Model study of protein unfolding by interfaces

S. D. Chakarova
Department of Applied Physics, Chalmers University of Technology and Göteborg University, SE-412 96, Sweden

A. E. Carlsson
Department of Physics, Washington University, St. Louis, Missouri 63130-4899, USA

(Received 29 July 2003; revised manuscript received 4 November 2003; published 26 February 2004)

We study interface-induced protein unfolding on hydrophobic and polar interfaces by means of a two-dimensional lattice model and an exhaustive enumeration ground-state structure search, for a set of model proteins of length 20 residues. We compare the effects of the two types of interfaces, and search for criteria that influence the retention of a protein's native-state structure upon adsorption. We find that the unfolding proceeds by a large, sudden loss of native contacts. The unfolding at polar interfaces exhibits similar behavior to that at hydrophobic interfaces but with a much weaker interface coupling strength. Further, we find that the resistance of proteins to unfolding in our model is positively correlated with the magnitude of the folding energy in the native-state structure, the thermal stability (or energy gap) for that structure, and the interface energy for native-state adsorption. We find these factors to be of roughly equal importance.

DOI: 10.1103/PhysRevE.69.021907 PACS number(s): 87.14.Ee, 87.15.Aa, 87.15.He, 68.47.Pe

I. INTRODUCTION

Protein adsorption at interfaces is a very common phenomenon which can often be important in technological applications. The adsorption can alter the protein structure and therefore influence or destroy the function of the protein. Such interface-induced unfolding may present a problem, for instance if it is important that the adsorbed protein retains its function, as in some biosensors based on adsorbed protein layers. Interface-induced unfolding also presents problems when one wants to study a protein's native-state properties by an experimental method that requires an interface contact. These issues motivate both experimental and theoretical studies of the influence of interfaces on protein structure, in particular the development of criteria which could determine whether or not a particular protein will denature on a specific interface.

Both the experimental and theoretical understanding of these phenomena are limited at present. Experimental results show that some interfaces affect protein structure substantially. Conformational changes have been monitored by several probes which show that the extent of a protein's ordered secondary structure can be diminished by adsorption. This holds for both hydrophobic and hydrophilic interfaces [1]. For example, the circular dichroism spectrum of adsorbed immunoglobulin G (IgG) at the hydrophobic Teflon-water interface [2,3] indicates that the relative fractions of α-helix and random coil content are increased, whereas the β-sheet content is strongly diminished. In an infrared spectroscopy study of Ribonuclease A adsorption at the hydrophilic germanium-water interface, a decrease of the content of β-sheet was observed together with an increase in the signal for turns and unordered structure [4]. Both the α-helix and β-sheet content were diminished for bovine serum albumin at the hydrophilic silica-water interface studied by circular dichroism [5]. Other experimental studies using a variety of techniques have shown that the first adsorbed protein layer is often denatured [6–8].

The effect of adsorption is especially strong for hydrophobic interfaces. For example, IgG is not denatured when adsorbed on a hydrophilic quartz interface, but it does denature on a hydrophobic teflon-coated interface [2,3]. Since it is generally believed that the hydrophobic forces are dominant in protein folding [9], it is indeed expected that a hydrophobic interface should affect a protein more strongly than a hydrophilic interface. Proteins in solution tend to have a hydrophobic core and a polar shell. A hydrophobic interface would therefore extract the amino acids from the core of the folded protein, while a polar interface might bind to the polar surface parts of the protein; the latter would be more likely to leave the protein intact. In addition, the interactions between a polar interface and the polar protein shell are weaker due to the competitive presence of (hydrophilic) water molecules. One thus expects a polar interface to be more like the normal water-rich environment for a protein than a hydrophobic interface, which would render the polar interface less disruptive. Nevertheless, experiments [4,5] have shown that polar interfaces can have a large effect on protein structure.

No theoretical calculations have been performed to our knowledge which directly treat interface-induced unfolding of proteins. A study of homopolymers at interfaces has been performed using a three-dimensional lattice model [10]. This work identified two types of adsorption behavior—''docking'' and ''flattening.'' It treated the effect of the interface on the amount of secondary structure, but could not treat denaturation in the sense of the loss of native structure since homopolymers do not have a well-defined native structure. Other relevant existing studies are those of protein folding in confined geometries and protein unfolding by tension, which are somewhat analogous to the present problem because they treat the effect of an external perturbation on protein structure. A simple exact model has been used by Chan and Dill [11] to show that a class of chaperonins having hydrophobic interiors, modeled by hydrophobic interfaces confining the protein, can stabilize some of its important transitional structures on the pathway to the native fold. This is achieved by

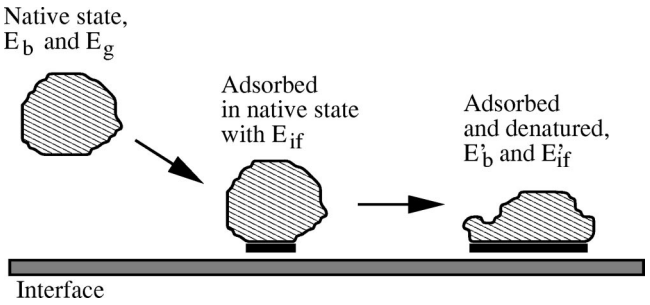

FIG. 1. A protein, characterized by its native-state bulk energy E_b and the energy gap to the next-lowest-energy structure E_g, can either retain its native-state structure or denature when it interacts with the interface. E_{if}: interaction energy of native-state protein with interface. Primed quantities are evaluated for a denatured structure.

the protein being partly unfolded while interacting with the chaperonin, which makes it less prone to misfolding. Unfolding of proteins by an external force field has been treated by Shen and co-workers [12]. They used a residue-level model with springlike interactions that favor the native-structure nearest-neighbor contacts, distances, and orientations. They found that the external force causes a sharp and cooperative unfolding transition. It is of interest to establish whether unfolding by interfaces exhibits similar behavior.

In this paper we present simple model calculations which aim to clarify the generic effects of interface adsorption on protein structure (schematic picture shown in Fig. 1), and the main factors preventing denaturation. We use a simplified model because this type of problem cannot be fully treated by any existing methods with quantitative accuracy. There is continuing progress in the use of force fields [13,14] to describe protein-protein interactions, but these force fields do not have parameters for treating interfaces. In addition, such methods cannot reliably evaluate the lowest-free-energy structures because the calculations cannot be run long enough.

The outline of the remainder of the paper is as follows. Section II introduces and motivates our choice of model and method of solution. Section III presents the numerical results and discusses their potential relation to experiments. The last section summarizes our findings.

II. MODEL AND METHOD OF SOLUTION

A. Model

The specific structures formed by proteins are determined by a delicate competition between several types of forces including hydrogen bonds, van der Waals interactions, and ion pairing. Recently it has been demonstrated that also the effect of the polarity of the protein backbone is of some importance [15] and can be added to this list. It is generally believed, however, that the dominant folding force is the hydrophobic interaction [9]. A commonly accepted picture of protein folding is that the hydrophobic force selects a limited range of conformations from which the other forces choose the native-state structure. A model that captures this dominant force of folding is the HP model proposed by Lau and Dill [16] and developed in protein folding research. In the model the 20 amino acids are divided into two types: hydrophobic (H) and polar (P). The dominant interaction is the one between two amino acids of type H, and the remaining interactions are often ignored. This model is most applicable to globular proteins, since they have a compact native-state structure, a core of hydrophobic residues, and significant secondary structure. It has been widely explored and has provided basic insights into the protein folding process.

Our model, a HP_1P_2 model, is a modification of the HP model designed to provide a somewhat improved description of the hydrophobicity of the amino acids, and to reduce the ground-state structure degeneracy that is inherent to the HP model. H stands for hydrophobic, P_1 for slightly polar, and P_2 for polar amino acids. This choice was motivated by the clean division of the amino acids into three groups with respect to hydrophobicity, shown in Table I. We consider a two-dimensional lattice model with nearest-neighbor interactions between amino acids that are not nearest neighbors in the chain. The Hamiltonian for an adsorbed protein is written as a sum of bulk and interface contributions:

$$\mathcal{H}=\mathcal{H}_b+\mathcal{H}_{if}, \tag{1}$$

where

TABLE I. Relative abundance in *E. coli* proteins [29], hydropathy index [30], and the assigned hydrophobicity type for the HP_1P_2 model for all amino acids. Asterisks denote that the relative abundance in the reference was given jointly for Glu and Gln as well as for Asp and Asn. We assume equal distribution, i.e., 50%-50% in each pair. The total relative abundances for the three types H, P_1, and P_2 are 45%, 22%, and 34%.

Amino acid	Ala	Glu	Gln	Asp	Asn	Leu	Gly	Lys	Ser	Val
Relative abundance [29]	13.0	5.4*	5.4*	4.95*	4.95*	7.8	7.8	7.0	6.0	6.0
Hydropathy [30]	1.8	−3.5	−3.5	−3.5	−3.5	3.8	−0.4	−3.9	−0.8	4.2
Assigned type	H	P_2	P_2	P_2	P_2	H	P_1	P_2	P_1	H
Amino acid	Arg	Thr	Pro	Ile	Met	Phe	Tyr	Cys	Trp	His
Relative abundance [29]	5.3	4.6	4.6	4.4	3.8	3.3	2.2	1.8	1.0	0.7
Hydropathy [30]	−4.5	−0.7	1.6	4.5	1.9	2.8	−1.3	2.5	−0.9	−3.2
Assigned type	P_2	P_1	H	H	H	H	P_1	H	P_1	P_2

TABLE II. Interaction parameters: The averaged values obtained from the Miyazawa-Jernigan (MJ) interaction matrix [20], E^{MJ} in RT units, were modified by the formula $(E^{\mathrm{MJ}}/E^{\mathrm{MJ}}_{HH}-0.3)/1.3$ in order to obtain normalized and negative interactions with units of E_{HH}. We quote our values only to four significant figures although more were used in the calculations.

	E_{HH}	E_{HP_1}	E_{HP_2}	$E_{P_1P_1}$	$E_{P_1P_2}$	$E_{P_2P_2}$
E^{MJ}	−0.4969	−0.2368	0.1136	−0.1647	0.0394	0.1389
Normalized	−1	−0.5972	−0.0550	−0.4857	−0.1698	−0.0157

$$\mathcal{H}_{\mathrm{b}}=\sum_{i,j} n_{i,j}E_{w_iw_j} \tag{2}$$

and

$$\mathcal{H}_{\mathrm{if}}=\sum_{i} n_iE_{Iw_i}. \tag{3}$$

Here i,j denote site indices, w_i is the residue type of site i (H, P_1, or P_2), and $n_{i,j}$ equals 1 if i and j are nearest neighbors (not along the chain), and zero otherwise. The $E_{w_iw_j}$ are effective interactions that include both direct interactions between residues and indirect interactions mediated by water molecules in the solution; for an attractive effective interaction $E_{w_iw_j}<0$. Similarly, I denotes the residue type corresponding to the interface, to be discussed shortly; n_i is 1 if site i contacts the interface and zero otherwise, and E_{Iw_i} is the effective interaction between an interface and residues of type w_i.

It should be noted that recent studies [17–19] have shown that models which use additive contact energies similar to ours have limited capabilities of reproducing proteinlike thermodynamics in a quantitative fashion. These studies pointed out that the latter requires, among other conditions, that the van't Hoff to calorimetric enthalpy ratio $\Delta H_{vH}/\Delta H_{cal}$ is near to unity, while for a pair-interaction model such as ours it is much less than unity.

Our model is very simple, but even in two dimensions it is difficult to treat exactly chains longer than 20 amino acids with explicit enumeration. We argue below that some general features of our results would hold in more realistic models as well.

B. Choice of bulk parameters

The main criterion behind our choice of parameters is to obtain a spread of interaction strengths roughly comparable to that of real proteins. To accomplish this, we begin with the interaction parameters developed by Miyazawa and Jernigan [20] (MJ), which are based on observed pair frequencies in native-state structures of proteins. We obtain the interactions $E_{w_iw_j}$ in our three-letter alphabet model from their parameters in three steps. First, we divide the 20 residues into the three hydrophobicity classes given in Table I. The distribution for the amino acid types used was obtained from relative abundances in *E. coli* proteins: 45%, 22%, and 34% for the respective amino acid types. Rounding off to the nearest integer for a chain of length 20 amino acids gives nine residues of type H, four of type P_1, and seven of type P_2. Second, we average the MJ parameters over the residues in each class, giving the first row in Table II. Finally, we normalize and shift them uniformly in such a fashion that the energy unit becomes the HH interaction. This gives the results shown in the second row of Table II. The reason for the shift is that the observed pair frequencies used to evaluate the MJ parameters do not determine the overall zero, but only the differences between the parameters. We choose our shift so that the structures that are obtained are fairly compact, although not absolutely compact, which is consistent with observed protein structures. A similar strategy of using a range of negative interaction energies has been used previously [21,22]. The final values were not rounded off to the number of significant figures in the original MJ parameters, in order to avoid parameter-induced degeneracies. Our procedure resulted in roughly 60% of the chains having unique ground-state structures. The results presented in this paper are for the nondegenerate sequences only.

Other kinds of averages from the MJ values are possible. However, we will show that reasonable variations in the parameters do not introduce any major changes in the results.

C. Interface coupling strength

A hydrophobic interface, which is our main interest, can in a first approximation be thought of as consisting of hydrophobic residues similar to the ones included in the protein chain. The interactions of the amino acids in the chain with the interface, E_{Iw_i}, are thus set proportional to E_{Hw_i} by a factor depending on the density of hydrophobic residues on the interface:

$$E_{Iw_i}=\lambda_hE_{Hw_i}, \tag{4}$$

where $w_i=\{H,P_1,P_2\}$ and the dimensionless parameter λ_h determines the interface coupling strength. This parallels the experimental technique used by Kandori *et al.* [23] who studied protein adsorption on an interface with varying hydrophobicity. They started with an initially hydrophilic interface, which became more and more hydrophobic as it was gradually covered with oleyl phosphate (OP) molecules that had their hydrophobic ends pointing into the protein solution [24].

Similarly, one can think of a polar interface as consisting of polar molecules similar to the ones in the protein chain, distributed over the interface:

$$E_{Iw_i}=\lambda_pE_{P_2w_i}, \tag{5}$$

where λ_p is the degree of polarity. In constructing the polar interface we use the parameters for P_2 since it is the most polar of the residue types.

D. Method of solution

Our method is to find the lowest-energy structure using an exhaustive enumeration of all self-avoiding configurations in two dimensions for chains of length 20 amino acids. This procedure is rigorous, but limits us to two dimensions (2D) because of the vast increase in CPU time required to treat longer chains and more possible jump directions. Similar models in three dimensions have to use shorter chains, or abandon the exhaustive enumeration search and rely on more approximate methods. Monte Carlo searches for the ground-state structure have been performed in three dimensions for chains up to 36 amino acids. However, the available structure space has been limited to absolutely compact structures only, see for instance Ref. [25]. Such a limitation is naturally not acceptable when treating interface-induced unfolding.

Our enumeration task is equivalent to performing a 20-step self-avoiding walk on a 2D-lattice, which has three possible directions at each step. By limiting the initial conditions (first step always in the same direction and the first turn always to the left) one can eliminate most of the mirror degeneracies. We calculate the ground-state structure for a range of interface coupling strengths, assuming that the system always finds this lowest-energy structure at the interface.

The ordering of the H, P_1, and P_2 residues in the 20-monomer chain was chosen in a random fashion and the results are presented for a sample set of 1163 such sequences that also fulfilled the condition of having a unique native-state structure.

III. RESULTS

The distribution of native-state bulk energies E_b for the studied sequence set is fairly Gaussian with mean $E_\mathrm{b} \approx -7.9$ (units of E_{HH}, see Table II), and standard deviation $\sigma_\mathrm{b} \approx 0.54$. The native-state structure contained on average about 6.0, 2.5, 0.5, 0.4, 1.0, and 0.4 contacts of types HH, HP_1, HP_2, P_1P_1, P_1P_2, and P_2P_2, respectively. In the presence of an interface the total energy for the adsorbed-state structure can be split into bulk and interface components. The bulk component E'_b, defined as the value of $\mathcal{H}_\mathrm{b}$, is originally equal to the native-state bulk energy E_b and decreases in magnitude as the protein changes to a new ground-state structure. The interface component $E'_\mathrm{if} = \Sigma_{w_i} N_{w_i} E_{Iw_i}$ is defined as the sum of all energies due to interface contacts, and increases in magnitude upon adsorption. It becomes proportional to the interface coupling strength λ at large λ when all amino acids have interface contact, i.e., when the numbers N_{w_i} of residues with particular types of surface contacts reach their maximum possible values: $\max(N_H)=9, \max(N_{P_1})=4, \max(N_{P_2})=7$ for the studied 20-monomers.

Figure 2 shows the ground-state structures for a sample sequence on hydrophobic interfaces with varying interface coupling strengths λ_h. The protein is allowed to rotate with respect to the interface in order to minimize the total energy. The case $\lambda_h=0$ is equivalent to the case of no interface, i.e., the protein is in its native-state structure in solution. (The protein in our calculations lies in an infinite solution and is generally away from the interface for zero interaction. Therefore, the interface causes no excluded volume effect.) With increasing λ_h the model protein gradually loses its native contacts, until it eventually becomes entirely flattened out on the interface. Thus the identified types of adsorption behavior in the homopolymer study in Ref. [10], first step being docking to the interface and the last steps flattening, are present. As shown, even at $\lambda_h=1$ the sample protein is not completely unfolded. The interface coupling strength required to bring all of the residues into contact with the interface will of course vary from protein to protein. However, an upper bound is reached when the coupling is strong enough to make the strongest bulk contact (the dominant HH contact) less favorable than two of the weakest interface contacts. For our interaction parameter set this corresponds to $\lambda_h \geqslant 9.1$ for the hydrophobic interface and $\lambda_p \geqslant 31.8$ for the polar interface.

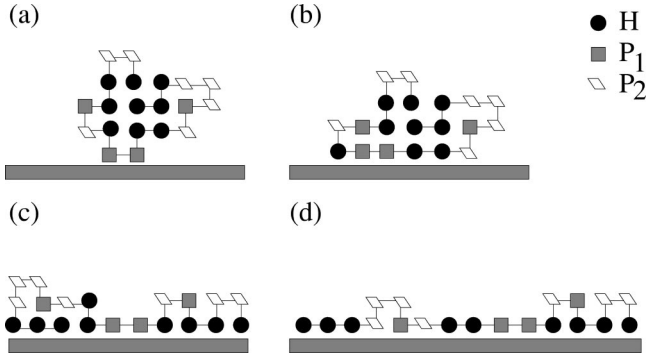

FIG. 2. The ground-state structure for the sample sequence $HP_2P_2HHP_1P_2HP_1P_1HHP_2P_1P_2P_2P_2HHH$, adsorbed at hydrophobic interfaces with different interface coupling strengths λ_h: (a) native-state structure with ten native contacts; (b) 20% of the native contacts lost at interface with $\lambda_h=0.3$; (c) 50% of the native contacts lost at $\lambda_h=0.8$; (d) nearly flat conformation with only three (native) contacts at $\lambda_h=1$.

A. Effect of interface coupling strength on quantitative measures of protein structure

We measure the interface-induced structure change for the model proteins by means of the fraction of preserved native contacts, Q, a measure used in earlier protein folding studies [26]. We also monitor the bulk energy E'_b, i.e., the energy of all native and non-native interresidue contacts. Q tells us how a given structure differs from the native one, while E'_b describes the degree of unfolding relative to a fairly compact starting point.

Figure 3 shows averaged curves for Q and E'_b (the latter normalized with respect to native-state bulk energy) for the adsorbed-state structure, as functions of the interface coupling strength λ. As expected, polar interfaces have a weaker effect on protein structure. For weak interface interactions $\langle Q \rangle$ varies linearly, while $\langle E'_\mathrm{b} \rangle$ changes much more slowly.

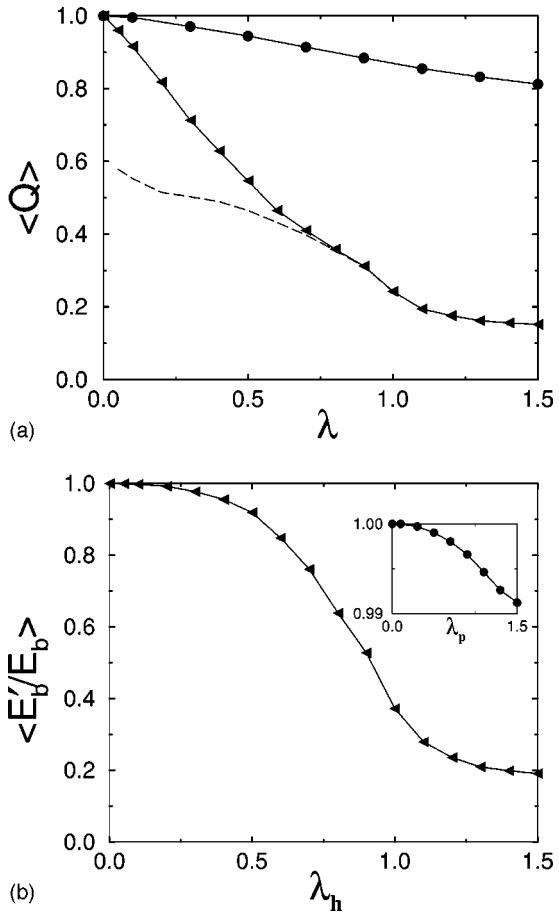

FIG. 3. Average fraction of native contacts $\langle Q \rangle$ (a), and average bulk energy, normalized with respect to the native-state bulk energy, $\langle E_b'/E_b \rangle$ (b), as functions of interface coupling strength λ for hydrophobic (triangles) and polar interfaces (circles). The initial slopes in $\langle Q \rangle$ are ≈ -0.93 and -0.12, respectively. The dashed line shows the soft set component of $\langle Q \rangle$ for the hydrophobic interface. (See Sec. III A.)

Thus a large change in Q can be accompanied by a very small change in E_b'. This suggests that the protein structure is still well described by a hydrophobic core, even though the fold is very different from the native one.

For strong interface interactions ($\lambda_h > 1$ in the case of a hydrophobic interface) the behavior of $\langle Q \rangle$ deviates from the initial linear trend. This also occurs for polar interfaces for $\lambda_p \gg 1$ (not shown). This is because the last few bulk contacts are very hard to break. As discussed above, the last native contact might continue to exist up to high values of λ_h, giving a $\langle Q \rangle$ value of about 10% since for the studied sequence set there are on average about ten native contacts. The entire unfolding process, in the case of a hydrophobic interface, can be thought of involving three major steps. The first involves extracting all the amino acids of type H from the bulk to the interface, and the second and third involve extracting types P_1 and P_2. This order is given by the ordering of the interaction strengths. For $\lambda_h = 1$ in Fig. 3, approximately eight out of nine H residues in the chain have interface contact, corresponding roughly to completion of the first step.

Our definition of E_b would correspond to some extent to the number of topological contacts in the structure. We can therefore qualitatively compare the $\langle E_b' \rangle$ vs λ curve in Fig. 3(b) to the results in Fig. 12(b) of Ref. [10], which shows the number of topological contacts for a 3D homopolymer of length $n = 14$ at a planar interface, as a function of the chain-interface contact energy. It was found that increasing contact energy results in a loss of topological contacts, with a profile similar to $\langle E_b' \rangle$ (flat region, transition, and another flat region). However while we see a drop of roughly 80% in $\langle E_b' \rangle$, the loss of topological contacts was about 30%. The transition in the number of topological contacts was also sharper, with a 90–10% drop occurring over an interval of roughly 15% around the midpoint value for the contact energy; the corresponding interval for $\langle E_b' \rangle$ is roughly 40%. Further, in the beginning of the transition, the topological contact curve is flatter.

From an experimental point of view, it is more important to understand the beginning of the structural changes in protein denaturation than complete unfolding. Experimental results [4,5] show that proteins denatured by adsorption often contain a substantial amount of α-helix and β-sheet secondary structure, i.e., they are far from completely unfolded. Figure 3(a) shows that on average denaturation begins already at the first sampled point, $\lambda_h = 0.05$. Even though some sequences do retain their native-state structure up to $\lambda_h = 0.9$, for that interface coupling strength most sequences are almost completely unfolded. If a real interface corresponds to one that can denature proteins but nevertheless leaves a certain amount of bulk contacts present, we can conclude that strong interface interactions corresponding to $\lambda_h \gg 1$ in our model are not physically relevant.

One could define a typical $\lambda_{\text{critical}}$ as the interface coupling strength needed for denaturation to occur, for instance by assuming that an average loss of one native contact leads to denaturation. However, the fluctuations between sequences are large enough that a better measure than the average values given in Fig. 3(a) is required. If we define a sequence as hard if its native-state structure is retained for a given λ, and soft otherwise, we find that the sequences divide into ''hard'' and ''soft'' groups with respect to denaturation in the manner shown in Fig. 4. Here the Q distribution for interfaces with two different degrees of hydrophobicity, $\lambda_h = 0.1$ and $\lambda_h = 0.4$, is presented. The height of the column $Q = 1$ shows the number of hard sequences which is by far the larger group for $\lambda_h = 0.1$. The softer sequences spread over $Q < 1$ values in a Gaussian-like distribution. The number of soft sequences is greater for $\lambda_h = 0.4$. (A very large increase in λ_h would of course give only a peak at $Q = 0$.) The shape of the distribution is fairly similar in the $\lambda_h = 0.1$ and $\lambda_h = 0.4$ plots. The main difference is that there is somewhat less weight at values of Q near unity in the $\lambda_h = 0.4$ case, leading to reduced values of $\langle Q \rangle$ over the soft sequences. The gradual decrease of $\langle Q \rangle$ for the soft sequences is better seen in Fig. 3(a) where the dashed curve shows $\langle Q \rangle$ as a function of λ_h for the soft sequence set only. ($\langle Q \rangle$ for the hard sequences set is by definition 1.) We find that even for small λ_h values $\langle Q \rangle$ is reduced to about 0.5, i.e., the change in conformation

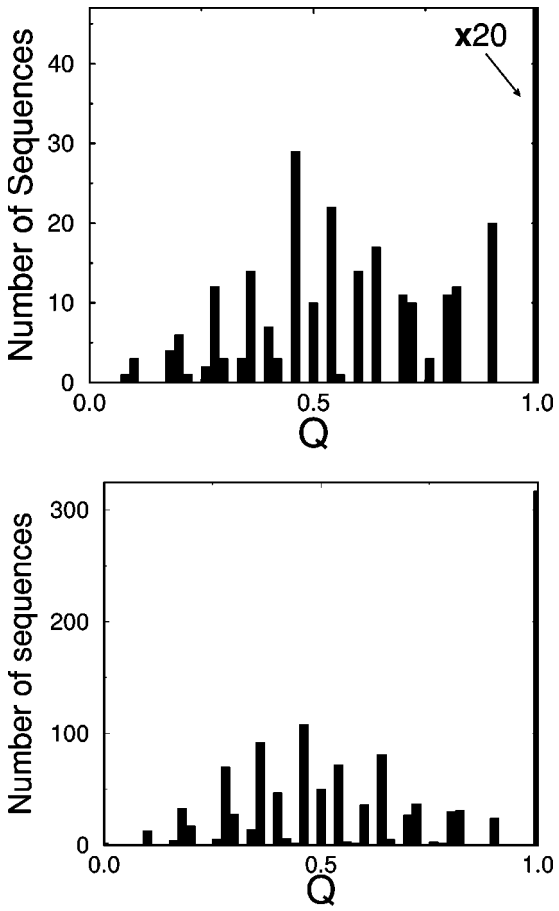

FIG. 4. Distribution of the fraction of preserved native contacts Q for our sequence set, at a hydrophobic interface with interface coupling strengths $\lambda_h = 0.1$ (a) and $\lambda_h = 0.4$ (b).

for soft sequences is very large even for small interface interactions. This we believe suggests that denaturation induced by the interface is a highly cooperative process, similar to solvent-induced denaturation.

Figure 5 shows the dependence of the fraction of hard sequences, F_{native}, on λ. Up to about $\lambda_h \simeq 0.5$, it shows a

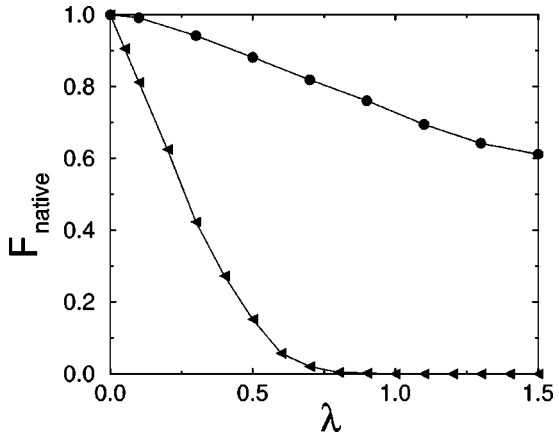

FIG. 5. The fraction of sequences adsorbed in native-state structure F_{native}, as a function of the interface coupling strength λ, for polar (circles) and hydrophobic interfaces (triangles).

linear drop. The linear behavior of $\langle Q \rangle$ vs λ_h in Fig. 3(a) then results from combination of this linear drop with the near constancy in $\langle Q \rangle$ for the soft set of sequences (dashed line) for $\lambda_h < 0.5$. We note that the curves for unfolding at polar interfaces (Figs. 3 and 5) show behavior very similar to the ones for hydrophobic interfaces, which makes polar interfaces comparable to hydrophobic interfaces with very weak interface coupling strengths. This is consistent with the experiments comparing hydrophobic and polar interfaces, discussed in Sec. I.

It is not possible to further compare our results directly to existing experimental data, both because our model is very simplified, and because the correct values of the interface coupling strengths for particular interfaces are not known. However, a crude estimate of a typical interface coupling strength λ_h can be obtained by calculating its value for the hydrophobic interface used in Ref. [23]. The density of the hydrophobic OP molecule adsorbed on the interface was up to 2.6 molecules/nm^2. Approximating the hydrophobicity of the OP molecule to that of H residues and taking as a lattice constant in our model the typical distance between two residues, roughly 4 Å, we estimate that for the most hydrophobic interface studied in that experiment, $\lambda_h \approx 0.4$. The coupling strength of a typical hydrophobic interface is therefore expected to lie in the range $\lambda_h \leqslant 0.4$. For polar interfaces it is more difficult to give an estimate.

In summary, we find that adsorbed model proteins that change from their native-state structure do so by a large change in their native contacts which, however, can correspond to only small changes in their bulk energy. Thus there is a clear division into a soft and a hard set of sequences. We also find that polar interfaces have an effect on protein structure comparable to that of very weakly hydrophobic interfaces.

B. Hardness criteria

As mentioned above, it is important to develop methods for predicting in statistical fashion whether a particular protein or class of proteins will denature on a particular interface. We define the hardness of a protein as the maximum interface interaction strength for which the protein remains adsorbed in its native-state structure. For hydrophobic interfaces the hardness is thus specified by the degree of hydrophobicity λ_h at which the first conformational change occurs.

1. A priori analysis of parameters relevant to interface denaturation behavior

The hardness of a protein is of course determined solely by the amino acid sequence. However, this relation is very complex, so we focus on establishing some properties of the protein that correlate more directly with the hardness. We define the energy gap E_g as the energy difference between the native-state structure and the next-lowest-energy structure for the protein in solution. We expect both E_g and the interaction E_{if} of the optimally oriented native-state protein with the interface to be important for the behavior of the adsorbed protein. This expectation is supported by noting that

$$|E_b'| \leqslant |E_b| - |E_g|, \tag{6}$$

where E_b' is the bulk energy for an arbitrary structure other than the native-state structure in solution. A sufficient condition for retaining the native-state structure at the interface is then that (see also Fig. 1)

$$|E_b| + |E_{if}| \geqslant |E_b| - |E_g| + |E_{if}'| \tag{7}$$

or

$$|E_{if}'| \leqslant |E_{if}| + |E_g|, \tag{8}$$

for all nonnative-state structures, where E_{if}' is the interface energy for a nonnative adsorbed-state structure. Large values of $|E_g|$ and $|E_{if}|$ will enhance the likelihood of inequality (8) holding, and should thus lead to enhanced hardness. We also expect the native bulk (folding) energy $|E_b|$ to be important. When this energy is large, the hydrophobic core residues are well coordinated by other residues. This results in a high-energy cost for extracting them to make contacts with the interface.

Of the three properties that we have considered, E_b and E_g are familiar from previous studies of protein folding, while E_{if} has not been treated. $|E_{if}|$ would be largest when the native-state structure of the protein in solution has an exposed hydrophobic patch, i.e., when the hydrophobic core in the native-state structure is only semicovered by a polar shell. This would normally give a destabilizing contribution to the native-state structure. However, since there are frustrations in the folding, such hydrophobic patches can and do exist in the native-state structures. $|E_{if}|$ includes indirectly topographical effects since it is favored by a flat hydrophobic patch matching the flat model interface. Such effects could be included in more detail in potential applications to experiment treated below.

In order to facilitate potential experimental tests of our results, we have also investigated the effect of the bulk thermal stability of the model protein on its hardness at an interface. The thermal stability is determined by the transition midpoint temperature T_m for which the probability of finding the protein in native state is 50%. This factor is highly correlated to the energy gap, and is measurable, by calorimetric measurements or by other methods (see for example, Ref. [27]). For our model proteins the thermal stability measure was expressed in terms of T_m, which was obtained by assigning Boltzmann weights to the exactly enumerated conformations.

2. Numerical results

In order to study the effect of E_b, E_g, T_m, and E_{if} on the hardness, we compare the average behavior for all studied sequences to the behavior of sequences that have extreme values of these parameters or combinations of them. We defined above F_{native} as the fraction of sequences that retain their native-state structure upon adsorption at a particular value of λ_h. Figure 6 shows F_{native} for all sequences, along with F_{native} for the sequences with values of $|E_b|$, T_m, or $|E_{if}|$ that are in the top 5%. We chose to concentrate the analysis on T_m rather than on E_g, because T_m is measurable. Further, the correlation coefficient for these two properties is high, 0.84 for the studied sequence set. In all cases the "top" sequences are harder than average, i.e., all factors seem to be positively correlated to the hardness. As discussed earlier, we believe $\lambda_h = 0.3$ is in the range of values achievable for real interfaces. At this value of λ_h, a sequence with $|E_b|$ in the top 5% has ≈60% better chance of retaining its native-state structure than an average sequence, and nearly twice the average chance if it instead has T_m (or $|E_g|$) in the top 5%. The effect of $|E_{if}|$ is similar to that of T_m. These results suggest that E_b, T_m, and E_{if} are indeed relevant factors for the protein hardness, and that T_m and E_{if} are slightly more important than E_b.

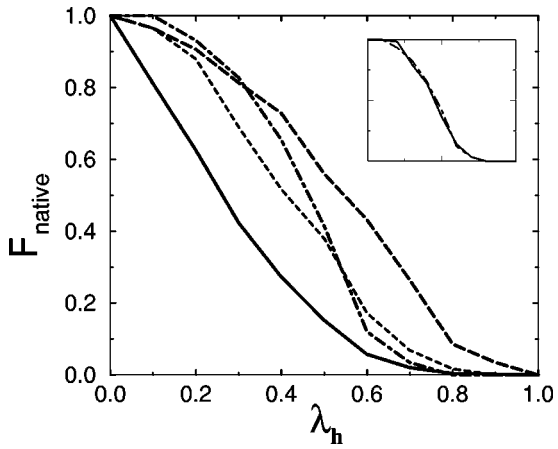

FIG. 6. Comparison of the fraction of all sequences adsorbed in native state, F_{native} (solid line), to that obtained for the sequences with 5% highest native-state bulk energy $|E_b|$ (dashed), thermal stability T_m (long-dashed), and interface interaction for the native-state structure $|E_{if}|$ (dot-dashed), respectively. The inset shows that the curve for 5% highest thermal stability T_m is very similar to the one for the energy gap $|E_g|$ (solid line).

The protein hardness is undoubtedly determined by a combination of factors, including those discussed here. For this reason we have investigated the effects of combining pairs of the three factors, and also combining all three. The results are presented in Fig. 7. In the combinations, we require that the number of selected top sequences is 5% of the total number of studied sequences so that the graphs can be easily compared with the graphs for a single factor. This means, however, that the requirements for the individual criteria are less restrictive. For example in order to obtain a number of top sequences that is 5% of the total number, we chose sequences with $|E_b|$ and T_m within the best 9%. In this case the combination criteria select sequences of similar hardness as selected by the top 5% of T_m. On the other hand, we find that model proteins selected for the other two pairs of the factors are slightly harder than those that are top sequences with respect to only one factor. In spite of the less strict criteria for individual factors (within the top 24% and 26%), the best results are obtained for the combination of large $|E_b|$ and $|E_{if}|$, and for $|E_{if}|$ and T_m ($F_{native} = 0.93$ for both compared to the average 0.42 at $\lambda_h = 0.3$). Even though

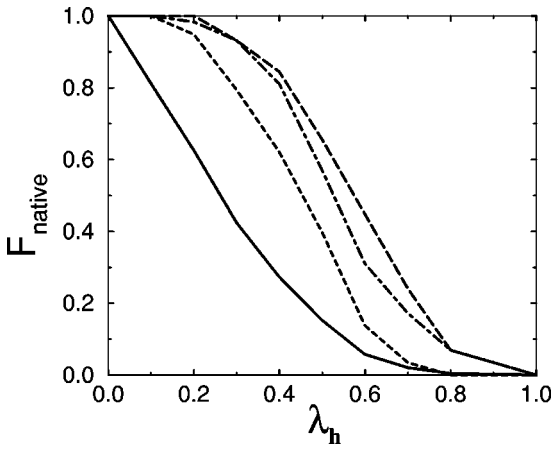

FIG. 7. Fraction of all sequences adsorbed in native state, F_{native} (solid line), compared to F_{native} for sequences satisfying combinations of criteria. The selected number of sequences is always required to be 5% of the total number of sequences. The criteria for each of the three factors (native-state bulk energy E_b, thermal stability T_m, and interface interaction for the native-state structure E_{if}) are therefore lowered as indicated. Dashed line, $|E_b|$ and T_m (sequences within top 9%); long-dashed line, $|E_b|$ and $|E_{if}|$ (within top 24%), dot-dashed line, T_m and $|E_{if}|$ (within top 26%). The curve for the all three criteria (within top 33%) is very similar to the curve shown for T_m and E_{if}.

E_b is less effective than the other two factors in the single-factor picture, it thus seems to be of similar importance when combined with other factors.

The effect of the factors is weaker but still present when we treat the top 10% instead of the top 5% for the three factors. For the combination of all three factors no new information is obtained, since the curve is similar to the one for the combination T_m and E_{if}. The similarity between the effects of T_m and E_g is shown in the inset in Fig. 6 and holds also for combinations of factors. We have also considered the hardness of sequences in the bottom 5% for the same factors and their combinations, and we find that the hardness is substantially reduced.

We have thus found that large values of $|E_b|$, T_m (or $|E_g|$), and $|E_{if}|$ are all correlated with increased hardness of the model proteins. One would expect these characteristics to be related to each other. For example, the native bulk energy is enhanced by having H residues in the protein core, while the interface energy is enhanced by having more H residues in the protein shell. This could make them competing characteristics. However, we found no strong correlations between the three properties for the studied set, apart from some correlation between high thermal stability and large native-state bulk energy. The correlation coefficients were 0.60 for $|E_b|$ and T_m, 0.10 for $|E_b|$ and $|E_{if}|$, and 0.17 for T_m and $|E_{if}|$.

C. Robustness analysis

Our results, the sharp transition from native-state to denatured structure, the division into hard and soft groups, as well as the importance of E_b, T_m (or E_g), and E_{if}, seem robust with respect to both chain length of the model proteins studied, and to variations in the interaction parameters between H, P_1 and P_2 residues. In order to check the dependence on chain length we performed calculations for a shorter chain (18-monomer, of which $8H$, $4P_1$ and $6P_2$). Even though we were forced to have a slightly different distribution of the amino acid types we observed the same general characteristics as reported for the 20-monomer chains. We evaluated the parameter dependence of the results by studying systems closer to the *HP*-model. All interactions apart from the dominant interaction E_{HH} were diminished to 50% and then also to 10% of the interaction values in Table II. In the former case we reproduced closely the general results from the original parameter set. However, when the non-*HH* interactions were reduced to only 10% of their starting values, the bulk energy of the native-state structure became very discretely distributed, which resulted in a sharper initial decrease in the $\langle Q \rangle$ diagram in Fig. 3(a). Typical changes in the quantities plotted in Figs. 3(a) and 5 were less than 10% in both cases.

We also believe that extension of the computations to three dimensions would yield similar conclusions. The suddenness of the unfolding seen in Figs. 3(a) and 4 is probably related to cooperativity. Since cooperativity increases in higher dimensions, the effect seen here should be enhanced in three dimensions. As discussed in Sec. II A, we also expect that using a model with nonadditive interactions would enhance the observed cooperativity [17–19]. Further, all three hardness factors we consider would remain important in three dimensions. Large $|E_g|$ (corresponding to high T_m) and $|E_{if}|$ would still enhance hardness, since the condition for retaining native-state structure on adsorption, Eq. (8), would still hold. The correlation with $|E_b|$ would continue to be important since a bulk hydrophobic core residue would still have more hydrophobic neighbors than it could at the interface.

D. Relation to experiments

Here we propose possible experiments to test our predicted correlations between protein properties and hardness. Due to the model chosen, our results would be most applicable to single-domain proteins without disulphide bonds. Since we present statistical results, a systematic study of a protein set, of order at least ten proteins, would be required. E_b could be identified with the folding free energy, which is measurable. We believe that, in comparing proteins of different lengths, the relevant variable is the folding free energy per residue. Furthermore, E_{if} could be approximated with the help of molecular mechanics code. In such a calculation, one would dock the protein on the interface, keeping the protein structure fixed, as in protein-protein docking [28]. One would thus perform a search over the six-dimensional position orientation space rather than having to fold the protein *de novo*. As discussed earlier, T_m can be measured straightforwardly [27].

To establish the validity of the predicted hardness correlations, one would first prepare a given hydrophobic inter-

face, then evaluate the structure of a large number of adsorbed proteins, and establish to what extent their denaturation or lack thereof correlates with E_b, E_{if}, and T_m. Extending these experiments to measure the hardness of individual proteins would involve preparing interfaces of varying hydrophobicity. This could be accomplished by the methods of Ref. [23], for instance using different types of adsorbed molecules or layers of molecules on a hydrophilic interface. One could then evaluate the fraction of proteins retaining their native fold as a function of λ_h, to see if the linear behavior predicted by Fig. 5 for $\lambda_h < 0.5$ is observed. These experiments with a range of values of λ could also be used to confirm the dependence of the protein hardness on E_b, E_{if}, and T_m.

It is hard to evaluate the magnitude of Q experimentally, and therefore our suggestion that the proteins are clearly divided into a hard and soft set (see Fig. 4) is difficult to confirm. To a certain extent this could be accomplished by existing methods for measuring a secondary structure profile. However, comparison between different proteins, for instance one containing primary α-helices to one containing primary β-sheets, would be difficult.

IV. CONCLUSIONS AND OUTLOOK

The aims of this study were to understand the nature of interface-induced unfolding and to clarify the factors which determine protein stability to interface-induced denaturation. We have found that the unfolding proceeds by a large loss of native contacts, making it sudden rather than gradual. Our comparison of unfolding at hydrophobic and polar interfaces has shown that their generic behavior is similar, but the latter has a much weaker interface coupling strength. Finally, we have found that the hardness of proteins in our model is correlated with the magnitude of the folding energy in the native-state structure, the thermal stability (or energy gap) for that structure, and the interface energy for native-state adsorption. We find these factors to be of roughly equal importance.

ACKNOWLEDGMENTS

We appreciate informative discussions with David Sept. We gratefully acknowledge financial support from the Swedish Foundation for Strategic Research through the Materials-Theory Consortium ATOMICS, and Jubilee Grant support from Chalmers University of Technology to A.E.C.

[1] Even though the term ''surface'' is used in most experimental and theoretical papers, we will here use the more correct term ''interface'' instead, leaving the former term to denote the protein surface.
[2] A.W. Vermeer, M.G. Bremer, and W. Norde, Biochim. Biophys. Acta **1425**, 1 (1998).
[3] A.W.P. Vermeer and W. Norde, J. Colloid Interface Sci. **225**, 394 (2000).
[4] A. Bentaleb, A. Abele, Y. Haikel, P. Schaaf, and J.C. Voegel, Langmuir **14**, 6493 (1998).
[5] C.E. Giacomelli and W. Norde, J. Colloid Interface Sci. **233**, 234 (2000).
[6] M.C. Coen, R. Lehmann, P. Groning, M. Bielmann, C. Galli, and L. Schlapbach, J. Colloid Interface Sci. **233**, 180 (2001).
[7] J.R. Lu, T.J. Su, P.N. Thirtle, R.K. Thomas, A.R. Rennie, and R. Cubitt, J. Colloid Interface Sci. **206**, 212 (1998).
[8] F. Hook, M. Rodahl, B. Kasemo, and P. Brzezinski, Proc. Natl. Acad. Sci. U.S.A. **95**, 12 271 (1998).
[9] K. Dill, Biochemistry **29**, 7133 (1990).
[10] H.S. Chan, M.R. Wattenbarger, D.F. Evans, V.A. Bloomfield, and K.A. Dill, J. Chem. Phys. **94**, 8542 (1991).
[11] H.S. Chan and K.A. Dill, Proteins: Struct., Funct., Genet. **24**, 345 (1996).
[12] T. Shen, L.S. Canino, and J.A. McCammon, Phys. Rev. Lett. **89**, 068103 (2002).
[13] D.J. Price and C.L. Brooks, III, J. Comput. Chem. **23**, 1045 (2002).
[14] W. Wang, O. Donini, C.M. Reyes, and P.A. Kollman, Annu. Rev. Biophys. Biomol. Struct. **30**, 211 (2000).
[15] Q.W. Bolen, and I.V. Baskakov, J. Mol. Biol. **310**, 955 (2001).
[16] K.F. Lau and K.A. Dill, Macromolecules **22**, 3986 (1989).
[17] H.S. Chan, Proteins: Struct., Funct., Genet. **40**, 543 (2000).
[18] H. Kaya, and H.S. Chan, Proteins: Struct., Funct., Genet. **40**, 637 (2000).
[19] H. Kaya, and H.S. Chan, Phys. Rev. Lett. **85**, 4823 (2000).
[20] S. Miyazawa and R.L. Jernigan, Proteins: Struct., Funct., Genet. **34**, 49 (1999).
[21] K.A. Dill, S. Bromberg, K. Yue, K.M. Fiebig, D.P. Yee, P.D. Thomas, and H.S. Chan, Protein Sci. **4**, 561 (1995).
[22] N. Socci and J.N. Onuchic, J. Chem. Phys. **101**, 1519 (1994).
[23] K. Kandori, M. Mukai, A. Fujiwara, A. Yasukawa, and T. Ishikawa, J. Colloid Interface Sci. **212**, 600 (1998).
[24] Kandori *et al.*, measured the saturated adsorbed amount for lysozyme and bovine serum albumin as a function of the degree of hydrophobicity of the interface. The interface in this case is initially pure calcium hydroxyapatite (hydrophilic), but becomes more hydrophobic as oleyl phosphate molecules are adsorbed on it [23] . The work focused on the adsorbed amount of protein and did not study the conformational changes.
[25] H. Cejtin, J. Edler, A. Gottlieb, R. Helling, H. Li, J. Philbin, N. Wingreen, and C. Tang, J. Chem. Phys. **116**, 352 (2002).
[26] N.D. Socci, J.N. Onuchic, and P.G. Wolynes, Proteins: Struct., Funct., Genet. **32**, 136 (1998).
[27] A. Sinha, S. Yadav, R. Ahmad, and F. Ahmad, Biochem. J. **345**, 711 (2000).
[28] I. Halperin, B. Wolfson, and R. Nussinov, Proteins: Struct., Funct., Genet. **47**, 409 (2002).
[29] G.E. Schulz and R.H. Schirmer, *Principles of Protein Structure* (Springer, New York, 1979).
[30] D.L. Nelson and M.M. Cox, *Lehninger Principles of Biochemistry* (Worth, New York, 2000).