Exploring the potential of the monobody scaffold: effects of loop elongation on the stability of a fibronectin type III domain

Vincent Batori1, Akiko Koide1,2 and Shohei Koide1,2,3

1Department of Biochemistry and Biophysics, University of Rochester School of Medicine and Dentistry, Rochester, NY 14642 and 2Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL 60637, USA
3To whom correspondence should be addressed at Chicago address.
E-mail: skoide@uchicago.edu

The tenth fibronectin type III domain of human fibronectin (FNFn10) is a small, monomeric β-sandwich protein, similar to the immunoglobulins. We have developed small antibody mimics, ‘monobodies’, using FNFn10 as a scaffold. We initially altered two loops of FNFn10 that are structurally equivalent to two of the hypervariable loops of the immunoglobulin domain. In order to assess the possibility of utilizing other loops in FNFn10 for target binding, we determined the effects of the elongation of each loop on the conformational stability of FNFn10. We found that all six loops of FNFn10 allowed the introduction of four glycine residues while retaining the global fold. Insertions in the AB and FG loops exhibited very small degrees of destabilization, comparable to or less than predicted entropic penalties due to the elongation, suggesting the absence of stabilizing interactions in these loops in wild-type FNFn10. Insertions in the BC, CD and DE loops, respectively, resulted in modest destabilization. In contrast, the EF loop elongation was highly destabilizing, consistent with previous studies showing the presence of stabilizing interactions in this loop. These results suggest that all loops, except for the EF loop, can be used for engineering a binding site, thus demonstrating excellent properties of the monobody scaffold.

Keywords: antibody mimic/combinatorial library/ligand binding/peptide display/protein stability

Introduction

A major challenge in protein engineering is the development of binding proteins that exhibit specificity and affinity to a defined target. The most common and successful approach has been to select novel binding proteins from libraries containing a large number of protein variants. Besides numerous reports on antibody engineering (Barbas et al., 1991; Hanes and Pluckthun, 1997; Rader et al., 1998; Sieber et al., 1998; Cattaneo and Biocca, 1999), a number of smaller proteins have been exploited as scaffolds in engineering novel binding proteins (McConnell and Hoess, 1995; O’Neil and Hoess, 1995; Nygren and Uhlen, 1997; Skerra, 2001). Our group has established the tenth type III domain of human fibronectin (FNFn10) as such a scaffold to produce novel binding proteins (Koide et al., 1998, 2002). FNFn10 is a small β-sandwich protein with seven β-strands. Its overall fold is similar to that of the immunoglobulin domains. In the initial trials we prepared a combinatorial phage display library, in which residues in two surface loops (the BC and FG loops; Figure 1) were diversified. FNFn10 variants isolated from this library were found to bind to a test target with modest affinity and specificity, while retaining the overall structure of FNFn10 (Koide et al., 1998). We have termed novel binding proteins engineered on the FNFn10 scaffold ‘monobodies’.

To engineer monobodies, interactions need to be introduced by an insertion of residues in a loop or by redesign of an existing loop. Both elongation and redesign of a loop risk destabilizing the scaffold, since interactions stabilizing the fold may be disrupted and elongation of a loop is generally destabilizing (Helms and Wetzel, 1995; Ladurner and Fersht, 1997; Nøgå and Regan, 1997). FNFn10 has been subject to extensive characterization of its structure (Main et al., 1992; Dickinson et al., 1994; Leahy et al., 1994, 1996) as well as its dynamics and folding (Carr et al., 1997; Plaxco et al., 1997; Cota and Clarke, 2000; Cota et al., 2000, 2001; Koide et al., 2001). Cota et al. previously found FNFn10 to be extremely stable and that it accommodates point mutations in its hydrophobic core with minimal effect on protein stability (Cota et al., 2000). However, effects of altering the loops in FNFn10 have not been systematically investigated. In our initial trial, we chose the BC and FG loops to be diversified solely based on the structural homology of FNFn10 with the antibody variable domain (Koide et al., 1998). In this study, we compared the effect of elongation of all loop regions of FNFn10 to determine the potential of each loop as a site to introduce mutations for target binding. Our results revealed that all but one loop could accommodate mutations without significant loss in the conformational stability.

Materials and methods

Protein expression and purification

The expression vector for wild-type FNFn10 has been described (Koide et al., 1998). For each loop of FNFn10, a four-glycine insertion was designed. Mutant proteins are designated as AB+4, etc., to indicated identity of the mutated loop. The genes for all the mutants were prepared using standard polymerase chain reactions and then inserted in the wild-type FNFn10 expression vector. The genes for all the mutants were confirmed by DNA sequencing to contain only the intended mutations. In addition, an eight-glycine insertion was prepared for the FG loop (FG+8) and a protein harboring both a four-glycine mutation in the BC loop and the eight-glycine insertion in the FG loop was prepared (BC+4:FG+8). All the mutant proteins were expressed as soluble protein and subsequently purified using metal affinity chromatography, as previously described for the wild-type protein (Koide et al., 1998).

Chemical denaturation measurements

Proteins were dissolved to a final concentration of 2 μM in 20 mM sodium citrate buffer at pH 6.0 containing 100 mM sodium chloride. Guanidine hydrochloride (GuHCl)-induced unfolding experiments were performed by monitoring changes in the fluorescence emission from the single tryptophan residue...
in FNFn10 using a Spectronics AB-2 spectrofluorometer equipped with an automated titrator as described previously (Koide et al., 1998, 1999). The GuHCl concentrations before and after a titration experiment were determined using an Abbe refractometer (Spectronic Instruments) as described (Pace and Sholtz, 1997). Data were fitted to the standard two-state model (Santoro and Bolen, 1992) using the program Igor Pro 3.0 (Wavemetrics, Lake Oswego, Oregon). Errors for the free energy of unfolding and m-value were estimated both from the standard deviations determined from the non-linear least-squares routine in Igor Pro and from Monte Carlo simulations. In the latter, we generated 100 synthetic data sets by adding appropriate noise to the fit to the experimental data (Press, 1992), and then the standard deviations for the parameters were determined from results of 100 curve fittings. The errors reported in Table I are the greater of those estimated by these two methods.

**NMR spectroscopy**

NMR experiments were performed at 30°C on a Varian INOVA 600 spectrometer. The 1H, 15N-heteronuclear single-quantum correlation (HSQC) spectra were taken as described previously (Kay et al., 1992). NMR data were processed using the NMR Pipe package (Delaglio et al., 1995), and analyzed using the NMRView software (Johnson and Blevins, 1994).

**Molecular graphics**

Schematic drawings (Figures 1 and 3) were made using the program MOLSCRIPT (Kraulis, 1991).

**Results and discussion**

**Design of insertion mutants**

The structure of FNFn10 contains seven β-strands and six loop regions (Figure 1). We first analyzed the FNFn10 crystal structure (Dickinson et al., 1994) using the program PROMOTIF (Hutchinson and Thornton, 1996) to objectively define the location and extent of the six loops. We designated residues to be in a loop when they are not assigned as in a β-strand, based on dihedral angles and hydrogen bonding patterns. There are large differences among the defined loop lengths from two residues in the AB loop to 11 residues in the FG loop (Table I). Three (BC, DE and EF) loops connect strands in opposite sheets of FNFn10, while the remaining three (AB, CD and FG) loops form a β-hairpin between structurally adjacent strands. For each loop, a mutant protein was designed featuring an insertion of four glycines placed near the center of the loop (Figure 1). Compared to the other amino acids, glycine offers the most conformational freedom, while displaying only limited opportunities to introduce potentially stabilizing interactions with the rest of the protein or within the newly introduced loop. Thus, the glycine insertions should provide the simplest data for identifying stabilizing and destabilizing interactions that are present in the wild-type loop.

**Effects of loop insertion mutations on the conformational stability of FNFn10**

In GuHCl-induced chemical unfolding, both wild-type and all mutant proteins underwent a reversible transition consistent with the two-state approximation (Figure 2). All the mutants showed a lower midpoint of unfolding compared to that of wild-type FNFn10. However, the decrease in the free energy of unfolding varied significantly among mutants. For further analysis, we used the free energy of unfolding at 3 M GuHCl, which is within the unfolding transition region for most of the proteins, to minimize errors caused by long extrapolation (Table I). As measured by the net change of free energy of unfolding (ΔG3M GuHCl in Table I), AB+4 and FG+4 only exhibited a small degree of destabilization followed by BC+4, CD+4, DE+4 and EF+4 in the order of decreasing ΔG3M GuHCl. EF+4 was destabilized by 4.4 kcal/mol, significantly more than the other mutations. Significant differences in the m-value were observed for these mutants, which may have been caused by changes in the structure of the unfolded state (Smith et al., 1996).

We performed NMR measurements on the most destabilized mutant, EF+4, to verify it retained the overall structure of FNFn10. The 1H, 15N-HSQC spectrum correlates the chemical shifts of amide 1H and 15N, thus providing a fingerprint of a protein. The majority of cross peaks in the HSQC spectrum of EF+4 were found to have a corresponding cross peak in the wild-type spectrum (Figure 3A). Sixteen cross peaks clearly shifted between the two spectra, indicating structural changes around these amides. These residues were either in the mutated EF loop itself (six residues) or in adjacent regions (five residues in the AB loop, two residues in the CD loop and three residues at the C-terminal end of strand G). These results indicate that structural changes in this mutant are confined to the vicinity of the insertion and that the overall conformation of the EF+4 protein is similar to that of the wild-type. Since the other mutations exhibited smaller degrees of destabilization than EF+4, they should also have smaller degrees of structural changes.

In addition to these mutants with an insertion of four glycines, a protein with eight glycines inserted in the FG loop was prepared. This FG+8 protein showed an unfolding profile almost identical to that of FG+4 (Figure 2B and Table I). A double mutant featuring both the BC+4 and the FG+8 mutations was examined to test if additional destabilization occurs when two loops in close proximity are simultaneously mutated. The BC+4:FG+8 protein had only a ΔG3M GuHCl of -1.2 kcal/mol, which was the same as the ΔG3M GuHCl for the BC+4 protein with the wild-type FG loop. If we assume that the two loop elongations are independent and thus the decrease in stability additive, a ΔG3M GuHCl of -1.7 kcal/mol would be expected. Thus, these results indicate that the elongation of these two adjacent loops has no synergistic destabilizing effects.

**Comparison of observed effects with predicted destabilization due to entropy change**

We then compared observed changes in the free energy of unfolding with predictions using a simple polymer model (Chan and Dill, 1988; Ladurner and Fersht, 1997; Nagi and Regan, 1997). This model assumes that the examined loop region is highly flexible. For a sufficiently flexible loop, an elongation does not change enthalpic contribution to the free energy since accommodation of additional residues in such a flexible loop does not disrupt existing interactions. Thus, these existing models should predict reasonably well the observed changes in free energy upon elongation due to the loss of entropy. The change in entropy, ΔΔS for an elongated-loop mutant is approximated by:

\[
\Delta \Delta S = cRT \ln \frac{l}{l_{\text{reference}}}
\]

where l represents the number of amino acid residues in a
Loop elongation of fibronectin type III scaffold

Fig. 1. Schematic drawing of the structure of FNfn10 (A) and the amino acid sequence in its secondary structure context (B). An arrow marks the site at which glycine residues were inserted in this study. Loop residues as assigned by the program PROMOTIF are shown shaded. Residues of the β-strands whose side chain forms the hydrophobic core are enclosed in circles with thicker ring.

Fig. 2. (A) GuHCl-induced unfolding curves of the four-glycine insertion mutants and the wild-type FNfn10. The fraction of unfolded protein as determined from Trp fluorescence is plotted as a function of GuHCl concentration. Wild-type FNfn10 is depicted with a cross, AB+4 with closed squares, BC+4 with open diamonds, CD+4 with closed circles and FG+4 with open circles. (B) GuHCl unfolding curves of the BC and FG loop insertion mutants. FG+8 is shown with filled triangles, BC+4:FG+8 is shown with closed diamonds. The inset shows the unfolding curve for the wild-type monitored by fluorescence emission.

loop after a modification, $l_{\text{reference}}$ is the original length and R and T are the gas constant and temperature, respectively. The value of the constant $c$ depends on the model used (Chan and Dill, 1988; Ladurner and Fersht, 1997; Nagi and Regan, 1997). A value of $c = 2.41$ has previously been found most accurate to describe protein elongation (Chan and Dill, 1988). Using the length of each loop in the wild-type protein as the reference length ($l_{\text{reference}}$), the expected change in free energy was calculated (Table I and Figure 4). Conversely, we also calculated apparent loop lengths in the wild-type protein from the destabilization data, as an alternative measure of the accuracy of the prediction (Table I).

The experimentally obtained changes in free energy for the FG loop mutants showed good agreement with the calculated entropic change upon loop elongation. This suggests that these insertions were incorporated into a flexible loop and they had minimal effect on the existing interactions that stabilize FNfn10. The observed penalty for the FG+8 mutant on the stability was even less than the expected value, further supporting that the FG loop can easily accommodate alterations. These results are consistent with NMR studies showing that the FG loop is highly flexible (Main et al., 1992; Carr et al., 1997).

Mutation in the AB loop showed less destabilization than expected from the entropic penalty. Since the AB loop was the shortest loop, its predicted entropic penalty is the highest. An underestimation of the loop size would result in a decreased penalty. Nevertheless, the apparent length of the AB loop (10 residues; Table I) is unrealistically large, and thus an underestimation of the loop size cannot sufficiently explain the discrepancy. An alternative explanation is that the loop in the wild-type protein contains an unfavorable interaction that is relieved due to the insertion. However, examination of the wild-type structure did not reveal uncommon dihedral angles or bad contacts. A previous report indicated that mutations of residues on the A and B strands belonging to the hydrophobic core were less destabilizing than comparable mutations at other locations (Cota et al., 2000). They proposed a rearrangement of hydrophobic contacts upon mutations that compensated lost interactions with newly established ones. If such a repacking event occurred, the observed degree of destabilization of AB+4 might be smaller due to a rearrangement in a manner that effectively shortened the insertion and thereby stabilized the protein. It will be interesting to determine the high-resolution structure of the AB+4 protein to see how much structural arrangement has taken place.
Positions of amides whose chemical shifts were affected in the ER/H11001 hydrogen bond formed in the tyrosine corner motif is indicated by a dotted line. The loop positions are also shown with their respective names. (studies on FNfn10 revealed that there is signifi-
cant motion on the pico- to nanosecond timescale in the BC and CD loops but not as much as in the FG loop, and that the DE loop is as rigid as the surrounding β-strands (Carr et al., 1997). The apparent lengths of flexible segment in these loops estimated from the destabilization (Table I) are consistent with the notion that these loops are not completely flexible.

For the BC, CD and DE loops, the insertion mutations destabilized the protein more than expected from entropic penalty. This can be explained by underestimation of the penalty, which would occur if the length of a flexible loop (l_{reference}) is overestimated. In our calculation, l_{reference} was determined solely based on an inspection of the FNfn10 crystal structure. Thus, if only a part of a given loop is flexible, a smaller number for l_{reference} should be used, which would lead to a greater value for the entropic penalty. NMR relaxation studies on FNfn10 revealed that there is significant motion on the pico- to nanosecond timescale in the BC and CD loops but not as much as in the FG loop, and that the DE loop is as rigid as the surrounding β-strands (Carr et al., 1997). The apparent lengths of flexible segment in these loops estimated from the destabilization (Table I) are consistent with the notion that these loops are not completely flexible.

The destabilization caused by the EF loop insertion significantly exceeded the expected entropic penalty (Table I). As indicated by the unreasonably small ‘apparent loop length’ (0.2 residues), the destabilization was too large to be explained by an overestimation of the wild-type loop length, suggesting that the insertion disrupted stabilizing interactions that are present in the wild-type EF loop. The EF loop includes the ‘tyrosine corner’ motif (Cota et al., 2000) in which a side chain-backbone hydrogen bond is formed between Tyr68 and Lys63 (see Figure 3). It has been shown that the Y68F mutation destabilizes FNfn10 by 2 kcal/mol. In our EF+4 mutant, the elongation is located immediately following Lys63, and thus it is possible that the insertion compromised side chain locations leading to a reduction in the integrity of the hydrogen bond. Notably, if we assume a loss of 2 kcal/mol due to the compromised integrity of the hydrogen bond, the remaining

---

**Table I. Thermodynamic parameters of the unfolding reaction of FNfn10 mutants**

<table>
<thead>
<tr>
<th>Mutant name</th>
<th>ΔG^{298} (kcal/mol)</th>
<th>m^θ (kcal mol^{-1} M^{-1})</th>
<th>ΔG_{GM,GuHCl} (kcal/mol)</th>
<th>ΔG_{GM,GuHCl} (kcal/mol)</th>
<th>ΔG_{calc} (kcal/mol)</th>
<th>Defined loop length^b</th>
<th>Apparent loop length^c</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type</td>
<td>7.70 ± 0.08</td>
<td>1.60 ± 0.02</td>
<td>2.9 ± 0.07</td>
<td>0</td>
<td>0</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>AB+4</td>
<td>8.23 ± 0.07</td>
<td>1.94 ± 0.02</td>
<td>2.4 ± 0.09</td>
<td>−0.5</td>
<td>−1.6</td>
<td>2</td>
<td>9.8</td>
</tr>
<tr>
<td>BC+4</td>
<td>6.47 ± 0.09</td>
<td>1.58 ± 0.02</td>
<td>1.7 ± 0.15</td>
<td>−1.2</td>
<td>−0.7</td>
<td>7</td>
<td>3.1</td>
</tr>
<tr>
<td>CD+4</td>
<td>6.62 ± 0.18</td>
<td>1.81 ± 0.05</td>
<td>1.2 ± 0.19</td>
<td>−1.7</td>
<td>−0.7</td>
<td>7</td>
<td>1.7</td>
</tr>
<tr>
<td>DE+4</td>
<td>5.67 ± 0.27</td>
<td>1.70 ± 0.07</td>
<td>0.6 ± 0.18</td>
<td>−2.3</td>
<td>−1.2</td>
<td>3</td>
<td>1.0</td>
</tr>
<tr>
<td>EF+4</td>
<td>4.45 ± 0.09</td>
<td>1.99 ± 0.03</td>
<td>−1.5 ± 0.16</td>
<td>−4.4</td>
<td>−0.8</td>
<td>5</td>
<td>0.2</td>
</tr>
<tr>
<td>FG+4</td>
<td>6.93 ± 0.16</td>
<td>1.49 ± 0.03</td>
<td>2.5 ± 0.21</td>
<td>−0.4</td>
<td>−0.4</td>
<td>11</td>
<td>11.1</td>
</tr>
<tr>
<td>FG+8</td>
<td>6.62 ± 0.11</td>
<td>1.42 ± 0.02</td>
<td>2.4 ± 0.11</td>
<td>−0.5</td>
<td>−0.8</td>
<td>11</td>
<td>17.4</td>
</tr>
<tr>
<td>BC+4:FG+8</td>
<td>5.99 ± 0.30</td>
<td>1.43 ± 0.07</td>
<td>1.7 ± 0.30</td>
<td>−1.2^d</td>
<td>−1.4^d</td>
<td>7+11^d</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−0.7^e</td>
<td>−0.7^e</td>
<td>7^e</td>
<td>6.8^e</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−0.03^f</td>
<td>−0.8^f</td>
<td>11^f</td>
<td>375.6^f</td>
</tr>
</tbody>
</table>

^aValue obtained from fitting of the two-state model to the experimental data.

^bLoop length designated based on secondary structure found in the crystal structure (Dickinson et al., 1994) using PROMOTIF (Hutchinson and Thornton, 1996).

^cCalculated loop length in the wild-type protein that would result in the observed ΔΔG value.

^d,e,fResults for BC+4:FG+8 were analyzed using wild-type FNfn10 (d), FG+8 (e) or BC+4 (f) as the reference, respectively.
destabilization effect (2.4 kcal/mol) is close to that found for the DE loop. Similar to the DE loop, the NMR studies did not find increased motion in the EF loop, although limited data are available for this loop (Carr et al., 1997).

The FG+8 mutation did not affect the stability of the BC+4 mutant. Since the FG+8 mutation is destabilizing when introduced to the wild-type protein, the lack of destabilization on the BC+4 background may be due to stabilizing interactions between the elongated two loops.

Cota et al. have analyzed an extensive set of point mutations in FNfn10 focusing on residues forming the conserved hydrophobic core (Cota et al., 2000, 2001). They also measured the destabilization effects of an alanine substitution at a few residues in the loop regions. The P15A mutation in the AB loop showed almost no destabilizing effect (∆∆G = 0.02 kcal/mol). The P25A and V29A mutations in the BC loop destabilized FNfn10 by 0.48 and 1.45 kcal/mol, respectively. The P44A mutation in the CD loop showed a ∆∆G of 0.41 kcal/mol. The L62A and V66A mutations in the EF loop destabilized the protein by 2.94 and 0.64 kcal/mol, respectively. These data are generally consistent with our loop elongation data. However, some of alanine substitution mutants (e.g. P25A, P44A and V66A) show little destabilization, while our insertion mutant in the corresponding loop had a greater degree of destabilization. This observation suggests that the insertion mutations induce larger scale structural changes than alanine scanning mutagenesis where the changes are likely confined in the side chains, and thus insertion mutations may be more suitable for rapidly assessing the overall importance of a loop to stability.

Previous reports on loop elongation focused on the elongation of a single loop in a protein. Loop elongation mutants of chymotrypsin inhibitor-2 yielded a close match of the determined destabilization compared to entropy calculation for glycine inserts (Ladurner and Fersht, 1997), while insertion of alanine- or glutamine-rich peptides proved to be more destabilizing. Elongation of a loop connecting two helices in Rop revealed greater destabilization than expected from calculation (Nagi and Regan, 1997), although the degree of destabilization of these Rop mutations is not as much as that found in the EF+4 mutation of FNfn10. The wide distribution in the magnitude of destabilization found in our study demonstrates that loop elongation may have quite a different effect on protein stability, depending on the nature of a target loop.

The purpose of this study is to assess the possibility of engineering monobodies using loops in addition to the BC and FG loops used in our initial experiments. Although effects of a particular inserted sequence on conformational stability of a scaffold are somewhat unpredictable (Helms and Wetzel, 1995) and our data cannot address the dependence of destabilization on the location of loop elongation, it would be most logical to choose loops that can accommodate mutations with minimal loss in stability. Our results provide a clear ranking of the six loop regions: AB and FG loops are best suited for the introduction of additional residues, followed by the BC, CD, DE and EF loops in the decreasing order. The β-sandwich fold of FNfn10 presents three loops on each end of the protein (Figure 1). This study revealed that all loops except for the EF loop could accommodate alterations without drastic loss in stability. Thus, the end featuring the BC, DE and FG loops (the top end in Figure 1) of FNfn10 has three loops available, while the other end has two. Thus, the top end may be more desirable for engineering functional loops. This end of FNfn10 corresponds to that of the immunoglobulin variable domain that contains the three variable loops. This structural homology also suggests that the top end of FNfn10 is the favored site for engineering of a ligand-binding site. However, the immunoglobulin variable domains in which the other end is stericly blocked by the constant domain, both ends of FNfn10 should principally allow the presence of a binding site. For example, in the structure of human growth hormone receptor, AB and CD loop of its fibronectin type III domain are involved in ligand binding (de Vos et al., 1992). Indeed, we were able to isolate monobodies that use the AB loop for target binding (Koide et al., 2002), confirming the versatility of the FNfn10 scaffold.

Acknowledgements

We thank Shude Yan for the Monte Carlo program for error analysis and Scott Kennedy for assistance in NMR experiments. This work was supported in part by NIH grant GM55042 to S.Koide.

References


Received January 3, 2002; revised June 25, 2002; accepted October 4, 2002