Protein Design: Past, Present and Future

Journal:	*Biopolymers: Peptide Science*
Manuscript ID:	Draft
Wiley - Manuscript type:	Invited Review
Date Submitted by the Author:	n/a
Complete List of Authors:	Regan, Lynne; Yale University, Department of Molecular Biophysics and Biochemistry; Yale University, Department of Chemistry; Yale University, Integrated Graduate Program in Physical and Engineering Biology Caballero, Diego; Yale University, Department of Physics; Yale University, Integrated Graduate Program in Physical and Engineering Biology Virrueta, Alejandro; Yale University, Department of Mechanical Engineering & Materials Science; Yale University, Integrated Graduate Program in Physical and Engineering Biology Hinrichsen, Michael; Yale University, Department of Molecular Biophysics and Biochemistry Williams, Danielle; Yale University, Department of Molecular Biophysics and Biochemistry O'hern, Corey; Yale University, Department of Mechanical Engineering & Materials Science; Yale University, Department of Physics; Yale University, Physics; Yale University, Department of Applied Physics; Yale University, Applied Physics; Yale University, Integrated Graduate Program in Physical and Engineering Biology
Keywords:	Protein Design, Computation, Review, Nanotechnology

Protein Design: Past, Present and Future

Lynne Regan[1,2,3,*], Diego Caballero[4,3], Michael R. Hinrichsen[1], Alejandro Virrueta[5,3], Danielle M. Williams[1], Corey S. O'Hern[5,4,6,3]

[1]Department of Molecular Biophysics & Biochemistry, Yale University;

[2]Department of Chemistry, Yale University;

[3]Integrated Graduate Program in Physical and Engineering Biology, Yale University;

[4]Department of Physics, Yale University;

[5]Department of Mechanical Engineering & Materials Science, Yale University;

[6]Department of Applied Physics, Yale University;

[*]Corresponding Author: lynne.regan@yale.edu

Abstract: Building on the pioneering work of Bill DeGrado and colleagues in the late nineteen-eighties, protein design approaches have revealed many fundamental features of protein structure and stability. We are now in the era that the early work presaged – the design of new proteins with practical applications and uses. Here we briefly survey some past milestones in protein design, in addition to highlighting recent progress and future aspirations.

Introduction

Perhaps no one was more surprised than the first protein designers themselves at how easy it was to create novel proteins that adopted the desired fold – more or less. Two complementary approaches were initially employed. DeGrado and colleagues sought to create four-helix bundle proteins (an apparently 'simple fold' that is still a challenge today[1]) using the minimal number of amino acid types and a systematic approach. First helices were designed based on amino acid propensities. Helix-helix interaction interfaces were then introduced, and the four helices were linked together. At each stage the designs were checked to ensure the desired behavior **(Figure 1)**. This strategy allowed for evaluation, and correction if needed, of each component of the design[2,3].

The Richardsons and colleagues adopted a complementary approach. They too sought to create a four-helix bundle protein, but their design goal was to maximize the number of amino acids types used so that the sequence was as 'natural' as possible[4] **(Figure 2)**. Both groups created monomeric, compact helical proteins as evaluated by simple solution methods, predominantly circular dichroism (CD). It is important to note that both these approaches were considering the 'protein folding problem' in reverse. They were not trying to predict what 3-D structure a particular sequence would adopt, but rather identify a sequence (by no means the only sequence) that was compatible with a particular fold. In parallel with these early designs, considerable effort was also employed to delineate and experimentally quantify the thermodynamic contributions of intrinsic α-helix and β-sheet forming propensities. The results of such measurements significantly influenced future design[5-9].

Following these early successes, the interests of the protein design field transitioned to creating proteins with 'native-like' thermodynamic properties. The *zeitgeist* was that although it was possible to design associating helices, they were perhaps associating in a 'molten-globule like fashion', rather than truly recapitulating the packing and associated thermodynamic properties that are characteristic of natural proteins. Subsequently, Munson, Regan and colleagues explicitly tracked how the thermodynamic behavior of proteins changes with hydrophobic core redesigns[10].

At the same time, researchers employed coarse-grained models to computationally design protein cores, with the pervading concept being that the residues in the core must be hydrophobic and pack efficiently[11-13]. Ponder and Richards promoted the concept of amino acid side-chain rotamers – that side-chains adopt a limited subset of dihedral angles. They demonstrated that rotamer and hydrophobicity constraints plus strict limits on the free volume greatly restricted the number of amino acid combinations that are compatible in the core of a small protein. Desjarlais and Handel used this type of approach with a 'custom rotamer library' to redesign the core of small proteins and to subsequently make and determine the structure of repacked ubiquitin[14]. Dahiyat and Mayo also used a rotamer-based approach in their ORBIT design software. In addition, they classified every amino acid position into one of three categories: buried, surface, or boundary. Each class was given a different scoring function, which included an atomic solvation potential that favored the burial and

penalized the exposure of nonpolar surface area **(Figure 3 left)**[15].

Baker and colleagues designed and experimentally validated the first protein fold not found in nature, 'Top7,' using their RosettaDesign software[16,17]. Their strategy was to construct the protein scaffold using 3- and 9- residue fragments taken from the Protein Data Bank (PDB). The best combinations were then selected via Monte Carlo optimization of a number of energetic terms including hydrophobic burial, beta-strand hydrogen bonding and side-chain rotamers **(Figure 3 right)**.

In addition to designing 'native-like' proteins, protein scientists also began to introduce new functions into proteins, with much early emphasis placed on the design of metal ion binding sites. The reasons were both pragmatic and exciting: pragmatic because there are many spectroscopic methods that can be used to characterize binding geometry in solution, and exciting because many activities are associated with metal ions in proteins, including catalysis, electron transfer and enhanced stability[18,19].

The design of protein-based catalysts (retro-aldol, Kemp elimination, and stereoselective Diels-Alder reactions) followed, based on creating a binding site for the transition state of the desired reaction[20-23]. The resulting designs exhibited modest activity, and most have been subsequently improved by random mutagenesis and selection[21]. The power of randomization and selection to improve initial designs has been repeatedly demonstrated. For example, of 88 designs for hemagglutinin binding modules, only 2 displayed any binding activity, but the low binding affinity of those designs were increased substantially by affinity maturation[24]. Remarkably, DeGrado and colleagues used chemical intuition to create a Kemp eliminase, with activity comparable to that of initial computational designs, by introducing a single Glu residue into a hydrophobic cavity of a small protein, and thus perturbing its pKa significantly[25].

DeGrado and colleagues have advanced 'knowledge-based' approaches for the design of constructs for diverse applications: transmembrane-binding peptides, surface-organizing peptide superstructures, and protein crystals. Their CHAMP protocol enabled the design of α- and β-helical peptides that bind specifically to a transmembrane helix of a target protein **(Figure 4)**[26,27]. Yin *et al.* leveraged known membrane-protein structures to create backbone templates and a membrane depth-dependent knowledge-based potential to sample the 'relevant' sequence and rotamer space. This strategy has resulted in several designs whose binding has been experimentally verified[26,27]. Grigoryan *et al.* implemented a set of rule selections to assemble a superstructure of peptides that coat single-walled nanotubes (SWNTs)[28]. They matched the periodicity of an α-helix to the periodic pattern surface of a SWNT via Ala C_β methyl contacts to form a supercoil of α-helical coiled coils. In the presence of mixed types of SWNTs, the designed peptides preferentially sequestered the targeted nanotube species to produce stable aqueous suspensions. Lanci *et al.* implemented a similar methodology for the *de novo* design of a peptide that self-assembles into a P6 protein crystal[29]. After determining the optimal crystalline array for a homotrimeric parallel coiled-coil and designing the sequences, they obtained a protein crystal that matched the computational design to sub-Å precision.

Current computational methods in protein design

Why has most of the computational design work employed 'knowledge-based' potentials rather than potentials based solely on molecular mechanics? Classical molecular mechanics force-fields, which employ simplified interaction potentials, offer computational speed and are straightforward to implement. Typical simplifications include employing pairwise interaction potentials, treating covalent bonds as Hookean springs and using Lennard-Jones-like potentials to model van der Waals, hydrophobic, and hydrogen bonding interactions. Strengths of this approach are that it is easy to build upon and straightforward to apply in scalable computer applications. Disadvantages include the artificial separation of interactions that are deeply intertwined, including van der Waals interactions, hydrogen bonding and hydrophobic interactions. This can result in 'double counting', difficulty in calibrating the relative energetic contributions of different types of interactions and a large number of unknown parameters that must be determined. In this context, improvements or updates to widely used software packages[30-32] can be classified as one of two types: (1) tweaking the relative magnitudes of different energy terms - often driven by improved experimental data, and (2) the addition of new energy terms, for example 'knowledge-based' potentials that ensure that the main-chain and side-chain dihedral angles preferentially sample the observed distributions from the PDB[33].

Although the global results for protein simulations and the prediction of structure from sequence are ever improving[34], their limitations have been documented, and there have been a number of suggestions for their improvement. In a recent assessment of different force-fields and their performance in predicting peptide side-chain conformations[35], the authors demonstrated that different force-fields yield significantly different predictions for χ_1 side-chain dihedral angle distributions for virtually every amino acid **(Figure 5)**. In another study, Pande and colleagues used 524 experimentally-based metrics to assess the performance of different force-fields by running trajectories on the small protein ubiquitin. They found that different force-fields and different water models yield significantly different results[36] **(Figure 6)**. Hermans and colleagues compared the performance of different force-fields in reproducing the observed backbone dihedral angles of Ala and Gly[37]. They concluded that none of the current classical force-fields performed satisfactorily, and suggested that quantum mechanical (QM) effects must be included to properly predict backbone conformations. Currently, such an approach is impractical for protein design. Moreover, QM-based approaches require the initial input of knowledge-based potentials (such as CMAP) for them to reliably reproduce values observed in protein structures in a reasonable time[37,38].

Regan, O'Hern and colleagues have taken an alternative approach to computational protein design. They have shown that simple steric-based methods, as pioneered by Ramachandran and colleagues[39], predict the backbone and side-chain dihedral angle combinations observed in proteins better than more elaborate techniques **(Figure 7)**. They therefore advocate a 'back to basics' approach to protein structure analysis and design. Rather than including many different terms in the molecular mechanics force-field, they argue that only a minimal set of steric interactions and stereochemical constraints should be imposed to capture the defining features of protein structure. This method has proven very effective. Not only does it provide

predictions for the backbone and side-chain conformations that are observed in protein crystal structures and by NMR in solution[40-42], it also allows a mechanism to be proposed for transitions between different side-chain dihedral angle conformations[43].

Protein designs for nanotechnology

There is great potential to harness some of the defining properties of proteins for materials applications[44]. Proteins that can self-assemble into higher order structures are of particular interest and can be used to construct both amorphous materials and discrete assemblies. The unique features associated with protein-based materials make them attractive candidates for drug delivery and tissue engineering applications. Additionally, large, discrete protein structures that can be decorated at exact positions would facilitate several applications, including pathway engineering, sensors and vaccines.

Protein-based hydrogels confer a number of advantages over synthetic materials for biomedical applications: (1) the features required for 3-dimensional percolation and gelation are precisely encoded by the sequence, which specifies the structure; (2) genetic engineering to create virtually any desired sequence is relatively straightforward; (3) exquisite stimuli-responsiveness can readily be controlled by appropriately engineering the interactions between protein building blocks.

Olsen, Tirrell and colleagues designed hydrogels whose formation is based on the association of α-helices into coiled-coils[45]. Subsequently, Olsen and colleagues elaborated on these designs to create a thermosensitive hydrogel that is liquid at low temperatures (4°C) but which exhibits enhanced stiffness and durability at physiological temperature (37°)[46]. There are two key components to this design: the coiled-coil based shear thinning hydrogel midblock, and endblocks comprised of the thermoresponsive polymer poly(*N*-isopropylacrylamide) (PNIPAM) **(Figure 8)**. Such a shear thinning hydrogel that undergoes a transition to a more rigid, reinforced network at physiological temperatures could be used for injection, for example, in tissue repair.

Temperature is an attractive stimulus because it is straightforward to apply. The key is to engineer stimuli-responsiveness in a regime that is compatible with physiological temperature. Woolfson and colleagues designed hydrogels using α-helical peptides with thermosensitive properties encoded by the types of interactions between entangled helical fibrils[47]. The propensity for the hydrogel to become stronger or weaker after heat application is dependent on whether the fibril network is formed through hydrophobic (increase in strength with increasing temperature) or hydrogen bonding (decrease in strength with increasing temperature) interactions.

A major breakthrough in protein design that allows easy expression of branched protein building blocks was made by Howarth and colleagues. Protein expression is typically limited to linear topologies, but the development of SpyCatcher/SpyTag technology has changed that. SpyTag and SpyCatcher are peptide and protein components, respectively, that originated from splitting a fibronectin-binding domain of a bacterial adhesin. This fibronectin-binding domain spontaneously forms an intramolecular covalent bond in nature, and researchers were able to maintain this activity whilst dividing the protein into two separate parts[48] **(Figure 9, panel A)**. By

genetically encoding SpyTag and SpyCatcher in constructs of interest, diverse topologies can be created[49] **(Figure 9, panel B)**. Arnold, Tirrell and colleagues have exploited this technology for the construction of new protein-based materials[50]. A covalent hydrogel network was produced as a result of isopeptide bond formation by genetically encoding SpyTag and SpyCatcher into elastin-like protein (ELP) constructs **(Figure 10)**.

The development of "smart" protein-based hydrogels that can respond to various stimuli has unique potential for controlled substance release. Grove, Regan and colleagues demonstrated the reversible formation of self-assembling hydrogels using a protein-peptide binding interaction to encode the macroscopic properties of the gel[51,52]. Concatenated arrays of tetratricopeptide repeats (TPR) were mixed with multivalent 'star PEG' based arrays of cognate peptide to form non-covalently cross-linked gels **(Figure 11).** Because the binding interactions that form the cross-links are pH- and ionic strength-dependent, the gel dissolves and reforms in response to these stimuli. Moreover, the pH dependence of dissolution and cargo release is compatible with the decrease in pH associated with both the microenvironment of solid tumors[53] and the intracellular lysosomal pathway.

DNA origami has enabled the design of an impressive diversity of 2- and 3-dimensional structures[54]. Incorporating function has proven to be more difficult. By contrast, designing structures for 'protein origami' is more involved, but their functionalization is relatively straightforward. Jerala and colleagues took advantage of the specificity of association between coiled-coil building blocks to form a single-chain polypeptide structure that folds into a polyhedron[55]. The design employed six different pairs of coiled-coils **(Figure 12)**. A linker sequence was chosen that included residues that would enhance flexibility and disrupt helix formation - Ser-Gly-Pro-Gly. Another crucial component of the design is the orthogonality of the pairs - unintentional cross-reactivity between different coiled-coil monomers would prevent the proper assembly of the tetrahedron. The resulting 3-dimensional structure was imaged by atomic force microscopy (AFM), and the proximity of the N- and C-termini at the same vertex was confirmed by a split-fluorescent protein assay. Protein origami is attractive because such structures can be easily functionalized for use in pathway engineering, difference imaging, and novel vaccines.

Woolfson and colleagues created self-assembled cage-like particles (SAGEs) that form spheres of roughly 100 nm in diameter[56]. Non-covalent heterodimeric and heterotrimeric coiled-coils were employed as building blocks for the design, where different coils were connected by asymmetric disulfide bonds to form hubs that assemble into a hexagonal array upon mixing **(Figure 13, panel A)**. Interestingly, instead of forming the expected flat assembly based on the hexagonal design, the structures assembled into closed spheres **(Figure 13, panels B and C)**. Modeling suggested that the hubs are actually wedge-shaped instead of perfect tripods with arms angled at exactly 120°. The largest reported assembly with structural validation, a 24-subunit protein cube, was designed by Yeates and colleagues[57]. Their design strategy involved making fusions between natural dimeric and trimeric proteins. The particular proteins used were chosen so that the angle of the interface would satisfy the

requirements for cube formation when propagated. The trimeric protein 2-keto-3-deoxy-6-phosphogalactonate (KDPGal) aldolase and dimeric N-terminal domain of FkpA protein were connected by a flexible linker **(Figure 14)**. When mixed, they self-assembled into a porous cube with an outer diameter of 225 Å and an inner diameter of 132 Å, as determined by x-ray crystallography. The structure was additionally validated by negative stain electron microscopy and small angle x-ray scattering (SAXS) analysis. Like the SAGE particles *vide supra*, the large cavities in these protein assemblies have potential applications in delivery of molecular cargo.

Protein designs that function *in vivo*

To fully understand protein function in the cellular *milieu*, it is desirable to be able to study and manipulate protein activity in living cells. Since Green Fluorescent Protein (GFP) was first cloned two decades ago[58], fluorescent proteins have become a powerful and omnipresent tool in biology. The potential of GFP to study protein expression and localization was recognized early on, but several important limitations had to be overcome before it could be used routinely in biological applications. Protein design has played an important role in overcoming these obstacles and in extending the *in vivo* potential of fluorescent proteins.

Wild-type fluorescent proteins are often oligomers, a property that could interfere with the natural activity of a protein of interest. A common design strategy to prevent oligomerization is to introduce unfavorable electrostatic interactions to disrupt the subunit interfaces. This tactic was successfully used to create the monomeric mFruit series of fluorescent proteins[59], which was expanded recently with mPapaya[60], in addition to the monomeric version of the extremely bright green vivid *verde* fluorescent protein (mVFP)[61]. New fluorescent protein colors, which allow for more proteins to be tagged and followed in the same cell, have been created by randomly mutating GFP and screening for altered fluorescence emission characteristics[62] **(Figure 15)**. A similar method was also used to identify mutations that increase brightness and shift excitation peaks[63], allow GFP to fold faster[64], and introduce a number of additional properties useful for a wide range of applications.

Engineered versions of fluorescent proteins, such as ‘split GFP’ and ‘split dsRed’ have also been developed to study protein-protein interactions *in vivo*[65-67] **(Figure 16)**. As the name implies, these assays use versions of fluorescent proteins that have been split into N- and C-terminal halves. On their own, the two protein halves do not interact. Attaching proteins that bind to each other brings the two chains together, and the fluorescent protein is reconstituted. Libraries of potential protein-protein interacting partners can be screened by this assay. An advantage of the split GFP assay is that protein-binding partners can be examined in the context of their native cellular environment, whether it be *E. coli*, yeast, worm or mammalian cells.

In signaling networks a single protein may interact with multiple targets, and teasing apart the different activities can be a challenge. Several protein design strategies have been developed to address this issue. The simplest approach is to treat the binding and activating domains of a signaling protein as independent modules that can be mixed and matched. By fusing domains from different proteins, it is possible to

create novel networks that possess different input-output combinations. Park *et al*. transplanted new protein binding domains onto the MAP kinase scaffolding domain Ste5 to create a strain of yeast that responds to stimulation by mating hormone with an osmolarity response[68]. Similarly, Howard *et al.* created a novel adaptor protein by fusing protein binding domains from different signaling networks to produce a new strain of cells that underwent apoptosis in response to factors that normally trigger cell growth and survival[69]. These fusion proteins allow scientists to separate different activities of a signaling protein in order to systematically investigate each function.

Another approach to dissecting protein-protein interaction networks and pathways is to block the interaction of a protein with a particular binding partner. This can be accomplished without directly mutating the target proteins themselves by expressing a second protein that binds to a specific surface on the protein of interest. 'Native' antibodies are not appropriate for such intracellular applications because they contain disulfide bonds that are not stable in the reducing environment of the cell. As a result, other protein scaffolds that lack disulfide bonds have been used to create functional intracellular binding modules. Amstutz *et al.* used site-specific randomization in combination with an *in vitro* binding screen to isolate an ankyrin repeat protein that binds the kinase aminoglycoside phosphotransferase (APH) with high affinity. This ankyrin module binds and inhibits APH both *in vitro* and *in vivo*[70]. Cortajarena *et al.* randomized the substrate-binding residues of a TPR domain and sorted variants using a split GFP assay and FACS sorting in *E. coli* to isolate TPR variants that bind the human protein DSS1. Overexpressing the DSS1-binding modules in yeast phenocopied a Sem1 deletion mutant (Sem1 is the yeast homologue of Dss1). Because Sem1/DSS1 has been proposed to interact with a number of different partners, and its function in yeast is still unknown, having a TPR module that can inhibit a particular region of interaction is a powerful new tool[71].

An elegant example of a protein design approach to delineate protein function *in vivo* is seen in the work of Shokat and colleagues[72]. By designing a protein kinase with altered ATP-binding specificity they were able to identify substrates of that kinase. The strategy was to first introduce mutations in the kinase that enlarged the substrate-binding pocket such that bulky ATP derivatives could be bound. The binding pocket on the wild-type kinase is too small to bind the bulky ATP derivative. Thus, by supplying radioactively labeled ATP derivatives to cells expressing the mutant kinase, only substrates of that particular kinase would be labeled (**Figure 17)**.

A current goal in protein design is to create new methods to control protein activity *in vivo*. These methods would be a useful tool for studying processes that occur on fast timescales, as well as for rewiring existing pathways. Ideally, the stimulus would be fast, induce a high relative change in activity, and affect only the desired target. A creative strategy is to use small molecules to control protein localization. One way to achieve this is by attaching the ligand-binding domain of a nuclear hormone receptor to the protein of interest. Hormone receptors contain several conserved domains, among which are a ligand-binding domain (LBD) and a DNA-binding domain (DBD). In the absence of hormone, the LBD binds the chaperone Hsp90 and prevents entry of the receptor into the nucleus. Binding of hormone to the LBD induces a conformational

change that releases the receptor, allowing nuclear entry and binding of the DBD to its cognate sequence[73]. Hybrid receptors have been created that recognize a variety of hormones and DNA sequences[74,75]. Protein localization can also be controlled by taking advantage of proteins that interact only in the presence of a small molecule. Many of these strategies are based on rapamycin, a small molecule that can simultaneously bind the FK506 binding protein (FKBP) as well as the rapamycin binding domain of mTOR (FKBP-rapamycin binding domain or FRB) to form a ternary complex. By attaching either FRB or FKBP to the protein of interest and the other to a transmembrane protein, it is possible to induce protein localization to a specific subcellular compartment by adding rapamycin. This strategy can either be used to induce the activity of a protein whose substrate lies at the plasma membrane[76] or to sequester proteins away from cytosolic substrates[77].

Small molecule-based methods are effective, but they are restricted by molecular diffusion through the plasma membrane and cell walls. In principle, light would be an ideal stimulus because illumination can be rapidly switched on or off, resulting in essentially instantaneous addition or removal of signal. Moreover, it can reach any part of the cell, a property that is not necessarily true for small molecules. Most strategies modify natural plant photoreceptors to create light sensitive proteins. One popular choice is the light oxygen voltage domain from phototropin (LOV2), which consists of a core flavin mononucleotide-binding domain followed by a C-terminal Jα helix. Illumination with blue light triggers formation of a covalent bond between the excited flavin and a cysteine residue in the core domain, which induces a conformational rearrangement that results in unfolding of the Jα helix. Renicke *et al.* fused a short, synthetic, destabilizing domain from murine ornithine decarboxylase (cODC1) to LOV2 to create a photosensitive degron[78]. cODC1 is degraded through an ubiquitin independent mechanism, one of the requirements for which is exposure of a short unstructured region. Attaching cODC1 immediately after the Jα helix produced a protein that is only degraded when illuminated with blue light **(Figure 18)**. By varying the length of cODC1 and its attachment point to LOV2, Renicke *et al.* were ultimately able to isolate a module that rapidly and extensively reduced protein target concentrations upon illumination.

Strickland *et al.* also took advantage of the LOV2 domain to create TULIPs (tunable, light-controlled interacting protein tags)[79]. A short peptide recognized by Erbin PDZ (ePDZ, an engineered protein binding domain) was placed downstream of the Jα helix and designed so that a portion of the peptide would be incorporated in the α-helix under dark conditions. In the dark state, the peptide ligand is partially in a helical conformation, and binding to ePDZ is blocked. Illumination with blue light triggers Jα unfolding, and frees the peptide to interact with ePDZ. By fusing ePDZ to a transmembrane protein, it was possible to induce protein localization to the plasma membrane by illuminating cells with blue light. A number of peptide mutations were identified during the design process that varied in affinity for ePDZ. When combined with the ePDZ variants that bind peptides with different strengths, the authors were able to create a series of interaction pairs that covered a wide spectrum of binding affinities.

Future directions for protein design

We still do not have the theoretical or computational tools to design any protein structure or any protein-protein interaction interface on demand. Although different approaches have had some successes, it is not uncommon for a randomization plus screen or selection step to be required after the initial design to achieve the desired activity[23,24]. Currently, it is often easier to skip the design process entirely, and obtain the activity of interest solely by randomization and selection[71,80]. There are a number of factors underlying these struggles, the most significant of which are outlined below.

In computational protein design, a common approach is to add knowledge-based terms to existing force-fields (e.g. the cross-term energy correction map (CMAP) correction for the backbone dihedral angle distributions in CHARMM) so that they can recapitulate the experimentally observed backbone and side-chain dihedral angle distributions[33]**.** However, this strategy has disadvantages. For example the PDB includes many more α-helices than β-sheets, so the CMAP correction necessarily overemphasizes α-helical structures. To solve this issue, we need unbiased experimental measurements of intrinsic backbone and side-chain conformational propensities[81-84].

In a recent community-wide assessment of the current challenges in designing protein-protein interfaces, the inability to accurately model certain types of interactions, such as electrostatics and hydrogen bonding, was mentioned as a major limitation[85]. The issue of how to appropriately balance electrostatic and solvation effects has recently been discussed in detail[86]. A related question is whether or not quantum mechanical effects – such as the polarizability of electron clouds or solvation energies of compounds – are necessary to calibrate classical potentials.

How to best model the interplay between local steric interactions and backbone motion remains an unsolved problem. Consider the example of correctly balancing the energetics of introducing a Val residue versus an Ala residue at a certain position. Accommodating a Val necessitates a small movement of the backbone, but its side chain can interact favorably with a nearby hydrophobic patch. On the other hand, insertion of an Ala requires no backbone movement but abolishes the hydrophobic interaction. The correct balancing of such energetics continues to be a challenge. This scenario was encountered in state-of-the-art design work by Fleishman and colleagues[24], in which initial low-affinity designs for hemagglutinin-binding proteins were improved by randomization and selection. One of the mutations that resulted in a 25-fold increase in affinity was an Ala to Val mutation, in a situation similar to that described above. Other similar examples have been reported[12,87,88]**,** and several strategies are being developed to address this issue[87,89-93].

Correctly ranking computational designs remains a challenge for several reasons. First, when both physics- and knowledge-based terms are included in the force-field to generate and evaluate the designs, it is difficult to calibrate the relative strengths of the terms and determine the energy of the design in physical units. Thus, it would be

preferable to rank designs using several metrics rather than using the same force field that guided the design process. In addition, when using approaches that mix physics- and knowledge-based terms, it is difficult to ensure that all protein-protein and protein-water enthalpic contributions are properly accounted for and that the protein and solvent entropic contributions are included. Ideally, one would need to calculate the free energies of the designs to rank them. Privett *et al.* [23] addressed some of these issues by using all-atom MD simulations in addition to their standard design protocol **(Figure 3 left)**, to assess iterative designs of a Kemp eliminase, resulting in a functional enzyme after three rounds.

Despite these problems, there has been significant progress in developing computational techniques for predicting and designing protein structures *de novo*. This is in part because by examining 'static' protein structures, designers can delineate design goals in a relatively straightforward manner. By contrast, when designing functional proteins, the goal is by no means clear. Certain elements can be designed for, such as 'binding of the transition state' but the dynamics that accompany – and may be essential for - activity are not nearly so obvious. The connection between structure, dynamic protein-protein interactions and catalysis is not well understood.

The future of designed materials and assemblies lies in the creation of more diverse and robust protein building blocks. Orthogonality and specificity will be very important in creating new materials. For example, an expanded toolkit of coiled-coil interactions, in particular heterodimers, with minimal cross-reactivity would greatly benefit the construction of new self-assembling complexes[94-98]**.** Using building blocks that interact only with the desired ligands will allow for functionalization at discrete locations. The development of new types of specific protein-protein interaction modules with minimal cross-reactivity with each other or with cellular components would also be useful **(Speltz *et al.,* submitted)**. Additionally, new and improved "smart" materials will focus on enhanced responsiveness to molecular stimuli. By continuing to design protein building blocks that are more sensitive to stimuli such as pH, light, ionic strength, and temperature, we will expand the functionality of designed materials in the future.

For future *in vivo* applications, it would be useful to develop fluorescent proteins that are brighter and which mature faster than existing variants. In addition, proteins that emit at longer wavelengths would be beneficial to allow for deeper tissue penetration necessary for imaging in live multicellular organisms. Some progress towards this goal has been made with the discovery of several fluorescent proteins that fluoresce in the near IR region, although these variants suffer from low brightness[99,100], increased toxicity[101], and are limited by slow and/or incomplete fluorophore maturation[102]. Future efforts will also focus on developing new fluorescent protein-based biosensors that function *in vivo*[103], as well as identifying new variants that can be used for super-resolution microscopy[104]. Often when introducing new properties into fluorescent proteins, scientists select for a single attribute at the expense of other spectroscopic properties. For example, the mutations introduced to produce a monomeric VFP also decreased fluorescent brightness[61]. Future design efforts will also focus on improving the fluorescent properties of novel proteins. This will be achieved through iterative rounds of design, solving crystal structures of the new protein, and then using these

structures as a guide for further improvement. This approach has been used to improve the quantum yield of Cyan Fluorescent Protein (CFP) from 0.21 to most recently 0.93, the highest value to date for a monomeric protein[105]. Improved selection procedures, as well as past experience, will expedite this process.

To better understand signaling networks, tools will need to be developed that can distinguish between proteins with different post-translational modifications. Progress has been made with the design of protein domains that recognize phosphorylated substrates[106], and future efforts will be made easier by the recent creation of bacterial strains that can incorporate unnatural amino acids at specific positions[107].

New small molecule-inducible domains that respond to novel stimuli would also be useful. To create more intricate synthetic pathways it will be necessary to develop additional switches that can be used to control the activity of multiple proteins independently while minimizing interference with native cellular processes. It will also be useful to develop molecular switches that can respond to biologically relevant stimuli. These would greatly aid the creation of synthetic circuits for therapeutic, industrial, and detection applications. Another major focus will be to create new methods for controlling proteins with light. Light-sensitive domains will be developed that can control a wider range of protein activities, respond to light stimuli faster, and be induced by longer wavelengths than existing designs.

Efforts to design new proteins were first undertaken with the intent to increase our knowledge of structure and activity but also with the promise of creating new practical protein tools. Through the years, our basic understanding of proteins has increased greatly, and we have begun to enter the era when we will be able to produce functional proteins that will revolutionize medicine and technology.

We gratefully acknowledge the support of the Raymond and Beverly Sackler Institute for Biological, Physical, and Engineering Sciences; the National Science Foundation; the National Institutes of Health; and the Ford Foundation.

1. Murphy, G. S.; Sathyamoorthy, B.; Der, B. S.; Machius, M. C.; Pulavarti, S. V.; Szyperski, T.; Kuhlman, B. *Protein Sci* 2014.
2. Ho, S. P.; DeGrado, W. F. *Journal of the American Chemical Society* 1987, **109**, 6751-6758.
3. Regan, L.; DeGrado, W. F. *Science* 1988, **241**, 976-978.
4. Hecht, M. H.; Richardson, J. S.; Richardson, D. C.; Ogden, R. C. *Science* 1990, **249**, 884-891.
5. Padmanabhan, S.; Marqusee, S.; Ridgeway, T.; Laue, T. M.; Baldwin, R. L. *Nature* 1990, **344**, 268-270.
6. Kim, C. A.; Berg, J. M. *Nature* 1993, **362**, 267-270.
7. Smith, C. K.; Withka, J. M.; Regan, L. *Biochemistry* 1994, **33**, 5510-5517.
8. Minor, D. L., Jr.; Kim, P. S. *Nature* 1994, **367**, 660-663.
9. Smith, C. K.; Regan, L. *Science* 1995, **270**, 980-982.

10. Munson, M.; Balasubramanian, S.; Fleming, K. G.; Nagi, A. D.; O'Brien, R.; Sturtevant, J. M.; Regan, L. *Protein Science* 1996, **5**, 1584-1593.
11. Ponder, J. W.; Richards, F. M. *J Mol Biol* 1987, **193**, 775-791.
12. Lim, W. A.; Sauer, R. T. *Nature* 1989, **339**, 31-36.
13. Yue, K.; Dill, K. A. *Proceedings of the National Academy of Sciences* 1992, **89**, 4163-4167.
14. Desjarlais, J. R.; Handel, T. M. *Protein Sci* 1995, **4**, 2006-2018.
15. Dahiyat, B. I.; Mayo, S. L. *Protein Sci* 1996, **5**, 895-903.
16. Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. *Science* 2003, **302**, 1364-1368.
17. Butterfoss, G. L.; Kuhlman, B. *Annu Rev Biophys Biomol Struct* 2006, **35**, 49-65.
18. Regan, L.; Clarke, N. D. *Biochemistry* 1990, **29**, 10878-10883.
19. Regan, L. *Annu Rev Biophys Biomol Struct* 1993, **22**, 257-287.
20. 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. *Science* 2008, **319**, 1387-1391.
21. Rothlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D. *Nature* 2008, **453**, 190-195.
22. Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; St Clair, J. L.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L.; Houk, K. N.; Michael, F. E.; Baker, D. *Science* 2010, **329**, 309-313.
23. Privett, H. K.; Kiss, G.; Lee, T. M.; Blomberg, R.; Chica, R. A.; Thomas, L. M.; Hilvert, D.; Houk, K. N.; Mayo, S. L. *Proc Natl Acad Sci U S A* 2012, **109**, 3790-3795.
24. Fleishman, S. J.; Whitehead, T. A.; Ekiert, D. C.; Dreyfus, C.; Corn, J. E.; Strauch, E. M.; Wilson, I. A.; Baker, D. *Science* 2011, **332**, 816-821.
25. Korendovych, I. V.; Kulp, D. W.; Wu, Y.; Cheng, H.; Roder, H.; DeGrado, W. F. *Proc Natl Acad Sci U S A* 2011, **108**, 6823-6827.
26. Yin, H.; Slusky, J. S.; Berger, B. W.; Walters, R. S.; Vilaire, G.; Litvinov, R. I.; Lear, J. D.; Caputo, G. A.; Bennett, J. S.; DeGrado, W. F. *Science* 2007, **315**, 1817-1822.
27. Shandler, S. J.; Korendovych, I. V.; Moore, D. T.; Smith-Dupont, K. B.; Streu, C. N.; Litvinov, R. I.; Billings, P. C.; Gai, F.; Bennett, J. S.; DeGrado, W. F. *J Am Chem Soc* 2011, **133**, 12378-12381.
28. Grigoryan, G.; Kim, Y. H.; Acharya, R.; Axelrod, K.; Jain, R. M.; Willis, L.; Drndic, M.; Kikkawa, J. M.; DeGrado, W. F. *Science* 2011, **332**, 1071-1076.
29. Lanci, C. J.; MacDermaid, C. M.; Kang, S. G.; Acharya, R.; North, B.; Yang, X.; Qiu, X. J.; DeGrado, W. F.; Saven, J. G. *Proc Natl Acad Sci U S A* 2012, **109**, 7304-7309.
30. Jorgensen, W. L.; Tirado-Rives, J. *Journal of the American Chemical Society* 1988, **110**, 1657-1666.
31. 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.; 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. *J Phys Chem B* 1998, **102**, 3586-3616.
32. Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; III, T. E. C.; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.;

Janowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A.; University of California, San Francisco, 2014.
33. MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., 3rd. *J Am Chem Soc* 2004, **126**, 698-699.
34. Kryshtafovych, A.; Moult, J.; Bales, P.; Bazan, J. F.; Biasini, M.; Burgin, A.; Chen, C.; Cochran, F. V.; Craig, T. K.; Das, R.; Fass, D.; Garcia-Doval, C.; Herzberg, O.; Lorimer, D.; Luecke, H.; Ma, X.; Nelson, D. C.; van Raaij, M. J.; Rohwer, F.; Segall, A.; Seguritan, V.; Zeth, K.; Schwede, T. *Proteins: Structure, Function, and Bioinformatics* 2014, **82**, 26-42.
35. Vymětal, J.; Vondrášek, J. *J Chem Theory Comput* 2012, **9**, 441-451.
36. Beauchamp, K. A.; Lin, Y. S.; Das, R.; Pande, V. S. *J Chem Theory Comput* 2012, **8**, 1409-1414.
37. Hu, H.; Elstner, M.; Hermans, J. *Proteins* 2003, **50**, 451-463.
38. Zhu, X.; Lopes, P. E. M.; Shim, J.; MacKerell, A. D. *Journal of Chemical Information and Modeling* 2012, **52**, 1559-1572.
39. Ramakrishnan, C.; Ramachandran, G. N. *Biophys J* 1965, **5**, 909-933.
40. Zhou, A. Q.; O'Hern, C. S.; Regan, L. *Biophys J* 2012, **102**, 2345-2352.
41. Zhou, A. Q.; Caballero, D.; O'Hern, C. S.; Regan, L. *Biophys J* 2013, **105**, 2403-2411.
42. Zhou, A. Q.; O'Hern, C. S.; Regan, L. *Proteins* 2014, **82**, 2574-2584.
43. Caballero, D.; Maatta, J.; Zhou, A. Q.; Sammalkorpi, M.; O'Hern, C. S.; Regan, L. *Protein Sci* 2014, **23**, 970-980.
44. Grove, T. Z.; Cortajarena, A. L.; Regan, L. *Curr Opin Struct Biol* 2008, **18**, 507-515.
45. Olsen, B. D.; Kornfield, J. A.; Tirrell, D. A. *Macromolecules* 2010, **43**, 9094-9099.
46. Glassman, M. J.; Chan, J.; Olsen, B. D. *Advanced Functional Materials* 2013, **23**, 11.
47. Banwell, E. F.; Abelardo, E. S.; Adams, D. J.; Birchall, M. A.; Corrigan, A.; Donald, A. M.; Kirkland, M.; Serpell, L. C.; Butler, M. F.; Woolfson, D. N. *Nat Mater* 2009, **8**, 596-600.
48. Zakeri, B.; Fierer, J. O.; Celik, E.; Chittock, E. C.; Schwarz-Linek, U.; Moy, V. T.; Howarth, M. *Proc Natl Acad Sci U S A* 2012, **109**, E690-697.
49. Zhang, W. B.; Sun, F.; Tirrell, D. A.; Arnold, F. H. *J Am Chem Soc* 2013, **135**, 13988-13997.
50. Sun, F.; Zhang, W. B.; Mahdavi, A.; Arnold, F. H.; Tirrell, D. A. *Proc Natl Acad Sci U S A* 2014, **111**, 11269-11274.
51. Grove, T. Z.; Osuji, C. O.; Forster, J. D.; Dufresne, E. R.; Regan, L. *J Am Chem Soc* 2010, **132**, 14024-14026.
52. Grove, T. Z.; Forster, J.; Pimienta, G.; Dufresne, E.; Regan, L. *Biopolymers* 2012, **97**, 508-517.
53. Warburg, O. *Science* 1956, **123**, 309-314.
54. Sacca, B.; Niemeyer, C. M. *Angew Chem Int Ed Engl* 2012, **51**, 58-66.
55. Gradisar, H.; Bozic, S.; Doles, T.; Vengust, D.; Hafner-Bratkovic, I.; Mertelj, A.; Webb, B.; Sali, A.; Klavzar, S.; Jerala, R. *Nat Chem Biol* 2013, **9**, 362-366.

56. Fletcher, J. M.; Harniman, R. L.; Barnes, F. R.; Boyle, A. L.; Collins, A.; Mantell, J.; Sharp, T. H.; Antognozzi, M.; Booth, P. J.; Linden, N.; Miles, M. J.; Sessions, R. B.; Verkade, P.; Woolfson, D. N. *Science* 2013, **340**, 595-599.
57. Lai, Y. T.; Reading, E.; Hura, G. L.; Tsai, K. L.; Laganowsky, A.; Asturias, F. J.; Tainer, J. A.; Robinson, C. V.; Yeates, T. O. *Nat Chem* 2014, **6**, 1065-1071.
58. Prasher, D. C.; Eckenrode, V. K.; Ward, W. W.; Prendergast, F. G.; Cormier, M. J. *Gene* 1992, **111**, 229-233.
59. Shaner, N. C.; Campbell, R. E.; Steinbach, P. A.; Giepmans, B. N.; Palmer, A. E.; Tsien, R. Y. *Nature biotechnology* 2004, **22**, 1567-1572.
60. Hoi, H.; Howe, E. S.; Ding, Y.; Zhang, W.; Baird, M. A.; Sell, B. R.; Allen, J. R.; Davidson, M. W.; Campbell, R. E. *Chemistry & biology* 2013, **20**, 1296-1304.
61. Ilagan, R. P.; Rhoades, E.; Gruber, D. F.; Kao, H. T.; Pieribone, V. A.; Regan, L. *Febs Journal* 2010, **277**, 1967-1978.
62. Heim, R.; Prasher, D. C.; Tsien, R. Y. *Proceedings of the National Academy of Sciences* 1994, **91**, 12501-12504.
63. Heim, R.; Cubitt, A. B.; Tsien, R. Y. 1995.
64. Pédelacq, J.-D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. *Nature biotechnology* 2005, **24**, 79-88.
65. Ghosh, I.; Hamilton, A. D.; Regan, L. *Journal of the American Chemical Society* 2000, **122**, 5658-5659.
66. Hu, C.-D.; Chinenov, Y.; Kerppola, T. K. *Molecular cell* 2002, **9**, 789-798.
67. Magliery, T. J.; Wilson, C. G.; Pan, W.; Mishler, D.; Ghosh, I.; Hamilton, A. D.; Regan, L. *Journal of the American Chemical Society* 2005, **127**, 146-157.
68. Park, S.-H.; Zarrinpar, A.; Lim, W. A. *Science* 2003, **299**, 1061-1064.
69. Howard, P. L.; Chia, M. C.; Del Rizzo, S.; Liu, F.-F.; Pawson, T. *Proceedings of the National Academy of Sciences* 2003, **100**, 11267-11272.
70. Amstutz, P.; Binz, H. K.; Parizek, P.; Stumpp, M. T.; Kohl, A.; Grütter, M. G.; Forrer, P.; Plückthun, A. *Journal of Biological Chemistry* 2005, **280**, 24715-24722.
71. Cortajarena, A. L.; Liu, T. Y.; Hochstrasser, M.; Regan, L. *ACS chemical biology* 2010, **5**, 545-552.
72. Shah, K.; Liu, Y.; Deirmengian, C.; Shokat, K. M. *Proceedings of the National Academy of Sciences* 1997, **94**, 3565-3570.
73. Aranda, A.; Pascual, A. *Physiological reviews* 2001, **81**, 1269-1304.
74. Godowski, P. J.; Picard, D.; Yamamoto, K. R. *Science* 1988, **241**, 812-816.
75. Picard, D. *Methods in enzymology* 1999, **327**, 385-401.
76. Spencer, D. M.; Graef, I.; Austin, D. J.; Schreiber, S. L.; Crabtree, G. R. *Proceedings of the National Academy of Sciences* 1995, **92**, 9805-9809.
77. Robinson, M. S.; Sahlender, D. A.; Foster, S. D. *Developmental cell* 2010, **18**, 324-331.
78. Renicke, C.; Schuster, D.; Usherenko, S.; Essen, L.-O.; Taxis, C. *Chemistry & biology* 2013, **20**, 619-626.
79. Strickland, D.; Lin, Y.; Wagner, E.; Hope, C. M.; Zayner, J.; Antoniou, C.; Sosnick, T. R.; Weiss, E. L.; Glotzer, M. *Nature methods* 2012, **9**, 379-384.
80. Dreier, B.; Pluckthun, A. *Methods Mol Biol* 2012, **805**, 261-286.
81. Butterfoss, G. L.; Richardson, J. S.; Hermans, J. *Acta Crystallogr D Biol Crystallogr* 2005, **61**, 88-98.

82. Hagarman, A.; Measey, T. J.; Mathieu, D.; Schwalbe, H.; Schweitzer-Stenner, R. *J Am Chem Soc* 2010, **132**, 540-551.
83. Hansen, D. F.; Neudecker, P.; Kay, L. E. *Journal of the American Chemical Society* 2010, **132**, 7589-7591.
84. Hansen, D. F.; Neudecker, P.; Vallurupalli, P.; Mulder, F. A.; Kay, L. E. *Journal of the American Chemical Society* 2010, **132**, 42-43.
85. Fleishman, S. J.; Whitehead, T. A.; Strauch, E. M.; Corn, J. E.; Qin, S.; Zhou, H. X.; Mitchell, J. C.; Demerdash, O. N.; Takeda-Shitaka, M.; Terashi, G.; Moal, I. H.; Li, X.; Bates, P. A.; Zacharias, M.; Park, H.; Ko, J. S.; Lee, H.; Seok, C.; Bourquard, T.; Bernauer, J.; Poupon, A.; Aze, J.; Soner, S.; Ovali, S. K.; Ozbek, P.; Tal, N. B.; Haliloglu, T.; Hwang, H.; Vreven, T.; Pierce, B. G.; Weng, Z.; Perez-Cano, L.; Pons, C.; Fernandez-Recio, J.; Jiang, F.; Yang, F.; Gong, X.; Cao, L.; Xu, X.; Liu, B.; Wang, P.; Li, C.; Wang, C.; Robert, C. H.; Guharoy, M.; Liu, S.; Huang, Y.; Li, L.; Guo, D.; Chen, Y.; Xiao, Y.; London, N.; Itzhaki, Z.; Schueler-Furman, O.; Inbar, Y.; Potapov, V.; Cohen, M.; Schreiber, G.; Tsuchiya, Y.; Kanamori, E.; Standley, D. M.; Nakamura, H.; Kinoshita, K.; Driggers, C. M.; Hall, R. G.; Morgan, J. L.; Hsu, V. L.; Zhan, J.; Yang, Y.; Zhou, Y.; Kastritis, P. L.; Bonvin, A. M.; Zhang, W.; Camacho, C. J.; Kilambi, K. P.; Sircar, A.; Gray, J. J.; Ohue, M.; Uchikoga, N.; Matsuzaki, Y.; Ishida, T.; Akiyama, Y.; Khashan, R.; Bush, S.; Fouches, D.; Tropsha, A.; Esquivel-Rodriguez, J.; Kihara, D.; Stranges, P. B.; Jacak, R.; Kuhlman, B.; Huang, S. Y.; Zou, X.; Wodak, S. J.; Janin, J.; Baker, D. *J Mol Biol* 2011, **414**, 289-302.
86. Stranges, P. B.; Kuhlman, B. *Protein Sci* 2013, **22**, 74-82.
87. Davis, I. W.; Arendall, W. B., 3rd; Richardson, D. C.; Richardson, J. S. *Structure* 2006, **14**, 265-274.
88. Grove, T. Z.; Hands, M.; Regan, L. *Protein Eng Des Sel* 2010, **23**, 449-455.
89. Su, A.; Mayo, S. L. *Protein Sci* 1997, **6**, 1701-1707.
90. Farinas, E.; Regan, L. *Protein Science* 1998, **7**, 1939-1946.
91. Ross, S. A.; Sarisky, C. A.; Su, A.; Mayo, S. L. *Protein Sci* 2001, **10**, 450-454.
92. Friedland, G. D.; Linares, A. J.; Smith, C. A.; Kortemme, T. *J Mol Biol* 2008, **380**, 757-774.
93. Georgiev, I.; Keedy, D.; Richardson, J. S.; Richardson, D. C.; Donald, B. R. *Bioinformatics* 2008, **24**, i196-204.
94. O'Shea, E. K.; Lumb, K. J.; Kim, P. S. *Curr Biol* 1993, **3**, 658-667.
95. Gurnon, D. G.; Whitaker, J. A.; Oakley, M. G. *J Am Chem Soc* 2003, **125**, 7518-7519.
96. Woolfson, D. N. *Adv Protein Chem* 2005, **70**, 79-112.
97. Grigoryan, G.; Reinke, A. W.; Keating, A. E. *Nature* 2009, **458**, 859-864.
98. Thompson, K. E.; Bashor, C. J.; Lim, W. A.; Keating, A. E. *ACS Synth Biol* 2012, **1**, 118-129.
99. Shcherbo, D.; Shemiakina, I. I.; Ryabova, A. V.; Luker, K. E.; Schmidt, B. T.; Souslova, E. A.; Gorodnicheva, T. V.; Strukova, L.; Shidlovskiy, K. M.; Britanova, O. V. *Nature methods* 2010, **7**, 827-829.
100. Filonov, G. S.; Piatkevich, K. D.; Ting, L.-M.; Zhang, J.; Kim, K.; Verkhusha, V. V. *Nature biotechnology* 2011, **29**, 757-761.
101. Davidson, M. W.; Campbell, R. E. *nature methods* 2009, **6**, 713-717.
102. Subach, F. V.; Verkhusha, V. V. *Chemical reviews* 2012, **112**, 4308-4327.

103. Ibraheem, A.; Campbell, R. E. *Current opinion in chemical biology* 2010, **14**, 30-36.
104. Nienhaus, K.; Nienhaus, G. U. *Chemical Society Reviews* 2014, **43**, 1088-1106.
105. Goedhart, J.; von Stetten, D.; Noirclerc-Savoye, M.; Lelimousin, M.; Joosen, L.; Hink, M. A.; van Weeren, L.; Gadella Jr, T. W.; Royant, A. *Nature communications* 2012, **3**, 751.
106. Sawyer, N.; Gassaway, B. M.; Haimovich, A. D.; Isaacs, F. J.; Rinehart, J.; Regan, L. *ACS chemical biology* 2014, **9**, 2502-2507.
107. Park, H. S.; Hohn, M. J.; Umehara, T.; Guo, L. T.; Osborne, E. M.; Benner, J.; Noren, C. J.; Rinehart, J.; Soll, D. *Science* 2011, **333**, 1151-1154.

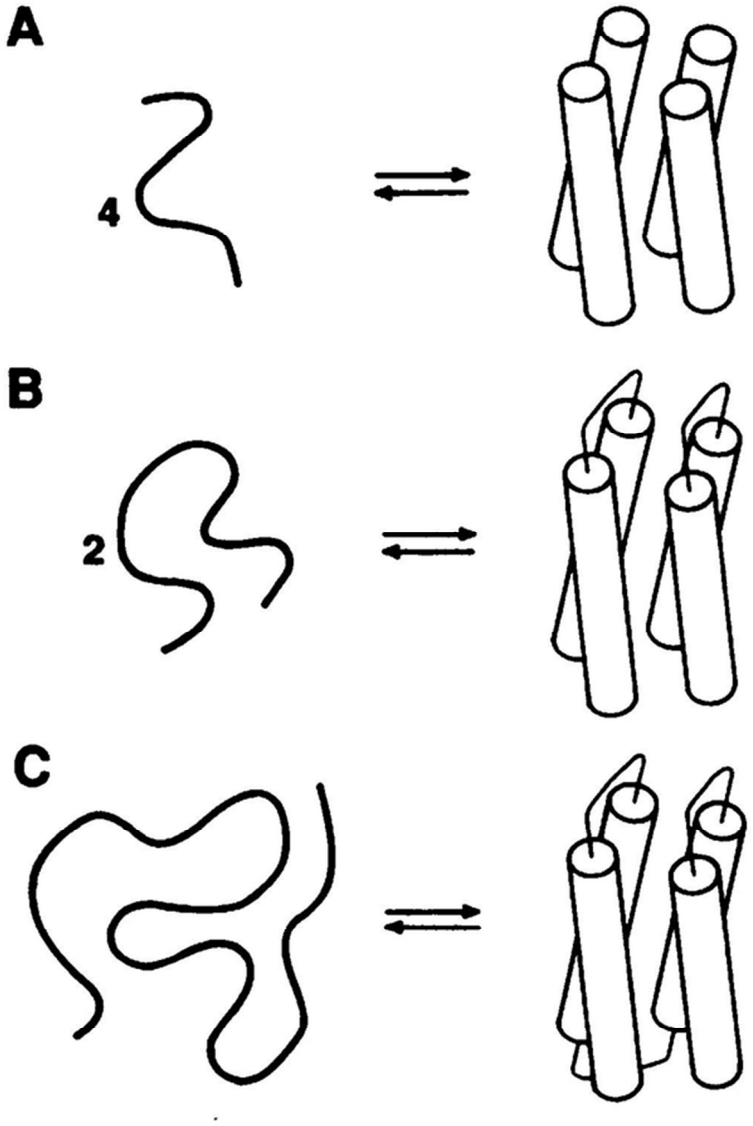

Figure 1: Schematic illustration of a systematic, minimal approach to the design of the four-helix bundle protein.

From Regan, L.; DeGrado, W. F. Science 1988, 241, 976-978. Reprinted with permission from AAAS.

67x79mm (300 x 300 DPI)

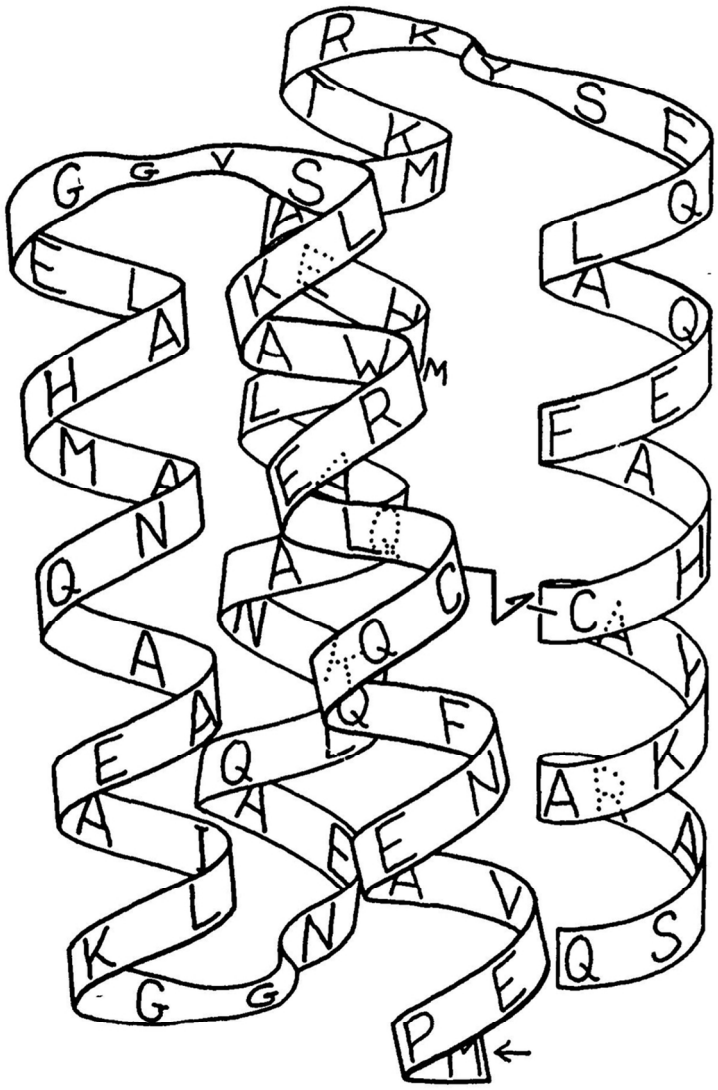

Figure 2: Schematic illustration showing both the sequence and proposed three-dimensional structure of the designed four-helix bundle protein, Felix.

From Hecht, M. H.; Richardson, J. S.; Richardson, D. C.; Ogden, R. C. Science 1990, 249, 884-891. Reprinted with permission from AAAS.

86x129mm (300 x 300 DPI)

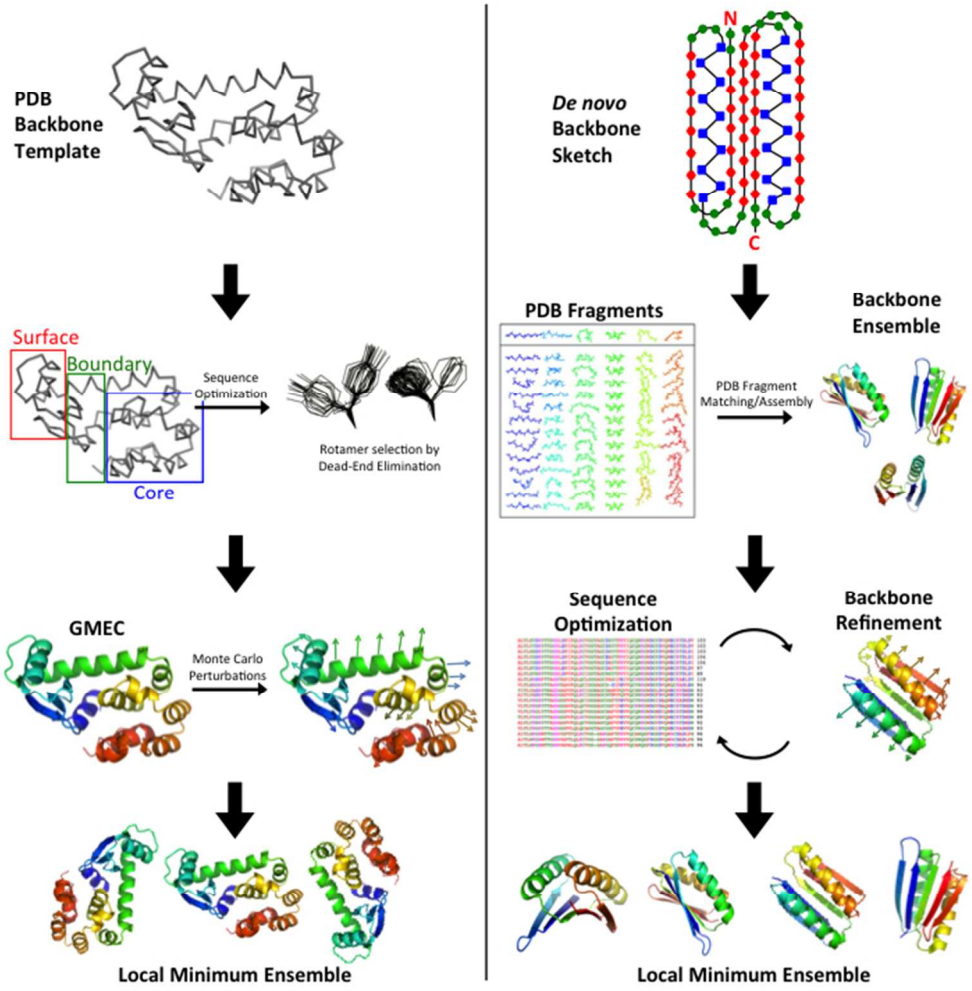

Figure 3: Comparison of ORBIT and RosettaDesign design strategies. Left: In ORBIT a backbone template with known coordinates is chosen. Amino acid positions are classified into three categories: core, boundary, and surface. Dead-End Elimination is implemented to reduce the combinatorial search of energetically allowed rotamers and to obtain the global minimum energy conformation (GMEC). Lastly, Monte Carlo is used to randomly change rotamers of the GMEC sequence to sample local low-energy configurations. Right: In RosettaDesign an ensemble of de novo backbones is generated using peptide fragments that match the desired backbone. This ensemble is subjected to iterative rounds of fixed backbone sequence optimization and flexible backbone energy minimization. Amino acid sequences are optimized by sampling different residues and rotamers with a Monte Carlo search protocol. Backbones are optimized by perturbing main-chain torsion angles, cycling through rotamers for side-chains with increased energies, and minimizing backbone energy at insertion sites according to the Metropolis criterion.

254x254mm (72 x 72 DPI)

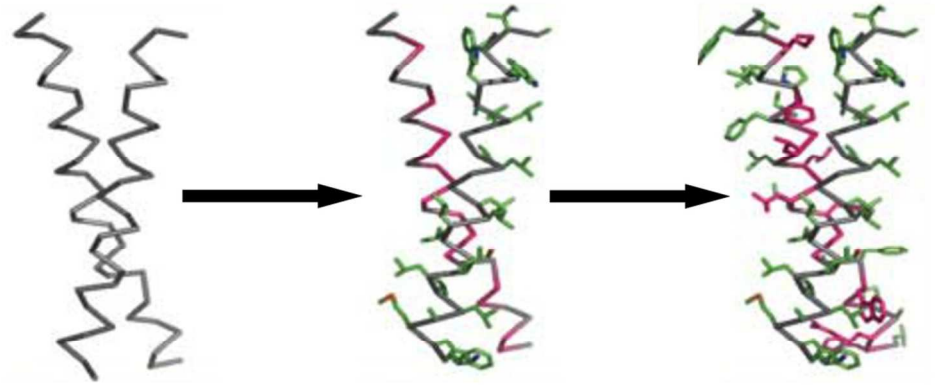

Figure 4: The CHAMP design process. A transmembrane helix-helix backbone pair with the desired geometry is selected from a database of membrane-protein structures (left). The original amino acid side-chains are discarded, and the helical backbones are extended to span the full length of a membrane. Next, the sequence from the target TM protein is 'threaded' onto either one of the helices – in this example, the right helix with green side-chains (middle). Lastly, the sequence and side-chain configurations for the anti-TM peptide (represented by the left helix) is chosen via iteration over amino acids and rotamer re-packing (right).

From Yin, H.; Slusky, J. S.; Berger, B. W.; Walters, R. S.; Vilaire, G; Litvinov, R.I.; Lear, J. D.; Caputo, G. A.; Bennett, J. S.; DeGrado, W. F. Science 2007, 315(5820), 1817-1822. Reprinted with permission from AAAS.

95x43mm (300 x 300 DPI)

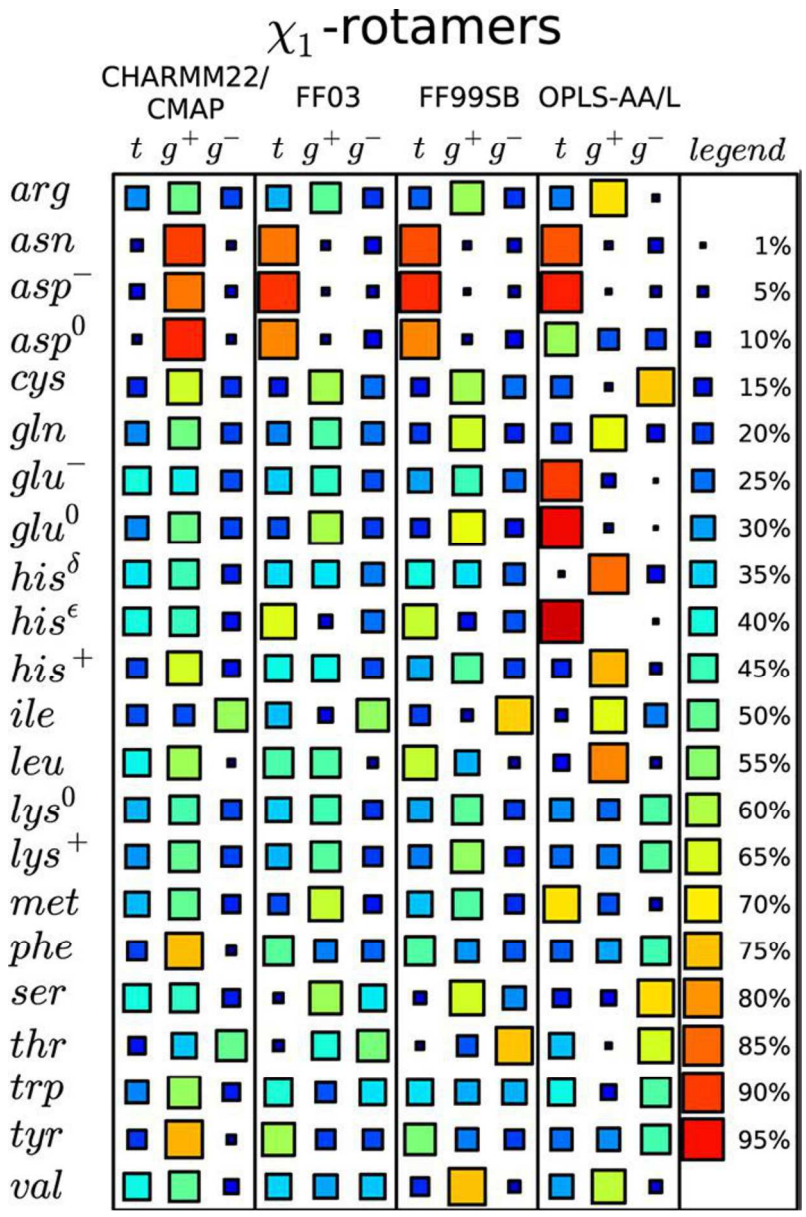

Figure 5: Comparing the performance of different force fields in predicting side-chain dihedral angle distributions. Comparison of the relative populations of χ1 dihedral angles (where t = 120º < χ1 < 240º, g+ = 240º < χ1 < 360º, and g- = 0º < χ1 < 120º) for different amino acids resulting from simulations of short peptides in solution, using different force fields (CHARMM22/CMAP; FF03; FF99SB; OPLS-AA/L). The relative occupancy of each side-chain dihedral is indicated by the size and color of the associated square (see legend). Each row shows the predictions for that particular amino acid type given by the specified force field. A good 'consistent' prediction would be if each force field gave the same result. Lys is quite good. By contrast, Val is particularly bad since its results differ with each force field.

299x452mm (300 x 300 DPI)

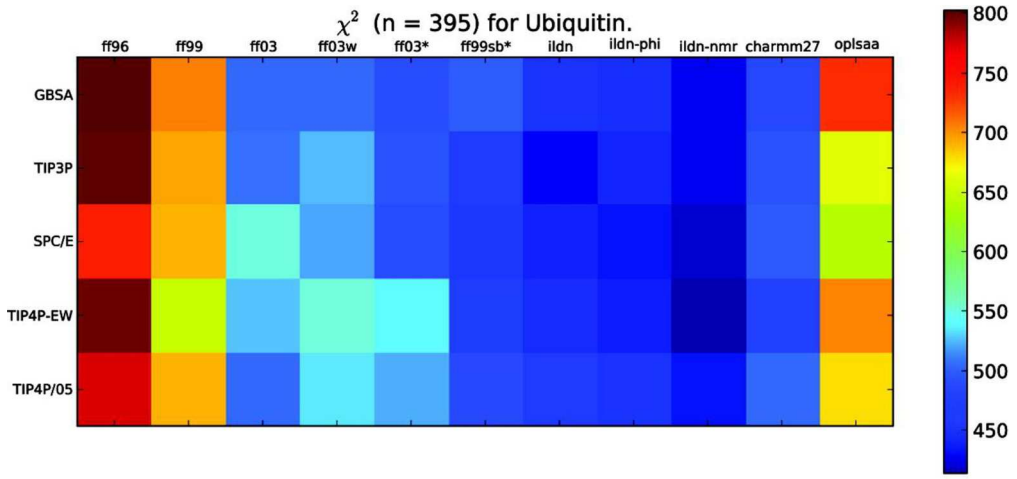

Figure 6: Comparison of force field performance in simulations of the 78 amino acid protein, ubiquitin. 524 different parameters were used in the assessment of the score. Each column corresponds to a given force field (as indicted) and each row corresponds to a different model for explicit solvent (as indicated). For each combination of force field and water model, χ 2 quantifies the agreement between simulation and experiment based on the 72 parameters, indicated by the color of the square. A smaller value of χ2 (darker blue) indicates a greater agreement between prediction and experiment. Note the differences between the different force fields with the same water models, and between different water models with the same force field.

173x85mm (300 x 300 DPI)

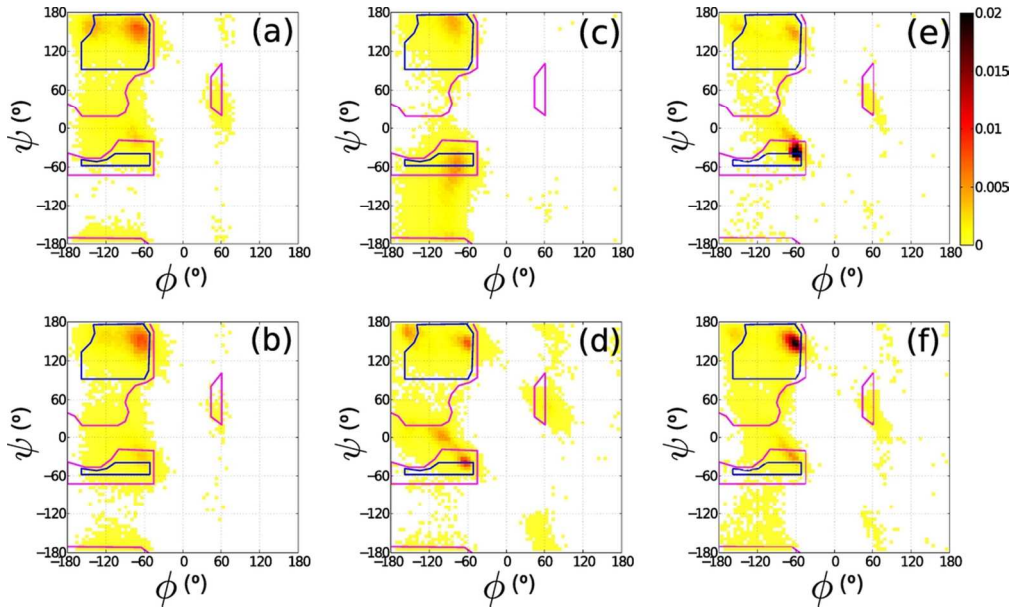

Figure 7: Comparison between different probability distributions P(ψ,ϕ) for the backbone dihedral angles φ and ψ obtained from molecular dynamics simulations of an Ala dipeptide mimetic using different versions of the CHARMM and Amber force fields, their associated optimized water models, and with and without the "ILDN-NMR" and "CMAP" dihedral angle potential corrections. Outlined are the Ramachandran hard sphere limits for 110º. (a) Amber99sb + TIP4P-Ew, (b) Amber99sb-ILDN-NMR + TIP4P-Ew, (c) CHARMM27 + TIP3SP, and (d) CHARMM27-CMAP + TIP3SP. Panels (e) and (f) correspond to the Alanine phi/psi distributions different subsets of the PDB. Note the different predictions in panels (a)-(d) compared with the experimental values in panels (e),(f).

Reprinted with permission from Caballero, D.; Määttä, J.; Zhou, A. Q.; Sammalkorpi, M.; O'Hern, C. S.; Regan, L. Protein Science 2014, 23, 970-980.

98x59mm (300 x 300 DPI)

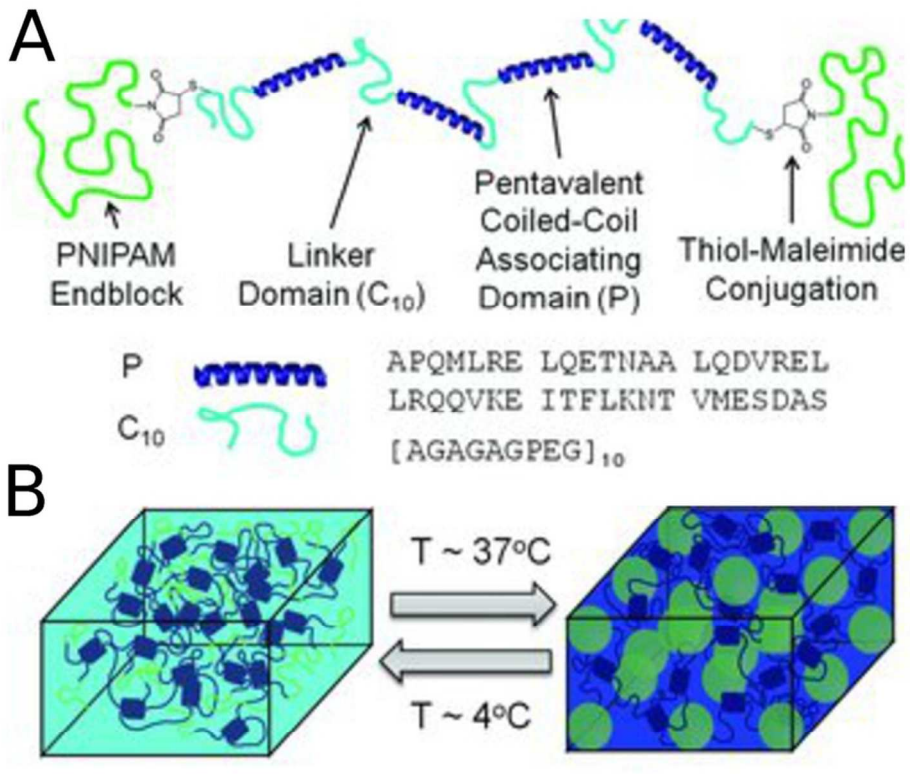

Figure 8: Creation of a temperature-responsive hydrogel, based on a 'dual network' design. (a) Illustration of the components of the dual polymer design comprising PNIPAM ends (green), helical coiled coils (dark blue), and linker regions (pale blue). (b) Schematic representation of the temperature-dependent reinforcement of shear-thinning hydrogels by the PNIPAM triblock copolymer domains. At 4°C, the coiled coils (dark blue) fold and associate, while the PNIPAM domains (green lines) do not interact. However, at 37o C the PNIPAM blocks associate (green spheres) and reinforce the hydrogel network.

Reprinted with permission from Glassman, M. J.; Chan, J.; Olsen, B. D. Advanced Functional Materials. 2013, 23, 1182–1193

83x70mm (300 x 300 DPI)

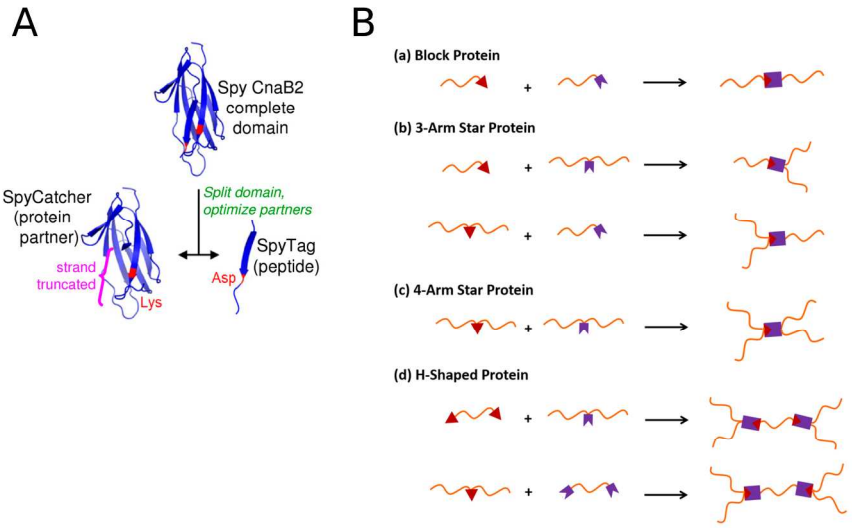

Figure 9: (A) Ribbon representation of the SpyTag-SpyCatcher complex assembly.The Lys residue (red) on the 12 kDa SpyCatcher protein spontaneously forms an isopeptide bond with the Asp (red) on the 13-reside SpyTag peptide. (B) Schematic illustration of the diverse protein topologies possible by the SpyTag (red triangle)/SpyCatcher (purple crown) technology.

Reprinted with permission from (A) Zakeri, B.; Fierer, J. O.; Celik, E.; Chittock, E. C.; Schwarz-Linek, U.; Moy, V. T.; Howarth, M. Proc Natl Acad Sci USA 2012, 109, E690-697, and (B) Zhang, W. B.; Sun, F.; Tirrell, D. A.; Arnold, F. H. Journal of the American Chemical Society 2013, 135, 13988-13997. Copyright 2013 American Chemical Society.

215x279mm (300 x 300 DPI)

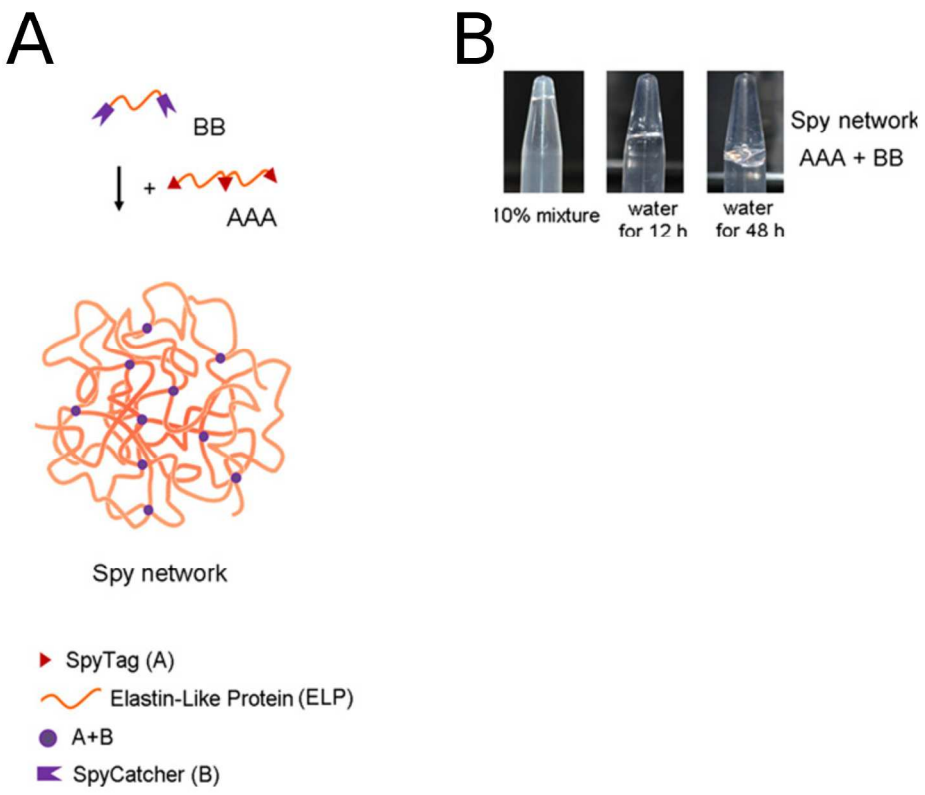

Figure 10: (A) Cartoon of covalent hydrogel formation using SpyTag/SpyCatcher technology. Three SpyTag sequences (red triangles) are connected by an elastin-like-protein (ELP) sequence (orange strand) to make the AAA construct. Two SpyCatcher units (purple crowns) are joined by an ELP linker to form the BB construct. Mixing these two proteins results in a covalent Spy network. (B) A photograph of the formed covalent Spy network. Mixing 10% wt aqueous solutions of AAA and BB in equimolar amounts of binding sites yields the hydrogel shown. Upon addition of water, the Spy network swells by 3,000% after 12 hours and continues to be swollen after 48 hours.
FROM PNAS: Sun, F.; Zhang, W. B.; Mahdavi, A.; Arnold, F. H.; Tirrell, D. A. Proc Natl Acad Sci USA 2014, 111, 11269-11274.

215x279mm (300 x 300 DPI)

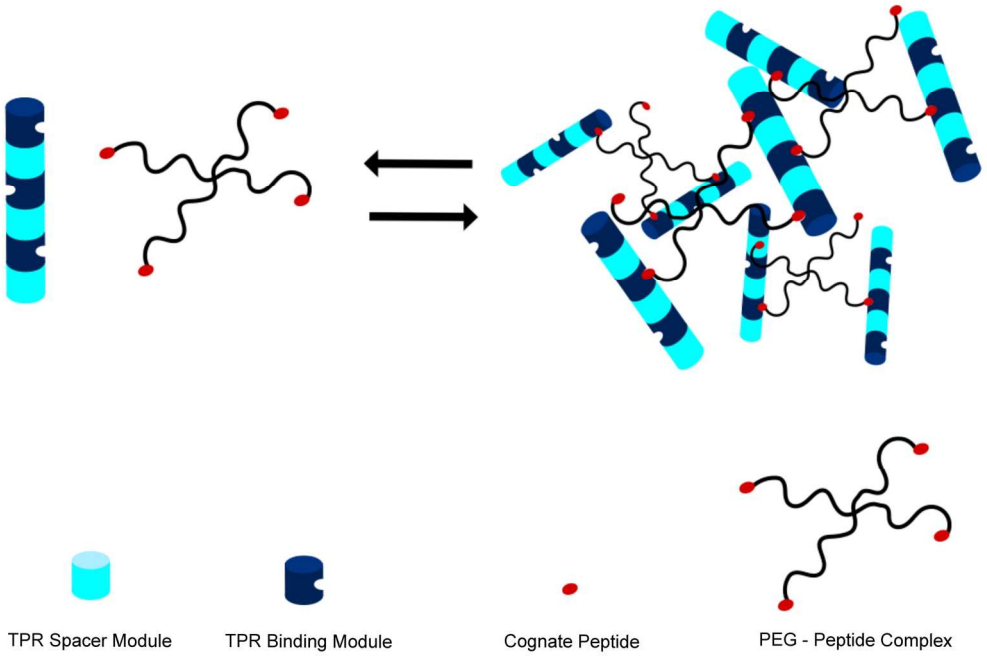

Figure 11: A schematic illustration of the reversible formation of a TPR-peptide based hydrogel. Consensus TPR "binding" modules (dark blue) that bind to the cognate peptide are concatenated with TPR "spacer" modules (pale blue) that do not bind the peptide so that binding sites are arrayed on different faces of the cylinder. Peptide cross-linkers were constructed by chemical attachment of the cognate peptide to functionalized 4-armed star PEG molecules (black lines with red termini). Mixing the TPR arrays with PEG-peptide cross-linkers in a stoichiometric ratio of 1:2 results in hydrogel formation, which can be reversed by increasing ionic strength or decreasing pH.

183x120mm (300 x 300 DPI)

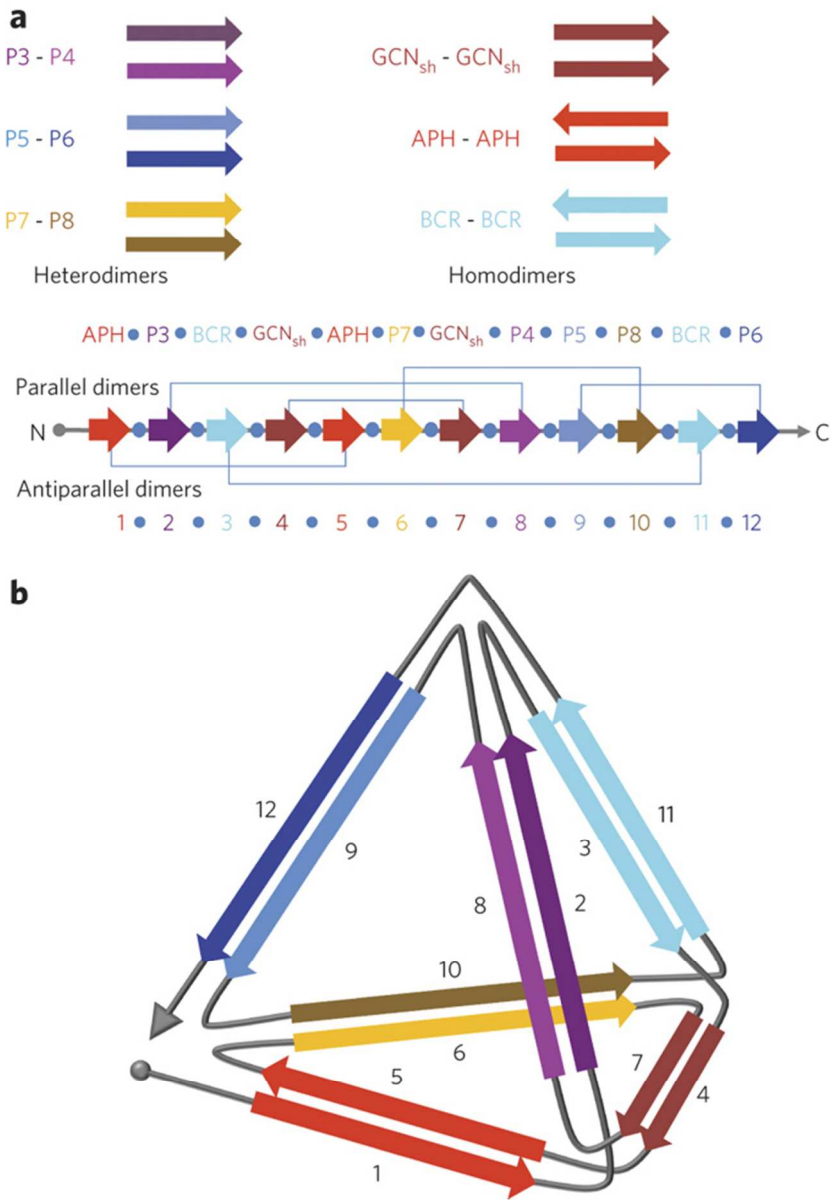

Figure 12: Design of a tetrahedron/trigonal pyramid using coiled coil assembly. (a) Cartoon illustrating the pyramid components – sets of heterodimeric and homodimeric parallel and antiparallel coiled coils. The twelve individual peptide sequences are concatenated in the indicated order, with each sequence separated by the flexible linker Ser-Gly-Pro-Gly. Grey lines indicate the interacting pairs. (b) Schematic of the desired tetrahedron structure. Arrows indicate the direction of the helices in the coiled-coil pairs.

238x346mm (72 x 72 DPI)

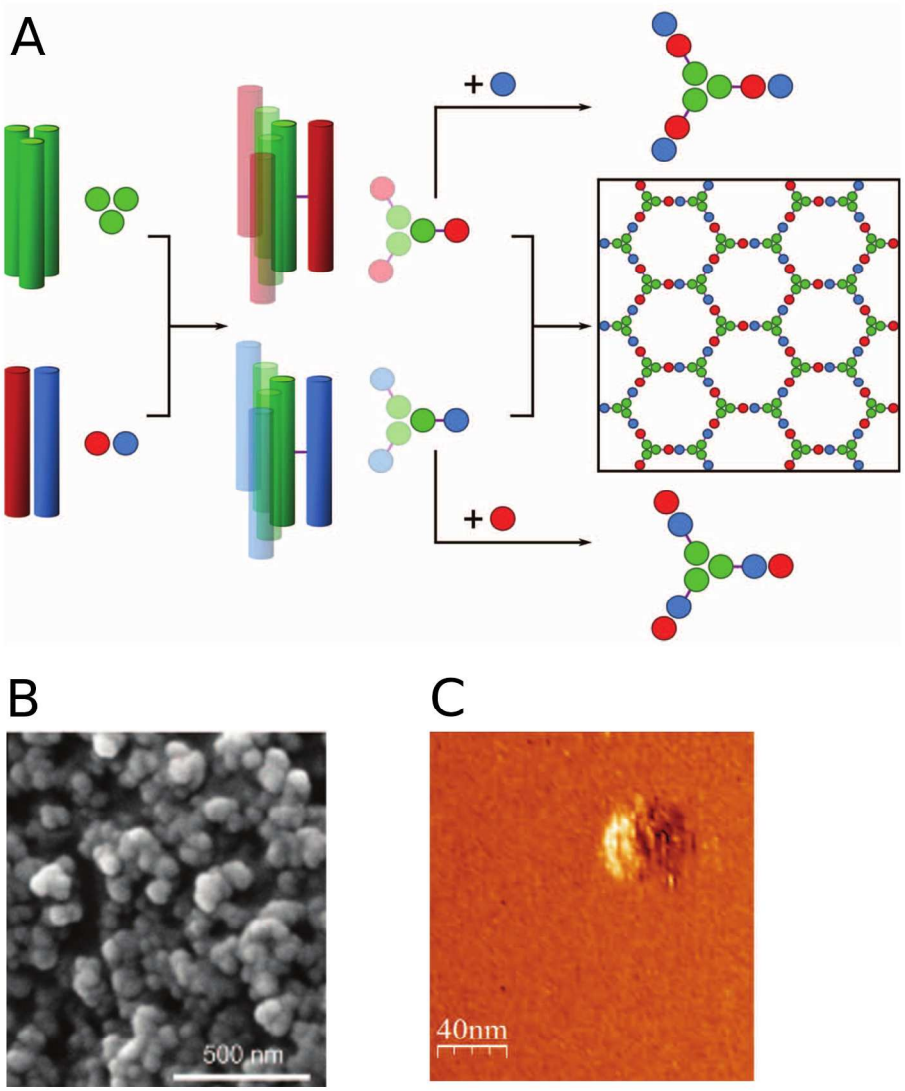

Figure 13:
Design of 'spheres' using coiled coil assemblies (SAGE). (A) Design of the hubs that self-assemble to form SAGE molecules. The heterotrimeric coiled-coil (CC-Tri3, green) connects to either CC-Di-A (red) and CC-Di-B (blue) via asymmetric sulfide linkages (purple lines) to form hub A (red-green) and hub B (blue-green). Mixing of hub A and hub B yields a hexagonal array by the formation of heterodimeric coiled-coils between CC-Di-A and CC-Di-B. EM (B) and LMFM (C) images of a hydrated SAGE molecule.

From Fletcher, J. M.; Harniman, R. L.; Barnes, F. R.; Boyle, A. L.; Collins, A.; Mantell, J.; Sharp, T. H.; Antognozzi, M.; Booth, P. J.; Linden, N.; Miles, M. J.; Sessions, R. B.; Verkade, P.; Woolfson, D. N. Science 2013, 340, 595-599. Reprinted with permission from AAAS.

215x279mm (300 x 300 DPI)

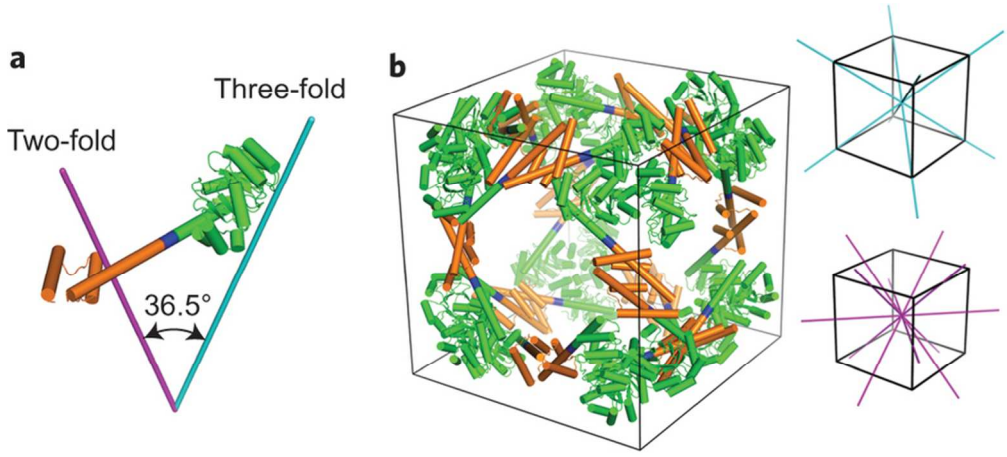

Figure 14: Design of a 24-subunit protein cube. (a) The designed fusion protein with the KDPGal aldolase trimer (green) connected to the dimeric domain of FkpA protein (orange) by the helical linker (blue). The purple and cyan lines represent the two-fold and three-fold axes of symmetry, respectively. (b) A cartoon model of the 24-subunit cage design. The two-fold and three-fold axes of symmetry in the cube are shown in purple and cyan, respectively.

333x150mm (72 x 72 DPI)

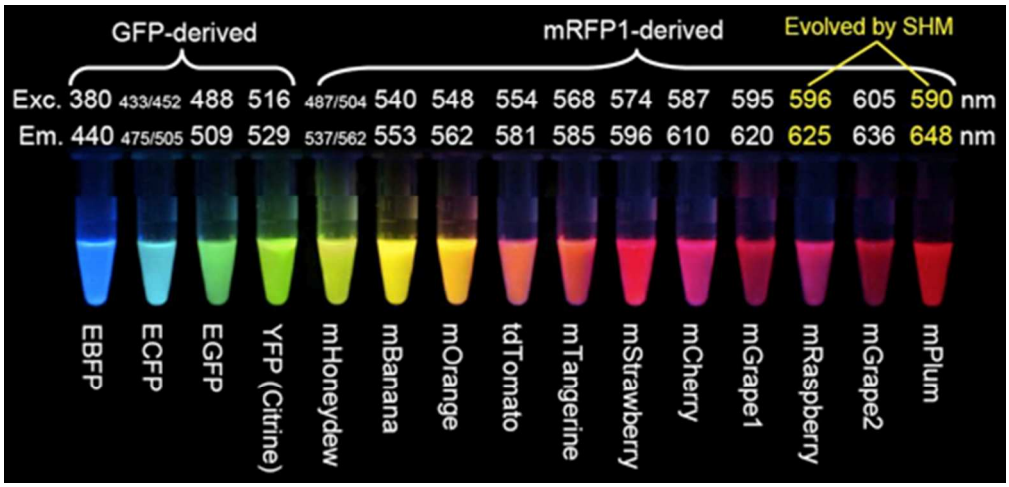

Figure 15: Fluorescent proteins with a wide spectral range of excitation and emission.

Reprinted with permission from Tsien, R. Y. Angewandte Chemie International Edition 2009, 48, 5612-5626.

141x67mm (96 x 96 DPI)

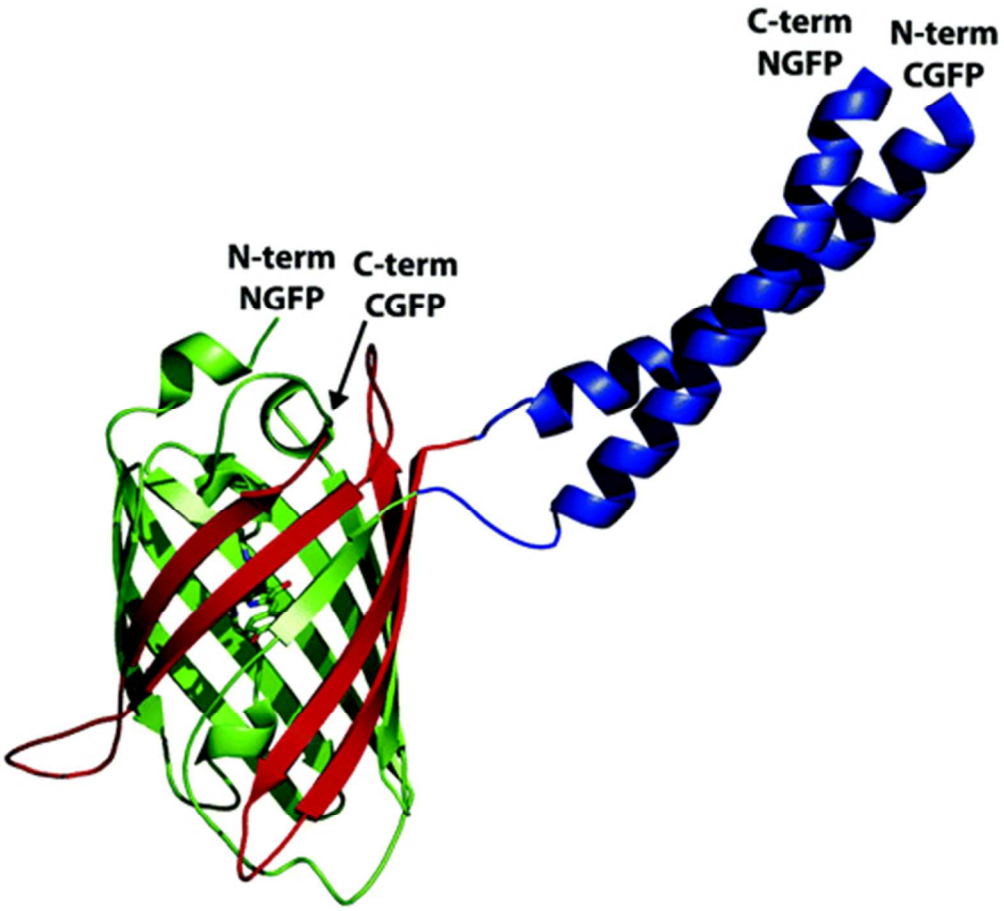

Figure 16: Schematic illustration of the split GFP system used to identify protein-protein interactions. GFP is split into N-terminal (green) and C-terminal (red) halves, which do not associate on their own. Attaching two interacting proteins (depicted here are a designed pair of coiled-coil dimers) forces the two halves to associate, producing the native fold and fluorophore.

176x160mm (72 x 72 DPI)

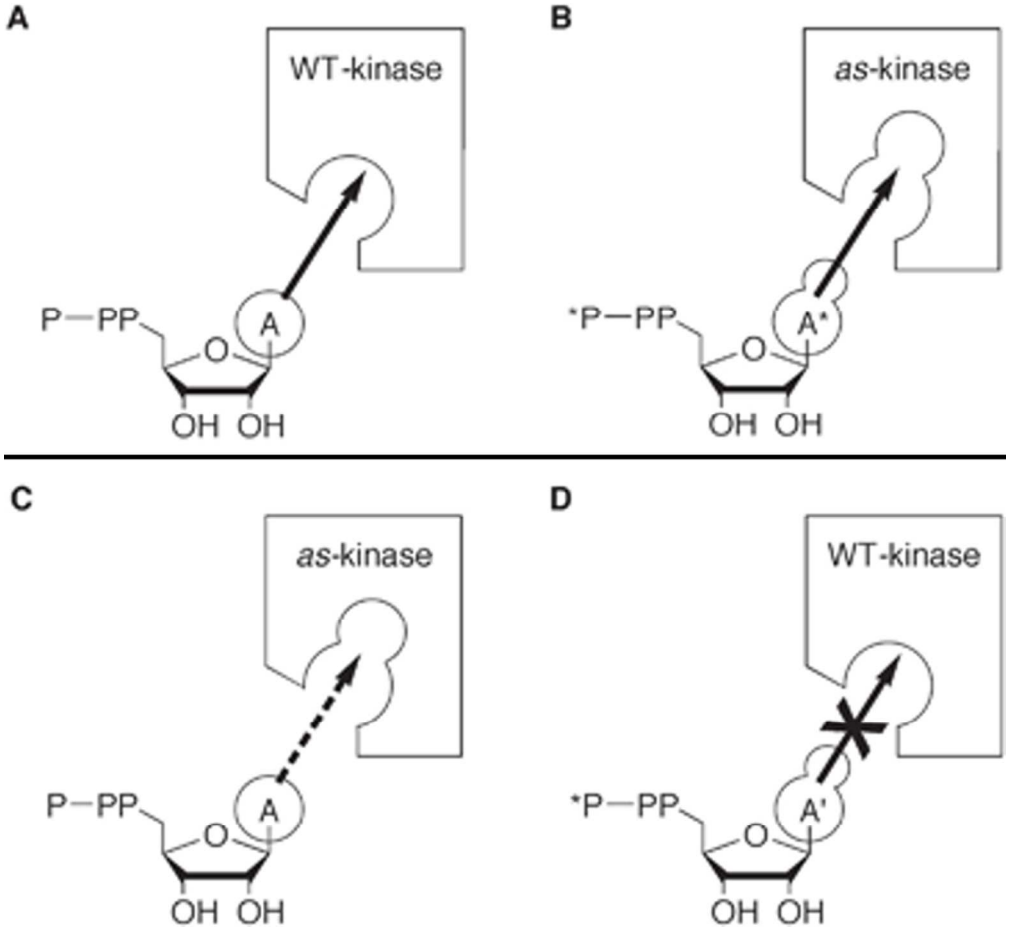

Figure 17: Schematic illustration of the method developed by Shokat et al. to identify kinase substrates. Analogue sensitive (As)-kinase contains a mutation in the ATP binding domain that enables it to function with both ATP (A) and a bulky ATP derivative (A*), while wild-type (WT) kinase can only functional with A.

Hodgson, D. R.; Schröder, M. Chemical Society Reviews 2011, 40, 1211-1223 - Reproduced by permission of The Royal Society of Chemistry.

148x136mm (150 x 150 DPI)

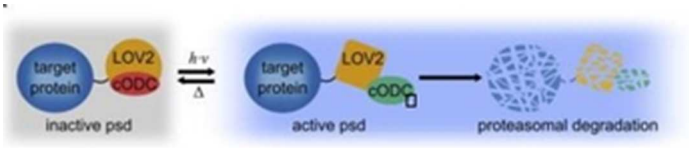

Figure 18: Cartoon depiction of the LOV2 based photosensitive degron (psd). Psd consists of a fusion between the LOV2 photosensitive domain and the proteasome binding peptide cODC. In the dark state, the Jα helix on the LOV2 domain is folded and associates with cODC, preventing it from interacting with the proteasome. Absorption of blue light triggers unfolding of Jα, releasing the cODC peptide to bind the proteasome and induce degradation of the psd module, along with any protein fused to it.

29x6mm (300 x 300 DPI)