{HB,MJ}}$, also with $V^{6\text{-}12}_{exp}$ and V^{A}, as shown in the Tables 6 and 7. $V^{6\text{-}12}_{const}$ also leads to good folding independent of the choice of E_{ij}. The folding is the most difficult for $E^{\mathrm{HB,MJ}}$ and $E^{\nu 1}$ with V^{C} almost for all choices of the potential.

Among the various choices of the contact potential, the 6-10-12 and 10-12 models have the highest $\tilde{T}_{min}$ and the narrowest U-curves independent of the nature of the local backbone stiffness as seen in the right hand panel of Fig. 3.

CONCLUSION

Our simulations show that the simplest Gō-like potential with uniform couplings, the chirality term and the M3 contact map is well suited to study mechanical unfolding and it also leads to reasonable folding kinetics and equilibrium properties. The values of T_{min} are less scattered among proteins when V^{C} is used instead of V^{A} to account for the backbone stiffness. This feature allows us to use single temperature in comparatory studies. We used this version of the Gō model to build a server designated for stretching studies of proteins. Its address is www.ifpan.edu.pl/BSDB/. Currently, it contains data on >7500 proteins. However, the V^{A} stiffness usually leads to a broader temperature range of folding, especially for the narrow potentials 10-12 or 6-10-12. It also improves the appearance of the F-d stretching curves. Even though our results favor the simplest model for studies of stretching, there are several other models, such as discussed in the literature (15,49) that should perform in a very comparable way. One considers thermodynamics and folding kinetics as providing selection criteria for a model then three models stand out: {6-12, A, M4, $E^{\mathrm{HB,MJ}}$}, {6-10-12, A, M4, E^{o}}, and {6-10-12, A, M4, $E^{\mathrm{HB,MJ}}$}. They come with T_f substantially larger than T_{min} and offer reasonably fast folding time, and when one considers thermodynamics and folding kinetics as providing selection criteria for a model, three models stand out: {6-12, A, M4, $E^{\mathrm{HB,MJ}}$}, {6-10-12, A, M4, E^{o}}, and {6-10-12, A, M4, $E^{\mathrm{HB,MJ}}$}. They come with T_f substantially larger than T_{min} and offer reasonably fast folding time, at least for the three proteins studied.

We appreciate useful discussion with J. R. Banavar, S. Niewieczerzal, P. Szymczak, and M. Wojciechowski.

This work was funded by Ministerstwo Nauki i Szkolnictwa Wyższego grants No. N202-021-31/0739 and N-N202-0852-33.

REFERENCES

1. Trylska, J., V. Tozzini, and J. A. McCammon. 2004. A coarse-grained model of the ribosome: molecular dynamics simulations. *Protein Sci.* 217(Suppl. 13):121.
2. Trylska, J., J. A. McCammon, and C. L. Brooks. 2005. Exploring assembly energetics of the 30S ribosomal subunit using an implicit solvent approach. *J. Am. Chem. Soc.* 127:11125–11133.
3. Trylska, J., V. Tozzini, and J. A. McCammon. 2005. Exploring global motions and correlations in the ribosome. *Biophys. J.* 89: 1455–1463.
4. Sułkowska, J. I., and M. Cieplak. 2008. Stretching to understand proteins—a survey of the Protein Data Bank. *Biophys. J.* 94:6–13.
5. Sułkowska, J. I., and M. Cieplak. 2007. Mechanical stretching of proteins—a theoretical survey of the Protein Data Bank. *J. Phys. Cond. Mat.* 19:283201.

6. Munoz, V., and W. A. Eaton. 1999. A simple model for calculating the kinetics of protein folding from free-dimensional structures. *Proc. Natl. Acad. Sci. USA*. 96:11311–11316.

7. Veitshans, T., D. Klimov, and D. Thirumalai. 1997. Protein folding kinetics: time scales, pathways and energy landscapes in terms of sequence-dependent properties. *Fold. Des.* 2:1–22.

8. Zhou, Y., and M. Karplus. 1999. Interpreting the folding kinetics of helical proteins. *Nature*. 401:400–403.

9. Dokholyan, N. V., S. V. Buldyrev, H. E. Stanley, and E. I. Shakhnovich. 1998. Discrete molecular dynamics studies of the folding of a protein-like model. *Fold. Des.* 3:577–587.

10. Hardin, C., Z. Luthey-Schulten, and P. G. Wolynes. 1999. Backbone dynamics, fast folding, and secondary structure formation in helical proteins and peptides. *Proteins Struct. Funct. Genet.* 34:281–294.

11. Chang, C.-E., T. Shen, J. Trylska, V. Tozzini, and J. A. McCammon. 2006. Gated binding of ligands to HIV-1 protease: Brownian dynamics simulations in a coarse-grained model. *Biophys. J.* 90:3880–3885.

12. Paci, E., M. Vendruscolo, and M. Karplus. 2002. Validity of Gō models: comparison with a solvent-shielded empirical energy decomposition. *Biophys. J.* 83:3032–3038.

13. Hoang, T. X., and M. Cieplak. 2000. Molecular dynamics of folding of secondary structures in Gō-like models of proteins. *J. Chem. Phys.* 112:6851–6862.

14. Cieplak, M., and T. X. Hoang. 2003. Universality classes in folding times of proteins. *Biophys. J.* 84:475–488.

15. Clementi, C., H. Nymeyer, and J. N. Onuchic. 2000. Topological and energetic factors: what determines the structural details of the transition state ensemble and ''on-route'' intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* 298:937–953.

16. Cieplak, M., T. X. Hoang, and M. O. Robbins. 2004. Thermal effects in stretching of Gō-like models of titin and secondary structures. *Proteins Struct. Funct. Biol.* 56:285–297.

17. Cieplak, M., and P. E. Marszalek. 2005. Mechanical unfolding of ubiquitin molecules. *J. Chem. Phys.* 123:194903.

18. Abe, H., and N. Gō. 1981. Noninteracting local-structure model of folding and unfolding transition in globular proteins. II. Application to two-dimensional lattice proteins. *Biopolymers*. 20:1013–1031.

19. Takada, S. 1999. Gō-ing for the prediction of protein folding mechanism. *Proc. Natl. Acad. Sci. USA*. 96:11698–11700.

20. Rief, M., M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub. 1997. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science*. 276:1109–1112.

21. Carrion-Vasquez, M., A. F. Oberhauser, S. B. Fowler, P. E. Marszalek, S. E. Broedel, J. Clarke, and J. M. Fernandez. 1999. Mechanical and chemical unfolding of a single protein: a comparison. *Proc. Natl. Acad. Sci. USA*. 96:3694–3699.

22. Watanabe, K., C. Muhle-Goll, M. S. Z. Kellermayer, S. Labeit, and H. L. Granzier. 2002. Different molecular mechanics displayed by titin's constitutively and differentially expressed tandem Ig segments. *Struct. Biol. J.* 137:248–258.

23. Watanabe, K., P. Nair, D. Labeit, M. S. Z. Kellermayer, M. Greaser, S. Labeit, and H. L. Granzier. 2002. Molecular mechanics of cardiac titins PEVK and N2B spring elements. *J. Biol. Chem.* 277:11549–11558.

24. Li, H. B., and J. M. Fernandez. 2003. Mechanical design of the first proximal Ig domain of human cardiac titin revealed by single molecule force spectroscopy. *J. Mol. Biol.* 334:75–86.

25. Yang, G., C. Cecconi, W. A. Baase, I. R. Vetter, W. A. Breyer, J. A. Haack, B. W. Matthews, F. W. Dahlquist, and C. Bustamante. 2000. Solid-state synthesis and mechanical unfolding of polymers of T4 lysozyme. *Proc. Natl. Acad. Sci. USA*. 97:139–144.

26. Lenne, P. F., A. J. Raae, S. M. Altmann, M. Saraste, and J. K. H. Horber. 2000. States and transition during unfolding of a single spectrin repeat. *FEBS Lett.* 476:124–128.

27. Brockwell, D. J., E. Paci, R. C. Zinober, G. Beddard, P. D. Olmsted, D. A. Smith, R. N. Perham, and S. E. Radford. 2003. Pulling geometry defines mechanical resistance of β-sheet protein. *Nat. Struct. Biol.* 10:731–737.

28. Carrion-Vazquez, M., A. F. Oberhauser, T. E. Fisher, P. E. Marszalek, H. Li, and J. M. Fernandez. 2000. Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering. *Prog. Biophys. Mol. Biol.* 74:63–91.

29. Lee, G., K. Abdi, Y. Jiang, P. Michaely, V. Bennett, and P. E. Marszalek. 2006. Nanospring behavior of ankyrin repeats. *Nature*. 440:246–249.

30. Li, L. W., S. Wetzel, A. Pluckthun, and J. M. Fernandez. 2006. Stepwise unfolding of ankyrin repeats in a single protein revealed by atomic force microscopy. *Biophys. J.* 90:30–32.

31. Best, R. B., B. Li, A. Steward, V. Daggett, and J. Clarke. 2001. Can non-mechanical proteins withstand force? Stretching barnase by atomic force microscopy and molecular dynamics simulation. *Biophys. J.* 81:2344–2356.

32. Brockwell, D. J., S. Godfrey, S. Beddard, E. Paci, D. K. West, P. D. Olmsted, D. A. Smith, and S. E. Radford. 2005. Mechanically unfolding small topologically simple protein L. *Biophys. J.* 89:506–519.

33. Schwaiger, I., A. Kardinal, M. Schleicher, A. A. Noegel, and M. Rief. 2004. A mechanical unfolding intermediate in an actin-crosslinking protein. *Nat. Struct. Mol. Biol.* 11:81–85.

34. Schlierf, M., and M. Rief. 2005. Temperature softening of a protein in single-molecule experiments. *J. Mol. Biol.* 345:497–503.

35. Cecconi, C., E. A. Shank, C. Bustamante, and S. Marqusee. 2005. Direct observation of the three-state folding of a single protein molecule. *Science*. 309:2057–2060.

36. Chyan, C. L., F. C. Lin, H. Peng, J. M. Yuan, C. H. Chang, S. H. Lin, and G. Yang. 2003. Reversible mechanical unfolding of single ubiquitin molecules. *Biophys. J.* 87:3995–4006.

37. Carrion-Vazquez, M., H. Li, H. Lu, P. E. Marszalek, A. F. Oberhauser, and J. M. Fernandez. 2003. The mechanical stability of ubiquitin is linkage dependent. *Nat. Struct. Biol.* 10:738–743.

38. Dietz, H., and M. Rief. 2004. Exploring the energy landscape of the GFP by single-molecule mechanical experiments. *Proc. Natl. Acad. Sci. USA*. 101:16192–16197.

39. Dietz, H., and M. Rief. 2006. Protein structure by mechanical triangulation. *Proc. Natl. Acad. Sci. USA*. 103:1244–1247.

40. Li, L., H. H.-L. Huang, C. L. Badilla, and J. M. Fernandez. 2005. Mechanical unfolding intermediates observed by single-molecule force spectroscopy in fibronectin type III module. *J. Mol. Biol.* 345:817–826.

41. Oberhauser, A. F., C. Badilla-Fernandez, M. Carrion-Vazquez, and J. M. Fernandez. 2002. The mechanical hierarchies of fibronectin observed with single-molecule *AFM. J. Mol. Biol.* 319:433–447.

42. Oberdorfer, Y., H. Fuchs, and A. Janshoff. 2000. Conformational analysis of native fibronectin by means of force spectroscopy. *Langmuir*. 16:9955–9958.

43. Kwiecinska, J. I., and M. Cieplak. 2005. Chirality and protein folding. *J. Phys. Cond. Mat.* 17:S1565–S1580.

44. Cieplak, M., T. X. Hoang, and M. O. Robbins. 2004. Stretching of proteins in the entropic limit. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 69:011912.

45. Szymczak, P., and M. Cieplak. 2007. Influence of hydrodynamic interactions on mechanical unfolding of proteins. *J. Phys. Cond. Mat.* 19:258224.

46. Szymczak, P., and M. Cieplak. 2007. Proteins in a shear flow. *J. Chem. Phys.* 127:155106.

47. Cao, Y., and H. Li. 2007. Polyprotein of GB1 is an ideal artificial elastomeric protein. *Nat. Mater.* 6:109–114.

48. Wojciechowski, M., and M. Cieplak. 2007. Coarse-grained modeling of pressure-related effects in staphylococcal nuclease and ubiquitin. *J. Phys. Cond. Mat.* 19:285218.

49. Karanicolas, J., and C. L. Brooks III. 2002. The origins of the asymmetry in the folding transition states of protein L and G. *Protein Sci.* 11:2351–2361.

50. West, D. K., P. Olmsted, and E. Paci. 2006. Mechanical unfolding revisited through a simple but realistic model. *J. Chem. Phys.* 124:154909–1.

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

52. Scheinerman, F. B., and C. L. Brooks III. 2001. Molecular picture of folding of a small α/β protein. *Proc. Natl. Acad. Sci. USA.* 95:1562–1567.

53. Cheung, M., A. E. Gracia, and J. Onuchic. 2002. Protein folding mediate by the solvation: water expulsion and formation of the hydrophobic core occur after structural collapse. *Proc. Natl. Acad. Sci. USA.* 99:685–690.

54. Onuchic, J. N., H. Nymeyer, A. E. Garcia, J. Chahine, and N. D. Socci. 2000. The energy landscape theory of protein folding: insights into folding mechanisms and scenarios. *Adv. Protein Chem.* 53:87–152.

55. Settanni, G., T. X. Hoang, C. Micheletti, and A. Maritan. 2004. Folding pathways of prion and doppel. *Biophys. J.* 83:3533–3541.

56. Clementi, C., M. Vandruscolo, A. Maritan, and E. Domany. 1999. Folding Lennard-Jones proteins by a contact potential. *Proteins Struct. Funct. Gen.* 37:544–553.

57. Tsai, J., R. Taylor, C. Chothia, and M. Gerstein. 1999. The packing density in proteins: standard radii and volumes. *J. Mol. Biol.* 290: 253–266.

58. Cieplak, M., A. Pastore, and T. X. Hoang. 2005. Mechanical properties of the domains of titin in a Gō-like model. *J. Chem. Phys.* 122:054906.

59. Sobolev, V., A. Sorokine, J. Prilusky, E. E. Abola, and M. Edelman. 1999. Automated analysis of interatomic contacts in proteins. *Bioinformatics.* 15:327–332.

60. Srinivasan, R., and G. D. Rose. 1995. LINUS: a hierarchic procedure to predict the fold of a protein. *Proteins Struct. Funct. Genet.* 22:81–99.

61. Cecconi, F., C. Guardiani, and R. Livi. 2008. Stability and kinetic properties of C5-domain from myosin binding protein C and its mutants. *Biophys. J.* 94:1403–1411.

62. Kabsch, W., and C. Sander. 1983. Dictionary of the protein secondary structure: pattern recognition of the hydrogen-bonded and geometrical features. *Biopolymers.* 22:2577–2637.

63. Kolinski, A., and J. Skolnik. 1994. Monte Carlo simulations of protein folding. I. Lattice model and interaction scheme. *Proteins.* 18:338–352.

64. Miyazawa, S., and R. L. Jernigan. 1996. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.* 256: 623–644.

65. Micheletti, C., J. R. Banavar, A. Maritan, and F. Seno. 1999. Protein structures and optimal folding from a geometrical variational principle. *Phys. Rev. Lett.* 82:3372–3375.

66. Marszalek, P. E., H. Lu, H. Li, M. Carrion-Vazquez, A. F. Oberhauser, K. Schulten, and J. M. Fernandez. 1999. Mechanical unfolding intermediates in titin modules. *Nature.* 402:100–103.

67. Oberhauser, A. F., P. E. Marszalek, H. P. Erickson, and J. M. Fernandez. 1998. The molecular elasticity of the extracellular matrix protein tenascin. *Nature.* 14:181–185.