Computer-based Redesign of a β-Sandwich Protein Suggests that Extensive Negative Design is Not Required for *De Novo* β-Sheet Design

Xiaozhen Hu[1], **Huanchen Wang**[1], **Hengming Ke**[1], and **Brian Kuhlman**[1,*]

[1]*Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, NC, 27599-7260, USA*

Summary

The *de novo* design of globular β-sheet proteins remains largely an unsolved problem. It is unclear if most designs are failing because the designed sequences do not have favorable energies in the target conformations or if more emphasis should be placed on negative design, i.e. explicitly identifying sequences that have poor energies when adopting undesired conformations. We tested if we could redesign the sequence of a naturally occurring β-sheet protein, tenascin, with a design algorithm that does not include explicit negative design. Denaturation experiments indicate that the designs are significantly more stable than the wild type protein and the crystal structure of one design closely matches the design model. These results suggest that extensive negative design is not required to create well-folded β-sandwich proteins. However, it is important to note that negative design elements may be encoded in the conformation of the protein backbone which was preserved from the wild type protein.

Keywords

Computational Protein Design; *De Novo* Protein Design; β-sheet Design; Negative Design

Introduction

Approximately one quarter of all protein domains are made entirely from β-strands and connecting loops (Orengo et al., 1997). β-sheets and β-barrels form relatively rigid structures that serve as excellent scaffolds for loops that can evolve new molecular recognition capabilities; antibodies are an excellent example of this. Despite the obvious importance of β-sheet proteins, we still do not understand them well enough to design them from first principles. Most *de novo* designed β-sheet proteins are prone to aggregation, and there are no *de novo* designs of an all β-sheet protein with more than three β-strands that have been validated with a NMR or crystal structure (Hughes and Waters, 2006; Kortemme et al., 1998; Kraemer-Pecore et al., 2003; Ramirez-Alvarado et al., 1999; Searle and Ciani, 2004). In contrast, several *de*

*corresponding author. Emai: E-mail: bkuhlman@email.unc.edu, Phone: 919 843 0188, Fax: 919 966 2852.

We declare no conflict of interest.

novo designs of all helical or mixed α/β proteins have been validated with high resolution structures (Harbury et al., 1998; Kuhlman et al., 2003; Walsh et al., 1999; Wei et al., 2003).

There may be several reasons why designed globular β-sheet proteins are prone to misfolding and aggregation. Many β-sheet proteins have greater sequence separation between contacting residues (high contact order) and therefore fold more slowly than helical and mixed α/β proteins (Plaxco et al., 1998). Slower folding rates may allow more time for misfolding, domain swapping and aggregation. β-sheet proteins (designed and naturally occurring) are generally enriched in amino acids with a high intrinsic propensity to form β-strands (Chou and Fasman, 1974; Koehl and Levitt, 1999; Minor and Kim, 1994a, b; Nagano, 1973; Smith et al., 1994). While these amino acids are energetically favorable for the target β-sheet structure, they also have a high propensity to aggregate into fibrils or form undesired strand-strand interactions (Fernandez-Escamilla et al., 2004; Garcia-Castellanos et al., 2005; Pawar et al., 2005). β-strands in two-layer β-sheet proteins often have an alternating repeat of hydrophobic and hydrophilic residues; this type of repeat is known to promote undesired strand-strand interactions (Hecht, 1994). β-sheet proteins that do not form barrels have exposed β-strands that may be well suited for forming edge-to-edge interactions. Indeed, it has been observed that naturally occurring β-sheet proteins contain negative design elements that protect them from unwanted edge-edge interactions (Richardson and Richardson, 2002). These include placing charged residues on both sides of the edge strand, using bulges and prolines to prevent optimal hydrogen bonding, and protecting the edge with other portions of the protein.

How many negative design elements are needed to create a well-folded globular β-sheet protein? Is it necessary to explicitly destabilize associations between non-native strand pairings or does the identification of a low free energy sequence for a target structure implicitly destabilize most competing states? In one study on *de novo* designed β-sheet proteins, the placement of a charged residue on the inward side of putative edge strands was shown to stabilize the monomer versus the aggregated state (Wang and Hecht, 2002). This result suggests that negative design elements may not need to be spread throughout the entire sequence. However, high resolution structures have not been solved for these designs, so it is not known if they are adopting the target structure. Other studies in *de novo* β-sheet design have also produced monomeric proteins, but in these cases it is also not certain if the proteins are adopting the target topology (Lim et al., 2000; Quinn et al., 1994; Yan and Erickson, 1994). A recent design of a Rubredoxin mimic is most likely adopting the target fold, but in this case the energy gained from metal binding may preclude the need for extensive negative design (Nanda et al., 2005).

In a previous study we used the design module of the molecular modeling program Rosetta to design a new amino acid sequence for the third FNIII domain of the protein tenascin (Dantas et al., 2003). This domain has 89 residues and forms a Greek Key fold with three β–strands in one sheet and four β-strands in the second sheet. Sheet 1 is formed by strands 1, 2 and 5. Sheet 2 is formed by strands 3, 4, 6 and 7. The side chains were removed from the protein and computational protein design was used to redesign the protein with no explicit knowledge of the wild type sequence. The only energy gap that was explicitly optimized was between the folded state and a reference energy that models the unfolded state and is based on amino acid composition. Rosetta's energy function is dominated by terms that model van der Waals forces, steric repulsion, desolvation energies, torsion energies and hydrogen bonds (Kuhlman et al., 2003; Rohl et al., 2004). Unfortunately, the designed protein, called TEN-D1, aggregated and we were not able to characterize it. This design may have failed because we did not identify a favorable sequence for the target state, or it may have failed because we did not sufficiently destabilize misfolded and aggregated states. Here, we further pursue this question by characterizing a new set of redesigns for the third FNIII domain of tenascin, but with an energy

function that has been specifically parameterized for β-sheet design. As before, we do not include any explicit negative design in the protocol.

Results

Reparameterizing the Rosetta Energy Function

The energy function used by Rosetta for protein design is a weighted sum of a damped 12-6 Lennard-Jones term, an implicit solvation model, an orientation dependent hydrogen bonding term, knowledge-based torsion energies and a set of reference values that control the relative favorability of the 20 amino acids (Rohl et al., 2004). The weights on these terms have been set to maximize the native sequence recovery during the complete redesign of whole proteins (Kuhlman and Baker, 2000). Our standard training set has a mixture of all helical, mixed α/β proteins, and all β proteins. For these studies we assembled a set of 121 high-resolution structures of all β-sheet proteins. The standard Rosetta energy function was used to design sequences for the proteins in the training set and the sequences were compared to the wild type sequences. Overall sequence identity was similar to what we have observed previously, but the fraction of hydrophobic residues in the redesigned sequences was higher than in the naturally occurring sequences (67% versus 53%, Supplementary Table S1 and Table S2). To create more native-like sequences, iterative rounds of perturbing the amino acid reference values and redesigning the proteins were used to arrive at a set of reference values that accurately reproduce the hydrophobic/hydrophilic preferences of the naturally occurring β-sheet proteins (Supplementary Table S1 and Table S2). The goal of our fitting procedure is to improve our ability to perform positive design and find low energy sequences for target structures. However, by adjusting the amino acid reference values and therefore perturbing the overall amino acid composition of the protein we may be implicitly including negative design in our protocol. In this regard, our experiments are testing the importance of explicit negative design with the constraint that overall amino acid composition has been set to resemble naturally occurring β-sheet proteins.

Computational Redesign of Tenascin

Tenascin (pdbcode : 1ten) was used as the starting model for fixed backbone design. All the sidechains were removed from the protein except Tyrosine 869. Tyrosine 869 was not allowed to vary because it forms a sidechain backbone hydrogen bond that is important for the stability of the protein (Hamill et al., 2000). Rosetta prefers to put a phenylalanine at this position because the tyrosine rotamers used during the simulation do not allow for a low energy hydrogen bond. This residue was mutated to a phenylalanine in our previously published redesign of tenascin TEN-D1. 100 independent design trajectories were used to look for low energy sequences. The Rosetta full atom energies in the redesigned models varied between -220 and -215 kcal / mol. The lowest energy model, called TEN-D2, was chosen for experimental characterization.

A second round of design simulations were performed with an additional surface area-based packing score (SASAprob) included in the optimization procedure (Leaver-Fay et al., 2007). The SASAprob score examines the difference in solvent accessibility computed with a 0.5 Å probe and a 1.4 Å probe (the size of water). The difference in these two terms will be greater for underpacked proteins. The score is formulated as a probability based on average values measured for naturally occurring proteins. To optimize this score during a design simulation we have developed a rapid algorithm for computing solvent accessible surface areas during protein design simulations. Our design picked from the first round of simulations, TEN-D2, has a SASAprob score of 0.46, indicating that it is more tightly packed than 46% of the proteins in the PDB. From the second round of simulations, we chose a design called TEN-D3, with a SASAprob score of 0.52 and a total score of -216 kcal / mol.

TEN-D2 has 53 mutations and TEN-D3 has 51 mutations when compared to the wild type sequence (Figure 1, Table 1). Our previously characterized sequence, TEN-D1, had 58 mutations. Highest sequence similarity is seen in the protein core; out of 20 buried residues, 9 were mutated in TEN-D2, and 8 were mutated in TEN-D3. The number of charged residues in the redesigns is significantly different than in the wild type protein. 20% of the wild type residues are negatively charged (Asp or Glu), while only 8% of the redesigns are negatively charged. The most highly conserved amino acids in the redesigns are proline, glycine and threonine. Four out of five prolines, three out of five glycines and ten out of 12 threonines are conserved.

Experimental Characterization

Both TEN-D2 and TEN-D3 were expressed in bacteria and experimentally characterized using a variety of biophysical methods. Size-exclusion chromatographies of the two redesigns suggest they are both monomeric (data not shown). There is good dispersion in the one-dimensional ^{1}H NMR spectra indicating that both redesigns are well-folded (Figure 2), and there are amide protons with chemical shifts above 8.5 ppm, indicative of β-sheet structure. Additionally, the circular dichroism (CD) spectra of the proteins are consistent with β-sheet structure. To probe the stability of the redesigns CD signal was monitored as a function of temperature and concentration of chemical denaturant at a single wavelength. Both TEN-D2 and TEN-D3 unfold at temperatures that are significantly higher than the wild type protein, the proteins unfold above 90 °C and 80 °C respectively (Figures 3 and 4, Table 2). The T_m for wild type tenascin is 58 °C. However, unlike the wild type protein, the thermal unfolding curves for the redesigns are not reversible at pH 7. It has been shown that high net charges can help solubilize proteins in the unfolded state (Lawrence et al., 2007). Consistent with this hypothesis, TEN-D2 refolds reversibly when the pH is dropped below the pK_a of the acidic side chains, increasing the net charge of the design (Figure 3.D).

Denaturation induced by guanidine hydrochloride was monitored with circular dichroism to measure the stability. Both redesigns fold reversibly in chemical denaturant and are significantly more stable than the wild type protein. The extrapolated free energies of folding are -11.9 and -8.7 kcal / mol respectively. The wild type protein has a free energy of folding of -5.1 kcal / mol. Interestingly, the m-values (slope of free energy versus [GuHCl]) are larger for the redesigns. This suggests that the redesigns bury more hydrophobic surface area upon folding than the wild type protein (Myers et al., 1995).

Structure Determination

The crystal structure of the TEN-D3 was determined at 2.4 Å resolution by X-ray crystallography (Supplementary Table S3) to verify that the structure matches the design model. Overall, there is a good match between the crystal structure and the design model, the root-mean-square deviation (RMSD) between the crystal structure and the design model is less than 0.8 Å for all heavy atoms of the protein (Figure 5). 82 percent of the sidechains have the same chi1 rotamer as designed and all the rotamers in the core have the same conformation (chi1 and chi2) as designed. Greater differences were seen on the surface; although several designed salt bridges on the protein surface were observed in the crystal structure. These include interacting pairs, Arg 74 and Asp 48, Asp 43 and Arg 37, and Glu 62 and Arg 37.

Discussion

60% of the residues in the tenascin redesigns are not a direct reflection of natural protein evolution, but rather were chosen solely based on a calculated free energy difference between the target structure and a reference state that only depends on amino acid composition. Despite the simplicity of this design criterion, the proteins fold into the target structure. Similar findings

have been reported for all helical, mixed α/β proteins, and small three stranded β-sheet proteins (Dahiyat and Mayo, 1997; Dantas et al., 2003; Kraemer-Pecore et al., 2003; Scalley-Kim and Baker, 2004). Our result suggests that the majority of amino acids in tenascin have not been explicitly selected to prevent misfolding, but rather selection for a low free energy target structure is sufficient to destabilize alternative folds. This result is not obvious *a priori*, given the fact that small stretches of sequence rich in β-sheet propensity are prone to association and the possible number of non-native strand pairings is much greater than native pairings.

Our results do not indicate that negative design is not important for *de novo* β–sheet design, but they do suggest that it may be sufficient to only focus on a limited number of negative design elements. For instance, the backbone conformation of tenascin appears to include negative design elements. Unwanted edge-to-edge β-strand interactions are most likely destabilized by a β-bulge in strand 1, the shortness of strand 5 and prolines in strand 7. All of these elements are preserved in our redesigns. Additionally, negative design elements may be encoded in the residues that are preserved from the wild type sequence. It is interesting that our designs do not include charged residues on the inward pointing face of the edge strands. Other design elements, such as the prolines in strand 7, must be preventing association between edge strands.

It is striking that the redesigned sequences are considerably more stable than the wild type sequences. Similar results have been observed when redesigning other protein folds with computational protein design software (Dantas et al., 2003; Malakauskas and Mayo, 1998). An increase in the m-values for chemical denaturation suggests that the designs bury more hydrophobic surface area upon folding. This increase is consistent with the addition of extra hydrophobic residues in the redesigns and may explain the increase in protein stability.

Our results are encouraging in that they suggest that the *de novo* design of a β-sandwich protein may be possible without extensive consideration of strand mispairings. Despite this fact, *de novo* design is still a very challenging problem. To create a protein from scratch, it is necessary to identify a protein backbone that allows for tight packing of the side chains and allows for hydrogen bonding to buried polar groups. It is especially challenging to ensure that backbone polar groups in the connecting loops have hydrogen bond partners. Many of these polar groups are removed from solvent, and in naturally occurring proteins are engaged in sidechain-backbone hydrogen bonds. It will be exciting to see if new techniques in computational protein design that allow for backbone sampling and sequence design will allow these hurdles to be overcome.

Experimental Procedures

Sequence Optimization Simulations

Fixed backbone design simulations were performed with svn version 9242 of Rosetta. The standard full atom energy function was used except for the following changes: the reference values were reparameterized to maximize the native sequence recovery test, the desolvation penalty for histidine was increased by varying the ddGfree parameter for histidine nitrogens from -4.0 to -9.0, and the Lennard-Jones potential was set to a linear slope at 0.85 of the van der Waals radius (instead of 0.6). Dunbrack's backbone dependent rotamer library was used with extra chi 1 torsion angles for all residues and extra chi 2 torsion angles for aromatic residues. The command line used for the simulations was: *Rosetta.gcc -s 1ten.pdb -design -fixbb -use_bw -ex1 -ex2aro_only -extrachi_cutoff 1 -resfile resfile -ndruns 100 (-use_sasa_pack_score)*.

Protein Expression and Purification

Genes for the redesigned proteins were synthesized in-house with PCR extension of commercially purchased overlapping oligonucleotides from Operon (Stemmer et al., 1995). The genes were inserted into *E. coli.* expression vector pET21b, with a linker "GSLE" followed by C terminal 6x His tag. The proteins were expressed in the *E. coli.* BL21 strain at 37 °C with 0.5 mM IPTG used for induction. The proteins were purified with a Ni^{++} affinity column followed by size-exclusion chromatography (Superdex-75).

NMR

The two redesigned proteins (~0.4 mM) were equilibrated in 20 mM sodium phosphate, 0.15 M NaCl, pH 7.2 buffer and one-dimensional ^1H NMR spectra were recorded at 25 °C on a Varian Inova 600 MHz spectrometer. NMR data were processed with NMRPipe (Delaglio et al., 1995).

Circular Dichroism

CD data were collected on a JASCO J-810/815 CD spectrometer using a 0.1 cm cuvette with 40 uM proteins. The CD signal was monitored at 215 nm as a function of temperature (4 – 96 °C). The fraction of unfolded protein was calculated assuming that the CD signal of the unfolded and folded protein varies linearly with temperature. GuHCl induced chemical denaturation experiments were recorded at 222 nm. The free energy calculations were obtained with a two-state assumption.

Crystallization, X-ray Diffraction and Structure Determination

The hanging-drop vapor diffusion method was used for crystallization trials. TEN-D3 with the concentration of 12 mg / mL in 100 mM NaCl, 20 mM Tris buffer at pH 7.4, was mixed with an equal volume of well buffer of 0.1 M sodium dihydrogen phosphate, 0.1 M potassium dihydrogen phosphate, 0.1 M MES, pH 6.5, 2.2M NaCl and100 mM urea. 20% glycerol was used as the cyroprotectant. Diffraction data of TEN-D3 were collected at the Beamline x29A at Brookhaven National Laboratory.

The data were indexed and processed with the program HKL2000 (Otwinowski, 1997). The structure of TEN-D3 was solved by molecular replacement using the programs MolRep (Vagin and Teplyakov, 2000) and Phaser (Storoni et al., 2004). Wild type tenascin (PDB code 1TEN) was used as the initial search model. The model was then refined against the synchrotron data to 2.4 Å resolution. O (Vagin and Teplyakov, 2000) was used to build the model and CNS (Brunger et al., 1998) was used to refine the structure. The geometry of the final model was assessed with the program PROCHECK (Laskowski et al., 1993).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, et al. Crystallography & NMR system: A new software suite for

macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 1998;54:905–921. [PubMed: 9757107]

Chou PY, Fasman GD. Conformational parameters for amino acids in helical, beta-sheet, and random coil regions calculated from proteins. Biochemistry 1974;13:211–222. [PubMed: 4358939]

Dahiyat BI, Mayo SL. De novo protein design: fully automated sequence selection. Science 1997;278:82–87. [PubMed: 9311930]

Dantas G, Kuhlman B, Callender D, Wong M, Baker D. A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins. J Mol Biol 2003;332:449–460. [PubMed: 12948494]

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 1995;6:277–293. [PubMed: 8520220]

Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nat Biotechnol 2004;22:1302–1306. [PubMed: 15361882]

Garcia-Castellanos R, Bonet-Figueredo R, Pallares I, Ventura S, Aviles FX, Vendrell J, Gomis-Rutha FX. Detailed molecular comparison between the inhibition mode of A/B-type carboxypeptidases in the zymogen state and by the endogenous inhibitor latexin. Cell Mol Life Sci 2005;62:1996–2014. [PubMed: 16091843]

Hamill SJ, Cota E, Chothia C, Clarke J. Conservation of folding and stability within a protein family: the tyrosine corner as an evolutionary cul-de-sac. J Mol Biol 2000;295:641–649. [PubMed: 10623553]

Harbury PB, Plecs JJ, Tidor B, Alber T, Kim PS. High-resolution protein design with backbone freedom. Science 1998;282:1462–1467. [PubMed: 9822371]

Hecht MH. De novo design of beta-sheet proteins. Proc Natl Acad Sci U S A 1994;91:8729–8730. [PubMed: 8090714]

Hughes RM, Waters ML. Model systems for beta-hairpins and beta-sheets. Curr Opin Struct Biol 2006;16:514–524. [PubMed: 16837192]

Koehl P, Levitt M. Structure-based conformational preferences of amino acids. Proc Natl Acad Sci U S A 1999;96:12524–12529. [PubMed: 10535955]

Kortemme T, Ramirez-Alvarado M, Serrano L. Design of a 20-amino acid, three-stranded beta-sheet protein. Science 1998;281:253–256. [PubMed: 9657719]

Kraemer-Pecore CM, Lecomte JT, Desjarlais JR. A de novo redesign of the WW domain. Protein Sci 2003;12:2194–2205. [PubMed: 14500877]

Kuhlman B, Baker D. Native protein sequences are close to optimal for their structures. Proc Natl Acad Sci U S A 2000;97:10383–10388. [PubMed: 10984534]

Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. Design of a novel globular protein fold with atomic-level accuracy. Science 2003;302:1364–1368. [PubMed: 14631033]

Laskowski RA, Moss DS, Thornton JM. Main-chain bond lengths and bond angles in protein structures. J Mol Biol 1993;231:1049–1067. [PubMed: 8515464]

Lawrence MS, Phillips KJ, Liu DR. Supercharging proteins can impart unusual resilience. J Am Chem Soc 2007;129:10110–10112. [PubMed: 17665911]

Leaver-Fay A, Butterfoss GL, Snoeyink J, Kuhlman B. Maintaining solvent accessible surface area under rotamer substitution for protein design. J Comput Chem 2007;28:1336–1341. [PubMed: 17285560]

Lim A, Makhov AM, Bond J, Inouye H, Connors LH, Griffith JD, Erickson BW, Kirschner DA, Costello CE. Betabellins 15D and 16D, de Novo designed beta-sandwich proteins that have amyloidogenic properties. J Struct Biol 2000;130:363–370. [PubMed: 10940239]

Malakauskas SM, Mayo SL. Design, structure and stability of a hyperthermophilic protein variant. Nat Struct Biol 1998;5:470–475. [PubMed: 9628485]

Minor DL Jr, Kim PS. Context is a major determinant of beta-sheet propensity. Nature 1994a;371:264–267. [PubMed: 8078589]

Minor DL Jr, Kim PS. Measurement of the beta-sheet-forming propensities of amino acids. Nature 1994b; 367:660–663. [PubMed: 8107853]

Myers JK, Pace CN, Scholtz JM. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. Protein Sci 1995;4:2138–2148. [PubMed: 8535251]

Nagano K. Logical analysis of the mechanism of protein folding I Predictions of helices, loops and beta-structures from primary structure. J Mol Biol 1973;75:401–420. [PubMed: 4728695]

Nanda V, Rosenblatt MM, Osyczka A, Kono H, Getahun Z, Dutton PL, Saven JG, Degrado WF. De novo design of a redox-active minimal rubredoxin mimic. J Am Chem Soc 2005;127:5804–5805. [PubMed: 15839675]

Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM. CATH--a hierarchic classification of protein domain structures. Structure 1997;5:1093–1108. [PubMed: 9309224]

Otwinowski, zaMW. Processing of X-ray Diffraction Data Collected in Oscillation Mode. Methods in Enzymology 1997;276:307–326.

Pawar AP, Dubay KF, Zurdo J, Chiti F, Vendruscolo M, Dobson CM. Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases. J Mol Biol 2005;350:379–392. [PubMed: 15925383]

Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. J Mol Biol 1998;277:985–994. [PubMed: 9545386]

Pokala N, Handel TM. Energy functions for protein design: adjustment with protein-protein complex affinities, models for the unfolded state, and negative design of solubility and specificity. J Mol Biol 2005;347:203–227. [PubMed: 15733929]

Quinn TP, Tweedy NB, Williams RW, Richardson JS, Richardson DC. Betadoublet: de novo design, synthesis, and characterization of a beta-sandwich protein. Proc Natl Acad Sci U S A 1994;91:8747–8751. [PubMed: 8090717]

Ramirez-Alvarado M, Kortemme T, Blanco FJ, Serrano L. Beta-hairpin and beta-sheet formation in designed linear peptides. Bioorg Med Chem 1999;7:93–103. [PubMed: 10199660]

Richardson JS, Richardson DC. Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. Proc Natl Acad Sci U S A 2002;99:2754–2759. [PubMed: 11880627]

Rohl CA, Strauss CE, Misura KM, Baker D. Protein structure prediction using Rosetta. Methods Enzymol 2004;383:66–93. [PubMed: 15063647]

Scalley-Kim M, Baker D. Characterization of the folding energy landscapes of computer generated proteins suggests high folding free energy barriers and cooperativity may be consequences of natural selection. J Mol Biol 2004;338:573–583. [PubMed: 15081814]

Searle MS, Ciani B. Design of beta-sheet systems for understanding the thermodynamics and kinetics of protein folding. Curr Opin Struct Biol 2004;14:458–464. [PubMed: 15313241]

Smith CK, Withka JM, Regan L. A thermodynamic scale for the beta-sheet forming tendencies of the amino acids. Biochemistry 1994;33:5510–5517. [PubMed: 8180173]

Stemmer WP, Crameri A, Ha KD, Brennan TM, Heyneker HL. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. Gene 1995;164:49–53. [PubMed: 7590320]

Storoni LC, McCoy AJ, Read RJ. Likelihood-enhanced fast rotation functions. Acta Crystallogr D Biol Crystallogr 2004;60:432–438. [PubMed: 14993666]

Vagin A, Teplyakov A. An approach to multi-copy search in molecular replacement. Acta Crystallogr D Biol Crystallogr 2000;56:1622–1624. [PubMed: 11092928]

Walsh ST, Cheng H, Bryson JW, Roder H, DeGrado WF. Solution structure and dynamics of a de novo designed three-helix bundle protein. Proc Natl Acad Sci U S A 1999;96:5486–5491. [PubMed: 10318910]

Wang W, Hecht MH. Rationally designed mutations convert de novo amyloid-like fibrils into monomeric beta-sheet proteins. Proc Natl Acad Sci U S A 2002;99:2760–2765. [PubMed: 11880628]

Wei Y, Kim S, Fela D, Baum J, Hecht MH. Solution structure of a de novo protein from a designed combinatorial library. Proc Natl Acad Sci U S A 2003;100:13270–13273. [PubMed: 14593201]

Yan Y, Erickson BW. Engineering of betabellin 14D: disulfide-induced folding of a beta-sheet protein. Protein Sci 1994;3:1069–1073. [PubMed: 7920252]

```
TEN-WT   LDAPSQIEVKDVTDTTALITWFKPLAEIDGIELTYGIKDVPGDRTTIDLTEDENQYSIGN
TEN-D1   LPPPYNITVTNIGPTTAVLVYVRSESPSDGYNITFGTKNDDSDRVTVTLPSENTSYVITN
TEN-D2   LQPPFNITVTNITLTTAVVKWLPAQLPVEGYLVTYGRKNDPSDETTVDLTSSITSLTLTN
TEN-D3   LQPPFNIKVTNITLTTAVVTWQPPILPIEGILVTFGRKNDPSDETTVDLTSSITSLTLTN


TEN-WT   LKPDTEYEVSLISRRGDMSSNPAKETFTT
TEN-D1   LKPNTTFQITIRSQNGDKSSPPVSTYFTL
TEN-D2   LTPNTEYEVRIVARNGNLYSPPVSTTFKT
TEN-D3   LEPNTTYEIRIVARNGQQYSPPVSTTFTT
```

Figure 1.
Sequences of the wild type and three redesigned proteins. (TEN-WT: wild type; TEN-D1,TEN-D2,TEN-D3: redesigned sequences). The TEN-D1 sequence is from a previously published study (Dantas et al., 2003).
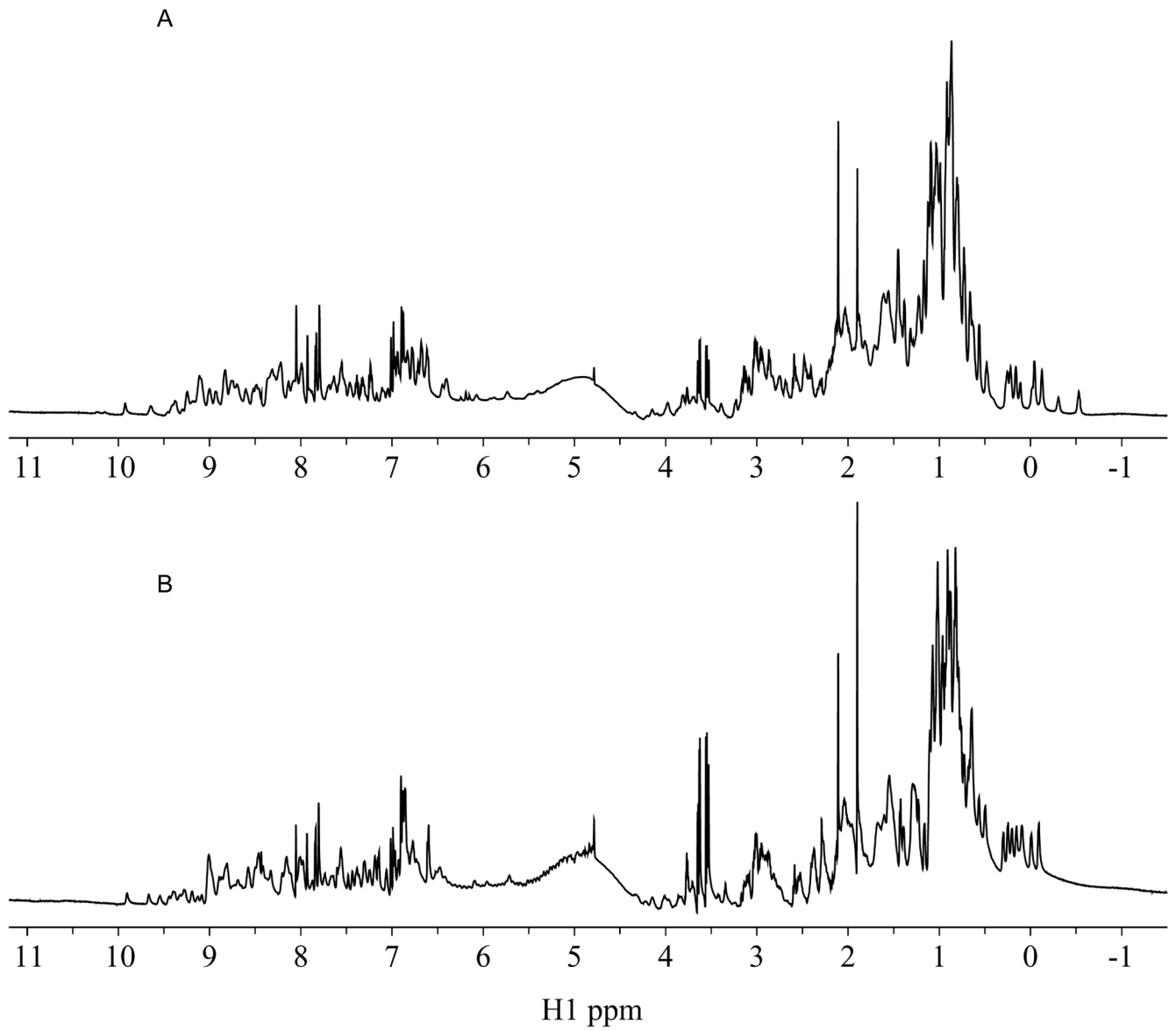
A

B

H1 ppm

Figure 2.
One-dimensional ^{1}H spectra of the redesigned proteins. A: TEN-D2. B: TEN-D3.
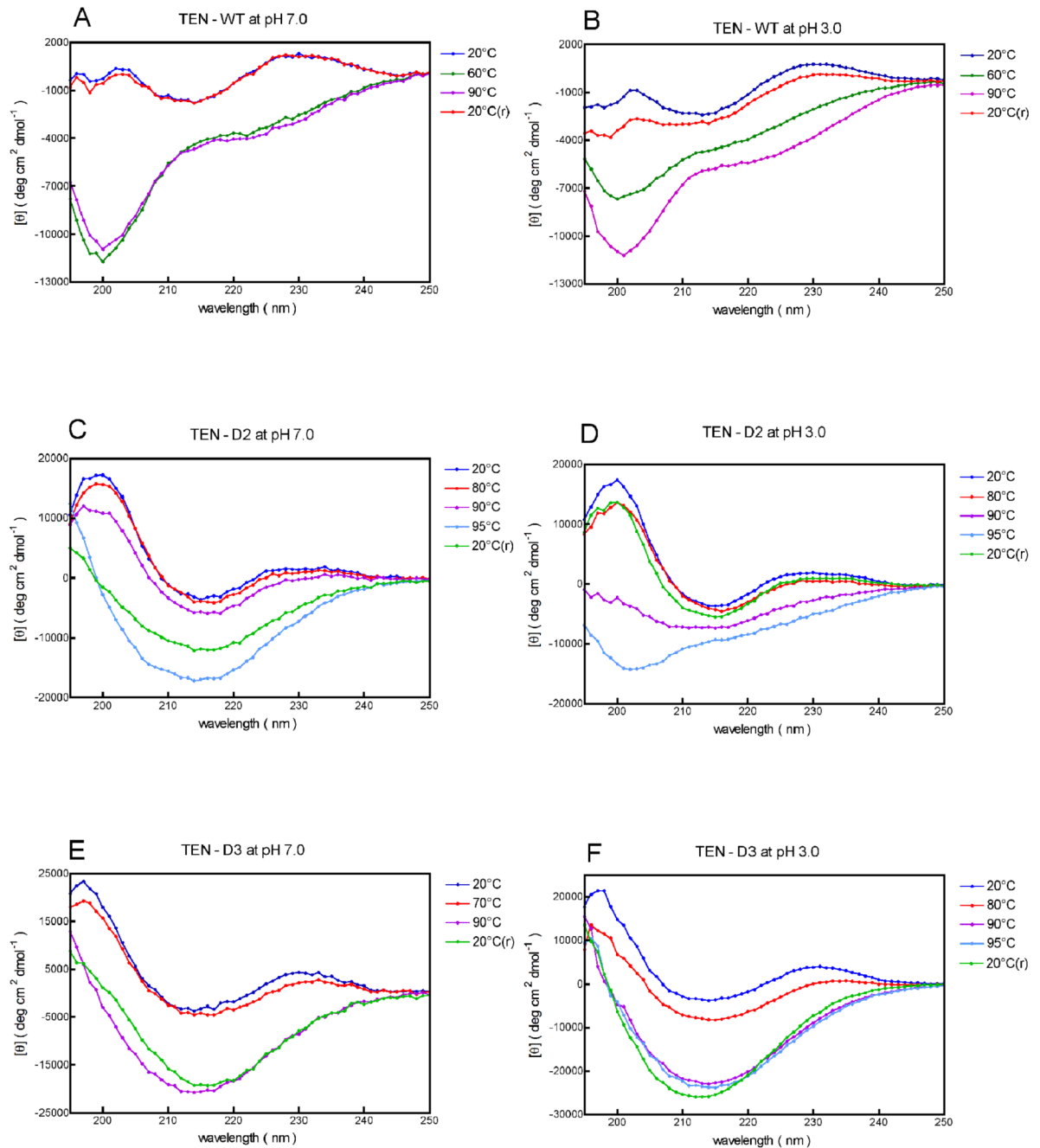
Figure 3.
Circular dichroism spectra of the wild type tenascin and the redesigned proteins at neutral and acidic pH with different temperatures. 20 °C(r) represents that the temperature was cooled back to 20 °C. A: TEN-WT at pH 7.0, B: TEN-WT at pH 3.0, C: TEN-D2 at pH 7.0, D: TEN-D2 at pH 3.0, E: TEN-D3 at pH 7.0, and F: TEN-D3 at pH 3.0.
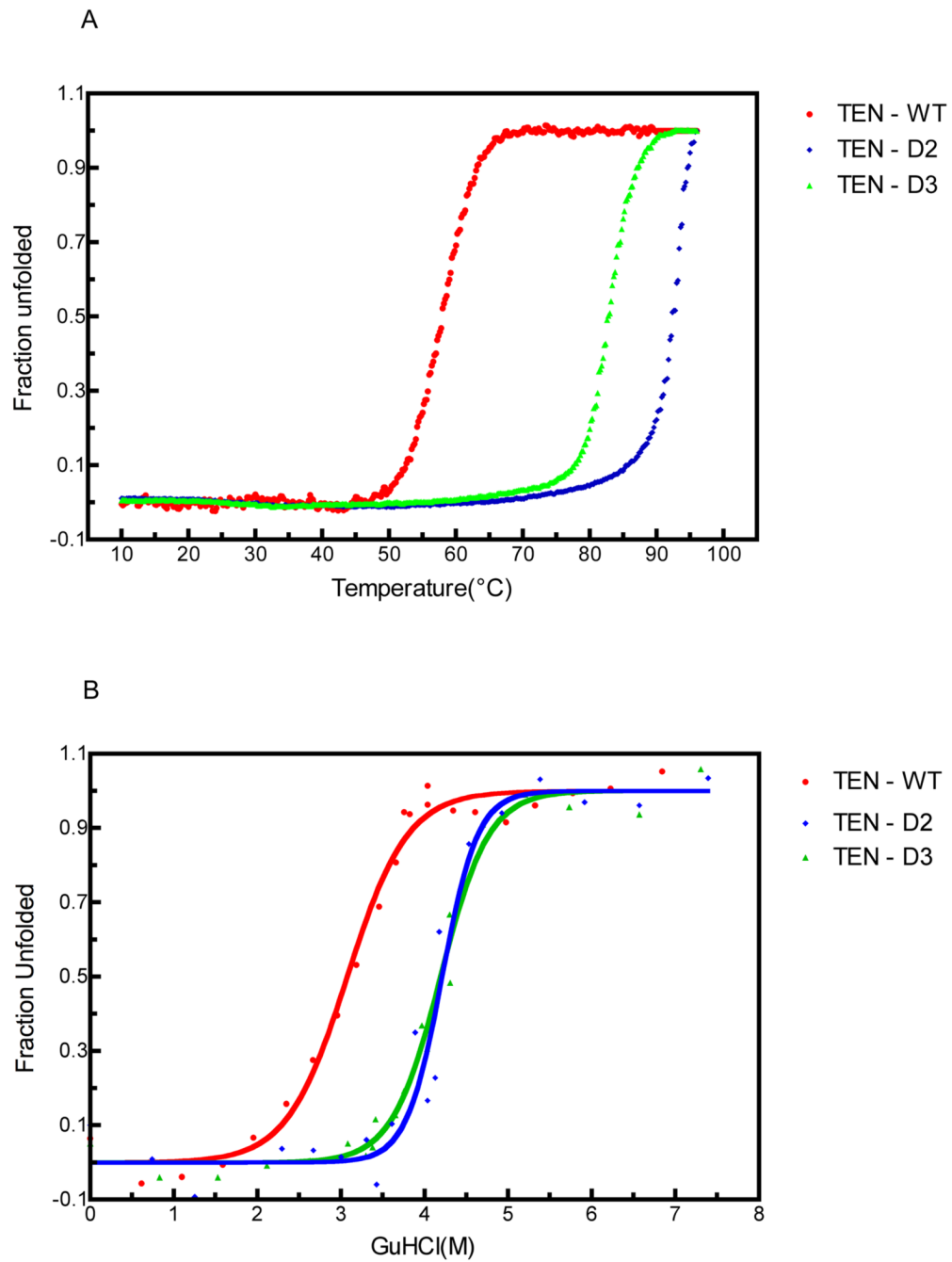
A

B

Figure 4.
Temperature and chemical denaturation as monitored by circular dichroism. A: Thermal unfolding of the wild type tenascin and the redesigned proteins. B: Chemical denaturation of the wild type tenascin, TEN-D2 and TEN-D3.
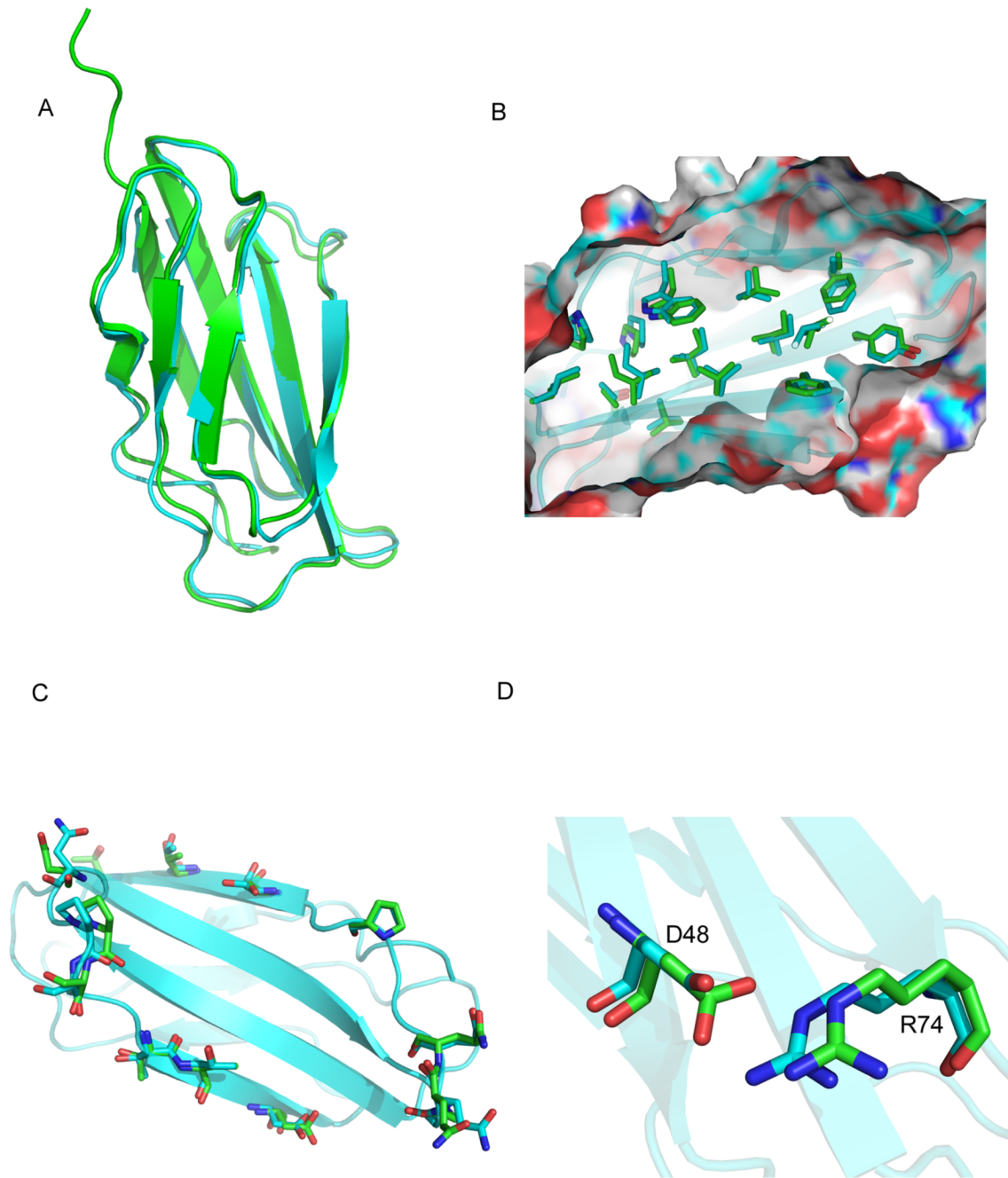
Figure 5.
Structure alignment between the designed model (cyan) and the crystal structure of TEN-D3 (green). A: backbone only, B: buried residues, C: selected surface residues, D: a designed salt bridge between Asp 48 and Arg 74.

Table 1

Sequence features of wild type and redesigned tenascin.

Protein	TEN-WT	TEN-D1	TEN-D2	TEN-D3
MW (Da)	9895.9	9729.7	9800.0	9790.1
Theoretical PI	4.15	4.99	5.10	4.72
Fraction of positively charged residues	0.09	0.07	0.07	0.06
Fraction of negatively charged residues	0.20	0.08	0.08	0.08
Fraction of hydrophobic residues	0.38	0.37	0.42	0.42
Sequence identity to WT (overall)	/	31/89	36/89	38/89
Sequence identity to WT (buried[*])	/	9/20	11/20	12/20
Sequence identity to TEN-D1 (overall)	31/89	/	45/89	48/89

[*] buried – Buried residues have more than 19 neighbors within 10Å.

Table 2

Thermodynamic parameters of wild type and redesigned tenascin.

Protein	Tm(°C)	ΔG_U^{H2O} (kcal mol^{-1})	m-GuHCl (kcal mol^{-1} M^{-1})
TEN–WT	58	5.1±1.2	1.7±0.3
TEN–D1	/	/	/
TEN–D2	>90	11.9±4.7	2.8±1.1
TEN–D3	>80	8.7±2.0	2.1±0.4