Streptavidin Tetramerization and 2D Crystallization: A Mean-Field Approach

T. Coussaert,* A. R. Völkel,† J. Noolandi,† and A. P. Gast*

*Department of Chemical Engineering, Stanford University, Stanford, California 94305-5025 USA, and †Xerox Research Center of Canada, Mississauga, Ontario L5K 2L1, Canada

ABSTRACT A mean-field theoretical approach is applied to streptavidin tetramerization and two-dimensional (2D) crystallization. This theory includes, in particular, solvent–residue interactions following the inhomogeneous Flory–Huggins model for polymers. It also takes into account residue–residue interactions by using tabulated pair interaction parameters. This theory allows one to explicitly calculate the entropy of the inhomogeneous system. We show that hydrophobic interactions are responsible for the stability of tetramerization. Within the present theory, the equilibrium distance between the two dimers is the same as that determined experimentally. The free energy of tetramerization (i.e., dissociation of the two dimers) is 50 k_BT. Unlike tetramerization, hydrophobic interactions alone are not sufficient to stabilize the 2D crystal C_{222}, but solvent-mediated residue–residue interactions give the most important contribution.

INTRODUCTION

The main goal of this paper is to establish a theoretical tool to study the effect of intermolecular interactions on the mechanism of two-dimensional (2D) protein crystallization. Protein tetramerization is concurrently studied to emphasize the difference between both processes. A better molecular understanding of the interactions underlying these mechanisms could lead to the development of improved technologies for crystallizing proteins or designing ordered protein layers for biomaterial applications (Durbin and Feher, 1996).

In the present work, we focus on the interaction between streptavidin molecules. Streptavidin forms 2D crystals on a biotinylated lipid monolayer, which are exactly one molecule thick. Three different crystal structures have been experimentally observed, including the C_{222} lattice analyzed in the present paper (Wang et al., 1999a). We are interested in understanding the relative stability of these crystals in relation with the different crystal contacts. This relative stability depends in particular on the pH of the solution and the length of the polypeptidic chains (Wang et al., 2000). Specific intermolecular interactions have also been manipulated via mutagenesis to investigate the effect of these interactions on the phase behavior (Wang et al., 1999b). All these studies show that streptavidin crystallization is mainly energetically driven (and not entropically). Indeed lattice parameters seem not to depend on the concentration of proteins in the bulk phase.

Received for publication 5 October 2000 and in final form 8 January 2001.

Address reprint requests to Tamara Coussaert, Unite de Physique des Polymeres, CP 223, Campus Plaine, Boulevard du Triomphe, Université Libre de Bruxelles, 1050 Bruxelles, Belgium. Tel.: 32-2-650-57-41; Fax: 32-2-650-56-75; E-mail: tcouss@ulb.ac.be.

Dr. Coussaert's present address is CP 223, Université Libre de Bruxelles, Campus Plaine, 1050 Brussels, Belgium.

0006-3495/01/04/2004/07 $2.00

Protein–protein contacts in crystals are complex and involve a delicate balance of interactions depending on solution conditions (Durbin and Feher, 1996). Recent studies have successfully related osmotic second virial coefficient measurements to favorable protein crystallization conditions (Rosenbaum and Zukoski, 1996; Rosenberger et al., 1996). A model of adhesive hard spheres has also been used to describe protein interparticle interactions near crystallization conditions (Rosenbaum and Zukoski, 1996; Piazza et al., 1998).

However, these models are not able to probe the effect of a local change in the polypeptide chain (by mutagenesis for example) on the crystallization. We therefore propose the use of a mean-field approach in which proteins are coarse-grained by introducing "effective atoms" which carry the main feature of the part they represent (either backbone or side chain). Effective interactions are introduced between residues (or effective atoms) and between residues and the solvent. A challenge with such a theory is the choice of effective interactions. However, such a model presents the possibility of studying, for example, the effect of mutagenesis or change of the polypeptide length on crystallization and especially on the relative stability of crystals.

In the next section, we explain the theoretical model, and we introduce the considered interactions. The application to streptavidin tetramerization is discussed in the following section. The streptavidin 2D crystallization case is presented in the section by that name. Our conclusions are gathered in the last section, and the details of the mean-field approach are given in the Appendices.

THEORETICAL MODEL

Crystallization and tetramerization are believed not to change the internal structure of the protein except maybe in the interaction domains. We therefore take, as a starting point, the experimentally determined three dimensional (3D) structure of streptavidin found in the Brookhaven

Protein Database. The position of the protein backbone is considered to be fixed to its experimental position, and to the side chains that are not directly in the interaction domains. In contact domains, on a particular protein (or dimer for tetramerization), each side chain interacting with other proteins (or dimers, respectively) will be allowed to move in the mean field created by all the other residues and the solvent (moving the other side chains will give a constant contribution to the total free energy in the present studies. See the next two sections).

Each amino acid residue is described by two effective atoms (i.e., groups of atoms), one sitting at the C_α position, representing its backbone part, and the other sitting at the side chain center of mass (cms) position, carrying the side chain properties such as hydrophobicity. To evaluate the free energy of the protein–solvent system, we use the mean-field theory established in Völkel and Noolandi (1996, 2000). This theory allows us to explicitly calculate the entropy of the inhomogeneous system (Appendix A).

The solvent–effective atom interactions are evaluated following the inhomogeneous Flory–Huggins model for polymers by defining contact interaction parameters that depend on the relative hydrophobicity of the side chains (Appendix B). Concerning the interactions between each two of the effective atoms, we use tabulated pair interaction parameters deduced from a statistical study of the number of residue–residue contacts in experimentally determined protein structures (Kolinski et al., 1993, Skolnick et al., 1997) (Appendix C).

STREPTAVIDIN TETRAMERIZATION

The present theory has already been applied to *p*53, where the effect of site-directed mutagenesis on the stability of its tetramerization domain has been investigated (Noolandi et al., 2000). In this Section, we study streptavidin tetramerization. Streptavidin can be seen as a dimer of dimers. We use the known streptavidin structure determined by Katz (1995) for residues 13–133 (PDB, 1sle). The existence of a minimum in the reduced total free energy βF with respect to the separation distance d between the two dimers (illustrated in Fig. 1) will tell us about the stability of the tetramer.

Residues interacting across the dimer–dimer interface include Ser-122, Thr-123, Leu-124, Val-125, Gly-126, His-127, Asp-128 and Thr-129. Corresponding side chains are allowed to move in the mean-field created by other residues and the solvent. All other side chains and the backbone are kept fixed at their experimentally determined position. Actually, explicitly including these side chains into the mean-field calculations adds a constant contribution to the total free energy, independent of the separation distance between the dimers.

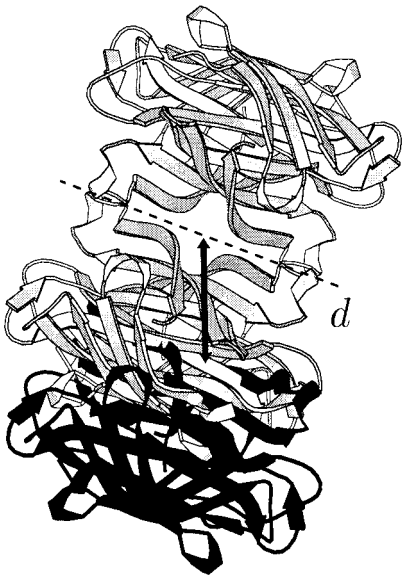

FIGURE 1 The two dimers forming streptavidin are represented in grey (equilibrium structure of streptavidin from PDB file). We translate one of them (*black*) by a separation distance d from the centrum of the protein (*dashed line*), i.e., from the other dimer. This figure was obtained with MolScript (Kraulis, 1991).

To limit the motion of the moving side chain n within a physically relevant area, we impose the constraint

$$\int \mathrm{d}\mathbf{x}\rho_{\mathrm{n}}(\mathbf{x})\mathbf{x} = \mathbf{x}_{\mathrm{n}}^{0}, \tag{1}$$

where $\mathbf{x}_{\mathrm{n}}^{0}$ is the cms position found in the PDB file. Alternatively, we can bias the cms density toward the known rotameric states of the side chain by including an appropriate potential (Noolandi et al., 2000). The constraint Eq. 1 is equivalent to forcing the side chain to remain in the same rotameric state as the one found in the PDB file.

Figure 2 shows that a minimum in the total free energy with respect to the separation distance exists. Clearly solvent–residue interactions are responsible for the existence of this minimum: this interaction gives the only increasing contribution to the free energy with increasing d. That means that the streptavidin tetramer is stable because of hydrophobic interactions between the two dimers. This is not surprising, knowing that the majority of hydrophobic residues are buried inside the protein. Moreover, within the tetramer, the estimated equilibrium separation distance between the two dimers is the one found in the experimentally determined structure in the PDB file. The dissociation free energy of the two dimers is evaluated as 50 $k_{\mathrm{B}}T$. This is in good agreement with the theoretical result from Sano et al. (1997).

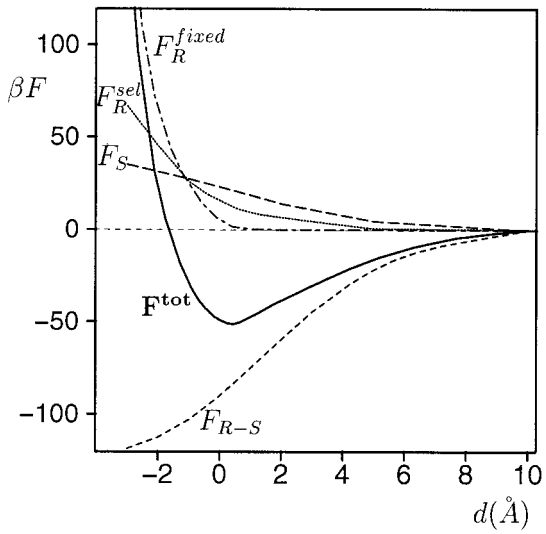

FIGURE 2 The reduced free energy βF as a function of the separation distance d between the two dimers. The lines on the figure represent the total free energy (F^{tot}, *solid line*), the free energy due to fixed effective atoms (F_R^{fixed}, *dot-dashed line*), the free energy due to the selected (moving) effective atoms (F_R^{sel}, *dotted line*), the entropic part of the free energy due to the solvent (F_S, *long dashed line*) and the solvent-residue free energy (F_{R-S}, *dashed line*).

STREPTAVIDIN 2D CRYSTALLIZATION

Experimental studies (Wang et al., 1999a,b, 2000) have shown that the 2D crystal structure C_{222} is observed at neutral pH (cf. Fig. 3). Crystallization seems to be driven mainly by energetic constraints and not entropic ones. Actually, the lattice parameters deduced experimentally for C_{222} always have the same value, independent of the density of the system. In this section, the present theory, where energetic terms dominate (cf. Eq. A13), is applied to C_{222} crystallization. For each pair of nearest neighbors, the contacts or interactions are equivalent in this 2D crystal (cf. Fig. 3). Interactions within the crystal are therefore characterized by the free energy of dissociation of a pair of proteins and the equilibrium separation distance d between these two proteins (cf. Fig. 4).

The selected residues are Glu-14, Ala-15, Gly-16, Thr-18, Gly-19, Thr-20, Tyr-22, Thr-32, Gly-98, Gly-99, Ala-100, Glu-101, Ala-102, Lys-132 (cf. Fig. 4). They all interact with each other, and their cms are allowed to move in the mean fields defined within the present theory. We still impose the constraint Eq. 1. Unlike the tetramerization described above, considering only the residue–solvent interactions based on hydrophobicity indices (See Appendix B) and residue–residue interactions based on Skolnick's interaction parameters (See Appendix C) is not sufficient to create a minimum in the free energy with respect to the separation distance d between the two proteins. Actually, unlike the tetramerization case, hydrophobic interactions give a decreasing free energy contribution with respect to the separation distance d as do all the other contributions.

Because the relative orientation of proteins within C_{222} crystal comes from an experimental study of the electron density projection map of the crystal, this relative orientation is only approximately known. However, even when varying the mutual alignment of the two proteins within the uncertainty of their published position, we do not observe the appearance of a well-defined minimum in the free energy. On the surface of streptavidin, there are mainly hydrophilic residues, and exposing them to the solvent does not lead to a distinct energy penalty within the interaction scheme of the model presented so far. Note that hydropho-

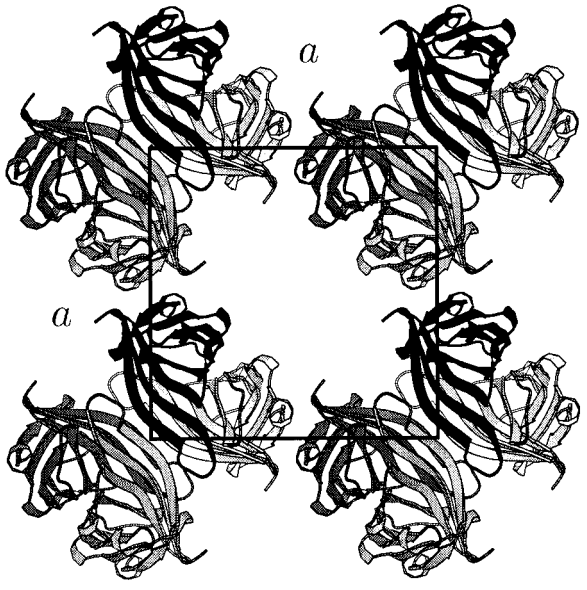

FIGURE 3 C_{222} crystal structure. The lattice parameter a is about 58 Å. The orientation of streptavidin within the crystal corresponds to experimental observations. The four crystal contacts are all equivalent. This figure was obtained with MolScript (Kraulis, 1991).

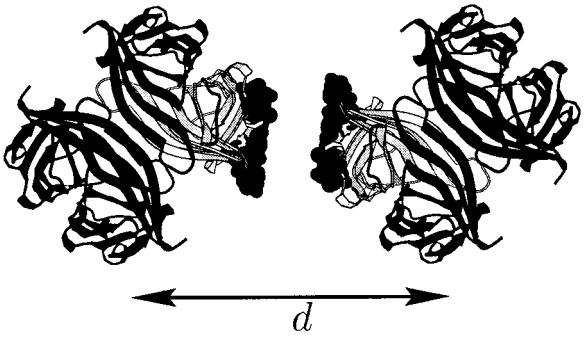

FIGURE 4 Two streptavidin proteins separated by a distance d. In the equilibrium structure, this distance is about $d = 58$ Å. Filled circles represent the selected residues allowed to change their cms position compared to the one found in PDB file. This figure was obtained with MolScript (Kraulis, 1991).

bicity can be a major factor in stabilizing protein–protein association (Chothia and Janin, 1975; Olson, 1998; Young et al., 1994). However, it appears to be more relevant for the association within subunits of a protein (e.g., in tetramerization) or between a protein and its ligand.

Note that electrostatic interactions are thought to be screened in the present experimental situation (Wang et al., 1999a,b, 2000). Note also that several studies report that direct protein–protein contacts are reinforced by well-ordered solvent molecules. Extended patterns of hydrogen bonds can therefore be formed to supplement and strengthen the few direct interactions between residues (Frey et al., 1988; Salemme et al., 1988). This contribution to the interaction energy potential is not included in the Skolnick interaction parameters. They have been evaluated with a statistical study made on the internal structure of individual proteins where interactions reported in Frey et al. (1988) and Salemme et al. (1988) are less likely to exist. Skolnick and associates introduced hydrogen bonds explicitly into their model for protein folding, but only between backbone elements related to C_α positions. The motivation for introducing these interactions was to enhance secondary structure formation during the folding process (Skolnick et al., 1997; Kolinski et al., 1993), but not the formation of bonds between different molecules during crystallization.

We therefore introduce a new contribution to the mean-field model presented in the second section. This contribution can be seen as an addition to the Skolnick interaction parameters. This new contribution βF^{H} to the reduced free energy of the system is

$$\beta F^{\mathrm{H}} = \frac{1}{2}\sum_{\nu=1}^{\mathrm{N}}\sum_{\substack{\mu=1;\\ \nu\neq\mu}}^{\mathrm{N}}\int \mathrm{d}\mathbf{x}\int \mathrm{d}\mathbf{x}'\rho_{\mathrm{n}}(\mathbf{x})\beta V^{\mathrm{H}}_{\mathrm{nm}}(\mathbf{x}-\mathbf{x}')\rho_{\mathrm{m}}(\mathbf{x}')$$

$$= \frac{1}{2}\sum_{\nu=1}^{\mathrm{N}}\sum_{\substack{\mu=1;\\ \nu\neq\mu}}^{\mathrm{N}}\chi_{\mathrm{nm}}\frac{1}{V}\int \mathrm{d}\mathbf{x}\phi_{\mathrm{n}}(\mathbf{x})\phi_{\mathrm{m}}(\mathbf{x}), \qquad (2)$$

following the simple idea of Eq. B5 for the residue–solvent contribution. The new parameters χ_{nm} with

$$\chi_{\mathrm{nm}} = \begin{cases} 0; & \text{if no possible H-bond} \\ -\alpha; & \text{if possible H-bond,} \end{cases} \qquad (3)$$

are introduced to favor attractions between residues that allow hydrogen bonds (direct or mediated by the solvent). In Eq. 3, α is a free parameter and sets the scale of this energy contribution.

The additional term Eq. 2 does not bring any relevant change to the tetramerization problem. The location of the minimum in the free energy with respect to the separation distance between the two dimers is not changed by introducing this additional term. That is because residues allowing hydrogen bonds do not interact with each other through the dimer–dimer interface.

By contrast, addition of the hydrogen bond attraction to the model now causes a minimum to appear in the contact free energy with respect to the separation distance between the two proteins in the C_{222} lattice as shown in Fig. 5. It is located at $d = 59$ Å in good agreement with the experimental lattice parameter a of C_{222} (cf. Fig. 4). The free energy of contact is estimated to be around 10 $k_{\mathrm{B}}T$.

CONCLUSIONS

The mean-field analysis of streptavidin tetramerization presented in this paper shows clearly that hydrophobicity of the dimer–dimer interface plays the main role in stabilizing the tetramer against dissociation into two dimers.

Unlike tetramerization, the existence of the C_{222} lattice in 2D streptavidin crystals cannot be explained with arguments based on hydrophobicity alone. Residue–residue interactions, which could be mediated by the solvent, give the main contribution to the stability of C_{222}. Hydrogen bonds could play an important role in this stability. A further study of the validity of the effective interactions we have used should be carried on to confirm that fact.

APPENDIX A: MEAN-FIELD APPROACH

For clarity, in this section, we will write the guidelines of the mean-field theory established in Völkel and Noolandi (1996). The main goal of this section is to determine the probability (density) of finding the cms of each of the effective atoms at a given position and the associated mean fields in

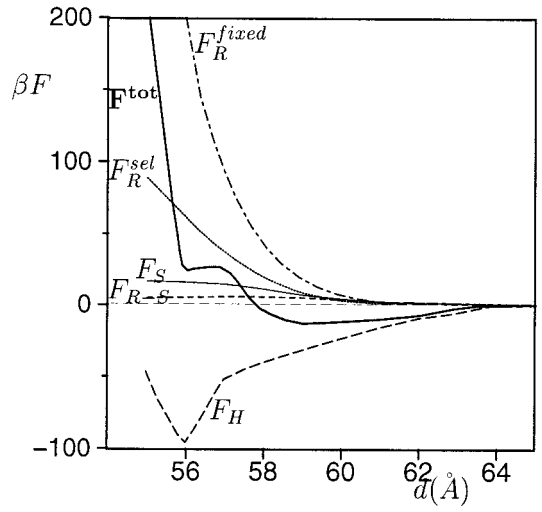

FIGURE 5 The reduced free energy βF as a function of the separation distance d between the two proteins. The lines on the figure represent the total free energy (F^{tot}, *thick solid line*), the free energy due to fixed effective atoms ($F_{\mathrm{R}}^{\mathrm{fixed}}$, *dot-dashed line*), the free energy due to the selected (moving) effective atoms ($F_{\mathrm{R}}^{\mathrm{sel}}$, *dotted line*), the entropic part of the free energy due to the solvent (F_{S}, *thin solid line*), the solvent–residue free energy ($F_{\mathrm{R-S}}$, *dashed line*) and the additional term Eq. 2 due to hydrogen bonds (F^{H}, *long dashed line*).

which the effective atoms are moving. We will also evaluate the Helmholtz free energy of the protein–solvent system.

First, we rewrite the partition function of a protein–solvent system in such a way that its free energy can be approximated in a systematic and consistent way (Völkel and Noolandi, 2000). The starting point of our approach is the partition function,

$$Z = \mathcal{L}\int \prod_{n=1}^{N} d\mathbf{x}_n \prod_{j=1}^{N_s} d\mathbf{r}_j e^{-\beta V(\{\mathbf{x}_n\},\{\mathbf{r}_j\})}, \tag{A1}$$

where N is the number of effective atoms representing proteins, N_s is the number of solvent atoms, $\mathbf{x}_n$ is the position of effective atom n, and $\mathbf{r}_j$ is the position of solvent molecule j. $V(\{\mathbf{x}_n\}, \{\mathbf{r}_j\})$ denotes the potential energy of the protein–solvent system and $\beta = 1/k_B T$ with the temperature T, and Boltzmann's constant k_B. $\mathcal{L}$ is the partition function related to the total kinetic energy,

$$\mathcal{L} = \frac{\Lambda_s^{-3N_s}}{N_s!}\prod_{n=1}^{N}\Lambda_n^{-3}, \tag{A2}$$

where $\Lambda_\nu = \sqrt{\beta h^2/2\pi m_\nu}$ is the thermal de Broglie wavelength of particles ν with mass m_ν (the index ν is used to label either effective atoms or solvent molecules). First, we characterize each effective atom by a density $\hat{\rho}_n(\mathbf{x})$,

$$\hat{\rho}_n(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{x}_n), \tag{A3}$$

and the solvent by the density $\hat{\rho}_s(\mathbf{x})$,

$$\hat{\rho}_s(\mathbf{x}) = \sum_{j=1}^{N_s}\delta(\mathbf{x} - \mathbf{r}_j). \tag{A4}$$

Using standard density functional techniques, we generalize the densities from the delta distributions in Eqs. A3 and A4 to continuous functions $\rho_\nu(\mathbf{x})$, $\nu = 1, 2, \ldots, N, S$ with $\int d\mathbf{x}\rho_\nu(\mathbf{x}) = N_\nu$ ($N_n = 1$ for $n = 1, 2, \ldots, N$). The density $\rho_n(\mathbf{x})$ represents the probability of finding the cms of the effective atom n at position $\mathbf{x}$. $\rho_s(\mathbf{x})$ is the usual solvent density. Introducing these functions $\rho_\nu(\mathbf{x})$, we can rewrite the exponential in Eq. A1 as

$$e^{-\beta V(\{\mathbf{x}_n\},\{\mathbf{r}_j\})} = \int \prod_{\nu=1}^{N+1}\mathcal{D}[\rho_\nu]\delta(\rho_\nu(\mathbf{x}) - \hat{\rho}_\nu(\mathbf{x}))e^{-\beta W[\{\rho_\nu\}]}, \tag{A5}$$

with

$$\beta W[\{\rho_\alpha\}] = \frac{1}{2}\sum_{\nu=1}^{N+1}\sum_{\substack{\mu=1;\\ \nu\neq\mu}}^{N+1}\int d\mathbf{x}\, d\mathbf{x}'\, \rho_\nu(\mathbf{x})\beta V_{\nu\mu}(\mathbf{x} - \mathbf{x}')\rho_\mu(\mathbf{x}'), \tag{A6}$$

where the notation $W[\{\rho_\alpha\}]$ denotes that W is a functional of the densities $\rho_\alpha(\mathbf{x})$ and where we assume pair-additivity for the interaction potential energy βV. Respectively, βV_{nm} and βV_{ns} represent the potential energy of atoms n and m interacting with each other and of effective atom n interacting with the solvent. The delta functions $\delta(\rho_\nu(\mathbf{x}) - \hat{\rho}_\nu(\mathbf{x}))$ can be replaced by an exponential function with the help of auxiliary fields $\omega_\nu(\mathbf{x})$,

$$\delta(\rho_\nu(\mathbf{x}) - \hat{\rho}_\nu(\mathbf{x})) = \int \mathcal{D}[\omega_\nu]\exp\left\{\int d\mathbf{x}\, \omega_\nu(\mathbf{x})(\rho_\nu(\mathbf{x}) - \hat{\rho}_\nu(\mathbf{x}))\right\}. \tag{A7}$$

Defining

$$Q_\nu = \int d\mathbf{x}\, e^{-\omega_\nu(\mathbf{x})}, \tag{A8}$$

we can rewrite the partition function Eq. A1 as

$$Z = \int \prod_{\nu=1}^{N+1}\{\mathcal{D}[\rho_\nu]\mathcal{D}[\omega_\nu]\}e^{-\beta\mathcal{F}[\{\rho_\nu\},\{\omega_\nu\}]}, \tag{A9}$$

with

$$\beta\mathcal{F}[\{\rho_\nu\},\{\omega_\nu\}] = \beta W[\rho_\nu] - \sum_{\nu=1}^{N+1}\left\{\int d\mathbf{x}\, \rho_\nu(\mathbf{x})\omega_\nu(\mathbf{x}) + N_\nu \ln(Q_\nu\Lambda_\nu^{-3})\right\} + \ln(N_s!). \tag{A10}$$

Eqs. A9 and A10 give an exact representation of the partition function of the system and are a convenient starting point for developing systematic approximations for calculating the partition function. In particular, we use a mean-field approximation where Eq. A9 is replaced by the maximum of its integrand. In this case, Eq. A10 must fulfill the conditions,

$$\frac{\partial\mathcal{F}[\{\rho_\nu\},\{\omega_\nu\}]}{\partial\rho_\nu(\mathbf{x})} = 0 \quad \frac{\partial\mathcal{F}[\{\rho_\nu\},\{\omega_\nu\}]}{\partial\omega_\nu(\mathbf{x})} = 0, \tag{A11}$$

which lead to the coupled set of equations

$$\rho_\nu(\mathbf{x}) = \frac{N_\nu}{Q_\nu}e^{-\omega_\nu(\mathbf{x})}$$

$$\omega_n(\mathbf{x}) = \sum_{m=1,\, m\neq n}^{N}\int d\mathbf{x}'\, \beta V_{nm}(\mathbf{x} - \mathbf{x}')\rho_m(\mathbf{x}') + \int d\mathbf{x}'\, \beta V_{ns}(\mathbf{x} - \mathbf{x}')\rho_s(\mathbf{x}') \tag{A12}$$

$$\omega_s(\mathbf{x}) = \sum_{n=1}^{N}\int d\mathbf{x}'\, \beta V_{sn}(\mathbf{x} - \mathbf{x}')\rho_n(\mathbf{x}').$$

By doing this approximation, we replace the detailed interactions between the different components of the system by the average of all the interactions that each of the individual components sees. Substituting Eqs. A12 in Eq. A10 yields (using Stirling's approximation for large N_s)

$$\beta F = \sum_{\nu=1}^{N+1}\int d\mathbf{x}\, \rho_\nu(\mathbf{x})\left\{\frac{1}{2}\omega_\nu(\mathbf{x}) + \ln(\Lambda_\nu^3\rho_\nu(\mathbf{x}))\right\} - \int d\mathbf{x}\, \rho_s(\mathbf{x}), \tag{A13}$$

where βF is the reduced Helmholtz free energy of the protein–solvent system. Besides the two-particle interactions and the usual kinetic energy contributions, we obtain an additional entropic part, $\Sigma_{n=1}^{N} \int d\mathbf{x}\, \rho_n(\mathbf{x}) \ln \rho_n(\mathbf{x})$, which arises from the finite width of the densities $\rho_n(\mathbf{x})$.

APPENDIX B: RESIDUE–SOLVENT INTERACTIONS

Interactions between the solvent (water in the present case) and proteins take place at their mutual interfaces (Völkel and Noolandi, 2000). We characterize each effective atom n by $\rho_n(\mathbf{x})$, i.e., the probability of finding the cms at position $\mathbf{x}$ (See Appendix A). Because the effective atom represents a group of atoms, we map the cms density $\rho_n(\mathbf{x})$ to the interface of the effective atom. Assuming each effective atom to be a sphere with an effective atom-dependent radius R_n, we can define a volume fraction $\phi_n(\mathbf{x})$ as

$$\phi_n(\mathbf{x}) = \int d\mathbf{y}\, K_n(\mathbf{x} - \mathbf{y}) \rho_n(\mathbf{y}) \tag{B1}$$

with

$$K_n(\mathbf{x}) = \begin{cases} K_{n0}; & |\mathbf{x}| \le R_n \\ 0; & \text{else,} \end{cases} \tag{B2}$$

where K_{n0} is a normalization constant, which is chosen such that $0 \le \phi_n(\mathbf{x}) \le 1$ everywhere in the system. Figure 6 shows schematically how this mapping works: $\phi_n(\mathbf{x}) = 1$ at places that are always occupied by the effective atom n; the interface is defined as the region that is occupied by the effective atom with a finite probability less than 1. Because of the incompressibility of the system, i.e.,

$$\sum_{n=1}^{N} \phi_n(\mathbf{x}) + \phi_s(\mathbf{x}) = 1, \tag{B3}$$

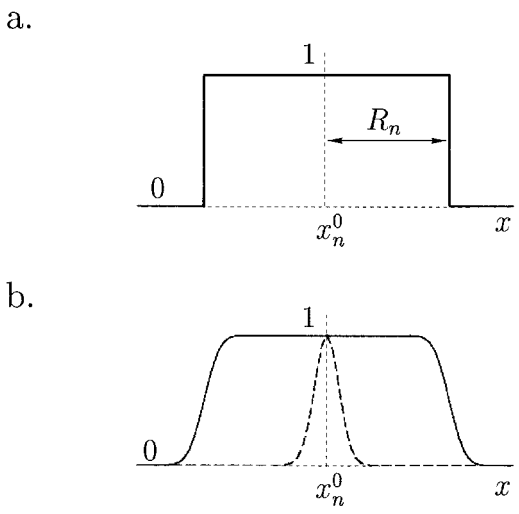

FIGURE 6 Schematic plot of the mapping of $\rho_n(\mathbf{x})$ to $\phi_n(\mathbf{x})$ (cf. Eq. B1). (*A*) Full line represents $\phi_n(\mathbf{x})$ for a situation of a fixed effective cms to position $\mathbf{x}_n^0$ ($\rho_n(\mathbf{x})$ is therefore a delta function centered on $\mathbf{x}_n^0$, not illustrated). (*B*) $\phi_n(\mathbf{x})$ (*full line*) for a nonfixed cms characterized by $\rho_n(\mathbf{x})$ (*dashed line*).

we calculate the solvent volume fraction as

$$\phi_s(\mathbf{x}) = 1 - \sum_{n=1}^{N} \phi_n(\mathbf{x}). \tag{B4}$$

Introducing volume fractions $\phi_\nu(\mathbf{x})$ and following the Flory–Huggins model for polymers, we rewrite the contribution of solvent–protein interactions to the free energy (cf. Eq. A13) as

$$\int d\mathbf{x}\, \rho_s(\mathbf{x}) \omega_s(\mathbf{x}) = \sum_{n=1}^{N} \chi_n \frac{1}{V} \int d\mathbf{x}\, \phi_n(\mathbf{x}) \phi_s(\mathbf{x}), \tag{B5}$$

where V is the volume of the system. Within this model the solvent–residue interaction is replaced by a contact interaction characterized by parameters χ_n, which describe the hydrophobic/hydrophilic nature of the effective atom n. Note that the incompressibility of the system Eq. B3 leads to nonzero contributions to Eq. B5 only where $\phi_n(\mathbf{x})$ and $\phi_s(\mathbf{x})$ overlap, i.e., at the mutual interfaces of the effective atoms and the solvent. That means that, within the present model, there is no interaction between the solvent and the backbone or the fixed side chains, because these residues volume fractions have no overlap with the solvent volume fraction ϕ_S. For the effective atom n (for nonfixed side chains), we define the parameters χ_n as

$$\chi_n = a V^{2/3} \frac{\beta G_n}{S_n}, \tag{B6}$$

where S_n is the free surface area of the effective atom (estimated by the method suggested by Lee and Richards (1971), and βG_n is the reduced solvation free energies per particle (Eisenberg and McLachlan, 1986), $\beta G_n = 2.3 \Pi_F(n)$. Hydrophobic indices $\Pi_F(n)$ were estimated at neutral pH by Fauchère and Pliska (1983) by comparing the distribution coefficients of amino acid amides in water and octanol and fixing the absolute scale such that $\Pi_F(\text{Gly}) = 0$. The constant a allows us to calibrate the numerical protein–solvent interaction strength from experimental results. More details about the way the different parameters introduced in the present theory are fixed can be found elsewhere (Völkel and Noolandi, 2000).

APPENDIX C: RESIDUE–RESIDUE INTERACTIONS

For the interactions between the different effective atoms, we use tabulated two-body interaction parameters obtained by Skolnick and coworkers (Kolinski et al., 1993; Skolnick et al., 1997), who use a similar coarse-graining procedure (two effective atoms per residue) for their study of protein folding. Because we always start with an experimentally determined protein structure in the present model, we only use two-body interactions and neglect higher-order interactions that were introduced by Skolnick and coworkers to enhance the formation of the secondary structure or to correct for the discretization effects of their lattice model. We also neglect the single-body interactions, because they were originally introduced to mimic solvent–protein interactions, which are explicitly included in the present model. Because of the continuous nature of our field theoretical approach we use the two-body potential (Völkel and Noolandi, 2000) between residues i and j at distance r_{ij}

$$\beta V_{ij}(r_{ij}) = \begin{cases} \beta V_{ij}^{K}(r_{ij}) (1/\mathcal{N}_{ij}) \{(1/r_{ij}^2) - (1/R_{ij}^2)\}; & R_{ij}^{\text{rep}} < r_{ij} < R_{ij} \\ \beta V_{ij}^{K}(r_{ij}); & \text{else,} \end{cases} \tag{C1}$$

which multiplies a $1/r^2$ dependence to the otherwise constant Skolnick two-body potential (Skolnick et al., 1997),

$$\beta V_{ij}^{K}(r_{ij}) = \begin{cases} \beta V_{rep}; & r_{ij} < R_{ij}^{rep}, \\ \beta\epsilon_{ij}; & R_{ij}^{rep} < r_{ij} < R_{ij} \quad \text{and} \quad \epsilon_{ij} \geq 0, \\ f_{ij}\beta\epsilon_{ij}; & R_{ij}^{rep} < r_{ij} < R_{ij} \quad \text{and} \quad \epsilon_{ij} < 0, \\ 0; & r_{ij} > R_{ij}. \end{cases} \tag{C2}$$

The two-body interaction parameters ϵ_{ij} between residues of type i and type j are estimated as

$$\beta\epsilon_{ij} = -\ln \frac{N_{ij}}{\langle N_{ij} \rangle}, \tag{C3}$$

where N_{ij} and $\langle N_{ij} \rangle$ are, respectively, the number of contact between effective atoms i and j as found in experimentally determined structures of proteins and in a completely random protein.

The normalization constants $\mathcal{N}_{ij}$ in Eq. C1 are chosen such that the average interaction energies $\langle V_{ij}^{K} \rangle$ and $\langle V_{ij} \rangle$ are the same within the spherical shell $R_{ij}^{rep} < r_{ij} < R_{ij}$. For close encounters $r_{ij} < R_{ij}^{rep}$ repulsion corresponding to an energy of $\beta V_{rep} = 6$ is considered. The contact distance (Kolinski et al., 1993) R_{ij} is estimated as the distance between the two effective atoms i and j with their heavy atoms not more than 4.2 Å separated from each other. For $r_{ij} > R_{ij}$ the interaction is zero. The factor,

$$f_{ij} = 1 - \{\cos^2(\mathbf{u}_i, \mathbf{u}_j) - \cos^2(\pi/9)\}^2, \tag{C4}$$

reflects the average angle between elements of secondary structure seen in globular proteins ($\mathbf{u}_i = \mathbf{r}_{i+2} - \mathbf{r}_{i-2}$ where $\mathbf{r}_i$ is the coordinate of the ith C_α on the backbone).

We thank Dr. Szu-Wen Wang for helpful discussions.

REFERENCES

Chothia, C., and J. Janin. 1975. Principles of protein–protein recognition. *Nature.* 256:705–708.

Durbin, S. D., and G. Feher. 1996. Protein crystallization. *Annu. Rev. Phys. Chem.* 47:171–204.

Eisenberg, D., and A. D. McLachlan. 1986. Solvation energy in protein folding and binding. *Nature.* 319:199–203.

Fauchère, J.-L., and V. Pliska. 1983. Hydrophobic parameter Π of amino-acid side chains from the partitioning of N-acetyl-amino-acid amides. *Eur. J. Med. Chem. Chim. Ther.* 18:369–375.

Frey, M., J.-C. Genovesio-Taverne, and J. C. Fontecilla-Camps. 1988. Application of the periodic bond chain (PBC) theory to the analysis of the molecular packing in protein crystals. *J. Cryst. Growth.* 90:245–258.

Lee, B., and F. M. Richards. 1971. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* 55:379–400.

Katz, B. A. 1995. Binding to protein targets of peptidic leads discovered by phage display: crystal structures of streptavidin-bound linear and cyclic peptide ligands containing the HPQ sequence. *Biochemistry.* 34: 15421–15429.

Kolinski, A., A. Godzik, and J. Skolnick. 1993. A general method for the prediction of the three dimensional structure and folding pathway of globular proteins: Application to designed helical proteins. *J. Chem. Phys.* 98:7420–7433.

Kraulis, J. 1991. A program to produce both detailed and schematic plots of protein structures. *J. Applied Crystallog.* 24:946–950.

Noolandi, J., T. S. Davison, A. R. Völkel, X.-F. Nie, C. Kay, and C. H. Arrowsmith. 2000. A Meanfield approach to the thermodynamics of a protein–solvent system with application to the oligomerization of the tumor suppressor P53. *Proc. Natl. Acad. Sci. U.S.A.* 97:9955–9960.

Olson, M. A. 1998. Mean-field analysis of protein–protein interactions. *Biophys. Chem.* 75:115–128.

Piazza, R., V. Peyre, and V. Degiorgio. 1998. "Sticky hard spheres" model of proteins near crystallization: a test based on the osmotic compressibility of lysozyme solutions. *Phys. Rev. E.* 58:R2733–R2736.

Rosenberger, F., P. G. Vekilov, M. Muschol, and B. R. Thomas. 1996. Nucleation and crystallization of globular proteins—what we know and what is missing. *J. Cryst. Growth.* 168:1–27.

Rosenbaum, D. F., and C. F. Zukoski. 1996. Protein interactions and crystallization. *J. Cryst. Growth.* 169:752–758.

Salemme, F. R., L. Genieser, B. C. Finzel, R. M. Hilmer, and J. J. Wendoloski. 1988. Molecular factors stabilizing protein crystals. *J. Cryst. Growth.* 90:273–282.

Sano, T., S. Vajda, C. L. Smith, and C. R. Cantor. 1997. Engineering subunit association of multisubunit proteins: a dimeric streptavidin. *Proc. Natl. Acad. Sci. U.S.A.* 94:6153–6158.

Skolnick, J., L. Jaroszewski, A. Kolinski, and A. Godzik. 1997. Derivation and testing of pair potentials for protein folding. When is the quasichemical approximation correct? *Protein Sci.* 6:676–688.

Völkel, A., and J. Noolandi. 1996. Mean-field approach to the thermodynamics of complex solids. *J. Comp. Aided Mat. Design.* 3:289–295.

Völkel, A., and J. Noolandi. 2000. Meanfield approach to the thermodynamics of protein–solvent systems with application to p53. *Biophys. J.* In press.

Wang, S.-W., C. R. Robertson, and A. P. Gast. 1999a. Molecular arrangement in two-dimensional streptavidin crystals. *Langmuir.* 15: 1541–1548.

Wang, S.-W., C. R. Robertson, and A. P. Gast. 1999b. Two-dimensional crystallization of streptavidin mutants. *J. Phys. Chem. B.* 103: 7751–7761.

Wang, S.-W., C. R. Robertson, A. P. Gast, S. Koppenol, T. Edwards, V. Vogel, and P. Stayton. 2000. Role of N- and C-terminal amino acids in 2D streptavidin crystal formation. *Langmuir.* 16:5199–5204.

Young, L., R. L. Jernigan, and D. G. Covell. 1994. A role for surface hydrophobicity in protein–protein recognition. *Protein Sci.* 3:717–729.