Chapter 2

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 ββα 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.

<table>
<thead>
<tr>
<th>Novelty</th>
<th>Target</th>
<th>Computational methods</th>
<th>Main characterization methods</th>
<th>PDB (Å resolution)</th>
<th>Year, first and last authors, journal, and references</th>
<th>Protein structure (first PDB only)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. De novo design of a four-helix bundle (<em>Felix</em>)</td>
<td>Antiparallel four-helix bundle (79 residues)</td>
<td>Design rules followed by MD and structural modeling</td>
<td>CD, SEC, Fluorescence</td>
<td>Ifx (theoretical model)</td>
<td>1990 Hecht, Ogden [10]</td>
<td></td>
</tr>
<tr>
<td>4. Enzyme design for a non-natural substrate</td>
<td>α-lytic protease</td>
<td>Single mutation free energy perturbation (FEP) with molecular mechanics (AMBER) and rotamer library (PROPAK)</td>
<td>Kinetic measurements of design and specificity</td>
<td>–</td>
<td>1991 Wilson, Agard, <em>JMB</em> [12]</td>
<td></td>
</tr>
<tr>
<td>5. Hydrophobic core design</td>
<td>Bacteriophage T4 lysozyme</td>
<td>Modified rotamer library and scoring function including standard force-field and conformational entropy components</td>
<td>X-ray, CD</td>
<td>1l77 (2.05 Å), 2l78 (2.0 Å), 1l79 (1.9 Å), 1l80 (1.8 Å), 1l81 (2.0 Å), 1l82 (2.1 Å)</td>
<td>1992, Hurley, Matthews, <em>JMB</em> [13]</td>
<td></td>
</tr>
<tr>
<td>6. β-sandwich de novo design (<em>Betadoublet</em>)</td>
<td>β-sandwich protein</td>
<td>SYBYL (Tripos), CHAOS, GCG</td>
<td>CD, NMR H/D exchange</td>
<td>1btd (theoretical)</td>
<td>1994 Quinn, Richardson, <em>PNAS</em> [14]</td>
<td></td>
</tr>
</tbody>
</table>
7. Hydrophobic core design

<table>
<thead>
<tr>
<th>Description</th>
<th>Design</th>
<th>Methodology</th>
<th>Techniques</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phage 434 cro protein (five to eight core residues)</td>
<td>Rotamer library + genetic algorithm within Repacking of Core (ROC)</td>
<td>CD, NMR</td>
<td>–</td>
<td>1995 Desjarlais Hnadel, Prot Sci [15]</td>
</tr>
</tbody>
</table>

8. Hydrophobic core design

<table>
<thead>
<tr>
<th>Description</th>
<th>Design</th>
<th>Methodology</th>
<th>Techniques</th>
<th>References</th>
</tr>
</thead>
</table>

9. Full sequence design

<table>
<thead>
<tr>
<th>Description</th>
<th>Design</th>
<th>Methodology</th>
<th>Techniques</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>ββα motif typified by the zinc finger DNA binding</td>
<td>20 of 28 motif positions were designed avoiding metal and Cys. Template = 1zaa</td>
<td>CD, NMR structure backbone root mean square deviation (RMSD) = 1.04 Å</td>
<td>1FSD, 1FSV (NMR)</td>
<td>1997 Dahiyat, Mayo, Science [18, 19]</td>
</tr>
</tbody>
</table>

10. Hyper-thermophile protein

<table>
<thead>
<tr>
<th>Description</th>
<th>Design</th>
<th>Methodology</th>
<th>Techniques</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein G—β1 domain (Gβ1)</td>
<td>7 mutations selected via ORBIT</td>
<td>CD, AUC, NMR, kinetics (Biacore)</td>
<td>1gb4 (NMR)</td>
<td>1998, Malakauskas, May, Nat Struct Biol [20]</td>
</tr>
</tbody>
</table>

11. Coiled-coil oligomers

<table>
<thead>
<tr>
<th>Description</th>
<th>Design</th>
<th>Methodology</th>
<th>Techniques</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right-handed dimer, trimer, and tetramer coiled-coils</td>
<td>Hydrophobic-polar residue patterning and side-chain packing. Comparison to mutations in other locations (representing a less folded state)</td>
<td>X-ray, CD</td>
<td>1rh4 (1.9 Å)</td>
<td>1998, Harbury, Kim, Science [21]</td>
</tr>
</tbody>
</table>

12. Three-helix bundle

<table>
<thead>
<tr>
<th>Description</th>
<th>Design</th>
<th>Methodology</th>
<th>Techniques</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Novelty</td>
<td>Target</td>
<td>Computational methods</td>
<td>Main characterization methods</td>
<td>PDB (Å resolution)</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>-----------------------</td>
<td>-------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>14. Dimerization of a monomer (VPA)</td>
<td>Dimerizing protein L-based on a strand-swapped dimer</td>
<td>RosettaDesign 3 positions were design in a eight-residue redesign</td>
<td>CD, X-ray</td>
<td>1jml (1.9 Å)</td>
</tr>
<tr>
<td>15. β-sheet peptide automatic redesign</td>
<td>Automatic redesign of one to three residues in the betanova β-sheet peptide design producing an array of peptides</td>
<td>PERLA</td>
<td>CD, NMR</td>
<td>NMR available (not deposited to PDB)</td>
</tr>
<tr>
<td>17. Heterdimeric coiled coils</td>
<td>Six heterodimeric coiled coils</td>
<td>Knowledge-based coiled coil rules, designated rotamer library (sampled via DEE/A*) and minimization with solvation correction</td>
<td>X-ray, CD, AUC, 1kd8 (1.9 Å), 1kd9 (2.1 Å), 1kdd (2.14 Å)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>18.</td>
<td>Diiron binding (DFtet)</td>
<td>A$_2$B$_2$ four-helix bundle binding diiron</td>
<td>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</td>
<td>CD, AUC, Co$_{2+}$ titration, ferroxidase activity</td>
</tr>
<tr>
<td>19.</td>
<td>Hydrophobic core 13 Spectrin SH3 core redesigns</td>
<td>PERLA, rotamers with 5$^\circ$ sub-rotameric states for $\chi_1$ and $\chi_2$</td>
<td>X-ray, fluorescence</td>
<td>1e6h (2.0 Å), 1e6g (2.3 Å), 1h8k (1.8 Å)</td>
</tr>
<tr>
<td>20.</td>
<td>Binding specificity Calmodulin binding to smMLCK</td>
<td>Eight residues were mutated via ORBIT modified for protein–protein interactions</td>
<td>CD, binding</td>
<td>–</td>
</tr>
<tr>
<td>21.</td>
<td>Antibiotic resistance redesign</td>
<td>Changing β-lactamase specificity from ampicillin- to cefotaxime-resistance</td>
<td>19 residues at the interface region were redesigned with PDA, MC/SA and experimental screening of a ~200,000 sequence library</td>
<td>Antibiotic-resistance minimum inhibitory concentration (MIC) assay</td>
</tr>
<tr>
<td>22.</td>
<td>Core stabilization Human growth hormone (hGH) core stabilization</td>
<td>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</td>
<td>CD, cell proliferation assay</td>
<td>–</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Novelty</th>
<th>Target</th>
<th>Computational methods</th>
<th>Main characterization methods</th>
<th>PDB (Å resolution)</th>
<th>Year, first and last authors, journal, and references</th>
<th>Protein structure (first PDB only)</th>
</tr>
</thead>
<tbody>
<tr>
<td>24.</td>
<td>Artificial specific endonuclease (E-Drel)</td>
<td>Engineered I-Dmol/I-Crel endonuclease where these domains are fused and which targets a long chimeric DNA target</td>
<td>Three-residue inter-domain linker and 14 interface residues identified by computational Ala scanning and redesigned with RosettaDesign</td>
<td>X-ray, lacZ-based solubility assay, biochemical DNA cleavage assay</td>
<td>1 mow (2.4 Å)</td>
<td>2002, Chevalier, Stoddard, <em>Mol Cell</em> [36]</td>
</tr>
<tr>
<td>25.</td>
<td>MHC-I binding peptides</td>
<td>Automatic design of nine-residue peptides inhibiting MHC-I</td>
<td>DESIGNER, correlation with a dataset of known MHC-I binding peptides</td>
<td>Fluorescence (peptide-protein binding), T-cell response inhibition (ELISPOT)</td>
<td>–</td>
<td>2003, Ogata, Wodak, <em>JBC</em> [37]</td>
</tr>
<tr>
<td>26.</td>
<td>Water solubilization of a membrane protein</td>
<td>Water-soluble pentameric phospholamban</td>
<td>Ten water-facing residues were mutated optimizing pairwise interactions with MC/SA sampling</td>
<td>X-ray, CD, AUC, SEC</td>
<td>Iyod (1.8 Å)</td>
<td>2003, Slovic, DeGrado <em>Prot Sci</em> [38] (design) 2005 <em>JMB</em> [39] (structure)</td>
</tr>
<tr>
<td>28.</td>
<td>De novo metal-binding design</td>
<td>Due ferro four-helix-bundle</td>
<td>88 out of 114 residues were designed by SCADS</td>
<td>AUC, CD, metal-binding, NMR</td>
<td>2h28 (NMR + QM/MM refinement)</td>
<td>2003 Calhoun, Saven, <em>JMB</em> [41] 2008 <em>Structure</em> [42]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>29.</td>
<td>Novel fold with atomic-level accuracy (Top7)</td>
<td>93-residue novel αβ fold with five β-strands and two α-helices</td>
<td>Rosetta package: 3D backbone via fragments followed by side-chain MC</td>
<td>CD, X-ray, GuHCl denaturation, NOESY &amp; HSQC NMR</td>
<td>1qys (2.5 Å)</td>
<td>2003 Kuhlman, Baker, <em>Science</em> [43]</td>
</tr>
<tr>
<td>31.</td>
<td>Protein redesign (increased folding kinetics and thermostability)</td>
<td>Human procarboxypeptidase A2 and 8 other folds</td>
<td>RosettaDesign</td>
<td>X-ray, NMR, CD, AUC, GuHCL</td>
<td>1yjq (2.1 Å) 2gf (NMR)</td>
<td>2003, 2007, Dantas, Baker <em>JMB</em> [45, 46]</td>
</tr>
<tr>
<td>33.</td>
<td>PPI specificity</td>
<td>PPI specificity in DNase (colicin E7)-inhibitor protein (immunity protein Im7) pairs</td>
<td>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</td>
<td>X-ray, SPR, fluorescence, toxicity plate assay, DNase activity</td>
<td>1ujz (2.1 Å), 2erh (2.0 Å)</td>
<td>2004, Kortemme, Baker <em>NSMB</em> and 2006 Joachimiak, Baker <em>JMB</em> [49, 50]</td>
</tr>
<tr>
<td>34.</td>
<td>Water-soluble potassium channel</td>
<td>Water-soluble KcsA potassium channel</td>
<td>SCADS, 35 solvent-exposed residues were designed with an environmental energy fit to solvent-exposure</td>
<td>CD, AUC, SEC, toxin-binding</td>
<td>–</td>
<td>2004, Slovic, DeGrado, <em>PNAS</em> [51]</td>
</tr>
<tr>
<td>Novelty</td>
<td>Target</td>
<td>Computational methods</td>
<td>Main characterization methods</td>
<td>PDB (Å resolution)</td>
<td>Year, first and last authors, journal, and references</td>
<td>Protein structure (first PDB only)</td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
<td>-----------------------</td>
<td>-------------------------------</td>
<td>-------------------</td>
<td>------------------------------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>35.</td>
<td>Thermostabilization of an enzyme</td>
<td>Thermostabilization of homodimeric hydrolase enzyme yeast cytosine deaminase (yCD)</td>
<td>Rosetta redesign of residues not in the dimer interface or catalytic site</td>
<td>X-ray, CD, kinetic measurements</td>
<td>1ysd (1.9 Å), 1ysb (1.7 Å)</td>
<td>2005, Korkegian, Stoddard, <em>Science</em> [52]</td>
</tr>
<tr>
<td>36.</td>
<td>Negative design for reengineering a homodimer to a heterodimer</td>
<td>SspB homodimer</td>
<td>ORBIT, modified by capping unfavorable vdW energetics and adding an MC negative-design module</td>
<td>X-ray, GuHCl-denaturation, chromatography dimerization assay</td>
<td>1zsz (2.0 Å)</td>
<td>2005, Bolon, Sauer, <em>PNAS</em> [53]</td>
</tr>
<tr>
<td>37.</td>
<td>Redox-active rubredoxin mimic</td>
<td>Minimal (40-residue) de novo designed rubredoxin mimic</td>
<td>SCADS</td>
<td>UV-vis, CD</td>
<td>–</td>
<td>2005, Nanda, DeGrado, <em>JACS</em> [54]</td>
</tr>
<tr>
<td>38.</td>
<td>Nonbiological cofactor binding</td>
<td>Four-helix bundle that selectively binds two DPP-Fe</td>
<td>Backbone design following different constrains followed by three rounds of SCADS sequence design</td>
<td>CD, SEC, AUC, UV-Vis, EPR</td>
<td>–</td>
<td>2005, Cochran, DeGrado, <em>JACS</em> [55]</td>
</tr>
<tr>
<td>40.</td>
<td>Binding selectivity design</td>
<td>TRAIL-variant for selective DR5 binding</td>
<td>WHATIF, FOLD-X</td>
<td>SPR, FACS</td>
<td>–</td>
<td>2006, van der Sloot, Quax <em>PNAS</em> [57]</td>
</tr>
<tr>
<td>41.</td>
<td>Endonuclease DNA binding and specificity redesign</td>
<td>Redesign of intron-encoded homing endonuclease I-MsoI</td>
<td>RosettaDesign protocol of MC optimization of discrete side-chain conformations and identities with continuous minimization of protein and DNA dihedral angles</td>
<td>X-ray, competitive cleavage assay, binding assay</td>
<td>2f1d (2.0 Å)</td>
<td>2006, Ashworth, Baker, Nature [58]</td>
</tr>
<tr>
<td>------</td>
<td>--------------------------------</td>
<td>--------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>42.</td>
<td>Antibody Fc variants</td>
<td>Antibody Fc variants with enhances Fcγ-receptor-mediated effector function. Four sites were mutated</td>
<td>Protein design automation (PDA), Sequence prediction algorithm (SPA)</td>
<td>SPR, ADCC assay, B-cell depletion (in macaques), AlphaScreen</td>
<td>–</td>
<td>2006, Lazar, Dahiyat, PNAS [59]</td>
</tr>
<tr>
<td>43.</td>
<td>Protein–protein interface</td>
<td>Protein G—β1 domain (Gβ1)</td>
<td>24 interface residues, side-chain-pruned docking of 18 complexes to derive docking parameterization, ORBIT, RESCLASS</td>
<td>HSQC NMR, AUC, dissociation kinetics</td>
<td></td>
<td>2007, Huang, Mayo, Prot Sci [60]</td>
</tr>
<tr>
<td>44.</td>
<td>Anti-transmembrane helix peptide (CHAMP)</td>
<td>Computed Helical Anti Membrane Protein Peptide binding αIIb and αv, integrin subunits</td>
<td>Protcad, scwrl3, GROMOS minimization, HELANAL</td>
<td>CD, ATR-IR, FRET, DNTOCAT, Hemolysis, platelet aggregation, and adhesion</td>
<td>–</td>
<td>2007, Yin, DeGrado, Science [61]</td>
</tr>
<tr>
<td>45.</td>
<td>De novo loop design</td>
<td>Tenascin-C ten-residue loop</td>
<td>Ten-residue loop design via Rosetta, loop fragments, and filtering</td>
<td>X-ray, 1H NMR, CD</td>
<td>2rb8 (1.45 Å), 2rbl (2.1 Å)</td>
<td>2007, Hu, Kuhlman, PNAS [62]</td>
</tr>
<tr>
<td>46.</td>
<td>De novo design of a erythropoietin (EPO) receptor binder</td>
<td>Design by grafting three binding residues from the EPO—EPO receptor complex and grafting them on a new scaffold</td>
<td>PDB search for scaffold supporting the grafted residues, filtering by protein–protein interface native-like measurables</td>
<td>CD, SPR, Luciferase Reporter Assay</td>
<td></td>
<td>2007, Liu, Lai, PNAS [63]</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Novelty</th>
<th>Target</th>
<th>Computational methods</th>
<th>Main characterization methods</th>
<th>PDB (Å resolution)</th>
<th>Year, first and last authors, journal, and references</th>
<th>Protein structure (first PDB only)</th>
</tr>
</thead>
<tbody>
<tr>
<td>47. De novo thermostable redesign</td>
<td>51-residue <em>Drosophila</em> engrailed homeodomains (two designs)</td>
<td>1. MC with simulated annealing. 2. FASTER with HERO search</td>
<td>CD, NMR</td>
<td>2p6j (NMR)</td>
<td>2007, Shah, Mayo, <em>JMB</em> [64]</td>
<td></td>
</tr>
<tr>
<td>49. Antibody affinity</td>
<td>Electrostatics-based ranking of antibody affinity improvement validated on anti-epidermal growth factor and anti-lysozyme antibodies</td>
<td>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</td>
<td>Antibodies displayed on yeast surface with binding affinity measured by incubating cells with varying antigen concentrations</td>
<td>–</td>
<td>2007, Lippow, Tidor, <em>Nature Biotech</em> [66]</td>
<td></td>
</tr>
<tr>
<td>50. PPI interface</td>
<td>TEM-BLIP</td>
<td>PDBmodDesign</td>
<td>SPR, double-mutant cycles</td>
<td>–</td>
<td>2008, Potapov, Schreiber, <em>JMB</em> [67]</td>
<td></td>
</tr>
<tr>
<td>51. Nobel metal binding in Ferritin</td>
<td>Human H ferritin designed to bind Noble metals that form nanoparticles</td>
<td>Four inner-surface and four outer-surface mutations in each of the 24 subunits. SCADS modified for symmetry</td>
<td>X-ray, MS, CD, DLS, SPR, transmission electron microscopy (TEM)</td>
<td>3erz (3.06 Å, Hg²⁺ complex), 2z6m (2.72 Å, apoprotein), 3es3 (2.8 Å Au³⁺ complex)</td>
<td>2008, Butts, Dmochowski, <em>Biochemistry</em> [68]</td>
<td></td>
</tr>
<tr>
<td>No.</td>
<td>Category</td>
<td>Model</td>
<td>Description</td>
<td>Assays</td>
<td>Methods</td>
<td>Year, Authors/References</td>
</tr>
<tr>
<td>-----</td>
<td>-----------------------------------</td>
<td>-------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>52</td>
<td>PPI interface</td>
<td>SHV-1-BLIP</td>
<td>EGAD two-state minimizer DEE for bound and unbound with MC followed by heuristic step and minimization</td>
<td>X-ray, kinetics measurements</td>
<td>3c4p (1.7 Å) 3c4o (1.7 Å)</td>
<td>2008, Renolds, Handel, <em>JMB</em> [69]</td>
</tr>
<tr>
<td>53</td>
<td>Coiled-coils with four metallo-porphyrin arrays</td>
<td>55-residue four-helix coiled-coil binding four nonbiological iron porphyrins</td>
<td>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</td>
<td>SEC, CD, binding stoichiometry, AUC, EPR</td>
<td>–</td>
<td>2008, McAllister, DeGrado, <em>JACS</em> [70]</td>
</tr>
<tr>
<td>54</td>
<td>Enzyme-design</td>
<td>Retro-aldolase</td>
<td>RosettaDesign &amp; RosettaMatch</td>
<td>X-ray, array of enzyme kinetics assays</td>
<td>3b5v, superseded by 3boj (2.2 Å) and 3bsl (1.8 Å)</td>
<td>2008, Jiang, Baker, <em>Science</em> [71]</td>
</tr>
<tr>
<td>56</td>
<td>Loop remodeling for altering enzyme specificity</td>
<td>Guanine deaminase</td>
<td>Two to five residue loop modeling fitting a hypothetical enzyme-ligand transition state</td>
<td></td>
<td>3e0l (2.37 Å)</td>
<td>2008, Murphy, Baker, <em>PNAS</em> [73]</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Novelty</th>
<th>Target</th>
<th>Computational methods</th>
<th>Main characterization methods</th>
<th>PDB (Å resolution)</th>
<th>Year, first and last authors, journal, and references</th>
</tr>
</thead>
<tbody>
<tr>
<td>57.</td>
<td>Altering binding specificity</td>
<td>Calmodulin–CamKII interaction</td>
<td>18 residue redesign with ORBIT focusing on inter-molecular interactions</td>
<td>CD, SPR</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2009, Yosef, Shifman, JMB [74]</td>
</tr>
<tr>
<td>58.</td>
<td>Binding peptides with high specificity</td>
<td>bZIP-binding peptides including against c-Jun, c-Fox, and c-Maf</td>
<td>CLASSY, negative design</td>
<td>Protein arrays</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2009, Grigoryan, Keating, Nature [75]</td>
</tr>
<tr>
<td>59.</td>
<td>Protein–DNA interaction specificity</td>
<td>Monomeric homing endonuclease I-ÂniI</td>
<td>Eight designs, each with up to six residues redesigned by Rosetta with flexible backbone and a genetic algorithm</td>
<td>Kinetic assays, fluorescence competition binding assay</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2009, Thyme, Baker, Nature [76]</td>
</tr>
<tr>
<td>60.</td>
<td>Altered nonribosomal peptide synthetase enzyme gramicidin S synthetase A (GrsA-PheA) activity</td>
<td>Phenylalanine adenylation domain of GrsA-PheA for a set of noncognate substrates (Leu and charged residues)</td>
<td>K* including active-site (two mutations) and nonactive site “bolstering” mutations (up to two mutations)</td>
<td>Spectrophotometric activity assay</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2009, Chen, Donald, PNAS [77]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2010, Fry, Therien, JACS [78]</td>
</tr>
<tr>
<td>62.</td>
<td>Protein–DNA interaction specificity</td>
<td>I-MsoI homing endonuclease</td>
<td>Six residues redesigned by Rosetta, loop-closure algorithm, and a genetic algorithm</td>
<td>X-ray, DNA-cleavage assay 3mip (2.4 Å), 3mis (2.3 Å), and 3ko2 (2.9 Å)</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2010, Ashworth, Baker, NAR [79]</td>
</tr>
<tr>
<td>63.</td>
<td>Generation of longer emission wavelength in red fluorescent protein</td>
<td>Red fluorescent protein mCherry</td>
<td>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</td>
<td>3nez (1.7 Å), 3ned (0.95 Å)</td>
<td>2010, Chica, Mayo, <em>PNAS</em> [80]</td>
</tr>
<tr>
<td>64.</td>
<td>TM diporphyrin-binding complex PRIME (PorPorphyrins in Membrane)</td>
<td>24-residue TM helices assembling to a four-helix-bundle binding two diphenyl-porphyrins (Fe^{3+}DPP’s)</td>
<td>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</td>
<td>Characterization in a micelle using absorption spectroscopy, CD, and EPR</td>
<td>–</td>
</tr>
<tr>
<td>65.</td>
<td>Protein–protein interaction (PPI)</td>
<td>Hyperplastic discs protein binding to P21-activated kinase 1 kinase domain PAK1</td>
<td>Rosetta-based DDMI protocol (Dock, Design, Minimize Interface) NMR, CD, fluorescence polarization (Chemical shifts ID 1670 in BMRB DB)</td>
<td>–</td>
<td>2010, Jha, Kuhlman, <em>JMB</em> [82]</td>
</tr>
<tr>
<td>66.</td>
<td>Multistate design for stabilization of a protein core</td>
<td>Gβ1, the β1 domain of <em>Streptococcal</em> protein G</td>
<td>Ten core positions underwent multistate CPD using NMR-derived ensemble templates. FASTER, CLEARSS library design</td>
<td>Combinatorial library GdmCl-based stability determination</td>
<td>–</td>
</tr>
<tr>
<td>67.</td>
<td>Prediction of resistance mutations</td>
<td>MRSA DHFR resistance mutations (maintaining function without inhibitor binding)</td>
<td>10 active-site residues subjected to K* CPD algorithm. Four mutants experimentally validated</td>
<td>X-ray, CD, enzyme kinetics 3LG4 (V31Y/F92I double mutant, 3.15 Å)</td>
<td>2010, Frey, Anderson, <em>PNAS</em> [84]</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Novelty</th>
<th>Target</th>
<th>Computational methods</th>
<th>Main characterization methods</th>
<th>PDB (Å resolution)</th>
<th>Year, first and last authors, journal, and references</th>
<th>Protein structure (first PDB only)</th>
</tr>
</thead>
<tbody>
<tr>
<td>70.</td>
<td>De novo design of a binding pair</td>
<td>RosettaDesign, PatchDock, motif search, experimental affinity maturation</td>
<td>X-ray, ELISA, SPR, fluorescence polarization, DLS, CD, NMR</td>
<td>3q9u (2.0 Å), 3q9u (2.3 Å), 3q9u (3q9u only, 1.9 Å)</td>
<td>2011, Karanicolas, Baker, <em>Mol Cell</em> [87]</td>
<td></td>
</tr>
<tr>
<td>71.</td>
<td>PPI: β-strand assembly <em>(βdimer1)</em></td>
<td>Rosetta-based DDMI including five rounds of interface sequence design and minimization</td>
<td>X-ray, AUC, size-exclusion chromatography (SEC), binding-affinity</td>
<td>3zy7 (1.09 Å)</td>
<td>2011, Stranges, Kuhlman, <em>PNAS</em> [88]</td>
<td></td>
</tr>
<tr>
<td>72.</td>
<td>Antigen design</td>
<td>Flexible backbone remodeling and resurfacing</td>
<td>CD, SPR, ELISA</td>
<td>–</td>
<td>2011, Correia, Schief, <em>JMB</em> [89]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Achievements and Challenges in Computational Protein Design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>73.</strong> Thermostable terpene synthase</td>
<td>SCADS, 12 mutations which were all &gt;12 Å from the active site</td>
<td>CD, GC-MS activity assay</td>
<td>2011, Diaz, Weiss, <em>Proteins</em> [90]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>74.</strong> Collagen heterotrimer</td>
<td>Positive and negative design for Pro and hydroxyl-Pro inclusion in triplets and energy gap between model and competing structures. MC/SA</td>
<td>CD</td>
<td>2011, Xu, Nanda, <em>JACS</em> [91]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>75.</strong> Enzyme-design by single mutation (<em>AlleyCat</em>)</td>
<td>Single-site scanning mutagenesis, docking, and super-rotamer modeling of transition state</td>
<td>NMR, CD, enzyme kinetics with pH-profile, CD, 2kz2 (NMR)</td>
<td>2011, Korendovych, DeGrado, <em>PNAS</em> [92]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>76.</strong> Enzyme-design by few mutations</td>
<td>Mutation modeling and docking</td>
<td>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)</td>
<td>2012, Merski, Shoichet, <em>PNAS</em> [93]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>77.</strong> Solubilization of the TM domain of nicotinic acetylcholine receptor α1 subunit</td>
<td>SCADS and loop modeling via MODELLER. Overall 21 residues in the TM region and loops were redesigned</td>
<td>LC/MS, CD, NMR 2lkg (NMR, major conformation), 2lhk (NMR, minor conformation)</td>
<td>2012, Cui, Xu, <em>BBA</em> [94]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>78.</strong> Redesign of a mononuclear Zn metalloenzyme for organo-phosphate hydrolysis</td>
<td>PDB scaffold search, RosettaMatch, RosettaDesign</td>
<td>X-ray, absorption spectra kinetic measurements 3t1g (2.35 Å)</td>
<td>2012, Khare, Baker, <em>Nat Chem Bio</em> [95]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Novelty</th>
<th>Target</th>
<th>Computational methods</th>
<th>Main characterization methods</th>
<th>PDB (Å resolution)</th>
<th>Year, first and last authors, journal, and references</th>
</tr>
</thead>
<tbody>
<tr>
<td>79.</td>
<td>High-affinity epitope scaffold</td>
<td>HIV 2F5 epitope design using backbone-grafting onto scaffold proteins</td>
<td>Six- and seven-residue epitope design via Rosetta, Epitope fragment search in the PDB, loop closure via cyclic coordinate descent (CCD) and MC</td>
<td>X-ray, CD, static light-scattering, SPR</td>
<td>3rij (2.3 Å), 3ri0 (2.25–2.8 Å), 3rhu (Å), 3rin (1.8 Å)</td>
</tr>
<tr>
<td>80.</td>
<td>Protein–protein orthogonal pair within a signaling circuitry</td>
<td>GTPase (Cdc42) redesigned to be activated by a designed GEF intersectin partner</td>
<td>RosettaBackrub, KIC, Rosetta</td>
<td>X-ray, CD, SPR, WASP fluorescence titration, GAP assay, ELISA, live-cell fluorescence microscopy, and XCELLigence Assay</td>
<td>3qbv (2.65 Å)</td>
</tr>
<tr>
<td>Page</td>
<td>Enzyme design ((HG-1, HG-2, HG-3 and directed evolution HG-3.17))</td>
<td>Kemp eliminase</td>
<td>QM, MD, iterative approach</td>
<td>CD, MS, X-ray, enzyme kinetics</td>
<td>3O2L (2.0 Å), 3NYD (1.23 Å), 3NZY (1.51 Å), 3NZ1 (1.56 Å), 4bs0 (1.09 Å)</td>
</tr>
<tr>
<td>------</td>
<td>-------------------------------------------------</td>
<td>-----------------</td>
<td>-----------------------------</td>
<td>--------------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>83.</td>
<td>Iron-sulfur ([4\text{Fe-4S}]) cluster protein</td>
<td>Four-helix Coiled coil iron-sulfur protein (CCIS) monomer</td>
<td>ProtCad with metal-binding first approach</td>
<td>CD, EPR, SEC</td>
<td>–</td>
</tr>
<tr>
<td>84.</td>
<td>Crystal structure space group de novo design</td>
<td>P6-space group crystal design from a three-helix 26-residues coiled-coil</td>
<td>Construction of a ‘grid-like’ ensemble of coiled coil P6-structures followed by side-chain design</td>
<td>X-ray, CD</td>
<td>4dac (2.1 Å), 3v86 (2.91 Å)</td>
</tr>
<tr>
<td>85.</td>
<td>Altered functionality of a de novo design</td>
<td>Altering the oxidation of hydroquinones to hydroxylation of arylamines in a designed four-helix bundle</td>
<td>MSL</td>
<td>NMR, CD</td>
<td>2i1d (NMR)</td>
</tr>
<tr>
<td>86.</td>
<td>Conformational switch ((UniRapR))</td>
<td>Rapamycin-regulated single chain of FKB12 and FKB12-Rapamycin binding protein</td>
<td>Medusa, replica exchange and discrete molecular dynamics</td>
<td>Immunoprecipitation, live-cell and live zebrafish imaging</td>
<td>–</td>
</tr>
<tr>
<td>87.</td>
<td>Binder of a hyperpolarizable abiological chromophore ((SCRPZ-1, SCRPZ-2, SCRPZ-3))</td>
<td>Four helix bundle binding a ruthenium-zinc abiological hyperpolarizable chromophore</td>
<td>SCADS, loop grafting, MD</td>
<td>CD, Pump-probe absorption electronic spectroscopy, Hyper-Rayleigh light scattering, X-ray reflectivity</td>
<td>–</td>
</tr>
<tr>
<td>Novemeber</td>
<td>Target</td>
<td>Computational methods</td>
<td>Main characterization methods</td>
<td>PDB (Å resolution)</td>
<td>Year, first and last authors, journal, and references</td>
</tr>
<tr>
<td>-----------</td>
<td>--------</td>
<td>-----------------------</td>
<td>------------------------------</td>
<td>-------------------</td>
<td>-------------------------------------------------</td>
</tr>
<tr>
<td>89.</td>
<td>GPCR solubilization and application as a biosensor</td>
<td>μ-opioid receptor</td>
<td>SCADS, comparative modeling, solubilized by mutating 53 membrane-exposed residues</td>
<td>CD, MS, homogeneous time-resolved fluorescence (HTRF)</td>
<td>–</td>
</tr>
<tr>
<td>90.</td>
<td>De novo lysozyme protein inhibitor</td>
<td>Hen egg lysozyme hot-spot centric active site protein binder</td>
<td>Target-surface scaffold selection, RosettaDock, RosettaDesign + directed evolution</td>
<td>X-ray, SEC, 19F NMR, SEC, HEL activity assay, Yeast surface display</td>
<td>3vb8 (2.9 Å)</td>
</tr>
<tr>
<td>91.</td>
<td>De novo high-affinity and selective ligand-binding protein</td>
<td>Selective, high-affinity binder of the steroid digoxigenin</td>
<td>Ligand conformer library, RosettaMatch, RosettaDesign, PDB scaffold scan, CCP4 (shape complementarity)</td>
<td>X-ray, SEC, CD, AUC, fluorescence polarization, isothermal titration calorimetry (ITC), yeast surface display</td>
<td>4j8t (2.05 Å), 4q9a (3.2 Å)</td>
</tr>
<tr>
<td>92.</td>
<td>Hot-spot centric de novo pH-sensitive IgG binding protein</td>
<td>Fc-domain His-433 pH-sensitive IgG binding protein</td>
<td>RosettaMatch, RosettaDesign, PDB scaffold scan</td>
<td>CD, ELISA, directed evolution</td>
<td>–</td>
</tr>
<tr>
<td>93.</td>
<td>Statistical energy function (SEF) boosted by experimental selection for foldability for automated CPD</td>
<td>Novel SEF exemplified by de novo CPD of a helical and mixed topology protein</td>
<td>New SEF with selection of structure neighbors with adaptive criteria (SSNAC) to address the multidimensional properties jointly, MC/SA minimization</td>
<td>CD, NMR, antibiotic resistance</td>
<td>2m1b (NMR), 2mn4 (NMR)</td>
</tr>
<tr>
<td></td>
<td>Achievements and Challenges in Computational Protein Design</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-----------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>94.</td>
<td><strong>Helical bundles</strong> 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</td>
<td>Parametric backbone generation and Rosetta</td>
<td>CD, X-ray, EM</td>
<td>4uot (five-helix bundle, 1.69 Å), 4tql (three-helix bundle, 2.8 Å), 4uos (four-helix bundle, 1.63 Å)</td>
<td>2014, Huang, Baker, <em>Science</em> [113]</td>
</tr>
<tr>
<td>95.</td>
<td><strong>Water-soluble α-helical barrels</strong> Pentameric, hexameric and heptameric Water-soluble α-helical barrels</td>
<td>CCBuilder, SOCKET, PoreWalker, coiled-coil assembly rules</td>
<td>X-ray, CD, AUC</td>
<td>4pna (2.0 Å), 4pn8 (2.0 Å), 4pn9 (2.0 Å), 4pnb (2.0 Å), 4pnd (2.0 Å)</td>
<td>2014, Thomson, Woolfson, <em>Science</em> [114]</td>
</tr>
<tr>
<td>96.</td>
<td><strong>Synthetic coiled-coils</strong> Antiparallel homodimeric coiled-coils with no cross-specificity</td>
<td>DFIRE, CCCP structure generator, CLASSY</td>
<td>AUC, disulfide-exchange, CD</td>
<td>–</td>
<td>2014, Negrod, Keating, <em>JACS</em> [115]</td>
</tr>
<tr>
<td>97.</td>
<td><strong>Epitope-focused vaccine design</strong> RSV helix-turn-helix epitope in a three-helix bundle scaffold</td>
<td>Fold-from-loops (FFL) protocol in Rosetta</td>
<td>X-ray, NMR-HSQC, CD, SPR, ELISA (on immunized macaques)</td>
<td>4L8I (2.0 Å), 4JLR (2.71 Å), 4N9G (2.5 Å)</td>
<td>2014, Correia, Schief, <em>Nature</em> [116]</td>
</tr>
<tr>
<td>98.</td>
<td><strong>TM transporter (Rocker)</strong> De novo TM Zn$^{2+}$-transporting four-helix-bundle</td>
<td>MaDCaT, Ez, MD, Volocidy</td>
<td>X-ray, ssNMR ($^1$H, $^13$C cross-polarization MAS, $^{19}$F-CODEX), AUC, Liposome flux</td>
<td>4p6j (2.8 Å), 4p6k (2.7 Å), 4p6l (2.8 Å), 2muz (NMR)</td>
<td>2014, Joh, DeGrado <em>Science</em> [117]</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Novelties</th>
<th>Target</th>
<th>Computational methods</th>
<th>Main characterization methods</th>
<th>PDB (Å resolution)</th>
<th>Year, first and last authors, journal, and references</th>
<th>Protein structure (first PDB only)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100.</td>
<td>Self-assembling symmetry</td>
<td>β-propeller with perfect six-bladed symmetry</td>
<td>Visual template datamining, RosettaDock, phylogenetic ancestral reconstruction, circular permutation for the ‘Vekro’ strap</td>
<td>X-ray, CD, AUC, MS, Differential Scanning Fluorimetry</td>
<td>3ww7 (1.7 Å), 3ww8 (1.4 Å), 3ww9 (1.33 Å), 3wwa (1.99 Å), 3wwb (1.7 Å), 3wwf (1.6 Å)</td>
<td>2014, Voet, Tame, <em>PNAS</em> [119]</td>
</tr>
<tr>
<td>101.</td>
<td>Repeat protein</td>
<td>Leucine-rich repeat protein from ribonuclease inhibitor family</td>
<td>Rosetta, new code for adjusting repeat geometry</td>
<td>CD, FTIR, NMR HSQC, AUC</td>
<td>–</td>
<td>2014, Ramisch, Andre, <em>PNAS</em> [120]</td>
</tr>
</tbody>
</table>
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 ΔΔG 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 α-helical regions, no net increase in the number of β-branched amino acids (Val and Ile) was allowed. While each addition of a β-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 β-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 β-sheet dimer designed from scratch included an intersubunit disulfide bridge. Internal side chains were chosen for their statistical preference for β-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 β-sheets.

### 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, β-sheets, mixed α-β 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 β-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 α-helical 434 Cro protein, ubiquitin is mainly composed of β-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 ββα 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 Cα-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 β1 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, α3C, 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$$\beta$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. β-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 ~25-fold ($K_M = 170 \pm 20 \muM, k_{cat} = 4.6 \pm 0.2 \times 10^{-4} s^{-1}$) 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$_\delta$ 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.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Template protein</th>
<th>Catalysis $k_{\text{cat}}/K_M$ [M$^{-1}$ min$^{-1}$]</th>
<th>Catalysis $[(k_{\text{cat}}/K_M)/k_{\text{uncat}}]$</th>
<th>pH optimum (catalytic residue)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalytic antibodies</td>
<td>Immunglobulin</td>
<td>5500</td>
<td>$-5 \times 10^9$ (pH 7.5)</td>
<td>pH &gt; 7.0 (Glu)</td>
<td>[137]</td>
</tr>
<tr>
<td>CPD iterative approach (QM, MD, + directed evolution)</td>
<td>Xylanase</td>
<td>430 ($2.3 \times 10^5$)</td>
<td>$4 \times 10^8$ (pH 7.25)</td>
<td>pH = 7.0 (Asp) 18</td>
<td>[100, 101]</td>
</tr>
<tr>
<td>CPD (Rosetta, KE70 + directed evolution)</td>
<td>Aldolase</td>
<td>126 (54,800)</td>
<td>$1 \times 10^6$ ($5 \times 10^{10}$, pH 7.25)</td>
<td>pH &gt; 7.0 (His-Asp dyad)</td>
<td>[138]</td>
</tr>
<tr>
<td>CPD (Rosetta, KE07 + directed evolution)</td>
<td>Hisf protein</td>
<td>12.2 (2600)</td>
<td>$1 \times 10^7$ ($2 \times 10^6$, pH 7.25)</td>
<td>pH &gt; 7.0 (Glu) 7,20</td>
<td>[72, 139]</td>
</tr>
<tr>
<td>CPD (Rosetta, KE59 + directed evolution)</td>
<td>Glycerophosphate synthase</td>
<td>$\sim 160$ (60,430)</td>
<td>$1 \times 10^8$ ($5 \times 10^{10}$, pH 7.25)</td>
<td>pH &gt; 8.0 (Glu) 7,21</td>
<td>[72, 140]</td>
</tr>
<tr>
<td>AlleyCat—CPD single-site mutation (super-rotamer minimization)</td>
<td>Calmodulin C-terminal domain</td>
<td>5.8</td>
<td>$4 \times 10^5$ (pH 8.0)</td>
<td>pH &gt; 8.0 (Glu) 22</td>
<td>[92]</td>
</tr>
<tr>
<td>CPD artificial cavity (most active construct L99A/M102H)</td>
<td>T4-lysozyme</td>
<td>1.8</td>
<td>$7 \times 10^7$ (pH 5.0)</td>
<td>pH = 5.0 (His)</td>
<td>[93]</td>
</tr>
</tbody>
</table>
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 DFitet [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 β-barrel. This was further validated by conducting Ile → 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 χ_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 (ε) 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 40 to 4. Third, a romater library that contained rotamers representing expansion about the χ_1 and χ_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 β-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 Å² 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 β-turn formation. All interface positions were redesigned using RosettaDesign
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 Statistical 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 ±5° perturbation of the backbone Φ and Ψ angles till a predetermined (0.3 Å) RMSD. Three antiparallel strands fold into a β-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 (A2B2) 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) α/β-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—β1 domain (Gβ1) 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 b-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 D2 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 Fcγ-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 Fcγ-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 αIIbβ3 and αvβ3 integrins were the subject of the design by replacing the β3 subunit with a new designed helix. The two subunits form a parallel GASRight motif [147] which was structurally modeled with the β3 subunit was redesigned. Five and 15 template backbones were tested for the design of the CHAMP against the αIIb and αv 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 β-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 PLCδ1-PH (pleckstrin homology domain of phospholipase C-δ1) 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\(_2\)-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 β-lactamase and its inhibitor β-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 Au^{3+} and 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, 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 Ag^{0} and 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 α-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 \(k_{cat}/K_M\) 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-AnI 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 A2B2 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^3+. 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 π-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 (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 $\mu$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 ~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 β-strand assembly in which two solvent-exposed strands were designed to form an antiparallel β-strand pairing [88]. Looking for solvent exposed β-strands, automatic homodimer docking (similar to the DDMI protocol) was applied with the β-strand part designed with five rounds of symmetric sequence optimization and minimization at the interface; searching for an $>850 \text{ Å}^2$ buried interface and minimizing unsatisfied buried polar atoms. Of the 5500 structures scanned, 1100 had an exposed β-strand. One structure, γ-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 βdimer1 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 ($k_{cat}/K_M = 5.8 \pm 0.3 \text{ M}^{-1} \text{ s}^{-1}$) 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 Cβ atoms within 8 Å of the amino acids: for residues with 0–6 Cβ atoms were constrained to charged, polar, and small residues. For those with 7–8 Cβ 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 Cβ 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 (α1 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<sub>P</sub> 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 \( k_{\text{cat}} \) 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 con-
sidered 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
\( \frac{k_{\text{cat}}}{K_M} = 1.8 \text{ M}^{-1} \text{ min}^{-1} \). 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 exper-
imental, is key to the design efforts. In this iterative approach the
first designs had low stability of \( \Delta \Delta G \approx 7 \text{ kcal/mol} \) relative to
wild-type T4 lysozyme while subsequent designs increased stability
to \( \Delta \Delta G \approx 2 \text{ kcal/mol} \) 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 exist-
ing cellular circuitry requires consideration of certain design cri-
teria: Not only must the redesigned GTPase be activated by its
redesigned GEF partner, but it must also be protected from inad-
vertent activation by the wild-type GEF and all other endogenous
GEFs. Further, the redesigned GTPase must also preserve interac-
tions 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 simul-
ations 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

Achievements and Challenges in Computational Protein Design 71
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 later 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_{3}G motif interfacing the coiled-coil interhelical contact and an antiparallel GX_{3}GX_{3}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_{2}-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^{6}-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 FKB12 were removed making the N- and C-terminii 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 SCRPZ-1 and SCRPZ-2 for the dimeric and monomeric form, respectively. The protein binds RuPZn, a hyperpolarizable super-molecular chromophore that features highly conjugated (porphnato)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 SCRPZ-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 μ-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 (naltraxone) binding affinity. Interestingly, the CPD was not conducted on a high-resolution known structure but rather on a comparative model using the β2 adrenergic receptor as a model with the subsequent structure of the murine μ-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 Cβ 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 ΔΔG 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 to 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 SEF 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-β-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 $>95 \, ^\circ C$ 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 Zn\textsuperscript{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 β-propeller protein [119]. Visual examination of 174 β-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 α/β-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]

<table>
<thead>
<tr>
<th>Year</th>
<th>Lab</th>
<th>Software</th>
<th>Publication of main software</th>
<th>Lab publications discussed in the chapter</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>1985–</td>
<td>DeGrado ProtCad for few studies but most CPD case-studies with numerous other tools.</td>
<td>[160]</td>
<td>[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]</td>
</tr>
<tr>
<td>2.</td>
<td>1987</td>
<td>Ponder PROPAK</td>
<td></td>
<td>[126]</td>
</tr>
<tr>
<td>3.</td>
<td>1991</td>
<td>Hellinga DEZYMER</td>
<td>[125]</td>
<td>[11, 125, 166, 167, 171–173]</td>
</tr>
<tr>
<td>4.</td>
<td>1995</td>
<td>Desjardins Repacking of Core (ROC) &amp; SPA (Sequence Prediction Algorithm)</td>
<td>[15, 128]</td>
<td>[15–17, 22, 44, 128, 185]</td>
</tr>
<tr>
<td>5.</td>
<td>1996, 1997</td>
<td>Mayo PDA (Protein Design Automation) and ORBIT (Optimization of Rotamers by Iterative Techniques), Other programs include Phoenix, Triad, Faster</td>
<td>[18, 141]</td>
<td>[18–20, 27, 31, 32, 60, 64, 80, 83, 100, 101, 141]</td>
</tr>
<tr>
<td>7.</td>
<td>2000</td>
<td>Kuhlman RosettaDesign</td>
<td>[134]</td>
<td>[25, 43, 46, 62, 82, 88, 134]</td>
</tr>
<tr>
<td>8.</td>
<td>2001</td>
<td>Saven SCADS (Statistical Computationally Assisted Design Strategy)</td>
<td>[143]</td>
<td>[41, 51, 54–56, 68, 70, 78, 90, 94, 103, 107, 121, 143, 146, 164]</td>
</tr>
<tr>
<td>10.</td>
<td>2014</td>
<td>Chen &amp; Liu Statistical energy function and boosted by experimental selection for foldability</td>
<td>[112]</td>
<td>[112]</td>
</tr>
</tbody>
</table>
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 β-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 undesired, 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

of a designed hydrophobic core variant of ubiquitin. Structure 7(8):967–976


146. 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


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