Kent, T., Spartalian, K., & Lang, G. (1979) *J. Chem. Phys. 71*, 4899.
Kuriyan, J., Wilz, S., Karplus, M., & Petsko, G. (1986) *J. Mol. Biol. 192*, 133.
Moore, J., Hansen, P., & Hochstrasser, R. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 5062.
Morikis, D., Champion, P. M., Springer, B. A., & Sligar, S. G. (1989) *Biochemistry 28*, 4791.
Morikis, D., Li, P., Bangcharoenpaurpong, O., Sage, J. T., & Champion, P. M. (1991) *J. Phys. Chem. 95*, 3391.
Ormos, P., Braunstein, D., Frauenfelder, H., Hong, M., Lin, S., Sauke, T., & Young, R. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 8492.
Ormos, P. Ansari, A., Braunstein, D., Cowen, B. R., Frauenfelder, H., Hong, M. K., Iben, I. E. T., Sauke, T. B., Steinbach, P. J., & Young, R. D. (1990) *Biophys. J. 57*, 191.
Perutz, M., Fermi, G., Luisi, B., Shaanan, B., & Liddington, R. C. (1987) *Acc. Chem. Res. 20*, 309.
Powers, L., Chance, B., Chance, M., Campbell, B., Friedman, J., Khalid, S., Kumar, C., Naqui, A., Reddy, K., & Zhou, Y. (1987) *Biochemistry 26*, 4785.
Rousseau, D., & Argade, P. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 1310.
Sassaroli, M., & Rousseau, D. (1987) *Biochemistry 26*, 3092.
Schomacker, K., & Champion, P. M. (1986) *J. Chem. Phys. 84*, 5314.
Shimada, H., & Caughy, W. (1982) *J. Biol. Chem. 257*, 11893.
Šrajer, V. (1991) Ph.D. Thesis, Northeastern University.
Šrajer, V., Schomacker, K. T., & Champion, P. M. (1986) *Phys. Rev. Lett. 57*, 1267.
Šrajer, V., Reinisch, L., & Champion, P. M. (1988) *J. Am. Chem. Soc. 110*, 6656.
Šrajer, V., Reinisch, L., & Champion, P. M. (1991) *Biochemistry 30*, 4886-4895.
Steinbach, P., Ansari, A., Berendzen, J., Braunstein, D., Chu, K., Cowen, B., Ehrenstein, D., Frauenfelder, H., Johnson, J., Lamb, D., Luck, S., Mourant, J., Nienhaus, G., Ormos, P., Philipp, R., Scholl, R., Xie, A., & Young, R. (1991) *Biochemistry 30*, 3988-4001.
Warshel, A. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 1789.
Young, R. D., & Bowne, S. F. (1984) *J. Chem. Phys. 81*, 3730.

Sequence-Specific ^{1}H NMR Assignments, Secondary Structure, and Location of the Calcium Binding Site in the First Epidermal Growth Factor Like Domain of Blood Coagulation Factor IX†

Linda H. Huang,‡ Hong Cheng,‡ Arthur Pardi,§ James P. Tam,∥ and William V. Sweeney*,‡

Department of Chemistry, Hunter College, 695 Park Avenue, New York, New York 10021, Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309-0215, and The Rockefeller University, 1230 York Avenue, New York, New York 10021

Received December 19, 1990; Revised Manuscript Received April 2, 1991

ABSTRACT: Factor IX is a blood clotting protein that contains three regions, including a γ-carboxyglutamic acid (Gla) domain, two tandemly connected epidermal growth factor like (EGF-like) domains, and a serine protease region. The protein exhibits a high-affinity calcium binding site in the first EGF-like domain, in addition to calcium binding in the Gla domain. The first EGF-like domain, factor IX (45-87), has been synthesized. Sequence-specific resonance assignment of the peptide has been made by using 2D NMR techniques, and its secondary structure has been determined. The protein is found to have two antiparallel β-sheets, and preliminary distance geometry calculations indicate that the protein has two domains, separated by Trp^{28}, with the overall structure being similar to that of EGF. An NMR investigation of the calcium-bound first EGF-like domain indicates the presence and location of a calcium binding site involving residues on both strands of one of the β-sheets as well as the N-terminal region of the peptide. These results suggest that calcium binding in the first EGF-like domain could induce long-range (possibly interdomain) conformational changes in factor IX, rather than causing structural alterations in the EGF-like domain itself.

Factor IX is a vitamin K dependent blood clotting protein important in the intrinsic clotting cascade. This protein has a domain structure consisting of an N-terminal γ-carboxyglutamic acid (Gla)[1] domain, two epidermal growth factor like (EGF-like) domains, and a C-terminal serine protease domain.

A wide range of blood clotting proteins contain tandem repeats of an epidermal growth factor like domain, including factor IX, factor X, factor XII, factor VII, protein S, and protein C (Furie & Furie, 1988). The function of these EGF-like domains is not understood. It has been shown that no EGF-receptor binding or mitogenic activity is present in a peptide consisting of only the first EGF-like domain of factor IX (Huang et al., 1989). However, calcium binding is required

† This research was supported by U.S. Public Health Service National Institutes of Health Grants HL41935 (J.P.T. and W.V.S.) and AI27026 (A.P.).

* Author to whom correspondence should be addressed.
‡ Hunter College.
§ University of Colorado.
∥ The Rockefeller University.

[1] Abbreviations: EGF, epidermal growth factor; TGF, tumor growth factor; Gla, γ-carboxyglutamic acid; 2D, two dimensional; NMR, nuclear magnetic resonance; COSY, two-dimensional correlation spectroscopy; DQF-COSY, two-dimensional double-quantum-filtered correlation spectroscopy; RELAY-COSY, two-dimensional relayed coherence-transfer spectroscopy; TOCSY, two-dimensional total correlation spectroscopy; TQ-COSY, triple-quantum coherence-transfer spectroscopy; t_1, evolution time; FID, free induction decay.

0006-2960/91/0430-7402$02.50/0

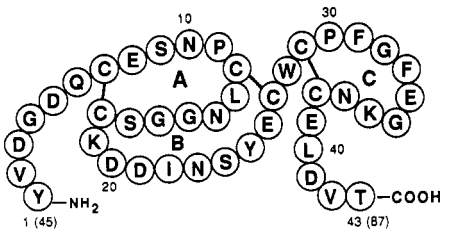

FIGURE 1: Primary structure of the first epidermal growth factor like domain human factor IX (45–87). The numbering system used is based on the synthetic peptide, with the corresponding factor IX sequence position shown in parentheses.

for activity in factor IX, and Gla-independent calcium binding has been noted (Morita et al., 1984). Recently the Gla-independent calcium binding site has been shown to occur in the first EGF-like domain (Huang et al., 1989, 1990; Handford et al., 1990). Similar Gla-independent calcium binding sites have been demonstrated in the EGF-like domain regions of factor X (Persson et al., 1989) and protein C (Ohlin et al., 1986).

In human factor IX, residues 45–87 (Figure 1) constitute the first EGF-like domain. As in EGF and TGF-α, the disulfide bond structure creates three closed loops terms in sequential order A-loop, B-loop, and C-loop. This peptide has been synthesized and can be obtained with disulfide bond pairing analogous to that found in EGF (Savage et al., 1972). In this paper, we describe the sequence-specific assignments of the ^{1}H resonances using standard 2D NMR techniques. In addition, two secondary structures consisting of two antiparallel β-sheet structures are described, and a preliminary model of the structure of the peptide is given. Furthermore, we have used NMR studies of the peptide in the presence of 20 mM Ca^{2+} to define the location of a calcium binding site.

Materials and Methods

Peptide Synthesis. The first EGF-like domain of human factor IX (45–87) was synthesized manually by the stepwise solid-phase method. The synthesis started with N^{α}-*tert*-Boc-Thr(Bzl)-OCH_2-Pam resin (0.76 mmol/g of resin; Applied Biosystem, Inc.). The Boc group was used for the protection of thhe N^{α}-amino terminus of all amino acids. Protecting groups for amino acid side chains were as follows: Asp(OBzl), Cys(4-MeBzl), Glu(OBzl), Lys(2-ClZ), Ser(Bzl), Thr(Bzl), and Tyr(2-BrZ). Deprotection of the Boc group and neutralization and coupling procedures have been described elsewhere (Huang et al., 1989). The ninhydrin test gave satisfactory results for each coupling. The peptide was cleaved by the low–high HF method (Tam et al., 1983). After HF cleavage, organic scavengers were removed by ethyl ether extraction. The peptide was fully reduced in 8 M urea/0.1 M dithiothreitol in 0.1 M Tris-HCl buffer at pH 8.4. The urea concentration of the peptide solution was slowly lowered to about 1 M by gradual dialysis against 8, 5, and 1.5 M urea. The peptide refolded in 0.1 mM reduced and oxidized glutathione solution after about 2 days at room temperature and slow stirring. The homogeneous refolded product was obtained after preparative C_{18} HPLC purification. Cf-252 mass spectrometry data showed the correct mass species for $(M + H)^+$ is 4750.6, and $(M + 2H)^{2+}$ is 4749.8, with the difference between the measured and the calculated mass being +0.7 and −0.1 mass units, respectively. Confirmation of correct disulfide refolding was achieved by thermolysin digestion (Huang et al., 1989).

NMR Experiments. Approximately 6 mg of purified peptide was dissolved in 0.6 mL of buffer containing 10 mM deuterated sodium succinate at pD 4.2. The pD values are the uncorrected pH meter readings. For the NMR experiments in D_2O, the sample was lyophilized and redissolved in 99.8% D_2O (Stohler Isotoped) several times, before the final NMR sample was dissolved in 99.996% D_2O (Stohler Isotopes). For the experiments in H_2O, the peptide solution was lyophilized and redissolved in 90% H_2O/10% D_2O. The peptide was found to dissolve more easily if it was diluted by a factor of 3–4 before lyophilization. For determination of the slowly exchanging amide protons, a peptide solution in H_2O was lyophilized and redissolved in D_2O, and a COSY spectrum was immediately acquired.

All NMR spectra were acquired on a Varian VXR-500S spectrometer operating at 499.84 MHz at a temperature of 25 ± 0.2 °C. All 2D spectra were recorded in the phase-sensitive absorption mode using the hypercomplex method (States et al., 1982). The carrier frequency was set on the water resonance with sweep widths of 7000 Hz in both dimensions except for the triple-quantum experiment which had a sweep width of 10500 Hz in the t_1 dimension. The 2D NMR spectra were collected in the following form: 2048 complex points were acquired in t_2, 300–350 complex FIDs in t_1, 32–64 transients for each FID with a recycle time of 1.3–1.9 s between each transient. In all the spectra, low-power continuous-wave irradiation was applied to the water frequency in order to suppress the residual water signal. The following 2D NMR experiments were performed according to standard procedures (Wüthrich, 1986): NOESY, TOCSY, COSY, and RELAYED-COSY in 90% H_2O/10% D_2O; and NOESY, TOCSY, DQF-COSY, TQ-COSY, and RELAYED-COSY in D_2O. The NOESY spectra were acquired with mixing times of 1, 40, 80, 120, and 200 ms, and the TOCSY spectra were acquired with a spin lock time of 70 ms. The 2D NMR data were transferred to a VaxStation 3100 or SUN 4 computer where spectral processing was performed by methods similar to those previously described (Bach et al., 1987) using the FTNMR and FELIX programs (Hare Research, Inc.).

Three-Dimensional Structure Calculations. Proton–proton distance constraints were estimated from the 200-ms NOESY spectra of the peptide in H_2O and D_2O. No intraresidue or sequential proton–proton distances were used in the calculations performed here, so these preliminary structures will simply serve as a set of possible structures that are consistent with the overall folding of the peptide. A set of 37 proton–proton distance constraints between nonsequential amino acids was used as input for a distance geometry algorithm. In these calculations, no attempt was made to account for spin diffusion in the peptide (Wüthrich, 1986), so these distances were entered as loose constraints with a lower bound of 1.9 Å and an upper bound of 4.5 Å for all proton pairs except for several α-proton–α-proton distances in the β-sheets, where the upper bound was set to 3.0 Å (because these interactions gave very strong cross-peaks in the 2D NOE spectra). A set of 10 structures were generated by the DSPACE (Hare Research, Inc.) distance geometry program using embed, annealing, and refinement procedures similar to those previously discussed (Pardi et al., 1988). These structures were then used as input for a constrained energy minimization program in order to generate energetically favorable structures consistent with the NMR distance constraints.

Results

Sequence-Specific Assignments. Amino acid spin types were identified from analysis of cross-peak patterns in standard COSY, TOCSY, and RELAY-COSY experiments in D_2O and 90% H_2O/10% D_2O solution. The D_2O and H_2O DQF-COSY spectra were the starting point for identification of spin

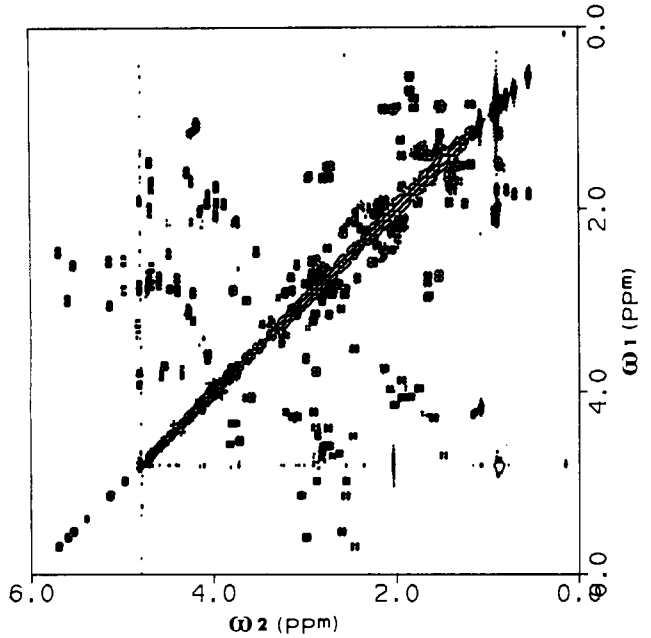

FIGURE 2: Contour plot of a portion of the D_2O DQF-COSY spectrum (25 °C, pD 4.2) showing the aliphatic region of the spectrum. Both positive and negative contour levels are plotted.

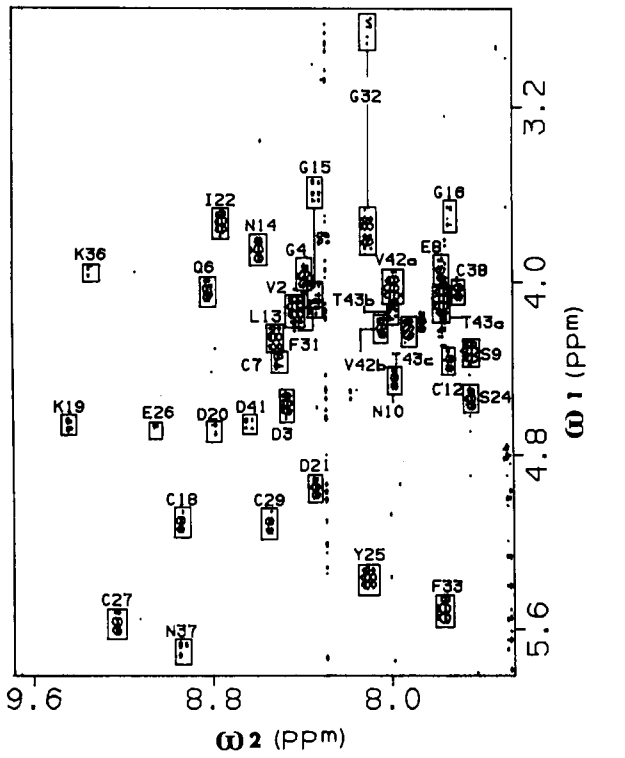

FIGURE 3: Contour plot of the fingerprint region of the H_2O COSY spectrum (25 °C, pH 4.2), showing the amide/α-proton cross-peaks. The amide/α-proton cross-peak of Glu[39] was observed (10.38 ppm/4.11 ppm) but is not shown.

systems. The quality of the data can be judged from the contour plots of portions of the D_2O DQF-COSY and H_2O COSY spectra shown in Figures 2 and 3. Spin systems identified by using D_2O COSY, RELAY-COSY, and TOCSY experiments were confirmed, and in some cases completed, by correlation with the corresponding H_2O spectra. A D_2O triple-quantum COSY experiment also proved useful in determining the chemical shifts of nearly equivalent β-protons. Determination of some amino acid types was possible from their unique spin system, as was the case, for example, with

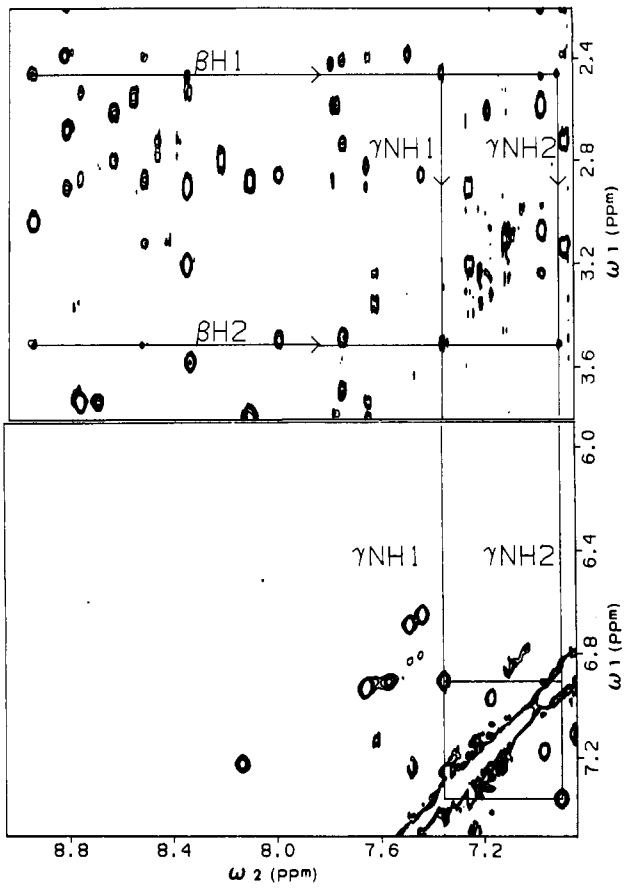

FIGURE 4: Two sections of the H_2O NOESY spectrum (25 °C, pH 4.2) showing the connectivities between the γ-amide protons and β-CH_2 protons for Asn[37].

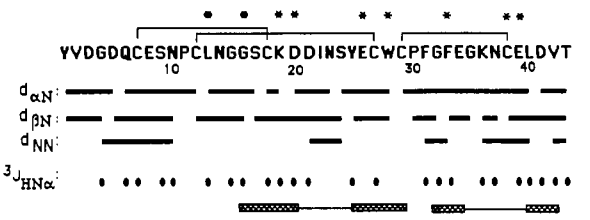

FIGURE 5: Summary of the sequential connectivities and antiparallel β-sheet segments found for the first EGF-like domain in factor IX. Disulfide bridge connections are shown, and slowly exchanging amide resonances are indicated by an asterisk. Sequential NOE connectivities are shown in three rows below the amino acid sequence. The row labeled $d_{\alpha N}$ shows the connectivities between an α-proton and the amide proton of the following amino acid (except when the following amino acid is proline, in which case a $d_{\alpha\delta}$ connection is indicated). Residues with observable amide/α coupling constants greater than 8 Hz are indicated by (●). The antiparallel β-sheet regions are specified by heavy checkered lines with thinner lines indicating the associated bends.

the lysines. In other cases, the NOESY spectrum was useful in identification. For example, side chain amide peaks for the one glutamine and all four asparagine residues were identified by strong amide/amide and amide/aliphatic cross-peaks in the H_2O NOESY spectrum (Figure 4). For all the aromatic residues, intraresidue NOE connectivities between ring protons and the β-carbon protons were used to make assignments of the aromatic proton resonances. The amine protons of Lys[19] and Lys[36] were seen in the H_2O TOCSY spectrum. These protons are not usually detected at pH 4.2 as a consequence of fast exchange rates (Wüthrich, 1986). Their observation suggests that they are involved in hydrogen bonds.

Figure 5 illustrates the sequential connectivities which were determined by using through-space NOESY interactions. The distance between the α-proton of one amino acid and the amide

Table I: Chemical Shifts for Human Factor IX (45–87), pH 4.2, 25 °C

amino acid residue	chemical shifts (ppm)			
	NH	αH	βH	others
Tyr[1]		4.25	3.10, 3.13	2,6H 7.10; 3,5H 6.85
Val[2]	8.41	4.14	2.01	γCH_3 0.86, 0.90
Asp[3]	8.46	4.58	2.74, 2.82	
Gly[4]	8.38	3.89, 4.00		
Asp[5]	8.22	4.71	2.82, 2.86	
Gln[6]	8.81	4.06	1.93, 2.19	γCH_2 2.24, 2.40; δNH_2 6.69, 7.49
Cys[7]	8.49	4.37	2.89, 3.13	
Glu[8]	7.77	3.97	2.04, 2.07	γCH_2 2.42, 2.42
Ser[9]	7.64	4.34	3.74, 3.81	
Asn[10]	7.98	4.47	2.87, 2.90	γNH_2 6.65, 7.44
Pro[11]		4.24	1.58, 1.69	γCH_2 0.64, 1.31; δCH_2 3.5
Cys[12]	7.73	4.38	2.42, 2.75	
Leu[13]	8.51	4.28	1.38, 1.59	γCH 1.80; δCH_3 0.79, 0.89
Asn[14]	8.58	3.87	1.23, 1.92	γNH_2 7.17, 7.22
Gly[15]	8.33	3.56, 4.06		
Gly[16]	7.73	3.69, 4.53		
Ser[17]	8.78	4.81	3.89, 3.93	
Cys[18]	8.94	5.12	3.05, 3.02	
Lys[19]	9.43	4.66	1.70, 1.77	γCH_2 1.28; δCH_2 1.53; ϵCH_2 2.73; ϵNH_3 7.37
Asp[20]	8.81	4.68	2.70, 2.92	
Asp[21]	8.33	4.97	2.55, 2.89	
Ile[22]	8.74	3.74	2.12	γCH_2 1.16, 1.51; γCH_3 0.88
Asn[23]	8.68	4.79	2.83, 2.91	γNH_2 6.93, 7.66
Ser[24]	7.64	4.56	3.79, 3.85	
Tyr[25]	8.10	5.39	2.74, 3.14	2,6H 6.88; 3,5H 6.84
Glu[26]	9.05	4.69	1.99, 2.04	γCH_2 2.24, 2.33
Cys[27]	9.22	5.59	2.81, 2.98	
Trp[28]	9.54	4.83	3.25, 3.37	2H 7.22; 4H 7.62; 5H 7.12; 6H 7.22; 7H 7.48; NH 10.07
Cys[29]	8.54	5.13	2.57, 2.93	
Pro[30]		4.47	2.15, 2.49	γCH_2 2.02, 2.21; δCH_2 3.26, 3.48
Phe[31]	8.35	4.22	2.92, 3.22	2,6H 7.25; 3,5H 7.35; 4H 7.29
Gly[32]	8.10	2.89, 3.79		
Phe[33]	7.77	5.52	2.60, 3.07	2,6H 6.97; 3,5H 7.17
Glu[34]	9.25	4.80	1.90, 2.12	γCH_2 2.08, 2.17
Gly[35]	8.13	3.79, 4.11		
Lys[36]	9.33	3.96	1.76, 1.94	γCH_2 1.42; δCH_2 1.65; ϵCH_2 2.74, 2.80; ϵNH_3 7.34
Asn[37]	8.93	5.69	2.44, 3.52	γNH_2 6.91, 7.36
Cys[38]	7.69	4.05	2.99, 3.60	
Glu[39]	10.38	4.11	1.95, 2.15	γCH_2 2.25, 2.56
Leu[40]	8.78	4.69	1.35, 1.49	γCH 1.18; δCH_3 0.88
Asp[41]	8.63	4.66	2.62, 2.81	
Val[42]	7.99	4.05	1.85	γCH_3 0.53, 0.68
Thr[43]	7.76	4.10		γCH_3 1.07

of the next amino acid in sequence ($d_{\alpha N}$) varies approximately between 2.20 and 3.55 Å (Wüthrich, 1986). In this 43-residue peptide, only 40 sequential $d_{\alpha N}$ peaks are expected in the H_2O NOESY spectrum because of the presence of Pro[11] and Pro[30]. Of these, 33 were observed. The remaining seven $d_{\alpha N}$ connectivities could not be unambiguously determined, either because of overlap with intraresidue cross-peaks or as a result of radio-frequency irradiation used for suppression of the water resonance. For six of the seven pairs with unobserved $d_{\alpha N}$ connections, $d_{\beta N}$ connections were seen. Thus, only for Trp[28]/Cys[29] was the connection dependent on knowledge of the sequence. Sequential connections between the two proline residues and their preceding amino acids were determined from the $d_{\alpha\delta}$ peak between the δ-protons of *trans*-proline and the α-proton of the preceding residue. The proton chemical shifts

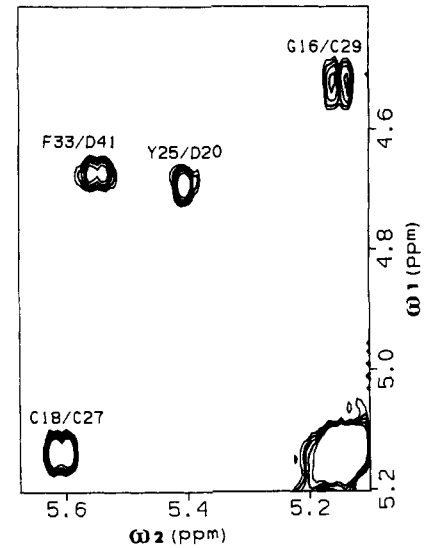

FIGURE 6: Plot of the D_2O NOESY spectrum (25 °C, pD 4.2) displaying the $d_{\alpha\alpha}$ peaks arising from the antiparallel β-sheet structures.

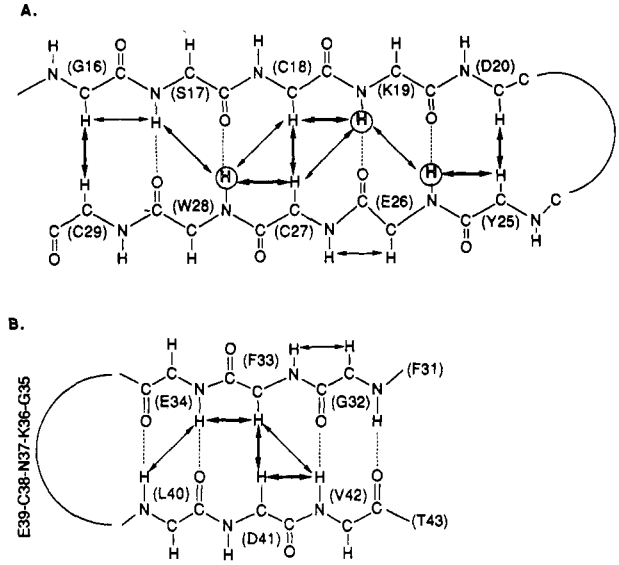

FIGURE 7: Schematic diagram of the antiparallel β-sheets found in the first EGF-like domain of factor IX. NOE connectivities are depicted by arrows, with thicker arrows denoting strong NOEs. Slowly exchanging protons detected in a COSY spectrum of a sample freshly dissolved in H_2O are encircled. Putative hydrogen bonds are shown by dashed lines.

from the sequential resonance assignment procedure are shown in Table I. A small number of other resonances were observed but not assigned to the peptide, and were presumed to arise either from impurities or from minor alternate conformations of the peptide.

Secondary Structure. Residues involved in β-sheet structures were identified by the presence of nonsequential NOESY peaks between α-protons (Figure 6). The presence and length of the two sheets were then determined from additional cross-strand cross-peaks and from the pattern of slowly exchanging amide protons. The following short interstrand proton–protons are expected in an antiparallel β-sheet: $d_{\alpha\alpha}$ = 2.3 Å, $d_{\alpha N}$ = 3.2 Å, and d_{NN} = 3.3 Å (Wüthrich, 1986). The backbone structures of two sections of proposed antiparallel β-sheets are schematically shown in Figure 7, along with the observed NOE interactions.

Figure 7A shows an antiparallel β-sheet between residues 16–20 and 25–29. Proposed hydrogen bonds are illustrated by dotted lines. With two exceptions, all the interchain NOE's

expected from an ideal antiparallel β-sheet are unambiguously detected.

The proposed β-sheet structure is confirmed by additional NMR data. The extended peptide conformation necessary for a β-sheet leads to very short sequential amide proton/α-proton distances (≈2.2 Å) and to intraresidue amide proton/α-proton spin–spin coupling constants ($^3J_{HN\alpha}$) near 9 Hz (Wüthrich, 1986). The observed sequential $d_{\alpha N}$ connectivities are consistent with an antiparallel β-sheet (Figure 7A), and the amide proton/α-proton coupling constants were greater than 8 Hz for all residues in the sheet where the coupling constant could be measured (some were not detected as a result of suppression of the solvent resonance). Additionally, the amide protons of Lys-19, Glu-26, and Trp-28 were detected as slowly exchanging, and this observation is consistent with the proposed hydrogen-bond network, and provides further support for the β-sheet.

A second shorter antiparallel β-sheet between residues 32–34 and 40–42 is found (Figure 7B). With the exception of two cross-peaks whose existence could not be unambiguously determined, all the NOE interactions expected for this sheet are found, and the detectable $^3J_{HN\alpha}$ coupling constants and strong sequential $d_{\alpha N}$ NOE peaks are consistent with the extended β-sheet structure.

A tight turn is observed at one end of each of the sheets, with a four amino acid turn in the first β-sheet (residues 21–24) and a five amino acid turn in the second (residues 35–39). The length of the second turn precludes its identification as a classical type I or II turn. The four amino acid turn does not contain glycine, reducing the probability of it being a type I′, type II, or type II′ turn (Creighton, 1984). In addition, the magnitude of the amide/α coupling constants in the turn does not conform to the expected patterns for classical type I, I′, II, or II′ turns (Wüthrich, 1986). Thus it is not possible to classify the type of β-turn for residues 21–24 from qualitative analysis of the NMR data.

The turn in the second sheet (residues 35–39) is connected to the end of the first β-sheet by a disulfide bond between Cys^{29} and Cys^{38}. Since the second β-sheet begins only two amino acids away from the end of the first sheet, the result of this is to cause the turn to fold over the second sheet. Thus, the presence of five amino acids in the second sheet β-turn may reflect the structural demands imposed by the primary and secondary structure.

Tertiary Structure. The structure of this peptide is relatively highly constrained. The peptide contains three disulfide bonds, and 58% of the peptide is tied up in the two β-sheets and associated turns. Furthermore, as was described in the preceding paragraph, the second β-sheet and turn assembly is constrained to fold over on itself, which appears to limit the relative positions of the two sheets. The two sheets then effectively form two subdomains at opposite ends of the molecule.

Although a definitive structure has not yet been obtained, distance geometry and molecular dynamics calculations, including constraints from disulfide bonding, secondary structure, and a small number (37) of other unambiguous NOE connectivities, have yielded a set of preliminary structures. These structures differ from each other in a number of important ways, but several features are present in all the structures. For the purpose of illustration, one of these structures is shown in Figure 8.

One prominent and consistent feature is the presence of two subdomains (Figure 8). These subdomains are joined by Trp^{28}, the single amino acid between loops B and C. Thus, one domain is composed of the N-terminal tail and the A and B

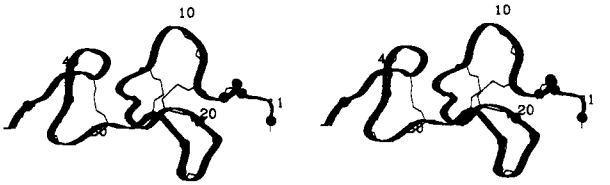

FIGURE 8: One of the energy-minimized distance geometry structures of the first EGF-like domain of human factor IX. The conformation of the N-terminal tail (residues 1–6) is undefined, as there are no nonsequential NOEs connecting any of these residues to the rest of the peptide. Dots on the ribbon indicate the α-carbon atoms of residues 1, 4, 5, 6, 20, 21, 24, 25, 34, and 42. The α- or β-proton chemical shifts of these residues change significantly upon addition of calcium ion (see text).

loops, while the other domain is made of loop C and the C-terminal tail. A second prominent feature is the general orientation of the β-sheet in the first subdomain (residues 16–29, including the β-turn), which forms a protrusion pointing away from the second subdomain. The calcium binding site appears to be located near the end of this protrusion. In the calcium-free form, the N-terminal tail (residues 1–6) is not connected by any NOEs to the rest of the first domain. However, the data presented below suggests that as a result of binding calcium, this tail folds across the end of the β-sheet, completing the binding site and enveloping the calcium ion.

Calcium Binding. The peptide binds calcium ion significantly only at pD values near 5 or above (Huang et al., 1989). Comparison of DQF-COSY spectra of the peptide in D_2O at pD values of 4.2, 4.7, and 5.4 indicated that while only small changes in chemical shifts are observed between pD 4.2 and 4.7, significant differences, particularly in the chemical shifts of Glu and Asp resonances, are seen at pD 5.4. Thus, pD 4.7 was chosen for the calcium binding studies as it allows confident extrapolation of the assignments obtained at pD 4.2, while still providing conditions close to those previously shown to be relevant for calcium binding.

A DQF-COSY spectrum of the first EGF-like domain of factor IX was acquired in D_2O at pD 4.7, and another spectrum was acquired after dissolving the peptide in buffer containing 20 mM $CaCl_2$. Assignments of shifted resonances obtained from this spectrum were confirmed by using an H_2O TOCSY spectrum.

The chemical shifts of most resonances are almost unchanged by the addition of calcium. This demonstrates that calcium ion does not induce any major change in protein conformation, consistent with the highly constrained conformation. Only 11 of 43 residues have aliphatic proton resonances which are shifted by more than 0.03 ppm by the addition of calcium. Figure 9 shows the calcium-induced changes in the resonances of protons bound to the β- and α-carbons.

Assuming that calcium binding alters the chemical shifts of the protons on the residues which act as calcium ligands, amino acids in three segments of the peptide are suggested as potential binding sites: residues 4–7, 20–21, and 24–26 (Figure 9). Asp^{20} has been previously implicated in calcium binding of factor IX (Rees et al., 1988). A model consistent with these NMR results is where a binding pocket for calcium is formed by residues 20–21 and 24–25, all of which flank the turn in the first β-sheet. Upon binding calcium, the N-terminal tail would fold across the β-sheet (Figure 8), allowing one or both of residues 5–6 to complete the binding site. However, it is not possible to rule out other adjacent residues as calcium ligands. In support of residue 6 being at or near the binding site, factor IX New London (Lozier et al., 1990), which contains Pro^6 in place of the glutamine found in our peptide, causes severe hemophilia B with no detectable factor IX ac-

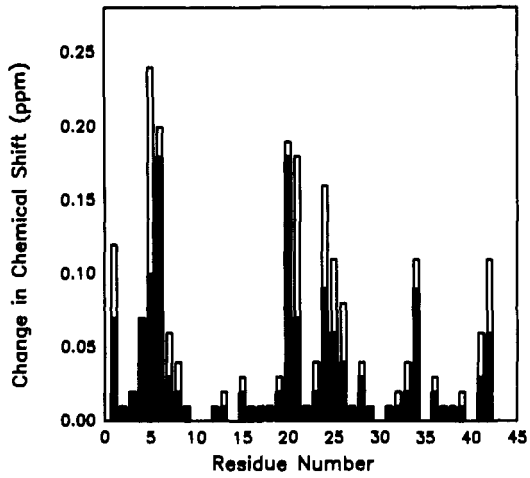

FIGURE 9: Bar graph showing the changes in chemical shift for individual amino acids as a result of the addition of calcium (20 mM $CaCl_2$, pD 4.7, 25 °C). The solid bar shows the magnitude of the maximum shift observed among the α- and β-proton resonances, while the open bar shows the sum of the absolute values of the chemical shift changes for these resonances. It was not possible to determine the chemical shift changes for either of the two proline residues, and so the values for positions 11 and 30 are unknown.

tivity. To confirm this potential binding site, NOE studies of the calcium-bound peptide are currently in progress.

Calcium is typically coordinated by oxygen atoms. In proteins, Asp or Glu are the most common ligands, but coordination through backbone carbonyl oxygens or side chain alcohol oxygens is also possible (Williams, 1986). Comparison of the amide chemical shifts for the calcium-bound peptide to those of the calcium-free peptide at pD 4.2 (data not shown) indicates that the resonance positions of the amide protons on Cys^7 and Tyr^{25} have shifted by almost 0.2 and 0.3 ppm, respectively. This suggests that Gln^6 and Ser^{24} act to bind calcium through their backbone carbonyl oxygen atoms or that there is some conformational change in the peptide upon calcium binding which leads to these chemical shift changes.

The NMR data give some evidence for a second calcium binding site formed by Glu^{34} and Val^{42}, possibly with the involvement of backbone carbonyl oxygens from several other residues. Using a shorter peptide (corresponding to residues 2–40 in Figure 1), Handford et al. (1990) noted the presence of a second aromatic resonance (corresponding to Phe^{33}) which showed some sensitivity to calcium ion. This same resonance, however, was also sensitive to magnesium, indicating a binding site of lower selectivity. However, since the Handford et al. peptide lacks Val^{42}, the significance of these calcium-dependent chemical shifts is unclear. On the basis of the set of structures obtained from distance geometry calculations (e.g., see Figure 8), it does not appear possible for either of these amino acids to be involved in the binding site in the first β-sheet. No conclusion on the presence of a second calcium binding site can be confidently drawn at this time.

Discussion

The peptide described in this paper has aspartic acid at residue 20 in place of the β-hydroxyaspartic acid usually found in this position. However, it has been shown that a considerable fraction of the aspartate at this position is unhydroxylated in human factor IX (Fernlung & Stenflo, 1983). Morita and Kisiel have also demonstrated (1985) that a β-hydroxylated aspartic acid may not be required for calcium binding. This result has been confirmed by the calcium dissociation constants of 0.4 mM (Huang et al., 1990) and 0.2–0.3 mM (Handford et al., 1990) reported for synthetic and yeast-secreted copies of the first EGF-like domain of factor IX, neither of which contain β-hydroxylated aspartic acid.

Role of Asp^3 in Calcium-Dependent Activity. Mutation studies which probed Gla-independent calcium binding have suggested a binding site in the first EGF-like domain involving Asp^3, Asp^5, and Asp^{20} (Rees et al., 1988). A molecular modeling study of the first EGF-like domain of factor IX (McCord et al., 1990) also suggested a calcium binding site involving (using the numbering system of our peptide) Asp^3, Asp^5, Asp^{20}, and Asp^{21}. Furthermore, Asp^3 appears to be important to the clotting function of factor IX (Davis et al., 1987). A mutant factor IX which has a glycine in place of Asp^3 is deficient in a number of calcium-dependent assays (McCord et al., 1990). Although this evidence would seem to strongly implicate Asp^3 in the binding of calcium, the NMR evidence appears inconsistent with this conclusion. Virtually no shift in the α- or β-protons of Asp^3 is observed upon addition of calcium ion at pH 4.7, indicating that this residue is not directly involved in calcium binding. Furthermore, the high-affinity calcium binding constant of a factor IX mutant where Asp^3 is replaced by glycine is similar to that of native factor IX (McCord et al., 1990). Thus, the role of Asp^3 in the clotting function of factor IX is likely to involve a mechanism other than direct binding of calcium.

Role of Calcium. Calcium can act in proteins to stabilize a conformation or to act as a link in the assembly of a larger complex. The absence of a major calcium-induced effect on the conformation of the first EGF-like domain, as evidenced by the peptide's constrained structure and the overall small change in the proton chemical shifts in the presence of calcium, would seem to rule out a significant conformational role for calcium. On the basis of the apparent location of the calcium binding site in our factor IX (45–87) fragment, the only possible major conformational effect of calcium appears to be the positioning of the N-terminal portion (Tyr^1–Gln^6) over the first β-sheet. While this would not be a major effect in the 43 amino acid fragment, in factor IX the implications of this potential conformation change may be significant.

The Gla domain is fastened to a membrane surface (Furie & Furie, 1988; Schwalbe et al., 1989), possibly through the interaction of Ca^{2+} and the phospholipid. The N-terminal domain of our peptide is the connection between the Gla domain and the first EGF-like domain. The consequence of calcium binding to the EGF-like domain results in calcium-dependent conformation changes that could have long-range conformational effects. These conformation changes, perhaps in concert with other molecular interactions, may be responsible for at least part of the configuration necessary at the membrane surface for activation and activity. Acting as a hinge, the residues constituting the N-terminal portion of the synthetic peptide could alter the relative orientation to the membrane surface of not only the first EGF-like domain but also perhaps the entire factor IX as well. McCord et al. (1990) have also proposed a similar calcium-dependent conformation change in factor IX, although the details of their proposal are different.

Evidence supporting a long-range interdomain effect on conformation can be found in studies on a related protein. Protein C has a structural motif very similar to that of factor IX, including a Gla domain and two tandem EGF-like domains. The first EGF-like domain of this protein has been shown to be a site of high-affinity calcium binding (Ohlin et al., 1988), as in factor IX. Studies of protein C using monoclonal antibodies directed against the N-terminal portion have shown that the high-affinity calcium binding of the first

EGF-like domain enhances the antibody affinity (Orthner et al., 1989). These results suggest that the calcium binding in the first EGF-like domain of protein C may alter the conformation of at least a part of the Gla domain.

Previous NMR Study on a Related Human Factor IX Peptide. Handford et al. (1990), using a shorter human factor IX peptide (corresponding to residues 2–40), reported that the resonances between 4.9 and 5.7 ppm broadened as the pD was raised from 3 to 7.4. This broadening effect observed near pD 7 was reversed by the addition of calcium ion. In this study, the broadening may arise from the presence of multiple conformations in equilibrium, and the binding of calcium leads to a stabilization of a single conformation. However, in our slightly larger 43-residue peptide, we do not observe a similar broadening. The peptide studied by Handford et al. (1990) is shorter by three amino acids at the C-terminal end, and therefore lacks two of the amino acids involved in the second β-sheet. The structural rigidity gained by the second β-sheet which is present in our peptide, but perhaps not in the shorter peptide, may account for the difference in pH dependence of the two peptides.

Relationship of the Secondary Structure to Other Proteins. Our results showing the two-β-sheet structure for the factor IX peptide are generally similar to structures for EGF (Montelione et al., 1987; Cooke et al., 1987; Kohda et al., 1988) and TGF-α (Tappin et al., 1989, 1990; Kohda, 1989; Katz et al., 1989). However, the structure of this factor IX domain is most similar to that of the equivalent EGF-like domain in coagulation factor X (Selander et al., 1990), which shares high sequence homology (>60%) with factor IX.

The antiparallel β-sheet structures observed here for the factor IX peptide are comparable to those seen in factor X (Selander et al., 1990), with the exception that the second sheet in factor X includes only four amino acids, rather than the six amino acid sheet seen here. In factor X, a short segment of parallel β-sheet, involving residues homologous to Cys^{12}–Leu^{13} and Asn^{37}–Cys^{38}, is reported which brings together the disulfide bridges of the B- and C-loops. In factor IX, these residues are also close, as demonstrated by an observed NOESY peak between the α-protons of Cys^{12} and Asn^{37}. Further, the amide of Leu^{13} is slowly exchanging, which is consistent with a hydrogen bond to the carbonyl oxygen of Asn^{37}. Our present data, however, do not permit identification of this third β-sheet structure.

In addition to the β-sheet structures, three turns are reported in the factor X domain, at positions homologous to residues 7–10, 13–16, and 30–33. These turns are identified as types I, I, and II, respectively. In factor IX, it is not possible to definitively classify the types of any of these turns, as the observed $^3J_{NH\alpha}$ coupling constants in factor IX are not consistent with classical type I or II turns (Wüthrich, 1986).

Conclusions

The first EGF-like domain of factor IX contains two antiparallel β-sheets, similar to the structures for factor X, EGF, and TGF-α. Preliminary distance geometry studies indicate that there are two domains in the peptide, separated by Trp^{28}, and that the conformation has a similar chain folding to that of EGF or TGF-α.

Calcium binding appears to cause very little conformation change in this first EGF-like domain, with the exception of the N-terminal tail, which may fold across the first β-sheet to complete the calcium binding site. This suggests that the role of this previously observed high-affinity calcium binding may be to participate in the positioning of factor IX, rather than to effect a conformation change in the first EGF-like domain itself. Thus, the calcium may lead to a repositioning of the full factor IX relative to the membrane surface, with the N-terminal region acting like a hinge.

Acknowledgments

We thank Dr. Ron Levy for providing a copy of the IMPACT program and Dr. Ping Yip for help in running the distance geometry and energy minimization calculations.

Registry No. FactorIX, 9001-28-9; factorIX (45–87), 119939-01-4; Ca, 7440-70-2.

References

Bach, A. C., II, Selsted, M. E., & Pardi, A. (1987) *Biochemistry 26*, 4389–4397.

Cooke, R. M., Wilkinson, A. J., Baron, M., Pastore, A., Tappin, M. J., Campbell, I. D., Gregory, H., & Sheard, B. (1987) *Nature 327*, 339–341.

Creighton, T. E. (1984) *Proteins: Structures and Molecular Properties*, pp 236–237, W. H. Freeman and Company, New York.

Davis, L. M., McGraw, R. A., Ware, J. L., Roberts, H. R., & Stafford, D. W. (1987) *Blood 69*, 140–143.

Fernlung, P., & Stenflo, J. (1983) *J. Biol. Chem. 258*, 12509–12512.

Furie, B., & Furie, B. C. (1988) *Cell 53*, 505–518.

Handford, P. A., Baron, M., Mayhew, M., Willis, A., Beesley, T., Brownlee, G. G., & Campbell, I. D. (1990) *EMBO J. 9*, 475–480.

Huang, L. H., Ke, X.-H., Sweeney, W., & Tam, J. P. (1989) *Biochem. Biophys. Res. Commun. 160*, 133–139.

Huang, L. H., Sweeney, W., & Tam, J. P. (1990) in *Peptides: Chemistry, Structure, and Biology* (Rivier, J. E., & Marshall, G. R., Eds.) pp 97–98, Escom, Leiden, The Netherlands.

Katz, B. A., Seto, M., Harkins, R., Jenson, J. C., & Sykes, B. D. (1989) in *Techniques in Protein Chemistry* (Hugli, T. E., Ed.) Academic Press, New York.

Kline, T. P., Brown, F. K., Brown, S. C., Jeffs, P. W., Kopple, K. D., & Mueller, L. (1990) *Biochemistry 29*, 7805–7813.

Kohda, D., Go, N., Hayashi, K., & Inagaki, F. (1988) *J. Biochem. 103*, 741–743.

Kohda, D., Shimada, I., Miyake, T., Fuwa, T., & Inagaki, F. (1989) *Biochemistry 28*, 953–958.

Lozier, J. N., Monroe, D. M., Stanfield-Oakley, S., Lin, S. W., Smith, K. J., Roberts, H. R., & High, K. A. (1990) *Blood 75*, 1097–1104.

McCord, D. M., Monroe, D. M., Smith, K. J., & Roberts, H. R. (1990) *J. Biol. Chem. 265*, 10250–10254.

Montelione, G. T., Wüthrich, K., Nice, E. C., Burgess, A. W., & Scheraga, H. A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 5226–5230.

Morita, T., Issacs, B. S., Esmon, C. T., & Johnson, A. E. (1984) *J. Biol. Chem. 259*, 5698–5704.

Morita, T., & Kisiel, W. (1985) *Biochem. Biophys. Res. Commun. 130*, 841–847.

Ohlin, A.-K., Linse, S., & Stenflo, J. (1988) *J. Biol. Chem. 263*, 7411–7417.

Orthner, C. L., Madurawe, R. D., Velander, W. H., Drohan, W. N., Battey, F. D., & Strickland, D. K. (1989) *J. Biol. Chem. 264*, 18781–18788.

Pardi, A., Hare, D. R., Selsted, M. E., Morrison, R. D., Bassolino, D. A., & Back, A. C., II (1988) *J. Mol. Biol. 201*, 625–636.

Persson, E., Selander, M., Linse, S., Drakenberg, T., Öhlin, A.-K., & Stenflo, J. (1989) *J. Biol. Chem. 264*, 16897–16904.

Rees, D. J. G., Jones, I. M., Handford, P. A., Walter, S. J., Esnouf, M. P., Smith, K. J., & Brownlee, G. G. (1988) *EMBO J. 7*, 2053–2061.
Savage, C. R., Jr., Inagami, T., & Cohen, S. (1972) *J. Biol. Chem. 247*, 7612–7621.
Schwalbe, R. A., Ryan, J., Stern, D. M., Kisiel, W., Bahlbäck, B., & Nelsestuen, G. L. (1989) *J. Biol. Chem. 264*, 20288–20296.
Selander, M., Persson, E., Stenflo, J., & Drakenberg, T. (1990) *Biochemistry 29*, 8111–8118.
States, D. J., Harberkorn, R. A., & Reuben, D. J. (1982) *J. Magn. Reson. 48*, 286–292.
Tam, J. P., Heath, W. F., & Merrifield, R. B. (1983) *J. Am. Chem. Soc. 105*, 6442–6455.
Tappin, M. J., Cooke, R. M., Fitton, J. E., & Campbell, I. D. (1989) *Eur. J. Biochem. 179*, 629–637.
Williams, R. J. P. (1986) in *Calcium and the Cell* (Evered, D., & Whelan, J., Eds.) Wiley, New York.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley-Interscience, New York.

^{13}C Magic-Angle Spinning NMR Studies of Bathorhodopsin, the Primary Photoproduct of Rhodopsin†

Steven O. Smith,*,‡ Jacques Courtin,§ Huub de Groot,§ Ronald Gebhard,§ and Johan Lugtenburg§
Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06511, and Department of Chemistry, Rijksuniversiteit te Leiden, Leiden, The Netherlands
Received December 11, 1990; Revised Manuscript Received May 7, 1991

ABSTRACT: Magic-angle spinning NMR spectra have been obtained of the bathorhodopsin photointermediate trapped at low temperature (<130 K) by using isorhodopsin samples regenerated with retinal specifically ^{13}C-labeled at positions 8, 10, 11, 12, 13, 14, and 15. Comparison of the chemical shifts of the bathorhodopsin resonances with those of an *all-trans*-retinal protonated Schiff base (PSB) chloride salt show the largest difference (6.2 ppm) at position 13 of the protein-bound retinal. Small differences in chemical shift between bathorhodopsin and the all-trans PSB model compound are also observed at positions 10, 11, and 12. The effects are almost equal in magnitude to those previously observed in rhodopsin and isorhodopsin. Consequently, the energy stored in the primary photoproduct bathorhodopsin does not give rise to any substantial change in the average electron density at the labeled positions. The data indicate that the electronic and structural properties of the protein environment are similar to those in rhodopsin and isorhodopsin. In particular, a previously proposed perturbation near position 13 of the retinal appears not to change its position significantly with respect to the chromophore upon isomerization. The data effectively exclude charge separation between the chromophore and a protein residue as the main mechanism for energy storage in the primary photoproduct and argue that the light energy is stored in the form of distortions of the bathorhodopsin chromophore.

Absorption of light by the visual pigment rhodopsin initiates a photochemical reaction of the protein's 11-*cis*-retinylidene prosthetic group. The first step in this reaction is an 11-cis ⇒ trans isomerization of the chromophore to form the bathorhodopsin intermediate (Figure 1) (Yoshizawa & Wald, 1963). The light energy absorbed in this process is channeled into the protein, where it initiates a biochemical chain of events leading to the closing of sodium channels in the plasma membrane (Stryer, 1986; Liebman et al., 1987). The amino acid sequence of rhodopsin has been determined and is thought to be folded into seven transmembrane helices (Nathans & Hogness, 1983; Hargrave et al., 1983; Ovchinnikov et al., 1982). The retinylidene chromophore is found in the interior of the protein attached to lysine 296 via a protonated Schiff base linkage. Recent mutagenesis studies have shown that glutamate 113 is the Schiff base counterion (Zhukovsky & Oprian, 1989; Sakmar et al., 1989; Nathans, 1990). The final step in the rhodopsin photoreaction is hydrolysis of the Schiff base linkage and release of *all-trans*-retinal from the protein.

Various approaches have been taken to study the structure and environment of the retinal chromophore in rhodopsin and bathorhodopsin with the goal of trying to understand how the energy is stored in the primary photoproduct, what factors determine the color change upon photoexcitation, and how retinal isomerization is coupled to the activation of transducin, a GTP-binding regulatory protein. Bathorhodopsin has a 6-s-cis,11-trans chromophore whose ground state energy is ~33 kcal/mol above that of rhodopsin and whose absorption maximum (λ_{max} = 543 nm) is shifted to a longer wavelength relative to rhodopsin (λ_{max} = 498). Several mechanisms have been proposed for the observation that ~60% of the light energy absorbed by rhodopsin is stored in the batho intermediate (Cooper, 1979; Schick et al., 1987) and for the red shift in the visible absorption band of bathorhodopsin. The largest contributions are calculated to come from conformational distortions of the retinal and charge separation between the positively charged Schiff base and its protein counterion [see Birge et al. (1988) for a recent review]. It is thus crucial for the understanding of the primary process of vision to

† This research was supported by the National Institutes of Health (GM-41212), the Netherlands Foundation for Chemical Research (SON), the Royal Netherlands Academy of Sciences, the Netherlands Organization for the Advancement of Pure Research (NWO), and the Searle Scholars Program/Chicago Community Trust.
‡ Yale University.
§ Rijksuniversiteit to Leiden.

0006-2960/91/0430-7409$02.50/0