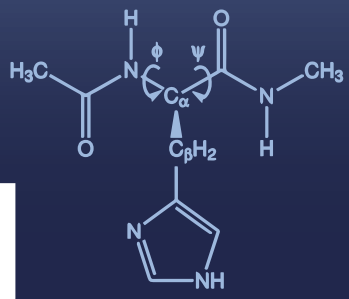
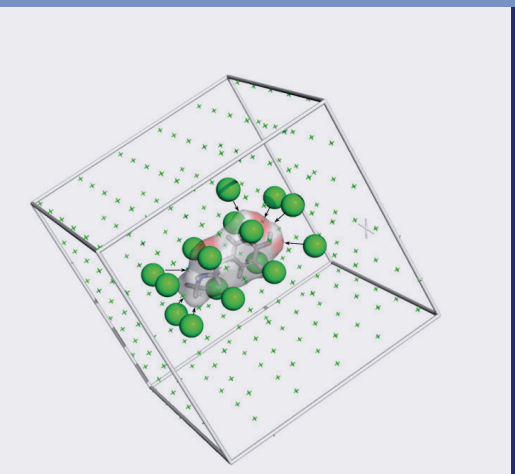
CRC Press
Taylor & Francis Group

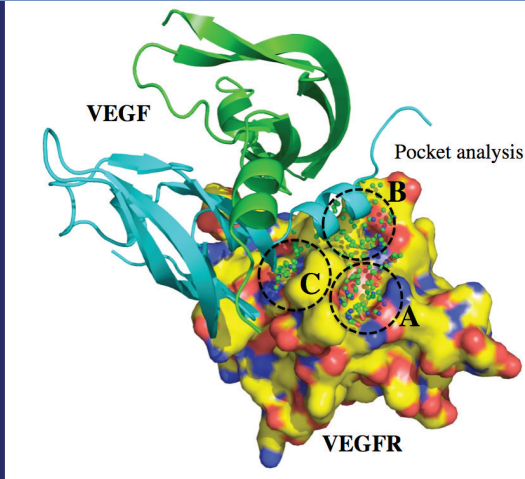


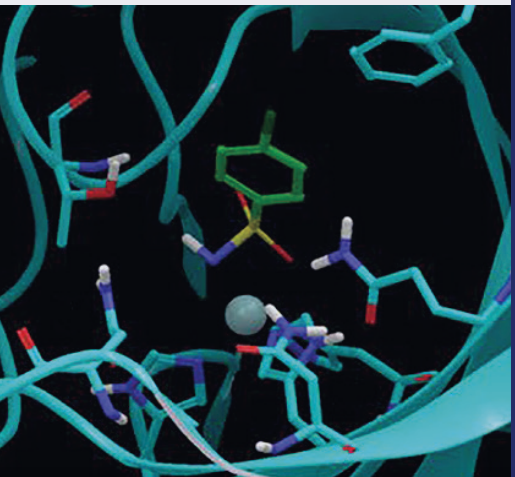
In Silico Drug Discovery and Design

Theory, Methods, Challenges, and Applications

Edited by Claudio N. Cavasotto







In Silico Drug Discovery and Design

Theory, Methods, Challenges, and Applications

In Silico Drug Discovery and Design

Theory, Methods, Challenges, and Applications

Edited by **Claudio N. Cavasotto**

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20150303

International Standard Book Number-13: 978-1-4822-1785-8 (eBook - PDF)

Contents

Section II Advanced Techniques

Section III Challenges

Preface

In silico methods are today a solidly established key component in early drug lead discovery (Jorgensen 2009). A plethora of computational tools are routinely used to identify and select therapeutic relevant targets, study the molecular basis of ligand–protein interactions, structurally characterize binding sites, develop target-specific compound libraries, model target proteins, identify hits by ligand- and structure-based virtual screening, estimate binding free energy, and optimize lead compounds, all of which can be used to rationalize and increase the efficiency, speed, and cost-effectiveness of the drug discovery process.

The ever-increasing availability, and decreasing cost of computational power, algorithmic and software development, and the large number of web servers have contributed to the success of computational drug lead discovery. However, it should be acknowledged that progress in some areas seems to have slowed down, and rethinking and innovation both in terms of the perspective of the problem as well as in the computational tools themselves are still needed. Moreover, Peter Goodford's perspectives after the "3D Molecular Structure and Drug Action" meeting in Erice, Italy, in 1989, on where the field should go, may sound surprisingly current (Van Drie 2007).

It is reasonable to think that more accurate and reliable methods would surely help to overcome the stagnation in the number of approved drugs in recent years, especially if *in silico* discovery and optimization of high-affinity ligands is coupled with druggability assessments early in the drug discovery process. In particular, a deeper understanding of the forces governing macromolecular interaction, a more accurate estimation of the enthalpic and entropic contributions to ligand binding, and accounting for protein dynamics should have a strong impact in computer-aided drug design.

It should, however, be noted that the ease of access of computational tools in drug design (programs, databases, web servers) has also come at a price. Too often, programs have become black-boxes, where the user has little or almost no knowledge of the underlying physical basis of the methods used, which clearly compromises the understanding and interpretation of the results thus obtained. In several published studies, the use of methods outside their range of application casts serious doubts as to whether meaningful results could indeed be obtained. It would be possible, for example, that docking hits could be experimentally validated, while the ligand–receptor interaction pattern is not accurately described: good results would be obtained, though not for the right reasons, which would clearly jeopardize lead optimization efforts.

These facts led me to consider a volume that provides a comprehensive, unified, and in-depth overview of current methodological strategies in

computer-aided drug discovery and design. Its main aims are to introduce the theoretical framework and algorithms, discuss the range of validity, strengths and limitations of each methodology, and present applications to real world problems in the drug discovery arena. Special emphasis has been given to the emerging and most pressing methodological challenges in *in silico* drug discovery and design. This approach should clearly facilitate a better interpretation of the simulation results, and should give the reader the adequate background to face the current challenges of the field.

This book is divided into three sections. Section I titled, "Theory, Methods, and Applications," presents the core methodology used in computational drug discovery, together with selected applications in Chapters 1 through 9. Chapters 1 and 2 set the tone by addressing the physical basis of ligand binding, and the force field representation of biomolecular systems. Chapter 3 discusses the concepts of chemical library representation and design, fragment libraries, drug-likeness, and filtering.

Chapters 4 through 7 cover the major *in silico* drug discovery methods: ligand-based chemical library screening, pharmacophore modeling and screening, and ligand–protein docking. Chapter 4 introduces the concepts of chemical space and molecular similarity in the context of structure–activity relationships (SARs) and hit identification, discusses the advantages and limitations of ligand-based methods, while also providing some recent examples from the literature. Chapter 5 describes the techniques of pharmacophore model generation, validation, virtual screening using various software, also including the specific requirements for those tasks. Due to its importance in characterizing ligand–protein interactions, and in structure-based virtual screening and lead optimization, two chapters are devoted to ligand–protein docking: Chapter 6 introduces the topic, focusing on binding mode prediction rather than hit discovery or ranking. Chapter 7, instead, introduces ligand docking in the context of virtual screening, describing the general workflow and the basic steps of this technique, and also reporting successful applications of docking-based virtual screening in drug discovery.

Three-dimensional protein structures have multiple uses in the computer-aided drug design scenario. Whenever experimental structures are not available, *in silico* characterized structures play a key role in drug discovery. In Chapter 8, the theoretical framework of homology or comparative modeling is presented, the individual steps of the entire process are discussed, and the use of homology models in structure-based drug discovery is reviewed, with a special focus on G protein-coupled receptors (GPCRs).

Implicit solvent methods for studying macromolecular interaction are described in Chapter 9. The theoretical foundations are presented, together with the practical aspects of their application in the context of ligand–receptor interaction, focusing on the Poisson–Boltzmann and generalized Born methods in the framework of molecular mechanics; the limitations of classical force field–based implicit solvent models are also discussed, and recent applications of quantum mechanics–based calculations in structure-based

drug design are discussed together with their advantages, progresses, and limitations. Chapter 9 lays the ground for the application of generalized Born methods for rescoring in docking-based virtual screening (Chapter 11).

Section II titled "Advanced Techniques" (Chapters 10 through 14) presents the theory, algorithms, and applications of those methods which either require a more skilled theoretical background, or their use is not as common or established compared to methods from Section I. Chapter 10 introduces the topic of druggability prediction, critically important to expand the target space beyond current limits, and from there toward the notion of the cellular "pocketome" and predictive polypharmacology.

The strengths and limitations of methods for postprocessing hits from structure-based virtual screening are presented in Chapter 11; these techniques have emerged as important computational approaches in structure-based lead optimization, since they provide for congeneric molecules superior correlations with experimental binding data than the traditional high-throughput docking scores. Free-energy calculations, presented in Chapter 12, represent a more accurate way to calculate ligand-binding free energies; this approach is not suited for evaluating binding affinities of large chemical databases of small molecules, but is rather invaluable in lead optimization scenarios, where accurate free-energy calculations are sought; this chapter provides the statistical mechanical basis of these methods, a description of available techniques, and discusses advantages and shortcomings of various approaches; future directions of free-energy calculations in the context of drug design are outlined. Chapter 13 is dedicated to molecular mechanics/coarse-grain approaches for structural prediction, a still not-too-explored avenue in drug discovery.

The general background, methodologies, and applications of small molecular fragments, instead of larger whole molecules, in virtual screening are presented in Chapter 14, highlighting that fragment screening not only improves hit rates but could offer a more balanced property profile for lead candidates developed from fragments; case studies are presented, including those targeting GPCRs.

Section III titled "Challenges," (Chapters 15 through 18) introduces the reader to the most pressing issues, where advances are sought to improve the performance and/or predictability of *in silico* methods in drug discovery and design. In Chapter 15 the role of water molecules and hydration properties in modeling ligand–protein interaction is presented, including the consideration of explicit water molecules in biomolecular interfaces, the description of methods to distinguish between bound and displaceable water molecules in the binding site of protein–ligand complexes, and applications of incorporating explicit water molecules in the context of drug design.

A major challenge, accounting for protein flexibility in structure-based drug discovery, is presented in Chapter 16, discussing the trade-off between incorporating protein degrees of freedom and computational affordability, depicting the most common approaches used by docking programs to

incorporate ligand and/or protein flexibility, exploring the use of molecular dynamics techniques to sample the conformational space of a target protein, and presenting real case examples.

The emerging challenge of targeting protein–protein interaction sites as pharmaceutical targets is thoroughly discussed in Chapter 17; the field is introduced, *in silico* tools and databases that can aid in the design of low molecular weight protein–protein interface modulators are described, key challenges are discussed, and finally, how *in silico* methods can be used and combined with experimental information to identify those modulators is illustrated.

Early-phase drug discovery has traditionally focused on optimizing drug-binding affinity, overlooking drug-binding kinetics; however, mounting evidence suggests that considering drug-binding kinetics early in the drug discovery process may increase the odds of success. Chapter 18 presents recent views on how drug-binding kinetics could impact drug discovery, introducing fast and approximate computational methods for aiding the design of drug candidates with favorable binding kinetics.

Throughout the book, particular attention has been paid to outline the theoretical basis of the described methods, thus providing the necessary background to avoid a "black-box" approach. In each self-contained chapter, the methodology is presented together with the latest developments and applications, and the challenges that lie ahead. This book constitutes both a desktop reference for academic and industrial researchers in the field, and a textbook for students in the area of molecular modeling and drug discovery.

I express my deep gratitude to all the contributors to this book, for their commitment, hard work, and outstanding chapters. I am grateful to my colleague Dr. Mario Rossi for insightful discussions. And finally, I thank Michael Slaughter from CRC Press/Taylor & Francis for his invitation to edit this book, and for his support throughout this project.

References

Jorgensen, W. L. 2009. Efficient drug lead discovery and optimization. *Acc. Chem. Res.* 42(6): 724–733.

Van Drie, J. H. 2007. Computer-aided drug design: The next 20 years. *J. Comput. Aided Mol. Des.* 21(10–11): 591–601.

Editor

Claudio N. Cavasotto, PhD, earned his MSc and PhD in physics from the University of Buenos Aires. He conducted his postdoctoral training at The Scripps Research Institute after which in 2002 he moved to MolSoft LLC, La Jolla, California, as senior research scientist, where he remained until 2007.

Dr. Cavasotto was then assistant and associate professor in the School of Biomedical Informatics at the University of Texas Health Science Center at Houston. In 2012, he moved to the Biomedicine Research Institute of Buenos Aires—Partner Institute of the Max Planck Society, where he is head of computational chemistry and drug design. His research interests are primarily biomolecular simulation, computer-aided drug discovery, and cheminformatics. His group develops and applies computational methods to study molecular interactions in biological systems, and to design molecules that modulate targets of pharmaceutical relevance.

Contributors

Ruben A. Abagyan
Skaggs School of Pharmacy and Pharmaceutical Sciences
University of California
San Diego, California

Muhammad Akram
Institute of Pharmacy/ Pharmaceutical Chemistry
and
Center for Molecular Biosciences Innsbruck (CMBI)
University of Innsbruck
Innsbruck, Austria

Rainer Bomblies
Physics Department T38
Technical University Munich
Garching, Germany

Josep M. Campanera
Departament de Fisicoquímica
and
Institut de Biomedicina (IBUB)
Universitat de Barcelona
Barcelona, Spain

Paolo Carloni
Computational Biomedicine (IAS-5/ INM-9)
Institute for Neuroscience and Medicine
and
Institute for Advanced Simulation
Forschungszentrum Jülich GmbH
and
Computational Biophysics
German Research School for Simulation Sciences GmbH
Jülich, Germany

Claudio N. Cavasotto
Biomedicine Research Institute of Buenos Aires—CONICET—Partner Institute of the Max Planck Society
Buenos Aires, Argentina

Antonella Ciancetta
Department of Pharmaceutical and Pharmacological Sciences
University of Padova
Padova, Italy

Álvaro Cortés-Cabrera
Department of Biomedical Sciences
Alcalá University
Madrid, Spain

Pietro Cozzini
Molecular Modeling Lab
Department of Food Chemistry
University of Parma
Parma, Italy

Luca Dellafiora
Molecular Modeling Lab
Department of Food Chemistry
University of Parma
Parma, Italy

Federico Gago
Department of Biomedical Sciences
Alcalá University
Madrid, Spain

Alfonso T. García-Sosa
Institute of Chemistry
University of Tartu
Tartu, Estonia

Tiziana Ginex
Molecular Modeling Lab
Department of Food Chemistry
University of Parma
Parma, Italy

Alejandro Giorgetti
Computational Biophysics
German Research School for Simulation Sciences and Computational Biomedicine
Institute for Advanced Simulation IAS-5
and
Institute of Neuroscience and Medicine INM-9
Forschungszentrum Jülich
Jülich, Germany

and

Department of Biotechnology
University of Verona
Verona, Italy

Cristiano R. W. Guimarães
Department of Research, Development and Innovation
Aché Laboratórios Farmacêuticos
São Paulo, Brazil

Teresa Kaserer
Institute of Pharmacy/ Pharmaceutical Chemistry
and
Center for Molecular Biosciences Innsbruck (CMBI)
University of Innsbruck
Innsbruck, Austria

György Miklós Keserű
Research Centre for Natural Sciences
Budapest, Hungary

Melaine A. Kuenemann
INSERM
Université Paris Diderot
Paris, France

David Lagorce
INSERM
Université Paris Diderot
Paris, France

Manuel Luitz
Physics Department T38
Technical University Munich
Garching, Germany

F. Javier Luque
Departament de Fisicoquímica
and
Institut de Biomedicina (IBUB)
Universitat de Barcelona
Barcelona, Spain

Alexander D. MacKerell, Jr.
Department of Pharmaceutical Sciences
School of Pharmacy
Computer Aided Drug Design Center
University of Maryland
Baltimore, Maryland

Gergely Makara
ComInnex Inc.
Budapest, Hungary

Maria A. Miteva
INSERM
Université Paris Diderot
Paris, France

Stefano Moro
Department of Pharmaceutical and Pharmacological Sciences
University of Padova
Padova, Italy

Antonio Morreale
REPSOL Technology Center
Madrid, Spain

Pedro A. Sánchez Murcia
Department of Biomedical Sciences
Alcalá University
Madrid, Spain

Francesco Musiani
Scuola Internazionale Superiore di Studi Avanzati (SISSA/ISAS)
Trieste, Italy

Damián Palomba
Biomedicine Research Institute of Buenos Aires—CONICET—Partner Institute of the Max Planck Society
Buenos Aires, Argentina

Daniela Schuster
Institute of Pharmacy/ Pharmaceutical Chemistry
and
Center for Molecular Biosciences Innsbruck (CMBI)
University of Innsbruck
Innsbruck, Austria

Thomas Simonson
Laboratoire de Biochimie
Department of Biology
Ecole Polytechnique
Palaiseau, France

Meagan C. Small
Department of Pharmaceutical Sciences
School of Pharmacy
Computer Aided Drug Design Center
University of Maryland
Baltimore, Maryland

Christoph A. Sotriffer
Institute of Pharmacy and Food Chemistry
University of Würzburg
Würzburg, Germany

Olivier Sperandio
INSERM
Université Paris Diderot
Paris, France

Francesca Spyrakis
Department of Life Science
University of Modena and Reggio Emilia
Modena, Italy

Bryn Taylor
Skaggs School of Pharmacy and Pharmaceutical Sciences
University of California
San Diego, California

Márton Vass
Gedeon Richter Plc.
Budapest, Hungary

Hugo O. Villar
Altoris, Inc.
San Diego, California

Bruno O. Villoutreix
INSERM
Université Paris Diderot
Paris, France

Chung F. Wong
Department of Chemistry and Biochemistry
and
Center for Nanoscience
University of Missouri-Saint Louis
St. Louis, Missouri

Martin Zacharias
Physics Department T38
Technical University Munich
Garching, Germany

William Zamora
Departament de Fisicoquímica
Institut de Biomedicina (IBUB)
Universitat de Barcelona
Barcelona, Spain

Section I

Theory, Methods, and Applications

1

The Physical Basis of Ligand Binding

Thomas Simonson

CONTENTS

1.1 Introduction

Noncovalent binding among molecules, such as enzymes/substrates, ligands/receptors, or proteins/nucleic acids, is an important element of the biochemistry and information flow in cells (Böhm and Schneider, 2003, Gohlke, 2012, Pawson, 1995). Specificity is needed to preserve the correctness of the biochemical pathways and the integrity of the information. Binding affinity and specificity are often provided by noncovalent interactions

among neighboring chemical groups, through hydrogen bonds, salt bridges, tight packing of complementary molecular surfaces, and hydrophobic forces mediated by solvent, although longer-range electrostatic interactions also play a role, particularly in the formation of encounter complexes (Fersht, 1999, Israelachvili, 1992, Jeffrey, 1997, Saenger, 1984).

In the crowded cellular environment, the number and variety of binding partners enormous, ranging from small ions to large cellular machines. With drug design as a goal, our scope is more limited, but still enormous. Even small, drug-like molecules can have complex energy surfaces, with polar, nonpolar, and polarizable groups, hard and soft degrees of freedom, multiple protonation states, possibly co-bound ions, all of which can reorganize on binding. They must recognize dynamic, fluctuating, macromolecular targets, displace water molecules, and compete with a host of other molecules. In addition, to engineer small ligands that interfere with protein/protein or protein/RNA complexes, we should understand the forces that govern such large complexes.

Only a few of these topics are covered, briefly, in this introductory chapter; various other aspects are covered in the remaining chapters. We focus on small molecule solvation and binding, mostly in the framework of the equilibrium thermodynamics of dilute solutions. In reality, the cell is crowded, stochastic, chemically open, and out of equilibrium. Nevertheless, this is the most basic and important framework with which to start an analysis of biological ligand binding, not only because it is relevant for the *in vitro* biochemistry that goes on in drug design, but also because the *in vitro* picture very often carries over in a qualitative or quantitative way to the cell.

Many of the concepts presented are general, and can be applied to any macromolecular receptor. However, RNA and DNA have some specific properties as receptors, including a high density of ionic phosphate groups, a corresponding ion cloud, their particular tertiary organization, and the high flexibility of some weakly structured RNAs. These aspects are not detailed; only for proteins do we sometimes go into specifics and detailed examples. This is partly due to space, and partly due to the prime importance of proteins as drug targets until now.

We assume a basic knowledge of molecular modeling and statistical mechanics. When we discuss molecular interactions, we treat the solutes and solvent at about the same level of theory as a molecular mechanics force field, using classical mechanics. We do not develop force field modeling, which is covered in Chapter 2, but we speak of atomic charges, point polarizabilities, van der Waals interactions, and so on. In contrast, we do not have the space or the need to discuss the shape of orbitals, spin states, tunneling, or other quantum effects.

We begin by discussing the definition of the bound state, the concept of chemical potential, and the law of mass action. Next, we discuss contributions to the binding free energy that are specifically associated with the solute

degrees of freedom: external rotations/translations and internal vibrations. We then turn to effects associated more specifically with aqueous solvent. To isolate some of the free energy contributions more clearly, we introduce a multistep binding path, where the ligand is first uncharged, then moved into the binding site, and then recharged. This allows us to separate (mostly) the discussion of electrostatic and hydrophobic effects. We also consider the displacement of water molecules from the binding pocket. Next, we discuss the separate enthalpic and entropic components of the binding free energy and their correlation or compensation. Finally, we discuss the role of conformational selection (CS) and induced fit (IF). Closely related topics that are taken up in more detail in later chapters include the many roles of solvent, receptor flexibility, and the kinetics of binding.

1.2 Defining the Bound State

To study receptor/ligand binding theoretically, one must partition the conformational space into "bound" and "unbound" states (Gallicchio and Levy, 2011, Jorgensen et al., 1988). There is no unique way to do this, but in practical situations there is often a natural choice. Thus, conformations where the ligand is within a well-defined binding pocket would be labeled "bound."* In some cases, the binding pocket will correspond to a deep energy well, so that ligand conformations near the boundary of the pocket will have high energies and low statistical weights. Thus, they will not contribute much to the thermodynamic properties, such as the binding constant, which will be robust with respect to the exact definition of the pocket (Gallicchio and Levy, 2011). In addition, when the binding of two similar ligands to a receptor is compared, there will be some cancellation of the boundary region contributions of each ligand. Even if two definitions of the binding site volume differ by a factor of two, the two definitions of the binding free energy would typically differ by kT log 2, just 0.4 kcal/mol at room temperature (where kT is the thermal energy). Such a change is not too important for a nanomolar binder at micromolar doses (a few grams in the bloodstream).

When simulations are compared to experiments, the problem is slightly different. The experiments measure a physical signal, such as heat release or optical energy absorption, and we should consider which conformations contribute to the experimental signal and use them as the basis for comparison. The most direct approach is to compute the physical signal directly from a simulation. Signals that can be directly modeled include NMR chemical shifts, pK_a shifts for protonation of a reporter group, fluorescence spectra,

* Here, a conformation is defined by the positions of all the atoms in the system, including the overall translation and rotation of the ligand relative to the receptor.

shifts in vibrational infrared bands, and so on. If the experimental signal is a local one, like a transfer of magnetization to a specific group in the binding site, then the full range of spectroscopically active conformations can be sampled in a molecular dynamics simulation, albeit with obvious limitations (imperfect force field and sampling). In general, modeling of the physical signal itself is not perfectly accurate, introducing further errors and uncertainty.

Other experimental techniques give a more global signal, like equilibrium dialysis or titration calorimetry. With these signals, minor binding sites can also contribute, including nonspecific sites located on the receptor surface. In theory, *all* the conformations may contribute, even ones where the ligand is separated from the receptor (Mihailescu and Gilson, 2004). Separating the contribution of specific and nonspecific binding modes is not straightforward. In practice, the simulation will (usually) not try to sample all possible conformations, but will focus on one or a few local regions and binding modes, and neglect the others. Furthermore, a signal like heat release usually cannot be modeled directly, and we must adopt a different route, computing the binding free energy for one or a few specific sites. This is the most common approach in free energy simulations. As mentioned, the results will often be robust with respect to the precise delimitation of the binding pocket(s); for more details, see a recent review (Gallicchio and Levy, 2011).

1.3 Chemical Potentials and Mass Action

The thermodynamic quantity that governs binding equilibria in solution is the free energy per molecule, or chemical potential (Fowler and Guggenheim, 1939, Hill, 1962, Landau and Lifschitz, 1980)

$$\mu_X = \frac{\partial G}{\partial n_X} = -kT \frac{\partial \log Q}{\partial n_X} \tag{1.1}$$

where X is a component of the solution (solvent, ligand L, receptor R, or complex RL), G is the Gibbs free energy, Q the partition function, n_X the number of molecules, and kT the thermal energy. We assume for now that X is not ionic. In what follows, we do not distinguish between Gibbs and Helmholtz free energies (NpT vs. NVT ensemble), because they differ by a negligible pV term (about 0.0005 kcal/mol for a volume change corresponding to a single water molecule at atmospheric pressure). If the solution is dilute with respect to X, μ_X has a simple, logarithmic dependence on concentration

$$\mu_X = v_X(p,T) + kT \log \frac{[X]}{[S]} = \mu_X^\circ(p,T) + kT \log \frac{[X]}{[X]^\circ} \tag{1.2}$$

where $[S]$ is the solvent concentration, p is the pressure, v_X is concentration-independent, and the superscript "o" indicates a "standard" reference state, arbitrary except that it must also be dilute (or "ideal-dilute," see below). For the binding free energy, if we choose $[R]^\circ = [L]^\circ = [RL]^\circ \stackrel{\text{def}}{=} C^\circ$ for simplicity, we have

$$\Delta G = \mu_{RL} - \mu_L - \mu_R = \Delta G^\circ + kT \log \frac{C^\circ [RL]}{[R][L]} \tag{1.3}$$

The concentration dependence arises from the loss of translational entropy upon binding.

In Equation 1.3, the free energy is defined for a peculiar equilibrium state, where the concentrations are held fixed through some kind of constraints. If the constraints are removed, the system relaxes into a more usual equilibrium. Being at a minimum, the free energy is stationary with respect to small fluctuations in the concentrations, like those produced by a single binding event. Therefore, the reaction free energy $\mu_{RL}^{\text{eq}} - \mu_L^{\text{eq}} - \mu_R^{\text{eq}} = 0$, which gives

$$\Delta G^\circ = -kT \log \frac{C^\circ [RL]_{\text{eq}}}{[R]_{\text{eq}}[L]_{\text{eq}}} \stackrel{\text{def}}{=} -kT \log C^\circ K_{\text{eq}}(p,T) \tag{1.4}$$

where K_{eq} is the equilibrium constant. Equation 1.4 is known as the law of mass action. It allows us to convert free energies into concentrations, and vice versa.

It is worth a small additional effort to see just how general are Equations 1.2 through 1.4. A very concise derivation can be found in Section 9.87 of Landau and Lifschitz (1980).* Remarkably, the only assumptions in this derivation are infinite dilution, nonionic solutes, and the validity of classical statistical mechanics. The derivation holds if the solute is not dilute but inter-solute interactions are absent, as in the usual 1 M "ideal-dilute" standard state (Ben Naim, 1973).

Biochemical applications routinely involve ionic ligands and/or receptors, so it is essential to generalize Equations 1.2 through 1.4 to this case. With an ionic solute, the derivation above* breaks down: solute/solute interactions occur at large distances, the environment of any particular solute is no longer uniform, and the free energy $\delta G(n, N)$ depends on the details of the solute positions. Thus, each anion lowers its free energy by preferentially surrounding itself by cations, and vice versa, and this alters the form of the chemical

* Suppose we add n molecules of a nonionic solute X to a large collection of N solvent molecules. At very high dilution, the solute molecules do not interact, and the free energy changes from the pure solvent value by $\delta G(n,N) = n\alpha(p,T,N) + kT \log n! \approx nkT \log((n/e)e^{\alpha/kT})$, where α is an unknown function and we use Stirling's approximation for $n!$. The $n!$ term appears because the n solute molecules are indistinguishable, which introduces a factor $1/n!$ into the partition function (Fowler and Guggenheim, 1939). δG must be a first-order homogeneous function of n and N, so that $e^{\alpha/kT}$ has the form $f(p,T)/N$. Taking $\mu_X = \partial \delta G/\partial n$ gives Equation 1.2.