Effect of Spatially Distributed Hydrophobic Surface Residues on Protein−Polymer Association

Malin Jönsson,*,† Marie Skepö,‡,§ Folke Tjerneld,† and Per Linse‡

Biochemistry, Lund University, Box 124, SE-221 00 Lund, Sweden, and Physical Chemistry 1, Lund University, Box 124, SE-221 00 Lund, Sweden

Received: October 28, 2002; In Final Form: January 22, 2003

The effect of the spatial distribution of hydrophobic surface residues on the adsorption of a weakly hydrophobic polymer to proteins has been examined using a coarse-grain model solved by Monte Carlo simulations. A given number of surface sites were distributed randomly on the protein surface subjected to a distance constraint. Five protein classes for which the minimum distance between the sites was 2, 3, 4, 5, and 6 Å were considered, and for each class 10 proteins with different site distributions randomly generated were examined. As the strength of the hydrophobic interaction was increased, the onset of the polymer adsorption to the protein appeared first for proteins with a more *heterogeneous* site distribution. This holds both for the systematic variation of the site distribution for proteins of different classes and for the random variation among proteins within a class. The degree of heterogeneity of the site distributions was quantified using the variance of the number of sites located within randomly positioned circles placed on the protein surface. The conformational changes of the polymer at the adsorption were also studied.

Introduction

Protein−polymer interactions are of great interest in various biological as well as industrial applications such as precipitation, chromatography, and aqueous two-phase systems.[1−3] The protein−polymer interaction and complexation have been investigated extensively both experimentally and theoretically.[4−13] The Coulombic (electrostatic) interaction is generally considered to be the major driving force behind protein−polyelectrolyte complexation. For uncharged macromolecules, however, the hydrophobic interaction is often the dominant factor. Nevertheless, in systems containing charged macromolecules the hydrophobic interaction still often plays an important role.

Hydrophobic interactions in protein−polymer complexation are generally more important when dealing with slightly hydrophobic polymers like poly(propylene oxide) (PPO) or hydrophobically modified polymers. For instance, Tribert et al. have shown the existence of complexation of a hydrophobically modified polyacrylate (HMPA) to bovine serum albumin (BSA), whereas no complexation occurred for the nonmodified polyacrylate.[7] In a further study, Porcar et al. investigated the importance of hydrophobicity in the association between BSA and HMPA.[8] An increased association was found as the density or the length of the hydrophobic motifs of HMPA was increased. Furthermore, Hattori et al. have found an association between the negatively charged β-galactosidase and a negatively charged polyelectrolyte; hence, other types of interaction such as hydrophobic interactions are here most likely operating in the formation of protein−polymer complexes.[10]

Partitioning of proteins in aqueous two-phase systems of random ethylene oxide−propylene oxide copolymers (EOPO) and dextran has been considered by Berggren et al. They observed a correlation between the amount of hydrophobic (aliphatic and aromatic) residues on several different mutations of cutinase and the affinity for the EOPO-enriched polymer phases.[11] Moreover, they also partitioned different globular and similar sized proteins and found that proteins with more hydrophobic surface residues displayed higher preference for the more hydrophobic EOPO-rich phase than for the dextran-rich phase.[13] Furness et al. have used NMR to investigate which amino acids on the surface of lysozyme are involved in the nonspecific interaction with water-soluble polymers like EOPO and poly(ethylene glycol) (PEG).[5] In the investigation by Furness et al., it was concluded that aromatic residues like tryptophans are important in the interaction with slightly hydrophobic polymers such as EOPO and that the interactions are primarily of hydrophobic character. Finally, Carlsson et al. showed that, in addition to electrostatic interactions, a short-range protein−polymer attraction could to bring simulation results in agreement with experimental ones.[12] Thus, it has been demonstrated that hydrophobic interactions play an important role in many different types of protein−polymer systems.

We expect that not only the number of hydrophobic surface residues but also their actual *distribution* on the protein surface may affect the strength of the protein−polymer association. However, to our knowledge there are yet no experiments examining the effect of the distribution of hydrophobic surface residues. Nevertheless, on a more general level some computer simulations as well as theoretical studies have considered pattern recognition of heteropolymers.[14−18]

In this report, the effect of an attractive interaction between proteins with different spatial distributions of hydrophobic surface residues and a slightly hydrophobic polymer has been investigated by Monte Carlo simulations. In particular, the adsorbed amount of polymer and polymer configurations as a function of (i) the interaction strength and (ii) the distribution of the hydrophobic sites were considered. Our aim is to gain some fundamental insight into the interaction between globular

† Biochemistry.
‡ Physical Chemistry 1.
§ Formely Marie Jonsson.
* Corresponding author: e-mail Malin.Jonsson@biokem.lu.se.

10.1021/jp0223253 CCC: $25.00
Published on Web 05/14/2003

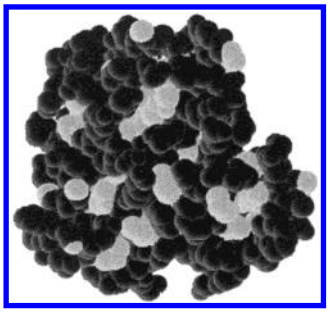

Figure 1. Crystal structure of myoglobin, 1YMB, with the side chains of hydrophobic amino acids at the surface colored in light gray. Amino acid residues with aliphatic side chains (Ala, Val, Pro, Leu, and Ile) and aromatic side chains (Tyr, Phe, and Trp) were considered to be hydrophobic. Data taken from Brookhaven Protein Data Bank (PDB).

proteins with different statistical patterns of their surface residues and a flexible and slightly hydrophobic polymer, rather than investigating a particular protein−polymer association. Thus, a coarse-grain model has been employed, in which the protein is represented by a hard sphere with discrete interaction sites at the surface and the polymer by a chain of beads connected by harmonic bonds.

The hydrophobic interaction is an effective potential with a complicated background appearing between two nonpolar units in an aqueous solution. The hydrophobic interaction between two methane molecules or two benzene molecules in water constitutes the prototype of the hydrophobic interaction between alkanes and aromatic groups, respectively, and these have been examined theoretically by several groups.[19−28] Considering hydrophobic groups embedded in curved surfaces, the situation becomes even more complicated. Therefore, a simple short-range and attractive term representing the hydrophobic interaction between the hydrophobic protein surface residues and a polymer segment has been employed.

Model

The aqueous solution of a globular protein and a slightly hydrophobic polymer is described using a simple model. The protein is represented by a large sphere with hydrophobic surface sites and the polymer by a sequence of connected hard spheres (segments). Water is treated as a continuum and enters the model only indirectly by the appearance of short-range hydrophobic interactions between hydrophobic sites of the protein and polymer segments.

Protein. The construction of the protein model has been guided by the structure of several small to medium sized globular water soluble proteins. Such a protein (myoglobin) is shown in Figure 1 with ∼20−30 hydrophobic surface residues brightly colored. Amino acids with aliphatic side chains (Ala, Val, Pro, Leu, and Ile) and with aromatic side chains (Tyr, Phe, and Trp) were considered to be hydrophobic. The same set of amino acids were found to enhance the interaction with random EOPO copolymers.[11,13]

In our model, the protein is described by a hard sphere with radius $R_p = 16$ Å containing $N_{hs} = 30$ hydrophobic interaction sites. The sites are placed at 13 Å from the center of the sphere, and hence 3 Å below the surface, merely to produce hydrophobic patches rather than hydrophobic points on the protein surface. All hydrophobic sites are equivalent, which of course is a simplification of a real protein where the hydrophobic amino acids have side chains of different hydrophobicity and different degree of surface exposure.

The hydrophobic sites were randomly distributed on the protein surface using a uniform distribution with the restriction

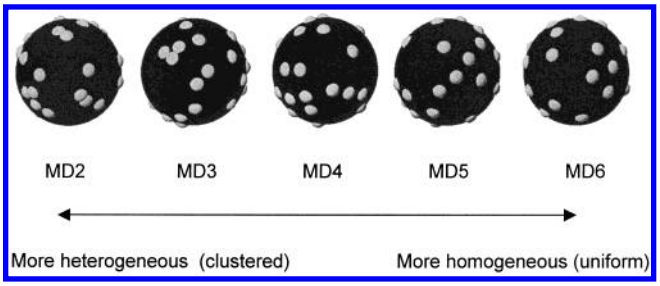

Figure 2. Pictures of one typical protein from each class (MD2, MD3, MD4, MD5, and MD6) illustrating how the hydrophobic surface sites (bright spots) become more homogeneously distributed as the minimum distance between the sites is increased. For increased visibility of the positions of the hydrophobic sites, the radius representing the protein has here been reduced from 16 to 14 Å, and each site is represented by a sphere of radius 2 Å (same is applied in Figures 3, 6, and 8).

that any two sites should be separated by at least the distance R_{min}. Five classes of proteins were considered characterized by $R_{min} = 2$, 3, 4, 5, and 6 Å, and the classes are referred to as MD2 (minimum distance 2 Å), MD3, MD4, MD5, and MD6, respectively. Ten different distributions were generated using different sequence of random numbers for each class, and the different proteins within a class are identified as protein 1, protein 2, ..., and protein 10, respectively. Consequently, we end up with a total of $5 \times 10 = 50$ proteins. Although the positions of the hydrophobic sites were selected randomly, the sites become more homogeneously distributed as R_{min} is increased. Figure 2 shows one typical protein from each class, demonstrating how the surface sites become more homogeneously distributed as R_{min} is increased. To improve the visualization of the position of the sites, the radius of the protein is decreased from 16 to 14 Å in Figure 2. The sites do actually not protrude from the surface.

Polymer. The polymer is described by a sequence of $N_{seg} = 40$ hard spheres (segments) connected with harmonic bonds. The polymer is fully flexible, and a segment radius of $R_{seg} = 2$ Å is used.

Potential Energy. All interactions are taken as pairwise additive. The total potential energy, U_{tot}, can be divided into two terms according to

$$U_{tot} = U_{nonbond} + U_{bond} \quad (1)$$

The nonbond potential energy, $U_{nonbond}$, is moreover divided into a hard-sphere potential energy, U_{hs}, and a hydrophobic potential energy, $U_{hydrophob}$, according to

$$U_{nonbond} = U_{hs} + U_{hydrophob} = \sum_i \sum_{j>i} u_{ij}^{hs} - \sum_i \sum_j \frac{\epsilon}{r_{ij}^6} \quad (2)$$

with

$$u_{ij}^{hs} = \begin{cases} 0, & r_{ij} \geq R_i + R_j \\ \infty, & r_{ij} < R_i + R_j \end{cases} \quad (3)$$

where R_i denotes the radius of particle i. Thus, the hydrophobic potential is described by an attractive $1/r^6$ potential. In the hard-sphere term the summation extends over protein center−polymer segment pairs as well as among polymer segments, whereas the hydrophobic potential term extends over hydrophobic site−polymer segment pairs. Thus, the properties of the polymer will remain the same as the hydrophobic potential changes, simplifying the evaluation of the results. The harmonic bond potential energy, U_{bond}, is described as

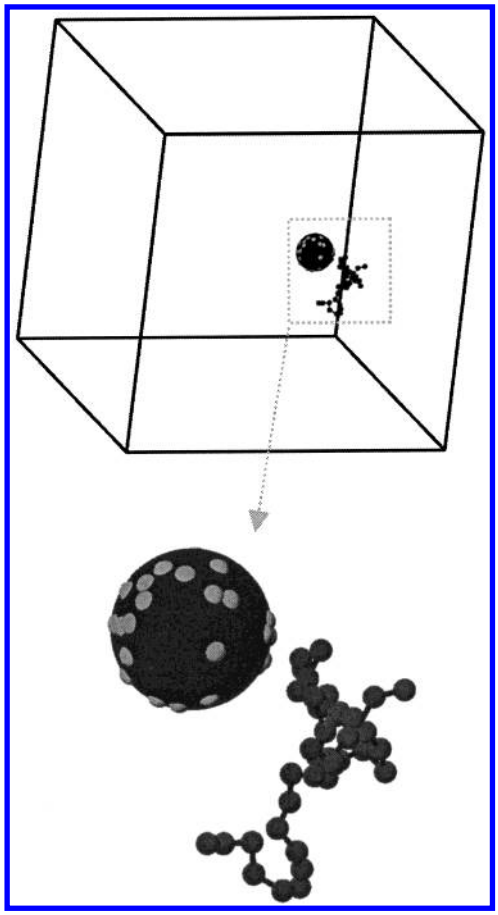

Figure 3. Snapshot of the simulation box containing the protein and the polymer and an enlargement of the two macromolecules. The large sphere with bright spots represents the protein with hydrophobic surface sites and the connected small spheres the polymer.

$$U_{\text{bond}} = \sum_{i=1}^{N_{\text{seg}}-1} \frac{k_{\text{bond}}}{2}(r_{i,i+1} - r_0)^2 \qquad (4)$$

where $k_{\text{bond}} = 2.4$ kJ/(mol Å^2) (0.4 N/m) is the force constant, $r_{i,i+1}$ is the distance between segment i and $i+1$, and $r_0 = 5.0$ Å is the equilibrium separation between adjacent segment. For the connected hard-sphere segments, the root-mean-square (rms) segment−segment separation becomes $\langle R_{\text{seg,seg}}^2\rangle^{1/2} \approx 5.5$ Å.

Systems. In total, 50 different proteins with different distributions of hydrophobic sites divided into five protein classes were considered. The hydrophobic interaction strength ϵ, entering in eq 2, was varied from 0.9×10^5 to 2.4×10^5 kJ Å^6/mol, corresponding to an attractive potential of $2.3kT$−$6.1kT$ at closest contact between a hydrophobic site and a polymer segment.

The protein and the polymer were placed in a cubic box with the box length $L = 250$ Å (see Figure 3), and the system volume was constant throughout. The size of the box was selected such as the polymer should not be able to interfere with itself due to the periodic boundary conditions (the length of an extended polymer is ≈220 Å). The temperature of the systems was $T =$ 298 K. General data of the model are collected in Table 1.

Central aspects in this study are whether (i) a polymer segment and (ii) the polymer are adsorbed to the protein or not. We have applied a geometrical condition and consider a polymer segment to be adsorbed if the distance between it and the protein $|\mathbf{r}_{\text{seg}} - \mathbf{r}_{\text{p}}|$ does not exceed 4 Å from their contact separation, i.e.

$$|\mathbf{r}_{\text{seg}} - \mathbf{r}_{\text{p}}| \le R_{\text{c}} \equiv R_{\text{seg}} + R_{\text{p}} + 4 \text{ Å} \qquad (5)$$

TABLE 1: General Data of the Model

protein radius	$R_{\text{p}} = 16$ Å
number of hydrophobic sites	$N_{\text{hs}} = 30$
number of polymer segments	$N_{\text{seg}} = 40$
segment radius	$R_{\text{seg}} = 2$ Å
box length	$L = 250$ Å
temperature	$T = 298$ K

and the polymer to be adsorbed if at least one of its segments is adsorbed. The results do not qualitatively depend on the precise value of R_{c}.

Simulation Method

The properties of the model system was obtained by performing Metropolis Monte Carlo simulations in the canonical ensemble.[29] The protein and the polymer were randomly positioned in the cubic box, and periodical boundary conditions were applied. The examination of the configurational space was accelerated by employing several different kinds of trial displacements.[29] The protein was subjected to translational and rotational displacements. The polymer was subjected to trials moves involving (i) translation of a single segment, (ii) pivot rotation of a part of the chain, (iii) slithering move, and (iv) translation of the entire chain with the relative probabilities 0.80, 0.10, 0.05, and 0.05, respectively.

The simulations were performed by employing 2×10^6 trial moves per particle. Reported uncertainties of a quantity, x, calculated from the simulation is the standard deviation, $\sigma(\langle x\rangle)$, based on a division of a run into $n_m = 20$ subdivisions. The variance, $\sigma^2(\langle x\rangle)$, was calculated according to

$$\sigma^2(\langle x\rangle) = \frac{1}{n_m(n_m-1)} \sum_{m=1}^{n_m} (\langle x\rangle_m - \langle x\rangle)^2 \qquad (6)$$

where $\langle x\rangle$ is the average of the quantity x from the entire simulation and $\langle x\rangle_m$ the average of subdivision m. The simulation was performed using the program MOLSIM, a Monte Carlo/molecular dynamics/Brownian dynamics simulation package.[30]

Results

The protein−polymer association and polymer conformations at various strengths of the hydrophobic potential and different spatial distribution of the hydrophobic sites have been investigated. The adsorption at different interaction strengths will first be considered. Thereafter, an examination of the variation within a protein class will be given using proteins of class MD2 having the most heterogeneous distribution of the hydrophobic surface sites. The results from the other classes will then be compared to those of MD2 and will be used to determine how the protein−polymer association is related to the distribution of the hydrophobic sites.

Effect of the Interaction Strength ϵ. Some features of the adsorption of the polymer to a protein will be given using protein 3 of class MD2. Two different properties to characterize the adsorption will be employed, and the dependence on the interaction strength will be examined.

Adsorption Probability. The probability of protein−polymer association has been examined in two different ways. We will initially consider (i) $P_{\text{ads}}^{\text{seg}}$ denoting the fraction of polymer segments within 4 Å from the protein surface according to eq 5 and (ii) P_{ads} denoting the probability that at least one segment being adsorbed to the protein surface, again using the same distance criterion. The limits $P_{\text{ads}}^{\text{seg}} = 0$ and 1 correspond to that no and all segments, respectively, are adsorbed, whereas

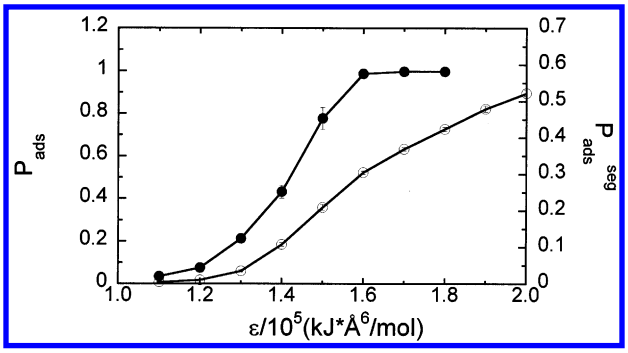

Figure 4. Adsorption probability, P_{ads}^{seg} (open circles) and P_{ads} (filled circles), as a function of the interaction strength, ϵ, for protein 3 of class MD2. Estimated uncertainties are given as error bars.

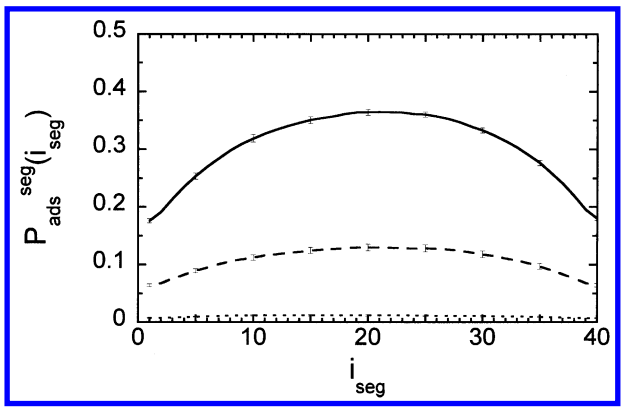

Figure 5. Segment contact probability function, $P_{ads}^{seg}(i_{seg})$, as a function of the segment number, i_{seg}, for protein 3 of class MD2 at interaction strengths $\epsilon = 1.2 \times 10^5$ (dotted curve), 1.4×10^5 (dashed curve), and 1.6×10^5 kJ Å^6/mol (solid curve). Estimated uncertainties are given as error bars at selected points.

$P_{ads} = 0$ and 1 to that the polymer is never and always, respectively, adsorbed to the protein.

Figure 4 shows the adsorption probabilities, P_{ads}^{seg} (open circles) and P_{ads} (filled circles), as a function of the interaction strength, ϵ, of protein 3 of class MD2. At low ϵ, P_{ads}^{seg} and P_{ads} are nearly zero, implying no protein−polymer segment or protein−polymer association, whereas at high ϵ, P_{ads} is one demonstrating the existence of a protein−polymer association. However, P_{ads}^{seg} does not reach one for the ϵ range considered; i.e., not all polymer segments are near the protein surface. At intermediate ϵ, P_{ads} is S-shaped with a fairly sharp transition, showing that only a modest change in ϵ may induce an adsorption of the polymer to the protein, whereas the rise of P_{ads}^{seg} is more gradual.

Since the main interest in this study is to determine whether the polymer is associated or not to the protein, the focus will be on P_{ads} in the following presentation. Nevertheless, the $P_{ads}^{seg}-\epsilon$ relations for proteins belonging to all five protein classes resemble the $P_{ads}-\epsilon$ relations given below.

Contact Probability Functions. It is expected that segments with different positions in the chain display different degrees of adsorption. This has been examined by considering segment contact probability functions, $P_{ads}^{seg}(i_{seg})$, denoting the probability of segment i_{seg} being adsorbed to the protein, and again eq 5 has been employed. Similar to P_{ads}^{seg}, $P_{ads}^{seg}(i_{seg}) = 0$ denotes that segment i_{seg} is not adsorbed at all and $P_{ads}^{seg}(i_{seg}) = 1$ that the segment is always adsorbed.

Figure 5 shows $P_{ads}^{seg}(i_{seg})$ at the interaction strengths $\epsilon = 1.2 \times 10^5$, 1.4×10^5, and 1.6×10^5 kJ Å^6/mol, corresponding to a (nearly) nonadsorbed, slightly adsorbed, and fully adsorbed polymer, respectively (cf. Figure 4), for protein 3 of class MD2. Not surprisingly, $P_{ads}^{seg}(i_{seg}) \approx 0$ at the lowest ϵ, consistent with a nonadsorbed polymer. As ϵ is increased, $P_{ads}^{seg}(i_{seg})$ increases for all i_{seg}, showing an increased probability for all segments to be adsorbed. At the largest ϵ, there is a considerable difference in adsorption between end segments and center segments. At this condition, the central segments are those more likely to be adsorbed.

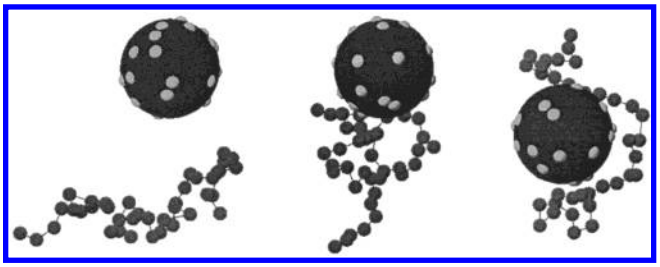

Figure 6. Typical snapshots from simulations involving protein 3 of class MD2 at the interaction strengths $\epsilon = 1.2 \times 10^5$ (left), 1.4×10^5 (center), and 1.6×10^5 kJ Å^6/mol (right).

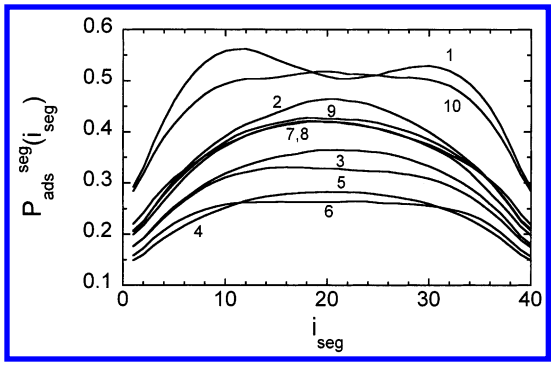

Figure 7. Segment contact probability function, $P_{ads}^{seg}(i_{seg})$, as a function of the segment number, i_{seg}, for all 10 proteins of class MD2 (protein numbers indicated) at the interaction strength $\epsilon = 1.6 \times 10^5$ kJ Å^6/mol.

Moreover, the contact probability functions are symmetric with respect to reflection at the central polymer segment, $i_{seg} = (N_{seg} + 1)/2$, as anticipated by symmetry arguments. This indicates sufficient mobility of the polymer in the configurational space and a satisfactory sampling of the adsorbed state.

Three snapshots taken from the simulations for conditions corresponding to the three contact probability functions in Figure 5 are shown in Figure 6. In the case of the lowest ϵ (left), the two macromolecules are separated. At intermediate ϵ, there are typically only a few protein−polymer segment contacts, and the polymer is hence only weakly attached to the protein. At the largest ϵ, a large number of polymer segments are in contact with the protein, and the polymer displays a large contact surface toward the protein. Moreover, ca. 8 segments at both polymer ends are detached from the surface, consistent with their lower segment adsorption probabilities given in Figure 5.

Variation within the Protein Class MD2. The adsorption of the polymer to 10 different proteins belonging to class MD2 will now be considered. As mentioned above, the proteins are generated using the same minimum distance condition, but using different sequence of random numbers.

Contact Probability Functions. Figure 7 displays contact probability functions, $P_{ads}^{seg}(i_{seg})$, for the 10 proteins of class MD2. The interaction strength $\epsilon = 1.6 \times 10^5$ kJ Å^6/mol corresponds to a rather strong protein−polymer interaction. As before, the probability functions are symmetric with respect to reflection of the central polymer segment. Noticeable is,

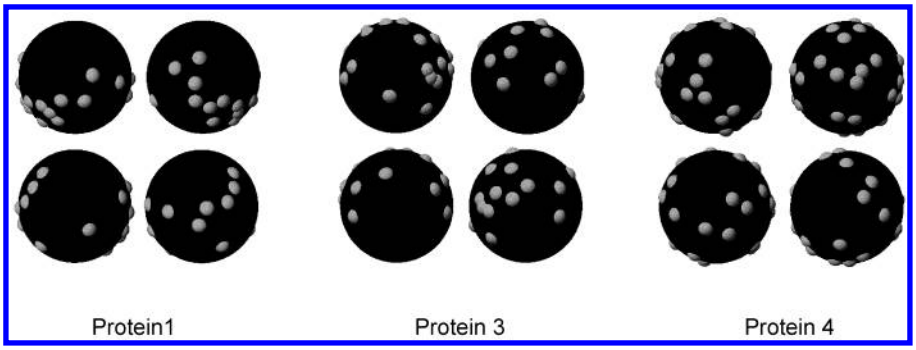

Figure 8. Pictures of proteins 1, 3, and 4 of class MD2. Each protein is rotated 90° around the vertical axis between each of the four images (top left, top right, bottom left, and bottom right). The hydrophobic sites are represented by bright spots.

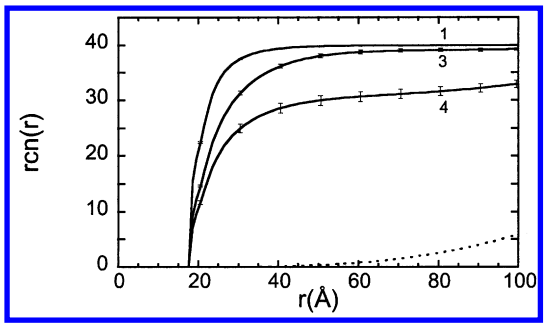

Figure 9. Running coordination number, $rcn(r)$, denoting the number of polymer segments within a distance r from the protein center for proteins 1, 3, and 4 (protein number indicated) of class MD2 at the interaction strength $\epsilon = 1.6 \times 10^5$ kJ Å^6/mol (solid curves). Estimated uncertainties are given as error bars at selected points. The corresponding $rcn(r)$ for a random distribution of the 40 polymer segments in the simulation box [$rcn(r) = 1.074 \times 10^{-5}$ $(r/\text{Å} - 18)^3$] is also shown (dotted curve).

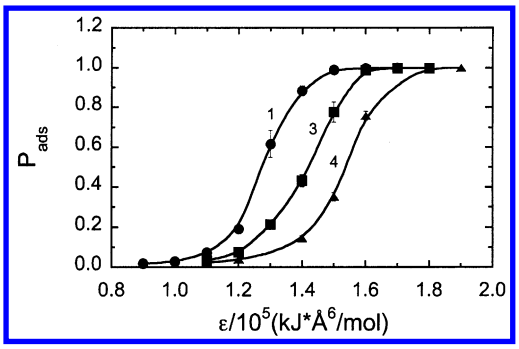

Figure 10. Adsorption probability, P_{ads}, as a function of the interaction strength, ϵ, for protein 1 (circles), protein 3 (squares), and protein 4 (triangles) of class MD2. Estimated uncertainties are given as error bars.

however, the considerable variation of the contact probability functions between the polymer segments and the 10 different proteins. A fairly small probability for segment adsorption appears for proteins 4 and 6, whereas it is much larger for proteins 1 and 10. Moreover, for protein 1, $P_{ads}^{seg}(i_{seg})$ possesses a minimum for central segments, indicating the formation of a loop conformation, while the adsorption to the other nine proteins is strongest for the central segments.[31] The corresponding probabilities of finding polymer segment i_{seg} within 9 Å from the protein surface were also calculated (data not shown), qualitatively giving the same picture.

In the following, the adsorption to proteins 1, 3, and 4 will be examined in more detail. These proteins represent one protein with a higher adsorption probability, one with an intermediate probability, and one with a lower adsorption probability (cf. Figure 7). Figure 8 shows four snapshots of each of these three proteins. It appears that the distribution of the hydrophobic sites is more heterogeneous (clustered) in protein 1 and less heterogeneous in protein 4. Later, this degree of clustering will be quantified and correlated to the adsorption probabilities.

Running Coordination Number. The running coordination number $rcn(r)$ denotes the number of polymer segments within the distance r from the protein center. Thus, $rcn(r)$ provides the accumulated amount of segments in the nearest neighborhood of the protein.

Figure 9 shows $rcn(r)$ for proteins 1, 3, and 4 of class MD2 at $\epsilon = 1.6 \times 10^5$ kJ Å^6/mol. We observe a strong increase of the running coordination number near contact separation and a leveling off at larger separations, supporting the notion that most of the polymer segments are close to the protein. Protein 1 has the strongest increase at short separation and reaches the largest plateau value, whereas protein 4 has the weakest increase and the lowest plateau value. Obviously, the order among the proteins is the same as in Figure 7. Moreover, Figure 9 shows that practically all $N_{seg} = 40$ segments are within 30 Å from the surface of protein 1.

Adsorption Probabilities. Figure 10 shows the adsorption, P_{ads}, as a function of the interaction strength, ϵ, for proteins 1, 3, and 4. It is clear that the onset of the polymer adsorption to a protein appears at different ϵ for the polymers belonging to the same protein class. The value of ϵ at $P_{ads} = 0.5$ is 1.27×10^5, 1.42×10^5, and 1.54×10^5 kJ Å^6/mol for proteins 1, 3, and 4, respectively. Thus, a weaker attraction is sufficient to adsorb the polymer to protein 1, whereas a larger attraction is necessary to adsorb the polymer to protein 4. Furthermore, at $\epsilon = 1.4 \times 10^5$ kJ Å^6/mol, we have almost a complete adsorption to protein 1 and a very weak adsorption to protein 4; thus, there appear conditions at which the distribution of the hydrophobic sites critically determine the degree of protein−polymer complexation. The qualitative features of P_{ads} and P_{ads}^{seg} are equivalent, giving an earlier onset of polymer adsorption to protein 1, followed by protein 3 and finally protein 4.

Hence, it has been established that the polymer adsorption to the protein does depend on the small differences within a statistical distribution of the hydrophobic sites.

Variation among Protein Classes. The adsorption properties for proteins belonging to different classes will now be considered. The natural question is, how does the minimum separation between the hydrophobic sites influence the adsorption pattern?

Polymer Adsorption at Fixed ϵ. Figure 11 shows the adsorption probability, P_{ads}, at the interaction strength $\epsilon = 1.6 \times 10^5$ kJ Å^6/mol for all the 50 proteins considered. Data for proteins belonging to the same class have the same horizontal position (10 proteins in each class). First, disregarding the variation within the classes, the polymer adsorption probability P_{ads} is reduced as the minimum separation between hydrophobic sites

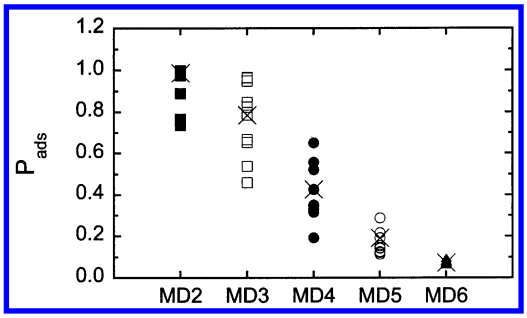

Figure 11. Adsorption probability, P_{ads}, as a function of protein class (MD2, MD3, MD4, MD5, and MD6) at the interaction strength $\epsilon = 1.6 \times 10^5$ kJ Å^6/mol. For each class, 10 points corresponding to each protein in that class are given. The crosses denote P_{ads} for one representative protein of each class further considered in Figures 12 and 14.

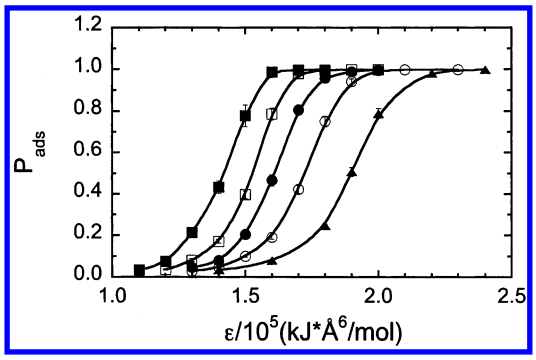

Figure 12. Adsorption probability, P_{ads}, as a function of the interaction strength, ϵ, for one representative protein from each class: MD2 (filled squares), MD3 (open squares), MD4 (filled circles), MD5 (open circles), and MD6 (filled triangles). Estimated uncertainties are given as error bars.

is increased from 2 to 6 Å. Second, the variation of P_{ads} within a class is largest for those classes displaying an intermediate P_{ads}. It is conceivable that at conditions where P_{ads} is either close to zero or one, the exact distribution of the hydrophobic sites should only marginally influence P_{ads}. An additional factor is, as the site distribution becomes more homogeneous, proteins within the same class become more alike, and hence the distribution of P_{ads} becomes smaller.

Adsorption Probabilities. Figure 12 displays adsorption probabilities for one representative protein from each class as a function of the interaction strength ϵ. The proteins selected display intermediate adsorption probabilities within its class (see crosses in Figure 11). All adsorption probabilities are S-shaped, displaying a rather abrupt adsorption behavior by small change in ϵ. The interaction strength at which $P_{ads} = 0.5$ are 1.42×10^5, 1.53×10^5, 1.61×10^5, 1.72×10^5, and 1.90×10^5 kJ Å^6/mol for the members of classes MD2, MD3, MD4, MD5 and MD6, respectively. Thus, it is evident that a larger ϵ is needed in order to adsorb the polymer as the protein type goes from a more heterogeneous distribution of the hydrophobic residues toward a more homogeneous one (left to right in Figure 2).

Cluster Analysis. From the results in previous subsections, it is obvious that the adsorption ability of the proteins is related to the spatial arrangement of the hydrophobic sites. A more heterogeneous site distribution gave rise to an improved adsorption as compared to a more homogeneous distribution (cf. Figures 2 and 12). Moreover, there was a different adsorption pattern among proteins belonging to the same class (Figure 10), and from the visual inspection of Figure 8 we conjecture that the difference in the adsorption behavior within

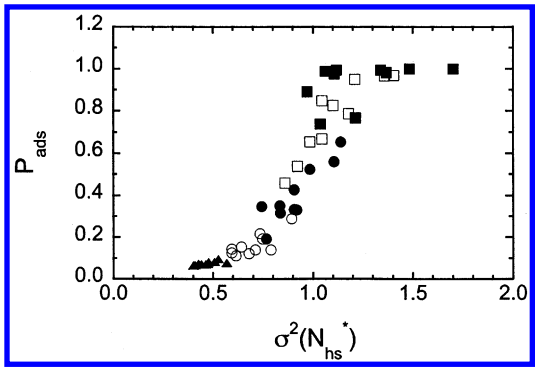

Figure 13. Adsorption probability, P_{ads}, as a function of the variance $\sigma^2(N^*_{hs})$ quantifying the degree of clustering (see text for further details) at $\epsilon = 1.6 \times 10^5$ kJ Å^6/mol for proteins of class MD2 (filled squares), MD3 (open squares), MD4 (filled circles), MD5 (open circles), and MD6 (filled triangles). All 50 proteins are considered.

a class could also be related to the degree of clustering of the hydrophobic sites.

Therefore, we will now relate the degree of clustering with the adsorbing capacity across *both* the variation within a class and between classes. In order to do so, we will consider the number of hydrophobic surface sites within a circle of radius of 6 Å placed on the protein surface, N^*_{hs}. The probability distribution $P(N^*_{hs})$ was formed by calculating N^*_{hs} for a large number of circles with random positions. Finally, the variance of the distribution $P(N^*_{hs})$, $\sigma^2(N^*_{hs})$, was calculated and taken as the measure of the degree of clustering of the hydrophobic sites. For a heterogeneous site distribution, the probability distribution $P(N^*_{hs})$ is broad and hence $\sigma^2(N^*_{hs})$ is large, whereas for a more homogeneous site distribution, $P(N^*_{hs})$ is narrower and $\sigma^2(N^*_{hs})$ becomes smaller.

Figure 13 displays P_{ads} as a function of $\sigma^2(N^*_{hs})$ for all the 50 proteins at $\epsilon = 1.6 \times 10^5$ kJ Å^6/mol. Results for proteins belonging to the same class are given by same symbols. First, it is observed that all data fall on a single master curve, although some scattering occurs. Data for proteins belonging to different classes overlap each other considerably, showing that there is a statistical variation within a class of the same order as the variation between two neighboring classes. Thus, $\sigma^2(N^*_{hs})$ appears to be a suitable variable able to *simultaneously* describe the variation of the adsorption within a protein class as well as across protein classes. Second, Figure 13 unambiguously shows that the probability of adsorption increases as the distribution of the hydrophobic sites becomes more heterogeneous.

It should also be noted that the master curve given in Figure 13 would be shifted to lower P_{ads} for smaller ϵ and to larger P_{ads} for higher ϵ, in both cases with a concomitant lower sensitivity. Finally, the radius of the circle probing the clustering has to be adapted to the average surface area per site for optimal resolution.

Chain Properties. The spatial extension and the shape of the polymer have been quantified by calculating the rms end-to-end distance, $\langle R_{ee}^2 \rangle^{1/2}$, and the shape ratio, $\langle R_{ee}^2 \rangle / \langle R_g^2 \rangle$, where $\langle R_g^2 \rangle$ denotes the mean-square radius of gyration. A large interaction energy ($\epsilon = 2.2 \times 10^5$ kJ Å^6/mol), at which proteins of all classes display complete polymer adsorption (see Figure 12), was used, and hence the properties of *adsorbed* polymers are examined. In Figure 14, these properties are shown for polymers adsorbed to one representative protein from each class (see crosses in Figure 11), and the corresponding properties for an isolated chain are given in Figure 14 (crosses). Figure 14a shows the polymer rms end-to-end distance, $\langle R_{ee}^2 \rangle^{1/2}$. First, it is noted that the polymer becomes considerably contracted upon

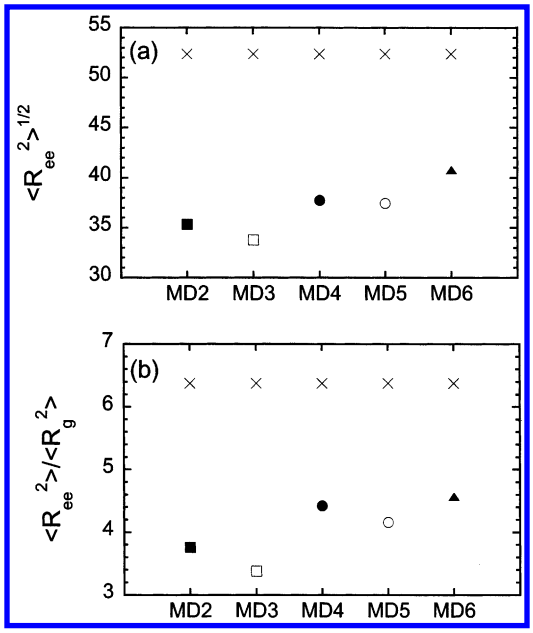

Figure 14. (a) Root-mean-square end-to-end distance $\langle R_{ee}^2\rangle^{1/2}$ and (b) shape ratio $\langle R_{ee}^2\rangle/\langle R_g^2\rangle$ with $\langle R_g^2\rangle$ denoting the mean-square radius of gyration for one representative protein from each class at the interaction strength $\epsilon = 2.2 \times 10^5$ kJ Å^6/mol. Properties of an unperturbed chain are given by the crosses.

adsorption to the protein. Second, the contraction is strongest for class MD2; thus, the polymer becomes stronger contracted when binding to a protein surface with a heterogeneous distribution of the hydrophobic sites. The corresponding shape ratio data are given in Figure 14b. It constitutes a simple measure of the shape of the polymer and becomes 12 for a rigid rod, 6 for a Gaussian chain, and 2 for a homogeneous sphere with the two polymer ends randomly positioned. For the present nonadsorbed chain, $\langle R_{ee}^2\rangle/\langle R_g^2\rangle$ becomes slightly larger than 6 due to hard-sphere repulsions. When being adsorbed, the shape ratio $\langle R_{ee}^2\rangle/\langle R_g^2\rangle$ reduces to ca. 3.5−4.5; hence, an adsorbed chain becomes less elongated than an nonadsorbed chain, and it is also much less nonspherical than a Gaussian chain. Moreover, the shape ratio is smallest for polymers adsorbed to proteins with a more heterogeneous distribution of the hydrophobic sites.

Discussion

Protein−polymer association is dependent on several factors such as size and shape of the protein, flexibility, length, and bulkiness of the polymer, and the nature of their intermolecular interaction. In experimental investigations of protein−polymer association, the exchange of one macromolecule with another one or a change of environmental conditions, such as pH, salt concentration, or solvent, often leads to simultaneous change of several system parameters, making the results difficult to interpret. In this respect, the use of model systems offers an advantage of disentangling these different effects from each other by performing systematic changes of individual parameters. In this work, a coarse-grain model has been used to examine the role of the distribution of the hydrophobic sites on the adsorption of polymers to the protein at various interaction strengths, keeping all other properties the same.

Effect of Interaction Strength. The adsorption of the polymer chain to the protein surface occurred rather abruptly as the potential strength was increased (Figures 4, 10, and 12), which is a general feature of polymer adsorption when short-range interactions are involved.[32] Such an adsorption is promoted by an increased number of favorable protein−polymer contacts and is hence enthalpically driven. However, the accumulation of polymer segments near the surface restricts the number of possible chain conformations constituting an entropic penalty. At low interaction strength, the entropic penalty prohibits the adsorption, but as the interaction strength increases, the entropic cost is overcome by the energetic advantage and the polymer adsorbs to the surface.

Effect of Site Distribution. The distribution of the hydrophobic surface sites affected the adsorption of the fully flexible and slightly hydrophobic polymer chain to the protein surface (Figures 7−12). Small differences in the distribution within a protein class as well as the systematic differences between protein classes affected the adsorption ability. It was shown that the adsorption was favored by a more heterogeneous distribution as compared to proteins with a more homogeneous distribution (Figure 13). The appearance of regions with several nearby hydrophobic sites or "hot spots" becomes more frequent in the protein classes characterized by a more heterogeneous site distribution. Experimentally, both hydrophobicity and location of hydrophobic residues have been suggested to be of importance in ligand/polymer binding.[33−35] It is often considered that so-called "hot spots" of hydrophobic residues form patterns matching protein and ligand, giving an efficient binding. This is a well-adopted idea within the closely related area of antibody−antigen binding.

In our analysis, we found that the largest sensitivity in distinguishing different proteins appeared for a rather narrow range of ϵ close to $P_{ads} \approx 0.5$. This was manifested in Figure 13, where proteins belonging to the intermediate classes showed an enhanced polymer adsorption as $\sigma^2(N^*_{hs})$ is increased. The same types of correlation would also appear for proteins of class MD2 or MD6 if the interaction strength, ϵ, had been selected differently. Similar results of a large variance of the adsorbed fraction of polymer close to the adsorption transition was found by Bratko et al.[15] In their study, the adsorption of heteropolymers with different statistical patterns to a surface with spatially distributed interactions sites was investigated as the number of interaction sites was increased.

Why Is a Heterogeneous Site Distribution More Effective? We will here discuss why the polymer adsorption to proteins should appear first for a heterogeneous site distribution as the interaction strength is increased in the present investigation. The procedure adopted is to argue that adsorption should be facilitated, as a homogeneous site distribution is made more heterogeneous. The argumentation is based on the simplification that a hydrophobic site can "bind" at most one single polymer segment.

Assume that polymer segments are bound to hydrophobic sites only. If the hydrophobic sites would be regularly distributed on the protein surface, the separation between two bound segments being neighbors would be $d_{ads} \approx [4\pi(R_p + R_{seg})^2/N_{hs}]^{1/2}$. At the present conditions, $d_{ads} \approx 11-12$ Å, and hence the separation between two bound segments would be twice the rms segment−segment separation ($\langle R_{seg,seg}^2\rangle^{1/2} \approx 5.5$ Å) between connected segments. Thus, such a polymer would typically have every second segment bound. This would not be particularly optimal for two reasons: (i) energetically only 20 of the 30 hydrophobic sites would be used, and (ii) entropically the chain would be forced to have a two-dimensional conformation following the protein surface or possibly forming many short loops involving very few segments each. A departure from the homogeneous site distribution would facilitate binding of the polymer to the protein surface by (i) making it possible for

more segments to be bound and (ii) allowing parts of the chain, preferentially tails, to be detached from the surface.

Increasing the heterogeneity of the site distribution also creates proteins where each segment can "bind" more than one hydrophobic site at a time. For such proteins the polymer chain would only have to involve few segments in order to achieve the same amount of contacts as for the homogeneous distribution where only the one-to-one binding stoichiometry can be applied, hence leaving larger parts of its chain protruding out into the solution as tails or loops.

The argumentation has its strongest validity for a one-to-one site-segment binding stoichiometry and in the limit of contact interactions. Nevertheless, it appears to provide a rational explanation for the observed behavior. For longer polymers with $N_{seg} > (d_{ads}/\langle R_{seg,seg}^2\rangle^{1/2})N_{hs}$, the situation could possibly be different. In this case, the polymer would be able to saturate all hydrophobic sites, even in the limit of a homogeneous site distribution.

In the study by Chakraborty and Bratko, the conformational entropies of heteropolymers adsorbed to planar surfaces have been evaluated.[16] Especially, the loop entropy for chains adsorbed to surfaces with increasing interaction site loading was analyzed, concluding that quenched loops are responsible for a first-order adsorption transition.

Measure of Site Clustering. The spatial distribution of the hydrophobic sites was quantified by $\sigma^2(N^*_{hs})$, denoting the variation of the distribution of the number of hydrophobic sites within a given area of the protein surface. In principle, the precise locations of all hydrophobic surface sites affect the adsorption of the polymer to the protein. Nevertheless, the projection of all the locations on this single quantity appeared to capture the variation of the adsorption ability for the proteins with different site distributions. Although the correlation between $\sigma^2(N^*_{hs})$ and the adsorption probability was not complete (Figure 13), we believe that this simple measure should be a useful initial approach to characterize the degree of site clustering.

Effect of the Interaction Range. The effect of the range of the protein−polymer interaction was not investigated. A more long-range protein−polymer interaction is expected to induce a more gradual binding behavior than observed here and hence less variation in the adsorption behavior between the different site distributions. Long-range electrostatics are, however, generally stronger than hydrophobic interactions, and thus the adsorption to proteins could still be modulated by the charge distribution of the proteins. Nevertheless, the general picture is that nonoriented long-range interactions are responsible for bringing two species together, whereas the specific binding is caused by short-range interactions.[36] Some recent studies have suggested the charge distribution on proteins to account for the complexation behavior to polyelectrolytes.[4,10] Furthermore, Takahashi et al. have shown that the surface charge distribution does affect the complexation behavior between positively charged proteins and a negative polyelectrolyte.[9]

Conclusions

Surface properties of proteins are crucial for recognition and interaction with other macromolecules such as polymers. In this study we have focused on how the spatial arrangement of hydrophobic surface sites affects the interaction with a slightly hydrophobic polymer. The degree of heterogeneity of the site distribution was measured using the variance of the number of sites located in randomly positioned circles on the protein surface. The polymer adsorption was promoted for proteins with more heterogeneous (clustered) site distributions as opposed to more homogeneous distributions of the hydrophobic sites. A strong correlation was found between the adsorption ability and the degree of heterogeneity both across the protein classes having different minimum site distances and within a class with only statistical differences. Upon adsorption the polymer is strongly contracted, the effect being largest for a more heterogeneous site distribution.

References and Notes

(1) Albertsson, P.-Å. *Partitioning of Cell Particles and Macromolecules*, 3rd ed.; Wiley: New York, 1986.
(2) Dubin, P.; Bock, J.; Davies, R. M.; Schultz, D. N.; Thies, C. *Macromolecular Complexes in Chemistry and Biology*; Springer-Verlag: New York, 1994.
(3) Galaev, I. Y.; Mattiasson, B. *Smart Polymers for Bioseparation and Bioprocessing*; Taylor & Francis: New York, 2002; p 55.
(4) Tsuboi, A.; Izumi, T.; Hirata, M.; Xia, J.; Dubin, P. L.; Kokufuta, E. *Langmuir* **1996**, *12*, 6295.
(5) Furness, E. L.; Ross, A.; Davis, T. P.; King, G. C. *Biomaterials* **1998**, *19*, 1361.
(6) Sato, T.; Kamachi, M.; Mizusaki, M.; Yoda, K.; Morishima, Y. *Macromolecules* **1998**, *31*, 6871.
(7) Tribert, C.; Porcar, I.; Bonnefont, P. A.; Audebert, R. *J. Phys. Chem. B* **1998**, *102*, 1327.
(8) Porcar, I.; Cottet, H.; Gareil, P.; Tribet, C. *Macromolecules* **1999**, *32*, 3922.
(9) Takahashi, D.; Kubota, Y.; Kokai, K.; Izumi, T.; Hirata, M.; Kokufuta, E. *Langmuir* **2000**, *16*, 3133.
(10) Hattori, T.; Hallberg, R.; Dubin, P. L. *Langmuir* **2000**, *16*, 9738.
(11) Berggren, K.; Egmond, M. R.; Tjerneld, F. *Biochim. Biophys. Acta* **2000**, *1481*, 317.
(12) Carlsson, F.; Linse, P.; Malmsten, M. *J. Phys. Chem. B* **2001**, *105*, 9040.
(13) Berggren, K.; Wolf, A.; Asenjo, J. A.; Andrews, B. A.; Tjerneld, F. *Biochim. Biophys. Acta* **2002**, *1596*, 253.
(14) Bratko, D.; Chakraborty, A. K.; Shakhnovich, E. I. *Chem. Phys. Lett.* **1997**, *280*, 46.
(15) Bratko, D.; Chakraborty, A. K.; Shakhnovich, E. I. *Comput. Theor. Polym. Sci.* **1998**, *8*, 113.
(16) Chakraborty, A. K.; Bratko, D. *J. Chem. Phys.* **1998**, *108*, 1676.
(17) Golumbfskie, A. J.; Pande, V. S.; Chakraborty, A. K. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11707.
(18) Genzer, J. *J. Chem. Phys.* **2001**, *115*, 4873.
(19) Smith, D. E.; Haymet, A. D. J. *J. Chem. Phys.* **1993**, *98*, 6445.
(20) Hummer, G.; Garde, S.; García, A. E.; Pohorille, A.; Pratt, L. R. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8951.
(21) Chipot, C.; Jaffe, R.; Maigret, B.; Pearlman, D. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1996**, *118*, 11217.
(22) Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; Tirado-Rives, J. *J. Chem. Phys.* **1988**, *89*, 3742.
(23) Jorgensen, W. L.; Severance, D. L. *J. Am. Chem. Soc.* **1990**, *112*, 4768.
(24) Linse, P. *J. Am. Chem. Soc.* **1992**, *114*, 4366.
(25) Linse, P. *J. Am. Chem. Soc.* **1993**, *115*, 8793.
(26) Pangali, C.; Rao, M.; Berne, B. J. *J. Chem. Phys.* **1979**, *71*, 2975.
(27) Pratt, L. R.; Chandler, D. *J. Chem. Phys.* **1980**, *73*, 3430.
(28) Rossky, P. J.; Friedman, H. L. *J. Phys. Chem.* **1980**, *84*, 587.
(29) Allen, M. P.; Tildesley, D. J. *Computer Simulations of Liquids*; Oxford University Press: New York, 1987.
(30) Linse, P. MOLSIM; 3.21 ed. Lund, 2001.
(31) Akinchina, A.; Linse, P. *Macromolecules* **2002**, *35*, 5183.
(32) Fleer, G. J.; Cohen Stuart, M. A.; Scheutjens, J. M. H. M.; Cosgrove, T.; Vincent, B. *Polymers at Interfaces*; Chapman & Hall: London, 1993.
(33) Xie, H.; Bolam, D. N.; Nagy, T.; Szabo, L.; Cooper, A.; Simpson, P. J.; Lakey, J. H.; Williamson, M. P.; Gilbert, H. J. *Biochemistry* **2001**, *40*, 5700.
(34) Kim, J. L.; Nikolov, D. B.; Burley, S. K. *Nature (London)* **1993**, *365*, 520.
(35) Sundberg, E. J.; Urrutia, M.; Braden, B. C.; Isern, J.; Tsuchiya, D.; Fields, B. A.; Malchiodi, E. L.; Tormo, J.; Schwarz, F. P.; Mariuzza, R. A. *Biochemistry* **2000**, *39*, 15375.
(36) Israelachvili, J. N. *Intermolecular & Surface Forces*, 2nd ed.; Academic Press: London, 1991.