JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Communication

pubs.acs.org/JACS

Potential of Fragment Recombination for Rational Design of Proteins

Simone Eisenbeis,[†,#] William Proffitt,[‡,#,§] Murray Coles,[†] Vincent Truffault,[†] Sooruban Shanmugaratnam,[†] Jens Meiler,*[,‡] and Birte Höcker*[,†]

[†]Max Planck Institute for Developmental Biology, Spemannstrasse 35, 72076 Tübingen, Germany

[‡]Department of Chemistry, Center for Structural Biology, Vanderbilt University, Nashville, Tennessee 37235, United States

Supporting Information

ABSTRACT: It is hypothesized that protein domains evolved from smaller intrinsically stable subunits via combinatorial assembly. Illegitimate recombination of fragments that encode protein subunits could have quickly led to diversification of protein folds and their functionality. This evolutionary concept presents an attractive strategy to protein engineering, e.g., to create new scaffolds for enzyme design. We previously combined structurally similar parts from two ancient protein folds, the $(\beta\alpha)_8$-barrel and the flavodoxin-like fold. The resulting "hopeful monster" differed significantly from the intended $(\beta\alpha)_8$-barrel fold by an extra β-strand in the core. In this study, we ask what modifications are necessary to form the intended structure and what potential this approach has for the rational design of functional proteins. Guided by computational design, we optimized the interface between the fragments with five targeted mutations yielding a stable, monomeric protein whose predicted structure was verified experimentally. We further tested binding of a phosphorylated compound and detected that some affinity was already present due to an intact phosphate-binding site provided by one fragment. The affinity could be improved quickly to the level of natural proteins by introducing two additional mutations. The study illustrates the potential of recombining protein fragments with unique properties to design new and functional proteins, offering both a possible pathway of protein evolution and a protocol to rapidly engineer proteins for new applications.

Today's protein world is extremely diverse. It evolved to facilitate a large variety of functions. However, careful analysis revealed that many proteins of different folds share fragments that are structurally similar.[1] This observation led to the proposition that protein domains evolved by combinatorial assembly of smaller gene fragments that encode intrinsically stable subunits.[2,3] Illegitimate recombination of such subunits could have quickly led to diversification of domain architecture, generating proteins from which new folds and functions could have emerged. Here, we present compelling experimental evidence for this hypothesis by demonstrating that fragments from contemporary proteins are easily adapted to form a new protein with selectable properties (Figure 1). Furthermore, this successful rational design is proof of principle that fragment recruitment from present-day proteins can be used to generate new scaffolds with ready-made and easily adaptable properties.

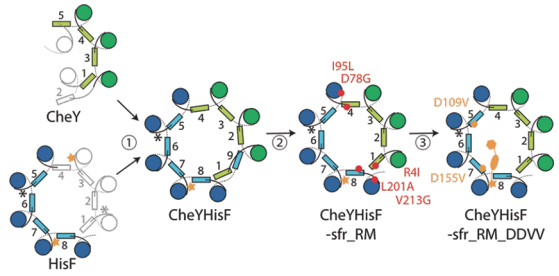

Figure 1. Schematic overview. Fragments from CheY (green) and HisF (blue) were combined (1) to form a new barrel-like protein (CheYHisF, β-strands are numbered). Removal of residues that formed an unexpected ninth strand and introduction of computationally predicted mutations (labeled in red) (2) led to a compact eight-stranded barrel (CheYHisF-sfr_RM). A high-affinity binding pocket for rCdRP (orange handle), a product analogue of the TrpF reaction, was established (3) by using a phosphate-binding site (orange star) contributed by the HisF fragment and introduction of two additional mutations (labeled in orange).

Recent successful approaches in computational enzyme design construct a new catalytic site into known protein scaffolds.[4,5] Thus, it would be advantageous to start with a protein that already has the propensity for a certain type of reaction, analogous to how evolution recruits protein scaffolds, or fragments thereof, that then evolve into specialized enzymes.

For the present study, protein fragments from two major folds were selected: the TIM- or $(\beta\alpha)_8$-barrel and the flavodoxin-like fold. The $(\beta\alpha)_8$-barrel, commonly found among enzymes, consists of a closed eight-stranded parallel β-sheet that forms a central barrel surrounded by eight α-helices.[6] The remarkable two-fold symmetry of two enzymes from histidine biosynthesis indicates that this fold evolved from an ancestral "half-barrel" fragment through duplication and fusion.[7−12] Diversification of $(\beta\alpha)_8$-barrel enzymes by exchange of "half-barrels" was further elucidated by combining halves from related $(\beta\alpha)_8$-barrel proteins.[8,13] In contrast, the flavodoxin-like fold is associated with a diverse mixture of functionalities such as response regulation in signaling systems, cofactor binding, or enzymatic activities. It is a three-layered fold made up of a parallel five-stranded β-sheet flanked by two α-helices on one side and three on the other, found both as an isolated domain and as part of multidomain proteins. Both the

Received: December 14, 2011
Published: February 13, 2012

dx.doi.org/10.1021/ja211657k | J. Am. Chem. Soc. 2012, 134, 4019−4022

flavodoxin-like fold and the $(\beta\alpha)_8$-barrel are considered to belong to the nine most ancient protein folds.[14]

We previously attempted to construct a chimeric $(\beta\alpha)_8$-barrel by combining fragments from these two different folds.[15] Based on a strong structural similarity observed between proteins of both folds,[16] we combined parts of the response regulator CheY and the enzyme imidazole glycerol phosphate synthase (HisF), both from *Thermotoga maritima*. The resulting protein CheYHisF was a stable monomer that unfolds cooperatively. Although its crystal structure confirmed a barrel-like fold and showed that the fragments retain their overall structures within this new context, it also revealed an unexpected additional β-strand embedded within the central barrel, formed by residues of the C-terminus including a histidine purification tag. Removal of the strand-forming residues yielded a variant (CheYHisF-sfr) that formed higher oligomers than CheYHisF. We hypothesized that this structural element relieves tension caused by nonoptimal packing at the interfaces of the combined fragments. Thus, we employed a computational protein design approach to predict a minimal set of mutations that relieve this tension and stabilize the intended $(\beta\alpha)_8$-barrel.

For the computational approach, we generated a model of an eight-stranded CheYHisF with the program Modeller[17] using an alignment of CheY (PDB 1TMY) and HisF (PDB 1THF). This initial model was minimized using the program Rosetta[18] to determine its predicted stability in the Rosetta energy function, expressed in Rosetta energy units per amino acid (REU/AA). Likewise, the parent proteins CheY and HisF were minimized. We then calculated a ΔREU value for each residue, representing the difference in thermodynamic stability between the native structure and the engineered native composite. While the energy of most residues did not change, an increase was observed for a number of residues at the interface of the two fragments (Figure 2A). We therefore used Rosetta to introduce mutations that decrease the overall energy of the model. A mutation had to overcome a threshold before it was considered for experimental validation (see Supporting Information). This threshold was introduced to determine the least number of residues needed to rescue the structure and to ensure that the predicted energy improvement is larger than the uncertainty inherent to the computational method. We chose not only to alter the identity of suboptimal interface residues but also to include all amino acids in the design simulation, as mutations in the second or even third shell might relieve tension at the interface. Altogether, 3600 redesigned models were generated, and the energy for each was plotted against the number of mutations involved (Figure S1). Introducing six or seven mutations caused the energy of the model to decrease significantly to −2.90 REU/AA, approaching the baseline energy of −2.94 REU/AA calculated for the minimized native composite. To determine a limited set of mutations, we chose to characterize these mutants before considering additional ones. Mutations at six positions were consistently predicted in silico to provide a significant improvement (Table S1). To ensure that these mutations do not also stabilize the nine-stranded barrel, they were tested in silico in the context of the nine-stranded CheYHisF crystal structure (PDB 3CWO). All were found to be neutral or destabilizing.

For experimental validation, we introduced the five most favorable mutations into CheYHisF-sfr: R4I, D78G, I95L, L201A, and V213G (CheYHisF-sfr_RM). The calculated energy of −2.90 REU/AA for this variant is significantly

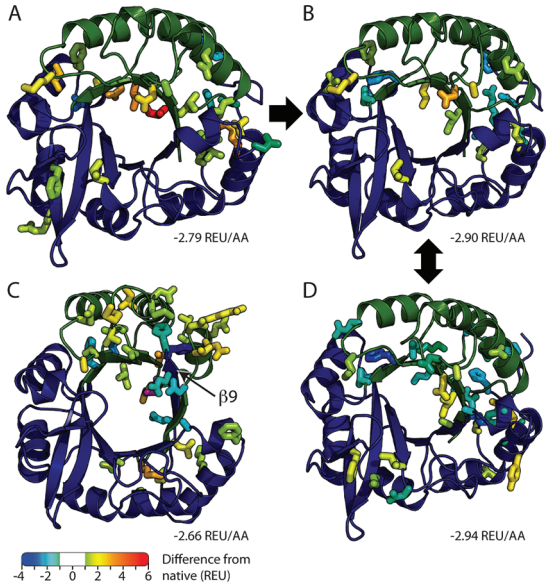

Figure 2. Comparison of predicted versus experimental structures. Models and experimentally determined structures of the chimera CheYHisF (A,C) and the in silico optimized CheYHisF-sfr_RM (B,D) shown as cartoon. Parts originating from CheY and HisF are colored green and blue, respectively. The side chains shown as sticks are color-coded based on the REU differences to the parent structures CheY and HisF (see text). An arrow indicates the position of the ninth β-strand in the crystal structure of CheYHisF (C). An enlarged version is provided as Figure S2.

improved compared to the energy for the CheYHisF crystal structure (−2.66 REU/AA) and also the CheYHisF model (−2.79 REU/AA, Figure 2A−C). M62P was not considered to limit the number of mutations further. The energy decrease predicted through this exchange was smallest, and visual inspection suggested that the mutation does not relieve tension at the interface. Biophysical characterization of the new variant showed improvements over CheYHisF-sfr in multiple protein characteristics: While CheYHisF-sfr was mainly found in the insoluble fraction of the cell extract, CheYHisF-sfr_RM was mostly expressed into the soluble fraction (Figure 3A). When testing the oligomerization state of the protein in solution by analytical gel filtration, CheYHisF-sfr_RM eluted as a sharp peak with an apparent molecular mass of 24.4 kDa, which corresponds well to the calculated molecular mass for the monomer of 25.5 kDa (Figure 3B). Furthermore, the protein elutes significantly later than CheYHisF, indicating higher compactness of the fold and thus a smaller radius as expected from an eight-stranded barrel. In addition, reversible unfolding by guanidinium chloride revealed a clear gain in stability associated with the introduction of the amino acid exchanges after removal of the strand-forming residues (Figure 3C).

For final validation, we determined the solution structure of CheYHisF-sfr_RM by NMR spectroscopy (PDB 2LLE) and compared it with the computational design. It confirms the overall arrangement of a classical $(\beta\alpha)_8$-barrel including the ellipsoid shape of the barrel, the tilt of the β-sheet, and the packing of the α-helices (Figure 4). Superposition of the solution structure with the computational model yields an rmsd of 0.91 Å over backbone atoms of 225 residues. Energy evaluation of the solution structure yielded −2.94 REU/AA for the best-scoring NMR model (Figure 2D). The difference from the CheYHisF_RM model (−2.90 REU/AA) results from the

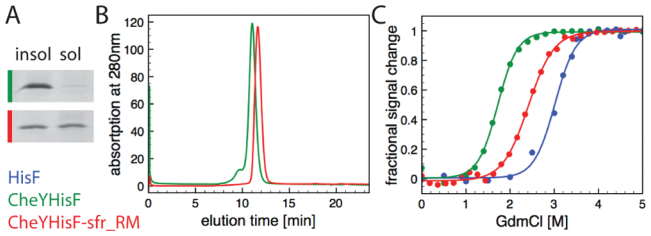

Figure 3. Characterization of the optimized chimera. (A) Soluble expression tested on SDS−PAGE. Equal aliquots from the soluble (sol) and insoluble (insol) fractions of cell extract were loaded. (B) Association states measured by analytical gel filtration. Equal amounts of protein were loaded. (C) Stability measured by GdmCl-induced denaturation. The loss of tertiary structure was followed by recording the decrease of the fluorescence emission.

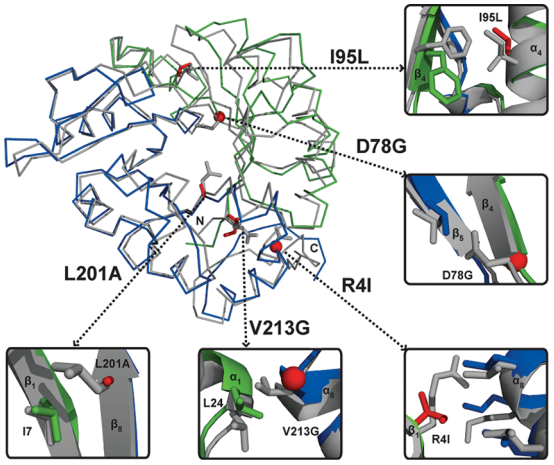

Figure 4. Superimposition of the eight-stranded CheYHisF model (in gray) and the experimental structure of CheYHisF-sfr_RM (*cheY* originating parts in green; *hisF* originating parts in blue) shown as ribbon diagram. The residues mutated based on our calculations are shown as sticks and spheres for glycine (experimental, red; model, gray). Close-up views are shown of the mutated areas.

removal of strand-forming residues and thus exclusion of the flexible histidine-tag from the calculations. Rosetta predicted the conformation of 71% of the side chains correctly when compared to the NMR structure. Alternative side chain conformations are observed for some solvent-exposed residues. When compared to the parent proteins, the structure exhibited 0.48 Å rmsd over the phosphate-binding site that is part of the HisF fragment. Overall, the HisF fragment is predicted with 0.66 Å rmsd, while the CheY fragment is predicted with 0.98 Å. The slight increase in rmsd for the CheY fragment might be connected to its altered curvature within CheYHisF-sfr_RM.

Interestingly, even though the nine-stranded CheYHisF structure was observed experimentally, energy calculations indicate a higher stability for the eight-stranded model (−2.66 vs −2.79 REU/AA, Figure 2A,C). We believe that both conformations of CheYHisF are thermodynamically frustrated; regions of both structures are outside energetic minima. For this reason, energy evaluation for these regions will be associated with larger error, as the Rosetta knowledge-based energy function is derived from low energy experimental structures. We speculate that this increased error causes the incorrect energy ranking of the two structures. It is further possible that kinetic foldability or crystallization conditions favor formation of the nine-stranded conformation.

When comparing the experimental structures of CheYHisF (3CWO) and CheYHisF-sfr_RM, the most drastic structural change is in β-strand 1, which is flipped compared to the one in the CheYHisF crystal structure. Previously solvent-exposed residues are now pointing toward the protein core. This is made possible by the R4I mutation, as the polar and basic arginine, which favors solvent exposure, is replaced with the apolar isoleucine that packs well with α-helix 8. The L201A and V213G mutations in β_8 and α_8, respectively, allow better packing at the HisF/CheY interface. L201A relieves steric hindrance with I7 in β_1, and V213G avoids clashes with L24 in α_1. Similarly, the mutations I95L and D78G optimize the interface at the opposite side of the barrel: I95L permits better packing of α_4 with β_4, while the change from a charged to a small neutral residue associated with the D78G change leads to better packing in the compact core of the barrel (Figure 4). The most unusual observation is the different conformations that the β1 strand can adopt. In CheYHisF, residues 4, 6, and 8 form hydrogen bonds with β9, while in CheY and CheYHisF-sfr_RM, the same residues hydrogen bond to β2. Observations of β-strand flips are rare, and it is unclear how often they might occur. However, sliding and flipping of a β-strand within an antiparallel sheet has been described in the argonaute silencing complex upon binding of a large target RNA,[19] and β-hairpin flips have been proposed as a mechanism in evolution to generate unusual topologies.[1]

Comparison of the β-barrels of CheYHisF-sfr_RM and HisF reveals the same radii and barrel shape. The curvature of the β-sheets is well conserved. The curvature is the highest where the CheY and HisF fragments are joined. This observation is consistent with the notion that the mutations mainly relieve steric hindrances at this interface.

After observing the positive effect of the combined mutations, we analyzed the contribution of each individual mutation. We constructed five variants of CheYHisF-sfr_RM, each having a single Rosetta mutation reverted. The effects are summarized in Table S2. The solubility is affected by the removal of the D78G, I95L, and L201A mutations: if one of these is missing, the protein is mainly found in the insoluble fraction of the cell extract. In contrast, removal of R4I has only a small effect, and for V213G, no change in solubility is observed. The influence of the mutations on stability was determined by chemical denaturation. The unfolding curves of all variants did not show any intermediates and therefore were analyzed based on two-state folding. Changes were noted in the transition midpoints $D_{1/2}$ as well as in the cooperativity of unfolding m^{app} (Table S2). Furthermore, ΔG in the absence of denaturant, $\Delta G(H_2O)$, was extrapolated from the transition and used to assess the conformational stability of the variants. Because of the extended extrapolation required, the calculated values have large errors. Nonetheless, the data clearly show effects upon removal of each mutation in either the transition midpoints or the cooperativity of unfolding. The largest change in stability is contributed by D78G and I95L.

Construction of a stable, well-folded protein from fragments corresponds, in evolutionary terms, to a gene fusion event followed by accumulation of up to five mutations. However, in evolution, a protein is selected for a specific trait that gives the organism an advantage; that is, it is driven by functionality. The chimera has such a selectable property in the form of an intact binding site for phosphate inherited from HisF.[15] The remainder of the newly constructed binding pocket can now be used to establish more complex functionality. We chose a

dx.doi.org/10.1021/ja211657k | J. Am. Chem. Soc. 2012, 134, 4019−4022

ligand similar to the HisF substrate but with a single phosphate moiety instead of two as a target, namely, reduced 1-(2-carboxyphenylamino)-1-deoxyribulose-5-phosphate (rCdRP). rCdRP is a product analogue of phosphoribosyl anthranilate isomerase (PRAI), a $(\beta\alpha)_8$-barrel that catalyzes a central step in tryptophan biosynthesis. Moreover, establishment of wild-type-like PRAI activity through combined directed evolution and docking studies on HisA,[13] a HisF paralog, suggests similar adjustments to the binding pocket of our chimera. Two crucial mutations in HisA, D127V and D169V, led to removal of negative charges, thus facilitating binding of the negatively charged substrate PRA due to relief of electrostatic repulsion. At equivalent positions in our chimera, we mutated two aspartate residues, D109 and D155, to valine (CheYHisF-sfr_RM_DDVV) and then tested rCdRP binding by fluorescence titration. Introducing the two mutations improved rCdRP binding 10-fold from 157 to 15 μM (Figure S3). This affinity is in the range of natural PRAI (*Escherichia coli* PRAI has K_d^{rCdRP} = 5 μM). Therefore, a stable and functional protein was created based on fragment recruitment with a few additional mutations, a pathway that could similarly occur during the course of natural evolution.

Our work has implications for the evolution as well as the design of protein folds. It demonstrates how a stable and functional protein domain can evolve through illegitimate gene recombination and few mutations. The order of events is not fixed. To form a thermodynamically stable protein, it is sufficient if at any point in evolution the fragments have sequences that can recombine without tension. The positions where we introduced stabilizing amino acid exchanges are not fully conserved in the parent proteins; thus, the mutations could already have accumulated through random drift[20] before the recombination event occurred. On the other hand, evolution could have sampled variations of the classical $(\beta\alpha)_8$-barrel, such as the earlier created nine-stranded barrel. These "hopeful monsters" (a term put forward by Goldschmidt in 1940 explaining sudden jumps in speciation[21,22]) will often be outcompeted by established folds and might converge again to one of these folds,[23] e.g., a proper $(\beta\alpha)_8$-barrel. However, in rare cases, they could become established in a population as seeds of a new fold.

By illustrating how fragments from different folds can participate in forming a new protein, it becomes apparent that there is plasticity between established proteins of different folds. These transcending evolutionary relationships cannot be captured by the hierarchical nomenclature of sequence and structure databases, but we expect that multiple such relationships exist. Many will have arisen early on in evolution through recombination in an ancestral pool of peptide modules,[3,24] or even later through recombination of sub-domain-size fragments.

We used fragments from contemporary proteins to show that recombining current genes can still lead to functional chimeras and thus is useful for the design of new proteins. A few targeted mutations at the fragments' interface yielded an extremely stable new protein scaffold in which properties of both parents are combined. One of these, namely a ligand binding site, was quickly adapted to an affinity level comparable to that of a natural enzyme. It can be concluded that imitating evolutionary mechanisms is an attractive strategy for designing proteins by recombining natural fragments, such that each contributes its own properties to the designed chimera.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Experimental and computational procedures as well as structure statistic. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

birte.hoecker@tuebingen.mpg.de; jens.meiler@vanderbilt.edu

Present Address

[§]Johns Hopkins University, Baltimore, MD

Author Contributions

[#]These authors contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by Deutsche Forschungsgemeinschaft grant HO 4022/1-1 (B.H.) and Defense Advanced Research Projects Agency grant 04 12, Protein Design Project (J.M.).

■ REFERENCES

(1) Grishin, N. V. *J. Struct. Biol.* **2001**, *134*, 167.
(2) Lupas, A. N.; Ponting, C. P.; Russell, R. B. *J. Struct. Biol.* **2001**, *134*, 191.
(3) Söding, J.; Lupas, A. N. *BioEssays* **2003**, *25*, 837.
(4) Nanda, V.; Koder, R. L. *Nat. Chem.* **2010**, *2*, 15.
(5) Richter, F.; Leaver-Fay, A.; Khare, S. D.; Bjelic, S.; Baker, D. *PLoS One* **2011**, *6*, e19230.
(6) Sterner, R.; Höcker, B. *Chem. Rev.* **2005**, *105*, 4038.
(7) Höcker, B.; Beismann-Driemeyer, S.; Hettwer, S.; Lustig, A.; Sterner, R. *Nat. Struct. Biol.* **2001**, *8*, 32.
(8) Höcker, B.; Claren, J.; Sterner, R. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16448.
(9) Lang, D.; Thoma, R.; Henn-Sax, M.; Sterner, R.; Wilmanns, M. *Science* **2000**, *289*, 1546.
(10) Höcker, B.; Lochner, A.; Seitz, T.; Claren, J.; Sterner, R. *Biochemistry* **2009**, *48*, 1145.
(11) Seitz, T.; Bocola, M.; Claren, J.; Sterner, R. *J. Mol. Biol.* **2007**, *372*, 114.
(12) Fortenberry, C.; Bowman, E. A.; Proffitt, W.; Dorr, B.; Combs, S.; Harp, J.; Mizoue, L.; Meiler, J. *J. Am. Chem. Soc.* **2011**, *133*, 18026.
(13) Claren, J.; Malisi, C.; Höcker, B.; Sterner, R. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 3704.
(14) Caetano-Anolles, G.; Kim, H. S.; Mittenthal, J. E. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 9358.
(15) Bharat, T. A.; Eisenbeis, S.; Zeth, K.; Höcker, B. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 9942.
(16) Höcker, B.; Schmidt, S.; Sterner, R. *FEBS Lett.* **2002**, *510*, 133.
(17) Sali, A.; Blundell, T. L. *J. Mol. Biol.* **1993**, *234*, 779.
(18) Rohl, C. A.; Strauss, C. E.; Misura, K. M.; Baker, D. *Methods Enzymol.* **2004**, *383*, 66.
(19) Wang, Y.; Juranek, S.; Li, H.; Sheng, G.; Wardle, G. S.; Tuschl, T.; Patel, D. J. *Nature* **2009**, *461*, 754.
(20) Bershtein, S.; Segal, M.; Bekerman, R.; Tokuriki, N.; Tawfik, D. S. *Nature* **2006**, *444*, 929.
(21) Goldschmidt, R. *The Material Basis of Evolution*; Yale Univ. Press: New Haven, CT, 1940.
(22) Dietrich, M. R. *Nat. Rev. Genet.* **2003**, *4*, 68.
(23) Lupas, A. N.; Koretke, .K. K. In *Computational Structural Biology: Methods and Applications*; Schwede, T. S., Peitsch, M., Eds.; World Scientific: Singapore, 2008; Chapter 6.
(24) Alva, V.; Remmert, M.; Biegert, A.; Lupas, A. N.; Söding, J. *Protein Sci.* **2010**, *19*, 124.