arXiv:cond-mat/0504785v1 [cond-mat.other] 29 Apr 2005

Finding pathways between distant local minima

Joanne M. Carr, Semen A. Trygubenko and David J. Wales*

University Chemical Laboratories, Lensfield Road,

Cambridge CB2 1EW, UK

February 2, 2008

Abstract

We report a new algorithm for constructing pathways between local minima that involve a large number of intervening transition states on the potential energy surface. A significant improvement in efficiency has been achieved by changing the strategy for choosing successive pairs of local minima that serve as endpoints for the next search. We employ Dijkstra's algorithm[1] to identify the 'shortest' path corresponding to missing connections within an evolving database of local minima and the transition states that connect them. The metric employed to determine the shortest missing connection is a function of the minimised Euclidean distance. We present applications to the formation of buckminsterfullerene and to the folding of the B1 domain of protein G, tryptophan zippers, and the villin headpiece subdomain. The corresponding pathways contain up to

*email: dw34@cam.ac.uk

163 transition states, and will be used in future discrete path sampling calculations.

1 Introduction

The discrete path sampling (DPS) approach[2–4] provides a means to sample pathways corresponding to 'rare events' using a coarse-grained framework based on the underlying potential energy surface (PES). The PES can be formally partitioned into the catchment basins of local minima,[5] while rate constants for transitions between these basins can be estimated using statistical rate theory[6–10] once transition states on the basin boundaries have been found. Here we adopt the geometrical definition of a transition state as a stationary point on the PES with a single negative Hessian eigenvalue, following Murrell and Laidler.[11] Minima are stationary points with no negative Hessian eigenvalues.

Two minima may be connected via a single transition state if they are sufficiently close to each other in configuration space. More generally, local minima may interconvert via any number of multi-step paths, and the minimum number of elementary steps will depend upon the isomers in question.[3] Here we consider an elementary step as a rearrangement that involves a single transition state. Locating multi-step pathways on a complex PES can be a difficult task. Double-ended methods[12–38] usually do not reveal all the transition states at once, mainly because of the existence of multiple barrier height and path length scales.[39] Consecutive double-ended searches have proved to be an efficient way of working around the above problem, and this strategy was adopted in our previous work.[2,38] An essential part of this approach is a mechanism to incorporate

the information obtained in all the previous searches into the next one. Various strategies can be adopted, the most general and effective being the one based on Euclidean separation, and we refer the reader to the original publications for details.[2,38]

A guess for the initial pathway is required for most double-ended searches. Linear interpolation is a straightforward way to automate this part of the calculation.[31–34,38] For large endpoint separations guessing the initial pathway can be difficult, and there is a large probability of finding many irrelevant stationary points at the beginning of the calculation.

Due to the discrete nature of the double-ended methods such as NEB[32,33] and DNEB,[38] which use a series of images to represent the path, the candidate transition states obtained from such searches usually need to be optimised further.[36,38] For this purpose we employ eigenvector-following methods.[40–52] The connectivity is obtained by following the two unique steepest-descent (SD) paths downhill from each transition state. In our calculations the refinement of transition states and calculation of approximate steepest-descent paths are the most time consuming steps.

The connection algorithm described in Ref. 38 uses one double-ended search per cycle. However, we have found that this approach can be overwhelmed by the abundance of stationary points and pathways for complicated rearrangements. In the present contribution we introduce the idea of an unconnected pathway and make the connection algorithm more focused by allowing more than one double-ended search per cycle. This approach, in combination with some other modifications described in the next section, greatly reduces the computational demands of the method, and has allowed us to tackle more complicated problems.

The new algorithm has been implemented within our `OPTIM` package, and a public domain version will be made available for download from the internet.[53] The examples presented below each involve the search for an initial connection between distant endpoints for use in subsequent DPS calculations.

2 Methods

2.1 Outline of the Connection Algorithm

A detailed description of the previous version of our connection algorithm was provided elsewhere.[38] Each iteration consists of one double-ended transition state search, followed by refinement of transition state candidates using (hybrid) eigenvector-following,[40–52] and calculation of approximate steepest-descent paths for each transition state to establish the connectivity.

We classify all the known local minima into three categories: S minima, which are connected to the starting structure, F minima, which are connected to the final structure, and U minima, which are so far unconnected to the S and F sets. U minima can be connected between themselves but not to any S or F minima. When a connection is established between the members of U and S or F sets, the unconnected minimum, and all the minima connected to it, become members of S or F, respectively. If a connection is found between members of the S and F sets then the algorithm terminates.

Before each cycle a decision must be made as to which minima to try and connect next. Various strategies can be adopted, for example, selection based on the order in which transition states were found,[2] or, selection of minima with the

minimal separation in Euclidean distance space.[38] However, when the endpoints are very distant in configuration space, neither of these approaches is particularly efficient. The number of possible connections that might be tried simply grows too quickly if the S, F and U sets become large. However, the new algorithm described below seems to be very effective.

2.2 A Dijkstra-based Selector

The modified connection algorithm we have used in the present work is based on a shortest path method proposed by Dijkstra.[1,54] We can describe the minima that are known at the beginning of each connection cycle as a complete graph,[55] $G = (M, E)$, where M is the set of all minima and E is the set of all the edges between them. Edges are considered to exist between every pair of minima u and v, even if they are in different S, F or U sets, and the weight of the edge is chosen to be a function of the minimum Euclidean distance between them:[56]

$$w(u,v) = \begin{cases} 0, & \text{if } u \text{ and } v \text{ are connected via a single transition state,} \\ \infty, & \text{if } n(u,v) = n_{max}, \\ f(D(u,v)), & \text{otherwise,} \end{cases} \tag{1}$$

where $n(u,v)$ is the number of times a pair (u,v) was selected for a connection attempt, n_{max} is the maximal number of times we may try to connect any pair of minima, and $D(u,v)$ is the minimum Euclidean distance between u and v. f should be a monotonically increasing function, such as $f(D(u,v)) = D(u,v)^2$. We denote the number of minima in the set $M = S \cup U \cup F$, as m, and the number of edges in the set E as $e = m(m-1)/2$.

Using the Dijkstra algorithm[1,54] and the weighted graph representation de-

scribed above, it is possible to determine the shortest paths between any minima in the database. The source is selected to be one of the endpoints. Upon termination of the Dijkstra algorithm, a shortest path from one endpoint to the other is extracted. If the weight of this pathway is non-zero, it contains one or more 'gaps'. Connection attempts are then made for every pair (u, v) of adjacent minima in the pathway with non-zero $w(u, v)$ using the DNEB approach.[38]

The computational complexity of the Dijkstra algorithm is at worst $O(m^2)$, and the memory requirements scale in a similar fashion. The most appropriate data structure is a weighted adjacency matrix. For the calculations presented in this paper, the single source shortest paths problem was solved at the beginning of each cycle, which took less than 10% of the total execution time for the largest database encountered. We emphasise here that once an initial path has been found, the perturbations considered in typical DPS calculations will generally involve attempts to connect minima that are separated by far fewer elementary rearrangements than the endpoints. It is also noteworthy that the initial path is unlikely to be contribute significantly to the overall rate constant. Nevertheless, it is essential to construct such a path to begin the DPS procedure.

The nature of the definition of the weight function allows the Dijkstra algorithm to terminate whenever a second endpoint, or any minimum connected to that endpoint via a series of elementary rearrangements, is reached. This observation reduces the computational requirements by an amount that depends on the distribution of the minima in the database among the S, U and F sets. One of the endpoints is always a member of the S set, while the other is a member of F set. Either one can be chosen as the source, and we have found it most efficient to select the one from the set with fewest members. However, this choice does

not improve the asymptotic bounds of the algorithm.

3 Results

3.1 Buckminsterfullerene

Various suggestions[57–75] have been made for the formation mechanism of buckminsterfullerene[76,77] in the gas phase. In total, there are 1812 different C_{60} fullerene isomers,[78,79] which probably interconvert via the 'pyracylene' or 'SW' rearrangement.[80] This process has been investigated in several previous studies,[81–88] and the most accurate calculations[89,90] yield a picture of the energy landscape that is quite similar to that obtained using a tight-binding potential.[86,91] The same model[92–94] was therefore used in this initial pathway calculation to minimise the computational expense, which is significantly greater than for analytical empirical potentials.[95] The present calculation therefore has most in common with the annealing study of Xu and Scuseria,[96] although the latter work did not involve transition state calculations.

An initial high-energy starting point was constructed by simply placing sixty carbon atoms in a container and minimising the energy. The resulting structure, which contains a number of large rings and chains, including polyacetylene fragments, is shown in Figure 1. Although the structure of the buckminsterfullerene endpoint is well known, it is not at all clear which would be the best permutational isomer to attempt a connection with, since the endpoints are so different. The distance between these structures was minimised with respect to permutation-inversion isomers, centre of mass, and orientational coordinates. However, this

connection still presents a significant challenge, and required 383 cycles of the Dijkstra-based algorithm to achieve a complete path (Figure 1).

An enlarged view of the pathway for the last two steps is also shown in Figure 1. Both rearrangements correspond to the SW process mentioned above.[80] The first step converts a patch containing seven, six, and two five-membered rings into a patch with three six-membered rings and one five-membered ring. The second step involves a more conventional process linking two patches that both contain two six-membered and two five-membered rings. The energy profile in this part of the pathway is consistent with the pattern of high barriers previously discussed for the low-energy region of the PES.[86,90,91] The present results suggest that further investigation of paths involving non-fullerene C_{60} isomers may be worthwhile, although we note once again that this initial path may not be dynamically significant.

3.2 GB1 Hairpin

The GB1 hairpin consists of residues 41-56 from the C-terminal fragment of the B1 domain of protein G. It forms a β-hairpin both in the complete protein,[97] and for the isolated fragment in solution.[98] A number of previous experimental and theoretical studies have been conducted for this system,[99–111] including a DPS investigation.[112] Obtaining an initial path between unfolded and hairpin conformations in the latter study was not an easy task, and the new algorithm speeded up this part of the calculation by at least a couple of orders of magnitude.

The `CHARMM` program[113] has previously been interfaced to our `OPTIM` code,[53] which includes a wide variety of algorithms for locating stationary points and

characterising pathways, and now includes the Dijkstra-based connection approach. As for the previous DPS study we employed the CHARMM19 force field with the EEF1 implicit solvation model.[114] The path illustrated in Figure 3 consists of 163 transition states, and was found in 126 cycles of the Dijkstra-based algorithm. It connects a partially collapsed non-native structure, with two turns, to a conformation from the native hairpin ensemble.

3.3 Tryptophan Zippers

Tryptophan zippers are stable fast-folding β-hairpins designed by Cochran *et al.*,[115] which have recently generated considerable interest.[116,117] In the present work we have obtained native to denatured state rearrangement pathways for five tryptophan zippers: trpzip 1, trpzip 2, trpzip 3, trpzip 3-I and trpzip 4. The notation is adopted from the work of Du *et al.*[117] All these peptides contain twelve residues, except for trpzip 4, which has sixteen. Tryptophan zippers 1, 2, 3 and 3-I differ only in the sequence of the turn. Experimental measurements of characteristic folding times for these peptides have shed some light on the significance of the turn sequence in determining the stability and folding kinetics of peptides with the β-hairpin structural motif.[117]

To model these molecules we used a modified CHARMM19 force field,[113] with symmetrised ASN, GLN and TYR dihedral angle and CTER improper dihedral angle terms, to ensure that rotamers of these residues have the same energies and geometries.[118] Another small modification concerned the addition of a non-standard amino acid, D-proline, which was needed to model trpzip 3. The implicit solvent model EEF1 was used to account for solvation,[114] with a small change to

the original implementation to eliminate discontinuities.[119]

We have used the Dijkstra-based connection algorithm to obtain folding pathways for all five trpzip peptides. In each case the first endpoint was chosen to be the native state structure, which, for 1, 2 and 4 trpzips, was taken from the Protein Data Bank (PDB).[120] There are no NMR structures available for 3 and 3-I, so for these peptides the first endpoint was chosen to be the putative global minimum obtained using the basin-hopping method.[121–123] The second endpoint was chosen to be an extended structure, which was obtained by simply minimising the energy of a conformation with all the backbone dihedral angles set to 180 degrees. All the stationary points (including these obtained during the connection procedure) were tightly converged to reduce the root-mean-squared force below $10^{-10}\,\mathrm{kcal\,mol^{-1}\,\AA^{-1}}$.

Each of the five trpzip pathway searches was conducted on a single Xeon 3.0 GHz CPU and required less than 24 hours of CPU time. The timings could certainly be improved by optimising the various parameters employed throughout the searches. However, it is more important that the connections actually succeed in a reasonably short time. It only requires one complete path to seed a DPS run, and we expect the DPS procedure to reduce the length of the initial path by a least a factor of two in sampling the largest contributions to the effective two-state rate constants. The results of all the trpzip calculations are shown in Figure 2.

3.4 Villin Headpiece Subdomain

The villin headpiece subdomain is the thermostable 35-residue C-terminal section of the headpiece domain of chicken villin protein. The sequence used for the NMR structure determination[124] included an additional methionine residue at the N-terminus from the expression system; thus, we are considering the 36-residue entity here, PDB code 1VII. The structure consists of a bundle of three short helices and a closely-packed hydrophobic core. The three helices are numbered from the N-terminus to the C-terminus. A turn connects helices 2 and 3, and helices 1 and 2 are joined by a loop.

Its small size and fast folding (a folding time on the order of tens of microseconds[125,126]) make the villin headpiece subdomain an attractive target for computational studies, including the 1 μs explicit-water MD simulation of Duan and Kollman.[127–139] Our DPS study of the villin headpiece subdomain employed the UNRES (united-residue) force field and model,[140–142] in which two interaction sites are assigned to each residue: one representing the main chain peptide group and one representing the side chain.

As for the GB1 hairpin, the construction of an initial path with the original connection algorithm proved difficult, and a considerable increase in both the efficiency and the success rate was attained with the Dijkstra-based strategy. The example path illustrated in Figure 4 contains 62 transition states and required 142 cycles of the Dijkstra-based algorithm. In this rearrangement, which connects a partially folded minimum to the locally minimised PDB structure, helices 2 and 3 are completed, the C-terminal residues pack correctly between helices 1 and 3, and the three helices adopt their native relative orientations.

4 Conclusions

Discrete path sampling calculations of effective two-state rate constants require an initial path consisting of local minima and the transition states that connect them.[2–4,112,143,144] For complex rearrangements the number of elementary steps involved may be rather large, and new methods for constructing an initial path are needed. This path does not need to be the shortest, or the fastest, but it does need to be fully connected. In the present work we have described a connection procedure based upon Dijkstra's shortest path algorithm, which enables us to select the most promising paths that include missing connections for subsequent double-ended searches. We have found that this approach, which is now implemented within the `OPTIM` package,[53] enables initial paths containing more than a hundred steps to be calculated automatically for a variety of systems. Some typical results have been presented for buckminsterfullerene, trpzip peptides, the GB1 hairpin, and the villin headpiece subdomain. These paths will be employed to seed future discrete path sampling calculations.

5 Acknowledgements

We gratefully acknowledge Dr David Evans for his Dijkstra-based utilities that locate shortest paths in connected databases of minima and transition states. S.A.T. is a Cambridge Commonwealth Trust/Cambridge Overseas Trust scholar. Most of the tryptophan zippers calculations were performed using computational facilities funded by the Isaac Newton Trust. J.M.C. is grateful to the EPSRC for financial support.

References

[1] E. W. Dijkstra, Numerische Math. **1**, 269 (1959).

[2] D. J. Wales, Mol. Phys. **100**, 3285 (2002).

[3] D. J. Wales, *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses*, Cambridge University, Cambridge (2003).

[4] D. J. Wales, Mol. Phys. **102**, 883 (2004).

[5] P. G. Mezey, Theo. Chim. Acta **58**, 309 (1981).

[6] H. Pelzer and E. Wigner, Z. Phys. Chem. **B15**, 445 (1932).

[7] H. Eyring, Chem. Rev. **17**, 65 (1935).

[8] M. G. Evans and M. Polanyi, Trans. Faraday Soc. **31**, 875 (1935).

[9] W. Forst, *Theory of Unimolecular Reactions*, Academic Press, New York (1973).

[10] K. J. Laidler, *Chemical Kinetics*, Harper & Row, New York (1987).

[11] J. N. Murrell and K. J. Laidler, Trans. Faraday Soc. **64**, 371 (1968).

[12] L. R. Pratt, J. Chem. Phys. **85**, 5045 (1986).

[13] R. Elber and M. Karplus, Chem. Phys. Lett. **139**, 375 (1987).

[14] R. S. Berry, H. L. Davis and T. L. Beck, Chem. Phys. Lett. **147**, 13 (1988).

[15] R. Czerminski and R. Elber, J. Chem. Phys. **92**, 5580 (1990).

[16] S. Fischer and M. Karplus, Chem. Phys. Lett. **194**, 252 (1992).

[17] L. L. Stachó and M. I. Bán, Theor. Chim. Acta **83**, 433 (1992).

[18] I. V. Ionova and E. A. Carter, J. Chem. Phys. **98**, 6377 (1993).

[19] L. L. Stachó and M. I. Bán, Theor. Chim. Acta **84**, 535 (1993).

[20] G. Dömötör, M. I. Bán and L. L. Stachó, J. Comput. Chem. **14**, 1491 (1993).

[21] C. Peng and H. B. Schlegel, Israeli J. Chem. **33**, 449 (1993).

[22] A. Matro, D. L. Freeman and J. D. Doll, J. Chem. Phys. **101**, 10458 (1994).

[23] O. S. Smart, Chem. Phys. Lett. **222**, 503 (1994).

[24] M. I. Bán, G. Dömötör and L. L. Stachó, J. Mol. Struct. (Theochem) **311**, 29 (1994).

[25] G. Mills and H. Jónsson, Phys. Rev. Lett. **72**, 1124 (1994).

[26] I. V. Ionova and E. A. Carter, J. Chem. Phys. **103**, 5437 (1995).

[27] C. Peng, P. Y. Ayala, H. B. Schlegel and M. J. Frisch, J. Comp. Chem. **17**, 49 (1996).

[28] H. Jónsson, G. Mills and W. Jacobsen, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, edited by B. J. Berne, G. Ciccotti and D. F. Coker, p. 385, World Scientific (1998).

[29] L. L. Stachó, G. Dömötör and M. I. Bán, Chem. Phys. Lett. **311**, 328 (1999).

[30] R. Elber and M. Karplus, Chem. Phys. Lett. **311**, 335 (1999).

[31] G. Henkelman and H. Jónsson, J. Chem. Phys. **111**, 7010 (1999).

[32] G. Henkelman, B. P. Uberuaga and H. Jónsson, J. Chem. Phys. **113**, 9901 (2000).

[33] G. Henkelman and H. Jónsson, J. Chem. Phys. **113**, 9978 (2000).

[34] G. Henkelman and H. Jónsson, J. Chem. Phys. **115**, 9657 (2001).

[35] R. A. Miron and K. A. Fichthorn, J. Chem. Phys. **115**, 8742 (2001).

[36] P. Maragakis, S. A. Andreev, Y. Brumer, D. R. Reichman and E. Kaxiras, J. Chem. Phys. **117**, 4651 (2002).

[37] B. Peters, A. Heyden, A. T. Bell and A. Chakraborty, J. Chem. Phys. **120**, 7877 (2004).

[38] S. A. Trygubenko and D. J. Wales, J. Chem. Phys. **120**, 2082 (2004).

[39] S. A. Trygubenko and D. J. Wales, The Journal of Chemical Physics **121**, 6689 (2004).

[40] G. M. Crippen and H. A. Scheraga, Arch. Biochem. Biophys. **144**, 462 (1971).

[41] J. Pancíř, Coll. Czech. Chem. Comm. **40**, 1112 (1974).

[42] R. L. Hilderbrandt, Comput. Chem **1**, 179 (1977).

[43] C. J. Cerjan and W. H. Miller, J. Chem. Phys. **75**, 2800 (1981).

[44] J. Simons, P. Jørgenson, H. Taylor and J. Ozment, J. Phys. Chem. **87**, 2745 (1983).

[45] A. Banerjee, N. Adams, J. Simons and R. Shepard, J. Phys. Chem. **89**, 52 (1985).

[46] J. Baker, J. Comp. Chem. **7**, 385 (1986).

[47] D. J. Wales, J. Chem. Phys. **101**, 3750 (1994).

[48] D. J. Wales and J. Uppenbrink, Phys. Rev. B **50**, 12342 (1994).

[49] L. J. Munro and D. J. Wales, Faraday Discuss. **106**, 409 (1997).

[50] L. J. Munro and D. J. Wales, Phys. Rev. B **59**, 3969 (1999).

[51] Y. Kumeda, L. J. Munro and D. J. Wales, Chem. Phys. Lett. **341**, 185 (2001).

[52] D. J. Wales and T. R. Walsh, J. Chem. Phys. **105**, 6957 (1996).

[53] The `OPTIM` program, along with other software for exploring and visualising potential energy surfaces, are available for download at http://www-wales.ch.cam.ac.uk/software.html.

[54] T. H. Cormen, C. E. Leiserson, R. L. Rivest and C. Stein, *Introduction to Algorithms*, MIT Press, Cambridge, Mass., 2nd edn. (2001).

[55] http://mathworld.wolfram.com/CompleteGraph.html (2004).

[56] Y. M. Rhee, J. Chem. Phys. **113**, 6021 (2000).

[57] J. R. Heath, S. C. O'Brien, R. F. Curl, H. W. Kroto and R. E. Smalley, Comments Condens. Matt. Phys. **13**, 119 (1987).

[58] H. W. Kroto, Science **242**, 1139 (1988).

[59] R. E. Smalley, Acc. Chem. Res. **25**, 98 (1992).

[60] K. G. Mckay, H. W. Kroto and D. J. Wales, J. Chem. Soc., Faraday Trans. **88**, 2815 (1992).

[61] J. R. Heath, in *Fullerenes–synthesis, properties, and chemistry of large carbon clusters*, edited by G. S. Hammond and V. J. Kuck, no. 481 in ACS Symposium Series, p. 1, Washington D. C. (1992), American Chemical Society.

[62] T. Wakabayashi and Y. Achiba, Chem. Phys. Lett. **190**, 465 (1992).

[63] C. Yeretzian, K. Hansen, F. Diederich and R. L. Whetten, Nature **359**, 44 (1992).

[64] S. W. McElvany, M. M. Ross, N. S. Goroff and F. Diederich, Science **259**, 1594 (1993).

[65] J. M. Hunter, J. L. Fye and M. F. Jarrold, Science **260**, 784 (1993).

[66] T. Wakabayashi, H. Shiromaru, K. Kikuchi and Y. Achiba, Chem. Phys. Lett. **201**, 470 (1993).

[67] G. von Helden, N. G. Gotts and M. T. Bowers, Nature **363**, 60 (1993).

[68] G. von Helden, M.-T. Hsu, N. Gotts and M. T. Bowers, J. Phys. Chem. **97**, 8182 (1993).

[69] J. M. Hunter, J. L. Fye and M. F. Jarrold, J. Chem. Phys. **99**, 1785 (1993).

[70] D. Babic and N. Trinajstic, Fullerene Sci. Technol. **2**, 343 (1994).

[71] J. M. Hunter and M. F. Jarrold, J. Am. Chem. Soc. **117**, 10317 (1995).

[72] D. E. Manolopoulos and P. W. Fowler, in *The far reaching impact of the discovery of* C_{60}, edited by W. Andreoni, vol. 316 of *NATO ASI Series, Series E.*, p. 51, Dordrecht, Netherlands (1996), NATO, Kluwer Academic Publ.

[73] D. Babic and N. Trinajstic, J. Mol. Struct. **376**, 507 (1996).

[74] D. L. Strout and G. E. Scuseria, J. Phys. Chem. **100**, 6492 (1996).

[75] K. R. Bates and G. E. Scuseria, J. Phys. Chem. A. **101**, 3038 (1997).

[76] H. W. Kroto, J. R. Heath, S. C. O'Brien, R. F. Curl and R. E. Smalley, Nature **318**, 162 (1985).

[77] W. Krätschmer, L. D. Lamb, K. Fostiropoulos and D. R. Huffman, Nature **347**, 354 (1990).

[78] X. Liu, D. J. Klein, W. A. Seitz and T. G. Schmalz, J. Comp. Chem. **12**, 1265 (1991).

[79] D. E. Manolopoulos, J. C. May and S. E. Down, Chem. Phys. Lett. **181**, 105 (1991).

[80] A. J. Stone and D. J. Wales, Chem. Phys. Lett. **128**, 501 (1986).

[81] J. Y. Yi and J. Bernholc, J. Chem. Phys. **96**, 8634 (1992).

[82] J. Baker and P. W. Fowler, J. Chem. Soc., Perkin Trans. 2 p. 1665 (1992).

[83] R. L. Murry, D. L. Strout, G. K. Odom and G. E. Scuseria, Nature **366**, 665 (1993).

[84] R. L. Murry, D. L. Strout and G. E. Scuseria, Int. J. Mass Spec. Ion. Proc. **138**, 113 (1994).

[85] K. Honda and E. Osawa, Fullerene Sci. Technol. **4**, 925 (1996).

[86] T. R. Walsh and D. J. Wales, J. Chem. Phys. **109**, 6691 (1998).

[87] E. Osawa, Z. Slanina and K. Honda, Mol. Cryst. Liq. Cryst. Sci. Technol. Sect. C **10**, 1 (1998).

[88] E. Osawa, Z. Slanina, K. Honda and X. Zhao, Fullerene Sci. Technol. **6**, 259 (1998).

[89] H. F. Bettinger, B. I. Yakobsen and G. E. Scuseria, J. Am. Chem. Soc. **125**, 5572 (2003).

[90] Y. Kumeda and D. J. Wales, Chem. Phys. Lett. **374**, 125 (2003).

[91] D. J. Wales, M. A. Miller and T. R. Walsh, Nature **394**, 758 (1998).

[92] J. Widany, T. Frauenheim, T. Koehler, M. Sternberg, D. Porezag, G. Jungnickel and G. Seifert, Phys. Rev. B **53**, 4493 (1996).

[93] D. Porezag, T. Frauenheim, T. Köhler, G. Seifert and R. Kaschner, Phys. Rev. B **51**, 12947 (1995).

[94] G. Seifert, D. Porezag and T. Frauenheim, Int. J. Quant. Chem. **58**, 185 (1996).

[95] P. A. Marcos, M. J. López, A. Rubio and J. A. Alonso, Chem. Phys. Lett. **273**, 367 (1997).

[96] C. H. Xu and G. E. Scuseria, Phys. Rev. Lett. **72**, 669 (1994).

[97] A. M. Gronenborn, D. R. Filpula, N. Z. Essig, A. Achari, M. Whitlow, P. T. Wingfield and G. M. Clore, Science **253**, 657 (1991).

[98] F. J. Blanco, G. Rivas and L. Serrano, Nat. Struct. Biol.**1**, 584 (1994).

[99] F. J. Blanco and L. Serrano, Eur. J. Biochem. **230**, 634 (1995).

[100] V. Muñoz, P. A. Thompson, J. Hofrichter and W. A. Eaton, Nature **390**, 196 (1997).

[101] V. Muñoz, E. R. Henry, J. Hofrichter and W. A. Eaton, Proc. Natl. Acad. Sci. USA **95**, 5872 (1998).

[102] A. R. Dinner, T. Lazaridis and M. Karplus, Proc. Natl. Acad. Sci. USA **96**, 9068 (1999).

[103] D. Roccatano, A. Amadel, A. Di Nola and H. J. C. Berendsen, Protein Sci. **8**, 2130 (1999).

[104] B. Ma and R. Nussinov, J. Mol. Biol. **296**, 1091 (2000).

[105] A. E. García and K. Y. Sanbonmatsu, Proteins: Struct., Func. and Gen. **42**, 345 (2001).

[106] R. Zhou, B. Berne and R. Germain, Proc. Natl. Acad. Sci. USA **98**, 14931 (2001).

[107] B. Zagrovic, E. J. Sorin and V. S. Pande, J. Mol. Biol. **313**, 151 (2001).

[108] R. Zhou and B. Berne, Proc. Natl. Acad. Sci. USA **99**, 12777 (2002).

[109] P. G. Bolhuis, Proc. Natl. Acad. Sci. USA **100**, 12129 (2003).

[110] G. Wei, P. Derreumaux and N. Mousseau, J. Chem. Phys. **119**, 6403 (2003).

[111] S. V. Krivov and M. Karplus, Proc. Nat. Acad. Sci. USA **101**, 14766 (2004).

[112] D. A. Evans and D. J. Wales, The Journal of Chemical Physics **121**, 1080 (2004).

[113] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, J. Comp. Chem. **4**, 187 (1983).

[114] T. Lazaridis and M. Karplus, Proteins: Struct., Func. and Gen. **35**, 133 (1999).

[115] A. G. Cochran, N. J. Skelton and M. A. Starovasnik, Proc. Natl. Acad. Sci. USA **98**, 5578 (2001).

[116] C. D. Snow, L. Qiu, D. Du, F. Gai, S. J. Hagen and V. S. Pande, Proc. Natl. Acad. Sci. USA **101**, 4077 (2004).

[117] D. Du, Y. Zhu, C. Huang and F. Gai, Proc. Natl. Acad. Sci. USA **101**, 15915 (2004).

[118] *http://www.trygub.com/charmm/rotamers/* (2004).

[119] J. D. Bloom, *Computer simulations of protein aggregation*, Master's thesis, University of Cambridge (2002).

[120] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, Nucleic Acids Research **28**, 235 (2000), the Protein Data Bank home page is http://www.rcsb.org/pdb.

[121] Z. Li and H. A. Scheraga, Proc. Natl. Acad. Sci. USA **84**, 6611 (1987).

[122] D. J. Wales and J. P. K. Doye, J. Phys. Chem. A. **101**, 5111 (1997).

[123] D. J. Wales and H. A. Scheraga, Science **285**, 1368 (1999).

[124] C. J. McKnight, P. T. Matsudaira and P. S. Kim, Nat. Struct. Biol.**4**, 180 (1997).

[125] J. Kubelka, W. A. Eaton and J. Hofrichter, J. Mol. Biol. **329**, 625 (2003).

[126] M. Wang, Y. Tang, S. Sato, L. Vugmeyster, C. J. McKnight and D. P. Raleigh, J. Am. Chem. Soc. **125**, 6032 (2003).

[127] Y. Duan and P. Kollman, Science **282**, 740 (1998).

[128] Y. Duan and P. Kollman, Proc. Natl. Acad. Sci. USA **95**, 9897 (1998).

[129] M. R. Lee, Y. Duan and P. A. Kollman, Proteins: Struct., Func. and Gen. **39**, 309 (2000).

[130] D. C. Sullivan and I. D. Kuntz, J. Phys. Chem. B **106**, 3255 (2002).

[131] M. Y. Shen and K. F. Freed, Proteins: Struct., Func. and Gen. **49**, 439 (2002).

[132] E. Z. Wen, M.-J. Hsieh, P. A. Kollman and R. Luo, J. Mol. Graphics & Modelling **22**, 415 (2004).

[133] S. Jang, E. Kim, S. Shin and Y. Pak, J. Am. Chem. Soc. **125**, 14841 (2003).

[134] A. Fernández, M.-Y. Shen, A. Colubri, T. R. Sosnick, R. S. Berry and K. F. Freed, Biochemistry **42**, 664 (2003).

[135] A. Mukherjee and B. Bagchi, J. Chem. Phys. **118**, 4733 (2003).

[136] A. Mukherjee and B. Bagchi, J. Chem. Phys. **120**, 1602 (2004).

[137] S. A. Islam, M. Karplus and D. L. Weaver, J. Mol. Biol. **318**, 199 (2002).

[138] V. S. Pande, I. Baker, J. Chapman, S. P. Elmer, S. Khaliq, S. M. Larson, Y. M. Rhee, M. R. Shirts, C. D. Snow, E. J. Sorin and B. . Zagrovic, Biopolymers **68**, 91 (2003).

[139] B. Zagrovic, C. D. Snow, M. R. Shirts and V. Pande, J. Mol. Biol. **323**, 927 (2002).

[140] A. Liwo, S. Ołdziej, M. R. Pincus, R. J. Wawak, S. Rackovsky and H. A. Scheraga, J. Comput. Chem. **18**, 849 (1997).

[141] A. Liwo, M. R. Pincus, R. J. Wawak, S. Rackovsky, S. Ołdziej and H. A. Scheraga, J. Comput. Chem. **18**, 874 (1997).

[142] A. Liwo, R. Kaźmierkiewicz, C. Czaplewski, M. Groth, S. Oldziej, R. J. Wawak, S. Rackovsky, M. R. Pincus and H. A. Scheraga, J. Comp. Chem. **19**, 259 (1998).

[143] D. A. Evans and D. J. Wales, J. Chem. Phys. **119**, 9947 (2003).

[144] F. Calvo, F. Spiegelman and D. J. Wales, J. Chem. Phys. **118**, 8754 (2003).

Figure Captions

1. Top: connected path between a high-energy C_{60} cluster and buckminsterfullerene. The path contains 82 transition states, and required 383 cycles of the Dijkstra-based connection algorithm, including 1620 DNEB searches. V is the potential energy in hartree and s is the integrated path length in bohr. Bottom: enlargement of the above plot for the last two SW rearrangements. The atoms mainly involved in these two steps are shaded, and both local minima and transition state structures are indicated at appropriate points along the path.

2. Energy profiles for native to denatured state rearrangements of tryptophan zippers found by the Dijksta-based connection algorithm. For each profile the number of steps in the pathway, the number of connection algorithm cycles, the total number of DNEB searches and the total number of stationary points in the database (recorded upon termination of the algorithm) are shown. The total number of stationary points is presented in the form (n, m), where n is the number of minima and m is the number of transition states. The potential energy, V, is given in the units of kcal/mol, and the integrated path length, s, is given in the units of Å.

3. Connected path between non-native and β-hairpin conformations for the GB1 peptide. The path contains 163 transition states, and required 126 cycles of the Dijkstra-based connection algorithm, including 1610 DNEB searches. V is the potential energy in kcal/mol and s is the integrated path length in Å.

4. Connected path between non-native and native conformations for the villin headpiece subdomain. The two endpoint minima are shown with the N-terminal helix (helix 1) on the left. The other two helices and the C-terminus are also labelled. The path contains 62 transition states, and required 142 cycles of the Dijkstra-based connection algorithm, including 294 DNEB searches. V is the potential energy in kcal/mol and s is the minimised Euclidean distance between consecutive stationary points in Å. For this system we simply join the stationary points with straight lines.

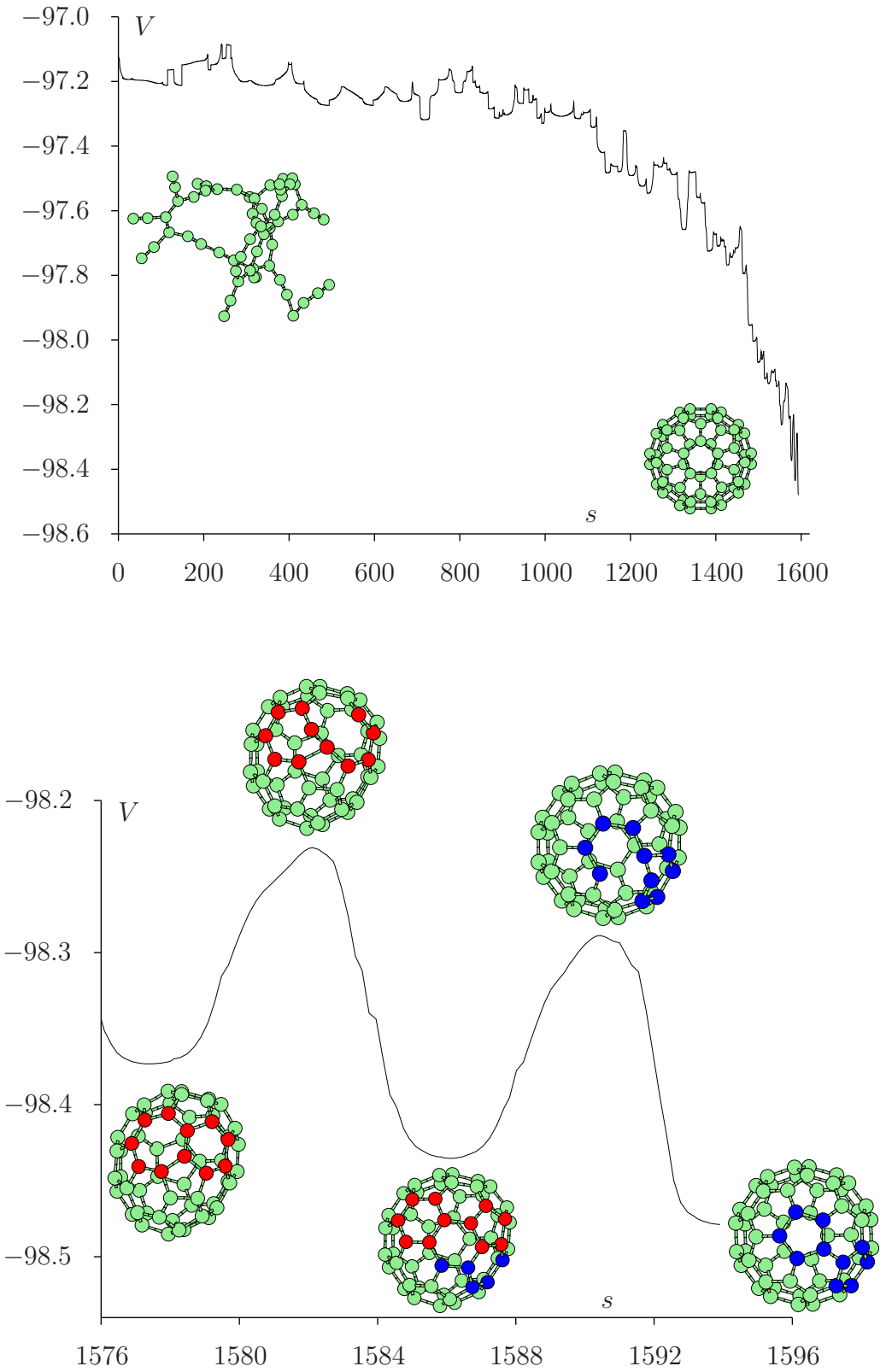
−97.0
−97.2
−97.4
−97.6
−97.8
−98.0
−98.2
−98.4
−98.6
V
s
0
200
400
600
800
1000
1200
1400
1600
−98.2
−98.3
−98.4
−98.5
V
s
1576
1580
1584
1588
1592
1596

Figure 1:

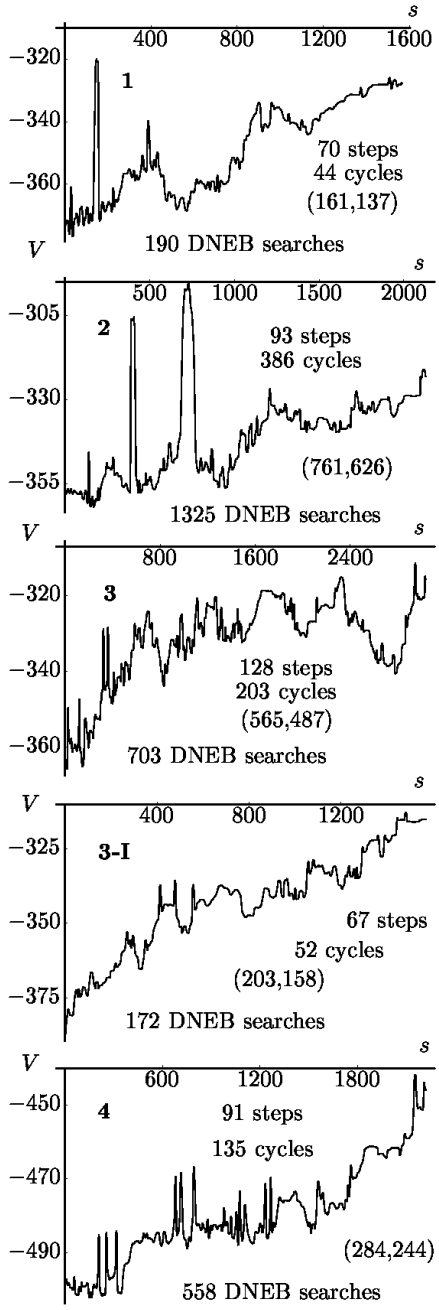
s
400
800
1200
1600
−320
−340
−360
V
1
70 steps
44 cycles
(161,137)
190 DNEB searches
s
500
1000
1500
2000
−305
−330
−355
2
93 steps
386 cycles
(761,626)
1325 DNEB searches
V
s
800
1600
2400
−320
−340
−360
3
128 steps
203 cycles
(565,487)
703 DNEB searches
V
s
400
800
1200
−325
−350
−375
3-I
67 steps
52 cycles
(203,158)
172 DNEB searches
V
s
600
1200
1800
−450
−470
−490
4
91 steps
135 cycles
(284,244)
558 DNEB searches

FIGURE 2:

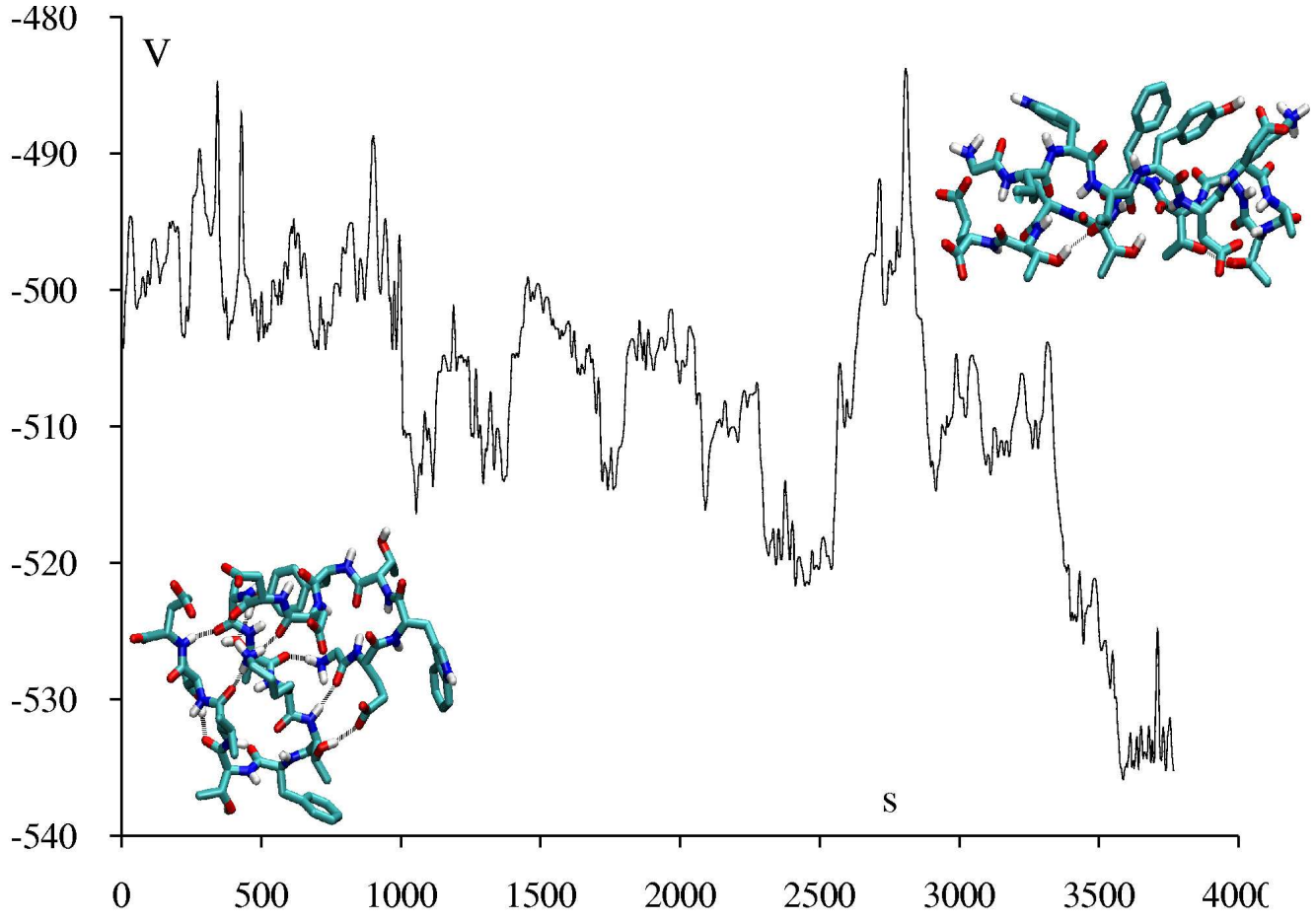
-480
-490
-500
-510
-520
-530
-540
V
s
0
500
1000
1500
2000
2500
3000
3500
4000

Figure 3:

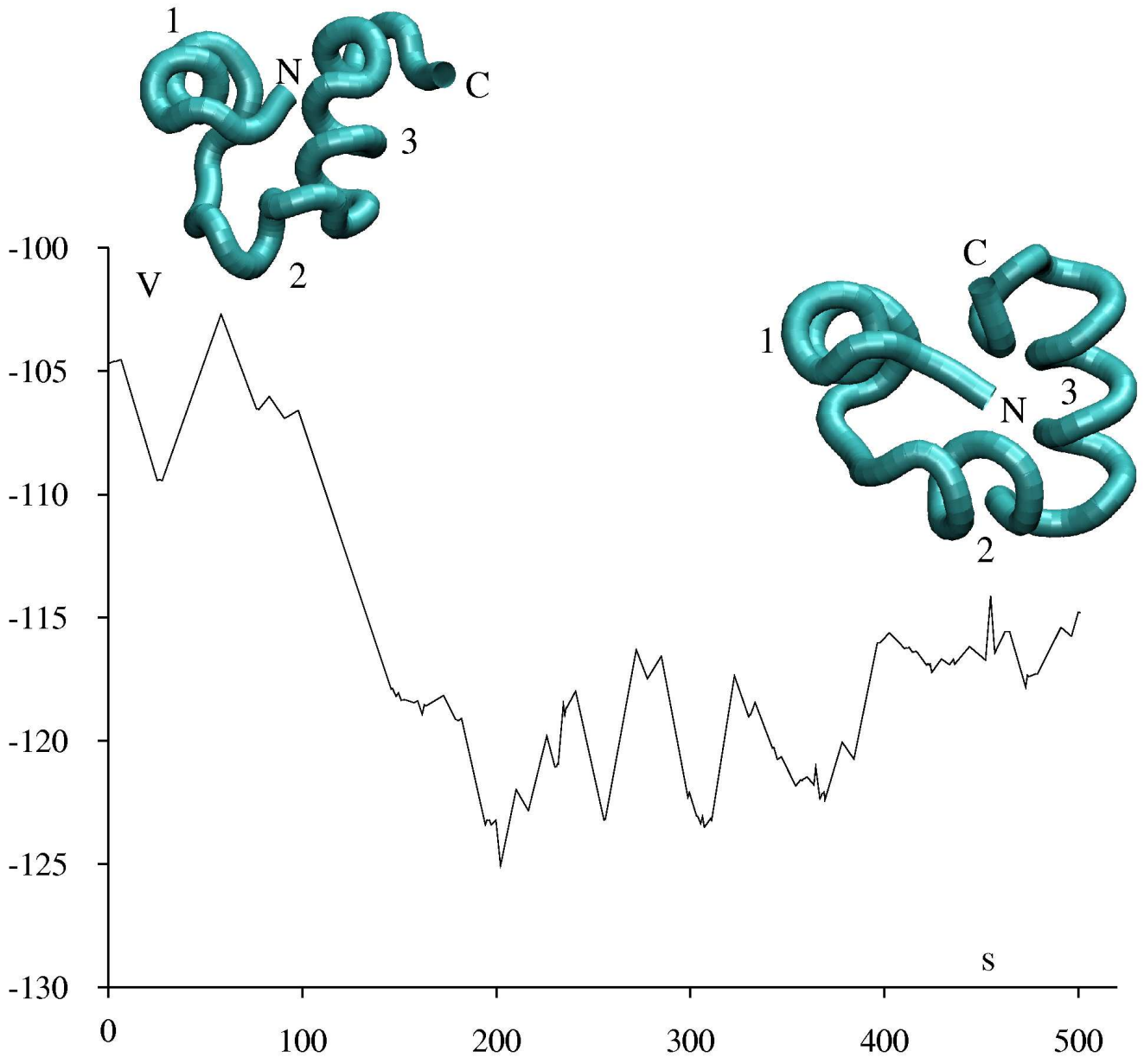

Figure 4: