Designing protein structures and complexes with the molecular modeling program Rosetta

Brian Kuhlman[‡§1]

From the ‡Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, North Carolina 27599-7260 and the §Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina 27599

Edited by Wolfgang Peti

Proteins perform an amazingly diverse set of functions in all aspects of life. Critical to the function of many proteins are the highly specific three-dimensional structures they adopt. For this reason, there is strong interest in learning how to rationally design proteins that adopt user-defined structures. Over the last 25 years, there has been significant progress in the field of computational protein design as rotamer-based sequence optimization protocols have enabled accurate design of protein tertiary and quaternary structure. In this award article, I will summarize how the molecular modeling program Rosetta is used to design new protein structures and describe how we have taken advantage of this capability to create proteins that have important applications in research and medicine. I will highlight three protein design stories: the use of protein interface design to create therapeutic bispecific antibodies, the engineering of light-inducible proteins that can be used to recruit proteins to specific locations in the cell, and the *de novo* design of new protein structures from pieces of naturally occurring proteins.

Essential to the function of all proteins is their capacity to specifically bind other molecules. Regardless of whether the binding partner is a single atom or a chromosome, binding specificity is derived from the ability of proteins to fold into unique three-dimensional structures with well-defined binding pockets and surfaces. Because of the strong link between protein structure, binding, and function, there is longstanding interest in learning how to design proteins that adopt predetermined structures and complexes (1). In addition to providing an approach for creating molecules with important applications in medicine, protein design tests our understanding of the critical determinants of protein structure and function (2).

For the last 20 years, my research has focused on the development and application of computational methods for protein design. This work began while I was a postdoctoral fellow in David Baker's laboratory at the University of Washington and has continued in the context of the RosettaCommons consortium, a large set of research groups (now over 60 groups) that collaborate to develop and extend the capabilities of the molecular modeling program Rosetta. Whereas Rosetta was first developed for predicting protein structure (3), it is now widely used for a variety of protein design problems, including stabilizing naturally occurring proteins (4), the *de novo* design of new protein structures (2, 5), and the design of protein complexes (6).

Protein design with Rosetta

All computational methods that have been developed for protein design include protocols for optimizing amino acid sequences for a specified protein structure (Fig. 1) (7). These protocols aim to identify sequences that form tightly packed hydrophobic cores, satisfy the hydrogen bond capacity of all backbone and side chain polar groups, and minimize torsional strain. The sequence optimization protocol in Rosetta contains two primary components: an energy function for ranking the fitness of alternative sequences for a particular protein structure and a search protocol to find lower-energy sequences.

The energy function in Rosetta is a linear combination of a Lennard–Jones potential that models van der Waals forces and steric repulsion, an implicit solvation model that penalizes the burial of polar groups and favors the burial of hydrophobic groups, an orientation-dependent hydrogen-bonding term, a short-range electrostatics potential, and knowledge-based potentials for scoring side-chain and backbone torsion energies (5, 8, 9). We use an implicit solvation model as it is computationally too expensive to model bulk water during a design simulation. Since first creating the energy function, we have placed a strong emphasis on parameterizing it to produce protein models that reproduce structural and sequence features observed in naturally occurring proteins (10). For instance, the hydrogen bond potential and electrostatics potential in Rosetta were parameterized simultaneously so that the combined model favors hydrogen bond distances and orientations that closely resemble those in high-resolution crystal structures (11). To explicitly tune the energy function for protein design, we make frequent use of sequence recovery tests. In these benchmarks, Rosetta is used to redesign the sequences of a large set of naturally occurring proteins and complexes. Although it is not expected that the naturally occurring sequence for a protein will necessarily be the most stable, we have found that parameterizing the energy function to produce sequences that
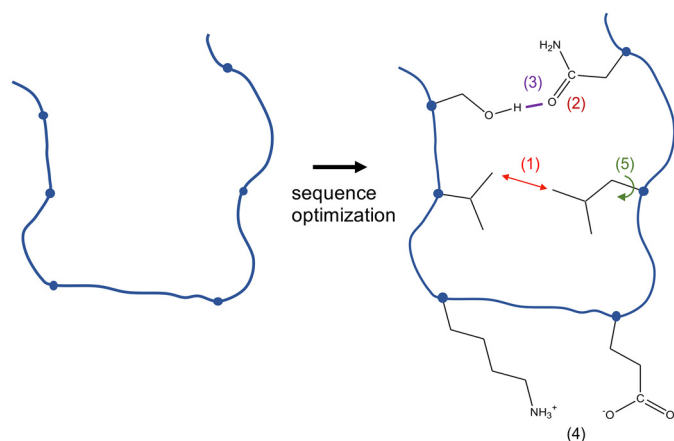
Figure 1. Computational protein design with Rosetta. Given a protein structure (*e.g.* the one shown to the *left*), the sequence optimization protocol in Rosetta searches for amino acid side chains that will maximize the stability of the protein. Protein stability is evaluated using an energy function that favors tight van der Waals contacts (*1*), penalizes the burial of polar groups (*2*), rewards hydrogen bonds with good geometries (*3*), favors interactions between unlike charges (*4*), and prefers side-chain conformations ("rotamers") that are frequently observed in naturally occurring proteins (*5*).

are similar to native sequences generates designs with high stabilities and good solubilities. Recently, Frank DiMaio and Hahnbeom Park have further improved the Rosetta energy function by parameterizing it on features from small molecules as well as macromolecules (9).

With an energy function in place, it is possible to search for alternative sequences with lower energy. Because the size of sequence space even for a small protein is astronomical, the search protocol cannot explicitly enumerate all possible sequences and conformations of the amino acid side chains. Instead, researchers have developed a variety of deterministic (converges to the same answer each time) and stochastic (can give different answers each time it is run) algorithms for identifying low-energy sequences (7). Deterministic methods like dead-end elimination are guaranteed to find the optimal solution if they converge but are computationally expensive and usually can only be applied to a region of a protein. Rosetta uses a stochastic approach based on Monte Carlo sampling to optimize the sequence of a protein (12). Single amino acid substitutions are automatically accepted if they lower the energy of the protein and accepted with some probability if the energy is raised. The probability for accepting an unfavorable change is governed by the Metropolis Criterion, which lowers the probability of acceptance when the energy changes are large or the system is at a lower temperature. Simulations are started with a high temperature to allow energy barriers to be crossed, and then the temperature is slowly cooled to let the system settle into an energy minimum. Despite the simplicity of this optimization scheme, independent trajectories starting from different random sequences but with the same protein backbone converge on sequences that are very similar to each other, suggesting that the system is not being trapped in false minima far from the global minima.

Some advantages of the Monte Carlo–based sequence optimization scheme in Rosetta are that it can be applied to large systems (hundreds to thousands of residues) and that it runs quickly; a sequence can be designed for a 100-residue protein in

a few minutes on a single computer processor. It is important that sequence optimization be rapid, as for many projects in protein design, it is necessary to search through both backbone conformational space and sequence space to find low-energy sequence/structure pairs (5). For instance, when designing a protein *de novo*, there is no guarantee that the initial model of the protein backbone will be designable (*i.e.* that there is any combination of the 20 naturally occurring amino acids that can stabilize the structure). We have found that it is usually necessary to make small adjustments to the protein backbone to find a structure where the side chains can pack tightly and buried polar groups can form hydrogen bonds. One of the strengths of Rosetta is that it has a variety of protocols for sampling alternative backbone conformations (13–15). This is a consequence of all of the structure prediction capabilities that have been introduced into the software and the collaborative nature of the team developing Rosetta.

With effective protocols for sequence optimization and backbone sampling, it is possible to address a wide range of problems in protein design. The sequence optimization protocols in Rosetta can be used to closely pack protein cores (2, 16), create favorable interactions at a protein-protein interface (17), and form tight interactions with ligands (18). Here, to convey the broad range of problems that can be solved with computational protein design, I will summarize three recent design projects from our laboratory at the University of North Carolina.

Creating bispecific antibodies with interface design

Bispecific antibodies are engineered antibodies that can bind to two antigens simultaneously. There is strong interest in bispecific antibodies because they can be used to gain higher-specificity binding to particular cell types (*i.e.* cells that display both antigens recognized by the bispecific) and they can be used to co-localize different cell types (19). Bispecific antibodies that simultaneously bind receptors on the surface of T cells and receptors on the surface of tumor cells are currently used in the clinic to treat cancer (20). Because of their utility, a number of strategies have been developed for creating bispecific antibodies. Most approaches involve fusing fragments of one antibody with fragments from another antibody. This can be effective, but these antibodies typically lose some of the features that make antibodies effective therapeutics. The antibody fragments often have poor stability and short serum half-lives. To circumvent these problems, we collaborated with Steve Demarest's group at Lilly to develop a method for creating bispecific antibodies that more closely resemble naturally occurring IgG antibodies (21–23). Given two monoclonal antibodies (call them A and B), we wanted to create antibodies where one arm of the antibody had the light chain of A paired with the heavy chain of A, and the second arm of the antibody had the light chain of B paired with the heavy chain of B. The challenge is that if one simultaneously expresses two light chains and two heavy chains in a single cell, there is no strong reason why the light chain from the first antibody should not assemble with the heavy chain from the second antibody and vice versa. In fact, there are 10 different ways that the four chains can assemble, with only one of them being the desired assembly (Fig. 2).
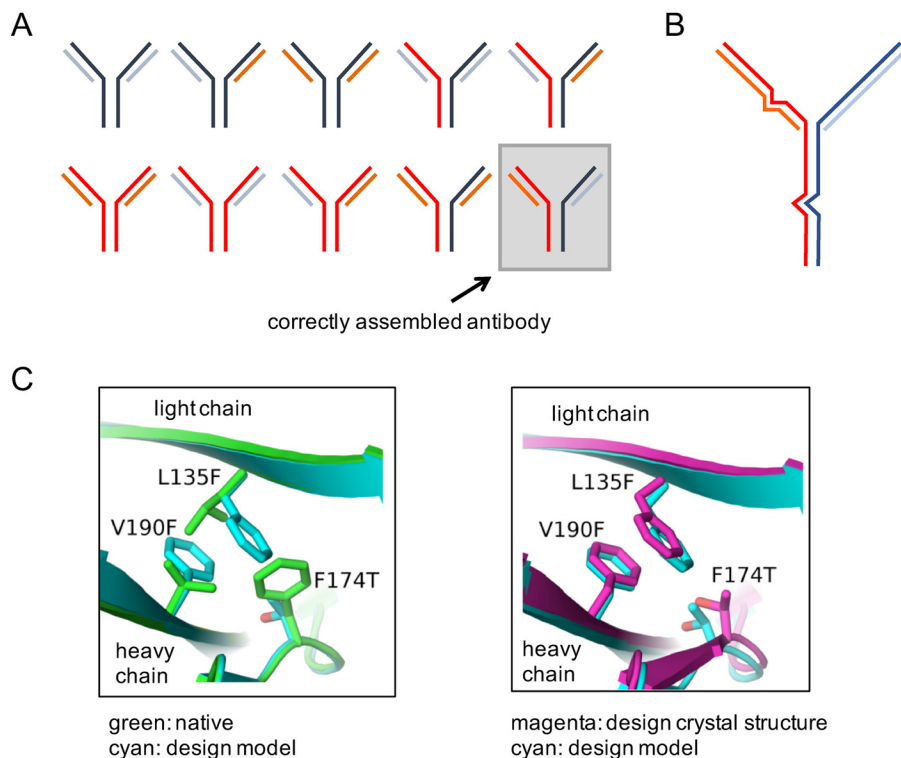
Figure 2. Design of altered specificity protein-protein interfaces promotes assembly of bispecific antibodies. *A*, there are 10 different ways that two antibody light chains and two antibody heavy chains can assemble if co-expressed in a single cell. Only one of these possibilities matches the desired bispecific IgG, where one arm of the antibody binds to one antigen and the other arm binds to a different antigen. *B*, to promote assembly of properly formed bispecific antibodies, we used Rosetta to redesign interactions between the antibody constant domains (shown as *bumps* and *holes*) so that the heavy chains would heterodimerize and each light chain would partner with the correct heavy chain. *C*, structures of a redesigned interface between the antibody light and heavy chains. The *left panel* shows that the redesign (*cyan*) introduces a bump in the light chain with leucine 135 mutated to phenylalanine and that this is accommodated with a hole in the heavy chain by mutating phenylalanine 174 to a threonine. An additional valine to phenylalanine at position 190 in the heavy chain further stabilizes the interface by making a π-stacking interaction with phenylalanine 135. The *right panel* shows a close match between the design model and a crystal structure of the redesign.

A solution to the assembly problem was proposed by a research team at Genentech over 20 years ago (24). They realized that by redesigning the interfaces between antibody chains, it would be possible to change how they assemble. In their case, they focused on breaking symmetry between the two antibody heavy chains to allow the formation of heavy-chain heterodimers (in naturally occurring antibodies, the heavy chains homodimerize). This is a necessary step in creating bispecific IgG antibodies, but it still does not solve the pairing problem between the light and heavy chains. Inspired by Genentech's approach, we used Rosetta to create altered specificity interfaces between antibody heavy and light chains so that we could direct the proper assembly of bispecific IgG antibodies.

In an IgG antibody, the constant domain of the light chain (C_L) makes a tight interface with the first constant domain (C_H1) of the heavy chain. Our goal was to make mutations to the two domains, the C_L and the C_H1, so that the redesigned variants would interact with each other but would no longer interact with the WT domains. To identify such mutations, we used a protocol for multistate protein design that we have created in Rosetta (25). During multistate design, an amino acid sequence is optimized for more than one feature. In this case, we wanted to simultaneously favor the interaction between the redesigned C_L and C_H1 domains while disfavoring interactions between the redesigns and the WT sequences. Designing

against a particular outcome is often referred to as "negative design" in the protein engineering field.

Twenty redesigned C_L-C_H1 interfaces predicted to be orthogonal to the WT interface were chosen for experimental characterization (23). Approximately half of the designs failed to express at levels similar to the WT protein. Most of these failures were cases where a large number of mutations were made to the interface (>10 mutations). Of the designs that expressed well, two of them (CRD1 and CRD2) displayed strong specificity for the redesign over the WT partners. Crystal structures of the designs showed that the Rosetta modeling correctly predicted how the redesigned side chains would interact. CRD1 relies on both redesigned hydrophobic packing and redesigned electrostatic interactions to create specificity, whereas CRD2 relies primarily on perturbed packing between hydrophobic amino acids.

Interestingly, creating an altered specificity interaction between the C_L and C_H1 domains was not fully sufficient for forcing the desired light chain/heavy chain pairing when assembling bispecific antibodies. Design simulations identified additional mutations in the framework regions of the variable domains (V_H and V_L) that further enhanced specificity and when combined with the mutations in the constant domains promoted the proper assembly of bispecific IgG antibodies. Particularly exciting was the finding that the mutations worked for a variety of antibody pairs that we tested and that the *in vivo*
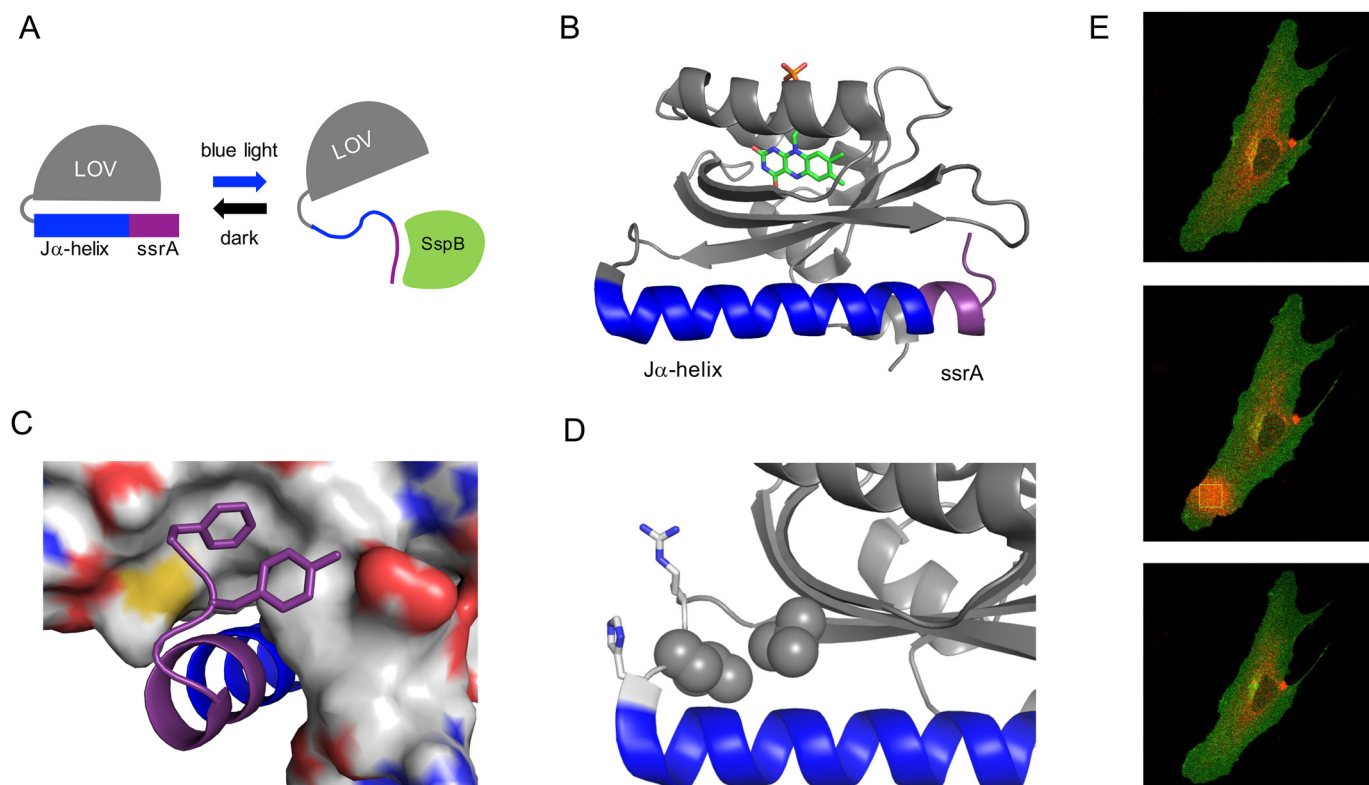
Figure 3. Design of a light-inducible protein dimer. *A*, the peptide, ssrA, was embedded in the Jα-helix of the LOV2 domain so that in the dark, it is sterically blocked from binding SspB, but when the Jα-helix undocks from the LOV2 domain upon stimulation with blue light, the ssrA peptide can bind SspB. *B*, a crystal structure of the engineered switch (called iLID, for improved light-inducible dimer) with the flavin cofactor shown in *stick format*. The crystal structure of iLID shows that a phenylalanine introduced at the C terminus of the ssrA peptide packs against the surface of the LOV2 domain and stabilizes the closed state (*C*) and that redesigned residues in the hinge connecting the Jα-helix to the LOV2 domain also form stabilizing interactions (*D*). *E*, the designed switch can be used along with a laser (illumination of the *yellow box* of the *middle panel*) to recruit protein to a specific region of a cell. In this experiment, iLID was fused to a protein that anchors it in the plasma membrane, and SspB was fused to a red fluorescent protein. The *three panels* show before illumination (*top*), during illumination (*middle*), and after illumination (*bottom*). Recruitment of SspB to the illuminated spot is readily apparent.

pharmacokinetic parameters and the thermodynamic stability of the bispecific IgGs were similar to their parental monoclonal antibodies. We have demonstrated that the bispecific IgG antibodies can be used to bind to two cell surface receptors simultaneously and that they can be used to redirect T cells to kill tumor cells by engineering one arm of the antibody to bind the CD3 receptor on the surface of T cells and the other arm to recognize the epidermal growth factor receptor, which is over-expressed on some cancers. The designed mutations are now being used by Lilly and the start-up company Dualogics to create novel bispecific antibodies with therapeutic potential.

Designing light-sensitive proteins for optogenetics

Optogenetics is a field of research based on using light to control biological processes in living systems. The advantage of using light to control cell signaling is that it affords precise spatial and temporal control. With a laser, it is possible to activate a protein within a specific region of a cell or within a specific set of cells in an organism, and with some light-sensitive systems, it is possible to rapidly and reversibly turn the system on and off by simply turning the laser on and off. Optogenetics first gained traction with the use of naturally occurring light-sensitive proteins such as channelrhodopsin (26), but recently scientists have also engineered a variety of proteins that allow control over biological processes (27).

In our laboratory, we have focused on redesigning the light-sensitive LOV2 domain from the *Avena sativa* protein phototropin to be a general tool for controlling cell signaling pathways. The LOV2 domain binds a flavin mononucleotide cofactor that when activated with blue light forms a metastable covalent bond with a cysteine in the core of the protein. This new covalent bond leads to structural perturbations throughout the protein and destabilizes interactions with a long α-helix (Jα-helix) at the C terminus of the protein (Fig. 3) (28). Once the interactions are disrupted, the Jα-helix rapidly undocks from the protein and unfolds. We and others have shown that this undocking event can be used to control the accessibility of proteins that are fused near the end of the Jα-helix (29–31). In particular, partially embedding a peptide in the last few residues of the Jα-helix has proven to be a robust strategy for regulating the binding of the peptide to other proteins. For instance, by placing a nuclear localization signal in the end of the Jα-helix, we created a switch that would only localize to the nucleus after activation with blue light (32–34).

One challenge we have faced with some peptides we have attempted to cage with the LOV2 domain is that even in the dark, they have exhibited appreciable affinity for their binding partner. This was true of our first attempt to control binding between the ssrA peptide and its binding partner SspB (29).

Because the ssrA motif and the SspB protein are bacterial and do not interact with eukaryotic proteins, we hypothesized that placing this interaction under the control of light would provide a powerful system for controlling protein-protein interactions in eukaryotic systems. By fusing the LOV2-ssrA construct to protein X and SspB to protein Y, it should be possible to regulate the proximity of X and Y with light. However, a simple fusion of the ssrA peptide near the end of the Jα-helix only results in a 2-fold change in affinity for SspB with light stimulation. The change was so small because in the dark LOV2-ssrA bound to SspB with an affinity that was only 2-fold less than affinity of the uncaged peptide.

We used Rosetta in a couple of ways to improve the dynamic range of the LOV2-ssrA switch (35). In both cases, the goal was to increase how tightly the ssrA peptide was held against the LOV2 domain in the dark. First, we modeled extensions off the C terminus of LOV2-ssrA to see whether any residues added to the end of the peptide could help hold it against the LOV2 domain. We observed that adding a single phenylalanine to the C terminus created favorable hydrophobic contacts with a small hydrophobic pocket on the surface of the LOV2 domain. This interaction was later confirmed with a crystal structure (Fig. 3) of the final engineered protein.

Our second approach to stabilize the dark state of LOV2-ssrA was to combine computational protein design with high-throughput selection experiments. We performed computational site saturation mutagenesis with Rosetta to identify point mutations likely to stabilize interactions between the core region of LOV2 domain and the Jα-helix. We then constructed a combinatorial library encompassing these mutations and used phage display to select library members that bind more tightly to SspB in the dark than in the light. The final selected sequence included several mutations on the surface of a β-sheet on the LOV2 domain that pack against the Jα-helix as well as mutations in a hinge-like loop that connects the Jα-helix to the rest of the domain. Our final designed sequence, named iLID (for improved light-inducible dimer), shows a 58-fold change in affinity between the dark state (47 μM) and the lit state (800 nM).

In the last few years, iLID has been used to control a diverse set of biological processes (36). It has been used to induce protrusions at specific regions on a cell surface by recruiting guanine exchange factors to that region of the cell (37). The Kiyomitsu laboratory has used the switch for spatial control of spindle-pulling forces in a dividing cell by inducing an interaction between NuMA and dynein (38), and the Brangwynne laboratory has used iLID to control localized liquid-liquid phase separation in living cells by recruiting intrinsically disordered macromolecules to each other (39).

Requirement-driven protein design with SEWING

The photoswitch and antibody projects that I have described are examples of template-based design, which is where the sequence and structure of a naturally evolved protein are modified to create a new function. Another type of protein design is *de novo* protein design (2). In *de novo* design, novel protein backbones and sequences are generated from scratch, guided by physiochemical constraints. Most projects in *de novo* protein design begin with defining the target fold/topology that one

would like to create, and then molecular modeling is used to create a protein backbone (or an ensemble of backbones) that adopt that fold (5). This is followed by rotamer-based sequence optimization to find a sequence that will stabilize the desired structure. This process is excellent for testing our understanding of protein folding and stability, but if the end goal of a project is to create a protein with a particular function, there may be many alternative protein folds that can provide the desired functionality. For this reason, we have been working on an approach to *de novo* protein design that does not require the user to predefine the target protein fold. Instead, the design process begins with a set of requirements that one wishes to satisfy. These requirements can be quite general (*e.g.* creating proteins that have a particular size groove on their surface), or they can be more specific (*e.g.* the inclusion of a particular functional motif). One problem that we are currently working on is the design of metal-binding proteins; in this case, the design requirement is that the protein include a metal-binding motif.

To enable requirement-driven design, we have created a design approach called SEWING that allows for the rapid generation of alternative protein structures that satisfy the design requirements (40). SEWING works by piecing together portions of naturally occurring protein structures (Fig. 4). To date, we have focused on using supersecondary structures—two secondary structures connected by a loop—as the building blocks for SEWING. We refer to these pieces of naturally occurring proteins as substructures. Starting from a single substructure, new proteins are assembled by sequentially adding substructures off the N or C terminus. Substructures are candidates for being pieced together if the C-terminal portion of one substructure structurally aligns with the N-terminal portion of another substructure (or vice versa) and if they do not clash when superimposed. A Monte Carlo assembly process is used to add and remove substructures from the design model, favoring models that optimize a low-resolution score function that rewards placing α-helices near each other at distances similar to those seen in naturally occurring proteins. The assembly process is rapid, and tens of thousands of alternative backbone models can be created in a few hours on a single computer processor.

As a first experimental test for SEWING, we aimed to create diverse helical bundle proteins with five or six helices. Eleven proteins were selected for experimental study. Four of the proteins were monomeric and were helical as evidenced by CD, and they unfolded cooperatively upon thermal denaturation. One of the designs was exceptionally stable and did not unfold even at 100 °C. A crystal structure of the design showed subangstrom agreement with the design model. The protein incorporates several interesting features, including a kinked helix and a small cavity in the protein core.

To perform requirement-driven protein design with SEWING, we have added several features to the protocol (41). SEWING can now be used to build proteins around small functional motifs. To design metal-binding proteins, we are starting with a single helix that contains two amino acid side chains coordinating a metal, and then during backbone assembly with SEWING, tertiary structures are searched for that bring in additional residues that can coordinate metal. Similarly, we are building proteins around peptide motifs that are known to interact with other proteins. The design
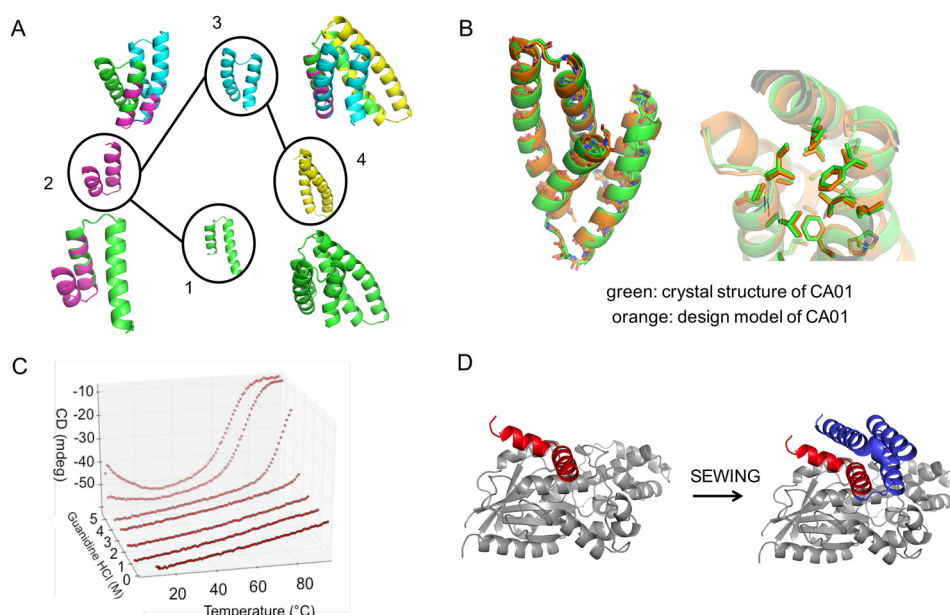
Figure 4. *De novo* protein design with SEWING. *A*, during SEWING, pieces of naturally occurring proteins (shown in the *circles*) are sequentially chimerized to create a larger structure. The design, CA01, was created from pieces of four proteins (1–4). The final assembly (*green*) is then input into the sequence optimization protocol of Rosetta to design a sequence for the protein. *B*, superposition of the design model and crystal structure of CA01. *C*, denaturation experiments followed by CD show that CA01 is exceptionally stable and only unfolds at high temperatures if chemical denaturant (guanidine hydrochloride) is also present. *D*, we are now using SEWING to perform requirement-driven design. In this example, SEWING is being used to build a protein around a naturally occurring peptide (*red*) and form additional contacts with the binding partner (*gray*).

process results in an expanded protein interface with the binding partner, providing an opportunity to increase binding affinity and specificity.

Learning from past design successes and failures

I have summarized results from three design projects that were a success and produced well-folded and functional proteins. However, in these projects many design sequences did not express or perform as intended, and in some design projects we have pursued in the laboratory none of the designed proteins behaved as desired. From a single design project, it is difficult to pinpoint the key factors that give rise to success or failure, but as our laboratory and others continue to work on a diversity of protein design problems, it is possible to gain some insight into the existing challenges in the field (42). A few years ago, our laboratory analyzed structural features in design models from several projects aimed at designing *de novo* protein-protein interactions (43). We observed that almost all of the design successes relied on protein-protein interfaces that were dominated by hydrophobic interactions. Attempts to design more polar interfaces (despite their prevalence in nature) generally resulted in proteins that did not interact. This observation spurred us and others to improve how we evaluate the energy of hydrogen-bonding interactions and develop methods for effectively sampling buried hydrogen bond networks (11, 44, 45). The design of buried polar interactions continues to be a significant challenge for the field, but there have been some impressive successes designing coiled-coils with buried hydrogen bond networks (44).

One interesting result that has been obtained in many *de novo* and template-based design projects is the generation of proteins with exceptionally high thermostabilities (5, 16, 40, 46). This indicates that naturally occurring proteins are gener-

ally not optimized for high stability. This could be for a variety of reasons. The function of many proteins requires them to adopt alternative conformations; overstabilization of one state could prevent switching to the alternative state. Once the thermostability of a protein reaches a certain level above the ambient temperature that an organism experiences, there may be little benefit or even a cost to raising stability further. For instance, placing hydrophobic amino acids at residues that are only partially buried often leads to a boost in protein stability (47), but this is likely to have deleterious effects on the solubility of the protein.

Conclusion

It is an exciting time for the field of structure-based protein design as the focus begins to shift from developing design methods toward applying these methods to important problems in biology and medicine (48). For instance, I have described our use of the modeling program Rosetta to create bispecific antibodies with therapeutic potential and our engineering of light-sensitive switches that control signaling pathways in living cells and animals. Our laboratory and others are exploring a variety of additional problems in protein design, including the creation of more effective vaccines (49), the generation of self-assembling protein materials (6), and the design of protein inhibitors and activators for disease-relevant signaling pathways (50). Future challenges for the field are numerous, including the rational design of enzymes with activities comparable with naturally occurring enzymes, the *de novo* design of proteins that incorporate allostery as an additional layer of regulation, and the design of functional loops such as those found in the antigen-binding sites of antibodies.

References

1. Betz, S. F., Raleigh, D. P., and DeGrado, W. F. (1993) *De novo* protein design: from molten globules to native-like states. *Curr. Opin. Struct. Biol.* **3,** 601–610 CrossRef

2. Huang, P.-S., Boyken, S. E., and Baker, D. (2016) The coming of age of *de novo* protein design. *Nature* **537,** 320–327 CrossRef Medline

3. Rohl, C. A., Strauss, C. E. M., Misura, K. M. S., and Baker, D. (2004) Protein structure prediction using Rosetta. *Methods Enzymol.* **383,** 66–93 CrossRef Medline

4. Goldenzweig, A., Goldsmith, M., Hill, S. E., Gertman, O., Laurino, P., Ashani, Y., Dym, O., Unger, T., Albeck, S., Prilusky, J., Lieberman, R. L., Aharoni, A., Silman, I., Sussman, J. L., Tawfik, D. S., and Fleishman, S. J. (2016) Automated structure- and sequence-based design of proteins for high bacterial expression and stability. *Mol. Cell.* **63,** 337–346 CrossRef Medline

5. Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L., and Baker, D. (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science* **302,** 1364–1368 CrossRef Medline

6. King, N. P., Bale, J. B., Sheffler, W., McNamara, D. E., Gonen, S., Gonen, T., Yeates, T. O., and Baker, D. (2014) Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **510,** 103–108 CrossRef Medline

7. Gainza, P., Nisonoff, H. M., and Donald, B. R. (2016) Algorithms for protein design. *Curr. Opin. Struct. Biol.* **39,** 16–26 CrossRef Medline

8. Alford, R. F., Leaver-Fay, A., Jeliazkov, J. R., O'Meara, M. J., DiMaio, F. P., Park, H., Shapovalov, M. V., Renfrew, P. D., Mulligan, V. K., Kappel, K., Labonte, J. W., Pacella, M. S., Bonneau, R., Bradley, P., Dunbrack, R. L., Jr., *et al.* (2017) The Rosetta all-atom energy function for macromolecular modeling and design. *J. Chem. Theory Comput.* **13,** 3031–3048 CrossRef Medline

9. Park, H., Bradley, P., Greisen, P., Jr., Liu, Y., Mulligan, V. K., Kim, D. E., Baker, D., and DiMaio, F. (2016) Simultaneous optimization of biomolecular energy functions on features from small molecules and macromolecules. *J. Chem. Theory Comput.* **12,** 6201–6212 CrossRef Medline

10. Leaver-Fay, A., O'Meara, M. J., Tyka, M., Jacak, R., Song, Y., Kellogg, E. H., Thompson, J., Davis, I. W., Pache, R. A., Lyskov, S., Gray, J. J., Kortemme, T., Richardson, J. S., Havranek, J. J., Snoeyink, J., *et al.* (2013) Scientific benchmarks for guiding macromolecular energy function improvement. *Methods Enzymol.* **523,** 109–143 CrossRef Medline

11. O'Meara, M. J., Leaver-Fay, A., Tyka, M. D., Stein, A., Houlihan, K., DiMaio, F., Bradley, P., Kortemme, T., Baker, D., Snoeyink, J., and Kuhlman, B. (2015) A combined covalent-electrostatic model of hydrogen bonding improves structure prediction with Rosetta. *J. Chem. Theory Comput.* **11,** 609–622 CrossRef Medline

12. Kuhlman, B., and Baker, D. (2000) Native protein sequences are close to optimal for their structures. *Proc. Natl. Acad. Sci.* **97,** 10383–10388 CrossRef Medline

13. Das, R., and Baker, D. (2008) Macromolecular modeling with Rosetta. *Annu. Rev. Biochem.* **77,** 363–382 CrossRef Medline

14. Mandell, D. J., Coutsias, E. A., and Kortemme, T. (2009) Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nat. Methods* **6,** 551–552 CrossRef Medline

15. Tyka, M. D., Jung, K., and Baker, D. (2012) Efficient sampling of protein conformational space using fast loop building and batch minimization on highly parallel computers. *J. Comput. Chem.* **33,** 2483–2491 CrossRef Medline

16. Murphy, G. S., Mills, J. L., Miley, M. J., Machius, M., Szyperski, T., and Kuhlman, B. (2012) Increasing sequence diversity with flexible backbone protein design: the complete redesign of a protein hydrophobic core. *Structure* **20,** 1086–1096 CrossRef Medline

17. Fleishman, S. J., Whitehead, T. A., Ekiert, D. C., Dreyfus, C., Corn, J. E., Strauch, E.-M., Wilson, I. A., and Baker, D. (2011) Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* **332,** 816–821 CrossRef Medline

18. Tinberg, C. E., Khare, S. D., Dou, J., Doyle, L., Nelson, J. W., Schena, A., Jankowski, W., Kalodimos, C. G., Johnsson, K., Stoddard, B. L., and Baker, D. (2013) Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* **501,** 212–216 CrossRef Medline

19. Labrijn, A. F., Janmaat, M. L., Reichert, J. M., Parren, P. W. H. I. (2019) Bispecific antibodies: a mechanistic review of the pipeline. *Nat. Rev. Drug Discov.* **18,** 585–608 CrossRef Medline

20. Krishnamurthy, A., and Jimeno, A. (2018) Bispecific antibodies for cancer therapy: a review. *Pharmacol. Ther.* **185,** 122–134 CrossRef Medline

21. Froning, K. J., Leaver-Fay, A., Wu, X., Phan, S., Gao, L., Huang, F., Pustilnik, A., Bacica, M., Houlihan, K., Chai, Q., Fitchett, J. R., Hendle, J., Kuhlman, B., and Demarest, S. J. (2017) Computational design of a specific heavy chain/κ light chain interface for expressing fully IgG bispecific antibodies. *Protein Sci.* **26,** 2021–2038 CrossRef Medline

22. Leaver-Fay, A., Froning, K. J., Atwell, S., Aldaz, H., Pustilnik, A., Lu, F., Huang, F., Yuan, R., Hassanali, S., Chamberlain, A. K., Fitchett, J. R., Demarest, S. J., and Kuhlman, B. (2016) Computationally designed bispecific antibodies using negative state repertoires. *Structure* **24,** 641–651 CrossRef Medline

23. Lewis, S. M., Wu, X., Pustilnik, A., Sereno, A., Huang, F., Rick, H. L., Guntas, G., Leaver-Fay, A., Smith, E. M., Ho, C., Hansen-Estruch, C., Chamberlain, A. K., Truhlar, S. M., Conner, E. M., Atwell, S., *et al.* (2014) Generation of bispecific IgG antibodies by structure-based design of an orthogonal Fab interface. *Nat. Biotechnol.* **32,** 191–198 CrossRef Medline

24. Ridgway, J. B., Presta, L. G., and Carter, P. (1996) "Knobs-into-holes" engineering of antibody CH3 domains for heavy chain heterodimerization. *Protein Eng.* **9,** 617–621 CrossRef Medline

25. Leaver-Fay, A., Jacak, R., Stranges, P. B., and Kuhlman, B. (2011) A generic program for multistate protein design. *PLoS One* **6,** e20937 CrossRef Medline

26. Govorunova, E. G., Sineshchekov, O. A., Li, H., and Spudich, J. L. (2017) Microbial rhodopsins: diversity, mechanisms, and optogenetic applications. *Annu. Rev. Biochem.* **86,** 845–872 CrossRef Medline

27. Johnson, H. E., and Toettcher, J. E. (2018) Illuminating developmental biology with cellular optogenetics. *Curr. Opin. Biotechnol.* **52,** 42–48 CrossRef Medline

28. Harper, S. M., Neil, L. C., and Gardner, K. H. (2003) Structural basis of a phototropin light switch. *Science* **301,** 1541–1544 CrossRef Medline

29. Lungu, O. I., Hallett, R. A., Choi, E. J., Aiken, M. J., Hahn, K. M., and Kuhlman, B. (2012) Designing photoswitchable peptides using the AsLOV2 domain. *Chem. Biol.* **19,** 507–517 CrossRef Medline

30. Strickland, D., Lin, Y., Wagner, E., Hope, C. M., Zayner, J., Antoniou, C., Sosnick, T. R., Weiss, E. L., and Glotzer, M. (2012) TULIPs: tunable, light-controlled interacting protein tags for cell biology. *Nat. Methods* **9,** 379–384 CrossRef Medline

31. Wu, Y. I., Frey, D., Lungu, O. I., Jaehrig, A., Schlichting, I., Kuhlman, B., and Hahn, K. M. (2009) A genetically encoded photoactivatable Rac controls the motility of living cells. *Nature* **461,** 104–108 CrossRef Medline

32. Yumerefendi, H., Dickinson, D. J., Wang, H., Zimmerman, S. P., Bear, J. E., Goldstein, B., Hahn, K., and Kuhlman, B. (2015) Control of protein activity and cell fate specification via light-mediated nuclear translocation. *PLoS One* **10,** e0128443 CrossRef Medline

33. Yumerefendi, H., Lerner, A. M., Zimmerman, S. P., Hahn, K., Bear, J. E., Strahl, B. D., and Kuhlman, B. (2016) Light-induced nuclear export reveals rapid dynamics of epigenetic modifications. *Nat. Chem. Biol.* **12,** 399–401 CrossRef Medline

34. Niopek, D., Benzinger, D., Roensch, J., Draebing, T., Wehler, P., Eils, R., and Di Ventura, B. (2014) Engineering light-inducible nuclear localization signals for precise spatiotemporal control of protein dynamics in living cells. *Nat. Commun.* **5,** 4404 CrossRef Medline

35. Guntas, G., Hallett, R. A., Zimmerman, S. P., Williams, T., Yumerefendi, H., Bear, J. E., and Kuhlman, B. (2015) Engineering an improved light-induced dimer (iLID) for controlling the localization and activity of signaling proteins. *Proc. Natl. Acad. Sci. U.S.A.* **112,** 112–117 CrossRef Medline

36. Yu, D., Lee, H., Hong, J., Jung, H., Jo, Y., Oh, B.-H., Park, B. O., and Do Heo, W. (2019) Optogenetic activation of intracellular antibodies for direct modulation of endogenous proteins. *Nat. Methods* **7,** 1010–1016 CrossRef Medline

37. Zimmerman, S. P., Asokan, S. B., Kuhlman, B., and Bear, J. E. (2017) Cells lay their own tracks: optogenetic Cdc42 activation stimulates fibronectin deposition supporting directed migration. *J. Cell Sci.* **130,** 2971–2983 CrossRef Medline

38. Okumura, M., Natsume, T., Kanemaki, M. T., and Kiyomitsu, T. (2018) Dynein-dynactin-NuMA clusters generate cortical spindle-pulling forces as a multi-arm ensemble. *eLife* **7,** e36559 CrossRef Medline

39. Shin, Y., Berry, J., Pannucci, N., Haataja, M. P., Toettcher, J. E., and Brangwynne, C. P. (2017) Spatiotemporal control of intracellular phase transitions using light-activated optoDroplets. *Cell* **168,** 159–171.e14 CrossRef Medline

40. Jacobs, T. M., Williams, B., Williams, T., Xu, X., Eletsky, A., Federizon, J. F., Szyperski, T., and Kuhlman, B. (2016) Design of structurally distinct proteins using strategies inspired by evolution. *Science* **352,** 687–690 CrossRef Medline

41. Guffy, S. L., Teets, F. D., Langlois, M. I., and Kuhlman, B. (2018) Protocols for requirement-driven protein design in the Rosetta modeling program. *J. Chem. Inf. Model.* **58,** 895–901 CrossRef Medline

42. Rocklin, G. J., Chidyausiku, T. M., Goreshnik, I., Ford, A., Houliston, S., Lemak, A., Carter, L., Ravichandran, R., Mulligan, V. K., Chevalier, A., Arrowsmith, C. H., and Baker, D. (2017) Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science* **357,** 168–175 CrossRef Medline

43. Stranges, P. B., and Kuhlman, B. (2013) A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds. *Protein Sci.* **22,** 74–82 CrossRef Medline

44. Boyken, S. E., Chen, Z., Groves, B., Langan, R. A., Oberdorfer, G., Ford, A., Gilmore, J. M., Xu, C., DiMaio, F., Pereira, J. H., Sankaran, B., Seelig, G., Zwart, P. H., and Baker, D. (2016) *De novo* design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. *Science* **352,** 680–687 CrossRef Medline

45. Maguire, J. B., Boyken, S. E., Baker, D., and Kuhlman, B. (2018) Rapid sampling of hydrogen bond networks for computational protein design. *J. Chem. Theory Comput.* **14,** 2751–2760 CrossRef Medline

46. Huang, P.-S., Oberdorfer, G., Xu, C., Pei, X. Y., Nannenga, B. L., Rogers, J. M., DiMaio, F., Gonen, T., Luisi, B., and Baker, D. (2014) High thermodynamic stability of parametrically designed helical bundles. *Science* **346,** 481–485 CrossRef Medline

47. Kim, D. N., Jacobs, T. M., and Kuhlman, B. (2016) Boosting protein stability with the computational design of β-sheet surfaces. *Protein Sci.* **25,** 702–710 CrossRef Medline

48. Gainza-Cirauqui, P., and Correia, B. E. (2018) Computational protein design-the next generation tool to expand synthetic biology applications. *Curr. Opin. Biotechnol.* **52,** 145–152 CrossRef Medline

49. Correia, B. E., Bates, J. T., Loomis, R. J., Baneyx, G., Carrico, C., Jardine, J. G., Rupert, P., Correnti, C., Kalyuzhniy, O., Vittal, V., Connell, M. J., Stevens, E., Schroeter, A., Chen, M., Macpherson, S., *et al.* (2014) Proof of principle for epitope-focused vaccine design. *Nature* **507,** 201–206 CrossRef Medline

50. Silva, D.-A., Yu, S., Ulge, U. Y., Spangler, J. B., Jude, K. M., Labão-Almeida, C., Ali, L. R., Quijano-Rubio, A., Ruterbusch, M., Leung, I., Biary, T., Crowley, S. J., Marcos, E., Walkey, C. D., Weitzner, B. D., *et al.* (2019) *De novo* design of potent and selective mimics of IL-2 and IL-15. *Nature* **565,** 186–191 CrossRef Medline

Designing protein structures and complexes with the molecular modeling program Rosetta

Brian Kuhlman

Access the most updated version of this article at doi: 10.1074/jbc.AW119.008144

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 50 references, 10 of which can be accessed free at
http://www.jbc.org/content/294/50/19436.full.html#ref-list-1