THE JOURNAL OF PHYSICAL CHEMISTRY B

ARTICLE

pubs.acs.org/JPCB

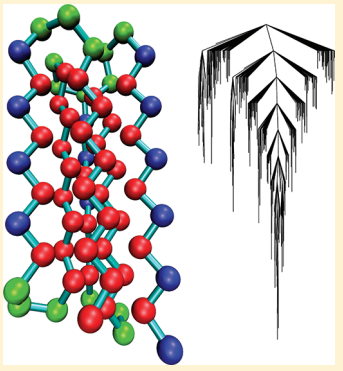
Energy Landscape and Global Optimization for a Frustrated Model Protein

Mark T. Oakley,[†] David J. Wales,[‡] and Roy L. Johnston*[,†]

[†]School of Chemistry, University of Birmingham, Edgbaston, Birmingham, B15 2TT, U.K.

[‡]University Chemical Laboratories, Lensfield Road, Cambridge CB2 1EW, U.K.

ABSTRACT: The three-color (BLN) 69-residue model protein was designed to exhibit frustrated folding. We investigate the energy landscape of this protein using disconnectivity graphs and compare it to a Gō model, which is designed to reduce the frustration by removing all non-native attractive interactions. Finding the global minimum on a frustrated energy landscape is a good test of global optimization techniques, and we present calculations evaluating the performance of basin-hopping and genetic algorithms for this system. Comparisons are made with the widely studied 46-residue BLN protein. We show that the energy landscape of the 69-residue BLN protein contains several deep funnels, each of which corresponds to a different β-barrel structure.



INTRODUCTION

The potential energy landscape of a protein is a high-dimensional object, where locating the global potential energy minimum, the "protein folding problem", is often still considered a "grand challenge". Since calculations using models where all degrees of freedom of all atoms in the protein are included are computationally demanding, coarse-grained models with simple interaction potentials are often used.

The BLN model,[1−4] in which protein residues are represented by hydrophobic (B), hydrophilic (L), and neutral (N) beads, is a widely studied representation.[5−21] In the present work we employ a version of the BLN potential with stiff harmonic terms to restrain the bond lengths and bond angles:[6]

$$V_{\mathrm{BLN}} = \frac{1}{2}K_{\mathrm{r}}\sum_{i=1}^{N-1}(R_{i,i+1}-R_{\mathrm{e}})^2 + \frac{1}{2}K_{\theta}\sum_{i=1}^{N-2}(\theta_i-\theta_{\mathrm{e}})^2 + \varepsilon\sum_{i=1}^{N-3}[A_i(1+\cos\phi_i)+B_i(1+3\cos\phi_i)] + 4\varepsilon\sum_{i=1}^{N-2}\sum_{j=i+2}^{N}C_{ij}\left[\left(\frac{\sigma}{R_{ij}}\right)^{12}-D_{ij}\left(\frac{\sigma}{R_{ij}}\right)^{6}\right] \quad (1)$$

where R_{ij} is the distance between two beads i and j. The first term is a harmonic bond restraint with $K_{\mathrm{r}} = 231.2\varepsilon\sigma^{-2}$ and $R_{\mathrm{e}} = \sigma$, for consistency with previous work.[6,9,14,17] The main features of the landscape are independent of the precise value chosen for K_{r} over quite a wide range of values. The second term is an angle restraint with $K_{\theta} = 20$ rad^{-2} and $\theta_{\mathrm{e}} = 1.8326$ rad. The third term involves torsional angles, ϕ, defined by four successive beads. If two or more of these beads are N, then $A = 0$ and $B = 0.2$. For all other sequences, $A = B = 1.2$. Thus, the main chains of the β-barrel prefer all-*trans* conformations, but the turn regions are more flexible. The final term introduces pairwise nonbonded interactions. If both residues are B, then $C = D = 1$. If one residue is L and the other is L or B, then $C = 2/3$ and $D = -1$. If either of the residues is N, then $C = 1$ and $D = 0$.

The 46-residue sequence (BLN-46) $B_9N_3(LB)_4N_3B_9N_3(LB)_5L$ was designed to exhibit a frustrated[22,23] energy landscape,[1,5,6] with a four-strand β-barrel as the global minimum. This protein has several alternative β-barrel structures with energies close to the global minimum, but separated from it by large barriers, compared to experimentally relevant thermal energies.[9] When all non-native attractive interactions are removed to make a Gō potential,[24] most of the frustration is removed and the potential energy landscape forms a single folding funnel.[9] In the Gō model all of the attractive interactions that are not present in the native state are set to zero. This potential is identical to the original BLN potential except that $D_{\mathrm{BB}} = 0$ for any pair of residues where $R_{ij} > 1.167\sigma$ in the global minimum of the unperturbed model. Placing salt bridges in key locations can also reduce the frustration.[16,17]

The 69-residue sequence (BLN-69) $B_9N_3(LB)_4N_3B_9N_3(LB)_4N_3B_9N_3(LB)_5L$ folds into a six-strand β-barrel. Studies using automated histogram filtering,[25] replica exchange Monte Carlo,[26] statistical temperature molecular dynamics,[27] and conformational space annealing[28] indicate that BLN-69 has a frustrated[22,23] potential energy surface.

Received: July 29, 2011
Revised: August 25, 2011
Published: August 26, 2011

The energy landscape of the BLN-69 protein is less well understood than that of its 46-residue counterpart. With more degrees of freedom, it has a much larger conformational space, but how frustrated is it? To answer this question, we analyze the energy landscape of BLN-69 and the corresponding Gō model using disconnectivity graphs,[29,30] where minima are connected by nodes representing the highest transition state on the lowest pathway between them. Finding the global minima of frustrated systems, such as the BLN proteins, provides a useful benchmark for global optimization algorithms. We assess the performance of basin-hopping and genetic algorithms on the BLN model proteins.

METHODOLOGY

Energy Landscape Mapping. The disconnectivity graphs for the 69-residue BLN and Gō model proteins were constructed from databases of stationary points generated using the PATHSAMPLE program,[31] which organizes independent pathway searches using the OPTIM program.[32] The databases of stationary points include 298 856 minima and 258 477 transition states for the BLN potential, and 53 901 minima and 75 389 transition states for the Gō potential. All the transition state searches in OPTIM were conducted in Cartesian coordinates[33] using a quasi-continuous interpolation scheme to avoid chain crossings, with local maxima accurately refined to transition states by hybrid eigenvector-following.[34−36] Successive pairs of local minima were selected for connection attempts within OPTIM using the missing connection algorithm.[37]

Global Optimization. Performance of a local energy minimization after structural perturbations, in combination with an accept/reject test, transforms the potential energy surface into the basins of attraction of local minima,[38] and removes downhill barriers.[39] The resulting basin-hopping (BH) algorithm[40,41] has proved very effective for locating the global minimum in a wide range of systems. In the present work all BH searches were performed using the GMIN program.[42] To improve the efficiency, a restart procedure was used with the NEWRESTART keyword. If the lowest minimum did not change for a fixed number of steps (30 000 in the present work), a new search was initiated from a random starting structure. A cyclic taboo list[43] of 10 structures was also employed to prevent revisits to configuration space that had already been explored, using the AVOID keyword with a distance tolerance of 2.0 for reseeding, in reduced units.

Genetic algorithms (GA) are another popular type of global optimization algorithm.[44,45] If all structures are subjected to energy minimization before fitness evaluation, the resulting "Lamarckian" genetic algorithm operates on the same transformed energy landscape as the BH algorithm. Lamarckian genetic algorithms have also been very effective in optimizing the structures of clusters[46] and proteins.[47,48] In our GA, each structure is represented by a genome comprising the sequence of torsional angles describing the protein backbone conformation. Tournament selection was used to choose the parents for mating. Offspring structures are generated by one-point crossover where the genomes from both parents are cut at the same random point and the offspring's genome comprises one section from each parent. Mutants were generated by making copies of the parent and offspring structures and replacing one randomly selected torsion angle. We use an elitist genetic algorithm, where parent structures are retained in the population. After each generation, all structures (parents, offspring, and mutants) are sorted in order of fitness and the fittest are taken forward to the next generation. Thus, the energy of the least fit member of the population is always less than or equal to that of the least fit in the previous generation. If no better structures are found in a generation, a new epoch is initiated and all members of the population are replaced with random structures. Optionally, a small number of the fittest structures can survive from one epoch to the next. To maintain population diversity, a duplicate predator operator is used.[49] If two members of the population have the same energy (within 0.1ε) and the same torsional angles in the backbone (within 10°), the least stable one is removed from the population.

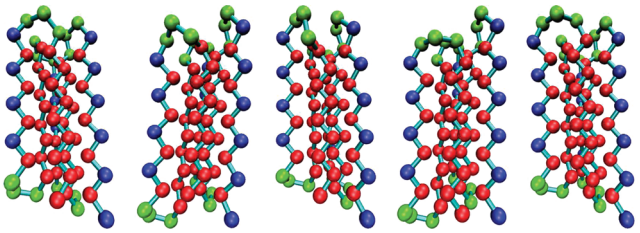


Figure 1. The five most stable structures of the 69-residue BLN protein, ordered in increasing energy with the global minimum on the left. Residues are colored red (B), blue (L), and green (N). The third structure is in the same basin as the global minimum and differs only by the position of the turn residues.

RESULTS AND DISCUSSION

Energy Landscapes of BLN-69 and Gō-69. The global minimum structure for the 69-bead BLN protein has the three B_9 chains forming a hydrophobic core, which is surrounded by the three $(LB)_4$ chains (Figure 1), as characterized in previous studies.[25,28] The global minimum with the Gō potential has the same structure, as expected. The BLN-69 disconnectivity graph is frustrated[22,23] (Figure 2), with a number of deep funnels, each of which leads to a structure close in energy to the global minimum. There are at least 33 other structures within ε of the global minimum. Of the 10 lowest-energy structures, four lie in the same basin as the global minimum and only differ by changes in the turn regions. The other six low-lying minima lie at the bottom of distinct funnels and have different arrangements of the β-strands (Figure 1). Although the overall landscape is frustrated and multifunnel in character, local relaxation within individual funnels to the alternative low-lying minima is expected to be quite rapid. This structure is associated with multiple relaxation time scales for the global minimum and features in the heat capacity.[50−52]

The 69-residue Gō model (Gō-69) has a funneled energy landscape, where most of the local minima are linked to the global minimum by barriers less than 4ε (Figure 3). There are only three other structures within ε of the global minimum, and they are all connected by low-energy transition states. These structures only differ from the global minimum by small rearrangements of the turn regions. It is interesting to note that the disconnectivity graphs for both the 69-residue Gō and BLN proteins are similar for structures accessible by transition states lower than 6ε from the global minimum (Figure 4). This structure implies that the energy landscape for the BLN protein in this region is described by the making and breaking of native contacts, with little influence from non-native contacts.

Global Optimization for 46- and 69-Residue BLN and Gō Model Proteins. The performance of the search algorithms was

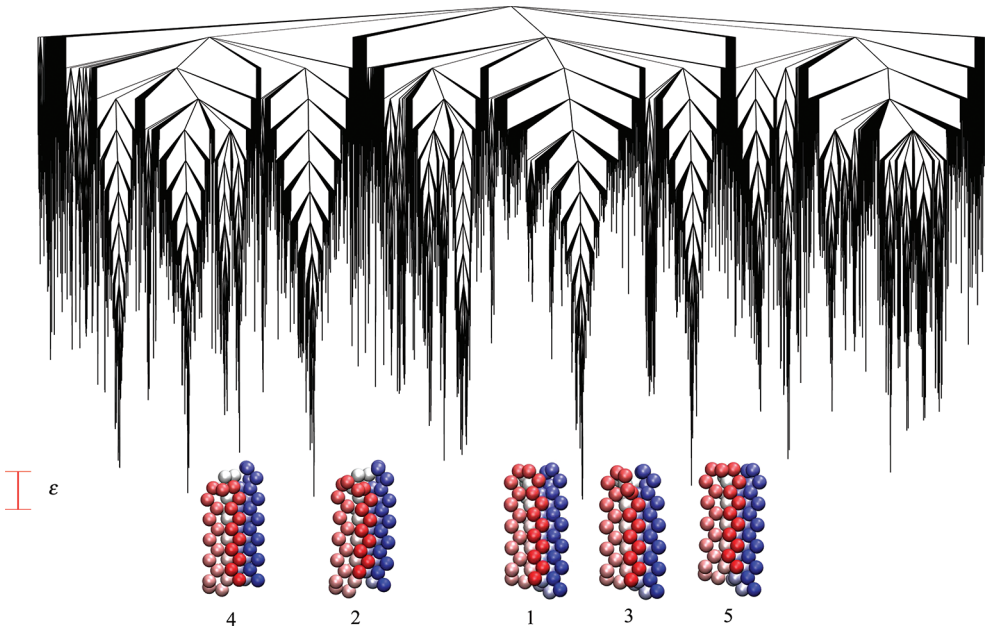


Figure 2. Disconnectivity graph of the 69-residue BLN protein showing the 11 343 minima accessible from the global minimum by transition states lower than 8ε. The five lowest minima are illustrated close to the bottom of the branches to which they correspond. These representations were generated using the VMD program (ref 53) with a coloring scheme for the beads that varies from red to blue (N-terminus to C-terminus).

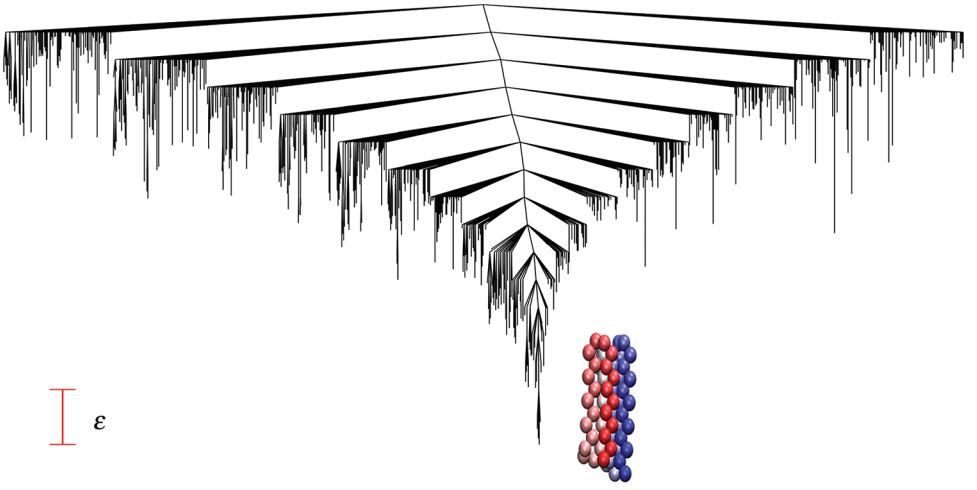


Figure 3. Disconnectivity graph of the 69-residue Gō protein including the 1061 minima accessible from the global minimum by transition states lower than 8ε relative to the global minimum. The structure of the global minimum is illustrated using the VMD program (ref 53) with a coloring scheme for the beads that varies from red to blue (N-terminus to C-terminus).

assessed for both the 46- and 69-residue BLN proteins. For each system, 100 searches were performed from random starting points and the mean time to the first encounter of the global minimum structure was recorded. For GA, the first-encounter time was taken at the end of the generation where the global minimum was found. We present the mean first-encounter time in terms of the number of energy evaluations and the number of minimization steps. The mean number of generations and epochs is also recorded for the GA searches. We consider the number of minimization steps to be the fairest test of the performance of the search algorithm because the number of energy evaluations depends on the convergence criteria used in the local energy minimization. Both of the search algorithms have a number of parameters that can be optimized to improve performance, and the best sets of parameters for the BH and GA searches are listed in Table 1. Searches were also performed for the Gō model proteins using the same search parameters.

The most effective optimization algorithm for BLN-46 reported in previous work is BH, which required an average of 6000 energy minimizations to find the global minimum.[17] The introduction of a taboo list and a restart operator reduce this figure to 4400 minimizations in the present work (Table 2), which required an average of 79 s of CPU time on an Intel Xeon E5405 CPU (single core, clock speed 2.0 GHz).

The epoch operator has a substantial effect on the performance of the GA. Only 82% of the GA searches converge to the global minimum within 100 generations. The other searches prematurely converge to other minima, and after 75 000 generations 10% have

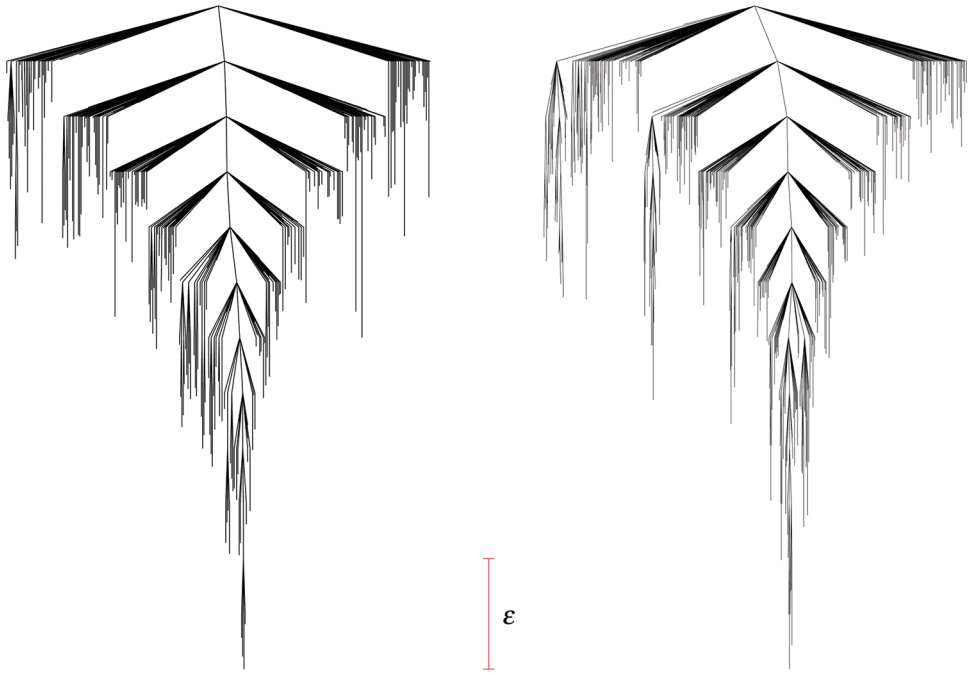


Figure 4. Disconnectivity graphs of the 69-residue Gō (left) and BLN (right) proteins including the minima accessible by transition states lower than 6ε relative to the global minimum.

Table 1. Best Parameters for the Two Optimization Strategies

	BLN-46	BLN-69
	BH	
temperature/ε	2.3	3.4
step size/σ	0.65	0.70
	GA	
population size	140	200
offspring rate	0.9	0.9
mutation rate	0.05	0.05

Table 2. Mean First-Encounter Times for 100 Global Optimization Runs from Random Starting Points of the BLN-46 Protein[a]

	mean first-encounter time		
method	energy evaluations	minimizations	generations
	BLN Model		
BH	6.7×10^5 (5.6×10^5)	4.4×10^3 (3.8×10^3)	n/a
GA	1.4×10^6 (9.3×10^5)	8.3×10^3 (5.7×10^3)	59 (41)
	Gō Model		
BH	1.7×10^5 (9.4×10^4)	5.6×10^2 (3.0×10^2)	n/a
GA	7.4×10^5 (9.8×10^4)	3.3×10^3 (7.3×10^2)	23 (5)

[a] Values in parentheses are the standard deviations. The 46 beads were randomly placed in a sphere of radius 3.0 to generate the initial configurations.

not found the global minimum. Introducing the epoch operator improves the performance of the GA, with all searches finding the

Table 3. Mean First-Encounter Times for 100 Global Optimization Runs from Random Starting Points of the BLN-69 Protein[a]

	mean first-encounter time		
method	energy evaluations	minimizations	generations
	BLN Model		
BH	4.8×10^6 (4.0×10^6)	2.6×10^4 (2.3×10^4)	n/a
GA	5.3×10^6 (2.8×10^6)	2.5×10^4 (1.5×10^4)	115 (68)
	Gō Model		
BH	5.1×10^5 (9.0×10^5)	2.0×10^3 (3.7×10^3)	n/a
GA	1.9×10^6 (2.7×10^5)	7.3×10^3 (2.3×10^3)	33 (11)

[a] Values in parentheses are the standard deviations. The 69 beads were randomly placed in a sphere of radius 3.0 to generate the initial configurations.

global minimum within three epochs (210 generations). The best performance is obtained when no structures are carried over from one epoch to the next.

For the BLN-69 protein the BH algorithm requires an average of 26 000 energy minimizations to locate the global minimum (Table 3), which corresponds to an average of 1100 s CPU time on an Intel Xeon E5405 CPU (single core, clock speed 2.0 GHz). Our GA requires a similar number of minimizations to find the global minimum (Table 3). The performance of the new epoch operator is improved if the fittest structure from each epoch is transferred to the next. The conformational space annealing algorithm has the best published performance for this system, requiring an average of 9.4×10^6 energy evaluations to find the global minimum.[28] However, the number of minimizations is not given, and therefore it is difficult to separate the performance of the conformational space annealing algorithm from the choice of

minimization algorithm. Our GA and BH results both show a significant improvement in efficiency.

As expected, the global minima of the Gō proteins are significantly easier to find than for the BLN proteins. With BH, there is a 10-fold reduction in the number of energy minimizations required when moving from the BLN potential to the Gō potential for both chain lengths. With GA, the number of minimizations required is a factor of 3 smaller for the Gō proteins.

CONCLUSIONS

We have shown that the BLN-69 protein has a frustrated[22,23] potential energy landscape, with multiple low-energy minima lying at the bottom of funnels separated by high barriers. Nevertheless, both the GA and BH algorithms locate the global minimum quite efficiently for both BLN-46 and BLN-69, with significant improvements over previous work. Searches for BLN-69 require between four and six times more energy minimizations than for BLN-46. In future work we plan to examine this scaling in terms of metric disconnectivity graphs and measures of landscape complexity.[19] Another area of interest is to investigate the energy landscapes and ease of global optimization for potentials that are intermediate in character between the BLN and Gō limits,[17,21] in order to identify the transition from a poor to a good folder. This analysis would also allow us to understand the effect of non-native interactions with different strengths on the thermodynamics and kinetics of protein folding. These studies are currently in progress.

AUTHOR INFORMATION

Corresponding Author

*E-mail: r.l.johnston@bham.ac.uk.

ACKNOWLEDGMENT

The authors acknowledge the Engineering and Physical Sciences Research Council, U.K. (EPSRC) for funding under Programme Grant EP/I001352/1. The calculations described in this paper were performed using the University of Birmingham's BlueBEAR HPC service, which was purchased through HEFCE SRIF-3 funds (see http://www.bear.bham.ac.uk).

REFERENCES

(1) Honeycutt, J. D.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 3526–3529.
(2) Honeycutt, J. D.; Thirumalai, D. *Biopolymers* **1992**, *32*, 695.
(3) Guo, Z. Y.; Thirumalai, D. *Biopolymers* **1995**, *36*, 83.
(4) Guo, Z. Y.; Thirumalai, D. *J. Mol. Biol.* **1996**, *263*, 323.
(5) Guo, Z.; Brooks, C. L., III. *Biopolymers* **1997**, *42*, 745.
(6) Berry, R. S.; Elmaci, N.; Rose, J. P.; Vekhter, B. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 9520–9524.
(7) Nymeyer, H.; Garca, A. E.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5921.
(8) Shea, J. E.; Nochomovitz, Y. D.; Guo, Z. Y.; Brooks, C. L. *J. Chem. Phys.* **1998**, *109*, 2895–2903.
(9) Miller, M. A.; Wales, D. J. *J. Chem. Phys.* **1999**, *111*, 6610–6616.
(10) Vekhter, B.; Berry, R. S. *J. Chem. Phys.* **1999**, *110*, 2195.
(11) Vekhter, B.; Berry, R. S. *J. Chem. Phys.* **1999**, *111*, 3753.
(12) Elmaci, N.; Berry, R. S. *J. Chem. Phys.* **1999**, *110*, 10606–10622.
(13) Shea, J.-E.; Onuchic, J. N.; Brooks, C. L. *J. Chem. Phys.* **2000**, *113*, 7663.
(14) Evans, D. A.; Wales, D. J. *J. Chem. Phys.* **2003**, *118*, 3891–3897.
(15) Brown, S.; Fawzi, N. J.; Head-Gordon, T. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 10712–10717.
(16) Stoycheva, A. D.; Onuchic, J. N.; Brooks, C. L. *J. Chem. Phys.* **2003**, *119*, 5722–5729.
(17) Wales, D. J.; Dewsbury, P. E. J. *J. Chem. Phys.* **2004**, *121*, 10284–10290.
(18) Komatsuzaki, T.; Hoshino, K.; Matsunaga, Y.; Rylance, G. J.; Johnston, R. L.; Wales, D. J. *J. Chem. Phys.* **2005**, *122*, 084714.
(19) Rylance, G. J.; Johnston, R. L.; Matsunaga, Y.; Li, C.-B.; Baba, A.; Komatsuzaki, T. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18551–18555.
(20) Kim, J.; Keyes, T. *J. Phys. Chem. B* **2007**, *111*, 2647–2657.
(21) Kim, J.; Keyes, T.; Straub, J. E. *Phys. Rev. E* **2009**, *79*, 030902.
(22) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins: Struct., Funct., Genet.* **1995**, *21*, 167–195.
(23) Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545–600.
(24) Ueda, Y.; Taketomi, H.; Gō, N. *Biopolymers* **1978**, *17*, 1531–1548.
(25) Larrass, S. A.; Pegram, L. M.; Gordon, H. L.; Rothstein, S. M. *J. Chem. Phys.* **2003**, *119*, 13149–13158.
(26) Pan, P. W.; Gordon, H. L.; Rothstein, S. M. *J. Chem. Phys.* **2006**, *124*, 024905.
(27) Kim, J.; Straub, J. E.; Keyes, T. *J. Chem. Phys.* **2007**, *126*, 135101.
(28) Kim, S. Y. *J. Chem. Phys.* **2010**, *133*, 135102.
(29) Becker, O. M.; Karplus, M. *J. Chem. Phys.* **1997**, *106*, 1495–1517.
(30) Wales, D. J.; Miller, M. A.; Walsh, T. R. *Nature* **1998**, *394*, 758–760.
(31) Wales, D. J. *PATHSAMPLE: A program for generating and analysing kinetic transition networks*. http://www-wales.ch.cam.ac.uk/software.html.
(32) Wales, D. J. *OPTIM: A program for geometry optimization and pathway calculations*. http://www-wales.ch.cam.ac.uk/software.html.
(33) Wales, D. J. *J. Chem. Soc., Faraday Trans.* **1993**, *89*, 1305–1313.
(34) Munro, L. J.; Wales, D. J. *Phys. Rev. B* **1999**, *59*, 3969–3980.
(35) Henkelman, G.; Jónsson, H. *J. Chem. Phys.* **1999**, *111*, 7010–7022.
(36) Kumeda, Y.; Munro, L. J.; Wales, D. J. *Chem. Phys. Lett.* **2001**, *341*, 185–194.
(37) Carr, J. M.; Trygubenko, S. A.; Wales, D. J. *J. Chem. Phys.* **2005**, *122*, 234903.
(38) Mezey, P. G. *Theor. Chim. Acta* **1981**, *58*, 309.
(39) Wales, D. J.; Scheraga, H. A. *Science* **1999**, *285*, 1368–1372.
(40) Li, Z.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 6611–6615.
(41) Wales, D. J.; Doye, J. P. K. *J. Phys. Chem. A* **1997**, *101*, 5111–5116.
(42) Wales, D. J. *GMIN: A program for basin-hopping global optimization*. http://www-wales.ch.cam.ac.uk/software.html.
(43) Cvijović, D.; Klinowski, J. *Science* **1995**, *267*, 664–666.
(44) Pedersen, J. T.; Moult, J. *Proteins* **1995**, *23*, 454–460.
(45) Rabow, A. A.; Scheraga, H. A. *Protein Sci.* **1996**, *5*, 1800–1815.
(46) Johnston, R. L. *Dalton Trans.* **2003**, 4193–4207.
(47) Unger, R. The Genetic Algorithm Approach to Protein Structure Prediction. In *Applications of Evolutionary Computation in Chemistry*; Johnston, R. L., Ed.; Springer: Berlin/Heidelberg, Germany, 2004; Vol. 110, pp 2697−2699.
(48) Del Carpio, C. A. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 258–269.
(49) Cox, G. A.; Mortimer-Jones, T. V.; Taylor, R. P.; Johnston, R. L. *Theor. Chim. Acta* **2004**, *112*, 163–178.
(50) Wales, D. J. *Philos. Trans. R. Soc. London, Ser. A* **2005**, *363*, 357–377.
(51) Wales, D. J.; Bogdan, T. V. *J. Phys. Chem. B* **2006**, *110*, 20765–20776.
(52) Wales, D. J. *Curr. Opin. Struct. Biol.* **2010**, *20*, 3–10.
(53) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.