Eur. Phys. J. D **16**, 47–50 (2001)

THE EUROPEAN PHYSICAL JOURNAL D

EDP Sciences

Connecting cluster dynamics and protein folding

R.S. Berry[1,a], A. Fernandez[2], and K. Kostov[1]

[1] The University of Chicago, Chicago, Illinois 60637, USA
[2] Instituto de Matematica, Universidad Nacional del Sur, Bahia Blanca, Argentina

Received 4 December 2000

Abstract. The relaxation dynamics of clusters can be interpreted in terms of the topographies of their potential surfaces. Systems with short-range potentials have sawtooth-like potential surfaces with small drops in energy from one local minimum to the next and few-body motions as the clusters move from one minimum to another; such systems readily take on amorphous structures. These are called "glass-formers". Systems with long-range forces have potentials whose topographies are like rough staircases, with some large drops in energy from one minimum to the next; their well-to-well passages involve very collective motions and such systems are excellent structure-seekers. They find their way to well-ordered, highly selective structures under almost all circumstances. These characteristics generalize to describe the potential surfaces and folding behavior of polypeptides and proteins. The forces are *effective* long-range forces due to the polymer chain. Staircase topographies emerge both from direct sampling of potential surfaces and from the inversion of the kinetics generated by a much more abstract topological model, from which folding pathways can be inferred.

PACS. 36.40.-c Atomic and molecular clusters – 87.10.+e General theory and mathematical aspects – 87.15.-v Biomolecules: structure and physical properties

1 Introduction

A central but relatively ill-studied issue in the science of clusters and other pico-scale systems is finding the relationships among the interparticle forces, the topographies of the multidimensional potential surfaces and the dynamics and relaxation behavior of these systems. Put in very coarse terms, we may ask, "Why do some systems, such as argon clusters, tend to assume amorphous structures while others, such as alkali halide clusters, find their way just to any of the very small fraction of locally-stable structures that are rock-salt crystals? Why are some systems "glass-formers" and others "structure-seekers"?" What is it about the interparticle forces that makes a topography that leads the system to a select few possible structures, or, on the other hand, that leaves a system in any of a very large number of locally-stable structures with no particular selectivity or structure type?

A question that follows naturally from these is the consequence of the attention currently focused on the folding of proteins: Are the reasons that explain why some clusters are structure-seekers the same reasons that underlie the remarkable selectivity that many natural proteins exhibit when they fold into physiologically active structures? Do proteins that are good, efficient folders have potential surfaces with topographies similar to those of clusters that are good structure-seekers?

[a] e-mail: berry@uchicago.edu

We shall see here how, at least qualitatively, we can relate interparticle forces to topographies, dynamics and relaxation properties of clusters. We then go on to see how these concepts extend as well to proteins, whether the potential surface is postulated or inferred from an abstract model of folding dynamics.

2 Atomic clusters: Glass-formers and structure-seekers

There is a dramatic difference between the behavior of argon clusters and alkali halide clusters when they, in simulations, are cooled from their liquid states. The alkali halide clusters find their way to rocksalt structures even when cooled at rates up to 10^{13} K/s, roughly five vibrational periods, although their rocksalt minima are outnumbered by amorphous minima roughly 10^{12}:1 [1]. By contrast, argon clusters cooled even at 10^9 K/s find themselves distributed among many amorphous structures; the global minimum for Ar_{19} can be the most populous geometric state under such conditions but never collects even half the population of the simulated sample [2]. These results stimulated the realization that it could be worthwhile to investigate what attainable characteristics of the topographies of multidimensional potential surfaces would tell us the extent to which a pico-scale system would be a glass-former or a structure-seeker.

Pursuing this problem had, by 1996, become a tractable problem because relatively efficient methods had become available to locate not only local minima but saddle points as well, if one were given a moderately simple form for a potential surface, even for a 50- or 100-particle cluster. Fortunately, it had become evident that the Lennard-Jones potential gave a rather good representation of the potential for argon clusters, and the Born-Mayer potential or others similar to it [3], for alkali halide clusters.

The format that made the collected sets of minima and saddles linking them particularly useful was a projection onto a plane of the linked successions of stationary points, with the minima in monotonic sequences of increasing energy, from the lowest minimum in a basin [4, 5]. This representation immediately provided new insights: a) it revealed that the multidimensional surface could be thought of as consisting of large, rough-walled basins with some cross-links between local minima in different monotonic sequences; b) it showed that the patterns of these monotonic sequences of stationary points were quite different for glass-formers and structure-seekers. The patterns for glass-formers are sawtooth-like; the patterns for structure-seekers are like rough staircases [2].

This analysis, carried a bit further, revealed even richer information. By following the structures of the clusters at the successive stationary points along arbitrarily-selected sequences from high to low energies, it was easy to see the amorphous structures in which the glass-formers became trapped, and equally easy to see how the large drops in energy along the staircases of the structure-seekers were associated with the formation of crystalline nuclei. Then, by applying a criterion developed by Stillinger and Weber [6], essentially the summed fourth powers of the distances all the particles move, divided by the square of the sum of the corresponding squares, one could evaluate the effective number of particles moving when the system passes from one minimum to the next. The rare gas clusters, with particles bound by Lennard-Jones forces, exhibited almost exclusively few-body motions, two to four particles out of 19, for example, in those interwell passages, while the alkali halide clusters, bound by long-range Coulomb forces, exhibited extremely collective, many-body motions in their interwell transformations. As many as half the particles of a $(KCl)_{32}$ cluster may take part in the motion [2].

This set of relations gives a consistent qualitative picture of the relations among the interparticle forces, the topography of the multidimensional potential surface, the dynamics of motion and relaxation on that surface and the tendency of a system to be a glass-former or a structure-seeker. While a systematic quantification of these ideas still lies ahead, some steps have been made in that direction [7,8].

The relationship between the range of the potential and the structure-seeking or glass-forming character also remains to be quantified systematically. Some results from simulation give us the first insights into this question. If the long-range Coulomb potential of the alkali halide cluster is replaced by a shielded Coulomb potential, $q_1 q_2 e^2 e^{-\beta R}/R$, then the surface becomes sawtooth-like and the system becomes a glass-former if the shielding distance β^{-1} is about the distance of the second-nearest neighbor [1].

3 Topographies for protein models

The foregoing discussion implies that a simple, structure-seeking cluster such as $(KCl)_{32}$ finds a crystalline local-minimum structure against statistical odds even greater than might be expected for a peptide of the same number of atoms "seeking" a specific folded structure. This observation immediately suggests the next step of our investigation: do the same generalizations that describe the behavior of clusters apply also to proteins? To investigate this, we first examined a model system, a 46-bead model that had been introduced as a lattice model by Skolnick and his collaborators [9] and then extended to a continuum model by Honeycutt and Thirumalai [10]. This system consists of four rather rigid strands, two made of only hydrophobic beads and two, of alternating hydrophobic and hydrophilic beads; the strands are linked by three segments of "neutral" beads, that have soft bending angles to enable the strands to align in antiparallel sequence. The forces between nonbonded (non-neighboring) beads simulate forces that make hydrophobic beads attract one another and hydrophobic beads to turn outward, as if to an aqueous solvent. The system has a global minimum with β-barrel structure. If the system is simulated with an initial state chosen at random or with the strands stretched into something of a linear form, then it can find its way to the global minimum *or to any of several other, very similar β-barrel structures*. These structures all have very similar energies.

With sufficiently slow annealing, the system can find its way to the global minimum, but it resembles the alkali halides insofar as it does not distinguish readily among the β-barrel structures, just as the alkali halide cluster does not distinguish readily among its rocksalt structures. In the sense that this system does not go naturally and efficiently to one single "native" structure, it is sometimes said to be not a good folder; in the sense that it almost invariably goes to one of the low-energy, β-barrel structures in any isothermal or moderate-speed annealing simulation, it is a good folder. (This system raises a question that seems to arise with increasing frequency in the field of protein structure and dynamics: Are "native" structures of proteins unique, or can they exhibit some variety of structures, especially in the scaffolding distant from active sites, and remain physiologically active?)

The potential surface of this 46-bead model can be searched, just like those of clusters, for minima and the saddles that link them, and, from these, one can construct statistical samples of sequences of stationary points that are monotonic in the energies of their local minima. Such sequences give staircase patterns very much like those of alkali halide clusters with about the same numbers of particles. The patterns of topographies leading into the various basins, with their slightly different

β-barrel minimum-energy structures, are virtually indistinguishable. While the explicit forces between nonbonded beads are Lennard-Jones-like in this model, the presence of the polymer chain makes all the forces in this system effectively long-range. The effective numbers of moving particles in inter-well passages are almost always large, as in the structure-seeking clusters. There are a few motions that involve only "flopping" of the few beads in the soft neutral links, but most well-to-well motions involve about half the particles in the system, or more. In short, the same generalizations that distinguish structure-seeking clusters characterize a structure-seeking protein model as well [11].

4 Protein folding and topography: A backward inference

It is quite feasible to examine topographies of small polypeptides with force field models that include all the atoms and even, if one chooses, surrounding water molecules. Such examinations show, for example, that small alanine polymers such as the tetramer have potential surfaces with moderately good staircase topographies – not as staircase-like as the surface of the 46-bead model but nonetheless more staircase-like than sawtooth. However because of the way complexity increases with the size of the system, and the uncertainty of the validity of only a small statistical sample of the sequences of stationary points, it is useful at this point to turn to an approach that reveals the character of the topography of a protein's potential surface, albeit usually averaged over a number of trajectories, and links that information to the folding character of the protein.

This is a method that begins with a topological representation of the protein's backbone in terms of the torsion angles ϕ and ψ that can line only in specific regions – the Ramachandran basins – for each amino acid residue. Each of the four possible basins corresponds to a configuration, crudely *cis* or *trans*, for each angle. Most amino acids have locally stable forms in three of the four Ramachandran basins; glycine, the simplest, with no side groups to create steric hindrance, can be in any of the four possible configurations. Proline, with its rigid ring structure, can exist in only two basins, and any residue following proline can also exist in only two basins, one of which is the same as that of proline. The model designates the nature of the side group on each residue and then, at each instant, assigns a basin – a torsional configuration – to each residue in the chain. The configuration is allowed to change, undergoing "flips" at a rate with a mean value of 10^{11}/s, and is examined every 64 ps to see whether any sequence of six or more residues has a configuration that forms a pattern consistent with a bit of secondary structure. Whenever such a pattern appears, the mean flipping rate of the residues in that group drops to 10^7/s. Likewise, when tertiary structure forms, the mean rate drops to 10^3/s. If, by contrast, any time 30% or more of a pattern loses its pattern, the entire unit goes to the faster flipping rate. These rates were all inferred from empirical generalizations from experiment. The method has been summarized in detail elsewhere [12–14].

One of the first proteins to which this approach was applied was bovine pancreatic trypsin inhibitor (BPTI) [13, 14]. The analysis showed the molecule folding to its native structure, with a slow intermediate period and with disulfide bonds that apparently assist the folding process but that break, so that the sulfur atoms find new partners in the late stages to achieve the native structure. Moreover there is one very brief intermediate period after a few milliseconds of slow folding, of order 10^{-7} s, in which a great deal of structure forms, in a sort of fast, highly collective collapse.

The occurrence of identifiable stages during the folding process make it possible to estimate forward rates for passage from one stage to the next. This information is complemented by the corresponding reverse rates, which are based on the premise that if 30% of a secondary-structured region dismantles, the entire structured region returns to random coil. From consideration of detailed balance and the knowledge of forward and backward rates between two stages, it becomes straightforward to construct the effective potential barrier connecting the two stages of folding [15]. In this way, a mean potential topography emerged from the computations of the "dihedral flipping" or topological model for the folding of BPTI. Characteristic of a good folder, this effective potential is indeed staircase-like.

One further example that has been analyzed in a similar way is the rather special case of β-lactoglobulin, which passes from a random coil to a stage with considerably more α-helix than occurs in the final, native state. This was the inference of transient circular dichroism measurements. Later, there were objections to this inference because no stable intermediate could be found with a large excess of α-helical structure. The topological simulation shows [16] that there is a period of about 0.1 μs during which there is considerable *transient* α-helix in rapid dynamic equilibrium with random coil. This persists until a bit of tertiary structure forms to stabilize the α-helix long enough for it to pass to the entropically more stable β-sheet of the native, final structure. The topography of the effective potential surface for this system was also determined as it had been for BPTI, in this case for both the most frequent set of paths and for the least frequent of the paths that reach the native structure. Both are very staircase-like, and are strikingly similar – not identical, by any means, but with relatively flat regions in roughly the same places along their time paths. As with clusters and with BPTI, the sharp drops at the staircase "risers" are associated with the nucleation of regions of ordered structure.

5 Conclusions

The central conclusion of this work is a qualitative, even semiquantitative set of connections linking the interparticle forces, the topographies of complex potential surfaces, the modes of nucleation and relaxation, and the tendency of a picoscale system to be a glass-former or a

structure-seeker. The systems with short-range forces have sawtooth potentials, only few-body motions in their well-to-well passages, little tendency to nucleate, and are, as a result, glass-formers. Systems with long-range or effective long-range forces have staircase potentials, highly collective, many-body motions in their well-to-well passages, considerable tendency to nucleate ordered regions and are structure-seekers. It remains ahead to determine how to quantify the extent of sawtooth or staircase character of a potential surface, and to make a quantitative link between that and the ranges of the interparticle forces.

We would like to acknowledge the contributions of Andrés Colubri, Ralph Kunz and Keith Ball to this work, and to thank Tobin Sosnick for helpful discussions. The research was supported by a Grant from the National Science Foundation.

References

1. J. Rose, R.S. Berry, J. Chem. Phys. **98**, 3262 (1993).
2. K.D. Ball, R.S. Berry, R.E. Kunz, F.-Y. Lin, A. Proykova, D.J. Wales, Science **271**, 963 (1996).
3. M.P. Tosi, F.G. Fumi, J. Phys. Chem. Sol. **25**, 45 (1964).
4. R.S. Berry, R.E. Breitengraser-Kunz, Phys. Rev. Lett. **74**, 3951 (1995).
5. R.S. Berry, R.E. Kunz, J. Chem. Phys. **103**, 1904 (1995).
6. F.H. Stillinger, T.A. Weber, J. Chem. Phys. **80**, 4434 (1984).
7. J.P. Doye, D.J. Wales, Phys. Rev. Lett. **80**, 1357 (1998).
8. J.P. Doye, D.J. Wales, M.A. Miller, J. Chem. Phys. **109**, 8143 (1998).
9. J. Skolnick, A. Kolinsky, R. Yaris, Proc. Natl. Acad. Sci. USA **85**, 5057 (1988).
10. J.D. Honeycutt, D. Thirumalai, Proc. Natl. Acad. Sci. USA **87**, 3526 (1990).
11. R.S. Berry, N. Elmaci, J.P. Rose, B. Vekhter, Proc. Natl. Acad. Sci. USA **94**, 9520 (1997).
12. A. Fernández, A. Colubri, J. Math. Phys. **39**, 3167 (1998).
13. A. Fernández, K. Kostov, R.S. Berry, Proc. Natl. Acad. Sci. USA **96**, 12991 (1999).
14. A. Fernández, R.S. Berry, J. Chem. Phys. **112**, 5212 (2000).
15. A. Fernández, K. Kostov, R.S. Berry, J. Chem. Phys. **112**, 5223 (2000).
16. A. Fernández, A. Colubri, R.S. Berry, Proc. Natl. Acad. Sci. USA **97**, 14062 (2000); A. Fernández, A. Colubri, R.S. Berry, J. Chem. Phys. **114**, 5871 (2001).