

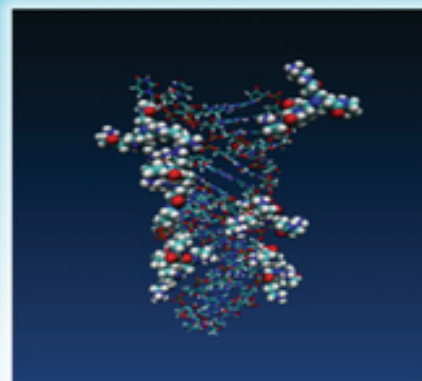
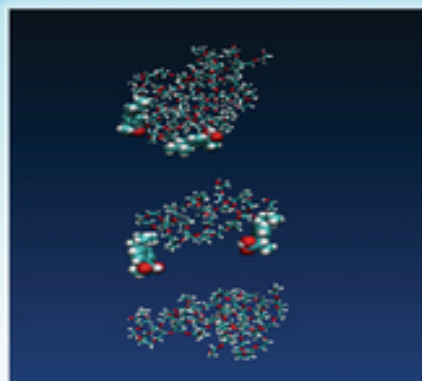
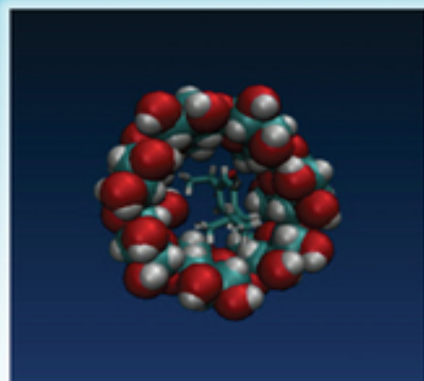
Editors

Defang Ouyang

Sean C. Smith

Computational Pharmaceutics

Application of Molecular
Modeling in Drug Delivery



Preface

While computer-aided drug design (or rational drug design) has been practised for half a century, the application of computational modeling to drug delivery and pharmaceutical formulations “computational pharmaceuticals” has emerged only in recent years. In combination with existing branches of pharmaceuticals, it offers rapidly growing potential for developing rational, deductive, and knowledge-based strategies in pharmaceuticals. Thus, this discipline has emerged and grown enormously in importance. Exploiting the exponential growth in power of high performance computing systems, computational pharmaceuticals has the ability to provide multiscale lenses to pharmaceutical scientists, revealing mechanistic details ranging across chemical reactions, small drug molecules, proteins, nucleic acids, nanoparticles, and powders to the human body.

Given the advances in the field, it is becoming increasingly important that pharmaceutical scientists have cognizance of computational modeling, which may be anticipated to become a more common part of postgraduate curricula. While the fundamental theories and technical implementations of many of the methods in this book are complex, the computational software is becoming much more readily accessible and usable. Hence, it is no longer a prerequisite to understand in detail how the methods work at the theory and computer coding levels in order to gain substantial insights and benefits from carrying out simulations. Written for an audience with little experience in molecular modeling, with a focus on applications, this book will prove an excellent resource for researchers and students working in pharmaceutical sciences. We anticipate

it will be useful not only for pharmaceutical scientists but more broadly for computational chemists looking to move into the pharmaceutical domain, for those working in medicinal chemistry, materials science, and nanotechnology.

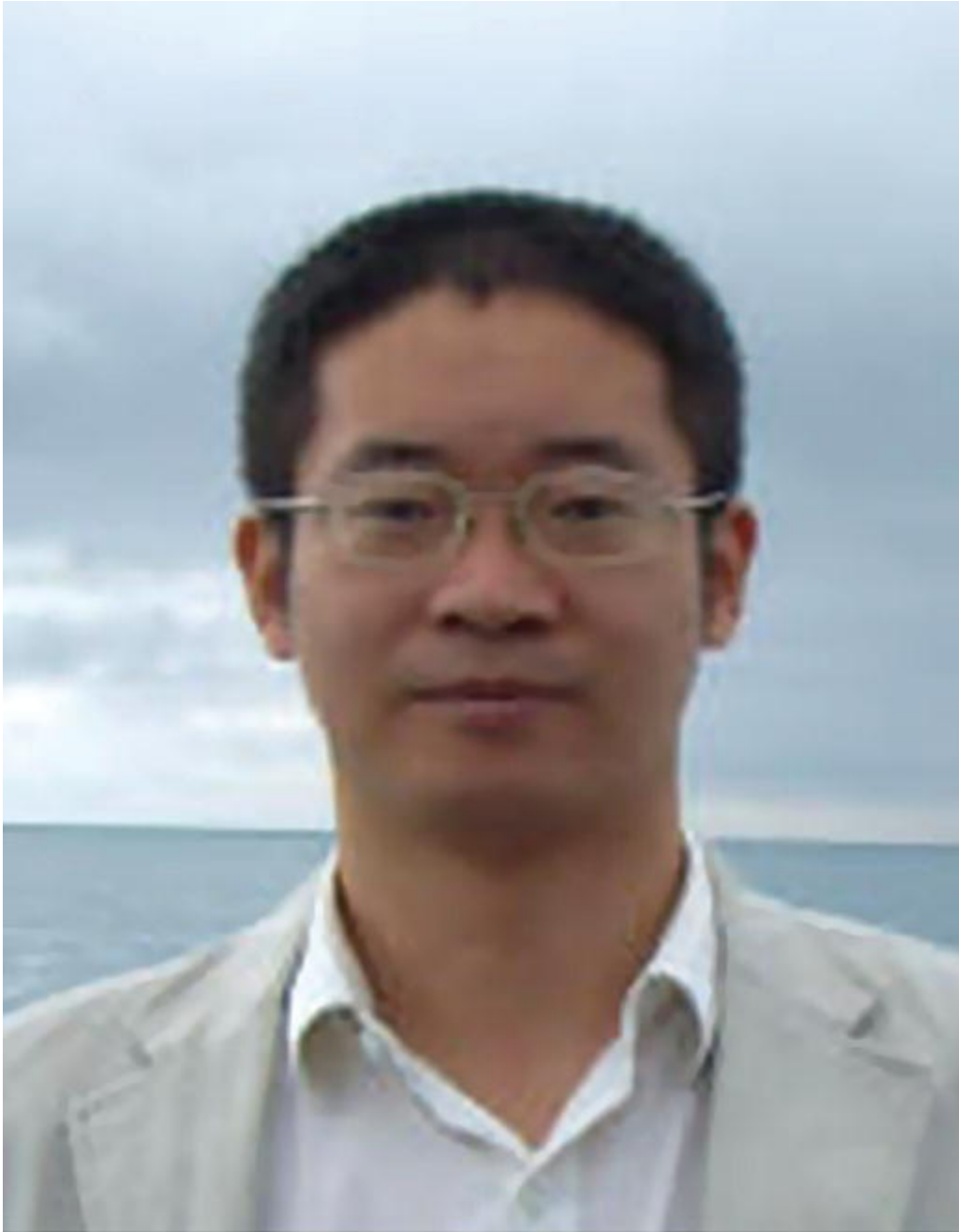
Defang Ouyang

Sean C. Smith

March 2015

Editors' Biographies

Defang Ouyang



Dr Defang Ouyang is an Assistant Professor in the University of Macau, China. He has a multidisciplinary

background in pharmaceuticals and computer modeling, with experience in academia and industry. He completed his PhD in pharmacy at The University of Queensland, Australia, in 2010 and progressed directly to a faculty position at Aston University, UK in January 2011. At the end of 2014, he moved to the University of Macau. Since 2011, he has pioneered the integration of molecular modeling techniques and experimental approaches in the field of drug delivery – “computational pharmaceuticals”. He has published over 20 refereed journal papers in this area. He chaired the Computational Pharmaceuticals Workshop at the 2014 Controlled Release Society Annual Meeting (Chicago, USA). Currently he supervises one PhD student and two MSc students. He has also successfully supervised one PhD student and 18 Master research projects.

Sean C. Smith



Professor Sean C. Smith received his PhD in theoretical chemistry at the University of Canterbury, New Zealand, in 1989. Following an Alexander von Humboldt Fellowship at the University of Göttingen, Germany and postdoctoral research at the University of California, Berkeley, USA he accepted a faculty position at The University of Queensland (UQ), Australia in 1993. He became Professor and Founding Director of the UQ Centre for Computational Molecular Science in 2002 and in 2006 his laboratory moved to the Australian Institute for Bioengineering and Nanotechnology at UQ. He was a founding chief

investigator, computational program leader, and Deputy Director in the Australian Research Council Center of Excellence for Functional Nanomaterials 2002–2010. In 2011, he undertook the role of Director of the Department of Energy funded Center for Nanophase Materials Sciences (CNMS) at Oak Ridge National Laboratory, USA. In 2014 he moved to the University of New South Wales (UNSW) Australia as Professor of Computational Nanomaterials Science and Engineering and Foundation Director of the Integrated Materials Design Centre. His specific research involves theoretical and computational studies of structure, reactivity and catalysis, photochemistry, electrochemistry and transport phenomena within nanomaterials, proteins, and hybrid nano-bio systems as well as his long-time passions of chemical kinetics and reaction dynamics. He has published over 230 refereed journal papers. In 2006 he was recipient of a Bessel Research Award from the Alexander von Humboldt Foundation in Germany and in 2012 he was elected Fellow of the American Association for the Advancement of Science (AAAS).

1

Introduction to Computational Pharmaceutics

Defang Ouyang¹ and Sean C. Smith²

¹Institute of Chinese Medical Sciences, University of Macau, Macau

²School of Chemical Engineering, University of New South Wales, Australia

1.1 What Is Computational Pharmaceutics?

It is a given that active pharmaceutical ingredients (APIs) should be made into safe and effective dosage forms or formulations before administration to patients.

Pharmaceutics is the discipline to make an API into the proper dosage form or medicine, which may then be safely and effectively used by patients [1]. Pharmaceutics also relates to the absorption, distribution, metabolism, and excretion of medicines in the body. Branches of pharmaceutics include formulation development, pharmaceutical manufacture and associated technologies, dispersing pharmacy, physical pharmacy, pharmacokinetics, and biopharmaceutics [1–4]. Today there are various dosage forms, such as tablets, capsules, solutions, suspensions, creams, inhalations, patches, and recently nanomedicines (e.g., liposomes, nanoparticles, nanopatches). Although numerous new techniques have been developed for the form of dosage, current development of drug formulations still strongly relies on personal experience of pharmaceutical scientists by trial and error in the laboratory [1]. The process of formulation

development is laborious, time-consuming, and costly. Therefore, the simplification of formulation development becomes more and more important in pharmaceutical research. *Computational pharmaceuticals* involves the application of computational modeling to drug delivery and pharmaceutical nanotechnology. In combination with existing branches of pharmaceuticals, it offers rapidly growing potential for developing rational, deductive and knowledge-based strategies in pharmaceuticals.

With stunningly rapid advances in hardware, theory and software algorithms, computer simulation is now able to model complex systems, which may be difficult, costly, or even impractical to measure or monitor directly by experiment [5, 6]. The first example of computer modeling was the simulation of the nuclear bomb process in the Manhattan Project, World War II. With the development of high performance computing, multiscale modeling techniques have been widely pursued, from quantum mechanics (QM) and molecular dynamics (MDs) to stochastic Monte Carlo methods, coarse grained dynamics, discrete element methods (DEMs), finite element methods as well as advanced analytical modeling. In principle, all properties of all systems are able to be described by QM. However, first principle calculations are limited to small systems, <1000 atoms, which is impractical for solving applications of large molecules or systems [5, 6]. MD simulations mimic the physical motion of atoms and molecules under Newton's laws of physics, which is applicable to larger systems containing millions of atoms [5, 6]. MD simulation is based on molecular mechanics, which models the interactions between atoms with force fields. Monte Carlo (MC) simulation uses the same empirical force field as MD simulation. However, MC simulation features playing and recording the results in casino-like conditions by repeated random sampling [5, 6].

Thus, unlike MD simulation, MC simulation cannot offer dynamical information with time evolution of the system in a form suitable for viewing. MC methods are especially useful for modeling systems with significant uncertainty and high degrees of freedom, such as polymer chains and protein membranes. For much larger systems, coarse-grain models do further classical approximations by treating functional groups as rigid bodies of constrained particles [5, 6]. The DEM is one of the numerical methods for computing the motion and effect of a large number of small particles, which is widely used in the pharmaceutical process and manufacturing [5, 6].

In the past three decades, the application of computational modeling approaches in the field of drug design (e.g., QSAR, ligand docking) has been intensively developed to the point of being a mature field [7]. Pharmaceutical research is, however, a far broader field than drug design alone. Proceeding beyond drug design, the application of computational modeling to drug delivery and pharmaceutical nanotechnology, *computational pharmaceuticals*, is a very new field with great potential for growth [8]. As shown in [Figure 1.1](#), computational pharmaceuticals has the ability to provide multiscale lenses to pharmaceutical scientists, revealing mechanistic details ranging across the chemical reactions of small drug molecules, proteins, nucleic acids, nanoparticles, and powders with the human body. The aim of this book is to provide a contemporary overview of the application of computational modeling techniques to problems relating to pharmaceuticals (drug delivery and formulation development) that will be of great relevance for pharmaceutical scientists and computational chemists in both industry and academia. Contributions from leading researchers cover both computational modeling

methodologies and various examples where these methods have been applied successfully in this field.

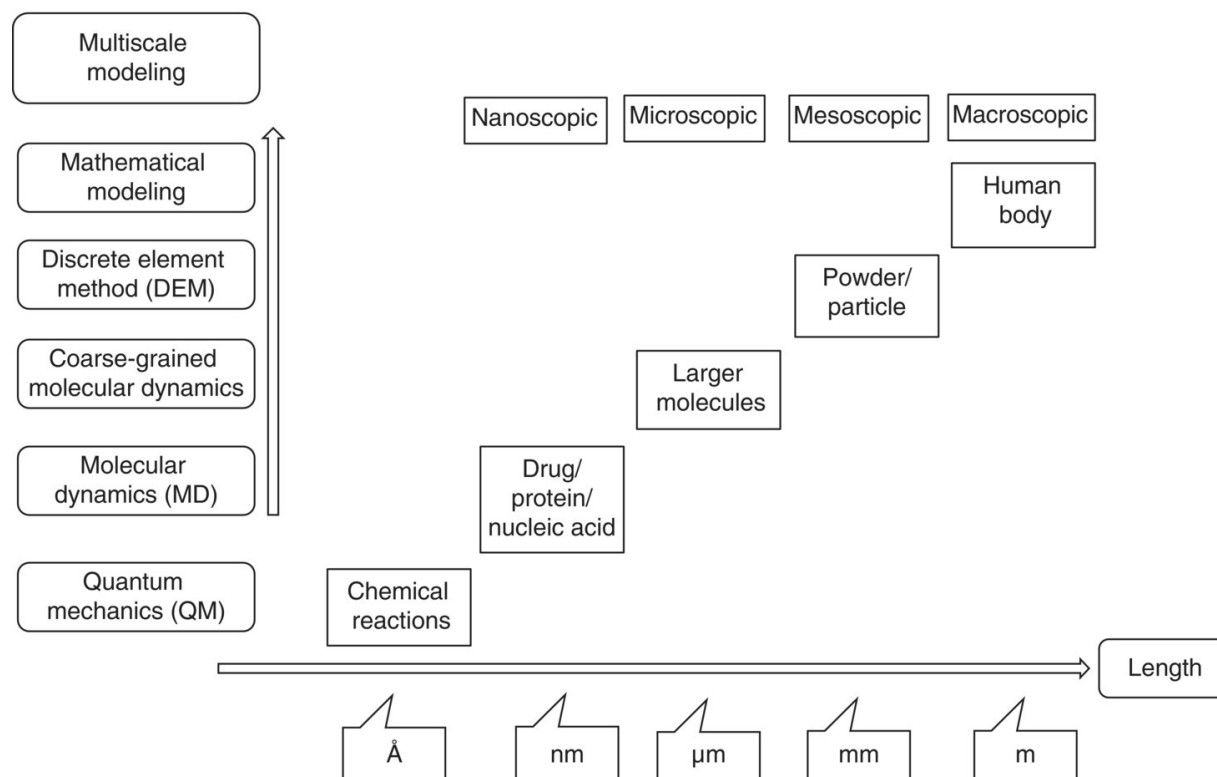


Figure 1.1 Application of computational pharmaceuticals.

1.2 Application of Computational Pharmaceuticals

Polymorphism of small drug molecules (e.g., crystal, hydrate, solvate, salt, cocrystal) plays a very important role in pharmaceutical research because it greatly influences the dissolution behavior and the bioavailability of pharmaceutical products. Thus, crystal structure prediction (CSP) methods have gained wide attention from pharmaceutical scientists. [Chapter 2](#) discusses the general principles of CSP and recent progress in solid form screening by the crystal energy landscape method.

Cyclodextrins are a family of cyclic oligosaccharides, which are widely used in drug delivery and formulations for the solubilization of poorly soluble drugs. In [Chapter 3](#), the physicochemical principles of cyclodextrin/drug complexation, experimental characterization and recent theoretical progress of drug/cyclodextrin modeling are discussed.

Polymeric-based micellar vehicles have been widely used in pharmaceuticals for the delivery of both hydrophilic and lipophilic drugs. Multiscale modeling for polymeric-based vehicles for drug delivery is discussed in [Chapter 4](#), including different computational approaches, micellar self-assembly and stability, the interaction of Taxol with model cellular membranes, and Taxol-tubulin association.

Solid dispersion refers to the dispersion of drug molecules in carriers in a solid state, prepared by the hot melting method. Poor physical stability has strongly hindered the commercialization of this technique. [Chapter 5](#) discusses the possible molecular structure of amorphous solid dispersions and the mechanism of physical stability of this technology.

Biological lipid membranes are the key to drug absorption and bioavailability. Liposomes, artificially prepared vesicles with a phospholipid bilayer, are widely employed in drug delivery for cancer and other diseases. [Chapter 6](#) reviews the theoretical progress of lipid membrane models, small-molecular uptake and permeation across lipid membranes, nanoparticle-membrane interaction, and the mechanism of chemical penetration enhancers.

Protein/peptide drugs are becoming increasingly important in the pharmaceutical market. However, the development of stable and effective formulations for biopharmaceuticals is still quite challenging for pharmaceutical scientists. [Chapter 7](#) summarizes the diverse modeling results on the

solution behavior of protein formulation, including protein aggregation pathways in liquid formulations, protein-cosolvent interactions, and protein-protein interactions.

Inorganic nanoparticles had been increasingly utilized for drug/gene delivery in recent decades. One of the main advantages of nanoparticle drug delivery systems is the targeting effect to specific organs and tissues. [Chapter 8](#) discusses recent progress in computational modeling of inorganic nanoparticle drug delivery systems: carbon nanotubes, graphene/graphene oxide, silica, and gold nanoparticles.

Although the concept of nanodiamonds (diamond nanoparticles) for drug delivery is still in its infancy, recently nanodiamonds have been widely studied for bio-imaging and drug targeting for improved chemotherapeutic effect. [Chapter 9](#) reviews the structure of the individual nanodiamond, its surface chemistry and interactions, and nanodiamond drug delivery systems.

Layered double hydroxides (LDHs) are composed of nanoscale cationic brucite-like layers and exchangeable interlayer anions. LDH nanoparticles are an efficient drug delivery system for anionic chemicals, such as small drug molecules (e.g., methotrexate, heparin) and nucleic acids (RNA and DNA). In [Chapter 10](#), different computational approaches are discussed to investigate the properties and interactions of LDH/anion systems.

The structure of particles or powders plays an important role in many dosage forms, such as tablets, granules, and capsules. However, the microstructure of these dosage forms is less well investigated. [Chapter 11](#) reviews the principles of synchrotron radiation-based microtomography (SR- μ CT) and its application to determine the particulate architecture of granules, osmotic pump tablets, and HPMC matrix tablets.

Physiology-based pharmacokinetics plays an important role in pre-clinical drug development and formulation development. [Chapter 12](#) discusses the principles of pharmacokinetic modeling and simulation and commercially available models for pharmaceutical scientists.

1.3 Future Prospects

“Today the computer is just as important a tool for chemists as the test tube” (Karplus, Levitt, and Warshel, Nobel Prize in Chemistry 2013). Analogous to the paradigm shift of drug development in the past three decades by computer-aided drug design, computational pharmaceuticals also has great potential to shift the paradigm of drug delivery research in the near future [8].

References

- [1] Aulton, M.E. and K. Taylor, *Alton's Pharmaceuticals - The Design and Manufacture of Medicines* 4th edn. 2013, Elsevier.
- [2] Sinko, P.J., *Martin's Physical Pharmacy and Pharmaceutical Sciences* 6th edn. 2010, Philadelphia, PA: Lippincott Williams and Wilkins.
- [3] Florence, A.T. and D. Attwood, *Physicochemical Principles of Pharmacy* 5th edn 2011, London: Pharmaceutical Press.
- [4] Allen, L.V., N.G. Popovich, and H.C. Ansel, *Ansel's Pharmaceutical Dosage Forms and Drug Delivery Systems* 9th edn. 2010, Philadelphia, PA: Lippincott Williams and Wilkins.
- [5] Leach, A.R., *Molecular Modelling: Principles and Applications*. 2nd edn 2001, New York: Prentice Hall.

- [6] Young, D.C., *Computational Chemistry: A Practical Guide for Applying Techniques to Real World Problems*. 2001, New York: John Wiley & Sons, Inc.
- [7] Seddon, G., V. Lounnas, R. McGuire, *et al.*, Drug design for ever, from hype to hope. *Journal of Computer-Aided Molecular Design*, 2012. **26**(1): 137-150.
- [8] Ouyang, D., A. Bunker, A. Rostami, and J. Zhang, Computational pharmaceuticals: the application of computer modelling to drug delivery. *CRS Newsletter*, 2014. **31**(2): 16-18.

2

Crystal Energy Landscapes for Aiding Crystal Form Selection

Sarah L. Price

*Department of Chemistry, University College London,
UK*

2.1 Introduction

The fundamental interrelationships between crystal structure, properties, performance and processing of a drug form the basis of “turning the development of pharmaceutical products from an art into a science” [1]. Thus, the development of an active pharmaceutical ingredient (API) into a pharmaceutical product requires detailed knowledge of its solid state forms and their physical properties. Even if the drug is not going to be delivered in a solid tablet, for example from a transdermal patch, there is a need to avoid crystallisation in the product. The manufacturing process will usually involve crystallisation as a purification step, possibly for separating out enantiomers as well as less similar synthetic impurities. Hence the design of the manufacturing process requires information about the solid forms that could be involved, including any hydrates or solvates [2].

An API rarely has only one crystalline phase. There may be polymorphs, hydrates, solvates, salts or cocrystals, though only cases where the solvent, counterion or coformer is pharmaceutically acceptable can be considered for development into a drug product. Salts and cocrystal forms can be used to improve solubility and dissolution rate [3]. Hence, a careful consideration of all alternatives is needed

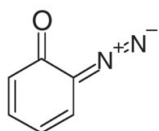
to select the physical form that will be used in the product, and the stability of multicomponent crystalline forms relative to their components is a key property [4].

Polymorphism [5, 6], the ability of a molecule (or defined stoichiometry hydrate, salt or cocrystal) to adopt more than one crystal structure, plays an important role in pharmaceutical development, because the physical properties of a crystal, most critically the solubility and dissolution rates, can differ between polymorphs. Thus, the pharmaceutical product must contain just the approved polymorphic form. A drug is most often formulated using the most stable polymorph to avoid potential problems of phase transformations during storage. However, since metastable polymorphs or even amorphous forms may have better properties, particularly being more soluble, a metastable form may be developed provided that transformation to the more stable form during production and storage can be excluded. Unfortunately, molecules do not necessarily readily crystallise in their most stable form: indeed Ostwald's Rule of Stages [7] encapsulates the observation that metastable forms tend to be the first observed. The problem is that, once a more stable form has been nucleated, it can prove very difficult if not practically impossible to produce the metastable crystal structure, giving rise to the phenomenon of disappearing polymorphs [8]. There was a huge expansion of industrial and academic research into organic polymorphism after the Norvir® crisis, when Abbott Laboratories had to reformulate ritonavir at considerable expense after a novel, thermodynamically more stable, considerably less soluble polymorph emerged two years after the launch [9]. This led to the development of many solid form screening and characterisation methods [10]. However, polymorphism still causes problems within the industry, such as the crystallisation of rotigotine within the transdermal patches

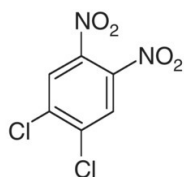
being used to deliver it to sufferers of Parkinson's disease in 2008 [11]. How can you ensure experimentally that the most stable crystalline form is known? One rare example of knowing that the most stable form has not yet been crystallised has been reported for a melatonin agonist, because the inactive enantiomer has two polymorphs and despite extensive efforts only the corresponding metastable form can be found for the active enantiomer [12]. The rationalisation is that the inactive enantiomer has come into contact with a chiral substance or surface capable of nucleating its most stable polymorph. It is the uncertainty in knowing that the most stable form has been crystallised that led to considerable practical interest in developing a computational method of predicting the most stable crystal structure of a given API [4].

Crystal structure prediction (CSP) methods gained their name decades ago, when people were trying to predict the crystal structure of a molecule, assuming that polymorphism was unlikely. The challenge was to predict the crystal structure of a molecule from just the chemical diagram. Success would indicate that all the factors which determined which crystal structure was adopted had been incorporated into the model. Most methods assumed that this involved searching for the most thermodynamically stable crystal structure, assuming that thermodynamics was the major determinant. The hope behind the first CSP programs was that there would be a significant energy difference between the observed structures and the hypothetical structures generated, a hope that was quickly dashed [13, 14]. The growing realisation that the prediction of crystal structures was industrially relevant, as well as being a fundamental scientific challenge, led to the Cambridge Crystallographic Data Centre organising blind tests of CSP [15–19]. In these tests, groups developing CSP methods are sent the chemical diagrams of the molecules

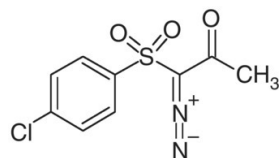
whose solved crystal structures are kept confidential until the groups have submitted three predictions by a deadline. The accounts of the discussion of the results form an excellent review of both the variety of methodologies and the progress in the field, as well as showing the low success rate ([Figure 2.1](#)). In the fourth blind test (CSP2007), one group had the remarkable success of correctly predicting all four targets [20]. In the following test, the variety of target systems was increased to include a molecular salt, polymorphs of a hydrate and a molecule of a size that is more relevant to pharmaceutical development ($C_{25}H_{22}N_2O_4S_2$, XX in [Figure 2.1](#)). The crystal structure of this “model pharmaceutical” required a significant adaption of the methodology and was successfully predicted by two groups [21] but not by the approach which had been successful for the smaller molecules. Hence, the methodology of CSP studies is still actively evolving and is a challenge to computational chemistry methods. Future blind tests will help to evaluate progress, with computational costs (currently measured in CPU years for larger molecules) being only part of the challenge.



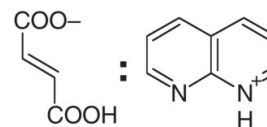
(XVI) - 2/14



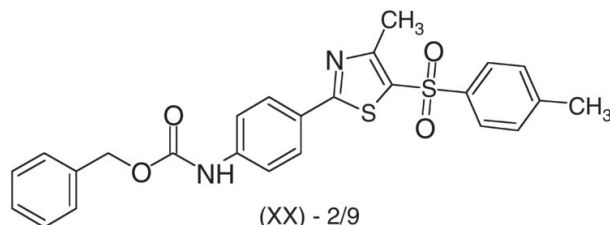
(XVII) - 2/12



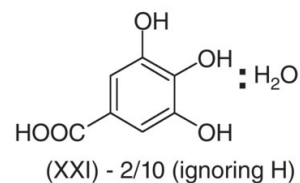
(XVIII) - 1/13



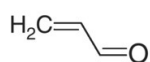
(XIX) - 2/11



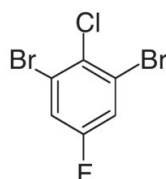
(XX) - 2/9



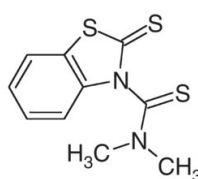
(XXI) - 2/10 (ignoring H)



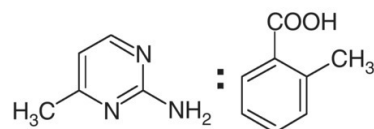
(XII) - 4/13



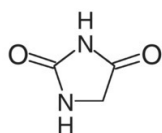
(XIII) - 4/14



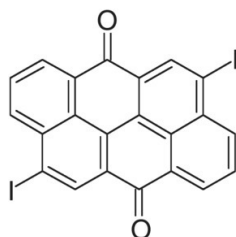
(XIV) - 3/12



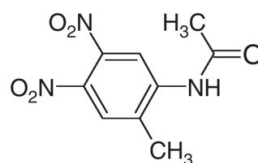
(XV) - 2/12



(VIII) - 4/14



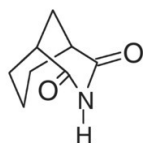
(IX) - 1/14



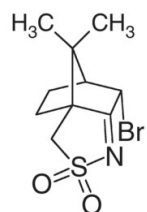
(X) - 0/15



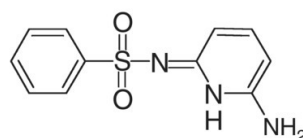
(XI) - 0/18



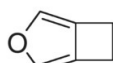
(IV) - 2/15



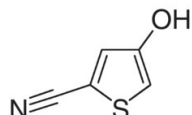
(V) - 4/15 (pure enantiomer)



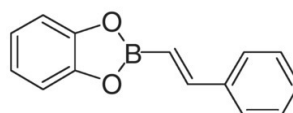
(VI) - 0/11



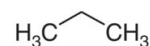
(I) -
0/11 (stable polymorph)
4/11 (metastable polymorph)



(II) - 1/8



(III) - 1/11



(VII) - 1/6

Figure 2.1 *Molecules used in the CCDC blind tests of crystal structure prediction. The Roman numerals denoting the targets are approximately in year order: I-III and VII CSP1999 [15], IV-VI CSP2001 [16], VIII-XI CSP2004 [17], XII-XV CSP2007 [18], and XVI-XXI CSP2010 [19]. The success rate is given as x/y, where x of the y groups which submitted three predictions included the experimental structure. For XXI the challenge was to predict further polymorphs.*

Polymorphism has always been acknowledged as a complicating factor in the interpretation of blind test results. Indeed further experimentation after blind tests found novel polymorphs for XXI [22], IV [23] and VI [24]. Whilst the thermodynamically most stable structure has to be found, which of the structures that are competitive in energy should correspond to polymorphs? CSP methods usually generate more plausible crystal structures than observed polymorphs, raising many questions about organic crystallisation [25]. This has led to the concept of a crystal energy landscape, a set of structures that are sufficiently low in energy to be potential polymorphs [26], and the question as to how these energy landscapes relate to the experimental crystal structures that the molecule could adopt. More practically, how can the calculated crystal energy landscapes help establish the true diversity of solid forms and act as a complement to other polymorph screening techniques?

Thus this chapter is divided into two sections. First, the general principles behind the computational approaches [27] that have been successful in blind tests will be outlined, with particular emphasis on the challenges for pharmaceutical molecules with their greater structural complexity. Pharmaceutical applications of CSP differ from those of computational materials design, where the aim is to suggest molecules whose crystals will have particular

properties and avoid the synthesis of unpromising molecules. For example, an isolated molecule can have a large nonlinear optical coefficient, but if it crystallises in a centrosymmetric space group, the crystalline material will not be optically active. Molecular electronics, energetics, pigments and porous frameworks are other classes of functional materials where CSP methods have been applied. The blind tests are more closely aligned with the materials design problem, whereas for pharmaceutical applications, there may already be relevant experimental information that can be used to focus computational efforts, or known problems that need to be understood at the molecular level. The second section will therefore review some recent work where the crystal energy landscapes have been used as a complement to detailed experimental work on model pharmaceuticals, which indicates the potential use of CSP studies to complement experimental solid form screening.

2.2 CSP Methods for Generating Crystal Energy Landscapes

CSP methods [28] based on searching for the most thermodynamically stable crystal structures have to compromise between the number and range of crystal structures considered and the approximations used in evaluating the crystal energy. Current methods are based on the lattice energy as an approximation to the crystal energy, that is the energy of a completely ordered crystal at 0 K ignoring zero-point vibrational effects. This is only seeking the group of energetically favourable crystal structures and does not necessarily give the correct relative ranking of enantiotropically related polymorphs, whose relative stability order changes with temperature. The lattice energy is defined as the energy of an infinite,

perfect, static crystal relative to the molecules in their lowest energy conformation at infinite separation.

Early work on CSP concentrated on simple rigid molecules, and in blind tests the molecular diagram is given ([Figure 2.1](#)). In applications, it is necessary to check whether you are modelling the appropriate molecules, and not their isomers or tautomers. The crystal energy landscapes are very different for different isomers [29], structural analogues or chemically closely related molecules [30]. A proton position can make a huge difference, for example the expected hydrogen bonding motif between a pyridine and a carboxylic acid is the same as between the corresponding pyridinium ion and carboxylate group, such that it would be reasonable to expect the crystal structures to be essentially the same apart from the acidic proton. However, comparing the crystal energy landscapes for three pyridinium carboxylate salts with those of the corresponding pyridine:carboxylic acid cocrystals showed considerable differences in the lower energy structures, except in the case where the proton is found to be disordered [31].

2.2.1 Assessment of Flexibility Required in Molecular Model

The definition of the lattice energy leads naturally to its separation into an intermolecular part, U_{inter} , and the intramolecular energy penalty paid for changing the conformation of the molecule to improve the molecular packing, ΔE_{intra} , giving $E_{\text{latt}} = U_{\text{inter}} + \Delta E_{\text{intra}}$. There is an important distinction between minor conformational changes and those that change the shape of the molecule and correspond to different conformers [32]. Minor changes of a few degrees in torsion angles, or amide pyramidalisation, methyl rotation or similar changes in

proton positions for given covalent bonding can have a significant effect on the lattice energy but are usually refined in later stages of generating the crystal energy landscapes. Major changes, such as rotations about torsions which result in a significant change in the molecular shape or the relative positions of hydrogen bonding donors and acceptors, need to be considered explicitly within the search to ensure that the possibility of any of these conformations generating low energy crystal structures is considered. In general, molecules adopt low energy conformations in their crystal structures [32, 33]. However, there are exceptions, for example when the low energy conformation of the isolated molecule has an intramolecular hydrogen bond, which does not appear in the crystal structure because an intermolecular hydrogen bond is more favourable. If the molecule has different conformations that are in deep energy wells, then separate searches with each conformation held rigid are adequate. However, when there are low energy barrier torsions which change the shape of the molecule, such as between aromatic rings, then the flexibility to change shape has to be included in the first stages of the search. One of the important lessons from the crystal structure of XX in the 2010 blind test ([Figure 2.1](#)) is that the conformations adopted by larger molecules, whilst low in energy, can correspond to rather different molecular shapes than the conformational energy minima [21]. [Figure 2.2](#) illustrates which torsions were considered as flexible in the search, and which only refined later, for a range of the molecules whose CSP studies are discussed in this chapter. The range of conformations to be considered in the search is one of the key limitations in the type of molecule that can be tackled or the computer time needed. Hence the development of strategies for CSP studies of flexible molecules depending on both the molecular size and flexibility and purpose of study is an active area, with

phenobarbital [34] and the tyrosine kinase inhibitor axitinib (Inlyta®) [4, 35] giving complementary examples to those outlined in this chapter. The critical conformational analysis may well have been carried out in the drug design process for pharmaceuticals. Hence a combination of conformer generators and analysis of the conformations of fragments in known crystal structures of related molecules in the Cambridge Structural Database (CSD) [36] can be used to suggest conformational regions to be included in the search. Care is needed to ensure that the flexibility and relative energies are sufficiently realistic to ensure that the search covers the required region and does not disfavour the likely structures. For example, a conformational energy scan for the low energy torsion of tolfenamic acid ([Figure 2.2](#)) performed with a routine *ab initio* molecular orbital method [self-consistent-field calculations with a 6-31+(d) basis set] had a minimum where more accurate methods that represented the correlation in electron motions, and hence modelled the intramolecular dispersion, had a low energy maximum. Although this maximum was only about 3 kJ mol⁻¹ for tolfenamic acid or 6 kJ mol⁻¹ for unsubstituted fenamic acid, the distribution of this torsion angle in crystal structures of related molecules correlated with the more realistic *ab initio* method [37]. Thus, the CrystalPredictor methodology uses an interpolated grid representing the conformational energy calculated at a sufficient level of theory to be qualitatively correct [38, 39], and the GRACE procedure uses a tailor-made force field [40] for the specific molecule derived by extensive dispersion-corrected density functional calculations [41]. The computational cost of treating flexibility at this level of accuracy is quite prohibitive, but has to be balanced by the possibility of the search procedure missing key structures if traditional force fields are used.

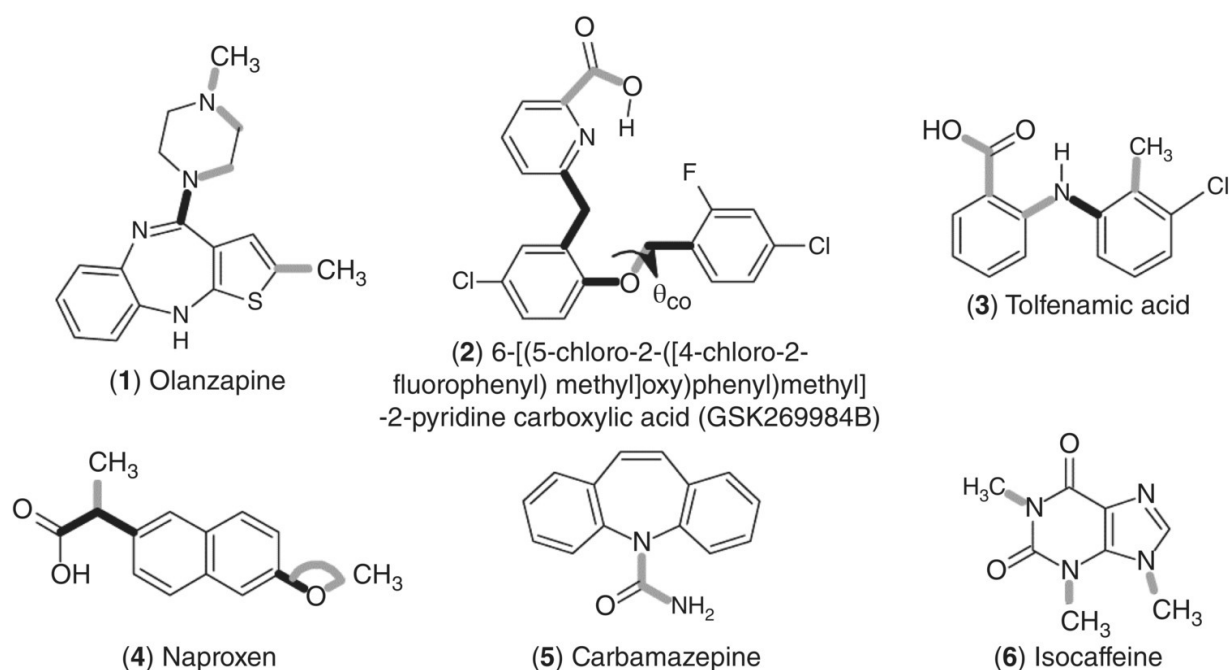


Figure 2.2 *Examples of the degree of flexibility considered during the generation of crystal energy landscapes. Changes to torsion angles marked in bold black are fundamental to the shape of the molecule and need to be considered at the outset of the search. The lighter grey bold angles only need to be refined at later stages of the structure optimisation. For (2) these approaches are combined in that the search was separated into nine subsearches in which one of three conformations of the CO₂H group and one of the three values of θ_{CO} was initially fixed and refined later.*

2.2.2 Search Method for Generating Putative Structures

A second choice which determines the computational cost and value of the CSP search is the coverage of crystallographic space. The number of independent molecules in the asymmetric unit cell (Z) is a critical choice, as the relative positions of these molecules vastly increase the number of search variables. This greatly adds

to the cost of generating crystal energy landscapes for multicomponent systems, such as salts, cocrystals and hydrates. However, the usual procedure of restricting the search to one molecule in the asymmetric unit cell for single component systems ($Z' = 1$) will miss the crystal structures with $Z' > 1$. Many crystal structures with $Z' > 1$ are sufficiently closely related to $Z' = 1$ structures that this restriction is unimportant. Such structures can be generated in the search when the $Z' = 1$ structures are found not to be true minima from examination of the second derivative properties such as elastic tensor or phonon modes. However, there are crystal structures in which different molecules have different hydrogen bonds or different conformations and so intrinsically could only be found in a $Z' > 1$ search. Usually only a $Z' = 1$ search is worthwhile, unless there is evidence, for example from solid state NMR, that a $Z' = 2$ form needs to be structurally characterised. All experimentally known crystal structures can be used to check the validity of the crystal energy landscape by calculating the corresponding lattice energy minima. The lattice energies of all known structures should be among the most stable. This comparison can reveal that the molecule adopts a $Z' > 1$ structure because it has a lower energy than $Z' = 1$ possibilities [29, 37].

The next choice is the range of space groups that are considered in the search. If the molecule is chiral and an enantiomerically pure sample is being crystallised, as is often the case for pharmaceuticals, then it can only adopt a limited number of space groups, the most likely being $P2_12_12_1$, $P2_1$, $P1$, $C2$ and $P2_12_12$. If the molecule is not chiral, or both enantiomers are present, it can adopt a far wider range of space groups that contain an inversion centre, mirror or glide plane. The most popular space group for structures in the CSD is $P2_1/c$, and for a $Z' = 1$ structure in this space group, there are the three cell

lengths, one cell angle and the position of the molecule relative to the unit cell axes as variables in the search. The ability to distinguish between an enantiopure and racemic crystal structure means that CSP is being developed to increase our understanding of the resolution of enantiomers by crystallisation, either through spontaneous resolution [42] or by diastereomeric salt formation [43]. For example, CSP was used [44] to solve the crystal structure of the racemic form of the nonsteroidal anti-inflammatory naproxen ([Figure 2.2](#)), which required consideration of closely related $Z' = 1$ and 2 structures, as well as confirmation that the known chiral structure was the most stable for the single enantiomer. Restricting the search to the chiral space groups will save a great deal of time for chiral APIs, but may exclude gaining valuable insights into the crystallisation process, particularly in cases when the chiral purity is suspect, for example if the barrier to racemisation is low because the chirality is conformational rather than from chiral sp^3 carbon atoms. In the case of tazofelone, there is a racemic solid solution that is isostructural with the enantiopure crystal that nucleates it, demonstrating that chiral impurities may be readily incorporated [45].

Evaluating the success of different search algorithms in covering defined areas of crystallisation space has been a major benefit of the blind tests [46]. In practical applications, the appropriate method will depend on the purpose of the study and timeline factors. For example, there is a rapidly growing use of CSP to aid the determination of the structure in cases where suitable single crystals cannot be grown, where the powder X-ray diffraction [47], solid state NMR [48] or electron diffraction data [49] may provide information that can be used to reduce the search space drastically. The algorithm choice also depends on the purpose of the search, and the blind

test papers give details of the ever-emerging range and the assumptions behind them. For example, MOLPAK is based on density searching and designed for energetic materials research [50], PROM has been designed for easy use by crystallographers for small rigid molecules [51] and PolymorphPredictor was designed for industrial use based on simulated annealing. The programs which are currently being used for published pharmaceutical applications are GRACE [52] and CrystalPredictor [35, 38, 39], with other programs being developed for use on high-performance computer clusters [53]. An indication of the data generation and handling requirements is that a general search for a small rigid molecule should produce at least a few thousand plausible structures to be credible, whereas searches for flexible pharmaceuticals, monohydrates of smaller molecules or cocrystals, can readily generate a million reasonable crystal structures and still not satisfy the criteria for a complete search.

2.2.3 Methods for Computing Relative Crystal Energies

The number of plausible crystal structures that are generated makes a hierarchical system of improving the relative lattice energies, using increasingly expensive and accurate methods on a reducing number of the most promising crystal structures, inevitable. For example, the GRACE procedure uses a tailor-made force field [40] to generate the structures and then uses periodic density functional lattice energy minimisation (with an empirical dispersion correction specifically developed for CSP studies [41]) to refine the structures, working up in energy until it is clear that the reranking produced by calculations on further structures will not generate any more structures within the energy range of likely polymorphism.

This immediately begs the question of what is the energy range of potential polymorphism, as well as what is the accuracy required in evaluating the relative energies of different crystal structures of the same molecule? The longevity and practical importance of a metastable polymorph is mainly determined by the barrier to transformation to the most stable form, as this determines whether or not a crystal form is so long lived that it appears to be stable. Thermal calorimetric measurements on polymorphic transformations, heats of fusion and melting temperatures can be used to construct energy temperature diagrams for known polymorphs, to give the order of stability at 0 K for comparison with lattice energy rankings. Measured enthalpies suggest that lattice energy differences between observed polymorphs are small, usually of order of a few kJ mol^{-1} . Various reviews of computational methods applied to solid form pharmaceutical modelling reveal that predicting the relative stability of known polymorphs is challenging [54, 55]. Moreover, even for smaller molecules, there is considerable research into developing periodic electronic structure methods capable of predicting lattice energies within the kcal mol^{-1} target that is usually considered “chemical accuracy” [56].

The main alternative approach to electronic structure calculations on the crystal structures requires electronic structure calculations on the molecule in each conformation within the crystal structures. As well as providing an estimate of the intramolecular energy penalty paid for changing the molecular conformation within the crystal structure, ΔE_{intra} , the molecular charge density can be analysed to provide an atomistic model for the intermolecular interactions and evaluate the intermolecular contribution to the lattice energy, U_{inter} , using DMACRYS