

Research



CrossMark  
click for updates

**Cite this article:** Booth J, Vazquez S, Martinez-Nunez E, Marks A, Rodgers J, Glowacki DR, Shalashilin DV. 2014 Recent applications of boxed molecular dynamics: a simple multiscale technique for atomistic simulations. *Phil. Trans. R. Soc. A* **372**: 20130384.  
<http://dx.doi.org/10.1098/rsta.2013.0384>

One contribution of 13 to a Theme Issue 'Multi-scale systems in fluids and soft matter: approaches, numerics and applications'.

**Subject Areas:**

computational chemistry, atomic and molecular physics

**Keywords:**

molecular dynamics, quantum mechanics, dynamics of proteins, chemical dynamics, kinetics

**Authors for correspondence:**

David R. Glowacki  
e-mail: [drglowacki@gmail.com](mailto:drglowacki@gmail.com)  
Dmitrii V. Shalashilin  
e-mail: [d.shalashilin@leeds.ac.uk](mailto:d.shalashilin@leeds.ac.uk)

# Recent applications of boxed molecular dynamics: a simple multiscale technique for atomistic simulations

Jonathan Booth<sup>1</sup>, Saulo Vazquez<sup>2</sup>, Emilio Martinez-Nunez<sup>2</sup>, Alison Marks<sup>3</sup>, Jeff Rodgers<sup>4</sup>, David R. Glowacki<sup>4,5</sup> and Dmitrii V. Shalashilin<sup>1</sup>

<sup>1</sup>School of Chemistry, University of Leeds, Leeds LS2 9JT, UK

<sup>2</sup>Departamento de Química Física and Centro Singular de Investigación en Química Biológica y Materiales Moleculares, Campus Vida, Universidad de Santiago de Compostela, 15782 Santiago de Compostela, Spain

<sup>3</sup>School of Life Sciences, University of Bradford, Bradford BD7 1DP, UK

<sup>4</sup>School of Chemistry, University of Bristol, Bristol BS8 1TS, UK

<sup>5</sup>Department of Chemistry, Stanford University, Stanford, CA 94305, USA

In this paper, we briefly review the boxed molecular dynamics (BXD) method which allows analysis of thermodynamics and kinetics in complicated molecular systems. BXD is a multiscale technique, in which thermodynamics and long-time dynamics are recovered from a set of short-time simulations. In this paper, we review previous applications of BXD to peptide cyclization, solution phase organic reaction dynamics and desorption of ions from self-assembled monolayers (SAMs). We also report preliminary results of simulations of diamond etching mechanisms and protein unfolding in atomic force microscopy experiments. The latter demonstrate a correlation between the protein's structural motifs and its potential of mean force. Simulations of these processes by standard molecular dynamics (MD) is typically not possible, because the experimental time scales are very long. However, BXD yields well-converged and physically meaningful results. Compared with other methods of accelerated MD, our

© 2014 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License <http://creativecommons.org/licenses/by/3.0/>, which permits unrestricted use, provided the original author and source are credited.

BXD approach is very simple; it is easy to implement, and it provides an integrated approach for simultaneously obtaining both thermodynamics and kinetics. It also provides a strategy for obtaining statistically meaningful dynamical results in regions of configuration space that standard MD approaches would visit only very rarely.

## 1. Introduction

Boxed molecular dynamics (BXD) [1–3] is a simple and straightforward technique that extends the time scale of atomistic molecular dynamics (MD) simulations and facilitates simulation of rare events. In BXD's simplest implementation, we assume that a chemical reaction or some other atomistic physical process can be described by some reduced description of the system's  $3N$  dimensional configuration space—e.g. a reaction coordinate or an appropriate order parameter. It is then possible to split this coordinate into several boxes and subsequently lock the dynamics in each box by inverting the velocity of the trajectory in the direction of the reaction coordinate (or order parameter) every time the trajectory hits a boundary between two subsequent boxes. Then, one can calculate the rate constant of exchange between the boxes simply by calculating average time between the 'hits'.

$$\text{and } \left. \begin{aligned} k_{m,m+1} &= \frac{h_{m,m+1}}{t_m} \\ k_{m,m-1} &= \frac{h_{m,m-1}}{t_m} \end{aligned} \right\} \quad (1.1)$$

where  $h_{m,m-1}$  and  $h_{m,m+1}$  are the number of 'hits' of the left and right boundary of the  $m$ th box, and  $t_m$  is the time spent in this box. After accumulating sufficient statistics, the trajectory is allowed into the neighbouring box where the procedure is repeated. Eventually, a set of rate constants for exchange between the boxes is accumulated, and the dynamics is reduced to a set of kinetic equations.

$$\text{and } \left. \begin{aligned} \frac{dn_1(t)}{dt} &= -(k_{12} + k_{10})n_1(t) + k_{21}n_2(t), \\ \frac{dn_2(t)}{dt} &= k_{12}(n_1(t) + k_{32}n_3(t) - (k_{21} + k_{23})n_2(t), \\ &\dots \\ \frac{dn_m(t)}{dt} &= k_{m-1,m}n_{m-1}(t) - k_{m,m-1}n_m(t), \end{aligned} \right\} \quad (1.2)$$

or in the matrix form

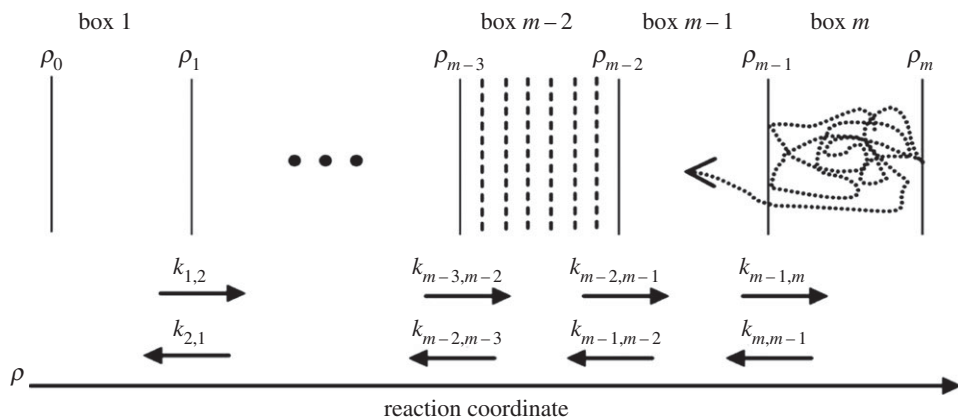
$$\frac{d\mathbf{n}(t)}{dt} = \mathbf{M}\mathbf{n}(t). \quad (1.3)$$

Then, the MD simulation of the system is replaced by the solution of the master equation (1.2)

$$\mathbf{n}(t) = U\Lambda U^{-1}\mathbf{n}(0), \quad (1.4)$$

where  $\mathbf{n}(0)$  contains the initial conditions for each box,  $U$  is the eigenvector matrix obtained from diagonalization of  $\mathbf{M}$ , and  $\Lambda$  is a diagonal matrix whose elements,  $\Lambda_{jj} = e^{\lambda_j t}$ , are determined by  $\lambda$ , the eigenvalue vector corresponding to  $\mathbf{M}$ .

BXD relies on the assumption that the motion within the box is stochastic and that sequential 'hits' and velocity inversions are uncorrelated—i.e. the time between the 'hits' must be bigger than the correlation time. In general, this is not the case. A trajectory reflected from a boundary can sometimes turn back rather quickly. One needs to use a simple technique described in [2] in order to remove these short-time-correlated events. The requirement of uncorrelated dynamics also imposes a restriction on the box size—i.e. it should be bigger than the so-called correlation length, which is the length at which a trajectory loses the memory of its initial conditions. For



**Figure 1.** Schematic of how BXD works with a reaction coordinate,  $\rho$ , split into  $m$  boxes.

large anharmonically coupled systems, the correlation time and correlation length are usually quite short. The procedure developed in [2] is based on analysis of the distribution of first passage times. It removes any contribution to the rate coefficients from velocity inversions separated by very short time and improves the quality of the kinetic results.

With the box-to-box rate coefficients in hand, it is possible to estimate the free energy profile along the reaction coordinate:

$$K_{n-1,n} = \frac{k_{n-1,n}}{k_{n,n-1}} = \exp\left(-\frac{\Delta G_{n-1,n}}{k_B T}\right). \quad (1.5)$$

A quantity which is closely related to the free energy is the potential of mean force (PMF). Using BXD, it has been shown [1] that the free energy (or PMF) profile can actually be estimated at a resolution much higher than the box size. The procedure involves splitting boxes into smaller regions and calculating the time spent within those smaller regions (enclosed by dashed lines in figure 1). This gives a high-resolution free energy profile within each box. These profiles can then be stitched together using an exact procedure that yields the global free energy profile along the entire reaction coordinate. We have also developed a post-analysis procedure for calculating kinetic rate coefficients between small boxes in the cases where that information is useful [3].

For us, BXD has its origin in the intramolecular dynamics diffusion theory (IDDT) [4–8], which has demonstrated that long-time reaction rates can be recovered from a set of short-time MD simulations [8]. Longer time-scale dynamics may be calculated from the set of rate constants obtained from short-time simulations within each box. This connection between short and long time scales is what makes BXD a multiscale method. In the IDDT, the dynamics was not restricted by the boundaries of the box. The idea of the box was introduced in [9], where the first BXD simulation of bond fission was carried out using only two boxes with a single boundary between them. Locking the trajectory within a box enables the system to visit the regions of high free energy and low probability, where an unrestricted trajectory would rarely visit.

Rare event acceleration is an important part of modern MD simulations. Over the past few years, a number of rare event acceleration strategies have been proposed, many of which share similarities with BXD. Milestoning [10–15] and particularly milestoning with Voronoi tessellations (MVT) [13,16] are the most closely related to the BXD technique, and it is worthwhile discussing connections between BXD and milestoning. Both use short-time dynamics to recover local kinetic information about the motion along the reaction coordinate in order to solve a kinetic state-to-state master equation. The original milestoning procedure required accurate sampling of the equilibrium distribution at the milestone surfaces located along the reaction coordinate. This is very similar to the earlier IDDT procedure [7,8], with the primary difference in the post-processing step: milestoning analysis relies on the milestone-to-milestone passage time,

and a IDDT analysis relies on a diffusional coefficient. In MVT [13,16], a modified milestone procedure has been proposed wherein the dynamics initiated between a set of milestones are then locked using a collision rule at the boundary so that they cannot escape, as in [9]. Perhaps the most important difference between the velocity inversion scheme used by BXD and the collision rule of MVT is that the BXD scheme conserves energy, linear momentum and angular momentum. So long as the distance between a BXD inversion boundary and transition state is larger than the characteristic system decorrelation length, then the accelerated dynamics obtained using BXD give meaningful statistics that allow detailed analysis of non-equilibrium dynamical properties in the post-transition state region of a given molecular system's phase space (see [2]). Indeed, these features of BXD have proven critical to many of the studies which are outlined in this paper, where BXD was used to investigate non-equilibrium energy relaxation in number of particles, volume, energy (NVE) microcanonical ensemble trajectories [17–19].

BXD and MVT also differ in the post-processing step—i.e. how they interpret the MD simulation results obtained from within the boxes. BXD uses the trajectory information to calculate box-to-box transition state theory rate constants and MVT uses the information to calculate transition times for moving from border to border. As has been shown in [2] calculating either of the two is not entirely straightforward and requires thinking about dynamical decorrelation. In BXD, the decay trace needs to be corrected by removing fast events where trajectories return to the boundary. This gives decay traces which are better approximated by a single exponent and box-to-box dynamics better described in terms of a single rate constant. MVT, on the other hand, gives flat non-exponential decays at short times. When these are removed, then the BXD and MVT decay traces are identical. The foundation for MVT is based on ideas from transition path sampling, and it is exact in the limit that the correct set of milestones is used [14]. BXD has its foundations in standard transition state theory. It is exact in the limit that the system is ergodic, and a correct set of dividing surfaces has been used. Based on BXD's origins in transition state theory, we have shown how to simply and exactly renormalize the BXD statistics to obtain thermodynamics [1] and kinetics [3] to arbitrary histogram resolution, distinct from milestone and MVT.

BXD also shares similarities with a number of other rare event acceleration methods. For example, BXD in its simplest form [9] is related to hyperdynamics [20,21], which similarly introduces constraints into a molecular system's configuration space to encourage it to visit regions of low probability. BXD also has similarities to umbrella sampling [22–25] with the primary difference that the former uses rectangular boxes instead of parabolic force restraints. Compared with these methods, BXD's main advantage is that it is the only technique which is capable of simultaneously extracting high-resolution thermodynamic and kinetic information. BXD's formulation is very straightforward, based upon a simple rewriting of the classical transition state theory formula, which is familiar in chemistry. In common with all of the aforementioned techniques, BXD does not address the question of finding a good reaction coordinate or order parameter; its usage relies on being able to define these in a sensible way. In many instances, definition of an appropriate reaction coordinate or order parameter is a substantial challenge, but there are a range of systems where these quantities are reasonably well defined. If a good reaction coordinate exists, then BXD provides long time-scale dynamics/kinetics and PMF profiles in regions of very high energies.

In this paper, we review several recent and ongoing applications of the BXD technique. First, we review the simulation of peptide loop formation dynamics where interesting and unusual power law kinetics observed in experiments [26–28] have been reproduced on time scales spanning several orders of magnitude—i.e. from picoseconds to microseconds. Second, we report new results from our simulations of ion diffusion through a monolayer on a metal surface, where residence times of hours were recovered in agreement with experiment [29–31]. Third, we give a brief summary of how BXD has been used to provide insights into the elementary chemical mechanisms of diamond etching. Fourth, we describe how BXD has been used to accelerate rare event sampling in studies of solution phase bimolecular reactions. Finally, we report preliminary BXD results of the free energies obtained from pulling apart several

small proteins, all of which were previously investigated experimentally using atomic force microscopy (AFM) [32–34]. For these systems, we observe that the slope of the preliminary PMFs calculated as a function of peptide extension (distance between the peptide's C and N termini) appear to be in reasonable agreement with the experimentally observed AFM forces required to unfold the proteins. For proteins composed of  $\alpha$ -helices, the force required along the extension coordinate is substantially lower than for those composed of  $\beta$ -sheets, in qualitative agreement with experimental observations. We also discuss future developments and applications.

## 2. Applications of boxed molecular dynamics

### (a) Peptide loop formation

Volk and Hochstrasser previously carried out a series of detailed studies investigating the kinetics of polypeptide cyclization. Using a laser pulse, their experiments photolytically cut a solvated polypeptide loop (in which the two termini of the polypeptide backbone were connected via a disulfide bond), generating a thiyl radical pair localized on each end of the peptide [26–28]. The radical concentration was monitored by time-resolved spectroscopy spanning picoseconds to milliseconds [27,28]. The measured absorbance of the peptide occurs predominantly from the thiyl radicals (a configuration denoted below as  $n_{\text{open}}$ ), decreasing when the radical pairs recombine to form the bond between the two ends of the polypeptide chain, i.e. when the peptide loop is reformed. In these experiments, the instantaneous rate constant [26–28]

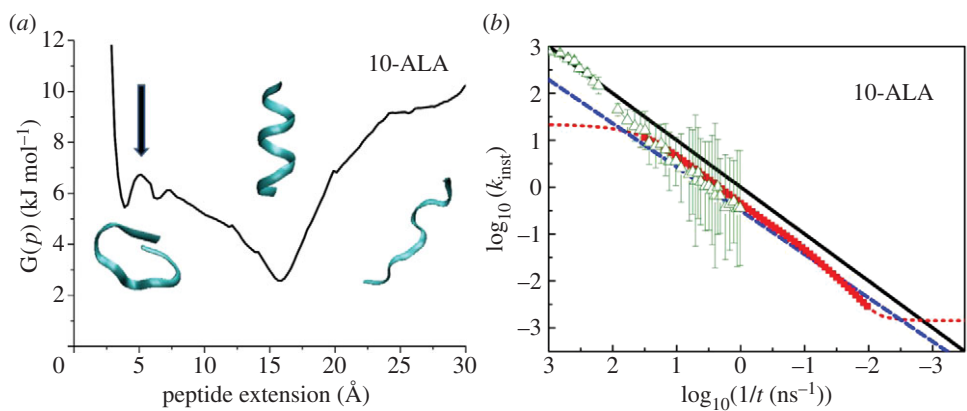
$$k_{\text{inst}} = -\frac{1}{n_{\text{open}}(t)} \frac{dn_{\text{open}}(t)}{dt} \quad (2.1)$$

was determined to have the form

$$k \approx At^{-\alpha}, \quad (2.2)$$

where  $k$  has units of  $\text{ns}^{-1}$ ,  $A = 3.2$  and  $\alpha = -0.93 \pm 0.02$ , with similar values observed for loop formation dynamics in a range of peptides. Remarkably, this time-dependence, which is very close to simple  $k = 1/t$  dependence, holds for nine orders of magnitude in time, from picoseconds to milliseconds, and was essentially the same for a short polypeptide [27] and a 100-residue protein [28]. To rationalize these experimental results, we performed the first fully atomistic simulations of the system, using BXD to investigate several model peptides with initial conditions corresponding to those of experiment, i.e. at  $t = 0$  the two ends of the peptide are in close contact. In our simulations, the reaction coordinate was taken to be the peptide extension. This coordinate was split into anywhere between 15 and 20 boxes, with sizes ranging from 1 to 2 Å. Box-to-box rate constants were obtained using a decorrelation algorithm which we have recently described [2]. The procedure developed in [3] was used to extract 'high-resolution' rate coefficients from the low resolution box-to-box rate coefficients. Having obtained the high-resolution rate coefficients, a solution (1.4) of the kinetic master equation (1.3) was obtained. Figure 2 shows BXD results obtained for one of the peptides investigated in this study, namely 10-ALA.

Frame (a) depicts the PMF as a function of the reaction coordinate. Frame (b) presents the instantaneous rate constant of cyclization of the 10-ALA peptide. Triangles show the rate constant obtained by unbiased MD simulation. Using a serial implementation of the CHARMM program suite [35] on a local computing cluster, these unbiased simulations only permit reaction events on a picosecond time scale. The red line demonstrates the outcome of the BXD simulation. The plateaus shown by the red dotted line indicate the limitations of BXD—i.e. the BXD results are only reliable for times  $t$ , where  $t_{\text{max}} > t > t_{\text{min}}$ , where  $t_{\text{max}}$  and  $t_{\text{min}}$  are the longest and shortest time scales obtained from the eigenvalue spectrum, respectively. As can be seen from figure 2, BXD reproduces a power law very similar to those obtained from fitting of the experimental data, which are shown by blue and black lines. Using BXD, atomistic simulations can be used to derive kinetic information on the microsecond time scale, far beyond time scales which are accessible by unbiased MD.



**Figure 2.** BXD results obtained from 10-ALA. (a) The PMF as a function of extension coordinate. The arrow indicates the transition state for going from the extended conformation to the closed loop form. (b) The results of BXD simulation (thick red line) together with the outcome of unbiased short-time MD simulations (triangles). The two plateaus shown by the red dashed line correspond to the smallest and largest eigenvalues of the matrix  $\mathbf{M}$  in equation (1.3) and therefore represent the limits between which the BXD master equation results are valid. The result is also compared with a  $1/t^\alpha$  power law, with  $A = 1$ ,  $\alpha = 1$  (black line) and  $A = 3.2$ ,  $\alpha = 0.93$  (blue dashed line). (Online version in colour.)

Similar results were obtained for a number of model peptides and it was shown that for all of them the instantaneous rate closely follows the power law (2.2). The behaviour appears to be largely independent of the structure of peptides and it can be explained on the basis of the solution (1.4) of the master equation (1.3), giving the following form for the time-dependence of the extended peptide

$$n_{\text{open}}(t) = 1 - n_0(t) = \sum_i F(\lambda_i) e^{\lambda_i t} = \sum_i F(k_i) e^{-k_i t}, \quad (2.3)$$

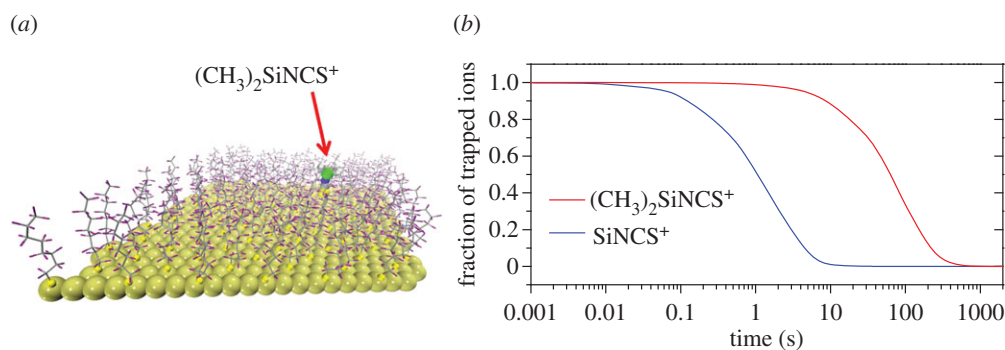
where  $n_0(t)$  is the population of the box 0, which corresponds to loop closure, and  $F(\lambda_i)$  is a pre-exponential factor that weights the contribution of a particular eigenvalue to the overall time-dependent population of extended peptide,  $n_{\text{open}}(t)$ . The rate constants  $k_i = -\lambda_i$  of the exponents in (1.4) are related to the eigenvalues of the kinetic matrix  $\mathbf{M}$  in (1.3). It was noted in [3] that the coefficients  $F(\lambda_i)$  depend strongly on the initial populations of the boxes  $n_i(0)$ . Our analysis focused on the eigenvalue spectrum weighted by the corresponding coefficients  $F(\lambda_i)$ . The shape of this weighted spectrum determines whether the phenomenological kinetics follow a power law or a standard form of a single exponent. In the limit that a single value in the weighted spectrum dominates, typical single exponential kinetics are observed. In the limit that there are no dominant values in the weighted spectrum, power law kinetics arise. In the immediate aftermath of experimental bond-breaking, all the population is concentrated in the ‘boxes’ adjacent to the loop formation transition state. This results in a scenario where the weighted spectrum has no dominant values, and all the rate coefficients  $k_i$  have characteristic times  $t_i = 1/k_i$ , that are equally probable, giving rise to the time-dependence (2.2) for the instantaneous rate (2.1).

These results can be understood with the help of a simple non-scientific analogy. Suppose that a large group of hikers randomly choose their routes through a landscape similar to that shown in figure 3, which provides a schematic representation of the energy landscape of a peptide. The randomness of each hikers’ path means that there will be a distribution of available paths, each of which has a unique characteristic time for returning home. At time  $t_2$ , for example, those hikers who took a route with characteristic time  $t_1$  have already returned, whereas those on a longer route  $t_3$  are still far away. Thus, at a given time  $t$ , only the routes with the rate constant close to  $k = 1/t$  contribute, which explains the origin of the  $1/t$  law in (2.3). An important feature of the scheme depicted in figure 3, which distinguishes the ‘returning hiker’ analogy from numerous





**Figure 3.** Routes selected randomly on a landscape. At time  $t_2$ , the hikers who took route  $t_3$  are still walking but the hikers on the short route  $t_1$  have already returned. Therefore, at the moment  $t_2$ , the instantaneous rate constant of return is equal to  $1/t_2$ . If the start and endpoints of all routes are in close proximity, then there is a large number of short routes, which is not the case when start and endpoint are distant from each other. (Online version in colour.)



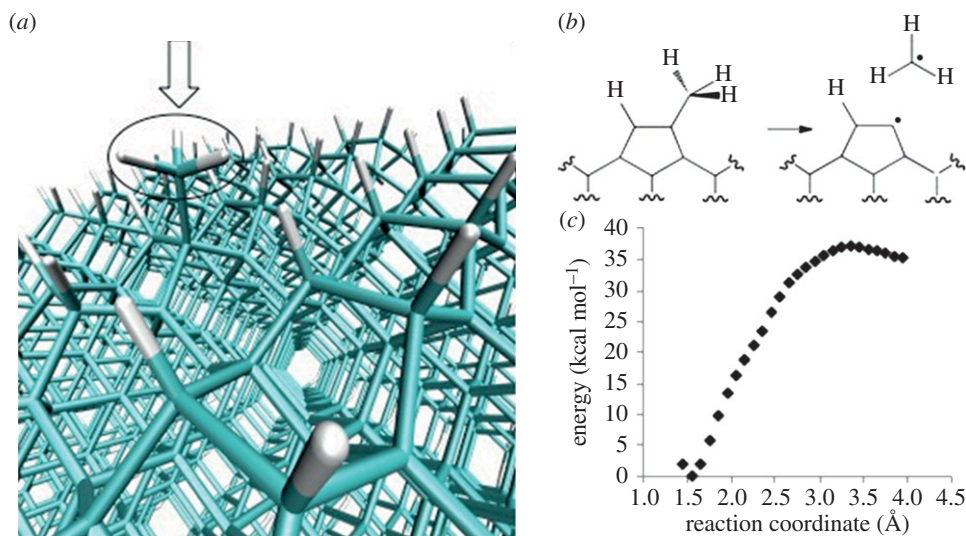
**Figure 4.** (a) Monolayer of perfluorinated alkanethiol on gold with adsorbed ion; (b) survival probability of  $\text{SiNCS}^+$  (red) and  $(\text{CH}_3)_2\text{SiNCS}^+$  (blue) in the monolayer calculated as a fraction of the trapped molecules. (Online version in colour.)

existing random walk models, is that the all routes are initialized and terminated in roughly the same place. This means that shorter routes contribute similarly to longer routes.

The example of peptide cyclization not only demonstrates that BXD can produce reliable thermodynamic and kinetic information on very long time scales (allowing one to reproduce experiments), but that it also helps to provide simple insights into the qualitative mechanisms that give rise to the experimental observations.

## (b) Desorption of ions from the monolayer

As has been shown in the previous example, BXD is capable of reproducing the dynamics of peptides on time scales of microseconds where the barriers along the PMF are relatively small. An important additional question is whether BXD can accurately treat processes with much larger energy barriers along the PMF, where the dynamical time scales are substantially slower. In a recent study [36], the dynamics of the desorption of silyl ions  $(\text{CH}_3)_2\text{SiNCS}^+$  and  $\text{SiNCS}^+$ , trapped in the self-assembled monolayer of perfluorinated alkanethiol on gold (F-SAM) has been investigated. The monolayer with  $(\text{CH}_3)_2\text{SiNCS}^+$  adsorbed on it is shown in figure 4a. In the experiment [29–31], a monolayer was bombarded with each of these ions (separately) during about 1 h, forming ions adsorbed just beneath the surface of the monolayer (figure 4a). Subsequently, the F-SAMs were analysed by high-energy collisions of  $\text{Xe}^+$  ions, which induce the ejection of surface species (sputtering). The sputtered ions were detected by mass spectrometry.



**Figure 5.** (a) The atomistic MD set-up upon which we carried out BXD sampling to determine PMFs for dissociation of the surface  $-\text{CH}_3$  group, which is highlighted by the arrow; (b) a schema for a simple C–C bond dissociation etching mechanism, creating a nascent methyl group; (c) the 1000 K PMF for methyl dissociation from a diamond surface, calculated using BXD. (Online version in colour.)

It was shown that the larger ion  $(\text{CH}_3)_2\text{SiNCS}^+$  was observed in the mass spectrum after several days, in stark contrast to the complete lack of the smaller  $\text{SiNCS}^+$  ion. The interpretation given in the experimental studies is that the methyl groups increase the attractive and steric interactions between the ion and the F-SAM, which facilitate the entrapment of the ion inside the monolayer and decrease its desorption rate.

To address whether atomistic MD simulations were able to provide any insights into these experimental observations, we carried out BXD simulation of the ions diffusing inside of the monolayer towards the surface with subsequent desorption. For a reaction coordinate, we used the distance from the ion centre of mass to the surface. The residence time was estimated using the kinetic master equation. The BXD module in the CHARMM code was used, and BXD box-to-box rate constants were obtained. The rate constants were corrected by the previously discussed decorrelation procedure, which excludes the correlated velocity inversions separated by very short time. The rate constants included an additional correction, designed to account for the presence of image charge at the gold surface. This procedure allows one to compute the survival probability for ions diffusing from the monolayer to the vacuum, which is shown at figure 4a, yielding the residence time of 91 s for  $(\text{CH}_3)_2\text{SiNCS}^+$  and 1.5 s for the more compact  $\text{SiNCS}^+$  ion, which explains why  $\text{SiNCS}^+$  is not seen in the mass spectrum at long times.

This system provides a nice demonstration of the fact that BXD is capable of reaching time scales which are many orders of magnitude longer than the time scales typically recoverable by atomistic simulations. This is because of BXD's multiscale approach: atomistic simulations yield box-to-box rate constants, which are used as subsequent input for kinetic analysis, covering much longer times.

### (c) Heterogeneous chemistry at the surface of diamond

BXD is also being applied to better understand the elementary chemical mechanisms that guide diamond etching—a process wherein solid-state diamond is converted to its gas-phase atomic constituents [37]. In particular, BXD is being used to understand the kinetics and thermodynamics for dissociation of methyl radicals from diamond surfaces, according to the scheme in figure 5a,b. Spontaneous thermal dissociation of C–C bonds at the surface of diamond has been suggested



as a possible etching mechanism. Preliminary PMFs along the C–CH<sub>3</sub> stretching coordinate using BXD in conjunction with a molecular dynamics simulation set-up shown in figure 5c (at 1000 K) suggest free energy barriers to methyl dissociation which are on the order of 38–46 kcal mol<sup>-1</sup>, depending on temperature. The magnitude of these free energy barriers is consistent with etching of diamond surfaces occurring as a result of spontaneous loss of methyl radicals.

#### (d) Solution phase dynamics studies

Another application where BXD has proved invaluable is in the area of solution phase reaction dynamics [38]. In particular, this method has permitted atomic level insights into solution phase bimolecular reaction dynamics for the following hydrogen abstraction systems: (i) CN + cyclohexane [CN + C<sub>6</sub>H<sub>12</sub> → HCN + C<sub>6</sub>H<sub>11</sub>] in dichloromethane solvent [17,18,39], and (ii) CN + tetrahydrofuran [CN + C<sub>4</sub>H<sub>8</sub>O → HCN + C<sub>4</sub>H<sub>7</sub>O] in tetrahydrofuran solvent [40].

For these studies, the utility of BXD arises from the fact that it provides an integrated methodology for (i) efficiently mapping the solution phase PMF, and (ii) accelerating reactive events that would otherwise not be possible to simulate. To obtain PMFs for these solution phase systems, ‘boxes’ were defined in terms of the distance between the approaching CN and an H atom on the co-reactant. Figure 7c shows results for CN + C<sub>6</sub>H<sub>12</sub>.

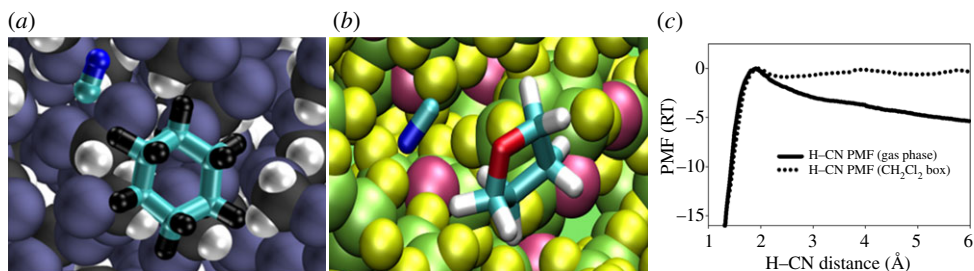
PMF sampling in BXD is particularly elegant in these sorts of systems; however, the principal advantage of BXD for the aforementioned solution phase systems is its ability to provide rare event acceleration that furnishes meaningful dynamics insights. These studies were carried out by placing a BXD box boundary to the reactant side of the transition state, and ‘locking’ the trajectory so as not to go beyond this boundary towards reactants. This approach [9] guarantees that the system does not stray too far towards isolated reactants, and simultaneously accelerates transition state passage. So long as the BXD box boundary is far enough away from the transition state, then the post-transition state dynamics are decorrelated from dynamical inversion events and provide statistically meaningful dynamics results [2,9]. Conveniently, this provides us dynamical insights into how the system behaves when unperturbed by BXD box.

The high-quality rare event statistics available with the BXD technique have permitted several new insights into solution phase dynamics. One of the most interesting is the direct observation of two distinct post-reaction non-equilibrium relaxation regimes [19,39]. For example, in the case of the CN + C<sub>6</sub>H<sub>12</sub> reaction in CH<sub>2</sub>Cl<sub>2</sub> solvent, the first relaxation regime follows in the immediate wake of the reaction, when the nascent HCN is in close proximity to its cyclohexyl coproduct, and strong interaction between co-products facilitates rapid energy transfer and correspondingly fast HCN relaxation. The second relaxation regime occurs once HCN and cyclohexyl have diffused away from one another and into the bulk solvent. HCN relaxation is considerably slower following diffusion, owing to weaker interactions between HCN and the CH<sub>2</sub>Cl<sub>2</sub> solvent.

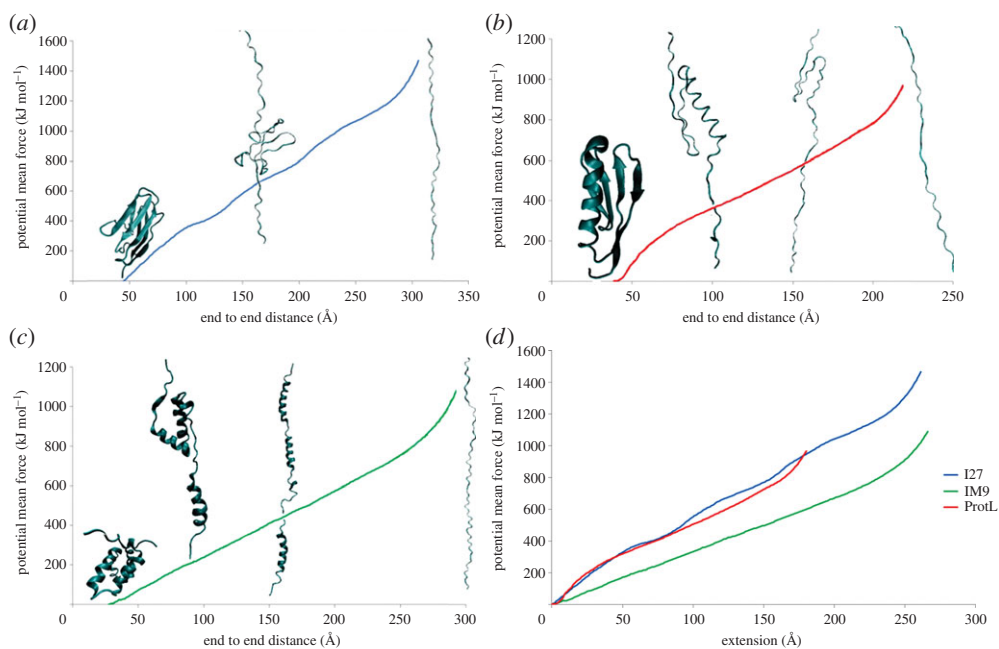
For the systems shown in figure 6, the potential energy surfaces for reaction were obtained using a multiconfigurational molecular mechanics approach, with Hamiltonian matrix elements fitted to configuration space points along the reaction path, calculated using high level CCSD(T) electronic structure theory results extrapolated to the infinite basis set limit. This approach provides a means for smoothly and accurately connecting molecular mechanics forcefields that describe both the reactant and product configuration. QM/MM approaches would have similarly allowed us to treat the reactive process, but at a considerably increased computational cost.

#### (e) Free energy of unfolding by atomic force microscopy

In a number of recent experiments [32–34], the unfolding of single protein molecules with mechanical force has been investigated using the atomic force microscope. In one class of these experiments, the proteins are pulled at a constant speed (anywhere between 40 and 4000 nm s<sup>-1</sup>); in another class, the proteins are pulled at a constant force. The net result of such experiments is a large quantity of experimental data with detailed information on force as a function of peptide extension.



**Figure 6.** (a) A schematic of CN + cyclohexane in dichloromethane solvent; (b) CN + THF in THF solvent; (c) the PMF calculated for CN + cyclohexane. (Online version in colour.)



**Figure 7.** Potential of mean force profile as a function of end-to-end distance for the proteins (a) I27, (b) IM9 and (c) L. (d) A comparison of the respective PMFs. The slope of free energy is significantly smaller for protein IM9, which has only  $\alpha$ -helical secondary structure. Snapshots of structures along the PMF profile are also shown. (Online version in colour.)

Because the time scale of the pulling experiment is quite long, atomistic MD simulations tend to use pulling speeds and forces which exceed those of experiment by many orders of magnitude. BXD may be easily applied to simulations of AFM due to the fact that the unfolding process is reasonably well described using a simple end-to-end distance order parameter (similar to the peptide cyclization example considered above). This allows us to investigate events that would otherwise be far too expensive to simulate. In what follows, we present the first preliminary results of BXD simulations applied to understand AFM unfolding experiments. We calculated free energies along the reaction coordinate for three proteins, all of which are shown in figure 7: IM9 is an all- $\alpha$  protein; I27 is an all- $\beta$  protein and protein L is an  $\alpha$ - $\beta$  protein. Experimentally, the AFM results showed a correlation between the mechanical strength of the proteins and their secondary structure, with  $\beta$ -proteins requiring a much higher unfolding force than  $\alpha$ -helical proteins.

Figure 7 shows PMF as function of end-to-end distance with results obtained using the previously discussed decorrelation algorithm [2]. The extension coordinate of proteins IM9, I27 and L were split into uniform boxes with sizes of 3, 2 and 1 Å, respectively. The average derivatives  $dG/dx$  (calculated by finite difference over an interval of 20 Å) in figure 7 are as follows: approximately 5 pN for protein IM9, approximately 170 pN for ProteinL, and approximately 170 pN for I27. These values are in broad agreement with trends reported for the experimental unfolding forces: approximately 20 pN for IM9, 185–200 pN for I27 and 100–200 pN for ProteinL (depending on the pulling speed) [32–34].

These results are preliminary, but nevertheless show qualitative agreement with the experimental trend. In future work, we will perform an extensive range of checks in an attempt to unravel the unfolding pathways, and determine how the dynamical pathways provided by BXD map into simulations compare with simulations carried out with an applied force. We will also perform kinetic studies of the extension dynamics and obtain time-dependent forces and distances under various experimental conditions—e.g. varying the pulling speed and the direction of pulling, which should permit a more direct comparison with experiment. For example, if a protein is pulled by the force  $F$ , then it is possible to calculate a force-modified PMF by adding the term  $-Fx$  to the PMF, and correcting the corresponding the box-to-box rate constants by the factor  $\exp(\pm F\Delta x)/2kT$ , where  $\Delta x$  is the box size. Initial kinetic simulations indicate that end-to-end distance starts increasing, and a protein begins unfolding when the force becomes comparable to the slope of the free energy plot.

### 3. Conclusion and future developments

In this paper, we presented a brief summary of the BXD method which allows the simultaneous calculation of thermodynamics (free energies along reaction coordinate) and kinetics of complicated molecular systems. We reviewed previous applications of BXD across a range of different chemical systems, such as peptide cyclization and solution phase bimolecular reaction dynamics. We also report first results obtained from simulating desorption of ions from the surface of a self-assembled monolayer, the kinetics of diamond etching, and preliminary results of simulations of protein unfolding in AFM. Atomistic simulations of these processes by standard MD are not possible, but BXD yields accurate and physically meaningful results. BXD is a simple method which is easy to implement, and which can accurately calculate free energies along a reaction coordinate or some appropriately partitioned region of molecular phase space. BXD allows one to recast the complicated atomistic dynamic evolution a given system as a set of kinetic processes occurring within different regions of configuration space. Although BXD is a fully atomistic technique, it is capable of describing dynamics many orders of magnitude longer than typical atomistic dynamics time scales. The systems outlined in this paper demonstrate how BXD has furnished considerable microscopic insights into the atomistic processes and mechanisms at work in a range of condensed phase and heterogeneous systems.

A number of BXD projects are currently underway in our groups, including the dynamics of diamond etching, additional solution phase reaction systems, the dynamics of AFM unfolding, the dynamics of pocket opening in biomedically important proteins (where access to the docking site is blocked by structural elements of the protein) and the folding dynamics of a range of small proteins. In a number of these systems, we obtain free energy/PMF profiles as well as the long-time kinetics/dynamics, furnishing considerable microscopic insights into a range of problems. The applications reported and reviewed in this paper have used relatively simple reaction coordinates, but we are presently exploring more sophisticated reaction coordinates as well as multidimensional extensions of BXD.

**Acknowledgements.** D.R.G. is grateful for funding from a Royal Society Research Fellowship. J.B. and D.V.S. acknowledge the support of EPSRC (grant no. EP/E009824/1). E.M.-N. and S.A.V. are grateful to the ‘Centro de Supercomputación de Galicia (CESGA)’ for the use of its computational resources, as well as to ‘Ministerio de Economía y Competitividad’ (grant no. CTQ2009-12588) for financial support. D.S. and E.M.-N.

## References

1. Glowacki DR, Paci E, Shalashilin DV. 2009 Boxed molecular dynamics: a simple and general technique for accelerating rare event kinetics and mapping free energy in large molecular systems. *J. Phys. Chem. B* **113**, 16 603–16 611. (doi:10.1021/jp9074898)
2. Glowacki DR, Paci E, Shalashilin DV. 2011 Boxed molecular dynamics: decorrelation time scales and the kinetic master equation. *J. Chem. Theory Comput.* **7**, 1244–1252. (doi:10.1021/ct200011e)
3. Shalashilin DV, Beddard GS, Paci E, Glowacki DR. 2012 Peptide kinetics from picoseconds to microseconds using boxed molecular dynamics: power law rate coefficients in cyclisation reactions. *J. Chem. Phys.* **137**, 165102. (doi:10.1063/1.4759088)
4. Guo Y, Shalashilin DV, Krouse JA, Thompson DL. 1999 Intramolecular dynamics diffusion theory approach to complex unimolecular reactions. *J. Chem. Phys.* **110**, 5521–5525. (doi:10.1063/1.478449)
5. Guo Y, Shalashilin DV, Krouse JA, Thompson DL. 1999 Predicting nonstatistical unimolecular reaction rates using Kramers' theory. *J. Chem. Phys.* **110**, 5514–5520. (doi:10.1063/1.478448)
6. Shalashilin DV, Thompson DL. 1997 Monte Carlo variational transition-state theory study of the unimolecular dissociation of RDX. *J. Phys. Chem. A* **101**, 961–966. (doi:10.1021/jp961755o)
7. Shalashilin DV, Thompson DL. 1997 Method for predicting IVR-limited unimolecular reaction rate coefficients. *J. Chem. Phys.* **107**, 6204–6212. (doi:10.1063/1.474285)
8. Shalashilin DV, Thompson DL. 1997 Intramolecular dynamics diffusion theory: nonstatistical unimolecular reaction rates. *ACS Symp. Ser.* **678**, 81–98. (doi:10.1021/bk-1997-0678.ch006)
9. Martinez-Nunez E, Shalashilin DV. 2006 Acceleration of classical mechanics by phase space constraints. *J. Chem. Theory Comput.* **2**, 912–919. (doi:10.1021/ct060042z)
10. Elber R. 2007 A milestone study of the kinetics of an allosteric transition: atomically detailed simulations of deoxy Scapharca hemoglobin. *Biophys. J.* **92**, L85–L87. (doi:10.1529/biophysj.106.101899)
11. Faradjian AK, Elber R. 2004 Computing time scales from reaction coordinates by milestoneing. *J. Chem. Phys.* **120**, 10 880–10 889. (doi:10.1063/1.1738640)
12. Kuczera K, Jas GS, Elber R. 2009 Kinetics of helix unfolding: molecular dynamics simulations with milestoneing. *J. Phys. Chem. A* **113**, 7461–7473. (doi:10.1021/jp900407w)
13. Maragliano L, Vanden-Eijnden E, Roux B. 2009 Free energy and kinetics of conformational transitions from voronoi tessellated milestoneing with restraining potentials. *J. Chem. Theory Comput.* **5**, 2589–2594. (doi:10.1021/ct900279z)
14. Vanden-Eijnden E, Venturoli M, Ciccotti G, Elber R. 2008 On the assumptions underlying milestoneing. *J. Chem. Phys.* **129**, 174102. (doi:10.1063/1.2996509)
15. West AMA, Elber R, Shalloway D. 2007 Extending molecular dynamics time scales with milestoneing: example of complex kinetics in a solvated peptide. *J. Chem. Phys.* **126**, 145104. (doi:10.1063/1.2716389)
16. Vanden-Eijnden E, Venturoli M. 2009 Markovian milestoneing with Voronoi tessellations. *J. Chem. Phys.* **130**, 194101. (doi:10.1063/1.3129843)
17. Glowacki DR, Rose RA, Greaves SJ, Orr-Ewing AJ, Harvey JN. 2011 Ultrafast energy flow in the wake of solution-phase bimolecular reactions. *Nat. Chem.* **3**, 850–855. (doi:10.1038/nchem.1154)
18. Greaves SJ *et al.* 2011 Vibrationally quantum-state-specific reaction dynamics of H atom abstraction by CN radical in solution. *Science* **331**, 1423–1426. (doi:10.1126/science.1197796)
19. Glowacki DR, Orr-Ewing AJ, Harvey JN. 2011 Product energy deposition of CN+ alkane H abstraction reactions in gas and solution phases. *J. Chem. Phys.* **134**, 214508. (doi:10.1063/1.3595259)
20. Voter AF. 1997 Hyperdynamics: accelerated molecular dynamics of infrequent events. *Phys. Rev. Lett.* **78**, 3908–3911. (doi:10.1103/PhysRevLett.78.3908)
21. Voter AF. 1997 A method for accelerating the molecular dynamics simulation of infrequent events. *J. Chem. Phys.* **106**, 4665–4677. (doi:10.1063/1.473503)
22. Torrie GM, Valleau JP. 1977 Monte Carlo study of a phase-separating liquid mixture by umbrella sampling. *J. Chem. Phys.* **66**, 1402–1408. (doi:10.1063/1.434125)

23. Torrie GM, Valleau JP. 1977 Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. *J. Comput. Phys.* **23**, 187–199. (doi:10.1016/0021-9991(77)90121-8)
24. Kaestner J. 2011 Umbrella sampling. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1**, 932–942. (doi:10.1002/wcms.66)
25. Frenkel D, Smit B. 2002 *Understanding molecular simulations*, 2nd edn. London, UK: Academic Press.
26. Metzler R, Klafter J, Jortner J, Volk M. 1998 Multiple time scales for dispersive kinetics in early events of peptide folding. *Chem. Phys. Lett.* **293**, 477–484. (doi:10.1016/S0009-2614(98)00758-1)
27. Volk M, Kholodenko Y, Lu HSM, Gooding EA, DeGrado WF, Hochstrasser RM. 1997 Peptide conformational dynamics and vibrational stark effects following photoinitiated disulfide cleavage. *J. Phys. Chem. B* **101**, 8607–8616. (doi:10.1021/jp971951a)
28. Milanese L, Waltho JP, Hunter CA, Shaw DJ, Beddard GS, Reid GD, Dev S, Volk M. 2012 Measurement of energy landscape roughness of folded and unfolded proteins. *Proc. Natl Acad. Sci. USA* **109**, 19 563–19 568. (doi:10.1073/pnas.1211764109)
29. Luo H, Miller SA, Cooks RG, Pachuta SJ. 1998 Soft landing of polyatomic ions for selective modification of fluorinated self-assembled monolayer surfaces. *Int. J. Mass Spectrom.* **174**, 193–217. (doi:10.1016/S0168-1176(97)00302-9)
30. Miller SA, Luo H, Jiang X, Rohrs HW, Cooks RG. 1997 Ion/surface reactions, surface-induced dissociation and surface modification resulting from hyperthermal collisions of  $\text{OCNCO}^+$ ,  $\text{OCNCS}^+$ ,  $(\text{CH}_3)_2\text{SiNCO}^+$ , and  $(\text{CH}_3)_2\text{SiNCS}^+$  with a fluorinated self-assembled monolayer surface. *Int. J. Mass Spectrom. Process.* **160**, 83–105. (doi:10.1016/S0168-1176(96)04502-8)
31. Miller SA, Luo H, Pachuta SJ, Cooks RG. 1997 Soft-landing of polyatomic ions at fluorinated self-assembled monolayer surfaces. *Science* **275**, 1447–1450. (doi:10.1126/science.275.5305.1447)
32. Brockwell DJ. 2007 Probing the mechanical stability of proteins using the atomic force microscope. *Biochem. Soc. Trans.* **35**, 1564–1568. (doi:10.1042/BST0351564)
33. Hann E, Kirkpatrick N, Kleanthous C, Smith DA, Radford SE, Brockwell DJ. 2007 The effect of protein complexation on the mechanical stability of Im9. *Biophys. J.* **92**, L79–L81. (doi:10.1529/biophysj.106.102475)
34. Sadler DP, Petrik E, Taniguchi Y, Pullen JR, Kawakami M, Radford SE, Brockwell DJ. 2009 Identification of a mechanical rheostat in the hydrophobic core of protein L. *J. Mol. Biol.* **393**, 237–248. (doi:10.1016/j.jmb.2009.08.015)
35. Brooks BR *et al.* 2009 Identification of a mechanical rheostat in the hydrophobic core of protein L. *J. Comput. Chem.* **30**, 1545–1614. (doi:10.1002/jcc.21287)
36. Nogueira JJ *et al.* 2014 Unraveling the factors that control soft landing of small silyl ions on fluorinated self-assembled monolayers. *J. Phys. Chem. C.* **118**, 10 159–10 169. (doi:10.1021/jp501841a)
37. Glowacki DR, Rodgers WJ, Harvey JN. In preparation. Chemistry of local hotspots on carbon nano-structures: a framework for treating energy dissipation and chemical reaction on surfaces.
38. Orr-Ewing AJ, Glowacki DR, Greaves SJ, Rose RA. 2011 Chemical reaction dynamics in liquid solutions. *J. Phys. Chem. Lett.* **2**, 1139–1144. (doi:10.1021/jz2002716)
39. Glowacki DR, Lightfoot R, Harvey JN. 2013 Non-equilibrium phenomena and molecular reaction dynamics: mode space, energy space and conformer space. *Mol. Phys.* **111**, 631–640. (doi:10.1080/00268976.2013.780100)
40. Rose RA *et al.* 2012 Reaction dynamics of CN radicals with tetrahydrofuran in liquid solutions. *Phys. Chem. Chem. Phys.* **14**, 10 424–10 437. (doi:10.1039/c2cp40158d)