

Design of novel FN3 domains with high stability by a consensus sequence approach

Steven A. Jacobs¹, Michael D. Diem, Jinqian Luo, Alexey Teplyakov, Galina Obmolova, Thomas Malia, Gary L. Gilliland and Karyn T. O'Neil

Janssen Research & Development, L.L.C., Radnor, PA 19087, USA

¹To whom correspondence should be addressed.
E-mail: sjacobs9@its.jnj.com

Received August 19, 2011; revised December 16, 2011;
accepted December 16, 2011

Edited by Jane Clarke

The use of consensus design to produce stable proteins has been applied to numerous structures and classes of proteins. Here, we describe the engineering of novel FN3 domains from two different proteins, namely human fibronectin and human tenascin-C, as potential alternative scaffold biotherapeutics. The resulting FN3 domains were found to be robustly expressed in *Escherichia coli*, soluble and highly stable, with melting temperatures of 89 and 78°C, respectively. X-ray crystallography was used to confirm that the consensus approach led to a structure consistent with the FN3 design despite having only low-sequence identity to natural FN3 domains. The ability of the Tenascin consensus domain to withstand mutations in the loop regions connecting the β -strands was investigated using alanine scanning mutagenesis demonstrating the potential for randomization in these regions. Finally, rational design was used to produce point mutations that significantly increase the stability of one of the consensus domains. Together our data suggest that consensus FN3 domains have potential utility as alternative scaffold therapeutics.

Keywords: consensus/crystal structure/fibronectin/FN3/scaffold/stability

Introduction

The conformational stability of recombinant proteins is of fundamental importance in many biotechnology and pharmaceutical applications. A variety of methods have been used to assess conformational stability, with thermal stability as one relatively simple method that has been widely applied. It is generally accepted that for industrial enzymes, increased thermal stability reduces process costs by allowing for higher reaction rates, lower enzyme turnover and lower levels of microbial contamination (Eijsink *et al.*, 2004). In the biopharmaceutical realm, thermal stability can have consequences on the formulation, storage conditions and susceptibility to aggregation of such products (Chen *et al.*, 1994a,b;

Tsai *et al.*, 1998; Remmele *et al.*, 1999; Chi *et al.*, 2003; Abdul-Fattah *et al.*, 2007; Kuelzto *et al.*, 2008). Several protein engineering methods have been applied to increase the apparent conformational stability of proteins and enzymes, including rational design based on structural and sequence comparisons to homologs from thermophilic organisms (Haney *et al.*, 1999; Mueller *et al.*, 2000; Perl *et al.*, 2000), alteration of the surface charge of the protein (Strickler *et al.*, 2006; Gribenko *et al.*, 2009), directed evolution (Giver *et al.*, 1998; Miyazaki *et al.*, 2000; Miyazaki *et al.*, 2006; Kotzia and Labrou, 2009), composition of consensus sequences (Lehmann *et al.*, 2000a,b), design of stabilizing salt bridges or disulfides and other methods (Lehmann and Wyss, 2001). The consensus sequence approach to protein stabilization is based on the hypothesis that during natural and *in vitro* evolution processes, random mutations that destabilize the protein have a high probability of occurring but are generally functionally neutral as they do not cause the stability of the protein to fall below a level that renders it inactive. Random mutations that increase the protein stability are assumed to be much less probable due to a lack of positive selection for such increases in stability. Thus, amino acid positions in homologs that have a strongly conserved consensus residue are thought to contribute more to the stability than those positions without a clear consensus (Steipe *et al.*, 1994). This approach has been used to increase the stability of antibodies (Steipe *et al.*, 1994; Ohage and Steipe, 1999; Wirtz and Steipe, 1999; Knappik *et al.*, 2000; Visintin *et al.*, 2002; Demarest *et al.*, 2004), SH3 domains (Maxwell and Davidson, 1998), enzymes (Lehmann *et al.*, 2000a,b), DNA-binding proteins (Nikolova *et al.*, 1998), a chaperone protein (Wang *et al.*, 1999), fluorescent proteins (Dai *et al.*, 2007), leucine-rich repeat proteins (Stumpff *et al.*, 2003) and ankyrin repeat proteins (Mosavi *et al.*, 2002; Binz *et al.*, 2003). In many of these cases, optimal stability was achieved not simply by producing a consensus sequence, but by combining consensus calculations with rational design based on covariation of amino acids, protein structure considerations, previously defined stability data, computational methods or activity data.

Optimization of conformational stability is particularly useful for antibodies or alternative scaffold molecules (Skerra, 2000), in which portions of the protein surface are randomized and unique clones selected to bind to target molecules. Since the introduction of diversity into these randomized regions often leads to decreased stability of the resulting library members, it is ideal to start with the most stable antibody framework or scaffold sequence possible. This principle is demonstrated with the designed ankyrin repeat proteins or DARPinS. DARPinS are composed of a 33-residue consensus sequence of 229 ankyrin repeats, resulting in a scaffold with a melting temperature of >85°C (Binz

et al., 2003). The high stability of the scaffold afforded by consensus design enables the generation of libraries composed of variants with relatively high melting temperatures. Ribosome display selection can be used to select high-affinity binders with good stability properties.

Alternative scaffold proteins, which combine the affinity and specificity properties of antibodies with the biophysical properties of smaller molecules, are being widely evaluated as new protein therapeutics. These proteins, because of their small size, biophysical properties and modular domain structure, are attractive candidates for studies designed to increase conformational stability. From a biotechnology perspective, these properties may also make alternative scaffolds more amenable to novel routes of drug delivery and formulations that require high stability (Skerra, 2000; Binz et al., 2005).

One such alternative scaffold is derived from the immunoglobulin (Ig)-fold family. This fold is found in the variable regions of antibodies, as well as thousands of non-antibody proteins. It has been estimated that the Ig-fold is found in 2% of all animal proteins (Bork and Doolittle, 1992) and often functions in establishing protein–protein interactions, such as the interactions of FN3 domains with integrins and cytokine/cytokine receptor interactions. The prototype alternative scaffold from this family is the 10th fibronectin type III (FN3) repeat from human fibronectin (Koide et al., 1998). It has been demonstrated that this individual domain can tolerate a number of mutations in surface-exposed loops while retaining the overall Ig-fold structure. Thus, libraries of amino acid variants have been built into these loops and specific binders selected for a number of different targets (Koide et al., 1998; Karatan et al., 2004; Hackel et al., 2008). Such engineered FN3 domains have been found to bind to targets with high affinity; however, isolated binders display a wide range biophysical properties (Parker et al., 2005). We sought to create a stable, novel alternative scaffold protein with an Ig-fold based on a consensus sequence as such designs have in some cases led to proteins that exhibit physical properties distinct from natural domains.

The human fibronectin protein contains 14 additional independently folded FN3 domains, all of which are predicted to adopt a fold similar to that for the 10th FN3 domain based on sequence similarity. Indeed, structural conservation for several of these domains has been confirmed by the comparison of high-resolution structures of fibronectin FN3 domains 1, 7, 8, 9, 10, 12, 13 and 14 (Dickinson et al., 1994; Leahy et al., 1996; Sharma et al., 1999; Vakonakis et al., 2007). Despite adopting similar folds, these FN3 domains share pair-wise sequence identities of only 22–47% with each other. Likewise, tenascin-C is an extracellular matrix protein composed of 15 FN3 domains with similar sequence similarities to one another as found in fibronectin. We sought to determine if a consensus sequence of 15 FN3 domains from either fibronectin or tenascin-C can adopt an independently folded, stable structure. As a first step, we present the sequence and biophysical characteristics of novel proteins designed from consensus sequences of these FN3 repeats. The consensus sequence proteins designed here are found to be highly stable, highly expressed in *Escherichia coli*, soluble and amenable to randomization in exposed loops. Furthermore, rationally designed mutations were introduced that significantly increase the stability of one of the consensus domains. The biophysical properties and structure of

these novel domains make them attractive candidate as an alternative scaffold for therapeutic and diagnostic applications.

Results

A simple multiple sequence alignment of the 15 FN3 domains from human fibronectin did not produce a structurally relevant model, since the length of the FG loop varies between family members and the conservation in the G-strand is weak (Leahy et al., 1996). Therefore, the initial alignment produced was altered manually to take into account the different lengths of the FG loops of these domains (Fig. 1). The available high-resolution structures in the PDB database were used to define the boundaries of this loop in order to properly align the G-strand as well as to define the N-terminal and C-terminal sequences of the individual FN3 domains. The resulting alignment, presented in Fig. 1, is consistent with previous multiple sequence alignments of FN3 domains (Halaby et al., 1999). The FN3 domains of fibronectin have sequence identities to one another of 18–61% (Supplementary Fig. S4). A new protein construct, Fibcon, was then made using the most frequently found residue at each position of the alignment. This new domain is 44% identical on average to the naturally occurring FN3 domains in pair-wise comparisons. His-tagged Fibcon expresses in *E.coli* at >350 mg/l of culture and remains completely within the soluble fraction after cell lysis. Purification by nickel affinity chromatography and size exclusion chromatography (SEC) produced a homogenous sample that was monomeric and monodisperse in solution (Supplementary Fig. S1). The stability of Fibcon was assessed by differential scanning calorimetry (DSC) and guanidine hydrochloride (GdmCl) denaturation monitored by tryptophan fluorescence. The data presented in Fig. 2 show Fibcon to have a melting temperature of 89°C when fit to a two-state unfolding model. Thermal-induced unfolding was reversible as determined by calculating the ΔH values for an initial and repeated scan. Tryptophan fluorescence experiments also demonstrate a folded domain with a free energy of unfolding of -11.4 ± 1.5 kcal/mol (Fig. 2, Table II).

We sought to determine if a new consensus sequence derived from a sequence alignment of only the most stable FN3 domains would provide added stability over the original Fibcon design. In order to determine which individual FN3 domains from fibronectin to include in such an alignment, 14 of the 15 FN3 domains from human fibronectin were expressed and purified individually from *E.coli*. We were unable to subclone the seventh FN3 domain from a universal cDNA source, so it was omitted from this analysis. All 14 constructs tested were purified from the soluble fraction of *E.coli* and appear to be independently folded by DSC, indicating that each domain is stable in the context of an isolated FN3 domain. None of the naturally occurring fibronectin domains were found to be as stable or as highly overexpressed as the Fibcon protein. Melting temperatures for these domains ranged from 52.3°C for FnFN9 to 86.5°C for FnFN15 (Table I). The results obtained were in general agreement with previous studies describing high thermal and chemical stability for FnFN10 while comparatively much lower stability for FnFN9 (Plaxco et al., 1997; Koide et al., 1998; Cota and Clarke, 2000). Table I shows domains 1, 2, 5, 8, 10, 12, 14 and 15 all have melting temperatures >70°C.

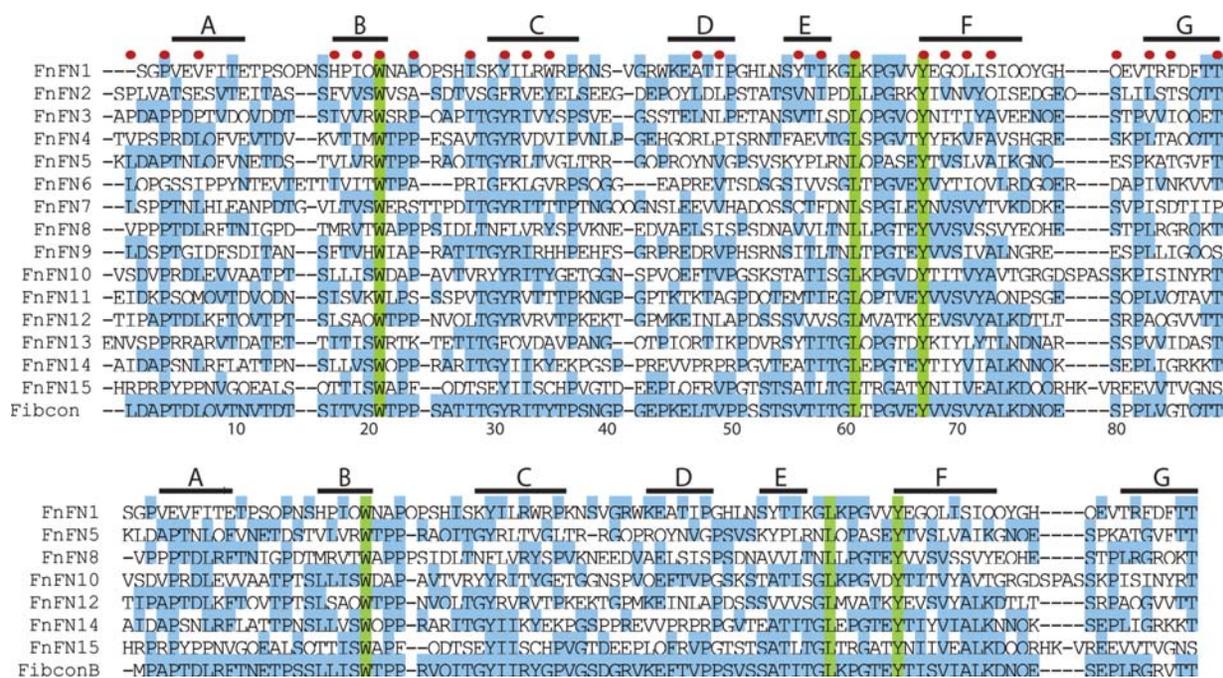


Fig. 1. Sequence alignment of Fibronectin FN3 domains. (A) Alignment of all FN3 domains used to produce Fibcon. (B) Alignment of most stable domains to produce FibconB. All alignments were produced with AlignX software. Green dots indicate positions of buried residues.

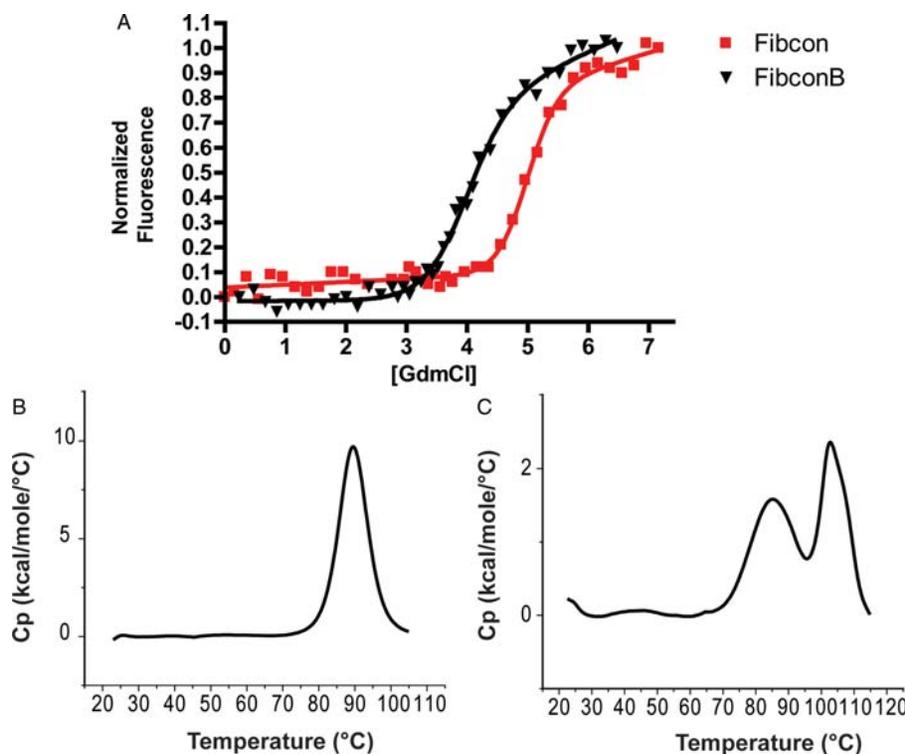


Fig. 2. Characterization of Fibcon and FibconB. (A) GdmCl denaturation monitored by tryptophan fluorescence. Fluorescence was observed by excitation at 280 nm and emission at 360 nm. Thermal denaturation measurements for Fibcon (B) and FibconB (C) by DSC. Samples were assayed at 1 mg/ml in PBS pH 7.4.

A multiple sequence alignment of domains 1, 5, 8, 10, 12, 14 and 15 was generated in order to design the alternate consensus sequence, FibconB (Fig. 1B). Domain 2 was excluded from this alignment as structural analysis has shown that the N-terminal 'A' strand is disordered in this domain (Vakonakis *et al.*, 2007). Similar to the larger set of FN3

domains (Fig. 1A), the most stable FN3 domains have only weak degrees of sequence similarity, 18–41% among pairs. FibconB was generated by selecting the most frequently used residue at each position of the alignment shown in Fig. 1B. FibconB is 40–64% similar to the individual domains aligned in Fig. 1B and 61% identical to Fibcon.

FibconB was expressed in *E.coli* at levels >120 mg/l in the soluble lysate fraction. The conformational stability of this domain was investigated by DSC and GdmCl denaturation monitored by tryptophan fluorescence. The DSC data obtained show a first peak at 82°C and a second peak at 105°C (Fig. 2). The transition at 105°C is most likely due to aggregation of FibconB at higher concentrations as the presence of this transition is concentration dependent (data not shown). GdmCl denaturation experiments are also consistent with a folded, stable domain that folds with a free energy of -6.7 kcal/mol. Finally, the oligomeric state of FibconB was investigated by size exclusion chromatography. FibconB elutes from a superdex 75 column at an elution volume consistent with that of a monomer (Supplementary Fig. S1).

We next investigated a consensus-driven design for other FN3 domain proteins, namely those from tenascin-C (Fig. 3). Far less structural information is available in the PDB regarding the tenascin FN3 domains compared with those of fibronectin. A multiple sequence alignment of the 15 tenascin FN3 domains was produced that predicts that TnFN4 has a slightly longer DE loop than the rest of the FN3 domains and that the FG loops are of approximately equal length (Fig. 3). The FN3 domains from tenascin have a wider range of sequence identity to one another (from 17 to 82% among pairs) than do the fibronectin FN3 domains (Supplementary Fig. S4). A tenascin consensus sequence, Tencon, was

generated by expressing a new protein comprised of the most common residue found at each position of the alignment in Fig. 3. Tencon has a sequence identity of 40% on average to the natural tenascin FN3 domains.

Tencon was expressed in *E.coli* and purified from the soluble fraction. Expression levels of >300 mg of soluble Tencon per liter of *E.coli* could be achieved. DSC experiments carried out in order to test the thermal stability Tencon show a single transition at 78°C in phosphate-buffered saline (PBS) (Fig. 4). Like Fibcon, Tencon thermal denaturation is completely reversible. GdmHCl-induced unfolding measurements monitored by tryptophan fluorescence resulted in a free energy of -10.6 ± 0.9 kcal/mol upon folding (Fig. 4, Table II), and SEC demonstrates that Tencon is monomeric in solution (Supplementary Fig. S1). The solubility of Tencon was measured by concentrating a stock solution in PBS to 130 mg/ml by ultrafiltration. Tencon remained monomeric after reaching this concentration as judged by the single peak eluting from the size exclusion column at a volume consistent with a 10 kDa protein (data not shown). No visible precipitates or aggregation were observed at any concentration.

In order to confirm that the protein domains derived from the consensus sequences are consistent with the FN3 fold, the crystal structures of Tencon and Fibcon were determined to 2.5 and 1.0 Å resolution, respectively (Table IV). Both consensus domains adopt the FN3 fold as designed, consisting of seven anti-parallel β -strands connected by surface-exposed loops (Fig. 5). Structural alignments with TnFN3 (1TEN) (Leahy et al., 1992) demonstrate the Tencon backbone structure is almost identical to this well-characterized FN3 domain, with an average root mean square deviation (RMSD) value of 1.4 Å for backbone C α atoms (Fig. 5). Similarly, Fibcon adopts a backbone confirmation with an average RMSD of 1.5 Å from the Fn10FN3 (PDB 1FNA). Tencon and Fibcon structures deviate only slightly from one another, with backbone RMSDs of only 1.6 Å. We also observed an interesting structural feature of Fibcon. Two neighboring β -strands, residues 10–12 of strand A and Ile18 and val20 of strand B, adopt alternative backbone conformations. Voronoia (Rother et al., 2003) was used to calculate average core packing densities of 0.830 (Fibcon) and 0.720 (Tencon), results that are similar to those of natural FN3 domains FnFN10 (0.724) and TnFN3 (0.715).

Table I. Melting temperatures of Fibronectin FN3 domains

Construct	T_m (°C)
FnFN1	81.2
FnFN2	70.7
FnFN3	66.1
FnFN4	52.3
FnFN5	72.3
FnFN6	68.3
FnFN7	N.D.
FnFN8	72.6
FnFN9	47.0
FnFN10	82.5
FnFN11	53.9
FnFN12	76.1
FnFN13	65.8
FnFN14	70.3
FnFN15	86.5

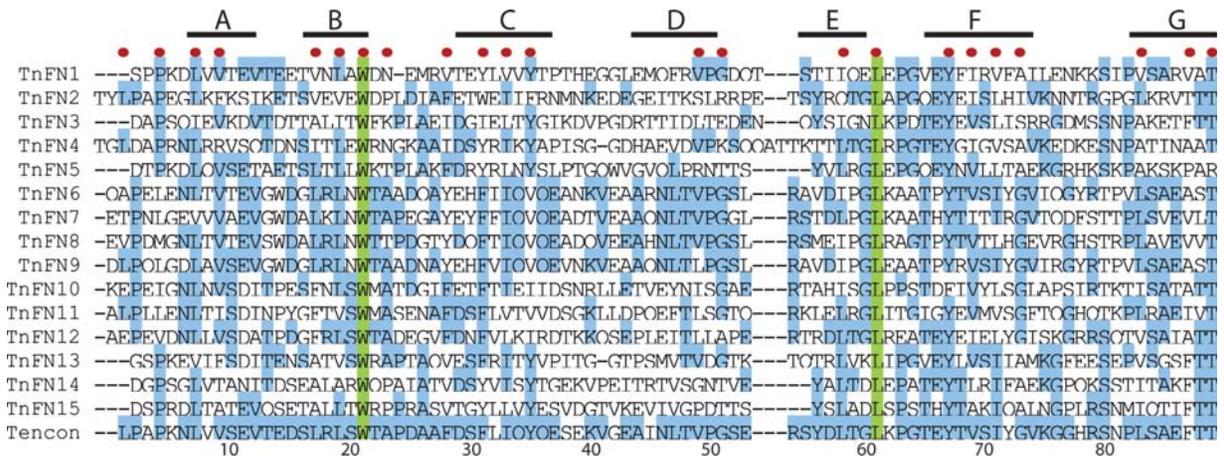


Fig. 3. Sequence alignment of Tenascin FN3 domains. Green dots indicate positions of buried residues.

Based on the biophysical properties of the consensus domains (expression level, conformational stability, reversible folding/unfolding, solubility and availability of a high-resolution structure) we sought to determine if these

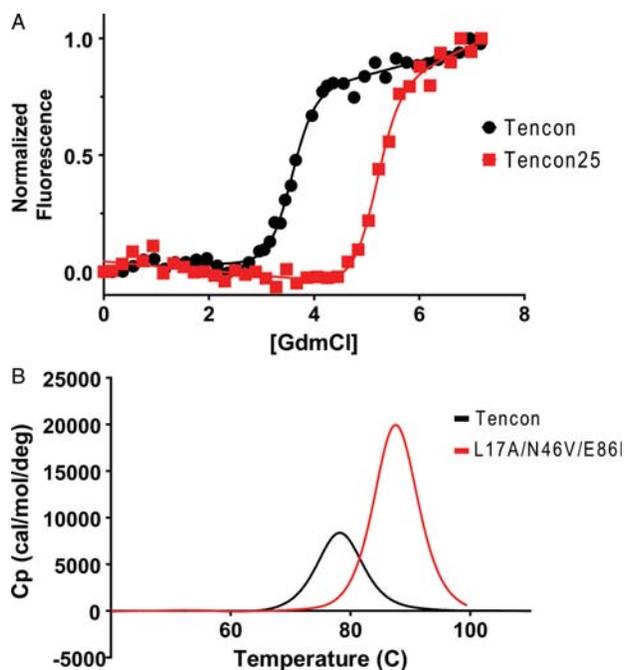


Fig. 4. Characterization of Fibcon and FibconB. Data for Tencon are shown in black and Tencon 25 (L17A/N46V/E86I) in red. (A) GdmCl denaturation monitored by tryptophan fluorescence. Fluorescence was measured by excitation at 280 nm and emission at 360 nm. (B) Thermal denaturation measurements by DSC. Samples were assayed at 1 mg/ml in PBS pH 7.4.

Table II. Summary of stability measurements for consensus domains. Errors associated with $[D]_{50\%}$ are based on two or three replicates

Construct	T_m	$[D]_{50\%}$	m	ΔG^{H_2O} (kcal/mol)
Fibcon	89.6	5.0 ± 0.01	2.29 ± 0.30	11.4 ± 1.5
FibconB	85.3/104.0	3.86 ± 0.09	1.72 ± 0.02	6.7 ± 0.2
Tencon	78.0	3.50 ± 0.07	3.03 ± 0.24	10.6 ± 0.9

Only a single replicate was measured for FibconB.

proteins could tolerate randomization of the loop residues. For the rest of the experiments described, we focused only on the Tencon domain. Alanine scanning mutagenesis was used to individually mutate each Tencon loop position, as well as positions just before or after the loops, to determine which residues are tolerant of mutation. Loop positions already occupied by alanine were mutated to serine. The comparative stability of each mutant was then estimated by GdmCl denaturation curves generated by high-throughput analysis using a plate reader (Fig. 6). Free energies of unfolding were not calculated for each mutant as m values obtained were not accurate when generated using this high-throughput method (Pace, 1986). Figure 6 shows that the majority of loop residues were tolerant of mutation. Loop positions most destabilized by mutation include F28, R54, L61 and S80. A number of residues in the framework (outside of loop regions depicted in Fig. 6) were also destabilized, including I45, V49, L58, T65 and Y67. Several mutants significantly increased stability including L17A, E53A and L83A. Mutagenesis of residue W22 was not investigated by this method as this tryptophan residue is responsible for the fluorescence signal that is monitored.

Although the consensus sequence approach used to generate Tencon produced a stable protein with excellent biophysical properties, we sought to determine if the conformational stability of this molecule could be further enhanced. Five specific mutations, N46V, E14P, E11N, E37P and G73Y, were predicted to improve conformational stability significantly using the program PoPMuSiC v2.0 (Dehouck *et al.*, 2009). In addition, position E86 from the homologous TnFN3 protein was previously found to increase the stability for this domain when mutated to alanine (Hamill *et al.*, 2000b). The E86I mutation in TnFN3 produced an even larger increase in conformational stability, leading to a 11.9°C increase in the melting temperature of TnFN3 (Jacobs, unpublished results). Thus, this E86I mutation was introduced into Tencon. The stability of each mutant described was measured by DSC and GdmCl denaturation (Supplementary Fig. S2, Table III). The N46V, E14P and E86I mutations significantly increased the stability of Tencon as measured in both assays. Mutants E11N and E37P had little effect on stability and mutant G73Y was significantly destabilizing. Two combinatorial mutants were made to examine if the stabilizing effects of the mutations described

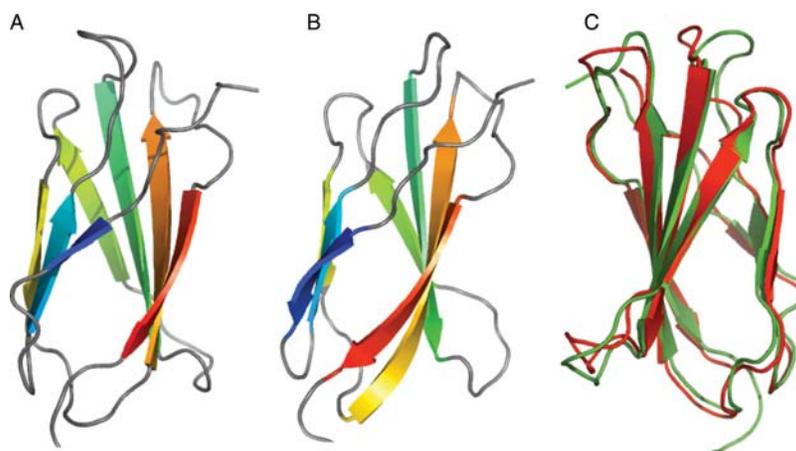


Fig. 5. Ribbon diagrams representing the crystal structures of Tencon (A) and Fibcon (B). A structural alignment of Tencon and TnFN3 (ITEN) is shown in panel C. All pictures were made with PyMol.

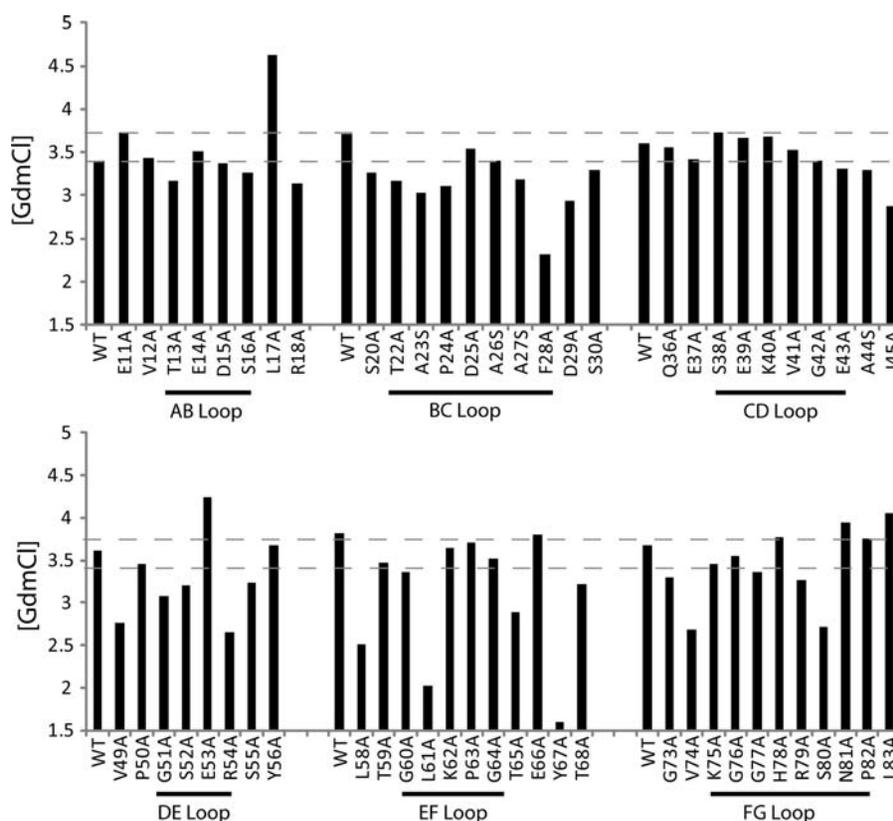


Fig. 6. Alanine scanning of Tencon loop regions. The midpoint concentration of GdmCl denaturation is presented for a number of mutants. Lines under the charts indicate the positions of the surface-exposed loops. A control sample of wild-type Tencon (WT) was run on each plate analyzed to serve as an internal standard. Dashed lines indicate the range of midpoint values obtained for WT Tencon across six plates.

Table III. Summary of stability measurements for Tencon mutants

Mutations	T_m	$[D]_{50\%}$	$\Delta\Delta G_{D-N}$ (kcal/mol)
WT	78.0	3.50 ± 0.07	
N46V	81.9	3.73 ± 0.13	-0.7 ± 1.8
E14P	82.8	3.69 ± 0.10	-0.6 ± 1.8
E11N	79.0	3.62 ± 0.18	-0.4 ± 1.8
E37P	77.4	3.40 ± 0.10	0.2 ± 1.7
G73Y	67.6	2.40 ± 0.01	3.2 ± 1.3
E86I	82.8	3.90 ± 0.08	-1.0 ± 1.8
N46V/E86I	86.7	4.13 ± 0.03	-1.8 ± 1.9
E14P/N46V/E86I	87.5	4.04 ± 0.01	-1.6 ± 1.9
L17A/N46V/E86I	92.7	5.17 ± 0.02	-4.9 ± 2.3

A single m value of 2.96 ± 0.41 was calculated by averaging the m value obtained for each mutant

above are additive. Mutations N46V and E86I are additive, increasing the T_m and ΔG calculated by GdmCl unfolding above that of either mutant alone (Table III). Addition of E14P to this background had little additional effect. Finally, a combination of N46V and E86I with L17A, discovered by alanine scanning experiments described above, led to the most stable Tencon construct with a melting temperature of 92.7°C (Fig. 4, Table III).

Discussion

A common approach for producing stable proteins is the generation of consensus sequences based on sequence and structural alignments. Here, we describe novel FN3 proteins

generated by determining consensus sequences from the FN3 domains of human fibronectin or human tenascin-C, with the ultimate goal of producing a FN3 protein with optimal biophysical properties to be used as an alternative scaffold template. The consensus sequence approach used here differs from that of previous studies in that the protein sequences used to form the consensus sequences are of modest sequence identity to one another. The immunoglobulin-like fold is perhaps the most common fold found in the proteome (Bork and Doolittle, 1992). A hallmark of this family is the relatively low-sequence similarity between members that, nevertheless, adopt the same overall fold. In fact, in many cases members of this family have no discernible sequence similarity to one another (Bork *et al.*, 1994; Smith and Xue, 1997). A subset of the Ig-fold is the FN3 domain. As this domain is small, compact and often expressed at high levels in *E. coli*, protein domains from this family have been used as model systems for studying the mechanisms of protein folding (Plaxco *et al.*, 1996; Clarke *et al.*, 1997). Results from these studies have led to the proposal that individual FN3 domains fold by a similar mechanism driven mainly by a common core nucleus of only a few residues (Hamill *et al.*, 2000a,b; Cota *et al.*, 2001). The consensus sequences generated in this study share only modest sequence identity to the domains that they were generated from and to one another (Supplementary Fig. S3). Despite only 45% sequence identity between Fibcon and Tencon, the backbone structures of these molecules are very similar. This may be explained by the conservation of a number of hydrophobic core residues, including those of the B, C, E and F strands which have previously been shown to be involved in FN3 folding

(Hamill *et al.*, 2000a). The property of FN3 domains that allows them to fold around a core of limited residues may explain the ability of these consensus domains to form highly stable, folded structures in the absence of strong sequence identity to naturally occurring FN3 domains.

The success of consensus engineering is dependent upon the sequences used to produce the design. The large number of FN3 (and Ig-like) domains represented in the proteome allows for almost infinite combinations of sequences that could be explored. We chose to use only subsets of FN3 domains from a single protein, human fibronectin or human tenascin-C. This path was chosen based on the idea that some of the domains from these proteins may have evolved in a concerted fashion (Odermatt *et al.*, 1985; Sipos *et al.*, 2008). As a group however, the FN3 domains from tenascin-C are almost as similar to the FN3 domains from tenascin-X and fibronectin as they are to each other. Thus, a larger set of sequences including the FN3 domains from tenascin-c, tenascin-X and fibronectin could be used to generate a different consensus sequence. Another approach is to use a guide protein as a template for positions in which no clear consensus exists (Dai *et al.*, 2007). The limited sequence similarity of FN3 domains may allow this approach to be fruitful for engineering stable FN3 domains by using one of the intrinsically very stable, naturally occurring domains (FnFN10 for example) as the guide protein. Interestingly, two consensus domains designed from the fibronectin FN3 domains exhibited different thermal stabilities. Fibcon, formed from a consensus of all 15 FN3 domains proved to be more stable than FibconB, which was derived from only the most stable 6 FN3 domains from fibronectin. Since the FN3 domains used to calculate the FibconB sequence share only 18–41% sequence identity, there may simply be too few sequences to form a true consensus. Although the FibconB domain is indeed a stable protein in its own right (T_m 82°C), these data suggest that when designing consensus sequences, a larger data set may produce more stable consensus designs than a smaller data set consisting of more stable molecules. Another explanation could be that optimal consensus design results from starting with less stable sequences as opposed to the more stable members.

Alanine scanning was used to demonstrate that the loop regions of Tencon are amenable to mutagenesis. These results are consistent with previous studies that have shown that the FnFN10 domain loops are tolerant to mutations, including the elongation of these loops, and that the EF loop is most sensitive to these changes (Batori *et al.*, 2002). Only a few of the mutants in this series significantly lower Tencon stability. Mutation of F28 (BC loop) and L61 (EF loop), which are buried in the core of Tencon forming several hydrophobic contacts, results in significant destabilization. Mutation of these residues to alanine most likely results in a loss of stability due to a loss of these contacts. S80, which is found at the C-terminal end of the FG loop, forms a hydrogen bond to the backbone of the F-strand, while R54 appears to form a salt bridge with the backbone of the BC loop. These results suggest that these residues should be fixed as is when designing libraries of variable amino acids in the Tencon loops. As expected, mutation of a number of framework positions resulted in decreased stability. Many of these residues, such as Y67, form hydrophobic contacts in the core of the structure and are thus not amenable to alanine

substitution. It may be possible, however, to retain the stability of the fold with more conservative substitutions.

Although we have demonstrated that Tencon has relatively high thermal stability, we felt that this stability was suboptimal when compared with other naturally occurring FN3 domains. With an original melting temperature of 78°C and ΔG of unfolding in GdmCl of -10.6 kcal/mol, we were able to introduce three mutations that significantly increased the stability of this domain. One of these mutations, N46V, was predicted by PoPMuSiC v2.0 (Dehouck *et al.*, 2009). It is unclear as to how this mutation contributes to such a large stabilization as this residue is exposed on the surface as part of the 'D' strand. One plausible explanation may be that the branched amino acid valine is energetically more favorable in the β -sheet, while asparagine has been shown to be unfavorable in the middle of β -strands (Janin and Wodak, 1978; Munoz and Serrano, 1994; Farzadfard *et al.*, 2008). A similar explanation is most likely for the stabilization generated by the E86I mutation. Alternatively, E86 is positioned in the middle of a cluster of polar residues with negative potentials: S86, S70, T68, S38, E66 and T88. Reduction of negative–negative repulsion with this patch may be the cause of increased stability of E86I. The L17A mutation, which had the largest positive effect on conformational stability for a single mutation, is most likely beneficial due to more efficient packing of the hydrophobic core. Comparison of the Tencon crystal structure with that of TnFN3 (PDB 1TEN) reveals that L17 displaces F87 (conserved in TnFN3) from the core of the protein. Mutation to alanine (also conserved in TnFN3) allows F87 to pack efficiently (Fig. 7). Inspection of the sequence alignment presented in Fig. 3 reveals that four of the TnFN3 domains are represented by alanine at position 17 (TnFN3, TnFN13, TnFN14 and TnFN15). In all of these examples, residue 87 is a phenylalanine. Conversely, none of the TnFN3 sequences having a leucine at position 17 have a phenylalanine at position 87. This provides evidence that incorporation of covariance of residues into the design of FN3 consensus sequences could further enhance the stability of the designed sequences.

In the work described here, three separate approaches were combined to improve the conformational stability of FN3 domains: consensus design, alanine scanning mutagenesis and

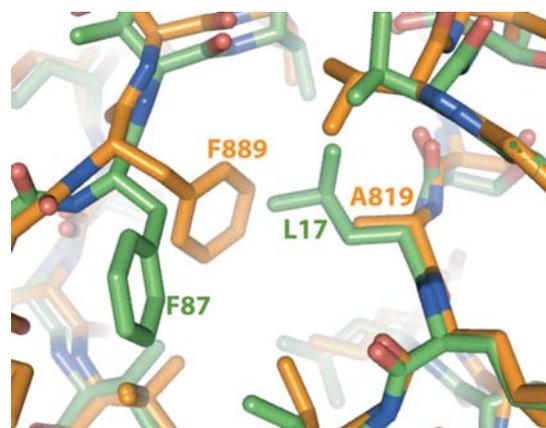


Fig. 7. Structural analysis of the increase in stability of Tencon by the L17A mutation. The structure of Tencon is shown in green and that of TnFn3 (1TEN) in yellow. Amino acids labeled are numbered according to the PDB files.

modeling via PoPMuSiC. It is difficult to compare the effectiveness of the consensus design approach to that of the others, as the point of reference for comparison is unclear due to the widely different stabilities exhibited by the FN3 domains used to calculate the consensus. For example, the Fibcon consensus protein is drastically more stable than the least stable domain, FnFN9, yet only marginally more stable than FnFN15, which has a melting temperature of 86.5°C. It is also unknown as to how stable Tencon is in comparison with the naturally occurring domains from which it was designed. In one sense, the data presented here suggest that consensus design is an efficient method of protein stabilization as two highly stable domains, Tencon and Fibcon, could be produced by designing only a single construct each whereas more constructs might need to be screened to find highly stable naturally occurring domains.

For stabilizing the Tencon molecule, alanine scanning mutagenesis produced the mutation (L17A) with the largest increase in conformational stability. Interestingly, this mutant was predicted to be significantly destabilizing by PoPMuSiC. The predictions made by PoPMuSiC agree in most other cases with the alanine scanning data, that is F28A, V49A, R54A, L58A, L61A, Y67A and S80A were all predicted to be destabilizing and N81A was predicted to be slightly stabilizing. On its own, PoPMuSiC was able to correctly predict stabilizing changes for three of the five mutants tested indicating that this is indeed a valuable tool. As only a small set of the beneficial mutations predicted by PoPMuSiC were investigated in this study and a much larger number of mutants were generated using alanine scanning, it is impossible to fairly compare the efficiency of these two methods for such engineering experiments at this time.

By improving the overall stability of Tencon with these mutations, it is expected that variants with overall higher conformational stability will be produced once the loop regions are randomized in order to select specific binders to target proteins. The resulting stability of the mutant Tencon domain is comparable with that of other engineered FN3 domains. Hu and colleagues used computational methods to redesign the third FN3 domain from human tenascin-C (Hu *et al.*, 2008). The result was a dramatic improvement in the melting temperature (>90°C) and subsequent increase in tolerance to GdmCl. The resulting constructs, however, did not retain the property of reversible thermal unfolding at pH 7.0 demonstrated by the parent molecule, presumably due to the change in pI of the new proteins, which the authors suggest results in a higher net charge that stabilizes the unfolded state. Consistent with this hypothesis, the redesigned FN3 domain was found to be reversible at pH 3.0. Dutta *et al.* (2005) used fragment complementation in yeast to screen a library of FnFN10 domains to isolate a mutant stabilized by -2 kcal/mol. An alternative rational design approach centered on the removal of unfavorable electrostatic interactions has also been used to significantly stabilize this particular domain (Koide *et al.*, 2001). The results from such studies suggest that rational design combined with higher-throughput screening methods may yield FN3 domains with even greater stability.

The overall goal of this work was to develop a template molecule for the generation of a new alternative scaffold platform. We chose to focus on Tencon due to the ideal biophysical properties of this protein. We have found Tencon to be expressed at high levels in *E. coli* (~300 mg/l in shake flask), monomeric, even at concentrations >12 mM and

tolerant of mutations in the CDR-like loops. Another interesting property of Tencon (and Fibcon) is the complete reversibility of this domain following heat denaturation. From a biotechnology point of view, this property may provide advantages to the design of large-scale purification processes and formulation strategies. Furthermore, we have found this domain to be stable in human serum, stable during room-temperature storage, resistant to a wide range of solution pH values, and to have low-predicted immunogenicity (Jacobs, unpublished results). Initial results suggest it is possible to create libraries in the surface-exposed loops of the Tencon scaffold and to select target-specific binders that maintain the overall FN3 architecture. Future communications will describe the design and generation of such libraries as well as the selection of Tencon variants that bind to specific targets.

Materials and methods

Sequence alignments

All multiple sequence alignments were performed using AlignX software (Invitrogen), which is based on a modified ClustalW algorithm (Thompson *et al.*, 1994).

Expression and purification

Genes encoding Fibcon, FibconB and Tencon were chemically synthesized and sub-cloned into a pET15 vector (Novagen) modified to include a ligase-independent cloning site. The resulting plasmids were transformed into BL21-GOLD(DE3) *E. coli* (Stratagene) for expression. A single colony was picked and grown overnight at 37°C in 2 ml of Terrific Broth containing 100 µg/ml ampicillin. This culture was used to seed 100 ml of autoinduction media (Overnight Express Instant TB media, Novagen) in a 500 ml baffled flask and grown at 37°C for 16 h.

The culture was harvested by centrifugation at $4000 \times g$ for 20 min and the pelleted cells resuspended 5 ml of BugBuster HT (Novagen) per gram of wet cell pellet. After 30 min of incubation at room temperature, lysates were clarified by centrifugation at $30\,000 \times g$ for 20 min and loaded onto a 3-ml Ni-NTA superflow column (Novagen) by gravity. After loading, each column was washed with 15 ml of a buffer containing 50 mM sodium phosphate pH 7.4, 500 mM NaCl and 10 mM imidazole. Bound protein was then eluted from the column using 10 ml of a buffer containing 50 mM sodium phosphate pH 7.4, 500 mM NaCl and 250 mM imidazole. Protein purity was assessed by sodium dodecyl sulfate polyacrylamide gel electrophoresis. Prior to biophysical analysis, each mutant was dialyzed thoroughly into PBS pH 7.4. 28–33 mg of purified protein was obtained for each sample from 100 ml of culture. For crystallization, each sample was further purified using a Superdex 75 16/60 column (GE Healthcare) in PBS followed by dialysis into 10 mM Tris pH 7.5, 50 mM NaCl.

Characterization of thermal stability

Thermal stability was measured by capillary DSC. Each sample was dialyzed extensively against PBS pH 7.4 and diluted to a concentration of 2–3 mg/ml. Melting temperatures were measured for these samples using a VP-DSC instrument equipped with an autosampler (MicroCal, LLC). Samples were heated from 10 to 95°C or 100°C at a rate of 1°C per minute. A buffer only scan was completed between each sample scan in order to calculate a baseline for

integration. Data were fit to a two-state unfolding model following subtraction of the buffer only signal. Reversibility of thermal denaturation was determined by repeating the scan for each sample without removing it from the cell. Reversibility was calculated by comparing the area under the curve from the first scan with the second scan.

Denaturation by Guandine Hydrochloride

For denaturation measurements, 200 μl of a solution containing 50 mM sodium phosphate pH 7.0, 150 mM NaCl, and variable concentrations of GdmCl from 0.48 to 6.63 M were pipetted into black, non-binding, 96-well plates (Greiner) in order to produce a 17-point titration. Tencon mutants (10 μl) were added to each well across the plate to make a final protein concentration of 12 μM and mixed by pipetting up and down gently. After incubation at room temperature for 24 h, fluorescence was read using a Spectramax M5 plate reader (Molecular Devices) with excitation at 280 nm and emission at 360 nm. Fluorescence signal was normalized using the equation (Pace, 1986):

$$f_u = (y_F - y)/(y_F - y_u)$$

where y_F and y_u are the fluorescence signals of the folded and the unfolded samples, respectively.

The mid-points of the unfolding transition and slope of the transition were determined by fitting to the equation below (Clarke *et al.*, 1997):

$$F = \frac{(\alpha_N + \beta_N[D]) + (\alpha_D + \beta_D[D]) \exp(m([D] - [D]_{50\%})/RT)}{1 + \exp(m([D] - [D]_{50\%})/RT)}$$

where F is the fluorescence at the given denaturant concentration, α_N and α_D are the y-intercepts of the native and denatured state, β_N and β_D are the slopes of the baselines for the native and denatured state, $[D]$ is the concentration of GdmCl, $[D]_{50\%}$ the GdmCl concentration at which point 50% of the sample is denatured, m the slope of the transition, R the gas constant, and T the temperature. The free energy of folding for each sample was estimated using the equation (Pace 1986; Clarke *et al.*, 1997):

$$\Delta G = m[D]_{50\%}$$

It is often difficult to accurately measure the slope of the transition, m , for such curves. Since the Tencon mutations described here are not expected to alter the folding mechanism of Tencon, the m value for each mutant was measured and the values averaged (Pace, 1986) to produce an $m = 2.96$ kcal/mol/M used for all free energy calculations.

Analytical SEC

SEC was used to assess the aggregation state of the consensus domains. of each sample (10 μl) was injected onto a Superdex 75 5/150 column (GE Healthcare) at a flow rate of 0.3 ml/min with a PBS mobile phase. Elution from the column was monitored by absorbance at 280 nm. In order to assess the aggregation state, the column was previously calibrated with globular molecular weight standards (Sigma).

Crystallization

Fibcon. Fibcon protein (50 mM Tris buffer pH 7.5, 50 mM NaCl) was concentrated to 20 mg/ml. Automated crystallization screening was performed using the Oryx4 automatic protein crystallization robot (Douglas Instruments) using Corning plate 3550 (Corning Incorporated). Initial screening was performed with In-house screen 1 & 2 (96 \times 2 crystallization conditions), Hampton Research HT screen (Hampton Research, Aliso Viejo, CA, USA) and The Qiagen poly(ethylene glycol) (PEG) screen. Micro seeds were prepared as above from the best crystallization hits (Acet 4.5, 34% PEG 8K). Microseed matrix screening (MMS) was performed using the Oryx4 robot. Crystals were obtained from 0.1 M sodium acetate pH 4.5, 23% PEG 8000 and 5% PEG 400.

Tencon. Tencon L33M variant was concentrated using Amicon-Ultra (3 kDa) device to 32 mg/ml in 20 mM Tris pH 7.5, 50 mM NaCl. The L33M mutant was used for crystallization studies in order to produce a selenomethionine modified version as initial efforts to solve the structure by molecular replacement methods were unsuccessful. The crystals from this mutant however were able to be solved by molecular replacement (described below) and thus a selenomet derivative was never produced. Initial crystallization was carried out by the vapor-diffusion method at 20°C using an Oryx4 robot (Douglas Instruments, East Garston, UK). The experiments were composed of equal volumes of protein and reservoir solution in a sitting drop format in 96-well Corning 3550 plates. The initial screening was performed with the Hampton HT (Hampton Research) crystallization screen. Microcrystals obtained from 12% PEG 4000 in 0.1 M Na citrate buffer pH 3.5 were harvested in 100 μL of reservoir solution, homogenized by vortexing for 3 min with a Teflon Seed Bead (Hampton Research) and stored at -20°C . The MMS was set up using the hanging-drop vapor-diffusion method in 24-well VDX-greased plates (Hampton Research). In each crystallization drop, 0.6 μL screening (reservoir) solution and 0.2 μL microseeds were added to 0.8 μL of 16 mg/ml protein solution. Diffraction quality crystals were obtained from 6% PEG 4000 in 0.1 M Na Citrate buffer pH 3.5.

X-ray diffraction data collection, structure determination and refinement

Fibcon crystals were soaked in 0.1 M sodium acetate pH 4.5, 20% PEG 8000 and 15% PEG 400 and flash frozen in liquid nitrogen. Diffraction data were collected at APS IMCA-CAT beamline ID-17BM. Diffraction data to 1 Å resolution were collected and processed with XDS (Kabsch, 2010). The crystals belong to space group $P2_12_12_1$ with cell dimensions $a = 28.39$, $b = 40.07$ and $c = 61.52$ Å. The Fibcon structure was solved by molecular replacement with Phaser using homology models based upon the PDB code 1FNF (Leahy *et al.*, 1996). Structure refinement was carried out with PHENIX (Adams *et al.*, 2002) and model adjustment in Coot (Emsley and Cowtan, 2004). In the final round, all atoms were refined anisotropically. Occupancies were also refined for all solvent molecules, several sidechains and the C-terminal His-tag residues. The final $R_{\text{cryst}}/R_{\text{free}} = 11.7/14.7\%$ for reflections to

Table IV. Crystallographic data and refinement statistics

	Tencon	Fibcon
Data collection		
Wavelength (Å)	1.5418	1.000
Space group	P2 ₁	P2 ₁ 2 ₁ 2 ₁
Unit cell axes (Å), angles (°)	48.72, 87.58, 56.22 90, 103.92, 90	28.39, 40.07, 61.52 90, 90, 90
Mol./ASU	4	1
Resolution (Å) ^a	30–2.5 (2.6–2.5)	50.00–1.00 (1.02–1.00)
Number of measured reflections	64 699 (5799)	217 081 (10 426)
Number of unique reflections	15 396 (1552)	37 669 (2528)
Completeness (%)	96.6 (97.7)	97.9 (90.3)
Redundancy ^a	4.2 (3.7)	5.76 (4.12)
R-merge ^a	0.078 (0.110)	0.047 (0.158)
<I/σI> (unaveraged)	10.5 (5.8)	24.3 (7.9)
B-factor (Wilson) (Å ²)	22.4	8.5
Structure refinement		
Resolution (Å)	26.50–2.5(2.86–2.50)	33.57–1.00 (1.03–1.00)
Number of reflections in refinement	15 313 (4 998)	37 663 (2495)
Number of atoms	2960	953
Number of solvent molecules	119	144
R _{cryst} (%)	24.2 (31.1)	11.7 (12.4)
R _{free} (5% data) (%)	29.7 (35.7)	14.7 (15.6)
RMSD bond lengths (Å)	0.003	0.022
RMSD bond angles (°)	0.79	2.07
RMSD B-factor main-chain (Å ²)	7.2	4.0
Mean B factors (Å ²)		
Proteins	39.7	14.9
Ramachdran plot ^b		
Favored (%)	97.0	96.9
Outliers (%)	0.0	0.0

^aValues for highest-resolution shell are in brackets.

^bThe Ramachdran plot was calculated with MolProbity (Davis et al., 2004, 2007).

1 Å resolution. The data collection and refinement statistics are given in Table IV.

Tencon crystals were soaked in 0.1 M sodium citrate, pH 3.5, 10% PEG 4000 with 25% glycerol and frozen in the liquid nitrogen stream at 95 K. The X-ray diffractions data to 1.8 Å were collected using a Rigaku MicroMax-007HF microfocus X-ray generator equipped with a Saturn 944 CCD detector and an X-stream 2000 cryocooling system. The Tencon structure was solved by MR with Phaser using a homology model based upon 1QR4 (Bisig et al., 1999). Structure refinement and model adjustment were done with PHENIX (Adams et al., 2002) and Coot (Emsley and Cowtan, 2004), respectively. During structure refinement, non-crystallographic symmetry (NCS) restraints for the four copies of tencon were initially imposed. A number of side-chains and parts of the backbone deviated from each other among the four molecules. In the final rounds, refinement strategies with (excluding deviating residues) and without the NCS restraints were tested and relaxing NCS straints led to lower R_{free} values. Thus, the final refinement was done without imposing NCS restraints. The final R_{cryst}/R_{free} = 24.2/29.7 for 15 313 reflections to 2.5 Å (Table IV). Structure quality assessments were done with MolProbity (Davis et al., 2004, 2007).

Coordinates were deposited in the RCSB Protein Data Bank having been assigned PDB ID codes 3TES (Tencon) and 3TEU (Fibcon).

Supplementary data

Supplementary data are available at PEDS online.

Acknowledgements

We thank Derrick Domingo, Ross Fellows, Ingeborg Feil and Winnie Chan for help with protein purification, Mimi Zhou and Ellen Chi for DNA synthesis, Brandy Strake for subcloning of fibronectin FN3 domains, Fang Yi for discussion of stability measurements, and the members of Centocor Discovery Research and the Centyrex Venture for helpful discussions and experimental support.

References

- Abdul-Fattah,A.M., Kalonia,D.S. and Pikal,M.J. (2007) *J. Pharm. Sci.*, **96**, 1886–1916.
- Adams,P.D., Grosse-Kunstleve,R.W., Hung,L.W., et al. (2002) *Acta Crystallogr. D Biol. Crystallogr.*, **58**, 1948–1954.
- Batori,V., Koide,A. and Koide,S. (2002) *Protein Eng.*, **15**, 1015–1020.
- Binz,H.K., Amstutz,P. and Pluckthun,A. (2005) *Nat. Biotechnol.*, **23**, 1257–1268.
- Binz,H.K., Stumpp,M.T., Forrer,P., Amstutz,P. and Pluckthun,A. (2003) *J. Mol. Biol.*, **332**, 489–503.
- Bisig,B., Weber,P., Vaughan,L., Winterhalter,K.H. and Piontek,K. (1999) *Acta Crystallogr. D Biol. Crystallogr.*, **55**, 1069–1073.
- Bork,P. and Doolittle,R.F. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 8990–8994.
- Bork,P., Holm,L. and Sander,C. (1994) *J. Mol. Biol.*, **242**, 309–320.
- Chen,B.L., Arakawa,T., Hsu,E., Narhi,L.O., Tressel,T.J. and Chien,S.L. (1994a) *J. Pharm. Sci.*, **83**, 1657–1661.
- Chen,B.L., Arakawa,T., Morris,C.F., Kenney,W.C., Wells,C.M. and Pitt,C.G. (1994b) *Pharm. Res.*, **11**, 1581–1587.
- Chi,E.Y., Krishnan,S., Randolph,T.W. and Carpenter,J.F. (2003) *Pharm. Res.*, **20**, 1325–1336.
- Clarke,J., Hamill,S.J. and Johnson,C.M. (1997) *J. Mol. Biol.*, **270**, 771–778.
- Cota,E. and Clarke,J. (2000) *Protein Sci.*, **9**, 112–120.
- Cota,E., Steward,A., Fowler,S.B. and Clarke,J. (2001) *J. Mol. Biol.*, **305**, 1185–1194.
- Dai,M., Fisher,H.E., Temirov,J., Kiss,C., Phipps,M.E., Pavlik,P., Werner,J.H. and Bradbury,A.R. (2007) *Protein Eng. Des. Sel.*, **20**, 69–79.
- Davis,I.W., Leaver-Fay,A., Chen,V.B., et al. (2007) *Nucleic Acids Res.*, **35**, W375–383.
- Davis,I.W., Murray,L.W., Richardson,J.S. and Richardson,D.C. (2004) *Nucleic Acids Res.*, **32**, W615–619.
- Dehouck,Y., Grosfils,A., Folch,B., Gilis,D., Bogaerts,P. and Rooman,M. (2009) *Bioinformatics*, **25**, 2537–2543.
- Demarest,S.J., Rogers,J. and Hansen,G. (2004) *J. Mol. Biol.*, **335**, 41–48.
- Dickinson,C.D., Veerapandian,B., Dai,X.P., Hamlin,R.C., Xuong,N.H., Ruoslahti,E. and Ely,K.R. (1994) *J. Mol. Biol.*, **236**, 1079–1092.
- Dutta,S., Batori,V., Koide,A. and Koide,S. (2005) *Protein Sci.*, **14**, 2838–2848.
- Eijssink,V.G., Bjork,A., Gaseidnes,S., Sirevag,R., Synstad,B., van den Burg,B. and Vriend,G. (2004) *J. Biotechnol.*, **113**, 105–120.
- Emsley,P. and Cowtan,K. (2004) *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
- Farzadfard,F., Gharaei,N., Pezeshk,H. and Marashi,S.A. (2008) *J. Struct. Biol.*, **161**, 101–110.
- Giver,L., Gershenson,A., Freskgard,P.O. and Arnold,F.H. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 12809–12813.
- Gribenko,A.V., Patel,M.M., Liu,J., McCallum,S.A., Wang,C. and Makhatadze,G.I. (2009) *Proc. Natl Acad. Sci. USA*, **106**, 2601–2606.
- Hackel,B.J., Kapila,A. and Wittrup,K.D. (2008) *J. Mol. Biol.*, **381**, 1238–1252.
- Halaby,D.M., Poupon,A. and Mormon,J. (1999) *Protein Eng.*, **12**, 563–571.
- Hamill,S.J., Cota,E., Chothia,C. and Clarke,J. (2000a) *J. Mol. Biol.*, **295**, 641–649.
- Hamill,S.J., Steward,A. and Clarke,J. (2000b) *J. Mol. Biol.*, **297**, 165–178.
- Haney,P.J., Badger,J.H., Buldak,G.L., Reich,C.I., Woese,C.R. and Olsen,G.J. (1999) *Proc. Natl Acad. Sci. USA*, **96**, 3578–3583.

- Hu,X., Wang,H., Ke,H. and Kuhlman,B. (2008) *Structure*, **16**, 1799–1805.
- Janin,J. and Wodak,S. (1978) *J. Mol. Biol.*, **125**, 357–386.
- Kabsch,W. (2010) *Acta. Crystallogr. D Biol. Crystallogr.*, **66**, 125–132.
- Karatan,E., Merguerian,M., Han,Z., Scholle,M.D., Koide,S. and Kay,B.K. (2004) *Chem. Biol.*, **11**, 835–844.
- Knappik,A., Ge,L., Honegger,A., et al. (2000) *J. Mol. Biol.*, **296**, 57–86.
- Koide,A., Bailey,C.W., Huang,X. and Koide,S. (1998) *J. Mol. Biol.*, **284**, 1141–1151.
- Koide,A., Jordan,M.R., Horner,S.R., Batori,V. and Koide,S. (2001) *Biochemistry*, **40**, 10326–10333.
- Kotzia,G.A. and Labrou,N.E. (2009) *FEBS J.*, **276**, 1750–1761.
- Kueltzo,L.A., Wang,W., Randolph,T.W. and Carpenter,J.F. (2008) *J. Pharm. Sci.*, **97**, 1801–1812.
- Leahy,D.J., Aukhil,I. and Erickson,H.P. (1996) *Cell*, **84**, 155–164.
- Leahy,D.J., Hendrickson,W.A., Aukhil,I. and Erickson,H.P. (1992) *Science*, **258**, 987–991.
- Lehmann,M., Kostrewa,D., Wyss,M., Brugger,R., D'Arcy,A., Pasamontes,L. and van Loon,A.P. (2000a) *Protein Eng.*, **13**, 49–57.
- Lehmann,M., Pasamontes,L., Lassen,S.F. and Wyss,M. (2000b) *Biochim. Biophys. Acta*, **1543**, 408–415.
- Lehmann,M. and Wyss,M. (2001) *Curr. Opin. Biotechnol.*, **12**, 371–375.
- Maxwell,K.L. and Davidson,A.R. (1998) *Biochemistry*, **37**, 16172–16182.
- Miyazaki,K., Takenouchi,M., Kondo,H., Noro,N., Suzuki,M. and Tsuda,S. (2006) *J. Biol. Chem.*, **281**, 10236–10242.
- Miyazaki,K., Wintrobe,P.L., Grayling,R.A., Rubingh,D.N. and Arnold,F.H. (2000) *J. Mol. Biol.*, **297**, 1015–1026.
- Mosavi,L.K., Minor,D.L., Jr and Peng,Z.Y. (2002) *Proc. Natl Acad. Sci. USA*, **99**, 16029–16034.
- Mueller,U., Perl,D., Schmid,F.X. and Heinemann,U. (2000) *J. Mol. Biol.*, **297**, 975–988.
- Munoz,V. and Serrano,L. (1994) *Proteins*, **20**, 301–311.
- Nikolova,P.V., Henckel,J., Lane,D.P. and Fersht,A.R. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 14675–14680.
- Odermatt,E., Tamkun,J.W. and Hynes,R.O. (1985) *Proc. Natl Acad. Sci. USA*, **82**, 6571–6575.
- Ohage,E. and Steipe,B. (1999) *J. Mol. Biol.*, **291**, 1119–1128.
- Pace,C.N. (1986) *Methods Enzymol.*, **131**, 266–280.
- Parker,M.H., Chen,Y., Danehy,F., Dufu,K., Ekstrom,J., Getmanova,E., Gokemeijer,J., Xu,L. and Lipovsek,D. (2005) *Protein Eng. Des. Sel.*, **18**, 435–444.
- Perl,D., Mueller,U., Heinemann,U. and Schmid,F.X. (2000) *Nat. Struct. Biol.*, **7**, 380–383.
- Plaxco,K.W., Spitzfaden,C., Campbell,I.D. and Dobson,C.M. (1997) *J. Mol. Biol.*, **270**, 763–770.
- Plaxco,K.W., Spitzfaden,C., Campbell,I.D. and Dobson,C.M. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 10703–10706.
- Remmele,R.L., Jr, Bhat,S.D., Phan,D.H. and Gombotz,W.R. (1999) *Biochemistry*, **38**, 5241–5247.
- Rother,K., Preissner,R., Goede,A. and Frommel,C. (2003) *Bioinformatics*, **19**, 2112–2121.
- Sharma,A., Askari,J.A., Humphries,M.J., Jones,E.Y. and Stuart,D.I. (1999) *EMBO J.*, **18**, 1468–1479.
- Sipos,B., Somogyi,K., Ando,I. and Penzes,Z. (2008) *BMC Bioinformatics*, **9**, 27.
- Skerra,A. (2000) *J. Mol. Recognit.*, **13**, 167–187.
- Smith,D.K. and Xue,H. (1997) *J. Mol. Biol.*, **274**, 530–545.
- Steipe,B., Schiller,B., Pluckthun,A. and Steinbacher,S. (1994) *J. Mol. Biol.*, **240**, 188–192.
- Strickler,S.S., Gribenko,A.V., Keiffer,T.R., Tomlinson,J., Reihle,T., Loladze,V.V. and Makhataidze,G.I. (2006) *Biochemistry*, **45**, 2761–2766.
- Stumpp,M.T., Forrer,P., Binz,H.K. and Pluckthun,A. (2003) *J. Mol. Biol.*, **332**, 471–487.
- Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) *Nucleic Acids Res.*, **22**, 4673–4680.
- Tsai,A.M., van Zanten,J.H. and Betenbaugh,M.J. (1998) *Biotechnol. Bioeng.*, **59**, 273–280.
- Vakonakis,I., Staunton,D., Rooney,L.M. and Campbell,I.D. (2007) *EMBO J.*, **26**, 2575–2583.
- Visintin,M., Settanni,G., Maritan,A., Graziosi,S., Marks,J.D. and Cattaneo,A. (2002) *J. Mol. Biol.*, **317**, 73–83.
- Wang,Q., Buckle,A.M., Foster,N.W., Johnson,C.M. and Fersht,A.R. (1999) *Protein Sci.*, **8**, 2186–2193.
- Wirtz,P. and Steipe,B. (1999) *Protein Sci.*, **8**, 2245–2250.