

Published in final edited form as:

Protein Sci. 2016 July ; 25(7): 1179–1187. doi:10.1002/pro.2892.

Why reinvent the wheel? Building new proteins based on ready-made parts

Olga Khersonsky^{1,*} and Sarel J Fleishman^{1,*}

¹Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel

Abstract

We protein engineers are ambivalent about evolution: on the one hand, evolution inspires us with myriad examples of biomolecular binders, sensors, and catalysts; on the other hand, these examples are seldom well-adapted to the engineering tasks we have in mind. Protein engineers have therefore modified natural proteins by point substitutions and fragment exchanges in an effort to generate new functions. A counterpoint to such design efforts, which is being pursued now with greater success, is to completely eschew the starting materials provided by nature and to design new protein functions from scratch by using *de novo* molecular modeling and design. While important progress has been made in both directions, some areas of protein design are still beyond reach. To this end, we advocate a synthesis of these two strategies: by using design calculations to both recombine and optimize fragments from natural proteins, we can build stable and as of yet un-sampled structures, thereby granting access to an expanded repertoire of conformations and desired functions. We propose that future methods that combine phylogenetic analysis, structure and sequence bioinformatics, and atomistic modeling may well succeed where any one of these approaches has failed on its own.

Keywords

protein design; evolution; bioinformatics; Rosetta; β/α barrel; β -propeller; antibody

Introduction

To quote Richard Feynman, “What I cannot create, I do not understand”. Protein design, then, must be the ultimate test of our understanding of the physical principles that underlie the basic building blocks of life. Over the past decades protein designers have made substantial progress in fold design. Starting with α -helix bundles (1–4), several folds have been successfully designed, including the Rossmann fold (5), the β/α barrel (6) and even novel folds not observed in nature (7). The guiding principle in all of these studies has been finding an energy minimum in sequence-conformation space (8). In general, the resulting designs have very high secondary-structure content and short loops. In some cases, designed proteins exhibit extreme stabilities not seen in natural proteins; for example, the $G_{\text{unfolding}}$ of designed helical bundles was found to be ~60kcal/mol (1) compared to typical unfolding free energies of natural proteins which range from 5 to 20kcal/mol (9). Taken as a whole,

*correspondence: olga.khersonsky@weizmann.ac.il, sarel@weizmann.ac.il.

these results demonstrate that we possess a fairly accurate understanding of the thermodynamics and design principles that underlie protein folds.

The design of protein function has met with more limited, albeit substantial, success. Algorithms that search for pre-organized structural elements and favorable interactions with the target were used to design small-molecule and protein binders. The resulting interfaces were found to be accurate -- in some cases, down to atomic details -- when compared to the subsequently determined experimental structures (10⁻¹⁴). Additionally, *de novo* designed binders and enzymes have been experimentally evolved to affinities or catalytic rates seen in nature (10⁻¹¹, 14⁻¹⁶). Although these results are promising, all the successful designs are reminiscent of the highly stable designed folds mentioned above, since they relied on rigid protein scaffolds with high secondary-structure content. Many natural proteins, by contrast, encode functional elements in regions lacking secondary structure. Recent analyses of design of function have therefore persistently pointed to the same problem: computational efforts fail to design loops or modify backbones for function (17⁻¹⁹).

Given these results, the question inevitably arises: should protein designers abandon backbone design in favor of more tractable systems, such as those with high secondary-structure content and rigid conformations? There are two arguments against this attitude: first, except in privileged cases, design of function demands positioning of multiple functional groups with high geometric accuracy. If we limit ourselves only to natural backbones and rigid scaffolds we will fail to address many real-world enzyme and binder-design problems. For instance, in enzyme design we have so far generated catalytic sites with up to four active-site residues; active sites of natural enzymes, by contrast, rely on a complex network of interactions, sometimes encompassing more than ten residues (19). Second, harking back to Feynman's quote, given that so many natural proteins encode functional elements on loops, we must address structural plasticity in design if we are to reach a deeper understanding of how function is encoded in nature.

Why has backbone and loop design for function proved so challenging? The answer is that design of function necessarily invokes tradeoffs between stability, foldability, and activity. Whereas fold design searches for the sequence and structure that optimize system energy, design of function must encode cavities, exposed hydrophobic groups for ligand binding, and desolvated polar and charged groups for increased reactivity. All of these molecular features decrease stability and may compromise foldability, especially in loop regions, since these often require backbone-side chain interactions to configure properly (20⁻²²). Furthermore, to ensure that the designed protein folds correctly and configures all functional groups in the desired orientations, it is necessary to encode second and third-shell stabilizing interactions around the active site, imposing additional design constraints. Considering all of the structure and sequence constraints that are a prerequisite to the design of function, it is not surprising that certain folds resist design efforts to adopt radically new functions that they were not naturally evolved to carry out. For instance, imposing the Kemp eliminase reactive groups on a native TIM barrel led to an unstable protein (16), and there are surely more such negative cases left unreported. Clearly, encoding all the interactions that are necessary to pre-organize the active site and its surroundings presents a critical challenge for design of function.

While improvements in the energy function (23, 24) and conformation sampling (25) will continue to make important contributions to backbone and function design, others and we are looking for clues from nature on the design principles of backbones involved in function. Instead of designing backbones from first principles – for which there are still formidable challenges with respect to both the energy function and the conformation search – our strategy is to determine the rules of backbone design within an individual fold and to use these rules as constraints in the design of novel functional backbones. In the following we describe the experimental, structural, bioinformatics, and atomistic-design studies that shed light on the principles of designing new backbones and functions within a protein fold.

Structural modularity facilitates the evolution of new functions in protein superfamilies

Many of the diverse molecular functions in nature are carried out by just a handful of superfamilies (26–28). Proteins within a superfamily share common ancestry as well as structural and mechanistic features (Fig.1), although some members of a superfamily might have diverged to the point that they do not show detectable sequence similarity (29). Structure and sequence analyses reveal that even among superfamily members that show no detectable sequence relationships, structure motifs and catalytic machineries can be conserved, indicating that evolution reuses structural elements as the functional repertoire of a family expands (30, 31). Recent bioinformatics analyses suggested that motifs can be conserved across different fold families (32), and even beyond the fold level (33).

The observation of common backbone segments in protein superfamilies provides an impressive demonstration of protein structural plasticity and its functional consequences. Exchanging backbone fragments among homologues is a powerful catalyst for the evolution of new or modified functions. Compared to the more gradual process of random substitution, incorporating an entirely new segment that has evolved to fit within the context of the fold has a higher chance of leading to considerable change in activity while conserving foldability (34–37). Certainly not all protein folds are modular (38); yet, a high degree of structural modularity is a distinguishing feature of many superfamilies, including globular symmetric folds (TIM barrels (39), β -propellers (40), β -trefoils (41), linear repeat proteins (42), and even asymmetric folds like antibodies). Indeed, modularity would seem to be a prerequisite for swapping fragments among homologues, and evolutionary forces that expand the functional repertoire may therefore prefer modular folds (43). Not surprisingly, modularity has attracted considerable attention from protein engineers in their search for ways to generate new backbones for function (44). Gene-shuffling techniques, which mimic natural recombination, have been widely used in laboratory-evolution experiments to produce proteins with enhanced stability and activity (45, 46). For instance, the SCHEMA method uses structures of homologous proteins to identify positions, where segmentation would abrogate the fewest possible physical contacts between segments; the resulting segments are then recombined experimentally and subjected to selection (47).

TIM barrels, β -propellers, and antibodies as examples of modular and functionally highly versatile protein classes

TIM barrels

The TIM barrel (or β/α barrel) fold is one of the most ubiquitous folds in nature and is capable of profound functional versatility, with representatives in 5 out of 6 enzyme classes (by EC numbers classification) (37). This fold can be schematically represented as an eightfold repetition of the β/α unit organized in two circular layers of secondary structures (48). The inner layer consists of eight parallel β -strands that form a closed barrel and is surrounded by an external layer of eight α -helices. The active sites of the TIM-barrel are located at the C-terminal ends of the β -strands. The structural modularity of TIM barrels has been demonstrated by sequence and structure studies on enzymes from the histidine-biosynthesis pathway, suggesting that the β/α barrel evolved by double duplication of ancestral $(\beta/\alpha)_2$ subunits, each of which comprises a quarter-barrel (39) (Fig. 2A).

The symmetric nature of TIM barrels has inspired multiple engineering experiments attempting to revisit ancestral forms along the gene-duplication and fusion evolutionary trajectory. Complete β/α barrels were built by duplication and fusion of half-barrel proteins (49–50), and subsequent sequence optimization by directed evolution yielded stable and catalytically active proteins (51–52). Remarkably, it was demonstrated that recombination of TIM barrel parts with proteins from entirely different folds, such as flavodoxin (belonging to the Rossmann fold), is possible, as long as the contributing and receiving proteins share similar structural fragments (53–54). These experiments not only demonstrate the modular nature of the β/α barrel fold, but also illustrate that sequence-structure space is readily traversable, especially through modular assembly.

Despite these promising results, *de novo* design of TIM-barrels and their engineering by loop grafting remain challenging. For example, swapping β - α loop 7 between the evolutionarily related phosphotriesterase (PTE) and lactonase enzymes shown in Figure 1 yielded mostly insoluble proteins, despite the high structural similarity among these proteins (55); several attempts to design *de novo* TIM-barrel structures resulted in unstable proteins with poorly defined tertiary structure (56–59), and only recently an idealized 4-fold symmetrical TIM-barrel was successfully designed from first principles (6). Although this designed protein lacks catalytic function it may in future open the way to designing complex active sites completely from scratch.

β -propellers

The β -propeller fold comprises between 4 and 10 repeats of 4-stranded β blades, which are arranged radially around a central pore. Despite their diversity, β -propellers show both sequence and structure similarity, and it has been demonstrated that they have evolved by duplication and fusion of single blades, followed by diversification (40). Several engineering experiments have proven the feasibility, as well as the challenges, of β -propeller construction from unrelated subunits. Consensus design and *in vitro* selection methods were used to generate β -propellers, but they formed molten globules (60), aggregated (61), or formed blade-swapped homodimers, a configuration that has not been observed in nature

(62, 63). More recently, an integrative modeling approach built several symmetrical 6-bladed propellers by ancestral-sequence reconstruction of a repeat unit and atomistic modeling (64). The structures were confirmed by X-ray crystallography, demonstrating the feasibility of β -propeller assembly from single blade fragments.

Antibodies

Antibodies are the major soluble component of the mammalian immune system (65). Although they are too closely related to be considered a superfamily, antibodies embody the same characteristics of high structural diversity, modularity, and functional versatility seen in the previously discussed symmetric superfamilies. Indeed, the modularity of the antibody variable domain is an essential and highly adaptive property, which allows the immune system to encode a repertoire of $\sim 10^9$ different potential binders using fewer than 500 gene segments (Fig. 2C).

The structural modularity of the antibody variable domain was known even before the advent of protein engineering (66, 67). In fact, the humanization of murine antibodies by grafting mouse-derived complementarity-determining regions (CDRs) onto a human antibody framework still represents a powerful example of leveraging modularity to effect new protein functions (68). Although early humanization approaches were successful, later studies revealed an unanticipated complication: CDR conformation and function are determined, in part, by interactions with the antibody framework (69). CDR grafting therefore often resulted in lower affinity and specificity relative to the parent antibody due to deformation of the CDR loops. These observations highlight that even subtle differences within the framework can have profound impact on the stability and function of the CDRs.

Combinatorial-backbone design: assembling natural backbone fragments and optimizing their sequences

Inspired by these bioinformatics and engineering lessons, our lab is developing design algorithms to use the plasticity and modularity of protein families to generate novel protein structures (70). While the studies above show the promise of recombining natural proteins to form new ones, they also bring to light a consistent problem: recombined proteins are often unstable, aggregation-prone, functionally impaired, poorly expressed, or completely inactive. Simply put, not all natural fragments from homologues can be fitted together. To effectively search the immense space of potential backbones that can be generated by modular assembly we therefore need both (1) a rational method for selecting fragments for recombination; and (2) a method to optimize the recombined protein sequence to obtain well-behaved and functional proteins.

Much engineering work suggests that loops are interchangeable across proteins belonging to a given fold, be it the CDRs on antibodies, or the β/α loops in TIM-barrels. Still, modeling shows that the scaffold and the loops, which sit on top of the scaffold, form complex networks of stabilizing contacts (Fig. 3A); ignoring these contacts in design poses the risk of impairing function, foldability, or stability. Our approach, which we call combinatorial-backbone design, instead uses segments that comprise the entire modular subunit of the

protein fold. Starting from a structural alignment of fold members, all structures are segmented into backbone fragments, where the segmentation points, or stems, are chosen such that they are structurally well aligned across family members and the intervening segment is structurally diverse. Antibodies are ideally suited to this segmentation procedure, since the fully conserved disulfide bonds in the light and heavy variable domains are aligned across all family members, and the intervening segment comprises the hyper-variable CDRs 1 and 2 (Fig. 4A). In TIM-barrels we can similarly find segmentation points in the core that are well aligned and encompass crucial stabilizing interactions between the α -helices and β -strands, as well as the β - α loops, which hold crucial substrate-recognition elements (Fig. 4B, C). The method recombines the resulting segments, irrespective of their evolutionary and functional origins, and uses Rosetta combinatorial sequence optimization to design low-energy structures. While studying antibodies and TIM barrels, we found that stabilizing contacts within segments are often highly conserved among structurally similar backbone conformations, but not across different conformation clusters, even in cases where the different conformations have the same sequence length (Fig. 3B). To guarantee that design simulations maintain these important sequence features, and yet are free to sample sequence choices in less constrained parts of the protein, we generate sequence profiles specifically for each conformation class and bias design calculations to sample amino acid choices from these profiles. Thus, the design algorithm samples different combinations of fragments from natural proteins belonging to the same family, and for each combination it optimizes the sequence, based on constraints derived from sequence alignments for each fragment. New combinations of backbones and designed sequences are selected if they encode favorable structure and energy characteristics, thus ensuring that the fragments are compatible with one another. This approach can generate, from a few hundred protein scaffolds on the order of 10^{14} distinct backbones, an unprecedented number for use in design of function.

To experimentally test combinatorial-backbone design we applied it to the long-standing problem of designing novel antibody binders (unpublished results). Similar to the early recombination experiments mentioned above, our first-generation designs expressed poorly and were unstable; however, through iterations of algorithm development and experimental testing we improved the stability of designed antibodies and obtained several binders. Binding was optimized by affinity maturation (71), and mutation analysis at the designed binding surface supports the designed binding mode. To demonstrate the generality of this approach, we have recently extended it to design novel TIM-barrel enzymes with detectable, though still weak, levels of phosphotriesterase activity, which we are now optimizing.

Conclusions

Designing backbones that lack secondary-structure elements for function remains an unsolved problem. The challenges are many: the conformation space of regions without secondary-structure elements is very large; loop regions need to be stabilized through a cooperative network of polar and hydrophobic interactions both within the loop and with the supporting scaffold structure; and designed active-site or binding residues must be compatible with the backbone conformation. Amino acid sidechain preorganization at binding sites has been successfully treated in past applications of fixed-backbone design on secondary-structure elements (10–12, 14, 72); yet, the number of constraints and their

complexity is much greater when the active site is positioned on unstructured backbones. At a more fundamental level, macromolecular energy functions used in design, though they are being improved, still suffer from inaccuracies in balancing contributions from solvation, hydrogen bonding, and electrostatics—all three of which are critical for accurate positioning of polar backbone groups and active-site residues.

A potential way of addressing some of these difficulties is to recombine fragments from natural proteins to generate new ones (see the recent review by Hocker (44) and the references within). The approach we are pursuing tries to bridge the engineering and design approaches, sampling fragment conformations from natural proteins and optimizing the sequences of the resulting proteins for stability and function (70). An iterative approach of algorithm development and experimental testing has been especially rewarding in teaching us important lessons on the extent and limits of modularity in protein folds. For instance, through repeated cycles of computational design and experimental characterization we could test the effects of modeling components in generating more stable and active proteins. Starting from an initial set of antibody and TIM-barrel designs that expressed very poorly we found that the use of segmentation stem points and sequence constraints from structurally homologous segments improved the designs considerably. There are many more lessons to learn about how to use information taken from protein families to encode new functions. We expect that future applications in design of function will take a synthetic approach, combining phylogenetic analysis, structure-based engineering, laboratory evolution, and atomistic simulations.

Acknowledgements

We thank Liam Longo, Gideon Lapidoth, and Maria-Gabriele Pszolla, for critical comments and discussions. The research was supported by the Human Frontier Science Program, a Starter's Grant from the European Research Council, the Israel Science Foundation (ISF) through an individual grant, the ISF-UGC framework, and the Center for Research Excellence in Structural Cell Biology, the Israel Chief Scientist's Office, the Marie Curie Reintegration Career Development Award, the Minerva Foundation, and a charitable donation from Sam Switzer and family. SJF is a Martha S Sagon Career Development Chair at the Weizmann Institute of Science.

References

1. Huang PS, Oberdorfer G, Xu C, Pei XY, Nannenga BL, Rogers JM, DiMaio F, Gonen T, Luisi B, Baker D. High thermodynamic stability of parametrically designed helical bundles. *Science*. 2014; 346(6208):481–485. [PubMed: 25342806]
2. Hill RB, Raleigh DP, Lombardi A, DeGrado WF. De novo design of helical bundles as models for understanding protein folding and function. *Acc Chem Res*. 2000; 33(11):745–754. [PubMed: 11087311]
3. Hecht MH, Richardson JS, Richardson DC, Ogden RC. De novo design, expression, and characterization of Felix: a four-helix bundle protein of native-like sequence. *Science*. 1990; 249(4971):884–891. [PubMed: 2392678]
4. Regan L, DeGrado WF. Characterization of a helical protein designed from first principles. *Science*. 1988; 241(4868):976–978. [PubMed: 3043666]
5. Koga N, Tatsumi-Koga R, Liu G, Xiao R, Acton TB, Montellione GT, Baker D. Principles for designing ideal protein structures. *Nature*. 2012; 491(7423):222–227. [PubMed: 23135467]
6. Huang PS, Feldmeier K, Parmeggiani F, Hocker B, Baker D. De novo design of an ideal TIM-barrel scaffold. 2015 in press.

7. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. Design of a novel globular protein fold with atomic-level accuracy. *Science*. 2003; 302(5649):1364–1368. [PubMed: 14631033]
8. Fleishman SJ, Baker D. Role of the biomolecular energy gap in protein design, structure, and evolution. *Cell*. 2012; 149(2):262–273. [PubMed: 22500796]
9. Fersht, A. Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding. W.H. Freeman and Company; New York: 1999.
10. Strauch EM, Fleishman SJ, Baker D. Computational design of a pH-sensitive IgG binding protein. *Proc Natl Acad Sci U S A*. 2014; 111(2):675–680. [PubMed: 24381156]
11. Tinberg CE, Khare SD, Dou J, Doyle L, Nelson JW, Schena A, Jankowski W, Kalodimos CG, Johnsson K, Stoddard BL, Baker D. Computational design of ligand-binding proteins with high affinity and selectivity. *Nature*. 2013; 501(7466):212–216. [PubMed: 24005320]
12. Procko E, Hedman R, Hamilton K, Seetharaman J, Fleishman SJ, Su M, Aramini J, Kornhaber G, Hunt JF, Tong L, Montelione GT, et al. Computational design of a protein-based enzyme inhibitor. *J Mol Biol*. 2013; 425(18):3563–3575. [PubMed: 23827138]
13. Der BS, Machius M, Miley MJ, Mills JL, Szyperski T, Kuhlman B. Metal-mediated affinity and orientation specificity in a computationally designed protein homodimer. *J Am Chem Soc*. 2012; 134(1):375–385. [PubMed: 22092237]
14. Fleishman SJ, Whitehead TA, Ekiert DC, Dreyfus C, Corn JE, Strauch EM, Wilson IA, Baker D. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science*. 2011; 332(6031):816–821. [PubMed: 21566186]
15. Blomberg R, Kries H, Pinkas DM, Mittl PR, Grutter MG, Privett HK, Mayo SL, Hilvert D. Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature*. 2013; 503(7476):418–421. [PubMed: 24132235]
16. Khersonsky O, Kiss G, Rothlisberger D, Dym O, Albeck S, Houk KN, Baker D, Tawfik DS. Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59. *Proc Natl Acad Sci U S A*. 2012; 109(26):10358–10363. [PubMed: 22685214]
17. Stranges PB, Kuhlman B. A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds. *Protein Sci*. 2013; 22(1):74–82. [PubMed: 23139141]
18. Khare SD, Fleishman SJ. Emerging themes in the computational design of novel enzymes and protein-protein interfaces. *FEBS Lett*. 2013; 587(8):1147–1154. [PubMed: 23262222]
19. Baker D. An exciting but challenging road ahead for computational enzyme design. *Protein Sci*. 2010; 19(10):1817–1819. [PubMed: 20717908]
20. Longo L, Lee J, Blaber M. Experimental support for the foldability-function tradeoff hypothesis: segregation of the folding nucleus and functional regions in fibroblast growth factor-1. *Protein Sci*. 2012; 21(12):1911–1920. [PubMed: 23047594]
21. Tokuriki N, Stricher F, Serrano L, Tawfik DS. How protein stability and new functions trade off. *PLoS Comput Biol*. 2008; 4(2):e1000002. [PubMed: 18463696]
22. Shoichet BK, Baase WA, Kuroki R, Matthews BW. A relationship between protein stability and protein function. *Proc Natl Acad Sci U S A*. 1995; 92(2):452–456. [PubMed: 7831309]
23. Li Z, Yang Y, Zhan J, Dai L, Zhou Y. Energy functions in de novo protein design: current challenges and future prospects. *Annu Rev Biophys*. 2013; 42:315–335. [PubMed: 23451890]
24. Leaver-Fay A, O'Meara MJ, Tyka M, Jacak R, Song Y, Kellogg EH, Thompson J, Davis IW, Pache RA, Lyskov S, Gray JJ, et al. Scientific benchmarks for guiding macromolecular energy function improvement. *Methods Enzymol*. 2013; 523:109–143. [PubMed: 23422428]
25. Ollikainen N, Smith CA, Fraser JS, Kortemme T. Flexible backbone sampling methods to model and design protein alternative conformations. *Methods Enzymol*. 2013; 523:61–85. [PubMed: 23422426]
26. Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, et al. The Pfam protein families database. *Nucleic Acids Res*. 2012; 40(Database issue):D290–301. [PubMed: 22127870]

27. Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM. CATH--a hierarchic classification of protein domain structures. *Structure*. 1997; 5(8):1093–1108. [PubMed: 9309224]
28. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol*. 1995; 247(4):536–540. [PubMed: 7723011]
29. Orengo CA, Thornton JM. Protein families and their evolution--a structural perspective. *Annu Rev Biochem*. 2005; 74:867–900. [PubMed: 15954844]
30. Glasner ME, Gerlt JA, Babbitt PC. Evolution of enzyme superfamilies. *Curr Opin Chem Biol*. 2006; 10(5):492–497. [PubMed: 16935022]
31. Gerlt JA, Babbitt PC. Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu Rev Biochem*. 2001; 70:209–246. [PubMed: 11395407]
32. Nepomnyachiy S, Ben-Tal N, Kolodny R. Global view of the protein universe. *Proc Natl Acad Sci U S A*. 2014; 111(32):11691–11696. [PubMed: 25071170]
33. Farias-Rico JA, Schmidt S, Hocker B. Evolutionary relationship of two ancient protein superfolds. *Nat Chem Biol*. 2014; 10(9):710–715. [PubMed: 25038785]
34. Tawfik DS. Biochemistry. Loop grafting and the origins of enzyme species. *Science*. 2006; 311(5760):475–476. [PubMed: 16439649]
35. Panchenko AR, Madej T. Structural similarity of loops in protein families: toward the understanding of protein evolution. *BMC Evol Biol*. 2005; 5:10. [PubMed: 15691378]
36. Voigt CA, Martinez C, Wang ZG, Mayo SL, Arnold FH. Protein building blocks preserved by recombination. *Nat Struct Biol*. 2002; 9(7):553–558. [PubMed: 12042875]
37. Nagano N, Orengo CA, Thornton JM. One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *J Mol Biol*. 2002; 321(5):741–765. [PubMed: 12206759]
38. Dellus-Gur E, Toth-Petroczy A, Elias M, Tawfik DS. What makes a protein fold amenable to functional innovation? Fold polarity and stability trade-offs. *J Mol Biol*. 2013; 425(14):2609–2621. [PubMed: 23542341]
39. Richter M, Bosnali M, Carstensen L, Seitz T, Durchschlag H, Blanquart S, Merkl R, Sterner R. Computational and experimental evidence for the evolution of a (beta alpha)8-barrel protein from an ancestral quarter-barrel stabilised by disulfide bonds. *J Mol Biol*. 2010; 398(5):763–773. [PubMed: 20363228]
40. Chaudhuri I, Soding J, Lupas AN. Evolution of the beta-propeller fold. *Proteins*. 2008; 71(2):795–803. [PubMed: 17979191]
41. Lee J, Blaber M. Experimental support for the evolution of symmetric protein architecture from a simple peptide motif. *Proc Natl Acad Sci U S A*. 2011; 108(1):126–130. [PubMed: 21173271]
42. Main ER, Jackson SE, Regan L. The folding and design of repeat proteins: reaching a consensus. *Curr Opin Struct Biol*. 2003; 13(4):482–489. [PubMed: 12948778]
43. Soding J, Lupas AN. More than the sum of their parts: on the evolution of proteins from peptides. *Bioessays*. 2003; 25(9):837–846. [PubMed: 12938173]
44. Hocker B. Design of proteins from smaller fragments--learning from evolution. *Curr Opin Struct Biol*. 2014; 27:56–62. [PubMed: 24865156]
45. Stemmer WPC. Rapid Evolution of a Protein in-Vitro by DNA Shuffling. *Nature*. 1994; 370(6488):389–391. [PubMed: 8047147]
46. Goldsmith M, Tawfik DS. Directed enzyme evolution: beyond the low-hanging fruit. *Curr Opin Struct Biol*. 2012; 22(4):406–412. [PubMed: 22579412]
47. Heinzelman P, Romero PA, Arnold FH. Efficient Sampling of SCHEMA Chimera Families to Identify Useful Sequence Elements. *Method Enzymol*. 2013; 523:351–368.
48. Sterner R, Hocker B. Catalytic versatility, stability, and evolution of the (betaalpha)8-barrel enzyme fold. *Chem Rev*. 2005; 105(11):4038–4055. [PubMed: 16277370]
49. Akanuma S, Yamagishi A. Experimental evidence for the existence of a stable half-barrel subdomain in the (beta/alpha)8-barrel fold. *J Mol Biol*. 2008; 382(2):458–466. [PubMed: 18674541]

50. Hocker B, Claren J, Sterner R. Mimicking enzyme evolution by generating new (betaalpha)8-barrels from (betaalpha)4-half-barrels. *Proc Natl Acad Sci U S A*. 2004; 101(47):16448–16453. [PubMed: 15539462]
51. Sperl JM, Rohweder B, Rajendran C, Sterner R. Establishing catalytic activity on an artificial (betaalpha)8-barrel protein designed from identical half-barrels. *FEBS Lett*. 2013; 587(17):2798–2805. [PubMed: 23806364]
52. Claren J, Malisi C, Hocker B, Sterner R. Establishing wild-type levels of catalytic activity on natural and artificial (beta alpha)8-barrel protein scaffolds. *Proc Natl Acad Sci U S A*. 2009; 106(10):3704–3709. [PubMed: 19237570]
53. Shanmugaratnam S, Eisenbeis S, Hocker B. A highly stable protein chimera built from fragments of different folds. *Protein Eng Des Sel*. 2012; 25(11):699–703. [PubMed: 23081840]
54. Eisenbeis S, Proffitt W, Coles M, Truffault V, Shanmugaratnam S, Meiler J, Hocker B. Potential of fragment recombination for rational design of proteins. *J Am Chem Soc*. 2012; 134(9):4019–4022. [PubMed: 22329686]
55. Afriat-Jurnou L, Jackson CJ, Tawfik DS. Reconstructing a Missing Link in the Evolution of a Recently Diverged Phosphotriesterase by Active-Site Loop Remodeling. *Biochemistry*. 2012; 51(31):6047–6055. [PubMed: 22809311]
56. Figueroa M, Oliveira N, Lejeune A, Kaufmann KW, Dorr BM, Matagne A, Martial JA, Meiler J, Van de Weerd C. Octarellin VI: Using Rosetta to Design a Putative Artificial (beta/alpha)(8) Protein. *PLoS One*. 2013; 8(8)
57. Offredi F, Dubail F, Kischel P, Sarinski K, Stern AS, Van de Weerd C, Hoch JC, Prosperi C, Francois JM, Mayo SL, Martial JA. De novo backbone and sequence design of an idealized alpha/beta-barrel protein: Evidence of stable tertiary structure. *Journal of Molecular Biology*. 2003; 325(1):163–174. [PubMed: 12473459]
58. Houbrechts A, Moreau B, Abagyan R, Mainfroid V, Preaux G, Lamproye A, Poncin A, Goormaghtigh E, Ruysschaert JM, Martial JA, Goraj K. 2nd-Generation Octarellins - 2 New De-Novo (Beta/Alpha)(8) Polypeptides Designed for Investigating the Influence of Beta-Residue Packing on the Alpha/Beta-Barrel Structure Stability. *Protein Engineering*. 1995; 8(3):249–259. [PubMed: 7479687]
59. Goraj K, Renard A, Martial JA. Synthesis, Purification and Initial Structural Characterization of Octarellin, a Denovo Polypeptide Modeled on the Alpha Beta-Barrel Proteins. *Protein Engineering*. 1990; 3(4):259–266. [PubMed: 2188263]
60. Nikkhah M, Jawad-Alami Z, Demydchuk M, Ribbons D, Paoli M. Engineering of beta-propeller protein scaffolds by multiple gene duplication and fusion of an idealized WD repeat. *Biomol Eng*. 2006; 23(4):185–194. [PubMed: 16651025]
61. Yadid I, Tawfik DS. Functional beta-propeller lectins by tandem duplications of repetitive units. *Protein Engineering Design & Selection*. 2011; 24(1-2):185–195.
62. Yadid I, Kirshenbaum N, Sharon M, Dym O, Tawfik DS. Metamorphic proteins mediate evolutionary transitions of structure. *P Natl Acad Sci USA*. 2010; 107(16):7287–7292.
63. Yadid I, Tawfik DS. Reconstruction of functional beta-propeller lectins via homo-oligomeric assembly of shorter fragments. *Journal of Molecular Biology*. 2007; 365(1):10–17. [PubMed: 17054983]
64. Voet ARD, Noguchi H, Addy C, Simoncini D, Terada D, Unzai S, Park SY, Zhang KYJ, Tame JRH. Computational design of a self-assembling symmetrical beta-propeller protein. *P Natl Acad Sci USA*. 2014; 111(42):15102–15107.
65. Janeway, CA.; Travers, P.; Walport, M. *Immunobiology: The Immune System in Health and Disease*. Garland Science; New York: 2001.
66. Padlan EA. Structural Basis for Specificity of Antibody-Antigen Reactions and Structural Mechanisms for Diversification of Antigen-Binding Specificities. *Q Rev Biophys*. 1977; 10(1):35–&. [PubMed: 17883]
67. Kabat EA, Wu TT, Bilofsky H. Unusual Distributions of Amino-Acids in Complementarity-Determining (Hypervariable) Segments of Heavy and Light-Chains of Immunoglobulins and Their Possible Roles in Specificity of Antibody-Combining Sites. *J Biol Chem*. 1977; 252(19):6609–6616. [PubMed: 408353]

68. Riechmann L, Clark M, Waldmann H, Winter G. Reshaping Human-Antibodies for Therapy. *Nature*. 1988; 332(6162):323–327. [PubMed: 3127726]
69. Kettleborough CA, Saldanha J, Heath VJ, Morrison CJ, Bendig MM. Humanization of a mouse monoclonal antibody by CDR-grafting: the importance of framework residues on loop conformation. *Protein Eng.* 1991; 4(7):773–783. [PubMed: 1798701]
70. Lapidoth GD, Baran D, Pszolla GM, Norn C, Alon A, Tyka MD, Fleishman SJ. AbDesign: An algorithm for combinatorial backbone design guided by natural conformations and sequences. *Proteins*. 2015; 83(8):1385–1406. [PubMed: 25670500]
71. Chao G, Lau WL, Hackel BJ, Sazinsky SL, Lippow SM, Wittrup KD. Isolating and engineering human antibodies using yeast surface display. *Nat Protoc.* 2006; 1(2):755–768. [PubMed: 17406305]
72. Fleishman SJ, Khare SD, Koga N, Baker D. Restricted sidechain plasticity in the structures of native proteins and complexes. *Protein Sci.* 2011; 20(4):753–757. [PubMed: 21432939]

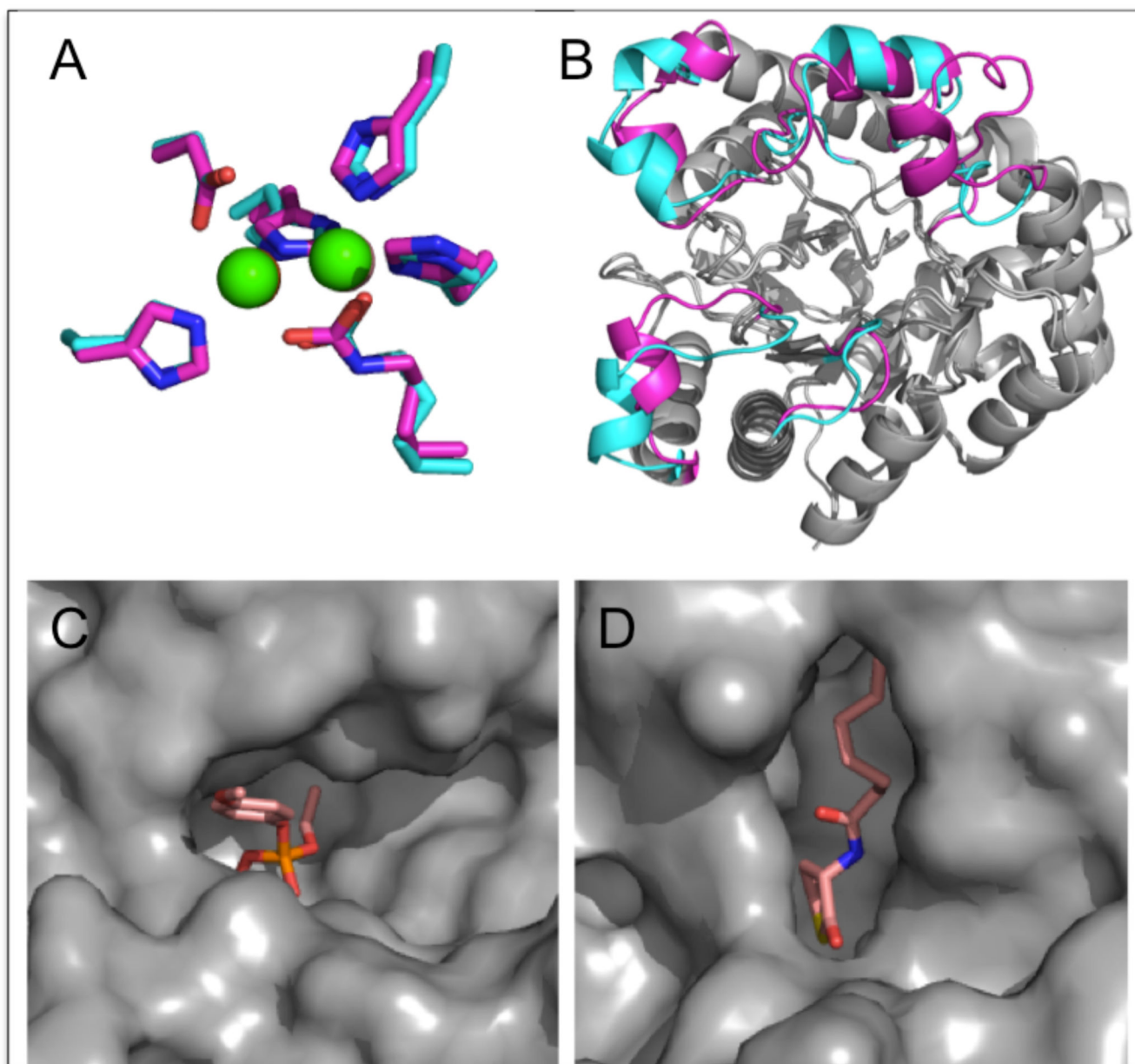


Figure 1. Conserved catalytic machinery on widely different active-site backbones

(A) Two members of the amidohydrolase family: phosphotriesterase (PTE, magenta) and lactonase (PLL, cyan) have remarkably conserved core catalytic groups, comprising two metal ions (green spheres) and chelating residues (sticks). (B) Divergence of the active site loops leads to quite different binding sites of PTE (C) and PLL (D). Substrates shown as sticks and protein in surface. Molecular representations were generated using PyMol [The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC.].

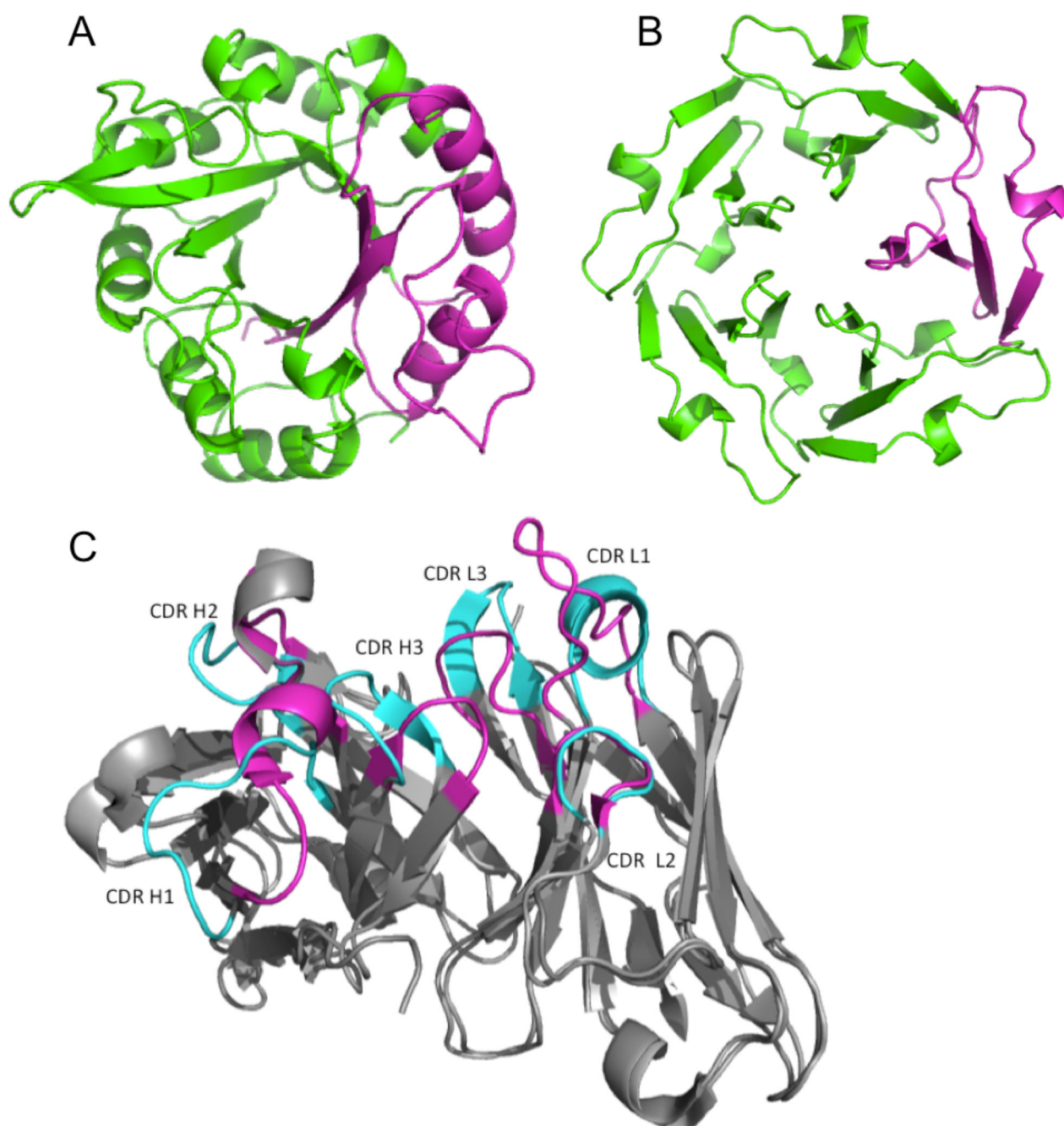


Figure 2. Examples of structural modularity in functionally versatile folds.

A. β/α barrel structure (PDB ID: 1THF), with a quarter-barrel highlighted in magenta. **B.** 5-bladed β -propeller structure (PDB ID: 1TL2), with a single blade subunit highlighted in magenta. **C.** Overlay of two antibodies, showing the conserved framework (gray) and divergent complementarity determining regions (CDRs). Molecular structures taken from the murine fluorescein-binding antibody (PDB ID: 1X9Q, purple) and an influenza hemagglutinin-binding antibody from human (PDB: 3GBN, cyan).

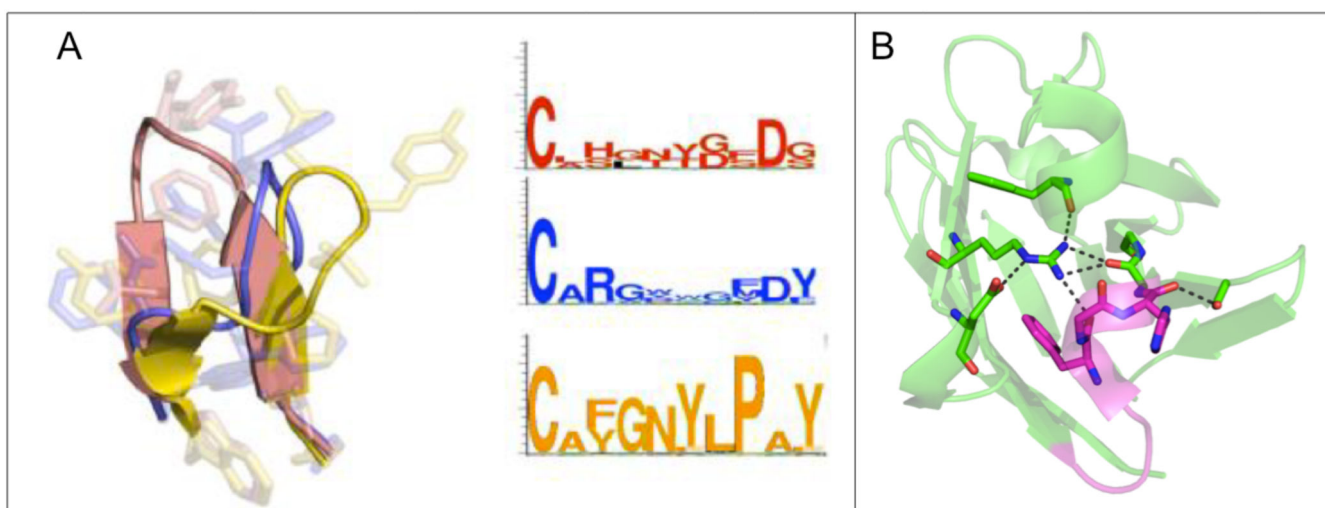


Figure 3.

A. Correlation of backbone segment conformations in antibodies with segment sequences (the conformations are colored according to their sequence profile).

B. Essential non-local stabilizing interactions of CDRs. Conformation of CDR H1 (magenta) of fluorescein binding antibody (PDB ID: 1X9Q) is supported by residues from adjacent framework and CDR regions, forming a complex network of nonlocal polar and hydrophobic interactions.

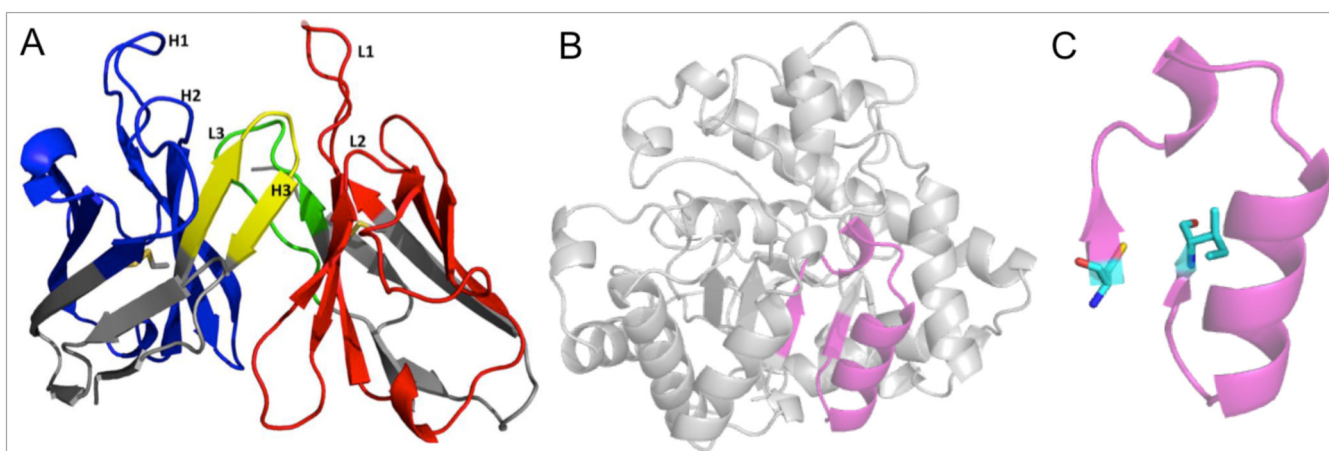


Figure 4.

A. Conformation segmentation used in *AbDesign*. *AbDesign* segments the antibody structure in places of highest structure conservation among antibodies (the disulfide cysteines, shown as sticks, and the stem positions of CDR3) to improve the potential of different conformation segments to be joined to form artificial combinations of backbones. The structure (PDB entry: 1X9Q) is color-coded by conformation segments (red: CDRs L1&L2, green: CDR L3, blue: CDRs H1 and H2, yellow: CDR H3). Grey segments are not subjected to backbone design and are left as in the template antibody. **B.** Conformation segmentation used in design of TIM barrels. The basic modular unit, $\beta-\alpha-\beta$, is shown in magenta on a phosphotriesterase (PTE, PDB entry: 1HZY). **C.** The stems, chosen as residues at the interface between β strands and α helices that are well-aligned across the family, are shown as cyan sticks.