A structure prediction for the ligand-binding region of the integrin β subunit: evidence for the presence of a von Willebrand factor A domain

Danny S. Tuckwell*, Martin J. Humphries

Wellcome Trust Centre for Cell-Matrix Research, Rm 2.205, School of Biological Sciences, University of Manchester, Oxford Road, Manchester M13 9PT, UK

Received 6 October 1996; revised version received 12 November 1996

Abstract The integrins are a family of cell surface receptors that mediate biologically important adhesive interactions. Integrin-ligand binding has been extensively studied because of the potential for the development of anti-adhesive therapies, but the molecular basis of this interaction is still poorly understood. A conserved region near the N-terminus of the β subunit appears to be of particular importance in ligand binding, but to date this domain has not been expressed in isolation. As a prelude to expression and potential structure determination, we have performed a detailed structure prediction for this region. Primary, secondary and tertiary structure analyses indicate that the region folds into a von Willebrand factor A-domain, thereby potentially placing a previously characterised module at the centre of a key functional region.

Key words: Integrin; Von Willebrand Factor A; Structure prediction

1. Introduction

The integrins are a family of cell-surface receptor proteins which mediate cell adhesion to a range of cell-surface, soluble and extracellular matrix proteins [1,2]. These interactions play key roles in a diverse range of biological processes including blood coagulation [3] and maintenance of tissue integrity [4], and the molecular basis of their function is therefore of considerable interest. While much is known about the sites on integrin ligands which interact with the receptors, the mechanism by which integrins bind their ligands is still poorly understood.

Each integrin is made up of a non-identical α and β subunit, of about 1000 and 800 amino acids in length respectively. It is well established that the N-terminal regions of both the α and β subunits are involved in ligand binding [5–9], but the lack of structural information for either region means that interpretation of the role of specific residues involved in ligand binding is very difficult. The exception to this lack of information is the subset of integrin α subunits which contain von Willebrand factor A-type domains (known as VWFA domains, sometimes referred to as A- or I-domains). Here, the VWFA domains have been shown to reproduce much of the ligand binding function of the parent integrins in which they occur [10–12]. VWFA domains are ~200 amino acids in length and are found in a large number of proteins [13]. Crystallographic studies have shown that the αM and αL VWFA domains have a Rossmann fold-like structure, with a central β sheet sandwiched between two layers of α helices [14,15]. This fold is formed by a secondary structure arrangement of an approximately alternating series of β strand-α helix-β strand etc. The VWFA domains contain a conserved Mg^{2+} binding site made up of a DxSxS motif, a serine residue, and an aspartate residue, from three non-contiguous regions [16].

Lee et al. [14], in their description of the structure of the αM VWFA domain, commented on the presence of a VWFA domain-like DxSxS motif in the region of the integrin β subunit known to be involved in ligand binding. Furthermore they were also able to show that consensus hydropathy plots for the VWFA domains and the β subunit could be favourably superposed, aligned around the DxSxS sequences. They therefore proposed that the ligand-binding region of the β subunit might comprise a VWFA domain. This would place a domain with known ligand-binding properties at the ligand-binding region of the β subunit, providing an explanation for the function of this region and permitting rational interpretation of studies. The suggestion of a VWFA domain at this location of the β subunit has therefore aroused considerable interest.

Any sequence similarity between the β subunit and the VWFA domains is in the equivocal 'twilight' region and obvious similarities are confined to the vicinity of the DxSxS sequence, thereby providing few residues which could be mutated to clearly indicate structural similarity. The traditional approach to identifying a VWFA domain in the β subunit would be by cloning and expression of the region followed by crystallography. However, the disulphide bonding of the β subunit, and its potential intermolecular interactions with other regions of the integrin, mean that expression of this domain in isolation has proved and may continue to prove complex. Consequently we have decided to use computer-based sequence analysis to test the hypothesis that the β subunit contains a VWFA domain.

2. Materials and methods

Sequence analysis was carried out either using the BBSRC SEQNET facility (Daresbury, Cheshire, UK) or by means of programs run on a Silicon Graphics 4D 240/GTX workstation. Multiple alignments were carried out using CLUSTALW [17], with default parameters, and manually adjusted where necessary using LINEUP [18], SOMAP [19] or VISTAS [20].

Amino acid matrices for data base scanning were generated and used to search the PIR database using the PROFILE group of programs [21]. Matrix generation and searches were carried out using default parameters.

Secondary structure predictions were carried out using PHD [22]. Multiple alignments were submitted by e-mail for prediction to PredictProtein@EMBL-Heidelberg.de.

Tertiary structure predictions were carried out using the THREADER algorithm [23] running on a Silicon Graphics 4D 240/GTX workstation. This program optimally threads the query sequence onto a library of template folds to give the folds with the best ener-

*Corresponding author. Fax: (44) (161) 275-5082.
E-mail: danny.tuckwell@man.ac.uk

PII S0014-5793(96)01368-3

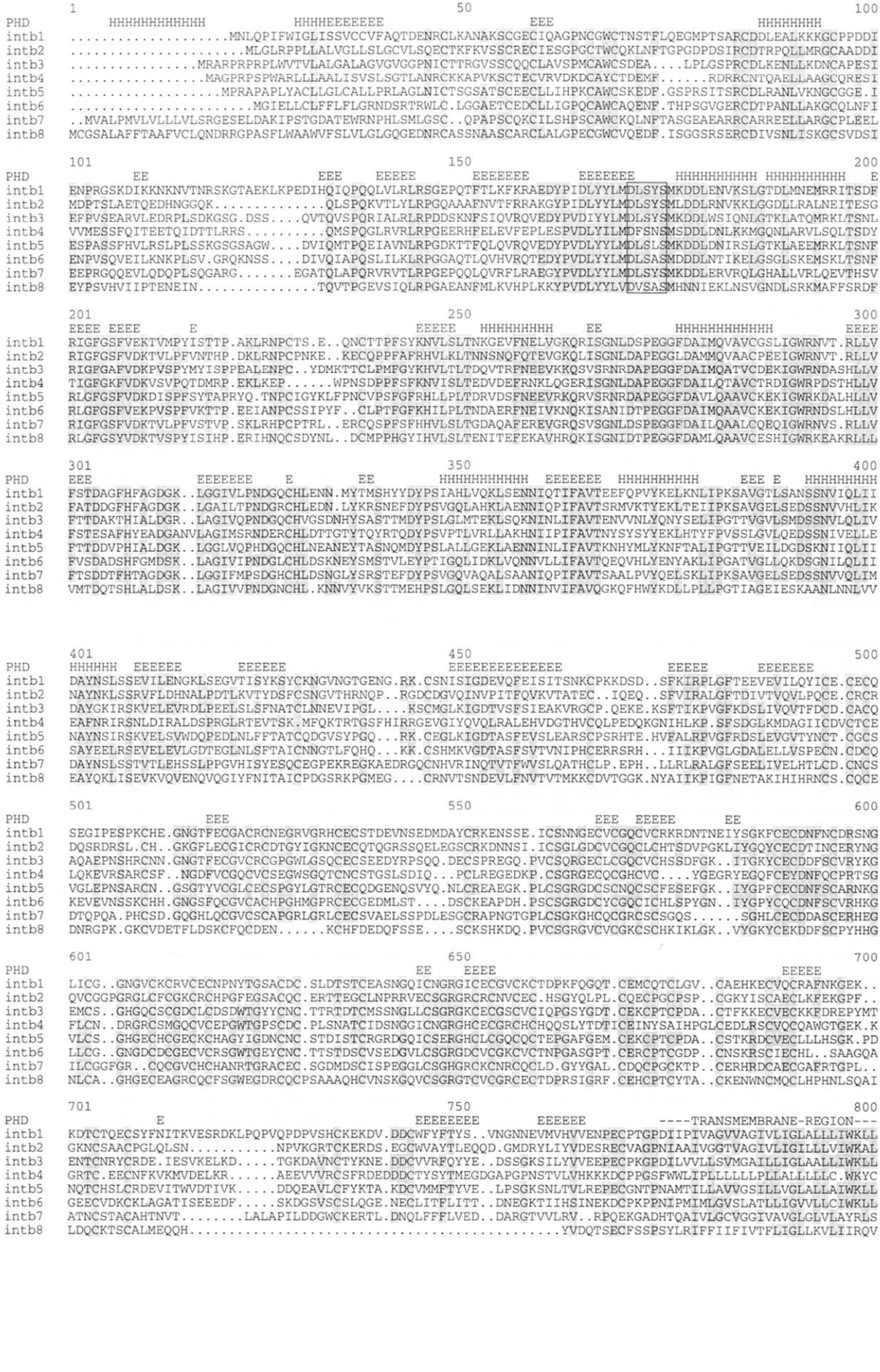

1 50 100
PHD HHHHHHHHHHHH HHHHEEEEEEE EEE HHHHHHHH
intb1 .....MNLQPIFWIGLISSVCCVFAQTDENRCLKANAKSCGECIQAGPNCGWCTNSTFLQEGMPTSARCDDLEALKKKGCPPDDI
intb2 .....MLGLRPPLLALVGLLSLGCVLSQECTKFKVSSCRECIESGPGCTWCQKLNFTGPGDPDSIRCDTRPQLLMRGCAADDI
intb3 .....MRARPRPRPLWVTVLALGALAGVGVGGPNICTTRGVSSCQQCLAVSPMCAWCSDEA....LPLGSPRCDLKENLLKDNCAPESI
intb4 .....MAGPRPSPWARLLLAALISVSLSGTLANRCKKAPVKSCTECVRVDKDCAYCTDEMF.....RDRRCNTQAELLAAGCQRESI
intb5 .....MPRAPAPLYACLLGLCALLPRLAGLNICTSGSATSCEECLLIHPKCAWCSKEDF.GSPRSITSRCDLRANLVKNGCGGE.I
intb6 .....MGIELLCLFFLFLGRNDSRTRWLC.LGGAETCEDCLLIGPQCAWCAQENF.THPSGVGERCDTPANLLAKGCQLNFI
intb7 ..MVALPMVLVLLLVLSRGESELDAKIPSTGDATEWRNPHLSMLGSC..QPAPSCQKCILSHPSCAWCKQLNFTASGEAEARRCARREELLARGCPLEEL
intb8 MCGSALAFFTAAFVCLQNDRRGPASFLWAAWVFSLVLGLGQGEDNRCASSNAASCARCLALGPECGWCVQEDF.ISGGSRSERCDIVSNLISKGCSVDSI
101 150 200
PHD EE EEE EEEEE EEEEEEE EEEEEEE HHHHHHHHHH HHHHHHHHHH E
intb1 ENPRGSKDIKKNKNVTNRSKGTAEKLKPEDIHQIQPQQLVLRLRSGEPQTFTLKFKRAEDYPIDLYYLMDLSYSMKDDLENVKSLGTDLMNEMRRITSDF
intb2 MDPTSLAETQEDHNGGQK.....QLSPQKVTLYLRPGQAAAFNVTFRRAKGYPIDLYYLMDLSYSMLDDLRNVKKLGGDLLRALNEITESG
intb3 EFPVSEARVLEDRPLSDKGSG.DSS....QVTQVSPQRIALRLRPDDSKNFSIQVRQVEDYPVDIYYLMDLSYSMKDDLWSIQNLGTKLATQMRKLTSNL
intb4 VVMESSFQITEETQIDTTLRRS.....QMSPQGLRVRLRPGEERHFELEVFEPLESPVDLYILMDFSNSMSDDLDNLKKMGQNLARVLSQLTSDY
intb5 ESPASSFHVLRSLPLSSKGSGSAGW....DVIQMTPQEIAVNLRPGDKTTFQLQVRQVEDYPVDLYYLMDLSLSMKDDLDNIRSLGTKLAEEMRKLTSNF
intb6 ENPVSQVEILKNKPLSV.GRQKNSS....DIVQIAPQSLILKLRPGGAQTLQVHVRQTEDYPVDLYYLMDLSASMDDDLNTIKELGSGLSKEMSKLTSNF
intb7 EEPRGQQEVLQDQPLSQGARG.....EGATQLAPQRVRVTLRPGEPQQLQVRFLRAEGYPVDLYYLMDLSYSMKDDLERVRQLGHALLVRLQEVTHSV
intb8 EYPSVHVIIPTENEIN.....TQVTPGEVSIQLRPGAEANFMLKVHPLKKYPVDLYYLVDVSASMHNNIEKLNSVGNDLSRKMAFFSRDF
201 250 300
EEEE EEEE E EEEEE HHHHHHHHH EE HHHHHHHHHHHH EEEE
intb1 RIGFGSFVEKTVMPYISTTP.AKLRNPCTS.E..QNCTTPFSYKNVLSLTNKGEVFNELVGKQRISGNLDSPEGGFDAIMQVAVCGSLIGWRNVT.RLLV
intb2 RIGFGSFVDKTVLPFVNTHP.DKLRNPCPNKE..KECQPPFAFRHVLKLTNNSNQFQTEVGKQLISGNLDAPEGGLDAMMQVAACPEEIGWRNVT.RLLV
intb3 RIGFGAFVDKPVSPYMYISPPEALENPC..YDMKTTCLPMFGYKHVLTLTDQVTRFNEEVKKQSVSRNRDAPEGGFDAIMQATVCDEKIGWRNDASHLLV
intb4 TIGFGKFVDKVSVPQTDMRP.EKLKEP.....WPNSDPPFSFKNVISLTEDVDEFRNKLQGERISGNLDAPEGGFDAILQTAVCTRDIGWRPDSTHLLV
intb5 RLGFGSFVDKDISPFSYTAPRYQ.TNPCIGYKLFPNCVPSFGFRHLLPLTDRVDSFNEEVRKQRVSRNRDAPEGGFDAVLQAAVCKEKIGWRKDALHLLV
intb6 RLGFGSFVEKPVSPFVKTTP.EEIANPCSSIPYF..CLPTFGFKHILPLTNDAERFNEIVKNQKISANIDTPEGGFDAIMQAAVCKEKIGWRNDSLHLLV
intb7 RIGFGSFVDKTVLPFVSTVP.SKLRHPCPTRL..ERCQSPFSFHHVLSLTGDAQAFEREVGRQSVSGNLDSPEGGFDAILQAALCQEQIGWRNVS.RLLV
intb8 RLGFGSYVDKTVSPYISIHP.ERIHNQCSDYNL..DCMPPHGYIHVLSLTENITEFEKAVHRQKISGNIDTPEGGFDAMLQAAVCESHIGWRKEAKRLLL
301 350 400
PHD EEE EEEEEEE E EE HHHHHHHHHHH EEEEEEE HHHHHHHHHH EEE E HHHHHHHHH
intb1 FSTDAGFHFAGDGK..LGGIVLPNDGQCHLENN.MYTMSHYYDYPSIAHLVQKLSENNIQTIFAVTEEFQPVYKELKNLIPKSAVGTLSANSSNVIQLII
intb2 FATDDGFHFAGDGK..LGAILTPNDGRCHLEDN.LYKRSNEFDYPSVGQLAHKLAENNIQPIFAVTSRMVKTYEKLTEIIPKSAVGELSEDSSNVVHLIK
intb3 FTTDAKTHIALDGR..LAGIVQPNDGQCHVGSDNHYSASTTMDYPSLGLMTEKLSQKNINLIFAVTENVVNLYQNYSELIPGTTVGVLSMDSSNVLQLIV
intb4 FSTESAFHYEADGANVLAGIMSRNDERCHLDTTGTYTQYRTQDYPSVPTLVRLLAKHNIIPIFAVTNYSYSYYEKLHTYFPVSSLGVLQEDSSNIVELLE
intb5 FTTDDVPHIALDGK..LGGLVQPHDGQCHLNEANEYTASNQMDYPSLALLGEKLAENNINLIFAVTKNHYMLYKNFTALIPGTTVEILDGDSKNIIQLII
intb6 FVSDADSHFGMDSK..LAGIVIPNDGLCHLDSKNEYSMSTVLEYPTIGQLIDKLVQNNVLLIFAVTQEQVHLYENYAKLIPGATVGLLQKDSGNILQLII
intb7 FTSDDTFHTAGDGK..LGGIFMPSDGHCHLDSNGLYSRSTEFDYPSVGQVAQALSAANIQPIFAVTSAALPVYQELSKLIPKSAVGELSEDSSNVVQLIM
intb8 VMTDQTSHLALDSK..LAGIVVPNDGNCHL.KNNVYVKSTTMEHPSLGQLSEKLIDNNINVIFAVQGKQFHWYKDLLPLLPGTIAGEIESKAANLNNLVV
401 450 500
PHD HHHHHH EEEEEE EEEEEE EEEEEEEEEEEEEE EEEEE EEEEEEE
intb1 DAYNSLSSEVILENGKLSEGVTISYKSYCKNGVNGTGENG.RK.CSNISIGDEVQFEISITSNKCPKKDSD..SFKIRPLGFTEEVEVILQYICE.CECQ
intb2 NAYNKLSSRVFLDHNALPDTLKVTYDSFCSNGVTHRNQP..RGDCDGVQINVPITFQVKVTATEC..IQEQ..SFVIRALGFTDIVTVQVLPQCE.CRCR
intb3 DAYGKIRSKVELEVRDLPEELSLSFNATCLNNEVIPGL....KSCMGLKIGDTVSFSIEAKVRGCP.QEKE.KSFTIKPVGFKDSLIVQVTFDCD.CACQ
intb4 EAFNRIRSNLDIRALDSPRGLRTEVTSK.MFQKTRTGSFHIRRGEVGIYQVQLRALEHVDGTHVCQLPEDQKGNIHLKP.SFSDGLKMDAGIICDVCTCE
intb5 NAYNSIRSKVELSVWDQPEDLNLFFTATCQDGVSYPGQ...RK.CEGLKIGDTASFEVSLEARSCPSRHTE.HVFALRPVGFRDSLEVGVTYNCT.CGCS
intb6 SAYEELRSEVELEVLGDTEGLNLSFTAICNNGTLFQHQ...KK.CSHMKVGDTASFSVTVNIPHCERRSRH...IIIKPVGLGDALELLVSPECN.CDCQ
intb7 DAYNSLSSTVTLEHSSLPPGVHISYESQCEGPEKREGKAEDRGQCNHVRINQTVTFWVSLQATHCLP.EPH..LLRLRALGFSEELIVELHTLCD.CNCS
intb8 EAYQKLISEVKVQVENQVQGIYFNITAICPDGSRKPGMEG....CRNVTSNDEVLFNVTVTMKKCDVTGGK.NYAIIKPIGFNETAKIHIHRNCS.CQCE
501 550 600
PHD EEE EEE EEEEE EE
intb1 SEGIPESPKCHE.GNGTFECGACRCNEGRVGRHCECSTDEVNSEDMDAYCRKENSSE.ICSNNGECVCGQCVCRKRDNTNEIYSGKFCECDNFNCDRSNG
intb2 DQSRDRSL.CH..GKGFLECGICRCDTGYIGKNCECQTQGRSSQELEGSCRKDNNSI.ICSGLGDCVCGQCLCHTSDVPGKLIYGQYCECDTINCERYNG
intb3 AQAEPNSHRCNN.GNGTFECGVCRCGPGWLGSQCECSEEDYRPSQQ.DECSPREGQ.PVCSQRGECLCGQCVCHSSDFGK..ITGKYCECDDFSCVRYKG
intb4 LQKEVRSARCSF..NGDFVCGQCVCSEGWSGQTCNCSTGSLSDIQ...PCLREGEDKP.CSGRGECQCGHCVC....YGEGRYEGQFCEYDNFQCPRTSG
intb5 VGLEPNSARCN..GSGTYVCGLCECSPGYLGTRCECQDGENQSVYQ.NLCREAEGK.PLCSGRGDCSCNQCSCFESEFGK..IYGPFCECDNFSCARNKG
intb6 KEVEVNSSKCHH.GNGSFQCGVCACHPGHMGPRCECGEDMLST....DSCKEAPDH.PSCSGRGDCYCGQCICHLSPYGN..IYGPYCQCDNFSCVRHKG
intb7 DTQPQA.PHCSD.GQGHLQCGVCSCAPGRLGRLCECSVAELSSPDLESGCRAPNGTGPLCSGKGHCQCGRCSCSGQS......SGHLCECDDASCERHEG
intb8 DNRGPK.GKCVDETFLDSKCFQCDEN.....KCHFDEDQFSSE....SCKSHKDQ.PVCSGRGVCVCGKCSCHKIKLGK..VYGKYCEKDDFSCPYHHG
601 650 700
PHD EE EEEE EEEEE
intb1 LICG..GNGVCKCRVCECNPNYTGSACDC.SLDTSTCEASNGQICNGRGICECGVCKCTDPKFQGQT.CEMCQTCLGV..CAEHKECVQCRAFNKGEK..
intb2 QVCGGPGRGLCFCGKCRCHPGFEGSACQC.ERTTEGCLNPRRVECSGRGRCRCNVCEC.HSGYQLPL.CQECPGCPSP..CGKYISCAECLKFEKGPF..
intb3 EMCS..GHGQCSCGDCLCDSDWTGYYCNC.TTRTDTCMSSNGLLCSGRGKCECGSCVCIQPGSYGDT.CEKCPTCPDA..CTFKKECVECKKFDREPYMT
intb4 FLCN..DRGRCSMGQCVCEPGWTGPSCDC.PLSNATCIDSNGGICNGRGHCECGRCHCHQQSLYTDTICEINYSAIHPGLCEDLRSCVQCQAWGTGEK.K
intb5 VLCS..GHGECHCGECKCHAGYIGDNCNC.STDISTCRGRDGQICSERGHCLCGQCQCTEPGAFGEM.CEKCPTCPDA..CSTKRDCVECLLLHSGK.PD
intb6 LLCG..GNGDCDCGECVCRSGWTGEYCNC.TTSTDSCVSEDGVLCSGRGDCVCGKCVCTNPGASGPT.CERCPTCGDP..CNSKRSCIECHL..SAAGQA
intb7 ILCGGFGR..CQCGVCHCHANRTGRACEC.SGDMDSCISPEGGLCSGHGRCKCNRCQCLD.GYYGAL.CDQCPGCKTP..CERHRDCAECGAFRTGPL..
intb8 NLCA..GHGECEAGRCQCFSGWEGDRCQCPSAAAQHCVNSKGQVCSGRGTCVCGRCECTDPRSIGRF.CEHCPTCYTA..CKENWNCMQCLHPHNLSQAI
701 750 800
PHD E EEEEEEEE EEEEEE ----TRANSMEMBRANE-REGION---
intb1 KDTCTQECSYFNITKVESRDKLPQPVQPDPVSHCKEKDV.DDCWFYFTYS..VNGNNEVMVHVVENPECPTGPDIIPIVAGVVAGIVLIGLALLLIWKLL
intb2 GKNCSAACPGLQLSN.....NPVKGRTCKERDS.EGCWVAYTLEQQD.GMDRYLIYVDESRECVAGPNIAAIVGGTVAGIVLIGILLLVIWKAL
intb3 ENTCNRYCRDE.IESVKELKD.....TGKDAVNCTYKNE.DDCVVRFQYYE..DSSGKSILYVVEEPECPKGPDILVVLLSVMGAILLIGLAALLIWKLL
intb4 GRTC.EECNFKVKMVDELKR.....AEEVVVRCSFRDEDDDCTYSYTMEGDGAPGPNSTVLVHKKKDCPPGSFWWLIPLLLLLLPLLALLLLLC.WKYC
intb5 NQTCHSLCRDEVITWVDTIVK.....DDQEAVLCFYKTA.KDCVMMFTYVE..LPSGKSNLTVLREPECGNTPNAMTILLAVVGSILLVGLALLAIWKLL
intb6 GEECVDKCKLAGATISEEEDF.....SKDGSVSCSLQGE.NECLITFLITT..DNEGKTIIHSINEKDCPKPPNIPMIMLGVSLATLLIGVVLLCIWKLL
intb7 ATNCSTACAHTNVT.....LALAPILDDGWCKERTL.DNQLFFFLVED..DARGTVVLRV..RPQEKGADHTQAIVLGCVGGIVAVGLGLVLAYRLS
intb8 LDQCKTSCALMEQQH.....YVDQTSECFSSPSYLRIFFIIFIVTFLIGLLKVLIIRQV

Fig. 1. Multiple alignment of human integrin β subunit sequences. Alignments were carried out using CLUSTALW [17]. Secondary structure prediction was carried out by means of PHD [22]; E = β strand, H = α helix, otherwise loop. Conserved residues are shaded; the region identified by Lee et al. [12] as resembling the DxSxS motif seen in VWFA domains is boxed. intb1, human integrin β1; intb2, human integrin β2; etc.
←

getic matches. Default parameters were used for all reported data. Tertiary structure prediction was also carried out using the algorithm of Fischer and Eisenberg [24]. This program used secondary structure information (predicted by PHD for an alignment of probe sequences) to inform the threading process, thereby improving the accuracy of predictions. Prediction was carried out by submitting the probe sequence to the web site server address http://www.mbi.ucla.edu/people/fischer/BENCH/benchmark1.html.

3. Results and discussion

3.1. Multiple alignment of the integrin β subunit

It is accepted that multiple alignments yield more structural information than individual sequences. A multiple alignment of the human integrin β1-β8 sequences was therefore constructed and used for all subsequent analyses. The alignment extended from the immature N-terminus of the β subunit to the transmembrane region (Fig. 1). Sequences from non-human species were not included in the alignment as preliminary data indicated that the inclusion of these did not make a major contribution to the spread of sequence diversity in the alignment.

3.2. Primary sequence analysis

The VWFA domain DxSxS motif which Lee et al. [14] observed in the integrin β subunit lies near the N-terminus of the VWFA domain, in a region conserved between domains [16]. In order to test this apparent similarity in detail, a database searching matrix was generated for a comparable conserved region of a putative β subunit VWFA domain, positions 163–174 in Fig. 1, which includes the DxSxS motif. The matrix was used to search the PIR database, and results of the search are shown in Table 1. Of the top 31 matches, 28 were against integrin β subunit sequences, as expected. Matches 23, 27 and 29 were against VWFA domain-containing proteins, as were matches 35–37. Examination of these matches (data not shown) demonstrated that they were against appropriate regions of the VWFA domain. Matches 32–34 were against proteins which do not have VWFA domains and were judged to be false positives by examination of the sequence match. Matrix scanning therefore demonstrates a primary sequence match between the integrin β subunit and the VWFA domain in this region, confirming and extending the similarity between the VWFA domain and the integrin β subunit.

3.3. Secondary structure prediction

The secondary structure of VWFA domains is an approximately alternating series of β strand-α helix-β strand etc. [14,15]. To examine the secondary structure of the integrin β subunit, secondary structure prediction was carried out using PHD [22], which is approximately 72.2% accurate for a 3-state prediction. The resulting prediction is given in Fig. 1. As discussed above, the region of the β subunit which was seen to resemble the VWFA domain lies at its beginning of the domain. A VWFA domain in the β subunit would therefore start at ~position 161 (Fig. 1). The secondary structure of the β subunit C-terminal to this point was studied and was found to be an approximately alternating series of β strand-α helix-β strand etc., resembling the VWFA domain. This pattern extended to position 408, giving a putative domain of 249 amino acids, comparing favourably with the ~200 amino acids seen for VWFA domains.

Since no programs currently exist for the determination of protein domain boundaries, the domain structure of the integrin β subunit cannot be specifically studied. However, in support of a domain boundary at the commencement of a putative β subunit VWFA domain, genome sequencing has

Table 1
Matches from searching PIR database with a matrix generated for positions 163–174 of the β subunit alignment

Number of match	Z score	Protein	Contains VWFA domain
1	8.68	integrin β7, mouse	
2	8.65	integrin β7, mouse	
3	8.63	integrin β_{PAT3}, *C. elegans*	
...	...	...	...
22	8.05	integrin β2, chicken	
23	7.97	complement component C2, guinea pig	+
24	7.93	integrin β subunit, *Drosophila*	
25	7.74	integrin β8, chicken	
26	7.24	integrin β8, rabbit	
27	7.21	integrin β8, human	
28	6.50	TRAP, *P. falciparum*	+
29	6.49	sporozoite surface protein, *P. falciparum*	+
30	6.20	integrin β4, human	
31	6.17	integrin β4, mouse	
32	6.12	probable olfactory receptor tpcr 120, human	−
33	5.89	E7 papilloma virus, human	−
34	5.84	hypothetical protein HI0593, *H. influenzae*	−
35	5.81	collagen α1(VI), human	+
36	5.80	collagen α1(VI), human	+
37	5.77	collagen α1(VI), human	+

Z score is a normalised measure of distance of a match from the background level of matches against the whole database, where the majority of matches are not genuine. Z score = ((normalised score for match minus average normalised score)/normalised standard deviation).

Table 2
Tertiary fold prediction for the integrin β subunit (positions 161–408) using the THREADER algorithm [23]

	β1 subunit				β3 subunit		
Brookhaven code	Protein	Z score	Protein structure	Brookhaven code	Protein	Z score	Protein structure
1DHR	dihydropteridin reductase	−3.08	Rossmann fold	1MFA	immunoglobulin L chain	−2.87	Ig fold
1NAR	narbonin	−2.76	Tim barrel	1DHR	dihydropteridin reductase	−2.82	Rossmann fold
A5TIM	triose phosphate isomerase	−2.37	Tim barrel	1AYH	glucanase	−2.60	β sandwich
1PPN	papain	−2.23	Segregated α+β	1SIM	sialidase	−2.59	β propeller
1GKY	guanylate kinase	−2.10	α-β-α sandwich	E1CSE	subtilisin	−2.49	α-β-α sandwich

Z score = ((energy of match minus mean energy)/standard deviation) for calculated pairwise energies for each fit. Z scores greater than −2.0 are very poor, those less than −3.0 are likely to be correct predictions.

identified a phase 1 intron in the codon for position 146 which is conserved between human β2, β3, β7, *C. elegans* βpat-3 and Drosophila βPS integrin subunits [25]. Conserved phase 1 introns are good indicators of ancestral domain boundaries [26] and there are no other conserved phase 1 introns in the extracellular region of the β subunit.

Regarding the C-terminus of a putative VWFA domain ending at ~position 408, the PHD prediction of β strands alone for the remaining C-terminal portion of the β subunit (positions 409–800) is consistent with the observation that positions 498–674 comprise 4 EGF-like domains [27]. However, this does not account for all the sequence in this region and the regions flanking the EGF repeats were therefore examined for similarities to known domains/proteins. Although none were found, the region 409–497 (between the putative VWFA domain and the EGF domains) was matched with a low Z score (< −3; i.e. a good match) to β sandwich structures using the tertiary fold matching program THREADER [23] (data not shown).

Secondary structure prediction therefore supports a domain with similar secondary structure to that of the VWFA extending from ~positions 161–408, and evidence can be presented to support the proposed domain boundaries.

3.4. Tertiary structure prediction

The secondary structure predicted for positions 161–408 could adopt a number of tertiary folds. Sequences from this region were therefore used as the probe sequence for the tertiary structure prediction algorithm THREADER [23]. Here, the probe sequence is optimally threaded onto a library of tertiary structure folds and the optimal folds taken as indicating likely three-dimensional structures for the probe sequence. To provide an appropriately controlled method, sequences from the two most divergent β subunits, β1 and β3, were used; results are shown in Table 2. The consensus prediction for positions 161–408 was a Rossmann fold/α-β-α sandwich, i.e. the structure of the VWFA domain (VWFA domain structures were not represented on the fold library used for these searches). Other predicted folds either ignored the α helix content (e.g. β sandwich) or represented alternative ways of organising the secondary structure (e.g. TIM barrel), and were not predicted for both probe sequences.

Tertiary structure prediction was also carried out by means of the algorithm of Fischer and Eisenberg [24]. This uses the PHD secondary structure prediction, derived from an alignment of the β3 probe sequence with related sequences, to improve the accuracy of the fold matching. The best match was to the αL VWFA domain [15], with a Z score of 5.53, which falls within the algorithm's confidence threshold of 4.8 ± 1.0. This algorithm therefore clearly matched the region 161–408 to the VWFA domain.

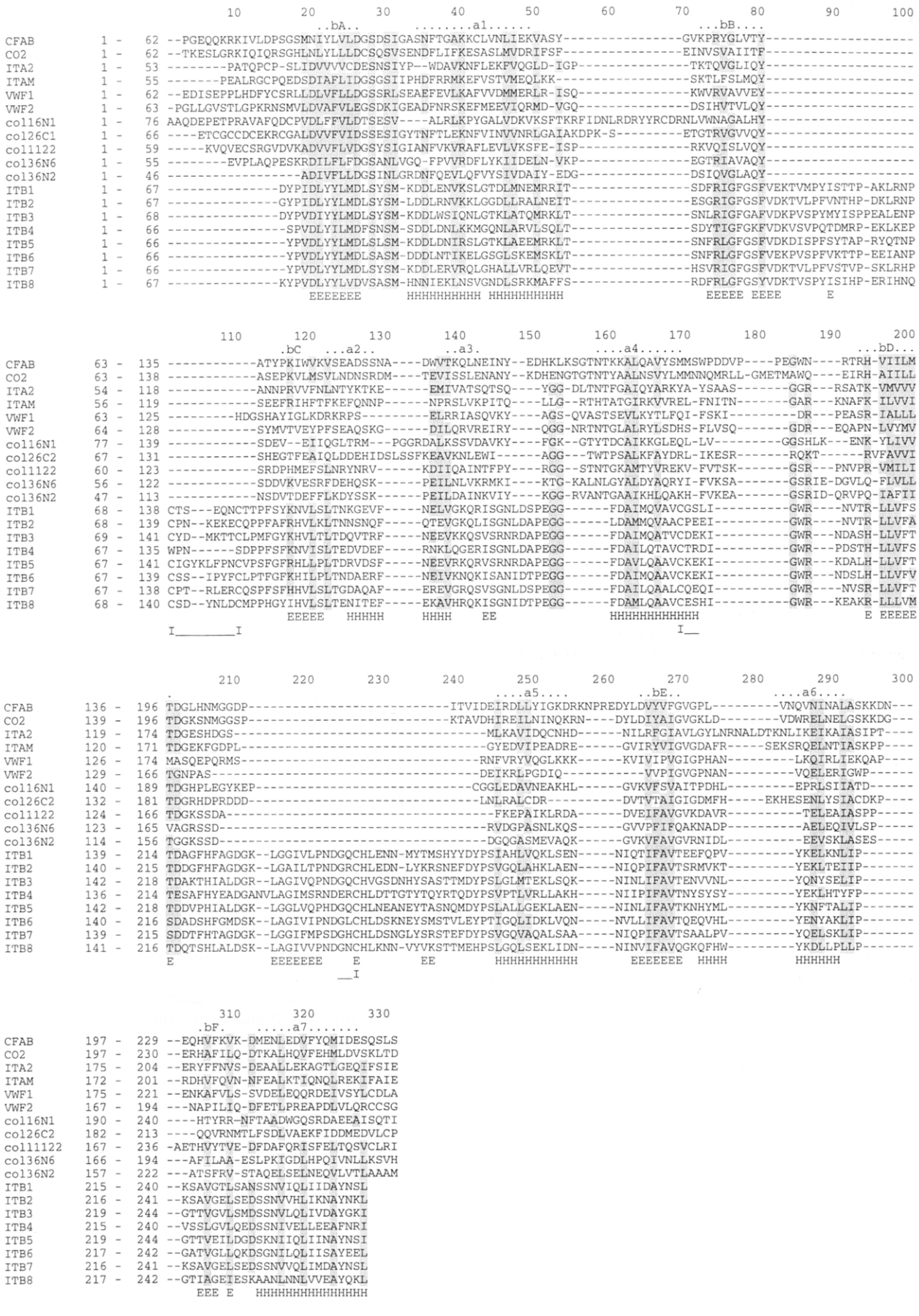
3.5. Aligning the putative β subunit VWFA domain with the VWFA domain family

From the above data it can be seen that primary, secondary and tertiary structure criteria predict that the region 161–408 of the integrin β subunit comprises a VWFA-like domain. Therefore, this region of the β subunit was aligned with the VWFA domain family. The resulting alignment is shown in Fig. 2. Based on this, a diagrammatic representation of the putative β subunit VWFA domain is given in Fig. 3. A representation showing the essential structural features of the αL and αM VWFA domains is shown for comparison.

All the structural elements seen in the αM and αL VWFA domains could be aligned with elements predicted by PHD for the β subunit. A number of regions of similarity between the β subunit putative VWFA domain and the VWFA domain family were observed in addition to positions 163–174 (corresponding to positions 19–30, β strand A, in Fig. 2) which have been commented on above. In particular, the VWFA domain sequence in the region of β strand D (Fig. 2,3) aligned well with β subunit sequence. These two regions are particularly important as they form the core of the VWFA domain structure [14,15] (Fig. 3), and the conserved DxSxS motif from β strand A and aspartate 201 (Fig. 2) are involved in cation coordination in the VWFA domain.

→

Fig. 2. Alignment of key divergent members of the VWFA domain family with the putative VWFA domain of the integrin β subunit. Divergent and representative members of the VWFA domain family are shown although initially a large alignment of VWFA sequences from a range of proteins was constructed to help define conserved residues. Shaded residues indicate regions of similarity. The secondary structural assignments determined from the integrin αM VWFA domain [12] are given above the alignment. The PHD secondary structure derived for the β subunit is given below it. I__I marks disulphide bonds [29]. CFAB, human complement factor B; CO2, human complement component C2; ITA2, human integrin α2; ITAM, human integrin αM; VWF1, human von Willebrand factor A1 domain; VWF2, human von Willebrand factor A2 domain; col16N1, human collagen α1(VI) N1 VWFA domain; col26C1, human collagen α2(VI) C1 VWFA domain; col1122, chicken collagen α1(XII), second VWFA domain; col36N6, human collagen α3(VI) N6 VWFA domain; col36N2, chicken collagen α3(VI) N2 VWFA domain; ITB1, human integrin β1, ITB2, human integrin β2, etc.


CFAB
CO2
ITA2
ITAM
VWF1
VWF2
col16N1
col26C1
col11122
col36N6
col36N2
ITB1
ITB2
ITB3
ITB4
ITB5
ITB6
ITB7
ITB8

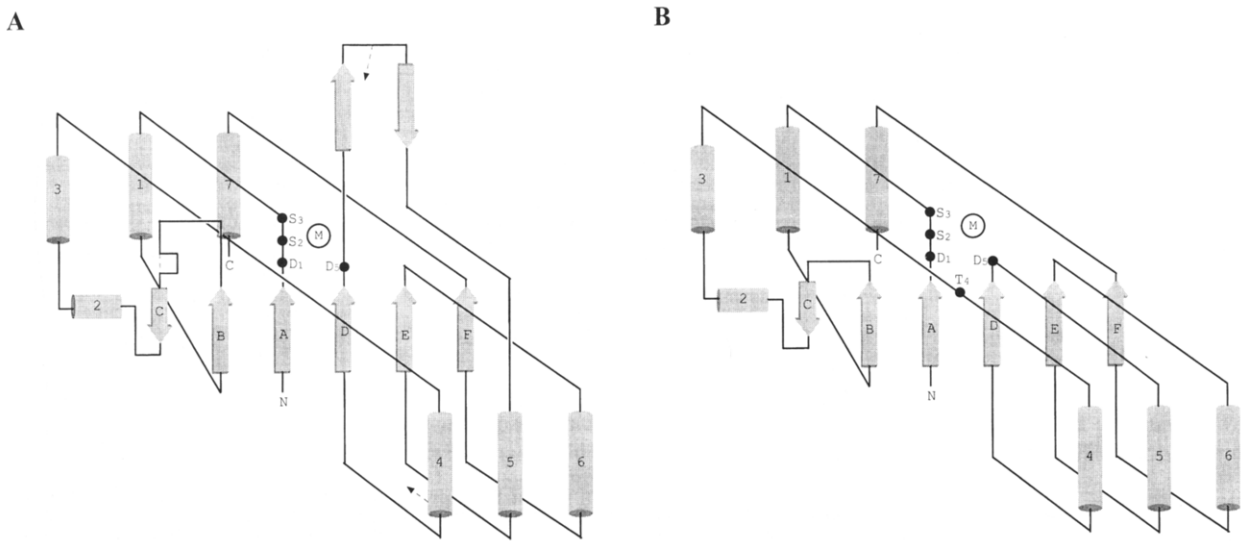


Fig. 3. Diagrammatic representation of the predicted VWFA of the integrin β subunit (A), based on the integrin αL and αM VWFA domain structures, with a representation of the essential structural features of the VWFA domains of integrin αL and αM for comparison (B). The diagram shows β strands (arrows), α helices (cylinders) and loops (lines); loop lengths are exaggerated for overall clarity. The metal ion (M) and coordinating residues are also marked.

Important differences between the putative β subunit VWFA domain and the VWFA domain family were seen:

(i) The coordinating threonine seen in most VWFA domains (position 157 in Fig. 2) is not matched by a similar residue in the β subunits. However, this may not be detrimental to the model as this residue is not always involved in direct coordination to the cation in all VWFA domains [15,28] and adjacent glutamate and aspartate residues in the β subunit could provide coordination.

(ii) Comparison of the β subunit and VWFA domain sequences shows two insertions, one between β strands B and C and one between β strand D and α helix 5. The first of these two insertions has little or no predicted structure but appears to be supported by a disulphide bond [29]. This loop lies near the extremity of the domain, near a region with poor sequence conservation (β strand C and α helix 2) and inserted sequence may therefore be tolerated at this point. The second loop represents a major departure from the VWFA domain, as it lies near the centre of the domain, projecting from it. Secondary structure prediction indicates that this region contains some β strands, perhaps forming a small sheet. This region appears to be stabilised by a disulphide bond to α helix 4.

3.6. Comparing integrin and VWFA domain function

Functional studies have identified positions ~161–411 of the integrin β subunit alignment (Fig. 1), which is well conserved between β subunits, as containing regions involved in ligand binding [6–8]. Since this region corresponds to the putative VWFA domain, recent findings on the function of VWFA domains can now be compared to what is known of integrin β subunit function:

(i) Most VWFA domains bind or are predicted to bind cations; cation binding for the β subunit has also been demonstrated [30]. Related to this, the DxSxS motif is important for VWFA domain function [12,31]; the DxSxS motif is also important for integrin function [6].

(ii) The helix α2–α3 region in VWFA domains is the site of an anti-functional mutation [31]. As this site is distal to the likely ligand binding site on the top face of the molecule, this mutation probably has an allosteric effect; activating and inhibitory anti-integrin β1 antibodies map to positions 124–139 (Fig. 3), which is predicted to correspond to the α2-α3 region in the β subunit [32].

The loop between β strand D-α helix 5 region has been shown to play a major role in VWFA binding to its ligands [33,34]; The insertion in this region in the β subunit therefore lies at a key site and this region may play a role in ligand binding.

4. Conclusions

The N-terminal region of the integrin β subunit is currently the subject of much interest because of its key role in integrin-ligand binding [35]. The complexities of expression and crystallography of integrin fragments have meant that these techniques have been slow to characterise functional units within integrins or to define their mechanisms. We have therefore used a range of computer-based sequence analysis tools to examine the structure of the integrin β subunit, in particular positions 161–408. Primary structure comparisons indicate that positions 163–174 are homologous to the N-terminal region of the VWFA domain. The predicted secondary structure for the β subunit indicates a region with VWFA domain-like secondary structure extending from positions 161–408, and tertiary structure predictions indicate that this region adopts the same fold as the VWFA domain. Finally, VWFA domains and positions 161–408 of the integrin β subunit can be aligned, demonstrating further primary sequence similarities. Positions 161–408 of the integrin β subunit are therefore predicted to comprise a VWFA domain. Since regions within positions 161–408 have been shown to play a role in integrin-ligand binding, our prediction places a known domain with established ligand-binding properties at the known ligand-binding site of the integrin. Our prediction allows inter-

pretation of existing data on the integrin β subunit and provides a framework for research into this key region of the integrin.

5. Note

While the manuscript was under review, we became aware of the papers of Puzon-McLaughlin, W. and Takada, Y. (1996) J. Biol. Chem. 271, 20438–20443, and Tozer, E.C., Liddington, R.C., Sutcliffe, M.J., Smeeton, A.H. and Loftus, J.C. (1996) J. Biol. Chem. 271, 21978–21984. Both papers describe mutagenesis studies which identify key functional residues in the region of the β subunit corresponding to the predicted VWFA domain. Tozer et al. also present a model for the putative VWFA domain which differs substantially from the one presented here. Examination of the model of Tozer et al. indicates that the difference in models is primarily due to less accuracy being credited to the PHD secondary structure prediction (Q_3 of 54.6% as opposed to the current value of 72.2%). In addition, our model is based on results from a number of predictive tools. The model presented here provides a framework for the interpretation of the mutagenesis data from these papers.

Acknowledgements: This work was supported by grants from the Wellcome Trust. The authors are grateful to Dr. D. Jones for supplying the THREADER program. We are also grateful to Peter Newham, Paul Mould, Michèle Clamp and Andy Brass for discussion, advice and assistance.

References

[1] Hynes, R.O. (1992) Cell 69, 11–25.
[2] Tuckwell, D.S. and Humphries, M.J. (1993) Crit. Rev. Oncol./Hematol. 15, 149–171.
[3] Corbi, A.L., Vara, A., Ursa, A., Rodriguez, M.C.G., Fontán, G. and Sánchez-Madrid, J. (1992) Eur. J. Immunol. 22, 1877–1881.
[4] Georges-Labousesse, E., Messaddeq, N., Yehia, G., Cadalbert, L., Dierich, A. and Le Meur, M (1996) Nature Genet. 13, 370–373.
[5] Loftus, J.C., Halloran, C.E., Ginsberg, M.H., Feigen, L.P., Zablocki, J.A. and Smith, J.W. (1996) J. Biol. Chem. 271, 2033–2039.
[6] Bajt, M.L. and Loftus, J.C. (1994) J. Biol. Chem. 269, 20913–20919.
[7] D'Souza, S.E., Ginsberg, M.H., Burke, T.A., Lam, S.C.T. and Plow, E.F. (1988) Science 242, 91–93.
[8] D'Souza, S.E., Haas, T.A., Piotrowicz, R.S., Byers-Ward, V., McGrath, D.E., Soule, H.R., Cierniewski, C., Plow, E.F. and Smith, J.W. (1994) Cell 79, 659–667.
[9] Tuckwell, D.S. and Humphries, M.J. (1996) in: Integrin-Ligand Binding (Eble, J.A. and Kühn, K., Eds.) Landes, Georgetown.
[10] Tuckwell, D.S., Calderwood, D.A., Green, L.J. and Humphries, M.J. (1995) J. Cell Sci. 108, 1629–1637.
[11] Randi, A.M. and Hogg, N. (1994) J. Biol. Chem. 269, 12395–12398.
[12] Kamata, T. and Takada, Y. (1994) J. Biol. Chem. 269, 26006–26010.
[13] Colombatti, A., Bonaldo, P. and Doliana, R. (1993) Matrix 13, 297–306.
[14] Lee, J.-O., Rieu, P., Arnaout, M.A. and Liddington, R.C. (1995) Cell 80, 631–638.
[15] Qu, A. and Leahy, D.J. (1995) Proc. Natl. Acad. Sci. USA 92, 10277–10281.
[16] Perkins, S., Smith, K.F., Williams, S.C., Haris, P.I., Chapman, D. and Sim, R.B. (1994) J. Mol. Biol. 238, 104–119.
[17] Higgins, D.G., Bleasby, A.J. and Fuchs, R. (1992) CABIOS 8, 189–191.
[18] Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.
[19] Parry-Smith, D.J. and Attwood, T.K. (1992) CABIOS 8, 451–459.
[20] Perkins, D.N. and Attwood, T.K. (1995) J. Mol. Graph. 13, 73–75.
[21] Gribskov, M., McLachlan, A.D. and Eisenberg, D. (1987) Proc. Natl. Acad. Sci. USA 84, 4355–4358.
[22] Rost, B. and Sander, C. (1993) J. Mol. Biol. 232, 584–599.
[23] Jones, D.T., Taylor, W.R. and Thornton, J.M. (1992) Nature 358, 86–89.
[24] Fischer, D. and Eisenberg, D. (1996) Protein Sci. 5, 947–955.
[25] Gettner, S.N., Kenyon, C. and Reichardt, L.F. (1995) J. Cell Biol. 129, 1127–1141.
[26] Patthy, L. (1991) Curr. Opin. Struct. Biol. 1, 351–361.
[27] Yuan, Q., Jiang, W.-M., Krissansen, G.W. and Watson, J.D. (1990) Int. Immunol. 2, 1097–1108.
[28] Lee, J.-O., Bankston, L.A., Arnaout, M.A. and Liddington, R.C. (1995) Structure 3, 1333–1340.
[29] Calvete, J.J., Henschen, A., González-Rodríguez, J. (1991) Biochem. J. 274, 63–171.
[30] Rivas, G.A. and González-Rodríguez, J. (1991) Biochem J. 276, 35–40.
[31] Edwards, C.P., Champe, M., Gonzales, T., Wessinger, M.E., Spencer, S.A., Presta, L.G., Berman, P.W. and Bodary, S.C. (1995) J. Biol. Chem. 270, 12635–12640.
[32] Takada, Y. and Puzon, W. (1993) J. Biol. Chem. 268, 17597–17601.
[33] Ueda, T., Rieu, P., Brayer, J. and Arnaout, M.A. (1994) Proc. Natl. Acad. Sci. USA 91, 10680–10684.
[34] Huang, C. and Springer, T.A. (1995) J. Biol. Chem. 270, 19008–19016.
[35] Humphries, M.J. (1996) Curr. Opin. Cell Biol. 8, 632–640.