PHYSICS-BASED DESIGN OF PROTEIN-LIGAND BINDING

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF BIOCHEMISTRY

AND THE COMMITTEE ON GRADUATE STUDIES

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

F. Edward Boas

May 2008

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____
Pehr Harbury

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____
Dan Herschlag

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____
Tom Wandless

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____
Axel Brünger

Approved for the Stanford University Committee on Graduate Studies

_____

Abstract

Different potential energy functions have been used in protein dynamics simulations, protein design calculations, and protein structure prediction. Clearly, the same physics applies in all three cases, so the variation in potential energy functions reflects differences in how the calculations are performed. With improvements in computer power and algorithms, the same potential energy function should be applicable to all three problems.

Here we show that a standard molecular-mechanics potential energy function without any modifications can be used to engineer protein-ligand binding. A molecular-mechanics potential is used to reconstruct the coordinates of various binding sites with an average root mean square error of 0.61 Å, and to reproduce known ligand-induced side-chain conformational shifts. Within a series of 34 mutants, the calculation can always distinguish weak (K_d > 1 mM) and tight (K_d < 10 μM) binding sequences. Starting from partial coordinates of the ribose binding protein lacking the ligand and the ten primary contact residues, the molecular-mechanics potential is used to redesign a ribose binding site. Out of a search space of 2×10^{12} sequences, the calculation selects a point mutant of the native protein as the top solution (experimental K_d = 17 μM), and the native protein as the second best solution (experimental K_d = 210 nM). The quality of the predictions depends on the accuracy of the generalized Born electrostatics model, treatment of protonation equilibria, high resolution rotamer sampling, a final local energy minimization step, and explicit modeling of the bound, unbound, and unfolded states.

After this initial proof of principle experiment, we next used a standard molecular mechanics potential energy function to redesign ribose binding protein to bind a series of ligands: L-arabinose, D-xylose, indole-3-acetic acid, and estradiol. The resulting proteins have 5 – 10 mutations from the native, are stable, the predicted structures have good hydrogen bonds and shape complementarity, and they use motifs similar to natural binding proteins. All of the designed proteins bind to their target ligands with measurable but weak affinity. The affinity was improved by random mutagenesis and screening.

The application of unmodified molecular-mechanics potentials to protein design links two fields in a mutually beneficial way. Design provides a new avenue to test molecular-mechanics energy functions, and future improvements in these energy functions will presumably lead to more accurate design results.

This is the first time a single model has been used to predict structures, binding constants, and to design new small-molecule binding sites. Using a standard model should improve the generality of protein design, which could enable the creation of custom proteins for a wide variety of applications, including sensors, enzymes, and protein therapeutics.

Acknowledgements

Pehr does the type of science that makes people say “wow, I had no idea you could do that” or “why didn’t *I* think of that.” He has the remarkable ability to invent crazy new ideas, and turn them into reality. He focuses on the key roadblocks in a field, and comes up with creative new approaches to solving them. That’s why I joined Pehr’s lab, and I have learned a lot from doing science with him.

The lab is a very social place, where everyone helps each other out. Lance has been my baymate for 5 years, and we’ve gone through many life changing events together. Since I met Lance, he got married and became a father. We’ve discussed the mysteries of the universe, and all of the popular culture that I missed while growing up. Jarrett and I are classmates in the MSTP program, and we joined the lab at the same time. He’s a remarkably multi-talented person, and I’ve appreciated his good insights into both scientific and personal topics. Becky always knows the inside story, and she knows how to negotiate the best price for anything. She is a loyal friend and also has been a wonderful collaborator, generating crystals of our designed small molecule binding proteins. Rebecca has a good stories to tell, and is always eager to hear my stories as well. She is smart, curious, and quirky in the best possible way. Becky and Rebecca have both made the lab a much nicer place to work, making sure that everyone’s birthday gets celebrated, and generally keeping the lab together socially. Jim and Erica also worked on computational protein design, and they were a good source of ideas. Jim developed the multi-state design framework that the lab

uses. Erica developed a continuous mean field algorithm for improving structural sampling.

Several collaborators have been indispensable for characterizing the designed binding proteins. Rebecca Fenn has spearheaded this effort, from collecting small angle X-ray scattering (SAXS) data to generating crystals. Pavel Strop in Axel Brunger's lab solved a crystal structure of one of the unbound designed proteins. Jan Lipfert from Seb Doniach's lab collected SAXS data on the designed proteins. John King in Suzanne Pfeffer's lab helped with protein purification.

I'd also like to thank my committee — Dan Herschlag, Axel Brunger, and Tom Wandless — for keeping me on track, and getting me focused on realistic goals. Dan Herschlag was particularly instrumental in this.

I've gotten great scientific advice from a few other people. Loren Looger and I talked broadly about protein design and he suggested ways to troubleshoot the design calculations. Michael Levitt and Buzz Baldwin provided encouragement and advice when I needed it.

My Mom and Dad have been a big source of wisdom over the years. My Mom encouraged my interest in science from an early age, and I remember all the times she helped me learn new things and meet interesting people. My mom taught me the value of family, hard work, and perseverence. She has a lot of willpower, doesn't take no for an answer, and I hope I have picked up some of that. My Dad taught me how to tinker and build things, and I enjoyed the times we used to spend together building things. Finally, Katie has been my best friend, source of encouragement, and she has taught me how to relax and enjoy the ride.

Thank you to everyone who helped make this possible! I couldn't have done it without you all.

Table of contents

Abstract iv

Acknowledgements vi

Table of contents ix

List of tables xii

List of figures xiii

Chapter 1: Introduction 1

Relationship to the protein folding problem 1

Design goal 2

Innovative aspects of the research 3

Models for understanding molecular recognition 5

Finding a protein's low energy conformations 7

Probabilistic description of protein conformation 10

Evolving binding sites 11

Chapter 2: Potential energy functions for protein design 13

Summary 13

Introduction 13

Potential energy functions 16

Conclusions and future directions 28

Acknowledgements 29

Chapter 3: Physics-based design of protein-ligand Binding 30

Summary 30

Introduction 31
Results 34
Discussion 53
Materials and methods 55
Acknowledgments 58
Chapter 4: The protein design algorithm 59
Potential energy function 59
Overview 59
Water 65
Generalized-Born energy 65
Pairwise approximation of solvent accessible surface area (SASA) 69
Deprotonation energy 71
Discrete sampling 74
Protein scaffold coordinates 74
Selection of design positions 74
Rotamer library 75
Ligand poses 77
Structural optimization 79
Unfolded state 81
Sequence optimization (genetic algorithm) 83
Appendix: Integrals 86
Chapter 5: Physics-based design of new binding proteins 87
Summary 87

Introduction 88

Results 90

Discussion 101

Materials and methods 103

Chapter 6: Conclusion 105

Designing for specificity 108

Energetic vs structural predictions 110

Comparison to more established branches of engineering 110

Application: Sensors 111

Application: Custom enzymes 115

Application: Therapeutic proteins 116

References 118

List of tables

Table 1. Predicted protonation states. 43

Table 2. Predicted and calculated arabinose dissociation energy of ABP mutants. ... 47

Table 3. High resolution rotamer library, gradient-based local minimization, and an accurate solvation model are required to successfully redesign the ribose binding site in RBP. 52

Table 4. Deprotonation energies for the titratable amino acids in the 6028-member rotamer library. 74

Table 5. The highest resolution rotamer library with 6028 rotamers. 76

Table 6. Intrinsic unfolded-state chemical potentials for the amino acids in the 6028-member rotamer library..... 83

Table 7. Genetic algorithm parameters. 84

Table 8. Sequences of RBP redesigned to bind other ligands. 91

Table 9. Ligand parameters. 92

Table 10. Designed estradiol binding site in RBP is more polar when VDW stretch = 0.3 Å. 95

Table 11. Properties of designed and natural binding proteins..... 98

Table 12. RBP-IAA library..... 101

Table 13. Sensor applications. 114

List of figures

Figure 1. Examples of computational protein design. 4

Figure 2. Qualitative model of molecular recognition. 5

Figure 3. Low energy protein conformations. 9

Figure 4. Probabilistic model of protein structure. 11

Figure 5. Proteins can be modeled at different levels of detail. 15

Figure 6. Molecular mechanics potential energy function with continuum solvent... 18

Figure 7. Knowledge-based potential energy function. 19

Figure 8. Simplified schematic of the protein design algorithm. 35

Figure 9. Higher rotamer resolution improves structural predictions for the RBP binding site (PDB code: 2DRI). 38

Figure 10. Prediction of binding site coordinates. 40

Figure 11. Prediction of binding site coordinates for bevacizumab-VEGF (1BJ1), unbound VEGF (2VPF), and unbound RBP (1URP). 41

Figure 12. Prediction of side chain conformational shifts in RBP upon binding ribose, or VEGF upon binding bevacizumab. 42

Figure 13. In ABP-arabinose, 14 Glu and 89 Asp must be protonated to maintain the crystal structure coordinates under local minimization. If they are deprotonated, then the coordinates for 89 Asp shift out of position. 43

Figure 14. Predicting dissociation energies. 45

Figure 15. Redesigning the ribose binding site in RBP. 50

Figure 16. Potential energy function. 61

Figure 17. Examples of protein behaviors treated by our model. Factors typically ignored in design calculations are highlighted. 63

Figure 18. Slice through ribose binding protein, showing generalized Born radii. The radii correlate with atom burial. 67

Figure 19. Comparison of generalized Born radii for protein tyrosine phosphatase 1B calculated using an integral formula (y-axis) with radii calculated using a finite-difference approach (x-axis). 67

Figure 20. Comparison of solvent polarization energies for a set of small molecules, peptides, and proteins calculated using the generalized-Born approach (y-axis) with values calculated using a finite-difference approach (x-axis). 69

Figure 21. Thermodynamic cycle used to evaluate the deprotonation energy for aspartate (A). 73

Figure 22. Ligand sampling and filters. 77

Figure 23. Discrete then continuous optimization of protein structure. 81

Figure 24. Genetic algorithm. 85

Figure 25. In vitro evolution vs computational protein design. 90

Figure 26. Target ligands. 91

Figure 27. Experimental[119,120] and calculated β-pyranose energy – α-pyranose energy (kcal/mol). 92

Figure 28. The Lennard-Jones potential is frequently softened in design calculations to compensate for low sampling resolution. 94

Figure 29. Structures of designed and natural binding proteins. 97

Figure 30. Summary figure. 107

Figure 31. Designing for specificity in arabinose binding protein. 108

Figure 32. Structural determinants of specificity. 109

Figure 33. Energetic vs structural predictions in an inaccurate energy model. 110

Figure 34. Biosensor. Binding induces a conformational change that results in a change in fluorescence or enzymatic activity. 112

Figure 35. Binding to the transition state of a reaction catalyzes that reaction. 115