A Generalised Mechano-Kinetic Model For Use in Multiscale Simulation Protocols

A Preprint

Benjamin S. Hanson
Department of Molecular & Nanoscale Physics
School of Physics & Astronomy
University of Leeds
Leeds, LS2 9JT
b.s.hanson@leeds.ac.uk

Daniel J. Read
Department of Applied Mathematics
School of Physics & Astronomy
University of Leeds
Leeds, LS2 9JT
d.j.read@leeds.ac.uk

July 17, 2020

Abstract

Many biophysical systems and proteins undergo mesoscopic conformational changes in order to perform their biological function. However, these conformational changes often result from a cascade of atomistic interactions within a much larger overall object. For simulations of such systems, the computational cost of retaining high-resolution structural and dynamical information whilst at the same time observing large scale motions over long times is high. Furthermore, this cost is only going to increase as ever larger biological objects are observed experimentally at high resolution.

With insight from the theory of Markov state models and transition state theory, we derive a generalised mechano-kinetic simulation framework which aims to compensate for these disparate time scales. The framework models continuous dynamical motion at coarse-grained length scales whilst simultaneously including fine-detail, chemical reactions or conformational changes implicitly using a kinetic model. The kinetic model is coupled to the dynamic model in a highly generalised manner, such that it can be applied to any defined continuous energy landscape, and indeed, any existing dynamic simulation framework. We present a series of analytical examples to validate the framework, and showcase its capabilities for studying more complex systems by designing a simulation of minimalist molecular motor.

***K*eywords** Kinetics, Dynamics, Statistical Mechanics, Multiscale, Simulation, Molecular Motors

1 Introduction

Over the past decade, experimental techniques have matured to such an extent that we can now observe biological systems on length scales never before seen. Cryo-electron microscopy is enabling us to determine the atomistic structures of macromolecular complexes without the need for crystallisation, providing us with significantly more realistic instances of biological functional pathways[1, 2, 3]. Additionally, the continued development of various forms of super resolution microscopy and high throughput data analysis are also enabling us to observe motions over ever shorter time scales[4, 5]. Many of these structures are being deposited into the Electron Microscopy Data Bank (EMDB), which has seen an almost exponential growth rate over the past decade[6, 7]. Whilst simulation techniques are being developed which utilise this data[8], further consideration is needed with respect to the meaning of physical models at these newly explored spatio-temporal scales[9], and to the practicalities of performing computer simulations of these almost microscopic objects.

Take the molecular motor cytoplasmic dynein as an example system. Hanson *et al.* recently used existing all-atom molecular dynamics simulations of a single monomer of the motor to parameterise a much coarser continuum mechanical model of the same object[10], and found that the dominant dynamical motion was well represented in the coarser model. Hence, to simulate the equilibrium dynamics of the motor, a significantly coarser model is appropriate. From

the associated reduction in computational overhead, much longer and larger simulations can now be performed, such as a simulation of the motor in its native dimeric complex and attached to a microtubule environment. Yet the equilibrium fluctuations of the motor, even in this new environment, are only a small constituent of the overall functionality of the motor.

Experiments show that it is over much longer time scales that dynein undergoes its so called 'powerstroke', generating the directed force required for the motor to do the useful physical work of transporting cargo throughout the cytoskeletal network[11, 12, 13]. *In vitro* this powerstroke cycle, whilst constituted by cascading atomistic processes[14], is so slow compared to the dynamic time scales associated with the molecule that it is often modelled purely kinetically[15, 16]. As such, the differences in mechanical energies between the defined conformational states (known as the pre- and post-powerstroke states) are taken to be constants that can then modify the transition rates of the kinetic network as constant Boltzmann energy factors. The inspiring recent work of Trott *et al.* strongly deviates from this trend, introducing a highly detailed mechanistic model of cytoplasmic dynein which combines a wealth of experimental evidence to deduce the continuous function of the motor[17]. Examination of their computational insights indicates that from an atomistic statistical mechanics perspective, the mesoscopic transition rates emerge from the underlying dynamical behaviour.

The Arrhenius equation represents an early attempt to connect dynamical behaviour with chemical kinetics[18], which evolved into a more analytical form with the work of Eyring[19]. From there developed the generalised framework of transition state theory, in which the reaction rates of chemical reactions (often first-order) are determined from the statistical mechanical behaviour of the underlying atomistic constituents[20]. Finally, this led to the further generalised theory of Markov state models, in which we are able to define the kinetic states themselves from an analysis of the underlying continuous energy landscape[21, 22]. The ReaxFF molecular dynamics forcefield[23], the SEEKR computational tools package[24, 25] and the method of kinetic monte carlo[26] provide further examples of combining dynamic and kinetic modelling.

A Markov state model (MSM) is, at its core, a memoryless kinetic model. The MSM comprises of a series of discrete states with associated transition rates between them (see Figure 1), and those rates satisfy the Markov property of being 'memoryless', in that the current transition rates are state occupation probabilities are not affected by any of the previous transitions. What gives an MSM its predictive power is that the states themselves are not arbitrarily defined, but instead are calculated through an exploration of a dynamic energy landscape, either analytically or via simulation[27]. Local energy minima can be isolated and identified as the central loci of a set of discrete kinetic states. The rates of transition between the states, or energy minima, are then related to the physical energy barriers separating them, and as such combine the effects of the depth of the minima itself, the distance between minima in phase space, and any forms of activation energy required, such as chemical energy, catalysis or temperature.

A core goal of the application of Markov state models to a given simulation is to reduce the disjunction in time scales between the high-resolution equilibrium dynamical behaviour and the much longer time scales over which protein function often occurs, such as we described earlier with cytoplasmic dynein. Indeed, MSMs have already been successfully applied to the analysis of protein conformational changes such as folding, ligand binding, and structural changes about the native folded state[28]. The have also been used to quantitatively reconcile the results of simulation and experimental techniques[29], which also often apply to disparate time scales. One may notice, however, that within a Markov state model the total number of states defined is somewhat arbitrary. As noted by Bowman[21], the number of states used is often determined by the need to have sufficient statistics on the transitions between states so as to have confidence in the emergent transition rates between them. This is a purely practical limitation, as theoretically the number of states could be anywhere between one, representing the entire phase space as a single state with a probability of occupation of unity, and infinite, where in the infinite limit the system simply becomes a dynamic system and 'transitions' occur as per appropriate equations of motion. For MSMs with states that are 'small' with respect to the energetic environment, such that the mechanical energies within defined kinetic states are on the order of the free energy differences between kinetic states themselves, pure discretisation may not be a valid approximation. In this regime, which may well describe the loosely defined biological mesoscale, mechanical variations within the defined states may themselves become important.

Bialek *et al.* noted early on that protein dynamics and their reaction rates are coupled[30], such that a majority of the reactive events in some proteins can be coupled to specific protein modes of motion, indicating dynamic specificity. Brockwell *et al.* also discovered that forced protein unfolding with an atomic force microscope yields different unfolding forces depending on the directions in which you apply that force, indicating dynamic anisotropy, and Berkovich *et al.* showed that the collapse dynamics of unfolded proteins indicates force-dependent free energies of unfolding[31]. For cytoplamic dynein itself, a stalling force has been experimentally measured[11]. Thus, when under significant load or when opposed by other molecular motors such as kinesin[32], it may be the case that the instantaneous mechanical energy, not simply the average or minimum, plays a role in the overall kinetic cycle either by gating or accelerating

certain kinetic transitions. We must therefore consider how the reaction rates within a defined kinetic state vary as a function of our location in a defined energy landscape.

To address this considerable challenge in mesoscale modelling, we have developed a generalised 'mechano-kinetic' simulation framework which utilises both explicit mechanical modelling at mesoscopic length scales, whilst at the same time accounting for events beyond the scope of the dynamic model with a coupled kinetic model similar to a Markov state model. This coupling is such that the kinetic rates themselves dynamically alter as the system moves through the dynamic energy landscape and hence, using only the principle of detailed balance, our simulation framework enables the mesoscopic dynamical simulation to instantaneously contribute towards kinetic events such as conformational changes and chemical reactions, and thus implicitly affect dynamics at smaller length scales and vice versa.

2 Theoretical Development

Our aim with this work is to derive the generalised mechano-kinetic simulation framework. By 'mechano-kinetic' simulation we mean the coupling of two traditionally independent types of modelling. First we require a dynamic simulation framework, such as molecular dynamics (MD), with a well-defined energy landscape and associated system parametrisation. In the present paper we will later use an extremely simplified 1D dumbbell model utilising Brownian dynamics as an example case for the dynamic simulation method. For context, the ideas behind this paper were conceived with the aim of model molecular motors such as cytoplasmic dynein using a dynamical modelling method at an intermediate level of coarse graining between all-atom MD and simple dumbbell models[10]. Any such dynamic simulation requires some parameterisation of an energy landscape, which is used to calculate forces and dynamics of the constituent objects.

Having specified a dynamical simulation method, we will then couple an additional kinetic framework that will enable the energy landscape and parametrisation themselves to discretely change, on average, with some pre-defined kinetic transition rate. These kinetic states may be deduced, as with MSMs, or empirically defined by experimentation. However, as discussed in Section 1, to properly represent certain mesoscopic biophysical systems we require these kinetic transition rates to vary as a function of the dynamical movement through the mechanical energy landscape. This variation must be constrained in that any equilibrium transition rate for a given kinetic process will be correctly reproduced as the sum of transitions between all locations within the defined mechanical energy landscape.

2.1 Microscale Dynamics

We begin from first principles at the kinetic level with a set of N discrete kinetic states, commonly represented as in Figure 1. Within each kinetic state, the system we are considering (e.g. a biological molecule) can explore a set of configurations, which are microstates within a phase space Ω. Let us define the physical phase space of microstates as $\Omega = \Omega\left(\vec{x}\right)$ for simplicity, and recognise that we could expand to momentum coordinates $\vec{p}$ if required.

Each kinetic state has an associated occupation probability P_i and each pair of states has an associated transition rate R_{ij} between them (this would be the experimentally measured transition rates between kinetic states). However, we will assume that kinetic transitions between any two states i and j, actually occur between specific microstates $\left(\vec{x}\right)_i$ and $\left(\vec{x}'\right)_j$ within two the kinetic states. Our goal is to determine what the instantaneous transition rates between microstates, $r_{ij}\left(\vec{x}, \vec{x}'\right)$, are as a function of the ensemble average transition rates between the defined kinetic states, R_{ij}.

At equilibrium, the Boltzmann distribution gives us the probability density function for each kinetic state, $p_i\left(\vec{x}\right)$,

$$p_i\left(\vec{x}\right) = \frac{A_i}{Z_T} \exp\left(-\frac{E_i(\vec{x})}{k_B T}\right), \tag{1}$$

where A_i is a scaling factor with units dependent upon the defined phase space, and Z_T is the total partition function for the entire system of kinetic states and associated microstates, which will be defined shortly. In principle, for any defined phase space, we can develop a dynamic simulation framework with which to explore the defined energy landscape E_i. By numerically integrating appropriate equations of motion, we could therefore use Eq. 1 to calculate the specific structure of p_i as a function of the location in phase space. At this stage, however, such a simulation would only be sufficient to study the system within a single kinetic state. Integrating over the entire phase space within a single kinetic state gives us the occupation probability of that state, P_i

$$P_i = \int_\Omega d\Omega p_i, \tag{2}$$

and so we may define the partition function corresponding only to the kinetic state i, Z_i, as

$$Z_i = A_i \int_\Omega d\Omega \exp\left(-\frac{E_i(\vec{x})}{k_B T}\right) = \exp\left(-\frac{F_i}{k_B T}\right), \tag{3}$$

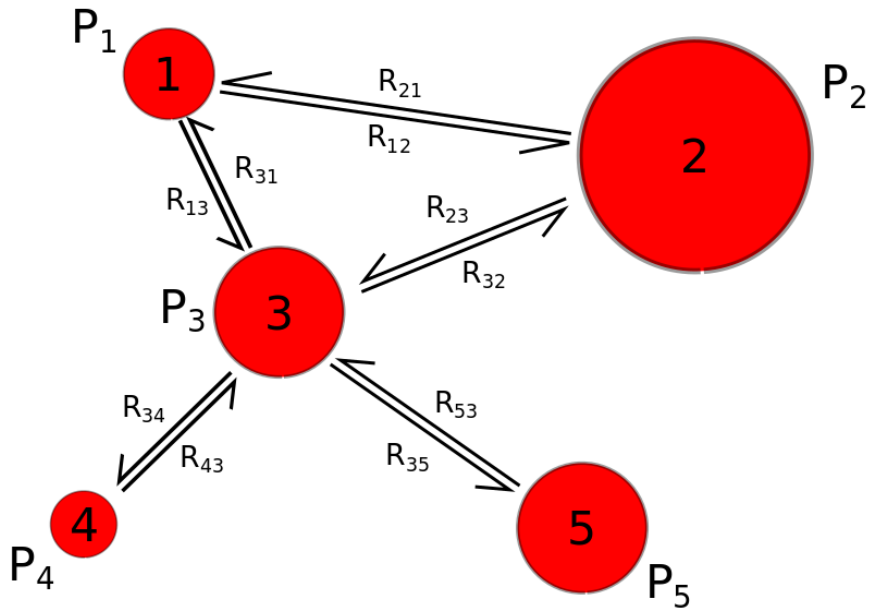


Figure 1: A dimensionless representation of a kinetic network with the number of states $N = 5$. Each state has an associated probability P_i which is determined by the number of available microstates within it. These probabilities then determine the relative transition rates R_{ij} between each pair of states i, j, as given by the principle of detailed balance.

where F_i is the free energy of the kinetic state i, and more specifically the Helmholtz free energy when considering mechanical systems at constant volume and temperature. Eqs. 1, 2 and 3 show that if there is more than a single kinetic state, such that $Z_i \neq Z_T$, then p_i is not normalised over Ω. This implies that in this framework, transitions between the defined kinetic states i and j do not correspond only to movement through phase space, but potentially to some other process of energy change. Depending on the specific system we are considering, and which spatio-temporal regime the kinetic states correspond to, these changes may be chemical changes, structural conformational changes or some other unknown internal or external change which alters the free energy landscape over the domain Ω.

For a complete set of states, where the probability of being somewhere within the defined kinetic system is equal to one, Z_T must act as a normalisation constant for the set of occupation probabilities P_i

$$P_i = \frac{1}{Z_T} \exp\left(-\frac{F_i}{k_B T}\right) = \frac{Z_i}{Z_T}, \tag{4}$$

and hence

$$Z_T = \sum_i^N \exp\left(-\frac{F_i}{k_B T}\right). \tag{5}$$

To emphasise the biological regime we are implicitly considering here, we will refer to the general set of discrete kinetic states we as 'mesostates' from this point forward.

2.2 Mesoscale Kinetics

Just as the dynamic simulation protocol enables movement through phase space within each mesostate, transitions can occur between the mesostates themselves. Whether these rates are empirically measured or deduced to form a Markov state model, the principle of detailed balance lets us relate the transition rates between these mesostates to the associated occupation probabilities at equilibrium

$$R_{ij} P_i = R_{ji} P_j, \tag{6}$$

where R_{ij} is the rate of transitioning from mesostate i to mesostate j, irrespective of the specific microstate. Substitution of Eq. 4 into Eq. 6 yields

$$\frac{R_{ij}}{R_{ji}} = \exp\left(-\frac{\Delta F_{ij}}{k_B T}\right), \tag{7}$$

where $\Delta F_{ij} = F_j - F_i$.

This would now be sufficient detail to set up a pure kinetic model. If we can measure the values of P_i experimentally for a given system, we can determine the free energy differences ΔF_{ij} between mesostates i and j. Further, if we can measure the independent values of R_{ij} then we can determine the individual free energy values F_i, which subsequently enables us to make predictions about different networks containing the same states. However, our interest is in simulating not only these mesostate networks, nor simply the dynamic behaviour local to each mesostate, but rather the transitions between any two microstates, whether dynamically within each mesostate or kinetically between mesostates. This requires a coupling between the two types of simulation framework.

2.3 Coupling the Regimes

The similarity in form between Eqs. 1 and 4 and the continuous scale over which we may define discrete kinetic states suggests that we may define further detailed balance conditions between any two specific microstates in different mesostates

$$r_{ij}\left(\vec{x},\vec{x}'\right)p_i\left(\vec{x}\right) = r_{ji}\left(\vec{x}',\vec{x}\right)p_j\left(\vec{x}'\right). \tag{8}$$

Substitution of Eq. 1 into Eq. 8 yields

$$\frac{r_{ij}}{r_{ji}} = \frac{A_j}{A_i}\exp\left(-\frac{\Delta E_{ij}\left(\vec{x},\vec{x}'\right)}{k_B T}\right), \tag{9}$$

where $\Delta E_{ij}\left(\vec{x},\vec{x}'\right) = E_j\left(\vec{x}'\right) - E_i\left(\vec{x}\right)$ is the change in total energy between microstates, and r_{ij} is the transition rate from mesostate i to mesostate j with the further specification of starting at location $\vec{x}$ in phase space and finishing in location $\vec{x}'$.

We now consider the sources of these energies. The total energy of the system in each microstate can be written as

$$E_i\left(\vec{x}\right) = E_i^c + E_i^m\left(\vec{x}\right), \tag{10}$$

where E_i^m is the mechanical energy corresponding to the specific location in Ω, which is used to determine the dynamics within the dynamical simulation framework. E_i^c is any other source of energy, e.g. an additional chemical energy of state i. Substitution of Eq. 10 into Eq. 9 allows us to separate all mechanical and non-mechanical components of the energy changes, and hence define rates r_{ij}^c that are specific to the non-mechanical components of energy

$$\frac{r_{ij}^c}{r_{ji}^c} = \frac{A_j}{A_i}\exp\left(-\frac{\Delta E_{ij}^c}{k_B T}\right), \tag{11}$$

where $\Delta E_{ij}^c = E_j^c - E_i^c$. Therefore

$$\frac{r_{ij}}{r_{ji}} = \frac{r_{ij}^c}{r_{ji}^c}\exp\left(-\frac{\Delta E_{ij}^m}{k_B T}\right), \tag{12}$$

where $\Delta E_{ij}^m = E_j^m - E_i^m$. With insight from the work of Sarlah and Vilfan[33, 15], we now introduce an additional term f_{ij}^m in order to to separate the detailed balance condition of Eq. 12 into individual rates

$$r_{ij} = r_{ij}^c\exp\left(-\frac{f_{ij}^m\Delta E_{ij}^m}{k_B T}\right), \tag{13}$$

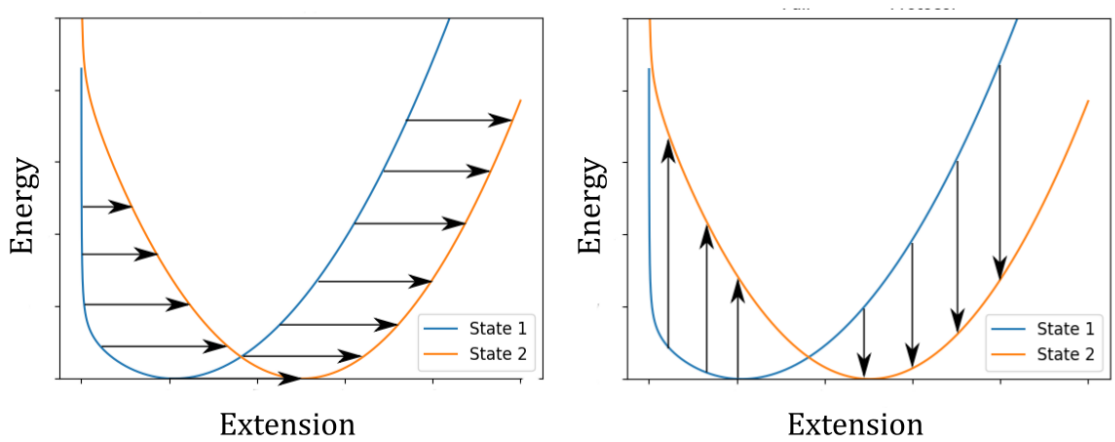
where, in order for Eq. 13 to be consistent with Eq. 12, we require that

$$f_{ij}^m + f_{ji}^m = 1. \tag{14}$$

The physical meaning of f_{ij}^m is related to form of the activation energy barrier for the transition between specific microstates $\left(\vec{x}\right)_i$ and $\left(\vec{x}'\right)_j$ and in principle, there may be an additional term of the same form for each 'type' of energy contained within ΔE_{ij}^m (kinetic, potential, work etc). This is explored in detail in the Supplementary Information Section S2. For now, we take them as necessary to recover Eq. 12 from Eq. 13.

Now that Eq. 13 has provided us with an expression for the individual microstate reaction rates, we need to couple this to the ensemble transition rate R_{ij}. As with a Markov state model, we do this by noting that the mesostate transition rate from mesostate i to mesostate j are the sum of all transitions from all points $\vec{x}$ within mesostate i to all points $\vec{x}'$ within mesostate j . Hence, in complete generality

$$R_{ij}P_i = \int_{\Omega} d\Omega \int_{\Omega'} d\Omega' r_{ij}\left(\vec{x},\vec{x}'\right)p_i\left(\vec{x}\right), \tag{15}$$



(a) (b)

Figure 2: Two possible methods for describing a transition between mesostates at the microstate level, where the black arrows indicate the 1-1 mapping between initial and final states. In principle, any set of transitions is possible so long as they collectively form a complete 1-1 mapping for all of phase space. a) A mesostate transition protocol with no associated microstate energy change. b) A mesostate transition protocol with no associated movement through phase space.

and substituting Eq. 2 into Eq. 15

$$R_{ij} = \frac{\int_\Omega d\Omega \int_{\Omega'} d\Omega' r_{ij} p_i}{\int_\Omega d\Omega p_i}. \tag{16}$$

Recall again that R_{ij} may be an experimentally measurable parameter, the reaction rate between mesostates irrespective of the location in phase space, whereas r_{ij} is the more fundamental transition rate between specific microstates in different mesostates as the system explores phase space. At this stage, r^c_{ij} may still be a function of the location in Ω and so would remain under the integrals when substituting Eqs. 1 and 13 into Eq. 16. This yields

$$R_{ij} = \frac{\int_\Omega d\Omega \int_{\Omega'} d\Omega' r^c_{ij} \exp\left(-\frac{f^m_{ij}\Delta E^m_{ij}}{k_B T}\right) \exp\left(-\frac{E_i(\vec{x})}{k_B T}\right)}{\int_\Omega d\Omega \exp\left(-\frac{E_i(\vec{x})}{k_B T}\right)}. \tag{17}$$

where the constant terms A_i and Z_T have cancelled. We now make the approximation that all non-mechanical energies comprising E^c_{ij} are independent of phase space, and thus r^c_{ij} is constant. For a simulation involving these rates, once the energy landscapes have been defined, the set of r^c_{ij} become the determining parameters for the entire mechano-kinetic framework. Expanding Eq. 17 in terms of explicit energies, making r^c_{ij} the subject and simplifying, we obtain

$$r^c_{ij} = R_{ij} \frac{\int_\Omega d\Omega \exp\left(-\frac{E^m_i(\vec{x})}{k_B T}\right)}{\int_\Omega d\Omega \int_{\Omega'} d\Omega' \exp\left(-\frac{f^m_{ji} E^m_i(\vec{x}) + f^m_{ij} E^m_j(\vec{x}')}{k_B T}\right)}. \tag{18}$$

Supplementary Information Section S3 discusses how Eq. 18 relates to the individual partition functions of each mesostate, Z_i, at this stage. We note that following a kinetic transition, Eq. 18 still enables us to have a target location in phase space different to the base location. Figure 2 shows two possibilities for such a transition, where a pure kinetic model would likely assume something like Figure 2a to be valid, with no mechanical energy change but some movement though phase space, perhaps between energy minima. However, dynamic movement through phase space is complex, and if relevant to the question at hand ought to be accounted for explicitly by a dynamic simulation. For this work, we consider only the case in which the system remains stationary in phase space as it undergoes the kinetic transition

between mesostates, corresponding to Figure 2b. This is equivalent to introducing delta functions $\delta\left(\vec{x}' - \vec{x}\right)$ for each phase space coordinate into the double integral in Eq. 16, yielding the core equation for this work

$$r_{ij}^{c} = R_{ij} \frac{\int_{\Omega} d\Omega \exp\left(-\frac{E_i^m(\vec{x})}{k_B T}\right)}{\int_{\Omega} d\Omega \exp\left(-\frac{f_{ji}^m E_i^m(\vec{x}) + f_{ij}^m E_j^m(\vec{x})}{k_B T}\right)}. \tag{19}$$

If R_{ij} is experimentally known, say, and we have sufficient knowledge of the energy landscape to calculate the integrals in Eq. 19, then we can now calculate r_{ij}^c. The integrals may be calculated analytically for simple energy landscapes, as we will see in Section 3, or if they are too complex then they may be calculated via Monte Carlo integration or explicit dynamic simulation of the individual mesostate energy landscapes as in Markov state models. Either way, with r_{ij}^c we can use Eq. 13 to calculate the instantaneous transition rate from mesostate i to mesostate j from any specific location in phase space. Thus we have derived the required coupling to generate a mechano-kinetic simulation framework. With respect to the instantaneous rate r_{ij} we can now consider r_{ij}^c to be an intrinsic, constant rate which is modulated by dynamic movement through phase space within the current energy landscape E_i^m and with a target energy landscape E_j^m. The collective set of kinetic transitions between mesostates i and j from each individual pair of microstates $(\vec{x})_i$ and $(\vec{x})_j$ subsequently results in the average, experimentally measurable rate R_{ij}. We can determine what the dynamical exploration of a mechanical energy landscape must be to result in the measured kinetic transition rates.

Eq. 19 is valid at equilibrium, allowing the calculation of r_{ij}^c. Having determined this, we can also consider situations where further external changes are made to the form of the mechanical energy (e.g. from an external force applied to a mesoscopic biomolecule), with the assumption that Eq. 13 still applies with the equilibrium values of r_{ij}^c. The scheme then becomes predictive, with modifications made to the mechanical energy landscape propagating through Eq. 13 and resulting in new mesostate transition rates R_{ij}.

In the following section we will show how the integrals in Eq. 19 can be solved analytically for a simple mechano-kinetic system. We present a series of example simulations with this system which showcase the capabilities of such a coupled framework, showing that a kinetic model appropriately coupled to a dynamic simulation creates a mechno-kinetic framework which can not only be used to infer dynamical behaviour, but that has predictive capabilities at the kinetic level.

3 Results

To validate the mechano-kinetic framework and investigate its capabilities, we consider a simple dumbbell as our example system. The object is made of two spherical nodes connected by a Hookean spring with elastic constant k_i and equilibrium length l_i. The system is immersed in a background viscous fluid which imparts a local drag $\lambda = 6\pi\mu R$ onto each node independently, where μ is the fluid viscosity and R is the node radius.

The dynamic simulation component used for this example simulation is a Brownian dynamics protocol. As such, inertia and the associated velocity degrees of freedom are neglected. For further simplicity, we will restrict the object to one dimension and consider only the single spatial degree of freedom separating the two dumbbell nodes. Hence, phase space contains all points along the real number line, $\Omega = \{x \in \mathbb{R}\}$, and $d\Omega = dx$. The associated equations of motion are

$$\lambda \frac{dx_1}{dt} - k_i\left((x_2 - x_1) - l_i\right) = f_1, \tag{20}$$

$$\lambda \frac{dx_2}{dt} + k_i\left((x_2 - x_1) - l_i\right) = f_2, \tag{21}$$

for nodes 1 and 2 respectively, where x_1 and x_2 are the specific positions of nodes 1 and 2, and f_1 and f_2 are the associated stochastic thermal noise terms with statistics in accordance with the fluctuation-dissipation theorem. Using a first-order integration approach, the relevant discrete-time equations are

$$x_1(t + \Delta t) = x_1(t) + \Delta t \frac{k_i}{\lambda}\left((x_2(t) - x_1(t)) - l_i\right) + \frac{\Delta t}{\lambda} f_1(t), \tag{22}$$

$$x_2(t + \Delta t) = x_2(t) - \Delta t \frac{k_i}{\lambda}\left((x_2(t) - x_1(t)) - l_i\right) + \frac{\Delta t}{\lambda} f_2(t), \tag{23}$$

where the discrete-time statistics of f_1 and f_2 are

$$\langle f_n(t) \rangle = 0, \tag{24}$$

$$\langle f_n(t) f_m(t') \rangle = \begin{cases} \frac{2k_B T \lambda}{\Delta t} \delta_{nm} & t = t' \\ 0 & t \neq t' \end{cases} \tag{25}$$

The repeated solution of Eqs. 22 and 23 constitutes our dynamic simulation component. We keep all simulations numerically stable and physically accurate by setting the integration timestep Δt to be a factor of ten smaller than the smallest dynamic time scale within whichever system we simulate.

The kinetic simulation component utilises the sets of internal spring parameters, k_i and l_i. These two parameters define the mesostate, i.e. k_i and l_i are the parameters for mesostate i, and will vary between mesostates. When altered, these parameters change the energy landscape independently of the location of the dumbbell nodes within phase space. This is akin to a biological conformational change of a highly coarse-grained molecule, such that all of the internal atomic degrees of freedom implicitly change and result in a different set of mesoscale parameters. In principal all parameters could vary between mesostates, but for clarity we will leave $R = 1$nm, $\mu = 0.001$Pa.s constant, and consider only kinetic alterations in the elastic energy landscape. λ is therefore also constant.

At any dynamic simulation timestep we include the probability of a mesostate transition as per Eq. 13, where we convert the rates to transition probabilities by multiplying by the simulation timestep. Eq. 13 therefore makes these transition probabilities explicitly dependent on the current microstate mechanical energy. If a transition occurs, we keep the two node positions constant, change the spring parameters, and allow the simulation to continue.

With these independent definitions of the dynamic and kinetic simulation components, we can now use the derivations of Section 2 to couple them together. For this system the mechanical energy term within Eq. 19 is greatly simplified

$$E_i^m = \frac{1}{2} k_i \left(x - l_i\right)^2, \tag{26}$$

where $x = x_1 - x_2$ defines the single elastic degree of freedom within the system. Introducing this energy function into Eq. 19 and expressing the denominator integral exponent as a polynomial in x gives

$$r_{ij}^c = R_{ij} \frac{\int_{-\infty}^{\infty} dx \exp\left(-\frac{k_i (x - l_i)^2}{2k_B T}\right)}{\int_{-\infty}^{\infty} dx \exp\left(-\frac{\left(f_{ji}^m k_i + f_{ij}^m k_j\right) x^2 - 2\left(f_{ji}^m k_i l_i + f_{ij}^m k_j l_j\right) x + f_{ji}^m k_i l_i^2 + f_{ij}^m k_j l_j^2}{2k_B T}\right)}. \tag{27}$$

Eq. 27 contains only Gaussian integrals in x and so can be solved exactly, yielding

$$R_{ij} = r_{ij}^c \frac{1}{\sqrt{1 + f_{ij}^m \left(\frac{k_j}{k_i} - 1\right)}} \exp\left(-\frac{f_{ij}^m \left(1 - f_{ij}^m\right) k_j \left(l_i - l_j\right)^2}{2k_B T \left(1 + f_{ij}^m \left(\frac{k_j}{k_i} - 1\right)\right)}\right), \tag{28}$$

where we have made R_{ij} the subject for clarity. Eq. 28 completes the coupling between dynamic and kinetic simulation components within our mechano-kinetic framework. The equation is analysed in detail in Supplementary Information Section S4 to provide a more intuitive understanding of what is happening when we use introduce r_{ij}^c into a mechano-kinetic simulation to obtain the variable instantaneous values of r_{ij}. Additionally, the introduction of instantaneous kinetic transitions within a dynamic simulation also introduces another timescale related to numerical accuracy and statistical equilibration, the derivation of which is provided in Supplementary Information Section S5. Finally, physical validation of each simulation component is provided in Supplementary Information Section S6.

From here, we show example simulations of two different mechano-kinetic systems to investigate the effect of instantaneous coupling between dynamic and kinetic frameworks.

3.1 A Reaction Coordinate

To investigate the behaviour of our model when applied to a realistic physical process, we have designed a series of kinetic transitions corresponding to a bidirectional reaction coordinate in which the average length of the dumbbell

increases with as the "reaction" progresses. We define two alternative parameterisations of such a set of kinetic mesostates, given by their mechanical parameters in Table 1, with an associated equilibrium kinetic transition rate matrix for both sets given in Table 2.

Mesostate (i)	k_i (pN/nm)	l_i(nm)
1	10	1
2	10	2
3	10	3
4	10	4

Mesostate (i)	k_i (pN/nm)	l_i(nm)
1	200	1
2	50	2
3	50	3
4	200	4

Table 1: The two families of material parameters used to define the mesostates in the reaction coordinate kinetic networks of the dumbbell model. The first (left) set contains equal and relatively low values of k_i whereas the second (right) set of k_i are much higher. The values of l_i linearly increase with the mesostate index, providing a well defined reaction coordinate.

Transition Rates, R_{ij} (μs^{-1})					P_i
From / To	1	2	3	4	
1	N/A	9.158	-	-	0.013
2	3.369	N/A	500	-	0.035
3	-	67.67	N/A	183.94	0.256
4	-	-	67.67	N/A	0.696

Table 2: The average transition rates between mesostates used in the reaction coordinate kinetic networks of the dumbbell model. The mesostate occupation probabilities have been analytically calculated for comparison with simulation.

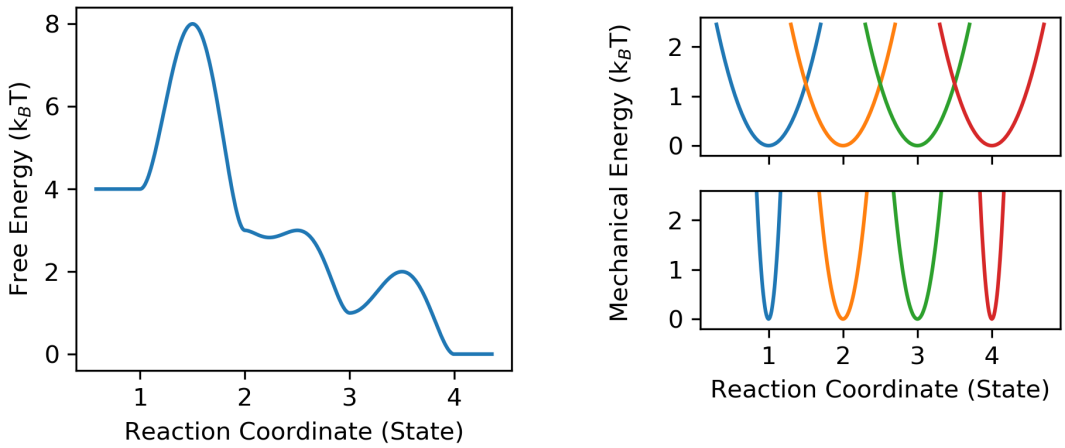
As Table 2 defines all of the expected transition rates, which hypothetically speaking may have been experimentally measured, it also implicitly defines the mesoscale free energy landscape and all of the resultant occupation probabilities. Thus our investigation will tell us nothing new at the mesoscale level at this stage. Instead, our aim is to infer the what the continuous underlying dynamical behaviour must be, given that we have measured the mesoscopic kinetic behaviour. Note that many of the transition rates shown in Table 2 are unspecified, indicating that direct transitions cannot occur between these states. However, transitions via other intermediate states can still occur (i.e. from mesostate 1 to mesostate 3 via mesostate 2).

The energy landscapes implied by this kinetic system are illustrated in Figure 3. We see in Figure 3b that as all of the stiffness parameters k_i are equal in the first parameterisation, no particular state is preferable over long periods of time with respect to mechanical energy. On the other hand, in the second parameterisation we include two high stiffness states (high k_i) and two low stiffness states (low k_i). However, irrespective of the mechanical energy environment, the free energy as a function of reaction co-ordinate implied by the average rates, shown in Figure 3a, shows that we have implicitly included two free energy barriers, a metastable state and an overall difference in the free energy minima between the final state and the initial state. This free energy landscape, the free energy barriers and associated rates, were calculated using the analysis presented in Supplementary Information Section S2 where $f_{ij}^m = 0.5$ for all i, j.

3.1.1 Reaction Coordinate - Results

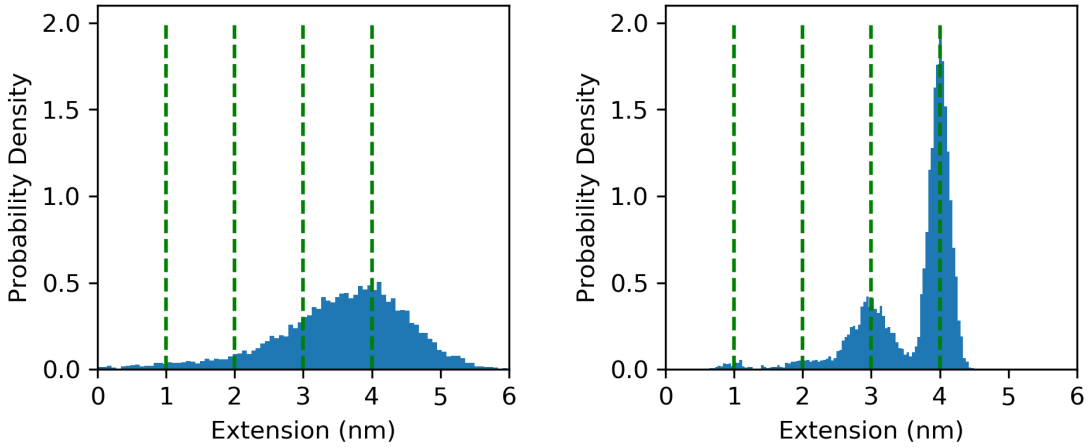
For each mechanical simulation parameterisation, Table 4 shows the rates r_{ij}^c which are modified by the mechanical energy landscape to obtain the R_{ij} values shown in Table 2. Although the r_{ij}^c values calculated and used in each simulation are different, Table 3 shows the equilibrium mesostate occupation probabilities for both sets of mechanical material parameterisations, calculated by averaging the expected occupation probabilities from a single, long-time simulation over many repeats beginning in different initial states. We can see that the mesostate occupation probabilities for the two simulations are approximately equivalent to the analytically calculated values given in Table 2, despite the large differences in the values of k_i. This indicates that the mesostate occupation probabilities within the model do indeed remain constant, being fixed by the R_{ij} values, whilst the underlying mechanical energy landscape is modified. Incidentally, the largest uncertainties correspond to mesostates with the largest associated kinetic rates shown in Table 4, showing that we may be approaching the numerical limits defined in Supplementary Information Section S5 with this system.

Figure 4 shows the probability density of length fluctuations at equilibrium from a single simulation of each set of mesostates. We see that whilst the kinetic rate modifications within our model take the mechanical energies into account in order to return the average rates R_{ij}, the mechanical energy landscape is still fully realised within the framework. In



(a) (b)

Figure 3: The two representations of energy landscape within the reaction coordinate kinetic networks of the dumbbell model. a) The form of the free energy as a function of reaction co-ordinate implied by the rates given in Table 2. b) The mechanical energy landscapes given in Table 1, where the occupation of each state is given a different colour. Top: Set 1 - All values of k_i are equal. Top: Set 2 - Values of k_i vary between mesostates.



(a) (b)

Figure 4: The microstate probability densities from simulation of the reaction coordinate kinetic networks of the dumbbell model. The mechanical equilibrium positions corresponding to each state are shown as dashed green lines. a) Mesostate set 1 b) Mesostate set 2.

	Mesostate Occupation Probabilities, P_i			
Mesostate	1	2	3	4
Set 1	0.0148(4)	0.0358(3)	0.256(1)	0.693(1)
Set 2	0.013(1)	0.037(1)	0.269(2)	0.681(1)

Table 3: The final expected mesostate occupation probabilities within the reaction coordinate kinetic networks of the dumbbell model. Each set corresponds to a different set of material parameters as defined in Table 1. Uncertainties in the last digit are shown in brackets.

Transition Rates, r^c_{ij} (μs^{-1})				
From / To	1	2	3	4
1	N/A	12.41	-	-
2	4.566	N/A	677.61	-
3	-	91.71	N/A	249.28
4	-	-	91.71	N/A

Transition Rates, r^c_{ij} (μs^{-1})				
From / To	1	2	3	4
1	N/A	82.38	-	-
2	60.61	N/A	2285.64	-
3	-	309.34	N/A	3309.08
4	-	-	608.69	N/A

Table 4: The instantaneous transition rates between mesostates used in the reaction coordinate kinetic networks of the dumbbell model corresponding to the two sets of mesostates given in Table 1.

particular, we see that the low mechanical stiffness of individual states in Figure 4a means that length fluctuations within each state are relatively large compared to the associated equilibrium lengths, resulting in a unimodal (albeit skewed) overall distribution centred around, but not exactly on, the most likely mesostate. Within a standard Markov state model, each of these states would likely be indistinguishable from one another, yet our implied knowledge of both the mesostate kinetics and mechanical energy landscape allows us to include them anyway, such that their indistinguishably is effectively a deduced property. On the other hand, for the second parameterisation in Figure 4b, the higher mechanical stiffness suppresses the length fluctuations within each mesostate so that the individual mesostates are distinguishable in the overall length distribution, with obvious Gaussian distributions. This is less apparent for mesostates 1 and 2, however, as their associated free energies are significantly higher, and occupation probabilities lower.

Overall we see that within the mechano-kinetic framework, no matter the structure of the mechanical energy landscape, the introduction of Eq. 28 into Eq. 13 enables the simulation to reproduce the R_{ij} values used as input parameters. As such, we have used the measured kinetic rates given in Table 2 together with the spring parameters in Table 1 to infer the specific form of the mechanical energy landscape. Further information on the interdependence of the kinetic and dynamic regimes with respect to the mechanical energy is given in Supplementary Information Section S6.3. However, we will see in the next section that we can go further than the simple reproduction of existing measurements.

3.2 External Forces

For a simulation framework to be of use, it ought to have predictive properties as well as being able to reproduce equilibrium results. Considering our example problem detailed in Section 1, we may wish to know the effect of applying external mechanical forces to biological objects such as molecular motors. Consider the case where single molecule experiments have been performed and both the free energy and mechanical energy landscapes of the object are well known[34, 10]. We can use our mechano-kinetic simulation framework to advance beyond this single molecule experimental system.

To explore this type of problem in a general sense, we have designed two kinetic networks, both of which use the set of mesostates defined in Table 5. The first network is similar to the first example in Section 3.1, whereas the second network is a fully connected kinetic network, where all transitions between states are possible. The associated transition rates defining these networks are given in Table 6, left and right tables respectively. In contrast to the reaction coordinate example, we observe here that because all of the rates are the same, this system initially ought to have an

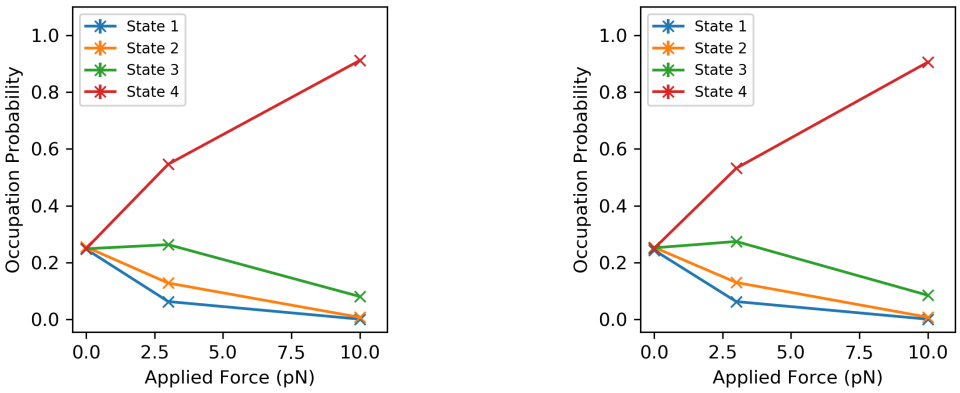
Mesostate (i)	k_i (pN/nm)	l_i(nm)
1	25	1
2	25	2
3	25	3
4	25	4

Table 5: The material parameters used to define the mesostates used in the kinetic network of the dumbbell model with external forces applied.

Transition Rates, R_{ij} (μs^{-1})					P_i
From / To	1	2	3	4	
1	N/A	50	-	-	0.25
2	50	N/A	50	-	0.25
3	-	50	N/A	50	0.25
4	-	-	50	N/A	0.25

Transition Rates, R_{ij} (μs^{-1})					P_i
From / To	1	2	3	4	
1	N/A	15	15	15	0.25
2	15	N/A	15	15	0.25
3	15	15	N/A	15	0.25
4	15	15	15	N/A	0.25

Table 6: The matrices defining the two sets of average transition rates between mesostates, together with the expected mesostate occupation probabilities, used in the kinetic network of the dumbbell model with external forces applied. Left: A single reaction coordinate through phase space. Right: A fully connected kinetic network.



(a) (b)

Figure 5: The expected occupation probability of each mesostate of the dumbbell model as a function of applied force. a) The reaction coordinate network. b) The fully connected network.

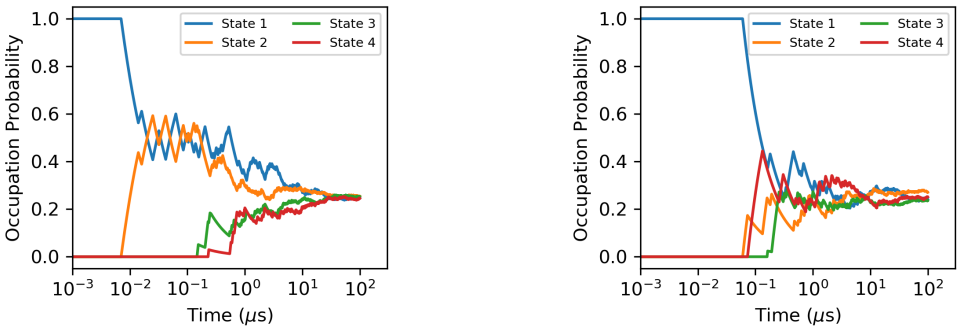
equal probability of being in each of the mesostates. However, under the application of external forces we would expect that the forces may bias the states in some manner. Within the dynamic simulation component Eqs. 20 and 21, we apply a constant external force f^{ex} to both nodes of the dumbbell but in opposite directions, therefore biasing the dumbbell to higher extensions. We therefore assume that the values of r_{ij}^c remain as given at equilibrium, but we modify the reaction rates via Eq. 13 to account for additional contributions to the mechanical energy from the external forces. With this approach we are able to use our mechano-kinetic model to predict not only how the application of this external force within the dynamical simulation affects the overall kinetic transition rates between mesostates, but also determine where those transitions are likely to occur in the mechanical energy landscape.

3.2.1 External Forces - Results

Figure 5 shows the steady state occupation probabilities of each mesostate as a function of applied force, for both sets of rates. These were again calculated by averaging the expected occupation probabilities from a single, long-time simulation over many repeats beginning in different initial states. Although all of the initial transition rates in Table 6 and values of k_i in Table 5 are equal, indicating no energetic preference for any of the mesostates (as shown with zero external force applied), the application of force immediately biases the system towards larger extensions. We also see that the steady-state occupation probabilities as a function of applied force are independent of the connectivity of the kinetic network, but are rather given simply by the Boltzmann distribution, modified to account for the extra energy contribution from the externally applied force. Detailed balance implies that this change is brought about within the model via a change in the resulting values of R_{ij}, propagated from the dynamic component into the kinetic component of the model.

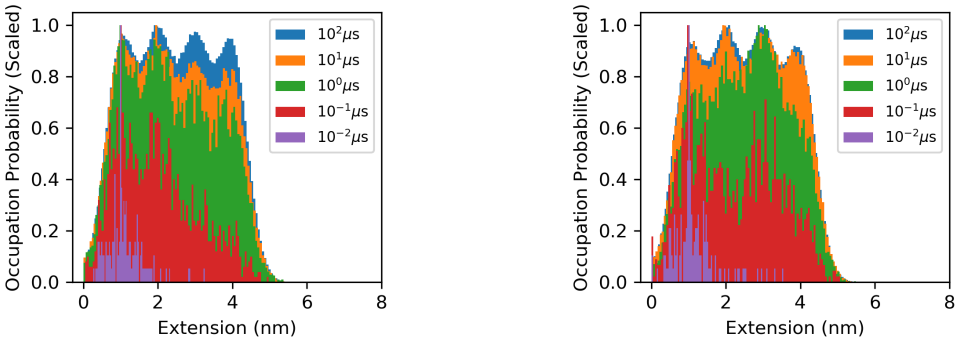
Forces such as this can be included in pure kinetic models as free energies, ΔF_{ij}^{ex}, using a term such as $\Delta F_{ij}^{ex} = f^{ex}\Delta x_{ij}$ where x_{ij} is some approximation of the expected extension. However, with our explicit inclusion of the mechanical energy environment, we can observe the dynamical evolution of both microstate and mesostate occupation probabilities.

First we consider Figure 6, which shows the running averages of occupation probabilities of each mesostate from two representative simulations under no external forces, the system having begun in mesostate 1. Figure 6a, representing the



(a) (b)

Figure 6: Running averages of the occupation probability of each mesostate of the dumbbell model as a function of applied force. These graphs show a single, representative simulation from a full set of repeats all beginning in mesostate 1. a) The reaction coordinate network. b) The fully connected network.



(a) (b)

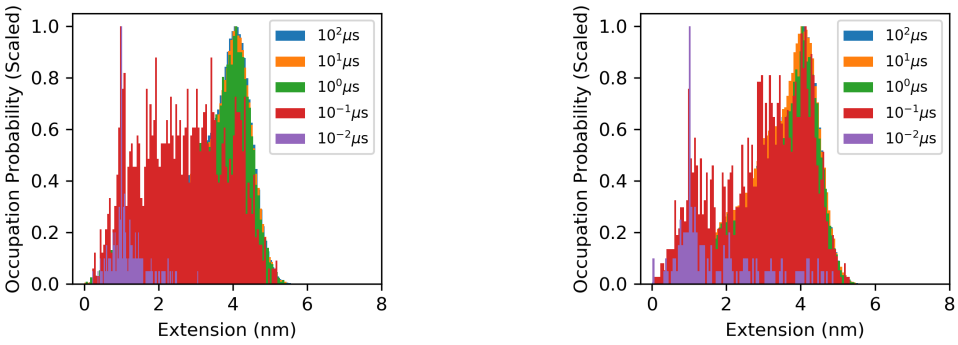
Figure 7: The microstate probability distributions of the dumbbell model at various points in time under the application of no external forces. a) The reaction coordinate. b) The connected network.

reaction coordinate network, shows that the system must pass through each of the intermediate states to get to mesostate 4, and hence equilibration proceeds continuously through the phase space. On the other hand, Figure 6b, representing the fully connected network, shows a much more chaotic equilibration.

Figure 7 shows this dynamic growth more clearly, visualising the (scaled) microstate probability distributions as they evolve over time. These distributions were obtained by combining the microstate probability distributions of multiple repeated simulations at the same after each length time, all repeats having been started in mesostate 1. We see that neighbouring peaks emerge one after another in Figure 7a, indicated the direction of the reaction pathway. In Figure 7b, however, whilst a single peak is visible at short times (time scales smaller than the average mesostate transition rate given in Table 6), once mesoscale equilibration begins to take place, the peaks grow much more evenly.

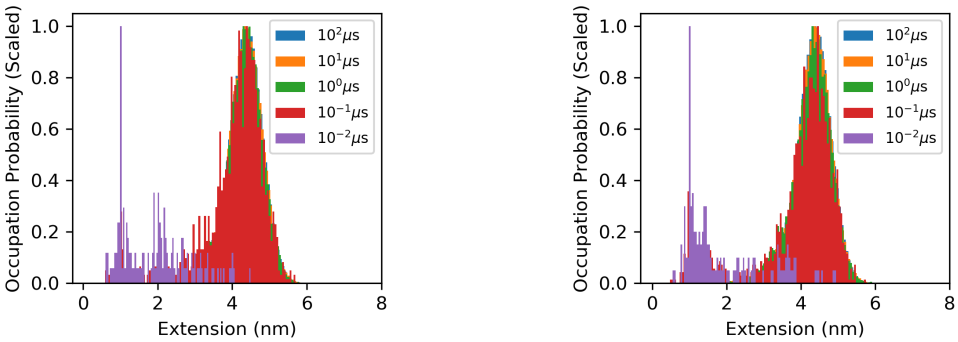
Finally, we observe these same microstate distributions under the application of an external forces. Figure 8 shows the systems under an external force, $f^{ex} = 3$pN, and Figure 9 with $f^{ex} = 10$pN.

Again we observe a single peak at short times, before the system under force has had a chance to relax, a single peak centred on the initial mesostate. At long times, however, we see very little difference between the fully connected network and the reaction coordinate network. The force across all mesostates biases the system to higher extensions in general, especially for the $f^{ex} = 10$pN simulations. The underlying exploration of phase space allowed at low force is prohibited at high force such that the majority of transition pathways are no longer possible. Interestingly at $f^{ex} = 3$pN, the intermediate times show start differences between the reaction coordinate and fully connected networks. In the reaction coordinate network, we observe a relatively flat distribution at intermediate times whereas in the fully connected network, we observe a much more rapid growth of the final peak. Here we see that the introduction of a kinetic process allowing a jumps through phases space (straight from mesostate 1 to mesostate 4, say) provides



(a) (b)

Figure 8: The microstate probability distributions of the dumbbell model at various points in time under the application of an external force $f^{ex} = 3$pN. a) The reaction coordinate network. b) The fully connected network.



(a) (b)

Figure 9: The microstate probability distributions of the dumbbell model at various points in time under the application of an external force $f^{ex} = 10$pN. a) The reaction coordinate network. b) The fully connected network.

more pathways which are positively biased by the force. In the context of a molecular motor, these may represent some conformational change or reaction that deviates from the most commonly observed pathway, and which may be exacerbated by the application of external force (detachment from the microtubule in the absence of ATP, for example [15]).

We have seen that the application of force across all mesostates effectively generates an additional contribution to the free energy, fixing the steady state distribution. However, the dynamical motion through phase space that we infer from the application of force at the continuous microstate level is dependent upon the structure and connectivity of the the kinetic landscape, i.e. which kinetic transitions are permitted. Further, we could in principle add force only in a subset of kinetic states, including it in the definition of a mesostate as we have been doing with spring parameters, showing the true generality of our model.

The above is suggestive of alternative methods to parameterise the model from experiments. It might be, for example, that it is not possible to obtain R_{ij} from equilibrium experimentation (e.g. if the rates are too slow). A viable alternative, then, is to measure transition rates accelerated under the application of external forces. Such forces modify the energy landscape, but if these modifications are included in 19 and used to derive an alternate version of Eq. 28, then the necessary r^c_{ij} values can be obtained from the measured R_{ij} under external force. Then on removal of the force from the model, the equilibrium R_{ij} (in the absence of force) may be inferred. This is a further example of how this mechano-kinetic framework can go beyond simply reproducing the initial parametrisation from experiment, extending the simulations to experimentally difficult and potentially infeasible scenarios for predictive purposes.

4 Discussion

Our mechano-kinetic model was developed with molecular motors in mind, with the intention of inferring the effect of explicit mechanical interactions on the emergent kinetics, and perform subsequent predictive simulations from there. However, we believe the model is significantly more general that that, and we described with more general application in mind. The most obvious is perhaps a chemical reaction, where we can measure average rates of reaction at the macroscale without considering specific bond fluctuations and mechanics at the microscale. However, from a fundamental chemical perspective, those specific specific interactions can allow us to determine reaction mechanisms and potentially new reaction pathways. The inclusion of free energies of reaction as chemical potentials in the deviation presented in Section 2 could allow similar inductions of the atomistic mechanical energy environment, given experimentally measured reaction rates. The application of external mechanical stimuli to complex reaction networks is of clear interest.

We must also appreciate that there remains a large gap between the upper limit of accessible time scales available to computer simulations, and the lower limit of accessible time scales accessible by experiments. The Markov state model allows for the definition of discrete kinetic states through the analysis of purely dynamic trajectories. However, as we have seen implicitly in this work, the difference in the characteristic time scales between dynamical motion through phase space and kinetic transitions is often large due to the size of the energy barriers involved (ATP hydrolysis is associated with a free energy change of k_BT, for example). It is only very recently with the advent of GPU technology[35, 36] and hardware-specific simulations[37] that fully atomistic simulations of biological objects have been able to reach microsecond time scales and beyond. On the other hand, the empirical measurement of these long-time kinetic events is often much more easily achieved due to the 'top-down' approach of the experimental procedures involved. With this in mind, we have provided a mechanism which is almost the reverse of a Markov state model, beginning with the kinetic information and inferring the underlying dynamical behaviour. In reality, of course, the difference in time scales between the two regimes is still wide. In this work we have vastly accelerated our kinetic transitions to compensate for this and subsequently verify the equilibrium behaviour of both the continuous dynamical phase space and the discrete kinetic space. However, for systems with disparate time scales, such acceleration is justified for analysis so long as the kinetic and dynamic processes remain separated, and dynamic equilibration can be reached following each kinetic transition. We have therefore provided a generalised mechanism for a true coupling of kinetic and dynamic processes, potentially encompassing the full span of the biological mesoscale in a physically consistent manner.

5 Conclusions

With insight from the theory of Markov state models, we have developed a generalised coupling of kinetic and dynamic simulations, dubbed a mechano-kinetic simulation framework. Using various test cases, we have shown that our model can be used to infer the underlying mechanical behaviour within a continuous phase space given a knowledge of the kinetic processes, and that the model can be used to predict new kinetic behaviours through the inclusion of additional external energies within the mechanical model.

Via Eq. 19, this protocol can be integrated into any existing dynamic simulation protocol and used to explore free energy landscapes which may be inaccessible to fully dynamic simulations. We note that direct integration of Eq. 19 may become analytically infeasible for more complex systems than those studied here. Yet, if a dynamic simulation protocol already exists to which this kinetic model can be applied, then the integrals involved could be calculated using the dynamical simulation protocol as a microstate sampling method, in the same manner as a standard Markov state model, or via Monte Carlo methods implementing the same energy functions. Here, the assumption is that the dynamical or Monte Carlo simulations samples states according to the Boltzmann distribution, so that the integrals over phase space in Eq. 19 are approximated by a sum over the states encountered in the simulation.

The biological mesoscale represents the frontier of both experimentation[3] and simulation[36], the regime in which biophysicists from both disciplines can truly meet and compare results. We envisage that our model will help to reconcile these commonly discrete regimes.

6 References

References

[1] M. Bates, B Huang, Michael J. Rust, Graham T. Dempsey, W. Wang, and X Zhuang. Sub-Diffraction-Limit Imaging with Stochastic Optical Reconstruction Microscopy. In Gräslund A., Rigler R., and Widengren J., editors,

Single Molecule Spectroscopy in Chemistry, Physics and Biology. Springer Berlin Heidelberg, 2010.

[2] Yifan Cheng. Single-particle Cryo-EM at crystallographic resolution. *Cell*, 161(3):450–457, 2015.

[3] Werner Kühlbrandt. The Resolution Revolution. *Science*, 343(6178):1443–1444, 2014.

[4] Joachim Frank. Time-resolved Cryo-Electron Microscopy: Recent Progress. *Journal of Structural Biology*, 200(3):303–306, 2017.

[5] Lothar Schermelleh, Alexia Ferrand, Thomas Huser, Christian Eggeling, Markus Sauer, Oliver Biehlmaier, and Gregor P.C. Drummen. Super-resolution microscopy demystified. *Nature Cell Biology*, 21(1):72–84, 2019.

[6] Catherine L. Lawson, Matthew L. Baker, Christoph Best, Chunxiao Bi, Matthew Dougherty, Powei Feng, Glen Van Ginkel, Batsal Devkota, Ingvar Lagerstedt, Steven J. Ludtke, Richard H. Newman, Tom J. Oldfield, Ian Rees, Gaurav Sahni, Raul Sala, Sameer Velankar, Joe Warren, John D. Westbrook, Kim Henrick, Gerard J. Kleywegt, Helen M. Berman, and Wah Chiu. EMDataBank.org: Unified data resource for CryoEM. *Nucleic Acids Research*, 39(SUPPL. 1):456–464, 2011.

[7] EMDB. EMDB Statistics. https://www.ebi.ac.uk/pdbe/emdb/statistics_releases.html/, 2020.

[8] Albert Solernou, Benjamin S. Hanson, Robin A. Richardson, Robert Welch, Daniel J. Read, Oliver G. Harlen, and Sarah A. Harris. Fluctuating Finite Element Analysis (FFEA): A continuum mechanics software tool for mesoscale simulation of biomolecules. *PLoS Computational Biology*, 14(3):1–6, 2018.

[9] Joanna Leng, Massa Shoura, Tom C.B. McLeish, Alan N. Real, Mariann Hardey, James McCafferty, Neil A. Ranson, and Sarah A. Harris. Securing the future of research computing in the biosciences. *PLoS Computational Biology*, 15(5):1–15, 2019.

[10] Benjamin S. Hanson, Shinji Iida, Daniel J. Read, Oliver G. Harlen, Genji Kurisu, Haruki Nakamura, and Sarah A. Harris. Continuum Mechanical Parameterisation of Cytoplasmic Dynein from Atomistic Simulation. *Methods*, 2020.

[11] Yoshimi Kinoshita, Taketoshi Kambara, Kaori Nishikawa, Motoshi Kaya, and Hideo Higuchi. Step Sizes and Rate Constants of Single-headed Cytoplasmic Dynein Measured with Optical Tweezers. *Scientific Reports*, 8(1):1–14, 2018.

[12] Lin Li, Joshua Alper, and Emil Alexov. Cytoplasmic dynein binding, run length, and velocity are guided by long-range electrostatic interactions. *Scientific Reports*, 6:31523, 2016.

[13] Samara L. Reck-Peterson, Ahmet Yildiz, Andrew P. Carter, Arne Gennerich, Nan Zhang, and Ronald D. Vale. Single-Molecule Analysis of Dynein Processivity and Stepping Behavior. *Cell*, 126(2):335–348, 2006.

[14] Shintaroh Kubo, Wenfei Li, and Shoji Takada. Allosteric conformational change cascade in cytoplasmic dynein revealed by structure-based molecular simulations. *PLoS Computational Biology*, 13(9):1–27, 2017.

[15] Andreja Šarlah and Andrej Vilfan. The winch model can explain both coordinated and uncoordinated stepping of cytoplasmic dynein. *Biophysical Journal*, 107(3):662–671, 2014.

[16] Kazuo Sasaki, Motoshi Kaya, and Hideo Higuchi. A Unified Walking Model for Dimeric Motor Proteins. *Biophysical Journal*, 115(10):1981–1992, 2018.

[17] Laurie E. Trott, M. Hafezparast, and A. Madzvamuse. A mathematical understanding of how cytoplasmic dynein walks on microtubules. *Royal Society Open Science*, 5(8), 2018.

[18] Svante Arrhenius. Über die Dissociationswärme und den Einfluss der Temperatur auf den Dissociationsgrad der Elektrolyte. *Zeitschrift für Physikalische Chemie*, 4(1):96–116, 1889.

[19] Henry Eyring. The Activated Complex in Chemical Reactions. *The Journal of Chemical Physics*, 3(2):107–115, 1935.

[20] Keith J. Laidler and M. Christine King. The development of transition-state theory. *Journal of Physical Chemistry*, 87(15):2657–2664, 1983.

[21] Gregory R Bowman, Vijay S Pande, and Frank Noe. *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*. Springer, 2014.

[22] Brooke E. Husic and Vijay S. Pande. Markov State Models: From an Art to a Science. *Journal of the American Chemical Society*, 140(7):2386–2396, 2018.

[23] Thomas P. Senftle, Sungwook Hong, Md Mahbubul Islam, Sudhir B. Kylasa, Yuanxia Zheng, Yun Kyung Shin, Chad Junkermeier, Roman Engel-Herbert, Michael J. Janik, Hasan Metin Aktulga, Toon Verstraelen, Ananth Grama, and Adri C.T. Van Duin. The ReaxFF reactive force-field: Development, applications and future directions. *npj Computational Materials*, 2, 2016.

[24] Lane W. Votapka and Rommie E. Amaro. Multiscale Estimation of Binding Kinetics Using Brownian Dynamics, Molecular Dynamics and Milestoning. *PLoS Computational Biology*, 11(10):1–24, 2015.

[25] Lane W. Votapka, Benjamin R. Jagger, Alexandra L. Heyneman, and Rommie E. Amaro. SEEKR: Simulation Enabled Estimation of Kinetic Rates, A Computational Tool to Estimate Molecular Kinetics and its Application to Trypsin-Benzamidine Binding. *Journal of Physical Chemistry B*, 121(15):3597–3606, 2017.

[26] Arthur F Voter. Introduction to the Kinetic Monte Carlo Method. In Kurt E. Sickafus, Eugene A. Kotomin, and Blas P Uberuaga, editors, *Radiation Effects in Solids*, chapter 1, pages 1–23. Springer Netherlands, 2007.

[27] Arti Bhoutekar, Susmita Ghosh, Swati Bhattacharya, and Abhijit Chatterjee. A new class of enhanced kinetic sampling methods for building Markov state models. *Journal of Chemical Physics*, 147(15), 2017.

[28] John D Chodera and Frank Noe. Markov state models of biomolecular conformational dynamics. *Current Opinion in Structural Biology*, 25:135–144, 2014.

[29] Simon Olsson, Hao Wu, Fabian Paul, Cecilia Clementi, and Frank Noé. Combining experimental and simulation data of molecular processes via augmented Markov models. *Proceedings of the National Academy of Sciences of the United States of America*, 114(31):8265–8270, 2017.

[30] W. Bialek and J. N. Onuchic. Protein dynamics and reaction rates: mode-specific chemistry in large molecules? *Proceedings of the National Academy of Sciences of the United States of America*, 85(16):5908–5912, 1988.

[31] Ronen Berkovich, Sergi Garcia-Manyes, Michael Urbakh, Joseph Klafter, and Julio M. Fernandez. Collapse dynamics of single proteins extended by force. *Biophysical Journal*, 98(11):2692–2701, 2010.

[32] Kazuka G. Ohashi, Lifeng Han, Brandon Mentley, Jiaxuan Wang, John Fricks, and William O. Hancock. Load-dependent detachment kinetics plays a key role in bidirectional cargo transport by kinesin and dynein. *Traffic*, 20(4):284–294, 2019.

[33] Andreja Šarlah and Andrej Vilfan. Minimum requirements for motility of a processive motor protein. *PLoS ONE*, 12(10):1–13, 2017.

[34] Helgo Schmidt. Dynein motors: How AAA+ ring opening and closing coordinates microtubule binding and linker movement. *BioEssays*, 37(5):532–543, 2015.

[35] John E Stone, James C Phillips, Peter L Freddolino, David J Hardy, Leonardo G Trabuco, and Klaus Schulten. Accelerating Molecular Modeling Applications with Graphics Processors. *Journal of Computational Chemistry*, 28(16):2618–2640, 2007.

[36] Peter Eastman, Jason Swails, John D. Chodera, Robert T. McGibbon, Yutong Zhao, Kyle A. Beauchamp, Lee Ping Wang, Andrew C. Simmonett, Matthew P. Harrigan, Chaya D. Stern, Rafal P. Wiewiora, Bernard R. Brooks, and Vijay S. Pande. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Computational Biology*, 13(7):1–17, 2017.

[37] David E Shaw, Martin M Deneroff, Ron O Dror, Jeffrey S Kuskin, Richard H Larson, John K Salmon, Cliff Young, Brannon Batson, Kevin J Bowers, Jack C Chao, Michael P Eastwood, Joseph Gagliardo, J P Grossman, C Richard Ho, Douglas J Ierardi, István Kolossváry, John L Klepeis, Timothy Layman, Christine Mcleavey, Mark A Moraes, Rolf Mueller, Edward C Priest, Yibing Shan, Jochen Spengler, Michael Theobald, Brian Towles, and Stanley C Wang. Anton, a special-purpose machine for molecular dynamics simulation. *Communications of the ACM*, 51(7):91–97, 2008.

7 Acknowledgements

We would like to thank Professors Haruki Nakamura and Genji Kurisu for conversations on the nature of cytoplasmic dynein that led to the conceptualisation of this work. Ben Hanson was supported by the EPSRC through a DTA.