

CHAPTER 5

CONCLUSIONS

Still round the corner there may wait
A new road or a secret gate,
And though we pass them by today,
Tomorrow we may come this way
And take the hidden paths that run
Towards the Moon or to the Sun.

J. R. R. Tolkien, A walking song

5.1 SUMMARY OF CONTRIBUTIONS

In this thesis I have tried to improve upon existing double-ended methods for finding rearrangement pathways as well as methods for extracting kinetic information from pathway ensembles. The main accomplishments are as follows:

- (I) We presented a graph transformation (GT) method, which can be used to calculate the total transition probabilities and mean escape times for arbitrary digraphs with arbitrary sets of sources and sinks that are allowed to overlap. At low temperatures the GT method becomes the method of choice, outperforming kinetic Monte Carlo and matrix multiplication methods.
- (II) We have suggested a version of the GT method (SDGT) that can take full advantage of the sparsity of the problem. Apart from switching to the standard sparse-optimised adjacency-list-based data structure, the modifications were the

implementation of Fibonacci-heap-based min-priority queue to ensure that nodes with smaller degrees are detached first, and an algorithm that monitors the graph density and switches to the dense-optimised version of the method when it becomes computationally cost-effective.

- (III) The stability of the NEB method was improved by introducing a portion of the spring gradient component perpendicular to the path back into the NEB gradient.
- (IV) The efficiency of the DNEB method was improved by eliminating the removal of the overall rotation and translation and employing a quasi-Newton method (L-BFGS) for minimisation of the band.
- (V) The efficiency and stability of the QVV minimiser was increased by finding the optimal point in time of quenching the velocity.
- (VI) We have devised a method for finding rearrangement pathways between distant local minima, which is based on the consecutive DNEB searches and uses the Euclidean distance as a measure of separation in configuration space. Part of this work that concerned the Dijkstra-based selector was performed in collaboration with Dr. Joanne M. Carr [167].
- (VII) A new cooperativity index, introduced in Chapter 3, enabled us to find a correlation between the cooperativity of an atomic rearrangement and the energy barrier. We showed that cooperative rearrangements of LJ clusters and the BLJ liquid have lower energy barriers irrespective of the degree of localisation.
- (VIII) We have demonstrated that it is possible to control the overall cooperativity of the pathway sample, and outlined a technique for sampling cooperative pathways using single-ended transition state searching methods.
- (IX) We have described the edge weight function that allows us to find the path with the largest DPS non-recrossing rate constant using the Dijkstra algorithm. We have also described an algorithm for sampling for the fastest paths.
- (X) We have devised a method for computing a recrossing contribution to the DPS rate constant exactly in linear time with constant memory requirements.

- (XI) We have obtained recursive expressions for the total transition probabilities from an arbitrary digraph by considering the corresponding pathway ensembles.

Because random walk has applications in many areas of sciences, from Brownian motion [120] and diffusion [121] in physics to dynamics of stock markets in economics [281] and tumour angiogenesis in medicine [282], we expect methods developed in Chapter 4 to be relevant to a much wider domain.

5.2 FUTURE RESEARCH

Avenues for future research based on the results of this thesis may be opened by trying to answer the following questions:

- (I) If a pathway sample that accurately reproduces the kinetics of the complete pathway ensemble is sought what is the best sampling strategy to use?
- (II) Do cooperative rearrangements start to dominate the relaxation processes at lower temperatures?
- (III) What is the relationship between the number of cooperative pathways supported by a PES and the form of the potential?
- (IV) Can cooperative moves improve existing methods for global optimisation and evolution of kinetic databases?

As a number of alternative double-ended approaches have been developed in the past few years, such as, for example, the string method [283–287], the growing string method [109, 288], and a super-linear NEB based on adopted basis Newton-Raphson minimiser [289], it would be interesting to make a detailed comparison of these methods on a set of problems we are likely to be solving in the future.

The connection algorithm and the algorithm for sampling the fastest path presented in Chapter 2 and Appendix E, respectively, are far from optimum because they operate on evolving databases but use the static Dijkstra algorithm to build the shortest path tree. When applying these methods to databases larger than these discussed in this thesis the use of dynamic graph algorithms may be of benefit.

It would be exciting to see more applications of the GT method that would allow us to come to a more detailed understanding of its strengths and weaknesses. Comparisons with sparse-optimised numerical approaches for solving the master equation and iterative solvers based on first-step analysis are also desirable.

From a theoretical point of view, it would be interesting to extend the approach of Section 4.3 to obtain higher moments of the escape time distribution function and maybe even the distribution function itself.

Although the DNEB method and connection algorithm presented in this thesis have seen a number of successful applications already the scope for further applications and development is ample. Potential areas include the design of better initial pathway guessing strategies for proteins, reduction of memory requirements of the connection algorithm, better understanding of the relationship of the optimal edge weight function and the form of the potential, and parallelisation of these methods for distributed-memory computing, to name but a few. Much ongoing work is now focused on attempts to construct an initial folding pathway for a medium-sized protein, barnase. Preliminary results showed that further improvements (with emphasis on the large number of degrees of freedom) to both double-ended and single-ended transition state searching methods are required to complete this task.

5.3 OUTLOOK

There is a need for future research to address the gap between the force field designers and the energy landscapes community. Success in the application of many methods discussed in this thesis is contingent on the potential energy function being continuous and well-defined. Some of the most promising potential energy functions developed recently can only be used with methods described here after serious modifications.

Ultimately, our ability to make valid predictions about the properties of any system is limited by the accuracy of the force field. It is thus hoped that more research will be performed in the direction of the development of realistic, cheap and friendly force fields.