



REVIEW

Open Access

Mathematical and computational modeling in biology at multiple scales

Jack A Tuszynski^{1*}, Philip Winter², Diana White², Chih-Yuan Tseng², Kamlesh K Sahu³, Francesco Gentile⁴, Ivana Spasevska⁵, Sara Ibrahim Omar², Niloofar Nayeibi³, Cassandra DM Churchill⁶, Mariusz Klobukowski⁶ and Rabab M Abou El-Magd⁷

* Correspondence: jackt@ualberta.ca
¹Department of Physics and
Department of Oncology, University
of Alberta, Edmonton, Canada
Full list of author information is
available at the end of the article

Abstract

A variety of topics are reviewed in the area of mathematical and computational modeling in biology, covering the range of scales from populations of organisms to electrons in atoms. The use of maximum entropy as an inference tool in the fields of biology and drug discovery is discussed. Mathematical and computational methods and models in the areas of epidemiology, cell physiology and cancer are surveyed. The technique of molecular dynamics is covered, with special attention to force fields for protein simulations and methods for the calculation of solvation free energies. The utility of quantum mechanical methods in biophysical and biochemical modeling is explored. The field of computational enzymology is examined.

Keywords: Mathematical biology, Epidemiological models, Cellular physiological models, Cancer models, Maximum entropy, Molecular dynamics, Force fields, Solvation free energies, Quantum mechanics, Computational enzymology

Introduction

Mathematical, computational and physical methods have been applied in biology and medicine to study phenomena at a wide range of size scales, from the global human population all the way down to the level of individual atoms within a biomolecule. Concomitant with this range of sizes between global to atomistic, the relevant modeling methods span time scales varying between years and picoseconds, depending on the area of interest (from evolutionary to atomistic effects) and relevance. This review will cover some of the most common and useful mathematical and computational methods. Firstly, we outline the maximum entropy principle as an inference tool for the study of phenomena at different scales, from gene evolution and gene networks to protein-drug molecular interactions, followed with a survey of the methods used for large scale systems—populations, organisms, and cells—and then zooming down to the methods used to study individual biomolecules—proteins and drugs. To study the large systems, the most common and reliable mathematical technique is to develop systems of differential equations. At the molecular scale, molecular dynamics is often used to model biomolecules as a system of moving Newtonian particles with interactions defined by a force field, with various methods employed to handle the challenge of solvent effects. In some cases, pure quantum mechanics methods can and should be used,

which describe molecules using either wave functions or electron densities, although computational costs in time and resources may be prohibitive, so hybrid classical-quantum methods are often more appropriate. Quantum methods can be particularly valuable in the study of enzymes and enzymatic reactions.

Maximum entropy in biology and drug discovery

Two reasoning methods, deduction and inductive inference, have been utilized in the development of theories to interpret phenomena we observe in nature, and to make predictions about complex systems. Deduction allows us to draw conclusions when sufficient information is available, and is contrasted with inductive inference (also known as inductive logic or probable inference). Inductive inference provides a least biased way to reason when the available information is insufficient for deduction. It is called “inference” when we make estimates of quantities for which we do not have enough information to use deductive reasoning, and “induction” when we are generalizing from special cases [1].

When we deal with complex systems, for example either many-body interactions at the microscopic level, complicated regulatory protein-protein networks at the mesoscopic level, or population genetics at the macroscopic level, we never have enough knowledge to completely understand the system. Therefore, we normally rely on inductive inference based on the available information to infer the most preferred solution to problems related to these systems. Particularly, we are interested in a mathematical tool for inductive inference based on the Bayesian interpretation of probability, the rules of probability theory, and the concept of entropy. Bayesian interpretation treats probability as a degree of our knowledge about a system of interest, rather than the frequency of appearance of an event. Cox demonstrated that this type of probability can be manipulated by the rules of standard probability theory [2]. This forms the building blocks of inductive inference, termed Bayesian inference. Moreover, Caticha and Giffin have shown that Bayesian inference is a special case of entropy-based inference [3]. Therefore, our discussion in this section will be founded on entropy-based inference.

First, we briefly address the basics of entropy-based inference, which includes using entropy as an information measure and a tool for inductive inference, then we provide examples in the fields of biology and drug discovery to demonstrate that these fields benefit from the application of inductive inference. Regarding using entropy as an information measure, we consider two examples. The first example provides a clue to investigate genomic evolution through appropriate genomic sequence analysis [4]. The second one discusses robustness of biological networks from information point of view [5]. Regarding using entropy as a tool for inductive inference, we consider another two examples. The first one demonstrates the benefit of introducing this inference in virtual screening for drug discovery [6]. The second one then shows an application of this scheme in fragment-based drug design [7]. These examples also illustrate in a straightforward way how to extract information and unveil the global characteristics of the biological systems. They show that there exists a universal reasoning platform to solve any problem of interest that is independent of the specifics of a given type of problem. The key in this platform lies in the answer to the question, “What are the constraints in the system?” Once the constraints are determined, the maximum entropy principle provides a robust, universal and

least biased prescription for information processing. Furthermore, it helps us to analyze problems and gain insights into the functioning of the underlying complex systems.

Entropy as an information measure

Shannon's pioneering work on the quantification of information loss during communication established a new viewpoint on entropy, which was until then only known as a measure of randomness in thermodynamics [8]. Since then, using entropy as an information measure has attracted much attention not only in signal processing, but also in the field of biology.

Information in genomic evolution

With the advance of genomic sequencing technology there is more and more genomic sequence data available for species across the three domains: Bacteria, Archaea, and Eukaryota. The question is, how do we compare complete genomes and extract useful information from the sequencing data?

To address the question of genome comparison, Chang et al. [4] proposed an entropy-based scheme for complete genome comparison. The foundation of their approach is to define the probability distribution that represents our current state of knowledge regarding the occurrence of different combinations of the four bases in DNA sequences. Chang et al. [4] specified that k -mer nucleotides in the sequence encode genetic information, where k is an arbitrary number. The occurrence of k -mers in a DNA sequence characterizes that sequence. Based on this definition, information in sequences can be quantified with Shannon information. Furthermore, Chang et al. [4] introduced the concept of reduced Shannon information, which is defined as the ratio of the Shannon information of the genome to the Shannon information of random sequences, so as to quantify to what extent the information contained in the genome is different from the information in a random DNA sequence. Note that this concept is similar to the concept of relative entropy, which is discussed in the next section. Based on reduced Shannon information (or relative entropy), a universal feature across three taxonomic domains was observed; namely, the effective root-sequence length of a genome, which is defined as the ratio of genome length and reduced Shannon information, linearly depended on k , and was a genome-independent constant. Furthermore, this study revealed a possible genome growth mechanism: at an early stage of evolution, a genome is likely to utilize random segmental duplication, which would maximize reduced Shannon information. These insights not only provide a clue to the origin of evolution but also may shed light on further questions, such as which genes are responsible for drug resistance.

Robustness of biological networks

The development of high throughput screening techniques, such as microarray technology, has generated numerous protein-protein interaction data to map out biological networks, and has revealed regulatory mechanisms of the biological entities involved in the networks. Many studies have suggested that the robustness of biological networks may be the key for identifying systems that can tolerate external perturbations and uncertainty triggered by external forces [9,10]. It has further been shown that the robustness of biological networks shares a global feature; these networks are scale-free networks, which means that one can observe a power-law degree distribution in these networks.

Therefore, there have been many endeavors to provide more insights into the origin of these power-law distributions. Bak et al. [8] proposed the mechanism of self-organized criticality, which leads to a scale-free structure in complicated systems. An entropy-based interpretation described by Dover [5] suggested a promising and intuitive way to understand the emergence of power-law distributions in complicated networks. According to Dover's studies on a toy model, the emergence of the power-law distributions is merely a consequence of the maximum entropy principle when the internal order of sub-networks of a complicated large network remained fixed. Note that the internal order was defined as the mean of the Boltzmann entropy over all sub-networks. In the framework of entropy-based inference, the power-law distributions of biological networks simply represent the most preferred choice that maintains the fixed internal order of the sub-networks.

Entropic scheme for inductive inference

In addition to the use of entropy as an information measure, the concept of entropy also plays a role in inductive inference. The inductive inference addressed here refers to two processes. The first process is the determination of the most likely state of knowledge about a system of interest based on the information in hand. The second process is the determination of the most likely updated state of knowledge when we acquire new information regarding the system. The foundation of inductive inference is the maximum entropy principle and relative entropy. The least biased inference one can make based on the information in hand is the one that maximizes the relative entropy of all possible new and old beliefs. For more details, the reader is referred to Caticha [1].

Entropy in molecular docking

Our first example of applying entropy for inductive inference is *in silico* drug discovery. Virtual screening has attracted much attention in the pharmaceutical industry [12,13]. It provides a more economical way to screen diverse chemicals as drug candidates compared with a wet-lab approach. Basically, it consists of the creation of a chemical library, followed by searching optimal ligand-receptor binding modes through docking algorithms, and finally the evaluation of binding affinities. There are three criteria that are required to successfully identify drug candidates. First, the chemical library needs to be large and contain diverse chemical structures. Second, conformational search algorithms need to be able to search possible binding modes within a reasonable time. Third, an appropriate scoring function needs to be utilized to correctly evaluate the binding affinity of the chemical structures. In the framework of information theory, the first and third criteria are the fundamental information required in virtual screening process. The second criterion then can be treated as an information processing guideline. The efficiency and accuracy of this step will depend on the methods of information processing.

Genetic algorithms, which borrow from the concept of genomic evolution processes to search conformations of complex targets and chemical structures, are commonly used in docking protocols, such as AutoDock [14]. Chang et al. have offered a better alternative, MEdock [6]. Although MEdock did not completely exploit entropic-based inductive inference for searching, it does utilize the maximum entropy principle as a guideline to make decisions during this process. The fundamental question asked in MEdock is "What is the probability of finding the deepest energy valley in a ligand-

target interaction energy landscape?" Maximum entropy provides a direction to update the initial guess of binding modes (described by an almost uniform distribution) to the optimal mode (a localized distribution around the global energy minimum).

Entropy in aptamer design

The second example of entropy for inductive inference is aptamer design. Aptamers are short nucleic acid sequences that are traditionally identified through an experimental technique, the Systematic Evolution of Ligands by Exponential Enrichment (SELEX) [15,16]. Aptamers can bind to specific molecular targets including small molecules, proteins, nucleic acids, and phospholipids, and can also be targeted to complex structures such as cells, tissues, bacteria, and other organisms. Because of their strong and specific binding through molecular recognition, aptamers are promising tools in molecular biology and have both therapeutic and diagnostic clinical applications [15-18]. Unfortunately, some limitations of SELEX have slowed the progress of discovering specific aptamers for various applications [18]. With the help of entropy-based inductive inference, a fragment-based approach has been developed to design aptamers given the structure of the target of interest [18].

The concept of the fragment-based approach to aptamer design is to ask the question "Given the structural information about the target, what is the preferred probability distribution of having an aptamer that is most likely to interact with the target?" The solution was found using entropy-based inductive inference [7]. This approach initially determines the preferred probability distribution of first single nucleotide that likely interacts with the target. Subsequently, the approach iteratively updates the probability distribution as more nucleotides are added to the growing aptamer. The maximum entropy principle allows us to determine to what extent this update is sufficient, and what is the sequence of nucleotides that is most likely to bind to the target. This method has been applied to design aptamers to bind specifically to targets such as thrombin, phosphatidylserine [19] and galectin-3 (under experimental confirmation).

The maximum entropy principle and inductive inference just provide one reasoning platform to make the most preferable inference based on all kinds of information for understanding biological systems at different scales. In the next section, a variety of mathematical and computational models addressing other aspects that have been developed for biological and medical problems are surveyed.

Mathematical and computational models for biological systems

In recent years, mathematical biology has emerged as a prominent area of interdisciplinary scientific research. It is not a new area of research, but with recent advances in medical and computational methods, it has grown extensively, being applied to solve many health related problems across a spectrum of life sciences. Areas of mathematical biology where modeling has made contributions to biology and medicine include epidemiology, cell physiology, cancer modeling, genetics, cellular biology, and biochemistry. Because there is such a broad range of topics and methods that can be discussed, we limit ourselves to a discussion of how differential equations have been used to solve important biological problems in epidemiology, cell physiology, and cancer modeling, and briefly discuss some of the clinical advances that have arisen from such efforts. For a more extensive review on mathematical modeling for each of these branches of

science, we refer the reader to recent books on these topics [20-23]. For the reader who is interested in learning more about mathematical biology from a beginner's perspective, books by Edelstein-Keshet [24], Murray [25,26], and Britton [27] are also recommended.

Here, we highlight only a few of the models that have been developed to study epidemiological, physiological, and cancer problems. The reader is encouraged to look more extensively into the literature regarding other models that have been developed and successfully applied to improve present medical treatments.

Mathematical models in epidemiology

Epidemiology describes the study of patterns, cause and effect, and treatment of disease within a given population [28]. Here, we provide a brief introduction to epidemiological models used for studying the spread of various types of disease, many of which are outlined in [25,27]. One of the first models to describe the dynamics of a disease caused by a viral infection is the so-called SIR model, an ordinary differential equation (ODE) model developed by Kermack and Mckendrick in 1927 [29].

$$\frac{dS}{dt} = -\beta IS \tag{1}$$

$$\frac{dI}{dt} = \beta IS - \mu I \tag{2}$$

$$\frac{dR}{dt} = \mu I \tag{3}$$

This model, given by equations (1), (2), and (3) (for all differential equation models we omit initial and boundary conditions for ease in reading), describes the rate of change of the number of susceptible (S), infected (I), and recovered (R) individuals in a population over time, where β describes the rate of transmission of disease, and μ describes the rate of removal of infected individuals (those that have recovered). An important feature of this model is that it incorporates recovered patients, meaning that an individual can acquire immunity, as is often the case for viral-type infections like influenza and measles. This model, although quite basic, provides important information to health care professionals interested in understanding how severe an outbreak is. For example, from these equations, the basic reproduction number given by

$$R_0 = \frac{\beta S(0)}{\mu} \tag{4}$$

describes the average number of secondary infections produced by one infected individual introduced into a completely susceptible environment. High values of R_0 , corresponding to high numbers of initially susceptible individuals $S(0)$, and/or high disease transmission rates β , result in the high probability of an outbreak. In particular, if R_0 is less than 1, the infection will not persist (and will eventually die out), whereas if R_0 is greater than 1, the infection will grow (and there will be an epidemic).

One key assumption of this model is that the total population N ($N = S + I + R$) is constant and that there is no death or birth. Many models have since been developed to include such population demographics [30-32], the first being completed by Soper [31] in an attempt to understand the dynamics of measles.

A number of extensions have been made to describe a wider class of infections. For example, the SIRS and SIS models allow for the movement of individuals back into a susceptible class S , meaning there may be no immunity to re-infection [30]. Such models are useful in studying bacterial-type infections like tuberculosis and gonorrhoea. Other models, referred to as SEIR and SEIRS models, where “E” stands for a latent class of individuals (exposed but not showing symptoms), can be used to describe a disease where a delayed time of infection may exist [33]. For example, this is often the case with individuals suffering from malaria.

A disadvantage of ODE-based modeling is that it assumes the well mixing of large populations of individuals. Also, such models are deterministic, meaning that the outcome is determined solely on the initial conditions and the parameters that govern the dynamics. For some populations, where contacts and transmission rates between individuals may vary, agent based [34,35] stochastic [36] or network type [36] models may be more useful. Also, age-structured models [37] may be more appropriate for diseases that depend on age, such as AIDS.

Another disadvantage of ODE-based modeling is that it does not describe the movement of individuals through space. This information is extremely important because a disease may not just spread within a single population, but may spread from one location to another. Examples of models that incorporate spatial dynamics include partial differential equations (PDEs). These models have been used to study the outbreak of rabies in continental Europe in the early 1940s [38], as well as to study the more recent outbreak of the West Nile Virus in 1999 in New York State [39]. Other models used to study the spatial spread of disease include patch models [40]. In the patch model of Lloyd and May [40] the authors consider an SEIR modeling approach. Here, the total population is broken up into subpopulations (patches), where S_i , E_i , I_i , and R_i denote the number of susceptible, exposed (latent), infected, and recovered individuals, in each patch i , respectively. The dynamics of each patch are governed by their own similar set of differential equations,

$$\frac{dS_i}{dt} = dN_i - dS_i - \sigma_i S_i, \quad (5)$$

$$\frac{dE_i}{dt} = \sigma_i S_i - (d + \epsilon) E_i, \quad (6)$$

$$\frac{dI_i}{dt} = \epsilon E_i - (d + \gamma) I_i, \quad (7)$$

$$\frac{dR_i}{dt} = \gamma I_i - dR_i. \quad (8)$$

In each patch, all model parameters are the same, except the infection rate σ_i , which depends on each connection between patches. Here, σ_i is called the force of infection, and is given by the mass action expression

$$\sigma_i = \sum_{j=1}^n \beta_{ij} I_j, \quad (9)$$

where n is the total patch number and β_{ij} is the rate of infection between patches i and j .

Mathematical models have influenced protocol in disease control and management. Now, such modeling is part of epidemiology policy decision making in many countries [41]. Some important modeling contributions include the design and analysis of epidemiology surveys, determining data that should be collected, identifying trends and forecasting outbreaks, as well as estimating the uncertainty in these outbreaks.

Physiological models at the cellular level: enzyme kinetics, ion channels, and cell excitability

The field of physiology is arguably the number one biological field where mathematics has had the greatest impact. Two broad areas of physiology where mathematics has made a profound impact are cell physiology and systems physiology. Here, we focus on cell physiology, and restrict ourselves to the topics of enzyme kinetics, ion channels, and cell excitability. For an excellent review on systems physiology, the reader is referred to Keener and Sneyd [22].

The rate of change of a simple chemical reaction can be described by the law of mass action, a law that describes the behavior of solutions in dynamic equilibrium [42]. That is, for a simple reaction given by



where k_+ is the reaction rate constant, the rate of the production of product molecules is given by

$$\frac{d[C]}{dt} = k_+[A][B], \quad (11)$$

where $[X]$ is the concentration of each species $X = A, B, C$. Equation (11) is commonly referred to as the law of mass action. The above formulation can only be used for the simplest of reactions involving a single step and only two reactants, although extensions are fairly straight-forward and have been developed over the past century to describe more complicated reactions [43]. One example is the model of Michaelis and Menten [44], used to describe reactions catalyzed by enzymes. Given the reaction scheme



where S is the substrate, E the enzyme, P the product concentration, and k_1 , k_{-1} , and k_2 are the reaction rate constants, Michaelis and Menten describe this reaction by the following four ODEs,

$$\frac{ds}{dt} = k_{-1}c - k_1se, \quad (13)$$

$$\frac{de}{dt} = (k_{-1} + k_2)c - k_1se, \quad (14)$$

$$\frac{dc}{dt} = k_1se - (k_2 + k_{-1})c, \quad (15)$$

$$\frac{dp}{dt} = k_2c. \quad (16)$$

Assuming that the substrate is at equilibrium ($ds/dt = 0$), one can simplify this system and find explicit solutions. Also, without solving these equations, one can gain useful

information about the process. For example, the velocity of the reaction (the rate at which the products are formed) is

$$V = \frac{dp}{dt} = \frac{V_{\max}s}{K_s + s}, \quad (17)$$

where $V_{\max} = k_2e_0$ and $K_s = k_{-1}/k_1$ (K_s is called the equilibrium constant). Equation (17) is often referred to as the Michaelis–Menten equation. Also, the steady-state approximation simplifies the above system of equations so that we can find explicit solutions [45]. This approximation requires that the rates of formation and breakdown of the complex c are essentially always equal ($dc/dt = 0$). Further extensions of this model have been developed to describe other types of enzyme activity, such as cooperativity [46] and enzyme inhibition [47], and have had similar success.

Computational systems biology has been creating a series of tools that are useful for application to enzyme kinetics. This is particularly true in the area of parameter estimation where several algorithms have been shown to be useful to enzyme kinetics. The availability of increasingly sophisticated and standardized modeling and simulation software will undoubtedly benefit enzyme kinetics [48].

Biochemical networks are sets of reactions that are linked by common substrates and products. The dynamics of biochemical networks are frequently described as sets of coupled ODEs, similar to those given by equations (13) through (16), that represent the rate of change of concentrations of the chemical species involved in the network [48]. The right-hand side in these ODEs is typically the algebraic sum of the rate laws of the reactions that produce or consume the chemical species (positive when it is produced, negative when consumed). There is formally no difference between a biochemical network and an enzyme reaction mechanism, as both conform to this description. For systems biology studies, it is sufficient to represent each enzyme-catalyzed reaction as a single step and associate it with an appropriate integrated rate law [49]. The systems biologist should be cautioned, though, that mechanistic details may indeed affect the dynamics, as is the case with competitive versus uncompetitive inhibitor drugs [50-52].

The Systems Biology Markup Language (SBML) [53] is a standard format to encode the information required to express a biochemical network model including its kinetics. SBML is based on the Extended Markup Language (XML), which is itself a standard widely adopted on the Internet. After a series of progressive developments, there are now several compatible software packages available to model biochemical networks. Some are generic and provide many algorithms, while others are more specialized. This includes not only simulators [54], but also packages for graphical depiction and analysis of networks [43,55,56], and databases of reactions and kinetic parameters [57], to name but a few examples. In some cases these packages can even work in an integrated way, such as in the Systems Biology Workbench (SBW) suite [58].

Another area of cell physiology where mathematical modeling has been used to describe complex molecular-scale dynamics is the study of ion channels. Molecules (both large and small) move back and forth across a cell membrane, to ensure that conditions for homeostasis are met [42]. Some molecules are small enough (and soluble to lipids) to diffuse across the membrane, while others require energy, working against electrochemical gradients between the outside and the inside of the cell [42]. For

example, differences in ionic potential across a cell membrane can drive ionic current. The Nernst equation,

$$V = \frac{RT}{zF} \ln\left(\frac{c_e}{c_i}\right), \quad (18)$$

describes the potential difference V across a cell membrane, where c_e and c_i are the external and internal ionic concentration (of a particular ion) respectively, R is the universal gas constant, T is the absolute temperature, F is Faraday's constant, and z is the ion's charge. To determine the ionic current across a membrane, one can write the total potential drop across the cell membrane, V_T , as the addition of the Nernst potential V to the potential drop due to an electrical current rI_c (where r is the resistance), so that $V_T = V + rI_c$. Solving for the ionic current we arrive at

$$I_c = g(V_T - V), \quad (19)$$

where $g = 1/r$ is the membrane conductance. Ions only travel through small pores, referred to as ion channels, and the control of such ionic current is vital for proper cellular function.

A second model for determining the ionic flow across a cell membrane is given by the Goldman–Hodgkin–Katz equation,

$$I = P \frac{z^2 F^2}{RT} V \frac{c_i - c_e \exp\left(\frac{-zFV}{RT}\right)}{1 - \exp\left(\frac{-zFV}{RT}\right)}, \quad (20)$$

where P is the permeability of the membrane to the ion [59]. Such an equation is derived under the assumption of a constant electric field. The use of either the linear equation (19) or nonlinear equation (20) in defining ionic current is often debated and depends on the underlying properties of the particular cell studied.

Using the fact that we can model the cell membrane as a capacitor (since it separates charge), and knowing that there is no net build-up of charge on either side of the membrane, the sum of the capacitive and ionic currents across a membrane should add up to zero,

$$C_m \frac{dV}{dt} + I_{\text{ion}}(V, t) = 0. \quad (21)$$

Here, C_m is the capacitance, I_{ion} is the ionic current, and $V = V_i - V_e$. In 1952, through a series of pioneering papers, Hodgkin and Huxley presented experimental data and a comprehensive theoretical model that fit experimental findings, to describe the action potential across the giant axon of a squid [60-64]. The model given by equation (22) (and based on equation (21)), was awarded the Nobel Prize in Physiology and Medicine in 1963, and is possibly one of the greatest mathematical results in physiology to date:

$$C_m \frac{dV}{dt} = -g_{\text{Na}}(V - V_{\text{Na}}) - g_{\text{K}}(V - V_{\text{K}}) - g_{\text{L}}(V - V_{\text{L}}) + I_{\text{app}}. \quad (22)$$

In equation (22), g_{Na} , g_{K} , and g_{L} describe the sodium, potassium, and leakage conductance (other ionic contributions, including the chloride current, that are small), respectively, V_{Na} , V_{K} , and V_{L} are their respective resting potentials, and I_{app} is a small applied current. Hodgkin and Huxley were able to measure the individual ionic

currents, and to use this information to determine the functional forms for each of the conductances.

Much of the work completed on ion channels and cell excitability has been used to study diseases that are associated with malfunction of ion channels. As a result, such channels have become new targets for drug discovery [47]. One example of a disease caused by the disruption of the action potential of cardiac myocytes is cardiac arrhythmia. Certain drugs used in the treatment of arrhythmias, such as lidocaine and flecainide, are sodium channel blockers, and so interfere with open sodium channels. Although these drugs have been used in treatment for cardiac arrhythmias, their exact mode of action is not well understood. Current computational models are being developed to understand the function of these, as well as other anti-arrhythmia drugs [65]. Another example of a disease caused by the disruption of ion channels is cystic fibrosis (CF), which has been found to be associated with malfunctions in chloride channel operation [66]. Although there is still no cure for CF, new directions for treatment protocols are being developed [67].

Models of cancer growth and spread: avascular tumor growth, tumor-induced angiogenesis, and tumor invasion

The primary mathematical modeling techniques used in the area of cancer modeling, include the study of avascular tumor growth, angiogenesis, and vascular tumor growth (tumor invasion), and so in what follows we limit ourselves to these topics.

Growth and development of a solid tumor occurs in two stages: avascular growth and vascular growth (leading to tumor invasion), and has been studied extensively in a biological framework [68]. The modeling of avascular tumor growth is one of the earliest approaches taken by mathematicians to study cancers [69]. The simplest type of model that can be used to describe how cancer cells of a solid tumor change over time is the exponential growth law, given by equation (23).

$$\frac{dN}{dt} = rN \quad (23)$$

Such an equation is limited in its application, since it suggests that tumors grow (with growth rate r) to a tumor of unbounded size. Other models, such as the logistic growth equation (24)

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{k}\right) \quad (24)$$

have been used to describe tumor size saturation (a type of volume constraint). In particular, equation (24) describes the rate of change of cancer cells N in a tumor, where r is the growth rate and k is the carrying capacity (the maximum number of cancer cells within the tumor).

A limiting case of the logistic equation is the Gompertz model, given by equation (25).

$$\frac{dN}{dt} = r \log\left(\frac{k}{N}\right)N \quad (25)$$

This model is one of the most commonly used tumor growth models to date, and was first used by Casey to fit real tumor growth in 1934 [70]. Such ODE-based models

are useful because they can be easily analyzed. Also, these models have the advantage that they can be expanded to incorporate other cell types (such as proliferating, quiescent, and dead cells) by the inclusion of more ODEs, one for each cell type. Since these models are similar to the compartmental SIR model described in the section concerning epidemiology, they have similar limitations, one being a lack of spatial information (such as the location of each cell type within a tumor). Some models for solid tumor growth have included such information by incorporation of reaction-diffusion type equations, like equation (26). This latter model, studied by Murray [26], was developed to describe the spatio-temporal invasion of gliomas. Equation (26) is read as follows: cancer cells C grow exponentially with rate ρ , and diffuse at a rate that depends on whether cells are moving in white brain matter or gray brain matter (diffusion given by $D(x)$).

$$\frac{dC}{dt} = \nabla \cdot (D(x)\nabla C) + \rho C \quad (26)$$

Other models by Cruywagen *et al.* [71] and Tracqui *et al.* [72] have used a more realistic logistic growth term for cancer cell growth. However, the survival times calculated for these models are only slightly different from those calculated using model (26), and so using either logistic growth or exponential growth is appropriate in this modelling framework. One limitation of the Cruywagen *et al.* [71] and Tracqui *et al.* [72] is that they are constructed under the assumption of constant diffusion, thus neglecting to distinguish between grey and white matter in the brain. Simulation of models similar to (26), such as those completed by Swanson *et al.* [73], show that incorporation of a spatially dependent diffusion coefficient $D(x)$ produces images that are in good agreement with real images of gliomas produced by magnetic resonance imaging (MRI).

Other spatial models, like that of equation (27), describe the movement of certain chemical signals, such as growth factors (GFs), which dictate directions for tumor growth [74,75].

$$\frac{dC}{dt} = D\nabla^2 C - \gamma C + \sigma S(r) \quad (27)$$

In equation (27), the first term on the right-hand side describes the diffusion of the concentration of GFs denoted by C , while the second term describes the degradation of C with rate γ . The third term describes production of C at a rate σ by a source $S(r)$. The source is a function of the tumor radius r . Some models incorporate integro-differential equations that describe the radius of tumors over time [75], providing a type of spatial information for each cell type within the tumor. For example, the nutrient concentration often dictates where each type of cell is most likely located within the tumor. Thus, in a radially-symmetric cell, there are typically zero nutrients located at the center of the tumor, and so cells there are dead, defining a necrotic core, whereas the outermost layer is generally nutrient rich, and so it is composed of proliferating cells. Other models, such as those incorporating systems of PDEs, are used to describe the spatial movement over time of proliferating and quiescent cancer cells in the presence of GFs [76].

Typically, solid tumors initially grow to about 2 mm in size. To grow larger, tumors require nutrients. Such nutrients are additionally acquired through the process of angiogenesis, the formation of blood vessels that connect the tumor to the circulatory

system [68]. The switch to this angiogenic stage can occur due to multiple factors, many of which are outlined in Semenza [77]. Tumor-induced angiogenesis, caused by the release of GFs or tumor angiogenic factors (TAFs) from the tumor, promotes the growth of endothelial cells (EC), which make up the linings of capillaries and other vessels. ECs then migrate towards the TAF source by chemotaxis. In other words, blood vessels grow towards the tumor.

The first model (given by equation (28)) to describe chemotaxis in a biological system using PDEs was developed by Keller and Segel [78]. Such an equation was initially developed to describe the chemotactic movement of slime molds towards an attracting source called cyclic adenosine monophosphate (cAMP):

$$\frac{du}{dt} = \nabla D \nabla u - \chi u \nabla v \quad (28)$$

In the context of cancer cells, the first term on the right-hand side of equation (28) describes random motion (diffusion) D of a cell population u , while the second term corresponds to chemotaxis of the cells u towards a GF. The chemotactic effect is given by χ and v is the concentration of the GF. The dynamics of the growth factor is typically modeled by reaction-diffusion equations similar to that given above in equation (27).

Many other models, typically PDE-type models, have extensions similar to those described above, incorporating not only the key interactions of the endothelial cells with angiogenic factors, but also the macromolecules of the extracellular matrix [79,80]. For an excellent review on mathematical approaches to studying tumor-induced angiogenesis we refer the reader to Mantzaris, Webb, and Othmer [81].

After a tumor has progressed to the stage of angiogenesis, by successfully recruiting blood vessels, the tumor eventually becomes vascular, thus connecting itself to the circulatory system [68]. Once this happens, tumor cells have the ability to enter the circulatory system, depositing themselves at secondary locations within the body and possibly promoting the growth of a secondary tumor, a process referred to as metastasis [68]. As soon as this occurs the cancer has progressed to the point where it is nearly impossible to cure the patient. One of the key steps in the progression to metastasis is the degradation of the extracellular matrix (ECM) by matrix-degrading proteins deposited by cancer cells, and the movement of cancer cells through the ECM. Newer mathematical approaches, using integro-PDEs, have been able to capture the qualitative movement of such cells through the ECM [82-84]. One such model was proposed by Hillen [82].

$$\frac{dp(\mathbf{x}, t, v)}{dt} + v \cdot \nabla p(\mathbf{x}, t, v) = -\mu \int_V \frac{q(\mathbf{x}, t, \tilde{v})}{w} p(\mathbf{x}, t, \tilde{v}) d\tilde{v} \quad (29)$$

Here, $p(\mathbf{x}, t, v)$ describes the density of tumor cells at location \mathbf{x} , time $t > 0$, and velocity v in $V = [v_{\min}, v_{\max}]$. The advection term on the left-hand side of equation (29) describes directed movement of cells along collagen fibers in the ECM with speed v . The right-hand side describes changes in the cell's velocity due to changes in the matrix orientation $q(\mathbf{x}, t, v)$, where w is an appropriate weighting function. This matrix can change over time due to cuts made by matrix degrading proteins. For ease in reading we do not give the evolution equation for the matrix distribution q .

Many different cancer treatment modalities are available, including the administration of chemotherapeutic drugs and their combinations [85] and radiation treatment [86]. Also, as is often the case with avascular tumors, the solid tumor may be surgically removed, if discovered early enough. With the advancement in imaging techniques, much work has been done to extend earlier modelling of glioma to better model tumor growth as it applies to treatment (removal and radiation). In particular, better models for predicting survival rates have been developed [87], as well as models that predict efficacy of radiotherapy for individual patients [88]. Typically, MRI is used to detect tumors, and provides the imaging information needed for validating mathematical models. Other imaging techniques, such as Diffusion Tensor Imaging (DTI) (a technique which measures the anisotropic diffusion of water molecules in a tissue), have been used to better establish diffusion parameters required for models to predict the appropriate boundary of a tumor. Such information can be used to describe the appropriate boundary required for either surgical removal of a tumor or radiation treatment [89,90].

Even though many treatment protocols exist, the probability of individuals surviving advanced stages of cancer is very low. As a result, many questions still remain as to how chemotherapeutic drugs work at a molecular level (e.g., which proteins they target by design and which by accident), and how radiation treatments should be delivered in order to maximize the death of cancer cells, while minimizing the harm to normal healthy tissue. Mathematical and computational modeling plays an important role in understanding radiation treatment protocols. For example, many models have been developed to define a tumor control probability (TCP) [91], where TCP is defined as the probability that no clonogenic cells survive radiation treatment [92]. TCP models have been incorporated into existing cancer models that describe cancer growth without treatment, such as those described above, to better understand the effects of radiation treatment on normal tumor dynamics [93].

The models discussed above describe the spatial and temporal changes of certain quantities that are of interest in various biological systems. In particular, the differential equations described above give temporally dependent solutions (and spatially dependent solutions in the case of the PDEs described) for various quantities, including the total populations of individuals, the total number/density of cells, or the total molecular concentration of a certain compound. Many of the successes and limitations of a differential equation modeling approach are highlighted above. One limitation, not highlighted in the above sections, is that such methods (those that use only a handful of differential equations) are not appropriate for describing the smaller scale movements of molecules. The movement and structure of an individual molecule is based on the many complex interactions between the individual atoms within a molecule, as well as its interactions with surrounding molecules. In order to follow the motions of every atom and molecule over extremely small timescales, computational methods such as molecular dynamic simulations (designed to solve extremely large systems of differential equations over very small timescales) can be applied. This technique is described in the following section.

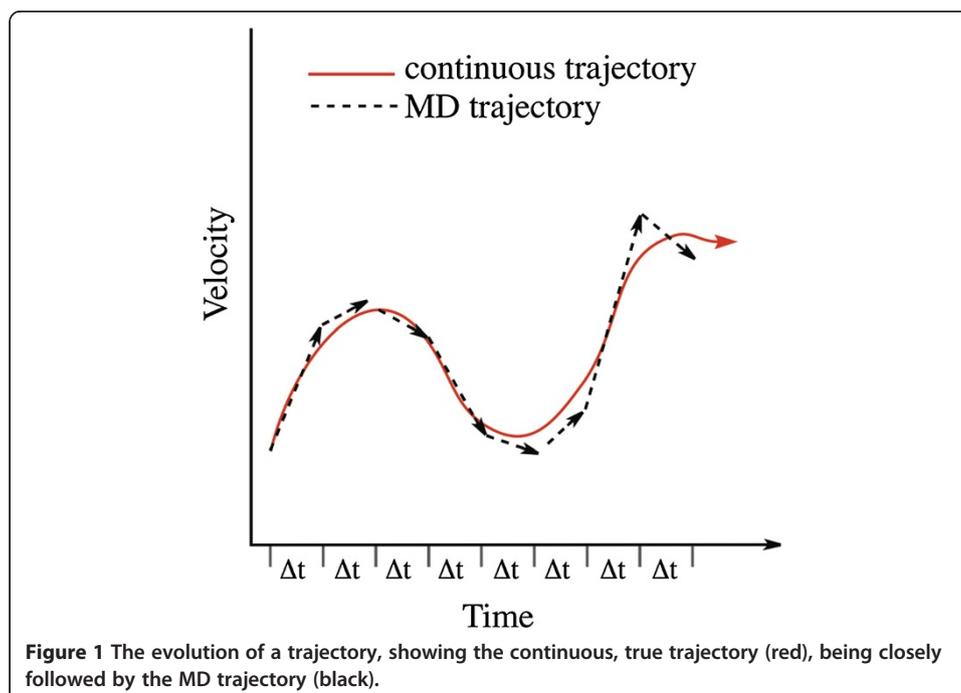
Molecular dynamics

The sophistication of the model used to study a given system depends on the property of interest. Often, a 3D model of a molecule or complex that shows the spatial relationships

between atoms is the best way to understand a system. Such computational models provide a means of observing the structure and motion of individual atoms within complex biomolecular systems. Although a physical model of a small molecule with less than 20 atoms can be easily made from plastic or wire in a few minutes, a similar model of a protein or an enzyme involves hundreds or thousands of atoms. Over the last decade improvements in a combination of computer graphics programs, and molecular modeling techniques and hardware have resulted in an unprecedented power to create and manipulate 3D models of molecules.

Molecular dynamics (MD) simulations follow the motions of atoms and molecules, and provide a means of investigating biological problems at a molecular level. This is achieved by solving Newton's equations of motion (EOM) for interacting atoms and evolving a system through time and space. Changes in atomic positions and velocities over time, usually ranging from nano- to milliseconds, result in a trajectory. In simple cases with few atoms, analytic solutions to Newton's EOM can be obtained, giving a trajectory that is a continuous function of time (Figure 1). However, in a computer simulation with many atoms, the EOM are solved numerically. Forces are evaluated for discrete intervals, or time steps (Δt , Figure 1) on the order of femtoseconds, where the forces are considered constant over a given time step. The goal is to follow the continuous function of time as closely as possible, which requires small time steps to ensure the motions of all atoms are resolved.

The forces on each atom are derived from the potential energy of the system, which can be described with quantum or classical mechanics. Since quantal descriptions are generally limited to small systems, classical descriptions are commonly used when studying biological systems and will be discussed in this section. It is worth noting that MD is a deterministic approach for exploring the potential energy surface of a system, while a stochastic approach can be obtained using Monte Carlo methods.



MD trajectories are analyzed to obtain information about the system, including structural changes as measured by atomic root-mean-square deviation (RMSD), non-covalent interactions, binding free energies [94], structural stability, short-lived reaction intermediates [95], conformational changes, flexibility, ligand binding modes [96], as well as ionic conductivity and mobility [97]. Numerous and diverse applications include the investigation of clinically important proteins such as HIV-1 gp120 [98], protein binding sites [99], drug resistance mechanisms of HIV-1 protease [100], protein folding [101,102] and the role of crystal water molecules in ligand-protein binding [103].

Force fields for protein simulations

Newton's second law, $m\ddot{x} = -\nabla V$, establishes the relation between mass (m) and acceleration (\ddot{x}), as well as force ($-\nabla V$), which is the negative of the gradient of the potential energy function (V). During an MD simulation, the forces acting on each atom of the system are calculated and atoms are moved according to those forces. In a classical MD simulation, the potential is calculated with a force field. The potential terms in a force field are a sum of contributions due to covalently-bonded interactions (V_{bond} , V_{angle} , V_{torsion}) and non-bonded interactions. Non-bonded interactions are calculated pairwise between two atoms, denoted i and j , and commonly include van der Waals ($V_{ij,\text{vdW}}$) and electrostatic ($V_{ij,\text{electrostatic}}$) contributions.

$$V = \sum_{\text{bonds}} V_{\text{bond}} + \sum_{\text{angles}} V_{\text{angle}} + \sum_{\text{torsions}} V_{\text{torsion}} + \sum_{i<j} V_{ij,\text{vdW}} + \sum_{i<j} V_{ij,\text{electrostatic}} \quad (30)$$

Due to the pairwise calculation of non-bonded interactions, such force fields scale as N^2 , where N is the number of atoms in the system. Each potential term contains parameters, where fitting to experiment is required to calculate these interactions [104]. Parameter fitting is done to reproduce the behavior of real molecules. This includes determining the van der Waals' radii, partial charges on atoms, bond lengths, bond angles and force constants. These parameters, along with the functional form of each potential term, collectively define a force field. Today, several types of force fields are available: (a) all-atoms force fields (parameters are considered for every atom), (b) united-atoms force fields (aliphatic hydrogen atoms are represented implicitly) and (c) coarse-grained force fields (groups of atoms are treated as super atoms). For a list of all force fields discussed, see Table 1.

Most all-atom force fields for proteins use relatively simple functions for modeling the potential energy surface [105], which correspond to the terms in Equation (30):

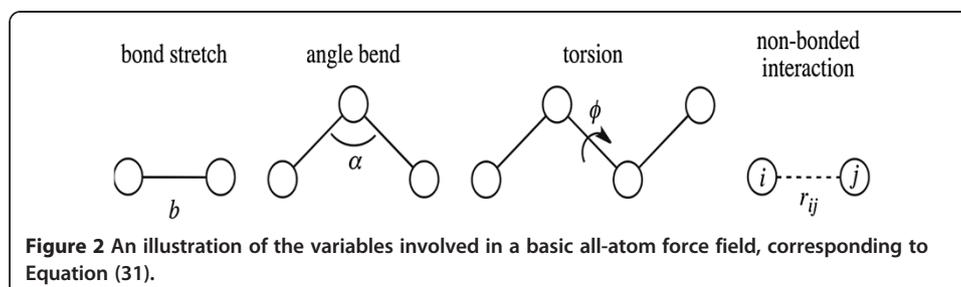
$$V = \sum_{\text{bonds}} K_b(b-b_0)^2 + \sum_{\text{angles}} K_\alpha(\alpha-\alpha_0)^2 + \sum_{\text{torsions}} K_\phi[\cos(n\phi + \phi_0) + 1] \\ + \sum_{i<j}^N \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + \sum_{i<j}^N \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (31)$$

The variables are indicated in Figure 2. The first term includes the stretching of bonds, where the potential is calculated using parameters K_b (related to force constant) and b_0 (related to equilibrium bond length), as well as the variable, b , which is the distance between bonded atoms. Similarly, the second term includes angle bending, where parameters K_α and α_0 , represent force constants and equilibrium angles, and the variable (α) is the angle obtained from the structure of interest. The third term, representing the

Table 1 List of force fields

Name	Family	Type	Reference
ff94	Amber	All-atoms	[104]
ff99	Amber	All-atoms	[107]
ff99SB	Amber	All-atoms	[106]
ff03	Amber	All-atoms	[112,113]
ff03.r1	Amber	All-atoms	
ff03ua	Amber	United-atoms	[114]
ff12SB	Amber	All-atoms	[110]
ff14SB	Amber	All-atoms	
GAFF	Amber	All-atoms	[115]
CHARMM	CHARMM	United-atoms	[117]
CHARMM19	CHARMM	United-atoms	[118]
CHARMM22	CHARMM	United-atoms	[119]
CHARMM22/CMAP	CHARMM	United-atoms	[120]
CHARMM27	CHARMM	United-atoms	[116]
CGenFF	CHARMM	United-atoms	[121]
OPLS-UA	OPLS	United-atoms	[117]
OPLS-AA	OPLS	All-atoms	[122]
GROMOS (A-version)	GROMOS	United-atoms	[125]
GROMOS (B-version)	GROMOS	United-atoms	[125]
43A1	GROMOS	United-atoms	[126]
45A3	GROMOS	United-atoms	[127]
53A5	GROMOS	United-atoms	[128]
53A6	GROMOS	United-atoms	[128]
54A7	GROMOS	United-atoms	[129,130]
54B7	GROMOS	United-atoms	[129,130]
54A8	GROMOS	United-atoms	[129,130]
MARTINI	MARTINI	Coarse-grained	[131-134]

potential from dihedral angles, involves parameters K_ϕ (barrier for rotation), n (number of maxima) and ϕ_0 (angular offset), as well as the variable ϕ , obtained from dihedral angles in the structure. The final two terms in Equation (31) represent non-bonded interactions between atoms i and j , which are summed over N atoms of the system. In this potential function, the van der Waals term is represented by the Lennard-Jones potential, where A_{ij} and B_{ij} are atom specific parameters related to atom size, and r_{ij} is a variable representing the distance between atoms i and j . The electrostatic term in Equation (31) is calculated



by a Coulomb potential, where parameters q_i and q_j are (fixed) charges on atoms i and j , respectively, and the constant ϵ_0 is the permittivity of free space. The force constants are obtained by empirical methods or by quantum mechanics calculations, depending on the force field. Here we present a summary of the most commonly used force fields in MD simulations, namely Amber, CHARMM, OPLS, GROMOS and MARTINI.

Amber force fields

The functional form from which most of the Amber (Assisted Model Building with Energy Refinement) force fields come uses Equation (31) and was developed by Cornell and co-workers (denoted ff94) [104]. However, various revisions have since been developed that vary in their parameterization, with the goal of improving results. One notable revision of Amber for proteins and nucleic acids is ff99SB [106], which was developed at Stony Brook University as a modification of the old ff99 force field [107] and improves on ff99 in its description of the ϕ and ψ dihedral angles of the protein backbone, resulting in a better balance between secondary structures, and improved treatment of glycine [108,109]. The ff12SB revision [110] reparameterizes backbone torsion angles, side chain torsions in select amino acids, and incorporates improved backbone torsions in DNA and RNA, with recent studies finding ff12SB performs better than ff99SB [111]. The current and most recent version, ff14SB, is recommended by Amber developers and minimizes dependencies of protein side chain conformations on backbone conformations by including side-chain corrections, and improves upon dihedrals in DNA and RNA, particularly χ . Another extensively-used Amber force field is ff03 [112,113] (the latest version is ff03.r1), which improves upon the charges calculated by ff99 by using charges derived from quantum calculations with a continuum dielectric to simulate the solvent polarization. A united-atom version of ff03, ff03ua, is also available [114]. Expanding the utility of the Amber force fields beyond peptides and nucleotides is the general Amber force field (GAFF) [115], which includes a complete set of parameters for a large number of small molecules while still remaining fully compatible with the other versions of the Amber force fields discussed above. This allows for Amber and GAFF to be combined to examine protein-ligand complexes, as well as modified proteins, DNA or RNA.

CHARMM force fields

The CHARMM (Chemistry at Harvard Macromolecular Mechanics) force fields are also a prominent set of force fields for studying biological systems. The functional form of the force field used by CHARMM is based on Equation (31), but also includes additional terms to treat improper torsions and atoms separated by two bonds (Urey-Bradley term) [116]. The CHARMM force fields use classical (empirical or semi-empirical) and quantum mechanical (semi-empirical or *ab initio*) energy functions for different types of molecular systems. They include parameters for proteins, nucleic acids, lipids and carbohydrates, allowing simulations on many common biomolecules. The initial version of the CHARMM force field was developed in the early 1980s, and used an atom force field with no explicit hydrogens [117]. However, in 1985, CHARMM19 parameters were developed in which hydrogen atoms bonded to nitrogen and oxygen were explicitly represented; hydrogens bonded to carbon or sulfur were still treated as extended atoms [118]. CHARMM19 parameters aimed to provide a balanced interaction between solute-water

and water-water energies. Although this force field was tested primarily on gas-phase simulations, it is now used for peptide and protein simulation with implicit solvent models. Newer versions of CHARMM, such as CHARMM22, include atomic partial charges that are derived from quantum chemical calculations of the interactions between model compounds and water [119]. Although CHARMM22 is parameterized for the TIP3P explicit water model, it is frequently used with implicit solvents. CHARMM27 parameters were developed for nucleic acids (RNA, DNA) and lipid simulations [116]. Since both CHARMM27 and CHARMM22 are compatible, it is recommended CHARMM27 be used for DNA, RNA and lipids, while CHARMM22 should be applied to protein components [116]. A more recent, dihedral-corrected version of CHARMM22 was developed, denoted CHARMM22/CMAP, which improves the parameters describing the ϕ and ψ dihedral angles of the protein backbone [120]. A general version of the CHARMM force field (CGenFF) also exists which allows to the treatment of drug-like while maintaining compatibility with other the CHARMM force fields [121].

OPLS force fields

For the OPLS (Optimized Potentials for Liquid Simulations) family of force fields the form of the potential energy function differs from Equation (28) in the dihedral term, and was parameterized by Jorgensen and co-workers [117]. OPLS force fields were parameterized simulate the properties of the liquid states of water and organic liquids [118]. For proteins, a united-atom version was first developed (OPLS-UA), followed by an all-atoms version (OPLS-AA) [122]. Charges and van der Waals terms were extracted from liquid simulations. The OPLS-AA force field uses the same parameters as the Amber force fields for bond stretching and angles. The torsional parameters were obtained by using data from *ab initio* molecular orbital calculations for 50 organic molecules and ions [123]. Several improvements and re-parameterizations were proposed in later years for this set of force fields [122,124], including for simulations of phospholipid molecules [125].

GROMOS force fields

The GROMOS (Groningen Molecular Simulation) force fields are united-atom force fields that that were developed in conjunction with the software package of the same name to facilitate research efforts in the field of biomolecular simulation in a university environment [125]. Its functional form varies from Equation (31) in the dihedral term, and differs from other force fields in its goal of reproducing enthalpies of hydration and solvation. The initial GROMOS force field (A-version) was developed for applications to aqueous or apolar solutions of proteins, nucleotides and sugars. However, a gas phase version (B-version) for the simulation of isolated molecules is also available [125]. Important versions of the GROMOS force fields include GROMOS 43A1 [126] (improved treatment of lipid bilayers), GROMOS 45A3 [127] (relevant to lipid membranes and micelles) and GROMOS 53A5 and 53A6 [128] (recommended for the simulation of biomolecules in explicit water). Recent releases such as GROMOS 54A7 and 54B7 involve modifications to ϕ and ψ protein dihedral angles, new lipid atoms types, new ion parameters and additional improper dihedral types, while GROMOS 54A8 accurately models the structural of lipid bilayers, proteins and electrolyte solutions [129,130].

MARTINI force fields

The semi-empirical MARTINI force field is the most commonly used coarse-grained force field for biomolecular system simulations; it was originally developed for lipid simulations [131]. In this potential energy surface, four heavy atoms in a molecule are considered as a single interaction site, and only four types of interaction have been considered, namely polar, non polar, apolar and charged; moreover, each particle has different subtypes for taking in account the underlying atomic structure; a special particle types has been introduced for the ring conformations, as the four-to-one mapping is not appropriate to represent small ring molecules. Initially developed for lipid simulations, this force field includes now extensions for the parameterization of proteins [132], carbohydrates [133] and more recently glycolipids [134]. With the approximations introduced by the coarse grain approach it is possible to increase both the time-scale and the size of systems compared with all-atoms or united-atoms force field simulations. The list of fields where MARTINI was employed includes characterization of lipid membranes [135], protein-protein interactions [136], self assembly of peptides and proteins [137] and interactions between nanoparticles and biological molecules, as for example the study of the mechanism by which fullerene can penetrate the cellular lipid membrane performed by Wong-Ekkabut et al. [138] and other applications [139]. However, MARTINI cannot be used for protein-folding studies [139], as the secondary structure is a required input parameter.

Calculation of solvation free energies

The calculation of solvation free energies is a challenging problems in MD simulations. Determining solvation free energy is especially difficult in aqueous bio-systems due to the size of the system [140]. Solvation free energy, ΔG_{solv} , is a thermodynamic property defined as the net energy change upon transferring a molecule from the gas phase into a solvent with which it equilibrates [141]. Solvation effects can change the physical and chemical properties of biomolecules including charge distribution, geometry, vibrational frequencies, electronic transition energies, NMR constants and chemical reactivity [142].

Several methods have been developed for modeling solvation and one can select the most advantageous choice among them based on the required accuracy and computational cost. To simulate effects of solvent on biomolecules, one can use explicit or implicit solvent models. While explicit solvent models include solvent molecules in the system, implicit models use a mean field approach [143,144]. Although explicit solvent simulations are computationally expensive because of the enormous numbers of atoms involved, they provide a more realistic picture of solute-solvent interactions, reflecting the molecular complexity of the biomolecule and its environment. In comparison, implicit solvent models increase the speed of the simulation since the Newtonian equations of motion are not solved for additional solvent molecules. Table 2 lists the solvation models discussed.

Implicit water models

The simplest approach to solvation is to treat the effects (e.g. electrostatic interactions, cavitation, dispersion attraction and exchange repulsion) of solvent on the solute with an implicit model. These methods represent the solvent as a continuum environment, where the quality of the results is most affected by the electrostatic and cavitation (the

Table 2 List of solvation models

Name	Type	Reference
Poisson–Boltzmann (PB)	Implicit	[145,146]
Generalized Born (GB)	Implicit	[140,141,147,149-151]
3D-RISM	RISM	[140,152-158]
3D-RISM-KH	RISM	[156,164]
MTS-MD/OIN/ASFE/3D-RISM-KH	RISM	[149]
3D-RISM-KH-NgB	RISM	[168]
SPC	Explicit	[174]
SPC/E	Explicit	[174]
POL3	Explicit	[175]
TIP3P	Explicit	[176]
TIP3P/F (TIP3P-PME/LRC)	Explicit	[177]
TIP4P	Explicit	[176,178]
TIP4P/Ew	Explicit	[179]
TIP5P	Explicit	[180]

size and shape of a cavity that the solute occupies) contributions. The solvation free energy of a molecule, ΔG_{solv} , can be divided into two parts: electrostatic (ΔG_{el}) and non-electrostatic (ΔG_{nonel}). The electrostatic energy is defined as the free energy required to remove all the charges in vacuum and add them back to the solute in the presence of continuum solvent [140,141]. The origin of the non-electrostatic energy is a combination of favorable solute-solvent van der Waals interactions and the unfavorable disruption of the water structure by solute molecules (cavitation), and corresponds to solvating the neutral solute. There are several different implicit solvent models discussed below: the Poisson–Boltzmann model and the generalized Born model, which differ in how ΔG_{el} is obtained.

Poisson–Boltzmann model Solving Poisson’s equation, which is valid under conditions where ions are absent, gives a second order differential equation describing the electrostatic environment that is modeled with a dielectric continuum model [145],

$$\nabla \cdot \epsilon(r) \nabla \phi(r) = -4\pi\rho(r), \quad (32)$$

where $\phi(r)$ is the electrostatic potential, $\epsilon(r)$ is the dielectric constant and $\rho(r)$ is the charge density.

Poisson’s equation cannot be solved analytically for most systems, and must be solved using computers and adopting numerical methods. The Boltzmann contribution, along with the assumptions of the Debye–Hückel theory, describes the charge density due to ions in solution. This results in the (non-linearized) Poisson–Boltzmann (PB) equation [146]:

$$\nabla \cdot \epsilon(r) \nabla \phi(r) - \frac{\epsilon_s k T}{e} S(r) \kappa^2 \sinh[e\phi(r)/kT] = -4\pi\rho(r), \quad (33)$$

where κ denotes the Debye–Hückel parameter, ϵ_s is the solvent dielectric constant, $S(r)$ is a “masking” function with value 1 in the region accessible to the ions in the solvent and value 0 elsewhere; e is the protonic charge; k is Boltzmann’s constant; T is the absolute temperature. Here, the charge density on the right represents the partial charges

in the cavity. When the ionic strength of the solution or the potential is low, Equation (33) can be linearized by expanding the second term on the left into a Taylor series and retaining only the first term:

$$\nabla \cdot \varepsilon(r) \nabla \phi(r) - \varepsilon_s \kappa^2 S(r) \phi(r) = -4\pi\rho(r), \quad (34)$$

The non-electrostatic contribution ($\Delta G_{\text{nonelect}}$) to the solvation free energy is calculated by empirical methods and is proportional to the solvent accessible surface area. This is added to the electrostatic part to yield the solvation free energy. Although the PB approach is mathematically rigorous, it is computationally expensive to calculate without approximations [141,147,148]. The generalized Born model provides a more efficient means of including solvent in biomolecular simulations.

Generalized born model The generalized Born (GB) model is based on the Born approximation of point charges, modeling solute atoms as charged spheres with an internal dielectric (generally equal to 1) that differs from the solvent (external) dielectric. The polarization effects of the solvent are represented by a dielectric continuum represents the polarization effects of the solvent. Numerical methods are used to determine the charges on the solute spheres that result in the same electrostatic potential on the cavity surface that mimics that of the solute in a vacuum.

By making approximations to the linear Poisson-Boltzmann equation (Equation 31), the electrostatic contribution of the generalized Born model is obtained:

$$\Delta G_{\text{el}} = -\frac{1}{2} \left(\frac{1}{\varepsilon_{\text{int}}} - \frac{1}{\varepsilon_{\text{ext}}} \right) \sum_{ij} \frac{q_i q_j}{\left[r_{ij}^2 + \alpha_i \alpha_j \exp\left(-r_{ij}^2 / 4\alpha_i \alpha_j\right) \right]^{1/2}}, \quad (35)$$

where α_i is effective Born radius of particle i , r_{ij} is the distance between atoms i and j , ε_{int} and ε_{ext} the internal and external dielectric constants, respectively, and q_i is the electrostatic charge on particle i . Like the PB method, $\Delta G_{\text{nonelect}}$ is calculated from the solvent-accessible surface area [140,141,147,149-151].

Reference interaction site model

Another type of solvation model is a probabilistic method known as the 3D reference interaction site model (3D-RISM) [140,152-158]. This molecular theory of solvation simulates the solvent distributions rather than the individual solvent molecules. However, the solvation structure and the associated thermodynamics are obtained from the first principles of statistical mechanics.

In this method, the 3D site density distributions of the solvent are obtained, which accounts for different chemical properties of the solvent and solute. These properties include hydrogen bonding, hydrophobic forces and solvation thermodynamics, such as the partial molar compressibility and volume. In addition, the solvation free energy potential and its energetic and entropic components can be calculated. The solvation free energy is calculated from the RISM equation as well as the closure relation [159-163].

Several additional advances have been made in formulating improved versions of the 3D-RISM theory including the hypernetted chain (HNC) closure approximation [152,153]. Another derivation came from the molecular Ornstein–Zernike integral equation [163] for the solute-solvent correlation functions [155,156,164]. Sometimes

the calculated solvation free energy for ionic and polar macromolecules involves large errors due to the loss of long-range asymptotics of the correlation functions. Work has been done to account for the analytical corrections of the electrostatic long-range asymptotics for the 3D site direct correlation functions as well as the total correlation functions [157,158,164]. Other developments include the closure approximation, 3D-RISM-KH closure, for solid-liquid interfaces, fluid systems near structural and phase transitions, as well as poly-ionic macromolecules [156,164].

Two methods have been developed to couple 3D-RISM with MD. The first method makes use of a multiple time step (MTS) algorithm [165,166] wherein the 3D-RISM equations are solved for a snapshot of the solute conformation, then solved again after a few MD steps. This method is limited by the requirement to re-solve the 3D-RISM equations every few MD steps, which is computationally expensive for large biomolecular systems. The second of these methods involves the contraction of the solvent degrees of freedom and the extrapolation of the solvent-induced forces. These methods are aimed at speeding up the calculations, which is useful for larger systems [149].

Other work has involved the development of the multi-scale method of multiple time steps molecular dynamics (MTS-MD) in a method referred to as MTS-MD/OIN/ASFE/3D-RISM-KH [149]. Specifically, this method converges the 3D-RISM-KH equations at large outer time steps and uses advanced solvation force extrapolation to calculate the effective solvation forces acting on the biomolecule at inner time steps. The integration between the inner and outer time steps is stabilized by the optimized isokinetic Nosé–Hoover chain (OIN) ensemble, which enables an increase of the outer time step. Furthermore, effort was expended on MTS-MD aimed at converging the 3D-RISM-KH integral once every few OIN outer time steps, and the solvation forces in between were obtained by using solvation force-coordinate extrapolation (SFCE) in the subspace of previous 3D-RISM-KH solutions [167]. Another developed model is the 3D-RISM-KH-NgB [168]. In this model the non-polar component of the hydration free energy obtained from 3D-RISM-KH is corrected using a modified Ng bridge function [169]. Calibration of this model is based on the experimental hydration free energy values of a set of organic molecules.

Lastly, work has been done to improve the performance of 3D-RISM calculations by running them on graphical processing units (GPUs). To overcome memory issues, a modification of the Anderson Method [170] that accelerates convergence was introduced [171]. This method was reported to be eight times faster on an NVIDIA Tesla C2070 GPU as compared to the time taken on an eight-core Intel Xeon machine running at 3.33 GHz.

Although 3D-RISM calculations are more computationally expensive than GB and PB based solvation methods [172], they overcome some of the inherent shortcomings of these empirical methods [173].

Explicit water models

Explicit solvation is characterized by modeling individual water molecules around a solute. Several explicit water models are available in the Amber, NAMD and Gromacs MD simulation packages. These include simple explicit solvent models (SPC [174], SPC/E [174]), polarizable models (such as POL3 [175]), and fixed-charge explicit solvent models (TIP3P [176], TIP3P/F [177], TIP4P [176,178], TIP4P/Ew [179] and TIP5P [180]).

Examples of explicit water models are the simple point charge (SPC) model and the extended simple point charge (SPC/E) model [174]. In both of these models the water molecules are rigid. A derivative of SPC with flexible water molecules has been developed [181]. Another simple explicit model is the POL3 water model, which is a polarizable model [175].

More complex explicit water models include the transferable intermolecular potential n point (TIP n P), where n represents the number of interaction sites on each model. These are the most common classes of explicit solvent models in use [147]. In the case of TIP3P, the most simple TIP n P model, the interaction sites includes the oxygen and two hydrogen atoms [176]. A re-parameterized model of TIP3P is the TIP3P-PME/LRC, also referred to as TIP3P/F [177], which calculates electrostatic contributions by particle mesh Ewald (PME) summation and includes a long-range van der Waals correction (LRC). TIP4P [176,178] introduced a fourth dummy atom bonded to the oxygen to improve the electron distribution in the water molecule. This model has been re-parameterized for use with Ewald sums: TIP4P/Ew [176,178]. The five interaction points in the TIP5P [180] model include two dummy atoms near oxygen, which further improves the charge distribution around the water molecule.

Molecular dynamics methods

Examining the dynamics of a system at an atomic scale requires beginning with a model having atomic-level resolution. For biological macromolecules, this may be experimentally obtained from nuclear magnetic resonance (NMR) spectroscopy or X-ray crystallographic data. Although electron microscopy data does not provide structures with atomic resolution, this data can be combined with structural data from NMR or crystallography to obtain a high-resolution structure. NMR, crystallographic and electron microscopy structures of bio-macromolecules can be downloaded from the Protein Data Bank (<http://www.pdb.org>). In the absence of experimental data, homology modeling may be used to generate the 3D structure of a protein (the target) using its amino acid sequence and an experimentally-available 3D structure of a homologous protein (the template). Homology modeling can produce high-quality structures of a target protein if the sequence identity of the target and template is sufficiently high (typically > 40%).

As with any computational modeling, performing MD simulations requires a balance between accuracy and efficiency. Ideally, a system should be allowed to evolve in time indefinitely so that all states of a system may be sampled. However, this is not possible in practice. Several techniques have been developed so improve simulation efficiency, while still maintaining accuracy. During a simulation, costly force calculations are performed in discrete time increments, known as time steps, typically on the order of femtoseconds. Although it is necessary to have small time steps to properly resolve the motion of atoms, this results in many force calculations and therefore large computational costs. To address this, one can restrain the fastest vibrations, which involve hydrogen, giving the SHAKE algorithm [182]. This allows a larger time step to be used during simulations [183]. Additionally, the number of time-consuming non-bonded potentials calculated can be limited by using a cut-off based method [184]. Here, interactions are calculated between pairs of atoms within the cutoff distance, but neglected for atom pairs that are far apart. Although this is generally appropriate for the short-range

nature of van der Waals interactions, using cutoffs to calculate long-range Coulomb interactions leads to instabilities [185]. Alternatively, the particle mesh Ewald (PME) approach [185-188] can be used to calculate electrostatic potentials, which involves the calculation of the short range electrostatic component in real space and the long-range electrostatic component in Fourier space.

The simplest way to enhance sampling during an MD simulation is to increase the time duration of a simulation, which is usually on the order of nano- to milli-seconds. Researchers may also conduct multiple simulations that begin with the same initial structures [189], providing denser sampling of the conformational space by utilizing multiple trajectories. Advanced techniques also exist to increase the number of states of a system that are sampled during an MD run [190]. Enhanced sampling can be achieved by metadynamics [191], which allows for the high-energy regions between minima to be explored. This helps the system to escape local free energy minima and explore metastable states separated by large free energy barriers [192]. Metadynamics, which is used to calculate static properties, can also be used to calculate dynamic properties by introducing a history-dependent biasing potential as a function of a few collective variables. Selection of collective variables is an essential part of a metadynamics run as these variables help sample different energy basins. The use of metadynamics, which is a combination of ideas involving coarse-grained dynamics in space and the introduction of a history-dependent bias, can overcome the problem of limited time scale exploration by existing sampling algorithms and computational resources. Replica-exchange molecular dynamics (REMD) has also been preferred over standard MD to enhance sampling by allowing systems of similar potential energies to sample conformations at different temperatures [193,194]. This overcomes the energy barriers on potential energy surfaces and helps explore more conformational space. REMD can effectively sample energy landscapes, including both high- and low-energy structures, which is especially important in the case of protein folding and unfolding processes [195].

In addition to obtaining structural information about a system, molecular dynamics provides information about energetics, particularly binding. Given the inherent flexibility of these biomolecules over time, and throughout an MD simulation, properties such as the free energy are calculated as the time-average of an ensemble of snapshots obtained from MD trajectories [196-198]. Binding calculations have applications in drug design, protein-protein interactions and DNA stability. For example, the binding free energy of a ligand to a protein is calculated as the difference in free energy between the complex, and the receptor and ligand ($\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{receptor}} - G_{\text{ligand}}$) [199]. The free energy is calculated as a sum of the energy contributions from the force field, free energy of solvation and entropy ($G = E_{\text{MM}} + G_{\text{sol}} + T\Delta S$). Commonly in MD, the solvation free energy is obtained from an implicit solvent model (Poisson-Boltzmann, generalized Born) and the solvent accessible surface area (denoted MM-PBSA and MM-GBSA, respectively) [199]. Decomposition of the binding free energies provides a means of obtaining information about the residues that significantly contribute to the binding affinity of a ligand. Pairwise decomposition may also provide insight into changes in binding that result from mutations, especially single point mutations [200].

Molecular dynamics is increasingly being used to solve a host of problems [201-207]. The simulation of bio-macromolecules, especially in conjunction with solvent, is very computationally demanding. This demand is being met by the increasing power and

speed of modern computers, including special purpose computers such as Graphical Processing Units (GPUs) [208]. Novel methods have been developed to enhance the exploration of conformational space. The accuracy of force fields continues to improve with recent reparameterizations. A variety of solvation models exist, which can also be used to calculate binding free energies, which are immensely important in drug discovery applications, identifying enzyme inhibitors or compounds that block protein-protein interactions. Although classical MD simulations have been proven useful in studying a variety of biomolecular systems, they are limited in their application to stable structures. The examination of systems involving chemical reactions or quantum phenomena requires treatment using quantum mechanics methods, or the combination of quantum mechanics and molecular mechanics methods (denoted QM/MM). The application of quantum methods to biomolecules is discussed in the next section.

Quantum mechanics in biophysical modeling

Quantum mechanics (QM) calculations, being highly accurate and rigorous, are an essential tool in computational chemistry studies. Unfortunately, the prohibitive size of many biological systems has limited theoretical and computational studies of them to the realm of classical mechanics, largely utilizing non-polarizable force fields in MD simulations. This has necessarily reduced the scope of studies to conformational or structural aspects of these bio-systems, rather than more complex problems such as chemical reactions or quantum phenomena (excited states and charge-transfer, for example). Although hybrid quantum mechanics/molecular mechanics (QM/MM) approaches have increased accuracy, recent improvements in software, hardware and theory have allowed for full quantum mechanical studies of biochemical systems. The QM methods and functionals discussed are listed in Table 3.

Quantum mechanics methods

The wavefunction, which contains all the information describing a quantum system, is obtained by solving the Schrödinger equation, usually in its time-independent, non-relativistic form, invoking the Born–Oppenheimer approximation.

$$\hat{H}\psi\left(\zeta_1, \zeta_2, \dots, \zeta_N; \vec{R}_A, \vec{R}_B, \dots, \vec{R}_M\right) = E\psi\left(\zeta_1, \zeta_2, \dots, \zeta_N; \vec{R}_A, \vec{R}_B, \dots, \vec{R}_M\right), \quad (36)$$

This wavefunction, ψ , is a function of N electronic coordinates and depends parametrically on M nuclear coordinates. The coordinates of an electron are denoted $\zeta_N = (x_N, y_N, z_N, \sigma_N)$, where x_N, y_N, z_N are the spatial coordinates of electron N , and σ_N is the spin of this electron. The spatial coordinates of nucleus M are indicated by \vec{R}_M .

The Hartree–Fock (HF) method is the simplest *ab initio* QM method for obtaining the wavefunction, where the N -electron wavefunction is approximated as a product of N one-electron wavefunctions, expressed in a Slater determinant. However, the HF theory is insufficient in describing electron correlation. Electron correlation is included in post-Hartree–Fock *ab initio* methods such as the Møller–Plesset perturbation theory (MP2) or coupled-cluster (CC) calculations, although at increased computational cost. While HF scales formally as K^4 , where K is the number of basis functions, MP2 scales as K^5 , and the gold-standard CCSD(T) scales as K^7 . A detailed

Table 3 List of QM methods and functionals

Name	Type	Reference
Hartree-Fock (HF)	<i>ab initio</i> method	[209]
Møller-Plesset (MP2)	<i>ab initio</i> method	[209]
Coupled-cluster (CC)	<i>ab initio</i> method	[209]
CCSD(T)	<i>ab initio</i> method	[209]
CCSD(T)/CBS	<i>ab initio</i> method	[220]
MNDO	Semi-empirical method	[210-212]
AM1	Semi-empirical method	[210-212]
PM3	Semi-empirical method	[210-212]
PM6	Semi-empirical method	[210-212]
OMx	Semi-empirical method	[210-212]
B3LYP	DFT functional	[214]
PBE	DFT functional	[214]
TPSS	DFT functional	[214]
DFT-D2	DFT method	[215,218]
DFT-D3	DFT method	[215,218]
Effective fragment potential (EFP)	Fragmentation method	[241-243]
Fragment molecular orbital (FMO)	Fragmentation method	[244]
Elongation (ELG)	Fragmentation method	[245]
Divide and conquer (DC)	Fragmentation method	[246]

description of these methods, and many others, is available in the *Encyclopedia of Computational Chemistry* [209].

An alternative approach to including electron correlation is using semi-empirical quantum methods, where expensive two-electron integrals in the HF method are eliminated and the method is supplemented with parameters derived from experimental data. Although these methods are computationally very efficient, their applicability is generally limited to systems similar to the parameterization set. Many variants of semi-empirical methods exist, with applicability to biological systems [210-212], including MNDO, AM1, PM3, PM6, and the OMx methods.

The electron density, a physical observable, also determines all properties of a system, and it is this fact that is utilized in the density functional theory (DFT) [213].

$$\rho\left(\vec{r}_1\right) = N \int \left| \psi\left(\vec{r}_1, \vec{r}_2, \vec{r}_3, \dots, \vec{r}_N\right) \right|^2 dr_2 \dots dr_N \quad (37)$$

Since DFT methods scale as K^3 , they are more efficient than the HF method while also containing electron correlation. Therefore, DFT has been successful in the study of a variety of biological systems [214]. Many functionals exist (for example, B3LYP, PBE and TPSS [214]), which vary based on the form of the exchange-correlation functional, and the performance of each functional is highly dependent on the system and properties of interest [213]. Even with the formal inclusion of electron correlation, many DFT functionals are inadequate in modeling dispersion forces, a dominant source of stabilization in bio-macromolecules. A common approach for improving the performance of DFT functionals is to add a dispersion correction, which can be derived either

empirically or from high-level *ab initio* calculations [215-217]. For example, the DFT-D2 and DFT-D3 methods are popular [215,218]. It is worth noting that the dispersion terms have also been applied to improve semi-empirical methods [210,212,217,219].

The electronic structure methods mentioned above, particularly the *ab initio* methods, have been largely limited to studying small molecules or truncated systems of biological and medicinal/pharmaceutical interest. For example, extensive benchmarking studies have examined non-covalent interactions to identify efficient methods of comparable accuracy to CCSD(T)/CBS [220]. QM methods are also routinely used to develop MD force field parameters and charges using small models, as well as to parameterize docking scoring functions [219]. Additionally, ligand strain may be evaluated with QM methods [219]. Quantitative structure–activity relationship (QSAR) models also utilize QM methods to calculate the predictor variables such as electrostatic potentials, orbital energies, charges, and dipoles in small molecules [219,221]. The QM methods, specifically the semi-empirical and DFT approaches, were combined with truncated protein models to examine protein-ligand interactions [211]. Nevertheless, the use of QM methods to study very large systems remains limited. In order for the extensive applications of QM methods to macromolecular systems of biological and medicinal interest to be feasible, significant advances in software, hardware, and theory must be achieved [222].

Acceleration of quantum mechanics with graphical processing units

Graphics processing units (GPUs) have led to substantial accelerations in high-performance computing, with significant advancements for costly QM calculations. GPU-accelerated code has been developed for the HF method [223], correlated *ab initio* methods [223-225], semi-empirical methods [226,227], and DFT [228]. In fact, GPU acceleration is now implemented in many quantum chemistry programs. NVIDIA reports the following programs as having GPU support: ABINIT, BigDFT, CP2K, GAMESS-US, GAMESS-UK, GPAW, LATTE, MOLCAS, MOPAC2012, NWChem, OCTopus, PEtot, Q-Chem, QMCPACK, Quantum Espresso, TeraChem, and VASP [208]. However, the most extensive use of GPUs in a QM code [229-234] has been implemented in TeraChem, an electronic-structure package specifically designed for use with GPUs. This has enabled a QM description of a protein to examine charge-transfer and polarization in a solvated environment [235], full QM optimizations of protein structures [233], and the examination of excitations in the green fluorescent protein (GFP) chromophore [236].

Fragment-based quantum mechanics methods

One approach to making QM methods more tractable to biological and medicinal applications is the modification of methods so that they scale linearly with system size [237,238]. One such approach is based on fragmentation methods, which have been developed to facilitate the application of wavefunction and density functional methods to macromolecular structures. These methods partition a macromolecular system and perform QM calculations on each fragment to obtain their wavefunctions and properties, which are then combined to arrive at properties of the macromolecular system as a whole. Fragmentation methods also benefit from their ability to be massively parallelized. A comprehensive summary of the many fragmentation methods available, with many applications to biological systems, can be found in a recent review [239].

Of the electronic structure software available, the GAMESS program includes the greatest variety of fragmentation methods [239,240], which include the effective fragment potential (EFP) methods [241-243], the fragment molecular orbital (FMO) method [244], the elongation (ELG) method [245], and divide and conquer (DC) approaches [246]. Although the EFP methods utilize intermolecular potentials, these are distinct from those in force fields since they are rigorously derived from *ab initio* calculations rather than empirical parameters. Of these fragmentation methods, the FMO approach is arguably the most robust and has been widely applied to biological systems [239,240,247-249], which include drug discovery, protein-ligand binding, protein-protein interactions, enzymatic catalysis, and DNA.

Application of quantum mechanics/molecular mechanics to computational enzymology

Enzymology investigates, among other topics, enzyme kinetics and mechanisms of inhibition in steady-state turnover. Advances in technology and methods have led to more detailed information about enzyme structures and mechanisms. With an explosion in the number of novel and uncharacterized enzymes identified from the vast number of genome sequences, it has become evident that the structural and functional properties of these enzymes need to be elucidated to establish precisely their mechanisms of action and how the enzymes fit into the complex webs of metabolic reactions found in even the simplest of organisms [250].

Vast changes have occurred in the science of enzymology since molecular simulations and modeling were first developed. Calculations can provide detailed, atomic-level insights into the fundamental mechanisms of biological catalysis. Computational enzymology was launched in the 1970s [251]. The pioneering studies of Warshel are particularly notable [252,253]. By the early 1990s the number of computational mechanistic studies of enzymes was still relatively small [254,255], but recently there have been a great number of computational studies of enzymatic reaction mechanisms published [253,256-258]. Currently, computational enzymology is a rapidly developing area, focused on testing theories of catalysis, challenging "textbook" mechanisms, and identifying novel catalytic mechanisms [259].

The choice of an appropriate method for the particular enzyme being modeled is vital. Quantitative predictions of reaction rates or the effects of mutations remain very challenging, but with appropriate methods, useful predictions can be made with some confidence. Careful testing and experimental validation are important. For example, a comparison of calculated barriers for a series of alternative substrates with experimentally determined activation energies demonstrated good correlation validating mechanistic calculations [260,261]. Some enzymes have become important model systems in the development and testing of computational methods and protocols; these include chorismate mutase [259,262-266], citrate synthase [267-269], P450 [257,259], para-hydroxybenzoate hydroxylase [257,260,265] and triosephosphate isomerase [255,270,271].

Modeling enzyme-catalyzed reactions

The usual starting point for modeling an enzyme-catalyzed reaction is an enzyme structure from X-ray crystallography. When this is not available, sometimes a model may be constructed based on homology to other structures that have been solved [272], though such models should be treated with much more caution. The first step in studying an

enzyme-catalyzed reaction is to establish its chemical mechanism. Its goal is to determine the functions of catalytic residues, which are often not obvious. Even the identities of many important groups may not be certain. Any specific interactions that stabilize transition states or reactive intermediates should also be identified and analyzed.

Enzymes, representing usually large molecules, need sophisticated modeling steps as the reactions that they catalyze are complex. This can be complicated further by the need to include a part of a particular enzyme's molecular environment, such as the surrounding solvent, cofactors, other proteins, a lipid membrane, or DNA. There are many practical considerations in simulating such complex systems, such as the proper interpretation of crystal structures and the choice of protonation states for ionizable amino acids [273]. Here, we illustrate these challenges with recent examples of modeling enzyme-catalyzed reactions.

The empirical valence bond method

The empirical valence bond (EVB) is considered to be one of the main methods used for modeling enzyme-catalyzed reactions [274]. In the EVB method, a few resonance structures are chosen to represent the reaction. The energy of each resonance form is given by a simple empirical force field, with the potential energy given by solving the related secular equation. The EVB Hamiltonian can be calibrated to reproduce experimental data for a reaction in solution, or *ab initio* results can be used [275]. The surrounding protein and solvent are modeled by an empirical force field, using proper long-range electrostatics. The free energy of activation is calculated from free energy perturbation simulations [276]. The free energy surfaces can be calibrated by comparison with experimental data for reactions in solution. The EVB method allows the use of a non-geometrical reaction coordinate, which helps to evaluate non-equilibrium solvation effects [274]. A mapping procedure gradually moves the system from the reactants to products. The simplicity of the EVB potential function allows extensive molecular dynamics (MD) simulations, giving good sampling [277]. The EVB method has been widely used for studying reactions in condensed phases, particularly in enzymes [278-284].

Quantum chemical methods

Another approach to the modeling of enzyme-catalyzed reactions is to study only the active site using quantum chemical methods. This methodology is usually named the cluster approach or the supermolecule approach. The active site model should contain molecules representing the substrate(s), any cofactors, and enzyme residues involved in the chemical reaction or in binding substrates. Important functional groups are represented by small molecules (e.g., acetate can represent an aspartate side chain). The initial positions of these groups are usually coordinates taken from a crystal structure, or from an MD simulation of an enzyme complex.

Quantum chemical methods can give excellent results for reactions of small molecules. Semi-empirical techniques, such as AM1 and PM3, can model larger systems that contain hundreds of atoms. However, semi-empirical methods must be applied with caution due to their sensitivity to parameterization, because typical errors may be over 10 kcal/mol for barriers and reaction energies [285,286]. DFT methods are considerably more accurate, while also allowing calculations on relatively large systems (e.g.,

active site models on the order of 100 atoms), larger than is feasible with correlated *ab initio* calculations. Many DFT methods, however, do not properly account for dispersion forces, which are important in the binding of ligands to proteins and can also be important in the calculation of energy barriers [287]. Dispersion corrections may be required in such cases [215-217]. Calculations on active site models can provide models of transition states and intermediates, which has proved particularly useful for studying metalloenzymes using DFT methods [288,289].

Combined quantum mechanics/molecular mechanics methods

Combined quantum mechanics/molecular mechanics (QM/MM) methods are widely applied to accurately model enzymes. This has been made possible by increased computer power, and improved software packages. QM/MM approaches treat a small part of the system quantum mechanically (describing the electronic structure of molecules) and the rest of the system with a molecular mechanics (MM) method (using a classical potential energy function [290]). The QM treatment accounts for the electronic rearrangements involved in bond breaking and bond making, while the MM treatment allows to include the effects of the environment on the reaction energetics.

There are two general types of QM/MM methods. The first is additive,

$$E_{\text{total}} = E_{\text{QM}}(\text{QM}) + E_{\text{MM}}(\text{MM}) + E_{\text{QM-MM,interaction}} \quad (38)$$

where $E_{\text{QM}}(\text{QM})$ is the energy of the QM region according to the QM method, $E_{\text{MM}}(\text{MM})$ is the energy of the MM region according to the MM method, and $E_{\text{QM-MM,interaction}}$ is the interaction energy between the two regions.

The second type of QM/MM method is subtractive,

$$E_{\text{total}} = E_{\text{MM}} + E_{\text{QM}}(\text{QM}) - E_{\text{MM}}(\text{QM}) \quad (39)$$

where E_{MM} is the energy of the total system as calculated by the MM method, $E_{\text{QM}}(\text{QM})$ is the energy of the QM region as calculated by the QM method, and $E_{\text{MM}}(\text{QM})$ is the energy of the QM region as calculated by the MM method. This is used, for example, in the ONIOM method [291].

In the subtractive approach, the active site region is modeled at the MM level, and the choice of suitable MM parameters (e.g., atomic charges) for all states of the reaction is an important and delicate consideration. Until recently, calculations using the subtractive approach typically used the more approximate “mechanical embedding” scheme, whereas currently most implementations of both approaches allow “electrostatic embedding” [291] which takes into account the electrostatic influence of the MM region on the QM region, i.e., polarization of the QM region by the atomic charges of the MM region. Five general aspects are important in a QM/MM calculation on an enzyme:

1. Choice of the QM method.
2. Choice of the MM force field (including the MM parameters for the QM region).
3. Partitioning of the system into QM and MM regions with due attention paid to any chemical bonds that straddle the two regions.
4. Type of simulation (e.g., an MD simulation, or calculation of potential energy profiles); whether extensive conformational sampling will be performed.
5. Construction (and testing) of an accurate molecular model of the enzyme complex.

The MM force field employed in a QM/MM study should be chosen to describe the part of the system outside the QM region and its interactions with the QM region. For proteins, standard all-atom force fields such as CHARMM27, AMBER ff99 or ff99SB, and OPLS-AA are commonly used. Apart from selecting suitable QM and MM methods and a QM/MM approach, modeling an enzyme reaction with a QM/MM method requires other important choices, such as deciding which atoms to include in the QM region and how to treat covalent bonds that cross the QM/MM boundary [292]. Another important choice is determining the protonation states of residues, and how (long-range) electrostatic interactions are treated. The influence of such choices on the results should be tested [267,293-296] in order to be able to draw reliable conclusions. Recent improvements allow relocating the QM-MM boundary on-the-fly (adaptive partitioning) [297].

Modeling enzyme reactions by calculating potential energy surfaces

With QM or QM/MM methods, potential energy surfaces of enzyme reaction mechanisms can be explored accurately enough to enable discrimination between different mechanisms, e.g., if the barrier for a proposed mechanism is significantly larger than that derived from experiment (using transition state theory), within the limits of accuracy of the computational method and experimental error, then that mechanism can be considered to be unlikely. A mechanism with a calculated barrier comparable to the apparent experimental barrier (for that step, or failing that for the overall reaction) is more likely. However, to calculate rate constants also requires reliable estimates of enthalpies, internal energies, and free energies of a given reaction and activation, given the potential energy surface. Traditional approaches to modeling reactions rely on the identification of stationary points (reactants, products, intermediates, transition states) via geometry optimization, followed by the computation of second derivatives to enable relatively simplistic evaluation of zero-point corrections, thermal and entropy terms. Algorithms developed for small molecules are often not suitable for large systems, storage and manipulation of Hessian matrices become extremely difficult. A basic means of modeling approximate reaction paths is the “adiabatic mapping” or “coordinate driving” approach. The energy of the system is found by minimizing the energy at a series of fixed (or restrained) values of a reaction coordinate, e.g., the distance between two atoms. This approach has been successfully applied to many enzymes [257], but it is only valid if one conformation of the protein can represent the state of the system at a particular value of the reaction coordinate.

Calculations of a potential energy surface may not consider significant conformational fluctuations of the enzyme. Conformational changes, even on a small scale, may cause significant chemical changes. Conformational changes of the active site can greatly affect the energy barrier. To take the fatty acid amide hydrolase as an example, conformational fluctuations do not affect the general shape of the potential energy surfaces, but consistency between experimental and calculated barriers is observed only with a specific infrequent arrangement of the enzyme-substrate complex [298]. These findings indicate that investigation of different protein conformations is essential for a meaningful determination of the energetics of enzymatic reactions for calculations of potential energy profiles or surfaces.

Calculating free energy profiles for enzyme-catalyzed reactions

According to transition state theory, the rate constant of a reaction is related to the free energy barrier. The techniques described previously calculate potential energy barriers for a particular conformation. Techniques that illustrate configurations along a reaction coordinate give a more sophisticated and extensive description by taking account of multiple conformations and estimating entropic effects, and can be essential for modeling enzyme reactions. Simulations of this type provide estimates of the free energy profile along a specific reaction coordinate, which is often referred to as the potential of mean force. MD and Monte Carlo methods allow such illustration, but do not provide a sufficiently detailed view of high energy regions, such as those in the vicinity of transition states. Conformational illustration of processes of chemical change requires specialized techniques, e.g., to bias the simulation to sample the transition state region. Umbrella sampling, which is widely used in MD simulations, when combined with QM/MM techniques, can be used to model enzymatic reactions [262]. QM/MM umbrella sampling simulations are possible with semi-empirical molecular orbital methods (e.g., AM1 or PM3). Often, such methods are highly inaccurate for reaction barriers and energies but their accuracy can be improved significantly by re-parameterization for a specific reaction.

Conclusions

This review paper has aimed to provide a comprehensive guide to a plethora of mathematical and computational methods developed in the past few decades to tackle key quantitative problems in the life sciences. Our critical overview covers methods used across the life sciences, starting from macroscopic systems such as those in evolutionary biology, and ending with atomic level descriptions of biomolecules including quantum mechanical or hybrid classical/quantum approaches. Particular attention was given to large-scale computational methods, such as molecular dynamics, which play pivotal roles in the development of our understanding of molecular mechanisms at the level of molecular, structural, and cell biology. Important applications in medicine and pharmaceutical sciences have been discussed, in particular in the context of extracting crucial conclusions about complex system behavior with information limitations. We hope the reader will be encouraged to explore particular topics at a deeper level using the information and references provided in this review.

Abbreviations

AIDS: Acquired immunodeficiency syndrome; AMBER: Assisted model building with energy refinement; AMP: Adenosine monophosphate; ASFE: Advanced solvation force extrapolation; cAMP: Cyclic adenosine monophosphate; CF: Cystic fibrosis; COPASI: Complex pathway simulator; DFT: Density functional theory; DNA: Deoxyribonucleic acid; DTI: Diffusion tensor imaging; EC: Endothelial cell; ECM: Extracellular matrix; EOM: Equations of motion; EVB: Empirical valence bond; GAMESS: General atomic and molecular electronic structure system; GB: Generalized born; GF: Growth factor; GFP: Green fluorescent protein; GPCR: G protein-coupled receptor; GPU: Graphical processing unit; GROMOS: Groningen molecular simulation; HNC: Hypernetted chain; LJ: Lennard-Jones; LRC: Long-range van der Waals correction; MARTINI: Coarse-grained force field; MD: Molecular dynamics; MM: Molecular mechanics; MM-GBSA: Molecular mechanics Generalized Born surface area; MM-PBSA: Molecular mechanics Poisson-Boltzmann surface area; MRI: Magnetic resonance imaging; MTS: Multiple time step; MTS-MD: Multiple time step molecular dynamics; NAMD: Not just another molecular dynamics software; NMR: Nuclear magnetic resonance; ODE: Ordinary differential equation; OIN: Optimized isokinetic Nosé-Hoover chain; ONIOM: Our own N-layered Integrated molecular Orbital and molecular Mechanics; OPLS: Optimized Potentials for Liquid Simulations; P450: Cytochrome P450 superfamily of monooxygenases; PB: Poisson-Boltzmann; PDB: Protein data bank; PDE: Partial differential equation; PME: Particle mesh Ewald; PMF: Potential mean force; QM: Quantum mechanics; QM/MM: Quantum mechanics/molecular mechanics; QSAR: Quantitative structure-activity relationship; REMD: Replica-exchange molecular dynamics; RISM: Reference interaction site model; RMSD: Root-mean-square deviation; RNA: Ribonucleic acid; SBML: Systems biology markup language; SBW: Systems biology workbench; SEIR: Susceptible-infected-exposed-recovered

model; SEIRS: Susceptible-infected-exposed-recovered-susceptible model; SELEX: Systematic Evolution of Ligands by Exponential Enrichment; SFCE: Solvation force-coordinate extrapolation; SHAKE: Algorithm for molecular dynamics; SIR: Susceptible-infected-recovered model; SIRS: Susceptible-infected-recovered-susceptible model; TAF: Tumor angiogenic factor; TCP: Tumor control probability; XML: Extensible markup language.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

JAT conceived of the review. JAT and PW edited the review and wrote various portions of the text. DW wrote the section on mathematical models. CYT wrote the section on maximum entropy. KS wrote the section on molecular dynamics. FG and IS wrote the section on force fields for protein simulations. SIO and NN wrote the section on the calculation of solvation free energies. CDMC and MK wrote the section on quantum mechanics. RMAEM wrote the section on computational enzymology, the subsection on enzyme kinetics and computational biology, and part of the section on molecular dynamics. All authors read and approved the final manuscript.

Acknowledgements

The authors are grateful to the Natural Sciences and Engineering Research Council of Canada, the Li Ka Shing Institute of Applied Virology and the Alberta Cancer Foundation for funding support.

Author details

¹Department of Physics and Department of Oncology, University of Alberta, Edmonton, Canada. ²Department of Oncology, University of Alberta, Edmonton, Canada. ³Department of Physics, University of Alberta, Edmonton, Canada. ⁴Department of Mechanical and Aerospace Engineering, Politecnico di Torino, Torino, Italy. ⁵Department of Biology, Ecole Normale Supérieure de Lyon, Lyon, France. ⁶Department of Chemistry, University of Alberta, Edmonton, Canada. ⁷Department of Biological Sciences, University of Alberta, Edmonton, Canada.

Received: 20 October 2014 Accepted: 25 November 2014

Published: 27 December 2014

References

1. Caticha A: *Entropic Inference and the Foundations of Physics (monograph Commissioned by the 11th Brazilian Meeting on Bayesian Statistics-EBEB-2012)*. Sao Paulo: USP Press; 2012.
2. Cox RT: *Probability, Frequency, and Reasonable Expectation*; 1946.
3. Caticha A, Giffin A: **Updating probabilities**. *ArXiv Prepr Physics* 2006, 0608185.
4. Chang C-H, Hsieh L-C, Chen T-Y, Chen H-D, Luo L, Lee H-C: **Shannon information in complete genomes**. *J Bioinform Comput Biol* 2005, **3**:587–608.
5. Dover Y: **A short account of a connection of power laws to the information entropy**. *Phys Stat Mech Its Appl* 2004, **334**:591–599.
6. Chang DT-H, Oyang Y-J, Lin J-H: **MEDock: a web server for efficient prediction of ligand binding sites based on a novel optimization algorithm**. *Nucleic Acids Res* 2005, **33**(suppl 2):W233–W238.
7. Tseng C-Y, Ashrafuzzaman M, Mane JY, Kapyt J, Mercer JR, Tuszynski JA: **Entropic fragment-based approach to aptamer design**. *Chem Biol Drug Des* 2011, **78**:1–13.
8. Shannon CE: **A mathematical theory of communication**. *Bell Syst Tech J* 1948, **27**:379–423.
9. Stelling J, Sauer U, Szallasi Z, Doyle FJ 3rd, Doyle J: **Robustness of cellular functions**. *Cell* 2004, **118**:675–685.
10. Li F, Long T, Lu Y, Ouyang Q, Tang C: **The yeast cell-cycle network is robustly designed**. *Proc Natl Acad Sci U S A* 2004, **101**:4781–4786.
11. Bak P, Tang C, Wiesenfeld K: **Self-organized criticality**. *Phys Rev A* 1988, **38**:364–374.
12. Honma T: **Recent advances in de novo design strategy for practical lead identification**. *Med Res Rev* 2003, **23**:606–632.
13. Schneider G, Fechner U: **Computer-based de novo design of drug-like molecules**. *Nat Rev Drug Discov* 2005, **4**:649–663.
14. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ: **AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility**. *J Comput Chem* 2009, **30**:2785–2791.
15. Nimjee SM, Rusconi CP, Sullenger BA: **Aptamers: an emerging class of therapeutics**. *Annu Rev Med* 2005, **56**:555–583.
16. James W: **Aptamer**. In *Encycl Anal Chem*. Hoboken, NJ: Wiley & Sons Inc; 2000:4848–4871.
17. Hamula CLA, Guthrie JW, Zhang H, Li X-F, Le XC: **Selection and analytical applications of aptamers**. *TrAC Trends Anal Chem* 2006, **25**:681–691. On-Site Instrumentation and Analysis.
18. James W: **Aptamers in the virologists' toolkit**. *J Gen Virol* 2007, **88**(Pt 2):351–364.
19. Ashrafuzzaman M, Tseng C-Y, Kapyt J, Mercer JR, Tuszynski JA: **A computationally designed DNA aptamer template with specific binding to phosphatidylserine**. *Nucleic Acid Ther* 2013, **23**:418–426.
20. Brauer F, Castillo-Chávez C: *Mathematical Models in Population Biology and Epidemiology*. New York: Springer; 2001.
21. Brauer F, Van den Driessche P, Wu J, Allen LJS: *Mathematical Epidemiology*. Berlin: Springer; 2008.
22. Keener JP, Sneyd J: *Mathematical Physiology*. New York: Springer; 1998.
23. Preziosi L: *Cancer Modelling and Simulation*. Boca Raton, Fla: Chapman & Hall/CRC; 2003.
24. Edelstein-Keshet L: *Mathematical Models in Biology*. New York: Random House; 1988.
25. Murray JD: *Mathematical Biology I: An Introduction*. 3rd edition. New York: Springer Verlag; 2003.
26. Murray JD: *Mathematical Biology II: Spatial Models and Biomedical Applications*. 3rd edition. New York: Springer Verlag; 2003.
27. Britton NF: *Essential Mathematical Biology*. 2nd edition. London: Springer Verlag; 2003.
28. Merrill RM: *Introduction to Epidemiology*; 2013.
29. Kermack MD, McKendrick AG: **Contributions to the mathematical theory of epidemics. Part I**. In *Proc R Soc A* 1927, **115**:700–721.
30. Hethcote HW: **Qualitative analyses of communicable disease models**. *Math Biosci* 1976, **28**:335–356.

31. Soper HE: **The interpretation of periodicity in disease prevalence.** *J R Stat Soc* 1929, **92**:34–73.
32. Kermack WO, McKendrick AG: **Contributions to the mathematical theory of epidemics. II. The problem of endemicity.** *Proc R Soc Lond Ser A* 1932, **138**:55–83.
33. Smith HL, Wang L, Li MY: **Global Dynamics of an SEIR Epidemic Model with Vertical Transmission.** *SIAM J Appl Math* 2001, **62**:58–69.
34. Dunham JB: **An agent-based spatially explicit epidemiological model in MASON.** *J Artif Soc Soc Simul* 2005, **9**:3.
35. Tian C, Ding W, Cao R, Jiang S: **Extensive Epidemic Spreading Model Based on Multi-agent System Framework.** *Lect Notes Comput Sci* 2007, **4490**:129–133.
36. Newman MEJ: **Spread of epidemic disease on networks.** *Phys Rev E Stat Nonlin Soft Matter Phys* 2002, **66**(1 Pt 2):016128.
37. Webb GF: *Theory of Nonlinear Age-Dependent Population Dynamics.* New York: M. Dekker; 1985.
38. Murray JD, Stanley EA, Brown DL: **On the spatial spread of rabies among foxes.** *Proc R Soc Lond Ser B Contain Pap Biol Character R Soc G B* 1986, **229**:111–150.
39. Wonham MJ, de-Camino-Beck T, Lewis MA: **An epidemiological model for West Nile virus: invasion analysis and control applications.** *Proc Biol Sci* 2004, **271**:501–507.
40. Lloyd AL, May RM: **Spatial heterogeneity in epidemic models.** *J Theor Biol* 1996, **179**:1–11.
41. Fraser C: *Epidemiology and Control of Infectious Diseases.* London: Department of Infectious Disease Epidemiology Imperial College London; 2014.
42. Lehninger AL, Nelson DL, Cox MM: *Principles of Biochemistry.* New York, NY: Worth Publishers; 1993.
43. Köhler J, Baumbach J, Taubert J, Specht M, Skusa A, Rüegg A, Rawlings C, Verrier P, Philipp S: **Graph-based analysis and visualization of experimental results with ONDEX.** *Bioinforma Oxf Engl* 2006, **22**:1383–1390.
44. Michaelis L, Menten ML, Johnson KA, Goody RS: **The original Michaelis constant: translation of the 1913 Michaelis-Menten paper.** *Biochemistry (Mosc)* 2011, **50**:8264–8269.
45. Briggs GE, Haldane JB: **A Note on the Kinetics of Enzyme Action.** *Biochem J* 1925, **19**:338–339.
46. Monod J, Wyman J, Changeux JP: **On the nature of allosteric transitions: a plausible model.** *J Mol Biol* 1965, **12**:88–118.
47. Hübner CA, Jentsch TJ: **Ion channel diseases.** *Hum Mol Genet* 2002, **11**:2435–2445.
48. Mendes P, Messiha H, Malys N, Hoops S: **Enzyme kinetics and computational modeling for systems biology.** *Methods Enzymol* 2009, **467**:583–599.
49. Chance B, Garfinkel D, Higgins J, Hess B: **Metabolic control mechanisms. 5. A solution for the equations representing interaction between glycolysis and respiration in ascites tumor cells.** *J Biol Chem* 1960, **235**:2426–2439.
50. Cornish-Bowden A: **Why is uncompetitive inhibition so rare? A possible explanation, with implications for the design of drugs and pesticides.** *FEBS Lett* 1986, **203**:3–6.
51. Abou El-Magd RM, Park HK, Kawazoe T, Iwana S, Ono K, Chung SP, Miyano M, Yorita K, Sakai T, Fukui K: **The effect of risperidone on D-amino acid oxidase activity as a hypothesis for a novel mechanism of action in the treatment of schizophrenia.** *J Psychopharmacol Oxf Engl* 2010, **24**:1055–1067.
52. Westley AM, Westley J: **Enzyme inhibition in open systems. Superiority of uncompetitive agents.** *J Biol Chem* 1996, **271**:5347–5352.
53. Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, Cuellar AA, Dronov S, Gilles ED, Ginkel M, Gor V, Goryanin II, Hedley WJ, Hodgman TC, Hofmeyr J-H, Hunter PJ, Juty NS, Kasberger JL, Kremling A, Kummer U, Le Novère N, Loew LM, Lucio D, Mendes P, Minch E, Mjolsness ED, et al.: **The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models.** *Bioinforma Oxf Engl* 2003, **19**:524–531.
54. Hoops S, Sahle S, Gauges R, Lee C, Pahle J, Simus N, Singhal M, Xu L, Mendes P, Kummer U: **COPASI—a COMplex PATHway Simulator.** *Bioinforma Oxf Engl* 2006, **22**:3067–3074.
55. Funahashi A, Morohashi M, Kitano H, Tanimura N: **Cell Designer: a process diagram editor for gene-regulatory and biochemical networks.** *Biosilico* 2003, **1**:159–162.
56. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** *Genome Res* 2003, **13**:2498–2504.
57. Rojas I, Golebiewski M, Kania R, Krebs O, Mir S, Weidemann A, Wittig U: **Storing and annotating of kinetic data.** *In Silico Biol* 2007, **7**(2 Suppl):S37–S44.
58. Sauro HM, Hucka M, Finney A, Wellock C, Bolouri H, Doyle J, Kitano H: **Next generation simulation tools: the Systems Biology Workbench and BioSPICE integration.** *Omics J Integr Biol* 2003, **7**:355–372.
59. Frankenhaeuser B: **A quantitative description of potassium currents in myelinated nerve fibres of *Xenopus laevis*.** *J Physiol* 1963, **169**:424–430.
60. Hodgkin AL, Huxley AF, Katz B: **Measurement of current-voltage relations in the membrane of the giant axon of *Loligo*.** *J Physiol* 1952, **116**:424–448.
61. Hodgkin AL, Huxley AF: **Currents carried by sodium and potassium ions through the membrane of the giant axon of *Loligo*.** *J Physiol* 1952, **116**:449–472.
62. Hodgkin AL, Huxley AF: **The components of membrane conductance in the giant axon of *Loligo*.** *J Physiol* 1952, **116**:473–496.
63. Hodgkin AL, Huxley AF: **The dual effect of membrane potential on sodium conductance in the giant axon of *Loligo*.** *J Physiol* 1952, **116**:497–506.
64. Hodgkin AL, Huxley AF: **A quantitative description of membrane current and its application to conduction and excitation in nerve.** *J Physiol* 1952, **117**:500–544.
65. Moreno JD, Zhu ZI, Yang P-C, Bankston JR, Jeng M-T, Kang C, Wang L, Bayer JD, Christini DJ, Trayanova NA, Ripplinger CM, Kass RS, Clancy CE: **A computational model to predict the effects of class I anti-arrhythmic drugs on ventricular rhythms.** *Sci Transl Med* 2011, **3**:98ra83.
66. Welsh MJ: **Abnormal regulation of ion channels in cystic fibrosis epithelia.** *FASEB J Off Publ Fed Am Soc Exp Biol* 1990, **4**:2718–2725.
67. Keiser NW, Engelhardt JF: **New animal models of cystic fibrosis: what are they teaching us?** *Curr Opin Pulm Med* 2011, **17**:478–483.
68. Knowles MA, Selby P: *Introduction to the Cellular and Molecular Biology of Cancer.* New York: Oxford University Press; 2005.

69. Vaidya VG, Alexandro FJ: Evaluation of some mathematical models for tumor growth. *Int J Biomed Comput* 1982, **13**:19–36.
70. Casey AE: The experimental alteration of malignancy with an homologous mammalian tumor material: I. results with intratesticular inoculation. *Am J Cancer* 1934, **21**:760–775.
71. Cruywagen GC, Woodward DE, Tracqui P, Bartoo GT, Murray JD, Alvord EC: The modelling of diffusive tumours. *J Biol Syst* 1995, **03**:937–945.
72. Tracqui P, Cruywagen GC, Woodward DE, Bartoo GT, Murray JD, Alvord EC: A mathematical model of glioma growth: the effect of chemotherapy on spatio-temporal growth. *Cell Prolif* 1995, **28**:17–31.
73. Swanson KR, Alvord EC, Murray JD: A quantitative model for differential motility of gliomas in grey and white matter. *Cell Prolif* 2000, **33**:317–329.
74. Chaplain MA: Avascular growth, angiogenesis and vascular growth in solid tumours: The mathematical modelling of the stages of tumour development. *Math Comput Model* 1996, **23**:47–87.
75. Chaplain MA, Ganesh M, Graham IG: Spatio-temporal pattern formation on spherical surfaces: numerical simulation and application to solid tumour growth. *J Math Biol* 2001, **42**:387–423.
76. Sherratt JA, Chaplain MA: A new mathematical model for avascular tumour growth. *J Math Biol* 2001, **43**:291–312.
77. Semenza GL: Angiogenesis in ischemic and neoplastic disorders. *Annu Rev Med* 2003, **54**:17–28.
78. Keller EF, Segel LA: Model for chemotaxis. *J Theor Biol* 1971, **30**:225–234.
79. Balding D, McElwain DL: A mathematical model of tumour-induced capillary growth. *J Theor Biol* 1985, **114**:53–73.
80. Anderson AR, Chaplain MA: Continuous and discrete mathematical models of tumor-induced angiogenesis. *Bull Math Biol* 1998, **60**:857–899.
81. Mantzaris NV, Webb S, Othmer HG: Mathematical modeling of tumor-induced angiogenesis. *J Math Biol* 2004, **49**:111–187.
82. Hillen T: M5 mesoscopic and macroscopic models for mesenchymal motion. *J Math Biol* 2006, **53**:585–616.
83. Preziosi L: Modeling cell movement in anisotropic and heterogeneous network tissues. *Netw Heterog Media* 2007, **2**:333–357.
84. Preziosi L: Modelling the motion of a cell population in the extracellular matrix. *DISCRETE Contin Dyn Syst Ser B* 2007, **Supplement**:250–259.
85. Collins I, Workman P: New approaches to molecular cancer therapeutics. *Nat Chem Biol* 2006, **2**:689–700.
86. Baskar R, Lee KA, Yeo R, Yeoh K-W: Cancer and radiation therapy: current advances and future directions. *Int J Med Sci* 2012, **9**:193–199.
87. Swanson KR, Rostomily RC, Alvord EC: A mathematical modelling tool for predicting survival of individual patients following resection of glioblastoma: a proof of principle. *Br J Cancer* 2008, **98**:113–119.
88. Rockne R, Rockhill JK, Mrugala M, Spence AM, Kalet I, Hendrickson K, Lai A, Cloughesy T, Alvord EC, Swanson KR: Predicting the efficacy of radiotherapy in individual glioblastoma patients in vivo: a mathematical modeling approach. *Phys Med Biol* 2010, **55**:3271–3285.
89. Clatz O, Sermesant M, Boudreau P-Y, Delingette H, Warfield SK, Malandain G, Ayache N: Realistic simulation of the 3-D growth of brain tumors in MR images coupling diffusion with biomechanical deformation. *IEEE Trans Med Imaging* 2005, **24**:1334–1346.
90. Painter KJ, Hillen T: Mathematical modelling of glioma growth: the use of Diffusion Tensor Imaging (DTI) data to predict the anisotropic pathways of cancer invasion. *J Theor Biol* 2013, **323**:25–39.
91. Hillen T, de Vries G, Gong J, Finlay C: From cell population models to tumor control probability: including cell cycle effects. *Acta Oncol Stockh Swed* 2010, **49**:1315–1323.
92. Munro TR, Gilbert CW: The relation between tumour lethal doses and the radiosensitivity of tumour cells. *Br J Radiol* 1961, **34**:246–251.
93. Zaider M, Minerbo GN: Tumour control probability: a formulation applicable to any temporal protocol of dose delivery. *Phys Med Biol* 2000, **45**:279–293.
94. Zhu T, Lee H, Lei H, Jones C, Patel K, Johnson ME, Hevener KE: Fragment-based drug discovery using a multidomain, parallel MD-MM/PBSA screening protocol. *J Chem Inf Model* 2013, **53**:560–572.
95. Fujihashi M, Ishida T, Kuroda S, Kotra LP, Pai EF, Miki K: Substrate distortion contributes to the catalysis of orotidine 5'-monophosphate decarboxylase. *J Am Chem Soc* 2013, **135**:17432–17443.
96. Tiwari G, Mohanty D: An in silico analysis of the binding modes and binding affinities of small molecule modulators of PDZ-peptide interactions. *PLoS One* 2013, **8**:e71340.
97. Zahn D: Molecular dynamics simulation of ionic conductors: perspectives and limitations. *J Mol Model* 2011, **17**:1531–1535.
98. Wood NT, Fadda E, Davis R, Grant OC, Martin JC, Woods RJ, Travers SA: The influence of N-linked glycans on the molecular dynamics of the HIV-1 gp120 V3 loop. *PLoS One* 2013, **8**:e80301.
99. Chaudhuri A, Sarkar I, Chakraborty S: Comparative analysis of binding sites of human meprins with hydroxamic acid derivative by molecular dynamics simulation study. *J Biomol Struct Dyn* 2013,
100. Leonis G, Steinbrecher T, Papadopoulos MG: A contribution to the drug resistance mechanism of darunavir, amprenavir, indinavir, and saquinavir complexes with HIV-1 protease due to flap mutation I50V: a systematic MM-PBSA and thermodynamic integration study. *J Chem Inf Model* 2013, **53**:2141–2153.
101. Yoda T, Sugita Y, Okamoto Y: Hydrophobic core formation and dehydration in protein folding studied by generalized-ensemble simulations. *Biophys J* 2010, **99**:1637–1644.
102. Yoda T, Sugita Y, Okamoto Y: Salt effects on hydrophobic-core formation in folding of a helical miniprotein studied by molecular dynamics simulations. *Proteins* 2014, **82**:933–943.
103. Fukunishi Y, Nakamura H: Improved estimation of protein-ligand binding free energy by using the ligand-entropy and mobility of water molecules. *Pharm Basel Switz* 2013, **6**:604–622.
104. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA: A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* 1995, **117**:5179–5197.
105. Ponder JW, Case DA: Force fields for protein simulations. *Adv Protein Chem* 2003, **66**:27–85.

106. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C: **Comparison of multiple Amber force fields and development of improved protein backbone parameters.** *Proteins* 2006, **65**:712–725.
107. Wang J, Cieplak P, Kollman PA: **How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?** *J Comput Chem* 2000, **21**:1049–1074.
108. Wickstrom L, Okur A, Simmerling C: **Evaluating the performance of the ff99SB force field based on NMR scalar coupling data.** *Biophys J* 2009, **97**:853–856.
109. Cino EA, Choy W-Y, Karttunen M: **Comparison of secondary structure formation using 10 different force fields in microsecond molecular dynamics simulations.** *J Chem Theory Comput* 2012, **8**:2725–2740.
110. Case DA, Babin V, Berryman JT, Betz RM, Cerutti T, Cheatham I, Darden TA, Duke RE, Gohlke H, Goetz AW, Gusarow S, Homeyer N, Janowski P, Anderson J, Kolossvary I, Kaus J, Kovalenko A, Lee TS, LeGrand S, Luchko T, Luo R, Madej B, Merz KM, Paesani F, Roe DR, Roitberg A, Sagui C, Salomon-Ferrer R, Seabra G, Simmerling CL, et al: *AMBER 14*; 2014.
111. Zhang Y, Sagui C: **The gp41(659-671) HIV-1 antibody epitope: a structurally challenging small peptide.** *J Phys Chem B* 2014, **118**:69–80.
112. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P: **A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations.** *J Comput Chem* 2003, **24**:1999–2012.
113. Lee MC, Duan Y: **Distinguish protein decoys by Using a scoring function based on a new AMBER force field, short molecular dynamics simulations, and the generalized born solvent model.** *Proteins Struct Funct Bioinforma* 2004, **55**:620–634.
114. Yang L, Tan C-H, Hsieh M-J, Wang J, Duan Y, Cieplak P, Caldwell J, Kollman PA, Luo R: **New-generation amber united-atom force field.** *J Phys Chem B* 2006, **110**:13166–13176.
115. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA: **Development and testing of a general amber force field.** *J Comput Chem* 2004, **25**:1157–1174.
116. MacKerell AD, Banavali N, Foloppe N: **Development and current status of the CHARMM force field for nucleic acids.** *Biopolymers* 2000, **56**:257–265.
117. Jorgensen WL, Tirado-Rives J: **The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin.** *J Am Chem Soc* 1988, **110**:1657–1666.
118. Kaminski G, Jorgensen WL: **Performance of the AMBER94, MMFF94, and OPLS-AA force fields for modeling organic liquids.** *J Phys Chem* 1996, **100**:18010–18013.
119. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M: **All-atom empirical potential for molecular modeling and dynamics studies of proteins.** *J Phys Chem B* 1998, **102**(18):3586–3616.
120. Feig M, MacKerell AD, Brooks CL: **Force field influence on the observation of π -helical protein structures in molecular dynamics simulations.** *J Phys Chem B* 2003, **107**:2831–2836.
121. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, Mackerell AD: **CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields.** *J Comput Chem* 2010, **31**:671–690.
122. Jorgensen WL, Maxwell DS, Tirado-Rives J: **Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids.** *J Am Chem Soc* 1996, **118**:11225–11236.
123. Maxwell DS, Tirado-Rives J, Jorgensen WL: **A comprehensive study of the rotational energy profiles of organic systems by ab initio MO theory, forming a basis for peptide torsional parameters.** *J Comput Chem* 1995, **16**:984–1010.
124. Kahn K, Bruice TC: **Parameterization of OPLS-AA force field for the conformational analysis of macrocyclic polyketides.** *J Comput Chem* 2002, **23**:977–996.
125. Siu SWI, Pluhackova K, Böckmann RA: **Optimization of the OPLS-AA force field for long hydrocarbons.** *J Chem Theory Comput* 2012, **8**:1459–1470.
126. Daura X, Mark AE, Van Gunsteren WF: **Parametrization of aliphatic CH_n united atoms of GROMOS96 force field.** *J Comput Chem* 1998, **19**:535–547.
127. Schuler LD, Daura X, van Gunsteren WF: **An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase.** *J Comput Chem* 2001, **22**:1205–1218.
128. Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF: **A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6.** *J Comput Chem* 2004, **25**:1656–1676.
129. Reif MM, Winger M, Oostenbrink C: **Testing of the GROMOS force-field parameter set 54A8: structural properties of electrolyte solutions, lipid bilayers, and proteins.** *J Chem Theory Comput* 2013, **9**:1247–1264.
130. Schmid N, Eichenberger AP, Choutko A, Riniker S, Winger M, Mark AE, van Gunsteren WF: **Definition and testing of the GROMOS force-field versions 54A7 and 54B7.** *Eur Biophys J EBJ* 2011, **40**:843–856.
131. Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, de Vries AH: **The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations.** *J Phys Chem B* 2007, **111**:7812–7824.
132. Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink S-J: **The MARTINI Coarse-Grained Force Field: Extension to Proteins.** *J Chem Theory Comput* 2008, **4**:819–834.
133. López CA, Rzepiela AJ, de Vries AH, Dijkhuizen L, Hünenberger PH, Marrink SJ: **Martini Coarse-Grained Force Field: Extension to Carbohydrates.** *J Chem Theory Comput* 2009, **5**:3195–3210.
134. López CA, Sovova Z, van Eerden FJ, de Vries AH, Marrink SJ: **Martini Force Field Parameters for Glycolipids.** *J Chem Theory Comput* 2013, **9**:1694–1708.
135. Marrink SJ, de Vries AH, Tieleman DP: **Lipids on the move: Simulations of membrane pores, domains, stalks and curves.** *Biochim Biophys Acta BBA - Biomembr* 2009, **1788**:149–168. Lipid Interactions, Domain Formation, and Lateral Structure of Membranes.
136. Yoo J, Cui Q: **Membrane-mediated protein-protein interactions and connection to elastic models: a coarse-grained simulation analysis of gramicidin A association.** *Biophys J* 2013, **104**:128–138.

137. Gautieri A, Russo A, Vesentini S, Redaelli A, Buehler MJ: **Coarse-Grained Model of Collagen Molecules Using an Extended MARTINI Force Field.** *J Chem Theory Comput* 2010, **6**:1210–1218.
138. Wong-Ekkabut J, Baoukina S, Triampo W, Tang J-M, Tieleman DP, Monticelli L: **Computer simulation study of fullerene translocation through lipid membranes.** *Nat Nanotechnol* 2008, **3**:363–368.
139. Marrink SJ, Tieleman DP: **Perspective on the Martini model.** *Chem Soc Rev* 2013, **42**:6801–6822.
140. Freedman H: *Solvation Free Energies from a Coupled Reference Interaction Site Model/simulation Approach (Thesis).* Department of Chemistry: University of Utah; 2005.
141. Chipman DM: **Vertical electronic excitation with a dielectric continuum model of solvation including volume polarization. I. Theory.** *J Chem Phys* 2009, **131**:014103.
142. Canuto S: *Solvation Effects on Molecules and Biomolecules: Computational Methods and Applications.* Sao Paulo, Brazil: Springer; 2008.
143. Arakawa T, Kamiya N, Nakamura H, Fukuda I: **Molecular dynamics simulations of double-stranded DNA in an explicit solvent model with the zero-dipole summation method.** *PLoS One* 2013, **8**:e76606.
144. Liu Y, Haddadian E, Sosnick TR, Freed KF, Gong H: **A novel implicit solvent model for simulating the molecular dynamics of RNA.** *Biophys J* 2013, **105**:1248–1257.
145. Yu Z, Jacobson MP, Friesner RA: **What Role Do Surfaces Play in GB Models? A New-Generation of Surface-Generalized Born Model Based on a Novel Gaussian Surface for Biomolecules.** *J Comput Chem* 2006, **27**:72–89.
146. Pang X, Zhou H-X: **Poisson-Boltzmann Calculations: van der Waals or Molecular Surface?** *Commun Comput Phys* 2013, **13**:1–12.
147. Salomon-Ferrer R, Case DA, Walker RC: **An overview of the Amber biomolecular simulation package.** *Wiley Interdiscip Rev Comput Mol Sci* 2013, **3**:198–210.
148. Höfinger S: **Solving the Poisson-Boltzmann equation with the specialized computer chip MD-GRAPE-2.** *J Comput Chem* 2005, **26**:1148–1154.
149. Luchko T, Gusarov S, Roe DR, Simmerling C, Case DA, Tuszynski J, Kovalenko A: **Three-dimensional molecular theory of solvation coupled with molecular dynamics in Amber.** *J Chem Theory Comput* 2010, **6**:607–624.
150. Orozco M, Luque FJ: **Theoretical methods for the description of the solvent effect in biomolecular systems.** *Chem Rev* 2000, **100**:4187–4226.
151. Simkin BI: *Quantum Chemical and Statistical Theory of Solutions: A Computational Approach.* London; New York: Ellis Horwood; 1995. Ellis Horwood Series in Physical Chemistry.
152. Chandler D, McCoy JD, Singer SJ: **Density functional theory of nonuniform polyatomic systems. I. General formulation.** *J Chem Phys* 1986, **85**:5971.
153. Chandler D, McCoy JD, Singer SJ: **Density functional theory of nonuniform polyatomic systems. II. Rational closures for integral equations.** *J Chem Phys* 1986, **85**:5977.
154. Beglov D, Roux B: **An integral equation to describe the solvation of polar molecules in liquid water.** *J Phys Chem B* 1997, **101**:7821–7826.
155. Kovalenko A, Hirata F: **Potential of mean force between two molecular ions in a polar molecular solvent: a study by the three-dimensional Reference Interaction Site Model.** *J Phys Chem B* 1999, **103**:7942–7957.
156. Kovalenko A, Hirata F: **Self-consistent description of a metal–water interface by the Kohn–Sham density functional theory and the three-dimensional reference interaction site model.** *J Chem Phys* 1999, **110**:10095.
157. Kovalenko A, Hirata F: **Potentials of mean force of simple ions in ambient aqueous solution. I. Three-dimensional reference interaction site model approach.** *J Chem Phys* 2000, **112**:10391.
158. Kovalenko A, Hirata F: **Potentials of mean force of simple ions in ambient aqueous solution. II. Solvation structure from the three-dimensional reference interaction site model approach, and comparison with simulations.** *J Chem Phys* 2000, **112**:10403.
159. Andersen HC: **Optimized cluster expansions for classical fluids. I. General theory and variational formulation of the mean spherical model and hard sphere Percus–Yevick equations.** *J Chem Phys* 1972, **57**:1918.
160. Rossky PJ, Friedman HL: **Accurate solutions to integral equations describing weakly screened ionic systems.** *J Chem Phys* 1980, **72**:5694.
161. Hirata F, Rossky PJ: **An extended rism equation for molecular polar fluids.** *Chem Phys Lett* 1981, **83**:329–334.
162. Hirata F: **Application of an extended RISM equation to dipolar and quadrupolar fluids.** *J Chem Phys* 1982, **77**:509.
163. Hansen J-P: *McDonald IR: Theory of Simple Liquids with Applications to Soft Matter.* Burlington: Elsevier Science; 2013.
164. Hirata F: *Molecular Theory of Solvation.* Dordrecht; Boston: Kluwer Academic Publishers; 2003. Understanding Chemical Reactivity, v. 24.
165. Tuckerman ME, Berne BJ, Martyna GJ: **Molecular dynamics algorithm for multiple time scales: Systems with long range forces.** *J Chem Phys* 1991, **94**:6811.
166. Tuckerman M, Berne BJ, Martyna GJ: **Reversible multiple time scale molecular dynamics.** *J Chem Phys* 1992, **97**:1990.
167. Omelyan I, Kovalenko A: **Generalised canonical–isokinetic ensemble: speeding up multiscale molecular dynamics and coupling with 3D molecular theory of solvation.** *Mol Simul* 2013, **39**:25–48.
168. Truchon J-F, Pettitt BM, Labute P: **A cavity corrected 3D-RISM functional for accurate solvation free energies.** *J Chem Theory Comput* 2014, **10**:934–941.
169. Ng K-C: **Hypernetted chain solutions for the classical one-component plasma up to $\Gamma = 7000$.** *J Chem Phys* 1974, **61**:2680.
170. Anderson DG: **Iterative procedures for nonlinear integral equations.** *J ACM* 1965, **12**:547–560.
171. Maruyama Y, Hirata F: **Modified Anderson method for accelerating 3D-RISM calculations using graphics processing unit.** *J Chem Theory Comput* 2012, **8**:3015–3021.
172. Miller BR, McGee TD, Swails JM, Homeyer N, Gohlke H, Roitberg AE: **MMPBSA.py : An Efficient Program for End-State Free Energy Calculations.** *J Chem Theory Comput* 2012, **8**:3314–3321.
173. Genheden S, Luchko T, Gusarov S, Kovalenko A, Ryde U: **An MM/3D-RISM Approach for Ligand Binding Affinities.** *J Phys Chem B* 2010, **114**:8505–8516.

174. Berendsen HJC, Grigera JR, Straatsma TP: **The missing term in effective pair potentials.** *J Phys Chem* 1987, **91**:6269–6271.
175. Caldwell JW, Kollman PA: **Structure and properties of neat liquids using nonadditive molecular dynamics: water, methanol, and N-methylacetamide.** *J Phys Chem* 1995, **99**:6208–6219.
176. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML: **Comparison of simple potential functions for simulating liquid water.** *J Chem Phys* 1983, **79**:926.
177. Price DJ, Brooks CL: **A modified TIP3P water potential for simulation with Ewald summation.** *J Chem Phys* 2004, **121**:10096.
178. Jorgensen WL, Madura JD: **Temperature and size dependence for Monte Carlo simulations of TIP4P water.** *Mol Phys* 1985, **56**:1381–1392.
179. Horn HW, Swope WC, Pitner JW, Madura JD, Dick TJ, Hura GL, Head-Gordon T: **Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew.** *J Chem Phys* 2004, **120**:9665.
180. Mahoney MW, Jorgensen WL: **A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions.** *J Chem Phys* 2000, **112**:8910.
181. Toukan K, Rahman A: **Molecular-dynamics study of atomic motions in water.** *Phys Rev B* 1985, **31**:2643–2648.
182. Ryckaert J-P, Ciccotti G, Berendsen HJC: **Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes.** *J Comput Phys* 1977, **23**:327–341.
183. Elber R, Ruymgaart AP, Hess B: **SHAKE parallelization.** *Eur Phys J Spec Top* 2011, **200**:211–223.
184. Ye X, Cai Q, Yang W, Luo R: **Roles of boundary conditions in DNA simulations: analysis of ion distributions with the finite-difference Poisson-Boltzmann method.** *Biophys J* 2009, **97**:554–562.
185. Crowley M, Darden T, Iii TC, Ii DD: **Adventures in Improving the Scaling and Accuracy of a Parallel Molecular Dynamics Program.** *J Supercomput* 1997, **11**:255–278.
186. Darden T, York D, Pedersen L: **Particle mesh Ewald: An N log(N) method for Ewald sums in large systems.** *J Chem Phys* 1993, **98**:10089–10092.
187. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG: **A smooth particle mesh Ewald method.** *J Chem Phys* 1995, **103**:8577–8593.
188. Darden T, Perera L, Li L, Pedersen L: **New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations.** *Struct Lond Engl* 1993 1999, **7**:R55–R60.
189. Adler M, Beroza P: **Improved ligand binding energies derived from molecular dynamics: replicate sampling enhances the search of conformational space.** *J Chem Inf Model* 2013, **53**:2065–2072.
190. Mitsutake A, Sugita Y, Okamoto Y: **Generalized-ensemble algorithms for molecular simulations of biopolymers.** *Biopolymers* 2001, **60**:96–123.
191. Laio A, Parrinello M: **Escaping free-energy minima.** *Proc Natl Acad Sci U S A* 2002, **99**:12562–12566.
192. Tiwary P, Parrinello M: **From metadynamics to dynamics.** *Phys Rev Lett* 2013, **111**:230602.
193. Hukushima K, Nemoto K: **Exchange Monte Carlo Method and Application to Spin Glass Simulations.** *J Phys Soc Jpn* 1996, **65**:1604–1608.
194. Hukushima K, Takayama H, Nemoto K: **Application of an extended ensemble method to spin glasses.** *Int J Mod Phys C* 1996, **07**:337–344.
195. Chu W-T, Zhang J-L, Zheng Q-C, Chen L, Zhang H-X: **Insights into the folding and unfolding processes of wild-type and mutated SH3 domain by molecular dynamics and replica exchange molecular dynamics simulations.** *PLoS One* 2013, **8**:e64886.
196. Patra MC, Rath SN, Pradhan SK, Maharana J, De S: **Molecular dynamics simulation of human serum paraoxonase 1 in DPPC bilayer reveals a critical role of transmembrane helix H1 for HDL association.** *Eur Biophys J EBJ* 2014, **43**:35–51.
197. Harris RC, Boschitsch AH, Fenley MO: **Influence of grid spacing in Poisson-Boltzmann equation binding energy estimation.** *J Chem Theory Comput* 2013, **9**:3677–3685.
198. Reddy MR, Reddy CR, Rathore RS, Erion MD, Aparoy P, Reddy RN, Reddanna P: **Free Energy Calculations to Estimate Ligand-Binding Affinities in Structure-Based Drug Design.** *Curr Pharm Des* 2014, **20**:3323–3337.
199. Hou T, Wang J, Li Y, Wang W: **Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations.** *J Chem Inf Model* 2011, **51**:69–82.
200. Kar P, Lipowsky R, Knecht V: **Importance of polar solvation and configurational entropy for design of antiretroviral drugs targeting HIV-1 protease.** *J Phys Chem B* 2013, **117**:5793–5805.
201. Osterberg F, Aqvist J: **Exploring blocker binding to a homology model of the open hERG K⁺ channel using docking and molecular dynamics methods.** *FEBS Lett* 2005, **579**:2939–2944.
202. Jenwithesuk E, Samudrala R: **Virtual screening of HIV-1 protease inhibitors against human cytomegalovirus protease using docking and molecular dynamics.** *AIDS Lond Engl* 2005, **19**:529–531.
203. Tatsumi R, Fukunishi Y, Nakamura H: **A hybrid method of molecular dynamics and harmonic dynamics for docking of flexible ligand to flexible receptor.** *J Comput Chem* 2004, **25**:1995–2005.
204. Hu S, Yu H, Liu Y, Xue T, Zhang H: **Insight into the binding model of new antagonists of kappa receptor using docking and molecular dynamics simulation.** *J Mol Model* 2013, **19**:3087–3094.
205. Hu W, Deng S, Huang J, Lu Y, Le X, Zheng W: **Intercalative interaction of asymmetric copper(II) complex with DNA: experimental, molecular docking, molecular dynamics and TDFT studies.** *J Inorg Biochem* 2013, **127**:90–98.
206. Huang X, Zheng G, Zhan C-G: **Microscopic binding of M5 muscarinic acetylcholine receptor with antagonists by homology modeling, molecular docking, and molecular dynamics simulation.** *J Phys Chem B* 2012, **116**:532–541.
207. Moitessier N, Henry C, Maigret B, Chapleur Y: **Combining pharmacophore search, automated docking, and molecular dynamics simulations as a novel strategy for flexible docking. Proof of concept: docking of arginine – glycine – aspartic acid-like compounds into the $\alpha\text{v}\beta\text{3}$ binding site.** *J Med Chem* 2004, **47**:4178–4187.
208. NVIDIA: *Computational Chemistry & Biology Benchmark Report.* NVIDIA: Santa Clara, CA; 2013.
209. von R SP: *Encyclopedia of Computational Chemistry.* Chichester: New York: J. Wiley; 1998.

210. Tuttle T, Thiel W: **OMx-D: semiempirical methods with orthogonalization and dispersion corrections. Implementation and biochemical application.** *Phys Chem Chem Phys PCCP* 2008, **10**:2159–2166.
211. Yilmazer ND, Korth M: **Comparison of molecular mechanics, semi-empirical quantum mechanical, and density functional theory methods for scoring protein-ligand interactions.** *J Phys Chem B* 2013, **117**:8075–8084.
212. Voityuk AA: **Semi-empirical methods: current status and future directions.** In *Drug Des Strateg Comput Tech Appl*. Edited by Banting L, Clark T. Cambridge: Royal Society of Chemistry; 2012:107–119.
213. Koch W, Holthausen MC: *A Chemist's Guide to Density Functional Theory*. Weinheim, New York: Wiley-VCH; 2001.
214. Putz MV, Mingos DMP: *Applications of Density Functional Theory to Biological and Bioinorganic Chemistry*. Berlin, New York: Springer; 2013.
215. Grimme S: **Density functional theory with London dispersion corrections.** *Wiley Interdiscip Rev Comput Mol Sci* 2011, **1**:211–228.
216. Klimeš J, Michaelides A: **Perspective: Advances and challenges in treating van der Waals dispersion forces in density functional theory.** *J Chem Phys* 2012, **137**:120901.
217. Korth M: **Error estimates for (semi-)empirical dispersion terms and large biomacromolecules.** *Org Biomol Chem* 2013, **11**:6515–6519.
218. Antony J, Grimme S: **Fully ab initio protein-ligand interaction energies with dispersion corrected density functional theory.** *J Comput Chem* 2012, **33**:1730–1739.
219. Mucs D, Bryce RA: **The application of quantum mechanics in structure-based drug design.** *Expert Opin Drug Discov* 2013, **8**:263–276.
220. Hobza P: **Calculations on noncovalent interactions and databases of benchmark interaction energies.** *Acc Chem Res* 2012, **45**:663–672.
221. De Benedetti PG, Fanelli F: **Computational quantum chemistry and adaptive ligand modeling in mechanistic QSAR.** *Drug Discov Today* 2010, **15**:859–866.
222. Haag MP, Reiher M: **Real-time quantum chemistry.** *Int J Quantum Chem* 2013, **113**:8–20.
223. Asadchev A, Gordon MS: **New multithreaded hybrid CPU/GPU approach to Hartree–Fock.** *J Chem Theory Comput* 2012, **8**:4166–4176.
224. Olivares-Amaya R, Watson MA, Edgar RG, Vogt L, Shao Y, Aspuru-Guzik A: **Accelerating correlated quantum chemistry calculations using graphical processing units and a mixed precision matrix multiplication library.** *J Chem Theory Comput* 2010, **6**:135–144.
225. Asadchev A, Gordon MS: **Fast and flexible coupled cluster implementation.** *J Chem Theory Comput* 2013, **9**:3385–3392.
226. Maia JDC, Urquiza Carvalho GA, Mangueira CP, Santana SR, Cabral LAF, Rocha GB: **GPU linear algebra libraries and GPGPU programming for accelerating MOPAC semiempirical quantum chemistry calculations.** *J Chem Theory Comput* 2012, **8**:3072–3081.
227. Wu X, Koslowski A, Thiel W: **Semiempirical quantum chemical calculations accelerated on a hybrid multicore CPU–GPU computing platform.** *J Chem Theory Comput* 2012, **8**:2272–2281.
228. Yasuda K: **Accelerating density functional calculations with graphics processing unit.** *J Chem Theory Comput* 2008, **4**:1230–1236.
229. Ufimtsev IS, Martínez TJ: **Quantum chemistry on graphical processing units. 1. Strategies for two-electron integral evaluation.** *J Chem Theory Comput* 2008, **4**:222–231.
230. Ufimtsev IS, Martínez TJ: **Quantum chemistry on graphical processing units. 2. Direct self-consistent-field implementation.** *J Chem Theory Comput* 2009, **5**:1004–1015.
231. Ufimtsev IS, Martínez TJ: **Quantum chemistry on graphical processing units. 3. Analytical energy gradients, geometry optimization, and first principles molecular dynamics.** *J Chem Theory Comput* 2009, **5**:2619–2628.
232. Luehr N, Ufimtsev IS, Martínez TJ: **Dynamic precision for electron repulsion integral evaluation on graphical processing units (GPUs).** *J Chem Theory Comput* 2011, **7**:949–954.
233. Kulik HJ, Luehr N, Ufimtsev IS, Martínez TJ: **Ab initio quantum chemistry for protein structures.** *J Phys Chem B* 2012, **116**:12501–12509.
234. Titov AV, Ufimtsev IS, Luehr N, Martínez TJ: **Generating efficient quantum chemistry codes for novel architectures.** *J Chem Theory Comput* 2013, **9**:213–221.
235. Ufimtsev IS, Luehr N, Martínez TJ: **Charge transfer and polarization in solvated proteins from ab Initio molecular dynamics.** *J Phys Chem Lett* 2011, **2**:1789–1793.
236. Mori T, Martínez TJ: **Exploring the conical intersection seam: the seam space nudged elastic band method.** *J Chem Theory Comput* 2013, **9**:1155–1163.
237. Kussmann J, Beer M, Ochsenfeld C: **Linear-scaling self-consistent field methods for large molecules.** *Wiley Interdiscip Rev Comput Mol Sci* 2013, **3**:614–636.
238. Leszczynski J: *Linear-Scaling Techniques in Computational Chemistry and Physics: Methods and Applications*. New York: Springer; 2011.
239. Gordon MS, Fedorov DG, Pruiett SR, Slipchenko LV: **Fragmentation methods: a route to accurate calculations on large systems.** *Chem Rev* 2012, **112**:632–672.
240. Alexeev Y, Mazanetz MP, Ichihara O, Fedorov DG: **GAMESS as a free quantum-mechanical platform for drug research.** *Curr Top Med Chem* 2012, **12**:2013–2033.
241. Day PN, Jensen JH, Gordon MS, Webb SP, Stevens WJ, Krauss M, Garmer D, Basch H, Cohen D: **An effective fragment method for modeling solvent effects in quantum mechanical calculations.** *J Chem Phys* 1996, **105**:1968–1986.
242. Gordon MS, Freitag MA, Bandyopadhyay P, Jensen JH, Kairys V, Stevens WJ: **The effective fragment potential method: a QM-based MM approach to modeling environmental effects in chemistry.** *J Phys Chem A* 2001, **105**:293–307.
243. Gordon MS, Slipchenko L, Li H, Jensen JH: **The effective fragment potential: a general method for predicting intermolecular interactions.** In *Annu Rep Comput Chem*, Volume 3. Edited by Spellmeyer DC, Wheeler R. Oxford: Elsevier Science; 2007:177–193.

244. Fedorov DG, Kitaura K: **The importance of three-body terms in the fragment molecular orbital method.** *J Chem Phys* 2004, **120**:6832–6840.
245. Imamura A, Aoki Y, Maekawa K: **A theoretical synthesis of polymers by using uniform localization of molecular orbitals: Proposal of an elongation method.** *J Chem Phys* 1991, **95**:5419–5431.
246. Kobayashi M, Nakai H: *Divide-and-Conquer Approaches to Quantum Chemistry: Theory and Implementation*; 2011.
247. Fedorov DG, Kitaura K: **Extending the power of quantum chemistry to large systems with the fragment molecular orbital method.** *J Phys Chem A* 2007, **111**:6904–6914.
248. Fedorov D, Kitaura K: *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*. Boca Raton: CRC Press/Taylor & Francis; 2009.
249. Fedorov DG, Nagata T, Kitaura K: **Exploring chemistry with the fragment molecular orbital method.** *Phys Chem Chem Phys* PCCP 2012, **14**:7562–7577.
250. Munro AW, Scrutton NS: **Enzyme mechanisms: fast reaction and computational approaches.** *Biochem Soc Trans* 2009, **37**(Pt 2):333–335.
251. Scheiner S, Lipscomb WN: **Molecular orbital studies of enzyme activity: catalytic mechanism of serine proteinases.** *Proc Natl Acad Sci U S A* 1976, **73**:432–436.
252. Warshel A, Levitt M: **Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonion ion in the reaction of lysozyme.** *J Mol Biol* 1976, **103**:227–249.
253. Kamerlin SCL, Haranczyk M, Warshel A: **Progress in ab initio QM/MM free-energy simulations of electrostatic energies in proteins: accelerated QM/MM studies of pKa, redox reactions and solvation free energies.** *J Phys Chem B* 2009, **113**:1253–1272.
254. Mulholland AJ, Grant GH, Richards WG: **Computer modelling of enzyme catalysed reaction mechanisms.** *Protein Eng* 1993, **6**:133–147.
255. Mulholland AJ, Karplus M: **Simulations of enzymic reactions.** *Biochem Soc Trans* 1996, **24**:247–254.
256. Friesner RA, Guallar V: **Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis.** *Annu Rev Phys Chem* 2005, **56**:389–427.
257. Senn HM, Thiel W: **QM/MM methods for biomolecular systems.** *Angew Chem Int Ed Engl* 2009, **48**:1198–1229.
258. Lin H, Truhlar DG: **QM/MM: what have we learned, where are we, and where do we go from here?** *Theor Chem Acc* 2007, **117**:185–199.
259. Lonsdale R, Ranaghan KE, Mulholland AJ: **Computational enzymology.** *Chem Commun Camb Engl* 2010, **46**:2354–2372.
260. Ridder L, Harvey JN, Rietjens IMCM, Vervoort J, Mulholland AJ: **Ab Initio QM/MM Modeling of the Hydroxylation Step in p-Hydroxybenzoate Hydroxylase.** *J Phys Chem B* 2003, **107**:2118–2126.
261. Ridder L, Mulholland AJ: **Modeling biotransformation reactions by combined quantum mechanical/molecular mechanical approaches: from structure to activity.** *Curr Top Med Chem* 2003, **3**:1241–1256.
262. Ranaghan KE, Mulholland AJ: **Conformational effects in enzyme catalysis: QM/MM free energy calculation of the 'NAC' contribution in chorismate mutase.** *Chem Commun* 2004, (10):1238–1239.
263. Martí S, Roca M, Andrés J, Moliner V, Silla E, Tuñón I, Bertrán J: **Theoretical insights in enzyme catalysis.** *Chem Soc Rev* 2004, **33**:98–107.
264. Warshel A, Sharma PK, Kato M, Xiang Y, Liu H, Olsson MHM: **Electrostatic basis for enzyme catalysis.** *Chem Rev* 2006, **106**:3210–3235.
265. Claeysens F, Harvey JN, Manby FR, Mata RA, Mulholland AJ, Ranaghan KE, Schütz M, Thiel S, Thiel W, Werner H-J: **High-accuracy computation of reaction barriers in enzymes.** *Angew Chem Int Ed Engl* 2006, **45**:6856–6859.
266. Woodcock HL, Hodošček M, Sherwood P, Lee YS, Iii HFS, Brooks BR: **Exploring the quantum mechanical/molecular mechanical replica path method: a pathway optimization of the chorismate to prephenate Claisen rearrangement catalyzed by chorismate mutase.** *Theor Chem Acc* 2003, **109**:140–148.
267. Van der Kamp MW, Zurek J, Manby FR, Harvey JN, Mulholland AJ: **Testing high-level QM/MM methods for modeling enzyme reactions: acetyl-CoA deprotonation in citrate synthase.** *J Phys Chem B* 2010, **114**:11303–11314.
268. Van der Kamp MW, Mulholland AJ: **Computational enzymology: insight into biological catalysts from modelling.** *Nat Prod Rep* 2008, **25**:1001–1014.
269. Mulholland AJ, Lyne PD, Karplus M: **Ab initio QM/MM study of the citrate synthase mechanism. A low-barrier hydrogen bond is not involved.** *J Am Chem Soc* 2000, **122**:534–535.
270. Garcia-Viloca M, Gao J, Karplus M, Truhlar DG: **How enzymes work: analysis by modern rate theory and computer simulations.** *Science* 2004, **303**:186–195.
271. Cui Q, Karplus M: **Quantum mechanics/molecular mechanics studies of triosephosphate isomerase-catalyzed reactions: effect of geometry and tunneling on proton-transfer rate constants.** *J Am Chem Soc* 2002, **124**:3093–3124.
272. Bjelic S, Aqvist J: **Computational prediction of structure, substrate binding mode, mechanism, and rate for a malaria protease with a novel type of active site.** *Biochemistry (Mosc)* 2004, **43**:14521–14528.
273. Van der Kamp MW, Shaw KE, Woods CJ, Mulholland AJ: **Biomolecular simulation and modelling: status, progress and prospects.** *J R Soc Interface R Soc* 2008, **5**(Suppl 3):S173–S190.
274. Warshel A: **Computer simulations of enzyme catalysis: methods, progress, and insights.** *Annu Rev Biophys Biomol Struct* 2003, **32**:425–443.
275. Bentzien J, Muller RP, Florián J, Warshel A: **Hybrid ab initio quantum mechanics/molecular mechanics calculations of free energy surfaces for enzymatic reactions: the nucleophilic attack in subtilisin.** *J Phys Chem B* 1998, **102**:2293–2301.
276. Warshel A: *Computer Modeling of Chemical Reactions in Enzymes and Solutions*. New York: Wiley; 1997.
277. Villà J, Warshel A: **Energetics and dynamics of enzymatic reactions.** *J Phys Chem B* 2001, **105**:7887–7907.
278. Warshel A, Sharma PK, Chu ZT, Aqvist J: **Electrostatic contributions to binding of transition state analogues can be very different from the corresponding contributions to catalysis: phenolates binding to the oxyanion hole of ketosteroid isomerase.** *Biochemistry (Mosc)* 2007, **46**:1466–1476.
279. Bjelic S, Aqvist J: **Catalysis and linear free energy relationships in aspartic proteases.** *Biochemistry (Mosc)* 2006, **45**:7709–7723.

280. Trobro S, Aqvist J: Mechanism of peptide bond synthesis on the ribosome. *Proc Natl Acad Sci U S A* 2005, **102**:12395–12400.
281. Sharma PK, Xiang Y, Kato M, Warshel A: What are the roles of substrate-assisted catalysis and proximity effects in peptide bond formation by the ribosome? *Biochemistry (Mosc)* 2005, **44**:11307–11314.
282. Hammes-Schiffer S: Quantum-classical simulation methods for hydrogen transfer in enzymes: a case study of dihydrofolate reductase. *Curr Opin Struct Biol* 2004, **14**:192–201.
283. Liu H, Warshel A: Origin of the temperature dependence of isotope effects in enzymatic reactions: the case of dihydrofolate reductase. *J Phys Chem B* 2007, **111**:7852–7861.
284. Sharma PK, Chu ZT, Olsson MHM, Warshel A: A new paradigm for electrostatic catalysis of radical reactions in vitamin B12 enzymes. *Proc Natl Acad Sci U S A* 2007, **104**:9661–9666.
285. Gao J, Truhlar DG: Quantum mechanical methods for enzyme kinetics. *Annu Rev Phys Chem* 2002, **53**:467–505.
286. Ranaghan KE, Mulholland AJ: Computer simulations of quantum tunnelling in enzyme-catalysed hydrogen transfer reactions. *Interdiscip Sci Comput Life Sci* 2010, **2**:78–97.
287. Lonsdale R, Harvey JN, Mulholland AJ: Inclusion of dispersion effects significantly improves accuracy of calculated reaction barriers for cytochrome P450 catalyzed reactions. *J Phys Chem Lett* 2010, **1**:3232–3237.
288. Himo F, Siegbahn PEM: Quantum chemical studies of radical-containing enzymes. *Chem Rev* 2003, **103**:2421–2456.
289. Siegbahn PEM, Himo F: Recent developments of the quantum chemical cluster approach for modeling enzyme reactions. *J Biol Inorg Chem JBIC Publ Soc Biol Inorg Chem* 2009, **14**:643–651.
290. Mackerell AD: Empirical force fields for biological macromolecules: Overview and issues. *J Comput Chem* 2004, **25**:1584–1604.
291. Vreven T, Byun KS, Komáromi I, Dapprich S, Montgomery JA, Morokuma K, Frisch MJ: Combining quantum mechanics methods with molecular mechanics methods in ONIOM. *J Chem Theory Comput* 2006, **2**:815–826.
292. Zhang Y: Pseudobond ab initio QM/MM approach and its applications to enzyme reactions. *Theor Chem Acc* 2006, **116**:43–50.
293. Meier K, Thiel W, van Gunsteren WF: On the effect of a variation of the force field, spatial boundary condition and size of the QM region in QM/MM MD simulations. *J Comput Chem* 2012, **33**:363–378.
294. Riccardi D, Schaefer P, Cui Q: pKa calculations in solution and proteins with QM/MM free energy perturbation simulations: a quantitative test of QM/MM protocols. *J Phys Chem B* 2005, **109**:17715–17733.
295. Rodríguez A, Oliva C, González M, van der Kamp M, Mulholland AJ: Comparison of different quantum mechanical/molecular mechanics boundary treatments in the reaction of the hepatitis C virus NS3 protease with the NS5A/5B substrate. *J Phys Chem B* 2007, **111**:12909–12915.
296. Sumowski CV, Ochsenfeld C: A convergence study of QM/MM isomerization energies with the selected size of the QM region for peptidic systems. *J Phys Chem A* 2009, **113**:11734–11741.
297. Pezeshki S, Lin H: Adaptive-partitioning redistributed charge and dipole schemes for QM/MM dynamics simulations: on-the-fly relocation of boundaries that pass through covalent bonds. *J Chem Theory Comput* 2011, **7**:3625–3634.
298. Lodola A, Mor M, Zurek J, Tarzia G, Piomelli D, Harvey JN, Mulholland AJ: Conformational effects in enzyme catalysis: reaction via a high energy conformation in fatty acid amide hydrolase. *Biophys J* 2007, **92**:L20–L22.

doi:10.1186/1742-4682-11-52

Cite this article as: Tuszynski et al.: Mathematical and computational modeling in biology at multiple scales. *Theoretical Biology and Medical Modelling* 2014 **11**:52.

Submit your next manuscript to BioMed Central
and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

