Chapter 6: Conclusion

Computer simulation tools play an important role in established branches of engineering, such as designing aircraft, bridges, buildings, and circuits. In many cases, the computer can tell us how good our design is before we even build it. Our goal is to produce tools that will move protein engineering towards the same level of sophistication. Specifically, we've developed an algorithm to engineer a new ligand binding site into a protein, or remodel an existing one to fit a different ligand. This algorithm could eventually be used to design custom sensors, enzymes, and protein therapeutics.

The design algorithm has two major components: a calculation to determine the ligand's binding affinity for a given amino acid sequence, and a genetic algorithm that "evolves" the amino acid sequence to optimize the calculated binding affinity. To calculate the energy of a given molecular conformation, we use a standard molecular mechanics potential with an accurate continuum solvent model. To model conformational changes and thermal fluctuations, we represent the protein / ligand system as a probabilistic ensemble of different backbone, side chain, and ligand conformations. To ensure stability and specificity, we compare the free energy of the bound state with several competing states, such as the unbound state or the protein bound to a related molecule.

Using this algorithm, we were able to predict binding constants, active site structures, and to design new small molecule binding proteins (Figure 30). This is the first successful redesign of an entire binding site based on an unmodified molecular-

mechanics potential energy function. It is also the first time a single model has been used to predict structures, binding constants, and to design new small-molecule binding sites.

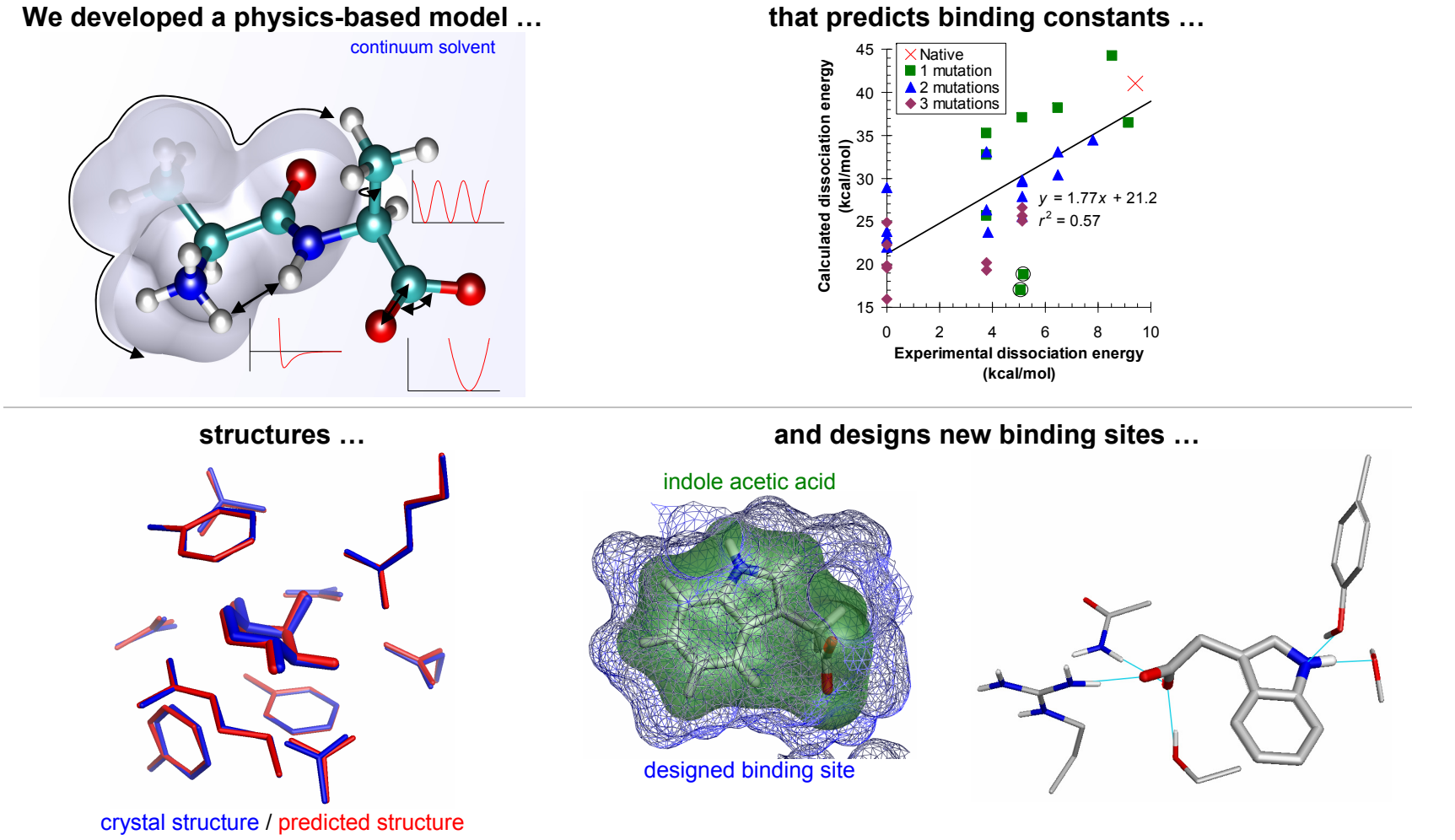

Figure 30. Summary figure.

Designing for specificity

In this thesis, the genetic algorithm was used to select for affinity and stability. Other types of selections are also possible. For example, we can select for specificity, hydrogen bonding, geometry of catalytic residues, etc. Figure 31 shows that we can select out mutants of arabinose binding protein (ABP) that have various combinations of predicted affinity and specificity for arabinose and galactose. Native ABP binds to both arabinose and galactose. Figure 32 shows that predicted specificity for arabinose is achieved by creating a steric clash with the extra CH_2OH group in galactose.

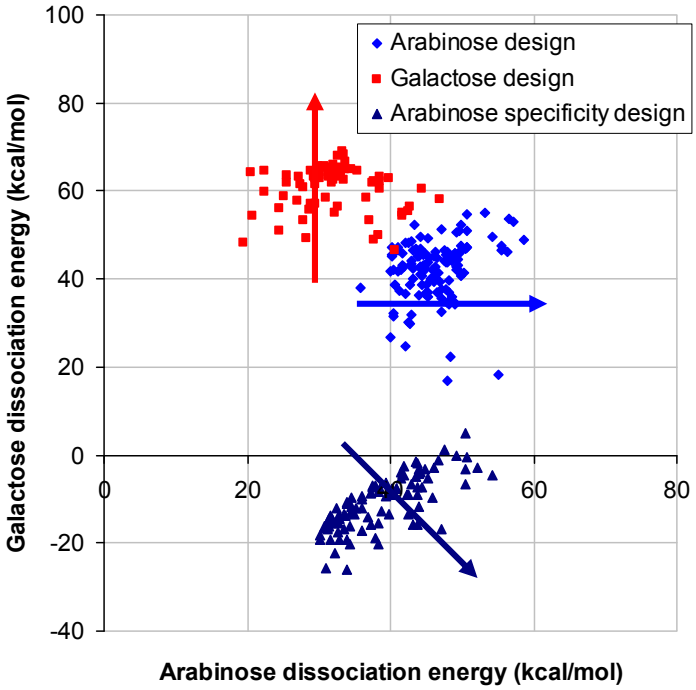

Figure 31. Designing for specificity in arabinose binding protein.
Each point represents a single sequence.

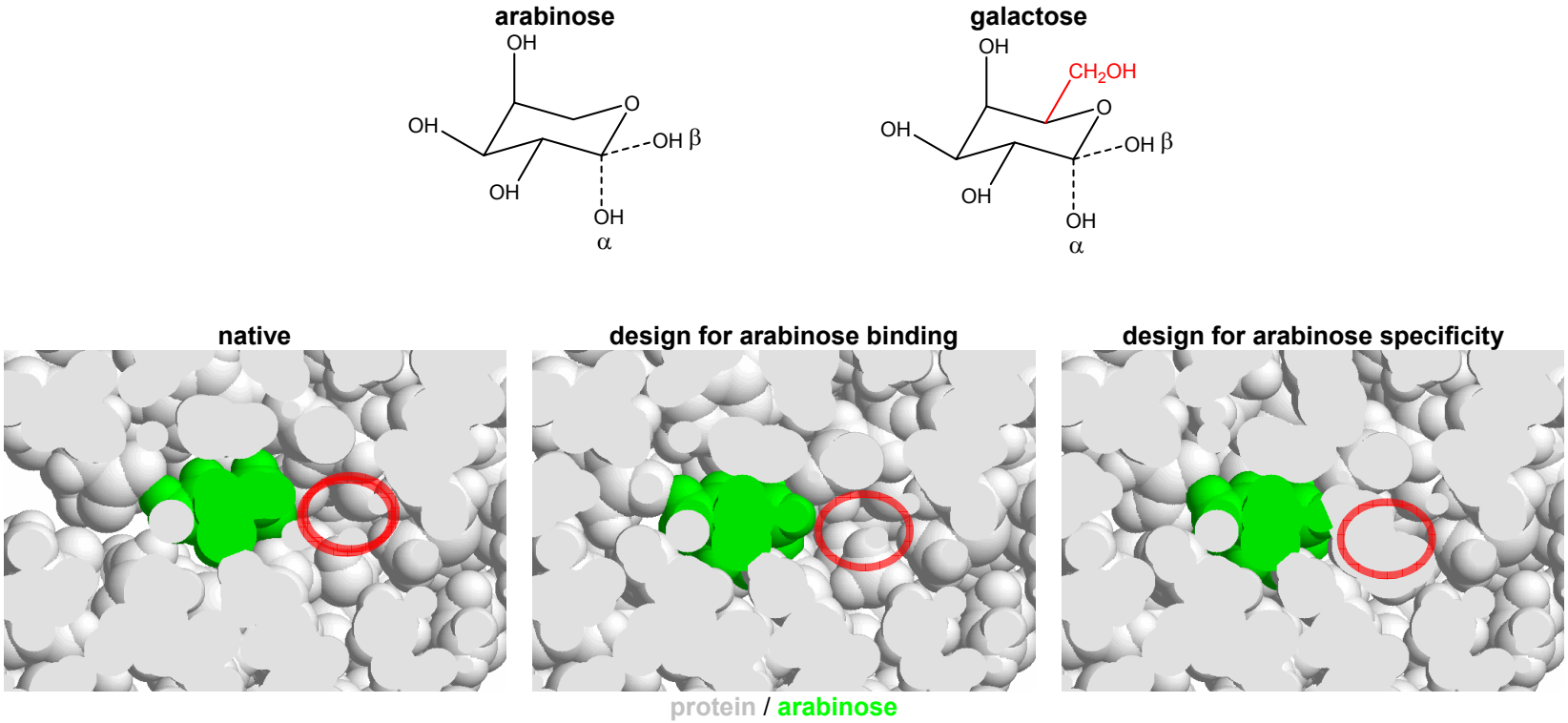

Figure 32. Structural determinants of specificity.

Space for galactose CH_2OH is seen in the native and arabinose binding design (both of which bind galactose), but not in the specificity design.

Energetic vs structural predictions

We have found that structures are easier to predict than energies. This can be understood as follows. If the bound state is a deep well in the energy landscape, then errors in the energy function will affect the well depth (dissociation energy) much more than the well position (bound structure). See Figure 33.

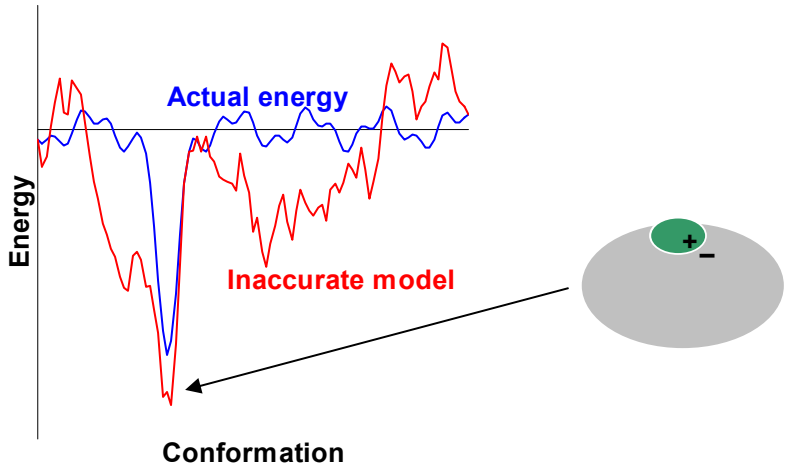

Figure 33. Energetic vs structural predictions in an inaccurate energy model.

Comparison to more established branches of engineering

Computational protein design is a young field. Can we take any clues from more established branches of engineering? For example, in electronic circuit design, if you connect a bunch of transistors in a random fashion, it will be very difficult to predict the behavior of the system without doing detailed computer simulations. Yet,

this is our approach to protein design: generate a bunch of random sequences and try to predict their behavior. Of course, circuits are not designed that way. Instead, circuits are built up from a set of modular components, with well defined rules for how these modules can be connected to each other. For example, logic inputs have to be at 0 or 5 volts, you can't switch them too fast, there's a maximum amount of current you should try to draw from certain outputs, and so on. If you follow these rules, then you can understand the behavior of the system just by thinking about it. However, if you violate the rules, then it's harder to predict what will happen, and you have to look inside each module to figure out what will happen.

Similarly, it might be possible to design a set of modular components for use in protein engineering. For example, specific protein-protein interaction motifs might be displayed side-by-side in a combinatorial fashion to create a larger set of interaction motifs.

Alternatively, there might be a set of rules for identifying amino acid sequences whose behaviour will be difficult to predict. This set of rules would be used to screen which sequences are run through a mean field calculation.

Application: Sensors

Binding is perhaps the simplest function that a protein can perform. However, with appropriate modifications to the scoring function that the genetic algorithm uses to select "good" sequences, the binding site design algorithm can be extended to engineer proteins with more sophisticated functions.

For example, if the binding event can be transduced into a detectable signal, this would produce a biosensor (Figure 34). One strategy involves adding a prosthetic fluorophore that occludes a binding site.[130] Then, when the ligand binds, the fluorophore will swing out into solvent, changing its fluorescence. A second strategy involves attaching two fluorophores or fusing two fluorescent proteins to a protein that undergoes a conformational change upon binding.[131-134] Then, ligand binding can be detected as a change in fluorescence resonance energy transfer (FRET). A third strategy involves taking an allosteric enzyme that produces a colored product, and engineering a binding site that stabilizes the enzyme's active conformation.[135,136]

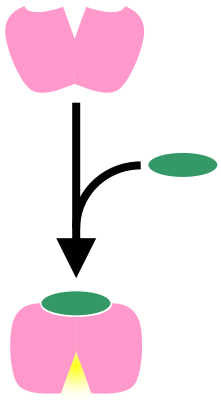

Figure 34. Biosensor. Binding induces a conformational change that results in a change in fluorescence or enzymatic activity.

There already are many systems for detecting small molecules from complex mixtures, including mass spectrometry, antibody-based assays, enzyme-coated electrodes, and arrays of materials whose electrical or physical properties change when molecules are adsorbed. Computationally engineered protein biosensors have some potential advantages over these competing technologies. Importantly, the signal readout is directly coupled to binding, does not require additional reagents or

expensive equipment, and the analyte does not need to be labeled. Furthermore, once a suitable sensor scaffold is identified, it can be engineered to detect a wide range of different molecules, enabling the creation of small-molecule microarrays that could detect a panel of biomarkers for medical diagnostic purposes (Table 13).

Application		Molecule
Subcellular fluorescence imaging of signaling molecules		IP_3, cAMP, leukotrienes
Detecting bacteria	Gram negative bacteria	lipid A (conserved portion of LPS)
	Staph. aureus	toluene
	Klebsiella pneumoniae	methyl ethyl ketone
	Pseudomonas aeruginosa	o-aminoacetophenone
	Bacillis	dipicolinic acid
	Bacterial vaginitis	putrescine
Blood / urine tests	Metabolic molecules	≈ 2500 metabolites in humans
	Hormones	
	Therapeutic drugs with a low therapeutic index	theophylline, digoxin, phenytoin, cyclosporine, methotrexate
	Illicit drugs	
	Toxins	
Cancer screening	Lung cancer	in breath: toluidine, acetophenone, benzothiazole
	Pheochromocytoma or neuroblastoma	in urine: vanillylmandelic acid
	Carcinoid tumors	in urine: 5-hydroxyindoleacetic acid
Fertility test	In axillary secretions during ovulation:	dehydroxyepiandrosterone sulfate

Table 13. Sensor applications.

Application: Custom enzymes

Modeling enzymes computationally is much more difficult than modeling non-covalent binding, because making and breaking covalent bonds needs to be treated quantum mechanically. However, simply binding the substrates in the correct orientation required for reactivity and hence stabilizing the transition state, can significantly speed up a reaction. For example, when the entire catalytic triad (Asp, His, Ser) of a serine protease is mutated, the mutant enzyme still produces 3 orders of magnitude of rate enhancement.[137] Published catalytic antibodies provide up to 8 orders of magnitude rate enhancement, although 4 orders of magnitude is more typical. Thus, even if we ignore covalent chemistry, we can still get significant catalysis with non-covalent stabilization of the transition state (Figure 35).

Furthermore, if we do know the desired geometry of the catalytic residues, we can ask the genetic algorithm to rank sequences based on both non-covalent transition state stabilization and proper predicted orientation of the catalytic residues.

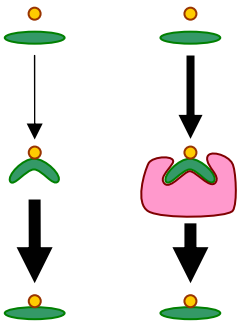

Figure 35. Binding to the transition state of a reaction catalyzes that reaction.

Custom enzymes could be used for chemical synthesis and pharmaceutical manufacturing. The extraordinary specificity of enzyme-catalyzed reactions stands in stark contrast to inorganic catalysts, which are typically very promiscuous. Thus, laboratory synthesis of complex molecules typically involves multiple steps, and multiple protecting groups that have to be added and removed at various points during the reaction. Having a customizable toolbox of enzymes that performs a desired set of reactions specifically could enable one-pot multi-step synthesis without protecting groups.

Other possible applications for custom enzymes include: custom proteases and restriction enzymes for molecular biology experiments, enzymes to degrade toxins and biofilms, and enzymes to remove antigens from transplanted cells.

Application: Therapeutic proteins

Design of better protein therapeutics will probably be the most significant commercial application of computational protein design. Therapeutic antibodies have been developed for a wide range of indications, from anti-cancer to anti-inflammatory. They currently have $5.1 billion in annual sales, and 30 antibodies are in late stage clinical trials.

Antibodies are the most widely used custom binding proteins, but they have several known limitations. Human and mouse antibodies are unable to bind to deep grooves,[128] and other targets have proven elusive as well.[138] Many antibodies are unstable or aggregation-prone.[139] Non-human antibodies are immunogenic in humans.

Therapeutic antibodies are glycosylated and thus more expensive to manufacture. Finally, the large size of whole antibodies may limit their tissue distribution.

Many of these problems can be addressed by moving to a small, stable scaffold protein that can be expressed in *E. coli*. Many such scaffolds have been proposed, including A domains, fibronectin, PDZ domains, ankyrin repeat proteins, and protein A.[140-142] Computational protein design, followed by experimental screening or selection experiments, could be used to engineer new binding sites in these scaffolds. The design framework described in this thesis can be used to select for affinity, stability, and specificity. Furthermore, computational techniques are available for predicting aggregation[143] and immunogenicity.[144,145]

References

1. Hegyi, H. & Gerstein, M. (1999). The relationship between protein structure and function: a comprehensive survey with application to the yeast genome. *J. Mol. Biol.* **288**, 147-64.
2. Looger, L. L., Dwyer, M. A., Smith, J. J. & Hellinga, H. W. (2003). Computational design of receptor and sensor proteins with novel functions. *Nature* **423**, 185-90.
3. Dwyer, M. A., Looger, L. L. & Hellinga, H. W. (2004). Computational design of a biologically active enzyme. *Science* **304**, 1967-71.
4. Dwyer, M. A., Looger, L. L. & Hellinga, H. W. (2008). Retraction. *Science* **319**, 569.
5. Dahiyat, B. I. & Mayo, S. L. (1997). De novo protein design: fully automated sequence selection. *Science* **278**, 82-7.
6. Harbury, P. B., Plecs, J. J., Tidor, B., Alber, T. & Kim, P. S. (1998). High-resolution protein design with backbone freedom. *Science* **282**, 1462-7.
7. Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L. & Baker, D. (2003). Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364-8.
8. Dahiyat, B. I. & Mayo, S. L. (1997). De novo protein design: Fully automated sequence selection. *Science* **278**, 82-87.
9. Havranek, J. J. & Harbury, P. B. (2003). Automated design of specificity in molecular recognition. *Nat. Struct. Biol.* **10**, 45-52.
10. Lazar, G. A., Dang, W., Karki, S., Vafa, O., Peng, J. S., Hyun, L., Chan, C., Chung, H. S., Eivazi, A., Yoder, S. C., Vielmetter, J., Carmichael, D. F., Hayes, R. J. & Dahiyat, B. I. (2006). Engineered antibody Fc variants with enhanced effector function. *Proc. Natl. Acad. Sci. USA* **103**, 4005-10.
11. Steed, P. M., Tansey, M. G., Zalevsky, J., Zhukovsky, E. A., Desjarlais, J. R., Szymkowski, D. E., Abbott, C., Carmichael, D., Chan, C., Cherry, L., Cheung, P., Chirino, A. J., Chung, H. H., Doberstein, S. K., Eivazi, A., Filikov, A. V., Gao, S. X., Hubert, R. S., Hwang, M., Hyun, L., Kashi, S., Kim, A., Kim, E., Kung, J., Martinez, S. P., Muchhal, U. S., Nguyen, D. H., O'Brien, C., O'Keefe, D., Singer, K., Vafa, O., Vielmetter, J., Yoder, S. C. & Dahiyat, B. I. (2003). Inactivation of TNF signaling by rationally designed dominant-negative TNF variants. *Science* **301**, 1895-8.
12. Kortemme, T., Joachimiak, L. A., Bullock, A. N., Schuler, A. D., Stoddard, B. L. & Baker, D. (2004). Computational redesign of protein-protein interaction specificity. *Nat. Struct. Mol. Biol.* **11**, 371-9.
13. Jiang, L., Althoff, E. A., Clemente, F. R., Doyle, L., Rothlisberger, D., Zanghellini, A., Gallaher, J. L., Betker, J. L., Tanaka, F., Barbas, C. F., 3rd, Hilvert, D., Houk, K. N., Stoddard, B. L. & Baker, D. (2008). De novo computational design of retro-aldol enzymes. *Science* **319**, 1387-91.
14. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D.,

Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D. & Karplus, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586-3616.

15. Bashford, D. & Case, D. A. (2000). Generalized born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* **51**, 129-152.
16. Lee, M. S., Salsbury, F. R. & Brooks, C. L. (2002). Novel generalized Born methods. *J. Chem. Phys.* **116**, 10606-10614.
17. Lovell, S. C., Word, J. M., Richardson, J. S. & Richardson, D. C. (2000). The penultimate rotamer library. *Proteins: Structure Function and Genetics* **40**, 389-408.
18. Katchalskikatzir, E., Shariv, I., Eisenstein, M., Friesem, A. A., Aflalo, C. & Vakser, I. A. (1992). Molecular-Surface Recognition - Determination of Geometric Fit between Proteins and Their Ligands by Correlation Techniques. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 2195-2199.
19. Koehl, P. & Delarue, M. (1996). Mean-field minimization methods for biological macromolecules. *Curr. Opin. Struct. Biol.* **6**, 222-6.
20. Forrest, S. (1993). Genetic algorithms: principles of natural selection applied to computation. *Science* **261**, 872-8.
21. Kortemme, T., Joachimiak, L. A., Bullock, A. N., Schuler, A. D., Stoddard, B. L. & Baker, D. (2004). Computational redesign of protein-protein interaction specificity. *Nature structural & molecular biology* **11**, 371-9.
22. Kauzmann, W. (1959). Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* **14**, 1-63.
23. Mackerell, A. D., Jr. (2004). Empirical force fields for biological macromolecules: overview and issues. *J. Comput. Chem.* **25**, 1584-604.
24. Jorgensen, W. L. & Tirado-Rives, J. (2005). Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc. Natl. Acad. Sci. USA* **102**, 6665-70.
25. Gordon, D. B., Marshall, S. A. & Mayo, S. L. (1999). Energy functions for protein design. *Curr. Opin. Struct. Biol.* **9**, 509-13.
26. Pokala, N. & Handel, T. M. (2001). Review: protein design--where we were, where we are, where we're going. *J Struct Biol* **134**, 269-81.
27. Lazaridis, T. & Karplus, M. (2000). Effective energy functions for protein structure prediction. *Curr. Opin. Struct. Biol.* **10**, 139-45.
28. Mohanty, D., Dominy, B. N., Kolinski, A., Brooks, C. L., 3rd & Skolnick, J. (1999). Correlation between knowledge-based and detailed atomic potentials: application to the unfolding of the GCN4 leucine zipper. *Proteins* **35**, 447-52.
29. Ben-Naim, A. (1997). Statistical potentials extracted from protein structures: Are these meaningful potentials? *J. Chem. Phys.* **107**, 3698-3706.
30. Simons, K. T., Kooperberg, C., Huang, E. & Baker, D. (1997). Assembly of protein tertiary structures from fragments with similar local sequences using

simulated annealing and Bayesian scoring functions. *Journal of molecular biology* **268**, 209-25.
31. Dehouck, Y., Gilis, D. & Rooman, M. (2006). A new generation of statistical potentials for proteins. *Biophys J.* **90**, 4010-7.
32. Kortemme, T. & Baker, D. (2002). A simple physical model for binding energy hot spots in protein-protein complexes. *Proc. Natl. Acad. Sci. USA* **99**, 14116-21.
33. Dahiyat, B. I., Gordon, D. B. & Mayo, S. L. (1997). Automated design of the surface positions of protein helices. *Protein Sci* **6**, 1333-7.
34. Looger, L. L. & Hellinga, H. W. (2001). Generalized dead-end elimination algorithms make large-scale protein side-chain structure prediction tractable: implications for protein design and structural genomics. *J. Mol. Biol.* **307**, 429-45.
35. Wisz, M. S. & Hellinga, H. W. (2003). An empirical model for electrostatic interactions in proteins incorporating multiple geometry-dependent dielectric constants. *Proteins* **51**, 360-77.
36. Lazaridis, T. & Karplus, M. (1999). Effective energy function for proteins in solution. *Proteins* **35**, 133-52.
37. Kortemme, T., Morozov, A. V. & Baker, D. (2003). An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *J. Mol. Biol.* **326**, 1239-59.
38. Dahiyat, B. I. & Mayo, S. L. (1997). Probing the role of packing specificity in protein design. *Proc. Natl. Acad. Sci. USA* **94**, 10172-7.
39. Chothia, C. (1974). Hydrophobic bonding and accessible surface area in proteins. *Nature* **248**, 338-9.
40. Street, A. G. & Mayo, S. L. (1998). Pairwise calculation of protein solvent-accessible surface areas. *Fold. Des.* **3**, 253-8.
41. Choudhury, N. & Pettitt, B. M. (2005). On the mechanism of hydrophobic association of nanoscopic solutes. *J. Am. Chem. Soc.* **127**, 3556-67.
42. Wagoner, J. A. & Baker, N. A. (2006). Assessing implicit models for nonpolar mean solvation forces: the importance of dispersion and volume terms. *Proc. Natl. Acad. Sci. USA* **103**, 8331-6.
43. Eisenberg, D. & McLachlan, A. D. (1986). Solvation energy in protein folding and binding. *Nature* **319**, 199-203.
44. Honig, B., Sharp, K. & Yang, A. S. (1993). Macroscopic Models of Aqueous Solutions: Biological and Chemical Applications. *J. Phys. Chem.* **97**, 1101-1109.
45. Marshall, S. A., Vizcarra, C. L. & Mayo, S. L. (2005). One- and two-body decomposable Poisson-Boltzmann methods for protein design calculations. *Protein Sci.* **14**, 1293-304.
46. Pokala, N. & Handel, T. M. (2005). Energy functions for protein design: adjustment with protein-protein complex affinities, models for the unfolded state, and negative design of solubility and specificity. *J. Mol. Biol.* **347**, 203-27.

47. Lee, M. S., Feig, M., Salsbury, F. R., Jr. & Brooks, C. L., 3rd. (2003). New analytic approximation to the standard molecular volume definition and its application to generalized Born calculations. *J. Comput. Chem.* **24**, 1348-56.
48. Yu, Z., Jacobson, M. P. & Friesner, R. A. (2006). What role do surfaces play in GB models? A new-generation of surface-generalized born model based on a novel gaussian surface for biomolecules. *J. Comput. Chem.* **27**, 72-89.
49. Feig, M., Onufriev, A., Lee, M. S., Im, W., Case, D. A. & Brooks, C. L., 3rd. (2004). Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J. Comput. Chem.* **25**, 265-84.
50. Schymkowitz, J. W., Rousseau, F., Martins, I. C., Ferkinghoff-Borg, J., Stricher, F. & Serrano, L. (2005). Prediction of water and metal binding sites and their affinities by using the Fold-X force field. *Proc. Natl. Acad. Sci. USA* **102**, 10147-52.
51. Jiang, L., Kuhlman, B., Kortemme, T. & Baker, D. (2005). A "solvated rotamer" approach to modeling water-mediated hydrogen bonds at protein-protein interfaces. *Proteins* **58**, 893-904.
52. Morozov, A. V., Kortemme, T., Tsemekhman, K. & Baker, D. (2004). Close agreement between the orientation dependence of hydrogen bonds observed in protein structures and quantum mechanical calculations. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 6946-51.
53. Morozov, A. V. & Kortemme, T. (2005). Potential functions for hydrogen bonds in protein structure prediction and design. *Adv. Protein. Chem.* **72**, 1-38.
54. Friesner, R. A. (2006). Modeling polarization in proteins and protein-ligand complexes: Methods and preliminary results. *Adv. Protein Chem.* **72**, 79-104.
55. Maple, J. R., Cao, Y. X., Damm, W. G., Halgren, T. A., Kaminski, G. A., Zhang, L. Y. & Friesner, R. A. (2005). A polarizable force field and continuum solvation methodology for modeling of protein-ligand interactions. *Journal of Chemical Theory and Computation* **1**, 694-715.
56. Friesner, R. A. (2005). Ab initio quantum chemistry: methodology and applications. *Proc. Natl. Acad. Sci. USA* **102**, 6648-53.
57. Cho, A. E., Guallar, V., Berne, B. J. & Friesner, R. (2005). Importance of accurate charges in molecular docking: quantum mechanical/molecular mechanical (QM/MM) approach. *J. Comput. Chem.* **26**, 915-31.
58. Zhou, H. X. (2003). Direct test of the Gaussian-chain model for treating residual charge-charge interactions in the unfolded state of proteins. *J. Am. Chem. Soc.* **125**, 2060-2061.
59. Waldburger, C. D., Schildbach, J. F. & Sauer, R. T. (1995). Are buried salt bridges important for protein stability and conformational specificity? *Nat. Struct. Biol.* **2**, 122-8.
60. Clark, L. A., Boriack-Sjodin, P. A., Eldredge, J., Fitch, C., Friedman, B., Hanf, K. J., Jarpe, M., Liparoto, S. F., Li, Y., Lugovskoy, A., Miller, S., Rushe, M., Sherman, W., Simon, K. & Van Vlijmen, H. (2006). Affinity enhancement of an in vivo matured therapeutic antibody using structure-based computational design. *Protein Sci.* **15**, 949-60.

61. Shifman, J. M. & Mayo, S. L. (2003). Exploring the origins of binding specificity through the computational redesign of calmodulin. *Proc. Natl. Acad. Sci. USA* **100**, 13274-9.
62. Boas, F. E. & Harbury, P. B. (2008). Design of protein-ligand binding based on the molecular-mechanics energy model. *J. Mol. Biol.* **In press.**
63. Kuhlman, B. & Baker, D. (2000). Native protein sequences are close to optimal for their structures. *Proc. Natl. Acad. Sci. USA* **97**, 10383-8.
64. Ashworth, J., Havranek, J. J., Duarte, C. M., Sussman, D., Monnat, R. J., Jr., Stoddard, B. L. & Baker, D. (2006). Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature* **441**, 656-9.
65. Ambroggio, X. I. & Kuhlman, B. (2006). Computational design of a single amino acid sequence that can switch between two distinct protein folds. *J. Am. Chem. Soc.* **128**, 1154-61.
66. Saunders, C. T. & Baker, D. (2005). Recapitulation of protein family divergence using flexible backbone protein design. *J. Mol. Biol.* **346**, 631-44.
67. Snow, C. D., Sorin, E. J., Rhee, Y. M. & Pande, V. S. (2005). How well can simulation predict protein folding kinetics and thermodynamics? *Annu. Rev. Biophys. Biomol. Struct.* **34**, 43-69.
68. Ren, P. Y. & Ponder, J. W. (2002). Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. *J. Comput. Chem.* **23**, 1497-1506.
69. Kuhn, B. & Kollman, P. A. (2000). Binding of a diverse set of ligands to avidin and streptavidin: an accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *J Med Chem* **43**, 3786-91.
70. Huo, S., Massova, I. & Kollman, P. A. (2002). Computational alanine scanning of the 1:1 human growth hormone-receptor complex. *J Comput Chem* **23**, 15-27.
71. Mobley, D. L., Graves, A. P., Chodera, J. D., McReynolds, A. C., Shoichet, B. K. & Dill, K. A. (2007). Predicting absolute ligand binding free energies to a simple model site. *J Mol Biol* **371**, 1118-34.
72. Wang, J., Kang, X., Kuntz, I. D. & Kollman, P. A. (2005). Hierarchical database screenings for HIV-1 reverse transcriptase using a pharmacophore model, rigid docking, solvation docking, and MM-PB/SA. *J Med Chem* **48**, 2432-44.
73. Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A. & Cheatham, T. E., 3rd. (2000). Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* **33**, 889-97.
74. Snow, C. D., Nguyen, H., Pande, V. S. & Gruebele, M. (2002). Absolute comparison of simulated and experimental protein-folding dynamics. *Nature* **420**, 102-6.

75. Barth, P., Alber, T. & Harbury, P. B. (2007). Accurate, conformation-dependent predictions of solvent effects on protein ionization constants. *Proc Natl Acad Sci U S A* **104**, 4898-903.
76. Chakrabarti, R., Klibanov, A. M. & Friesner, R. A. (2005). Computational prediction of native protein ligand-binding and enzyme active site sequences. *Proc Natl Acad Sci U S A* **102**, 10153-8.
77. Chakrabarti, R., Klibanov, A. M. & Friesner, R. A. (2005). Sequence optimization and designability of enzyme active sites. *Proc Natl Acad Sci U S A* **102**, 12035-40.
78. Boas, F. E. & Harbury, P. B. (2007). Potential energy functions for protein design. *Curr. Opin. Struct. Biol.* **17**, 199-204.
79. Bolon, D. N. & Mayo, S. L. (2001). Enzyme-like proteins by computational design. *Proc. Natl. Acad. Sci. USA* **98**, 14274-9.
80. Zanghellini, A., Jiang, L., Wollacott, A. M., Cheng, G., Meiler, J., Althoff, E. A., Rothlisberger, D. & Baker, D. (2006). New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci* **15**, 2785-94.
81. Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C. & Mainz, D. T. (2006). Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J Med Chem* **49**, 6177-96.
82. Jain, A. N. (2006). Scoring functions for protein-ligand docking. *Curr Protein Pept Sci* **7**, 407-20.
83. Lassila, J. K., Privett, H. K., Allen, B. D. & Mayo, S. L. (2006). Combinatorial methods for small-molecule placement in computational enzyme design. *Proc Natl Acad Sci U S A* **103**, 16710-5.
84. Quiocho, F. A. & Ledvina, P. S. (1996). Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes. *Mol. Microbiol.* **20**, 17-25.
85. Gilson, M. K. & Zhou, H. X. (2007). Calculation of protein-ligand binding affinities. *Annu Rev Biophys Biomol Struct* **36**, 21-42.
86. Declerck, N. & Abelson, J. (1994). Novel substrate specificity engineered in the arabinose binding protein. *Protein Eng.* **7**, 997-1004.
87. Vermersch, P. S., Lemon, D. D., Tesmer, J. J. & Quiocho, F. A. (1991). Sugar-binding and crystallographic studies of an arabinose-binding protein mutant (Met108Leu) that exhibits enhanced affinity and altered specificity. *Biochemistry* **30**, 6861-6.
88. Li, Y., Li, H., Yang, F., Smith-Gill, S. J. & Mariuzza, R. A. (2003). X-ray snapshots of the maturation of an antibody response to a protein antigen. *Nat. Struct. Biol.* **10**, 482-8.
89. Fersht, A. (1999). *Structure and mechanism in protein science: A guide to enzyme catalysis and protein folding*, W.H. Freeman and Company, New York.
90. Word, J. M., Richardson, J. S. & Richardson, D. C. (2000). *bndlst version 1.6 (http://kinemage.biochem.duke.edu/)*.

91. Lawrence, M. C. & Colman, P. M. (1993). Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946-50.
92. Qiu, D., Shenkin, P. S., Hollinger, F. P. & Still, W. C. (1997). The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *J. Phys. Chem. A* **101**, 3005-3014.
93. Monard, G. & Merz, K. M. (1999). Combined quantum mechanical/molecular mechanical methodologies applied to biomolecular systems. *Accounts Chem. Res.* **32**, 904-911.
94. Liu, H., Elstner, M., Kaxiras, E., Frauenheim, T., Hermans, J. & Yang, W. (2001). Quantum mechanics simulation of protein dynamics on long timescale. *Proteins* **44**, 484-9.
95. Still, W. C., Tempczyk, A., Hawley, R. C. & Hendrickson, T. (1990). Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **112**, 6127-6129.
96. Lim, C., Bashford, D. & Karplus, M. (1991). Absolute pKa Calculations with Continuum Dielectric Methods. *J. Phys. Chem.* **95**, 5610-5620.
97. Nocedal, J. & Wright, S. J. (1999). *Numerical Optimization*. Springer series in operations research, Springer, New York.
98. Ponder, J. W. (2004). *TINKER version 4.2 (http://dasher.wustl.edu/tinker/)*.
99. Slovic, A. M., Kono, H., Lear, J. D., Saven, J. G. & DeGrado, W. F. (2004). Computational design of water-soluble analogues of the potassium channel KcsA. *Proc. Natl. Acad. Sci. USA* **101**, 1828-33.
100. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallographica Section D-Biological Crystallography* **54**, 905-921.
101. Kunkel, T. A., Bebenek, K. & McClary, J. (1991). Efficient site-directed mutagenesis using uracil-containing DNA. *Methods Enzymol.* **204**, 125-39.
102. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. (1995). How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **4**, 2411-23.
103. Leach, A. R. (2001). *Molecular Modelling : Pinciples and Applications*. 2nd edit, Prentice Hall, Harlow, England ; New York.
104. Srinivasan, J., Trevathan, M. W., Beroza, P. & Case, D. A. (1999). Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. *Theoretical Chemistry Accounts* **101**, 426-434.
105. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1996). *Numerical Recipes in C: The Art of Scientific Computing*. 2nd edit, Cambridge University Press, Cambridge; New York.
106. Lide, D. R. (2000). *CRC Handbook of Chemistry and Physics*. 81st edit, CRC Press, Boca Raton; New York; Washington D.C.
107. Creighton, T. E. (1993). *Proteins: Structures and Molecular Properties*. 2nd edit, W.H. Freeman and Company, New York.

108. Word, J. M., Lovell, S. C., Richardson, J. S. & Richardson, D. C. (1999). Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* **285**, 1735-47.
109. McDonald, I. K. & Thornton, J. M. (1994). Satisfying hydrogen bonding potential in proteins. *J Mol Biol* **238**, 777-93.
110. Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A. A., Aflalo, C. & Vakser, I. A. (1992). Molecular-Surface Recognition: Determination of Geometric Fit between Proteins and Their Ligands by Correlation Techniques. *Proc. Natl. Acad. Sci. USA* **89**, 2195-2199.
111. Desmet, J., Demaeyer, M., Hazes, B. & Lasters, I. (1992). The Dead-End Elimination Theorem and Its Use in Protein Side-Chain Positioning. *Nature* **356**, 539-542.
112. Desmet, J., Spriet, J. & Lasters, I. (2002). Fast and accurate side-chain topology and energy refinement (FASTER) as a new method for protein structure optimization. *Proteins* **48**, 31-43.
113. Hu, X. & Kuhlman, B. (2006). Protein design simulations suggest that side-chain conformational entropy is not a strong determinant of amino acid environmental preferences. *Proteins* **62**, 739-48.
114. Lin, H. & Cornish, V. W. (2002). Screening and selection methods for large-scale analysis of protein function. *Angew. Chem. Int. Ed. Engl.* **41**, 4402-25.
115. Mowbray, S. L. & Cole, L. B. (1992). 1.7 A X-ray structure of the periplasmic ribose receptor from Escherichia coli. *J. Mol. Biol.* **225**, 155-75.
116. Bjorkman, A. J. & Mowbray, S. L. (1998). Multiple open forms of ribose-binding protein trace the path of its conformational change. *J. Mol. Biol.* **279**, 651-64.
117. Berthod, H., Giessner*, C. & Pullman, A. (1967). Sur les roles respectifs des electrons sigma et pi dans les proprietes des derives halogenes des molecules conjuguees: Application a letude de luracile et du fluorouracile. *Theoretica Chimica Acta* **8**, 212-&.
118. Matthew Clark, R. D. C. I. I. I. N. V. O. (1989). Validation of the general purpose tripos 5.2 force field. *Journal of Computational Chemistry* **10**, 982-1012.
119. Loudon, G. M. (1995). *Organic Chemistry*. 3rd edit.
120. Matsuo, K. & Gekko, K. (2004). Vacuum-ultraviolet circular dichroism study of saccharides by synchrotron radiation spectrophotometry. *Carbohydr Res* **339**, 591-7.
121. Kyte, J. & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105-32.
122. Hunter, C. A., Singh, J. & Thornton, J. M. (1991). Pi-pi interactions: the geometry and energetics of phenylalanine-phenylalanine interactions in proteins. *J. Mol. Biol.* **218**, 837-46.
123. Cadwell, R. C. & Joyce, G. F. (1992). Randomization of genes by PCR mutagenesis. *PCR methods and applications* **2**, 28-33.

124. Zhao, H., Giver, L., Shao, Z., Affholter, J. A. & Arnold, F. H. (1998). Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nature biotechnology* **16**, 258-61.
125. Cadwell, R. C. & Joyce, G. F. (1992). Randomization of genes by PCR mutagenesis. *PCR Methods Appl.* **2**, 28-33.
126. Morozov, A. V., Kortemme, T., Tsemekhman, K. & Baker, D. (2004). Close agreement between the orientation dependence of hydrogen bonds observed in protein structures and quantum mechanical calculations. *Proc. Natl. Acad. Sci. USA* **101**, 6946-51.
127. Lockless, S. W. & Ranganathan, R. (1999). Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* **286**, 295-9.
128. Desmyter, A., Transue, T. R., Ghahroudi, M. A., Thi, M. H., Poortmans, F., Hamers, R., Muyldermans, S. & Wyns, L. (1996). Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme. *Nat. Struct. Biol.* **3**, 803-11.
129. Fersht, A. (1999). *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, W.H. Freeman, New York.
130. Morii, T., Sugimoto, K., Makino, K., Otsuka, M., Imoto, K. & Mori, Y. (2002). A new fluorescent biosensor for inositol trisphosphate. *J. Am. Chem. Soc.* **124**, 1138-9.
131. Fehr, M., Frommer, W. B. & Lalonde, S. (2002). Visualization of maltose uptake in living yeast cells by fluorescent nanosensors. *Proc. Natl. Acad. Sci. USA* **99**, 9846-51.
132. Miyawaki, A. & Tsien, R. Y. (2000). Monitoring protein conformations and interactions by fluorescence resonance energy transfer between mutants of green fluorescent protein. *Methods Enzymol.* **327**, 472-500.
133. Muddana, S. S. & Peterson, B. R. (2003). Fluorescent cellular sensors of steroid receptor ligands. *Chembiochem* **4**, 848-55.
134. Bacskai, B. J., Hochner, B., Mahaut-Smith, M., Adams, S. R., Kaang, B. K., Kandel, E. R. & Tsien, R. Y. (1993). Spatially resolved dynamics of cAMP and protein kinase A subunits in Aplysia sensory neurons. *Science* **260**, 222-6.
135. Villaverde, A. (2003). Allosteric enzymes as biosensors for molecular diagnosis. *FEBS Letters* **554**, 169-172.
136. Ha, J. H., Butler, J. S., Mitrea, D. M. & Loh, S. N. (2006). Modular enzyme design: Regulation by mutually exclusive protein folding. *Journal of Molecular Biology* **357**, 1058-1062.
137. Carter, P. & Wells, J. A. (1988). Dissecting the catalytic triad of a serine protease. *Nature* **332**, 564-8.
138. Kwong, P. D., Doyle, M. L., Casper, D. J., Cicala, C., Leavitt, S. A., Majeed, S., Steenbeke, T. D., Venturi, M., Chaiken, I., Fung, M., Katinger, H., Parren, P. W., Robinson, J., Van Ryk, D., Wang, L., Burton, D. R., Freire, E., Wyatt, R., Sodroski, J., Hendrickson, W. A. & Arthos, J. (2002). HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature* **420**, 678-82.

139. Jespers, L., Schon, O., Famm, K. & Winter, G. (2004). Aggregation-resistant domain antibodies selected on phage by heat denaturation. *Nat. Biotechnol.* **22**, 1161-5.
140. Silverman, J., Liu, Q., Bakker, A., To, W., Duguay, A., Alba, B. M., Smith, R., Rivas, A., Li, P., Le, H., Whitehorn, E., Moore, K. W., Swimmer, C., Perlroth, V., Vogt, M., Kolkman, J. & Stemmer, W. P. (2005). Multivalent avimer proteins evolved by exon shuffling of a family of human receptor domains. *Nat. Biotechnol.* **23**, 1556-61.
141. Binz, H. K., Amstutz, P. & Pluckthun, A. (2005). Engineering novel binding proteins from nonimmunoglobulin domains. *Nat. Biotechnol.* **23**, 1257-68.
142. Binz, H. K. & Pluckthun, A. (2005). Engineered proteins as specific binding reagents. *Curr. Opin. Biotechnol.* **16**, 459-69.
143. Fernandez-Escamilla, A. M., Rousseau, F., Schymkowitz, J. & Serrano, L. (2004). Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.* **22**, 1302-6.
144. Brusic, V., Rudy, G. & Harrison, L. C. (1998). MHCPEP, a database of MHC-binding peptides: update 1997. *Nucleic Acids Res.* **26**, 368-71.
145. Kolaskar, A. S. & Tongaonkar, P. C. (1990). A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett.* **276**, 172-4.