PHYSICAL REVIEW E **67**, 050901(R) (2003)

Preferential adsorption of hydrophobic-polar model proteins on patterned surfaces

N.-K. Lee and T. A. Vilgis
Max-Planck-Institute for Polymer Research, P.O. Box 3148, D-55021 Mainz, Germany
(Received 21 August 2002; revised manuscript received 4 February 2003; published 28 May 2003)

We study the adsorption of a single hydrophobic-polar (HP) model protein under the influence of a flat but specially designed surface. A folded HP model protein is brought to the surface with a designed pattern consisting of certain attractive and repulsive sites for the different monomers (amino acids). In contrast to the deformation of a random sequence that is continuous, deformation of any proteinlike sequences is unlikely and an energy gap is associated with it. The surface with a certain wavelength of pattern attracts a certain type of folded structure preferentially and the free energy of the combined system is reduced. The model presented here represents a minimal theoretical model for protein recognition.

DOI: 10.1103/PhysRevE.67.050901 PACS number(s): 87.15.Nn, 36.20.Fz

Protein adsorption plays a major role in many technological applications and biological processes [1–3]. One important aspect is the immobilization of proteins, which is a necessary and basic step preceding any biotechnological manipulation. Moreover, in nature, certain macromolecules initiate a well defined biological activity through recognition. In particular, proteins adsorb selectively on structured surfaces. Typical examples of such surfaces can be binding sites of cells or designed antibodies which recognize the primary or secondary structure of proteins via their specific sequence of amino acids [1].

We present here a minimal theoretical model that captures some essential aspects of the physics of protein recognition. Coarse grained models for biomolecules consist of two or more monomer types which represent at the lowest order the amino acids [4,5]. These ''monomers'' are arranged in the sequence as a random heteropolymer (RHP), which represent in theoretical physics the so-called quenched disorder. Indeed, RHPs have been recognized as models for proteins since amino acids in proteins can be classified either hydrophobic-(H) and polar-(P) type monomers [6–8].

Biological activity and function in proteins are strongly related to their spatial conformation, which corresponds to a certain folded state (secondary structure). This is mainly determined by the sequence of the amino acids (primary structure), since they interact locally with the environment, i.e., a physiological salt solution (which acts as a selective solvent for the H- and P-type monomers). If a HP sequence carries ''relevant information,'' the RHP forms a frustration-free globular state, where the hydrophobic groups are located inside the globule and the hydrophilic groups are arranged on the surface to allow their solubility in water.

Most interesting are the interactions of RHPs with surfaces [2,3,9,10] or other ligands, whereby certain conformations are recognized. Naturally, the geometric matching between the protein and the ligand is an important factor for the adsorption. However, the binding affinity of the protein to the surface is determined by the arrangement of the H and P monomers, i.e., by the sequence in protein structure. At the binding sites of the proteins, hydrophobic monomers are exposed on the surface so that burial of the exposed hydrophobic surface is an important driving force for the protein binding [11]. The RHP becomes strongly adsorbed if the design (or pattern) of the surface affinity matches with the surface pattern of the RHP. Such a preferential adsorption and surface-induced deformations are essential characteristics of a theoretical model for antigen-antibody interactions in immunology.

We study here the adsorption of single-folded chains with various and deliberately chosen sequences. We select such RHP chains whose sequences form compact, folded proteinlike structures, i.e. (1) the chain is in a compact conformation and (2) there is free-energy gap between the native structure and the rest of the conformations. Although it is difficult to define one unique (native) structure within the simple HP model, we make use of force-extension relations (free-energy gap along the force-induced deformation) to identify sequences with one free-energy gap. Such chosen sequences (p1 and p2 in Fig. 1) show an abrupt unfolding behavior under the application of nanomechanical forces [8]. When these chosen sequences are brought near specially designed surfaces, the resulting conformational properties are different from the arbitrary random sequences (n1 in Fig. 1). While random sequences have a tendency to change their structure, the chosen sequences of globular forms either immobilize at the surface or are repelled depending on the choice of the surface design without drastic changes of their conformation.

We use a variational method [12–16] to find optimal conformations of single chains with specific sequences of H and P monomers under the influence of a preferentially attractive or repulsive surface. By estimating the free-energy change related to the adsorption or the adaptation of their conformation, we show that selective HP-protein adsorption of certain designed surface patterns can be attributed to the specific sequences. Earlier, we employed this technique successfully to the elastic response of RHPs in terms of a mechanical unfolding of HP proteins, where we could relate the primary structure in the polymers to the single-chain force-extension relation [8].

Formally, we consider a polymer chain consisting of N monomers of size b interacting with a surface:

$$H = H_{\mathrm{HP}} + \sum_i \int d\vec{r}_i V(\vec{r})\,\delta(\vec{r}_i - \vec{r})\,\delta(z_i)\,\theta(i), \qquad (1)$$

where $\vec{r}_i = (\rho_i, z_i)$ designates the spatial location of the ith

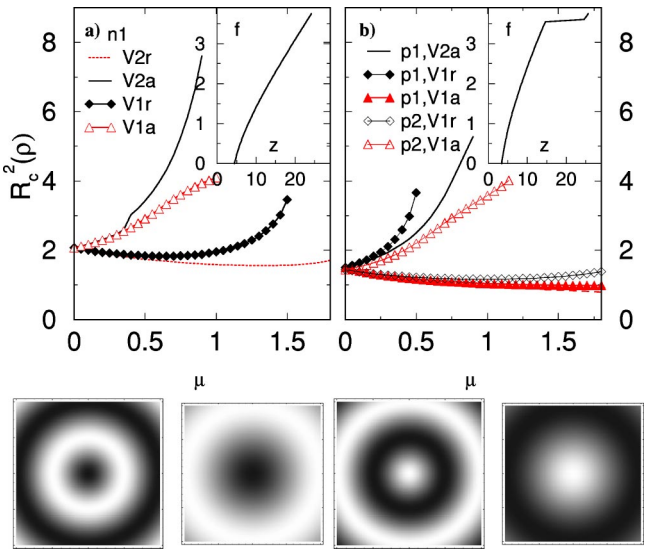

FIG. 1. The square of the radial size of cores for (a) a typical random sequence (n1) and (b) two proteinlike sequences (p1 and p2) interacting with patterned surfaces. The $8b\times 8b$ central sector of each pattern (V1a, V2a, V1r, and V2r) is shown below. The gray scale indicates the affinity to the H-type monomers (black: attractive, white: repulsive). The sequences of chains are n1] HPHPPH PPHPHP PPHHPH PPHHHH HPPPPP HHHPHH, p1] PHPHPP HPHHHH PHPHPH PHHPPH HHPHPP PHHPPP, p2] HPHPPH HHPPPP HHHHHH PHHHPP HPHPPH HPPPPP. Corresponding force-extension curves (n1 and p1) are shown in insets.

monomer, while $\theta(i)$ identifies the type of the monomer (H or P) located at the ith monomer. If the surface is attractive or repulsive only to H monomers, $\theta(i)=1$ if ith monomer is type H, otherwise $\theta(i)=0$. The set $\{\theta_i\}_1^N$ represents the sequence-surface interaction. The potential $V(\vec{r})$ contains all properties of the surface and its pattern on structure. Note that the potential scales like a typical surface-chain interaction, i.e., $\propto N/R_z$, where R_z is the perpendicular extension with respect to the surface. The ''bare'' term H_{HP} collects the connectivity and monomer interaction terms (two- and three-body interactions of the virial expansion) in the standard manner,

$$H_{\mathrm{HP}}/k_BT=\frac{d}{2b^2}\sum_{i=1}^{N}(\vec{r}_{i+1}-\vec{r}_i)^2+\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{v_{ij}}{2!}\delta(\vec{r}_i-\vec{r}_j)+\frac{w}{3!}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\delta(\vec{r}_i-\vec{r}_j)\delta(\vec{r}_j-\vec{r}_k). \quad (2)$$

The first term corresponds to the elastic properties (connectivity) of the polymer chain [17]. The inclusion of the three-body interactions $w\sim O(b^6)$ prevents the chain to collapse to a point and stabilizes the density of the compact globule. The two-body interactions v_{ij} can be attractive or repulsive depending on the type of the monomer pairs, and are conveniently written as $v_{ij}=v_0-\frac{1}{2}[\alpha(\sigma_i+\sigma_j)+\chi(\sigma_i\sigma_j)]$ [18]. The sequence of monomers is described by a set of Ising spin-type variables σ_i with their values $\sigma_i=1$ if monomer i is of type H and $\sigma_i=-1$ if it is of type P. These sequence information are directly translated to the initial conformation. The Flory parameter $\chi=v_{\mathrm{HP}}-(v_{\mathrm{HH}}+v_{\mathrm{PP}})/2$ is positive when similar monomers attract each other and $\alpha=\frac{1}{2}(v_{\mathrm{PP}}-v_{\mathrm{HH}})$. We use the standard values for HP proteins $v_{\mathrm{HH}}=-2.3k_BT$, $v_{\mathrm{HP}}=-1.0k_BT$, and $v_{\mathrm{PP}}=0k_BT$ [19]. Here, $\chi=0.15k_BT$ indicates that different monomer types are naturally separated, unless the connectivity prevents. For more realistic approaches, 20-letter-amino-acid models with preferred bond angles [20] could be alternatively chosen. However, we restrict ourselves here to the simplest level to avoid computational difficulties. As we will show below, the two-letter model is sufficient to catch the main features of preferential adsorption processes.

The variational principle [12–15] uses Gaussian trial Hamiltonians H_0 with well known properties and uses Feynman's inequality $F\leqslant F_V\equiv\langle H-H_0\rangle_0+F_0$, where $\langle\cdots\rangle_0$ stands for the average over the variational probability distribution, $P_V(\vec{r}_1,\ldots,\vec{r}_N)=Z_V^{-1}\exp\{-H_0(\vec{r}_1,\ldots,\vec{r}_N)/k_BT\}$, where Z_V is the normalization constant satisfying, and $F_0=-k_BT\ln Z_V$.

In the presence of a surface or interface, it is useful to describe the conformation of the chain in cylindrical coordinates by the use of two order parameters. We distinguish the chain deformation in radial and axial directions. In terms of two components of the correlation function, we choose consequently the trial Hamiltonian H_0 as

$$\frac{H_0(\vec{r}_1,\ldots,\vec{r}_N)}{k_BT}=\frac{1}{2}\sum_{j,l=1}^{N}[G_\rho^{-1}(j,l)\vec{\rho}^j\cdot\vec{\rho}^l+G_z^{-1}(j,l)\vec{z}^j\cdot\vec{z}^l], \quad (3)$$

where $G_\rho(j,l),G_z(j,l)$ are correlation functions for monomers j and l in the radial and axial directions, respectively (see also Refs. [8,16]). After minimization of the variational free energy with respect to G_z,G_ρ, we obtain coupled equations in an analytic form. We obtain the solution of these coupled equations self-consistently in a numerical way.

Before we proceed with the variational results, we provide some general arguments. It is important to realize that an isolated RHP has significantly broad variations in its primary structure, i.e., in sequences and block sizes. Naturally, the initial bulk conformation of the RHP depends on the sequences and, in turn, the conformations on the surface also.

For completely random sequences, the RHP forms, on an average, $N/4$ blocks with an average size $\overline{N}_{\mathrm{H}}\sim 2+2/\sqrt{N}$. Due to topological restrictions given by the connectivity, not every H monomer can join a single core. The optimal number of aggregation blocks per core p^* is mainly determined by the balance between the surface energy gain and elastic energy penalty of H blocks. The free energy of a core made of aggregation of p H blocks is $F/k_BT=pR_c^2/\overline{N}_{\mathrm{H}}+\gamma R_c^2$, with $R_c=(p\overline{N}_{\mathrm{H}})^{1/3}$ and $\gamma=k_BTv_{\mathrm{HH}}^2/b^8$. The minimizing of free energy provides the optimal aggregation number $p^*=\overline{n}_{\mathrm{H}}v_{\mathrm{HH}}^2/4$. Depending on the value of p^* for a given sequence and an interaction strength, there are two cases in globule conformation: (1) strong globule and (2) weak globule. If p^* is larger than the total number of H blocks n_{H}, all

H monomers form a single core (strong globule). In contrast, if p^* is less than n_H, H-block segments are distributed into several cores. The cores attract each other via P-block bridging ($v_{HP}<0$) and form an assembly of globules consisting of various sizes of smaller micelles [21]. The effective surface tension of this assembly is $\gamma^{eff}=k_BTv_{HP}^2/b^8$ [21]. The binding energy between the smaller cores is relatively weak ($|v_{PH}|<|v_{HH}|$). Proteinlike sequences (of single-domain protein) are a subset of the strong globules. The differences on the sequence are reflected in the behavior of RHP at attractive walls. It turns out that some sequences (weak globules) can easily deform and spread out on the surface, whereas other sequences (strong globules) are more resistant to any conformational changes.

First, we consider the deformation of the core when exposed on a plain attractive surface, $V(r)=-\mu$, $\mu>0$. The core forms roughly a half sphere with surface area $2\pi r_0^2$ on the wall to minimize the contact energy to the solvent. When μ is smaller than the effective surface tension of the core, the deformation is negligible. When the surface interaction strength μ increases, the spherical core deforms to a half ellipsoid. The size in the z direction is reduced to $c<r_0$, but the radial symmetry remains and the radial size is $a>r_0$. The surface area of such an ellipsoid is $(2/a+1/c)V_0$, where V_0 is constant volume, given by $V_0=(2\pi/3)a^2c=(2\pi/3)r_0^3$. The force acting on the circumference of the ellipsoidal droplet spreads the droplet. The surface energy cost by spreading is compensated by the energy gain due to the adsorption of more monomers on the surface. The equilibrium condition is obtained when this force is equal to the surface tension of the droplet. The free energy of the core block is $F/k_BT=\gamma[(2V_0/a)+(2\pi a^2/3)]-\mu a^2$. The spreading force in radial direction vanishes under the condition: $dF/da=0$. The radial size a is growing as the adsorption strength increases, $a\sim(\gamma V_o/[(2\pi\gamma/3)-\mu])^{1/3}$. For the deformation of a weak globule, we may then simply replace γ by γ^{eff}. This corresponds to the energy cost for adjusting p-block adsorption bridges between the cores.

We present now the adsorption of our model protein beyond the phenomenological description. The main result is that some surface patterns with characteristic length scales are able to recognize certain conformations and sequences of the HP model protein by matching the internal block size. The patterned surfaces, which will be used here, are simple models of substrates consisting of more than one chemical component. The H monomers are either attractive or repulsive to the substrates depending on the chemical components of substrates. For simplicity, we assume that all P monomers do not (or only weakly) interact with the surface. This assumption does not change the physical picture, but shifts only the energy scale by an irrelevant constant. The substrate which is preferentially attractive to the H-type monomer mimics the exposed hydrophobic monomers on binding sites. The repulsive sites can be considered as polar sites or alternatively as excluded volume effects which are unfavorable for the binding.

All internal correlation lengths are calculated by using the variational method for different patterns and different sequences. In Fig. 1, we show the radial core size of the H monomers R_c^2 for several sequences. The first sequence (n1) is an example for the weak globule which has the continuous force-extension curve reflecting the continuous energy spectrum. The second (p1) and the third (p2) sequence correspond to compact proteinlike globules.

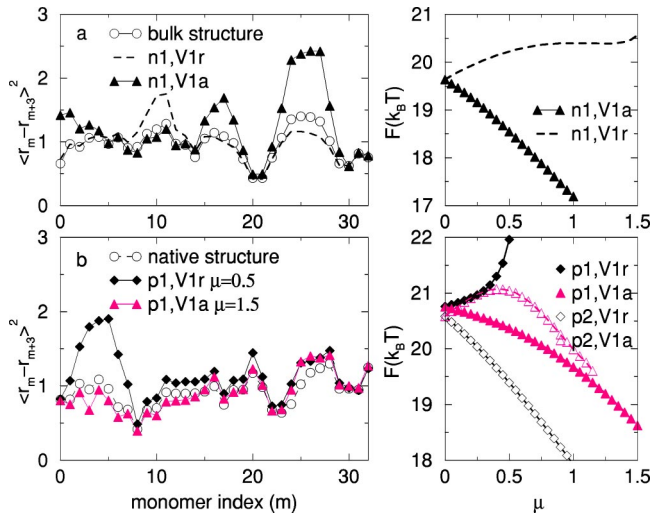

FIG. 2. The internal correlation functions (left pannels) and free energy (right pannels) of typical sequences as increasing interaction strength μ.

For calculational convenience, we use a surface potential that has its concentric center exposed close to the center of mass of the chain, such that $V(\rho)=\mu\cos(2\pi\rho/R_0)$. We identify the pattern by two parameters, its prefactor μ and the wavelength R_0. If the pattern has an attractive (repulsive) center for $\mu<0$ ($\mu>0$), with a wavelength R_0, indicated by $VR_0a(r)$ (see Fig. 1). At places where $V(\rho)<0$, the surface attracts the H monomers and vice versa. For the actual demonstration, we have chosen sequences of 36 monomers.

The minimization of free energy involves (1) the elastic contribution of a P loop with corresponding entropy $S\sim k_BTd^2/n_l^2$, more importantly (2) the energy gain or cost of core blocks of length n_h by combining with favorable or unfavorable substrate type at ρ, estimated to $\delta E=\pm\mu(\rho)n_h^{2/3}$, and (3) the surface energy cost $E_{surface}=\gamma(\delta A)$ by deforming the core structure.

If the diameter of the attractive center is large enough compared to the core size (V2a), all possible contacts are favorable. The radial size of the cores increases for all sequences, i.e., proteinlike sequences do not deform much from the original conformation. When the pattern has a relatively small wavelength, the chain segments experience the heterogeneity of the surface directly, and therefore, become selective.

This selection works for a sequence with a large internal binding energy for which the possibility of internal deformation is suppressed. The primary mechanism for the preferential adsorption relies on reducing the energy of the combined system with pattern matching. The total free energy of the chain strongly depends on the arrangement of the primary structure on the pattern. We observe selection by ''pattern

matching'' among proteinlike sequences when the typical wavelength of the pattern is of the order of monomer size, $R_0 = 1b$ (see Fig. 1). The pattern matching occurs by locating core blocks in adequate positions on the surface, while the neutral *P* blocks can accommodate on places that are H repulsive. When the length of the P-block size matches the wavelength of the repulsive pattern, the preference becomes obvious. The deformation in the internal structure becomes likely only after overcoming the energy barrier associated with the breaking of one single core to two or more, correspondingly, smaller cores. The adsorption that requires the conformational changes is suppressed for strong globules. Hence, a strong globule with a specific sequence can either become attracted to the favorable surface pattern or be repelled from the unfavorable surfaces. For weak globules (made out of random sequences), any necessary deformation into the optimal conformation according to the surface pattern can be achieved easily.

The internal structure change (local conformational changes), when the folded structure is forced closer to the patterned surface, is shown in Fig. 2. It exists an energy gain by adsorption in the pattern matching case (p1 and V1a, p2 and V1r). The pattern V1r is repulsive to p1. For a weak globule (n1), the conformation changes upon adsorption and this does not cost much energy.

In summary, we have shown that proteinlike sequences (corresponding to a single-core conformation) are more stable with respect to the influence of the external surface. For such sequences, we observe a preferential adsorption when the internal length scale matches with length scales in the given pattern. Moreover, we developed here an analytical method to calculate the general behavior of model-type proteins with specific sequences on surfaces. Since specific sequences can be addressed, this method enables further possibilities in future. These include specific calculations for HP proteins where special sequences may have defects.

N.-K.L. appreciated the financial support of the BMBF—Germany and the Nanocenter Mainz, and E. Jarkova and M. H. Hettler for the helpful discussions. T.A.V. acknowledges helpful discussions with H. Decker from the Biophysics Department of the University of Mainz.

[1] I. Roitt and P.J. Delves, *Essential Immunology* (Blackwell, Oxford, 2001).
[2] S. Srebnik, A.K. Chakraborty, and E.I. Shakhnovich, Phys. Rev. Lett. **77**, 3157 (1996); S. Srebnik, J. Chem. Phys. **112**, 9655 (2000).
[3] D. Bratko, A. Chakraborty, and E. Shakhnovich, Chem. Phys. Lett. **280**, 46 (1997).
[4] A. Irbäck, C. Peterson, and F. Potthast, Proc. Natl. Acad. Sci. U.S.A. **93**, 9533 (1996).
[5] V. Pande, A. Grosberg, and T. Tanaka, Biophys. J. **73**, 3192 (1997).
[6] T. Garel and H. Orland, Europhys. Lett. **6**, 597 (1988).
[7] E. Shakhnovich and A.M. Gutin, J. Phys. (France) **50**, 1843 (1989).
[8] N.-K. Lee and T.A. Vilgis, Europhys. Lett. **57**, 817 (2002); N.-K. Lee and T.A. Vilgis, Eur. Phys. J. B **28**, 451 (2002).
[9] J. Genzer, Macromol. Theory Simul. **11**, 481 (2002).
[10] A.J. Golumbfskie, V.S. Pande, and A.K. Chakraborty, Proc. Natl. Acad. Sci. U.S.A. **96**, 11 707 (1999).
[11] M. Averof, A. Rokas, K.H. Wolfe, and P.M. Sharp, Science (Washington, DC, U.S.) **287**, 1283 (2000).
[12] J.-P. Bouchaud, M. Mezard, G. Parisi, and J.-S. Yedida, J. Phys. A **24**, L1025 (1991).
[13] J. des Cloizeaux, J. Phys. (Paris) **31**, 715 (1970).
[14] B. Jönsson, C. Peterson, and B. Söderberg, Phys. Rev. Lett. **71**, 376 (1993).
[15] G. Migliorini, N.-K. Lee, V. Rostiashvili, and T.A. Vilgis, Eur. Phys. J. E **6**, 259 (2001).
[16] R.R. Netz and H. Orland, Eur. Phys. J. B **8**, 81 (1999).
[17] M. Doi and S.F. Edwards, *The Theory of Polymer Dynamics* (Clarendon Press, Oxford, 1986).
[18] S.P. Obukhov, J. Phys. A **19**, 3655 (1986).
[19] H. Li, R. Helling, C. Tang, and N. Wingreen, Science (Washington, DC, U.S.) **273**, 666 (1996).
[20] S. Miyazawa and R.L. Jernigan, Macromolecules **18**, 534 (1985).
[21] N.-K. Lee, A. Johner, and T.A. Vilgis, Macromolecules **35**, 6043 (2002).