

## Achievements and Challenges in Computational Protein Design

Ilan Samish

### Abstract

Computational protein design (CPD), a yet evolving field, includes computer-aided engineering for partial or full de novo designs of proteins of interest. Designs are defined by a requested structure, function, or working environment. This chapter describes the birth and maturation of the field by presenting 101 CPD examples in a chronological order emphasizing achievements and pending challenges. Integrating these aspects presents the plethora of CPD approaches with the hope of providing a “CPD 101”. These reflect on the broader structural bioinformatics and computational biophysics field and include: (1) integration of knowledge-based and energy-based methods, (2) hierarchical designated approach towards local, regional, and global motifs and the integration of high- and low-resolution design schemes that fit each such region, (3) systematic differential approaches towards different protein regions, (4) identification of key hot-spot residues and the relative effect of remote regions, (5) assessment of shape-complementarity, electrostatics and solvation effects, (6) integration of thermal plasticity and functional dynamics, (7) negative design, (8) systematic integration of experimental approaches, (9) objective cross-assessment of methods, and (10) successful ranking of potential designs. Future challenges also include dissemination of CPD software to the general use of life-sciences researchers and the emphasis of success within an in vivo milieu. CPD increases our understanding of protein structure and function and the relationships between the two along with the application of such know-how for the benefit of mankind. Applied aspects range from biological drugs, via healthier and tastier food products to nanotechnology and environmentally friendly enzymes replacing toxic chemicals utilized in the industry.

**Key words** Computational protein design, Inverse folding problem, De novo design, Directed evolution, Rational design, Synthetic biology, Negative design, Enzyme design, Protein–protein interaction

*“The abundance of substances of which animals and plants are composed of, the remarkable processes whereby they are formed and then broken down again claimed the attention of mankind, and hence from the early days they also persistently captivated the interest of chemists. . . . To determine the structure of the molecule the chemist proceeds in a similar way to the anatomist. By chemical actions he breaks the system down into its components and continues with this division until familiar substances emerge. Where this decomposition has taken different directions, the structure of the original system can be inferred from the decomposition products. Usually, however, the structure will only be finally elucidated by the reverse method, by building up the molecule from the decomposition products or similar substances, i.e. by what is termed synthesis. Nevertheless, the chemical enigma of Life will not be solved until organic chemistry has mastered another, even more difficult subject, the proteins, in the same*

*way as it has mastered the carbohydrates. It is hence understandable that the organic and physiological chemists are increasingly turning their attention to it. ...*

Emil Fischer, Nobel Lecture, December, 12th 1902

---

## 1 Introduction: The Birth of Computational Protein Design

In 1902 Emil Fischer's Nobel lecture [1] presented the idea of protein design (*see exert*). He emphasized that molecules can be elucidated only by the reverse method, namely, design from decomposition products, which in the case of proteins are the amino acids. At the time Fischer stated that proteins are far more difficult than carbohydrates, for which he received the Nobel. Indeed, it was only in 1972 that Chris Anfinsen received a Nobel Prize for the "connection between the amino acid sequence and the biologically active conformation." Anfinsen's famous experiment included denaturing and renaturing ribonuclease A; thus setting the stage for the sequence-structure-function relationships underlying protein science [2]. In 1981 Drexler speculated that it should be possible to design novel proteins and that such proteins could provide a general capability for molecular manipulation [3]. In 1983 Pabo wrote about designing proteins and peptides concluding that it may be difficult to design proteins which carry out a particular function but the use of pre-folded backbone configuration may be useful at this stage [4]. Pabo pointed at the so called *inverse folding problem* of using a known backbone conformation on which new sequences can be applied; thus modifying function. In agreement with Pabo, in 1987 Wodak reviewed the field with the title "computer-aided design in protein engineering" where the key features of CPD were laid out in a manner that is accurate till this very day, and not only in e.g. the Wodak lab's DESIGNER [5, 6] CPD software.

In 1985 DeGrado conducted what should be regarded as the first CPD: a design, synthesis, and characterization of a 17-residue helical peptide that was the tightest calmodulin-binding peptide produced [7]. This first CPD attempt, described in more detail below, includes many of the main features of current CPD including the need to produce and characterize the suggested design, the crosstalk between human and computer input and the iterative feedback process of the CPD scheme to learn and improve the design.

Other early attempts were "computer-aided" by visually inspecting the protein for suggesting specific point mutations. For example, in 1985 Rutter and coworkers replaced two glycines by alanines in the binding site of trypsin, thus altering binding specificity [8].

While DeGrado and others used computer-aided protein design in early days, according to PubMed, the term "protein design" was introduced only in 1986 by Vonderviszt, Matrai, and Simon [9]. As in the talk of Fischer, Simon's paper did not focus on the protein design per se. Rather, they implied the potential use of analysis of protein environment trends as parameterization required

for protein design. It took an additional decade for the term “computational protein design” to enter the literature. In 1997, Dahiyat, Sarisky, and Mayo introduced the term as part of a systematic design of a  $\beta\beta\alpha$  motif (Table 1) in which they designed 20 of the 28 motif residues [18, 19]. Early attempts of CPD often did not use this term despite describing science that is in the core of the CPD field till this very day. In parallel, numerous CPD publications refer to CPD with related terms that relate to protein design but do not focus on the related computational methodology. These include protein design, synthetic biology, rational design, and more.

Of special note is the fuzzy division between “protein design” and CPD as often there is a significant contribution from computational tools to protein designs that are conducted with an expert know-how that is formulated by computation. This review will emphasize attempts of computer-assisted designs but will focus on protein designs in which the computational part is central to the design methodology.

Thus, in a century since Fischer’s visionary Nobel lecture, science has moved from yearning to understanding protein structure by designing it from building blocks to applying a computational general design algorithm. Not less important, protein design is often termed “the inverse folding problem” as the success of using building blocks to fold a protein into a given structure and function is the true proof that folding is well-understood. Consequently, the know-how and success of CPD contribute directly to that of protein structure prediction in healthy and diseased proteins. Within these frameworks, the CPD field is constantly growing into new basic- and applied-scientific research.





Here, rather than providing an overview of methodological components [121, 122], the idea is to present CPD examples in chronological order showing the achievements and pending challenges in a timeline perspective. In other words, rather than providing a grocery list of available computationally assisted protein design, this review is aimed towards presenting the state of the field as it evolves on the chronological milestone road. Taken together, these case-studies encompass the breadth of the CPD field, the plethora of distinct flavors of it as well as the common threads of success and pitfalls computational protein designers are encountered with (Table 1). The concluding remarks focus on the latter; providing scientific questions for years to come.




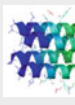


---

## 2 The First Decade of Computational Protein Design, 1985–1994

In 1985 DeGrado, a leading pioneer in protein design, designed with coworkers the tightest-binding peptide inhibitors of calmodulin known till then [7]. Computationally, the 17-residue helical peptide designs included computer-graphics based modeling of the calmodulin target as well as computer modeling [123] of the



**Table 1**  
**Examples of CPD case-studies including the main CPD method and experimental characterization and validation tools utilized to characterize the resulting design. All pictures were drawn with PyMol. In case of NMR models, only the first model is presented. In relevant cases the biological assembly rather than the asymmetric unit is presented. Secondary structures are shown by “dumbbell” cartoon highlighting their irregularities. For clarity, structural water was omitted**

Novelty	Target	Computational methods	Main characterization methods	PDB (Å resolution)	Year, first and last authors, journal, and references	Protein structure (first PDB only)
1. Helical calmodulin-binding peptides	Peptides binding calmodulin	Homology modeling, binding site characterization, iterative synthesis	Calmodulin binding to Melex, Trp fluorescence, CD	–	1985 DeGrado, Cox, <i>J Cell Biochem</i> [7]	
2. De novo design of a four-helix bundle ( <i>Felix</i> )	Antiparallel four-helix bundle (79 residues)	Design rules followed by MD and structural modeling	CD, SEC, Fluorescence	1lfx (theoretical model)	1990 Hecht, Ogden [10]	
3. Grafted metal-binding site	Thioredoxin	DEZYMER	In vivo activity, CD, EPR, absorption, and fluorescence spectroscopy	–	1991 Hellinga, Richards, <i>JMB</i> [11]	
4. Enzyme design for a non-natural substrate	α-lytic protease	Single mutation free energy perturbation (FEP) with molecular mechanics (AMBER) and rotamer library (PROPAK)	Kinetic measurements of design and specificity	–	1991 Wilson, Agard, <i>JMB</i> [12]	
5. Hydrophobic core design	Bacteriophage T4 lysozyme	Modified rotamer library and scoring function including standard force-field and conformational entropy components	X-ray, CD	1l77 (2.05 Å), 2l78 (2.0 Å), 1l79 (1.9 Å), 1l80 (1.8 Å), 1l81 (2.0 Å), 1l82 (2.1 Å)	1992, Hurley, Matthews, <i>JMB</i> [13]	
6. β-sandwich de novo design ( <i>Betadoublet</i> )	β-sandwich protein	SYBYL (Tripos), CHAOS, GCG	CD, NMR H/D exchange	1brd (theoretical)	1994 Quinn, Richradson, <i>PNAS</i> [14]	

7.	Hydrophobic core design	Phage 434 cro protein (five to eight core residues)	Rotamer library + genetic algorithm within Repacking of Core (ROC)	CD, NMR	–	1995 Desjarlais Hnadel, <i>Prot Sci</i> [15]	
8.	Hydrophobic core design	Ubiquitin	Nine designs, each with three to eight core-residue mutations, ROC	CD, Fluorescence, ANS binding, NMR	1ud7 (NMR)	1997, Lazar, Handel, <i>Prot Sci</i> [16] 1999, Johnson, Handel, <i>Structure</i> [17]	
9.	Full sequence design 1 ( <i>IFSD-1</i> )	$\beta\beta\alpha$ motif typified by the zinc finger DNA binding	20 of 28 motif positions were designed avoiding metal and Cys. Template = 1zaa	CD, NMR structure backbone root mean square deviation (RMSD) = 1.04 Å	IFSD, IFSV (NMR)	1997 Dahiyat, Mayo, <i>Science</i> [18, 19]	
10.	Hyper-thermophile protein	Protein G— $\beta$ 1 domain (G $\beta$ 1)	7 mutations selected via ORBIT	CD, AUC, NMR, kinetics (Biacore)	1gb4 (NMR)	1998, Malakauskas, May, <i>Nat Struct Biol</i> [20]	
11.	Coiled-coil oligomers	Right-handed dimer, trimer, and tetramer coiled-coils	Hydrophobic-polar residue patterning and side-chain packing. Comparison to mutations in other locations (representing a less folded state)	X-ray, CD	1rh4 (1.9 Å)	1998, Harbury, Kim, <i>Science</i> [21]	
12.	Three-helix bundle	De novo designed 73-residue, single chain antiparallel three-helix bundle $\alpha 3C$ , $\alpha 3D$	ROC and NBSEARCH for core design, InsightII for modeling, Discover for minimization	NMR, SEC, AUC, CD, Fluorescence, ANS-binding, H-D exchange	2a3d (NMR)	1998 [22] (Design). 1999, Walsh, DeGrado, <i>PNAS</i> [23] (structure)	

(continued)




**Table 1**  
**(continued)**

Novelty	Target	Computational methods	Main characterization methods	PDB (Å resolution)	Year, first and last authors, journal, and references	Protein structure (first PDB only)
13. Coiled-coil mimetic	GCN4-derived mimetic of interleukin-4 two-helix coiled coil	SMD, MD	CD, NMR, SPR	–	1999 Domingues, Serrano, <i>Nat Struct Bio</i> [24]	
14. Dimerization of a monomer (VPA)	Dimerizing protein L-based on a strand-swapped dimer	RosettaDesign 3 positions were design in a eight-residue redesign	CD, X-ray	1jml (1.9 Å)	2001, Kuhlman, Baker, <i>PNAS</i> [25]	
15. β-sheet peptide automatic redesign	Automatic redesign of one to three residues in the <i>betanona</i> β-sheet peptide design producing an array of peptides	PERLA	CD, NMR	NMR available (not deposited to PDB)	2001, Lopez de la Paz, Serrano, <i>JMB</i> [26]	
16. Protozyme enzyme-like protein	Thioredoxin-based PNPA hydrolase	94 non-glycine positions designed via ORBIT	Enzyme kinetics (product production), MS	–	2001, Bolon, Mayo, <i>PNAS</i> [27]	
17. Heterodimeric coiled coils	Six heterodimeric coiled coils	Knowledge-based coiled coil rules, designated rotamer library (sampled via DEE/A*) and minimization with solvation correction	X-ray, CD, AUC,	1kd8 (1.9 Å), 1kd9 (2.1 Å), 1kdd (2.14 Å)	2001, Keating, Kim, <i>PNAS</i> [28]	



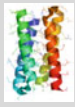

18. Diiron binding ( <i>DFret</i> )	A <sub>2</sub> B <sub>2</sub> four-helix bundle binding diiron	Genetic algorithm fitting to a previous smaller structure which was elongated following coiled-coil parameters. 700,000 sequence design runs and contact energetics relative to competing topologies	CD, AUC, Co <sup>2+</sup> titration, – ferroxidase activity	2002, Summa, DeGrado, <i>JMB</i> [29]
19. Hydrophobic core	13 Spectrin SH3 core redesigns	PERLA, rotamers with 5° sub-rotameric states for $\chi_1$ and $\chi_2$	X-ray, fluorescence 1e6h (2.0 Å), 1e6g (2.3 Å), 1h8k (1.8 Å)	2002, Ventura, Serrano, <i>Nat Struct Biol</i> [30]
20. Binding specificity	Calmodulin binding to smMLCK	Eight residues were mutated via ORBIT modified for protein–protein interactions	CD, binding	2002, 2003 Shifman, Mayo <i>JMB</i> [31], PNAS [32]
21. Antibiotic resistance redesign	Changing $\beta$ -lactamase specificity from ampicillin- to cefotaxime-resistance	19 residues at the interface region were redesigned with PDA, MC/SA and experimental screening of a ~200,000 sequence library	Antibiotic-resistance minimum inhibitory concentration (MIC) assay	2002, Hayes, Dahiyat, PNAS [33]
22. Core stabilization	Human growth hormone (hGH) core stabilization	11 mutations were applied with the 45 target core residues, PDA, new scoring function with backbone and side-chain entropy component and newly weighted other components	CD, cell proliferation assay	2002, Filikov, Dahiyat, <i>Prot Sci</i> [34]
23. Core stabilization	G-CSF core stabilization	10–14 mutations within the 35 redesigned core residues, PDA, homology modeling of template	CD, storage stability, cell-proliferation, assay, biological activity, pharmako-kinetics	2002, Luo, Dahiyat, <i>Prot Sci</i> [35]

(continued)

Table 1  
(continued)



Novelty	Target	Computational methods	Main characterization methods	PDB (Å resolution)	Year, first and last authors, journal, and references	Protein structure (first PDB only)
24. Artificial specific endonuclease ( <i>E-DreI</i> )	Engineered I-Dmol/I-CreI endonuclease where these domains are fused and which targets a long chimeric DNA target	Three-residue inter-domain linker and 14 interface residues identified by computational Ala scanning and re-designed with RosettaDesign	X-ray, lacZ-based solubility assay, biochemical DNA cleavage assay	1 mow (2.4 Å)	2002, Chevalier, Stoddard, <i>Mol Cell</i> [36]	
25. MHC-I binding peptides	Automatic design of nine-residue peptides inhibiting MHC-I	DESIGNER, correlation with a dataset of known MHC-I binding peptides	Fluorescence (peptide-protein binding), T-cell response inhibition (ELISPOT)	–	2003, Ogata, Wodak, <i>JBC</i> [37]	
26. Water solubilization of a membrane protein	Water-soluble pentameric phospholamban	Ten water-facing residues were mutated optimizing pairwise interactions with MC/SA sampling	X-ray, CD, AUC, SEC	1yod (1.8 Å)	2003, Slovic, DeGrado <i>Proat Sci</i> [38] (design) 2005 <i>JMB</i> [39] (structure)	
27. Binding specificity	GON4-derived dimerization specificity	Genetic algorithm for multi-state positive- and negative-design	CD, binding thermodynamics	–	2003, Havranek, Harbury, <i>Nat Struct bio</i> [40]	
28. De novo metal-binding design	De novo four-helix-bundle	88 out of 114 residues were designed by SCADS	AUC, CD, metal-binding, NMR	2hz8 (NMR + QM/MM refinement)	2003 Calhoun, Saven, <i>JMB</i> [41] 2008 <i>Structure</i> [42]	





29.	Novel fold with atomic-level accuracy ( <i>Top7</i> )	93-residue novel $\alpha\beta$ fold with five $\beta$ -strands and two $\alpha$ -helices	Rosetta package: 3D backbone via fragments followed by side-chain MC	CD, X-ray, GuHCl denaturation, NOESY & HSQC NMR	1 qys (2.5 Å)	2003 Kuhlman, Baker, <i>Science</i> [43]	
30.	De novo domain redesign	WW domain (three antiparallel $\beta$ -strands)	SPANS (SPA for numerous states)	CD, NMR	–	2003, Kraemer-Pecore, Desjarlais, <i>Prot Sci</i> [44]	
31.	Protein redesign (increased folding kinetics and thermostability)	Human procarboxypeptidase A2 and 8 other folds	RosettaDesign	X-ray, NMR, CD, AUC, GuHCL	1 vjq (2.1 Å) 2 gtf (NMR)	2003, 2007, Dantas, Baker <i>JMB</i> [45, 46]	
32.	Enzyme-design	Phenol-oxidase	Manual binding pocket sculpting in CPD protein	CD, size-exclusion chromatography, binding kinetics	1 jmb (2.2 Å) (template-like design)	2002, 2004, Kaplan, DeGrado <i>PNAS</i> [29, 47, 48]	
33.	PPI specificity	PPI specificity in DNase (colicin E7) -inhibitor protein (immunity protein Im7) pairs	1. Modeling point mutations at interface sights of both binding partners and then combining the single point mutations. 2. Sampling alternate rigid-body orientations followed by interface hydrogen bond design facilitated by previous structure	X-ray, SPR, fluorescence, toxicity plate assay, DNase activity	1 uiz (2.1 Å), 2 erh (2.0 Å)	2004, Kortemme, Baker <i>NSMB</i> and 2006 Joachimiak, Baker <i>JMB</i> [49, 50]	
34.	Water-soluble potassium channel	Water-soluble KcsA potassium channel	SCADS, 35 solvent-exposed residues were designed with an environmental energy fit to solvent-exposure	CD, AUC, SEC, toxin-binding	–	2004, Slovic, DeGrado, <i>PNAS</i> [51]	

(continued)


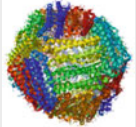
**Table 1**  
**(continued)**

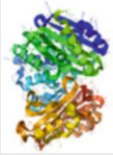
Novelty	Target	Computational methods	Main characterization methods	PDB (Å resolution)	Year, first and last authors, journal, and references	Protein structure (first PDB only)
35. Thermostabilization of an enzyme	Thermostabilization of homodimeric hydrolase enzyme yeast cytosine deaminase (yCD)	Rosetta redesign of residues not in the dimer interface or catalytic site	X-ray, CD, kinetic measurements	1ysd (1.9 Å), 1ysb (1.7 Å)	2005, Korkegian, Stoddard, <i>Science</i> [52]	
36. Negative design for reengineering a homodimer to a heterodimer	SspB homodimer	ORBIT, modified by capping unfavorable vdW energetics and adding an MC negative-design module	X-ray, GuHCl-denaturation, chromatography dimerization assay	1zzz (2.0 Å)	2005, Bolon, Sauer, <i>PNAS</i> [53]	
37. Redox-active rubredoxin mimic	Minimal (40-residue) de novo designed rubredoxin mimic	SCADS	UV-vis, CD	–	2005, Nanda, DeGrado, <i>JACS</i> [54]	
38. Nonbiological cofactor binding	Four-helix bundle that selectively binds two DPP-Fe	Backbone design following different constraints followed by three rounds of SCADS sequence design	CD, SEC, AUC, UV-Vis, EPR	–	2005, Cochran, DeGrado, <i>JACS</i> [55]	
39. Ferritin-like protein surface hydrophobicity	12-subunit Dps ferritin-like protein. Each four-helix bundle subunit underwent three to ten hydrophilic → hydrophobic mutations	SCADS modified for symmetry (homooligomers)	CD, MS, SEC, iron mineralization, Dynamic light scattering	–	2006, Dmochowski, Swift, <i>JACS</i> [56]	
40. Binding selectivity design	TRAIL-variant for selective DR5 binding	WHATIF, FOLD-X	SPR, FACS	–	2006, van der Sloot, Quax <i>PNAS</i> [57]	

41.	Endonuclease DNA binding and specificity redesign	Redesign of intron-encoded homing endonuclease I-MsoI	RosettaDesign protocol of MC optimization of discrete side-chain conformations and identities with continuous minimization of protein and DNA dihedral angles	X-ray, competitive cleavage assay, binding assay	2fld (2.0 Å)	2006, Ashworth, Baker, <i>Nature</i> [58]	
42.	Antibody Fc variants	Antibody Fc variants with enhanced Fcγ-receptor-mediated effector function. Four sites were mutated	Protein design automation (PDA), Sequence prediction algorithm (SPA)	SPR, ADCC assay, B-cell depletion (in macaques), AlphaScreen	–	2006, Lazar, Dahiya, <i>PNAS</i> [59]	
43.	Protein–protein interface	Protein G—β1 domain (Gβ1)	24 interface residues, side-chain-pruned docking of 18 complexes to derive docking parameterization, ORBIT, RESCLASS	HSQC NMR, AUC, dissociation kinetics	–	2007, Huang, Mayo, <i>Prot Sci</i> [60]	
44.	Anti-transmembrane helix peptide (CHAMP)	Computed Helical Anti Membrane Protein Peptide binding α <sub>1b</sub> and α <sub>2</sub> integrin subunits	Protead, scwrl3, GROMOS minimization, HELANAL	CD, ATR-IR, FRET, DN-TOXCAT, Hemolysis, platelet aggregation, and adhesion	–	2007, Yin, DeGrado, <i>Science</i> [61]	
45.	De novo loop design	Tenascin-C ten-residue loop	Ten-residue loop design via Rosetta, loop fragments, and filtering	X-ray, <sup>1</sup> H NMR, CD	2rb8 (1.45 Å), 2rb1 (2.1 Å)	2007, Hu, Kuhlman, <i>PNAS</i> [62]	
46.	De novo design of a erythropoietin (EPO) receptor binder	Design by grafting three binding residues from the EPO—EPO receptor complex and grafting them on a new scaffold	PDB search for scaffold supporting the grafted residues, filtering by protein–protein interface native-like measurables	CD, SPR, Luciferase Reporter Assay	–	2007, Liu, Lai, <i>PNAS</i> [63]	

(continued)

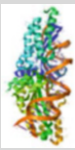
**Table 1**  
(continued)

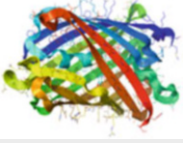

Novelty	Target	Computational methods	Main characterization methods	PDB (Å resolution)	Year, first and last authors, journal, and references	Protein structure (first PDB only)
47. De novo thermostable redesign	51-residue <i>Drosophila</i> engrailed homeodomains (two designs)	1. MC with simulated annealing. 2. FASTER with HERO search	CD, NMR	2p6J (NMR)	2007, Shah, Mayo, <i>JMB</i> [64]	
48. Single-chain porphyrin-binding four-helix bundle	108-residue asymmetric diphenyl-porphyrin metalloprotein	SCADS, STITCH, MC/SA, MD	NMR, CD, UV-vis, SEC, AUC	–	2007, Bender, DeGrado, <i>JACS</i> [65]	
49. Antibody affinity	Electrostatics-based ranking of antibody affinity improvement validated on anti-epidermal growth factor and anti-lysozyme antibodies	DEE/A* sampling of rigid backbone. Bound-case binding energetics and unbound-case approximating flexible binding. Poisson-Boltzmann continuum electrostatics and continuum solvent van der Waals were used for design success ranking	Antibodies displayed on yeast surface with binding affinity measured by incubating cells with varying antigen concentrations	–	2007, Lippow, Tidor, <i>Nature Biotech</i> [66]	
50. PPI interface	TEM-BLIP	PDBmodDesign	SPR, double-mutant cycles	–	2008, Potapov, Schreiber, <i>JMB</i> [67]	
51. Nobel metal binding in Ferritin	Human H ferritin designed to bind Noble metals that form nanoparticles	Four inner-surface and four outer-surface mutations in each of the 24 subunits, SCADS modified for symmetry	X-ray, MS, CD, DLS, SPR, transmission electron microscopy (TEM)	3erz (3.06 Å, Hg <sup>2+</sup> complex), 2z6m (2.72 Å, apoferritin), 3es3 (2.8 Å Au <sup>3+</sup> complex)	2008, Butts, Dmochowski, <i>Biochemistry</i> [68]	

52. PPI interface	SHV-1-BLIP	EGAD two-state minimizer DEE for bound and unbound with MC followed by heuristic step and minimization	X-ray, kinetics measurements	3c4p (1.7 Å) 3c4o (1.7 Å)	2008, Renolds, Handel, <i>JMB</i> [69]	
53. Coiled-coils with four metallo-porphyrin arrays	55-residue four-helix coiled-coil binding four nonbiological iron porphyrins	Three heptad repeats were added to the previous design of a 34-residue peptide assembling to a coiled coil. Specific mutations were introduced to maintain an antiparallel orientation	SEC, CD, binding stoichiometry, AUC, EPR	–	2008, McAllister, DeGrado, <i>JACS</i> [70]	
54. Enzyme-design	Retro-aldolase	RosettaDesign & RosettaMatch	X-ray, array of enzyme kinetics assays	3b5v, superseded by 3hoj (2.2 Å) and 3b5l (1.8 Å)	2008, Jiang, Baker, <i>Science</i> [71]	
55. Enzyme-design	Kemp-eliminase	RosettaDesign & RosettaMatch	X-ray, array of enzyme kinetics assays, in-vitro evolution	2rkx (2.25 Å)	2008, Rothlisberger, Tawfik, Baker, <i>Nature</i> [72]	
56. Loop remodeling for altering enzyme specificity	Guanine deaminase	Two to five residue loop modeling fitting a hypothetical enzyme-ligand transition state		3c0l (2.37 Å)	2008, Murphy, Baker, <i>PNAS</i> [73]	

(continued)

**Table 1**  
**(continued)**

<b>Novelty</b>	<b>Target</b>	<b>Computational methods</b>	<b>Main characterization methods</b>	<b>PDB (Å resolution)</b>	<b>Year, first and last authors, journal, and references</b>	<b>Protein structure (first PDB only)</b>
57. Altering binding specificity	Calmodulin–CamKII interaction	18 residue redesign with ORBIT focusing on inter-molecular interactions	CD, SPR	–	2009, Yosef, Shifman, <i>JMB</i> [74]	
58. Binding peptides with high specificity	bZIP-binding peptides including against c-Jun, c-Fox, and c-Maf	CLASSY, negative design	Protein arrays	–	2009, Grigoryan, Keating, <i>Nature</i> [75]	
59. Protein–DNA interaction specificity	Monomeric homing endonuclease I-AniI	Eight designs, each with up to six residues redesigned by Rosetta with flexible backbone and a genetic algorithm	Kinetic assays, fluorescence competition binding assay	–	2009, Thyme, Baker, <i>Nature</i> [76]	
60. Altered nonribosomal peptide synthetase enzyme gramicidin S synthetase A (GrsA-PheA) activity	Phenylalanine adenylation domain of GrsA-PheA for a set of noncognate substrates (Leu and charged residues)	K* including active-site (two mutations) and nonactive site “bolstering” mutations (up to two mutations)	Spectrophotometric activity assay	–	2009, Chen, Donald, <i>PNAS</i> [77]	
61. Binder of an abiological chromophore	A <sub>2</sub> B <sub>2</sub> four-helix bundle that selectively binds two emissive abiological (porphinato)/zinc chromophores of DPP-Zn	SCADS	CD, SEC, AUC, transient absorption spectroscopy, fluorescence lifetime measurements	–	2010, Fry, Therien, <i>JACS</i> [78]	
62. Protein–DNA interaction specificity	I-Msol homing endonuclease	Six residues redesigned by Rosetta, loop-closure algorithm, and a genetic algorithm	X-ray, DNA-cleavage assay	3mip (2.4 Å), 3mis (2.3 Å), and 3ko2 (2.9 Å)	2010, Ashworth, Baker, <i>NAR</i> [79]	

63.	Generation of longer emission wavelength in red fluorescent protein	Red fluorescent protein mCherry	13 residues redesigned via Phoenix and FASTER. The CPD was used to generate focused combinatorial libraries	mRojoA, mRouge, MS, site-directed mutagenesis, library-screening, spectroscopic characterization	3nez (1.7 Å), 3ned (0.95 Å)	2010, Chica, Mayo, <i>PNAS</i> [80]	
64.	TM diporphyrin-binding complex <i>PRIME</i> (PorPhyrins in Membrane)	24-residue TM helices assembling to a four-helix-bundle binding two diphenyl-porphyrins ( $\text{Fe}^{3+}\text{DPP}^{\text{r}}$ 's)	Backbone template of an idealized porphyrin-binding four-helix bundle, extended energy-based conformer library, DEE search followed by MC/SCMF. Pairwise energies calculated with the addition of the Lazaridis implicit membrane solvation (IMM1), model ranking by helix-helix interaction energies	Characterization in a micelle using absorption spectroscopy, CD, and EPR	–	2010, Korendovich, DeGrado, <i>JACS</i> [81]	
65.	Protein-protein interaction (PPI)	Hyperplastic discs protein binding to P21-activated kinase 1 kinase domain PAK1	Rosetta-based DDMI protocol (Dock, Design, Minimize Interface)	NMR, CD, fluorescence polarization	(Chemical shifts ID 1670 in BMRB DB)	2010, Jha, Kuhlman, <i>JMB</i> [82]	
66.	Multistate design for stabilization of a protein core	Gp1, the $\beta 1$ domain of <i>Sreppococcal</i> protein G	Ten core positions underwent multistate CPD using NMR-derived ensemble templates. FASTER, CLEARSS library design	Combinatorial library GdmCl-based stability determination	–	2010, Allen, Mayo, <i>PNAS</i> [83]	
67.	Prediction of resistance mutations	MRSA DHFR resistance mutations (maintaining function without inhibitor binding)	10 active-site residues subjected to K* CPD algorithm. Four mutants experimentally validated	X-ray, CD, enzyme kinetics	3LG4 (V31Y/F92I double mutant, 3.15 Å)	2010, Frey, Anderson, <i>PNAS</i> [84]	

(continued)

Table 1  
(continued)


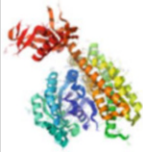

Novelty	Target	Computational methods	Main characterization methods	PDB (Å resolution)	Year, first and last authors, journal, and references	Protein structure (first PDB only)
68. Carbon nanotube surface coating ( <i>HexCoil-Gly</i> , <i>HexCoil-Ala</i> , <i>DSD-Gly</i> , <i>DSD-Ala</i> )	Single-walled carbon nanotube coating with hexameric coiled coil	Three selection rules for backbone choice followed by side-chain design of DEE/A*, minimization, MC/SA	X-ray, CD, AUC, SEC, Absorption spectroscopy, TEM	3s0r (2.44 Å)	2011, Grigoryan, DeGrado, <i>Science</i> [85]	
69. High-affinity protein binder ( <i>HB36</i> , <i>HB80</i> )	Influenza Hemagglutinin binder	Target-surface scaffold selection, RosettaDock, RosettaDesign + directed evolution	X-ray, SPR, Bio-layer interferometry	3z2x (3.1 Å)	2011, Fleishman, Baker, <i>Science</i> [86]	
70. De novo design of a binding pair	Ankryn-repeat-based Tyr-Tyr binding between Prb and Pdar.	RosettaDesign, PatchDock, motif search, experimental affinity maturation	X-ray, ELISA, SPR, fluorescence polarization, DLS, CD, NMR	3q9n (2.0 Å), 3q9u (2.3 Å), 3qa9 (Prb only, 1.9 Å)	2011, Karanicolas, Baker, <i>Mol Cell</i> [87]	
71. PPI; β-strand assembly ( <i>βdimer1</i> )	Homodimer of γ-adaptin	Rosetta-based DDMI including five rounds of interface sequence design and minimization	X-ray, AUC, size-exclusion chromatography (SEC), binding-affinity	3zy7 (1.09 Å)	2011, Stranges, Kuhlman, <i>PNAS</i> [88]	
72. Antigen design	HIV 4E10 epitope	Flexible backbone remodeling and resurfacing	CD, SPR, ELISA	–	2011, Correia, Schief <i>JMB</i> [89]	

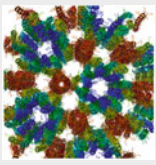
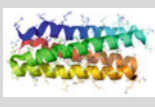


73.	Thermostable terpene synthase	Thermostable tobacco 5-epi-aristolochene synthase	SCADS, 12 mutations which were all >12 Å from the active site	CD, GC-MS activity assay	–	2011, Diaz, Weiss, <i>Prot Sci</i> [90]
74.	Collagen heterotrimer	Collagen A:B:C-type heterotrimer	Positive and negative design for Pro and hydroxyl-Pro inclusion in triplets and energy gap between model and competing structures. MC/SA	CD	–	2011, Xu, Nanda, <i>JACS</i> [91]
75.	Enzyme-design by single mutation ( <i>AllylCat</i> )	Kemp-eliminase	Single-site scanning mutagenesis, docking, and super-rotamer modeling of transition state	NMR, CD, enzyme kinetics with pH-profile, CD,	2kz2 (NMR)	2011, Korendovych, DeGrado, <i>PNAS</i> [92]
76.	Enzyme-design by few mutations	Kemp-eliminase	Mutation modeling and docking	CD, X-ray, enzyme kinetics	4e97 (1.3 Å), 4ekp (1.64 Å), 4ekq (1.54 Å), 4ekr (1.49 Å), 4eks (1.64 Å)	2012, Merski, Shoichet, <i>PNAS</i> [93]
77.	Solubilization of the TM domain of nicotinic acetylcholine receptor	TM domain of nicotinic acetylcholine receptor $\alpha 1$ subunit	SCADS and loop modeling via MODELLER. Overall 21 residues in the TM region and loops were redesigned	LC/MS, CD, NMR	2lkq (NMR, major conformation), 2lkh (NMR, minor conformation)	2012, Cui, Xu, <i>BBA</i> [94]
78.	Redesign of a mononuclear Zn metalloenzyme for organo-phosphate hydrolysis	De novo metalloenzyme design for the hydrolysis of the R <sub>p</sub> isomer of a coumarinyl analog of the nerve agent cyclosarin	PDB scaffold search, RosettaMatch, RosettaDesign	X-ray, absorption spectra kinetic measurements	3tlg (2.35 Å)	2012, Khare, Baker, <i>Nat Chem Bio</i> [95]

(continued)

Table 1  
(continued)

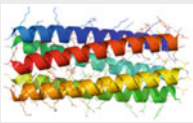
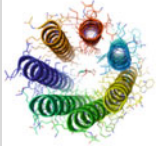


Novelty	Target	Computational methods	Main characterization methods	PDB (Å resolution)	Year, first and last authors, journal, and references	Protein structure (first PDB only)
79. High-affinity epitope scaffold	HIV 2F5 epitope design using backbone-grafting onto scaffold proteins	Six- and seven-residue epitope design via Rosetta, Epitope fragment search in the PDB, loop closure via cyclic coordinate descent (CCD) and MC	X-ray, CD, static light-scattering, SPR	3rj (2.3 Å), 3ri0 (2.25–2.8 Å), 3rhu (Å), 3rfn (1.8 Å)	2012, Azóitei, Schief, <i>JMB</i> [96]	
80. Protein–protein orthogonal pair within a signaling circuitry	GTPase (Cdc42) redesigned to be activated by a designed GEF intersectin partner	RosettaBackrub, KIC, Rosetta	X-ray, CD, SPR, WASP fluorescence titration, GAP assay, ELISA, live-cell fluorescence microscopy, and XCELLigence Assay	3qbv (2.65 Å)	2012, Kapp, Kortemme, <i>PNAS</i> [97]	
81. Ideal folds from secondary structure patterns	Five different-topology folds	New secondary structure connectivity rules, negative rules, and RosettaDesign (MC/SA) and RosettaHoles	NMR, CD, SEC-MALS	NMR structures: 2kl8, 2lv8, 2ln3, 2lvb, 2lta	2012, Koga, Baker, <i>Nature</i> [98]	
82. Self-assembling, register-specific collagen hetero-trimers	Minimalistic de novo design of self-assembling, register-specific collagen hetero-trimers	Novel scoring function and genetic algorithm search procedure, focusing on axial inter-strand ionic interactions	CD, NMR	–	2002, Fallas, Hargernic, <i>Nat Comm</i> [99]	

83.	Enzyme design ( <i>HG-1</i> , <i>HG-2</i> , <i>HG-3</i> and directed evolution <i>HG-3.17</i> )	Kemp eliminates	QM, MD, iterative approach	CD, MS, X-ray, enzyme kinetics	3O2L (2.0 Å), 3NYD (1.23 Å), 3NYZ (1.51 Å), 3NZ1 (1.56 Å), 4bs0 (1.09 Å)	2012, Privett, Mayo <i>PNAS</i> [100] Directed evolution to HG-3.17 by Blomberg, Hilvert <i>Nature</i> [101]	
84.	Iron-sulfur [4Fe-4S] cluster protein	Four-helix coiled coil iron-sulfur protein ( <i>CCIS</i> ) monomer	ProtCad with metal-binding first approach	CD, EPR, SEC	–	2012, Gryb, Noy, <i>Biochim Biophys Acta</i> [102]	
85.	Crystal structure space group de novo design	P6-space group crystal design from a three-helix 26-residues coiled-coil	Construction of a 'grid-like' ensemble of coiled coil P6-structures followed by side-chain design	X-ray, CD	4dac (2.1 Å), 3v86 (2.91 Å)	2012, Lanci, Saven, <i>PNAS</i> [103]	
86.	Altered functionality of a de novo design	Altering the oxidation of hydroquinones to hydroxylation of arylamines in a designed four-helix bundle	MSL	NMR, CD	2lfid (NMR)	2012, Reig, DeGrado, <i>Nature Chem</i> [104]	
87.	Conformational switch ( <i>UniRapR</i> )	Rapamycin-regulated single chain of FKB12 and FKB12-Rapamycin binding protein	Medusa, replica exchange and discrete molecular dynamics	Immunoprecipitation, live-cell and live zebrafish imaging	–	2013, Dagliyan, Dokholyan, <i>PNAS</i> [105]	
88.	Binder of a hyperpolarizable abiological chromophore ( <i>SCRPPZ-1</i> , <i>SCRPPZ-2</i> , <i>SCRPPZ-3</i> )	Four helix bundle binding a ruthenium-zinc abiological hyperpolarizable chromophore	SCADS, loop grafting, MD	CD, Pump-probe absorption electronic spectroscopy, Hyper-Rayleigh light scattering, X-ray reflectivity	–	2013, Fry, Therien, <i>JACS</i> [106]	

(continued)


Table 1  
(continued)

Novelty	Target	Computational methods	Main characterization methods	PDB (Å resolution)	Year, first and last authors, journal, and references	Protein structure (first PDB only)
89. GPCR solubilization and application as a biosensor	μ-opioid receptor	SCADS, comparative modeling, solubilized by mutating 53 membrane-exposed residues	CD, MS, homogeneous tim-resolved fluorescence (HTRF)	–	2013, Perez-Aguilar, Liu, <i>PLoS one</i> [107], (2014 biosensor application [108])	
90. De novo lysozyme protein inhibitor	Hen egg lysozyme centric active site protein binder	Target-surface scaffold selection, RosettaDock, RosettaDesign + directed evolution	X-ray, SEC, <sup>19</sup> F NMR, SEC, HEL activity assay, Yeast surface display	3vb8 (2.9 Å)	2013, Procko, Baker, <i>JMB</i> [109]	
91. De novo high-affinity and selective ligand-binding protein	Selective, high-affinity binder of the steroid digoxigenin	Ligand conformer library, RosettaMatch, RosettaDesign, PDB scaffold scan, CCP4 (shape complementarity)	X-ray, SEC, CD, AUC, fluorescence polarization, isothermal titration calorimetry (ITC), yeast surface display	4j8t (2.05 Å), 4j9a (3.2 Å)	2013, Tinberg, Baker, <i>Nature</i> [110]	
92. Hot-spot centric de novo pH-sensitive IgG binding protein	Fc-domain His-433 pH-sensitive IgG binding protein	RosettaMatch, RosettaDesign, PDB scaffold scan	CD, ELISA, directed evolution	–	2014, Strauch, Baker, <i>PNAS</i> [111]	
93. Statistical energy function (SEF) boosted by experimental selection for foldability for automated CPD	Novel SEF exemplified by de novo CPD of a helical and mixed topology protein	New SEF with selection of structure neighbors with adaptive criteria (SSNAC) to address the multidimensional properties jointly, MC/SA minimization	CD, NMR, antibiotic resistance	2mlb (NMR), 2mn4 (NMR)	2014, Xiong, Liu, <i>Nat Comm</i> [112]	

94. Helical bundles	Helical bundles including an antiparallel, untwisted three-helix bundle with 80-residue helices, an antiparallel right-handed four-helix bundle and a parallel left-handed five-helix bundle	Parametric backbone generation and Rosetta	CD, X-ray, EM	4not (five-helix bundle, 1.69 Å), 4tql (three-helix bundle, 2.8 Å), 4uos (four-helix bundle, 1.63 Å)	2014, Huang, Baker, <i>Science</i> [113]	
95. Water-soluble α-helical barrels	Pentameric, hexameric and heptameric Water-soluble α-helical barrels	CCBuilder, SOCKET, PoreWalker, coiled-coil assembly rules	X-ray, CD, AUC	4pna (2.0 Å), 4pn8 (2.0 Å), 4pn9 (2.0 Å), 4pnb (2.0 Å), 4pnd (2.0 Å)	2014, Thomson, Woolfson, <i>Science</i> [114]	
96. Synthetic coiled-coils	Antiparallel homodimeric coiled-coils with no cross-specificity	DFIRE, CCCP structure generator, CLASSY	AUC, disulfide-exchange, CD	–	2014, Negrod, Keating, <i>JACS</i> [115]	
97. Epitope-focused vaccine design	RSV helix-turn-helix epitope in a three-helix bundle scaffold	Fold-from-loops (FFL) protocol in Rosetta	X-ray, NMR-HSQC, CD, SPR, ELISA (on immunized macaques)	4L8I (2.0 Å), 4JLR (2.71 Å), 4N9G (2.5 Å)	2014, Correia, Schief, <i>Nature</i> [116]	
98. TM transporter ( <i>Rocker</i> )	De novo TM Zn <sup>2+</sup> transporting four-helix bundle	MaDCaT, Ez, MD, Volocity	X-ray, ssNMR ( <sup>1</sup> H, <sup>13</sup> C cross-polarization MAS, <sup>19</sup> F-CODEX), AUC, Liposome flux	4p6j (2.8 Å), 4p6k (2.7 Å), 4p6l (2.8 Å), 2muz (NMR)	2014, Joh, DeGrado <i>Science</i> [117]	

(continued)

Table 1  
(continued)

Novelty	Target	Computational methods	Main characterization methods	PDB (Å resolution)	Year, first and last authors, journal, and references	Protein structure (first PDB only)
99. De-immunization	T-cell epitope removal from different proteins	T-cell epitope identification by SVM, redesign by Rosetta	Flow cytometry, fluorescence	–	2014, King, Baker, <i>PNAS</i> [118]	
100. Self-assembling symmetry	β-propeller with perfect six-bladed symmetry	Visual template datamining, RosettaDock, phylogenetic ancestral reconstruction, circular permutation for the ‘Velcro’ strap	X-ray, CD, AUC, MS, Differential Scanning Fluorimetry	3vw7 (1.7 Å), 3vww8 (1.4 Å), 3vww9 (1.33 Å), 3vwwa (1.99 Å), 3vwwb (1.7 Å), 3vwwf (1.6 Å)	2014, Voet, Tame, <i>PNAS</i> [119]	
101. Repeat protein	Leucine-rich repeat protein from ribonuclease inhibitor family	Rosetta, new code for adjusting repeat geometry	CD, FTIR, NMR HSQC, AUC	–	2014, Ramisch, Andre, <i>PNAS</i> [120]	

calmodulin–peptide interaction focusing on electrostatic potential surfaces and structural modeling. These included side-chain positioning using geometries taken from a known homologous structure of an intestinal calcium-binding protein, interactive computer graphics, and minimization using the AMBER [124] force-field. The acquired know-how of the calmodulin-peptide structure and binding characterization was tested by iterative peptide synthesis and characterization. Hence, this early attempt of CPD underscores the need to integrate all available know-how and methods for the requested target as well as the need to combine theory and experiment in an interactive and iterative feedback loop.

In 1990 Hecht and Ogden and the Jane and David Richardson lab designed a de novo four-helix bundle, termed *Felix* [10]. This is an example in which protein design rather than CPD was the leading method. Even for designing the hydrophobic core, the authors write that: “*Space-filling models of Felix were constructed and the sequence was then modified to remove lumps or fill holes. This is easier to do with physical models than on the computer.*” Computationally, several structures were modeled followed by application of molecular dynamics (MD). Positive- and negative-design rules were conducted manually, including for residues preferring helicity, for the radial distribution of hydrophobicity along the helices and for helix capping. Hence, this case-study proves that it is required not only to focus on the requested design combining existing and newly found parameterization, but rather attention should be devoted to the so-called negative design of avoiding unwanted designs.

In 1991 Hellinga and coworkers used CPD software aimed at sites with predefined geometry (DEZYMER [125]). They introduced a copper-binding site into thioredoxin by mutating four amino acids [11]. In the analysis of the design they concluded that two residues are pivotal for the metal ligation while the two other are pivotal for removing alternative modes of binding, thus highlighting the need to focus on negative design.

In 1991 Wilson, Mace and Agard presented a generalized model for altering substrate specificity [12]. Using a  $\Delta\Delta G$  free energy perturbation approach, the free energy of the free substrate, free enzyme, and complex were computed separately as to non-bonded and solvation energetics over the different potential conformations suggested by the PROPAK [126] rotamer-library based CPD software. The approach was tested using a protease in which the specificity for cleaving leucine was raised by three orders of magnitude following a single mutation. While this CPD example entails merely a single mutation, the components of the approach include many of the later CPD methodology.

In 1992 Hurley and Matthews redesigned the core of bacteriophage T4 lysozyme [13]. This case-study, coming from the lab most known for thoroughly studying the effect of mutation on

protein structure and function, includes several insights. Only nine solvent inaccessible amino acids were subjected to redesign. Moreover, a core valine residue was not part of the redesign as it binds structural water. The repacking was limited to residues that are more hydrophobic compared to the *wild-type* residues. In addition, as all potential sites occur in  $\alpha$ -helical regions, no net increase in the number of  $\beta$ -branched amino acids (Val and Ile) was allowed. While each addition of a  $\beta$ -branched amino acid to a helix has a small energetic cost of less than 0.5 kcal/mol, it was feared that the accumulation of such residues will destabilize the structure. For packing calculations, the Ponder and Richards rotamer library [126] was used truncating rare (<5 %) rotamer conformations. Hydrogens were omitted and reduced van der Waals radius was applied to account for local relaxation. The free energy was calculated with a standard local minimization as well as a component accounting for the loss of side-chain conformational entropy. Four amino-acids were mutated showing a similar stability compared to the template structure (0.5 kcal/mol destabilization). The destabilization of each single mutation was much larger thus showing the overall cooperative nature of the overall core repacking design.

In 1994 Jane and David Richardson, de novo designed *beta-doublet*, a  $\beta$ -sandwich protein [14]. It is no surprise that such an endeavor came from pioneers in visualization (Richardson diagram, also known as ribbon diagram), parameterization, and quality control of protein structures. A four-stranded  $\beta$ -sheet dimer designed from scratch included an intersubunit disulfide bridge. Internal side chains were chosen for their statistical preference for  $\beta$ -sheet formation and their ability to tightly pack in a protein core. This knowledge-based parameterization was corroborated by side-chain repacking of rotamers. This design scheme focused on negative design, specifically disfavoring the Greek Key topology. To minimize alternative folding modes, turns were shortened as much as possible. Binding of 1-anilinonaphthalene-8-sulfonate (ANS) was higher, compared to binding to well-folded proteins. Along with low unfolding cooperativity and poor NMR characteristics, this may indicate a loosely packed hydrophobic core or even a molten-globule structure; highlighting the challenge of obtaining thermostable de novo designed proteins, let alone those composed of  $\beta$ -sheets.

---

### 3 The Second Decade of CPD, 1995–2004

Setting the framework for CPD, in 1995 DeGrado and coworkers reviewed the hierarchic approach to protein design including helix stabilization, coiled coils, four helix bundles,  $\beta$ -sheets, mixed  $\alpha$ - $\beta$  structures, DNA-binding proteins, and functional proteins [127].



The presented approach emphasized the need for quantitative parameterization of the various levels of structure and function within the design target. Such parameterization can be either physics-based or knowledge-based. In either ways, it should be integrated into quantitative potential (scoring) functions.

In 1995 Desjarlais and Handel presented a novel computational framework for the de novo design of hydrophobic cores [15]. The CPD was conducted via the Repacking of Core (ROC) program, later developed to their Sequence Prediction Algorithm (SPA) [128]. The approach included two steps—a custom-made rotamer library for hydrophobic residues (Val, Ile, Leu, Phe, and Trp) and a genetic algorithm (GA) for optimizing sequence and structure space of the designed protein. The method was exemplified on the phage 434 Cro helical protein with five to eight amino acid changes in the hydrophobic core. Two of the three attempted designs resulted in a stable protein. This first study into a pivotal protein region helped to substantiate the notion that the noncore residues of a protein play a role in determining the uniqueness of the folded structure [15].

In 1997 Desjarlais and Handel applied their ROC program for the stabilization of a mainly  $\beta$ -sheet protein, ubiquitin [16]. Nine designs with three to eight mutations each were experimentally characterized. Unlike their 434 Cro [15] redesign, all ubiquitin designs were less stable relative to the *wild-type* protein. The authors postulate that this may be due to the fact that in contrast to the  $\alpha$ -helical 434 Cro protein, ubiquitin is mainly composed of  $\beta$ -strand secondary structures which may dictate more stringent packing requirements. One of the designs was structurally elucidated confirming that the core side-chains had less favorable conformations and higher flexibility compared to the *wild-type* [17].

In 1997 Dahiyat and Mayo opened the field of full-protein fully automated computational de novo protein design [18, 19]. The CPD scheme was termed ORBIT [18] for Optimization of Rotamers by Iterative Techniques. The so called full sequence design 1 (*FSD-1*) was not a typical protein of over 200 amino acids, but rather a small, 28-residue sequence; a length considered a peptide rather than a protein. Nevertheless, the remarkable achievement included a complex  $\beta\beta\alpha$  motif based on the polypeptide structure of a zinc finger domain in which 20 of the 28 residues were subjected to design. Moreover, while such a small DNA-binding motif is folded in nature with the aid of a zinc ion, the zinc-ligating residues (two cysteines and two histidines) were replaced in the design with two phenylalanines, an alanine, and a lysine without the need for the metal ion. As a side-remark, the use of a charged lysine in such a core position highlights the need to take caution in stigmatizing amino acids as “hydrophobic” or “hydrophilic” as in this case the long hydrophobic neck of this charged residue filled the hydrophobic requirement within this position. The  $1.9 \times 10^{27}$  possible

amino acid sequences were searched by application of the Dead End Elimination (DEE) theorem [129]; highlighting the intertwined connected between CPD and search and sampling methods [130]. FSD-1 displayed low identity to any other existing sequence, thus establishing it as a ‘de novo’ design. In this fixed-backbone design, an existing crystal-structure template was utilized in which eight residues were left as is and the remaining 20 were subjected to design. The hierarchical approach of confining key positions was further confined by considering 7, 10, and 16 optional amino acids for each core, surface, and boundary position, respectively. The backbone dihedral angle further confined two positions to glycine, thus de facto leaving 18 positions for CPD. The combined structure space defined by the accessible backbone-dependent Dunbrack rotamer library [131] applied over the accessible fold space, resulted in  $1.1 \times 10^{62}$  possible rotamer sequences. The experimental validation included Nucleic Magnetic Resonance (NMR) structural elucidation exhibiting 1.98 Å and 0.98 Å C $\alpha$ -atom root means square deviation (RMSD) between the design and the template structure for the full and the core residues (residues 8 to 26), respectively. The difference between these two numbers highlights the intrinsic flexibility and disorder associated with nonsecondary structure elements, especially when positioned at the edge of the protein sequence.

In 1998 the Mayo lab applied the ORBIT [18] for the design of a hyperthermophilic Streptococcal protein G  $\beta$ 1 domain [20]. The stability enhancement stemmed from seven mutations which optimized core packing, increased burial of hydrophobic surface area, more favorable helix dipole interactions, and improvement of secondary structure propensity. The resulting protein displayed a melting temperature above 100 °C and a 4.3 kcal/mol thermodynamic stabilization compared to the *wild-type* at 50 °C. Structure, activity, and binding to an antibody were similar to the *wild-type* structure thus changing only the thermal stability of the protein.

In 1998 the Kim lab designed right-handed coiled coils applying backbone flexibility, hydrophobic-polar residue patterning for the superhelical axis and the hydrophobic core along with modeling of packing [21]. Backbone coordinates were determined by exploring a parametric family of superhelical backbones described originally by Francis Crick. Negative design was applied by mimicking a less-folded state via permutations on the mutation location and calculating the energy gaps to such permutations. Dimeric, trimeric, and tetrameric bundles were designed. The tetramer was structurally resolved exhibiting a striking 0.2 Å RMSD for the core residues.

In 1998 the DeGrado lab de novo designed an antiparallel three-helix bundle,  $\alpha$ 3C, in an iterative process with specific interactions added incrementally [22]. In this design many steps were designed rationally without the aid of the computer. Two rounds of

core design were conducted fully by CPD. A previously designed dimer (CoilSer) that was found to be a trimer was the initial template for the design. In this structure, some hydrophobic Leu residues adopt a less likely rotamer suggesting the availability of better core packing. The trimer was trimmed by one turn. In the first round, GlyAsn and ProGlyAsn loops were added to turn the discrete helices into a single subunit. In the second round, helix capping was introduced and in the third round nonnative characteristics were eliminated by negative design. Specifically, the 17 residues of the hydrophobic core were repacked using 30 runs of ROC followed by 30 runs of ROC for a subset of six residues. Further, to avoid both clockwise and counterclockwise turning of the helices within the trimer, charged residues were designed to cause electrostatic repulsion and favor only one conformation. This is a direct negative design step. Thus, the designed helix capping interactions and electrostatic interactions between partially exposed residues assisted in achieving a unique, native-like structure. In 1999, three surface exposed residues were changed thus designing  $\alpha 3D$  in which the homology between the helices was decreased thus simplifying structural elucidation [23].

In 1999 the Serrano lab redesigned the two-helix coiled-coil interleukin-4 using GCN-4 as a template [24]. This is not a classical CPD case-study but rather a computer-aided sequential rational design where deep understanding of the binding interface enabled grafting of the positive electrostatic convex binding site shape from the four-helix-bundle protein to a new two-helix template. The side-chains of the mutated positions were structurally predicted via the rotamer-library-based software SMD [132]. Interestingly, MD simulations were applied as in silico screening of the mutations prior to decision on experimental characterization. Depending on the size of the interleukin-4 binding site (to interleukin-4 receptor alpha) grafted on the GCN4 template, the binding affinities ranged from 2 mM to 5  $\mu$ M.

In 2001 the Baker lab applied CPD to convert the monomeric protein L to an obligate dimer by just three mutations [25]. The design relied on a  $\beta$ -hairpin single mutation domain swapped dimer in which a  $\beta$ -turn straightens and the C-terminal strand inserts into the  $\beta$ -sheet of the partner. The Rosetta [133] module RosettaDesign [134] focused on an eight-residue region and added just two mutations to the domain swapping mutation resulting in an obligate dimer.

In 2001 the Serrano lab applied PERLA [135] for the redesign of their previously designed 20-residue  $\beta$ -sheet protein *betanova* [136] aiming to create a set of double- and triple-mutations with different folding stabilities so as to compare predicted and experimental folding stabilities [26]. Briefly, PERLA includes a custom-made rotamer library, an all-atom force-field, and a combination of statistical terms including solvation and entropy. Relaxation of the

local strains is achieved by sub-rotamer states and most parameters are balanced with respect to a reference denatured state. DEE is applied to prune the search space and then side-chain conformations are weighted using a mean-field approach. Here, two CPD schemes were applied: First, four positions adjacent to aromatic residues were discretely redesigned aiming at utilizing the Nuclear Overhauser effects (NOEs) between the aromatic residues and the new mutations for evaluating structural effects. Second, multiple-residue mutations were designed with the most promising designed experimentally characterized. Increase in core hydrophobicity or van der Waals contacts stabilized the design. At one site the algorithm did not predict a hairpin destabilization, possibly due to alternative conformations. Alternatively, the sequence of folding events should be taken into account along with the balance between long-range electrostatic interactions and short-range van der Waals interactions.  $\beta$ -sheet propensities were also shown to correlate with stabilization. Some of the mutants stabilized the design by 1 Kcal/mol. Taken together; this early study displays the usage of CPD algorithms for the study of structure–stability relationships and parameterization of their underlying causes.

In 2001 Bolon and Mayo applied ORBIT [18] to computationally design protozymes which are enzyme-like proteins exemplified on a thioredoxin scaffold catalyzing a nucleophilic hydrolysis of *p*-nitrophenol acetate [27]. ORBIT applies a force-field and DEE theorem to compute sequences that are optimal for a given scaffold. The use of an inert scaffold required the design of a new cleft, which was relatively open to the surrounding milieu, thus possibly affecting efficiency. The 94 non-glycine positions reflected  $10^{101}$  rotamer sequences that were scanned using the DEE algorithm within ORBIT [18]. The rate enhancement of  $\sim 25$ -fold ( $K_M = 170 \pm 20 \mu\text{M}$ ,  $k_{\text{cat}} = 4.6 \pm 0.2 \times 10^{-4} \text{ s}^{-1}$ ) is comparable to that of early catalytic antibodies (Table 2).

In 2001 the Kim lab designed six dimeric coiled coils with a range of stabilities by combining knowledge-based rules (specifically the *a* and *d* hydrophobic positions in the heptad repeat), rotamer selection and sampling followed by minimization [28]. The first two parts address the large accessible search space while the last one assists in achieving quantitative estimates of interaction energies. For example, a hydrophobic Val was constrained to the gauche (–) rotamer, which is known to be favored in this position. In parallel to choosing a small subset of rotamers, subrotamers were introduced by including  $\pm 110^\circ$  of the  $\chi_1$  and  $\chi_2$  rotamer positions. Interestingly, to address the difficulty of modeling solvent-exposed charged residues, residues at the *e* and *g* positions of the heptad repeat were truncated beyond the C<sub>8</sub> position. Minimization was carried out without electrostatics but with an explicit hydrogen-bonding term and the overall solvent-exposed residue energetics were later fixed by an empirical solvation correction.

Table 2

CPD approaches towards engineering improved Kemp eliminases. The kinetic data displayed is modified after [100]. An array of protein templates and design schemes were applied towards a common goal, often involving feedback from previous research on this new enzyme target. Likewise, the scope of the CPD ranged from a single mutation to a QM-optimized catalytic site and full-protein template re-engineering

Approach	Template protein	Catalysis $k_{\text{cat}}/K_M$ [ $\text{M}^{-1} \text{min}^{-1}$ ]	Catalysis [( $k_{\text{cat}}/K_M$ )/ $k_{\text{uncat}}$ ]	pH optimum (catalytic residue)	References
Catalytic antibodies	Immunoglobulin	5500	$\sim 5 \times 10^9$ (pH 7.5)	pH > 7.0 (Glu)	[137]
CPD iterative approach (QM, MD, + directed evolution)	Xylanase	430 ( $2.3 \times 10^5$ )	$4 \times 10^8$ (pH 7.25)	pH = 7.0 (Asp) 18	[100, 101]
CPD (Rosetta, KE70 + directed evolution)	Aldolase	126 (54,800)	$1 \times 10^8$ ( $5 \times 10^{10}$ , pH 7.25)	pH > 7.0 (His-Asp dyad)	[138]
CPD (Rosetta, KE07 + directed evolution)	Hisf protein	12.2 (2600)	$1 \times 10^7$ ( $2 \times 10^9$ , pH 7.25)	pH > 7.0 (Glu) 7,20	[72, 139]
CPD (Rosetta, KE59 + directed evolution)	Glycerophosphate synthase	$\sim 160$ (60,430)	$1 \times 10^8$ ( $5 \times 10^{10}$ , pH 7.25)	pH > 8.0 (Glu) 7,21	[72, 140]
AlleyCat—CPD single-site mutation (super-rotamer minimization)	Calmodulin C-terminal domain	5.8	$4 \times 10^5$ (pH 8.0)	pH > 8.0 (Glu) 22	[92]
CPD artificial cavity (most active construct L99A/M102H)	T4-lysozyme	1.8	$7 \times 10^7$ (pH 5.0)	pH = 5.0 (His)	[93]

The propensity of residues to be in helices was also added to the equation. The designed structures displayed an impressive  $<0.7$  Å for all non-hydrogen atoms.

In 2002 the DeGrado lab computationally designed an  $A_2B_2$  four-helix bundle protein binding diiron called *DueFerro tetramer* or *DFtet* [29]. The de novo design focused on the gap between the requested fold and alternative folds thus explicitly incorporating positive- and negative-design considerations. The design was built using a template of a previous design which was then elongated to increase stability by extending the four-helix bundle Crick parameters. Residues were chosen to increase helical propensity, stabilize one of the competing topologies via computing contact energetics. The best four designs following 700,000 iterations of sequence design were modeled structurally and the best design was validated experimentally.

In 2002 the Serrano lab de novo designed 13 divergent spectrin SH3 core sequences to determine their folding properties [30]. The PERLA-based [135] redesign included nine nonconsecutive positions resulting in a larger buried hydrophobic volume. The computational design over-packed the core resulting in an expansion of the  $\beta$ -barrel. This was further validated by conducting Ile  $\rightarrow$  Val mutations which all resulted in strain removal and stabilization. Eleven of the 13 designs folded well with similar characteristics to the folded *wild-type*. Two structurally resolved designs were similar to the *wild-type* with small changes at a loop region following discrepancies at the  $\chi_2$  side-chain positions relative to the design.

In 2002 Shifman and Mayo modulated calmodulin binding specificity by CPD [31]. The calmodulin binding interface was optimized to improve binding specificity towards one of its natural targets, smooth muscle myosin light chain kinase (*smMLCK*). ORBIT [18] considered  $10^{22}$  sequences to optimize the calmodulin–*smMLCK* interface. Thus, without considering negative design explicitly, a design with eight mutations enabled similar binding affinity to the target and 1.5- to 86-fold decreased affinity to six other targets. In 2003 a follow-up included optimization of the CPD for PPI [32]. First, the pairwise portion of the energy function was weighted to enhance intermolecular interactions and attenuate intramolecular ones. Second, the large dielectric constant ( $\epsilon$ ) routinely used, effectively underemphasized the long-range electrostatics term in the energy function relative to more local terms such as van der Waals and hydrogen bonding interactions. Consequently, the dielectric constant at the boundary- and surface-optimization region was lowered from  $40r$  to  $4r$ . Third, a rotamer library that contained rotamers representing expansion about the  $\chi_1$  and  $\chi_2$  angles was applied. Six designs were tested on eight targets of which the best showed a specificity change of 0.9- to 155-fold. Hence, by optimizing the protein– protein binding, the

natural promiscuous binding was decreased. Yet, without direct incorporation of negative design, this decrease displayed large variation among the alternative targets.

In 2002 Xencor applied the Protein Design Automation CPD software (PDA [141]) and demonstrated it by redesigning 19 residues in the vicinity of  $\beta$ -lactamase's active site to confer resistance against antibiotic cefotaxime [33]. The PDA defines a library of mutant sequences at specific positions. After finding the global minimum energy conformation (GMEC) an MC/SA search algorithm is applied to find near-optimal sequences which are then processed to generate a probability table of mutations at each designed position. The CPD reduced the large sequence space to a library of ~200,000 sequences which were experimentally screened obtaining variants exhibiting a 1280-fold increase in cefotaxime resistance along with a 40-fold decrease in ampicillin resistance.

In 2002 Xencor applied CPD to stabilize solubility and improve thermostability of the human growth hormone (hGH) [34] and to stabilize the granulocyte-colony stimulation factor (G-CSF) [35]. In both cases, only core residues were redesigned. As the CPD scheme of the two targets was similar, they are described here together. In both cases, the DEE-based PDA CPD scheme was applied. Interestingly, new terms for side-chain and backbone entropies were added to the scoring function as a combined measurable reflecting the loss of conformational entropy during core packing of the designed core residues. Other scoring function components such as polar hydrogen burial, dielectric constant, and surface-based nonpolar exposure penalty were weighted into a new scoring function. The 45 core residues were redesigned resulting in 11 mutations. Three designs were tested experimentally achieving thermostabilization of 13–16 °C without compromising biological activity. Similarly, the G-CSF was redesigned to improve pharmacological properties for the prevention of chemotherapy-related neutropenia [35]. Here, a homology model based on the bovine structure was used as a template with 25–34 core residues redesigned with PDA. Several mutants with 10–14 mutations were experimentally characterized. Without compromising biological activity, a thermostabilization of 13 °C and a tenfold improvement in shelf-life was obtained.

In 2002 the Baker, Monnat and Stoddard labs designed an artificial endonuclease by fusing the N-terminal domain of homing endonuclease I-Dmol to an I-Crel monomer, creating a new 1400 Å<sup>2</sup> interface between the domains [36]. The design, termed *E-Drel*, for engineered I-Dmol/I-Crel, was initially modeled by superimposing a single helix from the N-terminal domain of I-Dmol on the same helix in I-Crel and linking the two domains using a three-residue linker –NGN– which encourages  $\beta$ -turn formation. All interface positions were redesigned using RosettaDesign



[134]. The relative contribution of side chains to the interface free energy were evaluated by computational alanine scanning [142]. The CPD focused on six residues exhibiting steric clashes in the original model and extended to eight additional residues predicted to contribute substantially to the interface free energy. One thousand separate designs were conducted over two backbone models eliminating results that may affect the active site and reducing redundant results. The 16 top-scoring designs, each with 8–12 interface mutations were screened *in vivo*. The resulting structurally- and functionally characterized E-Drel enzyme bound the DNA target site with nanomolar affinity and cleaves it at precisely the same rate as the *wild-type* enzyme.

In 2003 the Wodak lab conducted automatic design of major histocompatibility complex class I (MHC-I) 9-residue binding peptides which impair CD8+ T-cell recognition [37]. While this is a 9-amino acid peptide design rather than a protein design, it is presented here as an early example of computationally designing peptide–protein interactions. DESIGNER [5, 6], which combines a fitness function with an optimization procedure selecting highly scoring sequences. To select amino acid sequences with lowest free energies, a DEE procedure was applied as well as a heuristic procedure with 250,000 iterations. In an early ensemble-like approach, DESIGNER was run on all six representative MHC-peptide complexes available in the PDB. In addition, the top-scoring peptides were scanned against peptides known to bind the same MHC allele. The six strongest binders not only bound MHC but also formed stable complexes and three displayed significant inhibition of CD8+ T-cell recognition.

In 2003 the Saven and DeGrado labs designed a water-soluble analog of the pentameric phospholamban membrane protein [38]. Solubilization enables to study the protein, including ligand or drug interaction, in the much friendlier soluble milieu. Here, 11 solvent-exposed residues were identified in the transmembrane (TM) helix. Ten residues were redesigned using a pairwise potential including intrahelical pairwise residue interactions, contribution to the helix macrodipole, interhelical electrostatic interactions, solubility, and sequence entropy. The water-soluble analog mimicked all the TM protein characteristics including oligomerization state, helical structure, and stabilization upon phosphorylation. A truncated version of the helix bundle was resolved crystallographically [39] displaying a parallel tetramer, rather than an antiparallel pentamer; suggesting that buried and interfacial hydrogen bonds are pivotal for oligomerization.

In 2003 Havernek and Harbury approached molecular recognition by entwining positive- and negative-design using a multi-state framework for engineering specificity in GCN4-based coiled-coils [40]. Their approach selects sequences maximizing the transfer free energy of a protein from a target conformation to a set of undesired



competitor conformations. The algorithm identified three specificity motifs that have not been observed in naturally occurring coiled coils. Their genetic algorithm (GA) considered four states including homodimer, heterodimer, aggregated-, and unfolded-state which focus on homospecificity, solubility, and stability. Unlike previous CPD approaches, they selected sequences that maximize the transfer free energy from a target state to an ensemble of competitors, thus requiring separate structural optimization for each state. Further, they evaluated prediction by molecular mechanics with continuum solvent allowing for direct prediction of observed free energies. Seven of the eight engineered pairs showed  $\Delta G_{\text{specificity}}$  values exceeding the largest control value that was obtained fortuitously.

In 2003 the Saven and DeGrado labs designed a de novo monomeric helical dinuclear metalloprotein [41]. The 114-residue four-helix-bundle due ferro single-chain ( $DF_{SC}$ ) was modeled in the backbone level using previous oligomeric structures and inter-helical turns. While 26 residues were predetermined including ligand-binding residues and one of the turns, all other 88 residues were computationally designed using the Statistically Computationally Assisted Design Strategy (SCADS [143]). The fixed positions relied on previous designs of due ferro peptide ensembles [47, 144]. The software provides site-specific amino acid probabilities, which are then used to guide sequence design. This successful design was the first realization of complete de novo design, where backbone structure, activity, and sequence are specified in the design process. Several years later, the structure was solved combining NMR and unrestrained MD using nonbonded force-field for the metal shell, followed by quantum mechanical/ molecular mechanical dynamics used to relax the NMR-apparent local frustration at the metal-binding site [42].

In 2003 Kuhlman, Dantas and coworkers at the Baker lab presented a milestone in CPD—the first systematic de novo CPD of a 93-residue  $\alpha/\beta$  novel topology protein, which folded in atomic-level accuracy (1.2 Å RMSD) to the design template [43]. The so called *TOP7* protein includes four  $\beta$ -strands flanked by two  $\alpha$ -helices. The loops connecting the secondary structure elements are very short thus contributing to the atomic-level accuracy of the design. The starting models for the design were assembled from three- and nine-residue fragments via the Rosetta package [133]. 172 backbone-only models were generated, forming an ensemble of structures that all fit the requested fold. The sequences were generated using RosettaDesign [134] via a Monte Carlo (MC) search protocol focusing on van der Waals and hydrogen-bonding interactions within an implicit solvent. An additional reduction of search complexity was attained by restricting the  $\beta$ -strand positions to polar residues. With the Dunbrack rotamers [145] considered for each position, the procedure included  $>10^{186}$  rotamer

combinations. A simultaneous optimization of sequence and structure was conducted by using the Rosetta approach for backbone optimization with each starting structure followed by 15 cycles of sequence design and backbone optimization.

In 2003 the Desjarlais lab de novo designed a WW domain using fully automated CPD emphasizing backbone flexibility [44]. Here, the labs' SPA [128] CPD software was coupled to a sampling procedure integrating information from an ensemble of backbone structures, thus setting the stage to multistate CPD. The new procedure was termed SPANS for sequence prediction algorithm for numerous states. The ensemble was generated by a simple MC expansion of  $\pm 5^\circ$  perturbation of the backbone  $\Phi$  and  $\Psi$  angles till a predetermined (0.3 Å) RMSD. Three antiparallel strands fold into a  $\beta$ -sheet WW domain. The 34–40 amino acid WW domain folds autonomously with two-state kinetics and is utilized as a module to bind proline-containing regions. Two CPD approaches were used, each with methods applied in many other applications. First, SPANS-WW1 applied multiple “sub-rotamer” states which were sampled stochastically. The Boltzmann weights of these states were combined into one “super-rotamer” and included in the partition function. Alternatively, SPANS-WW2 optimized each canonical rotamer by torsion-space steepest-descent minimization. Both designs exhibited WW domain biophysical characteristics yet with decreased stability relative to the template, especially for SPANS-WW1 which included a less-dispersed hydrogen-bond network.

In 2003 the Baker lab applied RosettaDesign for the redesign of nine different globular folds achieving, on average 65 % deviation in sequence space with biochemical characteristics comparable with their natural templates [45]. One of these designs, human procarboxypeptidase A2, was structurally resolved in 2007 enabling to discretely analyze residues contributing to different types of hydrophobic packing: interhelical, inter-strand, and helix-strand packing [46]. While the original redesign had numerous mutations and 10 kcal/mol increased stability, relative to the *wild-type*, mutating merely four residues yielded a 5 kcal/mol stability increase.

In 2004 Kaplan and DeGrado designed a phenol-oxidase from first principles [48] using a computationally designed four-helix-bundle scaffold made out of four peptides of two kinds ( $A_2B_2$ ) that assemble in a noncovalent manner [29]. Specifically, positions 15 and 19 were mutated to small amino acids thus sculpting the diiron binding pocket to bind the 4-aminophenol substrate. The resulting quinone monoamine product was produced with a  $k_{\text{cat}}/K_M = 1500 \text{ M}^{-1} \text{ min}^{-1}$  with efficiency sensitive to the size of the binding pocket, thus reporting on design specificity. Herein, although the three-dimensional structure of the backbone and sequence of the de novo designed scaffold protein was designed computationally,

the subsequent introduction of catalytic activity was accomplished without methods or by screening large number of variants.

In 2004 the Baker lab redesigned specificity of a protein–protein interaction between a bacterial nonspecific DNase (colicin E7) and its tightly bound inhibitor protein (immunity protein Im7) pairs [49]. The structurally resolved binding pairs offer straightforward activity assays and the computational design focused on destabilizing interactions with the *wild-type* partner while stabilizing the mutant complex. Interface positions on both binding partners were mutated and assessed as to their binding free energies and specificity changes between cognate and noncognate binding partners. Three positions were chosen for redesign in the DNase and nine in the inhibitor. The designed cognate pairs displayed low affinity relative to the *wild-type* pair, presumably due to a new water network, which was not part of the modeling. This suggests focusing on explicit modeling of bound water in interface design. Nevertheless, the redesigned interface was structurally resolved displaying 0.62 Å RMSD between the model and the actual structure. Focusing on the hydrogen bond network and water therein, a 2006 follow-up study sampled alternate rigid body orientations to optimize the interface interactions and then utilized the resolved structure to further optimize the hydrogen bonding network, thus increasing the specificity difference between cognate to noncognate complexes by 300-fold [50].

In 2004 the DeGrado and Saven labs applied CPD to design a water-soluble analog of the potassium channel KcsA [51]. Using SCADS [143] and the previous solubilization application [38], 35 solvent-exposed residues were identified and subjected to mutation. The first round of the water-soluble K-channel (Denoted WSK-1) displayed high oligomers and thus additional mutations were applied on two solvent-exposed hydrophobic patches. The resulting WSK-3 structure mimics the TM structure in secondary structure, tetrameric quaternary structure, and tight binding of a toxin and a channel blocker.

---

## 4 The Current Decade of CPD, 2005–2014: From Enzymes to Membrane Proteins

In 2005 the Stoddard and Baker labs conducted thermostabilization of the homodimeric hydrolase enzyme yeast cytosine deaminase (yCD), which converts cytosine to uracil [52]. Only three mutations enabled an increase of 10 °C in the melting temperature. All residues that were more than 4 Å from the active site and were not involved in the dimer interface were subjected to CPD. Half of the 65 residues were left unchanged following the redesign and half of the remaining suggested mutations were solvent exposed. The remaining suggested mutations were experimentally characterized individually suggesting a triple mutant as the most thermostable one.

In 2005 Sauer and coworkers compared positive- and negative-design strategies for reengineering a homodimer into a heterodimer [53]. Using the Stringent Starvation Protein B (*SspB*)  $\alpha/\beta$ -fold homodimer as a model system, stability-focused (positive design) using the DEE search algorithm as implemented in ORBIT [18] and specificity-focused (negative design) were applied aiming to reengineer the homodimer into a heterodimer. While the positive design yielded a more stable heterodimer, only the incorporation of negative design yielded exclusive heterodimerization. Eight interface residues (four from each subunit) were subjected to design allowing for ten out of the 20 amino acids in each position. The authors note that the greatest challenge was modeling the energetic effects of destabilizing mutations in competing state. This challenge was approached by capping unfavorable van der Waals energies as an approximation for conformational relaxation that would alleviate atomic overlaps. Notably, in 2007 the Mayo lab used ORBIT [18] to design 13 and 11 residues on two monomer variants of streptococcal protein G— $\beta$ 1 domain (*G $\beta$ 1*) that were designed to heterodimerize [60]. Of the 24 positions, 15 “core” positions were restricted to seven hydrophobic residues and the rest to polar and charged residues. Applying such hydrophobic patches serves as negative designs destabilizing the monomer state. This specific design was successful in shifting a monomer to a dimer, albeit with a low binding constant. Overall, these studies showed the challenges of PPI design along with the importance of negative design, even at the expense of stability.

In 2005 the DeGrado, Saven and Dutton lab de novo designed a 40-residue redox-active minimal rubredoxin mimic [54]. This is one of the first  $\beta$ -sheet CPD, let alone with the rubredoxin tetrahedral metal-binding motif. The last three strands of the *Pyrococcus furiosus* rubredoxin were transformed using a twofold symmetric axis containing the metal ion. A hairpin motif (tryptophan zipper) was used to fuse the two sides. Other than the hairpin motif, active-site Cys, two Gly and an Ile residue, all amino acids were designed using SCADS [143]. The apoprotein and holoproteins were stable with 16 Fe(II/III) functional cycles under aerobic conditions.

In 2005 the DeGrado lab applied CPD for a de novo four-helix bundle protein that selectively binds two nonbiological cofactors termed *DPP-Fe* for 5, 15-Di[(4-carboxymethylene-oxy)phenyl]porphinato iron(III)-chloride [55]. Herein, the apoprotein folds upon binding the cofactors. The four-helix bundle was designed to maintain 17–19 Å between the metals, His-Fe coordinative interactions, second shell hydrogen-bonding, minimal steric clashes and  $D_2$  symmetry with sampling via MC/SA. Then, three rounds of SCADS [143] sequence calculations were applied to 28 residues.

In 2006 Dmochowski, Saven, and coworkers designed ferritin-like proteins (*Dps*) with increasingly hydrophobic cavities [56]. The

resilience of the self-assembling complex to mutation which intuitively should denature the protein is striking. As many as 120 hydrophilic residues were mutated to hydrophobic or small amino-acids. The Dps complex is a 12-subunit iron warehouse in which each subunit is a four-helix-bundle with two helices facing the interior large iron-binding cavity. The SCADS [143] software extended for symmetric homo-oligomeric quaternary structures [146] was applied forming Dps3, Dps7, and Dps10, each with three, seven, and ten mutations in each of the dozen subunits. Not only was the mutation per se taken into account but also how much each residue is prone to an acceptable mutation. Amino acids participating in salt bridges within the hydrophobic core were not subjected to mutagenesis. The mutations increased the percent of hydrophobic surface within the iron-binding cavity from 52 % to 86 %. The high melting temperature of the complex as well as iron-mineralization function were largely unchanged for Dps3 and Dps3 and even Dps10 folded and assembled properly. Taken together, this study questions the importance of the hydrophilic surface for proper folding of proteins, let alone protein complexes; thus opening the door for CPD of hydrophobic surface regions.

In 2006 Quax, Serrano, and coworkers designed tumor necrosis factor-related apoptosis-inducing ligand (TRAIL) variants which initiated apoptosis exclusively via the DR5 receptor [57]. The DR5-selective TRAIL variants represent a reduced binding promiscuity CPD approach which in this case potentially permits tumor-selective therapies. The CPD scheme was straightforward including protein modeling via WHATIF followed by refinement via FOLD-X. Residues binding to nonconserved positions in the different four potential receptors were mutated via FOLD-X to all other amino acids obtaining 2720 models for the 34 designed sites. The binding energy of the models was used to assess selectivity yielding seven single-site variants for experimental validation.

In 2006 the Baker lab redesigned a cleavage specificity of the intron-encoded homing endonuclease I-MsoI [58]. The CPD aimed at changing one base pair in each recognition half site. The CPD approach used as input the *wild-type* crystallographic structure and considered (in turn) all symmetric base pair changes. New side chains next to these base pairs were attempted listing the predicted discrimination energy between the previous and new recognition sites. The modeling of the DNA-protein interface is challenging not only due to the highly charged electrostatic environment possibly requiring bound water molecules, but also as the binding may involve conformational changes in both binding constituents. The redesigned enzyme cleaves the new recognition site ~10,000 more effectively compared to the *wild-type* protein.

In 2006 Xencor Inc. designed antibody Fc variants with enhanced Fc $\gamma$ -receptor-mediated effector function [59]. A

combination of “directed” diversity and “quality” diversity strategies were applied within the CPD scheme of optimizing the IgG Fc region for Fc $\gamma$ -receptor affinity and specificity. Four positions were mutated in different combinations. Where structural information was available, substitutions that provide favorable interactions were designed, and where such information was incomplete, calculations provided a quality set of variants enriched for stability and solubility. At some positions, only residues with high propensity to the core, surface and boundary of the protein were allowed, thus focusing the search space sampled. The designed variants displayed over 2 orders of magnitude enhancement of in vitro effector function, enabled efficacy against cells with low levels of target antigens and resulted in increased cytotoxicity *in vivo*.

In 2007 the DeGrado lab designed a TM peptide that specifically targets a membrane protein [61]. The peptide was named *CHAMP* for Computed Helical Anti Membrane-Protein Peptide. The TM helices of the  $\alpha_{\text{IIb}}\beta_3$  and  $\alpha_v\beta_3$  integrins were the subject of the design by replacing the  $\beta_3$  subunit with a new designed helix. The two subunits form a parallel GAS<sub>Right</sub> motif [147] which was structurally modeled with the  $\beta_3$  subunit was redesigned. Five and 15 template backbones were tested for the design of the *CHAMP* against the  $\alpha_{\text{IIb}}$  and  $\alpha_v$  helices, respectively. In the inner half of the membrane only eight residues were considered. Repacking of proximal positions was accomplished with a linearly dampened Lennard-Jones potential with van der Waals radii scaled to 90 %, as implemented in PROTCAD [29] and a membrane-depth dependent knowledge-based potential. 10,000 iterations of an MC with simulated annealing (MC/SA) were applied for the sequence and rotamer space search and sampling, with the rotamers optimized using DEE followed by exhaustive enumeration. The new designs were tested in micelles, bacterial membranes, and mammalian cells.

In 2007 the Kuhlman lab focused on high-resolution design of a protein loop [62]. Within the Rosetta software package a loop design protocol was developed. The protocol iterates between optimizing the sequence and conformation of a loop in search of low-energy sequence–structure pairs. 10-residue loops were designed for connecting the 2nd and 3rd strand of  $\beta$ -sandwich protein tenascin-C. Loop templates were datamined from 142 12-residue loops found in the protein databank (PDB) that superimpose the backbone atoms of the design target. These backbone templates were redesigned with many undergoing four to five mutations. Loops were filtered by searching for solvent accessible surface area to a 0.5 Å radii probe and by searching for unsatisfied hydrogen bonds. Two of three experimentally tested loop designs were solved showing similar structures compared to the design while a third design appeared in a significantly different structure; thus highlighting the potential for loop design along with the unique challenge in designing loops.



In 2007 Lai and coworkers de novo designed a protein that binds the erythropoietin receptor [63]. The CPD was based on grafting discontinuous interaction epitopes. The erythropoietin (EPO)—EPO-receptor complex structure was studied; identifying three key residues in EPO which were searched in the PDB - yielding 1756 potential scaffold proteins onto which the keystone residues were grafted. These were filtered for RMSD, shape-complementarity, packing density, and high buried accessible surface area yielding 15 potential scaffolds for further analysis. A fourth mutation was designed to eliminate a steric clash. The novel triple mutant, composed of an unrelated protein, rat PLC $\delta$ 1-PH (pleckstrin homology domain of phospholipase C- $\delta$  1) bound the EPO receptor with a  $K_D$  of 24 nM in vitro and gave an  $IC_{50}$  of 5.7  $\mu$ M in a cell-based assay.

In 2007 the Mayo lab redesigned a 51-residue homeodomain aiming at thermostability [64]. Different sequence optimization algorithms were compared of which two were characterized. Amino acids were divided into buried and solvent-exposed, and further restricted at helix-capping sites. MC/SA yielded the best solution. The successful design had a thermal denaturation midpoint temperature of  $>99$  °C.

In 2007 the DeGrado, Saven and Roder labs applied CPD for the de novo design of a single-chain asymmetric diphenylporphyrin four-helix bundle metalloprotein [65]. An MC/SA protocol was applied given five constraints: (a) a metal-metal distance of 17–19 Å, (b) optimal His to Fe bonding interactions, (c) second-shell His-Thr hydrogen bonding, (d) minimal steric clashes, and (e) D<sub>2</sub>-symmetry. A previous four-chain design [55] was shortened by four residues at each end and replaced by loops. A new program, STITCH, identified loops within a nonredundant PDB set that superimposed well on five amino acids at the helical ends. Iterative cycles of SCADS [143] CPD chose the sequence for 100 of the 108 amino-acids, with eight keystone His and Thr residues fixed as part of the cofactor ligation. The experimentally characterized single-chain design demonstrated higher stability compared to the four-chain previous design both apo- and holo-forms with the latter increasing stability significantly.

In 2007 the Tidor computational lab and the Wittrup experimental lab joined forces to apply CPD for the improvement of antibody affinity [66]. The iterative CPD cycle focused on electrostatic binding contributions and single mutations. By combining multiple designed mutations, a tenfold and 140-fold affinity improvement was engineered to an anti-epidermal growth factor antibody and to an anti-lysozyme antibody, respectively. Interestingly, this study began by a general CPD approach that was in general not successful and led to the understanding that for antibody designs the calculated electrostatic term (using Poisson-Boltzmann continuum electrostatics calculations) for binding was

a better predictor for affinity improvement compared to the total calculated binding free energy. Thus, a full side-chain conformational search was maintained but only the electrostatic component was applied for affinity improvement.

In 2008 the Schreiber and Edelman-Sobolev labs redesigned a protein–protein interface between TEM1  $\beta$ -lactamase and its inhibitor  $\beta$ -lactamase inhibitor protein (BLIP) for high-affinity and binding specificity using a novel method [67]. Their novel PDBmodDesign method included replacing structural interface modules with fragments taken from nonrelated proteins and ranking the  $10^7$  starting templates with an accurate atom–atom contact surface scoring function. The resulting high affinity and specificity affirms their modularity approach.

In 2008 the Dmochowski, Saven and Christianson labs joined forces to design a human H ferritin protein that will bind noble metal ions  $\text{Au}^{3+}$  and  $\text{Ag}^+$ , reduce the ions and form nanoparticles within the protein's cavity [68]. The study followed up on the ferritin-like protein hydrophobic cavity design [56] and applied a similar CPD methodology. Here, 192 mutations were designed in the 24-subunit complex including four external- and four internal-surface mutations for each subunit. Two His and two Cys on the external surface were mutated to charged, polar, or small residues. In parallel, three Glu and a Lys on the internal surface were all mutated to Cys as an ion-binder residue. Combining positive- and negative-design this was aimed to promote noble metal ion binding in the cavity while avoiding such binding on the outside surface as well as minimizing protein aggregation. Following experimental difficulties of crystallization with gold ions,  $\text{Hg}^{2+}$  was used to probe the metal–thiol interactions. Probably due to decrease in aggregation, the outer-surface mutations stabilized the protein. Strikingly, the internal-surface mutations kept this high stability and exhibited  $\text{Ag}^0$  and  $\text{Au}^0$  nanoparticles upon soaking with their respective ions. Indeed, the crystal structure proved the CPD structure and requested function.

In 2008 Handel and coworkers redesigned BLIP to increase affinity to SHV-1 which unlike TEM (presented in the previous example), displays micromolar affinity, thus providing space for affinity improvement [69]. The EGAD design software succeeded to stabilize the interface by 10- to 1000-fold. The experimental structures generally agreed with the computational designs, except for salt-bridges. Additionally, the authors claim that the off-rotamer conformational sampling could be improved by adding a short minimization following the DEE rotamer search.

In 2008 the Saven, Therien, Blasie and DeGrado labs from the University of Pennsylvania designed nanostructured metalloporphyrin arrays from coiled coils [70]. Following a previous design of a  $D_2$ -symmetric  $\alpha$ -helical coiled coil (34 residues for each helix) that binds two nonbiological porphyrin cofactors [55], the four-helical



coiled-coil was extended by three-heptad repeats, enabling the binding of four iron porphyrins. Three charge patterning mutations were introduced to enforce an antiparallel orientation and two additional mutations were introduced to improve electrostatic interactions with the cofactor carboxylates. The resulting four-porphyrin complex was experimentally characterized. The modular addition of heptad repeats between the helical capping sections demonstrates the robustness of the coiled-coil structure, as defined by the Crick parameters. This design introduces the feasibility of engineering electrically and optically responsive multiporphyrin arrays.

In 2008 the Baker lab presented two computational enzyme designs—a group of retro-aldolases [71] and a Kemp eliminase [72], the latter with Tawfik. Both designs applied a similar scheme for enzyme design without cofactors [148]. These computational enzyme designs followed an algorithm presented in 2006, which was successful in targeting ten different enzymes and identifying the native site in the native scaffold and ranking it within the top five designs for six of the ten reactions [149].

The retro-aldolase CPD strategy is described over 12 pages in the supplementary material of the publication highlighting the many aspects that must be addressed [71]. These range from the quantum-mechanical (QM) structural description of the catalytic sites to the computational and experimental ranking and validation of the designs. Briefly, composite active-site descriptions of transition states were applied to generate candidate catalytic sites via RosettaMatch [150] which fills a hash-table with catalytic amino-acid rotamers for the proposed catalytic site constraints. The remaining positions are redesigned to optimize the transition-state binding affinity using RosettaDesign [134]. Following structural refinement, the potential designs are ranked based on the total binding energy to the composite transition state as well as satisfaction of specific catalytic geometry. Designs were filtered if the van de Waals energetics was too high ( $>-5$  kcal/mol), the binding pocket was too buried or was not sufficiently accessible. This CPD scheme resulted in 72 designs of which 32 displayed retro-aldolase activity of up to 4 orders of magnitude kinetic acceleration.

The 2008 Kemp eliminase CPD by the labs of Baker and Tawfik [72] achieved a  $10^5$  rate enhancement. In vitro evolution further enhanced  $k_{\text{cat}}/K_{\text{M}}$  by  $>200$ -fold. The CPD scheme was similar to the one of for the retro-aldolase. The successful designs showed high shape-complementarity with several polar or charged catalytic residues: out of 59 designs, 39 used Asp or Glu as a general base while 20 used His-Asp or His-Glu as a catalytic dyad. Such variation highlights the robustness of the CPD strategy which, in this case, exhibits variability in the functionally accessible set of catalytic residues.  $\pi$ -stacking interactions contributed towards stabilizing the transition state. The collaboration between the CPD approach

provided by the Baker lab and the directed evolution approach provided by the Tawfik lab continued with subsequent directed evolution efforts conducted by Khersonsky et al. [138–140]. Cumulatively, the latter efforts showed that CPD designs are highly evolvable and can be optimized for catalytic efficiency, reduced thermodynamic stability (which is often too high in computational designs), optimization of the catalytic site microenvironment for the required transition state preorganization, and the presentation of key changes that provide feedback for deciphering mechanism and further CPD efforts. While directed evolution is not the focus of this chapter, the collaboration highlights the need to embed within the CPD approach other fields in a multiple dimension feedback approach. Fortunately for the CPD field, this Kemp eliminase computational design sparked an array of follow-up research of which some is highlighted below [92, 93, 100, 151] with the key kinetic parameters summarized in Table 2.

In 2009 the Baker lab focused on loop remodeling to alter enzyme specificity [73]. Following benchmark tests on eight native protein-ligand complexes, a critical loop in guanine deaminase was redesigned such that it became 100-fold more active on ammelide and 25,000-fold less active on guanine. The two to five residue loop modeling succeeded in altering specificity. Nevertheless, it should be noted that the absolute activity towards the new substrate ( $k_{\text{cat}}/K_M = 0.15 \text{ s}^{-1} \text{ M}^{-1}$ ) is still 7 orders of magnitude lower than the activity of the *wild-type* enzyme towards its innate substrate; highlighting the comprehensive evolution of enzymes towards their functionality, which is likely to include far more than one loop.

In 2009 the Shifman lab applied CPD for increasing the binding specificity of calmodulin 900-folds [74]. Relying on the promiscuous binding of calmodulin to both CaM-dependent protein kinase II (CaMKII) and calcineurin (CaN), calmodulin was optimized to bind the former. The ORBIT-based [18] CPD emphasized intermolecular interactions and showed that the specificity increase was largely due to a decrease in binding to CaN.

In 2009 the Keating lab applied a computational framework for design of protein-interaction specificity allowing for CPD of selective basic-region leucine zipper (bZIP) binding peptides [75]. The 20 bZIP transcription factor family share high sequence similarity challenging specificity design. As shown by protein arrays, the CPD succeeded in designing selectivity by optimizing the affinity and specificity trade-off e.g. by sacrificing the stability score and by introducing negative design to disfavor complexes with undesired bZIP competitors. The bZIP microarray assay benefits from reversible folding of short coiled coils, and data from previous array measurements of many bZIP transcription factor pairs were critical for developing predictive models. Their CPD framework is denoted CLASSY for cluster expansion and linear programming-based

analysis of specificity and stability [75]. The CLASSY multi-state CPD applies integer linear programming followed by cluster expansion in which a structure-based interaction model is converted into a quick-to-evaluate sequence-based scoring function. Negative design is integrated by applying CLASSY to the design-target and to design-off-target states.

In 2009 the Baker lab conducted CPD on the monomeric homing endonuclease I-AniI which cleaves at the center of a 20-base-pair DNA target site [76]. The pseudo-symmetrical enzyme's N- and C-terminal domains bind to the left (−) and right (+) DNA target sites in very different manners as reflected by causes of CPD-based altered specificity: specificity on the (−) side was achieved by modulating single-turnover conditions ( $K_M$ ) while that in the (+) side was achieved by modulating turnover number ( $k_{cat}$ ). The Rosetta-based CPD scheme tailored for DNA–protein interactions relied on their previous study [58]. Loop rebuilding was used to model backbone shifts. In a feedback loop, the best designs were reverted position by position to the *wild-type* sequence to identify mutations that did not contribute significantly to the energy or specificity. Multi-state design [40] to assess the specificity offset between the altered and *wild-type* DNA target structure. Further, a genetic algorithm was applied to evolve sequence for preference of the target state compared to competitor states.

In 2009 the Donald lab conducted computational structure-based redesign of the phenylalanine adenylation domain of the nonribosomal peptide synthetase enzyme gramicidin S synthetase A (GrsA-PheA) for a set of noncognate substrates for which the *wild-type* enzyme has little or virtually no specificity [77]. Here the aim was increased specificity with the leading design exhibiting 1/6 of the enzyme/*wild-type* substrate activity. The  $K^*$  algorithm [152] was applied on the active site, a generally considered optimized region which is not the classical target for most CPD attempts. The double mutant selected showed a 19-fold increase of  $k_{cat}/K_m$  for the new Leu substrate and a 27-fold decrease of this measurable for the *wild-type* Phe substrate. On top of two active-site mutations, so called “bolstering” mutations were designed outside the active site aiming to stabilize the -active-site mutant. Indeed, such mutations gave an additional twofold increase in  $k_{cat}/K_m$  for the Leu substrate. Similarly, further designs for charged substrates were also successful experimentally.

In 2010 the DeGrado, Saven and Therien labs applied CPD for the design of an  $A_2B_2$  four-helix bundle that selectively binds two emissive abiological (porphinato)zinc chromophores of DPP-Zn [78]. The positive and negative ligand-directed CPD is selective and did not bind related chromophores such as DPP-Fe<sup>3+</sup>. To achieve the selective Zn-cofactor binding, a pentacoordinate environment with one His ligand was designed, yielding  $C_2$  symmetry. One peptide chain included a His ligand while the other

included a Thr ligand; thus applying a negative design element that allows only the heterotetramer to bind the chromophore. SCADS [143] was applied for the recursive design of 62 variable positions. Cys (potentially making disulfide bridges), His (potentially ligand binding) and Pro (potential helix-breaker) were excluded at all positions, Met at interior positions. Three sequential rounds of sequence CPD were applied and the resulting design was validated experimentally.

In 2010 the Baker lab altered the cleavage specificity of the I-Msol homing endonuclease for three contiguous base pair substitutions [79]. Using a CPD scheme previously applied to the protein [58, 76], concerted design for all simultaneous substitutions was more successful than a modular approach against individual substitutions, highlighting the importance of context-dependent redesign and optimization of protein–DNA interactions. In a CPD and structure determination feedback loop, a structure of the CPD effort and its associated unanticipated shifts in DNA conformation was utilized to create an endonuclease that specifically cleaves a site with four contiguous base pair substitutions.

In 2010 the Mayo lab changed the emission wavelength of red fluorescent protein by CPD [80]. Herein, CPD was combined with small experimental combinatorial libraries of mCherry mutants. The library design procedure takes as input a list of scored sequences, and two sets of constraints: a list of allowed sets of amino acids, and a range of desired library sizes. The algorithm generates a list of the combinatorial libraries that satisfy these constraints, and then ranks the libraries by the degree to which they reflect the energetic preferences present in the list of scored sequences. Thus, CPD was used to perform an *in silico* prescreen to eliminate sequences incompatible with the protein fold and generate combinatorial libraries amenable to rapid experimental screening. The successful 20–26 nm red-shifted mutants found included targeted stabilization of the excited state via H-bonding and  $\pi$ -stacking interactions as well as destabilization of the ground state via hydrophobic packing. Overall, 13 residues were involved in the design.

In 2010 Warshel suggested that the current computational enzyme design approaches reflect incomplete understanding of the details of the enzymatic system and/or inaccurate modeling by the CPD algorithm [151]. Using his empirical valence bond (EVB) simulations of the Baker and Tawfik Kemp eliminase [72], his group showed that the attempt to predict the proper transition state stabilization and related overall preorganization effect are not likely to be achieved by gas phase models. Warshel showed that the transition state design displays a charge distribution that makes it hard to exploit the active site polarity, even with the ability to quantify the effect of different mutations. Further, the directed

evolution led to reduction of the solvation of the reactant state rather than to the expected transition-state stabilization applied by naturally evolved enzymes. This study highlights the need to carefully design the preorganized environment such that it will exploit the small changes in charge distribution during the formation of the transition state.

In 2010 the DeGrado, Therien, Blasie and Walker labs de novo designed a TM diporphyrin-binding protein complex [81]. The design, termed *PRIME* (PoRphyrins In MEMbrane), positions two non-natural iron diphenylporphyrins ( $\text{Fe}^{3+}$  DPP's) sufficiently close to provide a multicentered pathway for TM electron transfer. Unlike previous TM to soluble solubilization efforts, here the opposite path was applied with a four helix  $D_2$ -symmetrical bundle adapted for the membrane milieu. First, keystone cofactor-binding residues (His and Thr) were designed within an idealized four-porphyrin binding soluble four-helix bundle backbone template [70]. Then, an all side-chain DEE followed by MC/Self-consistent mean field (SCMF) approach was applied to explore the reduced search space along with the Lazaridis implicit membrane solvation (IMM1). The 24 positions were divided to four categories (buried, mostly buried, mostly exposed and completely exposed). These were given different degrees of side-chain conformational sampling with conformations selected from a conformer library. Models were ranked by oligomerization energy, i.e. the difference between the energy of the complex and that of the monomeric state (a membrane solvated helical state, with relaxed side chain conformations), and the lowest energy model was extensively experimentally characterized validating the design.

In 2010 the Kuhlman lab redesigned the binding of hyperplastic discs protein to P21-activated kinase 1 kinase (PAK1) domain [82]. The Iterative Rosetta-based DDMI (Dock, Design, Minimize Interface) protocol was used for docking the scaffold on a chosen hotspot. Next, loops of an MC-based sequence optimization and backbone optimization by minimization were conducted. This resulted with potential redesigned interfaces that were filtered by knowledge-based criteria including binding energy density and the number of unsatisfied polar interface residues. Of six experimentally characterized designs, two aggregated and the rest had binding affinities of up to 100  $\mu\text{M}$ .

In 2010 the Mayo lab combined CPD with experimental library screening demonstrating the successful synergism of the two approaches for thermostabilization of core positions of *G $\beta$ 1*, the  $\beta$ 1 domain of Streptococcal protein G [83]; a protein previously designed by the lab to dimerize [60]. The lab's previous Fast and Accurate Side-chain Topology and Energy Refinement (FASTER) CPD software for single-state design was expanded here for the multistate design case. The combination enables the application of multistate design methods to large conformational libraries,

transformation of semi-rational CPD results to combinatorial mutation libraries, and the experimental stability determination of the designed libraries. The novel protein library design method took into account the library size and possible sets of amino-acids to best reflect the experimental results. The library design procedure was called *CLEARSS* for Combinatorial Libraries Emphasizing And Reflecting Scored Sequences. Five experimental crystallographic and NMR structures were used, each resulting in a 24-member design library. The results enabled to characterize the sequence space available for the multistate design.

In 2010 the Anderson and Donald lab applied CPD for the prediction of drug resistance mutations in methicillin-resistant *Staphylococcus aureus* (MRSA) dihydrofolate reductase (DHFR) [84]. Using ensemble-based CPD algorithm K\* which includes DEE search followed by energy minimization [152], potential resistance mutations were predicted. The process incorporated positive design to maintain catalytic function and negative design to interfere with binding of a lead inhibitor. Interestingly, the *wild-type* sequence was ranked low for both the natural ligand and the inhibitor; suggesting that numerous sequences may have improved binding to these ligands. Four of the ten top-ranking designs were experimentally evaluated, of which three were shown to maintain activity while lowering binding affinity 9- to 18-fold for the inhibitor. The top-ranked double-mutant was crystallized; validating the design by showing reduced hydrophobic interactions in one locus and introducing a steric bulk in another.

In 2011 the DeGrado lab applied CPD to design virus-like protein assemblies on carbon nanotube surfaces [85]. The surface properties and symmetry were used to define the sequence and superstructure of the designed surface-organizing peptides. Single-walled carbon nanotubes were covered with virus-like coating converting the smooth surface into a highly textured assembly with long-scale order, thus capable of e.g. directing the assembly of gold nanoparticles into helical arrays along the nanotube axis. Three selection rules were applied for the design, defining the intrinsic recognition motif and its packing into higher-order assembly in accord with the long-range order of the underlying surface. First, a group compatible with the target surface was identified, in this case avoiding a hydrophobic motif and using small residues Gly or Ala. Second, intersubunit packing was defined in accordance with the surface symmetry. The cylindrical nanotube suggested rotational-screw symmetry in the form of coiled coils with a radius of  $\sim 9$  Å defining five to seven subunits. Third, designability of the coiled coils was assessed by searching existing tertiary motifs. Four designs were tested, sequences based on an existing protein (domain swapped dimer) and a de novo coiled coil, each with Gly or Ala as the nanotube-facing residue. Adding gold particles to the



outer surface enabled transmission electron microscopy (TEM) validation.

In 2011 the Baker lab took the challenge of PPI and designed a protein that binds to the conserved stem surface of influenza hemagglutinin [86]. The strategy focused on the design of shape-complementarity with hot-spot-like residue interactions, with the latter serving as anchors to the former. 865 potential scaffold proteins were searched to support the disembodied hot-spot residues and the shape complementarity. The coarse-grain binding modes were then refined by docking followed by scaffold redesign. Selected designs included 51 and 37 designs with two and three hot-spot residues, respectively. Designs that presented binding were subjected to directed evolution for increased binding; resulting in mutations supporting interactions of filling a void in the binding interface, favorable interactions in the unbound state, electrostatic complementarity, and desolvation. Two binding proteins displayed nanomolar affinity.

In 2011 the Baker lab applied a motif-based method to computationally design protein-protein complexes with native-like interface composition and interaction density as exemplified on the Prb-Pdar heterodimer [87]. The tight dimer was further optimized by directed evolution which surprisingly rotated one of the complex partners by 180°, showing that the specificity of the binding patch was not sufficient yet the binding hot-spot was sufficient to facilitate the binding within a noncrowded pure protein environment. The motif-based approach focused on a key polar aromatic residue (Trp or Tyr) which facilitate packing and hydrogen bonding followed by shape-complementarity. Here, the ankryn repeat which naturally associates with an array of proteins served as one scaffold (redesigned to Pdar). Each of several ankryn repeat protein structures was paired with a set of 37 structurally diverse thermostable proteins applying a surface feature-matching approach, PatchDock [153], followed by rigid-body docking to generate a set of bound orientations with shape-complementarity. The interface design started from screening a well-packed hydrogen-bond containing aromatic pair followed by expanding it to include a hydrophobic first shell of residues and a polar secondary shell of residues protecting the hydrophobic patch from the solvent. RosettaDesign was used to optimize residue identities at the interface periphery holding the hydrophobic inner layer fixed. Further, global long-range electrostatic complementarity was aimed at by biasing one partner to acidic residues and the other to basic ones. Finally, natural parameterizations of native interfaces, e.g. size, packing, void volume, and lack of steric clashes were used to filter the suggested designs. Notably, negative design was not applied in any step, possibly facilitating the 180° flip of binding orientation in an experimentally validated pair. Twelve designed pairs were experimentally screened of which five displayed a signal >2-fold over nonspecific binding. Finally, a combination of phage and yeast display was applied

to evolve tighter binding of the leading pair. Two mutations introduced in this step improved binding from a  $K_d$  of 130 nM to 180 pM.

In 2011 the Kuhlman lab designed a symmetric homodimer using  $\beta$ -strand assembly in which two solvent-exposed strands were designed to form an antiparallel  $\beta$ -strand pairing [88]. Looking for solvent exposed  $\beta$ -strands, automatic homodimer docking (similar to the DDMI protocol) was applied with the  $\beta$ -strand part designed with five rounds of symmetric sequence optimization and minimization at the interface; searching for an  $>850 \text{ \AA}^2$  buried interface and minimizing unsatisfied buried polar atoms. Of the 5500 structures scanned, 1100 had an exposed  $\beta$ -strand. One structure,  $\gamma$ -adaptin was chosen. Two mainly hydrophobic and two mainly polar interface homodimers were characterized of which the former were more successful emphasizing the difficulty in designing hydrogen-bond networks. One promising structure  *$\beta$ dimer1* was structurally resolved showing that the design was successful.

In 2011 William Schief and coworkers applied CPD with flexible backbone remodeling and resurfacing for designing antigens [89]. In this intriguing approach, an HIV 4E10 epitope structure was implanted onto a new scaffold enabling antigen optimization. The remodeling refers to replacing a backbone segment by a new design. The resurfacing refers to redesigning the antigen surface outside the target epitope to obtain variants that maintain only the epitope. Briefly, their six-stage protocol includes segment selection (length, secondary structure), de novo backbone CPD of the segment followed by sequence design and minimization. Next, designs that did not meet energy, packing, and unsatisfied polar-atoms were filtered and surface hydrophobic residues were replaced by polar ones. Three designs of 16–17 remodeled segment were experimentally characterized showing a viable epitope while maintaining solubility and binding affinity.

In 2011 Korendovych and DeGrado applied an alternative minimalist approach to the Kemp eliminase design challenge [92]. Rather than conducting a comprehensive design of a full protein from the QM-optimized active site to the rest of the enzyme, they applied a single mutation in a minimal 75-residue allosterically regulated catalyst, termed *AlleyCat* (for ALLostERICally Controlled cATalsyt), with activity ( $k_{\text{cat}}/K_M = 5.8 \pm 0.3 \text{ M}^{-1} \text{ s}^{-1}$ ) comparable to the original [72] Kemp eliminase design. The rationale was that protein folding energetics can dehydrate a carboxylate side-chain rendering it from the weakly basic aqueous state to a strongly basic dehydrated state. The computational design scheme applied on calmodulin C-terminal domain included in silico single-site Asp or Glu mutagenesis scanning of the C-terminal domain cavity, which naturally binds aromatic side-chains, suggesting that it can bind the benzisoxazole substrate. Low energy models including the point mutation which facilitated a cavity were next docked to the substrate. This determined whether the C-H hydrogen would be



appropriately positioned in the Michaelis complex. Finally, the Glu carboxylate was virtually fused to the substrate and the resulting “superrotamer” was optimized. Alternative mutations were used as control.

In 2011 the Weiss and Saven labs applied SCADS [143] to design a thermostable terpene synthase, an enzyme involved in the synthesis of antibiotics, flavorings, and fragrances [90]. A dozen mutations were selected for design in the tobacco 5-epi-aristolochene synthase (TEAS) for the catalysis of carbocation cyclization. All mutations were  $>12$  Å from the substrate binding site so as to minimize an effect on the functional site. Amino acids identities were prepatterned at the mutated sites based on the number of C $\beta$  atoms within 8 Å of the amino acids: for residues with 0–6 C $\beta$  atoms were constrained to charged, polar, and small residues. For those with 7–8 C $\beta$  atoms, aliphatic and aromatic residues were added to the potential mutations enabling mutation to all residues except Cys, Pro, His, and Thr. Last but not least, buried residues with 10 or more C $\beta$  atoms were allowed to mutate to eight relatively hydrophobic residues. Mutations included both buried and surface-exposed positions with the latter eliminating surface-exposed hydrophobic patches and introducing salt bridges. The design retained activity in 65 °C and denatured in 80 °C, which is twice the temperature relative to the *wild-type*.

In 2011 the Nanda lab computationally designed an A:B:C-type heterotrimer collagen [91]. They applied positive and negative design constraints. A compositional constraint was used where all triplets in the design contained Pro or hydroxy-Pro. The energy score was constrained to allow the melting temperature to be above 26 °C. Specificity was enforced by optimizing the energy gap between the design and the best competing stoichiometry. The resulting empirical design displayed two of the nine available stoichiometries (B:2C and 2B:C). The ABC design indicated multiple species (due to permutations) which were removed upon increasing the salt concentration to 100 mM.

In 2012 several labs from the University of Pennsylvania and University of Pittsburgh applied Saven’s SCAD CPD software to produce a water-soluble TM domain ( $\alpha 1$  subunit) of the nicotinic acetylcholine receptor [94]. The template used for the CPD was a 4-Å low-resolution cryo-electron microscope (EM) structure in which hydrophobic residues with  $>40$  % exposure to the membrane region were redesigned using a molecular mechanics force field entwined with an energy function that constrained the average hydrophobicity of surface-exposed residues to that expected for an average soluble protein of a similar size. In order to avoid spectral over-crowding in NMR spectra used to solve the structure, residues which were not highly favorable in a given site underwent an additional round of CPD with an additional constraint imposed so as to increase sequence diversity. In addition, a polyglycine linker

was designed between the C-terminus of helix-4 and the N-terminus of helix-1 using the loop builder in MODELLER [154]. The design was structurally resolved by NMR displaying high resemblance to the TM domain of the bacterial pentameric ligand-gated ion channel (GLIC); demonstrating the robustness and general applicability of the CPD scheme. Two conformations were resolved with overall dynamics that may be due to the dynamic loops. Moreover, anesthetics were bound to the same residue as in the bacterial GLIC validating the functionality of the solubilized protein.

In 2012 Baker and coworkers redesigned a mononuclear zinc adenosine deaminase metalloenzyme for organophosphate hydrolysis of the  $R_p$  isomer of a coumarinyl analog of the nerve agent, cyclosarin [95]. First, a set of mononuclear zinc enzyme scaffolds with at least one open coordinate state was extracted from the PDB. The open coordinate state was utilized to ensure that structural zinc is excluded from the set. Previous gas-phase quantum-mechanical calculations of organophosphate hydrolysis were used to construct models of the reaction transition state bond lengths and angles. RosettaMatch [150] was used to search for hydrogen-bonding interactions to the phosphoryl oxygen, the nucleophilic hydroxyl moiety, and the leaving group oxygen. Next, RosettaDesign was used for shape-complementarity interactions to the transition state. These parameters along with the presence of a docking funnel timed the results to 12 potential proteins, of which a redesigned adenosine deaminase hydrolyzed the substrate 7-hydroxycoumarinyl phosphate (DECP). The eight-mutation design exhibited activity that was sevenfold higher than that of the buffer background. Directed evolution at eight positions increased activity kinetics to levels identical to the *wild-type* deaminase with over 140 catalytic turnovers per enzyme and high stereospecificity. The directed evolution improvement of  $k_{cat}$  was *post factum* realized as an increase in the basicity of an active site Glu residue.

In 2012 the Schief lab followed up on their previous epitope grafting research [89, 155] and applied CPD with Rosetta to design a new 2F5 HIV epitope with improved biophysical characteristics followed by transplanting the linear epitope onto different scaffolds [96]. Here, the epitope design used side-chain grafting while backbone-grafting was applied to transplant the design onto the new scaffold. Potential scaffolds were identified by searching the PDB for the core Asp-Lys-Trp sequence of the epitope. Side-chain grafting was conducted by binding interface optimization followed by sequence design for epitope accommodation and removal of extraneous interfacial interactions. The latter was facilitated also by initially changing the identity of all non-interacting scaffold residues to glycines. During the automated CPD, residues

within 4 Å of the epitope were allowed to change to any non-cysteine residue while other residues were allowed to change to small residues Gly, Ala, Ser, or Thr. For the backbone grafting, both N-terminal to C-terminal and C-terminal to N-terminal were considered with a 3 Å-RMSD threshold of the epitope to the scaffold set as an initial filter followed by a steric-clash filter. Loop closure utilized a Rosetta low-resolution scoring function, cyclic coordinate descent (CCD [156]) and MC sampling. Next, a high-resolution scoring function was applied to catch problematic conformations. Finally, a full-atom refinement was applied. For two of the three cases tested experimentally, binding to the antibody was increased 9- and 30-fold compared to side-chain grafting alone.

In 2012 Merski and Shoichet applied an alternative minimalist approach by engineering a Met102 → His mutation to the Leu99 → Ala cavity in T4 lysozyme [93]. Here, CPD was applied to engineer subsequent mutations that increased activity fourfold to  $k_{\text{cat}}/K_{\text{M}} = 1.8 \text{ M}^{-1} \text{ min}^{-1}$ . The absence of ordered water or hydrogen bonds and the presence of a common catalytic histidine base in complexes of the enzyme with product analogs facilitated detailed analysis of the reaction mechanism and its optimization. Notably, in this design some of the stabilizing mutations followed previous studies on the T4 lysozyme showing that deep knowledge-based understanding of the template, whether theoretical or experimental, is key to the design efforts. In this iterative approach the first designs had low stability of  $\Delta\Delta G = \sim -7 \text{ kcal/mol}$  relative to *wild-type* T4 lysozyme while subsequent designs increased stability to  $\Delta\Delta G = \sim -2 \text{ kcal/mol}$  with a significant increase in catalytic activity.

In 2012 the Kortemme lab applied CPD to control protein signaling by designing a GTPase/guanine nucleotide exchange factor (GEF) orthogonal (non-cross-reacting) pair [97]. A new interaction was designed while maintaining correct interface with existing machinery. Integrating such a new protein pair into existing cellular circuitry requires consideration of certain design criteria: Not only must the redesigned GTPase be activated by its redesigned GEF partner, but it must also be protected from inadvertent activation by the *wild-type* GEF and all other endogenous GEFs. Further, the redesigned GTPase must also preserve interactions with both upstream regulators and downstream effectors. Here, the known interface between the GTPase Cdc42 and ITSN (GEF) was used as a template for the new design. Computational alanine scanning was used followed by backbone design using the computational second-site suppressor protocol [49]. These simulations identified substitutions in one protein that are significantly destabilizing to the complex formed with the *wild-type* partner but can be compensated for by complementary changes in the partner. Flexible backbone CPD used RosettaBackrub [157] and the robotics-inspired local loop reconstruction method for peptide

chains, called kinematic closure (KIC) [158]. One hundred resulting models were used as a backbone ensemble for interface redesign using one interaction pair as an anchor followed by backbone diversification. Then, soft and hard repulsive forces were applied iteratively aiming at modeling conformational changes that initially appear unfavorable but may be accommodated by subsequent refinement. The experimentally validated design was proven structurally and functionally. The interaction is activated exclusively by the engineered cognate partner while maintaining ability to interface with other GTPase signaling components *in vitro*. The orthogonality was also shown in mammalian cells.

In 2012 the Montelione and Baker labs applied new rules for designing ideal protein structures applying CPD for the design of five different folds [98]. Secondary structure connectivity rules were derived from simulation and from datamining available structures. For connecting two  $\beta$ -strands, 2- and 3-residue loops prefer L-hairpins while 5-residue loops give rise to R-hairpins. For connecting a  $\beta$ -strand to a  $\alpha$ -helix, a parallel orientation is preferred for 2-residue loops while an antiparallel one is preferred for 5-residue loops. For the reverse connectivity ( $\alpha\beta$ ), the general preference is for parallel connectivity, especially for short 2-residue loops and longer loops providing helix-capping. Similar rules were applied for connecting three secondary structures. Negative design was applied for local interactions and for the edge of  $\beta$ -strands, the protein surface and high core packing. Five new folds were designed, almost all with short 2- and 3-residue loops, 7-residue  $\beta$ -strands, and 18-residue  $\alpha$ -helices. *Ab initio* simulations of 200,000–400,000 structure predictions were performed to map the folding energy landscape, selecting 10 % with well-funneled landscapes. Five folds were experimentally determined displaying 1.1–2.0 Å RMSD as compared to their respective designs.

In 2012 Fallas and Hartgerink applied CPD for the design of self-assembling, register-specific collagen heterotrimers focusing on sequence-specific axial salt-bridges [99]. A collagen composed of three distinct chains can trimerize in 27 unique combinations. Axial rather than lateral contacts, stabilize the heterotrimeric collagen target state. The energy score includes a component for the difference between the number of ionizable residues and the number of salt-bridges which was searched using a genetic algorithm. An automated sequence selection algorithm was successful as it balances between destabilization induced on triple helical assemblies by changing conformationally restricted imino acids (Pro) to ionizable residues and the stabilization conferred on the formation of axial interstrand ionic interactions. For each mutation, the gap between the target state and competing states was computed for all 27 states. Experimental validation showed that this minimalist function is sufficient, though could be optimized with the addition of components such as electrostatic repulsion and specific local energetic contributions.

In 2012 the Mayo lab published an interesting story of applying an iterative stepwise approach to computational enzyme design of Kemp eliminases termed *HG-1*, *HG-2*, and *HG-3* [100]. The paper highlights the evolution of the CPD process with increasing success following careful analysis of the result in the previous round, an approach named *the protein design cycle* [141]. The motivation for this study followed on the study of Warshel [151] showing that the Kemp eliminase design of Baker and Tawfik [72] was not an ideal enzyme and required a “shotgun” approach of selection, not to mention benefiting from *in vitro* evolution. Interestingly, for the case of *HG-3*, 17 rounds of directed evolution produced an enzyme which accelerated the reaction by  $6 \times 10^8$ -fold, thus approaching natural enzyme rates [101]. The directed evolution optimized substrate-enzyme shape-complementarity, substrate-catalytic base (Asp127) alignment and, above all, stabilization of a negative charge in the transition state which emerged over the course of the evolution, reminiscent of the serine-protease oxanion hole.

In 2012 four labs from four countries (Grzyb, Nanda, Lubitz, and Noy) joined forces to compare computational and empirical design of iron-sulfur cluster proteins [102]. Both approaches successfully yielded a cluster-binding helical bundle. The CPD of a several coiled coil iron-sulfur clusters (*CCIS*) aimed at increasing stability of the reduced state of the [4Fe-4S] cluster by improving packing, helix propensity, oligomerization prevention (by changing surface net charge), and charge pairing optimization. Each of these aims was tested in a different design. Structural modeling was conducted by multiple-threading alignment within I-Tasser [159], and CPD was conducted using ProtCad [160] using the metal-first approach [161]. All *CCIS* designs were helical. The design focusing on stabilizing the iron-sulfur cluster increased helicity upon binding the cluster, showing the success of the design within a marginally stable protein. In this case, attempts to improve the CPD by intuitive modifications had limited success as to improved stability of the [4Fe-4S] stability over redox cycling suggesting that a different backbone scaffold should be attempted.

In 2012 the Saven and DeGrado labs applied CPD for designing a protein crystal [103]. A three-helix coiled-coil was designed de novo to form a polar and layered P6-space group crystal. An ensemble of crystalline structure models consistent with the required space group was constructed of which designable structures were datamined. These include minima structures in the sequence-structure energy landscape. Within the 26-residue peptide forming the  $C_3$ -symmetry coiled coil, the eight interior positions (*a* and *d* in the heptad repeat) were hydrophobic Val and Leu residues. The other 16 amino acids (not including Pro and Cys) were allowed to be positioned in other places. 19,200 structures were designed to construct a grid over  $R$  and  $\theta$ , representing the inter-protein distance and the angle of rotation around the

superhelical angle, respectively. The final design included a parallel GX<sub>3</sub>G motif interfacing the coiled-coil interhelical contact and an antiparallel GX<sub>3</sub>GX<sub>3</sub>A motif between the coiled coils. Exploiting the symmetry of the honeycomb-like space group, the resulting structure had sub-Å RMSD relative to the designed model.

In 2012 the DeGrado lab altered the function of a de novo Due Ferri four-helix bundle from catalyzing the O<sub>2</sub>-dependent two-electron oxidation of hydroquinones to selectively catalyzing *N*-hydroxylation of arylamines [104]. This was conducted by remodeling the substrate access cavity and by introducing an additional His ligand to the metal-binding cavity. Further second- and third-shell CPD was applied using the Molecular Software Library (MSL [162]) to stabilize the catalytic core. The resulting design had a 10<sup>6</sup>-fold rate enhancement towards the altered function relative to the previous one.

In 2013 the Hahn and Dokholyan labs applied CPD for the rational design of a ligand-controlled protein conformational switch [105]. Their unique topology design of a rapamycin-regulated switch, denoted *uniRapR*, was utilized as a src kinase activator. A high-affinity binding pocket of FK506-binding protein and FKBP12-rapamycin were used with the two proteins connected by a double linker. The first 20 residues of FKBP12 were removed making the N- and C-termini close in space for insertion of the regulatory domain to the other protein. The conformational switching was assessed by replica-exchange and equilibrium discrete molecular dynamics.

In 2013 the Therien, Saven and DeGrado labs joined forces and computationally de novo designed a protein that selectively binds a highly hyperpolarizable abiological chromophore [106]. The 109-residue four-helix-bundle was designated *SCRPPZ-1* and *SCRPPZ-2* for the dimeric and monomeric form, respectively. The protein binds RuPZn, a hyperpolarizable super-molecular chromophore that features highly conjugated (porphyrinato)zinc and (poly-pyridyl) ruthenium. The antiparallel four helix bundle was designed to accommodate the size of the chromophore and ligate the metal ions. Loops for connecting the helices were selected from natural proteins and spliced to accommodate the structure. The SCADS [143] software was used in two rounds first placing the keystone residues and then the other positions. 17 residues were allowed in the helices. His and Cys were precluded as a negative design approach to avoid unwanted metal ligations and disulfide bonds, respectively. Likewise, Pro was precluded from the helices to avoid unwanted kinks. For *SCRPPZ-2* the surface was then redesigned to decrease hydrophobic patches and incorporate interhelical salt bridges to increase bundle stability. A third design included Cys, enabling binding onto functionalized silica surfaces. The protein



structure, stability, and nonlinear optical functional elements were proven with an array of experimental methods.

In 2013 the Liu and Saven labs applied CPD for the design of a solubilized G-protein coupled receptor (GPCR)—the  $\mu$ -opioid receptor [107]. The pain and addiction receptor underwent 53 mutations on the exterior surface solubilizing it completely without loss of structural characteristics and antagonist (naltrexone) binding affinity. Interestingly, the CPD was not conducted on a high-resolution known structure but rather on a comparative model using the  $\beta_2$  adrenergic receptor as a model with the subsequent structure of the murine  $\mu$ -opioid receptor validating the model. Amino acids with >40 % solvent accessible surface area that were within the TM region were targeted for redesign within the SCADS framework [143] and the previous solubilization protocol [38]. To account for solvation effects, an environmental effective energy was employed based on the local density of  $C_\beta$  atoms of each residue and parameterized using a dataset of soluble proteins having up to 288 residues, the size of the TM domain of the targeted receptor. In 2014 five labs from the USA and South Korea (Johnson, Lieu, Saven, Park, Xi) joined forces and implemented this solubilized opioid receptor within a graphene field effect transistor (GRET) biosensor [108]. The receptor exhibited high sensitivity and selectivity for an opioid receptor antagonist (naltrexone), with an impressive detection limit of 10 pg/mL. The approach is general and can be applied for any GPCR, the family of proteins which form most drug targets and which suffers from experimental challenges following their intrinsic dynamics and embedment in the membrane.

In 2013 Baker and colleagues applied CPD for the design of a de novo lysozyme inhibitor [109]. Unlike the dock and design approach, e.g. the CPD of a weak affinity binder for PAK1 [82], here a hot-spot centric CPD approach was applied. This approach was previously applied to design proteins that bind the erythropoietin receptor [63] or the influenza hemagglutinin [86]. Here, the challenge included targeting deeply recessed residues within the charged active site of hen egg lysozyme (HEL). First a dock-and design approach was pursued: Coarse-docking was conducted on the HEL active site from a library of scaffold followed by several rounds of refined docking using RosettaDesign. Designed potentially binding proteins were analyzed as to binding energetics, shape-complementarity, packing, and size, aiming at measurables similar to native HEL complexes. The top 24 designs were displayed in a yeast library assessing binding affinity and specificity. Interestingly, the top-binder appeared to bind via a patch that is different than the one designed computationally, as evident from error-prone PCR affinity maturation which yielded affinity increasing mutations in other regions. Following these rarely reported negative results, a hot-spot centric approach was applied: An existing HEL complex was studied with computational alanine scanning

finding residues significantly contributing to binding and targeting active site residues. The two binding residues (Arg and Tyr) were held fixed and scaffolds were docked on them using PatchDock [153] followed by RosettaDock refinement. The two binding residues were transplanted on the scaffold with the aid of rigid-body minimization and the surrounding residues were designed with RosettaDesign. The top 21 designs were experimentally tested for affinity and specificity and the top design was optimized by error-prone PCR in a yeast display framework. From analysis of the best binder displaying low nanomolar affinity, it was concluded that specific interactions across a rather large interface are pivotal. In addition, it seems that the directed evolution experimental approach corrected poor hydrogen-bonding and electrostatic repulsion that was not sufficiently optimized by the CPD, suggesting room for algorithmic improvement.

In 2013 Baker and coworkers applied CPD for the de novo design of selective binders to the steroid digoxigenin (DIG), an example of a small molecule to which a protein binder can be designed [110]. The CPD of small molecule binders is challenging and indeed only two of 17 designs bound the molecule. Deep sequencing and library selections optimized the binding to picomolar levels. Three characteristics of naturally occurring binding sites were aimed: shape complementarity, specific energetically favorable hydrogen-bonds and van der Waals protein–ligand interactions as well as a structural pre-organization in the unbound protein state, which minimized entropy loss upon ligand binding. RosettaMatch [150] was used to identify backbone constellations in 401 protein scaffold structures where a DIG molecule and side chain conformations interacting with DIG in a predefined geometry could be accommodated. Two successive rounds of sequence design were used. The purpose of the first was to maximize binding affinity for the ligand. The goal of the second was to minimize protein destabilization due to aggressive scaffold mutagenesis while maintaining the binding interface designed during the first round. During the latter round, ligand–protein interactions were up-weighted by a factor of 1.5 relative to intra-protein interactions to ensure that binding energy was preserved. No more than five residues were allowed to change from residue types observed in a multiple sequence alignment (MSA) of the scaffold if (a) these residues were present in the MSA with a frequency greater than 0.6, or (b) if the calculated  $\Delta\Delta G$  for mutation of the scaffold residue to alanine was large. Designs were evaluated as to their interface energy, ligand solvent exposed surface area, ligand orientation, shape-complementarity, and apo-protein binding site pre-organization. The latter was enforced by explicitly introducing second-shell amino acids. The binding affinity of the directed evolution optimized design is similar to those of anti-digoxin antibodies. As it is stable for extended periods and can be expressed



at high levels in bacteria, the design has the potential to provide a more cost-effective alternative for biotechnological and for therapeutic purposes as long as it can be made compatible with the human immune response.

In 2014 the Baker lab designed a pH-sensitive Fc-domain IgG binding protein using the hot-spot centric approach [111]. His-433 on the IgG domain was targeted as a pH-sensitive site that should bind only under a specific pH range. Ensembles of disembodied interaction residues were based on the IgG complex with protein A. Scaffolds with high bacterial expression and solubility that can host the keystone residues were then searched. The rest of the interface was designed with RosettaDesign with ranking assisted by shape-complementarity and computed binding energy. Nine of 17 designs exhibited binding signals. At pH 8.2 the design bound the target 500-fold more tightly compared to pH 5.5.

In 2014 Liu, Chen and coworkers presented a new CPD method with a comprehensive statistical energy function (SEF) and systematic integration of experimental selection for foldability which was proven experimentally on two de novo structurally resolved designs [112]. In this important paper they highlight some of the challenges of existing rule-based or general-CPD methods, the latter minimizing a general effective energy function. Challenges include low success-rate on common targets, insufficient reflection of the diversity in natural sequences sharing a common structure and lack of the rich functional conformational dynamics in CPD results. While SEFs are an integral part of numerous CPD methods, a full-scale SEF for automated CPD is not available as most general methods focus on physics-based energy functions. SEFs share the spirit of rule-based CPD, though the latter can include very few components which are not well calibrated between them. As such, the rule-based design, which often necessitates a human expert, receives here a systematic and coherent formalism. The SEF components including single-residue and pairwise terms with individual terms were determined by the probability distributions of rotamer types and pairs of rotamer types. Complementary, structure properties considered for single positions include secondary structure types, solvent accessibility, and backbone Ramachandran angles. Structural properties of pair terms also include the relative positioning in 3D space. Next, a general strategy for selecting structure neighbors with adaptive criteria (*SSNAC*) addressed the fact that some target properties are at the boundary of predefined boundary intervals and the need to treat multi-dimensional properties jointly. Small sample effects were corrected. Further, the publication aimed to establish the general applicability of an experimental approach assessing structural stability by linking it to antibiotic resistance in bacterial cells expressing an engineered TEM1- $\beta$ -lactamase fused to the protein of interest. Unstable proteins are prone to proteolysis leading to weak antibiotic resistance. Comparing the SEF to fixed-backbone to

RosettaDesign, the authors claim that the SEF captures energy contributions that favor native sequences. The authors note that the SEF approach cannot treat packing in the same level as physics-based approaches, but seems to do a better job in capturing topology-related features, especially for  $\beta$ -strand containing topologies. Four well-folded de novo proteins for three different targets were obtained and two were structurally resolved validating the promising approach.

In 2014 the Baker lab applied CPD for the design of hyper-stable helical bundles [113]. Specifically, using Rosetta along with parametric backbone generation an antiparallel, monomeric untwisted three-helix bundle with 80-residue helices (18-residue repeat) was designed as well as an antiparallel right-handed monomeric four-helix bundle and a parallel left-handed five-helix bundle. While the classical coiled-coil structure is considered as a side-chain ‘knobs-into-holes’ structure, here the focus was on the less-appreciated contribution of backbone strain. Within the coiled-coil Crick parameters, a change of  $2^\circ$  in the helical twist and the coupled supercoil parameter can dictate the coiled coil twisting or lack of it. Within RosettaDesign, finer grid searches were undertaken in the vicinity of these parameters, yielding optimized designs. The resulting designs denatured only in  $>95^\circ\text{C}$  with 0.4–1.1 Å RMSD between the crystallographically resolved structures and the designs.

In 2014 Woolfson applied CPD for designing water-soluble  $\alpha$ -helical barrels [114]. These are coiled-coils with more than four helices which form a central cavity. Within the *abcdefg* heptad repeat of coiled coils positions *gade* determine the oligomer state. As such, these positions were the focus of the design with specific positions relating to the requested coiled coil type. A bZIP scoring function was used to assess the fitness score of the homo-oligomer. Sequential rules were applied to reduce the set to be sampled and then Coiled Coil Builder (*CCBuilder*) was applied to construct the requested full-atom models. This includes the SOCKET knobs-into-holes packing assessment. Next, a genetic algorithm was applied to optimize radius, pitch, and inter-helical rotational offset. The designed pentamer, hexamer, and heptamer coiled coil were resolved crystallographically with RMSDs of 0.67–1.77 Å between the design and the actual structure.

In 2014 Negron and Keating combined the CLASSY [75] multi-state CPD and the distance-scaled, finite-gas reference (DFIRE [163]) state potential for de novo CPD of three coiled coils consisting three orthogonal antiparallel homodimers [115]. The heptad repeat coiled coil structure enabled the multi-state design scheme to provide a partition function between the stability and the specificity gap; allowing for the design of novel and experimentally prove 43-residue peptides folding into specific antiparallel homodimers. As such, a synthetic coiled-coil toolkit is provided for modular synthetic biology applications.

In 2014 the Schief lab collaborated with Baker and others to apply CPD for the important cause of epitope-focused vaccine design [116]. Their 27-author study focused on inducing potent neutralizing antibodies to small and stable CPD scaffolds which present a respiratory syncytial virus (RSV) epitope. The fold-from-loops (FFL) CPD Rosetta protocol starts by identifying a functional motif (epitope), which in this case was a helix-turn-helix motif in the RSV Fusion (F) glycoprotein, as identified from an antigen-antibody crystal structure. The epitope was placed on a target topology along with distance restraints of the scaffold, a thermally stable three-helix bundle. Then, *ab initio* folding was applied to build diverse backbone conformations consistent with the target topology. Successful low-resolution designs were subjected to an all-atom sequence design in which functional motif side chains were recovered followed by three rounds of sequence design and full-atom optimization. Last but not least, the 40,000 successful designs were evaluated by structural metrics and 8 designs were subjected to human-guided sequence design to correct potential flaws. These included replacing surface residues outside the epitope with the original template residues and designing larger hydrophobic residues at selected positions. One of the designs also underwent resurfacing (described above). The successful design induced neutralizing antibodies and was recognized by an existing antibody against the epitope.

In 2014 the DeGrado lab joined forces with three other labs, applying CPD for a *de novo* TM  $\text{Zn}^{2+}$ -transporting four-helix bundle [117]. The protein was named *ROCKER*. The first shell of the metal binding was inspired by a previous di-manganese four helix bundle while the second shell was adapted from that soluble structure for the TM milieu. A stochastic search over the helix-bundle Crick parameters was applied for a D2-symmetric anti-parallel tetrameric coiled-coil. A design alphabet was guided by the membrane depth (using the Ez potential [164]) and functional requirements of the different regions. Rotameric self and pair energies were computed with a van der Waals radii reduced to 90 % of their size with the optimal rotameric conformation searched using a DEE/A\* algorithm. 1008 resulting sequences had a preference for an asymmetric state, excluding the transporter from being filled with two ions. To confirm an asymmetric rather than symmetric conformation, each of these sequences was subjected to the two-state free-energy comparison evaluator algorithm VALOCIDY (Valuation of Local Configuration Integral with Dynamics [165]) using independent MD trajectories. The protein was extensively characterized structurally and functionally, confirming the CPD models.

In 2014 Baker and coworkers applied CPD for reducing immunogenicity by removing T-cell epitopes [118]. As proteins represent the fastest-growing class of pharmaceuticals, their deimmunization

is of growing need. MHC-II-binding short-sequence epitopes have been characterized. Herein, a sliding window of 15-residues was searched using a support vector machine (SVM) for T-cell epitopes. These were searched and potential epitope sites were redesigned without losing structure, stability, and function. As the deimmunization scores favor negatively charged residues, a net charge constraint was added. First, they computationally recapitulated a previous deimmunization effort. Second, the method was experimentally validated on the superfolder green fluorescent protein (sfGFP) by redesigning the top four predicted H-2-IAb epitopes. The deimmunized protein designs failed to isolate T cells in mice while maintaining function. Third, 5 mutations were aimed at removing 3 epitopes in the toxin domain of the cancer therapeutic HA22, a potential drug for refractory cell leukemia. Two of these mutants lost 80 % of the cytotoxic effect while other mutants displayed increased effect.

In 2014 Zhang, Tame and coworkers applied CPD for the design of a self-assembling sixfold perfectly symmetric  $\beta$ -propeller protein [119]. Visual examination of 174  $\beta$ -propeller proteins was applied to choose the most visually symmetric protein for design. Therein, ancestor reconstruction of one of the six blades was applied followed by reverse engineering of a 6-blade protein. The process included docking of the blades and side-chain design in which essential inter-blade interacting residues were left as is. The actual design was experimentally proven to have an excellent 0.68 Å-backbone RMSD to the designed model.

In 2014 the Andre lab designed a leucine-rich repeat from the ribonuclease inhibitor family with predefined geometry [120]. Designated software was utilized to determine the length, curvature, and twist geometrical features. The protocol first defined the desired protein geometry. Second, a library of structures of individual repeats was compiled from crystal structures of selected repeat proteins. Third, self-compatible repeats capable of symmetrical assembly were selected. Fourth, the inter-repeat interface was optimized by cycles of docking and sequence optimization. Fifth, consecutive repeats were connected by loops. Last, capping was added to most N- and C-terminal repeats. A five double-repeat protein was confirmed to fold into a novel ring for the cap-less design and to a well-defined repeat protein when the caps were included.

---

## 5 CPD Failed Efforts and Retractions

Description of achievements and challenges of CPD cannot be complete without mentioning cases in which CPD publications were retracted. Naturally, published science highlights success stories rather than failures. Nevertheless, in some cases the failed attempt to repeat a published study results in exposing an

erroneous or disputed scientific publication. The need to analyze and understand failed efforts was highlighted by Mayo [100] in his description of an iterative design cycle: “*Proteins from failed computational design efforts are typically discarded without comment or investigation into the cause of failure. This situation is unfortunate, because valuable information is lost when successful designs are reported. Without detailed computational and/or experimental analysis of failed designs, flaws in the design procedure cannot be identified and remedied.*”

The field of protein design had suffered from several such incidents, partly as the proof of the output protein is not always straightforward. The resulting retracted publications may be due to innocent mistakes, insufficient validation or potentially even cheating in reporting the research. This section aims to present key retractions without getting into the details underlying the retractions. Rather, such retractions remind us of the caution required in reporting CPD studies and the need to unequivocally validate the result of the CPD process.

In 2008 Dwyer, Looger, and Hellinga retracted [166] their 2004 *Science* [167] publication which attempted to describe the first computational enzyme design, a triose phosphate isomerase (TIM) in a computationally redesigned ribose-binding protein. The retraction states that this is following a report that the provided clones that were supposed to be clones of the designed enzyme were actually clones of *wild-type* TIM impurity. In addition, a *JMB* computational enzyme design publication by the same group was retracted [168]. Following these retractions questions arose [169, 170] including over the validity of a 2003 *Nature* paper describing computational redesign of ligand-binding specificities [171] and a 2004 *PNAS* paper describing the CPD of receptors for an organophosphate surrogate of the nerve agent soman [172]. Notably, these papers were not retracted. Importantly, Hellinga has acknowledged responsibility for the two retractions and asked his university to hold an inquiry regarding them [173].

Unfortunately, retractions in the field of protein design are not limited to CPD. For example, following cross-contamination, in 2002 Fersht and coworkers have retracted [174] their *Nature* paper [175] on the directed evolution of new catalytic activity using the  $\alpha/\beta$ -barrel scaffold.

In summary, these retractions following irreproducible results and the heated debate that followed should remind us of the special care required in experimentally characterizing and confirming that the CPD product is indeed the designed protein.

---

## 6 Concluding Remarks: Future Challenges

Many aspects of CPD has been reviewed in the past [121, 122, 176–182], yet a chronological case-study review of the field is presented here for the first time. The field of CPD has undergone a tremendous leap forward in the three decades in which it exists. CPD demonstrates the ability to design functional and extremophile complex proteins with great precision using a wide array of tailored methods as well as imported methods from other fields. Taken together, it seems that the achievements and challenges of the CPD field reflect that of the broader structural bioinformatics and computational biophysics [183] field.

Some of the pending challenges include:

1. Accessibility to the general relevant scientific community. Thus far, the main efforts in the field of CPD were not distributed among a large community but rather clustered in a small number of labs (*see* Table 3 for list of main labs and software packages). Often, the CPD software packages are used solely ‘*in-house*’ and not utilized by the general community, even if the software is open-source. CPD requires multidisciplinary know-how in structural biology, biophysics, biochemistry, software engineering, and a general nontrivial combination of theory and experiment. As with other fields, it is expected that with time more and more scientists will apply CPD for their research and consequently use software developed by others.
2. Integration of knowledge-based and energy-based methods: Ideally, all design algorithms will rely on physics to address the enthalpic and entropic energetic contributions. Yet, within the complex protein milieu and within the foreseeable future of computer power, such a description is not practical in high resolution. Currently, it seems that each design lab selects a different method of integrating knowledge-based know-how into the design—from selection of hydrophobic or helix-forming amino acids to use of known structural motifs or structural fragments. A systematic and comparative analysis of the different design schemes may help determine better guidelines on this aspect.
3. Systematic differential approach towards different proteins levels of organization, different protein regions, and the relationships between such regions. While often the design is split to solvent-exposed and buried regions, the adaption of the CPD algorithm to the local milieu of the target site is still not optimized.
4. Assessment of electrostatics and solvation effects: Coupled to the previous item, the local dielectric milieu and long-range

**Table 3**

List of main labs discussed in this chapter and their CPD software packages. Notably, only the main software packages for CPD are described here. Further, only the main and most known software of each lab is mentioned with some of the specific case-studies of each lab not necessarily using that software. Additional software for different aspects of CPD are described throughout the book. For a brief description of the main early-stage software packages see [184]

Year	Lab	Software	Publication of main software	Lab publications discussed in the chapter
1. 1985–	DeGrado	ProtCad for few studies but most CPD case-studies with numerous other tools.	[160]	[7, 9, 22, 23, 29, 38, 39, 41, 42, 47, 48, 51, 54, 55, 61, 65, 70, 78, 81, 85, 92, 103, 104, 106, 117, 123, 127, 144, 147, 164]
2. 1987	Ponder	PROPAK	[126]	
3. 1991	Hellinga	DEZYMER	[125]	[11, 125, 166, 167, 171–173]
4. 1995	Desjarlais	Repacking of Core (ROC) & SPA (Sequence Prediction Algorithm)	[15, 128]	[15–17, 22, 44, 128, 185]
5. 1996, 1997	Mayo	PDA (Protein Design Automation) and ORBIT (Optimization of Rotamers by Iterative Techniques), Other programs include Phoenix, Triad, Faster	[18, 141]	[18–20, 27, 31, 32, 60, 64, 80, 83, 100, 101, 141]
6. 2000	Baker	Rosetta (including RosettaDesign, RosettaMatch etc.)	[133, 134, 150]	[25, 36, 43, 45, 46, 49, 50, 52, 53, 58, 71–73, 76, 79, 86, 87, 95, 98, 109–111, 113, 116, 118, 133, 134, 138–140, 142, 149, 150]
7. 2000	Kuhlman	RosettaDesign	[134]	[25, 43, 46, 62, 82, 88, 134]
8. 2001	Saven	SCADS (Statistical Computationally Assisted Design Strategy)	[143]	[41, 51, 54–56, 68, 70, 78, 90, 94, 103, 107, 121, 143, 146, 164]
9. 2001	Serrano	PERLA (protein Engineering Rotamer Library Algorithm)	[135]	[24, 26–30, 57, 135, 136, 182]
10. 2014	Chen & Liu	Statistical energy function and boosted by experimental selection for foldability	[112]	[112]



electrostatic interactions are still not sufficiently modeled within CPD software.

5. Integration of thermal plasticity and functional dynamics: While a generalization, the incorporation of dynamics into the design scheme is still not done, despite the hard-wired dynamic functional profile of every protein as e.g. depicted by quick Gaussian network models.
6. Negative design: Negative design, defined as a design aimed at avoiding unwanted conformations or functions, must be an explicit part of computational design. Since the 1991 thioredoxin redesign [11] and the betadoublet, a  $\beta$ -sandwich de novo design [14], the negative design aspect has been in the forefront of the field. While the importance of negative design is well acknowledged since early days of CPD [185], it is still not explicitly integrated into design algorithms. In this respect, the positive-design scheme explicitly or implicitly regards a reference state which can often be considered as a negative design element. However, too often insufficient emphasis is given to the definition of the reference state.
7. Systematic integration of experimental design approaches: the theoretical rational design is moving towards integration with experimental semi-rational design approaches such as directed evolution. Yet, currently the number of designs benefiting from the combination of approaches is still small. Moreover, there is no systematic protocol for combining the two approaches or even for reporting the stage to which each approach has advanced the target design.
8. Objective cross-assessment of methods: To date, there has not been an objective cross-assessment of the different available methods, as done for e.g. structure prediction via the Critical Assessment of Structure Prediction (CASP) competition [186] which is running since 1995. Therein, the community is given a mutual target to be submitted to assessors who are not part of the competitors thus enabling objective analysis of achievements and challenges in a method comparative manner. Without such a community-wide objective assessment the comparative analysis of CPD methods is often challenging relying solely on reports by the respective authors for each tool. Consequently, the identification of advantages and disadvantages of each method and the cross-dissemination of knowledge is hampered.
9. Definition of the reference state: In many cases the scoring function consists of scoring the gap between the desired state and the nondesired, e.g. denatured one. However, the reference state is still not sufficiently defined, let alone divided between protein and cellular regions.



10. *In vivo* CPD: Many designs are not stable and prone to aggregation [111]. As seen from the case-studies presented, the vast majority of designs were not characterized within an *in vivo* setting, which is the ultimate natural environment of proteins.

Each of the above items deserves a separate chapter. Yet, after highlighting some of the pending challenges, it is important to emphasize that the hierarchical approach to CPD has advanced in all levels—from large scaffold searches in the growing PDB to quantum-mechanical optimization of enzymatic catalytic sites. In parallel the richness in knowledge-based and physics-based methodology sets the stage to comparative analysis of methods and the dissemination of methods from the method creators to the general community of protein scientists.

## References

1. Fischer E (1966) In: Nobelstiftelsen (ed) Nobel lectures, chemistry 1901–1921, vol 1. Elsevier, Amsterdam, p 21–35
2. Anfinsen CB, Harrington WF, Hvidt A, Linderstrom-Lang K, Ottesen M, Schellman J (1989) Studies on the structural basis of ribonuclease activity. 1955. *Biochim Biophys Acta* 1000:200–201
3. Drexler KE (1981) Molecular engineering: an approach to the development of general capabilities for molecular manipulation. *Proc Natl Acad Sci U S A* 78(9):5275–5278
4. Pabo C (1983) Molecular technology. Designing proteins and peptides. *Nature* 301(5897):200
5. Jaramillo A, Wernisch L, Hery S, Wodak SJ (2002) Folding free energy function selects native-like protein sequences in the core but not on the surface. *Proc Natl Acad Sci U S A* 99(21):13554–13559. doi:10.1073/pnas.212068599
6. Wernisch L, Hery S, Wodak SJ (2000) Automatic protein design with all atom force-fields by exact and heuristic optimization. *J Mol Biol* 301(3):713–736. doi:10.1006/jmbi.2000.3984
7. DeGrado WF, Prendergast FG, Wolfe HR Jr, Cox JA (1985) The design, synthesis, and characterization of tight-binding inhibitors of calmodulin. *J Cell Biochem* 29(2):83–93. doi:10.1002/jcb.240290204
8. Craik CS, Largman C, Fletcher T, Rocznik S, Barr PJ, Fletterick R, Rutter WJ (1985) Redesigning trypsin: alteration of substrate specificity. *Science* 228(4697):291–297
9. Vonderviszt F, Matrai G, Simon I (1986) Characteristic sequential residue environment of amino acids in proteins. *Int J Pept Protein Res* 27(5):483–492
10. Hecht MH, Richardson JS, Richardson DC, Ogden RC (1990) De novo design, expression, and characterization of Felix: a four-helix bundle protein of native-like sequence. *Science* 249(4971):884–891
11. Hellinga HW, Caradonna JP, Richards FM (1991) Construction of new ligand binding sites in proteins of known structure. II. Grafting of a buried transition metal binding site into *Escherichia coli* thioredoxin. *J Mol Biol* 222(3):787–803
12. Wilson C, Mace JE, Agard DA (1991) Computational method for the design of enzymes with altered substrate specificity. *J Mol Biol* 220(2):495–506
13. Hurley JH, Baase WA, Matthews BW (1992) Design and structural analysis of alternative hydrophobic core packing arrangements in bacteriophage T4 lysozyme. *J Mol Biol* 224(4):1143–1159
14. Quinn TP, Tweedy NB, Williams RW, Richardson JS, Richardson DC (1994) Beta-doublet: de novo design, synthesis, and characterization of a beta-sandwich protein. *Proc Natl Acad Sci U S A* 91(19):8747–8751
15. Desjarlais JR, Handel TM (1995) De novo design of the hydrophobic cores of proteins. *Protein Sci* 4(10):2006–2018. doi:10.1002/pro.5560041006
16. Lazar GA, Desjarlais JR, Handel TM (1997) De novo design of the hydrophobic core of ubiquitin. *Protein Sci* 6(6):1167–1178. doi:10.1002/pro.5560060605
17. Johnson EC, Lazar GA, Desjarlais JR, Handel TM (1999) Solution structure and dynamics

- of a designed hydrophobic core variant of ubiquitin. *Structure* 7(8):967–976
18. Dahiyat BI, Mayo SL (1997) De novo protein design: fully automated sequence selection. *Science* 278(5335):82–87
  19. Dahiyat BI, Sarisky CA, Mayo SL (1997) De novo protein design: towards fully automated sequence selection. *J Mol Biol* 273(4):789–796. doi:[10.1006/jmbi.1997.1341](https://doi.org/10.1006/jmbi.1997.1341)
  20. Malakauskas SM, Mayo SL (1998) Design, structure and stability of a hyperthermophilic protein variant. *Nat Struct Biol* 5(6):470–475
  21. Harbury PB, Plescs JJ, Tidor B, Alber T, Kim PS (1998) High-resolution protein design with backbone freedom. *Science* 282(5393):1462–1467
  22. Bryson JW, Desjarlais JR, Handel TM, DeGrado WF (1998) From coiled coils to small globular proteins: design of a native-like three-helix bundle. *Protein Sci* 7(6):1404–1414. doi:[10.1002/pro.5560070617](https://doi.org/10.1002/pro.5560070617)
  23. Walsh ST, Cheng H, Bryson JW, Roder H, DeGrado WF (1999) Solution structure and dynamics of a de novo designed three-helix bundle protein. *Proc Natl Acad Sci U S A* 96(10):5486–5491
  24. Domingues H, Cregut D, Sebald W, Oschkinat H, Serrano L (1999) Rational design of a GCN4-derived mimetic of interleukin-4. *Nat Struct Biol* 6(7):652–656. doi:[10.1038/10706](https://doi.org/10.1038/10706)
  25. Kuhlman B, O'Neill JW, Kim DE, Zhang KY, Baker D (2001) Conversion of monomeric protein L to an obligate dimer by computational protein design. *Proc Natl Acad Sci U S A* 98(19):10687–10691. doi:[10.1073/pnas.181354398](https://doi.org/10.1073/pnas.181354398)
  26. Lopez de la Paz M, Lacroix E, Ramirez-Alvarado M, Serrano L (2001) Computer-aided design of beta-sheet peptides. *J Mol Biol* 312(1):229–246. doi:[10.1006/jmbi.2001.4918](https://doi.org/10.1006/jmbi.2001.4918)
  27. Bolon DN, Mayo SL (2001) Enzyme-like proteins by computational design. *Proc Natl Acad Sci U S A* 98(25):14274–14279. doi:[10.1073/pnas.251555398](https://doi.org/10.1073/pnas.251555398)
  28. Keating AE, Malashkevich VN, Tidor B, Kim PS (2001) Side-chain repacking calculations for predicting structures and stabilities of heterodimeric coiled coils. *Proc Natl Acad Sci U S A* 98(26):14825–14830. doi:[10.1073/pnas.261563398](https://doi.org/10.1073/pnas.261563398)
  29. Summa CM, Rosenblatt MM, Hong JK, Lear JD, DeGrado WF (2002) Computational de novo design, and characterization of an A(2)B(2) diiron protein. *J Mol Biol* 321(5):923–938
  30. Ventura S, Vega MC, Lacroix E, Angrand I, Spagnolo L, Serrano L (2002) Conformational strain in the hydrophobic core and its implications for protein folding and design. *Nat Struct Biol* 9(6):485–493. doi:[10.1038/nsb799](https://doi.org/10.1038/nsb799)
  31. Shifman JM, Mayo SL (2002) Modulating calmodulin binding specificity through computational protein design. *J Mol Biol* 323(3):417–423
  32. Shifman JM, Mayo SL (2003) Exploring the origins of binding specificity through the computational redesign of calmodulin. *Proc Natl Acad Sci U S A* 100(23):13274–13279. doi:[10.1073/pnas.2234277100](https://doi.org/10.1073/pnas.2234277100)
  33. Hayes RJ, Bentzien J, Ary ML, Hwang MY, Jacinto JM, Vielmetter J, Kundu A, Dahiyat BI (2002) Combining computational and experimental screening for rapid optimization of protein properties. *Proc Natl Acad Sci U S A* 99(25):15926–15931. doi:[10.1073/pnas.212627499](https://doi.org/10.1073/pnas.212627499)
  34. Filikov AV, Hayes RJ, Luo P, Stark DM, Chan C, Kundu A, Dahiyat BI (2002) Computational stabilization of human growth hormone. *Protein Sci* 11(6):1452–1461. doi:[10.1110/ps.3500102](https://doi.org/10.1110/ps.3500102)
  35. Luo P, Hayes RJ, Chan C, Stark DM, Hwang MY, Jacinto JM, Juvvadi P, Chung HS, Kundu A, Ary ML, Dahiyat BI (2002) Development of a cytokine analog with enhanced stability using computational ultrahigh throughput screening. *Protein Sci* 11(5):1218–1226. doi:[10.1110/ps.4580102](https://doi.org/10.1110/ps.4580102)
  36. Chevalier BS, Kortemme T, Chadsey MS, Baker D, Monnat RJ, Stoddard BL (2002) Design, activity, and structure of a highly specific artificial endonuclease. *Mol Cell* 10(4):895–905
  37. Ogata K, Jaramillo A, Cohen W, Briand JP, Connan F, Choppin J, Muller S, Wodak SJ (2003) Automatic sequence design of major histocompatibility complex class I binding peptides impairing CD8+ T cell recognition. *J Biol Chem* 278(2):1281–1290. doi:[10.1074/jbc.M206853200](https://doi.org/10.1074/jbc.M206853200)
  38. Slovic AM, Summa CM, Lear JD, DeGrado WF (2003) Computational design of a water-soluble analog of phospholamban. *Protein Sci* 12(2):337–348. doi:[10.1110/ps.0226603](https://doi.org/10.1110/ps.0226603)
  39. Slovic AM, Stayrook SE, North B, Degradó WF (2005) X-ray structure of a water-soluble analog of the membrane protein phospholamban: sequence determinants defining the topology of tetrameric and pentameric coiled coils. *J Mol Biol* 348(3):777–787. doi:[10.1016/j.jmb.2005.02.040](https://doi.org/10.1016/j.jmb.2005.02.040)

40. Havranek JJ, Harbury PB (2003) Automated design of specificity in molecular recognition. *Nat Struct Biol* 10(1):45–52. doi:[10.1038/nsb877](https://doi.org/10.1038/nsb877)
41. Calhoun JR, Kono H, Lahr S, Wang W, DeGrado WF, Saven JG (2003) Computational design and characterization of a monomeric helical dinuclear metalloprotein. *J Mol Biol* 334(5):1101–1115
42. Calhoun JR, Liu W, Spiegel K, Dal Peraro M, Klein ML, Valentine KG, Wand AJ, DeGrado WF (2008) Solution NMR structure of a designed metalloprotein and complementary molecular dynamics refinement. *Structure* 16(2):210–215. doi:[10.1016/j.str.2007.11.011](https://doi.org/10.1016/j.str.2007.11.011)
43. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science* 302(5649):1364–1368. doi:[10.1126/science.1089427](https://doi.org/10.1126/science.1089427)
44. Kraemer-Pecore CM, Lecomte JT, Desjarlais JR (2003) A de novo redesign of the WW domain. *Protein Sci* 12(10):2194–2205. doi:[10.1110/ps.03190903](https://doi.org/10.1110/ps.03190903)
45. Dantas G, Kuhlman B, Callender D, Wong M, Baker D (2003) A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins. *J Mol Biol* 332(2):449–460
46. Dantas G, Corrent C, Reichow SL, Havranek JJ, Eletr ZM, Isern NG, Kuhlman B, Varani G, Merritt EA, Baker D (2007) High-resolution structural and thermodynamic analysis of extreme stabilization of human procarboxypeptidase by computational protein design. *J Mol Biol* 366(4):1209–1221. doi:[10.1016/j.jmb.2006.11.080](https://doi.org/10.1016/j.jmb.2006.11.080)
47. Di Costanzo L, Wade H, Geremia S, Randaccio L, Pavone V, DeGrado WF, Lombardi A (2001) Toward the de novo design of a catalytically active helix bundle: a substrate-accessible carboxylate-bridged dinuclear metal center. *J Am Chem Soc* 123(51):12749–12757
48. Kaplan J, DeGrado WF (2004) De novo design of catalytic proteins. *Proc Natl Acad Sci U S A* 101(32):11566–11570. doi:[10.1073/pnas.0404387101](https://doi.org/10.1073/pnas.0404387101)
49. Kortemme T, Joachimiak LA, Bullock AN, Schuler AD, Stoddard BL, Baker D (2004) Computational redesign of protein-protein interaction specificity. *Nat Struct Mol Biol* 11(4):371–379. doi:[10.1038/nsmb749](https://doi.org/10.1038/nsmb749)
50. Joachimiak LA, Kortemme T, Stoddard BL, Baker D (2006) Computational design of a new hydrogen bond network and at least a 300-fold specificity switch at a protein-protein interface. *J Mol Biol* 361(1):195–208. doi:[10.1016/j.jmb.2006.05.022](https://doi.org/10.1016/j.jmb.2006.05.022)
51. Slovic AM, Kono H, Lear JD, Saven JG, DeGrado WF (2004) Computational design of water-soluble analogues of the potassium channel KcsA. *Proc Natl Acad Sci U S A* 101(7):1828–1833. doi:[10.1073/pnas.0306417101](https://doi.org/10.1073/pnas.0306417101)
52. Korkegian A, Black ME, Baker D, Stoddard BL (2005) Computational thermostabilization of an enzyme. *Science* 308(5723):857–860. doi:[10.1126/science.1107387](https://doi.org/10.1126/science.1107387)
53. Bolon DN, Grant RA, Baker TA, Sauer RT (2005) Specificity versus stability in computational protein design. *Proc Natl Acad Sci U S A* 102(36):12724–12729. doi:[10.1073/pnas.0506124102](https://doi.org/10.1073/pnas.0506124102)
54. Nanda V, Rosenblatt MM, Osyczka A, Kono H, Getahun Z, Dutton PL, Saven JG, DeGrado WF (2005) De novo design of a redox-active minimal rubredoxin mimic. *J Am Chem Soc* 127(16):5804–5805. doi:[10.1021/ja050553f](https://doi.org/10.1021/ja050553f)
55. Cochran FV, Wu SP, Wang W, Nanda V, Saven JG, Therien MJ, DeGrado WF (2005) Computational de novo design and characterization of a four-helix bundle protein that selectively binds a nonbiological cofactor. *J Am Chem Soc* 127(5):1346–1347. doi:[10.1021/ja044129a](https://doi.org/10.1021/ja044129a)
56. Swift J, Wehbi WA, Kelly BD, Stowell XF, Saven JG, Dmochowski IJ (2006) Design of functional ferritin-like proteins with hydrophobic cavities. *J Am Chem Soc* 128(20):6611–6619. doi:[10.1021/ja057069x](https://doi.org/10.1021/ja057069x)
57. van der Sloot AM, Tur V, Szegezdi E, Mullally MM, Cool RH, Samali A, Serrano L, Quax WJ (2006) Designed tumor necrosis factor-related apoptosis-inducing ligand variants initiating apoptosis exclusively via the DR5 receptor. *Proc Natl Acad Sci U S A* 103(23):8634–8639. doi:[10.1073/pnas.0510187103](https://doi.org/10.1073/pnas.0510187103)
58. Ashworth J, Havranek JJ, Duarte CM, Sussman D, Monnat RJ Jr, Stoddard BL, Baker D (2006) Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature* 441(7093):656–659. doi:[10.1038/nature04818](https://doi.org/10.1038/nature04818)
59. Lazar GA, Dang W, Karki S, Vafa O, Peng JS, Hyun L, Chan C, Chung HS, Eivazi A, Yoder SC, Vielmetter J, Carmichael DF, Hayes RJ, Dahiyat BI (2006) Engineered antibody Fc variants with enhanced effector function. *Proc Natl Acad Sci U S A* 103

- (11):4005–4010. doi:[10.1073/pnas.0508123103](https://doi.org/10.1073/pnas.0508123103)
60. Huang PS, Love JJ, Mayo SL (2007) A de novo designed protein protein interface. *Protein Sci* 16(12):2770–2774. doi:[10.1110/ps.073125207](https://doi.org/10.1110/ps.073125207)
  61. Yin H, Slusky JS, Berger BW, Walters RS, Vilaire G, Litvinov RI, Lear JD, Caputo GA, Bennett JS, DeGrado WF (2007) Computational design of peptides that target transmembrane helices. *Science* 315(5820):1817–1822. doi:[10.1126/science.1136782](https://doi.org/10.1126/science.1136782)
  62. Hu X, Wang H, Ke H, Kuhlman B (2007) High-resolution design of a protein loop. *Proc Natl Acad Sci U S A* 104(45):17668–17673. doi:[10.1073/pnas.0707977104](https://doi.org/10.1073/pnas.0707977104)
  63. Liu S, Liu S, Zhu X, Liang H, Cao A, Chang Z, Lai L (2007) Nonnatural protein-protein interaction-pair design by key residues grafting. *Proc Natl Acad Sci U S A* 104(13):5330–5335. doi:[10.1073/pnas.0606198104](https://doi.org/10.1073/pnas.0606198104)
  64. Shah PS, Hom GK, Ross SA, Lassila JK, Crowhurst KA, Mayo SL (2007) Full-sequence computational design and solution structure of a thermostable protein variant. *J Mol Biol* 372(1):1–6. doi:[10.1016/j.jmb.2007.06.032](https://doi.org/10.1016/j.jmb.2007.06.032)
  65. Bender GM, Lehmann A, Zou H, Cheng H, Fry HC, Engel D, Therien MJ, Blasie JK, Roder H, Saven JG, DeGrado WF (2007) De novo design of a single-chain diphenylporphyrin metalloprotein. *J Am Chem Soc* 129(35):10732–10740. doi:[10.1021/ja071199j](https://doi.org/10.1021/ja071199j)
  66. Lippow SM, Wittrup KD, Tidor B (2007) Computational design of antibody-affinity improvement beyond in vivo maturation. *Nat Biotechnol* 25(10):1171–1176. doi:[10.1038/nbt1336](https://doi.org/10.1038/nbt1336)
  67. Potapov V, Reichmann D, Abramovich R, Filchtinski D, Zohar N, Ben Halevy D, Edelman M, Sobolev V, Schreiber G (2008) Computational redesign of a protein-protein interface for high affinity and binding specificity using modular architecture and naturally occurring template fragments. *J Mol Biol* 384(1):109–119. doi:[10.1016/j.jmb.2008.08.078](https://doi.org/10.1016/j.jmb.2008.08.078)
  68. Butts CA, Swift J, Kang SG, Di Costanzo L, Christianson DW, Saven JG, Dmochowski IJ (2008) Directing noble metal ion chemistry within a designed ferritin protein. *Biochemistry* 47(48):12729–12739. doi:[10.1021/bi8016735](https://doi.org/10.1021/bi8016735)
  69. Reynolds KA, Hanes MS, Thomson JM, Antczak AJ, Berger JM, Bonomo RA, Kirsch JF, Handel TM (2008) Computational redesign of the SHV-1 beta-lactamase/beta-lactamase inhibitor protein interface. *J Mol Biol* 382(5):1265–1275. doi:[10.1016/j.jmb.2008.05.051](https://doi.org/10.1016/j.jmb.2008.05.051)
  70. McAllister KA, Zou H, Cochran FV, Bender GM, Senes A, Fry HC, Nanda V, Keenan PA, Lear JD, Saven JG, Therien MJ, Blasie JK, DeGrado WF (2008) Using alpha-helical coiled-coils to design nanostructured metalloporphyrin arrays. *J Am Chem Soc* 130(36):11921–11927. doi:[10.1021/ja800697g](https://doi.org/10.1021/ja800697g)
  71. Jiang L, Althoff EA, Clemente FR, Doyle L, Rothlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF 3rd, Hilvert D, Houk KN, Stoddard BL, Baker D (2008) De novo computational design of retro-aldol enzymes. *Science* 319(5868):1387–1391. doi:[10.1126/science.1152692](https://doi.org/10.1126/science.1152692)
  72. Rothlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O, Albeck S, Houk KN, Tawfik DS, Baker D (2008) Kemp elimination catalysts by computational enzyme design. *Nature* 453(7192):190–195. doi:[10.1038/nature06879](https://doi.org/10.1038/nature06879)
  73. Murphy PM, Bolduc JM, Gallaher JL, Stoddard BL, Baker D (2009) Alteration of enzyme specificity by computational loop remodeling and design. *Proc Natl Acad Sci U S A* 106(23):9215–9220. doi:[10.1073/pnas.0811070106](https://doi.org/10.1073/pnas.0811070106)
  74. Yosef E, Politi R, Choi MH, Shifman JM (2009) Computational design of calmodulin mutants with up to 900-fold increase in binding specificity. *J Mol Biol* 385(5):1470–1480. doi:[10.1016/j.jmb.2008.09.053](https://doi.org/10.1016/j.jmb.2008.09.053)
  75. Grigoryan G, Reinke AW, Keating AE (2009) Design of protein-interaction specificity gives selective bZIP-binding peptides. *Nature* 458(7240):859–864. doi:[10.1038/nature07885](https://doi.org/10.1038/nature07885)
  76. Thyme SB, Jarjour J, Takeuchi R, Havranek JJ, Ashworth J, Scharenberg AM, Stoddard BL, Baker D (2009) Exploitation of binding energy for catalysis and design. *Nature* 461(7268):1300–1304. doi:[10.1038/nature08508](https://doi.org/10.1038/nature08508)
  77. Chen CY, Georgiev I, Anderson AC, Donald BR (2009) Computational structure-based redesign of enzyme activity. *Proc Natl Acad Sci U S A* 106(10):3764–3769. doi:[10.1073/pnas.0900266106](https://doi.org/10.1073/pnas.0900266106)
  78. Fry HC, Lehmann A, Saven JG, DeGrado WF, Therien MJ (2010) Computational design and elaboration of a de novo heterotetrameric alpha-helical protein that selectively binds an emissive abiological (porphyrinato)



- zinc chromophore. *J Am Chem Soc* 132 (11):3997–4005. doi:[10.1021/ja907407m](https://doi.org/10.1021/ja907407m)
79. Ashworth J, Taylor GK, Havranek JJ, Quadri SA, Stoddard BL, Baker D (2010) Computational reprogramming of homing endonuclease specificity at multiple adjacent base pairs. *Nucleic Acids Res* 38(16):5601–5608. doi:[10.1093/nar/gkq283](https://doi.org/10.1093/nar/gkq283)
80. Chica RA, Moore MM, Allen BD, Mayo SL (2010) Generation of longer emission wavelength red fluorescent proteins using computationally designed libraries. *Proc Natl Acad Sci U S A* 107(47):20257–20262. doi:[10.1073/pnas.1013910107](https://doi.org/10.1073/pnas.1013910107)
81. Korendovych IV, Senes A, Kim YH, Lear JD, Fry HC, Therien MJ, Blasie JK, Walker FA, Degrado WF (2010) De novo design and molecular assembly of a transmembrane diporphyrin-binding protein complex. *J Am Chem Soc* 132(44):15516–15518. doi:[10.1021/ja107487b](https://doi.org/10.1021/ja107487b)
82. Jha RK, Leaver-Fay A, Yin S, Wu Y, Butterfoss GL, Szyperki T, Dokholyan NV, Kuhlman B (2010) Computational design of a PAK1 binding protein. *J Mol Biol* 400(2):257–270. doi:[10.1016/j.jmb.2010.05.006](https://doi.org/10.1016/j.jmb.2010.05.006)
83. Allen BD, Nisthal A, Mayo SL (2010) Experimental library screening demonstrates the successful application of computational protein design to large structural ensembles. *Proc Natl Acad Sci U S A* 107(46):19838–19843. doi:[10.1073/pnas.1012985107](https://doi.org/10.1073/pnas.1012985107)
84. Frey KM, Georgiev I, Donald BR, Anderson AC (2010) Predicting resistance mutations using protein design algorithms. *Proc Natl Acad Sci U S A* 107(31):13707–13712. doi:[10.1073/pnas.1002162107](https://doi.org/10.1073/pnas.1002162107)
85. Grigoryan G, Kim YH, Acharya R, Axelrod K, Jain RM, Willis L, Drndic M, Kikkawa JM, DeGrado WF (2011) Computational design of virus-like protein assemblies on carbon nanotube surfaces. *Science* 332(6033):1071–1076. doi:[10.1126/science.1198841](https://doi.org/10.1126/science.1198841)
86. Fleishman SJ, Whitehead TA, Ekiert DC, Dreyfus C, Corn JE, Strauch EM, Wilson IA, Baker D (2011) Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* 332(6031):816–821. doi:[10.1126/science.1202617](https://doi.org/10.1126/science.1202617)
87. Karanicolas J, Corn JE, Chen I, Joachimiak LA, Dym O, Peck SH, Albeck S, Unger T, Hu W, Liu G, Delbecq S, Montelione GT, Spiegel CP, Liu DR, Baker D (2011) A de novo protein binding pair by computational design and directed evolution. *Mol Cell* 42 (2):250–260. doi:[10.1016/j.molcel.2011.03.010](https://doi.org/10.1016/j.molcel.2011.03.010)
88. Stranges PB, Machius M, Miley MJ, Tripathy A, Kuhlman B (2011) Computational design of a symmetric homodimer using beta-strand assembly. *Proc Natl Acad Sci U S A* 108 (51):20562–20567. doi:[10.1073/pnas.1115124108](https://doi.org/10.1073/pnas.1115124108)
89. Correia BE, Ban YE, Friend DJ, Ellingson K, Xu H, Boni E, Bradley-Hewitt T, Bruhn-Johannsen JF, Stamatatos L, Strong RK, Schief WR (2011) Computational protein design using flexible backbone remodeling and resurfacing: case studies in structure-based antigen design. *J Mol Biol* 405(1):284–297. doi:[10.1016/j.jmb.2010.09.061](https://doi.org/10.1016/j.jmb.2010.09.061)
90. Diaz JE, Lin CS, Kunishiro K, Feld BK, Avrantinis SK, Bronson J, Greaves J, Saven JG, Weiss GA (2011) Computational design and selections for an engineered, thermostable terpene synthase. *Protein Sci* 20 (9):1597–1606. doi:[10.1002/pro.691](https://doi.org/10.1002/pro.691)
91. Xu F, Zahid S, Silva T, Nanda V (2011) Computational design of a collagen A:B:C-type heterotrimer. *J Am Chem Soc* 133 (39):15260–15263. doi:[10.1021/ja205597g](https://doi.org/10.1021/ja205597g)
92. Korendovych IV, Kulp DW, Wu Y, Cheng H, Roder H, DeGrado WF (2011) Design of a switchable eliminase. *Proc Natl Acad Sci U S A* 108(17):6823–6827. doi:[10.1073/pnas.1018191108](https://doi.org/10.1073/pnas.1018191108)
93. Merski M, Shoichet BK (2012) Engineering a model protein cavity to catalyze the Kemp elimination. *Proc Natl Acad Sci U S A* 109 (40):16179–16183. doi:[10.1073/pnas.1208076109](https://doi.org/10.1073/pnas.1208076109)
94. Cui T, Mowrey D, Bondarenko V, Tillman T, Ma D, Landrum E, Perez-Aguilar JM, He J, Wang W, Saven JG, Eckenhoof RG, Tang P, Xu Y (2012) NMR structure and dynamics of a designed water-soluble transmembrane domain of nicotinic acetylcholine receptor. *Biochim Biophys Acta* 1818(3):617–626. doi:[10.1016/j.bbame.2011.11.021](https://doi.org/10.1016/j.bbame.2011.11.021)
95. Khare SD, Kipnis Y, Greisen P Jr, Takeuchi R, Ashani Y, Goldsmith M, Song Y, Gallaher JL, Silman I, Leader H, Sussman JL, Stoddard BL, Tawfik DS, Baker D (2012) Computational redesign of a mononuclear zinc metalloenzyme for organophosphate hydrolysis. *Nat Chem Biol* 8(3):294–300. doi:[10.1038/nchembio.777](https://doi.org/10.1038/nchembio.777)
96. Azoitei ML, Ban YE, Julien JP, Bryson S, Schroeter A, Kalyuzhnyi O, Porter JR, Adachi Y, Baker D, Pai EF, Schief WR (2012) Computational design of high-affinity epitope scaffolds by backbone grafting of a linear epitope. *J Mol Biol* 415(1):175–192. doi:[10.1016/j.jmb.2011.10.003](https://doi.org/10.1016/j.jmb.2011.10.003)

97. Kapp GT, Liu S, Stein A, Wong DT, Remenyi A, Yeh BJ, Fraser JS, Taunton J, Lim WA, Kortemme T (2012) Control of protein signaling using a computationally designed GTPase/GEF orthologous pair. *Proc Natl Acad Sci U S A* 109(14):5277–5282. doi:[10.1073/pnas.1114487109](https://doi.org/10.1073/pnas.1114487109)
98. Koga N, Tatsumi-Koga R, Liu G, Xiao R, Acton TB, Montelione GT, Baker D (2012) Principles for designing ideal protein structures. *Nature* 491(7423):222–227. doi:[10.1038/nature11600](https://doi.org/10.1038/nature11600)
99. Fallas JA, Hartgerink JD (2012) Computational design of self-assembling register-specific collagen heterotrimers. *Nat Commun* 3:1087. doi:[10.1038/ncomms2084](https://doi.org/10.1038/ncomms2084)
100. Privett HK, Kiss G, Lee TM, Blomberg R, Chica RA, Thomas LM, Hilvert D, Houk KN, Mayo SL (2012) Iterative approach to computational enzyme design. *Proc Natl Acad Sci U S A* 109(10):3790–3795. doi:[10.1073/pnas.1118082108](https://doi.org/10.1073/pnas.1118082108)
101. Blomberg R, Kries H, Pinkas DM, Mittl PR, Grutter MG, Privett HK, Mayo SL, Hilvert D (2013) Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature* 503(7476):418–421. doi:[10.1038/nature12623](https://doi.org/10.1038/nature12623)
102. Grzyb J, Xu F, Nanda V, Luczkowska R, Reijerse E, Lubitz W, Noy D (2012) Empirical and computational design of iron-sulfur cluster proteins. *Biochim Biophys Acta* 1817(8):1256–1262. doi:[10.1016/j.bbabbio.2012.02.001](https://doi.org/10.1016/j.bbabbio.2012.02.001)
103. Lanci CJ, MacDermaid CM, Kang SG, Acharya R, North B, Yang X, Qiu XJ, DeGrado WF, Saven JG (2012) Computational design of a protein crystal. *Proc Natl Acad Sci U S A* 109(19):7304–7309. doi:[10.1073/pnas.1112595109](https://doi.org/10.1073/pnas.1112595109)
104. Reig AJ, Pires MM, Snyder RA, Wu Y, Jo H, Kulp DW, Butch SE, Calhoun JR, Szyperski T, Solomon EI, DeGrado WF (2012) Alteration of the oxygen-dependent reactivity of de novo De Ferri proteins. *Nat Chem* 4(11):900–906. doi:[10.1038/nchem.1454](https://doi.org/10.1038/nchem.1454)
105. Dagliyan O, Shirvanyants D, Karginov AV, Ding F, Fee L, Chandrasekaran SN, Freisinger CM, Smolen GA, Huttenlocher A, Hahn KM, Dokholyan NV (2013) Rational design of a ligand-controlled protein conformational switch. *Proc Natl Acad Sci U S A* 110(17):6800–6804. doi:[10.1073/pnas.1218319110](https://doi.org/10.1073/pnas.1218319110)
106. Fry HC, Lehmann A, Sinks LE, Asselberghs I, Tronin A, Krishnan V, Blasie JK, Clays K, DeGrado WF, Saven JG, Therien MJ (2013) Computational de novo design and characterization of a protein that selectively binds a highly hyperpolarizable abiological chromophore. *J Am Chem Soc* 135(37):13914–13926. doi:[10.1021/ja4067404](https://doi.org/10.1021/ja4067404)
107. Perez-Aguilar JM, Xi J, Matsunaga F, Cui X, Selling B, Saven JG, Liu R (2013) A computationally designed water-soluble variant of a G-protein-coupled receptor: the human mu opioid receptor. *PLoS One* 8(6), e66009. doi:[10.1371/journal.pone.0066009](https://doi.org/10.1371/journal.pone.0066009)
108. Lerner MB, Matsunaga F, Han GH, Hong SJ, Xi J, Crook A, Perez-Aguilar JM, Park YW, Saven JG, Liu R, Johnson AT (2014) Scalable production of highly sensitive nanosensors based on graphene functionalized with a designed G protein-coupled receptor. *Nano Lett* 14(5):2709–2714. doi:[10.1021/nl5006349](https://doi.org/10.1021/nl5006349)
109. Procko E, Hedman R, Hamilton K, Seetharaman J, Fleishman SJ, Su M, Aramini J, Kornhaber G, Hunt JF, Tong L, Montelione GT, Baker D (2013) Computational design of a protein-based enzyme inhibitor. *J Mol Biol* 425(18):3563–3575. doi:[10.1016/j.jmb.2013.06.035](https://doi.org/10.1016/j.jmb.2013.06.035)
110. Tinberg CE, Khare SD, Dou J, Doyle L, Nelson JW, Schena A, Jankowski W, Kalodimos CG, Johnsson K, Stoddard BL, Baker D (2013) Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* 501(7466):212–216. doi:[10.1038/nature12443](https://doi.org/10.1038/nature12443)
111. Strauch EM, Fleishman SJ, Baker D (2014) Computational design of a pH-sensitive IgG binding protein. *Proc Natl Acad Sci U S A* 111(2):675–680. doi:[10.1073/pnas.1313605111](https://doi.org/10.1073/pnas.1313605111)
112. Xiong P, Wang M, Zhou X, Zhang T, Zhang J, Chen Q, Liu H (2014) Protein design with a comprehensive statistical energy function and boosted by experimental selection for foldability. *Nat Commun* 5:5330. doi:[10.1038/ncomms6330](https://doi.org/10.1038/ncomms6330)
113. Huang PS, Oberdorfer G, Xu C, Pei XY, Nannenga BL, Rogers JM, DiMaio F, Gonen T, Luisi B, Baker D (2014) High thermodynamic stability of parametrically designed helical bundles. *Science* 346(6208):481–485. doi:[10.1126/science.1257481](https://doi.org/10.1126/science.1257481)
114. Thomson AR, Wood CW, Burton AJ, Bartlett GJ, Sessions RB, Brady RL, Woolfson DN (2014) Computational design of water-soluble alpha-helical barrels. *Science* 346(6208):485–488. doi:[10.1126/science.1257452](https://doi.org/10.1126/science.1257452)

115. Negron C, Keating AE (2014) A set of computationally designed orthogonal antiparallel homodimers that expands the synthetic coiled-coil toolkit. *J Am Chem Soc* 136 (47):16544–16556. doi:[10.1021/ja507847t](https://doi.org/10.1021/ja507847t)
116. Correia BE, Bates JT, Loomis RJ, Baneyx G, Carrico C, Jardine JG, Rupert P, Correnti C, Kalyuzhnyi O, Vittal V, Connell MJ, Stevens E, Schroeter A, Chen M, Macpherson S, Serra AM, Adachi Y, Holmes MA, Li Y, Klevit RE, Graham BS, Wyatt RT, Baker D, Strong RK, Crowe JE Jr, Johnson PR, Schief WR (2014) Proof of principle for epitope-focused vaccine design. *Nature* 507(7491):201–206. doi:[10.1038/nature12966](https://doi.org/10.1038/nature12966)
117. Joh NH, Wang T, Bhate MP, Acharya R, Wu Y, Grabe M, Hong M, Grigoryan G, DeGrado WF (2014) De novo design of a transmembrane Zn(2)(+)-transporting four-helix bundle. *Science* 346(6216):1520–1524. doi:[10.1126/science.1261172](https://doi.org/10.1126/science.1261172)
118. King C, Garza EN, Mazor R, Linehan JL, Pastan I, Pepper M, Baker D (2014) Removing T-cell epitopes with computational protein design. *Proc Natl Acad Sci U S A* 111 (23):8577–8582. doi:[10.1073/pnas.1321126111](https://doi.org/10.1073/pnas.1321126111)
119. Voet AR, Noguchi H, Addy C, Simoncini D, Terada D, Unzai S, Park SY, Zhang KY, Tame JR (2014) Computational design of a self-assembling symmetrical beta-propeller protein. *Proc Natl Acad Sci U S A* 111 (42):15102–15107. doi:[10.1073/pnas.1412768111](https://doi.org/10.1073/pnas.1412768111)
120. Ramisch S, Weininger U, Martinsson J, Akke M, Andre I (2014) Computational design of a leucine-rich repeat protein with a predefined geometry. *Proc Natl Acad Sci U S A* 111 (50):17875–17880. doi:[10.1073/pnas.1413638111](https://doi.org/10.1073/pnas.1413638111)
121. Samish I, MacDermaid CM, Perez-Aguilar JM, Saven JG (2011) Theoretical and computational protein design. *Annu Rev Phys Chem* 62:129–149. doi:[10.1146/annurev-physchem-032210-103509](https://doi.org/10.1146/annurev-physchem-032210-103509)
122. Pantazes RJ, Grisewood MJ, Maranas CD (2011) Recent advances in computational protein design. *Curr Opin Struct Biol* 21 (4):467–472. doi:[10.1016/j.sbi.2011.04.005](https://doi.org/10.1016/j.sbi.2011.04.005)
123. O'Neil KT, DeGrado WF (1985) A predicted structure of calmodulin suggests an electrostatic basis for its function. *Proc Natl Acad Sci U S A* 82(15):4954–4958
124. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S, Weiner P (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. *J Am Chem Soc* 106(3):765–784
125. Hellinga HW, Richards FM (1991) Construction of new ligand binding sites in proteins of known structure. I. Computer-aided modeling of sites with pre-defined geometry. *J Mol Biol* 222(3):763–785
126. Ponder JW, Richards FM (1987) Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J Mol Biol* 193 (4):775–791
127. Bryson JW, Betz SF, Lu HS, Suich DJ, Zhou HX, O'Neil KT, DeGrado WF (1995) Protein design: a hierarchic approach. *Science* 270 (5238):935–941
128. Raha K, Wollacott AM, Italia MJ, Desjarlais JR (2000) Prediction of amino acid sequence from structure. *Protein Sci* 9(6):1106–1119. doi:[10.1110/ps.9.6.1106](https://doi.org/10.1110/ps.9.6.1106)
129. Desmet J, De Maeyer M, Hazes B, Lasters I (1992) The dead-end elimination theorem and its use in protein side-chain positioning. *Nature* 356(6369):539–542
130. Samish I (2009) Search and sampling in structural bioinformatics. In: Bourne P, Gu J (eds) *Structural bioinformatics*. Wiley, Hoboken, NJ, pp 207–236
131. Dunbrack RL Jr, Karplus M (1993) Backbone-dependent rotamer library for proteins. Application to side-chain prediction. *J Mol Biol* 230(2):543–574. doi:[10.1006/jmbi.1993.1170](https://doi.org/10.1006/jmbi.1993.1170)
132. Tuffery P, Etchebest C, Hazout S, Lavery R (1991) A new approach to the rapid determination of protein side chain conformations. *J Biomol Struct Dyn* 8(6):1267–1289. doi:[10.1080/07391102.1991.10507882](https://doi.org/10.1080/07391102.1991.10507882)
133. Bowers PM, Strauss CE, Baker D (2000) De novo protein structure determination using sparse NMR data. *J Biomol NMR* 18 (4):311–318
134. Kuhlman B, Baker D (2000) Native protein sequences are close to optimal for their structures. *Proc Natl Acad Sci U S A* 97 (19):10383–10388
135. Fisinger S, Serrano L, Lacroix E (2001) Computational estimation of specific side chain interaction energies in alpha helices. *Protein Sci* 10(4):809–818. doi:[10.1110/ps.34901](https://doi.org/10.1110/ps.34901)
136. Kortemme T, Ramirez-Alvarado M, Serrano L (1998) Design of a 20-amino acid, three-stranded beta-sheet protein. *Science* 281 (5374):253–256
137. Debler EW, Ito S, Seebeck FP, Heine A, Hilvert D, Wilson IA (2005) Structural origins of

- efficient proton abstraction from carbon by a catalytic antibody. *Proc Natl Acad Sci U S A* 102(14):4984–4989. doi:[10.1073/pnas.0409207102](https://doi.org/10.1073/pnas.0409207102)
138. Khersonsky O, Rothlisberger D, Wollacott AM, Murphy P, Dym O, Albeck S, Kiss G, Houk KN, Baker D, Tawfik DS (2011) Optimization of the in-silico-designed kemp eliminase KE70 by computational design and directed evolution. *J Mol Biol* 407(3):391–412. doi:[10.1016/j.jmb.2011.01.041](https://doi.org/10.1016/j.jmb.2011.01.041)
  139. Khersonsky O, Rothlisberger D, Dym O, Albeck S, Jackson CJ, Baker D, Tawfik DS (2010) Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the KE07 series. *J Mol Biol* 396(4):1025–1042. doi:[10.1016/j.jmb.2009.12.031](https://doi.org/10.1016/j.jmb.2009.12.031)
  140. Khersonsky O, Kiss G, Rothlisberger D, Dym O, Albeck S, Houk KN, Baker D, Tawfik DS (2012) 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* 109(26):10358–10363. doi:[10.1073/pnas.1121063109](https://doi.org/10.1073/pnas.1121063109)
  141. Dahiyat BI, Mayo SL (1996) Protein design automation. *Protein Sci* 5(5):895–903. doi:[10.1002/pro.5560050511](https://doi.org/10.1002/pro.5560050511)
  142. Kortemme T, Baker D (2002) A simple physical model for binding energy hot spots in protein-protein complexes. *Proc Natl Acad Sci U S A* 99(22):14116–14121. doi:[10.1073/pnas.202485799](https://doi.org/10.1073/pnas.202485799)
  143. Kono H, Saven JG (2001) Statistical theory for protein combinatorial libraries. Packing interactions, backbone flexibility, and the sequence variability of a main-chain structure. *J Mol Biol* 306(3):607–628. doi:[10.1006/jmbi.2000.4422](https://doi.org/10.1006/jmbi.2000.4422)
  144. Lombardi A, Summa CM, Geremia S, Randaccio L, Pavone V, DeGrado WF (2000) Retrostructural analysis of metalloproteins: application to the design of a minimal model for diiron proteins. *Proc Natl Acad Sci U S A* 97(12):6298–6305
  145. Dunbrack RL Jr, Cohen FE (1997) Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci* 6(8):1661–1681. doi:[10.1002/pro.5560060807](https://doi.org/10.1002/pro.5560060807)
  146. Fu X, Kono H, Saven JG (2003) Probabilistic approach to the design of symmetric protein quaternary structures. *Protein Eng* 16(12):971–977. doi:[10.1093/protein/gzg132](https://doi.org/10.1093/protein/gzg132)
  147. Zhang SQ, Kulp DW, Schramm CA, Mravic M, Samish I, DeGrado WF (2015) The membrane- and soluble-protein helix-helix interactome: similar geometry via different interactions. *Structure* 23(3):527–541. doi:[10.1016/j.str.2015.01.009](https://doi.org/10.1016/j.str.2015.01.009)
  148. Smith MD, Zanghellini A, Grabs-Rothlisberger D (2014) Computational design of novel enzymes without cofactors. *Methods Mol Biol* 1216:197–210. doi:[10.1007/978-1-4939-1486-9\\_10](https://doi.org/10.1007/978-1-4939-1486-9_10)
  149. Zanghellini A, Jiang L, Wollacott AM, Cheng G, Meiler J, Althoff EA, Rothlisberger D, Baker D (2006) New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci* 15(12):2785–2794. doi:[10.1110/ps.062353106](https://doi.org/10.1110/ps.062353106)
  150. Richter F, Leaver-Fay A, Khare SD, Bjelic S, Baker D (2011) De novo enzyme design using Rosetta3. *PLoS One* 6(5), e19230. doi:[10.1371/journal.pone.0019230](https://doi.org/10.1371/journal.pone.0019230)
  151. Frushicheva MP, Cao J, Chu ZT, Warshel A (2010) Exploring challenges in rational enzyme design by simulating the catalysis in artificial kemp eliminase. *Proc Natl Acad Sci U S A* 107(39):16869–16874. doi:[10.1073/pnas.1010381107](https://doi.org/10.1073/pnas.1010381107)
  152. Georgiev I, Lilien RH, Donald BR (2008) The minimized dead-end elimination criterion and its application to protein redesign in a hybrid scoring and search algorithm for computing partition functions over molecular ensembles. *J Comput Chem* 29(10):1527–1542. doi:[10.1002/jcc.20909](https://doi.org/10.1002/jcc.20909)
  153. Schneidman-Duhovny D, Inbar Y, Polak V, Shatsky M, Halperin I, Benyamini H, Barzilai A, Dror O, Haspel N, Nussinov R, Wolfson HJ (2003) Taking geometry to its edge: fast unbound rigid (and hinge-bent) docking. *Proteins* 52(1):107–112. doi:[10.1002/prot.10397](https://doi.org/10.1002/prot.10397)
  154. Fiser A, Do RK, Sali A (2000) Modeling of loops in protein structures. *Protein Sci* 9(9):1753–1773. doi:[10.1110/ps.9.9.1753](https://doi.org/10.1110/ps.9.9.1753)
  155. Ofek G, Guenaga FJ, Schief WR, Skinner J, Baker D, Wyatt R, Kwong PD (2010) Elicitation of structure-specific antibodies by epitope scaffolds. *Proc Natl Acad Sci U S A* 107(42):17880–17887. doi:[10.1073/pnas.1004728107](https://doi.org/10.1073/pnas.1004728107)
  156. Canutescu AA, Dunbrack RL Jr (2003) Cyclic coordinate descent: a robotics algorithm for protein loop closure. *Protein Sci* 12(5):963–972. doi:[10.1110/ps.0242703](https://doi.org/10.1110/ps.0242703)
  157. Smith CA, Kortemme T (2011) Predicting the tolerated sequences for proteins and protein interfaces using RosettaBackrub flexible



- backbone design. *PLoS One* 6(7), e20451. doi:[10.1371/journal.pone.0020451](https://doi.org/10.1371/journal.pone.0020451)
158. Mandell DJ, Coutsiias EA, Kortemme T (2009) Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nat Methods* 6(8):551–552. doi:[10.1038/nmeth0809-551](https://doi.org/10.1038/nmeth0809-551)
  159. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 5(4):725–738. doi:[10.1038/nprot.2010.5](https://doi.org/10.1038/nprot.2010.5)
  160. Summa CM, Levitt M, Degradó WF (2005) An atomic environment potential for use in protein structure prediction. *J Mol Biol* 352(4):986–1001. doi:[10.1016/j.jmb.2005.07.054](https://doi.org/10.1016/j.jmb.2005.07.054)
  161. Antonkine ML, Maes EM, Czernuszewicz RS, Breitenstein C, Bill E, Falzone CJ, Balasubramanian R, Lubner C, Bryant DA, Golbeck JH (2007) Chemical rescue of a site-modified ligand to a [4Fe-4S] cluster in PsaC, a bacterial-like dicluster ferredoxin bound to Photosystem I. *Biochim Biophys Acta* 1767(6):712–724. doi:[10.1016/j.bbabi.2007.02.003](https://doi.org/10.1016/j.bbabi.2007.02.003)
  162. Kulp DW, Subramaniam S, Donald JE, Hannigan BT, Mueller BK, Grigoryan G, Senes A (2012) Structural informatics, modeling, and design with an open-source Molecular Software Library (MSL). *J Comput Chem* 33(20):1645–1661. doi:[10.1002/jcc.22968](https://doi.org/10.1002/jcc.22968)
  163. Yang Y, Zhou Y (2008) Ab initio folding of terminal segments with secondary structures reveals the fine difference between two closely related all-atom statistical energy functions. *Protein Sci* 17(7):1212–1219. doi:[10.1110/ps.033480.107](https://doi.org/10.1110/ps.033480.107)
  164. Schramm CA, Hannigan BT, Donald JE, Keasar C, Saven JG, Degradó WF, Samish I (2012) Knowledge-based potential for positioning membrane-associated structures and assessing residue-specific energetic contributions. *Structure* 20(5):924–935. doi:[10.1016/j.str.2012.03.016](https://doi.org/10.1016/j.str.2012.03.016)
  165. Grigoryan G (2013) Absolute free energies of biomolecules from unperturbed ensembles. *J Comput Chem* 34(31):2726–2741. doi:[10.1002/jcc.23448](https://doi.org/10.1002/jcc.23448)
  166. Dwyer MA, Looger LL, Hellinga HW (2008) Retraction. *Science* 319(5863):569. doi:[10.1126/science.319.5863.569b](https://doi.org/10.1126/science.319.5863.569b)
  167. Dwyer MA, Looger LL, Hellinga HW (2004) Computational design of a biologically active enzyme. *Science* 304(5679):1967–1971. doi:[10.1126/science.1098432](https://doi.org/10.1126/science.1098432)
  168. Allert M, Dwyer MA, Hellinga HW (2007) Local encoding of computationally designed enzyme activity. *J Mol Biol* 366(3):945–953. doi:[10.1016/j.jmb.2006.12.002](https://doi.org/10.1016/j.jmb.2006.12.002)
  169. (2008) Negative results. *Nature* 453(7193):258. doi:[10.1038/453258b](https://doi.org/10.1038/453258b)
  170. Hayden EC (2009) Key protein-design papers challenged. *Nature* 461(7266):859. doi:[10.1038/461859a](https://doi.org/10.1038/461859a)
  171. Looger LL, Dwyer MA, Smith JJ, Hellinga HW (2003) Computational design of receptor and sensor proteins with novel functions. *Nature* 423(6936):185–190. doi:[10.1038/nature01556](https://doi.org/10.1038/nature01556)
  172. Allert M, Rizk SS, Looger LL, Hellinga HW (2004) Computational design of receptors for an organophosphate surrogate of the nerve agent soman. *Proc Natl Acad Sci U S A* 101(21):7907–7912. doi:[10.1073/pnas.0401309101](https://doi.org/10.1073/pnas.0401309101)
  173. Hellinga HW (2008) In the wake of two retractions, a request for investigation. *Nature* 454(7203):397. doi:[10.1038/454397b](https://doi.org/10.1038/454397b)
  174. Altamirano MM, Blackburn JM, Aguayo C, Fersht AR (2002) Retraction. Directed evolution of new catalytic activity using the alpha/beta-barrel scaffold. *Nature* 417(6887):468. doi:[10.1038/417468a](https://doi.org/10.1038/417468a)
  175. Altamirano MM, Blackburn JM, Aguayo C, Fersht AR (2000) Directed evolution of new catalytic activity using the alpha/beta-barrel scaffold. *Nature* 403(6770):617–622. doi:[10.1038/35001001](https://doi.org/10.1038/35001001)
  176. Feldmeier K, Hocker B (2013) Computational protein design of ligand binding and catalysis. *Curr Opin Chem Biol* 17(6):929–933
  177. Wijma HJ, Janssen DB (2013) Computational design gains momentum in enzyme catalysis engineering. *FEBS J* 280(13):2948–2960. doi:[10.1111/febs.12324](https://doi.org/10.1111/febs.12324)
  178. Khare SD, Fleishman SJ (2013) Emerging themes in the computational design of novel enzymes and protein-protein interfaces. *FEBS Lett* 587(8):1147–1154. doi:[10.1016/j.febslet.2012.12.009](https://doi.org/10.1016/j.febslet.2012.12.009)
  179. Davey JA, Chica RA (2012) Multistate approaches in computational protein design. *Protein Sci* 21(9):1241–1252. doi:[10.1002/pro.2128](https://doi.org/10.1002/pro.2128)
  180. Tiwari MK, Singh R, Singh RK, Kim IW, Lee JK (2012) Computational approaches for rational design of proteins with novel functionalities. *Comput Struct Biotechnol J* 2, e201209002. doi:[10.5936/csbj.201209002](https://doi.org/10.5936/csbj.201209002)
  181. Senes A (2011) Computational design of membrane proteins. *Curr Opin Struct Biol* 21(4):460–466. doi:[10.1016/j.sbi.2011.06.004](https://doi.org/10.1016/j.sbi.2011.06.004)

182. Verschueren E, Vanhee P, van der Sloot AM, Serrano L, Rousseau F, Schymkowitz J (2011) Protein design with fragment databases. *Curr Opin Struct Biol* 21(4):452–459. doi:[10.1016/j.sbi.2011.05.002](https://doi.org/10.1016/j.sbi.2011.05.002)
183. Samish I, Bourne PE, Najmanovich RJ (2015) Achievements and challenges in structural bioinformatics and computational biophysics. *Bioinformatics* 31(1):146–150. doi:[10.1093/bioinformatics/btu769](https://doi.org/10.1093/bioinformatics/btu769)
184. Rosenberg M, Goldblum A (2006) Computational protein design: a novel path to future protein drugs. *Curr Pharm Des* 12(31):3973–3997
185. Desjarlais JR, Lazar GA (2003) Negative design for improved therapeutic proteins. *Trends Biotechnol* 21(10):425–427. doi:[10.1016/S0167-7799\(03\)00205-1](https://doi.org/10.1016/S0167-7799(03)00205-1)
186. Moult J, Pedersen JT, Judson R, Fidelis K (1995) A large-scale experiment to assess protein structure prediction methods. *Proteins* 23(3):ii–v. doi:[10.1002/prot.340230303](https://doi.org/10.1002/prot.340230303)



<http://www.springer.com/978-1-4939-6635-6>

Computational Protein Design

Samish, I. (Ed.)

2017, XII, 450 p. 144 illus., 111 illus. in color.,

Hardcover

ISBN: 978-1-4939-6635-6

A product of Humana Press