Solvation of Proteins: Linking Thermodynamics to Geometry

Hendrik Hansen-Goos,[1,2] Roland Roth,[1,2] Klaus Mecke,[3] and S. Dietrich[1,2]

[1]*Max-Planck-Institut für Metallforschung, Heisenbergstraße 3, 70569 Stuttgart, Germany*
[2]*Institut für Theoretische und Angewandte Physik, Universität Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany*
[3]*Institut für Theoretische Physik, Universität Erlangen-Nürnberg, Staudtstraße 7, 91058 Erlangen, Germany*

(Received 1 June 2007; published 17 September 2007)

We calculate the solvation free energy of proteins in the tube model of Banavar and Maritan [Rev. Mod. Phys. **75**, 23 (2003)] using morphological thermodynamics which is based on Hadwiger's theorem of integral geometry. Thereby we extend recent results by Snir and Kamien [Science **307**, 1067 (2005)] to hard-sphere solvents at finite packing fractions and obtain new conclusions. Depending on the solvent properties, parameter regions are identified where the β sheet, the optimal helix, or neither is favored.

DOI: 10.1103/PhysRevLett.99.128101 PACS numbers: 87.14.Ee, 05.70.−a, 61.20.−p

Understanding protein function requires the knowledge of native structures. This has for a long time motivated researchers to attempt the prediction of native structures of proteins in their cellular environment based on the given amino acid sequences. But to date no entirely satisfactory solution of the problem has been found [1]. Stunningly, tertiary structures of proteins in living organisms can be reduced to an estimated number of only about 1000 basic protein folds [2] which are composed according to a set of ''constructional rules'' and which are characterized by a particular stability against mutation. Backed by these observations it has been suggested to consider the basic folds as primary natural forms obeying physical laws in the spirit of a Platonic model of life [3]. Guided by this idea of certain robust motifs in protein folding, Banavar and Maritan introduced a simple geometrical model for folding which reproduces many of the basic building blocks of real proteins [4]. In their model, the protein backbone is viewed as an impenetrable, flexible tube with finite thickness. Locally, this results in an effective three body interaction due to the limited local curvature and, globally, in a specific interaction of *cylindrical* segments. While the first feature could also be modeled by a chain of tethered spheres with an appropriate three body potential, the second is genuine to the tube model. To complete the model, an attractive potential acting between different parts of the tube is introduced mimicking the effect of hydrophobic amino acids. These ingredients are sufficient to drive the model protein into the marginally compact regime where it displays conformations which resemble closely basic folds of real proteins. As the model does not include any chemical details, the results figure indeed as ''Platonian folds.''

Here we devise an efficient method for the calculation of solvent effects, which are crucial for protein folding [5], in particular, due to solvent *entropy* [6]. To this end we combine Hadwiger's theorem [7] of integral geometry with density functional theory (DFT) for classical fluids [8]. Our modified tube model allows us to include *directly* the cellular medium *without* resorting to the effective attractive potential of the Banavar-Maritan model. We consider mainly the purely entropic hard-sphere solvent and thereby extend a recent study by Snir and Kamien [9,10] to solvents beyond the limit of vanishing density. Depending on the solvent properties, conditions are discerned for which a tightly packed helix or the β sheet (see below) is favored. As an outlook we mention results for a simple solvent with intermolecular attraction.

Consider a body B corresponding to a closed and bounded set in $\mathbb{R}^3$ which, at a later stage, will represent the protein immersed in the solvent (see Fig. 1). We characterize B by its four Minkowski measures: the volume $V(B)$, the surface area $A(B)$, the integral mean curvature $C(B) = \int_{\partial B} \frac{1}{2}(\varkappa_1 + \varkappa_2) dA$, and the integral Gaussian curvature $X(B) = \int_{\partial B} \varkappa_1 \varkappa_2 dA$, where $\varkappa_1$ and $\varkappa_2$ are the local principal curvatures. $X(B)/4\pi$ equals the Euler characteristic χ of B. For instance, $\chi = 1$ for all B which are topologically equivalent to a sphere. The measures $M = V, A, C, X$ share the following properties [7]: (i) *motion invariance*, i.e., for every rotation and translation g one has $M(gB) = M(B)$; (ii) *conditional continuity*, i.e., for every sequence of convex bodies B_n which converges (with respect to the Hausdorff metric) to B for $n \to \infty$ one has that $M(B_n) \to M(B)$; and (iii) *additivity*, i.e., for the union

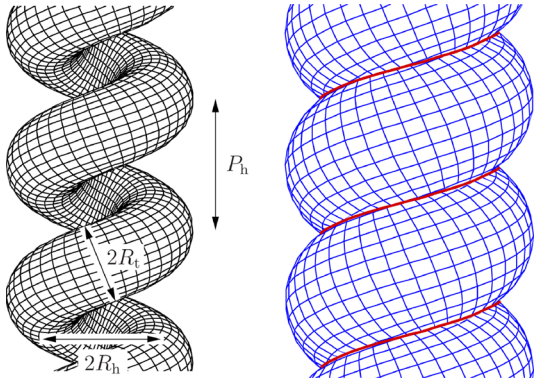


FIG. 1 (color online). Left: Helical conformation of a protein in the tube model. Right: A possible parallel surface for the same conformation. Because of self-intersection one or several helical intersection lines appear (in red; thicker lines) giving rise to additional contributions to C and X, and hence to F_{sol}.

$B_1 \cup B_2$ of two bodies B_1 and B_2 one has $M(B_1 \cup B_2) = M(B_1) + M(B_2) - M(B_1 \cap B_2)$ where $B_1 \cap B_2$ is the intersection of B_1 and B_2. The remarkable result obtained by Hadwiger is that *every* functional $\varphi(B)$ with properties (i), (ii), and (iii) is of the form $\varphi(B) = c_V V(B) + c_A A(B) + c_C C(B) + c_X X(B)$ with constant coefficients c_V, c_A, c_C, and c_X [7].

In the mid-1990s Hadwiger's theorem was introduced for the first time to the physics of complex fluids in a study of microemulsions [11]. Recently, detailed quantitative analyses have introduced the approach into the field of hard-sphere fluids [12]. The idea of this so-called morphological thermodynamics is to identify a thermodynamic quantity, which in the present context is the solvation free energy F_{sol}, with the above functional φ. In a physically motivated notation for the coefficients this reads

$$F_{\mathrm{sol}} = pV + \sigma A + \kappa C + \bar{\kappa} X. \tag{1}$$

Here V, A, C, and X are the geometric measures attributed to a protein in a given conformation. The thermodynamic coefficients are the solvent pressure p, the planar surface tension σ, the bending rigidity κ, and $\bar{\kappa}$ which couples to the topological invariant X. According to Hadwiger's theorem the coefficients do not depend on the protein conformation. They encode solvent properties depending on the temperature, the chemical potential, and, except for the pressure p, on the specific interaction between the solute and the solvent.

The morphological form [Eq. (1)] makes computations of F_{sol} very efficient even for complex protein conformations as it allows one to treat solvent and protein properties separately instead of performing time-consuming calculations for the solvent in complex protein geometries [13]. The applicability of Eq. (1) depends on whether F_{sol} obeys additivity [property (iii) of Hadwiger's theorem] in the physical system. It becomes invalid for highly confined portions of the solvent where additivity would basically imply that different segments of a protein would not experience any solvent mediated force. Equation (1) is also inappropriate for fluids with long-ranged correlations, e.g., near a critical point or at wetting and drying transitions [14]. However, for the solvents considered here correlations decay rapidly, i.e., within a few particle diameters. Therefore the assumption of additivity is expected to be a reliable approximation. In order to estimate the corresponding error, we have calculated the grand potential of a hard-sphere fluid (radius R, packing fraction $\eta = 0.38$) *inside* a cylinder (radius R_{cyl}) with DFT. We used the White-Bear functional Mark II [15] which is a recent version of fundamental measure theory [16] known to compare very well with simulation data. The overall agreement of the DFT results with Eq. (1) is excellent; only for $R_{\mathrm{cyl}} < 5R$ the DFT data start to oscillate around the morphological prediction due to packing effects. For a quantity as sensitive as the surface tension the amplitude of these oscillations is, however, below 20% even for strong confinement at $R_{\mathrm{cyl}} \simeq 2R$. Thus deviations from Eq. (1) do occur in extreme confinement but these are sufficiently small and do not affect the leading behavior. Previously, Eq. (1) has been shown to perform excellently for solutes of simple convex shapes [12].

We have calculated the measures V, A, C, and X for a broad range of protein conformations in the tube model covering two principal secondary motifs, namely, helices and β sheets. Helical tube conformations are given by all the points with a distance smaller than or equal to R_t from the centerline given by $(x, y, z) = (R_h \cos\tau, R_h \sin\tau, P_h\tau/2\pi)$, where $\tau \in (-\infty, +\infty)$ and P_h and R_h are the helix pitch and radius, respectively (see the left part of Fig. 1). Because of self-intersection constraints not all combinations of R_t, P_h, and R_h are admissible. The configurations belonging to the boundary of the admissible parameter region give rise to two regimes of close-packed helices [17]. The turn-to-turn distance limited close-packed (TTCP) regime is reached for given $R_h > R_h^* \simeq 0.8622R_t$ by minimizing P_h such that consecutive turns of the helix touch each other. For $R_h < R_h^*$ the radius of curvature of the TTCP helix centerline becomes smaller than R_t which would cause the tube to bend so strongly that it self-intersects locally. In this curvature limited close-packed (CCP) regime P_h must be chosen larger than according to the TTCP constraint in order to maintain the centerline curvature equal to (rather than above) $1/R_t$. The close-packed helix with $R_h = R_h^*$, which has $P_h = P_h^* \simeq 2.512R_h^*$, is termed optimal as it appears in packing problems of tubes subject to local compactness conditions [18]. Its geometry is closely related to that of helical conformations of many actual proteins [18]. If contributions from turns are neglected β sheets correspond to neighboring sections of the tube having straight centerlines and being aligned parallel. This is realized in the limit $R_h \to \infty$ in which the curvature of the centerline vanishes.

For the calculation of V, A, C, and X the parallel surface at the distance R (solvent particle radius) from the actual protein surface is the most suitable as it indicates the solvent-accessible regions (see the right-hand part of Fig. 1). Because of self-intersection the parallel surface is rather complicated such that the calculations of the geometric measures have to be carried out numerically. Intersection lines give rise to additional curvature contributions which can be calculated analogously to the case of intersecting spheres [19]: $C_{\mathrm{line}} = -(\pi/2 - \theta/2)\ell$ and $X_{\mathrm{line}} = -2\ell\varkappa_{\mathrm{line}}\cos(\theta/2)$. Here θ is the opening angle of the groove associated with the intersection line and $\varkappa_{\mathrm{line}}$ denotes the curvature of the intersection line of length ℓ.

Results for V, A, and C of the close-packed helices are shown in Fig. 2. X vanishes as the parallel body of a helix is topologically equivalent either to a solid or a hollow cylinder. We have subtracted the (R dependent) values of the volume V_β and of the surface area A_β corresponding to a β sheet. These are approached in the limit $R_h \to \infty$. For C one has $C_\beta \equiv 0$. The figure focuses on the TTCP regime

($R_h > R_h^*$). Only the onset of the CCP regime ($R_h < R_h^*$) is shown. Because of the rather strong increase of P_h in the CCP regime the geometric measures rapidly approach V_{str}, A_{str}, and C_{str} which are the values for a stretched tube ($R_h = 0$, $P_h \neq 0$). The values $\Delta V_{\mathrm{str}}^{\beta} = (V_{\mathrm{str}} - V_\beta)/(R_t^3 \tilde{L})$ and $\Delta A_{\mathrm{str}}^{\beta} = (A_{\mathrm{str}} - A_\beta)/(R_t^2 \tilde{L})$ are tabulated in the figure [$C_{\mathrm{str}}/(R_t\tilde{L}) = \pi$]. $\tilde{L} = L/R_t$ is the dimensionless length of the tube; contributions from the end points of the tube are neglected ($L \to \infty$). For $R < R_{\mathrm{uw}} \simeq 0.0835 R_t$, the curves for V have their global minimum at $R_h = R_h^*$. At $R = R_{\mathrm{uw}}$ the minimum starts to move to $R_h > R_h^*$. In the curves for A, if $R > R_s \simeq 0.0465 R_t$, an almost linear decrease is connected via a cusp to an interval in which A increases towards A_β. Exactly at the position of these cusps, the curves for C are discontinuous such that $C/(R_t\tilde{L})$ drops from a value between 1.5 and 2.0 to almost zero. The discontinuity in C marks the transition between two topologically different regimes: at small values of R_h the body parallel to the close-packed helix is equivalent to a cylinder; i.e., solvent particles can only probe the exterior of the helix. At large values of R_h the parallel body is equivalent to a hollow cylinder; i.e., solvent particles fit into the cavity formed by the helical tube conformation. Very small solvent particles ($R < R_s$) provide an exception. They can enter the cavity even if R_h is lowered towards R_h^*. Thus, while providing a very efficient packing of a tube, the optimal helix leaves space for spheres with $R \leq R_s$ entering its inside. For these spheres, C displays a jump at $R_h < R_h^*$. This corresponds to the formation of a connection between the interior of the cavity and the bulk, being established due to the increase of the pitch P_h in the CCP regime.

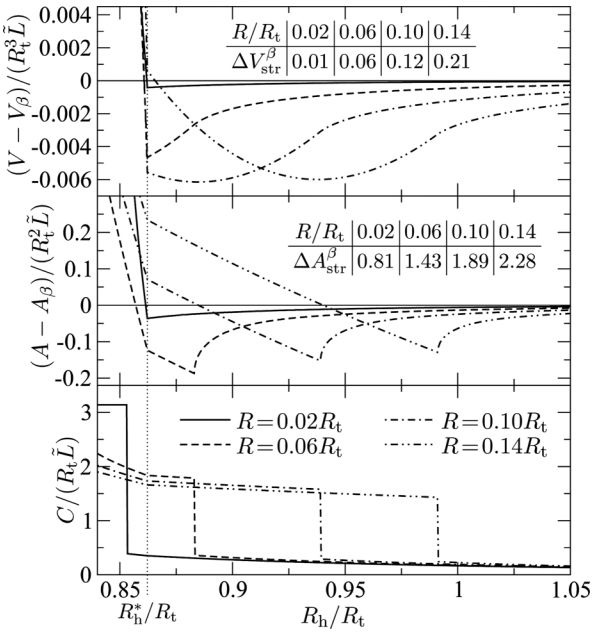


FIG. 2. Geometric measures V, A, and C for a range of close-packed helical tube conformations, which have minimal P_h. Calculations are performed for the parallel body at different distances R (radius of solvent hard core). The integral Gaussian curvature X vanishes, consistent with the given topology. $R_h >$ ($<$)R_h^* corresponds to the TTCP (CCP) regime.

Snir and Kamien have calculated F_{sol} [9] using the Asakura-Oosawa (AO) model [20] which corresponds to taking $\sigma = \kappa = \bar{\kappa} = 0$ due to the neglect of the solvent particle interaction. Thus within this model F_{sol} can be inferred directly from the results for V (see Fig. 2, which is restricted to the set of close-packed helices; detailed analysis shows that minima in V are always assumed on this set). Accordingly, for solvent particles with radius $R < R_{\mathrm{uw}}$ the optimal helix minimizes F_{sol}. For larger solvent particles, F_{sol} assumes its minimum for certain $R_h > R_h^*$. This was interpreted in Ref. [9] such that larger solvent particles lead to a favoring of sheetlike folding. We emphasize that the AO model is valid only for asymptotically low solvent packing fractions η at which, however, F_{sol} is very small because the pressure $p \propto \eta$. Accordingly, in this limiting case energetic contributions of origin other than the solvent dominate. Using Eq. (1) allows us to obtain reliable results for F_{sol} of the close-packed helices in a hard-sphere solvent at realistic, nonvanishing densities. For the thermodynamic coefficients we use accurate analytical results obtained recently from classical DFT [15,21]. As a function of R and η, we find three different regions (see Fig. 3) in which F_{sol} is minimal: (i) for $R_h = R_h^*$ (optimal helix), (ii) for $R_h \to \infty$ (β sheet), and (iii) for certain finite values $R_h > R_h^*$ (unwound helix), respectively. We have checked the stability of the close-packed conformations by slightly increasing P_h which always increases F_{sol}.

In the limit $\eta \to 0$ the result of Snir and Kamien is recovered. However, our results for $\eta > 0$ show that their conclusion that larger solvent particles favor β sheets is not valid. On the contrary, the β sheet minimizes F_{sol} for *small* solvent particles ($R_s \approx 0.05 R_t$) with $\eta \gtrsim 0.3$. Moreover, the picture conveyed in Refs. [9,10] that helices from the

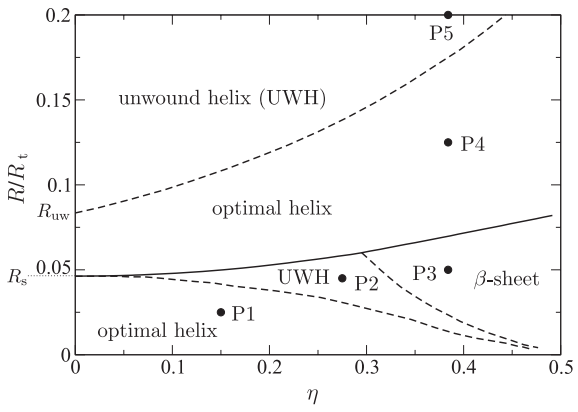
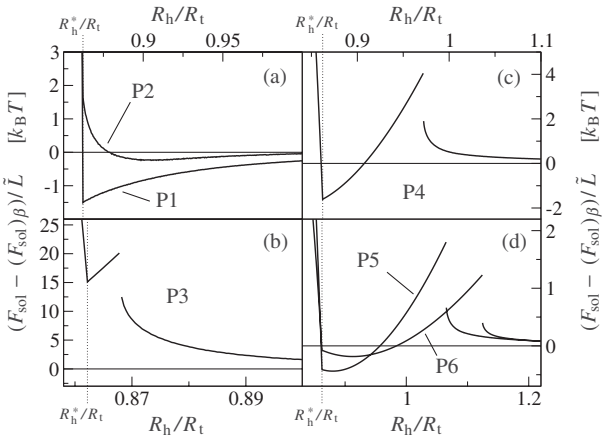


FIG. 3. Diagram of the different protein structures which minimize F_{sol} in a hard-sphere solvent with packing fraction η and solvent radius R. Transitions between the regions are either continuous (dashed lines) or discontinuous (full line) in R_h. In the limit $\eta \to 0$ there is no transition at $R = R_s$. The systems indicated by dots are discussed in Fig. 4.



	P1	P2	P3	P4	P5	P6
η	0.15	0.275	0.384	0.384	0.384	0.384
R/R_t	0.025	0.045	0.05	0.125	0.2	0.25
ΔF_{str}^{β}	43.73	39.13	49.36	13.87	7.30	5.39

FIG. 4. $F_{sol}(R_h)$ relative to its value at $R_h = \infty$ for the hard-sphere solvents indicated in Fig. 3, except P6. F_{sol} is shown for configurations along the line of close-packed helices with a focus on the turn-to-turn distance limited regime ($R_h > R_h^*$). In the curvature limited regime ($R_h < R_h^*$) the values ΔF_{str}^{β}, given in the table, are reached.

unwound regime at large R are *gradually* drifting to a β sheetlike geometry upon increasing R is not correct. This can be inferred from Fig. 4(d) which shows for two solvents that the minimum at $R_h > R_h^*$, generated by the volume contribution, is separated from the β sheetlike configuration by a *free energy barrier* arising from the surface area contribution (see Fig. 2). The discontinuity in these curves is due to the contribution of C. It corresponds to the special helix radius at which, upon decreasing R_h, the solvent is squeezed out from the inner part of the helix. Once this is completed, the system can easily relax to the optimal helix configuration at smaller R_h. The discontinuity in F_{sol} is a consequence of the hard body interaction between the protein tube and the solvent particles. It would be smeared out for soft interaction potentials. Other free energy curves for different solvent states are shown in Figs. 4(a)–4(c). From these the presence of continuous (in R_h) and discontinuous transition lines between the regions in Fig. 3 can be inferred. In the table of Fig. 4 we compare F_{sol} for a stretched tube with the value corresponding to the β sheet $[\Delta F_{str}^{\beta} = [(F_{sol})_{str} - (F_{sol})_\beta]/(k_BT\tilde{L})]$. The large values of ΔF_{str}^{β} demonstrate that in a hard-sphere solvent both the optimal helix and the β sheet are clearly favored free energetically compared with the stretched tube configuration which is devoid of any economic packing. We further note that the curves for F_{sol} become flatter with increasing R [compare the scales in Figs. 4(b)–4(d)]. Thus for large R intramolecular energy contributions of the protein are expected to significantly determine the native state. Therefore it appears to be problematic to follow Ref. [9] in invoking a *solvent* induced tendency to β sheetlike folding in this region.

In the general case, when F_{sol} is large, it has been argued that energy gains from intramolecular contributions such as hydrogen bonds between different amino acids are compensated to a large extent by the dehydration penalty which occurs upon folding of a protein [6]. For studying this interplay in more detail it is useful to extend the morphological approach Eq. (1) to a more realistic solvent with intermolecular square-well-like attraction and which interacts via repulsion (hydrophobicity) or attraction (hydrophilicity) with the protein. The corresponding *thermodynamic coefficients* can be obtained by fitting Eq. (1) to DFT calculations of F_{sol} for simple solutes such as spheres and cylinders which repel or attract the solvent, whereas the results for the *geometric measures* of the protein remain unchanged. Preliminary studies of F_{sol} confirm the role of hydrophobic side chains as a driving force for protein folding.

[1] O. Schueler-Furman, C. Wang, P. Bradley, K. Misura, and D. Baker, Science **310**, 638 (2005).
[2] C. Chothia, Nature (London) **357**, 543 (1992).
[3] M. Denton and C. Marshall, Nature (London) **410**, 417 (2001).
[4] J. R. Banavar and A. Maritan, Rev. Mod. Phys. **75**, 23 (2003), and references therein.
[5] C. M. Dobson, Nature (London) **426**, 884 (2003).
[6] Y. Harano, R. Roth, and M. Kinoshita, Chem. Phys. Lett. **432**, 275 (2006).
[7] H. Hadwiger, *Vorlesungen über Inhalt, Oberfläche und Isoperimetrie* (Springer, Berlin, 1957).
[8] R. Evans, Adv. Phys. **28**, 143 (1979).
[9] Y. Snir and R. D. Kamien, Science **307**, 1067 (2005).
[10] Y. Snir and R. D. Kamien, Phys. Rev. E **75**, 051114 (2007).
[11] C. N. Likos, K. R. Mecke, and H. Wagner, J. Chem. Phys. **102**, 9350 (1995).
[12] P.-M. König, R. Roth, and K. R. Mecke, Phys. Rev. Lett. **93**, 160601 (2004).
[13] R. Roth, Y. Harano, and M. Kinoshita, Phys. Rev. Lett. **97**, 078101 (2006).
[14] R. Evans, J. R. Henderson, and R. Roth, J. Chem. Phys. **121**, 12 074 (2004).
[15] H. Hansen-Goos and R. Roth, J. Phys. Condens. Matter **18**, 8413 (2006).
[16] Y. Rosenfeld, Phys. Rev. Lett. **63**, 980 (1989).
[17] S. Przybył and P. Pierański, Eur. Phys. J. E **4**, 445 (2001).
[18] A. Maritan, C. Micheletti, A. Trovato, and J. R. Banavar, Nature (London) **406**, 287 (2000).
[19] K. R. Mecke, T. Buchert, and H. Wagner, Astron. Astrophys. **288**, 697 (1994).
[20] S. Asakura and F. Oosawa, J. Chem. Phys. **22**, 1255 (1954).
[21] H. Hansen-Goos and R. Roth, J. Chem. Phys. **124**, 154506 (2006).