Chapter 1

Introduction

> The Dark Side of the Force is a pathway to
> many abilities some consider to be unnatural.
>
> ―――――――――――――――――
>
> George Lucas, Revenge of the Sith

Knowledge of the potential energy surface* (PES) and the ability to use this knowledge grant extraordinary powers of prediction about the structure, dynamics and thermodynamics of any molecular system [9]. A potential, also known as a force field, is used to formally specify the PES in theoretical studies.

1.1 The Force Field

A general force field can be written as a series of terms representing the interactions between increasingly large sets of atoms [34, 35]:

$$V = \epsilon^{(0)} + \sum_{\alpha}^{N} \epsilon_{\alpha}^{(1)} + \sum_{\alpha}^{N} \sum_{\beta<\alpha}^{N} \epsilon_{\alpha,\beta}^{(2)} + \sum_{\alpha}^{N} \sum_{\beta<\alpha}^{N} \sum_{\gamma<\beta}^{N} \epsilon_{\alpha,\beta,\gamma}^{(3)} + \cdots , \qquad (1.1)$$

where N is the total number of atoms, and the two-body term $\epsilon_{\alpha,\beta}^{(2)}$, for instance, describes the interaction of two atoms α and β.

Three-body and higher order terms in Equation 1.1 are often neglected, such as, for

―――――――――――――――――

*In this thesis terms 'potential energy landscape' and PES are used synonymously. At the time of writing searching for "potential energy surface" and "potential energy landscape" in Google® yielded 259,000 and 17,800 hits, respectively.

example, in the Lennard-Jones (LJ) pair potential [9, 36], which takes the form

$$V = 4\epsilon \sum_{\beta<\alpha}^{N} \left[\left(\frac{\sigma}{r_{\alpha,\beta}} \right)^{12} - \left(\frac{\sigma}{r_{\alpha,\beta}} \right)^{6} \right], \tag{1.2}$$

where $r_{\alpha,\beta}$ is the distance between atoms α and β, ϵ is the depth of the potential energy well, and $2^{1/6}\sigma$ is the pair equilibrium separation. This is an approximate potential as its form is a trade-off between the accurate reproduction of the interaction between closed-shell atoms and mathematical and computational simplicity. In this thesis we will use it to describe atomic clusters of various sizes.

1.2   CREATING A COARSE-GRAINED MODEL

It is often possible to gain new insight into the properties of a molecular system by expressing them in terms of stationary points of the PES, i.e. points where the gradient of the potential vanishes [9, 37]. Such a coarse-grained picture may be appropriate if the system spends most of its time in the vicinity of these points and the properties of interest can be expressed in terms of the properties of these points only. In realistic applications it may also be the only way forward, as the corresponding PES's are usually complex.

The most important stationary points are minima and the transition states that connect them. Here we define a minimum as a stationary point where the Hessian, the second derivative matrix, has no negative eigenvalues, while a transition state is a stationary point with precisely one such eigenvalue [38].

The number of stationary points on the PES generally scales exponentially with system size [39–43], which necessitates an appropriate sampling strategy of some sort for larger systems. In particular, to analyse kinetic properties a representative sample is usually obtained, which generally involves extensive use of single-ended and double-ended transition state searching techniques [7, 9, 30].

Locating transition states on a PES also provides an important tool in the study of dynamics using statistical rate theories [44–48]. Unfortunately, it is significantly harder to locate transition states than local minima, since the system must effectively 'balance on a knife-edge' in one degree of freedom. Many algorithms have been suggested for this purpose, and the most efficient method may depend upon the nature of the system. For

example, different considerations probably apply if second derivatives can be calculated relatively quickly, as for many empirical potentials [9]. Transformation to an alternative coordinate system may also be beneficial for systems bound by strongly directional forces [49–57].

Single-ended transition state searches [12–23, 54, 58–86] only require an initial starting geometry. The result of a single-ended search may be a transition state that is not connected to the starting point by a steepest-descent path, and such methods can be useful for building up databases of stationary points to provide a non-local picture of the potential energy surface, including thermodynamic and dynamic properties [9, 86]. However, double-ended searches [27–29, 53, 87–107] require two endpoint geometries, a mechanism to generate a set of configurations between them, and a suitable functional (or gradient) to be evaluated and minimised. The most successful single- and double-ended methods currently appear to be based upon hybrid eigenvector-following [12–24] and the nudged elastic band approach [7, 27, 29, 30, 108–110], respectively. The two search types are often used together, since double-ended transition state searches do not produce a tightly converged transition state and further refinement may be needed [7, 9].

1.3 WORKING WITH A COARSE-GRAINED MODEL

In Chapter 2 and Chapter 4 of this thesis coarse-grained models of a PES are discussed in graph-theoretical terms. Nowadays a flourishing branch of mathematics and computer science, graph theory arguably started in the year of 1736 with Leonhard Euler's paper on the seven bridges of Königsberg, where he abstracted from landmasses and bridges to highlight the connectivity, and proved that it is impossible to cross every bridge exactly once in a single walk that starts and ends at the same point.

A graph* is defined as a set of nodes with connections between them called edges [111]. A coarse-grained picture of a PES therefore naturally fits into this definition, with nodes representing minima and edges representing the transition states that connect them. This approach was adopted in a number of previous energy landscapes studies, examples being the characterisation of dynamics in a region of a PES [8, 112–114] and detailed topological analyses of semi-complete PES samples [115].

---

*In modern literature the term 'network' is often used synonymously with the term 'graph'.

A directed edge is defined by its origin and destination nodes and can be travelled in one direction only. A graph composed of directed edges is termed a directed graph (digraph). Although every transition state facilitates both forward and reverse reactions, these are usually inequivalent, which leads to a symmetric digraph representation of a PES, i.e. to a digraph that is composed of pairs of complementary edges that share the same endpoints. Studies aimed at elucidating the global topology of a PES usually do not make such a distinction and deal with undirected graphs.

Interesting properties of the PES's of small Lennard-Jones clusters were recently discovered by Doye and Massen in a study of the corresponding undirected graph representations [115]. It appears that such graphs have features similar to both small-world and scale-free graphs. (A graph is said to have a small-world property if most pairs of nodes are connected by relatively short paths [111, 116, 117]. Such a graph can be easily obtained via a random rewiring of regular grid or lattice. The degree of a node is defined as the total number of its neighbours, and a graph is termed scale-free if its distribution of degrees follows a power law [111, 116].) While scale-free graphs are usually obtained via preferential attachment during growth [118], the origin of scale-free topology in graphs corresponding to PES's may lie in an Apollonian-like [119] packing of the basins of attraction [115]. Following the 'inherent structure' PES partitioning due to Stillinger and Weber [41, 42], a basin of attraction of a minimum can be defined as a set of points in configuration space connected to that minimum via steepest-descent paths. Topological studies of PES's are important because they can provide further insights into the PES connectivity and, consequently, increase our understanding of the relationship between the structural organisation of the PES and the observed physical and chemical properties.

Describing physical phenomena such as, for example, Brownian motion [120] and diffusion [121], requires more sophisticated graph models that allow for different types of edges. In edge-weighted graphs a label (weight or cost) is associated with every edge. Such graphs will be used in this thesis for various purposes. For example, an important part of the path-finding method described in Chapter 2 is the undirected edge-weighted graph representation, where every node is connected to every other node via an edge with a weight that is a function of Euclidean distance. In Appendix E we explain how

a weighted graph approach can be used to identify the fastest reaction pathways.

Both Brownian motion and diffusion were extensively studied in the past with the help of stochastic processes such as the random walk [121]. In mathematics and physics, a random walk is a formalisation of the intuitive idea of taking successive steps, each in a random direction [33]. If the number of possible directions is predefined and finite, a random walk is very easy to realize on an edge-weighted graph. A walk on a graph is defined as a sequence of edges such that the second node of each edge (except for the last edge) is the first node of the next edge. If the weights of the outgoing edges for every graph node add up to unity the graph is called probabilistic. A single step of a random walk confined to a probabilistic graph constitutes choosing the next graph node from the neighbours of the current node with a probability that equals the weight of the corresponding edge.

In Chapter 4 we will use random walks to model unimolecular chemical reactions [122]. An elementary reaction pathway is described as a transition from one state to a neighbouring state via a single transition state. Valid predictions of the sequence of such steps, known as the reaction mechanism, and a time scale associated with it are the holy grails of modern theoretical chemistry. Node-weighted probabilistic graphs are needed to address the time scale issue, because the time spent by a system before the transition occurs is likely to vary from state to state. An alternative description based on a probabilistic graph where each node is allowed to have a connection to itself is a more general approach to achieve this objective [123]. Both these models are employed in Chapter 4 in calculations of the average time for the reaction known as the mean first passage time.

The aforementioned graph model with self-connections is known by scientists in the fields of probability and stochastic processes as a discrete-time Markov chain — a stochastic process with a discrete state space [123]. The question of the first moment of the distribution of first passage times in stochastic processes is one of the most basic ones and is over two hundred years old. The first mean first passage time calculation was probably done by Jacob Bernoulli in the beginning of eighteenth century. In his ground-breaking work on probability titled 'The art of conjecture', which was posthumously published by his nephew Nicholas Bernoulli in 1713, he describes the techniques for

calculating the duration of various games of chance [124]. A number of methods for efficient calculation of the mean first passage times was developed since then, some of which are yet to be fully utilised in chemical applications.

In this thesis I will attempt to address two important stages in an energy-landscapes-based approach to the analysis of chemical reactivity, namely, finding reaction (or rearrangement) pathways and extracting the kinetic information from the obtained pathway ensemble. As the meaning of the term pathway (or path) varies from chapter to chapter I will spend some time in the introductory sections clarifying the terminology.

1.4 THESIS OVERVIEW

The results of the work that I have carried out are presented in three chapters.

The main focus of Chapter 2 is on double-ended methods for finding transition states. A detailed review of one of the leading methods from that class is followed by the discussion of our modifications and improvements that allowed us to extend its applicability. Results for a model two-dimensional surface and Lennard-Jones clusters of several sizes are presented. The chapter culminates with an application to finding folding paths for a family of small peptides known as tryptophan zippers.

Chapter 3 is devoted to discussion of two exciting properties of rearrangement pathways — cooperativity and localisation. A new measure of cooperativity suitable for applications to atomic rearrangements is introduced and subsequently used to establish the links between cooperativity of a single-step rearrangement, the energy barrier height and the difficulty of locating the corresponding transition state with both single-ended and double-ended methods.

In Chapter 4 we deal with compact representations of large pathway ensembles borrowing ideas from graph theory and the theory of random processes. The main theme is the development of faster methods for calculation of mean escape times for graphs of increasing complexity. We devise a number of approaches for extracting this kinetic information and compare them to well-established techniques such as kinetic Monte Carlo and discrete path sampling.

Chapter 5 summarises the achievements of the work described in this thesis and suggests the directions for future research.