

## **Cell- and heparin-binding domains of the hexabrachion arm identified by tenascin expression proteins.**

**Ikramuddin Aukhil, Paritosh Joshi, Yingzhuo Yan, and Harold P. Erickson**  
**Supplementary material on construction of plasmids and protein purification.**

**J. Biol. Chem. 268: 2542-2553 (1993)**

*Revised version March 1999 incorporating some changes in purification protocols*  
**A word about SDS-PAGE.**

In our paper we reported that all of the expression proteins were the correct size on SDS-PAGE, but more recently we have had a closer look comparing the TNfn bands to the decade markers. We find that our TNfn proteins routinely run about 15% higher relative to these markers than their actual MW. The BioRad broad MW standards agreed pretty closely with the decade markers above 40 kDa, but the 31 and 21.5 kDa BioRad markers appeared at 28.5 and 17 kDa in the decade ladder. Just goes to indicate again the limitations of SDS-PAGE for accurate MW.

### **A word about storing plasmids.**

We store the expressing strains (BL21(DE3) transformed with the pET plasmid) as freezer stocks (in 15% glycerol, -80° C) for up to several years. To start a culture we scrape the frozen stock with a sterile pipet tip not letting it thaw. Some people recommend streaking the freezer stock to obtain a single colony before doing the prep.

The freezer stock should never be relied on for archival storage, as the plasmids can be damaged or lost. The best procedure for archival storage is to make a plasmid prep and store the DNA in 70% ethanol in the freezer. It is difficult to make a good plasmid prep from BL21, so we make a provisional plasmid prep from BL21, use this DNA to transform DH5 $\alpha$ , and then to a plasmid prep from the DH5 $\alpha$ . We use this plasmid to retransform BL21 and make a new freezer stock every year or so, or whenever the expression shows any problem.

### **Preparation of pET vectors, bacterial growth and protein expression**

We have used the pET expression vector system designed by Studier and colleagues (53) and available commercially from Novagen. For most work we have used the specific vector pET-11b (the 11a and 11c vectors would be equivalent). The DNA segments chosen for expression were cloned unidirectionally into the NdeI and BamHI sites of pET-11b.

All of our expression proteins have been made without fusion proteins, and have been designed to start and stop precisely at boundaries of folding domains (these can be determined precisely from sequence alignments). The table of primers used are in the Aukhil et al miniprint. Both the forward and reverse primers contained 18-21 bases corresponding to the 5' and 3' terminal sequences of the desired coding segment. The forward primer incorporated the NdeI restriction site immediately before the coding region; the final three bases of the NdeI site, ATG, code for the initiation methionine residue. The reverse primer added a TAG stop codon immediately after the coding segment, followed by the BamHI site. We always included a 3-7 base overhang at the free end of the restriction site to facilitate cutting. BamHI is happy with 3 bases, but NdeI is supposed to require a 7-base overhang. We have gotten by with a 3-base overhang on many occasions, but this cutting may be very inefficient. NdeI frequently gives problems, and sometimes it has been necessary to subclone PCR products first (into an A/T vector) and then cut the purified plasmid.

For PCR amplification we used 2.5 units of Taq polymerase (Promega) in a total volume of 100  $\mu$ l. *PFU polymerase is now highly recommended because of its lower error rate.* The reaction mixture contained 10  $\mu$ l of 10X buffer (supplied by Promega), 100  $\mu$ M each dNTP, 500 ng each primer, and variable amounts of template DNA (up to 500 ng of purified plasmid, or up to 1  $\mu$ g of cDNA). The reaction was cycled 40 times with denaturing, annealing and extension temperatures of 94°, 60-68° (depending on T<sub>m</sub> of primers), and 72° C. PCR products were gel purified, cut with the NdeI and BamHI, and ligated to cut pET11. *It is very important to transform the ligated product initially into DH5 $\alpha$  or JM109, do a miniprep and, use this large quantity of plasmid to transform BL21. BL21 requires at least ten times as much plasmid for transformation.*

The DNA for clones TNfn1-5 and FNfn7-10 each contain an internal BamHI site; the DNA for TNfbg contains an internal NdeI site. For these preparations the digestions were done with a limited amount of enzyme and for short times, testing for conditions that would give partial digestion. The larger fragments, not cut at the internal site, were then selected by gel purification.

Because PCR is known to generate errors, we sequenced the PCR derived DNA of all constructs. Out of 5,000 bases sequenced we discovered 8 base changes, giving an average of one error per 600 bases. Since the PCR was run for 40 cycles, this indicates a basic error frequency of  $4 \times 10^{-5}$ , which is in the range of reports from other laboratories (54). TNfn3 showed a single base change, but this was a silent mutation. FNfn10 also had a single silent mutation. FNfn7-10 had three base changes, but these have been corrected by Dan Leahy and only the corrected plasmid is used now. TNfn1-5 had three errors, all in domain 5: one was silent, one changed RLNY to RLDY, and the other changed PARV to PAHV (this latter change was designed into the primer based on the original sequence of Nies et al. (3)). TNfn3-5 should have these same errors in domain 5; TNfn1-3 should be without errors. In TNfnA-D (the second generation vector with PCR DNA only at the ends), there is one conservative change in domain D: RDIT is changed to RDLT. (TNfnA-D(-C) was not sequenced, since we use the full length TNfnA-D for most purposes.) TNfn6-8 showed one silent mutation. TNfbg was perfect.

### **Second generation expression vectors – replacing PCR product with cloned DNA.**

We have constructed five new plasmids, in which the majority of the coding DNA was cut directly from a cloned plasmid and ligated into the expression vector. This has two advantages: (1) it removes errors that might be introduced by the PCR amplification, and (2) it facilitates construction of large vectors and ones with specific splice variants. Vectors previously constructed by PCR were used as the basis for this splicing, supplying the pET-11 DNA as well as the beginning and end sequence of the expression proteins. The cloned DNA was obtained from the constructs pBS-HxB.L and pBS-HxB.S, described below in the section on mammalian expression vectors.

We will summarize here the splice points demarcating the PCR DNA at the ends from the cloned DNA insert. Details of the splicing will be supplied upon request.

TNfnALL contains cloned DNA from the BamHI @ 2250 in domain 2 to the XhoI @ 5624 in domain 7, including the seven alternatively spliced domains.

TNfn1-8 contains the same cloned DNA insert as TNfnALL, but is missing the alternatively spliced domains.

TNfn1-3 contains cloned DNA from the BSTX1 @ 2038 in domain 1 to the ScaI @ 2626 in domain 3. (Curiously, this plasmid TNfn1-3 would not grow in bacterial strain JM109; intermediate cloning and amplification used strain DH5 . )

TNfn3-5 is a second generation vector in that it was spliced from existing TNfn3 and TNfn1-5, but it contains only PCR DNA.

TNfnA-D, contains cloned DNA from the SacII @ 3344 in domain A1 to the HincII @ 4923 in domain D. The expression vector for the alternatively spliced domains, called HxB6-12 in our previous paper (45), was subsequently found by restriction and PCR analysis to be missing domain C (Y. Z. Yan, unpublished data). This vector is now designated TNfnA-D(-C).

### **Bacterial growth, induction and lysis.**

These steps were approximately the same for all of the recombinant proteins. 10-20 ml of L-broth/amp (50 µg/ml ampicillin) was infected with a freezer stock of e.coli B121(DE3), transformed with the pET-11 plasmid containing the appropriate insert, and grown overnight. The bacteria were centrifuged and resuspended in 10 ml fresh L-broth/amp, and added to 1.2 L of L-broth/amp in two 2L baffled flasks. The bacteria were grown for 2 – 3 hr shaking at 37° C (the  $A_{600} = 0.06 - 0.8$ ), and IPTG was added to 0.4 mM. The  $A_{600}$  increased about 70% during the following 30 min and then was stationary. Growth was continued for 3 – 4 hr following induction, and the bacteria were then collected by centrifugation (10 min at 5,000 rpm, Sorvall GSA rotor). Typically 4-5 g bacterial pellet was obtained per l of culture. The bacteria were resuspended in ~16 ml of lysis buffer (50 mM tris, pH 8, 1 mM EDTA, 0.1 M NaCl) per L of culture, giving a total volume of 20 ml. Lysozyme was added to 0.2 mg/ml and the culture was agitated at room temperature for 30 min. Triton X100 was added to 1% (Note: often we have problems with a part of the pellet floating and sticking to the side of the tube when we do  $(\text{NH}_4)_2\text{SO}_4$  precipitation. It was suggested that the floating is due to the high concentration of triton. We now routinely use 0.1 % triton, and have mostly eliminated this problem). PMSF was added to 0.1 mM, and agitation continued for 30 min, giving partial lysis. The sample was frozen, and upon thawing exhibited a very high viscosity indicating complete lysis. An additional 0.1 mM PMSF was added, Mg was added to 10 mM and DNAase I to 10 µg/ml. Agitation for 30 min led to a substantial but sometimes incomplete reduction in viscosity. Additional DNase and/or a brief sonication was used to eliminate the remaining viscosity. We sometimes add three lots of DNase at 30 min intervals, to reduce DNA contamination at later stages. The sample was then centrifuged at 10,000 rpm for 10 - 20 min, and the pellet and supernatant were analyzed by SDS-PAGE. With the exception of proteins containing Tnfbg, and sometimes TnfnA-D, the expressed proteins were mostly or entirely in the supernatant, and were usually the most prominent band.

**French Press.** We have recently found that we get much better yields of soluble protein if we lyse bacteria in a french pressure cell. The bacteria are suspended in the desired volume (generally 20 ml for 1 l culture) frozen overnight, thawed and lysed by three passes through the french press. The first pass is collected in PMSF. We use no detergent for this. Treat with Mg and DNase and centrifuge.

### **Purification of bacterial expression proteins.**

All steps in all purifications were performed at room temperature, except as noted for Tnfbg.

**TNfn3.** This protein was substantially purified from the bacterial supernatant by an  $(\text{NH}_4)_2\text{SO}_4$  cut between 50% (which precipitates almost all bacterial proteins) and 70% (which precipitates TNfn3). The pellet was resuspended in 12 ml of 0.02 M tris-HCl, pH 7.9 (this buffer was used for all mono Q chromatography), and chromatographed a mono Q 10/10 column.

Note: all buffers and protein solutions are filtered through 0.2  $\mu\text{m}$  filters before passing through the pumps and HPLC columns.

*In recent experiments we have found that 70% saturated ammonium sulfate leaves almost all the TNfn3 in the sup, and pellets the contaminating proteins completely, better than the 50% cut. So we now do a 70% satd ammonium sulfate cut, dialyze the sup vs mono Q buffer and load on the column. But check the pellet and sup, some TNfn3 may be pelleted at 70% in which case you should probably try 65%. This is an amazingly powerful purification step because it is a rare protein that remains soluble at 70% satd ammonium sulfate.*

The column was eluted with a linear gradient from 0.1 to 0.3 M NaCl (in tris buffer), and the protein eluted as a single peak at 0.22 M NaCl. Although the SDS-PAGE showed no contaminants, we wanted one more step of purification, especially for crystallography. The final step was chromatography on mono S or resource S, in 0.02 M formic acid, pH 3.5 or 4.2. The protein eluted as two sharp peaks, near 0.17–0.2 M NaCl. The exact elution position was very sensitive to the pH of the buffer, but the separation of the two peaks was reproducible, provided the pH was 3.5–4.2. The two peaks apparently arise from partial cleavage in the bacteria of the n-formyl methionine that initiates translation (53). The protein in the first peak contained the nf–met, and that in the second peak a free amide, and hence an extra positive charge. The amino acid composition confirmed two methionines per 91 residues in the first peak and one in the second. Up to 6 mg of protein could be separated on this small mono S column, but larger quantities overloaded it; the 6 ml Resource S column easily handled 12 mg TNfn3. Recovery from one 2 l preparation was 8.7 mg of peak 1 and 2.3 mg of peak 2. In a more recent preparation from cells grown in M9 medium (a 1 l culture supplemented with 2 g  $^{13}\text{C}$  glucose and 1 g  $^{15}\text{N}$   $\text{NH}_4\text{Cl}$ ) we recovered a total of 11.7 mg.

TNfn3 has been crystallized (peak 1 from the mono S) and its structure solved to atomic resolution (41).

**TNfn1-5.** The protein in the bacterial supernatant was first purified by an  $(\text{NH}_4)_2\text{SO}_4$  cut between 40 and 55% saturation. The next step was mono Q chromatography, where the protein eluted at 0.19–0.23 M NaCl. The protein appeared pure on SDS-PAGE at this step, but a final purification on heparin agarose (the protein eluted at 0.19 M NaCl) is recommended.

**TNfn1-3.** *This clone has been lost, it was unstable in the freezer stocks.* TNfn1-3 was precipitated by 33% saturated  $(\text{NH}_4)_2\text{SO}_4$ , leaving most of the bacterial proteins in the supernatant. The pellet was resuspended and chromatographed on Sephadex G–75. Attempts at mono Q chromatography resulted in complete loss of protein, so peak fractions were passed over the mono S column in 20 mM Tris-HCl (pH 7.9). TNfn1-3 did not bind to the column, but minor bacterial contaminants were removed.

**TNfn3-5.** The protein in the triton bacterial lysate was first purified by an  $(\text{NH}_4)_2\text{SO}_4$  cut between 30% (this 30% cut doesn't really get rid of much, could be eliminated) and 50%. The 50%  $\text{AmSO}_4$  pellet was substantially contaminated with bacterial proteins. The resuspended pellet was chromatographed on mono Q, where TNfn3-5 eluted in two peaks, at 0.065 M and 0.09–0.13 M NaCl. The first peak is probably monomeric, and appeared >95% pure on SDS-PAGE. The second peak is probably oligomeric (see discussion under TfnA-D) and should be discarded. (In several recent preparations, TNfn3-5 eluted from the mono Q as a single peak between 0.13 and 0.2 M NaCl.) TNfn3-5 was further purified on heparin-Sepharose column, where it eluted at 0.27 M NaCl. *The heparin step is highly recommended.*

**TNfn6-8.** The protein in the bacterial supernatant was first purified by an  $(\text{NH}_4)_2\text{SO}_4$  precipitation. In some preparations TNfn6-8 was soluble at 40% and precipitated at 50%  $(\text{NH}_4)_2\text{SO}_4$ ; in other preparations it precipitated at 40%  $(\text{NH}_4)_2\text{SO}_4$ . This should be checked on each purification. The pellet was suspended in tris buffer, dialyzed and chromatographed on mono Q chromatography. It eluted at 0.085 M NaCl, or 0.105 M NaCl (later preparation). In one preparation it eluted at .06 M NaCl, with a smaller peak at 0.09. In another it eluted in two peaks at 0.07 and 0.10 M NaCl. When re-chromatographed each of these fractions eluted at 0.12 M NaCl. The two peaks also sedimented at identical positions. We don't understand why the initial mono Q gives variable elution, but we consider all of these peaks to be identical protein. This material appeared pure on SDS-PAGE, and was generally used without further purification. Two preparations were further purified on the mono S column, eluting at 0.19 M NaCl (pH 3.5) or 0.06 M NaCl (pH 4.5). TNfn6-8 did not bind to a heparin-Sepharose column.

**TNfnA-D and TNfnA-D(-C).** The protein expression and purification of these two vectors (A-D containing all seven alternatively spliced domains, and A-D(-C) missing domain C) were virtually identical.

Because this protein was expressed at very high concentration, the amount of lysis buffer was doubled: ~7 g bacteria from 1.5 l culture was suspended in 30 ml lysis buffer. This protein was first purified by precipitation with 20% saturated  $(\text{NH}_4)_2\text{SO}_4$ ; this low concentration of  $(\text{NH}_4)_2\text{SO}_4$  completely precipitated the expression protein while leaving most bacterial proteins in solution. However the pellet from 20%  $(\text{NH}_4)_2\text{SO}_4$  is very soft and hard to collect; 25% sometimes gives a better pellet with little increase in bacterial proteins. Also, to obtain a good pellet it is important to use no Triton or no more than 0.1% Triton when lysing the bacteria. The pellet was resuspended in mono Q buffer, dialyzed, then loaded on the mono Q 10/10 column, and eluted with a linear gradient from 0.20 - 0.43 M NaCl. A cleanly separated peak eluted at 0.26 - 0.28 M NaCl, followed by a second peak at 0.37 M, and by 3-4 poorly resolved peaks at 0.39-0.41 M NaCl. All peaks showed predominantly or exclusively the 65 or 70 kDa bands on 10% SDS gels. The first peak was identified as monomeric protein, and the later peaks were oligomers comprising 3-10 or more subunits held together by stable non-covalent bonds. The oligomeric nature of these later peaks was demonstrated by sedimentation on glycerol gradients (they sedimented ahead of the monomers, in the range 6-12 S), and by electron microscopy of rotary shadowed specimens. Electron micrographs showed irregular oligomers with ~3-10 rods in each complex. Monomeric material could be recovered from the oligomeric fractions by adding urea to 6 M to dissociate the complexes, and slowly dialyzing out the urea (overnight, with no agitation). When the renatured protein was rechromatographed on mono Q, about 50% was recovered in the 0.31 M NaCl monomer peak, and the remainder in the 0.39 M peak.

In recent preps we have noticed that the A-D protein comes in two sizes as indicated by SDS-PAGE. The full length protein (presumably 70 kDa, but we haven't checked it by mass spec) comes off the mono Q at 0.27 M NaCl, and a smaller 65 kDa band comes off at 0.28 M. A mixed sample sent for mass spec gave a MW of 61,401 Da, which may be the smaller protein. We believe the smaller band is a proteolytic product, but the mass spec doesn't let us determine where the cut is.

In recent preparations, we have also noticed that the material from the mono Q is often contaminated with DNA, as indicated by the spectrum. We now always include a final purification by precipitating with 25% saturated  $(\text{NH}_4)_2\text{SO}_4$ , resuspending the pellet and dialyzing ( $(\text{NH}_4)_2\text{SO}_4$  could be toxic to cells).

We have also noticed that up to 50% of the total expressed TNfnA-D is sometimes in the bacterial pellet. In some cases we have recovered this precipitated material. The pellet is resuspended in lysis buffer, Mg and Dnase added (the pellet contains substantial DNA), and it is re-centrifuged. The pellet is washed once more in lysis buffer, pelleted and dissolved in 6 M urea. The urea is slowly dialyzed away (in the cold) and the soluble renatured protein purified on the mono Q. In at least one test this material seemed less active than the “native” TNfnA-D, but this might have been a concentration error. Or it might have been primarily the proteolyzed fragment in the pellet. Work on this step is still in progress. In our best preparations we still have ~100% of TNfnA-D in the bacterial supernatant, and consider pelleted material with some suspicion.

But overall we seem to be getting a much lower yield of 70 kDa protein in 1997-98 than we did years ago. We have even re-transformed BL21 using two separate isolates of the pETTNfnA-D, with similar results. Could it be a bad BL21?

**TNfnALL.** This protein comprises all 15 FN-III domains of the largest splice variant. The protein was precipitated from the bacterial supernatant at 35% saturated  $(\text{NH}_4)_2\text{SO}_4$ . Attempts to run the  $(\text{NH}_4)_2\text{SO}_4$  pellet directly on mono Q gave a broad peak with lots of contaminants. We suspected the column was being overloaded with DNA or other contaminants, so the  $(\text{NH}_4)_2\text{SO}_4$  pellet was first run over a Sephacryl 100 column. The peak Sephacryl fractions were pooled and run on a mono Q, eluting as a sharp peak at 0.31 M NaCl.

**TNfn1-8.** This protein comprises all of the FN-III domains of the smallest splice variant, specifically domains 1-5 and 6-8. The protein was first purified by an  $(\text{NH}_4)_2\text{SO}_4$  cut (30 – 50% saturated), then the 50% pellet was resuspended in tris buffer and chromatographed on mono Q, where it eluted at 0.2 – 0.25 M NaCl. The protein was finally purified on heparin-sepharose, where it eluted at 0.16 M NaCl.

**TNfbg.** *This section substantially revised 5/95.* Unlike the other expression proteins, this domain was entirely in the insoluble pellet of the bacterial lysate. We assumed this was due to inability of the disulfides to form correctly in the bacterial cytoplasm. The pellet from a 1 liter lysed culture was washed twice with 20 ml lysis buffer containing 2 M urea and repelleted, resulting in significant purification. In the first wash, Mg, DNase and triton were added (the extra DNase seems important to reduce DNA contamination, which has been substantial in some preparations). The pellet from the second wash was then suspended in 20 ml 4 M GuHCl in 0.02 M Tris-HCl, pH 8.0, containing 5 mM DTT. The resuspended pellet was centrifuged to remove insoluble material, and then dialyzed against nine volumes of 2 M GuHCl with no DTT. (We observed no formation of precipitate even upon rapid dilution to 2 M GuHCl, but precipitate formed if the GuHCl concentration were rapidly reduced to below 2 M. Similarly, for protein denatured in urea, we found no precipitate until the urea was reduced to 5.5 M, but substantial precipitate if urea was rapidly reduced to 5 M or below.) The 180 ml dialysis buffer (2 M GuHCl in 0.02 M tris, pH 8) was changed twice, reducing the DTT to 0.005 mM. A glutathione redox mixture (GSH/GSSG) consisting of 1 mM reduced and 0.3 mM oxidized glutathione, was added to this final dialysis mixture (this mixture was chosen as a consensus of several papers describing renaturation of different proteins with disulfide bonds). The concentration of GuHCl was then slowly reduced by pumping 0.02 M tris, containing the same GSH/GSSG mixture, into the dialysis container. Specifically, the 200 ml dialysis mixture was in a 400 ml beaker with a stir bar, and 200 ml of tris-glutathione was pumped into the beaker over a period of 12-20 hrs, reducing the GuHCl concentration to 1 M. At 1 M GuHCl a modest precipitate usually formed in the dialysis bag. The slow dialysis was continued by decanting 200 ml of the dialysis buffer, and pumping in another 200 ml of tris (this time without glutathione) over 12-20 hrs, reducing the GuHCl to 0.5

M. At this point the GuHCl buffer was discarded and the protein dialyzed against several changes of 0.02 M tris buffer.

The final step in the purification is chromatography on mono S, in 0.02 M tris pH 7.9, with a gradient from 0 – 0.2 M NaCl. We discovered that protein chromatographed directly from the final tris dialysis described above gave two peaks: peak 1 eluted at 0.04 M NaCl, and peak 2 at 0.075 M NaCl. Electrospray mass spectrometry indicated that peak 2 had the molecular weight expected for TNfbg plus an initiation met (without an N-formyl), and peak 1 was 607 Da heavier. The additional mass corresponds to two glutathione molecules, suggesting that the peak 1 protein had one incomplete disulfide, with the two cys residues forming disulfides to glutathione molecules. To test and correct this problem, we dialyzed peak 1 fraction against 20 mM tris containing 1 mM DTT (expecting to reduce the cys-glutathione pairs), then against three changes of ten volumes of tris, adding 0.1  $\mu$ M copper sulfate to the final dialysis to catalyze oxidation of the sulfhydryls. When this material was re-run on mono S it eluted entirely as peak 2. In our most recent purifications, we have therefore added the DTT reduction and copper catalyzed oxidation as a final dialysis step before the mono S purification. With this protocol we obtain a higher yield of peak 2 and no peak 1.

Recently we have explored mono Q chromatography as an alternative to mono S. The renatured protein is chromatographed on mono Q in 0.02 M tris, pH 8.0, and eluted with a linear gradient from 0-0.2 M NaCl. Protein equivalent to the mono S peak 2 eluted at 0.07 M NaCl, while peak 1 protein eluted somewhat earlier. The yield of protein from the mono Q was 1.7 times higher than that from the mono S in one experiment. The separation of peak 1 and peak 2 proteins was not as good on mono Q, but this is not a problem now that the final DTT treatment appears to have eliminated peak 1. We have not used the mono Q as extensively as mono S, but the the higher yield of protein makes mono Q the recommended final purification step.

{Our best preparation, in March 1995, gave about 40 mg per l Tnfbg expressed in the pellet from the lysed bacteria. Recovery as soluble protein after renaturation was better than 70%, and we recovered 10 mg of purified peak 2 protein from the mono S column. If we had used the mono Q we probably would have gotten 15-18 mg. In later preparations the initial expression levels were much less, and the recovery even less efficient. We have generated a new freezer stock by re-transforming BL21, and this has improved yield, but results with Tnfbg are still quite erratic.}

## **Purification of cell adhesion domains from fibronectin.**

**FNfn10.** The protein was first purified from the bacterial supernatant by an  $(\text{NH}_4)_2\text{SO}_4$  cut between 50 and 90% saturation. The protein pellet was resuspended and chromatographed on a Sephadex G-50 column: FNfn10 eluted near the included volume and was about 90% pure on SDS-PAGE. The peak fractions were pooled and passed over the mono Q column. FNfn10 did not bind to the mono Q in 0.02 M tris, pH 7.9, and this flow-through fraction was pure as judged by SDS-PAGE. This fraction was used for most biological assays. A final step of purification prior to crystallization was achieved by chromatography on mono S, eluting with a gradient from 0 – 0.3 M NaCl in 0.02 M sodium formate, pH 3.5. FNfn10 eluted as a single peak at 0.21 M NaCl in some preparations, and as two peaks, at 0.21 and 0.23 M NaCl, in others. The second peak is presumably missing the n-formyl methionine.

FNfn10 has been crystallized (H. C. Taylor et al., unpublished).

**FNfn7–10.** The protein was precipitated from the bacterial supernatant at 40% saturated  $(\text{NH}_4)_2\text{SO}_4$ , and chromatographed on mono Q, where it eluted at 0.2 M NaCl. SDS-PAGE indicated the protein was >90% pure following mono Q. In the original purification we did a final step of chromatography on mono S, which always results in substantial loss (we typically lose 50 % of any protein chromatographed on mono S). We now have a very simple final step of purification. Dialyze the protein from the mono Q vs 0.02 M sodium formate, pH 4.35 (make 0.02 M formic acid and add NaOH to raise the pH - if it doesn't precipitate, try lowering the pH to 4 or 3.7), in the cold. The protein will form a crystalline precipitate of fine needles. After several hours spin down the needles and dissolve in a higher pH buffer of your choosing. Check the pellet and supernatant on a gel to assure that most of the protein is in the pellet. In a recent prep we recovered 48 mg of highly purified protein from 2 l bacterial culture.

Our original version of FNfn7-10 had two amino acid errors. Dan Leahy (Columbia University) has prepared a corrected plasmid by site directed mutagenesis. The new “repaired” plasmid was originally called FNfn7-10/R, but we now have retired the earlier versions and refer to the new one as FN7-10. We have not noticed any differences in adhesion to the original and repaired protein. Anyone who has received this plasmid since 1993 should have the repaired protein with the native aa sequence.

The crystal structure of FN7-10 is published (Leahy, Aukhil and Erickson, Cell **84**:155-164, 1996).

## **Mammalian expression vectors for large, small and truncated hexabrachions.**

Clones covering the entire coding sequence of human tenascin were generously provided by Dr. L. Zardi, Istituto Nazionale per la Ricerca sul Cancro, Genoa, Italy (4). The clones used for our work were phage HT12, and M13 phages HT26 and HT24A,B, and C (the three M13 phages cover the phage HT24, which is cut at two internal EcoRI sites). To produce the smaller splice variant, we used the plasmid p31 (57), which was generously provided by Dr. K. Stefansson, Department of Neuropathology, University of Chicago, and which is missing the seven FN-III domains of the alternatively spliced segment.

Three plasmids were constructed by series of 3-piece ligations. First, the coding segments from each of the and M13 phages were ligated into plasmid Blue Script (pBS KS-, Stratagene), at the ecoRI site. Plasmid pBS-HxB.L, which contains the complete coding sequence for the

largest splice variant, was constructed by splicing together the five clones from Zardi. The restriction sites utilized for this construction were BstBI @ 770; XbaI @ 2253; SacII @ 3759; EcoRI @ 4771; NdeI @ 6177. Plasmid pBS-HxB.S, which codes for the smaller splice variant, used the Stefansson clone p31 to supply the segment between the XbaI and NdeI sites. Plasmid pBS-HxB.egf used the PCR derived clone TNegf1-14 (Table A), originally designed for the bacterial expression vector, to supply the stop codon at base pair 1912. The coding sequence between the BstBI @ 770 and an AvaI @ 1866 was replaced by cloned DNA from pBS-HxB.L, in order to eliminate possible errors introduced by PCR. HxB.egf, .L and .S all contain the 5' untranslated segment present in the Zardi clone HT24C.

The constructs in pBS were all confirmed by restriction analysis to have the proper size. They were then cut from pBS at the NotI and KpnI sites flanking the construct. The fragments were blunted with Klenow reagent. The mammalian expression vector pNUT (58), kindly supplied by Dr. Richard D. Palmiter, Department of Biochemistry, University of Washington, was cut with BamHI and blunted. The blunted tenascin constructs were ligated into the pNUT vector, and clones having the correct orientation were identified by restriction analysis. Transfection of BHK cells by the calcium phosphate procedure and selection of colonies in methotrexate followed standard protocols, as described in a previous application of the pNUT vector (59).

*The preferred reference for the sequence of human tenascin is X78565 (Ghezi...Zardi JBC 270:3429-3434), which has corrected a few errors in previous entries.*

**Table 1. Physical and chemical characterization of expression proteins**

(This table reproduced from the main manuscript, qv for notes. The second listing (*italics*) under yield gives results obtained in 1996)

<b>Protein name</b>	<b>1. M<sub>r</sub></b>	<b>2. ε</b>	<b>3. S</b>	<b>4. Size</b>	<b>5. Yield</b>
TNfn1-3	30,645	1.14			<b>7.5 mg/l</b>
TNfn1-5	50,388	1.22	3.0, 3.2	14.7nm	<b>6</b> (before hep) <i>4.6; 2.7 after hep</i>
TNfn1-8	78,872	1.17	3.2	24.6	<b>8</b> <i>8.9 (after hep)</i>
TNfnALL	148,288	1.22	4.3	47.9	<b>2.4</b>
TNfn3	10,142	0.97	~1.7 S	3.7	<b>6 - 12</b>
TNfn3-5	29,885	1.22			<b>18</b> (+18 oligomer) <i>13 (peak 1, after hep)</i>
TNfnA-D(-C)	59,459	1.34			<b>32</b> (+122 p)

TNfnA-D	69,434	1.27	3.1, 3.5	23.6	<b>38</b> (+25 p) <i>14 as mon from bact sup</i> <i>30 as mon from b pel</i> <i>30 as oligo from b pel</i>
TNfn6-8	28,502	1.08			<b>22</b> <i>23 in peak 1</i>
TNfbg	26,130	2.00		7.9	<b>28</b> <i>10 after mono Q</i>
FNfn10	9,928	1.41			<b>20</b>
FNfn7-10	39,902	1.05			<b>25</b> 24
HxB.egf	6x 66,243	0.34	9.8		
HxB.S	6x 168,967	0.99	10.7		
HxB.L	6x 238,392	1.07	12.5		

The numbers in the yield column indicate the mg of purified protein (determined from the extinction coefficient) per L of bacterial culture. The top number indicates the yield determined in our original study, the bottom number is the yield obtained in a more recent purification.

**Table A. PCR primers for cloning TN and FN expression proteins**

<b>Recombinant protein name</b>	<b>template <sup>1</sup></b>	<b>Forward Primer (and translation) Reverse Primer (sense strand, and translation) <sup>2</sup></b>
TNegf1-14 <i>Note 3</i>	(pBS-HxB.L)	Fwd: AAGGGATCC <u>CATATGTCTGAGCCCGAATGTCCAGGC</u> M S E P E C P G Rev: GGGCTACAGCGGAGAAGACTGCTAGGATCCGTC G Y S G E D C *
TNfn1-5	(P31)	Fwd: GCGGGATCC <u>CATATGTCAGAGGTGTCTCCTCCC</u> M S E V S P P Rev: CCGCACATG <u>TGAAGGCATCCACTTAGGATCCCC</u> A H V K A S T * <i>Note 4</i>
TNfn3	(P31)	Fwd: AAGAGGATCC <u>CATATGACACGCTTGGATGCC</u> M T R L D A P Rev: AGAGACCTTCACAACATAGGGATCCGCTCCC E T F T T *
TNfnA-D(-C)	(cDNA)	Fwd: TGAGGATCC <u>CATATGGAACAAGCCCCTGAGCTG</u> M E Q A P E L Rev: AGTGCTATAGCAACAACATAGGGATCCCTCCCCA S A I A T T *
TNfn6-8	(P31)	Fwd: CTAGGATCC <u>CATATGGCCATGGGCTCCCCAAAGG</u> M A N G S P K E Rev: CCAGACCATCTTCACCACATAGGATCCCCT Q T I F T T *
TNfbg	(cDNA)	Fwd: CCAGGATCC <u>CATATGATTGGACTCCTGTACCCCTTCCC</u> M I G L L Y P F P Rev: GGCAGGCGT <u>AAGCGGGCATAAGGATCCGGG</u> G R R K R A * <i>Note 4</i>
*****	*****	*****
FNfn7-10	(pFH60)	Fwd: ACAGGATCC <u>CATATGCCATTGTCTCCACCAACAAC</u> M P L S P P T N Rev: TCCATTAATTACCGAACATAGGATCCCAA S I N Y R T *
FNfn10	(pFH60)	Fwd: GCCGGATCC <u>CATATGGTTTCTGATGTTCCGAGG</u> M V S D V P R Rev: TCCATTAATTACCGAACATAGGATCCCAA S I N Y R T *
*****	*****	*****
TNfn1-3	---	Spliced from TNfn1-5 and TNfn3

TNfn3-5	---	Spliced from TNfn1-5 and TNfn3
TNfnA-D	---	Spliced from TNfnA-D(-C) and pBS-HxB.L
TNfnALL	---	Spliced from TNfn1-5, TNfn6-8, and pBS-HxB.L
TNfn1-8	---	Spliced from TNfn1-5, TNfn6-8, and pBS-HxB.S

**Notes to Table A:**

1. The templates were: cDNA prepared from U-251MG cells; plasmid p31 provided by Dr. K. Stefansson (57); pBS-HxB.L, constructed by us from clones provided by Dr. L. Zardi (4); phage pFH60 provided by Dr. K. Yamada (28).

2. The reverse primer is listed as the sense strand (with translation), but the anti-sense strand was the sequence actually used. BamHI restriction sites are indicated by single underline; NdeI sites are double underlined.

3. The expression protein TNeg1-14 has not yet been successfully produced. When expressed as a cytoplasmic protein it seemed to be quickly degraded. We also attempted to produce it as a secreted protein, using the vector pET-12 (from Novagen), and this was also unsuccessful.

4. Our earlier primers were based on sequence data that were incomplete or in one case incorrect in one base (our final sequence is based on ref 4, with corrections noted in ref 60). Errors in our primers are indicated by bases underlined and in italics. The PCR was apparently not affected by these small errors.