Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 14429–14432, December 1997
Biophysics

Folding thermodynamics of a model three-helix-bundle protein

(Lindemann criterion/two-state transition/disordered and ordered globules/surface molten solid)

YAOQI ZHOU† AND MARTIN KARPLUS†‡§

†Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138; ‡Laboratoire de Chimie Biophysique, ISIS, Institut le Bel, Université Louis Pasteur, 67000 Strasbourg, France

Contributed by Martin Karplus, October 27, 1997

ABSTRACT **The calculated folding thermodynamics of a simple off-lattice three-helix-bundle protein model under equilibrium conditions shows the experimentally observed protein transitions: a collapse transition, a disordered-to-ordered globule transition, a globule to native-state transition, and the transition from the active native state to a frozen inactive state. The cooperativity and physical origin of the various transitions are explored with a single "optimization" parameter and characterized with the Lindemann criterion for liquid versus solid-state dynamics. Below the folding temperature, the model has a simple free energy surface with a single basin near the native state; the surface is similar to that calculated from a simulation of the same three-helix-bundle protein with an all-atom representation [Boczko, E. M. & Brooks III, C. L. (1995) Science 269, 393–396].**

An understanding of the thermodynamics of proteins is required for a solution of the folding problem (1). Proteins are known to have complex phase behavior consisting of the denatured (random) coil state, often a more or less structured compact globule state and the well-defined native state (1, 2). In addition, certain proteins have been shown to undergo a transition from a native state that is active at physiological temperatures to a low-temperature "glassy" state that is not (3–5). In this paper, we use an off-lattice heteropolymer model that has a structure similar to a three-helix-bundle fragment of *Staphylococcus aureus* protein A (6, 7) and discrete molecular dynamics to investigate the phase behavior. Such simplified models, whose thermodynamics can be fully characterized computationally, are an important source of information about the essential factors involved in protein folding. The present model exhibits the most important transitions of proteins. The behavior is determined by a single model parameter. In particular, the collapsed globule state can be changed from being essentially disordered to one with significant native secondary structure by varying the parameter.

Theoretical Method

Model. The model consists of 46 freely jointed beads, each of which represents an amino acid residue that can interact with other residues via a square-well potential (8). To obtain the global minimum structure for the model (Fig. 1), the 46 beads were initially placed at the C_α positions of the protein (6). The bond length σ_b is chosen to be 3.80 Å, the average C_α–C_α virtual-bond distance. The bond length is allowed to vary freely between $0.9\sigma_b$ and $1.1\sigma_b$ (8, 10). The hard-core diameter, σ_c, is chosen to be 4.27 Å, which is the minimum distance between two "nonbonded" C_α atoms found in the original three-helix-bundle structure. The square-well diameter σ_d is $1.5\sigma_c = 6.41$ Å, an interaction cutoff distance that is close to 6.5 Å used by Miyazawa and Jernigan (11) to derive empirical pair contact energy parameters. Because no side chains are present, the model must contract from the actual structure to obtain reasonable packing. The structure was annealed from $T^* = 1$ to $T^* = 0.001$ for 200 million collisions (8, 10) during which the original 97 native contacts were kept intact. The final structure (Fig. 1) has reasonable packing with 254 contacts and a radius of gyration of 6.8 Å that is smaller than the original one (9.3 Å). To obtain a model with the minimum number of parameters, only two types of residue interactions were used. The 254 contacts that exist in the three-helix-bundle global minimum structure have a well-depth of $B_N\varepsilon$, and all other possible contacts are assigned the value $B_O\varepsilon$ ($B_O > B_N$), where ε determines the energy scale. This Gō-type potential (12) makes it possible to relate the thermodynamic properties of the model protein to the "bias gap", g ($g = 1 - B_O/B_N$) and the reduced temperature T^* ($T^* = T/\varepsilon$). The bias gap is a measure of the difference in stability between the global minimum contacts and other contacts; it has been used as an optimization parameter in the design of lattice models with stable structures (13). A value of g near or greater than 1 corresponds to a large stabilization energy of the global minimum structure, relative to other collapsed structures, and g near zero means that all contacts result in similar stabilization ($\varepsilon < 0$), as for a homopolymer (8).

Simulation. Constant-temperature collision-driven discontinuous molecular dynamics simulations (8, 10) were made at $g = 1.3$, 1, 0.7, 0.5, and 0.3. Each model was simulated for between 11 and 30 reduced temperatures ranging from 0.1 to 5. For every temperature, five simulations (different initial configuration and velocities) were made to estimate statistical errors. Depending on the temperature, the simulations involve from 10 million to 1 billion collisions, the first half of which (equilibration) was discarded. For low-temperature simulations, the random initial configuration is annealed from high temperature (e.g., $T^* = 1$) or slowly heated from the global minimum structure before starting an equilibrium average. This avoids the possibility that the system might be kinetically trapped in local free-energy minima. In the simulation, the time-averaged energy and squared energy are obtained. The heat capacity is calculated from the energy fluctuations (14). The thermodynamic properties at the temperatures that are not simulated are obtained by use of the weighted histogram method (10, 15), which is a least-square-optimization method for extracting the degeneracy factor of the energy levels from simulations at several temperatures so that the partition function can be determined. The radius of gyration and the average fraction of common global-minimum contacts were also obtained during the simulation.

Lindemann Parameter. The Lindemann parameter, that is often used for characterizing the solid-to-liquid transition (16, 17), is defined as $\Delta_L = (\Sigma_i \langle MSF_i \rangle_{cf}/N)^{1/2}/\sigma_c$, where N is 46, σ_c is the hard core diameter ($\sigma_c = 4.27$ Å), and MSF_i is the

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

§To whom reprint requests should be addressed. e-mail: marci@tammy.harvard.edu.

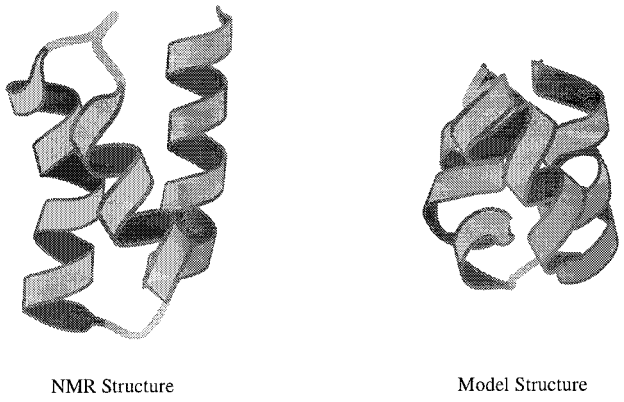

FIG. 1. (*Left*) The three-helix-bundle protein [residues 10–55 of the B domain of *Staphylococcus aureus* protein A (Protein Data Base accession no. 1bdd) (6)]; the same 46 residue protein was used in ref. 7. (*Right*) The global energy-minimum structure for the model. [Drawn with MOLSCRIPT (9).]

mean-squared fluctuation for bead i (after the translation and rotation of the entire molecule are removed). Because Δ_L is a sum over all atomic contributions, it is also useful to define partial Lindemann parameters for different portions of the system. In particular, we introduce Δ_L^{in}, the Lindemann parameter for the interior of the system, which is defined as a sphere located at the center of mass with a radius of 0.8 σ_c.

Results

Fig. 2 shows the heat capacity C_v and squared radius of gyration R_g^2 as a function of temperature for models with a bias energy gap equal to 0.3 and 1.3, respectively. There are four apparent peaks in the C_v curve that correspond to transitions between different states (18). Starting from high temperature, the first peak at $T^* = 0.9$ in the large-gap model ($g = 1.3$) coincides with the strong collapse transition shown in the R_g^2 curve. For the $g = 0.3$ model, the first well-defined peak occurs at a much lower temperature ($T^* = 0.55$) and the collapse transition is weak; it appears as a plateau (or hump) at $T^* \sim 1.5$ in the C_v curve. Analysis of the distribution of states indicates that the first transition for the $g = 1.3$ model is a cooperative two-state-like transition (i.e., a bimodal distribution of potential energies representing the coil and collapsed states) and the transition at $T^* = 0.55$ for the $g = 0.3$ model is continuous. The opposite is true for the transition represented by the second peak; i.e., it is first-order-like for the $g = 0.3$ model but not for the $g = 1.3$ model.

On the basis of the transition temperatures obtained from the C_v and R_g^2 curve (Fig. 2), a phase diagram as a function of the reduced temperature T^* and the bias gap g can be constructed (Fig. 3). The figure also shows structures associated with the states that are stable at different temperatures for the limiting cases, $g = 0.3$ and $g = 1.3$, whose C_v and R_g^2 variation is presented in Fig. 2. At high T^*, the random coil is stable and the transition temperature to a collapsed globule is sensitive to the bias gap. This is in accord with the fact that the contacts are stronger on average for small g so that the collapse transition moves to higher T^*. Depending on g, the globule resulting from the collapse transition has different structural properties. For small g ($g < 0.5$), a disordered globule with little native secondary structure results from collapse and the globule is stable over a wide temperature range. As g increases, a more ordered (molten) globule with much of the secondary structure (60–90% helical contents) and some of the tertiary structure (30–80%) of the three-helix-bundle results from the collapse transition.

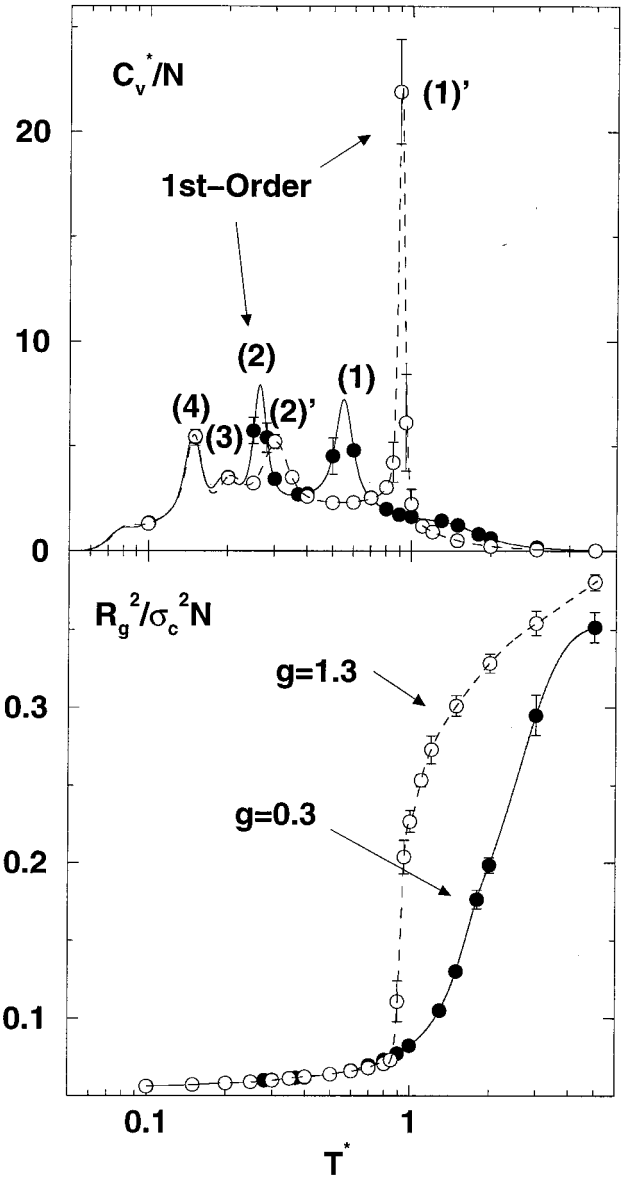

FIG. 2. The reduced heat capacity, C_v^* ($= C_v/k_B$) and the reduced squared radius of gyration per bead, $R_g^2/\sigma_c^2 N$, as a function of temperatures. The lines in C_v^* curves are obtained from the weighted histogram method (10, 15). The lines in R_g^2 (dashed line, $g = 1.3$; solid line, $g = 0.3$) are from a spline fit. Error bars are less than the size of points excepted as shown.

To analyze the dynamics of the collapsed globule, we use the Lindemann parameter Δ_L (16), which has been shown to be less than or equal to a critical value (0.10–0.14) for the solid state of many different substances and significantly greater than that for the liquid state (17). The values of Δ_L for the collapsed globule depend somewhat on g and the temperature; they are in the range 0.2–1.0 for the ordered globule and greater than 1 for the disordered globules. Thus, they are both liquid-like in their dynamics.

For proteins, both disordered (premolten) and ordered (molten) globules have been observed during folding and unfolding (2, 19–22). Many proteins have an ordered (molten) globule phase that is stable under appropriate conditions (e.g., low pH and denaturants) (2). A first-order-like transition from the ordered globule to a disordered globule or the random coil has been observed in the presence of denaturants for certain proteins (2, 22). In the present model, the transition from the random coil to an ordered globule (specific collapse) is

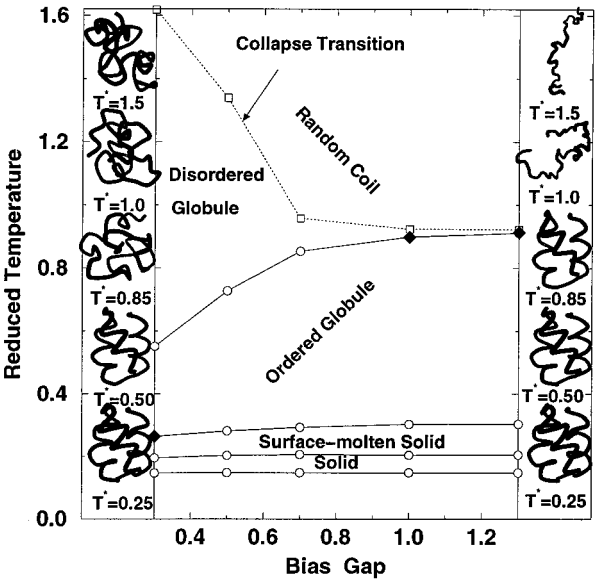

FIG. 3. Phase diagram of the three-helix-bundle protein along with the structures for $g = 0.3$ (*Left*) and $g = 1.3$ (*Right*) at selected temperatures. Four transitions (open circles connected by solid lines) are found from the peaks in the heat capacity. The collapse transition (open squares connected with dotted lines) was determined from the temperature at which dR_g^2/dT, the temperature derivative of the squared radius of gyration, is a maximum (see Fig. 2). All lines are drawn to serve as a guide. The first-order-like two-state transitions are indicated by a solid diamond. The structures shown for disordered globules and random coils are typical instantaneous structures and the structures for ordered globules and surface-molten solid are average structures; the average structures are obtained by averaging over 0.2–2 million configurations after removing translational and rotational motions by minimizing rms deviations with respect to the first configuration. [Drawn with MOLSCRIPT (9).]

first-order-like but the disordered globule to ordered globule transition is "continuous."

The higher temperature portion of the phase diagram described so far corresponds to the usual phase diagram of "protein-like" lattice models (23–25) and the predications of heteropolymer theories (26, 27); i.e., they show three states corresponding to a random coil, a disordered globule, and a highly ordered globule, the latter of which is often associated with the native state in such models (27). Further, the disordered globule phase has a triangular shape and the three phases meet at a triple point (Fig. 3 of this paper and figures 1–3 of ref. 23).

What is striking in the present off-lattice model and contrasts with most lattice model results is that the low-temperature portion of the phase diagram shows additional transitions. These transitions are relatively insensitive to the bias gap over the range $0.3 < g < 1.3$ (Fig. 3). This is true because there are no major changes in structure and most of the contacts, independent of the state, are strong contacts ($B_N\varepsilon$). The transition at $T^* \sim 0.3$ (see Fig. 3) goes from the collapsed globule state to a state with more than 80% of the nonlocal contacts ($|i - j| > 4$) in the global minimum structure and an rms deviation of less than 1 Å relative to that structure. This state corresponds to the native state of the protein. It has a solid-like interior and a liquid-like surface (10, 28), as measured by the Lindemann criterion (16, 17) (i.e., $\Delta_L^{in} < 0.1$ and $0.1 < \Delta_L < 0.2$). For the small bias gap model ($g \cong 0.3$), the transition to the native state is first-order-like, but it is not for the larger gap model (see also below).

At a lower temperature ($T^* \sim 0.2$), there is a much weaker transition (Fig. 2) that involves no structure changes (rms deviation < 1 Å), but the Lindemann parameter for the entire structure is now $\Delta_L \cong 0.1$; i.e., the system changes from one that has a solid core with a fluid surface to one that is frozen (solid) throughout (28). This corresponds to the inactive "glassy" state observed in a number of proteins (3–5). If one assigns an energy scale such that $T^* \sim 0.3$, the native to globule melting transition temperature, corresponds to 350 K ($\varepsilon = 2.3$ kcal/mol; 1 cal = 4.184 J) the transition to the "glassy" state is at about 230 K, in satisfactory agreement with experiment (3, 4). At an even lower temperature ($T^* = 0.15$, 175 K), there is an additional transition that involves subtle changes in the orientational order. A transition at 170 K in the dynamics of ligand binding to myoglobin has been observed (3). The transition also appears in a series of all-atom simulations of crambin and at different temperatures (Y.Z., D. Vitkup, M.K, unpublished results).

Fig. 4 shows that the average free energy at $T^* = 0.25$ (280 K), a temperature where the native state is stable, as a function of Q, the fraction of the global minimum contacts. For low g ($g = 0.3$) and high g ($g = 0.7$), there is a free energy minimum at $Q \cong 0.9$, corresponding to the fact that the average structure does not have all of the contacts of the minimum energy structure at a finite temperature. At $g = 0.7$, the curve decreases smoothly and there is no free energy barrier, but for $g = 0.3$, the free energy surface is somewhat rougher and there is a weak free energy barrier close to the native state near $Q = 0.87$ (the barrier is about 0.2 kcal/mol). These results can be compared with the free energy curve as a function of the radius of gyration obtained in a simulation by Boczko and Brooks (7) of the same three-helix-bundle protein with a fully solvated all-atom model. They found a similar, rather smooth, free energy surface with a small free energy barrier (~0.5 kcal/mol) close to the native state. In addition, there is a transition state with 30% tertiary and 50–70% helical contacts (29), which compares well with 30–50% tertiary and 60–70% helical contents in the present model at the disordered to ordered globule transition temperature.

Concluding Discussion

The success of the present model of a three-helix bundle in reproducing the transitions observed in proteins indicates that

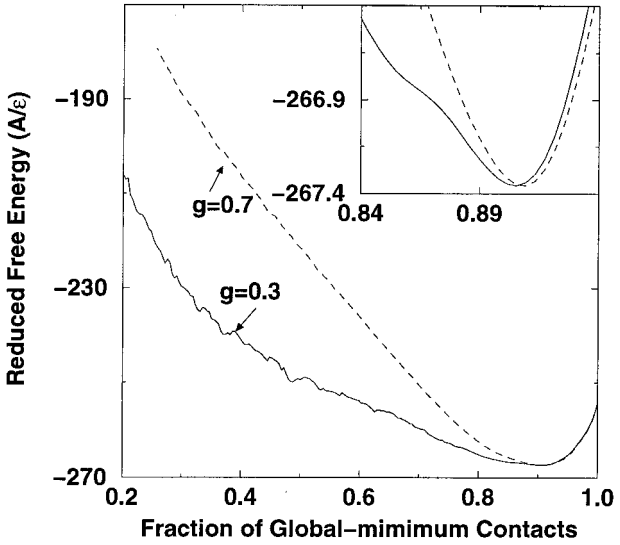

FIG. 4. Reduced free-energy landscape A^* ($A^* = A/\varepsilon$) at $T^* = 0.25$ (the surface-molten-solid phase) for $g = 0.3$ and $g = 0.7$, as a function of the fraction of global-minimum contacts Q. The results were obtained with the weighted histogram method (10, 15). The larger gap model has a smoother landscape. There is a small free-energy barrier for $g = 0.3$ at $Q \approx 0.87$; it yields a weakly bimodal potential energy distribution at the transition temperature (data not shown).

it may be possible to describe certain fundamental aspects of protein thermodynamics in relatively simple terms. For example, the model shows that the nature of the coil-to-globule transition is very sensitive to the bias gap, but that the other transitions (globule to active native or active native to "glassy" inactive state) are much less so, as is the overall free energy diagram. That the globule to native state transition is weaker than in real proteins may suggest that side-chain packing, which is not included in the present model, is important for this aspect of protein folding. Nevertheless, certain aspects of the essential relationship between the amino acid sequence and the structure of a protein are included, as illustrated by the specific protein-like structure of the native state.

We thank Aaron R. Dinner for helpful discussion and Charles L. Brooks, III for providing a copy of the paper (29) before publication. This work was supported by a grant from the National Science Foundation and a grant from Pittsburgh Supercomputing Center. The Protein Data Bank provided the structure (1bdd) before its release after obtaining approval from the authors (6). Y.Z. is a National Institutes of Health Postdoctoral Fellow.

1. Karplus, M. & Shakhnovich, E. I. (1992) in *Protein Folding*, ed. Creighton, T. (Freeman, New York), pp. 127–195.
2. Ptitsyn, O. B. (1995) *Adv. Protein Chem.* **47,** 83–230.
3. Austin, R. H., Beeson, K. W., Eisenstein, L., Frauenfelder, H. & Gunsalus, I. C. (1975) *Biochemistry* **14,** 5355–5373.
4. Rasmussen, B. F., Stock, A. M., Ringe, D. & Petsko, G. A. (1992) *Nature (London)* **357,** 423–424.
5. Ferrand, M., Dianoux, A. J., Petry, W. & Zaccai, G. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 9668–9672.
6. Gouda, H., Torigoe, H., Saito, A., Sato, M., Arata, Y. & Shimada, I. (1992) *Biochemistry* **31,** 9665–9672.
7. Boczko, E. M. & Brooks, C. L., III (1995) *Science* **269,** 393–396.
8. Zhou, Y., Hall, C. K. & Karplus, M. (1996) *Phys. Rev. Lett.* **77,** 2822–2825.
9. Kraulis, P. J. (1991) *J. Appl. Cryst.* **24,** 946–950.
10. Zhou, Y., Karplus, M., Wichert, J. M. & Hall, C. K. (1997) *J. Chem. Phys.*, in press.
11. Miyazawa, S. & Jernigan, R. L. (1985) *Macromolecules* **18,** 534–552.
12. Taketomi, H., Ueda, Y. & Go, N. (1975) *Int. J. Peptide Protein Res.* **7,** 445–459.
13. Abkevich, V. I., Gutin, A. M. & Shakhnovich, E. I. (1996) *Folding Design* **1,** 221–230.
14. Friedman, H. L. (1985) *A Course in Statistical Mechanics* (Prentice–Hall, Englewood Cliffs, NJ).
15. Ferrenberg, A. M. & Swendsen, R. H. (1989) *Phys. Rev. Lett.* **63,** 1195–1197.
16. Lindemann, F. A. (1910) *Physik. Z.* **11,** 609–612.
17. Stillinger, F. H. (1995) *Science* **267,** 1935–1939.
18. Beattle, J. A. & Oppenheim, I. (1979) *Principles of Thermodynamics* (Elsevier, Amsterdam).
19. Agashe, V. R., Shastry, M. C. R. & Udgaonkar, J. B. (1995) *Nature (London)* **377,** 754–757.
20. Sosnick, T. R., Mayne, L. & Englander, S. W. (1996) *Proteins* **24,** 413–426.
21. Jacob, M., Schindler, T., Balbach, J. & Schmid, F. X. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 5622–5627.
22. Uversky, V. N. & Ptitsyn, O. B. (1996) *J. Mol. Biol.* **255,** 215–228.
23. Dinner, A., Sali, A. Karplus, M. & Shakhnovich, E. I. (1994) *J. Chem. Phys.* **101,** 1444–1451.
24. Doniach, S., Garel, T. & Orland, H. (1996) *J. Chem. Phys.* **105,** 1601–1607.
25. Gutin, A. M., Abkevich, V. I. & Shakhnovich, E. I. (1995) *Biochemistry* **34,** 3066–3076.
26. Shakhnovich, E. I. & Gutin, A. M. (1989) *Biophys. Chem.* **34,** 187–199.
27. Plotkin, S. S., Wang, J. & Wolynes, P. G. (1997) *J. Chem. Phys.* **106,** 2932–2948.
28. Frauenfelder, H., Petsko, G. A. & Tsernoglou, D. (1979) *Nature (London)* **280,** 558–563.
29. Guo, Z., Brooks, C. L., III & Boczko, E. M. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 10161–10166.