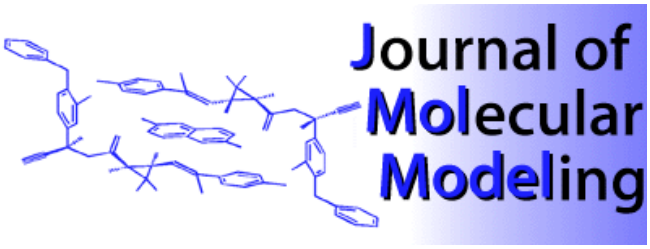
J. Mol. Model. **1998**, *4*, 268 – 275



FULL PAPER

Molecular Models for the two Discoidin Domains of Human Blood Coagulation Factor V

Bruno O. Villoutreix[1], Philipp Bucher[2], Kay Hofmann[2], Stefan Baumgartner[2,†], and Björn Dahlbäck[1]

[1]Lund University, The Wallenberg Laboratory, Dept. of Clinical Chemistry, University Hospital, Malmö, S-205 02 Malmö, Sweden. Fax: +46-40 33 70 44; E-mail: bruno.villoutreix@klkemi.mas.lu.se

[2]Swiss Institute for Experimental Cancer Research, Ch. des Boveresses 155, CH-1066 Epalinges, Switzerland

Received: 24 June 1998 / Revised: 29 July 1998 / Accepted: 31 July 1998 / Published: 11 August 1998

Abstract Discoidin (DS) domains occur in a large variety of proteins. We have recently reported the D1 domain of galactose oxidase (GOase), a copper-containing enzyme whose structure has been determined at 1.7 Å resolution, as distant member of the DS domain family. The D1 domain of GOase consists of a five-stranded antiparallel β-sheet packing against a three-stranded antiparallel β-sheet. We here show that it is possible to build 3D models for DS domains using GOase as initial template and propose a 3D structure for the C1 and C2 domains of factor V (residues 1879-2037 and 2038-2196). Factors V (FV) and VIII (FVIII) are essential and homologous non-enzymatic cofactors in the coagulation cascade. They share the domain organization A1-A2-B-A3-C1 and C2 and their C domains are members of the DS family. The C1 and C2 domains of FV are rich in positively charged residues. Several clusters of amino acids, most likely involved in inter-domain interactions, protein-protein interactions and/or phospholipid binding, are identified. Our report opens new avenues to study the structure-function relationships of DS domains.

Keywords Protein modeling, Blood coagulation, Thrombosis, FV, FVIII

Abbreviations APC, activated protein C; FV, factor V; FVa, activated FV; FVIII, factor VIII; FVIIIa, activated FVIII; DS domain, discoidin domain; 3D, three-dimensional; GOase, galactose oxidase; PT, prothrombinase; PHD, profile-fed neural network systems from Heidelberg

Running Title Molecular models for the C1 and C2 domains of FV

Introduction

It is known that the amino acid sequence of proteins can diverge significantly during evolution while the 3D fold can be conserved. It seems that partial conservation of a key set of residues is sufficient for conferring topological similari-

Correspondence to: B. O. Villoutreix
†Present address: Lund University, Dept. of Cell & Molecular Biology, Section for Developmental Biology, Box 94, S-22100 Lund, Sweden

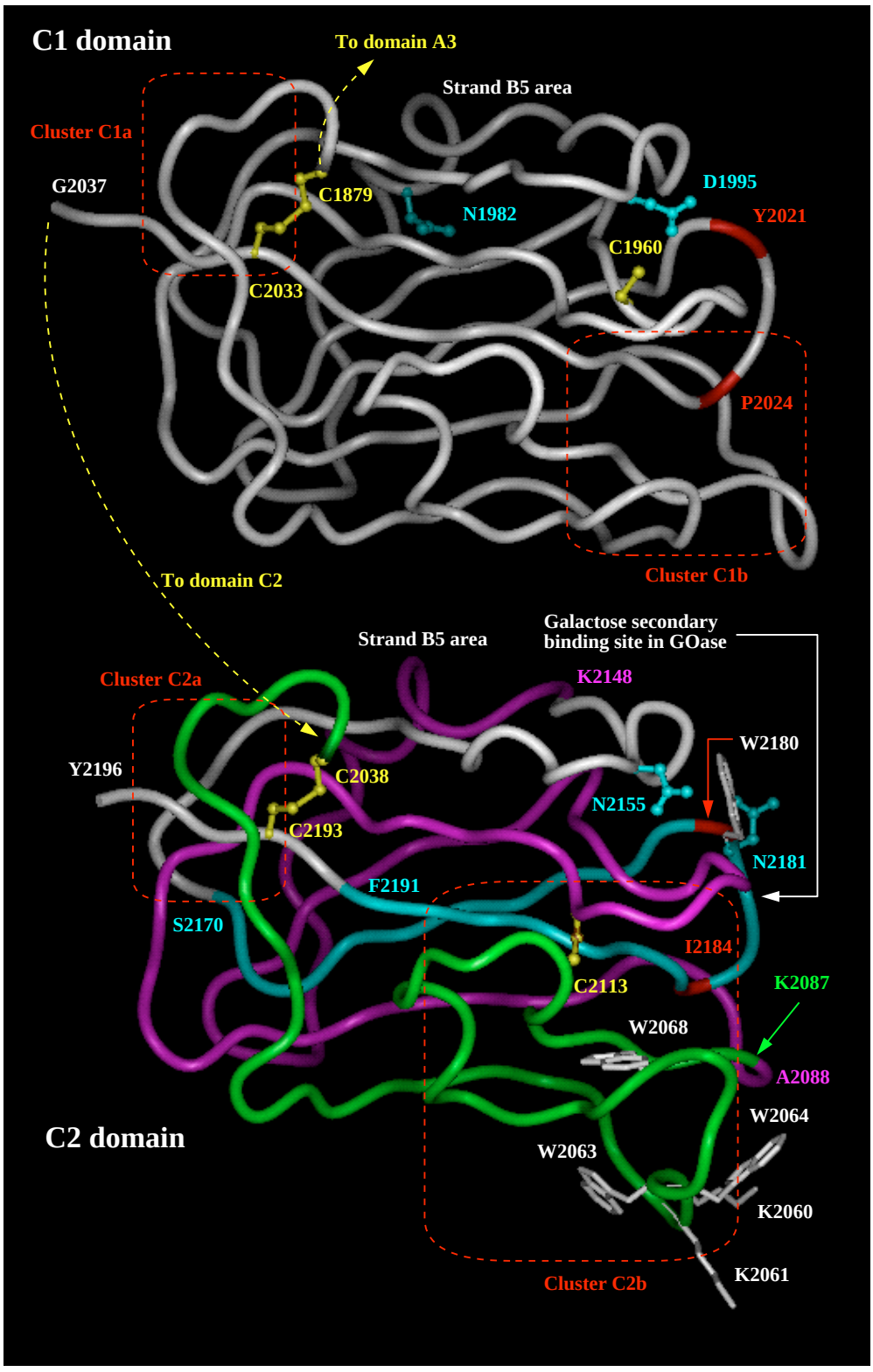


Figure 1 Ribbon diagram for the C1 and C2 domains of FV. *Some key residues discussed in the text are highlighted (e.g. glycosylation sites). A peptide segment from the C2 domain of FVIII has been suggested of importance for membrane binding and this region was expected to be α-helical. The homologous region in the C2 domain of FV is colored in blue and it is not helical (residues S2170-F2191). Residues Y2021 and P2024 or W2180 and I2184 would mimic cysteines that are known to form a S-S bond in the C1 domains of PAS-6/7 C1 (C234-C238). This extra disulfide bridge in PAS-6/7 is not compatible with the presence of an α-helix. The epitope of mAb H1 known to inhibit membrane binding is shown in green (residues C2038-K2087). Part of the epitope of mAb 6A5, antibody that is known to have no effect on membrane binding is presented in magenta (residues A2088-K2148). The putative secondary binding site for galactose within the GOase enzyme is also shown. Clusters of residues of potential functional importance are listed in the text.*

ties. Remote homologues are the result of divergent evolution that often share common fold and function despite of a low sequence identity (≈10-20%). They generally have strong signals in their linear amino acid sequence to be detected accurately. Several powerful approaches have been developed to detect remote homologies, like for instance, the generation of profiles or the use of hidden Markov models analysis [1]. Using generalized profile-based sequence search method it has been found that the DS domains are unambiguously related to the D1 domain of galactose oxidase (GOase) [2].

GOase is 639 residues long and is composed of three (D1-D2-D3) predominantly β-structure domains [3-4]. The first domain (D1, 155 residues), at the N-term of the protein, contains eight β-strands. A five-stranded β-sheet (B1-B2-B7-B4-B5) faces another three-stranded β-sheet (B8-B3-B6). The core interior of this domain is rich in hydrophobic residues coming from both β-sheets. The D1 domain of GOase possibly binds galactose within a hydrophobic/aromatic and hydrophilic pocket which is not the active site of the enzyme but a secondary binding site. This cavity could play a role in the binding of carbohydrates present in the cell wall of trees and therefore acts as an anchor to fix the protein at its appropriate location. One edge of the five-stranded β-sheet of D1 (B5) has numerous non-covalent interactions with the second domain (D2).

FVa, the activated form of FV (2196 residues), serves as cofactor within the prothrombinase (PT) complex, a multimolecular machinery (FVa, factor Xa, calcium, phospholipids) which enhances significantly the conversion of prothrombin to thrombin [reviewed in 5]. Structurally and functionally, FV is similar to FVIII, a key protein within the X-ase complex (FVIIIa, factor IXa, calcium, phospholipids). Detailed understanding of the FV and FVIII functions has been limited due to a lack of structural information. Only the 3D structure of the A domains of FV [6] and FVIII [7] have been recently predicted using the X-ray structure of ceruloplasmin [8] as initial framework. Starting from the N-terminus, FV (or FVIII) consists of domains A1, A2, B, A3, C1 and C2. FV is cleaved by thrombin within the B domain to give rise to the active cofactor (FVa), which is composed of a heavy chain (domains A1-A2) and a light chain (domains A3-C1-C2). Proteolytic inactivation of FVa by activated protein C (APC) is one of the key reactions in the regulation of thrombin formation. A congenital resistance to APC due to a point mutation in FV (R506Q) is the most common genetic risk factor for venous thrombosis [9-11]. Naturally occurring mutations in the FVIII gene can cause bleeding disorder (haemophilia A).

The roles of the C domains are not well characterized yet. However, it is known that both FV and FVa bind with high affinity to membranes that contain negatively charged phospholipids [12-14] and it has been proposed that the C2 domain of FVa interacts with the membrane through ionic forces [15,16]. Anti-FV antibodies binding to the C2 domain have been associated with hemorrhagic manifestations as they often do not allow for the formation of the PT complex [17]. Two forms of FV have been found in plasma, FV1 and FV2. FV1 has a lower affinity for negatively charged phospholipids as compared to FV2 and thus reduced cofactor activity for the generation of thrombin [18,19]. The difference between these two forms of FV is presumably due to differences in glycosylation of N2181 within the C2 domain with the FV1 variant being glycosylated [18-20]. A role for C domains in phospholipid binding is further suggested because a C domain found in a receptor tyrosine kinase present in breast carcinoma cells has been shown important for cell-cell interaction and recognition [21].

Methods

A sensitive multiple alignment-based search technique [1] was used to align a large family of DS domains [2]. Secondary structure prediction using the profile-fed neural network systems from Heidelberg (PHD) [22] was performed on this multiple sequence alignment and two threading methods [23,24] were used (giving the sequence of FV as query input) to gain further information about the DS module. The sequence identity between the C domains of FV and the GOase D1 domain is around 15 % but the multiple sequence alignment and structural analysis of GOase allow to align these sequences accurately. The (semi)-conservation of residues in the central core region in the DS domain family indicates clearly which residues of FV should be built using the amino acid residues of GOase. The sequences in the connecting regions (e.g., loops) are in some cases more difficult to align as the identity is low [2]. However, once the central core region of FV had been built it was easier to align the sequences within the loop segments. The X-ray structure of GOase (entry 1gof, [3,4]) was used as initial framework to build the two C domains of FV. Conservative side-chain replacements were modeled in conformations similar to the ones present in the GOase structure. Other residue replacements were modeled using the GOase coordinates as initial template but optimized, if needed, using low-energy rotamer conformations. The insertion regions were built using the random tweak algorithm which generates a peptide segment *de novo* by randomly searching the conformational space for a suitable backbone conformation while a screening for steric overlap violations and a set of distances between the two anchor residues are taken into account [25]. Deletions in FV when compared to GOase were effected by computationally removing the appropriate residues. Both models were energy minimized using Discover (Biosym-MSI). Calculations were carried out using the CVFF force field parameters [reviewed in 26], a dielectric constant of 1 and a 20 Å cutoff distance for non-bonded interactions. Hydrogen atoms were added to the models and partial charges were assigned to all atoms. Potentially charged residues were given appropriate parameters to obtain electrostatic neutrality as water molecules were not included in the calculations [26]. The stereochemistry of the models and X-ray template were analyzed using ProStat (Biosym-MSI). Electrostatic calculations were performed with DelPhi [27].

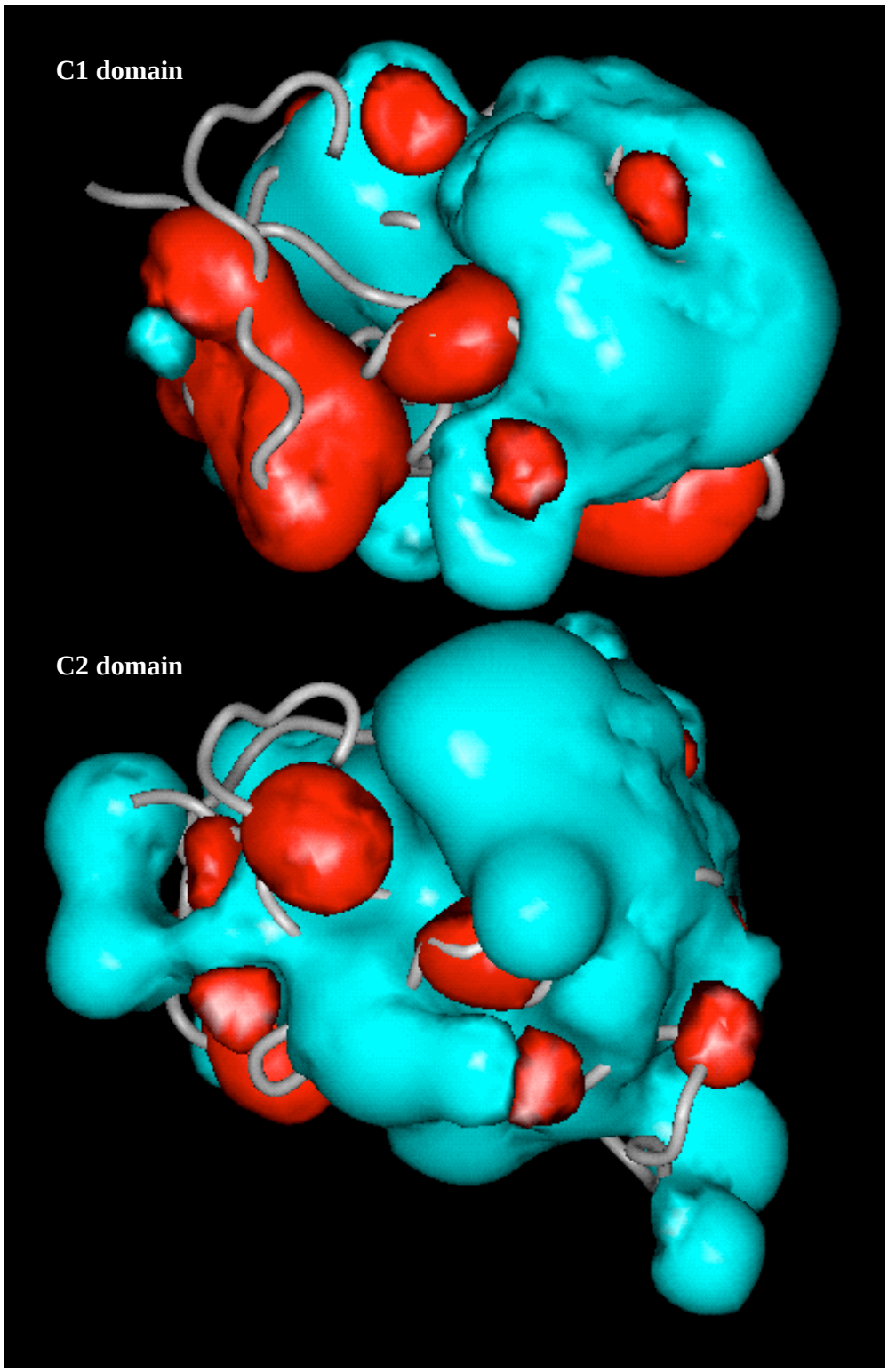


Figure 2 Electrostatic surface potential of the FV C1 and C2 models. *The models are shown with the same orientation than the one used in figure 1. These domains contain more positively charged residues than negative ones. The C1 and C2 domains have 17 and 24 positively charged residues respectively (counting only R and K but not H). The electrostatic isosurfaces are shown at a level of –1 (red) and +1 (blue) kcal/mol/e*

Results and discussion

It is known that fold recognition and structural prediction of remote homologues is basically achieved in 100% of the cases with the computational methods available at present (e.g., threading, secondary structure prediction) and a survey of the literature [28]. Errors can however occur, essentially in the loop regions [29], but even in this situation, the models are sufficiently accurate to predict binding sites or other functional features as well as to disqualify wrong interpretations of experimental data.

Overall validation of the FV models

The overall structure of the C1 and C2 models of FV shows the basic features of the GOase D1 domain (Fig. 1). Thus, the C domains are built of antiparallel β structures with a well defined hydrophobic core. Eight lines of evidence, which can not occur by chance, support strongly our structural predictions:

First. Analysis of the models (backbone angles, bond lengths and chilarity) using ProStat did not reveal any unusual structural feature.

Second. No major steric clashes occurred during the modeling process, supporting the fact that the D1 domain of GOase is an appropriate template to build the C domains.

Third. Secondary structure prediction [see Fig. 2 in ref. 2] with PHD [22] is in good agreement with the β-strand assignments in GOase (the accuracy of the PHD prediction is ≈62%). The sequence of the C domains of FV were also subjected to the prediction-based threading method TOPITS [23]. Although the Z-scores were low (+2.3), the D1 domain of GOase was reported as most similar fold (top score). The UCLA-DOE fold recognition server [24] was also used. The (top) Z-score (+5.8) identified also the D1 domain of GOase. With this last method the confidence threshold is a Z-score of 4.8 ±1. As our Z-score is above this threshold, our structural predictions should be reliable. The sequence alignments resulting from both threading experiments introduced slight shifts in some areas that seem clearly homologous in our original multiple sequence alignment [2]. Attempt to build a 3D model with these sequence alignments did not give satisfactory results (some steric clashes, which indicate that the sequences had to be re-aligned locally). Therefore, we used the sequence alignment reported in ref. [2] to build the FV models.

Fourth. In the models, all insertions-deletions as compared to GOase are in solvent exposed areas. Important amino acid conservation (or conservative substitution) is seen within the core interior of the models while non-conservative replacements tend to be solvent exposed. The ability to reproduce the tightly packed hydrophobic core, which should confer in part topological similarity in this family of modules, supports our structural predictions.

Fifth. Spatially correlated mutations are often observed in the models. For example, the disulfide bridge in GOase involves residues C18-C27 and nearby A21 is found. These residues are replaced in the C1 domain of FV by A1897, L1908 and F1900 respectively. F1900 would have severe steric clashes with residue 1897 if this latter had a longer side chain while L1908 points away from F1900.

Sixth. The few charged residues, which tend to be buried from the solvent make hydrogen bonds with surrounding amino acids (e.g., K1954-Q1951) or are involved in salt bridges (e.g., E1905-R1910, D1938-R2011, K2048-D2194, E2119-K2178). Interestingly, a salt bridge present in the D1 domain of GOase involving D58 (end of strand B2) and R122 (end of strand B6), highly conserved in the DS sequences reported in ref. 2 is also present in FV (residues D1938-R2011 and D2098-R2171). Most hydrophobic residues, as expected for a well-folded protein, are buried into the core.

Seventh. The disulfide bridges (S-S bonds) have been reported for the C domains of FVIII and FV [30,31]. Such bonds have also been found in two glycoproteins containing two DS domains called PAS-6 and PAS-7 (present in membranes of bovine milk fat globules) [32]. These disulfide bridges are not conserved in GOase and the ability to form S-S bonds was critical to the validation of our models. It is here important to note that the absence of conservation of disulfide bridge is often observed in a protein superfamily (e.g., serpin). In the C1 domain of FV, the S-S bond involves C1879-C2033 and in the C2 domain, C2038-C2193 (Fig. 1). It is possible to connect the appropriate cysteines in both domains of FV. These FV disulfides are conserved in PAS-6 and PAS-7 while the C1 domain of both proteins has an extra disulfide (PAS C234-C238). The residues homologous to PAS C234 and C238 are Y2021-P2024, in the FV C1 domain and W2180-I2184 in the FV C2 domain (Fig. 1). These residues in FV are sufficiently close in space (7-8 Å for the Cα atoms, the Cα atoms of two Cys involved in S-S bond are usually about 6 Å apart) to be consistent with the observation of an S-S bond in the corresponding position in PAS-6 and PAS-7 proteins. Furthermore, a DS domain is found in two mammalian receptor tyrosine kinases [21,33]. These proteins should have an extra disulfide bridge as compared to the C domains of FV. This disulfide would correspond to residues S1921-T2025 in the C1 domain of FV and to residues Q2085-T2185 in the C2 domain. Interestingly, the distance between the Cα atoms of these residues in the models is 6 Å, thus perfectly appropriate for the formation of a S-S bond. This data not only support the accuracy of our models but also the quality of our sequence alignment. A free cysteine is found in the C1 and C2 domains of FV (Fig. 1). These two residues in FV (C1960 and C2113) are buried into the core interior of the domains, which is consistent with their absence of functional role as free cysteine residues solvent exposed could react with surrounding molecules.

Eighth. Asparagine residues (N1982 in C1 and N2181 in C2), part of a consensus sequence for N-glycosylation, are identified in figure 1. In the C1 domain of human FV, N1982 has reduced solvent accessibility. In the C2 domain of FV, N2181, known to be glycosylated, is fully solvent exposed. Glycosylation of this Asn seems to interfere with membrane binding, suggesting that this region could be involved in the interaction with the phospholipids. N209 is glycosylated in PAS-6 [32] and this residue corresponds to FV D1995 (C1 domain) or N2155 (C2 domain) which are both solvent exposed and located in loop structures.

Structural analysis

Regions of potential functional interest were identified by scanning the surface of the models for unusual features such as partially exposed hydrophobic clusters (Fig. 1) and/or outstanding electrostatic potential (Fig. 2). On each C domain, two main clusters (**C1a** and **C1b**, **C2a** and **C2b**) presenting with solvent exposed hydrophobic/aromatic and positively charged residues were observed (Fig. 1). These regions could be involved in protein-protein or protein-membrane interactions. **Cluster C1a** contains M1881, P1882, M1883, L1885, I1944, I1978, P2006, I2008, V2009 and V2035; **cluster C1b:** Y1903, W1904, R1907, R1910, Y1917, K1924, L1925, Y1956, L1957, and R2023; **cluster C2a**: P2041, L2042, M2044, K2101, I2102, K2104, K2137, Y2139, R2140, L2141, P2166, I2168, I2195 and Y2196, and **cluster C2b**: I2049, K2052, F2059, K2060, K2061, W2063, W2064, Y2067, W2068, F2071, R2072, R2074, R2080, K2087, L2116 and W2180.

The presence of a solvent exposed insertion loop in the C2 segment around residues K2060-K2061-W2063-W2064, as compared to the homologous region of C1, provides a signal for an interaction site (most likely a membrane binding site) (Fig. 1). In FVIII, this segment within the C2 domain also displays an insertion when compared to the FVIII C1 domain, but the two K and two W are unique to FV. The C1 and C2 domains are essentially covered by a positive electrostatic potential isosurface (Fig. 2). The electropositive character of these two domains together with the fact that they present solvent exposed hydrophobic/aromatic patches could explain the ambiguity in differentiating which of the electrostatic and hydrophobic forces dominate during membrane binding.

Membrane binding

Not only electrostatic but also hydrophobic interactions can act in concert to anchor proteins to negatively charged phospholipids. The regions of a protein that are involved in membrane binding can have a helical conformation but other type of structures have been reported. A recombinant FV lacking the C2 domain lost its ability to bind the membrane suggesting that this domain is important for phospholipid interactions [16]. The epitope of a spontaneously arising FV inhibitor (this antibody is called H1) [16] that interacts with or near the phosphatidylserine (PhS) binding site in the C2 domain of FV involves some of the solvent exposed residues from the segment G2037 to K2087 (Fig. 1, green). The H1 antibody neutralized the procoagulant activity of FVa, inhibited membrane binding and has been associated with fatal hemorrhagic outcome. After analysis of different chimeras and monoclonal antibodies, Ortel et al. [16] have proposed a membrane binding site within the N-term region of the C2 domain and suggested that the structural organization of these FV and FVIII domains should be similar. They also pointed out that antibody 6A5 binds to an epitope that requires the presence of residues A2088 to K2148 (Fig. 1, magenta) but that this epitope does not participate in PhS binding. The epitopes for antibodies 6A5 and H1 are known to be different which is consistent with our model structure since the two peptide segments are on two different "sides" of the C2 domain (Fig. 1). Antibody 6A5 was also found to inhibit FVa procoagulant activity, most likely by interfering with factor Xa or prothrombin binding. These data help to identify the potential membrane binding site(s) as well as the regions of the C2 domain that could interact with approaching molecules (i.e., regions which are away from the other FV domains). The results obtained from the analysis of the H1 antibody are consistent with the structural analysis of the C2 model. We suggest that the C2b cluster is likely to play an important role in membrane interaction. This result would be also consistent with the FV variants presenting with different glycosylation of N2181 since this residue is in the area of cluster C2b (Fig. 1). This region in FV would be close to the putative secondary binding site for galactose in GOase.

Studies based on NMR spectroscopy of a peptide corresponding to the C-term region of the C2 domain of FVIII indicated this segment to be of importance for membrane binding [34]. Indeed, these authors suggested that this short peptide, which formed an amphipathic α-helix, interacted with the membrane via its hydrophobic face. This area would encompass region of FV C2 residues S2170 to F2191 (Fig. 1, blue). This α-helix can not be formed in the native FV and FVIII molecules because PAS-6/7 C1 domains have an extra disulfide bridge (PAS C234-C238) there, that is not compatible with the presence of such secondary structure element. Because these modules have a conserved overall 3D fold, it would be very surprising if an α-helix was present in the C2 domain of FVIII but not in the corresponding region of the PAS-6/7 proteins. This segment in the FV models is not helical. The structure of these areas involves instead β-strands and a loop, with the hydrophobic residues facing the core interior and the polar side chains pointing more toward the solvent. The fact that a short peptide in solution adopts a different conformation than the corresponding segment within a protein or undergoes structural changes depending on the experimental conditions is well known and could apply here. It is also important to note that the structure of this 22 residue peptide, as followed by circular dichroism spectroscopy, changed from a random coil structure to an α-helical conformation upon addition of micelles/SDS into the solution [34].

Secondary structure of the FV or FVIII light chain

Secondary structure evaluation for the human FVIII light chain (A3-C1-C2, 650 residues) has been performed by CD spectroscopy [35]. It was found that this region of FVIII contains about 22% of α-helix and 36% of β-strand structures. These investigators pointed out that the A domain was rich in helical structures as compared to the C domains. However, modeling studies for the A domains of FV [6] and FVIII [7] and the present investigation indicate that the light chain of both cofactors contain 6-7% of α-helix and 32% of β-strand structures. Departure of the FVIII A3 domain from the heavy chain (domains A1 and A2) could induce some conformational re-arrangements thereby increasing the α-helical content. It is known that FVIII is not stable [36]. This property could alter the secondary structure evaluation. Also, CD measurements are sensitive to many parameters (e.g., purity of the protein) that can easily induce an over- or under-estimation of the secondary structure content. We are confident in our structural prediction and suggest that the DS domain do not contain helical segments.

Conclusion

Several lines of evidence strongly suggest that the overall structure of the C domains of blood coagulation FV presented here is accurate. These modules are thus built essentially of β-strand and do not contain α-helices as hypothesized in previous reports. The region encompassing residues K2060-K2061-W2063-W2064 within the C2 domain should be involved in membrane binding. Because of the similarity between FV and FVIII, it will become possible in a near future to speculate on how missense mutations in these coding regions of FVIII can lead to dysfunction and to understand the mechanism by which some of the FV and FVIII inhibitory antibodies found in patients inactivate the molecules. Possibly, models for entire FVa and FVIIIa could be developed thus helping in the analysis of their interactions with other blood coagulation factors and cofactors.

Acknowledgements This work was supported by grants from the Swedish Medical Research Council (projects n. 07143 and 11793) and research funds from the University Hospital, Malmö; the Louis Jeantet Fondation de Médecine and the Swedish National Network for Cardiovascular Research (NNCR). Dr. S. Baumgartner acknowledges a grant from the Per-Eric and Ulla Schybergs Trust.

References

1. Bucher, P.; Karplus, K.; Moeri, N.; Hofmann, K. *Comput. Chem*. **1996**, *20*, 3.
2. Baumgartner, S.; Hofmann, K.; Chiquet-Ehrismann, R. ; Bucher, P. *Protein Sci*. **1998**, *7*, 1626.
3. Ito, N.; Phillips, S.E.; Yadav, K.D; Knowles, P.F. *J. Mol. Biol.* **1994**, *238*, 794.
4. Ito, N.; Phillips, S.E.; Stevens, C.; Ogel, Z.B.; McPherson, M.J.; Keen, J.N.; Yadav, K.D.; Knowles, P.F. *Nature* **1991**, *350*, 87.
5. Rosing, J.; Tans, G. *Thromb. Haemost*. **1997**, *78*, 427.
6. Villoutreix, B.O.; Dahlbäck, B. *Protein Sci*. **1998**, *7*, 1317.
7. Pemberton, S.; Lindley, P.; Zaitsev, V.; Card, G.; Tuddenham, E.G.D.; Kemball-Cook, G. *Blood* **1997**, *89*, 2413.
8. Zaitseva, I.; Zaitsev, V.; Card, G.; Moshkov, K.; Bax, B.; Ralph, A.; Lindley, P. *J. Biol. Inorg. Chem*. **1996**, *1*, 15.
9. Dahlbäck, B.; Carlsson, M.; Svensson, P.J. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 1004.
10. Bertina, R.M.; Koeleman, B.P.C.; Koster, T.; Rosendaal, F.R.; Dirven, R.J.; De Ronde, H.; Van der Velden, P.A.; Reitsma, P.H. *Nature* **1994**, *369*, 64.
11. Dahlbäck, B.; Hildebrand, B. *Proc. Natl. Acad. Sci. USA* **1994**, *94*, 1396.
12. Bloom, J.W.; Nesheim, M.E.; Mann, K.G. *Biochemistry* **1979**, *18*, 4419.
13. Pusey, M.L.; Mayer, L.D.; Wei, G.J.; Bloomfield, V.A.; Nelsestuen, G.L. *Biochemistry* **1982**, *21*, 5262.
14. van de Waart, P.; Bruls, H.; Hemker, H.C.; Lindhout, T. *Biochemistry* **1983**, *22*, 2427.
15. Ortel, T.L.; Devore Carter, D.; Quinn Allen, M.; Kane, W.H. *J. Biol. Chem.* **1992**, *267*, 4189.
16. Ortel, T.L.; Quinn-Allen, M.; Keller, F.G.; Peterson, J.A.; Larocca, D.; Kane, W.H. *J. Biol. Chem.* **1994**, *269*, 15898.
17. Ortel, T.L.; Moore, K.D.; Quinn-Allen, M.A.; Okamura, T.; Sinclair, A.J.; Lazarchick, J.; Govindan, R.; Carmagnol, F.; Kane, W.H. *Blood* **1998**, *91*, 4188.
18. Rosing, J.; Bakker, H.M.; Thomassen, M.C.; Hemker, H.C.; Tans, G. *J. Biol. Chem*. **1993**, *268*, 21130.
19. Koppaka, V.; Talbot, W.F.; Zhai, X.; Lentz, B.R. *Biophys. J.* **1997**, *73*, 2638.
20. Ortel, T.L.; Yoo, L.; Quinn-Allen, M.A.; Kane, W.H. *Blood* **1994**, *84*, 387a.
21. Johnson, J.D.; Edman, J.C.; Rutter, W.J. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 5677.
22. Rost, B.; Sander, C. *J. Mol. Biol*. **1993**, *232*, 584.
23. Rost, B. In: *The third International Conference on Intelligent Systems for Molecular Biology (ISMB)*; Rawlings, C., Clark, D., Altman, R., Hunter, L., Lengauer, T., Wodak, S., Eds; AAAI press, Menlo Park, California, 1995, pp 314-321.
24. Fisher, D.; Eisenberg, D. *Protein Sci*. **1996**, *5*, 947.
25. Shenkin, P.S., Yarmush, D.L., Fine, R.M., Wang, H., Levinthal, C. *Biopolymers*, **1987**, *26*, 2053.
26. Mackay, D. H. L., Cross, A. J., Hagler, A. T. In: *Prediction of protein structure and the principles of protein conformation*; Fasman, G.D., Ed.; Plenum Publishing Corporation: New York and London, 1989, pp 317-358.
27. Sharp, K., Honig, B. *Annu Rev Biophys Chem.* **1990**, *19*, 301.
28. Russel, B.R., Saqi, M.A.S., Bates, P.A., Sayle, R.A., Sternberg, M.J.E. *Prot. Engng*. **1998**, *11*, 1.

29. Bajorath, J., Linsley, P.S., Metzler, W.J. *J. Mol. Model*. **1997**, *3*, 287.
30. McMullen, B.A.; Fujikawa, K.; Davie, E.W.; Hedner, U.; Ezban, M. *Protein Sci*. **1995**, *4*, 740.
31. Xue, J.; Kalafatis, M.; Mann, K.G. *Biochemistry* **1993**, *32*, 5917.
32. Hvarregaard, J.; Andersen, M.H.; Berglund, L.; Rasmussen, J.T.; Petersen, T.E. *Eur. J. Biochem*. **1996**, *240*, 628.
33. Karn, T.; Holtrich, U.; Brauninger, A.; Bohme, B.; Wolf, G.; Rubsamen-Waigmann, H.; Strebhardt, K. *Oncogene* **1993**, *8*, 3433.
34. Gilbert, G.E.; Baleja, J.D. *Biochemistry* **1995**, *34*, 3022.
35. Bihoreau, N.; Fontaine-Aupart, M-P.; Lehegarat, A.; Desmadril, M.; Yon, J.M. *Biochem. J*. **1992**, *288*, 35.
36. Lamphear, B.J., Fay, P.J. *J. Biol. Chem*. **1992**, *267*, 3725.