
Preface

Introduction: Design and Creation of Ligand-Binding Proteins

The appropriate balance of ligand binding affinity and specificity is a fundamental feature of most if not all biological processes, including immune recognition, cellular metabolism, regulation of gene expression, and cell signaling. The ability to accurately predict and recapitulate the physical basis for ligand binding behavior is therefore a crucial part of understanding and manipulating such biological phenomena. It also represents a critical technical requirement in the reciprocal fields of drug design and protein engineering.

This book provides a collection of protocols and approaches, compiled and described by many of today's leaders in the field of protein engineering, that they apply to the problem of creating ligand-binding proteins that display desirable combinations of target affinity and specificity. The descriptions provided by each chapter's authors also provide a snapshot of their current "belief system" regarding the challenging problem of protein engineering and design, as it is applied to the creation of novel ligand binding functions.

The problem of how to effectively engineer novel binding properties onto protein scaffolds, and how to do so while exploiting the information that is provided by high-resolution protein structures, has been under investigation for almost 40 years if not longer. Such efforts date back at least to the design of small folded peptides and proteins capable of binding individual nucleosides and single-stranded DNA, followed by subsequent attempts to generate additional ligand binding functions using various protein scaffolds (*see* Refs. [1, 2] for early examples of such work). By the early 1990s, some of the first computational algorithms intended to design novel ligand binding sites into proteins of known structure had been described [3], and the field of structure-based protein engineering as it is known today was underway.

Although the field of protein engineering, including the specific problem of designing novel ligand binding capabilities onto engineered protein folds, now comprises an extensive and growing publication record, significant challenges regarding the accurate calculation or prediction of protein–ligand binding affinities (even when provided a high-resolution structure of the actual complex) still represent significant hurdles to the field's advancement. For example:

- Several recent studies have demonstrated that current methods for structure-based calculation of binding affinities display variable accuracies. At least three broad (and somewhat overlapping) classes of scoring functions for predicting binding affinities from high-resolution structures have been developed: *force-field* (formulated by calculating the individual energetic contributions of physical interactions between the protein and ligand) [4, 5], *knowledge-based* (produced by statistical mining of large databases of protein–ligand structures to deduce rules and models that govern binding affinity) [6–9], and *empirical* (in which binding energy is calculated to be a product of a collection of weighted energy terms fit to a training data set of known binding affinities, with the weighting coefficients calculated via linear regression analyses) [10–14]. Even with all these tools, the accuracy of many methods that are intended to calculate structure-based binding affinities (as well as the ability to identify and rank the most tightly bound ligands to a given protein) has been shown to often be somewhat poor

[15–17], leading to the conclusion by one group that “more precise chemical descriptions of the protein–ligand complex do not generally lead to a more accurate prediction of binding affinity” [17]. Therefore, the reliable prediction of affinity remains a significant challenge in biophysical chemistry [15].

- Even for the most thoroughly studied of ligand-binding proteins, the basis for tight, specific binding is not well understood. For example, avidin and streptavidin exhibit some of the highest known affinities to their cognate molecular ligand ($K_a \sim 10^{15} \text{ M}^{-1}$). Over 20 years of studies on these proteins have produced a wide range of hypotheses regarding their high affinities, including exceptional shape complementarity across a stabilized network of hydrophobic side chains and precisely arranged hydrogen bond partners [18], the precisely tuned dynamic behavior of the protein [19], a large free energy benefit upon ligand binding due to the strengthening of noncovalent interactions within the protein scaffold [20], or the induction of polarized moieties within the bound complex that create a cooperative effect between neighboring hydrogen bonds [21]. Not surprisingly, attempts to engineer altered binding properties onto avidin or streptavidin have yielded constructs with unexpected and unpredictable properties [22].
- Attempts to computationally engineer novel ligand-binding proteins have either been unsuccessful [23, 24] or have produced computationally designed constructs that display low affinities. Optimization of those designed proteins has then required laborious rounds of random mutagenesis and affinity maturation [25, 26].

The sources of error in calculating and modeling protein–ligand binding interactions and affinities are myriad, and their relative importance is still not entirely clear. These include: (1) Inaccuracies in the treatment of solvent and desolvation effects during binding [27–29]. (2) Limited consideration of protein dynamics [30–32]. (3) Difficulties incorporating the contribution of entropic changes into calculations of binding energies, leading to examples where modifications of ligand binding sites that lead to favorable enthalpic gains are confounded by substantial losses in entropy, with no improvement in overall binding affinity (recently reviewed extensively in Ref. [33]). Even for the most straightforward aspect of a protein–ligand interface (i.e., the observation of direct interatomic interactions and corresponding estimation of their enthalpic contributions to binding), uncertainties exist regarding interatomic distance cutoffs [17] and best strategies for estimating charge and protonation states [34].

Therefore, the creation of novel ligand-binding proteins that display tight binding affinity to their desired target and that also can discriminate between closely related targets remains an important goal, but is plagued by rather poor understanding of how to accurately calculate binding affinities or predict binding specificity, even when armed high structural information of protein–ligand complexes. As a result, the creation of highly specific ligand-binding proteins with high affinity remains extremely challenging and generally requires a substantial investment of time and effort to identify designed protein scaffolds that are actually active, and then to manually optimize their behavior. Nevertheless, studies from groups around the world have recently demonstrated that engineered proteins can, with considerable effort, be created that perform as desired, even in highly demanding *in vivo* applications. In this book, a series of 21 author groups present individual chapters that describe, in considerable detail, the types of overall thought processes and approaches, as well as very detailed computational and/or experimental protocols, that are used in their research groups as they attempt to address and resolve the difficulties associated with the design and creation of engineered ligand-binding proteins.

The reader will find a wide variety of technical issues and variables described in this volume. The first three chapters are largely concerned with a fundamental challenge that precedes actual protein engineering: identifying, characterizing, and modeling protein–ligand binding sites and predicting their corresponding modes and affinities of molecular interaction. Various strategies are shown to rely on both sequence-based and structure-based methods of analysis, and often utilize evolutionary information to determine the relative importance of positions within individual protein scaffolds that are important for form and function. With the development of controlled, blind binding site prediction challenges within the protein informatics and design community, the number of methods available to perform such analyses has exploded, as summarized in Chapter 2. Virtually all structure-based methods for binding site evaluation rely on accurate modeling of protein–ligand conformational sampling and scoring of individual docked solutions, which is further discussed in Chapters 3 and 4.

Beyond the basic ability to identify and model protein–ligand binding sites and their interactions, the field of protein engineering also now has at its disposal a number of increasingly powerful and robust computational platforms for structure-based engineering, including the widely used and rapidly evolving ROSETTA program suite as well as other programs such as POCKETOPTIMIZER and PROTEUS. Many of the fundamental features of these computational program suites, as well as individual examples of their utility and application for the design of a protein binding site for a defined small molecular ligand, are found in Chapters 5 through 7.

The output of even the most powerful structure-based computational design algorithms is usually augmented by considerable experimental time and effort, generally consisting of the preparation of combinatorial protein libraries or the systematic generation of large numbers of individual protein mutants on top of designed protein constructs, which are then subjected to selections or screens for optimal activity. While the ultimate goal of protein design is to eliminate the need for such manual intervention and effort, at this time many strategies for protein design involve combining information from computational design to the subsequent creation and screening of protein mutational libraries. Several examples of such approaches, which have resulted in particularly notable recent successes in protein engineering and the creation of designed ligand-binding proteins⁹, are outlined and described in Chapters 8–10 and can then be found at various points within the remaining chapters.

Finally, the exact technical hurdles and necessary approaches required for the creation of ligand-binding proteins obviously are dependent upon the chemical and structural nature of the ligand to be recognized and bound with high affinity and specificity. The remaining 12 chapters describe a variety of specific scenarios and methodological approaches, ranging from the design of metal-binding proteins and light-induced ligand-binding proteins, to the creation of binding proteins that also display catalytic activity, to binding of larger peptide, protein, DNA, and RNA ligands.

The continued development of approaches to design and create ligand-binding proteins, beyond enabling the creation of unique protein-based reagents and molecules for biotechnology and medicine, will continue to test and refine the ability of modern biophysical chemistry to fundamentally understand and exploit the forces and principles that drive molecular recognition. The behaviors and properties of designed ligand-binding proteins resulting from the types of methods described in this book (including the “failures”—those constructs that fail to bind their intended targets and those that bind to unintended ligands) will eventually be explained by systematically examining their structures and properties. As has been famously attributed to Richard Feynman, “That which I cannot create, I do not

understand.” The following volume provides detailed (although by no means complete and total) examples of the current approaches and methods by which the protein engineering and design community attempt to do both.

Seattle, WA, USA

Barry L. Stoddard

References

1. Gutte B, Daumigen M, Wittschieber E (1979) Design, synthesis and characterisation of a 34-residue polypeptide that interacts with nucleic acids. *Nature* 281:650–655
2. Moser R, Thomas RM, Gutte B (1983) Artificial crystalline DDT-binding polypeptide. *FEBS* 157:247–251
3. Hellenga HW, Richards FM (1991) Construction of new ligand binding sites in proteins of known structure. I. Computer-aided modeling of sites with pre-defined geometry. *J Mol Biol* 222:763–785
4. Huang N, Kalyanaraman C, Bernacki K et al. (2006) Molecular mechanics methods for predicting protein-ligand binding. *Phys Chem Chem Phys* 8: 5166–5177
5. Ewing T, Makino S, Skillman A et al. (2001) DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *J Comput Mol Des* 15:411–428
6. Gehlhaar DK, Verkhivker GM, Rejto PA et al. (1995) Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chem Biol* 2: 317–324
7. Muegge I, Martin Y (1999) A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J Med Chem* 42: 791–804
8. Mooij W, Verdonk M (2005) General and targeted statistical potentials for protein-ligand interactions. *Proteins* 61: 272–287
9. Hohlke H, Hendlich M, Klebe G (2000) Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol* 295: 337–356
10. Bohm H (1994) The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J Comput Mol Des* 8: 243–256
11. Eldridge M, Murray C, Auton T et al. (1997) Empirical scoring functions: the development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aid Mol Des* 11: 425–445
12. Friesner R, Al E (2004) Glide: a new approach for rapid, accurate docking and scoring. *J Med Chem* 47: 1739–1749
13. Krammer A, Kirchhoff P, Jiang X et al. (2005) LigScore: a novel scoring function for predicting binding affinities. *J Mol Graphics Model* 23: 395–407
14. Wang R, Lai L, Wang S (2002) Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J Comput Mol Des* 16: 11–26
15. Ross G, Morris G, Biggin P (2013) One size does not fit all: the limits of structure-based models in drug discovery. *J Chem Theory Comput* 9: 4266–4274
16. Ashtawy H, Mahapatra N (2012) A comparative assessment of ranking accuracies of conventional and machine-learning-based scoring functions for protein-ligand binding affinity prediction. *IEEE/ACM Trans Comput Biol Bioinform* 9: 1301–1312
17. Ballester P, Schreyer A, Blundell T (2014) Does a more precise chemical description of protein-ligand complexes lead to more accurate prediction of binding affinity? *J Chem Inform Model* 54: 944–955
18. Livnah O, Bayer EA, Wilchek M et al. (1993) Three-dimensional structures of avidin and the avidin-biotin complex. *Proc Natl Acad Sci U S A* 90: 5076–5080
19. Trong I, Wang Z, Hyre D et al. (2011) Streptavidin and its biotin complex at atomic resolution. *Acta Crystallogr D Biol Crystallogr* 67:813–821
20. Williams D, Stephens E, O’Brien D et al. (2004) Understanding noncovalent interactions: ligand binding energy and catalytic efficiency from ligand-induced reductions in motion within receptors and enzymes. *Angew Chem Int Ed Engl* 43:6596–6616
21. Dechancie J, Houk K (2008) The origins of femtomolar protein-ligand binding: hydrogen bond cooperativity and desolvation energetics in the biotin-(strept)avidin binding site. *JACS* 129: 5419–5429
22. Aslan FM, Yu Y, Mohr SC et al. (2005) Engineered single-chain dimeric streptavidins with an unexpected strong preference for biotin-4-fluorescein. *Proc Natl Acad Sci U S A* 102: 8507–8512
23. Schreir B, Stumpp C, Wiesner S et al. (2009) Computational design of ligand binding is not a solved problem. *Proc Natl Acad Sci U S A* 106: 18491–18496
24. Looger L, Dwyer M, Smith J et al. (2003) Computational design of receptor and sensor

- proteins with novel functions. *Nature* 423: 185–190
25. Procko E, Berguig G, Shen B et al. (2014) A computationally designed inhibitor of an Epstein-Barr viral Bcl-2 protein induces apoptosis in infected cells. *Cell* 157: 1644–1656
 26. Tinberg CE, Khare SD, Dou J et al. (2013) Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* 501: 212–216
 27. Leach A, Shoichet B, Peishoff C (2006) Prediction of protein-ligand interactions. Docking and Scoring: successes and gaps. *J Med Chem* 49: 5851–5855
 28. Schneider G (2010) Virtual screening: an endless staircase? *Nat Rev Drug Discov* 9: 273–276
 29. Huang S, Grinter S, Zou X (2010) Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. *Phys Chem Chem Phys* 12: 12899–12908
 30. Michel J, Esses J (2010) Prediction of protein-ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations. *J Comput Aid Mol Des* 24: 639–658
 31. Mobley D (2012) Let's get honest about sampling. *J Comput Aid Mol Des* 26: 93–95
 32. Guvench O, Mackerell A (2009) Computational evaluation of protein-small molecule binding. *Curr Opin Struct Biol* 19: 56–61
 33. Chodera J, Mobley D (2013) Entropy-enthalpy compensation: role and ramification in biomolecular ligand recognition and design. *Ann Rev Biophys* 42: 121–142
 34. Rocklin GJ, Boyce SE, Fischer M et al. (2013) Blind prediction of charged ligand binding affinities in a model binding site. *J Mol Biol* 425: 4569–4583



<http://www.springer.com/978-1-4939-3567-3>

Computational Design of Ligand Binding Proteins

Stoddard, B.L. (Ed.)

2016, XVI, 375 p. 94 illus., 76 illus. in color., Hardcover

ISBN: 978-1-4939-3567-3

A product of Humana Press