N-Terminal Domain Linkage Modulates the Folding Properties of Protein S Epidermal Growth Factor-like Modules†

Nyoman D. Kurniawan,‡ Joanne M. O'Leary,‡ Ann-Marie Thamlitz,§ Raphael Sofair,|| Jörn M. Werner,|| Johan Stenflo,§ and A. Kristina Downing*,‡

Division of Structural Biology, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, United Kingdom, Department of Clinical Chemistry, University of Lund, University Hospital, S-205 02 Malmö, Sweden, and School of Biological Sciences, University of Southampton, Bassett Crescent East, Southampton SO16 7PX, United Kingdom

Received April 20, 2004; Revised Manuscript Received May 24, 2004

ABSTRACT: Protein S interacts with activated protein C to play a crucial role in blood anticoagulation, and protein S deficiency is associated with increased risk of thrombosis. Despite the large volume of functional data available for this protein, no atomic resolution structure data have yet been reported. This is due at least in part to difficulties encountered when trying to produce fragments dissected from the intact protein; however, a few successful strategies have been described. In this research we have expressed a number of constructs containing protein S epidermal growth factor-like (EGF) domains 1 and 2 in *Escherichia coli* and *Pichia pastoris*. None of the proteins produced was stably folded as assayed by solution nuclear magnetic resonance spectroscopy. We therefore constructed a series of non-native protein S EGF concatemers to investigate the role of pairwise domain linkage in domain folding. Our results demonstrate that N-terminal domain linkage can either positively or negatively impact on the refolding of an adjacent domain. Furthermore, analysis of the NMR data for EGF3-4 reveals the expected interdomain NOEs that are characteristic of an extended arrangement of calcium-binding EGF domains and a similar average $[^1H]-^{15}N$ heteronuclear NOE value for each of the two domains. These results provide the first data in support of protein S EGF3-4 adopting the same extended domain orientation as observed for the functionally distinct proteins fibrillin-1 and the low-density lipoprotein receptor. The results also have important implications for future studies, particularly when a dissection approach is used, of tandem EGF domains from protein S and other proteins.

Protein S is a vitamin K-dependent plasma anticoagulant protein (*1*) that inhibits blood clotting by acting as a cofactor to activated protein C (APC)[1] in the degradation of factors Va and VIIIa (reviewed in ref *2*). Protein S has also recently been shown to stimulate the phagocytosis of apoptotic cells (*3*). Protein S exists in two forms in plasma: the free (approximately 40%), functionally active form and in complex with C4b-binding protein (C4BP, approximately 60%) (*4*, *5*). The complex has no APC cofactor activity. Deficiency of active protein S in the plasma is associated with increased risk of thrombotic disease (*6*).

Protein S is a modular protein comprising an N-terminal γ-carboxyglutamic acid-rich (Gla) domain, a thrombin-sensitive region (TSR), four epidermal growth factor-like (EGF) domains, and a sex hormone-binding globulin-like (SHBG-like) domain (*7*). As illustrated in Figure 1A, specific regions of protein S have been associated with distinct functional properties. The Gla domain, rich in carboxyglutamic acid residues, contains several calcium binding sites. Calcium binding to this domain stabilizes it in a conformation favorable for membrane binding (*8*, *9*). The thrombin-sensitive region and EGF1 interact with activated protein C (*10*) and also help to maintain the Gla domain in the correct orientation for phospholipid binding. Three of the EGF domains (2−4) contain a consensus sequence associated with calcium binding (*11*, *12*). For some time it has been known that the first EGF domain is also important for activated protein C cofactor function (*10*, *13*, *14*) and that the fourth EGF domain, containing the highest affinity calcium binding site of the EGF domains, enhances the interaction between protein S and APC (*13*). More recently it has been shown that protein S domain deletion or domain swap mutants lacking EGF2 also manifest dramatically reduced APC

† This work was supported by the Wellcome Trust (Grants 057725/Z/99/Z/CH/MB and 055640/Z/98/Z), the Swedish Medical Research Council (K2001-03X-04487-27A), and the Swedish Foundation for Strategic Research.

* Corresponding author: phone +44 (0) 1865 285322; fax +44 (0) 1865 285323; e-mail kristy.downing@bioch.ox.ac.uk.

‡ University of Oxford.

§ University Hospital, Malmö.

|| University of Southampton.

[1] Abbreviations: 2YT, 2× yeast−tryptone; APC, activated protein C; C4BP, C4b-binding protein; cb, calcium binding; cDNA, complimentary deoxyribonucleic acid; CHAPS, 3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate; dNTPs, deoxynucleotide triphosphates; DTT, dithio-1,4-threitol; EDTA, ethylenediaminetetraacetic acid; EGF, epidermal growth factor-like; Gla, γ-carboxyglutamic acid-rich; HSQC, heteronuclear single quantum correlation spectroscopy; IPTG, isopropyl β-D-thiogalactopyranoside; M9 medium, minimal medium; MES, 2-[*N*-morpholino]ethanesulfonic acid; Ni−NTA, nickel−nitrilotriacetic acid; NMR, nuclear magnetic resonance; NOESY, nuclear Overhauser effect spectroscopy; PCR, polymerase chain reaction; RPC, reverse-phase chromatography; SHBG-like, sex hormone-binding globulin-like; TSR, thrombin-sensitive region; TFA, trifluoroacetic acid; TOCSY, total correlation spectroscopy.

10.1021/bi0492105 CCC: $27.50
Published on Web 06/29/2004

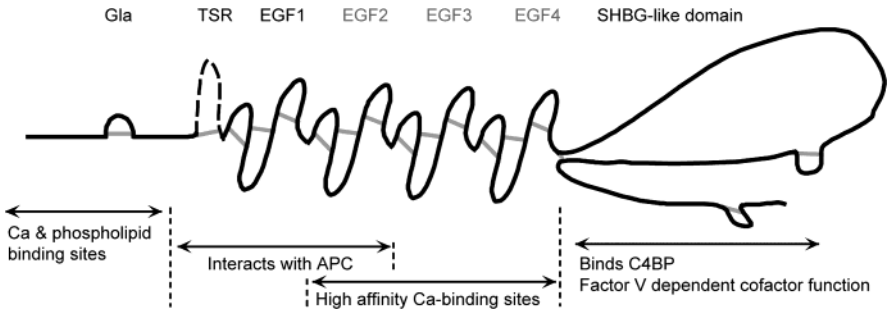


FIGURE 1: Modular structure of protein S: schematic diagram of the organization and description of the functional properties of protein S domains. Disulfide bonds are shown as gray bands.

cofactor activity (*15*). Last, the ability of protein S to complex with C4BP is regulated by the SHBG-like domain (*16*−*18*), and it has been suggested that this region is also important for factor V binding (*19*).

Despite the functional dissection of protein S, no atomic resolution structural data are available for this protein, although the structures of Gla, EGF, and SHBG-like modules have been solved for homologous proteins, and homology models have been constructed for the Gla-TSR-EGF1 (*20*) and the SHBG-like regions (*21*). We are particularly interested in the conformation of the protein S EGF domains, because of their functional importance in APC cofactor activity, and the common association of mutations to tandemly repeated calcium binding (cb) EGFs with human diseases including protein S deficiency (*22*). We have previously reported the structures of three tandem cbEGF pairs from fibrillin-1 (*23*, *24*) and the low-density lipoprotein receptor (*25*). These structures and the sequence similarity allowed us to predict that the domain arrangements of protein S EGF domains 2-3 and 3-4 are linear (*25*). However, the arrangement of EGF domains 1 and 2 cannot be predicted on the basis of sequence alone because of an additional residue in the linker between the two modules (*23*). Hence in this research we sought to recombinantly express a construct containing protein S EGF domains 1 and 2 in a system suited to cost-effective isotopic enrichment with nuclear magnetic resonance (NMR) sensitive isotopes, as a prerequisite for structural studies, and to validate the proposed model for a rodlike arrangement of domains 3 and 4.

Here we describe the recombinant expression and characterization of protein S constructs containing EGF domains 1 and 2 in the organisms *Escherichia coli* and *Pichia pastoris*. Despite testing of a large number of constructs containing different numbers of domains flanking EGF1-2 and the use of a range of different in vitro refolding strategies for the material produced in *E. coli*, no construct expressed from either system and purified to homogeneity produced NMR spectra characteristic of folded protein. On the basis of our earlier observations on the importance of pairwise domain linkage for calcium-binding EGF domain folding (*26*, *27*), we then sought to characterize the influence of pairwise domain interactions on the in vitro refolding of protein S EGF domains 1 and 2 in comparison to wild-type EGF3-4. Our results show that the folding properties of protein S EGF domains can be profoundly influenced by pairwise domain interactions. The data also provide important experimental support for an extended arrangement of cbEGF domains 3-4. We discuss the implications of this research for future studies of tandem EGF domains from protein S and other homologous proteins.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification from E. coli. Human protein S cDNA and pET16b-EGF3-4 vector have been described before (*7*, *28*). For cloning of native proteins**,** EGF1-2, EGF1-2-3, EGF1-2-3-4, TSR-EGF1-2, and Gla-TSR-EGF1-2-3 domains, DNA sequences were amplified by polymerase chain reaction (PCR) from the human protein S cDNA. The PCR reactions contained 10 pmol of forward/reverse primers, ∼100 ng of DNA template, 0.1 mM dNTPs, and 2 units of *Pfu* polymerase (Promega). The following cycles were used: a denaturation step of 94 °C for 5 min, followed by 25 cycles of amplification (94 °C for 30 s, 40 °C for 1 min, 72 °C for 1 min). PCR products were purified by use of a 1.5% agarose gel and Qiaquick Gel Extraction kit (Qiagen). The PCR products were A-tailed and TA cloned into a pGEM-3Zf(−)-based T-vector (Promega). Correct sequences were confirmed by automated DNA sequencing (Applied Biosystems). EGF1-2, EGF1-2-3, EGF1-2-3-4, and TSR-EGF1-2 were directionally subcloned into the pET16B vector (Novagen) at the *Nde*I/*Bam*HI sites downstream of the built-in decahistidine tag and factor Xa (FXa) recognition site. The Gla-TSR-EGF1-2-3 fragment was made by use of a sense PCR primer that contained a FXa recognition site and was directionally subcloned into the pQE30 vector (Qiagen) at the *Bam*HI/*Hin*dIII sites, downstream of the built-in hexahistidine tag.

For cloning of the domain swap mutants, EGF3-2, EGF3-1, EGF1-4, and EGF2-4 were made by PCR overlap extension methods (*29*). Briefly, first-stage PCR reactions were used to make single-domain fragments with overlapping DNA, in which reactions were carried out and purification was performed as described above. The swapped domain pair constructs were made in the second-stage PCR reactions, in which the purified PCR products from the previous reaction were mixed and amplified. This reaction contained ∼100 ng of PCR products, 10 pmol of extreme 5′ and 3′ primers, and 1 unit of Go-Taq polymerase mixture (Qbiogene) and used the following cycles: a denaturation step of 94 °C for 10 min, followed by 10 cycles of amplification (94 °C for 1 min, 40 °C for 1 min, 72 °C for 1 min). The final products were TA cloned, DNA sequenced, and subcloned into the pET16B vector as described above.

pET16B-EGF1-2, EGF1-2-3, EGF1-2-3-4, and TSR-EGF1-2 plasmids were transformed into *E. coli* BL21(DE3) pLysS cells (Novagen) and were grown in the presence of

100 μg/mL ampicillin and 25 μg/mL chloramphenicol. The pQE30-Gla-TSR-EGF1-2-3 plasmid was transformed into *E. coli* BL21 cells, which were pre-transformed with the repressor plasmid pREP4 (Qiagen), and grown in the presence of 100 μg/mL ampicillin and 25 μg/mL kanamycin. Cells were grown in 2YT or M9 minimal medium (without amino acids, supplemented with 1 g of $^{15}NH_4Cl$/L as nitrogen source) at 37 °C and induced with 1 mM IPTG at OD_{600} of 0.8 for 5 h. The cells were harvested by centrifugation (5000 rpm in Beckman JA-10 rotor for 25 min at 4 °C) and solubilized with denaturing lysis buffer (6 M guanidine hydrochloride, 20 mM Tris, pH 7.5, and 5 mM imidazole). Lysed cells were sonicated and cell debris was removed by ultracentrifugation (Beckman T-55 fixed angle rotor at 40 000 rpm for 30 min at room temperature). Protein supernatants were loaded onto a Ni−NTA column (Sigma), pre-equilibrated in the lysis buffer, washed with 20 column volumes of lysis buffer, and eluted with 5 column volumes of 67% denaturing buffer containing 300 mM imidazole.

Crude peptides were reduced with 0.1 M DTT (pH 8.5) overnight at room temperature, acidified with neat TFA, and dialyzed against 0.1% (v/v) TFA/H_2O prior to reverse-phase chromatography (RPC). Samples were run on a preparative Poros 10 R2 column [10 mm (D) × 100 mm (L), Applied Biosystems], with a 2−60% acetonitrile/TFA (80/0.1% v/v) gradient over 15 column volumes at 10 mL/min.

Protein refolding was performed by previously described methods (*30*) with modifications. Lyophilized, reduced peptides were dialyzed against refolding buffer (3.0/0.3 mM cysteine/cystine, 150 mM NaCl, 5 mM $CaCl_2$, and 20 mM Tris, pH 8.5). The peptide concentrations were ∼0.1 mg/mL and the refolding reactions were performed over 24 h at 4 °C. Alternative refolding conditions were also tested for EGF1-2 by use of 5.0/0.5 mM reduced/oxidized glutathione (Sigma) with 50 mM $CaCl_2$, 150 mM NaCl, and 20 mM Tris, pH 8.5, in the absence/presence of 4 mM Cu^{2+} as an oxidation catalyst, or mild denaturing agents (0.5 M arginine or 2 M guanidine hydrochloride), or 20 mM detergent CHAPS (Sigma) at 4 °C. Subsequently, a second preparative RPC was performed and the samples were lyophilized prior to $His_{10/6}$-tag cleavage. Except for the TSR-containing samples, refolded peptides (∼2 mg/mL) were cleaved with bovine FXa (Promega) (1 μg/mg of protein) in a buffer containing 100 mM Tris, pH 7.5, 1 mM $CaCl_2$, and 100 mM NaCl at 22−25 °C overnight. Cleaved protein was dialyzed against 20 mM MES buffer, pH 5.0, and purified on a cation-exchange Poros HS/20 column [4.6 mm (D) × 100 mm (L), Applied Biosystems] with a 0−60% gradient of 3 M NaCl in 20 mM MES, pH 5.0, over 30 column volumes at 5 mL/min. This purification was followed by a polishing step on an analytical Poros 10 R2 column [4.6 mm (D) × 100 mm (L), Applied Biosystems] with 2−60% acetonitrile/TFA (80/0.1% v/v) gradient over 15 column volumes at 3 mL/min. Eluted samples were freeze-dried prior to NMR. Typically, 2−4 mg of purified protein was obtained per liter of culture. The in vitro refolding properties of each of the peptides EGF3-1, EGF3-2, EGF1-4, EGF2-4, and EGF3-4 were assayed at least twice to confirm reproducibility of the results. The constructs were confirmed by DNA sequencing and Western blots with TSR, EGF1, and EGF3 antibodies, and the correct molecular weight of each peptide was confirmed by mass spectrometry.

NMR Analyses of Protein Folding and Dynamics. All NMR data were acquired on 500, 600, or 750 MHz GE/home-built spectrometers. Native proteins EGF1-2, EGF1-2-3, EGF1-2-3-4, His-tag-TSR-EGF1-2, His-tag-Gla-EGF1-2-3, *P. pastoris*-produced EGF1-2-3 (*p*EGF1-2-3), EGF3-4, and mutant concatemers EGF3-1, EGF3-2, EGF1-4, and EGF2-4 samples (∼0.1−0.4 mM) were initially assayed for spectral dispersion in the presence of 20 mM $CaCl_2$ and 90% H_2O/10% D_2O at pH 6.3−6.5 at 37 °C. One-dimensional spectra were acquired with 2048 complex points and a spectral width of 10 000 Hz at 600 MHz with gradients for water suppression. The starting-point condition above gave clear (soluble) NMR samples. However, many of the screened samples had poor spectral dispersion and the signals appeared broad. Thus, the microdrop method (*31*, *32*) with various buffering agents and additives was employed in an attempt to identify better sample conditions. The conditions tested included increasing $CaCl_2$ concentrations (up to 80 mM), additions of 50 or 150 mM NaCl, 3 or 5 mM MES, pH 5.0, 2 or 20 mM CHAPS, and 5 or 10 mM DTT and lowering temperatures to 32 and 25 °C.

Proteins with folded characteristics, i.e., showing dispersed amide chemical shifts and upfield-shifted methyl protons, were chosen for ^{15}N-labeling. These samples included EGF3-4 (wild type) and the mutant concatemers EGF3-2 and EGF3-1. Folding and calcium-dependent chemical shift changes of these samples (∼0.5 mM protein, 90% H_2O/10% D_2O, pH 6.3) were assayed by $^{15}N-^{1}H$ HSQC (*33*) spectroscopy. Briefly, calcium-free conditions were identified by titrating the samples with EDTA until no chemical shift changes were observed. Calcium saturation was established by titration, and the final calcium concentration for each sample was in the range of 22−25 mM. HSQC data were acquired with 1K complex data points and 128 or 256 increments in the indirect dimension, yielding acquisition times of 102 and 48 ms in the ^{1}H and ^{15}N dimensions, respectively.

Two- and three-dimensional $^{1}H-^{15}N$ HSQC−NOESY (150 and 250 ms mixing times) and two- and three-dimensional $^{1}H-^{15}N$ TOCSY (75 ms DIPSI mixing times) were recorded for sequential assignment purposes. Two-dimensional spectra were typically acquired with 2K complex data points in the direct dimension with 350−700 increments in the indirect dimension, yielding acquisition times of 205 and 29−58 ms in the direct and indirect dimensions, respectively. Three-dimensional spectra were typically acquired with 1K complex data points in the direct dimension with 25−64 increments in the indirect dimensions, yielding acquisition times of 102, 16−27, and 8.3−13.8 ms in the direct and the indirect ^{15}N and ^{1}H dimensions, respectively. All spectra were processed with Felix 97 (Accelrys, San Diego, CA), typically using shifted squared sine-bell or Gaussian window functions in the acquisition dimension and shifted squared sine-bell or Kaiser window functions in the indirect dimensions. Spectra were typically zero-filled twice of the number complex points to improve resolution. Sequence-specific ^{1}H and ^{15}N chemical shift assignments were made in SPARKY 3 (T. D. Goddard and D. G. Kneller, University of California, San Francisco) by conventional methods (*34*). The chemical shifts of EGF3-4 with one calcium bound in EGF4 (*35*) (BioMagResBank accession number 4729) were used as a reference.

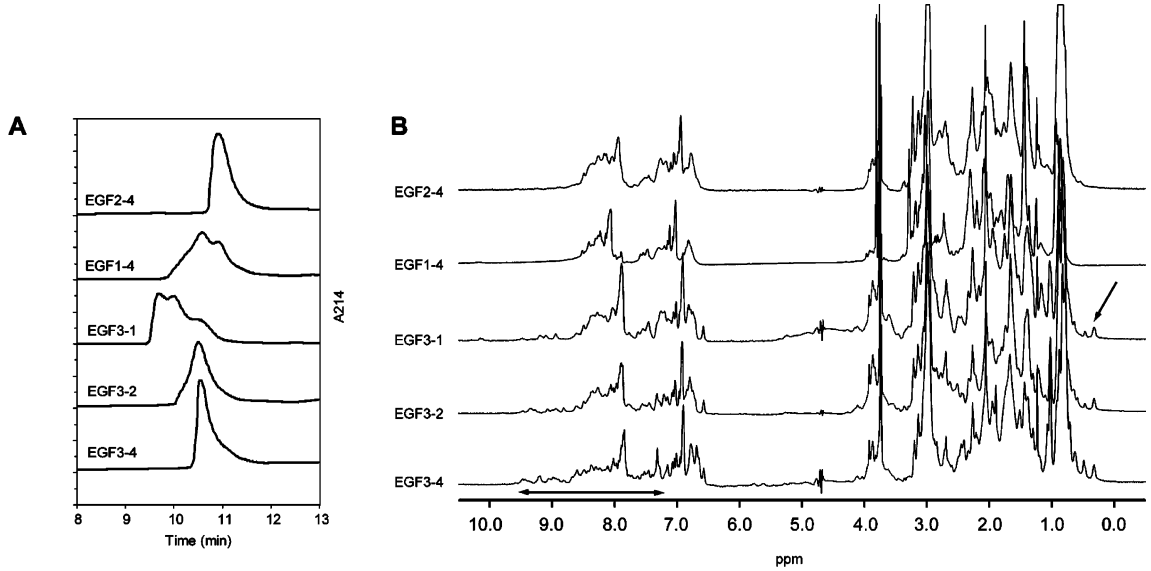


FIGURE 2: Refolding properties of native and non-native protein S EGF concatemers assayed by analytical RPC and one-dimensional NMR. (A) Analytical RPC chromatograms of refolded mutant concatemer His-tag proteins. For each construct, refolding was performed under identical conditions with the same amount of protein. The proteins were eluted by use of a 2−60% acetonitrile/TFA (80/0.1% v/v) gradient over 15 column volumes. (B) One-dimensional NMR analyses of refolded mutant concatemer His-tag proteins. All spectra were acquired in the presence of saturating 20 mM $CaCl_2$, pH 6.3 at 37 °C. Amide and methyl regions showing increased spectral dispersion for some constructs are indicated by double- and single-headed arrows, respectively.

All fingerprint chemical shifts (^{1}HN and ^{15}N) of the EGF3 domain in EGF3-4, EGF3-2, and EGF3-1 were assigned and verified by use of sequential NOEs, with the exception of residues D182 and E184, which have weak NOESY cross-peaks. Approximately 44% of the EGF2 domain fingerprint chemical shifts were assigned, including a fragment of EGF2 (T137-P138-G139) that forms the interaction region between the two domains, which was unambiguously assigned. EGF3-2 interdomain interactions were identified by medium-to-strong NOEs between the side chain of P138 and the G139 Hα protons of EGF2 to the aromatic protons of Y191 (EGF3), which are analogous pairwise domain contacts to those observed for fibrillin-1 and LDLR cbEGF domain pairs (*23*−*25*).

Relaxation data of calcium-saturated EGF3-2 and EGF3-4 were acquired at 36 °C on a Varian INOVA 600 MHz spectrometer using a 5 mm *z*-gradient triple-resonance probe as described previously (*36*). In the ^{15}N T_1 and ^{15}N T_2 experiments (*37*), the acquisition times were 64 ms (^{1}H) and 69.4 ms (^{15}N) for EGF3-2 and 64 ms (^{1}H) and 58.2 ms (^{15}N) for EGF3-4. The CPMG delay in the T_2 experiment was set to 587 μs. Relaxation delays were 20, 80, 240, 480, 700, 1000, and 1300 ms and 10, 30, 50, 70, 90, 110, and 150 ms for the T_1 and the T_2 series, respectively. T_1 and T_2 relaxation time constants were derived from two-parameter exponential fits to the intensity decays, and errors were estimated from the root-mean-square noise in the spectra.

Pairs of 2D ^{1}H-detected ^{1}H−^{15}N heteronuclear NOE experiments (*37*) were performed with acquisition times of 64 ms (^{1}H) and 45.7 ms (^{15}N) for EGF3-4 and 64 ms (^{1}H) and 69.7 ms (^{15}N) for EGF3-2. ^{1}H saturation was achieved in the experiment with NOE by a train of 120° flip-angle ^{1}H pulses at 5 ms intervals for 3 s in the case of EGF3-4 and 4 s in the case of EGF3-2. The [^{1}H]−^{15}N NOE was calculated as the ratio of the peak volumes in the experiments with and without NOE. The relaxation data were analyzed with NMRView (*38*).

Modeling of EGF3-4. The EGF3-4 structure was modeled by use of the fibrillin cbEGF32-33 structure (1EMN) as a template. Briefly, the amino acid sequence of EGF3-4 was aligned to 1EMN by use of the program Clustal_X (*39*) with the Gonnet 250 protein weight matrix and a pairwise gap opening/extension of 10.0/0.1. EGF3 residues G190 and Y191 were adjusted to align with cbEGF32 residues G2156 and Y2157, according to the conservation of the EGF core sequence (*23*). The model was calculated with the program MODELLER v 6.2 (*40*) implementing the model-default.top script with all (6) disulfide bonds explicitly included in the calculation. By use of the calculated model, the coordinates for calcium-binding oxygen ligands of each domain (for EGF3, D160, V161, E163, N178, and I179; for EGF4, D202, I203, E205, N217, and Y218) were averaged to obtain good starting coordinates for the positions of calcium ions to be included into the model. The final model was refined with CNS v 1.1 (*41*), by use of topology files protein-allhdg.top and ion.top and parameter files protein-allhdg.param and ion.parm. The calcium-containing EGF3-4 model was created with the generate.inp script, in which all hydrogen atoms of the model were rebuilt and subsequently refined in 0.25 ps of low-temperature molecular dynamics (300 K) and energy-minimized with 500 steps of Powell minimization.

RESULTS

One aim in this study was to produce a correctly folded protein S construct containing EGF1-2 in an expression system that is suitable for uniform isotopic labeling. Five constructs corresponding to EGF1-2, EGF1-2-3, EGF1-2-3-4, TSR-EGF1-2, and Gla-TSR-EGF1-2-3 were cloned, expressed in *E. coli*, and refolded in vitro under a variety of folding conditions (see Experimental Procedures). None of the proteins displayed spectral characteristics consistent with correctly folded EGF domains. On the basis of the earlier successful production of protein S constructs containing EGF domains in baculovirus (*13*, *42*), in vivo expression of EGF1-

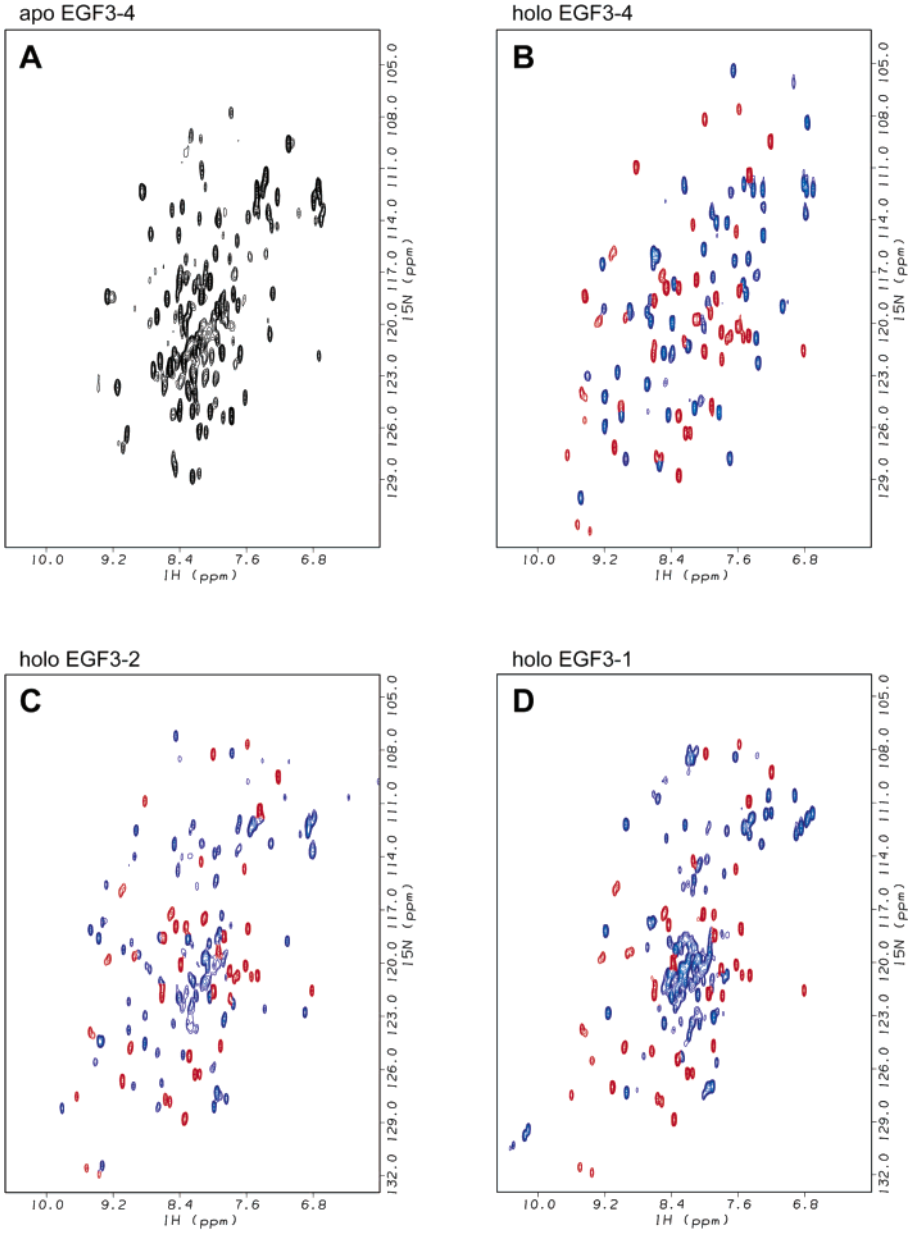


FIGURE 3: $^{1}H-^{15}N$ HSQC NMR of wild-type and mutant concatemers: (A) calcium-free EGF3-4 and calcium-saturated (B) EGF3-4, (C) EGF3-2, and (D) EGF3-1. EGF3 cross-peaks are shown in red, whereas those from EGF4, EGF2, and EGF1 are shown in blue. All spectra were acquired with ~0.5 mM sample in the presence of 100 μM EDTA (for calcium-free EGF3-4) or 20 mM $CaCl_2$, pH 6.3 at 37 °C.

2, EGF1-2-3, TSR-EGF1-2, and TSR-EGF1-2-3 was attempted in *P. pastoris*. In all cases, unreduced proteins formed oligomeric ladders, whereas reduced proteins ran as monomers of the expected molecular weight (data not shown). These observations indicated that the oligomers were formed by intermolecular disulfide bonds. Therefore, only one of these peptides, *p*EGF1-2-3, was selected for scale-up, purification, and further analysis. However, a folded component of the *p*EGF1-2-3 sample was not detectable by NMR.

Our inability to produce a sample containing correctly folded protein S EGF1-2 combined with earlier observations of the importance of pairwise domain linkage for the correct folding of a mutant cbEGF pair from fibrillin-1 (*27*) led us to test the in vitro refolding properties of non-native protein S EGF concatemers, which were produced in *E. coli*. Four constructs were designed, EGF3-1, EGF3-2, EGF1-4, and EGF2-4, to test the importance of both N- and C-terminal domain linkage. EGF3-4 was also made as a control and for characterization of the pairwise domain interface.

The refolding abilities of the wild-type and mutant concatemer proteins were assessed by analytical RPC and one-dimensional NMR. The chromatograms showed that EGF3-4 eluted as a major peak, whereas the EGF3-2 elution profile was broader than that of EGF3-4 (Figure 2A). In comparison, EGF3-1 and EGF1-4 produced multiple peaks, which may correspond to different disulfide-bonded isomers, and EGF2-4 eluted as a single broad peak. In the NMR spectra, both wild-type EGF3-4 and EGF3-2 showed good spectral dispersion in the amide region and upfield-shifted methyl protons, indicative of folded domains (Figure 2B). EGF3-1 also showed some characteristics of folded protein, although the number of shifted peaks was reduced in comparison to EGF3-4 and EGF3-2. In contrast, EGF1-4 and EGF2-4 did not appear to be folded on the basis of the NMR analysis.

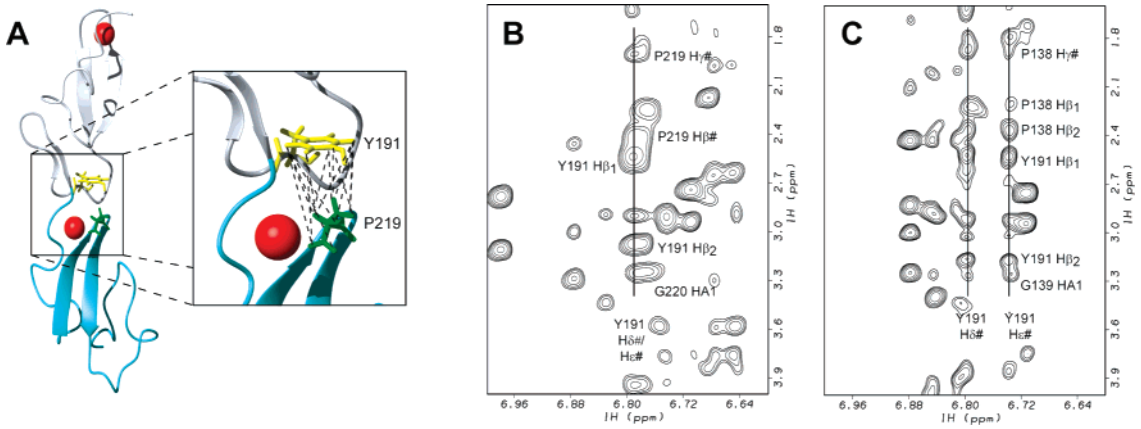


FIGURE 4: Conservation of interdomain interactions in wild-type EGF3-4 and concatemer EGF3-2. (A) EGF3-4 model structure, in which EGF3 is shown in gray and EGF4 is shown in cyan. Interdomain NOE interactions between EGF3 Y191 (yellow) and EGF4 P219 (green) are illustrated with dashed lines. (The interactions between EGF3 Y191 and EGF4 G220 are not shown in this diagram for clarity.) The interactions observed in the NOESY spectra of EGF3-4 are illustrated (B) and are conserved for the equivalent residues of EGF3-2, i.e., EGF3 Y191 and EGF2 P138 and G139 (C). The Y191 and P138 β-protons were not stereospecifically assigned.

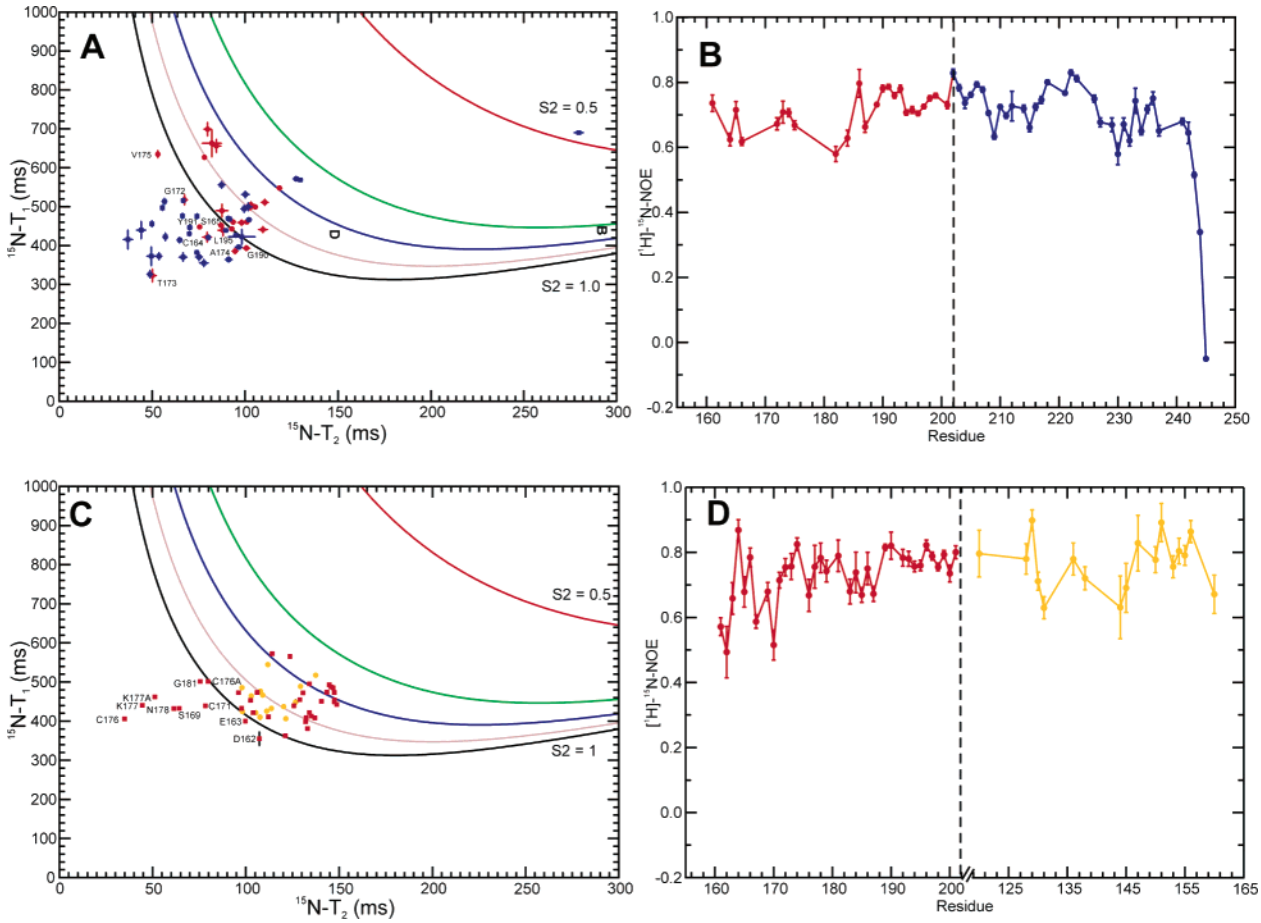
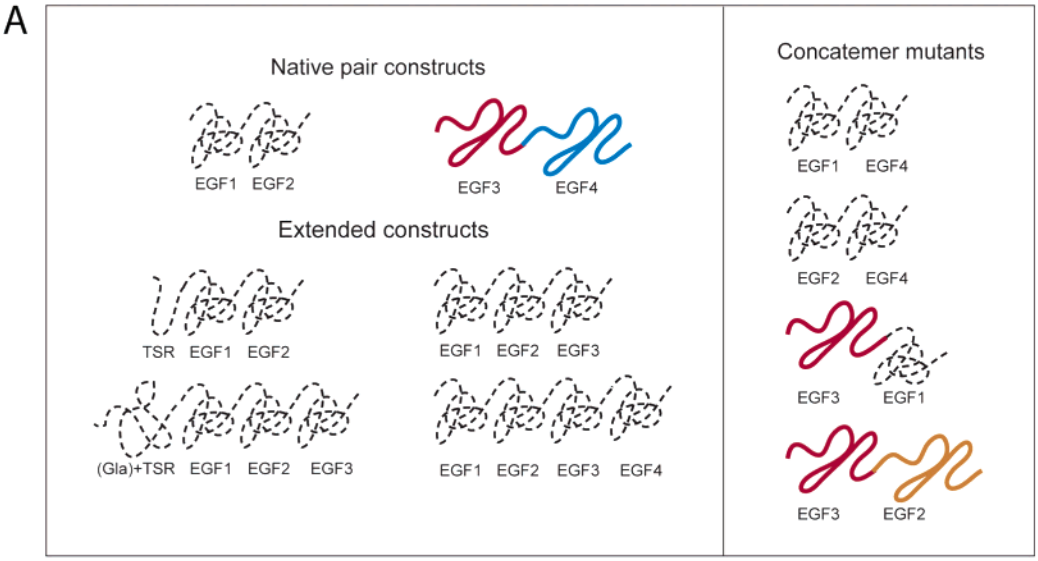


FIGURE 5: Heteronuclear relaxation data for wild-type EGF3-4 (A, B) and concatemer EGF3-2 (C, D). The data are colored according to the domains: EGF3 red, EGF4 blue, and EGF2 orange. (A) ^{15}N T_1 and T_2 values of EGF3-4 together with parametric T_1 T_2 graphs for order parameters S^2 between 1.0 and 0.5. Residues of EGF3 with low T_2 values indicative of exchange are labeled. (B) Heteronuclear ^{1}H−^{15}N NOE values of EGF3-4 along the amino acid sequence. The domain boundary between EGF3 and 4 is indicated by a vertical dashed line. (C, D) Same graphs as in panels A and B for EGF3-2. The residues are numbered according to the sequence of the native protein.

The constructs displaying chemical shift dispersion compatible with folded protein were labeled with ^{15}N and assayed for calcium-dependent chemical shift perturbations by HSQC (Figure 3). Calcium-free EGF concatemers, as represented by EGF3-4, showed reduced spectral dispersion. In the presence of calcium ions, domains EGF2, -3, and -4 of EGF3-4 and EGF3-2 concatemers showed calcium-dependent chemical shifts and improved dispersion, indicating that these domains are folded. However, only domain EGF3 but not EGF1 in the EGF3-1 concatemer showed good dispersion and calcium-dependent chemical shifts, indicating that EGF1 is not a functional calcium-binding EGF domain.

Two- and three-dimensional NOESY and TOCSY spectra were acquired and used for the sequential assignment of ^{15}N-



B

```
PrtS EGF1-2      IPDQCSP-----LPCNEDGYMSCKDGKASFTCTCKPGWQ-----GEKCEFDINECKDPSNINGGC-SQICDNTPGSYHCSCK--NGFVMLSNKKDC
PrtS EGF2-3      DINECKDPSNINGGCSQ----ICDNTPGSYHCSCKNGFVML-SNKKDCK-DVDECSLKP---SICGTAVCKNIPGDFECECP--EGYRYNLKSKSC
PrtS EGF3-4      DVDECSLK---PSICGTA---VCKNIPGDFECECPEGYRYN-LKSKSCE-DIDECSEN-----MC-AQLCVNYPGGYTCYCDGKKGFKLAQDQKSC
fib1 cbEGF12-13 DIDECRIS---PDLCGRG---QCVNTPGDFECKCDEGYESGFMMMKNCM-DIDECQRDP---LLCRGGVCHNTEGSYRCECP--PGHQLSPNISAC
fib1 cbEGF32-33 DMDECKE----PDVCKHG---QCINTDGSYRCECPFGYT---LAGNECV-DTDECSVG----NPCGNGTCKNVIGGFECTCE--EGFEPGP-MMTC
LDLR EGF-AB      GTNECLD---NNGGCSH----VCNDLKIGYECLCPDGFQL--VAQRRCE-DIDECQDP----DTC-SQLCVNLEGGYKCQCE--EGFQLDPHTKAC
```

FIGURE 6: Summary for the production of in vitro refolded protein S EGF domains (A) and comparison of EGF pair sequences (B). (A) N- and C-terminal extensions of EGF1-2 (left panel) failed to provide stable templates for chaperoning the refolding process. EGF3-4 refolded successfully under the same conditions. For the concatemer mutants (right panel), EGF3 at the N-terminus could aid the refolding of EGF2 but not EGF1. However, EGF4 at the C-terminus failed to recover the refolding of EGF1 and EGF2. In both native and mutant concatemers, the misfolding of EGF1 and EGF2 negatively influenced the folding of domains EGF3 and EGF4. Misfolded and folded domains are drawn with dashed and solid lines, respectively. The domains are color-coded as in Figure 5. (B) Multiple sequence alignment of EGF pairs from protein S and sequence homologues of known structure, fibrillin-1 cbEGF12-13 (*24*) and cbEGF32-33 (*23*) and the low-density lipoprotein receptor EGF-AB pair (*25*). Amino acid residues in the calcium binding consensus sequence are shaded cyan, and cysteines involved in disulfide bonding are in yellow. Residues at key positions that have been observed to engage in pairwise domain packing interactions in EGF3-4, fibrillin-1, cbEGF12-13, and cbEGF32-33 and the low-density lipoprotein receptor EGF-AB pair are colored orange. The interdomain linker residues are highlighted by white type against a black background.

labeled EGF3-4 and EGF3-2. The assignments permitted identification of pairwise domain NOE connectivities homologous to those observed for fibrillin-1 and LDLR EGF domain pairs (Figure 4). The presence of these NOEs supports the hypothesis that the protein S EGF3-4 pair also adopts an extended conformation, as illustrated by the homology model in Figure 6. However, the EGF3-4 pair is known to be flexible on multiple time scales (*43*), and therefore this structure may represent one populated conformer of a dynamic ensemble on the time scale of the NOESY experiment.

T_1, T_2, and heteronuclear NOE relaxation data for EGF3-4 and EGF3-2 are shown for comparison in Figure 5. For both constructs, the T_1 versus T_2 values cluster mainly around a single region of the plot, suggesting that both domains tumble with similar correlation times. Many residues in EGF3-4 fall outside of the Lipari−Szabo curves; an estimate of the rotational correlation time for this construct based on these data is 7.3 ns, versus 4.9 ns for EGF3-2. Therefore, it seems likely that the EGF3-4 data have a general line width contribution from aggregation and that the data do not necessarily reflect increased slow time scale motions for EGF3-4 relative to EGF3-2. Notably, there is a good correlation between residues with R_{ex} terms > 2 Hz for EGF3-4 in the $(Ca^{2+})_1$ form (*43*) and residues with high R_2 values in the $(Ca)_2$ form. $[^1H]-^{15}N$ heteronuclear NOE values for EGF3 are similar for the two constructs. In addition, for both constructs the average $[^1H]-^{15}N$ NOE value is similar for each of the two domains and across the interdomain linker region, also supporting a well-ordered interface.

DISCUSSION

The first and second EGF domains from protein S are an attractive structural target because of their importance in protein S interaction with APC and their homology to a large number of EGF domain pairs with two amino acid linkers (*23*). Although the literature suggested that this target might be difficult to produce for NMR study, some successes with chemical synthesis and constructs produced using baculovirus led us to attempt expression of a number of constructs in *E. coli* and *P. pastoris*. Of these, none was folded as judged by NMR analysis of calcium-dependent chemical shift perturbation and dispersion.

We have previously characterized the folding properties of two overlapping cbEGF domain pairs from human fibrillin-1 containing a G1127S point mutation (*26*, *27*). The results demonstrated that when the mutation was localized to the N-terminal domain, only the C-terminal domain refolded. However, when the mutation was in the C-terminal domain, both domains refolded and had native-like properties. These results suggested that pairwise domain interactions of EGF domains can be important determinants of folding. We therefore designed a series of EGF concatemers from protein S in order to analyze the importance of pairwise domain linkages for in vitro refolding. The domain pair constructs that were produced in *E. coli* included EGF3-1,

EGF3-2, EGF1-4, EGF2-4, and EGF3-4 as a control. Correct folding of domains was assessed by RPC, calcium-dependent NMR chemical shift perturbation, and analysis of NOESY data (Figures 2−4). The results of these and the other in vitro refolding experiments described in this paper are illustrated in Figure 6A.

Two important conclusions about the role of pairwise domain interactions in protein S EGF domain folding may be drawn from the results. The folding of EGF2 can be rescued by N-terminal domain linkage of EGF3 but not by C-terminal domain linkage of EGF4. Therefore, N-terminal domain linkage can have a positive effect on domain folding. However, because pairwise domain linkage has no detectable effect on the refolding ability of EGF1, the ability of an N-terminally linked domain to rescue is, not surprisingly, sequence-dependent.

Comparison of the amino acid sequences of tandem pairs of EGF domains shown in Figure 6B reveals differences that may be important in determining the refolding properties of EGF1. Most notably, EGF1 only contains a partial calcium binding consensus sequence, and calcium binding to this domain was not detected by chromophoric chelators (*44*). EGF1 also contains more residues between Cys 2 and 3, and fewer residues between Cys 5 and 6, than the other EGF domains. The conserved aromatic residue four positions after cysteine 5 is a Trp in EGF1, while typically Tyr or Phe is observed at this position (*22*), which may affect pairwise domain packing interactions when EGF1 is the N-terminal domain. On the basis of the results of Hackeng et al. (*45*), showing that a chemically synthesized construct containing Gla-TSR-EGF1 refolds spontaneously in the absence of calcium, it is also possible that EGF1 requires N-terminally linked Gla and TSR in order to refold in vitro. While the Gla domain was included in some of the constructs tested in this study, *E. coli* lacks the ability to modify Glu residues to Gla, which might explain why these peptides did not refold in vitro.

The second major conclusion to be drawn from the results is based on the observation that neither domain of the EGF1-4 or EGF2-4 construct refolded in vitro. The fact that EGF4 has been successfully refolded under similar conditions (*46*) suggests that N-terminal linkage of EGF1 or EGF2 has a negative effect on folding. So while EGF domains can exist as independently folding protein subunits, pairwise domain interactions of tandem repeats can have either a stabilizing or destabilizing effect on the refolding properties of the adjacent domain. These sorts of effects are likely to be of particular importance with respect to disease-associated mutations affecting tandem repeats of EGF domains.

Although the chemical shifts and secondary structure of EGF3-4 have been reported previously (*35*), our data also provide the first experimental evidence for a similar tertiary structure of EGF3-4 to other structures of cbEGF domain pairs from fibrillin-1 and the LDLR. Specifically, the expected pattern of interdomain NOEs was assigned, which has enabled the generation of the homology model of the EGF3-4 pair shown in Figure 6A. Confidence in the model is further provided by the analysis of heteronuclear relaxation data for EGF3-4 and small-angle X-ray scattering data for EGF1-2-3-4 and EGF2-3-4 that show clearly that the two constructs are elongated with end-to-end distances matching very well the expected distances for four and three linearly arranged EGFs (T. Drakenberg, personal communication). Finally, despite the difficulties encountered in expression of multi-EGF domain fragments from protein S in this research, the engineered EGF3-2 concatemer may be utilized to facilitate the folding and structural studies of the EGF2-3-4 triple.

ACKNOWLEDGMENT

A.K.D. is a Wellcome Trust Senior Research Fellow. We thank Torbjörn Drakenberg, Andreas Muranyi, and Björn Dahlbäck for critical reading of the manuscript and discussion.

SUPPORTING INFORMATION AVAILABLE

Four tables showing boundaries for the *E. coli* EGF1-2, EGF1-2-3, EGF1-2-3-4, TSR-EGF1-2, and Gla-TSR-EGF1-2-3 constructs, primers and templates used for PCR mutagenesis for the domain swap mutants EGF3-2, EGF3-1, EGF1-4, and EGF2-4, analytical RPC profiles for His-tag-EGF1-2 refolding under six different conditions, and summary of conditions and results of the microdrop screen to improve protein solubility; two figures showing representative 1-dimensional NMR spectra for constructs produced in *E. coli* and refolded in vitro and analyses of in vivo refolded proteins S domains produced by *Pichia pastoris*; and experimental procedures for protein expression and purification from *P. pastoris*. This information is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. DiScipio, R. G., Hermodson, M. A., Yates, S. G., and Davie, E. W. (1977) A comparison of human prothrombin, factor IX (Christmas factor), factor X (Stuart factor), and protein S, *Biochemistry 16*, 698−706.
2. Esmon, C. T. (2000) Regulation of blood coagulation, *Biochim. Biophys. Acta−Protein Struct. Mol. Enzymol. 1477*, 349−360.
3. Anderson, H. A., Maylock, C. A., Williams, J. A., Paweletz, C. P., Shu, H. J., and Shacter, E. (2003) Serum-derived protein S binds to phosphatidylserine and stimulates the phagocytosis of apoptotic cells, *Nat. Immunol. 4*, 87−91.
4. Dahlback, B., and Stenflo, J. (1981) High molecular-weight complex in human-plasma between vitamin-K-dependent protein-S and complement component C4b-binding protein, *Proc. Natl. Acad. Sci. U.S.A. 78*, 2512−2516.
5. Rezende, S. M., Simmonds, R. E., and Lane, D. A. (2004) Coagulation, inflammation, and apoptosis: different roles for protein S and the protein S−C4b binding protein complex, *Blood 103*, 1192−1201.
6. Makris, M., Leach, M., Beauchamp, N. J., Daly, M. E., Cooper, P. C., Hampton, K. K., Bayliss, P., Peake, I. R., Miller, G. J., and Preston, F. E. (2000) Genetic analysis, phenotypic diagnosis, and risk of venous thrombosis in families with inherited deficiencies of protein S, *Blood 95*, 1935−1941.
7. Lundwall, A., Dackowski, W., Cohen, E., Shaffer, M., Mahr, A., Dahlback, B., Stenflo, J., and Wydro, R. (1986) Isolation and sequence of the cDNA for human protein S, a regulator of blood-coagulation, *Proc. Natl. Acad. Sci. U.S.A. 83*, 6716−6720.
8. Burnier, J. P., Borowski, M., Furie, B. C., and Furie, B. (1981) Gamma-carboxyglutamic acid, *Mol. Cell. Biochem. 39*, 191−207.
9. Sunnerhagen, M., Forsen, S., Hoffren, A. M., Drakenberg, T., Teleman, O., and Stenflo, J. (1995) Structure of the Ca^{2+}-free Gla domain sheds light on membrane-binding of blood-coagulation proteins, *Nat. Struct. Biol. 2*, 504−509.
10. He, X. H., Shen, L., Villoutreix, B. O., and Dahlback, B. (1998) Amino acid residues in thrombin-sensitive region and first epidermal growth factor domain of vitamin K-dependent protein S determining specificity of the activated protein C cofactor function, *J. Biol. Chem. 273*, 27449−27458.

11. Rees, D. J., Jones, I. M., Handford, P. A., Walter, S. J., Esnouf, M. P., Smith, K. J., and Brownlee, G. G. (1988) The role of β-hydroxyaspartate and adjacent carboxylate residues in the first EGF domain of human factor IX, *EMBO J. 7*, 2053−2061.
12. Handford, P. A., Baron, M., Mayhew, M., Willis, A., Beesley, T., Brownlee, G. G., and Campbell, I. D. (1990) The first EGF-like domain from human factor IX contains a high-affinity calcium binding site, *EMBO J. 9*, 475−480.
13. Stenberg, Y., Drakenberg, T., Dahlback, B., and Stenflo, J. (1998) Characterization of recombinant epidermal growth factor (EGF)-like modules from vitamin-K-dependent protein S expressed in *Spodoptera* cells−the cofactor activity depends on the N-terminal EGF module in human protein S, *Eur. J. Biochem. 251*, 558−64.
14. Hackeng, T. M., Yegneswaran, S., Johnson, A. E., and Griffin, J. H. (1999) Conformational changes in activated protein C caused by binding of the EGF1 domain of protein S, *Thromb. Haemostasis*, 1355.
15. Mille-Baker, B., Rezende, S. M., Simmonds, R. E., Mason, P. J., Lane, D. A., and Laffan, M. A. (2003) Deletion or replacement of the second EGF-like domain of protein S results in loss of APC cofactor activity, *Blood 101*, 1416−1418.
16. Fernandez, J. A., Heeb, M. J., and Griffin, J. H. (1993) Identification of residues 413−433 of plasma protein S as essential for binding to C4b-binding protein, *J. Biol. Chem. 268*, 16788−16794.
17. Walker, F. J. (1989) Characterization of a synthetic peptide that inhibits the interaction between protein S and C4b-binding protein, *J. Biol. Chem. 264*, 17645−17648.
18. Evenas, P., de Frutos, P. G., Linse, S., and Dahlback, B. (1999) Both G-type domains of protein S are required for the high-affinity interaction with C4b-binding protein, *Eur. J. Biochem. 266*, 935−942.
19. Hackeng, T. M., van't Veer, C., Meijers, J. C., and Bouma, B. N. (1994) Human protein S inhibits prothrombinase complex activity on endothelial cells and platelets via direct interactions with factors Va and Xa, *J. Biol. Chem. 269*, 21051−21058.
20. Villoutreix, B. O., Teleman, O., and Dahlback, B. (1997) A theoretical model for the Gla-TSR-EGF-1 region of the anticoagulant cofactor protein S: From biostructural pathology to species-specific cofactor activity, *J. Comput.-Aided Mol. Des. 11*, 293−304.
21. Villoutreix, B. O., Dahlback, B., Borgel, D., Gandrille, S., and Muller, Y. A. (2001) Three-dimensional model of the SHBG-like region of anticoagulant protein S: New structure−function insights, *Proteins: Struct., Funct., Genet. 43*, 203−216.
22. Boswell, E. J., Kurniawan, N., and Downing, A. K. (2004) in *Handbook of Metalloproteins* (Messerschmidt, A., Ed.) p 2268, John Wiley & Sons, Weinheim, Germany.
23. Downing, A. K., Knott, V., Werner, J. M., Cardy, C. M., and Campbell, I. D. (1996) Solution structure of a pair of calcium binding epidermal growth factor-like domains: implications for the marfan syndrome and other genetic disorders, *Cell 85*, 597−605.
24. Smallridge, R. S., Whiteman, P., Werner, J. M., Campbell, I. D., Handford, P. A., and Downing, A. K. (2003) Solution structure and dynamics of a calcium binding epidermal growth factor-like domain pair from the neonatal region of human fibrillin-1, *J. Biol. Chem. 278*, 12199−12206.
25. Saha, S., Boyd, J., Werner, J. M., Knott, V., Handford, P. A., Campbell, I. D., and Downing, A. K. (2001) Solution structure of the LDL receptor EGF-AB pair: A paradigm for the assembly of tandem calcium binding EGF domains, *Structure 9*, 451−456.
26. Whiteman, P., Downing, A. K., Smallridge, R., Winship, P. R., and Handford, P. A. (1998) A Gly → Ser change causes defective folding in vitro of calcium-binding epidermal growth factor-like domains from factor IX and fibrillin-1, *J. Biol. Chem. 273*, 7807−7813.
27. Whiteman, P., Smallridge, R. S., Knott, V., Cordle, J. J., Downing, A. K., and Handford, P. A. (2001) A G1127S change in calcium-binding epidermal growth factor-like domain 13 of human fibrillin-1 causes short range conformational effects, *J. Biol. Chem. 276*, 17156−17162.
28. Stenberg, Y., Muranyi, A., Steen, C., Thulin, E., Drakenberg, T., and Stenflo, J. (1999) EGF-like module pair 3−4 in vitamin K-dependent protein S: modulation of calcium affinity of module 4 by module 3, and interaction with factor X, *J. Mol. Biol. 293*, 653−65.
29. Ho, S. N., Hunt, H. D., Morton, R. M., Pullen, J. K., and Pease, L. R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction, *Gene 77*, 51−59.
30. Knott, V., Downing, A. K., Cardy, C. M., and Handford, P. A. (1996) Calcium binding properties of an epidermal growth factor-like domain pair from human fibrillin-1, *J. Mol. Biol. 255*, 22−27.
31. Bagby, S., Tong, K. I., Liu, D., Alattia, J. R., and Ikura, M. (1997) The button test: a small scale method using microdialysis cells for assessing protein solubility at concentrations suitable for NMR, *J. Biomol. NMR 10*, 279−282.
32. Lepre, C. A., and Moore, J. M. (1998) Microdrop screening: a rapid method to optimize solvent conditions for NMR spectroscopy of proteins, *J. Biomol. NMR 12*, 493−499.
33. Bodenhausen, G., and Ruben, D. J. (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy, *Chem. Phys. Lett. 69*, 185−189.
34. Wuthrich, K. (1986) *NMR of proteins and nucleic acids*, Wiley-Interscience, New York.
35. Muranyi, A., Evenas, J., Stenberg, Y., Stenflo, J., and Drakenberg, T. (2000) ^{1}H, ^{15}N and ^{13}C assignments and secondary structure of the EGF− like module pair 3−4 from vitamin K-dependent protein S, *FEBS Lett. 475*, 135−8.
36. Werner, J. M., Campbell, I. D., and Downing, A. K. (2001) Shape and dynamics of a calcium binding egf domain pair investigated by ^{15}N NMR relaxation, *Methods Mol. Biol. 173*, 285−300.
37. Farrow, N. A., Zhang, O. W., Formankay, J. D., and Kay, L. E. (1994) A heteronuclear correlation experiment for simultaneous determination of N-15 longitudinal decay and chemical-exchange rates of systems in slow equilibrium, *J. Biomol. NMR 4*, 727−734.
38. Johnson, B. A., and Blevins, R. A. (1994) NMR View−a computer-program for the visualization and analysis of NMR data, *J. Biomol. NMR 4*, 603−614.
39. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, *Nucleic Acids Res. 25*, 4876−4882.
40. Sali, A., and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints, *J. Mol. Biol. 234*, 779−815.
41. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gross, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr. D Biol. Crystallogr. 54*, 905−921.
42. Stenberg, Y. (1999) in Department of Clinical Chemistry, Lund University, Malmö, Sweden.
43. Muranyi, A., Evenas, J., Stenberg, Y., Stenflo, J., and Drakenberg, T. (2000) Characterization of the EGF-like module pair 3−4 from vitamin K-dependent protein S using NMR spectroscopy reveals dynamics on three separate time scales and extensive effects from calcium binding, *Biochemistry 39*, 15742−15756.
44. Stenberg, Y., Linse, S., Drakenberg, T., and Stenflo, J. (1997) The high affinity calcium-binding sites in the epidermal growth factor module region of vitamin K-dependent protein S, *J. Biol. Chem. 272*, 23255−60.
45. Hackeng, T. M., Fernandez, J. A., Dawson, P. E., Kent, S. B. H., and Griffin, J. H. (2000) Chemical synthesis and spontaneous folding of a multidomain protein: Anticoagulant microprotein S, *Proc. Natl. Acad. Sci. U.S.A. 97*, 14074−14078.
46. Stenberg, Y., Julenius, K., Dahlqvist, I., Drakenberg, T., and Stenflo, J. (1997) Calcium-binding properties of the third and fourth epidermal-growth-factor-like modules in vitamin-K-dependent protein S, *Eur. J. Biochem. 248*, 163−70.

BI0492105