Homology modelling of integrin EF-hands

Evidence for widespread use of a conserved cation-binding site

Danny S. TUCKWELL,* Andy BRASS and Martin J. HUMPHRIES

Department of Biochemistry and Molecular Biology, University of Manchester, Stopford Building, Oxford Road, Manchester M13 9PT, U.K.

Integrin α -subunits contain three or four peptide sequences that are similar to the EF-hand, a 13-residue bivalent cation-binding motif found in calmodulin and parvalbumin. The integrin sequences differ from classical EF-hands in that they lack a co-ordinating residue at position 12. One hypothesis to explain integrin–ligand binding is that aspartate-containing recognition sequences in integrin ligands, which bind at or near to the EF-hand-like sequences, may take the place of the missing residue and co-ordinate directly to the bound cation. In this report, homology modelling of integrin EF-hand-like sequences has been performed using the X-ray structure of calmodulin as a template in order to assess the functional activity of the integrin sequences. In the calmodulin–integrin hybrid structures, integrin EF-hand-like sequences were able to retain cations whereas control sequences did not. Structural analyses demonstrated that the integrin sequences in the hybrid proteins closely resembled conventional EF-hands. The integrin sequences are therefore highly likely to bind Ca^{2+} ions *in vivo*, a prerequisite for the ligand-binding model. Database searching with a matrix derived from known integrin EF-hand-like sequences has been used to identify other proteins containing the integrin EF-hand-like motif. Annexin V (anchurin CII), atrial natriuretic peptide receptors and the 70 kDa heat-shock protein were identified by the matrix; the functions of these proteins are known from previous studies to be bivalent cation-dependent. These findings suggest that the integrin EF-hand-like sequence may be a more common motif than originally thought.

INTRODUCTION

The integrins are a family of cell membrane proteins that mediate both adhesion to extracellular matrix components, such as fibronectin, collagens, and laminin, and cell–cell interactions (Hynes, 1987; Ruoslahti & Pierschbacher, 1987; Ginsberg *et al.*, 1988; Akiyama *et al.*, 1990; Hemler, 1990; Springer, 1990). The mechanism by which integrins interact with their ligands is not yet established, but it is known that binding is dependent on the presence of bivalent cations such as Ca^{2+} or Mg^{2+} (Gailit & Ruoslahti, 1988; Kirchhofer *et al.*, 1991), and recently a functional association between the co-ordination sphere of the cation and ligand binding has been observed (Smith & Cheresch, 1991). Integrins are heterodimers, composed of non-identical α - and β -subunits (Humphries, 1990). Sequencing of the integrin α -subunits has identified repeating units (three or four in each α -subunit) that resemble the bivalent cation-binding 'EF-hands' found in calmodulin and parvalbumin (listed in Kirchhofer *et al.*, 1991), suggesting that these sites may be involved in ligand binding.

Studies of the EF-hands from calmodulin and other related Ca^{2+} -binding proteins have led to a consensus for the sequences involved in binding, as well as data for their secondary and tertiary structure (Figs. 1 and 2). The EF-hand motif is made up of 13 residues, with co-ordination typically supplied by residues 1, 3, 5, 7 and 12, and by a solvent molecule hydrogen-bonded to residue 9 (Bairoch, 1989; Strynadka & James, 1989).

Without exception, the integrin EF-hand-like sequences do not possess a suitable co-ordinating residue at position 12 (see Table 1). This residue is frequently a small hydrophobic amino acid rather than the aspartate or glutamate seen in conventional EF-hands. The presence of an aspartate residue in those regions of ligands that are known to be involved in binding to the integrins, e.g. RGDS (Pierschbacher & Ruoslahti, 1984; Ruoslahti & Pierschbacher, 1987), LGGAKQAGDV (Kloc-

zewiak *et al.*, 1984), LDV (Komoriya *et al.*, 1991) and KDGEA (Staatz *et al.*, 1991), has led to the proposal that the ligand aspartate provides the missing cation-co-ordinating group, thus enabling integrin–ligand binding (Corbi *et al.*, 1987; Edwards *et al.*, 1988; Loftus *et al.*, 1990). Support for this proposal has been provided by the observation that chemical cross-linking of RGD-containing peptides to the integrin $\alpha_v\beta_3$ showed the ligand to be attached in the vicinity of one of the EF-hand-like sequences (Smith & Cheresch, 1990). Similarly, cross-linking of the fibrinogen peptide LGGAKQAGDV to the platelet integrin $\alpha_{IIb}\beta_3$ revealed a binding site that contains one of the integrin EF-hand-like sequences (D'Souza *et al.*, 1990). In addition, a peptide encoding the EF-hand-like fibrinogen peptide-binding site, *TDVNGDGRHDL* (putative co-ordinating residues italicized) was found to block the binding of either platelets or purified $\alpha_{IIb}\beta_3$ to fibrinogen (D'Souza *et al.*, 1991). The EF-hand peptide was also shown to bind directly to fibrinogen in a Ca^{2+} -dependent manner. Taken together, these data provide evidence for the direct involvement of the integrin EF-hand-like sequences in bivalent cation-dependent ligand binding. Ca^{2+} -dependent ligand binding also appears to involve the integrin β -subunits (D'Souza *et al.*, 1988; Loftus *et al.*, 1990) and gangliosides (Cheresch, 1989). However, the exact contribution of these elements to binding is not yet clear. In addition, bivalent cations may also be involved in the regulation of integrin activity and specificity (Dransfield & Hogg, 1989; Kirchhofer *et al.*, 1990; Elices *et al.*, 1991).

In this report, we describe computer-based studies designed to investigate the integrin EF-hand-like sequences. The EF-hand model for integrin–ligand binding requires the integrin sequences to be able to chelate bivalent cations. In order to investigate the cation-binding ability of these sequences, and thereby the model itself, homology modelling of the integrin sequences has been carried out. The use of a protein of known structure as a template to generate a model of a related protein or domain is known as homology modelling (Greer, 1981). The structure of

Abbreviations used: GBP, galactose-binding protein; r.m.s., root mean squared.

* To whom correspondence should be addressed.

1 2 3 4 5 6 7
(D) - x - (DNS) - {ILVFIYW} - (DENSTG) - (DNQGHKR) - {GP} -
8 9 10 11 12 13
(ILVMC) - (DENQTSFGCA) - x - x - (DE) - (ILVMFIYW)

Fig. 1. Consensus for the 13-residue EF-hand motif

The consensus gives the preferred amino acids at each position (Bairoch, 1989); {...}, acceptable amino acids; {...}, unacceptable amino acids; x, any amino acid acceptable.

the integrin subunits is not known; however, X-ray-crystallographic data are available for calmodulin (Babu *et al.*, 1988). The integrin EF-hand-like sequences can therefore be substituted into a calmodulin EF-hand to give a hybrid molecule in which the integrin sequences will be initially in a conformation suitable for a functional Ca^{2+} -chelation site. Molecular dynamics simulations were then carried out on the hybrid proteins such that if the sequence substituted for the calmodulin sequence is functional, the Ca^{2+} will be retained (functional assays). Simulations were also carried out to examine the conformations adopted by the inserted sequences for comparison with the structure of the classical EF-hand (structural assays).

The integrin α -subunits are not alone in possessing EF-hand-like sequences defective in position 12, as a similar sequence has previously been identified in a galactose-binding protein (GBP) from *Escherichia coli* (Vyas *et al.*, 1987; Edwards *et al.*, 1988). This sequence is known to bind Ca^{2+} . In this protein the missing co-ordinating group is provided by a glutamate residue distant in the sequence from the EF-hand-like sequence. To investigate the possibility that the integrin sequences are present in other proteins, database searches have been carried out to identify other possible occurrences of this motif.

MATERIALS AND METHODS

Molecular dynamics simulations

Molecular dynamics simulations were carried out on a Silicon Graphics 4D 240/GTX graphics work station using the CHARMM (Brooks *et al.*, 1985) and Quanta packages from Polygen Corp. (Waltham, MA, U.S.A.). Hybrid structures were

based on the 0.22 nm (2.2 Å) X-ray structure of rat testis calmodulin (Brookhaven code 3CLN; Babu *et al.*, 1988) which includes a Ca^{2+} bound in each EF-hand.

Functional assays. The three EF-hand-like sequences from integrin α_4 (α_4 1-3, residues 320-332, 383-395, 445-457 respectively; Takada *et al.*, 1989) were compared with the four EF-hands of calmodulin (EF1-4, residues 16-28, 52-64, 89-101, 125-137 respectively). Calmodulin EF3 was found to show the greatest similarity at the sequence level to any of the integrin sequences and was therefore used as the site of mutation in these studies.

Hybrid proteins were generated by the substitution of a 16-residue sequence, comprising the 13 residues of calmodulin EF3 plus the two preceding and one following residues, for an equal number of residues comprising the sequence to be tested. Hybrid proteins were then minimized, to remove unfavourable contacts, using a steepest-descent minimizer. Simulations commenced at 0 K and structures were heated to the desired temperature at a rate of approx. 30 K/ps. Simulations were then run for 10 ps with temperature rescaling, to allow the average value of the kinetic and potential energies to equilibrate, and then for a further 10 ps without temperature rescaling, to allow the temperature fluctuations to reach equilibrium. After equilibrations the simulations were run for a further 40 ps during which the co-ordinates of each atom in the hybrid protein were sampled every 0.2 ps to give a record of the conformations adopted. From this it could be seen if the Ca^{2+} in the mutated site was retained or lost. The time step used in all the simulations was 1 fs and the SHAKE routine (van Gunsteren & Berendsen, 1977) was used to model the high-frequency hydrogen-bond-stretch motions. The cut-off distance for non-bonded interactions was set to 0.8 nm (8 Å) and the values of the electrostatic and Van der Waals potentials were set to zero at this distance using the CHARMM VSWITCH and SHIFT functions respectively. A distance-dependent dielectric constant of 80 was used. All simulations were run *in vacuo*.

Structural assays. Novotny *et al.* (1984) have demonstrated that, in a simulation *in vacuo*, substitution of a protein sequence with irrelevant residues does not necessarily lead to refolding of the protein, as the structures are trapped in local minima. In order

Fig. 2. Comparison of the structures of an EF-hand from calmodulin and an EF-hand-like sequence from integrin α_4 subunit

(a) An EF-hand (EF3) from calmodulin (residues 89-100). (b) An EF-hand-like sequence from integrin α_4 subunit (α_4 2) after insertion into calmodulin and simulated heating to 200 K. Potential co-ordinating residues are shaded and labelled in accordance with Fig. 1. Co-ordination from residues to the Ca^{2+} ions is shown. Note the missing co-ordination site in (b).

	T	C	A	G	N	S	P	F	L	Y	H	Q	V	K	D	E	I	W	R	M	-
1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
3	1	0	0	0	1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
4	0	0	0	0	0	0	0	-1	-1	-1	0	0	-1	0	0	0	-1	-1	0	0	0
5	1	0	0	1	1	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
6	1	0	0	1	1	1	0	0	0	0	1	1	0	1	1	0	0	0	1	0	0
7	0	0	0	-1	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	1	1	0	0	1	0	1	0	1	0	0	0	1	0	0	0	1	0	0	1	0
9	1	1	1	1	1	1	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0
10	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	1	0	0	0	0
11	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0
12	0	0	0	0	-1	0	0	0	1	0	-1	-1	1	-1	-1	1	0	-1	0	0	0
13	0	0	1	1	0	0	0	1	1	1	0	0	1	0	0	0	1	1	0	1	0

Fig. 3. Matrix specifying an integrin EF-hand-like sequence used for scanning of the OWL database

For each position (1–13) in the motif, values are assigned to score preferences for each amino acid; 1, favourable; 0, indifferent; –1, unfavourable.

to assess the structural similarity of the inserted sequences to the EF-hand family, hybrid proteins were heated to displace them from these minima as follows. The 300 K structures of a number of hybrid proteins (generated as above) were heated to 700 K in 10 ps and then equilibrated for 10 ps. The EF-hands of calmodulin are contained in the two globular domains, residues 1–68 and 82–143. Comparison was then made between these domains in the hybrid protein at 700 K and minimized unmutated calmodulin.

Database searches

Sequence searching and manipulation were carried out using the Daresbury Laboratory (Daresbury, Warrington, Cheshire, U.K.) SEQNET facilities. A consensus for the EF-hand (Bairoch, 1989) and 35 EF-hand-like sequences from the following integrin α -subunits (accession numbers given after each sequence), rat α_1 (P18614), human α_2 (P17301), hamster α_3 (P17852), human α_4 (P13612), human α_5 (P08648), human α_{IIb} (P08514), human α_L (P20701), mouse α_m (P05555), human α_m (P11215), *Drosophila* α_p (P12080) and human α_v (P06756), were combined to give a matrix (Fig. 3). This matrix was used to scan the OWL database via the LUPES package (Akrigg *et al.*, 1988). The matrix gives a score reflecting the preference for each amino acid at positions 1–13. The score for a given 13-residue section of sequence is given by the sum of the values for the residue at each position. This score of agreement with the matrix is calculated within a moving window (of 13 residues) which scans along every protein in the database. The 1000 regions of sequence with the best agreement are stored. Percentage agreement with the matrix, as given in Table 3, is calculated as (actual score $\times$ 100)/(maximum possible score). For this matrix the maximum possible score is 11.

Putative integrin EF-hand like sequences were screened by: (i) comparison of sequences with the specifications of the matrix; (ii) comparison of sequences with corresponding sequences in homologous proteins and/or the same protein from other species to check for conservation; (iii) scanning of the literature for functional data consistent with the presence of a cation-binding EF-hand-like motif.

RESULTS AND DISCUSSION

Molecular dynamics simulations of Ca^{2+} -binding and protein structure

The structures of unmutated calmodulin at 100 K, 200 K and 300 K were generated as described above. At 100 K and 200 K all four EF-hands retained their Ca^{2+} , whereas at 300 K, although a Ca^{2+} ion was retained in EF4, ions were lost from EF1–3

(results not shown). In simulations therefore those hybrid proteins in which the Ca^{2+} was retained in EF3, the site of mutation at 200 K would be exhibiting similar behaviour to unmutated calmodulin. Stable Ca^{2+} binding at 200 K by the inserted sequences in the hybrid proteins was thus used as an indicator that a sequence was likely to act as a functional Ca^{2+} -chelating site.

The EF-hand-like sequences from the integrin α_4 and α_5 subunits were substituted for EF-hand EF3 of calmodulin (see the Materials and methods section), and molecular dynamics simulations were carried out on the hybrid proteins to determine the Ca^{2+} -chelating ability of these sequences. The EF-hand-like sequences α_4 1, α_4 2 and α_5 1–3 exhibited Ca^{2+} binding at 200 K, thus resembling calmodulin (Table 1, Fig. 2).

α_4 3 did not bind Ca^{2+} at 200 K. The 13-residue sequence of α_4 3 closely resembled EF3 of calmodulin, but the flanking regions of α_4 3 differed from the corresponding sequences from calmodulin, α_4 1 and α_4 2. When the first (–2) residue of the α_4 3 sequence, a glutamine residue rather than the small hydrophobic residue found in calmodulin, α_4 1 and α_4 2, was substituted for an alanine, the Ca^{2+} ion was retained at 200 K. A detailed study of the α_4 3 hybrid protein suggested that the glutamine residue did not have a direct effect on the Ca^{2+} ion. The instability of this sequence may well be a secondary effect produced by the homology modelling, and it is unlikely that this has any implications for the functioning of this sequence *in vivo*.

Three putative control sequences derived from integrin α_4 , con1–3, were investigated, and all failed to bind Ca^{2+} at 200 K (Table 1). Con1 has a predicted secondary structure similar to that of α_4 1–3, con2 is predicted to be predominantly β -sheet and con3 is predicted to be predominantly α -helix. In addition, an EF-hand-like sequence from integrin α_5 (α_5 4) which has only limited similarity to the EF-hand consensus also failed to bind Ca^{2+} at 200 K (see Table 1).

Sequences of known EF-hands, when inserted into EF3, generally showed Ca^{2+} binding at 200 K (Table 1). EF1 from calmodulin and an EF-hand sequence from rabbit myosin regulatory light chain were stable at 200 K. EF2 and 4 from human calpain heavy chain bound Ca^{2+} at 200 K whereas EF1 and 3 did not. EF3 showed little similarity to the EF-hand consensus, thus accounting for its inability to bind. EF1, however, closely resembled the other EF-hands. The reason for the instability of this sequence is uncertain although it may, as in α_4 3, be due to an indirect effect from a neighbouring residue.

No gross changes in the structures of the complete hybrid proteins, compared with unmutated calmodulin, were observed in simulations at 200 K despite the widely differing sequences inserted into calmodulin EF3. The two globular domains of calmodulin contain the EF-hands. Comparison of these domains in the hybrid proteins at 200 K with those of unmutated calmodulin at 200 K showed that root mean squared (r.m.s.) deviations were all approximately 0.2 nm (2 Å) [0.243 ± 0.012 nm (2.43 ± 0.12 Å) mean $\pm$ s.e.m., for the two domains from six hybrid proteins]. Since the original calmodulin structure was determined to 0.22 nm (2.2 Å) resolution, the hybrid proteins did not therefore differ greatly from calmodulin itself. The absence of any effect on protein structure was also apparent in the retention of Ca^{2+} in EF-hands 1, 2 and 4, i.e. the unmutated sites, in almost all structures at 200 K regardless of the sequence inserted into EF3.

The lack of difference in the structures of the hybrid proteins was due to their being trapped in local minima. In order to study the effects of mutation on protein structure, the hybrids were heated to a high temperature to displace them from these minima and allow the effects of mutation to be seen (Novotny *et al.*, 1984; Bruczeroli & Karplus, 1990). Functional and structural

Table 1. Ca²⁺-binding activity of integrin EF-hand-like sequences

The sequences used in the molecular dynamics simulations and their Ca²⁺-binding ability after homology modelling. Upper-case letters indicate the residues corresponding to the EF-hand consensus, lower-case letters indicate those residues corresponding to the flanking amino acids, which were also included in the simulations. The first sequence (calmodulin EF3) is the sequence for which the test sequences were substituted. + indicates Ca²⁺ ion retained, - indicates Ca²⁺ ion lost. Accession numbers of non-integrin sequences; human calpain heavy chain, P07384; myosin regulatory light chain, P02608.

Sequence	Residue numbers	Sequence																Retention (+) or loss (-) of Ca ²⁺ ions at 200 K
		-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Calmodulin EF3 (normal)	87-102	v	f	D	K	D	G	N	G	Y	I	S	A	A	E	L	r	+
α_4 1	318-333	a	v	D	L	N	A	D	G	F	S	D	L	L	V	G	a	+
α_4 2	381-396	l	g	D	I	D	N	D	G	F	E	D	V	A	I	G	a	+
α_4 3	443-458	q	i	D	A	D	N	N	G	Y	V	D	V	A	V	G	a	-
α_4 3 (Gln → Ala)	443-458	a	i	D	A	D	N	N	G	Y	V	D	V	A	V	G	a	+
α_5 1	332-347	a	t	D	V	N	G	D	G	L	D	D	L	L	V	G	a	+
α_5 2	399-414	l	g	D	L	D	Q	D	G	Y	N	D	V	A	I	G	a	+
α_5 3	463-478	g	r	D	L	D	G	N	G	Y	P	D	L	I	V	G	s	+
α_5 4	278-293	v	g	E	F	S	G	D	D	T	E	D	F	V	A	G	v	-
Con1 (α_4)	302-317	e	m	K	G	K	K	L	G	S	Y	F	G	A	S	V	c	-
Con2 (α_4)	514-529	e	v	P	G	Y	I	V	L	F	Y	N	M	S	L	D	v	-
Con3 (α_4)	931-946	i	l	E	M	D	E	T	S	A	L	K	F	E	I	R	a	-
Calmodulin EF1	14-29	l	f	D	K	D	G	D	G	T	I	T	T	K	E	L	g	+
Calpain 1	596-611	l	m	D	R	D	G	N	G	K	L	G	L	V	E	F	n	-
Calpain 2	626-641	k	f	D	L	D	K	S	G	S	M	S	A	Y	E	M	r	+
Calpain 3	661-676	t	r	V	S	E	P	D	L	A	V	D	F	D	N	F	v	-
Calpain 4	691-706	t	l	D	T	D	L	D	G	V	V	T	F	D	L	F	k	+
Myosin	35-50	v	i	D	Q	N	R	D	G	I	I	D	K	E	D	L	r	+

Table 2. Comparison of domain structures of unmutated and hybrid proteins at 700 K

Comparison of the globular domains of hybrid structures at 700 K with minimized unmutated calmodulin. Values given are root mean squared deviations in Å. Heating to 700 K overcomes local minima, allowing the effects of mutation to be seen. The site of insertion (residues 87-102) is in domain II; increased distortion in domain I implies that the effects of mutation extend throughout the protein.

Origin of inserted sequence	Domain I: residues 1-68	Domain II: residues 82-143
Normal calmodulin	4.205	4.735
α_4 1	7.354	5.446
Calmodulin EF1	7.132	5.951
Con2	10.858	10.082

studies had to be carried out separately since at 200 K there was insufficient energy to permit refolding, whereas at 700 K Ca²⁺ ions were usually lost from both unmutated and mutated binding sites. Normal calmodulin and hybrids containing α_4 1, calmodulin EF1 or con2 were heated to 700 K (see the Materials and methods section; Table 2). Comparison of these structures with minimized unmutated calmodulin showed that insertion of an integrin EF-hand-like sequence had a similar effect on protein structure to insertion of a known EF-hand sequence. However, both of these mutations had less of an effect on structure than insertion of a non-EF-hand sequence. The integrin sequences therefore show structural similarity to the conventional EF-hands.

Water molecules are known to be important in EF-hands, typically providing one of the co-ordinating ligands for the chelated ion (Strynadka & James, 1989). However, the inclusion of water in simulations requires large amounts of computer time

and is therefore impractical in many studies. Although simulations *in vacuo* have the advantages of exploring phase space more efficiently, the absence of water molecules could conceivably affect cation chelation and also protein structure. For this reason, appropriate controls were included in both the functional and structural assays to ensure their validity. In the functional assays, sequences of known EF-hands generally retained Ca²⁺ ions, whereas irrelevant sequences did not. Similarly, in the structural assays, a hybrid protein containing a known EF-hand sequence showed less deviation from the structure of unmutated calmodulin than did a hybrid containing an irrelevant sequence. From these results it is clear that the lack of water did not affect the predictive power of the simulations. Simulations *in vacuo* have also been shown elsewhere to give structures that approximate to physiological data (Daggett *et al.*, 1991).

In order to test the homology modelling process against a protein of known properties and structure, a hybrid protein containing the *E. coli* GBP integrin-like sequence (see Table 3), which is known to bind Ca²⁺ *in vivo* (Vyas *et al.*, 1987), was constructed. The GBP sequence was found to bind Ca²⁺ at 200 K (results not shown). When the hybrid was heated to 700 K and the structure of the GBP sequence compared with the same residues in the X-ray structure of GBP itself (structure comparison was restricted to residues 1-9 of the Ca²⁺-binding region since this sequence is followed by a helix in the calmodulin hybrids and a β -strand in GBP), the α -residues showed a r.m.s. deviation of 0.1602 nm (1.602 Å) and comparison of all the atoms in this region gave a r.m.s. deviation of 0.2613 nm (2.613 Å) (Fig. 4). Since the X-ray structure of GBP was originally determined to 0.19 nm (1.9 Å), the structure of the GBP Ca²⁺-binding sequence as generated by the homology modelling method is therefore similar to the known structure of this protein. The homology modelling process used here is therefore shown to be a valid method for simulating the structure and properties of integrin and integrin-like sequences.

Table 3. Putative integrin EF-hand-like sequences identified by matrix search of the OWL database

Sequences of putative integrin EF-hand-like sequences detected by the matrix after screening. An EF-hand-like sequence from integrin α_4 is shown for comparison. Accession numbers are as follows; GBP, P02927; Annexin V, chicken, P17153, human, P08758, rat, P14668; Atrial natriuretic peptide receptor, type A, human, P16066, rat, P18910, mouse, P18293, type B, rat, P16067, type C, human, P17342, bovine, P10730; 70 kDa heat-shock protein, *Xenopus*, P02827.

Origin of sequence	Species	Sequence	Residues	Similarity to matrix (%)	
GBP	<i>E. coli</i>	D L N K D G Q I Q F V L L	134–146	90.91	
Annexin V/anchorin CII	Chicken	D R E T S G D L E K L L L	226–238	81.82	
	Human	D R E T S G N L E Q L L L	225–237	81.82	
	Rat	D R E T S G N L E N L L L	224–236	81.82	
Atrial natriuretic peptide receptor	Type A	Human	D S S G D R E T D F S L W	399–411	81.82
		Rat	D R N G D R D T D F S L W	395–407	81.82
		Mouse	D R N G D R D T D S P L W	395–407	72.73
	Type B	Rat	D K N N D R E T D F V L W	383–395	90.91
		Type C	Bovine	D A N G D R Y G D F S V I	403–415
	Human		D A N G D R Y G D F S V I	407–419	72.73
70 kDa heat-shock protein	<i>Xenopus</i>	D I D A N G I L N V S A V	480–492	72.73	
Integrin α_4	Human	D A D N N G Y V D V A V G	445–457	90.91

Fig. 4. Comparison of GBP EF-hand-like sequence after homology modelling with the known structure of this region

Comparison of the known X-ray structure of GBP (a) with the structure of the GBP Ca^{2+} -binding sequence as generated by homology modelling (b). The GBP sequence (residues 132–147) was inserted into calmodulin and heated to 700 K. Comparison was restricted to residues 134–142 (see the Results and discussion section), backbone residues are shown here and are labelled in accordance with Fig. 1. Note that the X-ray structure of GBP does not include backbone protons.

The overall similarity in behaviour of the integrin EF-hand-like sequences and calmodulin in both the functional and structural assays strongly suggests that the integrin sequences can function as Ca^{2+} -binding sites *in vivo*. This conclusion is in agreement with the findings of D'Souza *et al.* (1991), who observed that a peptide encoding an integrin EF-hand-like sequence was able to bind to its ligand in a Ca^{2+} -dependent fashion, implying co-ordination of the cation by the peptide, and of Smith & Cheresch (1991), who demonstrated that bivalent cations were physically associated with integrins and that the bivalent cations also interacted with the ligand, implying co-ordination of the cation by the integrin at the ligand-binding site.

Corbi *et al.* (1987), Edwards *et al.* (1988) and Loftus *et al.* (1990) have proposed that the integrins interact with their ligands via the EF-hand-like sequences, with the aspartate residue found in integrin ligands providing the co-ordination site missing at the 12th position in the integrin sequences. Ca^{2+} binding by the integrin sequences is therefore a prerequisite for the model for integrin–ligand binding. The studies of D'Souza *et al.* (1991) and Smith & Cheresch (1991) are consistent with the proposed binding mechanism. The data reported here therefore provide further support.

Database searching

Scanning of the OWL database with a matrix derived from integrin-EF-hand-like sequences and the EF-hand consensus of Bairoch (1989; Fig. 1) detected all the integrin EF-hand-like sequences on the database with the exception of two from human integrin α_4 . The integrin sequences showed 81.82–90.91% agreement with the matrix.

Scanning, followed by sequence and literature screens, identified putative integrin EF-hand-like sequences in GBP, annexin V (also known as anchorin CII), atrial natriuretic peptide receptors types A, B and C, and the *Xenopus* 70 kDa heat-shock protein (Table 3). In order to examine the ability of these sequences to act as Ca^{2+} -chelating domains, the function of the chicken annexin V sequence was tested in the molecular dynamics model described above. This sequence was found to retain a Ca^{2+} at 200 K, thus resembling calmodulin and the integrin sequences.

Proteins that were detected by the matrix but subsequently rejected by the screening stages included the cadherin family, thrombospondin, maltases and collagen $\alpha 3(\text{VI})$. These proteins showed agreement with the matrix but some sequences were not conserved in all members of the protein family or between species (e.g. cadherins). Where the sequences were conserved, for some proteins there were no data implicating cation-dependent function (e.g. maltases).

(i) *E. coli* GBP. This protein has previously been observed to contain an EF-hand-like sequence resembling that found in the integrins (Vyas *et al.*, 1987; Edwards *et al.*, 1988). However, here the missing co-ordination site is provided by an internal, but sequentially distant, glutamine residue. The identification of this sequence demonstrates that the matrix is sufficiently broad to detect likely sequences other than those in the integrins. The consequences of a broad matrix may be that some integrin sequences will not match the matrix. This may account for the

inability of the matrix to detect the two missing integrin α_v sequences in the best 1000 scores.

(ii) **Annexin V/anchoring CII.** The annexins are a family of Ca^{2+} - and phospholipid-binding proteins. Their precise physiological function is unclear but it has been proposed that they are involved in the regulation of inflammation and blood coagulation (Crompton *et al.*, 1988; Wirl & Schwartz-Albeiz, 1990). In addition, annexin V has been implicated in the binding of chondrocytes to collagen type II (Mollenhaur *et al.*, 1984). Annexin V is made up of four homologous domains, I–IV (Huber *et al.*, 1990). Domains I, II and IV each chelate a Ca^{2+} ion, but it is not clear whether or not domain III has this activity [compare results of Huber *et al.* (1990) and Marriott *et al.* (1990)]. In domains I, II and IV the Ca^{2+} ion is chelated by residues from two adjacent loops. The integrin EF-hand-like sequence identified by the matrix lies in domain III and is in one of the loops which, in the other domains, binds Ca^{2+} . These sequences in domains I, II and IV only provide a single co-ordinating residue to the Ca^{2+} , and show only a limited similarity to the EF-hand consensus. However, homology in these regions extends throughout the annexin family, and, although the close resemblance to the integrin sequence is only seen in domain III of annexin V (Pepinsky *et al.*, 1988), it has been previously noted that all these sequences do resemble an EF-hand defective in position 12 (Geisow, 1986). The annexin- Ca^{2+} binding motif may therefore be related to the integrin EF-hand-like sequence.

(iii) **Atrial natriuretic peptide receptors.** The atrial natriuretic peptide receptors, types A, B and C, are integral membrane proteins that mediate the effects of atrial natriuretic peptide on cardiovascular homeostasis (Fuller *et al.*, 1988). Under conditions of sodium depletion, some Ca^{2+} -dependent binding of atrial natriuretic peptide to its receptor has been observed. However, at normal sodium concentrations binding is Ca^{2+} -independent (Lyll & Morton, 1987; Morton *et al.*, 1989). Ca^{2+} dependence may therefore be mediated by the integrin EF-hand-like sequences.

(iv) **70 kDa heat-shock protein.** Comparison of the sequences of the 70 kDa heat-shock proteins and related proteins showed that the sequence given in Table 3 is well conserved. However, position 13 can contain a variety of residues other than the small hydrophobic amino acids seen in the conventional EF-hands and the integrins. In addition, the sequence from the *E. coli* 70 kDa heat-shock protein, D_{NAK} (Bardwell & Craig, 1984), shows rather poor homology to the EF-hand (*DIADGILHVS*AK; putative co-ordinating residues italicized). A Ca^{2+} -binding site does, however, appear to exist in these proteins, since the domain in D_{NAK} in which the integrin-like sequence lies is involved in the Ca^{2+} -dependent autophosphorylation of this protein (Cegielska & Georgopoulos, 1989) and Ca^{2+} -dependent autophosphorylation has also been observed in a 70 kDa protein from HeLa cells which resembles D_{NAK} (Leustek *et al.*, 1989).

The above data suggest that the integrin EF-hand-like sequences occur in a variety of proteins other than the integrins. It is likely that further studies will identify other occurrences of this domain.

Conclusion

In conclusion, it has been demonstrated that the integrin EF-hand-like sequences are highly likely to be Ca^{2+} -binding sites *in vivo*. This is a prerequisite for the model of integrin–ligand binding proposed previously (Corbi *et al.*, 1987; Edwards *et al.*, 1988; Loftus *et al.*, 1990). Our data therefore provide support for this model. In addition, we have identified putative integrin EF-hand-like sequences in a number of non-integrin proteins, suggesting that the occurrence of this motif may be more widespread than previously thought. In the future, X-ray crystal-

lography of the integrin α -subunits should give considerable information about the structure of the integrin EF-hand-like sequences. This should enable not only the Ca^{2+} - and ligand-binding mechanisms of the integrins to be determined but will also allow accurate assessment of the relationship between this motif and conventional EF-hands. Studies of GBP, which contains an integrin EF-hand-like sequence, should also be of value. Since the structure of this protein is known, it may be a useful source of information about the potential structure of the integrin EF-hand-like sequences.

This work was partly supported by the Wellcome Trust.

REFERENCES

- Akiyama, S. K., Nagata, K. & Yamada, K. M. (1990) *Biochim. Biophys. Acta* **1031**, 91–110
- Akrigg, D., Bleasby, A. J., Dix, N. I. M., Findlay, J. B. C., North, A. C. T., Parry-Smith, D. J., Wootton, J. C., Blundel, T. C., Gardner, S. P., Hayes, F., Islam, S., Sternberg, M. J. E., Thornton, J. M., Tickle, I. J. & Murray-Rust, P. A. (1988) *Nature (London)* **335**, 745–746
- Babu, Y. S., Bugg, C. E. & Cook, W. J. (1988) *J. Mol. Biol.* **203**, 191–204
- Bairoch, A. (1989) *PROSITE: A Dictionary of Protein Sites and Patterns*, 4th edn., University of Geneva, Geneva
- Bardwell, J. C. A. & Craig, E. A. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 848–852
- Brooks, B. R., Bruczeroli, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1985) *J. Comp. Chem.* **4**, 187–217
- Bruczeroli, R. E. & Karplus, M. (1990) *Biopolymers* **29**, 1847–1862
- Cegielska, A. & Georgopoulos, C. J. (1989) *J. Biol. Chem.* **264**, 21122–21130
- Cheresh, D. A. (1989) *Prog. Clin. Biol. Res.* **288**, 3–24
- Corbi, A. L., Miller, L. J., O'Conner, K., Larson, R. S. & Springer, T. A. (1987) *EMBO J.* **6**, 4023–4028
- Crompton, M. R., Moss, S. E. & Crompton, M. J. (1988) *Cell* **55**, 1–3
- Daggett, V., Kollman, P. A. & Kuntz, I. D. (1991) *Biopolymers* **31**, 285–385
- Dransfield, I. & Hogg, N. (1989) *EMBO J.* **8**, 3759–3765
- D'Souza, S. E., Ginsberg, M. H., Burke, T. A., Lam, S. C.-T. & Plow, E. F. (1988) *Science* **242**, 91–93
- D'Souza, S. E., Ginsberg, M. H., Burke, T. A. & Plow, E. F. (1990) *J. Biol. Chem.* **265**, 3440–3446
- D'Souza, S. E., Ginsberg, M. H., Matsueda, G. R. & Plow, E. F. (1991) *Nature (London)* **350**, 66–68
- Edwards, J. G., Hameed, H. & Campbell, G. (1988) *J. Cell Sci.* **89**, 507–513
- Elices, M. J., Urry, L. A. & Hemler, M. E. (1991) *J. Cell Biol.* **112**, 169–181
- Fuller, F., Porter, J. G., Arfsten, A. E., James, J. M., Schilling, J. W., Scarborough, R. M., Lewicki, J. A. & Schenk, D. B. (1988) *J. Biol. Chem.* **263**, 9395–9401
- Gailit, J. & Ruoslahti, E. (1988) *J. Biol. Chem.* **263**, 12927–12933
- Geisow, M. J. (1986) *FEBS Lett.* **203**, 99–103
- Ginsberg, M. H., Loftus, J. C. & Plow, E. F. (1988) *Thromb. Haemost.* **59**, 1–6
- Greer, J. (1981) *J. Mol. Biol.* **153**, 1027–1042
- Hemler, M. E. (1990) *Annu. Rev. Immunol.* **8**, 365–400
- Huber, R., Schneider, M., Mayr, I., Römisch, J. & Paques, E.-P. (1990) *FEBS Lett.* **275**, 15–21
- Humphries, M. J. (1990) *J. Cell Sci.* **97**, 585–592
- Hynes, R. O. (1987) *Cell* **48**, 549–554
- Kirchhofer, D., Gailit, J., Ruoslahti, E. & Grzesiak, J. (1990) *J. Biol. Chem.* **265**, 18525–18530
- Kirchhofer, D., Grzesiak, J. & Pierschbacher, M. D. (1991) *J. Biol. Chem.* **266**, 4471–4477
- Kloczewiak, M., Timmons, S., Lukas, T. & Hawiger, J. (1984) *Biochemistry* **23**, 1767–1774
- Komoriya, A., Green, L. J., Mervic, M., Yamada, S. S., Yamada, K. M. & Humphries, M. J. (1991) *J. Biol. Chem.* **266**, 15075–15079
- Leustek, T., Dalie, B., Amir-Shapira, D., Brot, N. & Weissbach, H. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7805–7808
- Loftus, J. C., O'Toole, T. E., Plow, E. F., Glass, A., Frelinger, A. L. & Ginsberg, M. H. (1990) *Science* **249**, 915–918
- Lyll, F. & Morton, J. J. (1987) *Clin. Sci.* **73**, 573–579

- Marriott, G., Kirk, W. R., Johnsson, N. & Weber, K. (1990) *Biochemistry* **29**, 7004–7011
- Mollenhaur, J., Bee, J. A., Lizarbe, M. A. & Von Der Mark, K. (1984) *J. Cell Biol.* **98**, 1572–1578
- Morton, J. J., Beattie, E. & Lyall, F. (1989) *Clin. Sci.* **76**, 9–11
- Novotny, J., Brucoleri, R. & Karplus, M. (1984) *J. Mol. Biol.* **177**, 787–818
- Pepinsky, R. B., Tizard, R., Mattaliano, R. J., Sinclair, L. K., Miller, G. T., Browning, J. L., Chow, E. P., Burne, C., Huang, K. S., Pratt, D., Wachter, L., Hession, C., Frey, A. Z. & Wallner, B. P. (1988) *J. Biol. Chem.* **263**, 10799–10811
- Pierschbacher, M. D. & Ruoslahti, E. (1984) *Nature (London)* **309**, 30–33
- Ruoslahti, E. & Pierschbacher, M. D. (1987) *Science* **238**, 491–497
- Smith, J. W. & Cheres, D. A. (1990) *J. Biol. Chem.* **265**, 2168–2172
- Smith, J. W. & Cheres, D. A. (1991) *J. Biol. Chem.* **266**, 11429–11432
- Springer, T. A. (1990) *Nature (London)* **346**, 425–434
- Staat, W. D., Fok, K. F., Zutter, M. M., Adams, S. P., Rodriguez, B. A. & Santoro, S. A. (1991) *J. Biol. Chem.* **266**, 7363–7367
- Strynadka, N. & James, M. (1989) *Annu. Rev. Biochem.* **58**, 951–998
- Takada, Y., Elices, M. J., Crouse, C. & Hemler, M. E. (1989) *EMBO J.* **8**, 1361–1368
- Van Gunsteren, W. F. & Berendsen, J. J. C. (1977) *Mol. Phys.* **34**, 1311–1327
- Vyas, N. K., Vyas, M. N. & Quirocho, F. A. (1987) *Nature (London)* **327**, 635–638
- Wirtl, G. & Schwartz-Albeiz, R. (1990) *J. Cell Physiol.* **144**, 511–522

Received 27 August 1991/7 November 1991; accepted 18 December 1991