Solution Structure of the Ca^{2+}-Binding EGF3−4 Pair from Vitamin K-Dependent Protein S: Identification of an Unusual Fold in EGF3[†,‡]

Torbjörn Drakenberg,*[,§] Houman Ghasriani,[§] Eva Thulin,[§] Ann-Marie Thämlitz,[||] Andreas Muranyi,[§,⊥] Arto Annila,[#] and Johan Stenflo[||]

Department of Biophysical Chemistry, University of Lund, P.O. Box 124, SE-221 00 Lund, Sweden, Department of Clinical Chemistry, University of Lund, University Hospital, Malmö, SE-2205 Malmö, Sweden, and Department of Physical Sciences, University of Helsinki, Helsinki, Finland

Received January 18, 2005; Revised Manuscript Received April 5, 2005

ABSTRACT: Vitamin K-dependent protein S is a cofactor of activated protein C, a serine protease that regulates blood coagulation. Deficiency of protein S can cause venous thrombosis. Protein S has four EGF domains in tandem; domains 2−4 bind calcium with high affinity whereas domains 1−2 mediate interaction with activated protein C. We have now solved the solution structure of the EGF3−4 fragment of protein S. The linker between the two domains is similar to what has been observed in other calcium-binding EGF domains where it provides an extended conformation. Interestingly, a disagreement between NOE and RDC data revealed a conformational heterogeneity within EGF3 due to a hinge-like motion around Glu186 in the Cys-Glu-Cys sequence, the only point in the domain where flexibility is allowed. The dominant, bent conformation of EGF3 in the pair has no precedent among calcium-binding EGF domains. It is characterized by a change in the ψ angle of Glu186 from 160° ± 40°, as seen in ten other EGF domains, to ≈0° ± 15°. NOESY data suggest that Tyr193, a residue not conserved in other calcium-binding EGF domains (except in the homologue Gas6), induces the unique fold of EGF3. However, SAXS data, obtained on EGF1−4 and EGF2−4, showed a dominant, extended conformation in these fragments. This may be due to a counterproductive domain−domain interaction between EGF2 and EGF4 if EGF3 is in a bent conformation. We speculate that the ability of EGF3 to adopt different conformations may be of functional significance in protein−protein interactions involving protein S.

Protein S is a vitamin K-dependent glycoprotein with a molecular mass of about 78 kDa. It functions as a cofactor to an anticoagulant serine protease, activated protein C (APC),[1] and is an important regulator of blood coagulation (*1*, *2*). Approximately 30% of protein S in plasma is free and active as an APC cofactor, whereas the rest is bound in a complex with C4b-binding protein and lacks APC cofactor activity (*1*, *3*). Protein S binds to negatively charged phospholipids with a nanomolar dissociation constant and interacts noncovalently with APC on the phospholipid surface, an interaction that enhances the rates of APC-mediated degradation of coagulation factors Va and VIIIa. Protein S deficiency is associated with an increased risk for venous thrombosis (*3*). Recently, protein S has been implicated in the phagocytosis of apoptotic cells (*4*).

Protein S consists of an N-terminal Gla domain (residues 1−46) followed by a region sensitive to cleavage by thrombin and factor Xa (residues 47−75) and four epidermal growth factor-like domains (EGF1−4; residues 76−242). The C-terminal part of protein S consists of a sex hormone-binding globulin-like region that contains two laminin G-type repeats (residues 243−635) (*5*, *6*). Posttranslational modifications include the vitamin K-dependent γ-carboxylation of 11 N-terminal Glu residues to form γ-carboxyglutamyl (Gla) residues, β-hydroxylation of one Asp or Asn residue in each of the four EGF-like domains, and N-linked glycosylations (*7*−*9*).

The EGF-like domain is one of the most widely distributed protein domains among extracellular and membrane proteins (*10*−*12*). It consists of around 45 amino acids, including six Cys residues that are disulfide-bonded in the arrangement 1 to 3, 2 to 4, and 5 to 6. Approximately 35% of all known EGF domains bind Ca^{2+} (*11*−*15*). Two consensus sequences characterize this group of domains. The first sequence (Asp/

[†] This work was supported by grants from the Swedish Medical Research Council, the Swedish Foundation of Strategic Research, the Kock Foundation, the Pålsson Foundation, and the Foundations of the University Hospital, Malmö. The 600 MHz NMR spectrometer was purchased with grants from the Knut and Alice Wallenberg Foundation and the Swedish Council for Planning and Coordination of Research.

[‡] The chemical shift data for EGF3 and EGF3−4 are available under the accession codes bmr4728 and bmr4729 in the shift database (www.bmrb.wisc.edu) and the coordinates of the EGF3−4 structure under the accession code 1Z6C in the structure database (www.rcsb.org/pdb).

* To whom correspondence should be addressed. Phone: +46-46 2224470. Fax: +46-46 2224543. E-mail: torbjorn.drakenberg@bpc.lu.se.

[§] Department of Biophysical Chemistry, University of Lund.

[||] Department of Clinical Chemistry, University of Lund.

[⊥] Present address: Protein Separations R&D, GE Healthcare, Björkgatan 30, SE-751 84 Uppsala, Sweden.

[#] Department of Physical Sciences, University of Helsinki.

[1] Abbreviations: 2D, two dimensional; 3D, three dimensional; APC, activated protein C; DSS, 2,2-dimethyl-2-silapentane-5-sulfonic acid; EGF, epidermal growth factor; HMQC, heteronuclear multiple-quantum coherence; HSQC, heteronuclear single-quantum coherence; IPAP, inphase−antiphase; NOE, nuclear Overhauser enhancement; NOESY, nuclear Overhauser enhancement spectroscopy; RDC, residual dipolar couplings; SAXS, small-angle X-ray scattering.

10.1021/bi050101f CCC: $30.25
Published on Web 05/25/2005

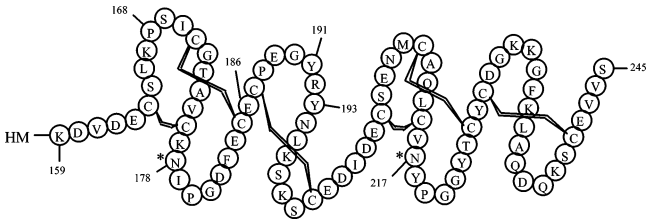

FIGURE 1: Schematic representation of EGF3−4. Some selected residues referred to in the text are labeled with their numbers in intact protein S. The two asterisks denote Asn residues that are hydroxylated. The two N-terminal amino acids (HM) remain from the His tag.

Asn-Xxx-Asp/Asn-Glu/Gln-Cys, where Xxx is typically a hydrophobic residue) ends with the N-terminal Cys residue, whereas the second sequence (Cys-Xxx-(Asp/Asn)*-Xxx-Xxx-Xxx-Xxx-Tyr/Phe-Xxx-Cys) is delineated by the third and fourth Cys residues and contains a β-hydroxylated Asp/Asn residue (denoted by an asterisk; *8*, *11*−*18*) (Figure 1). The latter sequence is folded into a two-stranded β-sheet with the hydroxylated Asp/Asn residue adjacent to the Tyr/Phe residue, a structural feature required for processing by the Asp/Asn β-hydroxylase (*18*).

The affinity of an isolated EGF domain for Ca^{2+} is too low to allow the site to be saturated at physiological extracellular Ca^{2+} concentrations, which are around 1.2 mM (*18*−*20*). The presence of an N-terminal Gla or EGF domain typically increases the Ca^{2+} affinity of the site 10−100-fold (*19*−*21*). In serine proteases such as factors VII, IX, and X and protein C, binding of Ca^{2+} to the N-terminal EGF domain orients the adjacent Gla domain in a manner that is a prerequisite for biological activity of the proteins (*22*). Moreover, proteins with multiple Ca^{2+}-binding EGF domains arranged in tandem, such as fibrillin, adopt an extended rodlike conformation in the presence of Ca^{2+}, whereas in the absence of the metal ion there is considerable interdomain mobility (*15*). The Ca^{2+}-binding site in EGF domains is a versatile structure, and its affinity can vary several orders of magnitude among proteins with tandem EGF domains. So far, the sites with the highest affinity have been found in protein S (*23*). It has also been shown that calcium-binding EGF domains can mediate long-range interactions in proteins such as the vitamin K-dependent serine proteases (*24*, *25*).

Studies employing site-directed mutagenesis have demonstrated that, in full-length protein S, the thrombin-sensitive domain and the first two EGF-like domains mediate the interaction with APC (*26*). This has been corroborated by studies performed with a chemically synthesized polypeptide, so-called microprotein S (containing the Gla domain, the thrombin-sensitive domain, and the first EGF-like domain) (*27*), and of recombinant protein S lacking the second EGF-like domain (*28*). Attempts to prepare the two N-terminal EGF-like domains of protein S as an isolated tandem pair have proven difficult as the recombinant protein is subject to misfolding, even when a baculovirus expression system is employed. In contrast, EGF1−4 expressed in the same system folds properly (*29*). Recently, it was shown that, during the folding process, interactions between adjacent EGF domains are crucially important for attaining a native conformation (*30*).

Here we describe the structure of the Ca^{2+}-saturated form of the EGF3−4 pair of protein S, which was successfully folded to a native conformation after being expressed in *Escherichia coli*. The structure reveals that the interdomain interactions between EGF3 and EGF4 conform to those previously described (*12*). However, the fold of EGF3 in the pair is unique among EGF-like domains due to rotation around the N−C^{α} and C^{α}−C(O) bonds of Glu186, i.e., the Glu residue located between the fourth and fifth Cys residues in EGF3. Hence, the tandem EGF domains of protein S do not necessarily have the typical elongated structure but can, under certain circumstances, adopt a bent conformation that might well have important physiological connotations considering the numerous protein−protein interactions in which protein S is involved.

MATERIALS AND METHODS

Protein Production. The production of the EGF3−4 fragment from protein S has been described elsewhere (*31*). In brief, EGF3−4 was expressed in *E. coli* BL21(DE3) pLysS cells as a fusion protein with a His tag. The recombinant protein was reduced and refolded in vitro. The correctly folded material was purified and the His tag removed. Ca^{2+}-free EGF3−4 was lyophilized and characterized by SDS−PAGE, reversed-phase HPLC, amino acid composition, and immunoblotting using a monoclonal antibody (HPS56) that recognizes a conformation-dependent epitope in EGF3 (*32*). Agarose gel electrophoresis in the presence and absence of Ca^{2+} was used to detect correctly folded EGF3−4. N-Terminal sequencing and mass spectroscopy indicated a homogeneous protein (M_w = 9843 Da). The sequence of the final product comprised residues 159−245 of human protein S with the addition of an N-terminal dipeptide (His-Met) derived from the His tag.

The human protein S EGF3 domain (residues Lys159−Ile203) was synthesized using Fmoc chemistry on a Milligen/Biosearch 9050 peptide synthesizer (*33*). The peptide was deprotected, cleaved from the resin, extracted, and dried. Thereafter, the peptide was purified by HPLC performed under reducing conditions. Fractions with the correct sequence and amino acid composition were refolded. Finally, the refolded peptide was purified by reversed-phase HPLC. Amino acid analysis, N-terminal sequencing, and electrospray mass spectrometric analysis gave the expected results.

Production and purification of the EGF 1−4 and EGF2−4 fragments used for SAXS measurements were performed as previously described (*29*).

NMR Samples and Experiments. Samples for NMR spectroscopy were prepared by dissolving unlabeled or ^{15}N-labeled protein S EGF3−4, or unlabeled EGF3, in 90%/10% H_2O/D_2O (100% D_2O), 0.1 mM NaN_3, and 1 mM DSS (2,2-dimethyl-2-silapentane-5-sulfonic acid). Protein concentrations were in the 0.5−2 mM range. The samples contained sufficient $CaCl_2$ to saturate the Ca^{2+}-binding sites.

All NMR experiments were performed at 36 °C on a Varian Unity Inova with a ^{1}H frequency of 599.89 MHz using inverse or triple resonance probes equipped with pulsed field gradient coils in the *Z*-direction. NOE intensities were collected from 2D NOESY spectra with 80 and 150 ms mixing times of samples in D_2O and from 3D NOESY-^{15}N-HSQC spectra with mixing times of 80 and 150 ms.

Data were processed with Varian software or FELIX97 (Molecular Simulations Inc., San Diego) and analyzed with

Sparky 3.1 (Goddard and Kneller, SPARKY 3.1, University of California, San Francisco, CA). Peak integration, in 2D as well as 3D spectra, was also performed with Sparky 3.1 assuming Gaussian line shapes. Slightly overlapping peaks could reliably be resolved in this way.

For determination of residual dipolar couplings (RDCs) 1.4 mg of lyophilized ^{15}N-labeled EGF3−4 was dissolved in 250 μL of 90%/10% H_2O/D_2O containing 0.1 mM NaN_3 and 1 mM DSS in a Shigemi microcell, giving a protein concentration of 0.5 mM. The pH was adjusted to 6.0. This Ca^{2+}- and phage-free sample was used to measure the one-bond ^{1}H−^{15}N spin coupling, *J*, using the IPAP sequence (*34*). To this sample was added 40 μL of a phage solution (ASLA Ltd., Riga, Latvia), yielding a phage concentration of 16 mg/mL. Alignment was confirmed by observing the quadrupolar splitting of the deuterium signal. The splitting of the signal in the ^{15}N dimension in the Ca^{2+}-free state caused by the combined effect from scalar, *J*, and dipolar, *D*, one-bond couplings was measured. $CaCl_2$ was added to the sample, and $J + D$ were measured at Ca^{2+} concentrations of 0.5 mM (1 site approximately saturated) and 20 mM (both sites saturated). In addition to the *J* couplings measured before any additions of phage or Ca^{2+} to this sample, *J* couplings were measured in a separate sample of 1 mM EGF3−4 in the $(Ca^{2+})_2$ state. Differences in the values of *J* couplings in the two samples were minimal, and the *J* values that were measured in the $(Ca^{2+})_2$ state could therefore be used to calculate the *D* couplings in the $(Ca^{2+})_1$ state.

Small-Angle X-ray Scattering. SAXS experiments were made by using a fine-focus Cu X-ray tube in the line-focusing mode. Cu Kα radiation was selected by using a Ni filter and a totally reflecting glass block monochromator (Huber small-angle chamber 701). The intensity curves were measured using a linear one-dimensional position-sensitive proportional counter (MBraun OED50M). The distance between the sample and the detector was 495 mm for EGF1−4 and EGF2−4 and the *k*-range was from 0.015 to 0.40 Å^{-1} for all samples. The magnitude of the scattering vector *k* is defined as $k = 4\pi \sin \Theta/\lambda$, where λ is the wavelength and 2Θ is the scattering angle. The instrument function had a full width at half-maximum of 0.35 and 0.005 Å^{-1} in vertical and horizontal directions, respectively. The concentrations of the samples were ≈1 mM in H_2O at pH 6, with no Ca^{2+} added for the apo form and with 4 mM $CaCl_2$ added for the Ca^{2+} form. The protein solution was placed in a sample cell with thin polyamide foil windows. Each data set was collected at 27 °C for 12 h. The background scattering caused by solvent was measured separately and subtracted from the intensity curves. The distance distribution function was calculated by the indirect Fourier transform method using the Program Gnom (*35*).

Structure Calculations. NOE-based distance restraints for the structure calculations were obtained from the NOESY cross-peak intensities in the 80 ms NOESY spectra from

$$d = d_{\mathrm{ref}} \sqrt[6]{I_{\mathrm{ref}}/I_{\mathrm{obs}}}$$

where I_{ref} and d_{ref} are intensity and distance obtained from a few cross-peaks with known distances. I_{obs} is the observed integrated intensity, and *d* is the distance to be used in the calculations. The lower limit is set to 1.8 Å for all distances, and the upper limit is 1.1*d*. NOESY cross-peaks from 150 ms spectra not present in the 80 ms spectra were given an upper limit of 5 Å.

The RDCs measured for a weakly oriented sample were used as additional restraints in the final refinement of the structure. The values for the axial, *D*, and rhombic, *R*, terms were obtained from a grid search to −15 Hz and 0.35, respectively. The grid search was done separately for residues 160−186 and residues 186−241. Only for the latter one did the grid search result in a clear minimum in total energy. The *D* (−15 Hz) and *R* (0.35) values were kept constant independent of the number of RDCs included in the various calculations. The final value of the force constant used for the RDCs was adjusted to 1.0 to obtain RDC violations approximately matching the experimental uncertainty for residues 189−241. The main chain ϕ and ψ dihedral restraints were predicted from backbone chemical shifts with the program TALOS (*36*). The angles obtained were used with 40° margins. Additional ϕ angle restraints were obtained from $^3J_{\mathrm{HNH}\alpha}$ coupling constants measured in HMQC-J spectra as described (*37*). A total of 76 angle restraints were used.

Structure calculations were performed with the CNS program (*38*) with 200 structures calculated in each cycle. The first structure calculations used the standard CNS protocol starting from extended structures and including the disulfide bonds first in the second slow cooling phase. Only a limited set of unambiguous NOE restraints and no RDCs were used. This resulted in only a few converged structures out of the 200 calculated. All structures had relatively high energies and several NOE violations in excess of 0.5 Å. A selected set of five structures with lowest total energies, however, had the same general structure. A second round of calculations using these five initial structures as starting structures and including the disulfide bond restraints throughout resulted in well-converged structures for the five sets. All structures formed a homogeneous family. These structures resulted in reasonably good overlays of residues 160−185 or 186−245. However, each overlay resulted in a rather ill-defined structure of the other part of the molecule. Including the RDCs did not improve this significantly but increased the total energy dramatically. The refinement proceeded, using RDCs for residues 189−245 and an increasing number of NOE restraints that became unambiguous when the definition of the structure got better. In the final structure calculations a total of 930 restraints were used. They included 756 NOE-derived distances (159 intraresidue, 261 sequential, 84 medium range, and 252 long range), 76 ϕ/ψ angles, 22 restraints between Ca^{2+} ligands, 30 hydrogen bonds, and 46 RDCs. Initially, the structures were calculated without including RDCs. Ca^{2+} restraints were initially simulated as restraints between Ca^{2+} ligands with no actual Ca^{2+} ions in place. In a final torsion dynamics refinement the Ca^{2+} ions were included, and bonds between Ca^{2+} and the protein ligands as well as a water molecule were used as calculated with the Hess2FF program (*39*).

RESULTS

Structure of EGF3−4 by NMR. All NMR experiments were performed with both calcium-binding sites of EGF3−4 saturated. An initial structure was calculated on the basis of a limited set of NOE restraints starting from an extended structure generated by the CNS program. Although the

convergence was quite poor, the five lowest energy structures showed the same general structure, albeit with a high RMSD. Starting from these structures more NOE restraints were included as well as 76 ϕ/ψ torsion angles and 22 restraints between Ca^{2+} ligands (calculations made without the Ca^{2+} ions) and 15 hydrogen bonds (30 constraints). In the initial structures there were too few long-range restraints to define the orientation of the two domains relative to each other. Moreover, EGF3 in the pair was less well defined than EGF4. A set of refined structures were calculated with the use of 756 NOEs, 76 ϕ/ψ torsion angles, and 15 hydrogen bonds (30 constraints) and with the Ca^{2+} ions ligated to the protein and to one water molecule each as described (*39*), but still no RDCs were rather well defined. These structures deviated, however, dramatically from what has been found for other EGF pairs. The difference could be localized mainly to the ψ torsion angle of Glu186, which was ca. 0° instead of 160° as observed in other EGF domains. The result of this is that the N-terminal 26 amino acids are oriented approximately perpendicular to the rest of the molecule. This result agrees to some extent with our preliminary finding that the RDC data from EGF3−4 modeled onto the structure of fibrillin EGF32−33 fits best if the two EGF domains are oriented perpendicular to each other (*40*). Including 66 $^{1}H-^{15}N$ RDCs in the calculation resulted in more than a doubling of the total energy for the structures with lowest energies (from 292 to 639 kcal/mol), showing a disagreement between the RDCs and the other restraints. Some RDC violations in excess of 2 Hz were observed (after optimizing the force final constant for the RDCs in the CNS calculations to have on average one RDC violation in excess of 1 Hz when including the 46 RDCs from residues 189−241), but none was constantly violated with more than 1 Hz in the 10 structures with lowest energies. However, a calculation including all RDCs, where all NOE restraints between amino acids 160−186 and the rest of the molecule were removed (in total 19), again resulted in low-energy structures but with high backbone atom RMSD (3.47 ± 1.87 for residues 160−244). The ψ torsion angle for residue Glu186 was −33° ± 22°. A structure calculation using neither RDCs nor NOEs between the two parts of the molecule resulted in a set of structures with no definition of the ψ torsion angle for residue Glu186. This set of structures also included elongated structures similar to those observed for other EGF domains. New structure calculations starting from the most elongated structure and including either all NOEs or all RDCs again resulted in the bent structures. Inclusion of 46 RDCs from amino acids 189−241 and all NOE restraints resulted in total energies (lowest total energy; 433 kcal/mol) only slightly higher than without the RDCs and no RDC violations in excess of 1.4 Hz. The backbone RMSD for the 20 lowest energy structures was 0.75 Å for residues 160−186 and 0.86 Å for residues 201−244. A set of those structures are shown in Figure 2. Inclusion of only the 20 RDCs for residues 161−187 and all NOE restraints in the structure calculation also resulted in low-energy structures (lowest total energy; 411 kcal/mol) and no RDC violations above 0.7 Hz.

Instead of comparing calculations based on different numbers of RDCs, we decided to compare calculations based on 46 RDCs either from residues 161−215 or from residues 161−187 and 209−241 with those based on 46 RDCs from residues 189−241. The results are summarized in Table 1. Many more RDC violations over 1 Hz were observed when using RDCs from both sides of residue 186. To pinpoint the problematic area in the sequence, we started from the calculation without RDCs and included the RDCs stepwise starting from either end of the molecule. The results are depicted in Figure 3 as total energy versus included RDCs. Clearly, there is a break point around amino acids 187−190, and RDCs from either residues 161−187 or residues 189−242 can be used, but not those from both parts together.

SAXS Measurements of EGF2−4 and EGF1−4. SAXS measurements were performed on EGF1−4 and EGF2−4 in order to establish whether the unusual fold observed in EGF3 of EGF3−4 is also present in these recombinant fragments. EGF3−4 was not included, as its small size would preclude conclusions about the fold of EGF3. Measurements were made in the presence of a saturating Ca^{2+} concentration (Figure 4). Both recombinant fragments appeared to be elongated, as has been reported before for other pairs of Ca^{2+}-binding EGF domains, for instance, those in fibrillin (*12*, *15*, *41*, *42*). The maximum length corresponded to 3- and 4-fold the length of a single EGF domain for EGF2−4 and EGF1−4, respectively. The shape of the $p(r)$ functions also clearly represented a rodlike structure. We can therefore safely assume that the Ca^{2+}-saturated forms of both EGF2−4 and EGF1−4 have approximately linearly oriented EGF domains in the dominant conformation. However, that a small population of molecules may have a more compact conformation cannot be ruled out.

DISCUSSION

Numerous extracellular and membrane proteins contain Ca^{2+}-binding EGF domains arranged in tandem. For instance, the low-density lipoprotein receptor contains three, protein S four, Notch 36, and fibrillin 47 (*12*, *18*). With the Ca^{2+}-binding sites saturated, the interdomain contacts give the EGF domain pairs an extended conformation with the N- and C-termini at opposite ends (*12*, *15*). Fibrillin is a typical example; with the Ca^{2+}-binding sites saturated it adopts an extended rodlike structure, whereas in the absence of Ca^{2+} the structure is flexible.

We have now found that the linkage between the calcium-binding EGF domains 3 and 4 of protein S is similar to what has been described for interdomain contacts in other members of this group, including fibrillin and the low-density lipoprotein receptor (*12*, *15*, *30*, *41*, *42*). Accordingly, there is a close contact between an aromatic ring in one domain (Tyr191 in EGF3; Figure 1) and the tip of the main β-sheet in the adjacent domain (EGF4). Yet, the structure of the EGF3−4 pair is very different from other pairs of calcium-binding EGF domains. In the final family of 20 structures the ψ angle of residue Glu186, located in the sequence Cys-Glu-Cys, is 0° ± 15° as compared to 160° ± 40° for the corresponding amino acid in 10 other pairs of calcium-binding EGF domains (http://www.rcsb.org/pdb). Similar results were also obtained in a calculation including all RDCs but no NOE restraints between residues 160−185 and the remainder of the protein. In none of these calculations were there any torsion angle restraints for residue 186. A hinge-like motion at Glu186 is possible without affecting the internal structures of the two subdomains. Furthermore, such a motion would not affect EGF4.

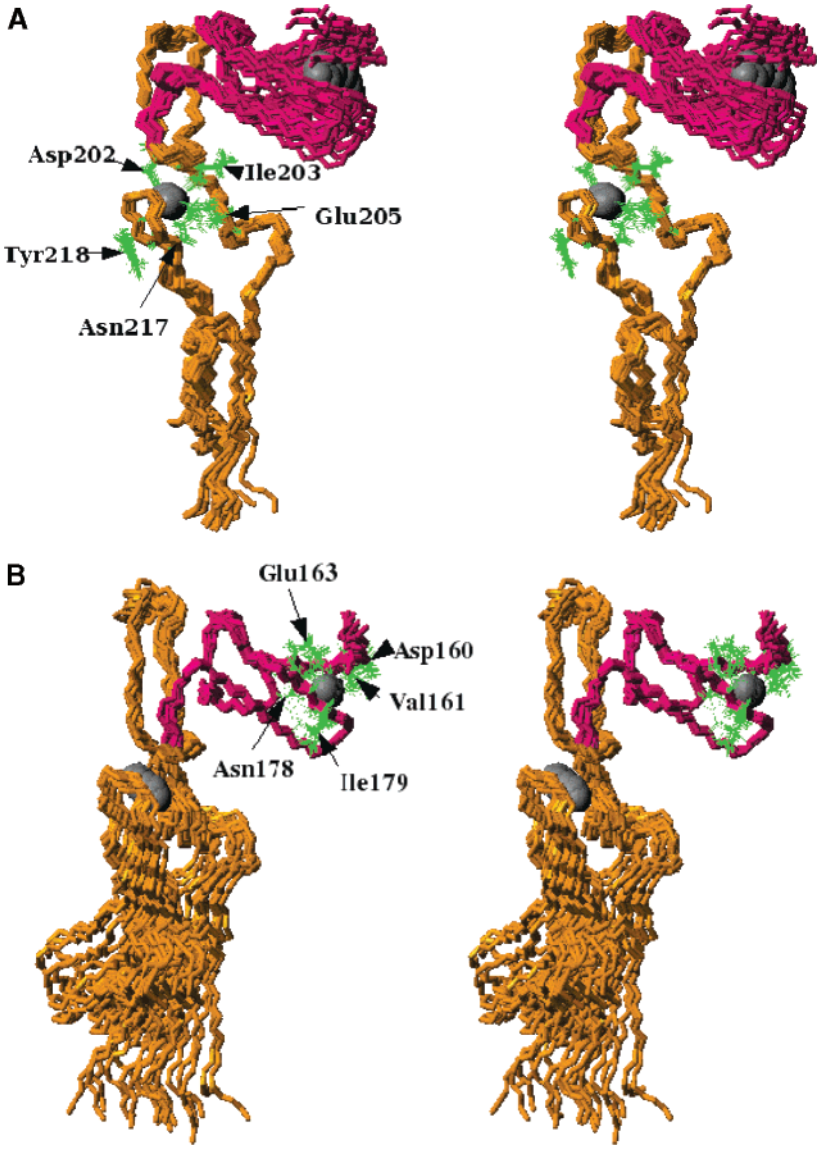

FIGURE 2: Stereoview of the structure of EGF3−4 from protein S. An overlay of the 20 lowest energy structures from the final refinement is shown as backbone traces with residues 159−186 in red and residues 187−245 in yellow. The side chains of the Ca^{2+} liganding are shown in green and the Ca^{2+} ions in gray. (A) Overlay of residues 189−244 for a refinement using 46 RDCs for residues 189−241. Note that Ile203 and Tyr218 are backbone ligands. (B) Overlay of residues 161−188 for a refinement using 20 RDCs from residues 161−187. Note that Val161 and Ile179 are backbone ligands.

Table 1: Energies and Violations for Structure Calculations Including 46 RDCs[a]

residues	total energy	NOE violation	RDC energy	RDC violation
189−241	444 (431)	96 (0.35)	7 (4.3)	13 (1.4)
161−215	554 (535)	130 (0.43)	27 (19)	65 (4.3)
161−187 and 209−241	547 (516)	110 (0.48)	16 (13)	60 (2.1)

[a] Total energies are the average of the 10 best with the best in parentheses. NOE violations are the total number from the 10 best structures with the maximum within parentheses. RDC energies are the average of the 10 best with the best in parentheses. RDC violations are the total number from the 10 best structures with the maximum within parentheses.

It is thus clear that EGF3 from protein S can adopt a structure not seen in any other Ca^{2+}-binding EGF domain so far (Figure 5). It is notable that this occurs at the only point in a calcium-binding EGF domain where some flexibility is allowed: disulfide bonds prohibit major variations in the rest of the domain. Hence, the main β-sheet in EGF3 is oriented roughly perpendicular to the other β-sheets and also perpendicular to the long axis of the molecule, resulting in a slightly less efficient ^{15}N relaxation than for the β-sheets oriented parallel to the long axis (*43*). However, modeling revealed that a structure close to the one typical for pairs of cbEGF domains could be obtained by rotation around the $N-C^{\alpha}$ and $C^{\alpha}-C(O)$ bonds of residue 186. In fact, in a structure calculation without RDCs and NOE restraints between amino acids 160−186 and the remainder of the protein, some low-energy structures were elongated. There was only a slight preference for structures similar to those in Figure 2.

The dynamics of EGF3−4, as based on ^{15}N relaxation studies, has been discussed previously (*43*, *44*). It was noted then that there is a slight difference between the N-terminal part, up to amino acid 187, and the remainder of the molecule. This difference can now be appreciated in the light of the structure of EGF3. The chemical shifts and the

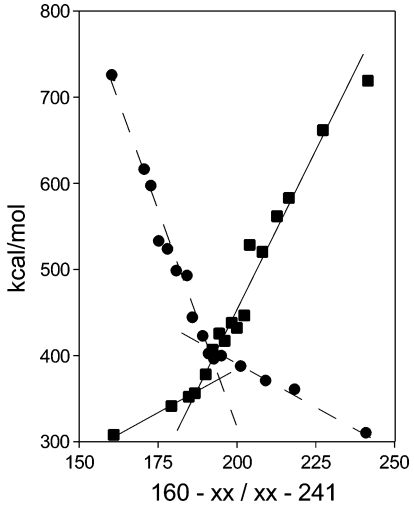

FIGURE 3: Total energy as a function of RDCs. Filled squares are derived from calculations where increasing numbers of RDCs were included from the N-terminus (*X*-axis; 160−*xxx*). Filled circles are derived from calculations where an increasing number of RDCs are included starting from the C-terminus (*X*-axis; *xxx*−241). Lines have been drawn through the data points only as an aid to localize the point of intersection.

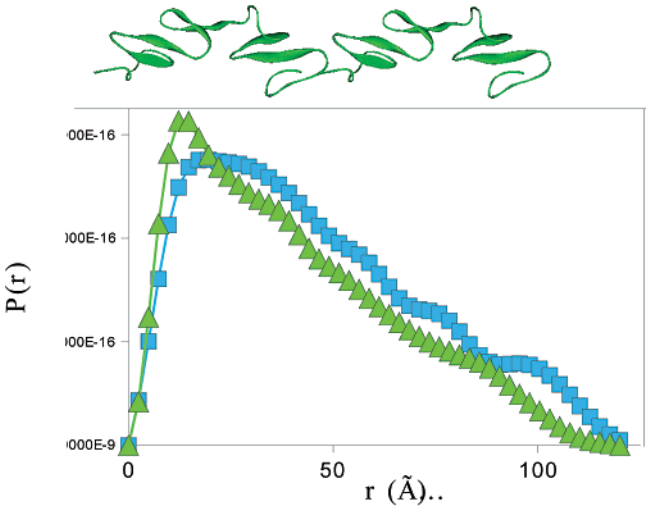

FIGURE 4: SAXS data for EGF1−4. At the top of the figure a model of two identical EGF pairs is shown, based on the structure of the LDL receptor. Filled green triangles correspond to *P*(*r*) values calculated for the model, whereas filled blue squares correspond to the experimental SAXS data for EGF1−4. Both curves are close to what would be expected for a 120 Å long cylinder.

selective broadening of signals from residues 183−186 indicate a similar conformational heterogeneity in free synthetic EGF3 and in EGF3 linked to EGF4. This is not surprising since there is no apparent contact between residues 160−185 and EGF4 (residues 201−245). In the very C-terminal part of EGF3 there are differences between the free and linked domain caused by interdomain contacts. In addition, there is a cis/trans equilibrium about the peptide bond between amino acids Lys167 and Pro168 (*43*). In this work we have only treated the dominant trans form, although signals from the cis form could be obtained from 25 residues, all in EGF3.

At this point the basis for the unique fold observed for EGF3 in protein S is uncertain. However, the position corresponding to Tyr193 in the human EGF3 domain is occupied predominantly by a Gly residue in calcium-binding

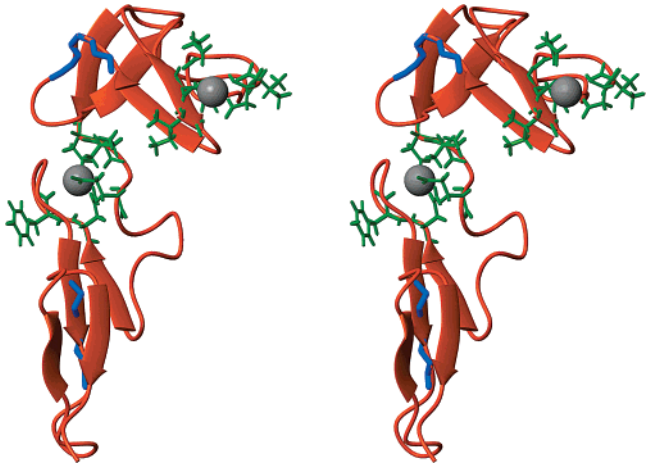

FIGURE 5: An unusual EGF conformation. A stereoview (crossed eye) of a ribbon structure of EGF3−4. The stretches Cys185-Glu186-Cys187 and Cys224-Tyr225-Cys226 are shown in purple to highlight the different conformations of the residues between the two cysteines. The side chains of the Ca^{2+}-binding residues are shown in green.

EGF domains from other proteins such as fibrillin, whereas it is completely conserved in the available protein S sequences. In Gas6, which is a homologue to protein S, the corresponding position is occupied by either Tyr, as in the human protein, or another hydrophobic residue (Phe), as in the porcine protein. Several NOESY cross-peaks between Tyr193 and other parts of EGF3 were observed, which could not be present in a typical calcium-binding EGF domain. It is, therefore, a distinct possibility that the unique fold of EGF3 is at least partly ordered by Tyr193. Additional studies employing site-directed mutagenesis may shed light on this.

A pair of EGF domains of the non-Ca^{2+}-binding type can adopt a compact structure in which the domains are packed in an antiparallel fashion with the N- and C-terminal residues on the same face of the molecule. Such a structure is entirely different from those described for pairs of Ca^{2+}-binding EGF domains. An example is the merozoite surface protein 1 from *Plasmodium falciparum* (*45*). In this case the domains have a seven-residue linker between the last Cys residue in the first EGF domain and the first Cys residue in the second domain, as compared to the five-residue linker typically found between tandem Ca^{2+}-binding EGF domains (*15*).

The bent form of EGF3 observed in this work could shed light on the results of some of our previous studies investigating Ca^{2+} binding to the EGF domains of protein S (*23*, *31*, *33*). The finding that high-affinity Ca^{2+} binding to EGF3 depends not only on the presence of EGF2 but also on the presence of EGF4 has been somewhat puzzling. In an elongated structure, such as that suggested by our SAXS results for EGF1−4 and EGF2−4, there is no interaction between the Ca^{2+}-binding site in the N-terminus of EGF3 and EGF4. However, in the bent conformation observed for EGF3 the possibility exists for an interaction to occur between EGF2 and EGF4, which may be contraproductive, resulting in an increased population of the elongated structure as observed for EGF2−4. If the Ca^{2+} affinity for an EGF domain with the bent structure is lower than for one with the canonical extended conformation, then the Ca^{2+} affinity of EGF3 could well be higher in EGF2−4 than in EGF2−3, which would be consistent with previous observations (*23*, *31*, *33*).

Although, as shown in Figure 5, the dominant conformation of EGF3−4 is the bent one, it is clear from the disagreement between NOEs and RDCs that it is not the only one. In a rigid molecule the NOEs and RDCs are assumed to agree. However, in a flexible molecule, or one exchanging between two rigid conformations, this is not the case, since NOEs and RDCs are averaged in very different ways. The RDCs are averaged linearly whereas, for the NOEs, an R^{-6} averaging is used. A conformational heterogeneity is also in agreement with our observation that signals from residues close to and including Glu186 are strongly broadened (*43*). Several residues within the N-terminal subdomain of EGF3 also experience some millisecond to microsecond time-scale motions, which indicates that the hinge-like motion at Glu186 may cause some rearrangements within the N-terminal subdomain (*43*). Our SAXS data show convincingly that both EGF1−4 and EGF2−4 are mainly elongated with a typical EGF structure also for EGF3. We therefore suggest that the second conformation of EGF3−4 also adopts the elongated conformation typical for other pairs of cbEGF domains. The ability of EGF3 in protein S to adopt different conformations may be important for modulating the protein's structure according to its varied biological roles.

ACKNOWLEDGMENT

The authors are grateful to Dr. M. A. Brown for critically reading the manuscript.

REFERENCES

1. Dahlbäck, B., and Stenflo, J. (2001) in *The molecular basis of blood diseases* (Stamatoyannopoulus, G., Majerus, P. W., Perlmutter, R. M., and Varmus, H., Eds.) pp 614−656, Saunders, Philadelphia.
2. Rezende, S. R., 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.
3. Dahlbäck, B. (1995) The protein C anticoagulant system: inherited defects as basis for venous thrombosis, *Thromb. Res. 77*, 1−43.
4. Andersson, H. A., Maylock, C. A., Williams, J. A., Paweletz, C. P., and Shacter, E. (2003) Serum-derived protein S binds to phosphatidylserine and stimulates the phagocytosis of apoptotic cells, *Nat. Immunol. 4*, 87−91.
5. Dahlbäck, B., Lundwall, Å., and Stenflo, J. (1986) Primary structure of bovine vitamin K-dependent protein S, *Proc. Natl. Acad. Sci. U.S.A. 83*, 4199−4203.
6. Lundwall, Å., Dackowski, W., Cohen, E., Shaffer, M., Mahr, A., Dahlbäck, B., Stenflo, J., and Wydro, R. (1986) Isolation and separation of the cDNA for human protein S, a regulator of blood coagulation, *Proc. Natl. Acad. Sci. U.S.A. 83*, 6716−6720.
7. Stenflo, J., and Suttie, J. W. (1977) Vitamin K-dependent formation of gamma-carboxyglutamic acid, *Annu. Rev. Biochem. 46*, 157−172.
8. Stenflo, J., Lundwall, Å., and Dahlbäck, B. (1987) beta-Hydroxyasparagine in domains homologous to the epidermal growth factor precursor in vitamin K-dependent protein S, *Proc. Natl. Acad. Sci. U.S.A. 84*, 368−372.
9. Kaufman, R. J. (1998) Post-translational modification required for coagulation factor secretion and function, *Thromb. Haemostasis 79*, 1068−1079.
10. Campbell, I. D., and Bork, P. (1993) Epidermal growth factor-like modules, *Curr. Opin. Struct. Biol. 3*, 385−392.
11. Stenflo, J., Stenberg, Y., and Muranyi, A. (2000) Calcium-binding EGF-like modules in coagulation proteases: function of the calcium ion in module interaction, *Biochim. Biophys. Acta 1477*, 51−63.
12. Boswell, E. J., Kurniawan, N. D., and Downing, A. K. (2004) Calcium-binding EGF-like domains, in *Handbook of Metalloproteins* (Messerschmidt, A., Bode, W., and Cygler, M., Eds.) Vol. 3, pp 1−18, John Wiley & Sons, NewYork.
13. Öhlin, A.-K., Linse, S., and Stenflo, J. (1988) Calcium binding to the epidermal growth factor homology region of bovine protein C, *J. Biol. Chem. 263*, 7411−7417.
14. Rees, D. J. G., Jones, I. M., Handford, P. A., Walter, S. J., Esnouf, M. P., Smith, K. J., and Brownlee, G. G. (1988) The role of beta-hydroxyaspartate an adjacent carboxylate residues in the first EGF domain of human factor IX, *EMBO J. 7*, 2053−2061.
15. Downing, A. K., Knott, V., Werner, J. M., Cardy, C. M., Campbell, I. D., and Handford, P. A. (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.
16. Selander-Sunnerhagen, M., Ullner, M., Persson, E., Teleman, O., Stenflo, J., and Drakenberg, T. (1992) How an epidermal growth factor (EGF)-like domain binds calcium. High-resolution NMR structure of the calcium form of the NH_2-terminal EGF-like domain in coagulation factor X, *J. Biol. Chem. 267*, 19642−19649.
17. Handford, P. A., Baron, M., Mayhew, M., Willis, A., Beestley, 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.
18. Stenflo, J. (1991) Structure−function relationships of epidermal growth factor modules in vitamin K-dependent clotting factors, *Blood 78*, 1637−1651.
19. Persson, E., Selander, M., Linse, S., Drakenberg, T., Öhlin, A.-K., and Stenflo, J. (1989) Calcium-binding to the isolated beta-hydroxyaspartic acid-containing epidermal growth factor-like domain of bovine factor-X, *J. Biol. Chem. 264*, 16897−16904.
20. Persson, E., Björk, I., and Stenflo, J. (1991) Protein structural requirements for Ca^{2+} binding to the light chain of factor X. Studies using isolated intact fragments containing the gamma-carboxyglutamic acid region and/or the epidermal growth factor-like domains, *J. Biol. Chem. 266*, 2444−2452.
21. Knott, V., Downing, A. K., Cardy, C. M., and Handford, P. (1996) Calcium binding properties of an epidermal growth factor-like domain pair from human fibrillin-1, *J. Mol. Biol. 256*, 22−27.
22. Sunnerhagen, M., Olah, G., Forsén, S., Stenflo, J., Drakenberg, J., and Trewhella, J. (1996) The relative orientation of Gla and EGF domains in coagulation factor X is altered by Ca^{2+} binding to the first EGF domain. A combined NMR−small-angle X-ray scattering study, *Biochemistry 35*, 11547−1155.
23. 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 proteins, *J. Biol. Chem. 272*, 23255−23260.
24. Lenting, P. J., Christophe, O. D., Maat, H. T., Rees, D. G., and Mertens, K. (1996) Ca^{2+} binding to the first epidermal growth factor-like domain of human blood coagulation factor IX promotes enzyme activity and factor VIII light chain binding, *J. Biol. Chem. 271*, 25332−25337.
25. Leonard, B. J. N., Clarke, B. J., Sridhara, S., Kelley, R., Ofosu, F. A., and Blaqjchman, M. A. (2000) Activation and active site occupation alter conformation in the region of the first epidermal growth factor-like domain of human factor VII, *J. Biol. Chem. 275*, 34894−34900.
26. He, X., Shen, L., Villoutreix, B. O., and Dahlbäck, 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.
27. Hackeng, T. M., Fernández, 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.
28. 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.
29. Stenberg, Y., Drakenberg, T., Dahlbäck, B., and Stenflo, J. (1998) Characterization of recombinant epidermal growth factor (EGF)-like modules from vitamin-K-dependent protein S expressed in *Spodoptera* cells. Cofactor activity depends on the N-terminal EGF module in human protein S, *Eur. J. Biochem. 251*, 558−564.

30. Kurinawan, N., O'Leary, J. M., Thämlitz, A.-M., Sofair, R., Werner, J. M., Stenflo, J., and Downing, A. K. (2004) N-Terminal domain linkage modulates the folding properties of protein S epidermal growth factor-like modules, *Biochemistry 43*, 9352−9360.
31. 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*, 653−665.
32. Dahlbäck, B., Hildebrand, B., and Malm, J. (1990) Characterization of functionally important domains in human vitamin K-dependent protein S using monoclonal antibodies, *J. Biol. Chem. 265*, 8127−8135.
33. Stenberg, Y., Julenius, K., 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*, 163−170.
34. Ottiger, M., Delaglio, F., and Bax, A. (1998) Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra, *J. Magn. Reson. 131*, 373−378.
35. Svergun, D. I., Semenyuk, A. V., and Feigin, L. A. (1988) Small-angle-scattering-data treatment by the regularization method, *Acta Crystallogr. A 44*, 244−250.
36. Cornilescu, G. F., Delaglio, G. F., and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology, *J. Biomol. NMR 13*, 289−302.
37. Kubinowa, H., Grzesiek, S., Delaglio, F., and Bax, A. (1994) Measurement of $H^N-H^\alpha J$ couplings in calcium-free calmodfulin using new 2D and 3D water-flip-back methods, *J. Biomol. NMR 4*, 871−878.
38. Brünger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Niges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (2000) Crystallography and NMR System CNS, Version 1.0, Yale University, New Haven, CT.
39. Hsiao, Y.-W., Ryde, U., and Drakenberg, T. (2004) NMR structure determination of proteins supplemented by quantum chemical calculations, *J. Biomol. NMR 31*, 97−114.
40. Muranyi, A. (2000) EGF-like modules in blood coagulation proteins. Ca^{2+} binding, module interactions, structure and dynamics as studied by NMR spectroscopy, Doctoral Thesis, Lund University, Lund, Sweden.
41. 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.
42. Kurniawan, N. D., Aliabadizadeh, K., Brereton, I. M., Kroon, P. A., and Smith, R. (2001) NMR structure and backbone dynamics of a concatemer of epidermal growth factor homology modules from the human low-density lipoprotein receptor, *J. Mol. Biol. 311*, 451−456.
43. Muranyi, A., Evenäs, 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. Muranyi, A., Evenäs, J., Stenberg, Y., Stenflo, J., and Drakenberg, T. (2000) ^1H, ^{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*, 135−138.
45. Morgan, W. D., Birdsall, B., Frenkiel, T. A., Gradwell, M. G., Burghaus, P. A., Syed, S. E. H., Uthaipibull, C., Holder, A. A., and Feeney, J. (1999) Solution structure of an EGF module pair from the *Plasmodium falciparum* merozoite surface protein 1, *J. Mol. Biol. 289*, 113−122.

BI050101F