Meanfield Approach to the Thermodynamics of Protein–Solvent Systems with Application to P53

Armin R. Völkel and Jaan Noolandi
Xerox Research Centre of Canada, Mississauga, Ontario L5K 2L1, Canada

ABSTRACT We present a meanfield theoretical approach for studying protein–solvent interactions. Starting with the partition function of the system, we develop a field theory by introducing densities for the different components of the system. At this point, protein–solvent interactions are introduced following the inhomogeneous Flory–Huggins model for polymers. Finally, we calculate the free energy in a meanfield approximation. We apply this method to study the stability of the tetramerization domain of the tumor suppressor protein p53 when subjected to site-directed mutagenesis. The four chains of this protein are held together by hydrophobic interactions, and some mutations can weaken this bond while preserving the secondary structure of the single protein chains. We find good qualitative agreement between our numerical results and experimental data, thus encouraging the use of this method as a guide in designing experiments.

INTRODUCTION

Understanding the relationship between the native conformation of a protein and its amino acid sequence ("protein folding problem") has been one of the great issues of natural sciences over the past decades (see e.g., Creighton, 1992). Though the possible number of conformations of a polypeptide chain is immense, protein sequences fold into their unique native structure within seconds (Levinthal, 1969). Predicting this native structure for an arbitrary amino acid sequence is still an unresolved problem.

It has been established that the amino acid side chains each exist in a finite number of conformations (rotamers), and early work has been devoted to the prediction of the optimum side chain packing of a given amino acid sequence and backbone conformation (Roitberg and Elbers, 1991; Koehl and Delaru, 1994) using the rotamer libraries of Ponders and Richards (1987) and Tuffery et al. (1991, 1993). Extensions of these approaches to homology models, where a finite number of possible backbone conformations are included into the packing algorithm are discussed by Koehl and Delaru (1995, 1996) and Vàsquez (1996). Recently, the rotamer libraries have been refined by including the backbone dependence of the different side chain conformations (Dunbrack and Karplus, 1993).

Early approaches for describing the protein-folding process used lattice models with artificial residues (Lau and Dill, 1989; Shakhanovich et al., 1991; Miller et al., 1992), however, there are now more realistic models available which try to predict the conformation of a given amino acid sequence (see e.g., Kolinski and Skolnick, 1992, 1994; Kolinski et al., 1993; Karplus and Weaver, 1994; Hinds and Levitt, 1994; Srinivasan and Rose, 1995; Cheng et al., 1996; Carlacci and Englander, 1996). It is beyond the scope of this article to discuss (or even reference) all the work published on protein folding, but extended reviews of this topic are published by Creighton (1992), Karplus and Šali (1995), and Levitt (1996).

A different problem is the prediction of the stability of a protein subjected to site-directed mutagenesis. Here, the starting point is a known protein structure and the task is to find the relative change arising from mutant. If the backbone conformation of the mutant does not deviate from the wild-type form, the problem reduces to finding the optimum side chain packing using one of the above-mentioned search algorithms. However, if the backbone structure changes significantly with the mutation, one ends up again with the protein-folding problem.

A systematic study of the stability of the λ repressor protein under a three-site mutation within its hydrophobic core has been presented by Lee and Levitt (1991) and Lee (1994), and compared to an exhaustive experimental study by Lim and Sauer (1991). In the numerical approach, the backbone was held fixed while the side chains were considered rather flexible with (almost) free rotating torsion angles (the bond lengths and angles were kept fixed). By comparing the energies of the mutants with the energy of the wild-type structure and correlating these energy differences to the protein activity, Lee and Levitt (1991) and Lee et al. (1994) found good agreement between their calculations and the experimental data. By confining their method to the hydrophobic core region of a single polypeptide chain, they ignored any solvent–protein interactions. This approximation fails if mutations at the solvent–protein interface or in the hydrophobic core of the quaternary structure of oligomeric proteins are considered. Here, interactions of the amino acids with the solvent can become very important for the protein stability, especially when the mutation has a different hydrophobicity value than the native residue.

Received for publication 13 September 2000 and in final form 27 November 2000.

Address reprint requests to Armin Völkel, Ph.D., Xerox Palo Alto Research Center, 3333 Coyote Hill Rd., Palo Alto, CA 94304. Tel.: 650-812-4198; Fax: 650-812-4140; E-mail: avolhel@pavc.xerox.com.

0006-3495/01/03/1524/14 $2.00

Protein–solvent systems have been studied on the atomistic level (see e.g., Dagget and Levitt, 1993; Karplus and Šali, 1995 and references therein). These calculations can reveal an accurate dynamics of these systems, but they are limited to small proteins and short simulation times (the time step for one integration is of the order of 1 fs). Approaches that use a coarser model for the protein structure commonly use a dielectric function to describe the solvent (Pappu et al., 1996; Ullner et al., 1996).

The goal of this approach is to develop a model that can describe the interactions of a single protein with solvent. To keep this model as simple as possible, we coarse-grain the protein by introducing "effective atoms," which carry the main features of the part they represent (either backbone or side chain). We then introduce a continuum approach where the coordinates of each of the effective atoms and the solvent are represented by a density. Applying a meanfield approach to the system allows us to write down the free energy of the system in a rather simple analytic form.

The advantages of this approach are: 1) The reduction of the degrees of freedom allows for a faster sampling of the possible conformations. It also smoothes the energy landscape, which makes it easier to move between the conformations corresponding to global free energy minima of the system. 2) The field theoretic approach allows for a self-consistent search of the optimum conformation in a properly weighted ensemble (field) of possible states. It also makes the introduction of protein–solvent interactions in the form of the inhomogeneous Flory–Huggins model for polymers (Flory, 1953) straightforward. Though the meanfield approach includes naturally forbidden conformations (due to packing constraints), this does not significantly reduce the accuracy of the predicted side chain conformations limited by the used intraprotein interactions (Vàsquez, 1996).

Stability studies of proteins can be performed by calculating the free energy of static protein conformations starting from the experimental known structure, then allowing for small changes in the total protein conformation. Stable structures are recognized by a local minimum in the free energy. In particular, if we consider the stability of oligomers, we can, in a first approximation, consider the backbones of the different polypeptide chains to be rigid and calculate the free energies of the protein with the single chains at different separations.

We compare our computational results with experimental data obtained from the protein p53 and mutations (C. H. Arrowsmith, private communication; T. S. Davison, private communication). Wild-type p53 is a sequence-specific DNA binding protein and is recognized as one of the key tumor suppressors in the human body (Friend, 1994). p53 consists of four equivalent polypeptide chains held together through hydrophobic interactions. The three-dimensional structure of this hydrophobic domain (or tetrameric domain) has been obtained recently by nuclear magnetic resonance (Clore et al., 1994, 1995a,b; Lee et al., 1994) and X-ray crystallography (Jeffrey et al., 1995)) (Fig. 1). The tetramer has three axes of two-fold symmetry and can be considered as a dimer of dimers. Wild-type p53 in its tetrameric form is most effective, but even monomeric and dimeric p53 has some tumor-suppressing abilities. To understand the role of monomeric/dimeric p53 in its function as tumor suppressor, and the interplay of all four monomers during this process, it is of interest to generate dimeric and monomeric p53. Because the DNA binding sites are separate from the tetramerization domain, mutations in the latter are not likely to change the tumor-suppressing properties of monomeric/dimeric p53. The most likely candidates for mutation are the amino acids in the hydrophobic core of the tetramer, e.g., Met340.

p53 is a good candidate to evaluate the effectiveness of our model. Its tetramerization domain can be viewed as a dimer of dimers (Fig. 1), and a separation of these two dimers (each of which consists of two α-helices at the interface, held together by a hairpin-like structure at their back) most likely will not destroy their backbone conformation (as long as the mutant amino acid preserves the local secondary structure), thus allowing us to focus on the protein–solvent interactions without first solving the protein-folding problem.

Replacing single amino acids in proteins (site-directed mutagenesis) is now a standard technique in molecular biology, but still requires some effort. A useful computer model could allow one to evaluate potentially disruptive mutation sites and candidates, thus limiting the number of mutations to be generated in the laboratory.

The manuscript is organized as follows: In the next section, we explain our model in some detail. The following section reviews the interactions used for the numerical calculations. The application to p53 is discussed in the

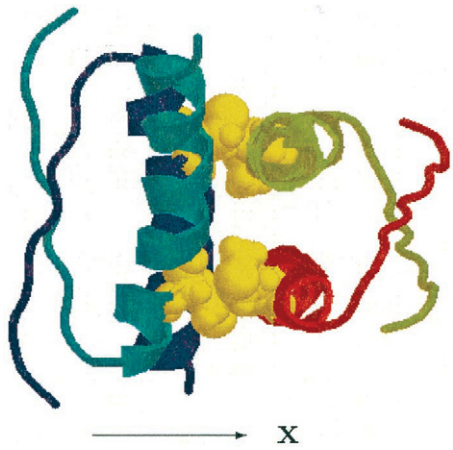

FIGURE 1 The tetramerization domain of p53 (amino acids Gly-325 through Ala-355) from a side view. The site 340 used for the single-point mutations is highlighted, showing the wild-type amino acid methionine (*yellow*). The arrow labeled x indicates the direction of the dimer–dimer separation.

section, "Application to p53," and, in the next section, we present our results and compare them to experimental data. The final section concludes with a short summary.

MODEL

The goal of this section is to rewrite the partition function of a polymer–solvent system in such a way that its free energy can be approximated in a systematic and consistent way. The starting point of our approach is the partition function,

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

where $\mathcal{L}$ is partition function related to the total kinetic energy, N is the total number of particles in the polymer, N_S is the number of solvent molecules, $\mathbf{x}_n$ is the position of particle n, and $\mathbf{r}_j$ is the position of solvent molecule j. $V(\{\mathbf{x}_n\}, \{\mathbf{r}_j\})$ denotes the potential energy, $\beta = 1/k_B T$ where T is the temperature and k_B is Boltzman's constant. The particles can be atoms or, more generally "effective atoms," i.e., sets of atoms within functional groups of the polymer. In the latter case, $\mathbf{x}_n$ would represent the center of mass (cms) of this effective atom. This approach to represent a group of real atoms as a single entity is especially useful for biopolymers where the building blocks are typically monomers with a distinct chemical or biological functionality. In the context of our goal, we group atoms together, which build up small domains of equal hydrophobicity/hydrophilicity.

In a first step we characterize each effective atom by a density,

$$\hat{\rho}_n(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{x}_n), \quad (2)$$

and the solvent by the density

$$\hat{\rho}_S(\mathbf{x}) = \sum_{j=1}^{N_s} \delta(\mathbf{x} - \mathbf{r}_j). \quad (3)$$

Using standard density functional techniques we can generalize the densities from the delta distributions in Eqs. 2 and 3 to continuous functions $\rho_\nu(x)$, $\nu = 1, \ldots, N, S$ (Appendix A). By doing this, we also introduce for each density $\rho_\nu(\mathbf{x})$ a conjugate field $\omega_\nu(\mathbf{x})$.

The resulting partition function, Eq. A5, is equivalent to the original form Eq. 1, but describes the system in terms of the densities $\rho_\nu(\mathbf{x})$ and their conjugate fields $\omega_\nu(\mathbf{x})$. Though we have not gained any advantage in solving the thermodynamics of the system exactly, we have now an expression that allows us to apply systematic approximations.

In particular, we replace the partition function, Eq. A5, by the extremum of its integrand, i.e., we consider only those fields $\rho_\nu(\mathbf{x})$ and $\omega_\nu(\mathbf{x})$, $\nu = 1, \ldots, N, S$, which fulfill the conditions

$$\frac{\delta\mathcal{F}([\{\rho_\nu\},\{\omega_\nu\}])}{\delta\rho_\nu(\mathbf{x})} = 0, \quad \frac{\delta\mathcal{F}([\{\rho_\nu\},\{\omega_\nu\}])}{\delta\omega_\nu(\mathbf{x})} = 0. \quad (4)$$

By doing this approximation, we replace the detailed interactions between the different components of the system by the average of all the interactions each of the single components sees. This approximation is called "Meanfield Approach" (Hong and Noolandi, 1981; Noolandi and Hong, 1982).

Using the equations derived in Appendix A, we can rewrite Eqs. 4 as

$$\begin{aligned}
\rho_n(\mathbf{x}) &= \frac{1}{Q_n} e^{-\omega_n(\mathbf{x})} \\
\omega_n(\mathbf{x}) &= \sum_{m=1}^{N} \int d\mathbf{x}' \beta V_{nm}(\mathbf{x} - \mathbf{x}')\rho_m(\mathbf{x}') \\
&\quad + \int d\mathbf{x}' \beta V_{nS}(\mathbf{x} - \mathbf{x}')\rho_S(\mathbf{x}') + \beta W_n(\mathbf{x}) \quad (5) \\
\rho_S(\mathbf{x}) &= \frac{1}{Q_S} e^{-\omega_S(\mathbf{x})} \\
\omega_S(\mathbf{x}) &= \sum_{n} \int d\mathbf{x}' \beta V_{nS}(\mathbf{x} - \mathbf{x}')\rho_n(\mathbf{x}'),
\end{aligned}$$

where Q_n is given by Eq. A4. V_{nm} and V_{nS} are, respectively, effective atoms and effective atom–solvent pair potential energies. W_n is a single-body potential energy for effective atom n.

Substituting Eqs. 5 back into Eq. A7, we find for the Helmholtz free energy $\mathcal{F}$,

$$\begin{aligned}
\beta\mathcal{F} &= \frac{1}{2}\sum_{n=1}^{N} \int d\mathbf{x}\rho_n(\mathbf{x})[\omega_n(\mathbf{x}) + \beta W_n(\mathbf{x})] \\
&\quad + \sum_{n=1}^{N} \int d\mathbf{x}\rho_n(\mathbf{x})\ln(\rho_n(\mathbf{x})\Lambda_n^3) \\
&\quad + \frac{1}{2}\int d\mathbf{x}\rho_S(\mathbf{x})\omega_S(\mathbf{x}) \\
&\quad + \int d\mathbf{x}\rho_S(\mathbf{x})\ln(\rho_S(\mathbf{x})\Lambda_S^3), \quad (6)
\end{aligned}$$

where $\Lambda_i^3 = \beta h^2/2\pi m_i$, $i = 1, \ldots, N, S$.

The different terms of the free energy have the following meaning:

a. The first and third terms describe the enthalpic energies arising from the particle–particle and particle–solvent interactions, respectively. Note that half of the particle–solvent interactions for each of the residues n is hidden in $\omega_n(\mathbf{x})$.
b. The second and fourth terms describe the entropy of the particles and the solvent, respectively. The entropy for the effective atoms arises because we consider the probability distribution of its cms position, which has a finite width due to the finite number of locations at which the particle can be found. In the case of a well-localized particle n, $\rho_n(\mathbf{x})$ reduces to a δ-function, and the entropy of the particle becomes zero.

The interaction between the solvent and the polymer takes place at their mutual interface. So far we have calculated the density distributions of the effective atom center-of-mass positions. If the size of the effective atoms is not negligible, as is the case for situations where they represent groups of real atoms, we have to map the cms density distribution to the interface of the effective atom. If we assume each effective atom to be a sphere with effective atom-dependent radius R_n, we can define a volume fraction as

$$\phi_n(\mathbf{x}) = \int d\mathbf{y} K_n(\mathbf{x} - \mathbf{y})\rho_n(\mathbf{y}), \tag{7}$$

with the kernel

$$K_n(\mathbf{x}) = \begin{cases} K_{n0}; & |\mathbf{x}| \le R_n \\ 0; & \text{else.} \end{cases} \tag{8}$$

Figure 2 shows schematically how this mapping works: $\phi_n(\mathbf{x})$ is 1 at places that are always occupied by effective atom n; the interface is defined as the region that is occupied by the effective atom with a finite probability less than 1.

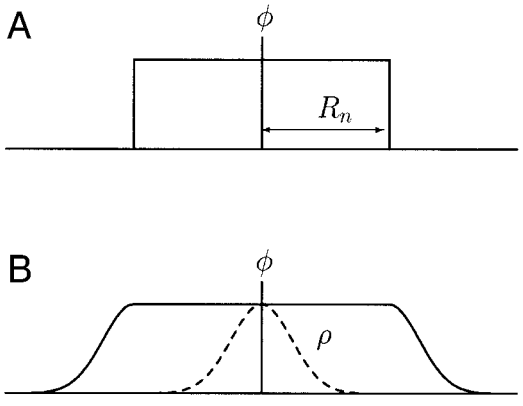

FIGURE 2 Mapping the cms density ρ to the effective atom volume fraction ϕ. (*A*) Volume fraction of an effective atom ($\rho_n = \delta(\mathbf{x} - \mathbf{x}_n)$). (*B*) Volume fraction of an effective atom with cms distribution ρ given by the dashed curve.

The K_{n0} are normalization constants that are chosen such that $0 \le \phi_n(\mathbf{x}) \le 1$ everywhere in the system. Different shaped effective atoms are easily represented by more complex kernels. For the solvent volume fraction we use

$$\phi_S(\mathbf{x}) = \frac{\rho_S(\mathbf{x})}{\rho_{0S}}, \tag{9}$$

where ρ_{0S} is the solvent bulk density.

By using volume fractions in Eqs. 5 and 6 instead of densities and following the inhomogeneous Flory–Huggins model for polymers (Flory, 1953) we can rewrite the contributions of the solvent–polymer interactions to the free energy as

$$\beta\mathcal{F}_S = \sum_{n=1}^{N} \frac{\chi_n}{V} \int d\mathbf{x}\phi_n(\mathbf{x})\phi_S(\mathbf{x}) + \rho_{0S} \int d\mathbf{x}\phi_S(\mathbf{x}) \ln \phi_S(\mathbf{x}). \tag{10}$$

Within this model, the effective atom–solvent interaction potential V_{nS} is replaced by a contact interaction described by the parameter χ_n. Because of the incompressibility of the system, i.e.,

$$\sum_{n=1}^{N} \phi_n(\mathbf{x}) + \phi_S(\mathbf{x}) = 1, \tag{11}$$

Eq. 10 has nonzero contributions only, where $\phi_n(\mathbf{x})$ and $\phi_S(\mathbf{x})$ overlap, i.e., at the mutual interface of the effective atom and the solvent.

INTERACTIONS FOR A PROTEIN–SOLVENT SYSTEM

We will now consider the special case of a protein–solvent system. Interprotein interactions and protein–solvent interactions have been studied by several groups at various levels of detail. For our purpose of studying the stability of the p53 tetramerization domain subjected to site-directed mutagenesis, we consider a coarse-graining of the protein by describing each amino acid by two effective atoms: one representing the backbone properties and sitting at the C_α position, the other representing the side chain properties and sitting at the side chain cms position. By doing this coarse-graining we lose information of the exact space-filling of the protein, but we keep the information about the solvability of the side chains which we believe is a major driving force of the p53 mutants to remain as a stable tetramer, or to dimerize.

Intraprotein interactions

For the interactions between the different effective atoms within the protein, we use tabulated force fields obtained by Kolinski and Skolnick (1992, 1994) and Vieth et al. (1994).

These authors consider a similar coarse-graining as we do for their protein-folding model. Because we always start with an experimentally determined protein structure, we use only the two-body interactions, but neglect the higher-order interactions that were introduced by Kolinski et al. (1993) to enhance the formation of secondary structures and to correct for the discretization effects of their lattice model. We also neglect their single-body interactions, because they were introduced to mimic the solvent interactions, which we include explicitly into our model. The two-body interactions are estimated as (Godzik et al., 1995)

$$\beta\epsilon_{ij} = -\ln\frac{N_{ij}}{\langle N_{ij}\rangle}, \tag{12}$$

where N_{ij} is the number of contacts between amino acids i and j as found in real proteins and $\langle N_{ij}\rangle$ is the average number of contacts between the same two amino acids in a completely random protein. Whereas Kolinski et al. (1993) use the ϵ_{ij} as constant interaction potentials between effective atoms (C'_α, side chains) i and j separated by distance r_{ij} within a certain contact distance R_{ij},

$$\beta V_{ij}^{K}(r_{ij}) = \begin{cases} \beta V_{rep} & r_{ij} < R_{ij}^{rep} \\ \beta\epsilon_{ij} & R_{ij}^{rep} < r_{ij} < R_{ij} \quad \text{and} \quad \epsilon_{ij} \geq 0 \\ f_{ij}\beta\epsilon_{ij} & R_{ij}^{rep} < r_{ij} < R_{ij} \quad \text{and} \quad \epsilon_{ij} < 0, \end{cases} \tag{13}$$

we use the ϵ_{ij} as amplitudes of $1/r^2$ potentials with the same range R_{ij}

$$\beta V_{ij}(r_{ij}) = \begin{cases} \beta V_{ij}^{K}(r_{ij}) & r_{ij} < R_{ij}^{rep} \\ \beta\dfrac{V_{ij}^{K}(r_{ij})}{\mathcal{N}_{ij}}\left\{\dfrac{1}{r_{ij}^2} - \dfrac{1}{R_{ij}^2}\right\} & R_{ij}^{rep} < r_{ij} < R_{ij}. \end{cases} \tag{14}$$

For close encounters $r_{ij} < R_{ij}^{rep}$, a repulsive "penalty" energy $\beta V_{rep} = 6$ is considered. The factor

$$f_{ij} = 1 - (\cos^2(\mathbf{u}_i, \mathbf{u}_j) - \cos^2(20°))^2, \tag{15}$$

with $\mathbf{u}_i = \mathbf{r}_{i+2} - \mathbf{r}_{i-2}$ ($\mathbf{r}_i$ is the coordinate of the ith C_α on the backbone) reflects the average angle between elements of secondary structure seen in globular proteins. The contact distances R_{ij} are estimated as the distance between two amino acids with their heavy atoms not more than 4.2 Å separated from each other. See Kolinski and Skolnick (1994) for more details. The normalization constants $\mathcal{N}_{ij}$ are chosen such that the average interaction energies $\langle V_{ij}^{K}(r_{ij})\rangle$ and $\langle V_{ij}(r_{ij})\rangle$ are the same within the spherical shell $R_{ij}^{rep} < r_{ij} < R_{ij}$.

So far, we have not included any bond-length and bond-angle constraints. In a first approximation, we assume that the cms position of each amino acid side chain remains at the same position $x_{0,n}$ (relative to its neighbors) as in the wild-type configuration, whereas its distribution function $\rho(\mathbf{x})$ may change from δ-function with changes applied to the protein. To accomplish this, we add to Eqs. 5 the constraint

$$\int d\mathbf{x}\, \mathbf{x}\rho(\mathbf{x}) = \mathbf{x}_{0,n}. \tag{16}$$

Because the side chains of the mutant amino acids differ in size and composition from the wild-type amino acid, we must relax this constraint. However, we know that the amino acid side chains can exist only in a finite number of conformations (so-called rotamers or rotameric states) due to geometric constraints (Ponders and Richards, 1987; Tuffery et al., 1991, 1992; Dunbrack and Karplus, 1993). We use the rotamer library of Tuffery et al. (1991) including the minor changes suggested by Koehl and Delaru (1994). Using tabulated torsion angles and well-known values for the bond lengths and angles within the 20 amino acids, we construct a table of cms positions for all the rotamers.

These rotameric states act like internal degrees of freedom for each amino acid side chain and can be introduced into the partition function as described in Appendix B. In the ideal case, where the side chains can assume only configurations corresponding to the rotameric states, the $f(t_n)$ in Eq. B1 are products of δ-functions. Because of the coarse-graining of the amino acid side chains into effective atoms, we replace these δ-functions by Gaussian functions with a finite width. The function $S_n(\mathbf{x})$ as defined in Eq. B10 then becomes

$$S_n(\mathbf{x}) = \sum_{j=1}^{N_n} \frac{f_{nj}}{(\pi w)^{3/2}} \exp\left\{-\frac{(\mathbf{x} - \mathbf{R}_{rot,nj})^2}{w}\right\}, \tag{17}$$

with f_{nj} describing the relative weight of rotamer j of side chain n, and w describing the width of the Gaussian.

Note: Some authors (Koehl and Delaru, 1994; Roitberg and Elbers, 1991) working on the optimization of the side chain packing for a given amino acid sequence and backbone conformation used the term *meanfield approach* when they calculated the interactions of a given side chain with the (weighted) sum of all the possible conformations of the neighboring side chains. In this paper, the meanfield approach is implemented when approximating the free energy functional Eq. A7 by its extremum (cf. Eq. 4). The weighted sum of all the rotameric states of all the amino acid side chains, which adds the harmonic potentials Eq. 17 to the interaction term (cf. Eq. B14), is an immediate consequence of counting all possible protein configurations in the partition function Eq. 1 at the start of our calculations.

Protein–solvent interactions

The interactions between the solvent and the amino acid side chains are characterized by the parameters χ_n in our model (Eq. 10). χ_n describes the hydrophobic/hydrophilic character of amino acid n. Hydrophobic indices $\Pi_F(n)$ have

been estimated by Fauchère and Pliska (1983) by comparing the distribution coefficients of amino acid amides in water and n-octanol and fixing the absolute scale such that $\Pi_F(\text{Gly}) = 0$. Eisenberg and McLachlan (1986) used these values to estimate reduced solvation free energies per amino acid $\beta G_n = 2.3RT\Pi_F(n)$(kcal/mol). For our model, we define

$$\chi_n = C_{PS} V^{2/3} \frac{\beta G_n}{S_n}, \tag{18}$$

where n labels the amino acid, S_n is the free surface area of side chain n (estimated by the method suggested by Lee and Richards, 1971), and C_{PS} is a factor common to all the χ_n, which allows us to calibrate the numerical protein–solvent interaction strength from experimental results.

APPLICATION TO p53

As a first application for the above-introduced meanfield approach, we study the stability of the tetramerization domain of the tumor suppressor protein p53, without and with single-point mutations. Figure 1 shows the tetramerization domain of p53 for amino acids Gly-325 through Ala-355 for each of the four polypeptide chains. Each two of the four polypeptide chains form a dimer. Each of these dimers has a large hydrophobic interface, and the tetramer forms by binding the two dimers at their hydrophobic interfaces. Single-site mutations of amino acids at this hydrophobic interface are very likely to destabilize the tetramer, while, by choosing the proper mutant amino acid, keeping structural changes to each of the dimers at a minimum.

To study the stability of the tetramerization domain for the wild-type and the different mutations, we calculate free energies for different configurations of the tetramerization domain and estimate their dissociation energies. In particular, we use the high symmetry of the tetramerization domain to calculate the energy as function of the mutual separation d of the two dimers (the left and right one, respectively in Fig. 1), where d measures the distance of the two dimers along the direction perpendicular to their plane of symmetry (direction of x in Fig. 1). The existence of a minimum in the free energy and its depth and width will tell us about the stability of this domain. Single-point mutations on either site 340 or site 344 are performed by replacing the hydrophobic methionine or leucine with the same set of more hydrophilic amino acids (Table 1) that were used in an experimental study (C. H. Arrowsmith, private communication; Noolandi et al., 2000). These amino acids represent a spectrum of different hydrophobicity indices χ (Fauchère and Pliska, 1983; Eisenberg and McLachlan, 1986), but all of them are considered to be reasonably good α-helix formers as indicated by their helical potentials P_α (Chou and Fasman, 1974) (Table 1). As can be seen from Fig. 3, both

TABLE 1 List of amino acids used as replacements for Met-340 and Leu-344 with their respective hydrophobicity index χ (Eisenberg and McLahan, 1986) and helical potential P_α (Chou and Fasman, 1974)

Met-340			Leu-344		
Amino Acid	χ	P_α	Amino Acid	χ	P_α
Met	1.68	1.20	Leu	2.32	1.34
His	0.42	1.24	Ala	0.18	1.45
Ser	−0.05	0.79	Arg	−0.05	0.79
Gln	−0.30	1.17	Gln	−0.30	1.17
Glu	−0.87	1.53	Lys	−1.35	1.07
Asp	−1.05	0.98			
Lys	−1.35	1.07			

Met-340 and Leu-344 are located roughly at the center of the long axis of the tetramerization domain. Destabilizing the tetramer at this position promises to be most efficient.

By moving the two dimers perpendicular to their plane of symmetry, we can assume that the internal structure of the two dimers does not change. Therefore, we simplify our calculations by keeping all backbone residues fixed at their experimentally determined positions. Furthermore, we assume that the side chains not located at the dimer–dimer interface will keep their structure and position during the dimer–dimer separation, hence, we keep them fixed at their experimentally determined positions as well. There are 32 amino acids, each interacting with the opposite dimer (see Fig. 3 and Table 2). Furthermore, there are the four Phe-341 amino acids that are located at the center of the interface. Though none of these four amino acids interacts directly with the opposite dimer, they may play an important role for the tetramerization domain due to their hydrophobic character, which prefers to avoid direct contact with the polar solvent (water). This leaves us with 36 side chains to which

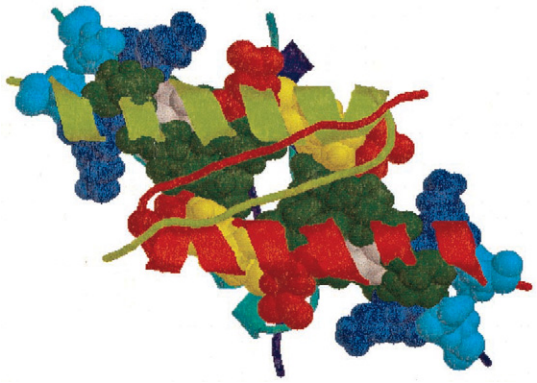

FIGURE 3 Top view of the tetramerization domain of p53 (amino acids Gly-325 through Ala-355) with all the amino acids at the dimer–dimer interface shown in a spacefilling mode. The four Phe-341 amino acids sit in the "hole" at the center of the interface region.

TABLE 2 List of all the amino acids at the dimer–dimer interface

Amino Acid	Position	χ	Color
Met	340	1.68	yellow
Glu	343	−0.87	bright red
Leu	344	2.32	green
Ala	347	0.42	dark grey
Leu	348	2.32	green
Leu	350	2.32	green
Lys	353	−1.35	blue
Gln	354	−0.30	cyan

The colors refer to Figs. 1 and 3.

we apply our meanfield approach to calculate their cms distribution and interaction energy with the solvent.

The calculation is done in several steps. First, we calculate the cms distribution of the 36 interface amino acid side chains by self-consistently solving Eqs. 5. In a second step, we map the cms densities onto volume fractions using Eqs. 7 and 8. The effective atoms are considered spherical with a radius $R_n = R_{ms,n}$ (cf. Appendix C), i.e., we chose R_n to be the radius of the minimal sphere into which the residuum n can be fitted (cf. Table 3). This results in a higher space-filling than observed in real proteins. In contrast, our procedure guarantees a complete overlap of the different amino acids within the α-helices, where one does not expect any solvent, but allows for sufficient interaction with the solvent at the interface as one would expect for the rather long side chains of Met, Lys, Glu, and Gln. The solvent volume fraction is obtained from Eq. 11 as

$$\phi_S(\mathbf{x}) = 1 - \sum_{n=1}^{N} \phi_n(\mathbf{x}). \tag{19}$$

Because the side chain cms distributions have been calculated without a constraint enforcing Eq. 11 at each space point, we set $\phi_S(\mathbf{x})$ to zero whenever $\Sigma_n \phi_n(\mathbf{x})$ exceeds 1. This guarantees that we get contributions to the free energy only at the solvent–side chain interface (where $0 < \phi_S(\mathbf{x}) < 1$), but not spurious ones from the interior of the protein. (It also means that the extensive and unphysical overlap of the side chains within the polypeptide chains introduced by our spherical approximation of the side chains only keeps the solvent out, but does not lead to any contributions to the free energy we try to estimate.)

RESULTS

The wild-type

The free energy calculated with the meanfield approach consists of two parts. The interprotein interactions are calculated self-consistently and represent the interactions between the different residues that are within contact distance. For the free energy as a function of dimer–dimer distance, we observe a strong increase in energy for small separations where the amino acids start to overlap. However, even for the wild-type, we do not observe a minimum in the free energy that is deep enough to account for the experimentally observed stability of this domain (Fig. 4 *A*) [*Note:* We are interested in relative changes of the free energy as a function of dimer–dimer distance d. For a better comparison, we move each of the free energy curves shown in Figs. 4–9 by a constant amount such that the free energy for large separations is zero.] The protein–solvent interactions add an enthalpic and an entropic contribution to the free energy.

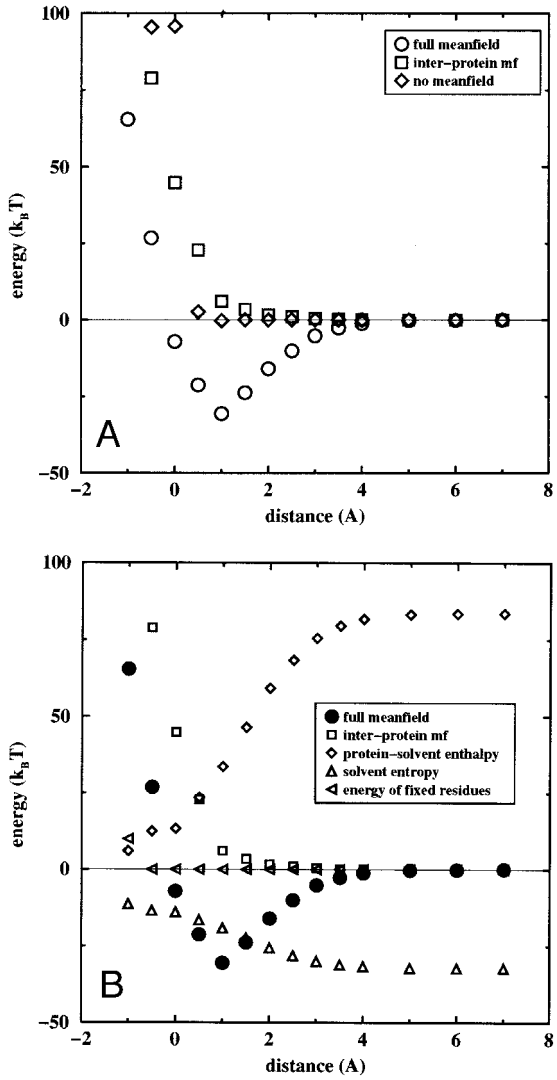

FIGURE 4 Reduced free energy $\beta\mathcal{F}$ of the p53 tetramerization domain as a function of dimer–dimer distance d. A distance of $d = 0$ corresponds to the experimental separation distance of the two dimers. (*A*) Comparison of the free energies calculated with the full meanfield approach, with the meanfield approach without protein–solvent interactions, and with no meanfield approach (i.e., all residues are kept at the experimentally determined positions), respectively. (*B*) The different contributions to the free energy calculated with the full meanfield approach.

The enthalpic part accounts for the hydrophilic/hydrophobic character of the amino acids in contact with the solvent and increases with increasing dimer–dimer distance. This is expected, because more and more of the hydrophobic amino acids become exposed to solvent during separation. The entropic term decreases with increasing d, but it changes much less than the enthalpic term. The sum of these two terms results in a net increase in solvent–protein interaction energy with increasing dimer–dimer distance (Fig. 4 *B*). The combination of interprotein and protein–solvent interactions leads to a minimum in the free energy of the wild-type tetramer of p53 that can account for the experimentally observed stability of this domain.

The meanfield approach introduces several length scales into the calculations. We first calculate the cms distributions $\rho_i(\mathbf{x})$ by self-consistently solving Eqs. 5. The cms densities $\rho_i(\mathbf{x})$ are expected to be well-localized functions around the average position of the effective atom. Hence, we can limit the space used for calculating the cms densities to a finite spherical volume centered at each effective atoms cms position. The effective atom-dependent radius of the spherical volumes are $C_V * R_{eff,i}$, where $R_{eff,i}$ is the effective radius of effective atom i defined in Appendix C. Figure 5 shows free energies for three different values of the constant C_V. Because the cms densities of the residues are very well localized, the free energy does not depend much on the change in the integration volume for $C_V > 1$. For all calculations presented here, we use a value of $C_V = 3.4$, if not mentioned otherwise.

We also introduce an excluded volume centered at each effective atom's cms position, into which no other effective atom is allowed to penetrate. This prevents the densities from overlapping too extensively with their neighbors. We have chosen a spherical volume with a radius of $C_X * R_{eff,i}$ to specify this region. Because of the localized nature of the residue's cms positions, we do not see any significant dependencies for values of $C_X < 0.5$, and chose a value of $C_X = 0.25$ for all the calculations presented in this report.

When we map the cms position density $\rho_i(\mathbf{r})$ into the volume fraction $\phi_i(\mathbf{r})$ of the residuum (cf. Eqs. 7 and 8), we use the radius $R_{ms,i} = C_{AA} * R_{eff,i}$ (cf. Appendix C). A value of $C_{AA} = 2.44$ has been estimated by using the size of the amino acids found in Lehninger (1977). Any change to C_{AA} has large effects on the free energy of the tetramerization domain, as shown in Fig. 6. By changing C_{AA}, the interaction energy with the solvent is modified, though it impacts the interactions most at larger distances. For all calculations presented here, we use a value of $C_{AA} = 2.44$, if not mentioned otherwise.

The interprotein interactions are defined with a finite interaction range (cf. Eq. 14). The free energy of the protein–solvent system depends significantly on the range of the interprotein interactions, both in the absolute values and in its dependence on the interdimer distance d. We have introduced a scaling factor C_{int} for the interprotein interaction cut-off distance (scaling of contact distances R_{ij} in Eq. 14). Figure 7 shows the free energy as a function of d for three different values of C_{int}. Only for values of C_{int} less than 1 can we obtain a minimum in the free energy that is close to $d = 0$, i.e., the experimentally observed interdimer distance. With increasing value of C_{int} the amino acids overlap much more and extend the repulsive part of the free energy to larger values of d. For all calculations presented here, we use a value of $C_{int} = 0.85$, if not mentioned otherwise.

The interaction parameters for the interprotein interactions and for the protein–solvent systems are determined by

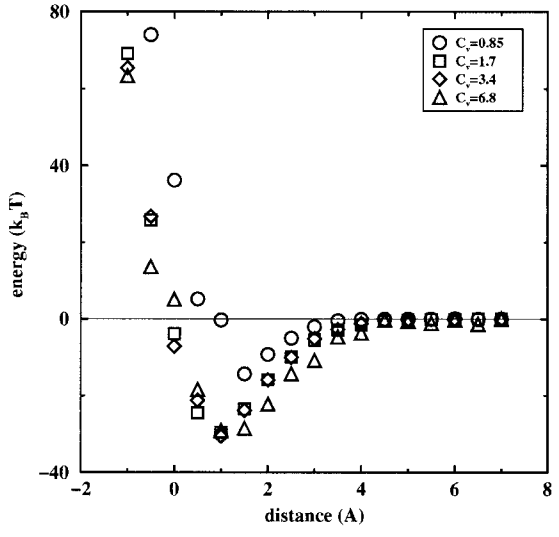

FIGURE 5 Reduced free energy $\beta\mathcal{F}$ of the p53 tetramerization domain as a function of dimer–dimer distance d for different values of the integration volume scale C_V.

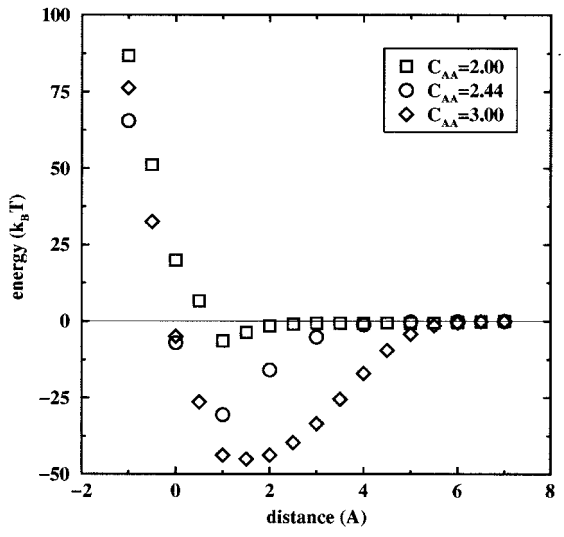

FIGURE 6 Reduced free energy $\beta\mathcal{F}$ of the p53 tetramerization domain as a function of dimer–dimer distance d for different values of the volume fraction mapping scale C_{AA}.

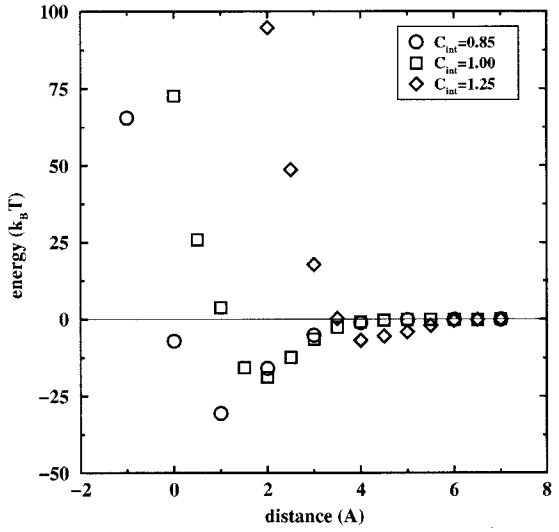

FIGURE 7 Reduced free energy $\beta\mathcal{F}$ of the p53 tetramerization domain as a function of dimer–dimer distance d for different values of the interaction cut-off distance scale C_{int}.

two completely different approaches (the interprotein interactions are statistically estimated from many proteins in the protein database, the protein–solvent interactions are estimated from solvability experiments of single amino acids). We introduce one additional factor that rescales the ratio of these two energy scales. In the current approach, we keep the interprotein interaction parameters as given in Kolinski and Skolnick (1992), Kolinski et al. (1993) and Vieth et al. (1994), but scale the protein–solvent interactions by a factor C_{PS}. Figure 8 shows the free energy for different values of C_{PS}. With increasing C_{PS}, the solvent–protein interactions become more dominant, deepening the minimum in the free energy. From experiments, we know the dissociation energy $E_{\text{D}} = 23$ kcal/mol of the wild-type, which translates into about $35k_{\text{B}}T$. From Fig. 8, we observe that a value of $C_{\text{PS}} = 6$ results in a minimum in the free energy of about $35k_{\text{B}}T$. We use this value for all the calculations presented here, if not otherwise mentioned.

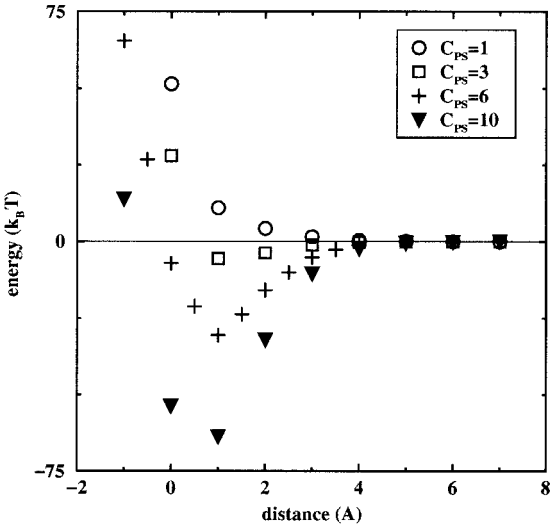

FIGURE 8 Reduced free energy $\beta\mathcal{F}$ of the p53 tetramerization domain as a function of dimer–dimer distance d for different values of the protein–solvent to interprotein energy ratio scale C_{PS}.

Mutations

Figure 9, *A* and *B*, show reduced free energies βF of the tetrameric region of p53 for different mutations of Met-340 and Leu-344 as function of the mutual distance d of the two dimers, respectively. d is the relative dimer–dimer distance compared to the wild-type experimental result ($d = 0$). If no

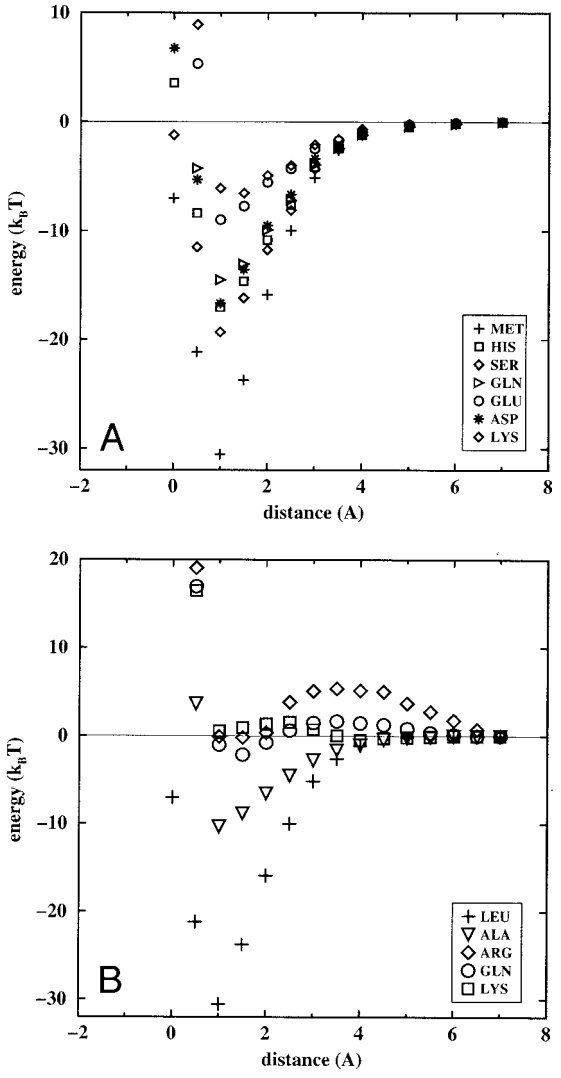

FIGURE 9 Reduced free energy $\beta\mathcal{F}$ as a function of the relative dimer–dimer distance for p53 mutants with energy ratio parameter $C_{\text{PS}} = 6$. (*A*) Mutations of Met-340; (*B*) Mutations of Leu-344. See the text for more details on the figure.

minimum in βF with respect to d is observed, the tetrameric state is not stable, and p53 forms dimers.

The calculations predict a decreasing stability of the tetrameric state for the following amino acids at residue 340: Met (tetramer) > Ser, Asp, His, Gln > Glu, Lys (tetramer), whereas experimental results (Noolandi et al., 2000) show the following order: Met (tetramer) > Ser > Gln > His, Lys > Asp, Glu (dimer). For residue 344, the calculated trend was Leu (tetramer) > Ala > Arg, Gln, Lys (dimer), and the experimental trend was Leu (tetramer) > Ala, Arg, Gln, Lys (dimer). It appears that the inclusion of protein–solvent interactions, which play an important role in the dimer–dimer dissociation, compensate for the coarse-grained description of the protein structure used in the meanfield calculations. The only major discrepancy between the simulation results and the experiments involves the Asp mutation at side 340: although the experiments show that this mutant dissociates into dimers, the meanfield calculation predicts stable tetrameric configurations for it. However, for the numerical calculations, we assume that all mutants retain the wild-type backbone structure. This might no longer be true for this mutant as indicated by its low melting temperatures and MD calculations (J. Wendling, private communication). We also find a discrepancy between the simulation results and the experiments involving the Lys mutation at residue 340, which is the least stable mutant according to the numerical calculations. This may be because lysine has a rather long and flexible sidechain with both hydrophobic and hydrophilic character. This dualism is not captured by our model that describes each amino acid sidechain by a single point coordinate and potential.

We also note that the calculations describe a system in infinite dilution (a single tetrameric molecule). Self association of proteins is highly concentration dependent and the relatively high concentrations used for CD and light scattering will drive the dimer–tetramer equilibrium toward the tetrameric state.

It is interesting to note that, although the trends in oligomeric state determined by meanfield calculations agree well with the experimental data, they do not directly correspond to the hydrophobicity index (Eisenberg and McLachlan, 1986). For example, His has a relatively high value on the hydrophobicity index (0.18) (compared to other mutations, cf. Table 1) yet His-340 is predicted to be more dimeric than Ser-340, which has a hydrophobicity value of -0.05. This trend is also observed experimentally, where His-340 is more dimeric than Ser-340 and Gln-340. In the case of His, the important feature is likely the relatively large and rigid shape of the His sidechain, which is not as easily accommodated at the dimer–dimer interface as Ser or Gln. Because each side chain in the meanfield calculation is treated as a single effective atom, sidechain geometrical constraints are not fully captured in the calculated free energy. However, the meanfield approach is still able to incorporate multiple properties of amino acid sidechain in most cases.

A comparison of the simulation and experimental (Noolandi et al., 2000) results suggests that p53tet is more sensitive to changes at residue 344 than at residue 340. For example, Lys-344 is a dimer with $T_M = 48°C$, but this same mutation at residue 340 has a T_m of 66°C and exists as an equilibrium mixture of dimers and tetramers. This difference in sensitivity can be rationalized by noting that the sidechains of residue 344 from all 4 subunits are closer to one another than those of residue 340 (Fig. 3). This will result in more charge repulsion at the dimer–dimer interface for Lys-344 than for Lys-340. The solvation of this interface in the dimer would relieve charge–charge repulsion between the two Lys sidechains in the dimer. It is also possible that the sidechain of Lys-340 can adopt a conformation such that the hydrophobic portion of the sidechain associates with the opposite dimer, while the charged portion has access to the solvated surface of the tetramer.

CONCLUSION

We have introduced a method to calculate protein–solvent interactions based on a meanfield approach to the partition function of the system. An advantage of this approach is the introduction of fields for the different residues of the system. This reduces the many solvent molecules needed for an atomistic calculation to a single field and allows the use of a Flory–Huggins type of contact interaction between the protein and the solvent. We have applied this method to study the stability of the tetramerization domain of the tumor suppressor p53 when subjected to single-point mutations at sites 340 or 344. Here, the secondary structure of the protein survives the mutation, but not necessarily the quaternary structure, leading to a dissociation of the protein into two dimers. This dissociation is mainly driven by the amino acid–solvent interactions, which prefer an exposed dimer–dimer interface for the more hydrophilic amino acids at sites 340 or 344.

Comparison with recent experiments on the stability of the p53 tetramerization domain shows that our meanfield representation of a solvent–protein system can predict the major features of a small self-associating protein system with reasonable accuracy.

Though we introduce several "free" parameters into our approach, we believe that our choice of parameters is general to other proteins besides p53: The length scales C_V (integration volume) and C_X (excluded volume around the center of each effective atom) have no significant impact on the meanfield calculation within the range defined in subsection, The wild-type. The length scale C_{AA} has been chosen such that the interface between the volume fraction and the surrounding solvent is at the location expected for the amino acid sidechains. Even though the free energy depends on the explicit value of this parameter, we believe that our choice is physically sound and general (at this level of approximation) for all proteins. The parameter C_{int} used to scale the range of the interprotein interactions has been

chosen such that we can obtain reasonable values of the interdimer distance of the wild-type tetramer. The value of $C_{int} = 0.85$ is in agreement with observations of other groups that use the Skolnick interaction parameters in their (continuous space) calculations (F. Cohen, private communication). The energy scale C_{PS} for the protein–solvent interactions is the only parameter that has been obtained using explicit data for p53 by fitting the calculated dimer–dimer dissociation energy to the experimental result. However, using this set of parameters for calculating dissociation energies of other quaternary proteins that are bound together primarily by hydrophobic interactions also gives good agreements with experimental results (e.g., for streptavidin, Coussaert et al., 2000).

Overall, the results presented here are encouraging for the application of the meanfield method toward other more complicated protein–solvent systems. Also, using more sophisticated intraprotein interaction schemes, this method is readily extended to situations where the secondary structure might also change significantly with the exchange of amino acids along the protein backbone.

APPENDIX A: DENSITY FUNCTIONAL APPROACH

Using the densities defined in Eqs. 2 and 3, we can rewrite the integrand of Eq. 1 as

$$e^{-\beta V} = \int \prod_{\nu=1}^{N+1} \mathcal{D}[\rho_\nu(\mathbf{x})]\delta(\rho_\nu(\mathbf{x}) - \hat{\rho}_\nu(\mathbf{x}))e^{-\beta W[\rho_\nu]}, \quad (A1)$$

where we introduced a single index ν to address the effective atoms ($\nu = 1, \ldots, N$) and the solvent ($\nu = N + 1$) together.

If we assume single- (W) and two-body interactions (V_{nm}) within the effective atoms, and two-body interactions between the solvent and the effective atoms (V_{nS}), we have

$$\begin{aligned}\beta W[\rho_\nu] = &\frac{1}{2}\sum_{n,m=1}^{N}\int d\mathbf{x}\, d\mathbf{x}'\rho_n(\mathbf{x})\beta V_{nm}(\mathbf{x}-\mathbf{x}')\rho_m(\mathbf{x}') \\ &+ \sum_{n=1}^{N}\int d\mathbf{x}\rho_n(\mathbf{x})\beta W_n(\mathbf{x}) \\ &+ \sum_{n=1}^{N}\int d\mathbf{x}\, d\mathbf{x}'\rho_n(\mathbf{x})\beta V_{nS}(\mathbf{x}-\mathbf{x}')\rho_S(\mathbf{x}'). \quad (A2)\end{aligned}$$

Rewriting these δ-functions, we introduce conjugate fields $\omega_\nu(\mathbf{x})$ for each density $\rho_\nu(\mathbf{x})$,

$$\begin{aligned}&\delta(\rho_\nu(\mathbf{x}) - \hat{\rho}_\nu(\mathbf{x})) \\ &\quad = \int \mathcal{D}[\omega_\nu(\mathbf{r})]\exp\left\{\int d\mathbf{x}\omega_\nu(\mathbf{x})(\rho_\nu(\mathbf{x}) - \hat{\rho}_\nu(\mathbf{x}))\right\}. \quad (A3)\end{aligned}$$

By defining

$$Q_\nu = \int d\mathbf{x}_\nu \exp\left\{-\int d\mathbf{x}\omega_\nu(\mathbf{x})\hat{\rho}_\nu(\mathbf{x})\right\} = \int d\mathbf{x}_\nu e^{-\omega_\nu(\mathbf{x}_\nu)}, \quad (A4)$$

we finally obtain the partition function,

$$Z = \int \prod_{\nu=1}^{N+1}\{\mathcal{D}[\rho_\nu]\mathcal{D}[\omega_\nu]\}e^{-\beta\mathcal{F}[\rho_\nu,\omega_\nu]}, \quad (A5)$$

with

$$\begin{aligned}&\beta\mathcal{F}[\rho_\nu, \omega_\nu] \\ &\quad = \sum_{\nu=1}^{N+1}\left\{\ln\left(\frac{\Lambda_\nu^3}{Q_\nu}\right) - \int d\mathbf{x}\rho_\nu(\mathbf{x})\omega_\nu(\mathbf{x}) + \beta W[\rho_\nu(\mathbf{x})]\right\}. \quad (A6)\end{aligned}$$

In the case where a subset $N_c < N$ of effective atoms is kept fixed at positions $\{\mathbf{x}_{k0}\}$, $k = 1, \ldots, N_c$, we start with the partition function

$$\begin{aligned}Z = &\prod_{k,l=1}^{N_c}\exp\left(-\frac{1}{2}\beta V_{kl}(\mathbf{x}_{k0} - \mathbf{x}_{l0})\right) \\ &\times\prod_{k=1}^{N_c}\exp(-\beta W_k(\mathbf{x}_{k0}))\mathcal{Z}_d \\ &\times\int\prod_{n=1}^{N_d}d\mathbf{x}_n\prod_{j=1}^{N_s}d\mathbf{r}_j\exp(-\beta\tilde{V}(\{\mathbf{x}_n\},\{\mathbf{r}_j\})), \quad (A7)\end{aligned}$$

where

$$\begin{aligned}&\tilde{V}(\{\mathbf{x}_n\},\{\mathbf{r}_j\})) \\ &\quad = \frac{1}{2}\sum_{n,m=1}^{N_d}V_{nm}(\mathbf{x}_n - \mathbf{x}_m) + \sum_{n=1}^{N_d}W_n(\mathbf{x}_n) \\ &\qquad + \sum_{n=1}^{N_d}\sum_{k=1}^{N_c}V_{nk}(\mathbf{x}_n - \mathbf{x}_{k0}) + \sum_{n=1}^{N_d}\sum_{j=1}^{N_s}V_{nS}(\mathbf{x}_n - \mathbf{r}_j) \\ &\qquad + \sum_{k=1}^{N_c}\sum_{j=1}^{N_s}V_{kS}(\mathbf{x}_{k0} - \mathbf{r}_j), \quad (A8)\end{aligned}$$

and $N_d = N - N_c$. The interactions of the effective atoms in subset N_d with those of subset N_c (the third term on the right-hand side of Eq. A8) and those of the effective atoms in subset N_c with the solvent (last term on the right-hand side of Eq. A8) are effective single-body energies and can be included into $W_k(\mathbf{x}_k)$. This leaves us with the same type of partition function as discussed above with the only change that N is replaced by the number of effective atoms in subset N_d.

APPENDIX B: PARTITION FUNCTION FOR PARTICLES WITH INTERNAL DEGREES OF FREEDOM

In this Appendix, we will consider a system of N particles with internal degrees of freedom. In particular, we assume particle n to have T_n internal

degrees of freedom, each of which is assumed with the position-dependent probability $f_{nk}(\mathbf{x})$, $k = 1, \ldots, T_n$. The partition function for this type of system is

$$Z = \mathcal{Z} \int \prod_{n=1}^{N} d\mathbf{x}_n \sum_{\{t_n\}} f(\{t_n\}) \exp(-\beta V(\{\mathbf{x}_n\}, \{t_n\})), \qquad \text{(B1)}$$

where $\Sigma_{\{t_n\}}$ denotes the sum over all possible permutations of the internal degrees of freedom, and $f(\{t_n\})$ is the total probability of permutation $\{t_n\}$.

Following the field theoretical approach described in Appendix A, we introduce densities

$$\hat{\rho}_{nk}(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{x}_n)\delta_{t_n k} \qquad \text{(B2)}$$

for each degree of freedom for each particle. Using the same steps as before, we can rewrite the partition function in generalized densities $\rho_{nk}(\mathbf{x})$ and their conjugate fields $\omega_{nk}(\mathbf{x})$ as

$$Z = \int \prod_{\nu=1}^{N+1} \prod_{k=1}^{T_n} \{\mathcal{D}[\rho_{nk}]\mathcal{D}[\omega_{nk}]\} \exp(-\beta \mathcal{F}[\rho_\nu, \omega_\nu]), \qquad \text{(B3)}$$

with the free energy functional

$$\beta \mathcal{F}[\rho_{nk}, \omega_{nk}] = \beta W[\rho_{nk}] - \sum_{n=1}^{N} \sum_{k=1}^{T_n} \int d\mathbf{x} \rho_{nk} \omega_{nk} + \sum_{n=1}^{N} \ln(\Lambda_n^3/\hat{Q}_n). \qquad \text{(B4)}$$

The integrals

$$\hat{Q}_n = \int d\mathbf{x}\, e^{-\omega_{eff,n}(\mathbf{x})} \qquad \text{(B5)}$$

contain now the weighted sum over all the conjugate fields of one particle

$$\omega_{eff,n}(\mathbf{x}) = -\ln\left(\sum_{k=1}^{T_n} f_k(\mathbf{x}) e^{-\omega_{nk}(\mathbf{x})} \right). \qquad \text{(B6)}$$

Within the meanfield approximation, the fields are given by

$$\omega_{nk}(\mathbf{x}) = \frac{\delta \beta W[\rho_{nk}]}{\delta \rho_{nk}(\mathbf{x})}$$
$$\rho_{nk}(\mathbf{x}) = \frac{e^{-\omega_{eff,n}(\mathbf{x})}}{\hat{Q}_n} \frac{f_k(\mathbf{x}) e^{-\omega_{nk}(\mathbf{x})}}{\sum_{l=1}^{T_n} f_l(\mathbf{x}) e^{-\omega_{nl}(\mathbf{x})}}. \qquad \text{(B7)}$$

In the special case when the interactions $W[\rho_{nk}]$ are independent of the internal degrees of freedom of the particles, the conjugate fields are all the same, i.e., $\omega_{nk}(\mathbf{x}) \equiv \omega_n(\mathbf{x})$, and are given as

$$\omega_n(\mathbf{x}) = \frac{\delta \beta W[\rho_{nk}]}{\delta \rho_n(\mathbf{x})}, \qquad \text{(B8)}$$

where

$$\rho_n(\mathbf{x}) = \sum_{k=1}^{T_n} \rho_{nk}(\mathbf{x}). \qquad \text{(B9)}$$

Furthermore, by defining

$$S_n(\mathbf{x}) = \sum_{k=1}^{T_n} f_{nk}(\mathbf{x}), \qquad \text{(B10)}$$

we obtain

$$\omega_{eff,n}(\mathbf{x}) = \omega_n(\mathbf{x}) - \ln S_n(\mathbf{x}) \qquad \text{(B11)}$$

and

$$\rho_n(\mathbf{x}) = \frac{e^{-\omega_n(\mathbf{x})}}{\hat{Q}_n}. \qquad \text{(B12)}$$

The Helmholtz free energy (Eq. 6) finally becomes

$$\beta \mathcal{F}[\rho_{nk}, \omega_{nk}] = \sum_{n=1}^{N} \int d\mathbf{x} \rho_n(\mathbf{x}) \left[\ln \rho_n(\mathbf{x}) + \frac{\omega_n(\mathbf{x}) + \beta W_n(\mathbf{x}) - 2 \ln S_n(\mathbf{x})}{2} \right]. \qquad \text{(B13)}$$

This system is equivalent to one with no internal degrees of freedom of the particles, but with the renormalized single-particle energy,

$$\beta W_n(\mathbf{x}) \rightarrow \beta W_n(\mathbf{x}) - \ln S_n(\mathbf{x}), \qquad \text{(B14)}$$

as can be seen directly from the initial partition function Eq. B1, when the $f_k(\mathbf{x})$ are added to the interactions in the exponent.

APPENDIX C: EFFECTIVE ATOMIC RADII

For each effective atom, we define a radius $R_{eff,i}$, (i counts the 20 amino acid side chains plus the backbone) as the radius of a sphere that has the same volume as all the heavy atoms (C, N, O, S) of the effective atom

$$R_{eff,i} = \left\{ \sum_{\text{heavy atoms (ha)}} R_{ha,i}^3 \right\}^{1/3}. \qquad \text{(C1)}$$

This radius correlates very well with the radius $R_{ms,i}$ of the minimal sphere needed to fit in effective atom i as estimated from size data given in Lehninger (1977) (Table 3). The ratio C_{AA} of the $R_{ms,i}$ and the effective

TABLE 3 List of the effective (R_{eff}) and "minimal sphere" (R_{ms}, Appendix C) radii of the residues in units of Å

Residue	R_{eff}	R_{ms}	Residue	R_{eff}	R_{ms}	Residue	R_{eff}	R_{ms}
Cyx*	1.15	3.2	Gly†	—	—	Ala	0.77	1.8
Ser	0.88	2.1	Cys	1.16	2.5	Val	1.11	2.8
Thr	1.04	2.5	Ile	1.23	3.2	Pro	1.11	3.35
Met	1.36	3.05	Asp	1.11	2.5	Asn	1.09	2.65
Leu	1.23	3.05	Lys	1.26	2.95	Glu	1.22	2.9
Gln	1.21	2.9	Arg	1.33	3.75	His	1.29	3.05
Phe	1.48	3.5	Tyr	1.51	3.35	Trp	1.63	4.15

*The symbol Cyx denotes the backbone effective atom.
†Because the side chain of glycine consists of a single hydrogen only no effective radius has been calculated.

radii $R_{\mathrm{eff,i}}$ is about the same for each effective atom, and its averaged value is

$$C_{\mathrm{AA}} = \frac{1}{N_{\mathrm{R}}} \sum_{\mathrm{i}=1}^{N_{\mathrm{R}}} \frac{R_{\mathrm{ms,i}}}{R_{\mathrm{eff,i}}} = 2.44 \pm 0.18, \qquad \text{(C2)}$$

where $N_{\mathrm{R}} = 21$ is the number of the possible effective atoms.

We use the $R_{\mathrm{eff,i}}$ as the basic size scale for each of the effective atoms, and use constant factors (which are the same for all effective atoms) to calculate specific size scales, such as integration volumes or excluded volumes.

This work was supported by grants from the Human Frontier Science Program (J.N.). One of us (A.R.V.) acknowledges useful discussions with Ulrich Suter and Jürgen Wendling from the Eidgenössische Technische Hochschule (ETH) Zürich and Tamara Coussaert from the Universite Libre de Bruxelles.

REFERENCES

Carlacci, L., S. W. Englander. 1996. Loop problem in proteins: developments on the Monte Carlo simulated annealing approach. *J. Comp. Chem.* 17:1002–1012.

Cheng, B., A. Nayeem, H. A. Scheraga. 1996. From secondary structure to three-dimensional structure: improved dihedral angle probability distribution function for use with energy searches for native structures of polypeptides and proteins. *J. Comp. Chem.* 17:1453–1461.

Chou, P. Y., G. D. Fasman. 1974. Prediction of protein conformation. *Biochemistry.* 13:222–245.

Clore, G. M., J. G. Omichinski, K. Sakaguchi, N. Zambrano, H. Sakamoto, E. Apella, A. M. Gronenborn. 1994. High resolution structure of the oligomerization domain of p53 by multidimensional NMR. *Science.* 265:386–391.

Clore, G. M., J. Ernst, R. Clubb, J. G. Omichinski, W. M. P. Kennedey, K. Sakaguchi, E. Apella, A. M. Gronenborn. 1995a. Refined solution structure of the oligomerization domain of the tumor suppressor p53. *Nature Struct. Biol.* 2:321–327.

Clore, G. M., J. G. Omichinski, K. Sakaguchi, N. Zambrano, H. Sakamoto, E. Apella, A. M. Gronenborn. 1995b. Interhelical angles in the solution structure of the oligomerization domain of p53: correction. *Science.* 267:1515–1516.

Coussaert, T., A. R. Völkel, J. Noolandi, A. P. Gast. 2000. Streptavidin tetramerization and 2D crystallization: a mean field approach. *Biophys. J.* in press.

Creighton, T. E. 1992. Protein Folding. W. H. Freeman and Company, New York.

Daggett, V., M. Levitt. 1993. Realistic simulations of native-protein dynamics in solution and beyond. *Annu. Rev. Biophys. Struct.* 22:353–380.

Dunbrack, R. L., Jr., M. Karplus. 1993. Backbone-dependent rotamer library for proteins: application to side-chain prediction. *J. Mol. Biol.* 230:543–547.

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

Fauchère, J.-L., 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.

Flory, P. J. 1953. Principles of Polymer Chemistry. Cornell University Press, Ithaca, NY.

Friend, S. 1994. p53: a glimpse at the puppet behind the shadow play. *Science.* 265:334–336.

Godzik, A., A. Kolinski, J. Skolnick. 1995. Are proteins ideal mixtures of amino acids? Analysis of energy parameter sets. *Protein Sci.* 4:2107–2117.

Hinds, D. A., M. Levitt. 1994. Exploring conformational space with a simple lattice model for protein structure. *J. Mol. Biol.* 243:668–682.

Hong, K. M., J. Noolandi. 1981. Theory of inhomogeneous multicomponent polymer systems. *Macromol.* 14:727–736.

Jeffrey, P. D., S. Gorina, N. P. Pavletich. 1995. Crystal structure of the tetramerization domain of the p53 tumor suppressor at 1.7 angstroms. *Science.* 267:1498–1502.

Karplus, M., D. L. Weaver. 1994. Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Sci.* 3:650–668.

Karplus, M., A. Šali. 1995. Theoretical studies of protein folding and unfolding. *Curr. Opin. Struct. Biol.* 5:58–73.

Koehl, P., M. Delaru. 1994. Application of a self-consistent mean field theory to predict protein side-chain conformation and estimate their conformational energy. *J. Mol. Biol.* 239:249–275.

Koehl, P., M. Delaru. 1995. A self consistent mean field approach to simultaneous gap closure and side-chain positioning in homology modelling. *Nature Struct. Biol.* 2:163–170.

Koehl, P., M. Delaru. 1996. Mean-field minimization methods for biological macromolecules. *Curr. Opin. Struct. Biol.* 6:222–226.

Kolinski, A., A. Godzik, 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.

Kolinski, A., J. Skolnick. 1992. Discretized model of proteins. I. Monte Carlo study of cooperativity in homopolypeptides. *J. Chem. Phys.* 97: 9412–9462.

Kolinski, A., J. Skolnick. 1994. Monte Carlo simulations of protein folding: II. Application to protein A, ROP, and Crambin. *Proteins: Struct. Funct. Genet.* 18:338–352.

Lau, K. F., K. A. Dill. 1989. A lattics statistical mechanics model of the conformational and sequence spaces of proteins. *Macromol.* 22: 3986–3997.

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

Lee, C., M. Levitt. 1991. Accurate prediction of the stability and activity effects of site-directed mutagenesis on a protein core. *Nature.* 352: 448–451.

Lee, C. 1994. Predicting protein mutant energetics by self consistent ensemble optimisation. *J. Mol. Biol.* 236:918–939.

Lee, W., T. S. Harvey, Y. Yin, P. Yau, D. Litchfield, C. H. Arrowsmith. 1994. NMR solution structure of the tetrameric minimum transforming domain of p53. *Nature Struct. Biol.* 1:877–890.

Lehninger, A. L. 1977. Biochemistry: The Molecular Basis of Cell Structure and Function. Worth Publishers, Inc., New York.

Levinthal, C. 1969. Levinthal's paradox. *In* Sci. Am. Mössbauer Spectroscopy in Biological Systems, P. Debrunner, J. C. M. Tsibris, E. Münck, editors. University of Illinois Press, Urbana, IL. 22–24.

Levitt, M. 1996. Theory and simulation: through the breach. *Curr. Opin. Struct. Biol.* 6:193–194.

Lim, W. A., R. T. Sauer. 1991. The role of internal packing interactions in determining the structure and stability of a protein. *J. Mol. Biol.* 219: 359–376.

Miller, R., C. A. Danko, M. J. Fasolka, A. C. Balazs, H. S. Chan, K. A. Dill. 1992. Folding kinetics of proteins and copolymers. *J. Chem. Phys.* 96:768–780.

Noolandi, J., K. M. Hong. 1982. Interfacial properties of immiscible homopolymer blends in the presence of block copolymers. *Macromol.* 15:482–492.

Noolandi, J., T. S. Davison, A. R. Völkel, X.-F. Nie, C. Kay, 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. USA.* 97:9955–9960.

Pappu, R. V., W. J. Schneller, D. L. Weaver. 1996. Electrostatic multipole representation of a polypeptide chain: an algorithm for simulation of polypeptide properties. *J. Comp. Chem.* 17:1033–1055.

Ponders, J. W., F. M. Richards. 1987. Tertiary templates for proteins. *J. Mol. Biol.* 193:775–791.

Roitberg, A., R. Elbers. 1991. Modeling side chains in peptides and proteins: application to the locally enhanced sampling and the simulated

annealing methods to find minimum energy conformations. *J. Chem. Phys.* 95:9277–9287.

Shakhnovich, E., G. Farztdinov, A. M. Gutin, M. Karplus. 1991. Protein folding bottlenecks: a lattice Monte Carlo simulation. *Phys. Rev. Lett.* 67:1665–1668.

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

Tuffery, P., C. Etchbest, S. Hazout, R. Lavery. 1991. A new approach to the rapid determination of protein side-chain conformation. *J. Biomol. Struct. Dynam.* 8:1267–1289.

Tuffery, P., C. Etchbest, S. Hazout, R. Lavery. 1993. A critical comparison of search algorithms applied to the optimization of protein side-chain conformations. *J. Comp. Chem.* 14:790–798.

Ullner, M., C. E. Woodward, B. Jönsson. 1996. A Debye–Hückel theory for electrostatic interactions in proteins. *J. Chem. Phys.* 105:2056–2065.

Vásquez, M. 1996. Modeling side-chain conformation. *Curr. Opin. Struct. Biol.* 6:217–221.

Vieth, M., A. Kolinski, C. L. Brooks, III, J. Skolnick. 1994. Prediction of the folding pathways and structure of the GCN4 leucine zipper. *J. Mol. Biol.* 237:361–367.