Protein Science (1992), *1*, 81-90. Cambridge University Press. Printed in the USA.

The three-dimensional structure of the first EGF-like module of human factor IX: Comparison with EGF and TGF-α

M. BARON,[1] D.G. NORMAN,[1] T.S. HARVEY,[1] P.A. HANDFORD,[2] M. MAYHEW,[2] A.G.D. TSE,[3,4] G.G. BROWNLEE,[2] AND I.D. CAMPBELL[1]

[1] Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK
[2] Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK
[3] MRC Cellular Immunology Unit, Sir William Dunn School of Pathology, South Parks Road, Oxford OX1 3RE, UK

(RECEIVED July 8, 1991; ACCEPTED August 8, 1991)

Abstract

The three-dimensional structure of the first epidermal growth factor (EGF)-like module from human factor IX has been determined in solution using two-dimensional nuclear magnetic resonance (in the absence of calcium and at pH 4.5). The structure was found to resemble closely that of EGF and the homologous transforming growth factor-α (TGF-α). Residues 60–65 form an antiparallel β-sheet with residues 68–73. In the C-terminal subdomain a type II β-turn is found between residues 74 and 77 and a five-residue turn is found between residues 79 and 83. Glu 78 and Leu 84 pair in an antiparallel β-sheet conformation. In the N-terminal region a loop is found between residues 50 and 55 such that the side chains of both are positioned above the face of the β-sheet. Residues 56–60 form a turn that leads into the first strand of the β-sheet. Whereas the global fold closely resembles that of EGF, the N-terminal residues of the module (46–49) do not form a β-strand but are ill-defined in the structure, probably due to the local flexibility of this region. The structure is discussed with reference to recent site-directed mutagenesis data, which have identified certain conserved residues as ligands for calcium.

Keywords: calcium binding; EGF-like module; epidermal growth factor; human factor IX; NMR; transforming growth factor-α

Sequence comparisons have shown that many extracellular and cell-surface proteins have a mosaic structure. They seem to have evolved from a common pool of different types of modular units that may be identified by their different consensus sequences (Doolittle, 1985; Patthy, 1987). These modules are independently folding structural scaffolds that have evolved many different functions and have been combined together in different ways to produce novel proteins (Baron et al., 1991; Day & Baron, 1991). One important type of module is the epidermal growth factor (EGF)-like unit that occurs in many extracellular proteins, for example in the blood coagulation proteins (Furie & Furie, 1988) and the complement system (Reid & Day, 1989). It is also found in a number of cell-surface proteins such as the low density lipoprotein (LDL) receptor (Scott, 1989) and the *Drosophila Notch* protein (Wharton et al., 1985).

Different functional classes of the EGF module are known. One class includes EGF, transforming growth factor-α (TGF-α), and several viral proteins. Members of this group activate the EGF receptor. They share a set of highly conserved residues (in addition to those that are thought to be required to define the global fold), and these are likely to be necessary for binding and activation of the receptor (Campbell et al., 1989). A second class of EGF-like module contains a calcium-binding site that is correlated with the presence of a consensus sequence of three Asp/Asn residues, a Gln/Glu residue, and a Tyr/Phe residue (Rees et al., 1988; Handford et al., 1990, 1991). This class of EGF-like module is found in a number of different proteins including serine protease components of the blood coagulation system. The blood coagulation protein factors VII, IX, X, and protein C consist of an N-terminal "Gla-domain" that contains γ-carboxy glutamate (Gla) residues (required for calcium-

Reprint requests to: I.D. Campbell, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK.
[4] Present address: Laboratory of Immunobiology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115.

dependent binding to cell membranes) (Stenflo & Suttie, 1977), a short hydrophobic sequence, two EGF-like domains (the first of which contains the calcium-binding consensus sequence), an activation peptide (cleaved on activation of the zymogen), and a serine protease domain.

The structures of a number of different examples of the growth factor class of module have been published previously (Cooke et al., 1987; Montelione et al., 1987) together with the secondary structure of an EGF-like module of blood clotting factor X (Selander et al., 1990). In this paper we describe the three-dimensional structure (in the calcium-free form) of the first EGF-like module from human factor IX. The structure is compared with those of EGF and TGF-α and the implications for a calcium-binding site are discussed.

Results

Choice of experimental conditions

The effects of pH and calcium on the nuclear magnetic resonance (NMR) spectrum of the EGF-like module of factor IX have been described previously (Handford et al., 1990). Around neutral pH, in the absence of calcium, α-proton resonances observed to low field of the water resonance are considerably broadened, implying an exchange process. Both the addition of calcium and the lowering of pH have been found to reduce the linewidth of these resonances. In the absence of calcium the optimum spectrum was thus obtained at pH 4.5. Working at this pH also had the benefit of relatively slow backbone amide proton exchange.

Sequence-specific assignment

Sequence-specific assignment of the ^{1}H NMR spectrum of the factor IX EGF-like module was carried out by established methodology (Wüthrich, 1986). The spin systems of the amino acid residues were identified using a combination of double quantum filtered correlated spectroscopy (DQF COSY) and homonuclear Hartman–Hahn (HOHAHA) spectroscopy. Nuclear Overhauser effect (NOE) connectivities were observed between protons that are close together (less than 0.5 nm) in space with two-dimensional (2D) NOESY spectra. Sequence-specific assignments were made using NOE connectivities between adjacent spin systems, together with the known sequence. Figure 1 shows a region of a NOESY spectrum recorded in 90% H_2O/10% D_2O at 20 °C with a mixing time of 200 ms, in which NOE connectivities between amide and C-α protons are observed. Sequential assignments on part of the spectrum are illustrated. The chemical shift assignments are given in Table 1.

The sequential NOEs observed in the spectrum are shown in Figure 2, along with those amide protons that are slowly exchanging and values of the $^3J_{HN\alpha}$ that were measured to be greater than 9 Hz. The sequential connectivities and information about slowly exchanging amide protons were used together with long-range backbone–backbone NOEs in order to define the secondary structure of the molecule (Wüthrich, 1986).

Calculation of the three-dimensional structure

A total of 469 distance restraints was determined from the NOESY spectra; 79 were long range ([i–j] > 4) and 390 were short range ([i–j] < 5). An additional 10 distance restraints were included to define the five inferred hydrogen bonds between residues 70–63, 63–70, 72–61, 61–72, and 84–78 (HN-O). ϕ angles of $-120° \pm 60°$ were restrained for residues 50, 66, 63, 65, 68, 70, 72, and 77 and χ^1 angles $-60° \pm 60°$ were restrained for residues 50, 51, 56, 57, 71, 73, and 77. Figure 2B shows the distribution of distance restraints throughout the sequence, and Figure 2C and Kinemage 1 give the spatial distribution of restraints.

Small numbers of structures with total NOE penalties greater than 369 kJ/mol were discarded after the first stage of molecular dynamics, before final energy minimization (Fig. 3). Inspection of structures above this cutoff showed clearly that they were misfolded forms. The energy statistics for 63 final structures are given in Table 2 and a backbone overlay of these structures is shown in Figure 4A. The root mean square deviation (RMSD) between the structures is shown in Figure 5. The large RMSD of residues 46–49 is due to the absence of NOEs in nonsequential residues from this region. The final structures had no NOE violation greater than 0.05 nm, and the RMSD of the NOE violations was 0.0036 nm. Figure 4B and Kinemage 2 show an energy-minimized average backbone structure (MinAv) upon which the locations of some key residues are indicated. Table 3 gives the mean and RMSDs of the backbone ϕ and φ angles calculated from the 63 final structures.

Discussion

The structure of the EGF-like module

The structure of the EGF-like module closely resembles that of TGF-α and EGF. Figure 6 and Kinemage 3 show a superposition of the backbone structures of the factor IX module and EGF. The major feature of secondary structure is a double-stranded antiparallel β-sheet comprised of residues 60–65 paired with residues 73–68. An NOE was observed between the backbone amide protons of Glu 78 and Leu 84, which is consistent with a minor antiparallel β-sheet at the C-terminus. Slowly exchanging amides were observed for residues 61, 63, 70, 72, and 84, consistent with hydrogen bonds predicted for the as-

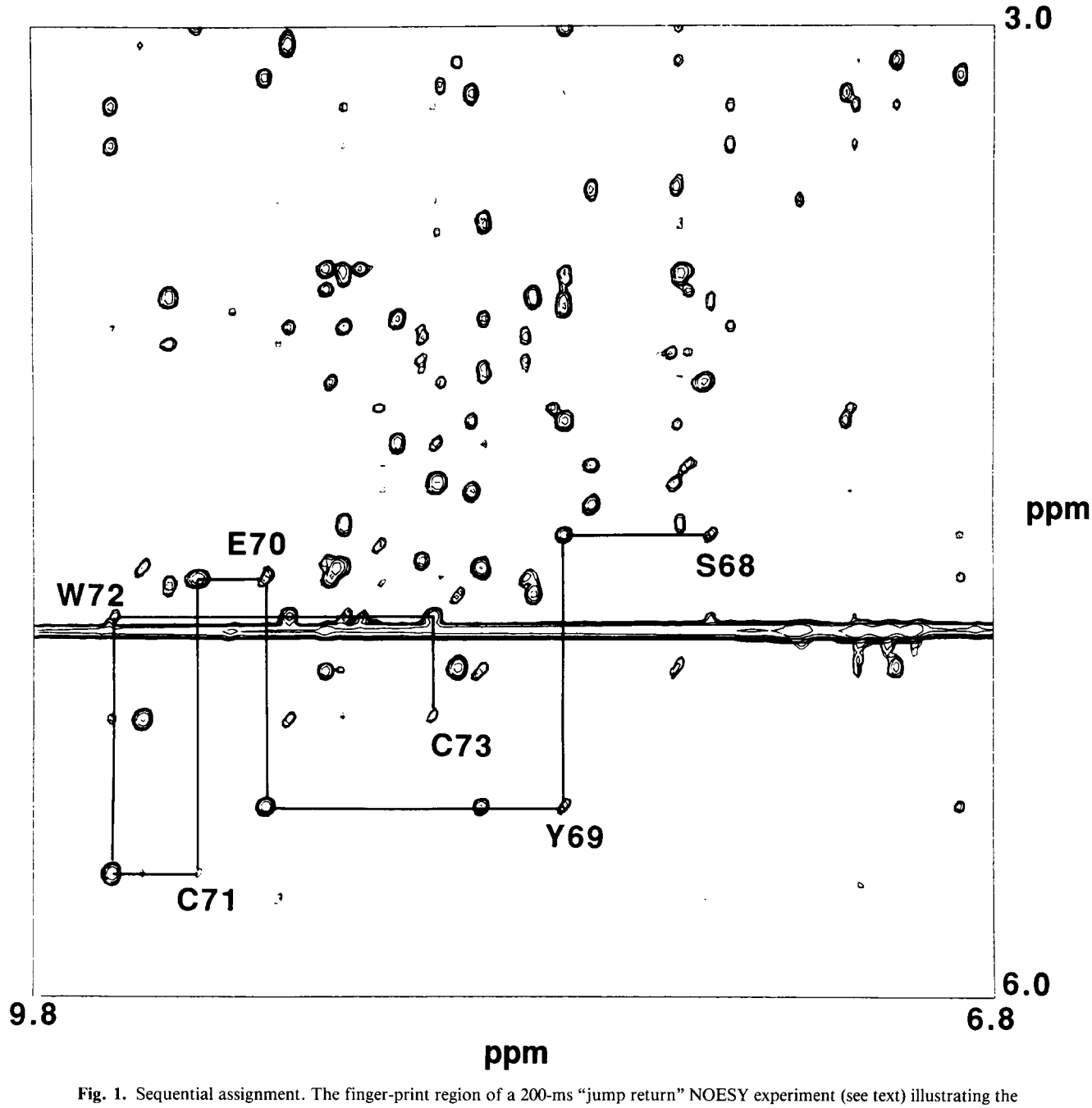

Fig. 1. Sequential assignment. The finger-print region of a 200-ms "jump return" NOESY experiment (see text) illustrating the sequential assignment of a segment of the sequence. Intra-residue HN–Hα NOEs are labeled and these are connected via horizontal and vertical lines to the sequential HN,H$\alpha_{(i,i-1)}$ NOES.

signed β-sheets. The turn between residues 65–68 appears to have a different conformation from that described for the homologous position in TGF-α (Harvey et al., 1991) and EGF (Harvey et al., unpubl.). In EGF and TGF-α the turn is type I, which has a "cobra head" conformation with the fourth residue of the turn position lying in the unusual αL region of a Ramachandran plot (Sibanda et al., 1989). In the FIX EGF-like module, however, the equivalent residue (Ser 68) has a $^3J_{HN\alpha}$ coupling constant, which is consistent with it being in the β conformation. The difference may be due to the sheet being one pair of residues shorter than in EGF and TGF-α and the absence of an Asp or Asn at the fourth turn position (these residues, along with Gly, favor the αL conformation) (Sibanda et al., 1989).

As has been reported previously for EGF and TGF-α,

Table 1. *Chemical shift assignments*[a]

Residue	NH	αH	βH	γH	Others	
Val 46			2.23		γCH_3	1.01*
Asp 47	8.78	4.66	2.82			
			2.70			
Gly 48	8.49	4.03				
		3.94				
Asp 49	8.21	4.71	2.82*			
Gln 50	8.79	4.11	2.24	2.42*	δHN	7.55
			1.95			6.75
Cys 51	8.48	4.43	3.18			
			2.92			
Glu 52	7.79	4.02	2.13	2.44*		
			2.08			
Ser 53	7.71	4.37	3.81			
			3.76			
Asn 54	8.00	4.49	2.88*		γHN	7.48
						6.74
Pro 55			1.72	1.21	δH	3.52
			1.59	0.9		3.50
Cys 56	7.72	4.42	2.76			
			2.47			
Leu 57	8.48	4.29	1.61	1.79	δCH_3	0.89
			1.40			0.78
Asn 58	8.57	3.94	1.91		γHN	8.66
			1.39			7.24
Gly 59	8.34	4.07				
		3.62				
Gly 60	7.75	4.71				
		3.75				
Ser 61	8.70	4.81	4.02*			
Cys 62	8.9	5.12	3.05*			
Lys 63	9.42	4.65	1.76	1.28*	δH	1.52*
			1.67		εH	2.74*
Asp 64	8.76	4.68	2.90			
			2.66			
Asp 65	8.36	4.98	2.89			
			2.56			
Ile 66	8.83	3.76	2.18	1.52	γCH_3	0.91
				1.19	δCH_3	0.87
Asn 67	8.70	4.81	2.91		γHN	7.72
			2.84			6.98
Ser 68	7.65	4.57	3.87			
			3.82			
Tyr 69	8.06	5.39	3.15		2,6H	6.89
			2.76		3,5H	6.85
Glu 70	9.03	4.67	2.08	2.34		
			1.98	2.24		
Cys 71	9.18	5.64	3.00			
			2.79			
Trp 72	9.51	4.83	3.36		2H	7.21
			3.24		4H	7.62
					5H	7.14
					6H	7.23
					7H	7.48
Cys 73	8.5	5.13	2.92			
			2.56			

(*continued*)

Table 1. *Continued*

Residue	NH	αH	βH	γH	Others	
Pro 74		4.45	2.46	2.18	δH	3.45
			2.05	1.96		3.25
Phe 75	8.33	4.23	3.18		2,6H	7.26
			2.94		3,5H	7.37
					4H	7.32
Gly 76	8.08	3.80				
		3.04				
Phe 77	7.77	4.95	3.10		2,6H	7.09
			2.68		3,5H	7.22
					4H	7.22
Glu 78	8.43	4.76	1.91*	2.19*		
Gly 79	8.18	4.70				
		3.85				
Lys 80	9.24	3.99	1.92	1.42	δH	1.65*
					εH	2.81
						2.76
Asn 81	9.02	5.03	3.52		γHN	7.40
			2.50			6.97
Cys 82	7.72	4.12	3.63			
			3.02			
Glu 83	10.35	4.18	2.14	2.55		
			1.95	2.31		
Leu 84	8.64	4.58	1.56	1.56	δCH_3*	0.89
			1.48			

[a] Assignments given are obtained from spectra recorded at 292 K and pH 4.5 and referenced to dioxane at 3.75 ppm. * Indicates where the chemical shifts of two protons bonded to the same carbon are the same.

the turn in the factor IX module between Pro 74 and Phe 77 resembles a type II turn with a conserved glycine at the third position having a positive ϕ angle. The amide proton of Phe 77 is slowly exchanging and is hydrogen bonded to the carbonyl of Pro 74 in 57 out of 63 structures. The segment between Gly 79 and Glu 83 forms a β-hairpin turn with both the conserved glycine at position 79 and Cys 82 (the sixth consensus cysteine) having positive ϕ angles. It is unusual for a cysteine residue to have a positive ϕ angle, and it was not possible to correlate this with the $^3J_{HN\alpha}$ coupling constant as the amide proton resonance of Cys 82 was too broad. However, the homologous residues in structures determined for both TGF-α and EGF also have positive ϕ angles. A β bulge is found at position 83. Both Leu 84 and Glu 83 have slowly exchanging amide protons; their amides are hydrogen bonded to the carbonyl of Glu 78 in over half of the calculated structures and the side chains of Glu 78, Glu 83, and Leu 84 all point from the same face. The homologous region has been described in detail for TGF-α where similar structural features are observed (Harvey et al., 1991). A slowly exchanging amide was observed for Cys 82. The only hydrogen bond found for this residue was to the carbonyl of Gly 79; however, this was

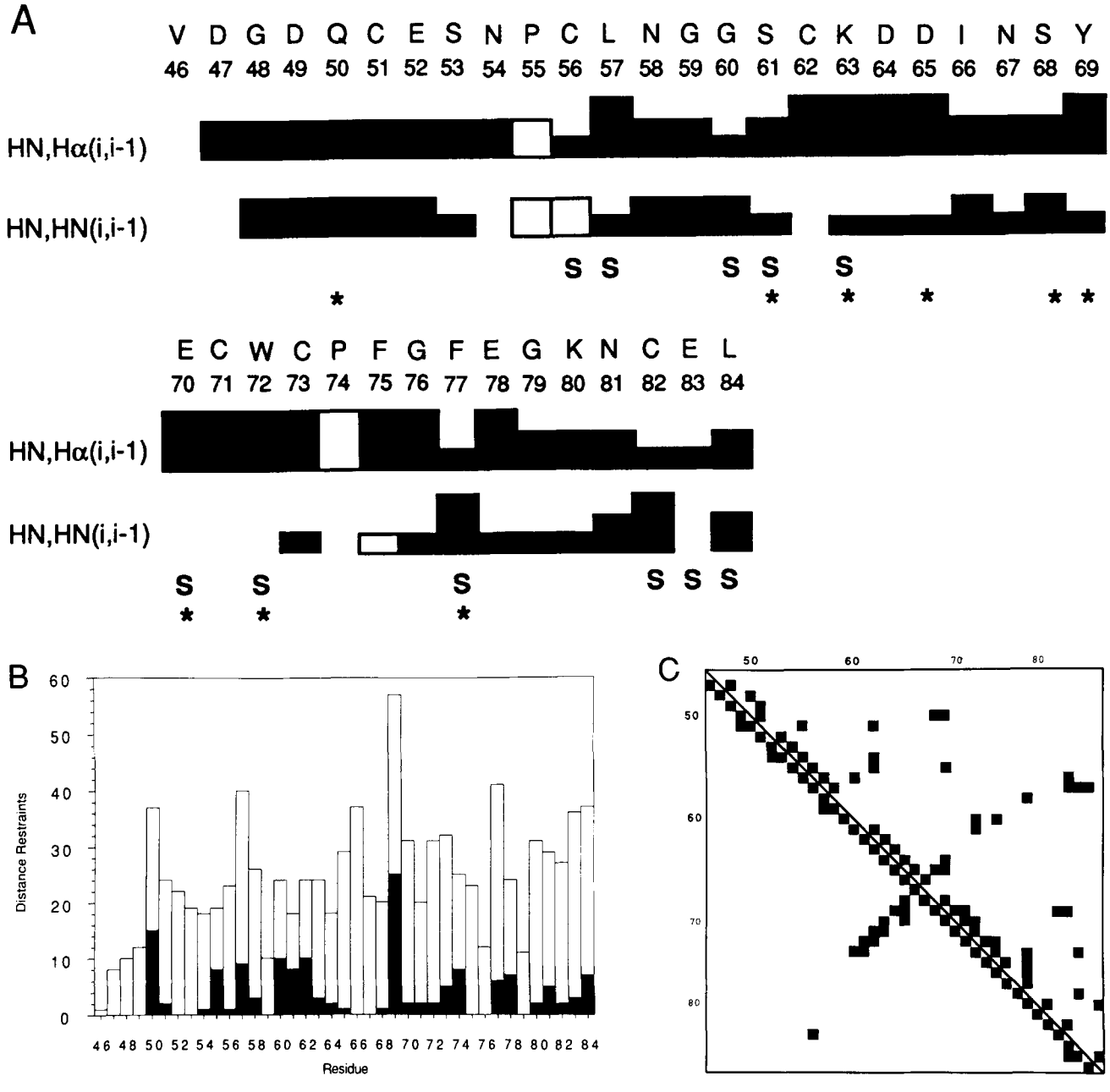

Fig. 2. Experimental restraints. **A**: Experimental data used to identify secondary structure. Black boxes represent $HN,HN_{(i,i-1)}$ and $HN,H\alpha_{(i,i-1)}$ NOEs. The height of the box relates to the size of the NOE (strong, medium, weak). Open boxes represent NOEs from proline Hδ protons. **S** indicates residues whose backbone amide protons were slow to exchange in D_2O, and * indicates $^3J_{HN\alpha}$ coupling constants greater than 9 Hz. **B**: The distribution of long-range and short-range distance restraints. Solid bars represent long-range distance restraints (>i − [i + 4]); open bars represent short-range distance restraints (<i − [i + 5]). **C**: The spatial distribution of distance restraints used in the structural determination. Solid squares indicate that one or more distance restraints were observed between the residues horizontally and vertically. The bottom left of the diagram shows the restraints involving only main-chain protons Hα and HN. The top right of the diagram shows restraints involving at least one side-chain proton.

only observed in 10 out of the 63 structures. The NMR peaks from residues of this region are broad, in particular those of Asn 81 and Glu 83. This implies that some form of chemical exchange of the millisecond timescale is occurring; similar broadening is found from residues at equivalent positions in both TGF-α and EGF. The amide proton of Glu 83 also has a large downfield chemical shift, which is a characteristic of the spectra of EGF,

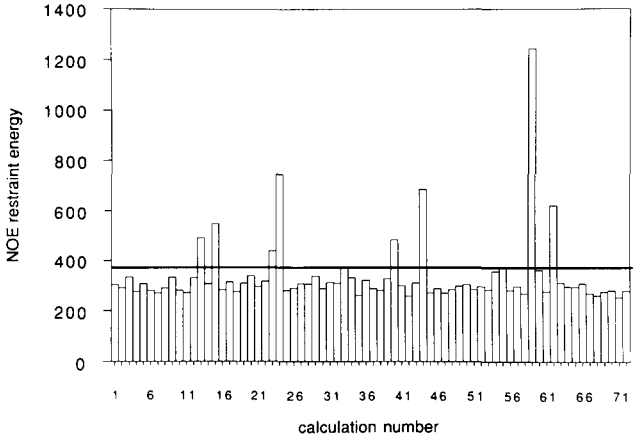

Fig. 3. Distribution of NOE restraint energy for structures produced by restrained simulated annealing (i.e., prior to final minimization). Restraint energies are in kJ/mol; selection cutoff at 369 kJ/mol, denoted by a line.

Table 2. *Mean, range, and standard deviations of potential energies in final minimized structures*[a]

	Min	Max	Mean	SD
F_{TOT}	325.89	619.92	475.99	59.18
F_{NOE}	26.00	80.01	42.62	9.45
F_{BOND}	22.89	40.99	31.58	3.53
F_{ANGLE}	404.04	475.78	432.73	16.71
F_{DIHED}	590.10	779.44	673.37	44.40
F_{IMPR}	4.49	9.53	6.93	1.09
F_{VDW}	−680.15	−558.01	−616.35	24.95
F_{ELEC}	−43.39	−29.99	−37.30	2.18
F_{HBOND}	−86.44	−29.99	−60.94	12.94
F_{CDIH}	0.00	5.71	0.21	0.76

[a] The potential energies are: F_{TOT}, total potential energy; F_{NOE}, energy due to distance restraint violations (with a force constant of 4,200 kJ/mol/nm^2); F_{CDIH}, energy due to violations of restrained dihedral angles (50 kJ/mol/rad^2); F_{VDW}, energy due to Van der Waals interactions calculated from a Lennard-Jones equation; F_{BOND}, F_{ANGLE}, F_{DIHED}, and F_{IMPR}, potential energies due to covalent geometry and interaction; F_{ELEC}, electrostatic potential energy calculated using a dielectric constant of 80; F_{HBOND}, potential energy due to an explicit hydrogen-bond term; only backbone HN and O atoms included.

Table 3. *Average backbone ϕ and ψ angles*[a]

Residue	Mean ϕ	RMSD	Mean ψ	RMSD
46			48	70
47	−98	49	−13	55
48	170	72	16	54
49	−163	87	17	65
50	−115	11	13	46
51	−128	51	43	31
52	−57	44	−33	51
53	−97	51	13	45
54	66	42	68	7
55	−77	14	31	46
56	−103	46	140	44
57	−76	62	−147	50
58	34	55	−3	46
59	101	39	59	25
60	−109	26	−175	8
61	−115	6	168	6
62	−116	7	136	9
63	−129	4	140	20
64	−103	31	160	6
65	−116	10	−77	13
66	−132	27	−61	41
67	−110	20	−37	13
68	−110	8	177	6
69	−124	10	152	7
70	−138	6	157	5
71	−111	5	129	12
72	−117	6	169	6
73	−119	11	169	6
74	−55	10	−174	3
75	−103	6	46	3
76	−137	6	−34	7
77	−113	4	152	9
78	−136	8	175	8
79	119	6	−179	13
80	−81	23	−49	8
81	−120	5	36	19
82	+50	18	63	22
83	−106	31	3	48
84	−67	46		

[a] ϕ and ψ angles shown represent the average calculated from 63 final structures. The RMSDs for these angles are also given.

TGF-α, and the human factor X EGF-like module (Cooke et al., 1990; Selander et al., 1990).

The N-terminal residues between 46 and 49 are poorly defined and show no nonsequential interresidue NOEs. Thus, they do not form a strand of β-sheet as is found for this region in TGF-α. In the factor IX EGF-like module, the region between the first and third cysteines appears to consist of two turns with Asn 54 and Gly 59, having backbone torsion angles lying in the αL conformation. Cys 56 and Gly 60 have slowly exchanging amide protons. Hydrogen bonds were found between the amide of Gly 60 and the carbonyls of 56 and 57 (in 10 and 5 structures, respectively). Two turns have also been identified in this region for the equivalent module from factor X (Selander et al., 1990). The segment between the first and second cysteines forms a loop with the side chains of both Gln 50 and Pro 55 positioned over the major β-sheet. The β-sheet has a pronounced right-hand twist such that the side chains of Gln 50 and Pro 55 give a number of NOEs to Tyr 69 on the second strand of the sheet. Residues 56 and 57 run parallel to the backbone between residues 80 and 83, whereas Asn 58 and Gly 59 form a turn that leads into the first strand of the major β-sheet. The amide proton of Leu 57 is slowly exchanging and is hydrogen bonded to the carbonyl of Asn 81 (in 54 out of 63 structures). A similar interaction between the homologous regions in EGF (Cooke et al., 1987),

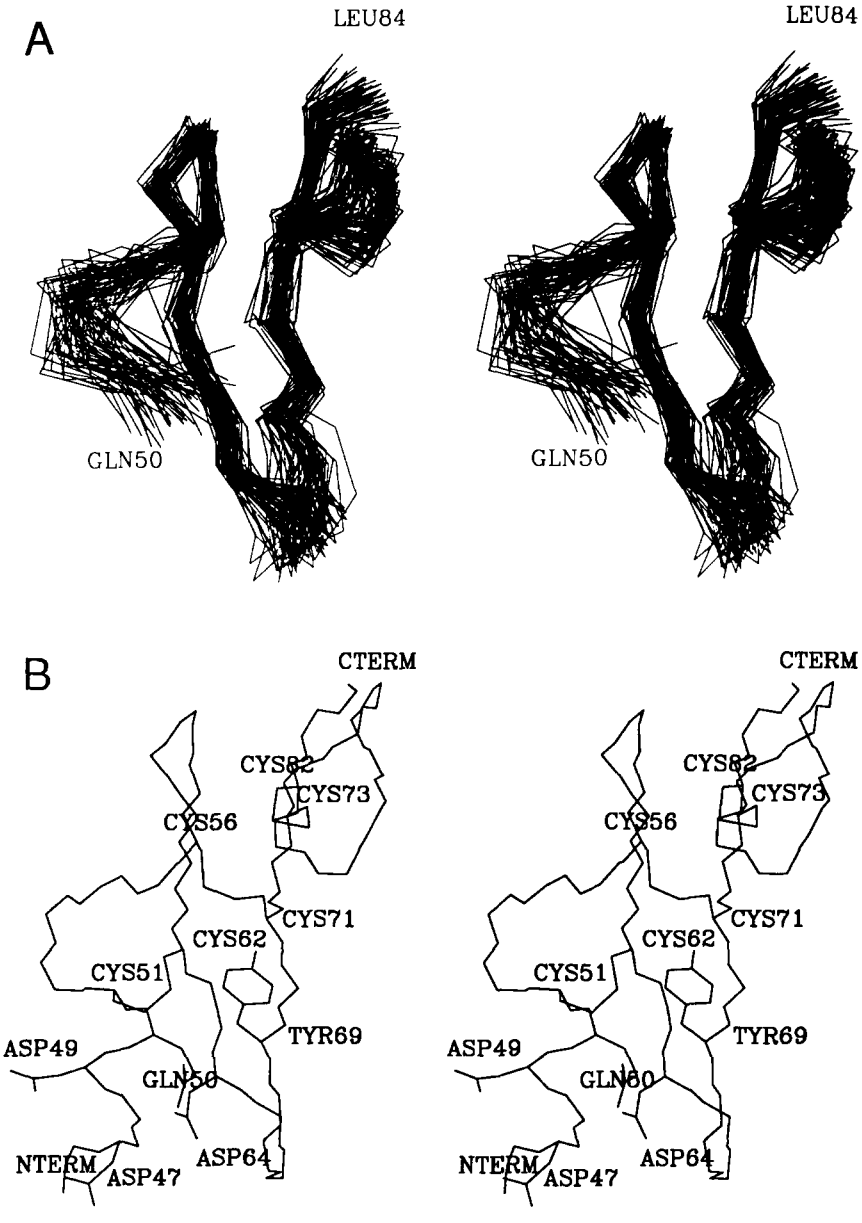

Fig. 4. Stereo overlay of the backbone atoms C, Cα, and N for the 63 final structures. **A**: Superimposition of all backbone atoms except residues 46–49. The N- and C-termini of the remainder are indicated. **B**: Backbone and selected side chains of the MinAv structure. Note that residues 46–49 have a high RMS coordinate difference (see also Fig. 5).

TGF-α (Harvey et al., 1991), and the EGF-like module of factor X (Selander et al., 1990) have been observed. The global fold and some features of the dynamics of the structure are thus highly conserved between the growth factors EGF and TGF-α and the first EGF-like modules of factors IX and X.

Implications for the calcium-binding site

Certain EGF-like modules including the first EGF-like modules of factor IX and factor X are known to contain a calcium-binding site (Persson et al., 1989; Handford et al., 1990). Conserved residues (Asp 47, Asp 49, Asp 64, and Tyr 69) are thought to be involved in forming this site as their presence is correlated with calcium binding. Binding of calcium causes significant shifts in the 2,6 ring proton of Tyr 69 and the homologous residue in the factor X module, while the resonances of other aromatic residues were unperturbed. Recent mutagenesis studies (Handford et al., 1991) have suggested that residues 47, 49, and 64 are ligands for the calcium ion, together with Gln 50, a conserved residue that had previously been overlooked. In the structures calculated for the factor IX EGF-like module, residues 46–49 are rather poorly de-

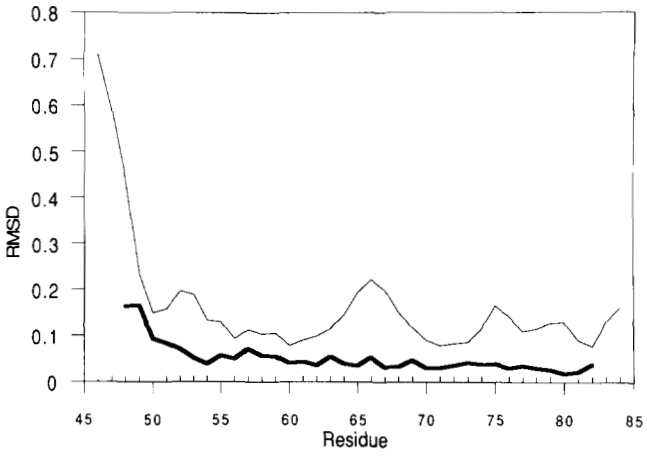

Fig. 5. Graph illustrating the RMS coordinate difference for main-chain atoms, per residue. The thin line shows the mean of the values calculated for each of the 63 final structures superimposed on the MinAv structure, using atoms C, Cα, and N of all residues. The thick line shows the results where the superimpositions were performed using only a five-residue window centered upon the residue calculated. Units are in nanometers.

fined by the data. However, from their average positions it can be seen (Fig. 4B; Kinemage 2) that Asp 47 and Asp 49 together with Asp 64 are in a location where they could form a calcium-binding site (Fig. 4B; Kinemage 2). The mutagenesis results suggesting that Gln 50 is a ligand for the calcium ion are supported by the three-dimensional structure as this residue is in close proximity to Tyr 69 and Asp 64.

Further structural studies of both wild-type and mutant EGF-like modules in the presence of calcium are in progress. It will be of interest to determine whether there is ordering of the N-terminal region in the presence of calcium and to compare the wild-type with mutated modules that are defective in calcium binding.

Materials and methods

Expression and purification of the EGF-like module

The first EGF-like module (corresponding to residues 46–84 of mature human factor IX) was expressed using a yeast-based secretion system and purified as previously described (Handford et al., 1990).

Peptide synthesis

The peptide corresponding to residues 46–84 was synthesized using standard solid-phase Fmoc chemistry on an Applied Biosystems synthesizer model 430A. The peptide was deprotected and removed from the resin simultaneously by treatment with 95% anhydrous trifluoroacetic acid and 5% scavenger cocktail consisting of anisole, ethyl methyl sulfide, and 1,2-ethanedithiol (3:1:1, respectively). The peptide was precipitated in ether and reduced for 2 h at room temperature by the addition of 0.1 M dithiothreitol in 6 M guanidinium hydrochloride, 0.1 M Tris-Cl, pH 8.3, 1 mM EDTA. The reduced peptide was acidified to pH 3.0 with HCl and dialyzed overnight at 4 °C against 0.01 M HCl, prior to purification on a C8 reversed-phase high performance liquid chromatography (HPLC) column. The purified reduced peptide was reoxidized overnight at 16 °C in a refolding buffer consisting of 0.1 M Tris-Cl, pH 8.3, 3 mM cysteine, 0.3 mM cystine, and 1 mM EDTA. The refolded material was purified on a C8 reversed-phase HPLC column.

NMR analysis

Purified EGF-like module derived from both yeast expression and peptide synthesis was dissolved in either D_2O or 90% H_2O/10% D_2O to a concentration of approximately 2 mM, and the pH was adjusted to 4.5 (uncorrected meter readings were used in the case of D_2O

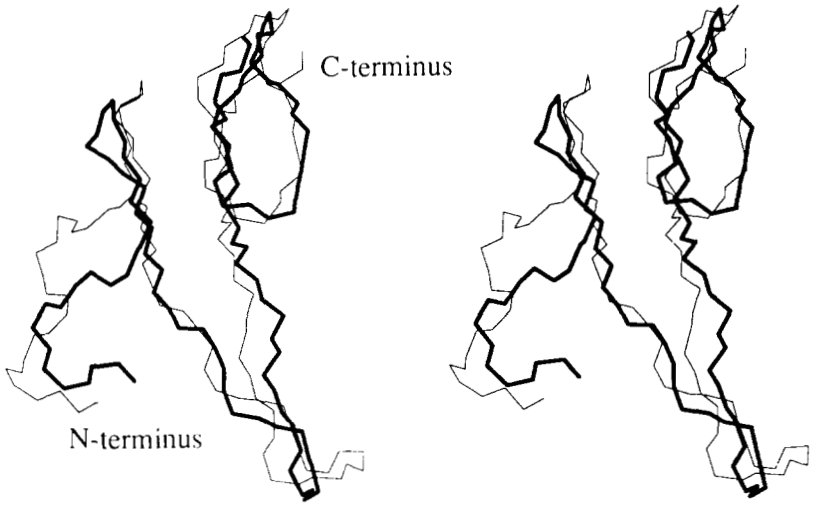

Fig. 6. Comparison of the structures of the factor IX EGF-like module and EGF. Residues before the first consensus cysteine are not shown. The N- and C-termini of the remainder are indicated. The thin line is EGF and the thick line is the factor IX EGF-like module.

samples). Experiments were recorded on Bruker AM 600 and AM 500 spectrometers at temperatures of 20 and 30 °C. Double quantum filtered correlation spectra (Rance et al., 1983) and NOE spectra (Kumar et al., 1981) were recorded in a phase-sensitive manner using the time proportional phase increment method. Mixing times of 100 and 200 ms were used for NOESY experiments. Homonuclear Hartman-Hahn (Braunschweiler & Ernst, 1983; Davis & Bax, 1985) spectra were collected in reverse mode with transfer of net magnetization by the WALTZ17 mixing sequence (Bax et al., 1987). The mixing time was 50 ms. For spectra collected in D_2O, irradiation of the solvent resonance was carried out during the relaxation delay and, for the NOESY spectra, during the mixing time. Lower decoupling powers, just sufficient to maintain saturation, were applied during the t_1 periods. In the case of NOE and HOHAHA spectra recorded in 90% H_2O/10% D_2O, a "jump return" sequence (Plateau & Gueron, 1982) was used to suppress the solvent resonance. The receiver phase was adjusted to eliminate baseline distortions. Coupling constants were measured using a PE COSY experiment (Mueller, 1987). Amide hydrogens that were relatively slow to exchange with the solvent were identified by lyophilizing a sample in 100% H_2O and resuspending in D_2O before recording a HOHAHA spectrum.

Restraints used in structure calculations

Proton-proton distance restraints were determined from NOE spectra. Mixing times of 100 and 200 ms were used, and peak volumes were estimated by volume integration within the program FTNMR (Hare, 1988). NOEs were then classified as strong (<0.25 nm), medium (<0.33 nm), or weak (<0.5 nm) in such a way as to make the data internally consistent with the observation of proton-proton distances of not more than 0.5 nm. Appropriate corrections were made for chemical shift degenerate methylene, methyl, and aromatic Hδ and Hϵ protons. In the case of nonstereospecifically assigned methylene protons, distance restraints for both protons were set as for the weaker NOE cross peak; where only one methylene proton was observed, center averaging, together with an appropriate restraint correction, was used. Restraints for ϕ angles of $-120°$ ($\pm 60°$) were included for residues with $^3J_{HN\alpha}$ of greater than 9 Hz.

χ_1 angles were calculated from derived coupling constants and Hα-Hβ and HN-Hβ NOE intensities and were used to restrain side chains to single rotamers when the appropriate conditions were satisfied (Wagner et al., 1987). Hydrogen bonds were predicted from the presence of slowly exchanging amide protons together with the identified secondary structure. Hydrogen bonds were incorporated into the structure calculations as paired restraints of 0.27–0.33 nm for N–O distances and 0.18–0.22 nm for H–O distances. Disulfide bonds (between cysteines 51–62, 56–71, and 73–82) were included as restraints of 0.2–0.22 nm for the S–S distances or as covalent bonds (depending on the stage of refinement) according to the consensus disulfide bridge pattern of EGF.

Calculation of structures

All structure refinement was carried out using the program XPLOR (Brünger, 1990). Initial structures were generated using random ϕ and ψ angles with side chains in an extended conformation and perfect covalent geometry. Simulated annealing was carried out using a protocol described previously (Brünger, 1990). The restraint energies after the first stage of refinement are artificially high as the protocol required that the distance between disulfide-bonded atoms be restrained to a distance that was less than their combined van der Waals radii. In order to form correct disulfide bonds the structures were finally energy minimized under a full CHARMM (Brooks et al., 1983) force field for 300 cycles with an NOE force constant of 4,200 kJ/mol/nm^2. Electrostatic interactions were made negligible by adjusting the dielectric constant to 80. A single representative structure was produced (MinAv), from the 63 refined structures, by superposition (not using residues 46–49) and coordinate averaging. The average structure was subjected to restrained energy minimization to produce good covalent geometry.

Acknowledgments

This work is a contribution from the Oxford Centre for Molecular Sciences, which is supported by the SERC and the MRC. P.A.H. is supported by an MRC postdoctoral grant awarded to G.G.B.; A.G.D.T. was supported by the Leukaemia Research Fund. We thank Paul Driscoll, Jonathon Boyd, and Nick Soffe for assistance with NMR and Tim Beesley for help with yeast fermentation.

References

Baron, M., Norman, D.G., & Campbell, I.D. (1991). Protein modules. *Trends Biochem. Sci. 16*, 13–17.

Bax, A., Sklenar, V., Clore, G.M., & Gronenborn, A.M. (1987). Water suppression in two-dimensional spin-locked nuclear magnetic resonance experiments using a novel phase-cycling procedure. *J. Am. Chem. Soc. 109*, 6511–6513.

Braunschweiler, L. & Ernst, R.R. (1983). Coherence transfer by isotropic mixing: Application to proton correlation spectroscopy. *J. Magn. Reson. 53*, 521–528.

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., & Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimisation, and dynamics calculations. *J. Comput. Chem. 4*, 187–217.

Brünger, A.T. (1990). *XPLOR v2.2 Manual.* Yale University, New Haven.

Campbell, I.D., Cooke, R.M., Baron, M., Harvey, T.S., & Tappin, M.J. (1989). The solution structures of epidermal growth factor and transforming growth factor alpha. *Prog. Growth Factor Res. 1*, 13–22.

Cooke, R.M., Tappin, M.J., Campbell, I.D., Kohda, D., Miyake, T., Fuwa, T., Miyazawa, T., & Inagaki, F. (1990). Nuclear magnetic resonance studies of human epidermal growth factor. *Eur. J. Biochem. 193*, 807–815.

Cooke, R.M., Wilkinson, A.J., Baron, M., Pastore, A., Tappin, M.J.,

Campbell, I.D., Gregory, H., & Sheard, B. (1987). The solution structure of human epidermal growth factor. *Nature 327*, 339–341.

Davis, D.G. & Bax, A. (1985). Assignment of complex ^{1}H NMR spectra via two-dimensional homonuclear Hartman-Hahn spectroscopy. *J. Am. Chem. Soc. 107*, 2820–2821.

Day, A.J. & Baron, M. (1991). Structure/function inter-relationships of mosaic proteins. *J. Biomed. Sci. 1*, 153–163.

Doolittle, R.F. (1985). The genealogy of some recently evolved vertebrate proteins. *Trends Biochem. Sci. 10*, 233–237.

Furie, B. & Furie, B.C. (1988). The molecular basis of blood coagulation. *Cell 53*, 505–517.

Handford, P.A., Baron, M., Mayhew, M., Willis, A., Beesley, T., Brownlee, G.G., & 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.

Handford, P.A., Mayhew, M., Baron, M., Winship, P., Brownlee, G.G., & Campbell, I.D. (1991). Key residues involved in calcium binding motifs in EGF-like domains. *Nature 351*, 164–167.

Hare, D.R. (1988). *FTNMR Manual v5.5.* Hare Research, Inc., Woodville.

Harvey, T.S., Wilkinson, A.J., Tappin, M.J., Cooke, R.M., & Campbell, I.D. (1991). The solution structure of human transforming growth factor alpha. *Eur. J. Biochem. 198*, 555–562.

Kumar, A., Ernst, R.R., & Wüthrich, K. (1981). A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun. 95*, 1–6.

Montelione, G.T., Wüthrich, K., Nice, E.C., Burgess, A.W., & Scheraga, H.A. (1987). Solution structure of murine epidermal growth factor: Determination of the polypeptide backbone chain fold by nuclear magnetic resonance and distance geometry. *Proc. Natl. Acad. Sci. USA 84*, 5226–5230.

Mueller, L. (1987). P.E. Cosy, a simple alternative to E.COSY. *J. Magn. Reson. 72*, 191–196.

Patthy, L. (1987). Intron-dependent evolution: Preferred types of exons and introns. *FEBS Lett. 214*, 1–7.

Persson, E., Selander, M., Linse, S., Drakenberg, T., Ohlin, A.K., & Stenflo, J. (1989). Calcium binding to the isolated β-hydroxyaspartic acid-containing epidermal growth factor-like domain of bovine factor X. *J. Biol. Chem. 264*, 16897–16904.

Plateau, P. & Gueron, M. (1982). Exchangeable proton NMR without baseline distortion, using new strong-pulse sequences. *J. Am. Chem. Soc. 104*, 7310–7311.

Rance, M., Sorensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R., & Wüthrich, K. (1983). Improved spectral resolution in COSY ^{1}H NMR spectra of proteins via double quantum filtering. *Biochem. Biophys. Res. Commun. 117*, 479–485.

Rees, D.J.G., Jones, I.M., Handford, P.A., Walter, S.J., Esnouf, M.P., Smith, K.J., & Brownlee, G.G. (1988). The role of β-hydroxyaspartate and adjacent carboxylate residues in the first EGF domain of human factor IX. *EMBO J. 7*, 2053–2061.

Reid, K.B.M. & Day, A.J. (1989). Structure function relationships of the complement system. *Immunol. Today 10*, 177–180.

Scott, J. (1989). Unravelling atherosclerosis. *Nature 338*, 118–119.

Selander, M., Persson, E., Stenflo, J., & Drakenberg, T. (1990). ^{1}H NMR assignment and secondary structure of the Ca^{2+} free form of the amino-terminal epidermal growth factor like domain in coagulation factor X. *Biochemistry 29*, 8111–8118.

Sibanda, B.L., Blundell, T.L., & Thornton, J.M. (1989). Conformation of β-hairpins in protein structures. *J. Mol. Biol. 206*, 759–777.

Stenflo, J. & Suttie, J.W. (1977). Vitamin K-dependent formation of γ-carboxyglutamic acid. *Annu. Rev. Biochem. 46*, 157–172.

Wagner, G., Braun, W., Havel, T.F., Schauman, T., Go, N., & Wüthrich, K. (1987). Protein structures in solution by nuclear magnetic resonance and distance geometry. *J. Mol. Biol. 196*, 611–639.

Wharton, K.A., Johansen, K.M., Xu, T., & Artavanis-Tsakonas, S. (1985). Nucleotide sequence from the neurogenic locus notch implies a gene product that shares homology with proteins containing EGF-like repeats. *Cell 43*, 567–581.

Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids.* Wiley-Interscience, New York.