Biochimica et Biophysica Acta 1836 (2013) 1–14

Contents lists available at SciVerse ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbacan

Review

Molecular modelling and simulations in cancer research

Ran Friedman [a,e,*], Kjetil Boye [b,c], Kjersti Flatmark [b,d]

[a] Computational Chemistry and Biochemistry Group, School of Natural Sciences, Linnæus University, 391 82 Kalmar, Sweden
[b] Department of Tumor Biology, Institute for Cancer Research, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway
[c] Department of Oncology, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway
[d] Department of Gastroenterological Surgery, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway
[e] Linnæus University, Biomaterials Chemistry Centre of Excellence, 391 82 Kalmar, Sweden

ARTICLE INFO

Article history:
Received 10 December 2012
Received in revised form 4 February 2013
Accepted 5 February 2013
Available online 14 February 2013

Keywords:
Molecular dynamics
AKR1B10
Tubulin polymerase
Pin1
ERK2 L-asparaginase
p53

ABSTRACT

The complexity of cancer and the vast amount of experimental data available have made computer-aided approaches necessary. Biomolecular modelling techniques are becoming increasingly easier to use, whereas hardware and software are becoming better and cheaper. Cross-talk between theoretical and experimental scientists dealing with cancer-research from a molecular approach, however, is still uncommon. This is in contrast to other fields, such as amyloid-related diseases, where molecular modelling studies are widely acknowledged. The aim of this review paper is therefore to expose some of the more common approaches in molecular modelling to cancer scientists in simple terms, illustrating success stories while also revealing the limitations of computational studies at the molecular level.

Contents

Abbreviations: AIMD, ab initio molecular dynamics; AEP, asparagine endopeptidase; ALK, anaplastic lymphoma kinase; BD, Brownian dynamics; CPMD, Car–Parrinello molecular dynamics; DFT, density functional theory; DPD, dissipative particle dynamics; ENM, elastic network model; ERK2, extracellular signal-regulated kinase 2; EXAFS, extended X-ray absorption fine structure; FRET, Förster resonance energy transfer; L-ASN, L-asparaginase; MD, molecular dynamics; NMA, normal mode analysis; QM, quantum mechanics; QM/MM, quantum mechanics/molecular mechanics; SAXS, small angle X-ray scattering; YAS, yttrium-alumino-silicate

* Corresponding author at: Linnæus University, Biomaterials Chemistry Centre of Excellence, 391 82 Kalmar, Sweden. Tel.: +46 480446290.
E-mail address: ran.friedman@lnu.se (R. Friedman).

0304-419X/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.bbcan.2013.02.001

1. Introduction

A large number of proteins have been shown to be involved in various types of cancer, from the ubiquitous tumour suppressor p53 to metastasis promoters such as S100A4. Accordingly, whereas drug design was once aimed primarily at the tumour level, many newer drugs are protein-specific. Such drugs often have less severe side effects, because they selectively aim at proteins that are related to cancer progression. Moreover, owing to the availability of affordable sequencing techniques and plethora of genetic markers, personalised treatments are beginning to be available [1]. On the other hand, sufficient knowledge about the biology of many cancer-related proteins and mutations is still missing. Very often, a marker of a certain cancer is identified but it is not clear how it is related to the specific disease.

Many details on a given protein's function(s), and how it is altered by different mutations, become clear once the protein's structure is known. Proteins, however, are not static and it is their dynamic landscape that determines their roles [2]. Nuclear magnetic resonance (NMR), Förster resonance energy transfer (FRET), X-ray Laue diffraction, extended X-ray absorption fine structure (EXAFS) and other biophysical methods can shed light on the dynamics of macromolecules, but suffer from some limitations in terms of sensitivity, applicability, and time scales. Computer-aided studies can complement molecular studies and yield details that are not available to the experiment. Molecular modelling approaches therefore become increasingly useful in many clinically-oriented studies, e.g., amyloid related diseases [3]. Such methods, however, are sometimes inaccessible to the cancer researcher owing to lack of understanding on their potential use, advantages and limitations.

This review article deals with molecular modelling in cancer research. It aims to acquaint the reader with molecular modelling and some of the most-commonly used methods in the field, in simple terms that would be accessible to non-experts. The scope is limited to molecular modelling methodologies, which are a relatively small part of computational biology. Bioinformatics and mathematical approaches to cancer research are extensively covered in the literature [4–7], and will not be discussed here.

This article is organised as follows. In the next section, some of the most relevant methods of molecular modelling are explained, with an emphasis on methods that have been used or can be of use in cancer research. The aim is to acquaint the reader with the molecular modelling methods and the jargon, rather than explain the governing mathematical and physical principles; details on the implementation can be found in the molecular modelling literature [8,9]. The third section describes a few interesting applications of the methods in cancer-related studies. In the fourth section, we give our view on the exertion of molecular modelling in cancer research and suggest how it can be routinely applied in molecular studies related to cancer. While the methods may be of use in many studies related to molecular biology and medicine, this article provides added insights with respect to use of the molecular modelling techniques within cancer research.

2. Methods of molecular modelling and simulation

"Molecular modelling" has gone a long way since its early days. Most of the readers are probably familiar with the plastic balls and sticks that connect them, which were used in chemistry classes to teach how molecules are made from atoms. Such models had also been used in research [10], but were replaced with computer-generated models. Today, "molecular modelling" refers to the application of computer-generated models in molecular studies ranging from a few atoms to multitude of biomolecules. These models are used to simulate processes that may be as fast as 10^{-15} s or as slow as a few seconds. Clearly, the accuracy and level of detail depend on the size and timescale of the system. Sub-Ångström differences between structures can be studied and have a large influence on the binding of a drug molecule to its receptor, whereas the size of protein complexes is three orders of magnitude larger. These differences necessitate the use of different methods. In general, the more accurate the method at hand, the more time and computational resources will be needed to get a meaningful result (for systems of similar size). The choice of the modelling method is therefore made based upon the problem at hand.

Although it is possible to model a collection of molecules, such as multiple peptides and lipids [11], most molecular modelling studies related to cancer research are carried out in atomistic details. These methods, however, are limited in scope to molecules for which structural data are available from experiments (X-ray crystallography or NMR) or can be accurately modelled (see below). Atomistic modelling is also limited in size and timescale. Soluble proteins can nowadays be studied for 10^{-7}–10^{-6} s, enough to shed light on some domain motions within a protein but not on large conformational changes such as protein folding or activation of channels. Longer processes and larger complexes can still be modelled, but necessitate sophisticated and more approximate methods. Several common molecular modelling approaches and applications are discussed in this section. The principles are explained in simple terms, and few examples of applications are given. In discussing the methods, we (somewhat artificially) group them into four categories: atomistic simulation and modelling methods, modelling of proteins and protein complexes, modelling of protein–drug interactions, and simplified approaches (that do not yield atomistic details). There is some overlap between the categories, and some methods belong to more than one category, in which case we have put the method where we feel it is (or has the potential to be) most frequently used in cancer research. The description of each method begins with an executive summary of its strengths, limitations and general applicability, and is followed by a more detailed (but still short) description. A graphical presentation of some of the most commonly used methods is given in Fig. 1. Several cancer-related studies involving these methods will be

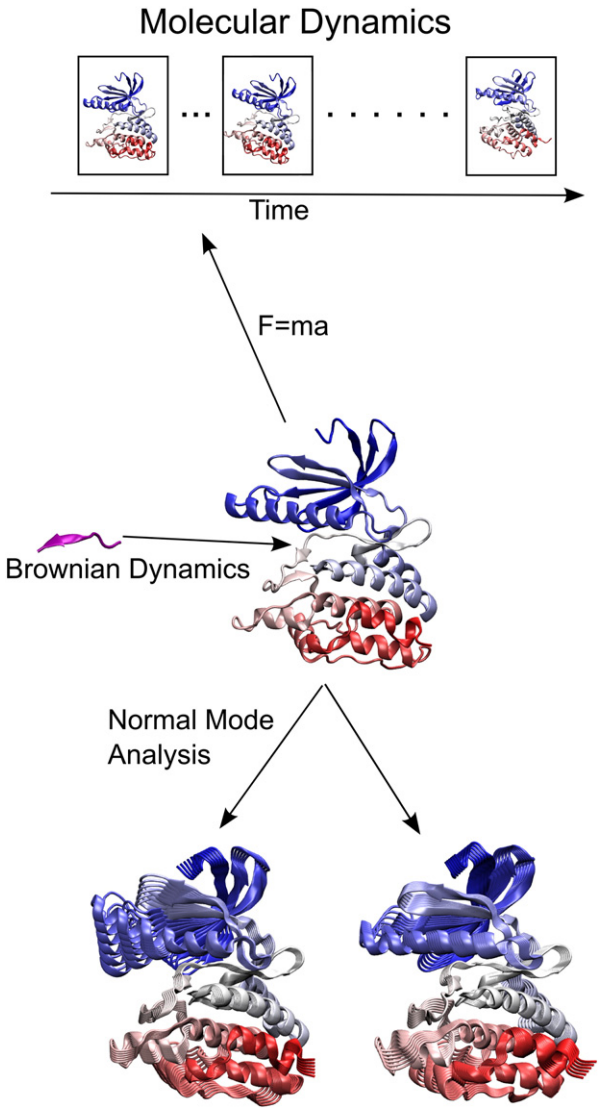

Fig. 1. Illustration of commonly used modelling and simulation methods. The dynamics of a protein can be studied by employing the Newton equations of motion, to yield an ensemble of structures that can be viewed as a movie (molecular dynamics). Brownian dynamics simulations can be used to study the formation of macromolecular complexes. Finally, normal mode analysis can yield insights into motions of protein domains. All protein figures in this article were prepared using VMD [12].

covered in the following section, and shall provide more light on successful applications of molecular modelling in cancer research.

2.1. Atomistic simulation and modelling methods

Molecular simulation methods deal with dynamic processes that are often difficult to follow experimentally. The applications usually deal with biomacromolecules, although some of the methods can be extended to larger systems. All of the methods mentioned here require structural information at atomic resolution when dealing with macromolecules. Ideally, this information should come from high-resolution protein crystallography or NMR, but models (see below, modelling of protein structures) can also be used.

2.1.1. Atomistic molecular dynamics simulations

2.1.1.1. Strengths. Relatively easy to use, yields dynamics data on the motions of atoms and molecules in atomistic details — almost as if you are watching a protein in an animation movie. Many structural observables can be easily extracted from the simulation. Protein modelling is often as realistic as it can get.

2.1.1.2. Limitations. Much less accurate when experimental data on the biomolecule in question are not available. Cannot be used to study large-scale or long-term processes without extensions (and at reduced accuracy). Cannot be used as is to study processes that involve breaking or formation of covalent bonds.

2.1.1.3. Example application. The structure of a protein bound to a drug molecule is known, and you want to understand whether a certain mutation may lead to drug resistance because it would reduce the affinity of the drug molecule.

Molecular dynamics (MD) simulations are widely used to address biological systems. The method is based on the centuries old Newton's equations of motion, namely $\mathbf{F}=m\mathbf{a}$, where **F** is the force operating on a particle, m is its mass and **a** its acceleration. Starting from a given coordination of the molecule(s), for example a crystallographic structure of an enzyme-inhibitor complex immersed in water, the forces that act on the atoms are calculated, and the system is propagated in time according to Newton's equations of motion. The force is calculated by differentiating the potential energy at the positions of every atom in the system. This potential energy will depend on all other atoms present and on their locations.

In practice, the potential energy of the system is estimated based on a set of equations and parameters, collectively known as a *force field*. Relatively short time steps, usually of 1–4 *femtoseconds* (fs, 10^{-15} *s*) are used in simulations at the atomic scale (i.e., where a molecule is represented as a collection of atoms). Accordingly, a simulation of 0.1 μs typically requires 50 million steps. An illustration of the method is given in Fig. 2.

The output of an MD simulation is a set of particle velocities and coordinates, which is termed *trajectory*. Free and commercial programs are available for viewing trajectory files. Using such programs, the researcher can watch the simulation just like a molecular movie, zooming in and out or following on the frames that appear to be more interesting. The simplicity of the approach and the availability of many different MD engines (software than can run MD, see Table A.1 in the Appendix A) and molecular viewers (Appendix Table A.2) are two of the most compelling reasons for the method's popularity. The trajectory can also be used for carrying out calculations that yield physical observables of a system, such as the prevalence of hydrogen bonds, distance between atoms or residues, modification of secondary structures and volumes of binding sites. Such observables often yield a biological insight.

What are the uses of MD simulations in a biological context? As a complete answer to this question cannot be given, a few recent examples follow. Transporter proteins are particularly amendable to MD

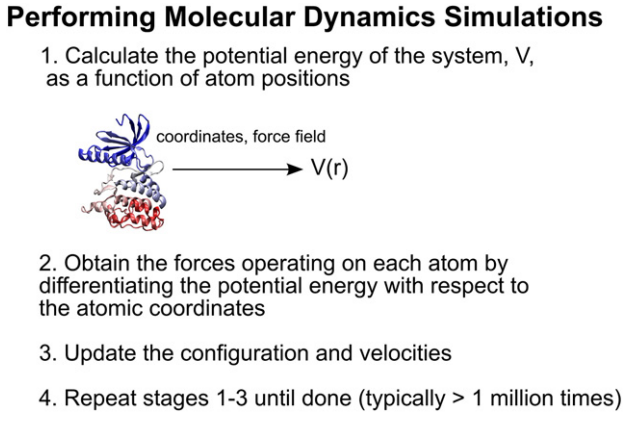

Fig. 2. Performing molecular dynamics simulations.

simulations, because structural and mutational data can only yield limited insights about the dynamics of the process (How is the molecule actually being transferred? Which residues play a major role? Why is a certain mutation important even when the residue is not involved in contacts with the substrates according to the known structural information?). In the field of computer-aided drug design, MD simulations can infer on correct and incorrect modes of binding of a molecule [13] and the energetics of the binding reaction, e.g., the interplay between electrostatic and hydrophobic forces. Post translational modification can also be studied, through a comparison of the modified and non-modified proteins; a similar comparison can be used to infer on the mechanism of action of a drug molecule.

The many strengths of MD should not obscure its limitations. First, out-of-the-box simulations, using established protocols, are generally limited to soluble proteins of known structures that are not modified or bound to any cofactors of inhibitors. Dealing with multivalent metal ions, cofactors, membranes or drug molecules requires additional expertise and often specific protocols or parameters must be devised. Second, MD simulations are limited to time-scales of less than 1 ms. Conformational changes involving whole protein domains necessitate the use of more elaborate or approximate methods or specialised hardware [14] and may suffer from numerical inaccuracies [15,16]. Third, MD simulations use an approximate function to calculate the potential energy, that fails to properly describe the system in some cases (but generally works well for biomolecules under ambient temperature and pressure). Finally, canonical MD simulations that employ interaction potentials that are based on classical mechanics are not suitable to follow on reactions that involve the breaking or formation of covalent bonds. Some of these limitations can be overcome by using related techniques, such as simulations based on quantum mechanics (QM/MM and AIMD) or using simplified models (coarse-grained MD), which are described below. Coarse-grained MD simulations on the one hand and QM/MM studies on the other are becoming increasingly popular and challenge the dominance of atomistic MD based on classical (=non-quantum) mechanics.

2.1.2. Brownian dynamics

2.1.2.1. Strengths. Physics-based yet fast to use and can deal with large biomolecular complexes.

2.1.2.2. Limitations. The molecules are treated as rigid bodies, i.e., no information on the internal dynamics of the molecules can be simulated or yielded.

2.1.2.3. Example application. You know the structures of two proteins and have some information of their complex, e.g., from cross-linking experiments. Brownian dynamics can then be used to follow on complex formation, for example when more than one binding mode is possible.

Brownian dynamics (BD) simulations in biomedicine are most often used to follow on the formation of complexes, e.g., between two or more proteins. Such systems are too large and evolve too slowly to be studied directly by MD, especially when several binding modes are possible. Therefore, more approximate methods must be used to infer how the complex is formed. MD simulations can be used at a later stage to discriminate between several potential complexes.

In the context of biomolecular BD simulations, the forces are extracted from the electrostatic potentials that surround the particles. Brownian dynamics simulations can yield trajectories and rates of encounter between the components thus allowing a comparison between different complexes. Applications are by no means limited to protein–protein interactions, and BD have been applied also to simulate the movement of ions in membrane channels [17], study enzyme–ligand association [18], and model the binding of proteins and DNA [19], to give a few examples. Several MD simulation packages can be used for BD simulations as well, but specialised software also exist (Table A.4).

2.1.3. Modelling based on quantum chemistry

2.1.3.1. Strengths. Accuracy and scope — can deal with processes that involve covalent bond making and breaking or excited states of molecules that cannot be studied by more approximate methods.

2.1.3.2. Limitations. Slow, difficult to master, available only for short timescales and small systems.

2.1.3.3. Example application. You want to understand how a certain carcinogen interacts with nucleic acid bases.

Quantum chemistry utilises quantum mechanical (QM) principles for chemical calculations. Thus, it can offer a better accuracy compared with molecular dynamics and other methods that rely on energy functions that do not involve quantum-mechanics, and can deal with processes involving bond-breaking and formation. Unfortunately, the computational cost of employing quantum chemistry programs can be prohibitive. Moreover, the underlying principles (quantum mechanics) may be non-intuitive and difficult to grasp by non-experts. To make things even more complicated, quantum chemists use a highly specific jargon that makes it even more difficult to follow the relevant publications compared with other molecular modelling methods.

Due to the high computational cost, QM calculations are very rarely performed on full-size proteins, although this situation is likely to change because methods to deal with large molecules, e.g., [20,21], are being actively developed. Instead, a model of the system of interest (e.g., a catalytic site of an enzyme) is constructed by taking into account a subset of atoms that directly participate in the reaction. The rest of the system is either ignored or approximated by use of faster methods (e.g., quantum mechanics/molecular mechanics, or QM/MM). Some of the limitations of traditional QM methods are alleviated by the use of the popular density functional theory (DFT). Developed by Hohenberg, Kohn and Sham in the 1960s [22,23], the theory has since then transformed into a field in itself. Today, it is commonly applied in a biological context. In connection with cancer research, quantum chemistry has been used to reveal how carcinogens interact with the DNA [24], for the design of chemotherapies [25], to study the mechanism of histone deacetylation [26], and in many additional applications.

2.1.4. Ab initio molecular dynamics

2.1.4.1. Strengths. Modelling dynamics processes as molecular dynamics with the accuracy offered by quantum chemistry.

2.1.4.2. Limitations. Limited to very small systems (tens of atoms) and very short time scales (10^{-11} s).

2.1.4.3. Example application. Following on an enzymatic reaction using a model system based on the active-site structure.

Ab initio MD (AIMD) methods aim to bridge molecular dynamics with quantum mechanics. Using these methods, a system is simulated by Newtonian mechanics, but the forces that operate on the atoms are calculated from quantum-mechanical principles. The advantage in accuracy is countered by the complexity of ab initio MD, rendering the method much less widely used for dealing with biological macromolecules. It is both more demanding from a computational aspect and more challenging to employ from a scientific point of view (i.e., it requires expertise and a deeper understanding). Nevertheless, ab initio molecular dynamics simulations have been used in the field of cancer research (see below for an example discussing radiotherapy). Car–Parrinello molecular dynamics (CPMD) [27] is a popular implementation of ab initio molecular dynamics. Path-integral molecular

dynamics [28] is another implementation, that enables the study of nuclear tunnelling as occurs e.g., in some enzymes [29].

2.1.5. Quantum mechanics/molecular mechanics

2.1.5.1. Strengths. Enables the application of quantum mechanics-based approaches to atomistically large systems such as proteins.

2.1.5.2. Limitations. As slow as quantum chemistry.

2.1.5.3. Example application. Following on an enzymatic reaction, taking into account all of the protein residues, not just the catalytic site.

Quantum mechanics/molecular mechanics (QM/MM) methods treat part of the system (e.g., the catalytic site of an enzyme and a substrate) quantum-mechanically and the rest in more approximate terms, usually similar to normal MD simulations (see Fig. 3 for an illustration of the scale of QM/MM and QM methods). Thus, macromolecules can be studied even if the reaction involves bond-making and breaking. The calculation is about as fast (or as slow) as the QM calculation of the small area of interest. From a practical point of view, employing QM/MM requires expertise in both methods and in their coupling. Nevertheless, QM/MM based methods are broadly used in biology. Some applications relevant to cancer research include modelling of drugs that covalently bind DNA [30], and the mechanism of mitogen-activated protein kinase [31].

2.1.6. Normal mode analysis

2.1.6.1. Strengths. Simple, fast and relatively easy to do — a quick way to get a glimpse on the dynamics at atomic resolution without running a simulation.

2.1.6.2. Limitations. No correlation to timescales, not as broad and accurate as MD.

2.1.6.3. Example application. You study a large protein with several domains and want to learn how the domains move one with respect to the other, because this is relevant to that protein's function.

In normal mode analysis the system (e.g., a protein or a biomolecular complex) is assumed to be in equilibrium, while it still undergoes some motions. These motions are modelled as harmonic (similar to a bead on a spring). Low frequency modes represent collective motions that involve large parts of the system (e.g., two protein domains moving towards and away from each other), whereas high frequency (fast) modes correspond to local fluctuations that are usually of smaller interest. Applying NMA to proteins is simpler and faster than running a molecular dynamics simulation and has the advantage that the global motions (that are usually more relevant from a physiological standpoint) are immediately separated from local deformations, something that requires an additional analysis for molecular dynamics trajectories. There are, however, several drawbacks of the method. First, motions of proteins are not entirely harmonic [32]. Second, the system needs a proper preparation before performing normal mode analysis rigorously (its potential energy must be brought very close to minimum), which is often a difficult task for systems such as proteins with thousands of degrees of freedom. Similar to MD, one does not have to use atomistic potentials in NMA, and simplified models often yield meaningful results (see also the section on Elastic network models below). Yet, it is included here under atomistic modelling and simulation methods, since the computational cost of atomistic NMA of proteins is usually not prohibitive.

NMA can have several applications related to cancer research, including a comparison between wild-type and mutant kinases [33] and modifications of the protein dynamics upon binding of other molecules (target proteins, DNA or inhibitors). In practice, normal mode analysis is performed by use of the same programs as used for carrying out molecular dynamics simulations (Appendix: Table A.1). In addition, several web servers can be used for NMA (Table A.5).

2.2. Modelling of structures of proteins and protein complexes

The protein data bank (PDB) contains about 87,700 biomolecular structures (January 2013), of which 97% include proteins, and almost all of the rest are nucleic acids. Many of these are redundant, and atomic-resolution structures of clinically important proteins are still missing (note that selecting only structures with sequence identity of 90% or less, the PDB contains only about 33,700 entries). Whereas some of the missing proteins are difficult to crystallise or resolve, others include intrinsically disordered regions rather than a defined three dimensional structure [34]. Moreover, proteins and peptides often alter their structure upon forming macromolecular complexes. Therefore, even if the individual structures are known, it may still be necessary to compute the structure of a complex. Fortunately, powerful methods for protein structure prediction exist (see below

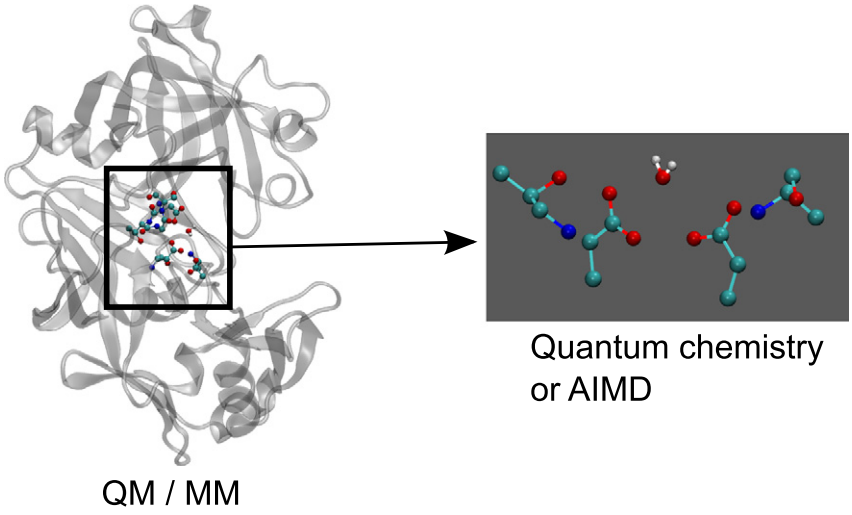

Fig. 3. The scale of QM/MM and QM methods. In QM/MM simulations, a small region, such as enclosed in a rectangle, is simulated at an accurate QM level and the other part of the macromolecule is approximated (left). Fully QM simulations such as ab-initio MD study only a small part of the macromolecule (right). The protein displayed here is cathepsin D, an aspartic proteases that is associated with poor prognosis in breast cancer [61]. The catalytic site of aspartic proteases was studied at a full QM level, revealing some organisational principles of these enzymes [62].

and Fig. 4), and are increasingly used to shed light on the conformations of proteins and biomolecular complexes.

2.2.1. Homology modelling

2.2.1.1. Strengths. Structure modelling based on evolutionary principles.

2.2.1.2. Limitations. Can only work when a structure of a related protein is known.

2.2.1.3. Example application. You want to simulate a human protein, whose structure is not available, based on the structure of the same (or a similar) protein but from another organism.

Homology modelling (also known as comparative or template-based modelling) is based on the notion that evolutionary conservation in sequence is correlated with conservation in structure. In other words, proteins that are similar in sequence are also similar in structure (whereas the latter is not always true). Thus, given a template protein with a known structure, which is somewhat similar in sequence to another protein whose structure is sought after, a computer program is used to thread the sequence of the protein with the unknown structure onto the template. Modern homology modelling software enables the use of multiple templates (for example, where a protein has two domains and each has a different homologue). The higher the sequence identity between a modelled protein and a template, the higher the accuracy of the model. Although many factors can contribute to the success of

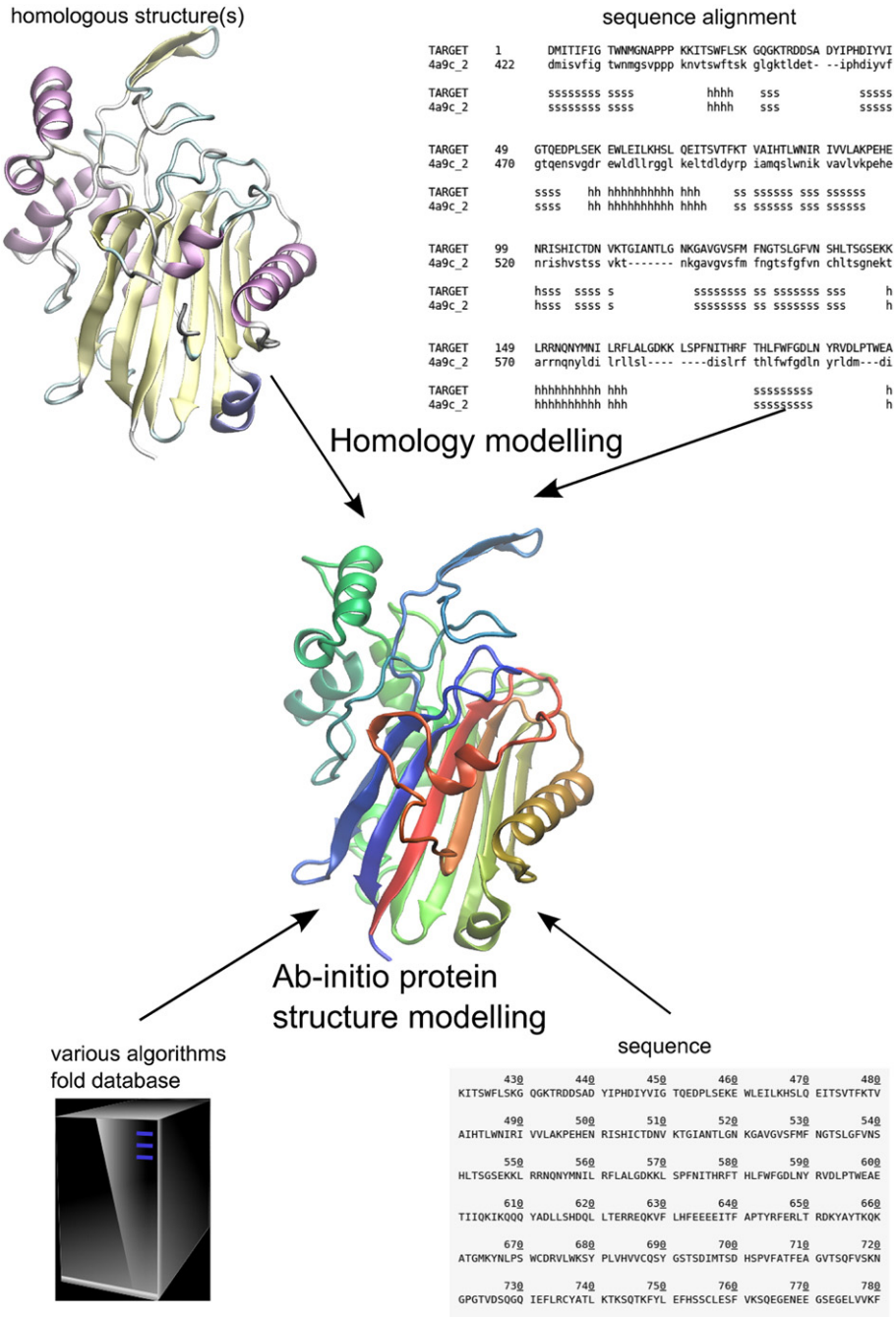

Fig. 4. Modelling of protein structures. Modelling of a protein structure is usually carried out based on sequence similarity to a related protein (homology modelling, top). An alternative approach is to use software that suggests a protein structure based on its sequence, often with the help of a database of short folded domains and their sequences.

modelling by homology, it is generally said that proteins that have >50% sequence identity to the target can be modelled accurately, whereas those with 25–30% sequence identity can still be modelled reasonably well. Proteins with lower similarity to their templates ('twilight zone' cases in the homology modelling terminology) pose a greater challenge, but may still be modelled.

Homology modelling can be performed by use of freely-available and commercial software or designated web-servers, some of which come with video tutorials or additional features such as modelling of proteins without close homologues. A list of several homology modelling servers and non-commercial programs is given in Table A.6. Commercial program packages (Table A.3) deal with homology modelling as part of their protein analysis capabilities, and the non-commercial viewer UCSD-Chimera can also be used to perform homology modelling with the Modeller program. Webservers that perform homology modelling are numerous. As is often the case with software that runs on a webserver (but even more for homology modelling), the default parameters may need adjustment for non-standard cases. Therefore, it is advised that users of modelling web-servers consult the documentation and make several trials to get the best model.

2.2.2. De-novo modelling of protein structures (fold recognition methods)

2.2.2.1. Strengths. Structure modelling when related structures are not available experimentally.

2.2.2.2. Limitations. Prediction accuracy.

2.2.2.3. Example application. Modelling of a protein with an unknown fold and no homologous or orthologous structures.

In principle, it should be possible to follow protein-folding in silico, using simulations starting from the protein sequence or a very crude model. Limitations in hardware and prediction accuracy (i.e., the quality of the potential energy function), however, prevent this from being commonplace [35], even if recent advances hint that ab-initio protein folding of some proteins (that fold relatively fast) can be followed by computer simulations [36].

Several homology modelling servers (Table A.6) can be used also for modelling of proteins that do not have a known orthologue. In addition, several research groups have developed servers or programs that use a combination of statistical methods based on the protein sequence, and/or physics-based energy functions to model protein structures. A list of servers and programs is given in Table A.7. The overall performance of many of the servers is evaluated by competitions such as the Critical Assessment of Structure Prediction (CASP). It is, however, difficult to predict their accuracy for a given protein with an unknown structure, and the results should therefore be validated by use of (indirect) experiments or additional simulations. Some programs allow the use of restraints derived from experiments (e.g., cross-linking), which can improve the prediction in many cases.

2.2.3. Modelling of protein complexes

The approach to generating a structural model of a complex involving a protein and another biological macromolecule depends on the existing data. If the structure of a similar complex is known, homology modelling can be used. Otherwise, one should first obtain structures or models of the participants. Then, a model can be generated by employing a program designated for the task. As is common in the field, several research groups have generated webservers that are aimed at modelling protein complexes, and a list of those is given in Table A.8. Typically, the servers produce a list of potential structures, sorted according to the underlying algorithm. It is eventually up to the user to select the model that makes most sense.

To generate models of protein complexes, servers employ different approaches. One of these is surface complementarity, where the proteins or macromolecules that interact are assumed to complement each other structurally (i.e., one can fit onto the other, similar to two pieces of a puzzle). The resulting structures can be refined by geometric or energetic considerations, e.g., removing clashes between the structures or calculating electrostatic interactions. Another class of methods relies on templates from the protein data bank, where a fitting of the interacting molecules to the template is based on sequence and structural similarities. The modelled complex is then threaded on to the template. A third class of programs utilises constraints from experimental measurements, such as NMR data, small angle X-ray scattering (SAXS), mutational analysis and cross-linking experiments.

2.3. Protein–drug interaction prediction

Identifying drug-like molecules that can bind to and inhibit the function of proteins is one of the most important objectives of computational medicinal chemistry. Kinases and various proteases in particular are interesting cancer-related targets of computer-aided drug design. The most straightforward approach is to use a structure of the target protein bound to an inhibitor and identify molecules that are structurally similar to that inhibitor. These candidate molecules are then fitted into the binding site, and the complexes are refined and studied by use of computer simulation methods such as molecular dynamics as a mean of validation. In many cases, however, this relatively straightforward approach does not yield any improvements or has already been exhausted, and it is necessary to identify chemically novel molecules, which may be more promising as leads to medication. This is when molecular docking approaches are employed.

2.3.1. Docking of drug-like molecules into protein targets

2.3.1.1. Strengths. High throughput, enabling the screening of large libraries of compounds.

2.3.1.2. Limitations. Accuracy. Further analysis is necessary before experimental validation. No analysis of ADMET (adsorption, distribution, metabolism, excretion and toxicity), which can be modelled separately. Not suitable for discriminating between similar compounds that bind the same target.

2.3.1.3. Example application. Discovery of novel kinase inhibitors [37].

Docking of drug-like molecules into target proteins has been studied for several decades. Advances in hardware and software have led to an impressive improvement, but no program can accurately predict the outcome of encounter between a protein and *any* drug-like molecule; docking programs try to outperform random selection of molecules, sometimes by orders of magnitude. Thus, if one wants to search for potential inhibitors for a given target out of a large dataset of drug-like molecules, careful application of a docking program will be of use (even if many of the suggested inhibitors will turn out to be false positives, and some real binders will not be detected). On the other hand, docking programs do not excel in discriminating between two inhibitors and judging which of them is more potent. In fact, even highly accurate and computationally demanding methods are not likely to be very successful in such an assignment.

Owing to the complexity of the task and the large amount of computational resources needed for ligand docking into proteins, it is more common to use a traditional computer program than an online server, although some servers do exist. Docking servers and programs are described in Tables A.9 and A.10, where we list several applications, commercial as well as free. Due to the vast use of docking programs in the pharmaceutical industry, commercial packages that deal with docking are quite prevalent and are often used also by academic researchers.

2.4. Simplified models that can yield biological insights

Modelling every atom in the molecule comes with a cost, and is not always meaningful: consider a complex of several proteins, large and small, that has a regulatory effect in the cell. It is hardly plausible that the regulatory effect is influenced by every amino acid side chain. In other cases, it is currently not possible to accurately model the conformation of a protein, because the protein has some unstructured regions (large flexible regions without any regular secondary structure). A prominent example of the latter case is p53, which has both folded and intrinsically disordered domains. Finally, we may need to deal with a protein with a known structure, but which undergoes a slow transition that cannot be modelled by atomistic methods because it takes too long. All of these cases necessitate the use of simplified models, often in conjugation with some of the methods that are used also (or predominantly) in atomistic studies such as MD simulations (Section 2.1.1) or normal mode analysis (Section 2.1.6).

2.4.1. Elastic network models

2.4.1.1. Strengths. Simple, fast and relatively easy to do — a quick way to get a glimpse on the dynamics without running a simulation. Does not require demanding preprocessing of the energy in the same way as atomistic normal mode analysis.

2.4.1.2. Limitations. No correlation to timescales. Depends on empirical parameters that need to be tuned and are not physics-based.

2.4.1.3. Example application. Applicable to large proteins or biomolecular complexes with known structure, e.g., myosin.

A great simplification of biological macromolecules can be achieved by considering interactions of residues or side-chains rather than atoms, thereby reducing the number of degrees of freedom. This is the basis for elastic network models (ENM), that model fluctuations in proteins by considering two adjacent Cα atoms or residues as being connected by springs. After building the elastic network, a normal mode analysis is carried out. The calculations are both simpler and faster than atomistic normal mode analysis (Section 2.1.6), and can therefore be routinely carried out by use of dedicated web servers (Table A.5), without the need for specialised software. However, ENM may still require some tuning, e.g., choosing the right distance between Cα atoms that are considered bonded in the chain (typically 5–7 Å).

2.4.2. Coarse-grained simulations

2.4.2.1. Strengths. Fast, yields dynamics data on the motions of the molecules, many structural observables can be easily extracted from the simulation, the simulated timescales can be orders of magnitude longer than atomistic simulations.

2.4.2.2. Limitations. Less accurate than atomistic simulations, most methods are not suitable for studies that involve changes in the secondary structure, and it is difficult to model drug molecules and cofactors such as heme. In addition, the simulation timescales do not correspond well with the experiment.

2.4.2.3. Example application. Simulations of large proteins or biomolecular complexes.

In coarse-grained simulations, molecules are represented by a collection of beads rather than atoms. Each bead represents several atoms (e.g., an amino acid side chain) or residues. Alternatively, in phenomenological models, a bead (or several beads) corresponds to a functional group. For example, a lipid can be represented by three beads — two for the tail and one for the hydrophilic head group [11]. Coarse-grained calculations are often carried out by running a molecular dynamics simulation with a regular MD program in combination with a specific set of parameters (coarse-grained force field). Other types of calculations, such as normal mode analysis can also be run with a coarse-grained system.

The main advantage of coarse-grained simulations is their speed. As each molecule is represented by a smaller number of particles, the simulations are faster to begin with because fewer degrees of freedom need to be sampled. Moreover, the timestep of a molecular dynamics simulation is dictated by the frequencies of the fastest motions (bond or angle stretching). These motions are slower for heavier particles, and therefore a larger timestep can be used between two successive iterations of the program (e.g., 50 fs [38], compared with 1–4 fs for an atomistic simulation), further speeding up the calculation. Third, water molecules are usually presented in a relatively crude way and not simulated explicitly, providing an additional speed-up factor. Thus, coarse-grained simulations can be used to study processes that cannot be followed by use of atomistic simulations because they are too slow or systems that cannot be handled because they are too large.

The gain in speed of coarse-grained simulations is countered by some drawbacks. The most obvious limitation of such simulations is that atomistic details are lost. Moreover, temporal and energetic estimations are also less accurate than those provided by an atomistic approach. Another caveat is the limited availability of coarse-grained representations and force fields. Methods for coarse-graining exist for protein residues and some types of lipids but generally not for ligands, cofactors, etc.

A hierarchical (or multiscale) approach may be used to overcome the first limitation. The system is studied at the more simplistic coarse-grained level first. Then, some parts of the trajectory are highlighted by performing additional simulations in full atomistic details to shed light on interesting parts of the trajectory. Using such a multiscale approach necessitates a method whereby the coarse-grained structures can be represented by atomistic details again (fine graining).

2.4.3. Dissipative particle dynamics

2.4.3.1. Strengths. Fast and flexible, suitable of simulations that range from protein complexes to cells and tissues.

2.4.3.2. Limitations. Less accurate, requires expertise and often also the ability to write or modify computer programs.

2.4.3.3. Example application. Simulations of multicomponent cell membranes (with proteins, cholesterol and attached sugars).

Dissipative particle dynamics (DPD) is a method for coarse-grained simulations that differs from both molecular and Brownian dynamics. Initially developed to deal with complex fluids [39], the method has since emerged as a powerful tool for simulations in biology, in systems that scale up to cells and tissues [40]. In DPD, the force acting upon a bead is calculated by summing up all of the interactions between a particle and all other particles within a certain cutoff. These are governed by forces that depend on the chemical bonds (or other rigid interactions), steric repulsions, the viscosity of the medium and the size of the interacting particles.

Being less widely used than molecular dynamics, DPD simulations are usually run by employing designated, home-written or specialised software such as DPDmacs (www.softsimu.net). Some MD packages (e.g., HoomD [41] and LAMMPS, http://lammps.sandia.gov) can also run DPD.

3. Examples of modelling and simulation studies as relevant to cancer-research

3.1. The catalytic activity of the aldo-keto-reductase tumour marker AKR1B10

Human small intestine aldose reductase, AKR1B10 is an $NADP^+$-dependent aldo-keto reductase. Aldo-keto reductases reduce a

variety of aldehydes and ketones. AKR1B10 is unique among aldo-keto reductases in its catalytic efficiency for reduction of retinaldehyde, and its elevated expression in non-small cell lung carcinoma. The physiological role of the enzyme, however, is not clear yet. Depletion of retinoic acid levels due to increased activity of the enzyme may be related to cancer development [42]. Structural details on the interaction between the enzyme and retinoids may therefore be useful for the design of specific inhibitors that target AKR1B10. Unfortunately, AKR1B10 and the structurally similar enzyme AKR1B1 have so far defied attempts for crysallising with retinoids. A molecular modelling approach was used to overcome this obstacle [43]. The structure of the enzyme was solved with an inhibitor bound instead of the native ligand. The native ligand was then docked into the binding site by a docking program. Molecular dynamics simulations were later applied to infer on the specificity of the enzyme to different forms of retinaldehyde (all-trans and 9-cis), in AKR1B10 and the less active AKR1B1. Mutational studies and simulations were used to explain the difference in the activity between the two enzymes [44]. Overall, these two studies contributed to the understanding on how AKR1B10 binds its substrates. Further modelling studies, e.g., with QM/MM can be used to shed light on its mechanism of action.

3.2. Understanding the mechanism of tubulin polymerisation inhibitors

Microtubules play a critical role in mitosis and are therefore important cancer drug targets. Microtubules are formed by polymerisation of tubulin, and inhibition of tubulin polymerisation leads to cell death. Colchicine, a natural inhibitor of tubulin polymerisation, is used for the treatment of gout and Familial Mediterranean fever but is toxic at the doses necessary for cancer treatment [45]. Other tubulin inhibitors, such as Taxol, have been proven useful in cancer treatment. Distant structural analogues of colchicine are sought after as novel and less toxic tubulin inhibitors. Zhang and co-workers synthesised a group of such inhibitors. Their most promising compound had anti-tubulin activity IC_{50} of 3.0 μM [46]. To understand its mechanism of action, the compound was docked to the crystal structure of tubulin bound to colchicine. Molecular dynamics simulations were then used to identify the protein residues that interact with the inhibitor. Similarly, Qian et al. have synthesised another group of anti-tubulin compounds, and have used molecular modelling to infer on the binding modes of several of them [47]. Both studies may be useful for further development of tubulin-binding agents.

3.3. Recognition of phosphorylated substrates by Pin1

Isomerisation of the ω-bond backbone angle of proline (peptidyl prolyl cis-trans isomerisation) initiates protein structural changes and is involved in biological signalling. Pin1 is an enzyme that catalyses peptidyl prolyl cis-trans isomerisation when a proline residue is located immediately after a phosphorylated serine or threonine. The enzyme is involved in different pathways that lead to oncogenesis, including the RTK/Ras/ERK, Wnt, TGF-β, NF-κB, Notch and others [48]. Atomistic details on how Pin1 interacts with its substrates are important for a better understanding of its mechanism of action, and may enable the design of inhibitors that mimic the transition state. In the lack of such data from experiments, Velazquez and Hamelberg [49] used molecular dynamics simulations. The authors have studies Pin1 in four states, namely free and bound to the substrate in cis, trans and transition-state conformations. Their analysis suggests that the structure of the Pin1 active site is modified in the presence of the specific backbone conformation of the phosphorylated substrate (conformational recognition), particularly in the transition state. These findings may be used for rational drug design.

3.4. Autoactivation of ERK2

Extracellular signal-regulated kinase 2 (ERK2) is a kinase involved in the MAPK signalling pathway. The enzyme is activated by cooperative phosphorylation events at Tyr^{185} and then Thr^{183}, which result in a conformational transition. Mutation of certain ERK2 residues leads to autophosphorylation and therefore to autoactivation. Barr and co-workers have studied the active and inactive forms of ERK2 and several mutants that cause autoactivation [50]. By use of molecular dynamics simulations, they have analysed the differences between the dynamics of mutants and wt proteins, and showed that the mutations do not mimic the active conformation. Rather, they lead to domain closure that is likely to promote autoactivation. Several mutations had been suggested to render autoactivation based on this study, and all but one were verified experimentally.

3.5. Engineering of a therapeutic protein

Escherichia coli L-asparaginase (L-ASN) is a protein drug that is used for treatment of the hematological malignancy acute lymphoblastic leukemia (ALL) [51], through hydrolysis of serum glutamine and asparagine, which leads to apoptosis. The lysosomal proteases cathepsinB and asparagine endopeptidase (AEP), which are produced by lymphoblasts, are involved in degradation of L-ASN. Offman and co-workers [52] have used structural and bioinformatic analysis to design L-ASN mutants that have lower affinity to being bound by AEP. Molecular dynamics simulations were then used to ensure that the active site of L-ASN is not modified by the mutations (otherwise L-ASN may become less active as a protein drug). The wild type (wt) protein and four promising mutants, as well as one mutant that was experimentally found to be less active (although it is not degraded by AEP) were then subject to extensive analysis by molecular dynamics simulations. The computer-aided protein engineering process was successful according to enzymatic and cell-toxicity assays. Furthermore, the authors could selectively inhibit glutamine hydrolysis by L-ASN, though a second mutation. Possible future studies include further optimisation of the protein, e.g., by modifications of the antigenic epitopes to avoid allergic reactions.

3.6. Computational approach discriminates functional activity of p53 mutants

Missense mutations of the tumour suppressor p53 often lead to loss-of-function, and are therefore termed 'cancer mutations'. The activity of p53 may be restored by an additional mutation in a different region ('rescue mutations') [53]. Demir et al. analysed a variety of mutants by functional experiments and measured their thermodynamic stability compared to the wt protein [54]. In parallel, they carried out molecular dynamics simulations and calculated the number of distinct conformations in the dynamic landscape of each mutant. The inactive mutants were found to be thermodynamically unstable, whereas a good correlation was found between the number of individual protein conformations in the simulation trajectory and the conformational stability. This suggests that mutants that destabilise and inactivate the protein lead to a larger number of distinct conformations, whereas secondary mutants that reduce the number of conformations stabilise the protein. The results of Demir et al. corroborate those of Boeckler and co-workers, who used a computational screening (docking) approach aimed at a crevice that is generated by a certain p53 mutation [55]. An identified compound bound to p53 and made it more stable, effectively mimicking a rescue mutation.

3.7. Atomistic structures of yttrium-containing glasses for cancer therapy

The radioactive isotope ^{90}Y of the rare earth metal yttrium is used in cancer therapy due to its cytotoxic effects. Its half-life of 64 h is long enough to allow treatment but not too long to yield excessive damage. A solution with ^{90}Y containing glass microspheres (brand name TheraSpheres) is injected into the blood vessels surrounding the tumour as a treatment for deeply sealed tumours such as hepatic neoplasia. To reduce side effects due to the toxicity of ^{90}Y, the glass should not release yttrium into the blood. Unfortunately, it is

difficult to analyse the local structure near the yttrium atom by experimental measurements. On the other hand, the periodic structure of the clinically important yttrium-alumino-silicate (YAS) glass is advantageous from a computational point of view, because a relatively small model can be studied due to the periodicity of the system (long term interactions are handled by an algorithm that replicates the small model in all directions). Moreover, the system is stable enough so that short-term quantum-mechanics based calculations could be performed and represent its equilibrium properties (something that cannot be assumed in the case of macromolecules). Accordingly, Christie and Tilocca have studied YAS glass by ab-initio molecular dynamics simulations [56]. The study elucidated the atomistic structure of YAS glass; future studies can deal with calculations of the energy needed to remove an yttrium atom from the glass or with the design of more stable yttrium binding compounds. Larger samples and perhaps more approximate methods will be necessary to achieve these aims. The authors have pinpointed another issue with YAS-microsphere treatment. The microspheres may reside in the patient's body many years after treatment, whereas their degree of biocompatibility is still unknown. Therefore, they have also modelled bioactive glasses that are known to have high biocompatibility, together with yttrium, using molecular dynamics simulations [57]. The simulations reveal two opposing effects of yttrium on the durability of bioactive glass. This calls for a combined experimental and computational approach for the design of novel types of ^{90}Y microspheres.

4. Potential applications for molecular modelling in cancer research

4.1. Prediction of functional consequences of molecular alterations

Next-generation sequencing and other high-throughput technologies are being extensively exploited to identify genomic, transcriptomic, proteomic or metabolomic characteristics of clinical samples from patients with cancer. In this context, novel molecular changes associated with cancer and cancer-related processes may be discovered, but in many cases the biological implications are unknown. Currently, major efforts are being invested in identifying genomic aberrations of tumour cells. When novel somatic mutations are detected in genes known to play important roles in tumour biology, the functional consequences of such mutations must be investigated. Computational strategies could be an invaluable tool in the initial stages of such studies, by structural modelling and molecular dynamics simulations, prior to experimental validation. The strength of using computational methods in attempts to predict biological functions lies in the ability to delineate complex biological systems into more simplified components. Subsequently, such individual components can be interconnected in more sophisticated models and thereby successively build a complex biological system. Through such approaches, otherwise complex, time-consuming and expensive biological experiments could be simplified by generating more specific hypotheses based on molecular modelling. In the future, as integrated data sets are generated that comprise information derived from the genomic, transcriptomic, proteomic and metabolomic levels, the need to perform more complex simulations with large amounts of data will be increasingly apparent. Effective modelling of such complex and large systems is at present not possible, but continued improvements in computational power and software will hopefully allow integration of more data into the simulations.

4.2. Therapeutic aspects

Molecular modelling is already being extensively used in the development of novel cancer drugs, as exemplified in the previous sections. The specific interactions between an existing drug and its target can be modelled (as outlined in Section 2.3.1), and modelling can be used in the generation of novel and more effective compounds (as exemplified for L-asparaginase in Section 3.5).

One setting that is particularly interesting in the perspective of targeted therapeutics is the possibility of predicting which compounds will be active in the presence of resistance mutations. Targeted therapies are being increasingly used in cancer treatment, and the mechanisms of action are often based on the inhibition of a single protein, such as BCR-ABL, KIT, EGFR, RAF or ALK. However, since these proteins are all part of complex signalling networks, modifying the activity of one pathway member will inevitably result in changed expression levels or activity of other proteins in the pathway. There is also an apparent redundancy of many important signalling networks in cancer that leads to alternate signalling if one protein is inhibited. Thus, in almost all cases the disease becomes resistant to the targeted therapy, and much effort is being devoted to the identification of resistance mechanisms.

A common resistance mechanism is the occurrence of a secondary mutation in the target protein, leading to decreased sensitivity towards the inhibitor. One example is the secondary resistance mutation T790M in the EGFR gene, which confers resistance to the EGFR inhibitors gefitinib and erlotinib in non-small cell lung cancer. Shortly after its identification it was suggested, based on structural modelling, that the introduction of a bulky methionine residue in position 790, which is located at the entrance to a hydrophobic pocket in the ATP binding cleft, could lead to steric hindrance and thus interfere with binding of erlotinib and gefitinib [58]. However, subsequent studies revealed that the T790M mutant confers resistance by a higher binding affinity for ATP [59], a mechanism that could be overcome by irreversible EGFR inhibitors. In this case, further development of irreversible EGFR inhibitors may have been halted due to incorrect conclusions in the former studies. A more recent example is the identification of secondary ALK mutations associated with resistance to the ALK inhibitor crizotinib in EML4-ALK positive non-small cell lung cancer patients [60]. Some of the discovered mutations were analogous to resistance mutations in other kinases (such as the T790M gate-keeper mutation), and some seem unique to ALK. Importantly, computational modelling may be able to predict the mechanisms of resistance of these new mutations, providing important information in the design and development of second generation ALK inhibitors (Fig. 5).

Both these examples illustrate the utility of molecular modelling approaches in the development of inhibitors designed to overcome resistance to targeted therapy. Similar sets of resistance mutations will probably be identified for other druggable kinases as new kinase inhibitors are developed, and we predict that molecular modelling will be crucial in this field in the years to come. One might argue that as next-generation sequencing techniques are becoming less expensive and more available, in silico prediction of novel resistance mutations would be less important. However, although high-throughput identification of resistance mutations in a clinical sample may be possible, this number is often quite high, and the challenge of predicting which mutations are relevant for function will still remain, leaving a role for molecular modelling in this setting.

5. Conclusions and outlook

One of the consequences of the post-genomic era is the realisation that biological macromolecules are much more complex than thought before. The straightforward relation sequence–structure–function is evidently too simplistic: many proteins carry out more than a single function or have a multitude of conformations rather than a single structure. Molecular modelling approaches can aid to the understanding of biomolecules by following them on the computer screen, and can save unnecessary experiments, but the

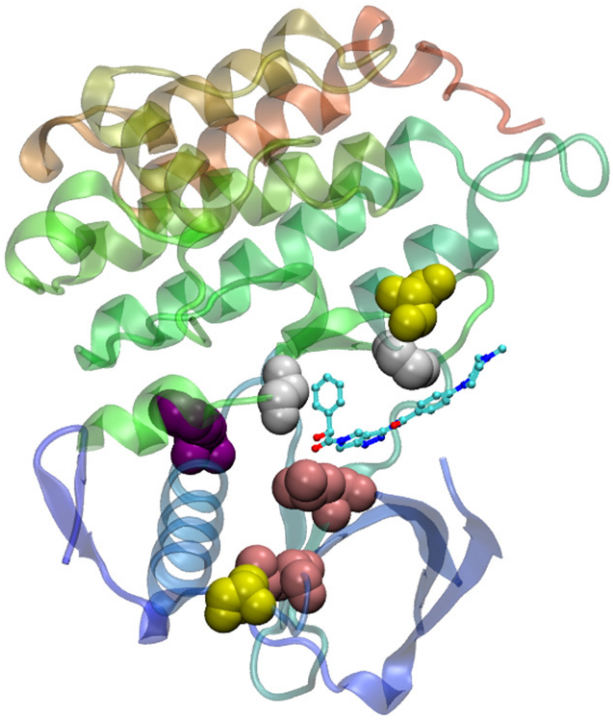

Fig. 5. Alk drug resistance. Several mutations of Alk that are associated with acquired resistance to therapy (reported in "My cancer genome" [63]) are shown on the structure of the protein with an inhibitor [64]. The protein is shown in a cartoon representation, the inhibitor is depicted with balls and sticks and the mutated residues are represented by coloured spheres (residues Gly in light gray, Phe in purple, Ser in dark yellow, Cys in light yellow, and Leu in mauve). Notably, some residues confer drug resistance although they are not adjacent to the inhibitor. Thus, the protein structure alone cannot explain the acquired resistance. Molecular dynamics simulations can be used to infer on modified dynamics of the protein or weakened protein–drug interactions upon mutation, which can later be followed by structural experiments, e.g., by using NMR.

limitations of molecular modelling methods should not be put aside. Importantly, whereas the computations become increasingly more user-friendly, and many calculations can be carried out online by users without deep understanding of the underlying methods, careful interpretation by an expert is almost always necessary to make sure that the results are indeed meaningful, which calls for a collaboration between molecular modellers and experimentalists. Additionally, graduate students with a background in experimental cancer-research and an inclination to quantitative studies may invest the time and effort to attend some of the (too few) courses that explain the methods as well as the underlying theories.

Realising the pros and cons of the approach is necessary to be able to validate its potential use in the field, and we have tried to provide some information that would enable cancer researchers to study the matter more closely. In our view, many of the tools described in this paper have the potential to become a useful part in the cancer-research arsenal, as they already are in other fields. We therefore have little doubt that molecular modelling approaches will become widely used for cancer research in the years to come, in particular for the design of drugs and other treatments and the understanding of biological networks.

Acknowledgement

We would like to thank Prof Gunhild M. Mælandsmo for fruitful discussions and critical reading of the manuscript, and two anonymous referees for their comments.

Appendix A. Computer programs and webservers

Table A.1
Molecular dynamics simulation packages. A list of several popular molecular dynamics simulation packages that are widely used in biology. [a] Reduced price. [b] Commercial licenses available separately.

Program	License	Particular strengths	Website
Amber [65]	Academic[a]/ commercial	A built-in method to deal with drug molecules or cofactors	http://ambermd.org
CHARMM [66,67]	Academic[a] [b]	Very broad functionality The charmm-gui web server (see text)	http://charmm.org
Gromacs [68,69]	Open source	Fast, relatively straightforward wide user support	http://gromacs.org
NAMD [70]	Free[b]	Specially designed for use with massive computer clusters	http://www.ks.uiuc.edu/Research/namd

Table A.2
Macromolecular viewers. Several popular and freely available programs to view macromolecular structures and MD trajectory files. All programs have a wide range of features to choose from and can carry out some calculations.

Program	Website	Comments
PyMol	pymol.org	Probably the most widely used
UCSF Chimera [71]	cgl.ucsf.edu/chimera	Relatively easy to use without vast knowledge
VMD [12]	ks.uiuc.edu/Research/vmd	Very comprehensive

Table A.3
Commercial molecular modelling packages. Several commercially available molecular modelling packages. The packages listed here can be used for molecular dynamics simulations, homology modelling, docking drug-like molecules into proteins and analysis of the results. The programs also include graphical user interface that can be used for preparation of the data and viewing of the results. See the distributors' sites for details.

Program	Distributer
Discovery Studio	Accelrys
Molecular Operating Environment (MOE)	Chemical Computing Group
SYBYL-X	Tripos

Table A.4
Brownian dynamics simulation software.

Program	Website
Browndye [72]	browndye.ucsd.edu
Brownmove [73]	gepard.bioinformatik.uni-saarland.de/services/brownmove
Macrodox [74]	iweb.tntech.edu/macrodox
SDA [75]	projects.villa-bosch.de/mcmsoft/sda

Table A.5
Normal mode analysis (NMA) and elastic network model (ENM) servers.

Server	Website	NMA/ENM	Special features
AD-ENM	enm.lobos.nih.gov	ENM	(1) Generates paths between protein conformations [76]. (2) Can be used in the context of protein structure prediction [77]. (3) Can deal with nucleic acids as well.
ANM [78]	ignmtest.ccbb.pitt.edu/cgi-bin/anm/anm1.cgi	ENM	
el-Némo [79]	www.igs.cnrs-mrs.fr/elnemo	ENM	Advanced options for comparing two conformations.
oGNM [80]	ignm.ccbb.pitt.edu/GNM_Online_Calculation-t.htm	ENM	Handles nucleic acids
NOMAD-Ref [81]	lorentz.immstr.pasteur.fr/nma	ENM & NMA	Analysis of collective motions [82]
WEBnm@ [83]	apps.cbu.uib.no/webnma	NMA	Comparative analysis (several protein structures).

Table A.6
Homology modelling servers and non-commercial programs (for commercial programs, see Table A.3).

Servers:

Server	Website	Additional features
CPHmodels [84]	http://www.cbs.dtu.dk/services/CPHmodels-3.2	Extended search for template sequences.
ESyPred3D [85]	http://www.fundp.ac.be/sciences/biologie/urbm/bioinfo/esypred	
Phyre^2 [86]	http://www.sbg.bio.ic.ac.uk/phyre2	(1) Video tutorials (2) Ab-initio modelling (no sequence). (3) Binding site prediction.
PS^2 [87]	ps2.life.nctu.edu.tw	
PS^2-v2 [88]	ps2v2.life.nctu.edu.tw	For use with low-identity templates
Swiss-Model [89]	http://swissmodel.expasy.org	(1) Automated modelling of homo-oligomeric assemblies. (2) Modelling of essential metal ions in protein structures. (3) Includes a repository of models.

Programs:

Program	Website	Additional features
Modeller [90]	salilab.org/modeller	Many additional tasks, e.g., ab-initio modelling of loops.
DeepView [91]	spdbv.vital-it.ch	Viewer, protein analysis.
Nest [92]	wiki.c2b2.columbia.edu/honiglab_public/index.php/Software:nest	

Table A.7
Web servers for de-novo protein structure prediction. [a] Combined de-novo and homology modelling.

Server	Website
chunk-TASSER [93]	cssb.biology.gatech.edu/skolnick/webservice/chunk-TASSER/index.html
Fobia [94]	bioinfo3d.cs.tau.ac.il/FOBIA
meta-TASSER	cssb.biology.gatech.edu/skolnick/webservice/MetaTASSER/index.html
pro-sp3-TASSER [95]	cssb.biology.gatech.edu/skolnick/webservice/pro-sp3-TASSER/index.html
I-TASSER [96]	zhanglab.ccmb.med.umich.edu/I-TASSER
Quark online [97]	zhanglab.ccmb.med.umich.edu/QUARK
Robetta[a] [98]	robetta.bakerlab.org

Table A.8
Web servers for protein complex structure prediction.

Server	Website	Molecules	Method(s)
3D-Garden [99]	www.sbg.bio.ic.ac.uk/3dgarden	Proteins, polynucleotides	Surface complementarity
ClusPro [100]	cluspro.bu.edu	Protein	Surface complementarity, energetics
COTH [101]	zhanglab.ccmb.med.umich.edu/COTH	Proteins	Threading
FoXS Dock [102]	modbase.compbio.ucsf.edu/foxsdock	Proteins	SAXS constraints
GRAMM-X [103]	vakser.bioinformatics.ku.edu/resources/gramm/grammx	Proteins	Surface complementarity, energetics
Haddock [104]	nmr.chem.uu.nl/haddock	Proteins, polynucleotides and other	Experimental constraints
PatchDock [102]	bioinfo3d.cs.tau.ac.il/PatchDock	Proteins	Surface complementarity

Table A.9
Web servers for docking of drug-like molecules into proteins.

Server	Website	Features/limitations
Dock Blaster [105]	blaster.docking.org	Includes a large database of drug-like molecules.
Docking@UTMB	docking.utmb.edu	Several databases available.
DockingServer	www.dockingserver.com	Commercial.
iScreen [106]	iscreen.cmu.edu.tw	Docking of compounds from traditional Chinese medicine.
SwissDock [107]	swissdock.vital-it.ch	Docks one or a few molecules, not a library.

Table A.10
Docking programs. Commonly used computer programs for docking of drug-like molecules into proteins. See also the commercial modelling packages in Table A.3.

Program	License	Website	Features/limitations
Autodock [108]	Open source	autodock.scripps.edu	External graphical user interface.
Autodock Vina [109]	Open source	vina.scripps.edu	Designed for ease of use.
Dock [110]	Academic	dock.compbio.ucsf.edu	Can deal with RNA molecules as targets.
FlexX [111]	Commercial	www.biosolveit.de/flexx	
FRED [112]	Commercial	www.eyesopen.com/oedocking	
Glide [113,114]	Commercial	www.schrodinger.com/products/14/5/	
GOLD	Commercial	www.ccdc.cam.ac.uk/products/life_sciences/gold	

References

[1] A. Awada, A.M. Vandone, P. Aftimos, Personalized management of patients with solid cancers: moving from patient characteristics to tumor biology, Curr. Opin. Oncol. 24 (2012) 297–304.
[2] K. Henzler-Wildman, D. Kern, Dynamic personalities of proteins, Nature 450 (2007) 964–972.
[3] R. Friedman, Aggregation of amyloids in a cellular context: modelling and experiment, Biochem. J. 438 (2011) 415–426.
[4] J.P. Mathew, B.S. Taylor, G.D. Bader, S. Pyarajan, M. Antoniotti, A.M. Chinnaiyan, C. Sander, S.J. Burakoff, B. Mishra, From bytes to bedside: data integration and computational biology for translational cancer research, PLoS Comput. Biol. 3 (2007) e12.
[5] A. Baudot, F.X. Real, J.M. Izarzugaza, A. Valencia, From cancer genomes to cancer models: bridging the gaps, EMBO Rep. 10 (2009) 359–366.
[6] H.M. Byrne, Dissecting cancer through mathematics: from the cell to the animal model, Nat. Rev. Cancer 10 (2010) 221–230.
[7] S.G. Baker, B.S. Kramer, Systems biology and cancer: promises and perils, Prog. Biophys. Mol. Biol. 106 (2011) 410–413.
[8] A. Leach, Molecular Modelling: Principles and Applications, Prentice Hall, New Jersey, 2001.
[9] D. Frenkel, B. Smith, Understanding Molecular Simulation, Second Edition: From Algorithms to Applications, Elsevier, Amsterdam, the Netherlands, 2002.
[10] M. Levitt, Detailed molecular model for transfer ribonucleic acid, Nature 224 (1969) 759–763.
[11] R. Friedman, R. Pellarin, A. Caflisch, Amyloid aggregation on lipid bilayers and its impact on membrane permeability, J. Mol. Biol. 387 (2009) 407–415.
[12] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38.
[13] R. Friedman, A. Caflisch, The protonation state of the catalytic aspartates in plasmepsin II, FEBS Lett. 581 (2007) 4120–4124.
[14] R.O. Dror, M.Ø. Jensen, D.W. Borhani, D.E. Shaw, Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations, J. Gen. Physiol. 135 (2010) 555–562.
[15] O.F. Lange, D. van der Spoel, B.L. de Groot, Scrutinizing molecular mechanics force fields on the submicrosecond timescale with NMR data, Biophys. J. 99 (2010) 647–655.
[16] E. Project, E. Nachliel, M. Gutman, Force field-dependent structural divergence revealed during long time simulations of calbindin d9k, J. Comput. Chem. 31 (2010) 1864–1872.
[17] W. Im, S. Seefeld, B. Roux, A Grand Canonical Monte Carlo–Brownian dynamics algorithm for simulating ion channels, Biophys. J. 79 (2000) 788–801.
[18] M. Kang, C. Roberts, Y. Cheng, C.-e.A. Chang, Gating and intermolecular interactions in ligand–protein association: coarse-grained modeling of HIV-1 protease, J. Chem. Theory Comput. 7 (2011) 3438–3446.
[19] C. Chen, B.M. Pettitt, The binding process of a nonspecific enzyme with DNA, Biophys. J. 101 (2011) 1139–1147.
[20] B. Aradi, B. Hourahine, T. Frauenheim, DFTB+, a sparse matrix-based implementation of the DFTB method, J. Phys. Chem. A 111 (2007) 5678–5684.
[21] D.G. Fedorov, T. Nagata, K. Kitaura, Exploring chemistry with the fragment molecular orbital method, Phys. Chem. Chem. Phys. 14 (2012) 7562–7577.
[22] P. Hohenberg, W. Kohn, Inhomogeneous electron gas, Phys. Rev. 136 (1964) B864–B871.
[23] W. Kohn, L.J. Sham, Self-consistent equations including exchange and correlation effects, Phys. Rev. 140 (1965) A1133–A1138.
[24] A. Kumar, M. Elstner, S. Suhai, SCC-DFTB-D study of intercalating carcinogens: benzo(a)pyrene and its metabolites complexed with the G–C base pair, Int. J. Quantum Chem. 95 (2003) 44–59.
[25] D.V. Deubel, The chemistry of dinuclear analogues of the anticancer drug cisplatin. A DFT/CDM study, J. Am. Chem. Soc. 128 (2006) 1654–1663.
[26] C. Corminboeuf, P. Hu, M.E. Tuckerman, Y. Zhang, Unexpected deacetylation mechanism suggested by a density functional theory QM/MM study of histone-deacetylase-like protein, J. Am. Chem. Soc. 128 (2006) 4530–4531.
[27] R. Car, M. Parrinello, Unified approach for molecular-dynamics and density-functional theory, Phys. Rev. Lett. 55 (1985) 2471–2474, (Times Cited: 3921 Article English Cited References Count: 12 Aun49).
[28] R.P. Feynman, A.R. Hibbs, Quantum Mechanics and Path Integrals, McGraw-Hill, New York, NY, USA, 1965.
[29] H. Engel, D. Doron, A. Kohen, D.T. Major, Momentum distribution as a fingerprint of quantum delocalization in enzymatic reactions: open-chain path-integral simulations of model systems and the hydride transfer in dihydrofolate reductase, J. Chem. Theory Comput. 8 (2012) 1223–1234.
[30] K. Spiegel, A. Magistrato, Modeling anticancer drug–DNA interactions via mixed QM/MM molecular dynamics simulations, Org. Biomol. Chem. (2006) 2507–2517.
[31] A.G. Turjanski, G. Hummer, J.S. Gutkind, How mitogen-activated protein kinases recognize and phosphorylate their targets: a QM/MM study, J. Am. Chem. Soc. 131 (2009) 6141–6148.
[32] S. Hayward, A. Kitao, N. Go, Harmonicity and anharmonicity in protein dynamics: a normal mode analysis and principal component analysis, Proteins 23 (1995) 177–186.
[33] E. Laine, I. Chauvot de Beauchene, D. Perahia, C. Auclair, L. Tchertanov, Mutation D816V alters the internal structure and dynamics of c-KIT receptor cytoplasmic region: implications for dimerization and activation mechanisms, PLoS Comput. Biol. 7 (2011) e1002068.
[34] L.M. Iakoucheva, C.J. Brown, J.D. Lawson, Z. Obradovi, A.K. Dunker, Intrinsic disorder in cell-signaling and cancer-associated proteins, J. Mol. Biol. 323 (2002) 573–584.
[35] P.L. Freddolino, C.B. Harrison, Y. Liu, K. Schulten, Challenges in protein folding simulations: timescale, representation, and analysis, Nat. Phys. 6 (2010) 751–758.
[36] K. Lindorff-Larsen, S. Piana, R.O. Dror, D.E. Shaw, How fast-folding proteins fold, Science 334 (2011) 517–520.
[37] H. Zhao, D. Huang, A. Caflisch, Discovery of tyrosine kinase inhibitors by docking into an inactive kinase conformation generated by molecular dynamics, ChemMedChem (in press), http://dx.doi.org/10.1002/cmdc.20120033.
[38] R. Friedman, A. Caflisch, Surfactant effects on amyloid aggregation kinetics, J. Mol. Biol. 414 (2011) 303–312.
[39] P.J. Hoogerbrugge, J.M.V.A. Koelman, Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics, Europhys. Lett. 19 (1992) 155.
[40] M. Basan, J. Prost, J.F. Joanny, J. Elgeti, Dissipative particle dynamics simulations for biological tissues: rheology and competition, Phys. Biol. 8 (2011), (026014–026014).
[41] J.A. Anderson, C.D. Lorenz, A. Travesset, General purpose molecular dynamics simulations fully implemented on graphics processing units, J. Comput. Phys. 227 (2008) 5342–5359.
[42] T.M. Penning, AKR1B10: a new diagnostic marker of non-small cell lung carcinoma in smokers, Clin. Cancer Res. 11 (2005) 1687–1690.
[43] O. Gallego, F.X. Ruiz, A. Ardèvol, M. Domínguez, R. Alvarez, A.R. de Lera, C. Rovira, J. Farrés, I. Fita, X. Parés, Structural basis for the high all-trans-retinaldehyde reductase activity of the tumor marker AKR1B10, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 20764–20769.
[44] F.X. Ruiz, A. Moro, O. Gallego, A. Ardèvol, C. Rovira, J.M. Petrash, X. Parés, J. Farrés, Human and rodent aldo-keto reductases from the AKR1B subfamily and their specificity with retinaldehyde, Chem. Biol. Interact. 191 (2011) 199–205.
[45] B.C. Baguley, Chapter 1 — a brief history of cancer chemotherapy, in: B.C. Baguley, D.J. Kerr (Eds.), Anticancer Drug Development, Academic Press, San Diego, 2002, pp. 1–11.
[46] Q. Zhang, Y. Peng, X.I. Wang, S.M. Keenan, S. Arora, W.J. Welsh, Highly potent triazole-based tubulin polymerization inhibitors, J. Med. Chem. 50 (2007) 749–754.
[47] Y. Qian, H.J. Zhang, P.C. Lv, H.L. Zhu, Synthesis, molecular modeling and biological evaluation of guanidine derivatives as novel antitubulin agents, Bioorg. Med. Chem. 18 (2010) 8218–8225.
[48] M. Theuerkorn, G. Fischer, C. Schiene-Fischer, Prolyl cis/trans isomerase signalling pathways in cancer, Curr. Opin. Pharmacol. 11 (2011) 281–287.
[49] H.A. Velazquez, D. Hamelberg, Conformational selection in the recognition of phosphorylated substrates by the catalytic domain of human Pin1, Biochemistry 50 (2011) 9605–9615.
[50] D. Barr, T. Oashi, K. Burkhard, S. Lucius, R. Samadani, J. Zhang, P. Shapiro, A.D. MacKerell, A. van der Vaart, Importance of domain closure for the autoactivation of ERK2, Biochemistry 50 (2011) 8038–8048.
[51] N.E. Labrou, A.C. Papageorgiou, V.I. Avramis, Structure-function relationships and clinical applications of L-asparaginases, Curr. Med. Chem. 17 (2010) 2183–2195.
[52] M.N. Offman, M. Krol, N. Patel, S. Krishnan, J. Liu, V. Saha, P.A. Bates, Rational engineering of L-asparaginase reveals importance of dual activity for cancer cell toxicity, Blood 117 (2011) 1614–1621.
[53] R. Baronio, S.A. Danziger, L.V. Hall, K. Salmon, G.W. Hatfield, R.H. Lathrop, P. Kaiser, All-codon scanning identifies p53 cancer rescue mutations, Nucleic Acids Res. 38 (2010) 7079–7088.
[54] Ö. Demir, R. Baronio, F. Salehi, C.D. Wassman, L. Hall, G.W. Hatfield, R. Chamberlin, P. Kaiser, R.H. Lathrop, R.E. Amaro, Ensemble-based computational approach discriminates functional activity of p53 cancer and rescue mutants, PLoS Comput. Biol. 7 (2011) e1002238.
[55] F.M. Boeckler, A.C. Joerger, G. Jaggi, T.J. Rutherford, D.B. Veprintsev, A.R. Fersht, Targeted rescue of a destabilized mutant of p53 by an in silico screened drug, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 10360–10365.

[56] J.K. Christie, A. Tilocca, Short-range structure of yttrium alumino-silicate glass for cancer radiotherapy: Car–Parrinello molecular dynamics simulations, Adv. Eng. Mater. 12 (2010) B326–B330.
[57] J.K. Christie, J. Malik, A. Tilocca, Bioactive glasses as potential radioisotope vectors for in situ cancer therapy: investigating the structural effects of yttrium, Phys. Chem. Chem. Phys. 13 (2011) 17749–17755.
[58] S. Kobayashi, T.J. Boggon, T. Dayaram, P.A. Jänne, O. Kocher, M. Meyerson, B.E. Johnson, M.J. Eck, D.G. Tenen, B. Halmos, EGFR mutation and resistance of non-small-cell lung cancer to gefitinib, N. Engl. J. Med. 352 (2005) 786–792.
[59] C.H. Yun, K.E. Mengwasser, A.V. Toms, M.S. Woo, H. Greulich, K.K. Wong, M. Meyerson, M.J. Eck, The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 2070–2075.
[60] R. Katayama, A.T. Shaw, T.M. Khan, M. Mino-Kenudson, B.J. Solomon, B. Halmos, N.A. Jessop, J.C. Wain, A.T. Yeo, C. Benes, L. Drew, J.C. Saeh, K. Crosby, L.V. Sequist, A.J. Iafrate, J.A. Engelman, Mechanisms of acquired crizotinib resistance in ALK-rearranged lung cancers, Sci. Transl. Med. 4 (2012), http://dx.doi.org/10.1126/scitranslmed.3003316.
[61] E. Liaudet-Coopman, M. Beaujouin, D. Derocq, M. Garcia, M. Glondu-Lassis, V. Laurent-Matha, C. Prébois, H. Rochefort, F. Vignon, Cathepsin D: newly discovered functions of a long-standing aspartic protease in cancer and apoptosis, Cancer Lett. 237 (2006) 167–179.
[62] R. Friedman, A. Caflisch, On the orientation of the catalytic dyad in aspartic proteases, Proteins 78 (2010) 1575–1582.
[63] C. Swanton, My cancer genome: a unified genomics and clinical trial portal, Lancet Oncol. 13 (2012) 668–669.
[64] R.T. Bossi, M.B. Saccardo, E. Ardini, M. Menichincheri, L. Rusconi, P. Magnaghi, P. Orsini, N. Avanzi, A.L. Borgia, M. Nesi, T. Bandiera, G. Fogliatto, J.A. Bertrand, Crystal structures of anaplastic lymphoma kinase in complex with ATP competitive inhibitors, Biochemistry 49 (2010) 6813–6825.
[65] D.A. Case, T.E. Cheatham, T. Darden, H. Gohlke, R. Luo, K.M. Merz, A. Onufriev, C. Simmerling, B. Wang, R.J. Woods, The amber biomolecular simulation programs, J. Comput. Chem. 26 (2005) 1668–1688.
[66] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM: a program for macromolecular energy, minimization, and dynamics calculations, J. Comput. Chem. 4 (1983) 187–217.
[67] B.R. Brooks, C.L. Brooks, A.D. MacKerell, L. Nilsson, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. Dinner, S. Fischer, J. Gao, M. Hodoscek, K. Kuczera, T. Lazaridis, J. Ma, E. Paci, R.W. Pastor, C.B. Post, M. Schaefer, B. Tidor, R.W. Venable, H.L. Woodcock, X. Wu, M. Karplus, CHARMM: the biomolecular simulation program, J. Comput. Chem. 30 (2009) 1545–1614.
[68] H.J.C. Berendsen, D. van der Spoel, R. Vandrunen, Gromacs — a message-passing parallel molecular-dynamics implementation, Comput. Phys. Commun. 91 (1995) 43–56.
[69] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J.C. Berendsen, Gromacs: fast, flexible, and free, J. Comput. Chem. 26 (2005) 1701–1718.
[70] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kalé, K. Schulten, Scalable molecular dynamics with NAMD, J. Comput. Chem. 26 (2005) 1781–1802.
[71] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera—a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612.
[72] G.A. Huber, J.A. McCammon, Browndye: a software package for Brownian dynamics, Comput. Phys. Commun. 181 (2010) 1896–1905.
[73] U. Winter, T. Geyer, Coarse grained simulations of a small peptide: effects of finite damping and hydrodynamic interactions, J. Chem. Phys. 131 (2009) 104102.
[74] S.H. Northrup, J.A. Luton, J.O. Boles, J.C. Reynolds, Brownian dynamics simulation of protein association, J. Comput. Aided Mol. Des. 1 (1988) 291–311.
[75] R.R. Gabdoulline, R.C. Wade, Brownian dynamics simulation of protein–protein diffusional encounter, Methods 14 (1998) 329–341.
[76] W. Zheng, B.R. Brooks, G. Hummer, Protein conformational transitions explored by mixed elastic network models, Proteins 69 (2007) 43–57.
[77] W. Zheng, B.R. Brooks, Modeling protein conformational changes by iterative fitting of distance constraints using reoriented normal modes, Biophys. J. 90 (2006) 4327–4336.
[78] E. Eyal, L.W. Yang, I. Bahar, Anisotropic network model: systematic evaluation and a new web interface, Bioinformatics 22 (2006) 2619–2627.
[79] K. Suhre, Y.H. Sanejouand, ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement, Nucleic Acids Res. 32 (2004) W610–W614.
[80] L.W. Yang, A.J. Rader, X. Liu, C.J. Jursa, S.C. Chen, H.A. Karimi, I. Bahar, oGNM: online computation of structural dynamics using the Gaussian Network Model, Nucleic Acids Res. 34 (2006) 24–31.
[81] E. Lindahl, C. Azuara, P. Koehl, M. Delarue, NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis, Nucleic Acids Res. 34 (2006) W52–W56.
[82] F. Tama, Y.H. Sanejouand, Conformational change of proteins arising from normal mode calculations, Protein Eng. 14 (2001).
[83] S.M. Hollup, G. Salensminde, N. Reuter, WEBnm@: a web application for normal mode analyses of proteins, BMC Bioinforma. (2005) 52.
[84] M. Nielsen, C. Lundegaard, O. Lund, T.N. Petersen, CPHmodels-3.0—remote homology modeling using structure-guided sequence profiles, Nucleic Acids Res. 38 (2010) W576–W581.
[85] C. Lambert, N. Leonard, X. De Bolle, E. Depiereux, ESyPred3D: prediction of proteins 3D structures, Bioinformatics 18 (2002) 1250–1256.
[86] L.A. Kelley, M.J. Sternberg, Protein structure prediction on the Web: a case study using the Phyre server, Nat. Protoc. 4 (2009) 363–371.
[87] C.C. Chen, J.K. Hwang, J.M. Yang, (PS)2: protein structure prediction server, Nucleic Acids Res. 34 (2006) W152–W157.
[88] C.C. Chen, J.K. Hwang, J.M. Yang, (PS)2-v2: template-based protein structure prediction server, BMC Bioinforma. 10 (2009) 366.
[89] L. Bordoli, F. Kiefer, K. Arnold, P. Benkert, J. Battey, T. Schwede, Protein structure homology modeling using SWISS-MODEL workspace, Nat. Protoc. 4 (2009) 1–13.
[90] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, J. Mol. Biol. 234 (1993) 779–815.
[91] N. Guex, M.C. Peitsch, SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling, Electrophoresis 18 (1997) 2714–2723.
[92] D. Petrey, Z. Xiang, C.L. Tang, L. Xie, M. Gimpelev, T. Mitros, C.S. Soto, S. Goldsmith-Fischman, A. Kernytsky, A. Schlessinger, I.Y. Koh, E. Alexov, B. Honig, Using multiple structure alignments, fast model building, and energetic analysis in fold recognition and homology modeling, Proteins 53 (Suppl. 6) (2003) 430–435.
[93] H. Zhou, J. Skolnick, Ab initio protein structure prediction using chunk-TASSER, Biophys. J. 93 (2007) 1510–1518.
[94] I. Kifer, R. Nussinov, H.J. Wolfson, Protein structure prediction using a docking-based hierarchical folding scheme, Proteins 79 (2011) 1759–1773.
[95] Z. Zhou, Y. Li, Molecular dynamics simulation of S100B protein to explore ligand blockage of the interaction with p53 protein, J. Comput. Aided Mol. Des. 23 (2009) 705–714.
[96] Y. Zhang, I-TASSER server for protein 3D structure prediction, BMC Bioinforma. 9 (2008) 40.
[97] D. Xu, Y. Zhang, Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field, Proteins 80 (2012) 1715–1735.
[98] D.E. Kim, D. Chivian, D. Baker, Protein structure prediction and analysis using the Robetta server, Nucleic Acids Res. 32 (2004) 526–531.
[99] V.I. Lesk, M.J. Sternberg, 3D-Garden: a system for modelling protein–protein complexes based on conformational refinement of ensembles generated with the marching cubes algorithm, Bioinformatics 24 (2008) 1137–1144.
[100] S.R. Comeau, D.W. Gatchell, S. Vajda, C.J. Camacho, ClusPro: a fully automated algorithm for protein–protein docking, Nucleic Acids Res. 32 (2004) 96–99.
[101] S. Mukherjee, Y. Zhang, Protein–protein complex structure predictions by multimeric threading and template recombination, Structure 19 (2011) 955–966.
[102] D. Schneidman-Duhovny, Y. Inbar, R. Nussinov, H.J. Wolfson, PatchDock and SymmDock: servers for rigid and symmetric docking, Nucleic Acids Res. 33 (2005) 363–367.
[103] A. Tovchigrechko, I.A. Vakser, GRAMM-X public web server for protein–protein docking, Nucleic Acids Res. 34 (2006) 310–314.
[104] S.J. de Vries, M. van Dijk, A.M. Bonvin, The HADDOCK web server for data-driven biomolecular docking, Nat. Protoc. 5 (2010) 883–897.
[105] J.J. Irwin, B.K. Shoichet, M.M. Mysinger, N. Huang, F. Colizzi, P. Wassam, Y. Cao, Automated docking screens: a feasibility study, J. Med. Chem. 52 (2009) 5712–5720.
[106] T.Y. Tsai, K.W. Chang, C.Y. Chen, iScreen: world's first cloud-computing web server for virtual screening and de novo drug design based on TCM database@Taiwan, J. Comput. Aided Mol. Des. 25 (2011) 525–531.
[107] A. Grosdidier, V. Zoete, O. Michielin, SwissDock, a protein-small molecule docking web service based on EADock DSS, Nucleic Acids Res. 39 (2011) 270–277.
[108] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, J. Comput. Chem. 19 (1998).
[109] O. Trott, A.J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem. 31 (2010) 455–461.
[110] P.T. Lang, S.R. Brozell, S. Mukherjee, E.F. Pettersen, E.C. Meng, V. Thomas, R.C. Rizzo, D.A. Case, T.L. James, I.D. Kuntz, Dock 6: combining techniques to model RNA–small molecule complexes, RNA 15 (2009) 1219–1230.
[111] B. Kramer, M. Rarey, T. Lengauer, Evaluation of the FLEXX incremental construction algorithm for protein–ligand docking, Proteins 37 (1999) 228–241.
[112] M. McGann, FRED pose prediction and virtual screening accuracy, J. Chem. Inf. Model. 51 (2011) 578–596.
[113] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, P.S. Shenkin, Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, J. Med. Chem. 47 (2004) 1739–1749.
[114] T.A. Halgren, R.B. Murphy, R.A. Friesner, H.S. Beard, L.L. Frye, W.T. Pollard, J.L. Banks, Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening, J. Med. Chem. 47 (2004) 1750–1759.