Chapter 2

Protein Design

In this chapter, I describe common approaches to protein design, and significant accomplishments in the field. Even so, review articles provide the best introduction. William DeGrado and colleagues wrote one of the most thorough reviews in 1999, although other reviews are more current. The major shortcoming of DeGrado’s review is the omission of recent advances made primarily by David Baker’s group using Rosetta, currently the most well-known and widely-used program for protein structure prediction and design. Rhiju Das and David Baker wrote a wonderful overview of Rosetta and its applications, which complements DeGrado’s review quite well. In regard to niche areas of protein design, several superb reviews have been published recently on the design of metalloproteins, biosensors, and enzymes.

2.1 Approaches to design

Before outlining the major accomplishments in protein design, a few definitions are useful.

The problem of protein design is to choose a sequence of amino acids which will fold into a given structure; the designed protein may also need to possess additional properties. The solution to a design problem is a design or amino acid sequence predicted to satisfy that problem. Additionally, the following classifications are often used to subdivide kinds of design problems: (a) Rational design is the selection of an amino acid sequence by human intuition and reasoning, while computational design uses algorithms to create that sequence. (b) In de novo design, one begins with desired 3D coordinates of backbone atoms and must select a sequence which folds into the given structure. In redesign, one begins with both the sequence and structure; the goal of redesign is instead to modify the starting sequence to achieve some end, such as adding a metal binding site or catalytic center. (c) In fixed backbone design, the coordinates of backbone atoms are static. The design problem is to choose a low-energy structure from all possible superpositions of amino acid rotamers on the immobilised backbone. In flexible backbone design, the backbone atoms are allowed to move slightly, and the goal is to create a sequence predicted to assume a structure nearly identical to the initial conformation.
2.1.1 This project

This proposal outlines a combined computational and rational approach to the design of minimal-disulfide knottins. I will manually select the general strategies, such as which disulfides to eliminate, what binding-site residues to preserve, and possibly force additional properties such as a well-packed aromatic cluster. After designs have been generated I will select the designs most likely to be successful.

While this proposal technically describes a redesign problem, the removal of multiple disulfides is a such profound structural change that the task is more akin to \textit{de novo} design. Native knottin disulfides may also lock backbone atoms into conformations impossible to replicate precisely in minimal disulfide redesigns. Nonetheless, a designed conformation nearly identical to that of a natural knottin, which also preserves the functional residues, seems likely to possess the same function. Furthermore, the typical NMR-determined knottin structures are not precise enough to justify a fixed backbone design. Hence, this project applies flexible backbone design.

2.1.2 Rational design

2.1.3 Computational design

Computational approaches are frequently applied in protein design and are well suited for creating good hydrophobic cores. Stable cores need two properties acceptable van der Waals interactions and absence of voids. The main component of van der Waals interactions is the repulsive force. The attractive force is quite minor, and consequently, unfavorable hydrocarbon-water interactions drive core formation. A well-packed core has no room for solvent molecules, and all water is expelled from the core during folding. But if a fully-formed core contains voids, water can remain in the core and destabilize the folded state. Computationally, estimating van der Waals forces is trivial while estimating the energetic cost of voids is expensive. Early research relied on proxies for voids such as the attractive portion of van der Waals interactions or volume of buried side chains. Far better methods now exist, but the development of inexpensive techniques to quantify the stability cost for voids remains a current area of research. So given a protein structure, the quality of packing is quantifiable, and design algorithms can produce beautifully-packed cores by generating an amino acid sequence that optimizes packing.

In contrast to the easily defined constraints of core design, surface design requires quantification of more than van der Waals interactions; hydrogen bonding, electrostatics, and solvation are especially important. These forces are often relatively weak and difficult to quantify, and computational methods are not nearly as well suited for quantifying these effects. Consequently, algorithms for the design of surfaces and, more importantly, the design of entire proteins must optimize a complex combination of measures. Although modern algorithms have successfully designed entire proteins, surface design remains a major shortcoming of computational design.

In fixed-backbone design, optimizing an amino acid sequence is a combinatorial optimization problem: given backbone coordinates and rotamers for all amino acids, select a sequence of rotamers which, when superimposed on the backbone, produce the...
best structure as determined by a quantitative measure of quality. When framed as an optimization problem, a scientist may apply standard optimization algorithms to design proteins.

With fixed-backbone design, computing a score for all possible sequences of roamers is impossible for any protein of reasonable length—including mini-AgRP. In general, design algorithms implement stochastic methods that instead seek a solution near the global optimum solution. In special cases, however, clever approaches can find the global optimum. In protein design, the most common approaches are genetic algorithms, simulated annealing, and dead-end elimination as discussed below.

Consider the following naïve deterministic optimization algorithm. Generate an initial structure by superimposing a random sequence of rotamers onto a fixed backbone. For the first residue, attempt replacement of the current rotamer with all possible rotamers. Keep the replacement that maximizes the objective function. Repeat for all remaining residues in the protein, and continue optimizing the entire sequence until no rotamer substitution can improve the score. This procedure will find a locally optimal solution, but because the objective function for protein design partially mimics the natural energy landscape, there are many locally optimal solutions—many of which score quite poorly relative to the global optimum.

Simulated annealing is a popular stochastic approach and is analogous to a natural process—protein folding. When an unfolded protein is too cold—near absolute zero—insufficient thermal energy exists to bump the structure out of local energy minima, and the protein will never fold. When a protein is too hot, it samples many conformations, some of which are likely near the folded state. But thermal motion causes too many random changes in structure and the protein never folds. If the correct folding temperature of a protein is unknown, one way to fold the protein is to begin at a temperature that is too hot and gradually cool the solution. Because the correct temperature is somewhere between too hot and too cold, the protein will fold at some point, and moving to colder temperatures will simply lock the structure in the folded state.

The objective function in protein design is analogous to the energy landscape and random rotamer changes are analogous to temperature. The naïve algorithm above always changes to better scoring sequences and is analogous to protein folding near absolute zero. But if rotamers are also occasionally changed at random—the equivalent of thermal motion—the sequence can migrate away from the local optimum and sample a more diverse set of rotamer sequences. Given sufficient sampling of sequences, some will be near the global optimum or, at least, a local optimum far superior to what would be found by the naïve approach. But if too many rotamer changes are random, the temperature is too great and no good solution will ever be found. The simulated annealing approach is therefore to use the naïve approach, but with random rotamer changes. Throughout the optimization process, the frequency of random rotamer changes—the temperature—decreases from frequent to never. The effective part of a search with simulated annealing happens with particular degree of noise, analogous to best folding temperature for a protein. Consequently, simulated annealing algorithms generally attempt to spend most of their time in near this optimal probability.

Genetic algorithms are another popular stochastic approach also analogous to a natural process. As one example, a design method might begin by generating 100 structures
through superposition of randomly selected rotamers onto a fixed backbone, creating an initial population. After removing the worst 80, the method produces 80 new conformations from the fittest 20. Some of the new conformations are constructed by rotamer replacement on an existing member. Others are created from pairs of existing designs, with some rotamer superpositions from the first and the remainder from the second. The process of selection, mutation, and hybridization continues until improvement from generation to generation becomes marginal. Members of the final pool will be near an optimum, and if the pool contained sufficient diversity at some point during the simulated evolution, members of the final pool will have scores close to the global optimum.

As previously mentioned, one can occasionally find the correct solution (global optimum) without scoring all possible rotamer sequences. The most common strategy is dead-end elimination. For example, an algorithm redesigning a protein core might start from a backbone structure and rotamers for all surface residues. The problem is then to select the best sequence of core rotamers to pack the core. All rotamer possibilities are considered for the first core residue. When a rotamer overlaps with the backbone or another rotamer, the van der Waals interaction is exceedingly unfavorable, and all structures with that rotamer for the first core position are clearly suboptimal. Given each possibly for the first core position, the algorithm then attempts all rotamers at the second core position—most of which can again be eliminated as dead ends. The process continues for the third core position, given all possibilities for the first two; the fourth, given the first three; and so on. After all potentially correct combinations are scored, the best-scoring is also the global optimum. Dead-end elimination is only useful in certain situations; unless the design problem is carefully constructed, the method will fail to eliminate a sufficiently large number of suboptimal solutions. In practice, a nearly optimal design is just as good as the global optimum, and stochastic methods such as simulated annealing and genetic algorithms are now favored due to their generalizability and practicality.

The largest problem with fixed-backbone design is that a compatible sequence of amino acids may not exist for a given backbone conformation, but a compatible sequence of amino acids may exist for backbone coordinates nearly identical to those given. For that reason, flexible backbone design is generally desirable, but is also a far more difficult problem. An algorithm must address the combinatorial complexity of fixed-backbone design for the infinite number of backbone conformations near the starting point. Although other strategies have been attempted, the most successful approach is the application of a normal fixed-backbone design strategy while periodically optimizing the backbone conformation, allowing for only tiny changes. The remarkable success of this approach is described in Sect 2.2.2

2.1.4 Molecular evolution

Molecular evolution refers to a class of techniques that start from a specific biomolecule, create a diverse library of mutants, and identify the most desirable members in the library. Once the most desirable mutants are identified, they may serve as a template for another round of diversification and selection. Molecular evolution is not a part of this proposal, but has been used extensively to solve the same types of problems
addressed by rational and computation redesign. These techniques are therefore relevant as alternative and occasionally complementary methods in protein design.

Eliciting antibodies is a common technique for creating proteins to bind to a particular chemical or biomolecule. Antibody production is essentially a method of molecular evolution, although not usually considered such. The immune system constantly produces a diverse collection of antibodies. To identify those that binds a substance of interest (antigen), the experimenter introduces the antigen into an animal and allows the immune system to identify the recognizing antibodies. After identification, the immune system produces large quantities of those antibodies, so the antigenic substance can be recognized and eliminated. The elicited antibodies are then isolated and identified by the experimenter, who now has proteins that bind the substance of interest. Antibodies are a common experimental tool extensively used in research, but even after identification of recognizing antibodies, production is fairly expensive, making antibodies poorly suited for industrial application.

### 2.2 Significant achievements

#### 2.2.1 Rational design

**Helix bundles**

The first \textit{de novo} designed protein—a two-helix bundle—was reported in 1976.\cite{1} Serge St-Pierre and Robert Hodges based their design on Francis Crick’s model of helix packing,\cite{2} the experimentally-derived structure of helix bundles in muscle,\cite{3} and the newly-discovered sequence of a primary muscle protein.\cite{4} They synthesized amphipathic \( \alpha \) helices—those having both hydrophilic and hydrophobic faces—and in solution, the helices associated to bury their hydrophobic surfaces. In general, helix bundles are relatively easy to design because \( \alpha \) helices are rigidly structured with 3.6 residues per turn. So St-Pierre and Hodges successfully designed amphipathic helices with a seven-residue amino acid repeat and did not require computational methods. Even so, their work is especially remarkable in that the reported peptides were painstakingly constructed through solution chemical synthesis.

Since the late 80s and early 90s, work on helix bundles has progressed rapidly, mostly thanks to widespread use of protein expression systems\cite{5} and solid-state chemical synthesis.\cite{6} William DeGrado’s group designed a four-helix bundle with a hydrophobic core and a hydrophilic surface.\cite{7} The core contained only leucine residues, and the surface contained positively and negatively charged amino acids—lysine and glutamic acid. The charged amino acids were selected to create stabilizing charge-charge on the surface. While the bundle was indeed quite stable, the group made no attempt to design a well-packed hydrophobic core, resulting in a loosely packed molten core, unlike most native proteins.\cite{8} A few years later, DeGrado’s group designed a three-helix bundle, but this time designed a well-packed hydrophobic core and complementary charge-charge
interactions between neighboring helices.\textsuperscript{15} The result was a crystallizable structure with a tightly packed core.

Also in the late 80s, Tomikazu Sasaki and Emil Kaiser reported another significant milestone.\textsuperscript{16} They designed a three-helix bundle with a hydrophobic pocket accommodating both a heme and an analine. Not only was this peptide the first \textit{de novo} designed protein capable of ligand binding, but the hemeprotein also catalyzed the formation of $p$-aminophenol, making it the first designed enzyme. Shortly thereafter, DeGrado’s group constructed the first \textit{de novo} designed metal binding protein by engineering a zinc binding site into their four-helix bundle.\textsuperscript{17}

Recent characterization of helix bundles provides well-defined strategies for their design. Mutagenesis experiments on a naturally occurring three-helix bundle has revealed patterns of amino acids controlling the number of helices the bundle.\textsuperscript{18} Helix bundles can now be designed by following a simple strategy\textsuperscript{19} (a) applying these patterns to determine the number of helices in the bundle, (b) running straightforward computational methods to create a well-packed hydrophobic core, and (c) creating complimentary patterns of charged amino acids on adjacent helices to promote the association of specific helix pairs. Because of their simplicity and stability, helix bundles have become a useful scaffold for a variety of applications\textsuperscript{20} such as designing metal binding sites with unusual geometry\textsuperscript{20} or engineering novel catalytic activity.\textsuperscript{21}

Other structures

After designing many helix bundles, chemists became interested in creating model systems for studying other structures, and the rational, \textit{de novo} design of $\beta$-sheets has been fairly successful. The unnatural amino acid \textit{cis}-proline\textsuperscript{22} favors a conformation compatible with $\beta$-hairpins, and initial three-stranded-$\beta$-sheet designs incorporated two \textit{cis}-proline residues to form hairpins between adjacent $\beta$-strands.\textsuperscript{23} Another designed three-stranded $\beta$-sheet, Betanova, used only natural amino acids.\textsuperscript{24} Although slightly unstable, that Betanova folded at all is attributable to an aromatic cluster inspired by a three-stranded $\beta$-sheet domain in a kinase-associated protein domain.\textsuperscript{25} The insight that aromatic interactions conferred significant stability permitted design of a fully stable, two-stranded $\beta$-sheet. The fully stable, 12-residue design contained a $\beta$-hairpin and two $\beta$-strands held together by two tryptophan–tryptophan interactions.\textsuperscript{26} With the exception of the helix bundles, the rational, \textit{de novo} design of hydrophobic cores is exceptionally challenging, so other methods have been used to move beyond helix bundles and $\beta$-sheets.

An extremely successful approach to peptide design is to redesign small domains of existing proteins. Perhaps the most studied of these is the villin headpiece,\textsuperscript{27,28} a self-folding, 35-residue, $\alpha$-helical domain of an actin binding protein. While fully structured in its original form, the headpiece has been redesigned for increased stability.\textsuperscript{29} And

\textsuperscript{*} \textit{Cis}-proline and other \textit{cis} amino acids do exist in nature, but the ribosome only incorporates \textit{trans}-proline into proteins. These \textit{trans}-prolines may be later converted to \textit{cis}-prolines, but when a peptide is initially synthesized with a \textit{cis}-proline, the \textit{cis}-proline is unnatural.

\textsuperscript{†} This domain is the WW domain, named after the two conserved tryptophans, one of which participated in an aromatic cluster. When Betanova was published, the WW domain was not known to fold independently.
because the villin headpiece is small and folds on microsecond timescales, folding can be simulated with molecular dynamics and compared with experimental data. An excellent model system for β-sheet formation is the 57-residue WW domain of Pin1, a cis-trans isomerase. The Pin1-WW domain is capable of folding independently into a three-stranded β-sheet and has been redesigned to fold on microsecond timescales. Another interesting peptide is derived from extendin-4, a 39-residue peptide in Gila monster venom. The structure consists of an α-helix packed against an extended, β-strand-like region. While only partially folded in water, extendin-4 has been shortened and extensively redesigned into a fully stable, 20-residue peptide that folds on microsecond timescales. Finally, the simplest imaginable α/β-structure is a two-stranded β-sheet and an α-helix, a domain found in the zinc-finger motif. A 23-residue redesign of Zif268 eliminated the zinc-binding site and contained cis-proline to promote β-hairpin formation, an aromatic cluster to stabilize the β-sheet, and hydrophobic surfaces on the α-helix and β-sheet to hold them together.

### 2.2.2 Computational design

Early research in protein structure prediction demonstrated that backbone conformation and van der Waals interactions almost completely define the identity of hydrophobic amino acids buried within the core. The predicted-best sequence was indeed more stable than the natural design, but while the two other designs still folded, they were less stable. Optimizing van der Waals interactions alone poorly account for the destabilizing effects of voids in the protein core, so the failure of two designs is not surprising. Another potential problem was that many of the redesigned core residues were partially solvent exposed, and bulky hydrophobic residues may protrude into the solvent also destabilizing the structure. A later attempt at redesigning the hydrophobic interface of a two-helix bundle accounted for voids by calculating buried side chain volume. Not surprisingly, the predicted-best design was

- redesign of only specific positions, fixed backbone, molecular mechanics force-field, predicted stabilizing mutations were actually stabilizing!

Through consideration of van der Waals forces alone, a redesign of phage 434 protein core residues resulted in sequences with five to seven changes in the core residues.

- redesign of natural proteins [stabilization of enzyme] [stabilized enzyme] [crystal structure of redesigned natural protein]

### 2.2.3 Hellinga controversy

The two most well-known groups in computational protein design are probably David Baker’s group at the University of Washington and Homme Hellinga’s group at Duke. Unfortunately, most researchers now view Hellinga’s research with suspicion due to accusations of misconduct within his group, and some scientists even question the validity of all of his work. A brief description of this controversy is presented here.

After the retraction of two high-profile papers on a computationally designed enzyme, Hellinga accused his graduate student, Mary Dwyer, of misconduct and re-
quested that Duke conduct a formal investigation\(^{56}\). Two months later, the University cleared Dwyer of all allegations\(^{57}\). In an interview with Nature, Dwyer claimed that Hellinga not only published the paper despite her objections but also made baseless accusations.\(^{58}\) Shortly thereafter and without the help or knowledge of Dwyer, a petition for the investigation of her claims was signed by eighteen current and former biochemistry graduate students and delivered to University officials.\(^{59}\) Less than a month later, Hellinga wrote to Nature that he himself had requested—and been granted—an impartial inquiry into his own actions.\(^{60}\) Two years passed before the investigation was finally put to rest\(^{61}\), but the University remains unwilling to disclose the outcome.\(^{62}\)

The Hellinga group also reported landmark success using DEZYMER to design ligand-binding proteins for use as biosensors.\(^{61,62}\) But results from another lab show that at least five of the designed proteins failed to bind to their intended target.\(^{63}\) David Baker, a highly respected member of the protein design community and spectator to this debacle commented, “it’s a little hard to see how [the reevaluated proteins] could have worked as designed.”\(^{64}\)

### 2.2.4 Metalloprotein design

[de novo design] [probability based de novo design]\(^{65}\) [dimetal helix]\(^{66}\) [helix redesign preferentially bind nonnatural cofactor]\(^{67,68}\) [design with beta sheets]\(^{69}\) - cheats a bit by designing explicit Trp-Trp interactions to stabilize hairpins

[METAL-SEARCH] \(^{70,71}\) [Cys2His2 zinc site]\(^{72}\) [Cys1His1 zinc site]\(^{72}\) DEZYMER\(^{73}\) [copper]\(^{74}\) [iron]\(^{75}\) [calcium]\(^{76}\) [lead]\(^{77}\)

In this section, I reference many DEZYMER-designed proteins reported by Hellinga’s group, but I believe that these results are believable for several reasons: (1) A group at another institution has published a DEZYMER-designed lead binding site,\(^{77}\) and no one from Hellinga’s group or Duke appears in the author list. (2) A second group has reported a DEZYMER-designed calcium binding site,\(^{76}\) and while Hellinga appears in the middle of the author list, he is the only author from Duke, suggesting only a minor contribution. This group has also published the crystal structure of the designed calcium binding protein.\(^{78}\) (3) Metal binding sites have been designed by other groups by both rational design\(^{14,17}\) and non-DEZYMER computational methods.\(^{70,72}\) (4) Hellinga reports extensive characterization of his designed copper binding protein\(^{74}\) using direct measurements of ligand binding; unless he is guilty of outright fraud, these results seem trustworthy.

### 2.2.5 Binding site design

[designed DNA binding; zinc fingers]\(^{79}\)

[designed protein sequence recognition]\(^{80}\)

[designed protein-protein interface]\(^{81}\)
2.2. SIGNIFICANT ACHIEVEMENTS

2.2.6 Biosensor design

The survival of cells depends on their remarkable ability to rapidly detect and respond to environmental factors, and harnessing this ability is the goal of biosensor design. One can envision small devices to detect minute quantities of molecules of medical significance, explosives, drugs, toxins, and more. The trick, of course, is discovering and applying elements in Nature’s remarkable toolbox.\(^6\)

The binding of proteins with other molecules often produces measurable differences in mass, charge, or optical characteristics.\(^8\) Unfortunately, these easily measured properties also change during nonspecific binding and are therefore ill suited to analyzing complex mixtures. The problem of nonspecific binding makes the design of biosensors quite challenging.

Instead of measuring differences in gross characteristics, natural biosensors detect substances through binding-induced structural changes.\(^8\) A protein containing a binding pocket contacts its substrate and change shape slightly to surround the bound molecule. Nonspecific binding still occurs, but differences in the charge or shape of the molecule typically prevent the protein from surrounding the substrate in the same way.

For example, consider the probable mechanism for activation of a melanocortin receptor. The MC3R receptor binds an extracellular substrate such as α-MSH, causing a conformation change at the binding pocket, moving a nearby transmembrane α-helix helix. Because α-helices are rigid rod-like structures, the conformational change is transmitted through the membrane and alters the intracellular portion of the receptor. This tiny structural change brings key catalytic residues into position, allowing the receptor to catalyze the chemical modification an intracellular signaling protein. The signal protein travels through the cell and similarly activates the next step in the signaling pathway. Ultimately, gene expression is altered and the correct response to the protein is initiated.

The green florescent protein (GFP) is extensively used in producing biosensors.\(^8\) GFP was first discovered in *Aequorea victoria*, a luminescent jellyfish. These animals produce blue light from chemical energy with the aequorin enzyme, but much of this energy is absorbed by GFP, which fluoresces and produces the characteristic blue-green glow. GFP is unusual because florescence at visible wavelengths is an extremely unusual phenomenon in typical proteins unless bound to a second fluorescent compound. GFP gains this unusual character during folding when the protein undergoes self-catalysis. The process alters the chemical structure of buried amino acids, producing a fluorophore in the hydrophobic core. Since the discovery of GFP, rational design has produced a wide variety of GFP mutants with various properties including increased stability and energy conversion as well as a wide selection of emitted wavelengths. Biologists now routinely monitor gene expression by appending GFP-coding DNA to a gene of interest. Modified cells fluoresce upon expression of GFP-tagged genes, and florescence microscopy can pinpoint the cellular locations tagged proteins.

To design biosensors, one possibility is to create a modified GFP which only fold in the presence of a ligand. To this end, Toshihisa Mizuno and coworkers used a designed four-helix bundle with a metal-binding site.\(^8\) In the absence of metal ions, the helix was unstructured. Using a GFP variant, they inserted two helices of the bundle into...
adjacent loops on the surface of GFP. Upon binding metal, the newly structured helices pulled the adjacent GFP loops together, allowing GFP to fold and fluoresce.

A more complex but practical solution is to exploit Förster resonance energy transfer (FRET) between two fluorophores. Cyan fluorescent protein (CFP) and yellow fluorescent protein (YFP) are both GFP mutants and commonly used for FRET signaling. CFP fluoresces upon exposure to violet light, but YFP does not. When the two proteins are in close proximity—on the order of several tens of angstroms—most of the energy normally emitted by CFP is transferred to YFP through FRET and YFP begins to fluoresce. The experimenter observes an emergence of YFP fluorescence and a corresponding decrease of CFP fluorescence. When applied to biosensors, the strategy is to design binding-induced structural changes that bring CFP and YFP into close proximity. The major advantage of FRET biosensors over simple fluorescence is observable signals in both bound and unbound states. Together, the two signals indicate the proportion of sensors in the bound state and, indirectly, the concentration of the protein-bound molecule.

Due to the relative ease of designing metal binding sites, zinc sensors were among the first applications. Kristian Jensen and coworkers designed zinc binding sites between the surfaces of CFP and YFP. In the absence of zinc, the redesigned CFP and YFP proteins were monomeric and seldom in close proximity. But in the presence of high concentrations of zinc, the proteins formed dimers and produced a detectable FRET signal. Later, the group linked the designed proteins together with a long, flexible peptide. The linker lowered the entropy cost of dimer formation and increased zinc sensitivity to the micromolar range.

With the exception of metal binding sites, the designing ligand-binding is extremely difficult. A common strategy is to circumvent the design problem by linking CFP and YFP to natural proteins that undergo large conformational changes upon ligand-binding. When combined with fluorescence microscopy, this class of biosensors have been used to observe maltose uptake in yeast, glutamate transmission between neurons, calcium localization in mammalian tissue-culture, and glucose metabolism in Arabidopsis.

The largest drawback of FRET signaling is that changes in the CFP-YFP distance must be quite large relative to structural changes of most natural ligand-binding proteins. A clever trick to introduce large conformational changes in a binding protein is to duplicate the C-terminal region of the protein which contains a portion of the binding site. The duplicated region is prepended to the N-terminus of the protein with a flexible linker. When expressed, the mutant gene will likely contain a disordered C-terminus and a naturally folded N-terminus regardless of whether the protein is bound or unbound. By rationally changing the binding site residues in the N-terminal region, the normally folded N-terminal portion can be rendered incapable of ligand-binding. In the presence of the appropriate substrate, the normally duplicated N-terminals replaces the modified C-terminus in the structure. This trick does require that the binding of the duplicated N-terminal region be competitive with the modified C-terminus, and rational design may be required to stabilize or destabilize alternative structures. Nonetheless, this technique permits the redesign of natural ligand binding proteins with their exquisite sensitivity and specificity and sidesteps the difficult problem of designing de novo ligand-binding sites.
An alternative solution is the use of other two-fluorophore signaling systems, such as attaching pairs of pyrene-based compounds. The p17 HIV protein readily elicits antibodies recognizing a 6-residue surface region (epitope). When the epitope is synthesized as six-residue peptide, the peptide is disordered in solution, but when antibody-bound, become extended as in the native p17 conformation. A biosensor for p17 antibodies can be formed by flanking the epitope with two cysteine-bound pyrene-based fluorophores. In the antibody-bound form, the distant pyrenes fluoresce. When separated, however, the pyrenes form dimer, called excimers, with red-shifted fluorescence. Unlike FRET pairs, excimer pairs are sensitive to changes in distance and geometry. The disadvantage is that excimer pairs must be chemically attached to proteins in vitro. One strategy is to engineer two cysteine residues near the binding site and attach pyrene-based compounds to exposed thiol (SH) groups, which makes the synthesis more difficult.

A general problem with fluorescent biosensors is that many samples, including many body fluids, contain fluorescent compounds that make accurate measurement of biosensor fluorescence challenging. Another approach is to design electrochemical sensors. Which has been tremendously successful when applied to nucleic acid-based sensors. And despite their simplicity, nucleic acid based sensors can have exquisite sensitivity and specificity. Even so, this technique has great promise for designing protein-based sensors. For example, the 6-residue p17 HIV epitope mentioned before was covalently attached to an electrode at one end and bound to methylene blue (MB; a redox indicator) at the other. The antibody-unbound peptide is normally disordered, and the MB is normally fully exposed to the sample but still able to contact the electrode. Because the majority of samples are oxidizing from dissolved atmospheric oxygen, the peptide-bound MB is normally oxidized and carries a positive charge. But when MB contacts the negatively charged electrode, it is readily reduced to a neutrally charged group. The sample reoxidizes MB, and the cycle repeats. So when the peptide is antibody-unbound, a low current flows from the negatively charged electrode to the oxidizing solvent, but when the antibody is present, the imposed structure prevents MB from contacting the electrode, producing a measurable reduction in current. This particular biosensor design is capable of measuring nanomolar antibody concentrations in realistically complex samples and is also reusable since the biosensor is permanently attached to a removable electrode.

The final strategy for designing biosensors is to link them to enzymatic activity, which presents interesting and unconventional possibilities. Linking a protease with its inhibitor via a segment of single stranded DNA can result in a sensor for the complementary strand of DNA. The flexible single-stranded DNA permits repression. When the complementary strand is present, it binds to the strand linking the inhibitor to the enzyme. The more rigid, rod-like structure of double-stranded DNA separates the enzyme and inhibitor and allows protease activity.

Another way to link bioactivity with substrate recognition is to insert a binding domain into a loop of an enzyme. When properly designed, the binding domain brings two halves of the enzyme together, forming the bioactive enzyme. For example, by inserting a maltose binding domain into a gene encoding ampicillin resistance, one can modulate the activity of that enzyme by altering the maltose concentration. Alternatively, by inserting a glucose-binding domain into a loop of aequorin, a light-producing enzyme,
one can design an enzyme that only generates light in the presence of glucose. Nearly all of the work cited in this section has employed traditional molecular biology techniques, suggesting that the computational design of biosensors remains a mostly unexplored area. One possible explanation is that Hellinga’s early reports of phenomenal success indicated that biosensor design was essentially a solved problem. At the time, researchers in the area chose between working on related problems or applying time and resources toward publishing me-too results in second-tier journals. After the publication of strong evidence contradicting Hellinga’s earlier findings, researchers now have a new dilemma: they must work in other areas or endure unusually harsh peer-review.

2.2.7 Enzyme design

<table>
<thead>
<tr>
<th>prior enzyme design</th>
<th>catalytic antibodies</th>
<th>2000–4-helix bundle library ORBIT</th>
<th>2004–4-helix bundle library</th>
<th>2007–4-helix bundle library</th>
</tr>
</thead>
<tbody>
<tr>
<td>modern enzyme design</td>
<td>rosettamatch</td>
<td>baker retro-aldol</td>
<td>baker kemp-elimination</td>
<td>baker diels-alder</td>
</tr>
</tbody>
</table>

2.3 Overall impressions

first glance is rather depressing

rational design precedes computational design
rational enzyme design precedes computational design

Zif268 redesign replicates chemistry w/ zero amino acids
Top7 is simply two Zif286 domains linked w/ a beta sheet
Diels alder replicates antibody design--so not quite new function

not all hope is lost
maybe not a magic bullet, but it does help with packing

* better link to biological data
* rational constraints
* multiple methods---neural networks, structure prediction, MD, and rational design

References


2.3. OVERALL IMPRESSIONS


48. Dantas G, Kuhlman B, Callender D, Wong M, Baker D. A large scale test of computational protein design: Folding and stability of nine com-


58. Hellinga HW. In the wake of two retractions, a request for investigation. Nature. 2008;454(7203):397. doi:10.1038/454397b


2.3. OVERALL IMPRESSIONS

64. Hayden EC. Key protein-design papers challenged. Nature. 2009;461(7266):859. doi:10.1038/461859a


70. Regan L, Clarke ND. A tetrahedral zinc(II)-binding site introduced into a designed protein. Biochemistry. 1990;29(49):10878–10883. doi:10.1021/bi00501a003


77. Shete VS, Benson DE. Protein design provides lead(II) ion biosensors for imaging molecular fluxes around red blood cells. Biochemistry. 2009;48(2):462–470. doi:10.1021/bi801777h


82. Walt DR. Ubiquitous sensors: When will they be here? 2009;3(10):2876–2880. \doi{10.1021/nn901295n}


91. Deuschle K, Chaudhuri B, Okumoto S, Lager I, Lalonde S, Frommer WB. Rapid metabolism of glucose detected with FRET glucose nanosensors in epidermal cells


