Conservation of the Human Integrin-Type Beta-Propeller Domain in Bacteria

Bhanupratap Chouhan[1,2], Alexander Denesyuk[1], Jyrki Heino[3], Mark S. Johnson[1], Konstantin Denessiouk[2]*

1 Department of Biosciences, Åbo Akademi University, Turku, Finland, 2 Turku Center for Biotechnology, University of Turku and Åbo Akademi University, Turku, Finland, 3 Department of Biochemistry and Food Chemistry, University of Turku, Turku, Finland

Abstract

Integrins are heterodimeric cell-surface receptors with key functions in cell-cell and cell-matrix adhesion. Integrin α and β subunits are present throughout the metazoans, but it is unclear whether the subunits predate the origin of multicellular organisms. Several component domains have been detected in bacteria, one of which, a specific 7-bladed β-propeller domain, is a unique feature of the integrin α subunits. Here, we describe a structure-derived motif, which incorporates key features of each blade from the X-ray structures of human αIIbβ3 and αVβ3, includes elements of the FG-GAP/Cage and Ca^{2+}-binding motifs, and is specific only for the metazoan integrin domains. Separately, we searched for the metazoan integrin type β-propeller domains among all available sequences from bacteria and unicellular eukaryotic organisms, which must incorporate seven repeats, corresponding to the seven blades of the β-propeller domain, and so that the newly found structure-derived motif would exist in every repeat. As the result, among 47 available genomes of unicellular eukaryotes we could not find a single instance of seven repeats with the motif. Several sequences contained three repeats, a predicted transmembrane segment, and a short cytoplasmic motif associated with some integrins, but otherwise differ from the metazoan integrin α subunits. Among the available bacterial sequences, we found five examples containing seven sequential metazoan integrin-specific motifs within the seven repeats. The motifs differ in having one Ca^{2+}-binding site per repeat, whereas metazoan integrins have three or four sites. The bacterial sequences are more conserved in terms of motif conservation and loop length, suggesting that the structure is more regular and compact than those example structures from human integrins. Although the bacterial examples are not full-length integrins, the full-length metazoan-type 7-bladed β-propeller domains are present, and sometimes two tandem copies are found.

Introduction

Integrins are large, heterodimeric cell-surface receptors that detect and transmit changes in mechanical forces resulting from interactions between a cell and the extracellular matrix [1]. Cell-cell and cell-matrix adhesion, mediated by integrins, play "key" roles in inflammation, cell development, and cell proliferation and differentiation [2,3].

Vertebrates express at least 18 different integrin α subunits and eight β subunits, forming 24 α/β heterodimeric receptors in human [4]. Half of the integrin α subunits in human, namely α1, α2, α10, α11, αL, αM, αX, αD, and αE, contain an additional domain – a von Willebrand factor type-A (vWA) domain, inserted between the second and the third blades of the N-terminal 7-bladed β-propeller domain, referred to as either the αA domain [5] or the αI domain [6]; the β-propeller domain is found in all integrin α subunits. All integrins with an inserted αI domain bind their natural ligands via the metal ion dependent adhesion site (MIDAS) of the αI domain and appear to have arisen within integrin α subunits around the divergence of the first chordates since they are found in some integrin α subunits from tunicates but not from the earliest-diverging deuterostomes, e.g. the echinoderms [7–9].

The remaining nine α subunits, α3, α4, α5, α6, α7, α8, α9, αV, and αIIb, do *not* contain the αI domain. The Arg-Gly-Asp (RGD) sequence, present in the integrin ligands fibronectin and laminin, is one of the very first motifs that was found to be recognized by integrins [10], and it was shown that all known RGD-recognizing integrins lack the αI domain. In these integrins, the carboxylate side-chain of the aspartate residue of the RGD peptide binds to the Mg^{2+} coordinated by MIDAS of the I-like βA domain present in the β subunit, while the arginine side chain directly binds to the N-terminal β-propeller domain of the α subunit [11]. Thus, the N-terminal β-propeller domain either directly participates in ligand recognition, as with all integrins having an α subunit lacking the αI domain, or it incorporates a ligand-binding αI domain budding from the β-propeller domain through which it recognizes the ligand [4,12,13]. In either case, the β-propeller domain of the α subunit forms key stabilizing interactions with the βA domain of the β subunit.

A study of 90 integrin α subunits and 57 β subunits from 26 different metazoan species, ranging from *C. elegans* to *H. sapiens* and including sequences from the tunicate *C. intestinalis* and the pufferfish *T. rubripes*, has shown that orthologues of the human α subunits and β subunits of the integrins are highly conserved in bony vertebrates [7], whereas earlier-diverging integrins from sponges, nematode, insects through to the earliest chordate integrins are not human orthologues [7,9]. However, both the α and β subunits of integrins have been detected throughout the metazoans, including sequences from some very early diverging species such as the sponge *Geodia cydonium*.

A cell adhesion system involving proteins with the RGD sequence may have already existed in protozoa since a GRGDSPK peptide but not the GRGESPK peptide caused substratum-detachment of cells of the marine amoeba *Neoparamoeba aestuarina* [14]. Recently, Sebé-Pedrós et al. have reported that genomes of two unicellular eukaryotic organisms, *Capsaspora owczarzaki* (a unicellular organism diverging just prior to metazoans) [15] and in *Thecamonas trahens* (formerly *Amastigomonas sp.*) contain domain sequences that appear to have the hallmarks of α and β subunits of integrins [16].

The first report of similarities between bacterial sequences and integrin sequences was made over ten years ago, when May and Ponting reported that an automated run of PSI-BLAST [17] showed sequence similarity between the cytoplasmic portion of the human integrin β4 subunit (residues 735–1125; the human β4 subunit has a large cytoplasmic domain in comparison to other integrin β subunits) and a region encoding a hypothetical protein from the cyanobacteria *Synechocystis* sp. PCC6803 [18]. The authors also reported sequence similarity between another sequence from the same bacterium and several β-propeller repeats from different integrin α subunits. Shortly after, Jenkins et al. showed homology between *Planctomycetale* bacteria *Gemmata obscuriglobus* and integrin αV, including the putative Ca^{2+}-binding region common to integrins [19].

In the course of our own search of the available sequence data for sequences matching human integrin α and β subunits, we identified several matching sequences in bacteria that aligned surprisingly well with portions of the integrin subunits [9]. For example, a sequence from the cyanobacterium *Trichodesmium erythraeum* is homologous with more than 450 residues of integrin β-subunit which includes the amino-terminal βA-domain but other domains such as the EGF repeats and transmembrane segment were not present in the sequence [9]. Sequences also matched with the integrin α subunits. On closer inspection of the alignments it is was clear that the bacterial sequences matched the repeating units corresponding to blades of the β-propeller domain, but that the reported sequences did not include the trans-membrane domain, and the Thigh, Calf-1, or Calf-2 domains were not recognizable in the sequences. However, it remained unclear whether the repeating units that were observed in bacteria would indicate the presence of an integrin-like 7-bladed β-propeller or a β-propeller formed from a different number of blades, or if they represented an entirely different fold.

As defined in the SCOP database, the N-terminal β-propeller domain of integrin α subunits has the 7-bladed β-propeller fold [20]. The integrin β-propeller domain is only one of the fourteen superfamilies of proteins with the 7-bladed β-propeller fold. In addition to the 7-bladed β-propeller fold, 4-, 5-, 6- and 8-bladed β-propeller folds also exist and each fold type contains one or more protein superfamilies (see SCOP) [20]. The 7-bladed β-propeller fold, however, is represented by the largest number of solved structures deposited with the Protein Data Bank (PDB), with more than 115 reported protein structures [21].

Structures of three different integrins containing the β-propeller domain, all from human, have now been solved: αVβ3 [11,22–24], αIIbβ3 [25–27], and αXβ2 [28]. Here, we have made an in-depth analysis based on the two highest-resolution X-ray structures: the ectodomains of human αVβ3 (3.10 Å resolution; PDB ID: 1JV2) [22] and αIIbβ3 (2.40 Å resolution; PDB ID: 2VDR) [25], as well as the sequences from the remaining 16 human α subunits. We identified the structural features that distinguish the human integrin-type 7-bladed β-propeller super-family from the other 13 superfamilies having the 7-bladed β-propeller fold and demonstrate that the structural features of the human integrin-type 7-bladed β-propeller superfamily uniquely specifies a pattern of sequence conservation that can be used to identify sequences fulfilling the requirements for forming the structures.

Previously, it was already shown, that at least four non-integrin type of 7-bladed β-propeller domains as well as several 8-, 6- and 4-bladed β-propeller domains were found in bacteria [29,30]. Herein we present strong evidence that bacterial sequences encode structures that are equivalent to the 7-bladed β-propeller found in human integrins.

Results and Discussion

Consensus sequence repeats define the seven blades of the integrin alpha N-terminal domain

There are two reported descriptions of a repeating consensus sequence for human integrin β-propeller domains, but both define the same basic consensus sequence: (1) the "FG-GAP repeat" sequence motif [31,32], and (2) the "Cage" sequence motif [22].

The FG-GAP repeat motif was defined based on sequence similarities, and a structural model of the integrin β-propeller domain proposed a repeating structure, where each "blade" contains a 4-stranded β-sheet with a Phe-Gly (FG) pair in the first strand *plus* a Gly-Ala-Pro (GAP) tripeptide in the second strand [31,32]. The Cage motif was defined by the $\phi\phi G\phi X_{13-20}PX_{2-15}GX_{5-8}$ (φ, aromatic residue; G, glycine; X, any residue; P, proline) consensus sequence reported with the X-ray structure of the human αVβ3 ectodomain [22]. The first glycine residue of the Cage motif and glycine of the FG pair of the FG-GAP repeat motif are the same residue, as are the proline residues. No 3D structural motif has, however, been defined.

When, in an earlier sequence-based search [9], bacterial sequences were identified with what appeared to be the repeated consensus sequence patterns, it was unclear whether they were truly of the integrin type, similar to one of the other 13 β-propeller fold superfamilies, or perhaps representing a novel superfamily having the same fold. Since 3D structure is well-known to be more conserved than sequence, and the information derived from analysis of known 3D structures places additional constraints on a consensus sequence motif, we made a detailed comparison of the two independent higher-resolution β-propeller domain structures from human integrins.

Defining a structure-based motif for blades of the human-type β-propeller domain

In addition to sequence data on all of the human integrin α and β subunits, we have considered the high-resolution X-ray structures for the human integrin ectodomains: αVβ3 [22] and αIIbβ3 [25]. Both of the αV and αIIb subunits contain an N-terminal domain with the 7-bladed β-propeller fold, with a 4-stranded β-sheet having the FG-GAP/Cage sequence motifs repeated in each of the 7 blades. Unlike the lower resolution αXβ2 structure [28], neither integrin has an inserted I domain in the β-

propeller domain. The "integrin alpha N-terminal domain" is only one superfamily out of 14 having the all-β class 7-bladed β-propeller fold (see SCOP) [20].

In this study, we considered only the sequences and structures of the β-propeller domain and excluded other regions including the I domain sequence that buds out of the β-propeller in 9 of the 18 human integrin α subunits. In order to more fully-characterize the repeating unit, we have made detailed comparisons of the sequences and structures in order to define a structural motif.

Any structural motif has two major features: (1) it includes segments of the polypeptide chain with identical secondary structure, which (2) interact in an identical way to fulfill their function. Thus, based on the definition of the motif we should be able to identify key essential features that define each blade of the integrin domain, differences with the other reported family members, and see if the sequence patterns in the bacterial sequences or other non-integrin sequences support the same structural requirements as the integrin β-propeller domains.

Each blade from the integrin-type β-propeller domains consists of four antiparallel β-strands and four loop regions; in Figure 1, segment A is part of loop 1 and segments B and C are part of loop 3. In every blade of the 7-bladed β-propeller, the key residues of the Cage motif reside within segments A, B and C, whereas those defined by the FG-GAP motif reside within segments A and B (Figure 1). Within a single blade, segments A, B and C interact with each other by means of a network of hydrogen bonds (Figure 2), which link five amino acids from A (residues A0–A4), five from B (residues B0–B4), and two from C (residues C1, C2).
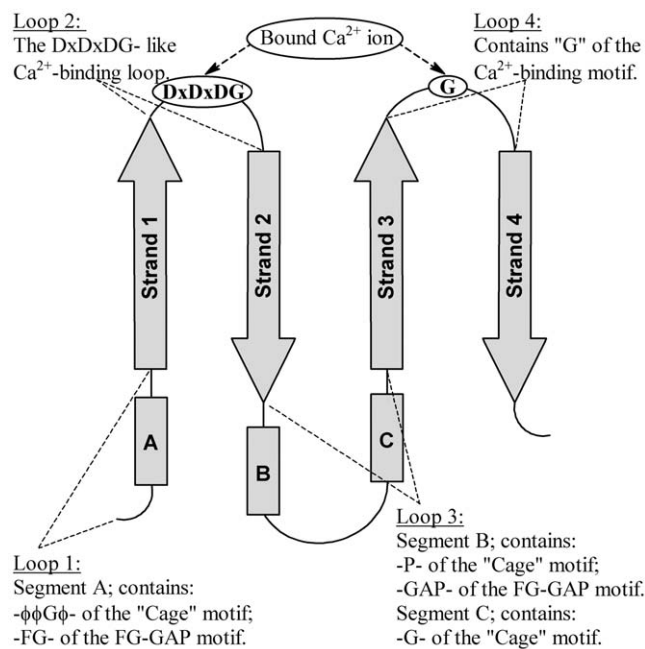
Figure 1. Schematic of the four-stranded super-secondary structure of a single blade, which is repeated seven times to form the β-propeller domain of the integrin α subunit. Segments A, B and C from the loops adjacent to β-strands 1, 2 and 3 show the location of key amino acids of the two defining sequence motifs: $\phi\phi G\phi X_{13-20}PX_{2-15}GX_{5-8}$ of the Cage motif, and the FG-GAP (pfam01839) motif. The position of the calcium-binding motif, which is found in blades 5 through 7 of all known integrins, and in blade 4 of integrins αIIb, αV, α5 and α8, is also shown in the loop regions between β-strands 1 and 2 and β-strands 3 and 4 (see Table S1 for reference). doi:10.1371/journal.pone.0025069.g001

These positions within the three segments, tabulated for all 7 blades from the β-propeller domains of the 18 human integrin α subunits (Table S1), are highly conserved (Figure 3). In addition to a high level of sequence conservation across the blades, the blades from the structures of the β-propeller domains from the human αV and αIIb subunits show that the number, geometry and orientation of the hydrogen bonds that join segments A, B and C, are identical in each of the seven blades from the two structures (Figure 2). The secondary structure of the segments A, B and C is also similar in all seven blades of the β-propeller domains of integrins αV and αIIb; segment A contains a classical type II β-turn, and segment B contains a classical type I β-turn.

A type II β-turn is a four-residue turn characterized by the following three criteria (for properties of β-turns see Richardson [33], and Hutchinson and Thornton [34]): (1) torsion angles $\varphi 2 = -60°$, $\psi 2 = +120°$, $\varphi 3 = +90°$, $\psi 3 = 0°$; (2) the residue at position 3 is almost always a glycine; and (3) a C-O⋯H-N hydrogen bond is formed between the main-chain oxygen atom of the residue at position one and the main-chain nitrogen atom of the residue at position four. The conservation of the type II β-turn in the β-propeller blades of integrins is shown via a Ramachandran plot for residues A2 and A3 (i.e. the second and third residues of the turn; Figure 4A), demonstrating that residues A2 and A3 from all of the blades in both structures except blade 6 from the integrin αV subunit (Figure 4A) have torsion angles corresponding to a classical type II β-turn. Residues A2 and A3 from blade 6 of the integrin αV subunit are among the residues noted by the authors to deviate from expected values; the structure was solved at a fairly low resolution (3.10 Å, PDB ID: 1JV2) (Figure 4C). The other two criteria of the type II β-turn, namely presence of a glycine and a stabilizing hydrogen bond, are also fulfilled with no exceptions.

As seen in the structures of the human αV and αIIb subunits, segment B corresponds to a classical type I β-turn with the consensus sequence APXX (A, alanine; P, proline; X, variable), with torsion angles near $\varphi_{B2} = -60°$, $\psi_{B2} = -30°$, $\varphi_{B3} = -90°$, $\psi_{B3} = 0°$ (Figure 4B). The conserved proline residue is found in most of the blades of the β-propeller domain in each of the 24 human integrin sequences (position B2 in Figure 2; Table S1). Of the 14 examples from the human αV and αIIb subunits, there is only one exception: blade 1 from the αV subunit has a non-standard conformation, $\varphi_{B2,Pro41} = -57°$, $\psi_{B2,Pro41} = 163°$, $\varphi_{B3,Lys42} = 64°$, $\psi_{B3,Lys42} = 32°$, which is more typical for a type II β-turn (Figure 4D).

As a consequence of the structural requirements of the type II and type I turns, two conserved residues common to the Cage and FG-GAP motifs, glycine and proline, are respectively found with high frequency at positions A3 and B2. It is due to the structural requirements that both residues are conserved in the sequence motif descriptions. In addition to the hydrogen bond stabilizing each turn type, there are 5 additional hydrogen bonds between main-chain amino and carbonyl groups that link segments A, B and C to form a rigid unit (Figure 2). These segments together with the β-sheet formed from strands 1–4 form a blade linked by a network of main-chain hydrogen bonds; the blade is further stabilized by binding Ca^{2+} at the opposite end of the blade to segments A–C.

Glycine in the type II turn and proline in the type I turn are conserved but not invariant (Figure 3). Our analysis shows that in human integrins, the probability for a glycine to occupy the third position in the first β-turn (position A3) is 89%, while the probability for a proline to occupy the third position in the second β-turn (position B2) is 71%. These probabilities show that these two types of amino acid are at least 100 times more frequent than any other type of amino acid taken separately. These data are in
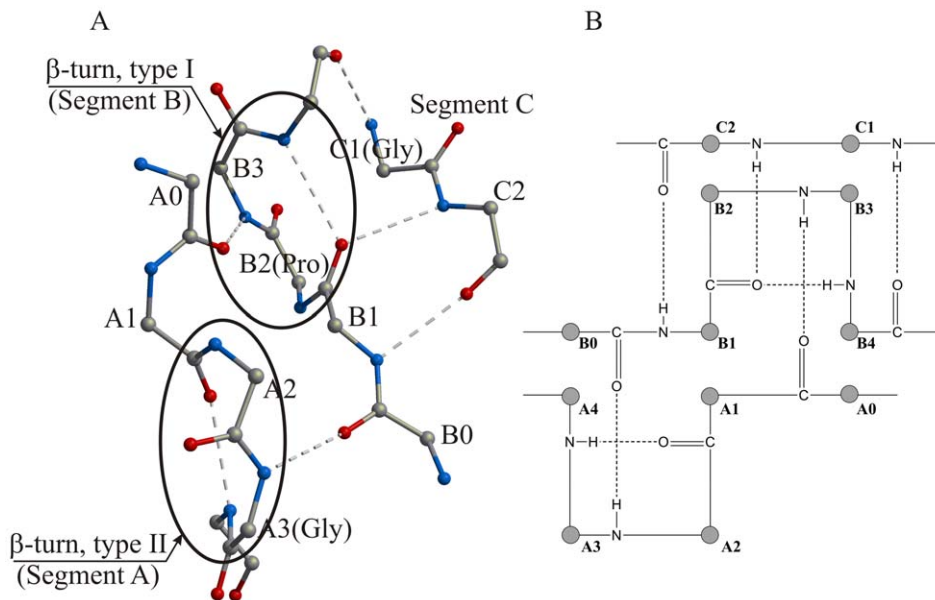
Figure 2. Three-dimensional (A) and schematic (B) representations of segments A, B and C (Table S1), which incorporate key amino acids from the Cage and the FG-GAP consensus repeat motifs. The three segments interact by means of a network of specific hydrogen bonds, which are identical for all blades from the β-propeller structures of αVβ3 and αIIbβ3. In addition, the three segments have a conservative secondary structure, where segments A and B contain β-turns of type II and I, respectively, thus requiring residue A3 (Gly170 in αIIb) to be nearly always glycine and residue B2 (Pro186 in αIIb) to be a proline. The network of conserved hydrogen bonds among segments A, B and C together with the calcium-binding motif join the loops from both sides of a four-stranded blade, resulting in a very compact blade structure. Numeration of amino acids A0–A4, B0–B4 and C1–C2 corresponds to those shown in Table S1. The structure in (A) is based on the structure of the αIIb subunit (PDB ID: 2VDR).

good agreement with the overall analysis of β-turns from unrelated protein structures [34], where among 405 type II β-turns and 1231 type I β-turns, the probability to find glycine at position A3 is 75% (13 times more frequent than any other amino acid) and the probability to find proline at position B2 is 17% (2 times more frequent). As a result, the consensus repeat sequence motif corresponds to a specific 3D H-bonded structural motif interacting by means of a network of hydrogen bonds (Figure 2).

Ca^{2+}-binding motifs stabilize the structure of several integrin blades

In the structures of the human αV and αIIb subunits, located at the opposite end of the blade from the FG-GAP/Cage consensus repeat, there are two loops joining the four β-strands that are involved in binding Ca^{2+} (Figure 1; Table S1). The Ca^{2+}-binding motif is not present in all seven blades of the β-propeller domain. However, in those blades where the motif is present it is conserved and consists of the well-known "DxDxDG-like Ca^{2+}-binding loop" (D, aspartate; in the 1,2-loop, Figure 1), present in a large number of unrelated Ca^{2+}-binding proteins [35], plus a glycine residue (in the 3,4-loop) that coordinates via main-chain atoms a water molecule bound to Ca^{2+}. Thus, in the integrins the motif for calcium binding is structural and is more complex than a DxDxDG-like Ca^{2+}-binding motif, and serves to bridge the two loops (Figure 5). The motif is conserved across human integrins for blades 5, 6 and 7, and additionally, is also present in integrins αV, αIIb, α8 and α5 in blade 4 (Table S1). In the X-ray structures of αV and αIIb, Ca^{2+} ions are bound to all four of the Ca^{2+}-binding loops. Consequently, in the integrins, the consensus repeat motif and the Ca^{2+}-binding motif form two anchors at opposite ends of the β-sheet of the blade that stabilize the structure of the integrin β-propeller blades from both sides.

The structural motif distinguishes the integrin β-propeller from 13 other superfamilies

The β-propeller domain of the integrin α subunits corresponds to one of the fourteen superfamilies of 7-bladed β-propeller fold classified in the SCOP database [20]. An examination of all 32 representative structures from the other 13 superfamilies was conducted and four specific characteristics distinguish the integrin β-propeller from the other superfamilies demonstrating that the structural consensus motif for a blade exists only in the integrin superfamily and is not present in the other 7-bladed β-propeller fold superfamilies:

Characteristic 1, the presence of the Cage motif: The Cage motif [22] is not present in any of the representative structures from the other 13 superfamilies. This can be seen by using a pattern value description of the Cage motif to check for the sequence motif with the PATTINPROT web-based service (PATTINPROT, http://npsa-pbil.ibcp.fr/). Moreover the conservation of the key glycine and proline residues was not observed in the 13 superfamilies.

Characteristic 2, the presence of type II and type I turns: Only in integrin α subunits, and not in the other 7-bladed β-propeller fold proteins, we observe the presence of two β-turns, one type II and one type I turn, within each blade. In the 32 representative structures from the other 13 superfamilies, 12 structures do not contain even one pair of β-turns within two adjacent segments A and B as seen in the integrins (see Figures 1 and 2). The remaining 20 representative structures (Table 1) contain a pair of β-turns in at least one of the seven blades, with a maximum of four examples seen only in the nucleoporin domain superfamily. Thus, with the exception of the integrin α subunits, there are no other structures from any of the superfamilies of 7-bladed β-propeller domains that have β-turns in the adjacent segments A and B in all seven blades.

Integrins without the I-domain:
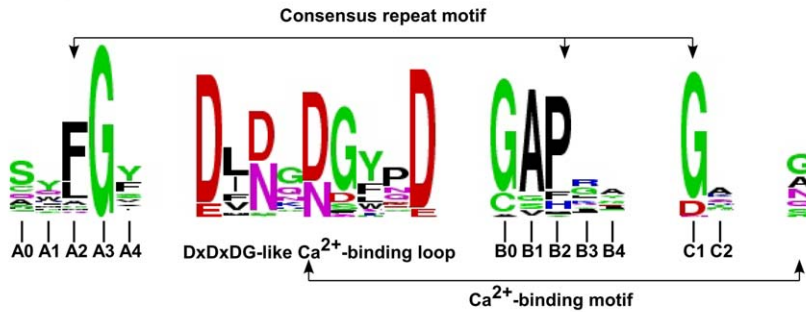
Integrins with the I-domain:
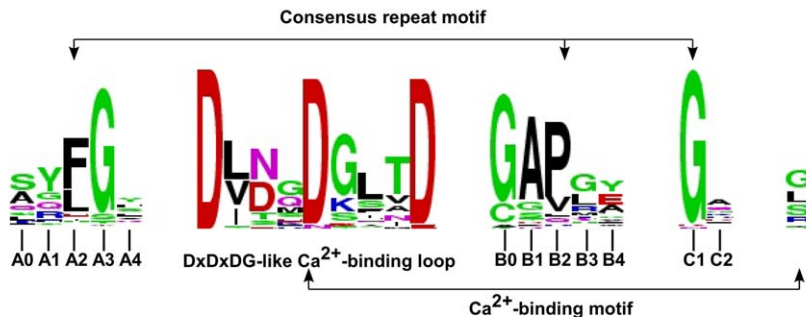
All the integrins (with and without the I-domain):
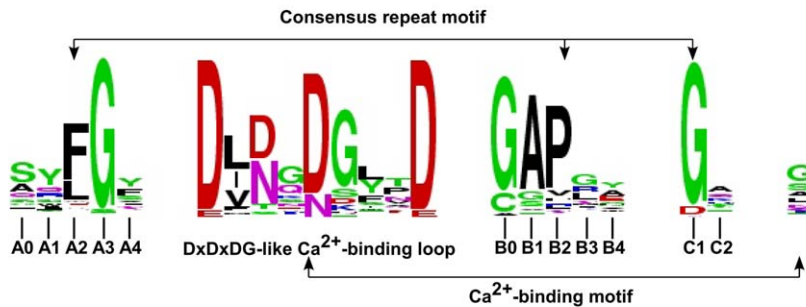
Figure 3. A WebLogo analysis of amino acid side chains of the consensus repeat motif incorporated into segments A, B and C and the calcium-binding motif from each of the seven blades of all of the human integrin sequences. Integrins with and without the I domains were analyzed together and separately.
doi:10.1371/journal.pone.0025069.g003

Characteristic 3, a hydrogen bonding network linking segments A, B and C: Only in the integrins, in each of the seven blades, the two β-turns residing in the segments A and B are closely aligned and interact with each other via a network of hydrogen bonds, as shown in Figure 2. Although two β-turns are found on adjacent segments A and B in some blades in the other representative structures (Table 1), in the X-ray structures these turns are without exception located beyond H-bonding distance.

Characteristic 4, the presence of a Ca^{2+}-binding motif: The Ca^{2+}-binding motif formed from two adjacent loops only found in the integrin-type β-propeller domain.

Characteristic 5 (non-specific), a non-sequential 4th strand in blade 7 arising from the sequence N-terminal to blade 1 of the domain. This is seen in the structures of the β-propeller domain

from the human αV and αIIb subunits but it is not a feature unique to the β-propeller domain from the human α subunits.

The presence of four characteristics specific for the integrin-type β-propellers and one characteristic specific to the β-propeller fold in general, appears to uniquely define the blades of the integrin-type β-propeller domain. Furthermore, these features demonstrate that the repeating unit is an extensively stabilized structural motif, defined by sequence and structural features, which include the FG-GAP/Cage sequence motif forming two interlocked β-turns, the Ca^{2+}-binding motif, as well as the four antiparallel strands and the extensive hydrogen bonding network.

Search for the consensus repeat motif in bacteria

For the purposes of this study, we established extremely restrictive criteria in order to identify bacterial sequences that
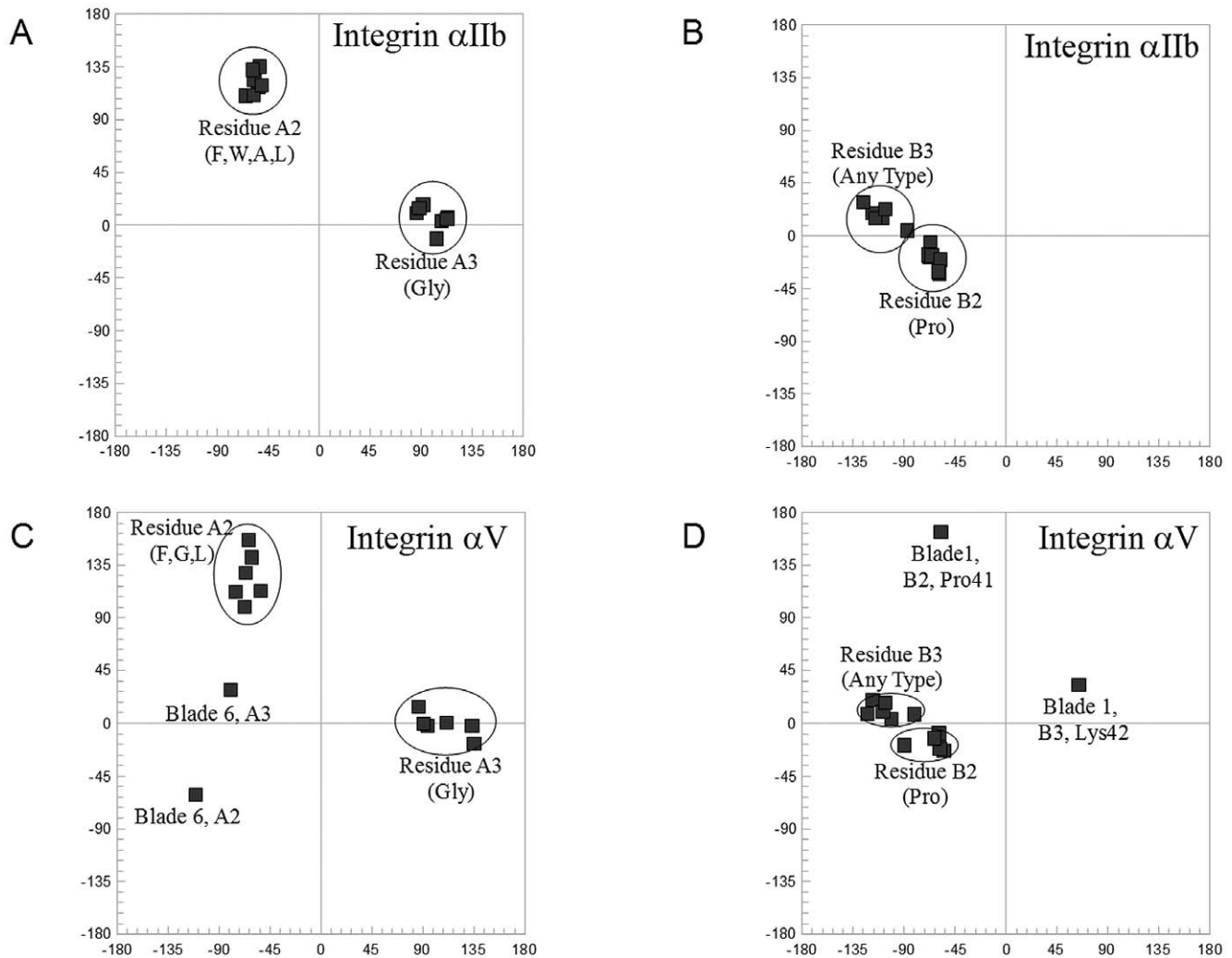
Figure 4. Fourteen pairs of torsion angles of amino acids A2 and A3 (segment A), and B2 and B3 (segment B), from the αV and αIIb subunits of two integrin structures are summarized on four Ramachandran plots (A) through (D). Values of the torsion angles φ and ψ and the amino acid composition for the second and third amino acids from the conserved β-turns within the segments A and B correspond to those of classical β-turns of type II (amino acids A2 and A3) and type I (amino acids B2 and B3). There are two exceptions: (C) in the 6th blade of integrin subunit αV segment A contains a non-standard β-turn instead of a type II β-turn; and (D) in the 1st blade of the αV subunit segment B contains a type II β-turn instead of a type I β-turn. Residue A3 is a glycine and residue B2 is a proline in almost all β-turns, which is in accordance with the β-turn type.
doi:10.1371/journal.pone.0025069.g004

are most similar to the human integrin β-propeller domain. First, we extracted all sequences matching at least one copy of the signatures for the human integrin-type β-propeller domain using Pfam (consensus sequence pfam01839) [36]. The pfam01839 signature motif defines a blade of a human integrin β-propeller domain and incorporates information on the secondary structure, as well as the signatures for the Ca^{2+}-binding site and the consensus repeat that includes the FG-GAP/Cage sequence. As the result, 1093 sequences were identified, of which 473 are eukaryotic sequences and 620 are bacterial sequences. Among the 1093 sequences, 9 eukaryotic and 58 bacterial ones were redundant or annotated as "obsolete" by the UniProtKB protein database (http://www.uniprot.org). The remaining 464 eukaryotic sequences and 562 bacterial sequences are given in the supplementary material section as two separate fasta files, eukaryote_464_init_record.fas and bacterial_562_init_record.fas, for eukaryotic and bacterial sequences respectively. The taxonomic groups and the numbers of identified sequences are shown in

Table 2. Each of the 562 bacterial sequences was inspected for the presence of at least seven repeats of the FG-GAP/Cage consensus signature, both significant and insignificant (as defined in Pfam), using the Pfam "sequence search" procedure. As the result, we identified 229 such sequences, which we give as the supplementary fasta file (bacterial_229_with_7_signatures.txt). Each of the 229 sequences in the file contains additional information about the number of identified signatures. Some sequences had 14 total signatures, suggesting the presence of tandem copies of the putative domain. Presence of a Pfam consensus signature does not automatically mean the presence of the FG-GAP/Cage motif. From the 229 identified sequences we removed all the sequences, where the Pfam consensus signature did identify protein subsequences, which were shorter than that is required to incorporate all structural elements of the structure-based motif described earlier. As the result, we obtained 35 sequences from 21 different bacteria (given in the supplementary material section as a file bacterial_35_with_7_full_length_motifs.txt), which contained sev-
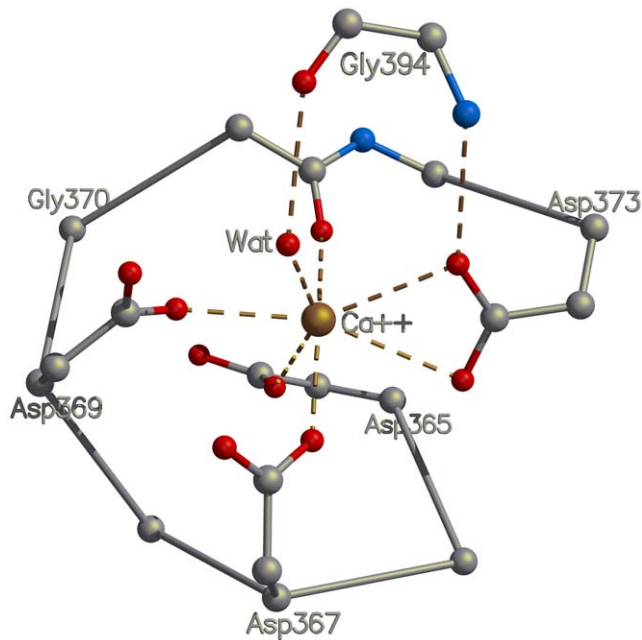
Figure 5. The calcium-binding motif present in several blades of the human integrin β-propeller domain is the well-known DxDxDG-like Ca²⁺-binding loop. In the integrin structures, the calcium-binding motif has an identical structure and similar mode of calcium binding. In every calcium-binding blade, conserved amino-acid side chains from the loop between β-strands 1 and 2 coordinate a calcium ion by means of a network of ionic interactions, while the main-chain oxygen and nitrogen atoms of the amino acid from the loop between β-strands 3 and 4 interacts with the calcium cation through a conserved water molecule and a conserved side chain of an aspartate residue.
doi:10.1371/journal.pone.0025069.g005

en full-length segments able to incorporate the structure-based motif for blades of the human integrin-type β-propeller domain described earlier, and where each of the seven full-length segments contained the Pfam-defined FG-GAP/Cage consensus signature. The 35 sequences were then scrutinized in order to identify and confirm the presence of the consensus FG-GAP/Cage motif in each of the seven repeat signatures. This led to 9 sequences from five different bacterial species, of which 5 have been aligned with the structural alignment of the repeats found in human integrin subunits αV and αIIb (Figure 6). The alignment contains one representative (A3VFV0_9RHOB) from *Rhodobacteriales* HTCC 2654 where there are 4 sequence entries, each with 14 repeats. *Nitrococcus oceani* ATCC 19707 has two sequence entries with seven repeats; one representative sequence, Q3JAB2_NITOC, is present in the alignment of Figure 6. The alignment of the repeats from all 9 sequences, including the two sets of seven repeats found in the four *Rhodobacteriales* HTCC 2654 entries and the sequence from *Synechococcus elongatus* PCC 7942 (Q31NK2_SYNP7), is presented in the Supplementary Material.

For the 5 sequence entries (seven sets of seven repeats), the individual secondary structure predictions and the consensus predicted secondary structure made using three different methods, PHD [37], PSIPRED [38] and PROF [39] (see Materials and Methods), coincided well with that of the known human β-propeller domain structures, with an identical distribution of β-strands as seen in the structures of integrins αV and αIIb (Figure 6). In the X-ray structures, the fourth β-strand in the final seventh blade of the β-propeller domain arises from sequence located just

before the sequence of the first blade. In this way, the N- and C-terminal ends of the β-propeller domain in the integrins are locked together. In each of the bacterial sequences a short β-strand was predicted adjacent to the N-terminus of the first β-strand of the first blade of every 7-blade repeat. The N-terminal sequences observed and predicted to be contributing to the fourth strand of the seventh blade are shown at the very end of the alignment in Figure 6.

Conserved residues in the bacterial sequences, labeled "Bact_Consensus" and computed for each individual blade, are shown in Figure 6 in comparison with the consensus Hidden Markov Model signature sequence of the "FG-GAP" (pfam01839) motif computed over all blades of the β-propeller domain from integrin α subunits. Both consensus sequences show that the three key glycine residues from the FG-GAP/Cage motif are highly conserved. The key proline present in many of the integrin repeats is also present in repeats from the bacteria. More surprisingly, the comparison of the two consensus sequences showed that the Ca²⁺-binding site exists and it is conserved in each of the bacterial sequences but, in contrast to the human integrin β-propeller domains, the signature is present in all 7 blades. This strongly suggests that the bacterial sequences form β-propeller domains, quite similar to the β-propeller domains of the human integrin α subunits, and that the Ca²⁺-binding motif is exploited even more frequently than in the integrins to form one of the two anchors that stabilizes each blade of the β-propeller. This in itself suggests that the bacterial sequences might form a more regular, disk-like three-dimensional than those of the human integrins. It also would be consistent with the idea of seven tandem repeats arising first in bacteria and then later being adapted for use in the integrins with a loss of some Ca²⁺ binding sites.

The analysis reported here demonstrates that the bacterial sequences do not share the features of the *other* 13 families of β-propeller fold proteins, but clearly display the sequence and implied structural features of the 7-bladed β-propeller domains of the integrins. The similarities are striking and suggest a common origin of sequences derived from an approximately 60-residue repeat unit. In comparison to human integrins, these bacterial sequences have retained more common features within each repeat, i.e. a complete set of seven Ca²⁺-binding sequences and more consistent lengths of the loop regions, which is consistent with the notion of the bacterial sequences still retaining features of an ancestral form of the β-propeller domain that was later exploited in the integrin α subunits apparently arising early in or prior to the origins of multicellular organisms.

One can not help but ask what possible functions the integrin-like β-propeller domains have in bacteria. In the integrins the β-propeller domains interact with the βA-like domain of the integrin β-subunit. Ligands bind either to this complex (see, e.g. the structure of αVβ3 ectodomain with RGD peptide) [11] or to an αI domain budding out from the 7-bladed β-propeller domains (see for example, the structure of the α2I domain with bound collagen-like peptide) [40]. The integrins are also membrane-bound proteins, and the bacterial sequences that we have retrieved here do not have hydrophobic sequences consistent with a transmembrane helix that could anchor the protein to a membrane. In addition to being more uniform across the 7 repeats than the integrin sequences, some bacterial sequences have two closely spaced sets of 7 repeats, suggesting they would correspond to tandem 7-bladed β-propeller domains.

For the bacterial proteins any functional role would be limited to the cytoplasm. In integrin β-propeller domains Ca²⁺ may play a stabilizing structural role and Ca²⁺ levels are high outside of cells, thus the sites should always be occupied. Within cells, Ca²⁺ is

Table 1. Presence of β-turns in segments A and B in the seven blades of 20 representative structures of proteins that have the 7-bladed β-propeller fold (*excluding* the integrins).

Protein Family	Blade/Turn type						
	Blade 1	Blade 2	Blade 3	Blade 4	Blade 5	Blade 6	Blade 7
1. YVTN repeat:							
1L0Q_A: Segment A	−	−	−	−	−	−	Type I
1L0Q_A: Segment B	+	+	+	+	+	+	Type I
2. Quinohemoprotein amine dehydrogenase B chain:							
1PBY_B: Segment A	Type II′	−	+	−	−	−	−
1PBY_B: Segment B	Type VIa1	+	−	−	−	−	−
3. Nitrous oxide reductase, N-terminal domain:							
1FWX: Segment A	Type II	−	−	−	−	−	−
1FWX: Segment B	Type I	+	−	+	+	+	+
4. WD40-repeat:							
1NR0_A(domain I): Segment A	−	−	−	−	−	Type I	−
1NR0_A(domain I): Segment B	+	−	+	+	+	Type I	+
1NR0_A(domain II): Segment A	−	−	−	−	−	−	Type II
1NR0_A(domain II): Segment B	+	+	+	+	+	+	Type I
5. RCC1/BLIP-II:							
1A12_A: Segment A	−	−	−	−	Type I	−	+
1A12_A: Segment B	+	+	+	+	N/S	+	−
6. Clathrin heavy-chain terminal domain:							
1UTC_A: Segment A							
1UTC_A: Segment B	Type I	−	−	Type II	−	Type I	−
	Type II′	+	−	Type I	+	Type VIb	+
7. Peptidase/esterase 'gauge' domain:							
1H2W_A: Segment A	+	Type I	−	+	−	−	−
1H2W_A: Segment B	−	Type I′	−	−	+	+	−
8. Tricorn protease domain 2:							
1K32_A: Segment A	Type I	−	−	−	−	−	−
1K32_A: Segment B	Type I′	+	+	+	−	−	+
9. 3-carboxy-cis, cis-mucoante lactonizing enzyme:							
1JOF_A: Segment A							
1JOF_A: Segment B	Type I	+	N/S	Type I	−	+	+
	Type II′	−	Type I	Type I	+	−	−
10. Putative isomerase YbhE:							
1RI6_A: Segment A	−	−	Type II	N/S	−	−	−
1RI6_A: Segment B	−	+	Type I	N/S	−	+	+
11. Sema domain:							
1OLZ_A: Segment A	−	−	Type I	+	−	−	−
1OLZ_A: Segment B	−	+	Type I	−	−	+	+
12. Oligoxyloglucan reducing end-specific cellobiohydrolase:							
1SQJ_A: Segment A							
1SQJ_A: Segment B	−	N/S	+	−	Type II	−	−
	−	Type I	−	−	Type VIb	−	−
13. Nucleoporin domain:							
1XKS_A: Segment A	Type I	N/S	−	−	Type I	+	Type I′
1XKS_A: Segment B	Type II′	Type I	+	−	Type I	−	Type I

"+", a β-turn is present; "−", a β-turn is absent. The β-turn type is indicated when simultaneously present in both segments A and B, similar to what is seen in integrins. N/S, the type of the β-turn is not specified.
doi:10.1371/journal.pone.0025069.t001

PLoS ONE | www.plosone.org   8   October 2011 | Volume 6 | Issue 10 | e25069

Table 2. Taxonomic groups and the identified number of sequences deposited to the UniProtKB database that contain domain architectures similar to (and including) the integrin α (β-propeller) superfamily.

Eukaryotes (464):			
	Chordates (323):		
		Vertebrates (308):	
			Mammals (232); Bony fishes (51); Amphibians (18); Birds (7);
		Lancelets (12); Tunicates (3).	
	Arthropods (77):		
		Insects (74); Ixodida (3).	
	Echinoderms (6); Cnidarians (8); Nematodes (6); Placozoans (2); Sponges (1); Other Eukaryotes (41).		
Bacteria (562):			
	Proteobacteria (123):		
		G-proteobacteria (33); D-proteobacteria (50); B-proteobacteria (6); A-proteobacteria (31); Other proteobacteria (3).	
	Actinobacteria (149); Cyanobacteria (96); Bacteroidetes (58); Planctomycetes (20); Acidobacteria (33); Spirochetes (9); Firmicutes (12); Chloroflexi (10); Lentisphaerales (5); Chlorobi (5); Verrucomicrobia (39); Other Bacteria (3).		

doi:10.1371/journal.pone.0025069.t002

generally at very low concentrations, thus giving it a unique role as a secondary messenger in signaling. In bacteria, calcium plays an increasingly recognized role in signaling [41,42] and other critical cellular processes [43–45]. The bacterial domains have 7 repeated highly conserved Ca^{2+}-binding motifs, sometimes duplicated in the same sequence, and it likely these sites are functional, but what role they might play in bacteria is unclear at this time. The high similarity between the bacterial domains and metazoan integrins and presence of the key features required for the structure, strongly suggest that the presence of this domain in modern bacteria belies the origin of β-propellers in the integrins. A key question concerns the basis for the apparently different functions in bacteria to those arising in the integrins in multicellular organisms.

Integrins, β-propellers and origins of multicellularity

Integrins are present in species that represent the earliest metazoans. We have made a comprehensive search for integrin-type β-propeller domains from sequences of unicellular eukaryotes, including those submitted to sequence and protein databases as well as those that are located with current genome-sequencing projects. There are 47 on-going genome projects of eukaryotic unicellular organisms with identified genomic sequences at different stages, from which there are 38 projects where protein

sequences have been deposited (NCBI Genome: http://www.ncbi.nlm.nih.gov/sites/genome), and nine others that include *C. owczarzaki* and *T. trahens* (Sebé-Pedrós et al. reported matches in these two species) [16].

The integrin α subunit-like sequence from *T. trahens* (formerly *Amastigomonas sp.*; sequence AMSG_06621.1) has three FG-GAP repeatsand the sequence record we obtained from the Broad Institute "Origin of Multicellularity" project did not have the transmembrane sequence. However, the sequences from *Capsaspora* have three repeats, a predicted transmembrane segment, and an adjacent KXGFFXR integrin motif. In contrast, sequence matches between true integrin β-propeller domains and domains observed in bacteria match all seven blades and the alignment is excellent throughout, albeit the latter examples additionally contain Ca^{2+}-binding motifs in all seven blades. Aside from the sequences reported by Sebé-Pedrós et al. no clear sequentially-repeated FG-GAP motifs were detected within the 47 genomes [16].

Ectocarpus siliculosus is a multicellular brown algae and phylogenetic studies place *this species* among the heterokonts (stramenophiles), which are eukaryotes that also include a large number of unicellular diatoms, and whose origin of multicellularity occurred *independently* from the metazoans [46]. The recent sequencing of the genome of *Ectocarpus* has identified several "integrin-like" sequenc-

Protein codes	Organism name	Protein name	PDB ID	UNIPROT	NCBI
2VDR	*H. sapiens*	Integrin αIIb	2VDR	-	-
1JV2	*H. sapiens*	Integrin αV	1JV2	-	-
A3VFV0	*Rhodobacterales bact* HTCC2654	Hyp. protein	-	A3VFV0_9RHOB	ZP_01013484
A5L751	*Vibrionales bact.* SWAT-3	Hyp. protein	-	A5L751_9GAMM	ZP_01816147
Q3JAB2	*N.oceani* ATCC 19707	Integrin α-like	-	Q3JAB2_NITOC	YP_343764
A0YNP0	*Lyngbya* sp. PCC 8106	Hyp. protein	-	A0YNP0_LYNSP	ZP_01620661
Q31NK2	*S.elongatus* PCC 7942	Integrin α-like	-	Q31NK2_SYNE7	YP_400354

```
Beta propeller repeat # 1
FG-GAP         HGSSVA-AGDLNGDGRHDLVVGAPGAD--GGTIGSVYLL
Consensus :D  G      S  GD LNGDG      D        G          G
2VDR      17:SQFGFSLD-FHK-DSHGRVAIVGAPRTL[4]EETGGVFLCPWRAEGG-QCPSLLFD-LRDET:76
1JV2      17:SYFGFAVD-FIVPSASSRMFLLVGAPKAN[6]VEGGQVLKCDW-SSTR-RCQPIEFDATGNPE:79
A3VFV0    26:DQAGYSVSNAGDVNGDGYDDVIVGAPYGDNGGGNSGEAYVIFGKANGL-ANIDLTSLTLTDGF:87
A3VFV0   520:VYLGRAVSGIGDVNGDGIDDVIVAEPYSTNGLSQNGAAYVIFGKAGGFSGDIDVDLTATDGF:582
A5L751   350:DSFGKVFS-IGDYNGDGVADLFVGYPTED[4]NDAGAVTVVYGGDS[7]MLSSRGFHQDSSGL:416
Q3JAB2    72:SSPGIAVSGVGDVNGDGIDDLIIGIPGADSGNGFSGASYVVFGSSGGLSPLELSSLNGSNGF:134
A0YNP0    25:DFSGGSVSSAGDVNGDGFDDVIIGSSSAD[4]SRAGESYVVFGSSVGFSPSLELSSLNGSNGF:88
Q31NK2    26:DYSGRSVSGAGDVNGDGIDDLIIGANYADPNGGYSGQSYVVFGSDQGFSADLNLADLDGSNGF:88
Q31NK2   668:DFSGGSVSNAGDVNGDGIDDLIIGARDAD[4]NSAGQSYVVFGSDQGFSADLNLSSLDGSNGF:731

Beta propeller repeat # 2
FG-GAP         HGSSVA-AGDLNHDGRHDLVVGAPGAD--GGTIGSVYLL
Consensus :D  G        GD NDG  D        G   V   G   G         L
2VDR      91:QGLGASVV-SNS------DVIVACAPWQH[13]PVGSCFLAQP------ESGRRAEYSPCRGN:149
1JV2      92:QWFGASVR-SKQ------DKILACAPLYH[10]PVGTCFLQDG------TKTVEYAPCRSQ:145
A3VFV0    96:DNAGQSVAAAGDVNGDGIDDLIIVGAPGGDDGGGAGEAYVIFGKSSGI-ANIDLSSLTLTDGF:157
Rba_Ia   593:LGDYASATGAGDVNGDGIDDVFIGARLWD[5]SAEGAGYVIYGKTSGH-GTIDLDTLTQTDGF:656
A5L751   425:DLFGYALA-SGDFNNDGFDDVAIGSPNED[4]SNVGMVTVLYGSNKGL-TTSSYQLLKEANIS:486
Q3JAB2   143:DNAGIAVSGVGDVNGDGIDDLIVGAPGVDSNGSGSGAGYVVFGSSGGFGPSLELSSLNGSNGF:205
A0YNP0    97:DLSGGSVSSAGDVNGDGFDDVIIGASEAS[4]RYVGESYVVFGSSGGFSSSLDLSSLNGSNGF:160
Q31NK2    97:DRSGISVSNAGDVNGDGIADLIVGANYADPNGNLSGQSYVVFGSDQGFSADLNLSTLNGSNGF:159
Q31NK2   740:DYSGRSVSGAGDVNGDGVDDLIIGAFGADPNGNLSGQSYVVFGSDQGFSADLNLSSLDGSNGF:802

Beta propeller repeat # 3
FG-GAP         HGSSVA-AGDLNGDGRHDLVVGAPGAD--GGTIGSVYLL
Consensus :CEAGF  SSV-VTQ-----AGELVLGAPGGY---YFLGLLAQAEVAD[12]HVSSQSLSFDSSNP:228
2VDR     167:CEAGFSSV-VTQ-----AGELVLGAPGGY---YFLGLLAQAEVAD[12]HVSSQSLSFDSSNP:228
1JV2     155:CQGGFSEL-FTK-----ADRVLGGPGSF---YWQGQLISDQVAE[12]KYNNQLATRTAQAI:216
A3VFV0   166:DQAGHSVAAAGDVNGDGIDDVIVGAYGGDDGGNAGEAYVIFGKSSGI-TNIDLSTLTLTDGF:227
A3VFV0   665:GAFASDVAGAGDFNGDGIDDVMFASPNLP[4]ANQGKTYVVFGVAGTR-ANTDVASLAPGSFI:727
A5L751   494:DKFGFALA-SLDFDGDGYDDLAIGSPYED[4]DNAGVVYVSYGTSNGLSKKRISRFHQNNSGL:556
Q3JAB2   214:ENAGISVSGAGDVNGDGLSDLIMGAYGASPNGSGSGAGYVVFGSSGGFGPSLELSGLDGSNGF:276
A0YNP0   169:DYSGRSVSGAGDVNGDGFDDLIIGASYAD[4]SFAGESYVVFGSDSGFSSSLELSTLDGSNGF:232
Q31NK2   168:DQSGFSVSNAGDVNGDGIDDLIIGAYFADPNGNLSGQSYVVFGSDQGFSAALNLADLDGSNGF:230
Q31NK2   811:DFSGFSVSSAGDVNGDGIDDLIIGASLADPNGARSGQSYVVFGSDQGFSADLNLADLNGTNGF:873

Beta propeller repeat # 4
FG-GAP         HGSSVA-AGDLNGDGRHDLVVGAPGAD--GGTIGSVYLL
Consensus :D  G      AGD NGDG   D   GA          G    V  FG    G            GF
2VDR     233:GYWGYSVA-VGEFDGD[4]EYVVGAPTWS----WTLGAVEIIDS------YYQRLHRLFGEQMA:286
1JV2     220:SYLGYSVA-VGDFNGDGIDDFVSGVPRAA----RTLGMVYIYLGK-----NMSSLYNFTGEQMA:273
A3VFV0   236:DSAGWSVSSAGDVNGDGYDDVVIGALNGDIGGINAGEAYVIFGKSSGI-ANIDLACLSLSDGF:297
A3VFV0   736:DYSGHGVAHAGDVNGDGIDDIIIGALGANGANGDSGAAYVIFGRSTPV-STIAAGSLTASEGF:797
A5L751   565:DRFGHELA-AGDFNGDGLGDLAIGAPYED[4]DDEGMVVVLFGTTQGLRTMEYREFYQGDRSGF:627
Q3JAB2   285:DSAGISVSGAGDVNGDGIDDLIVGAYGAGPDAYTNSGTSGAGYVVFGSRRGFAPSLELLNLNGNNGF:347
A0YNP0   241:DNSGRWVSSAGDVNGDGFDDLIIGANGAD[4]SSAGESYVVFGSDSGFSP-LELSSLNGSNGF:303
Q31NK2   239:DYSGRSVSGAGDVNGDGFDDFIIGASGADPNGSASGQSYVVFGSDQGFSAALNLADLDGSNGF:301
Q31NK2   882:DQSGFSVSGAGDVNGDGIDDLIIGAFLADPNGNYSGQSYVVFGKTDDFDTELNLADLDGSNGF:944

Beta propeller repeat # 5
FG-GAP         HGSSVA-AGDLNGDGRPDLVVGAPGAD--GGTIGSVYLL
Consensus :D  G             D   GA          G      V     G         G
2VDR     287:SYFGHSVA-VTDVNGDGRHDLLVGAPLYM[8]AEVGRVYLELQPRGPHALGAPSLLLTGTQLY:353
1JV2     274:AYFGFSVA-ATDINGDDYADVFIGAPLFM[8]QEVGQVSVSLGRASG---DFQTTKLNGFEVF:337
A3VFV0   306:DQAGISVSGAGDVNGDGYDDVIVGAFFGDDGGPNAGEAYVIFGKSGGF-SHIDLTGLSPSVGI:367
A3VFV0   806:DQFGFDVAGGGDVNGDGLDDVVVGARYNDENGTFSGAAYIVYGRSTPV-STIAAGSLTASEGF:867
A5L751   635:DHFGYTLA-AGDFNNDAIDELVIGVPYED[6]NNDGNVFVLRGTSSGPSRNSIREIHQDKSGV:699
Q3JAB2   356:DNAGISVSGMGDINGDGLGDLIVGAYGAGPNGRASGASYVVFGSRSGFAPSLELSSLNGSNGF:418
A0YNP0   312:DQAGISVSGAGDVNGDGFDDLIIGSRNAD[4]SNAGESYVVFGSDSGFSSSLELSSLSLDGSNGF:375
Q31NK2   310:DRSGFSVSGAGDVNGDGIDDLIIGAFGAD[4]NSAGQSYVVFGSDQGFSADLNLSSLDGSNGF:373
Q31NK2   953:DQSGFSVSGAGDVNGDGINDLIIGALGAS[7]TNAGQTYVVFGSDQGFSADLNLSALNGRNGF:101

Beta propeller repeat # 6
FG-GAP         HGSSVA-AGDLNGDGRHDLVVGAPGAD--GGTIGSVYLL
Consensus :D  G                 D   V  FG  G
2VDR     354:GRFGSAIAPLCDLDRDGYNDIAVAAPYGG--PSGRGQVLVHLGQSEGL-RSRPSQVLLSPFPT:413
1JV2     338:ARFGSAIAPLGDLDQDGFNDIAIAAPYGG--EDKKGIVYIFNGRSTGL-NAVPSQVLLEGQWAA:397
A3VFV0   376:DFAGYSVSAAGDVNGDGIDDVIVGAYSGDDGGVVHSGEAYVVFGKASGI-SHIDLSTLTAADGF:437
A3VFV0   876:DYAGLAIAIVGDMNGDGLSEVLVTAHGQDDSAVGAGAAYVIYGSTTGP-GTVNLDDLSPAVGF:937
A5L751   707:DKFGFALT-TGDFDGDGYQDLAIGSPTED[4]DNGGKVDVFFGGSGNGLRLSGSREYNQNKNNI:769
Q3JAB2   427:DASGISVSGVGDVSGDGLNDFLIGAPGAAPNGNFSGASYVVFGNSVGFGTSLELADLDGNNGF:489
A0YNP0   384:DRSGLSVSSAGDVNGDGFDDLIIGASAD[4]SNAGESYVVFGSSGGFSPSLELSTLDGSNGF:447
Q31NK2   382:DLSGFSVNGAGDVNGDGIDDLIIGARNADPNGGASGQSYVVFGKKGGFSADLNLSALNGNNGF:444
Q31NK2  1028:DESGRSVSGAGDVNGDGIDDLIIGASNADPNGNFSGQSYVVFGKKGGFAAELNLSDLNGVNGF:109

Beta propeller repeat # 7
FG-GAP         HGSSVA-AGDLNGDGRHDLVVGAPGAD--GGTIGSVYLL
Consensus :D  G       AGD  GDG DD          G      V   G
2VDR     415:SAFGFSLRGAVDIDDNGYPDLIVGAYGA-------NQVAVVRAQPLNLDPVQLTFYAGPNG-:16
1JV2     402:PSFGYSMKGATDIDKNGYPDLIVGAFGV-------DRAILYRARPFNLDVDSPAEYSGPEG-:16
A3VFV0   446:DGAGHGVSGAGDVNGDGFDDVIVGALWGAGDSGTDGAEAYVVYGGSMSARVIDLSFLTTAQGV:17
A3VFV0   946:DNFGIAAAAASDVNGDGLADIMIGANLNDDIAASSGKAYIVYGRGGFASEVDMSSLAPPYGV:509
A5L751   777:DKFGYSLM-AGDFDGDDKDDLVIGVPFED[6]NNDGTVHVLYGDS KNRKIIDFNLNNRDVMV:342
Q3JAB2   498:EASGDSVSGAGDVDGDGADDFIIGAYR-------SGTSVVVGFS VQAQTLTLSDLDGHNGF:63
A0YNP0   456:DRSGGSVSSAGDVNGDGFDDLIIGAYGAD[4]SFAGESYVVFGFS MT-SSINLSDLNGSNGF:16
Q31NK2   453:DNSGTSVSNAGDVNGDGVDDLLIGANYAD[4]TNAGQSYVVFGNA MAKASLNLSTLNGSNGF:17
Q31NK2  1099:DGFGSSVSGAGDVNGDGIDDLAIGAFTADPNGSQSGQSYVVFGNA ADKASVSLSSLDGGNGF:659
```

Figure 6. Sequence alignment of the β-propeller domains from the human integrin αV and αIIb subunits and seven sequences from five different bacteria, where seven FG-GAP consensus repeat motifs (pfam01839) were identified. The secondary structure (β-strands of the seven blades) from the structures of αV and αIIb is shown with dark shading, while the secondary structure of the bacterial sequences, predicted by three different methods, PHD, PSIPRED and PROF (see Materials and Methods) is shown with light shading. The upper sequence shown in the alignment is the consensus Hidden Markov Model (HMM) signature sequence of the FG-GAP (pfam01839) motif, followed by the consensus sequence (designated Consensus) derived from the bacterial sequences shown in the alignment. Identical residues in the two consensus sequences are boxed.

doi:10.1371/journal.pone.0025069.g006

es [46]. One sequence of 348 residues (CBJ33612.1) matches well the N-terminal half of the 788-residue human β3 subunit (e.g. 23% sequence identity over residues 99–358; human β3 numbering), with excellent matching of the MIDAS segment. An α subunit-like sequence (CBN77719.1; 1105 residues) aligns with few gaps to the 1048-residue human integrin αV subunit (ITAV_HUMAN): percentage identity ranging from 41% over residues 245–500 (human αV numbering) to 17% sequence identity over the entire 1012-residue alignment; the 7 blades of the β-propeller domain align with each other and even the 4 Ca^{2+}-binding motifs within *Ectocarpus* align to the 4 motifs found in the human αV subunit. Differences, however, appear towards the C-terminus: there is a glycine and alanine-rich low complexity region in the *Ectocarpus* protein, followed by an alanine and proline-rich region "matched" to the transmembrane segment of the αV subunit; the adjacent cytoplasmic KXGFFXR motif found in αV is not present in *Ectocarpus*, but this motif is also absent in some integrin α subunits, too.

Thus, in two independent lines of multicellular organisms, metazoans and heterokonts, the β-propeller is conserved to the point of having Ca^{2+}-binding sites in the corresponding blades, and there are strong hints of both α-like and β-like integrin subunits in the heterokont *Ectocarpus*. The current scope of genomic sequence data is obviously sparse and incomplete. Nevertheless, these observations do raise some fundamental questions with regard to the origin and evolution of integrins and their role in the origin of multicellularity, suggesting that a *common* ancestor of both multicellular lines contained some if not all of the features of the integrin α and β subunits, including integrin-like β-propeller domains and βA domains. The observation of complete and highly-conserved integrin-type β-propeller domains in bacteria, described here, suggest that the origin of domains in integrins are indeed prokaryotic, giving rise to the components of the present-day integrin subunits. However the link between having individual domains in prokaryotes and their functions, and their integration into integrin subunits in multicellular organisms remains obscure. In metazoan integrins, the β-propeller domain of the α subunit is directly or indirectly (via the inserted A domain) involved in binding external ligands, but it is always involved in binding the βI-like domain of the β subunit. The α subunit protrudes from the exterior membrane surface where Ca^{2+} levels are typically much higher than in the cytoplasm; the highly-conserved calcium-binding sites of the β-propeller and their key structural role near the ligand and subunit-subunit binding sites explain why the formation of the integrin αβ heterodimer capable of binding a specific ligand is calcium-dependent. Although the function and cellular location of the β-propeller domains in bacteria is currently unknown, if these domains are confined to the cytoplasm, then perhaps they have a role related to calcium signaling or storage. Clearly, further

genome sequencing combined with experimental studies aimed at establishing a functional role for these integrin-type domains in non-integrin sequences is needed in order to elucidate the linkage the establishment of the present-day integrin subunits from the putative origin in single-cell organisms.

Materials and Methods

Determination and Analysis of Protein Folds, Sequential and Structural motifs

The UniProt database (http://www.uniprot.org/) and the NCBI sequence database (http://www.ncbi.nlm.nih.gov/) were used as the initial sources of protein sequence and functional information. These two databases, particularly the UniProt database, are also widely used as sources of sequence data by structural databases and motif and pattern analysis computer programs, such as the Pfam (http://pfam.janelia.org/) [36]. All of the X-ray structures were taken from the PDB (http://www.rcsb.org) [21]. Protein folds, superfamilies and families were assigned according to the SCOP database (http://scop.mrc-lmb.cam.ac.uk/scop/) [20]. If a superfamily of 7-bladed β-propeller proteins contained several known structures, the structure with the best resolution has been used as a representative of the protein superfamily. Two different sequences, FG-GAP (pfam01839) and Cage [22], were used to define the consensus repeat sequence signature within the β-propeller domain of integrins. Secondary structure prediction of the retrieved bacterial sequences was based on the consensus alignment obtained using the results of three different computer-based secondary structure prediction methods: PHD from PredictProtein [37], PSIPRED [38] and PROF [39]. The φ and ψ angles of amino acids from the β-propeller repeats in the structures of the human integrins αV and αIIb subunits were calculated using SYBYL (Tripos Associates, Inc., St. Louis, MO). Structure visualization and structural analysis of geometrical parameters of interactions were done using SYBYL and BODIL [47]. The frequency of amino acids at positions within the β-propeller repeats were analyzed using the WebLogo server (http://weblogo.berkeley.edu/) [48]. Figures in this manuscript were produced with MOLSCRIPT V2.1 [49] and Raster3D V2.4b [50].

Creation of the Bacterial Data Set

Using two signature sequences of the consensus repeat motif, pfam01839 and Cage, we have analyzed the UniProt and NCBI databases and retrieved 562 bacterial sequences that contain one or more instances of the consensus repeat signature sequence (the

562 bacterial sequences are given in the supplementary material section as a fasta file bacterial_562_init_record.fas). Next, among the 562 bacterial sequences, we have selected 229 sequences that contain seven or more consensus repeat signature sequences (the 229 sequences are given in the supplementary material section as a file bacterial_229_with_7_signatures.txt). Then, from the 229 identified sequences we identified 35 sequences from 21 different bacteria that contained seven full-length segments able to incorporate the structure-based motif for blades of the human integrin-type β-propeller domain (see Section: Search for the consensus repeat motif in bacteria). The 35 sequences were then scrutinized in order to identify and confirm the presence of the consensus FG-GAP/Cage motif in each of the seven repeat signatures (the 35 sequences are given in the supplementary material section as a file bacterial_35_with_7_full_length_motifs. txt). This led to nine sequences from five different bacterial species, which have seven consensus repeat motifs, similar to the β-propeller domains in human integrin α subunits (Figure S1). These five representative bacterial sequences were further analyzed with several secondary structure prediction methods. Predicted secondary structure and position of the consensus repeat motifs within the secondary structure elements were compared with those of known structures of human integrins and shown in Figure 6.

Supporting Information

Figure S1 Alignment of nine sequences from five different bacterial species, which have seven full-length consensus repeat motifs, similar to those found in the β-propeller domains from the human integrin α subunits. For the nine sequences, the predicted secondary structure by three different methods, PHD, PSIPRED and PROF, coincided with the secondary structure of known human integrins.
(TIF)

Table S1 Localization of the FG-GAP/Cage motif and the Ca^{2+}-binding motif within each of the seven blades of β-propeller domains of the human integrin α subunits.
(DOC)

Author Contributions

Conceived and designed the experiments: AD MSJ KD. Performed the experiments: BC AD. Analyzed the data: AD MSJ KD. Contributed reagents/materials/analysis tools: JH MSJ. Wrote the paper: BC MSJ JH KD.

References

1. Takada Y, Ye X, Simon S (2007) The integrins. Genome Biol 8: 215.1–215.9.
2. Hynes RO (1992) Integrins: versatility, modulation, and signaling in cell adhesion. Cell 69: 11–25.
3. Arnaout MA, Goodman SL, Xiong JP (2007) Structure and mechanics of integrin-based cell adhesion. Curr Opin Cell Biol 19: 495–507.
4. Shimaoka M, Springer TA (2003) Therapeutic antagonists and conformational regulation of integrin function. Nat Rev Drug Discov 2: 703–716.
5. Lee JO, Rieu P, Arnaout MA, Liddington R (1995) Crystal structure of the A domain from the alpha subunit of integrin CR3 (CD11b/CD18). Cell 80: 631–638.
6. Larson RS, Corbi AL, Berman L, Springer T (1989) Primary structure of the leukocyte function-associated molecule-1 alpha subunit: an integrin with an embedded domain defining a protein superfamily. J Cell Biol 108: 703–712.
7. Huhtala M, Heino J, Casciari D, de Luise A, Johnson MS (2005) Integrin evolution: insights from ascidian and teleost fish genomes. Matrix Biol 24: 83–95.
8. Tulla M, Huhtala M, Jäälinoja J, Käpylä J, Farndale RW, et al. (2007) Analysis of an ascidian integrin provides new insight into early evolution of collagen recognition. FEBS Lett 581: 2434–2440.
9. Johnson MS, Lu N, Denessiouk K, Heino J, Gullberg D (2009) Integrins during evolution: Evolutionary trees and model organisms. Biochim Biophys Acta 1788: 779–789.
10. Pierschbacher MD, Ruoslahti E (1984) Cell attachment activity of fibronectin can be duplicated by small synthetic fragments of the molecule. Nature 309: 30–33.
11. Xiong JP, Stehle T, Zhang R, Joachimiak A, Frech M, et al. (2002) Crystal structure of the extracellular segment of integrin alpha Vbeta3 in complex with an Arg-Gly-Asp ligand. Science 296: 151–155.
12. Diamond MS, Garcia-Aguilar J, Bickford JK, Corbi AL, Springer TA (1993) The I domain is a major recognition site on the leukocyte integrin Mac-1 (CD11b/CD18) for four distinct adhesion ligands. J Cell Biol 120: 1031–1043.
13. Humphries MJ (2000) Integrin structure. Biochem Soc Trans 28: 311–339.
14. Custodio MR, Imsiecke G, Borojevic R, Rinkevich B, Rogerson A, et al. (1995) Evolution of cell adhesion systems: evidence for Arg-Gly-Asp-mediated adhesion in the protozoan *Neoparamoeba aestuarina*. J Eukaryot Microbiol 42: 721–724.
15. Ruiz-Trillo I, Lane CE, Archibald JM, Roger AJ (2006) Insights into the evolutionary origin and genome architecture of the unicellular opisthokonts *Capsaspora owczarzaki* and *Sphaeroforma arctica*. J Eukaryot Microbiol 53: 379–384.

16. Sebé-Pedrós A, Roger AJ, Lang FB, King N, Ruiz-Trillo I (2010) Ancient origin of the integrin-mediated adhesion and signaling machinery. Proc Natl Acad Sci U S A 107: 10142–10147.

17. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402.

18. May AP, Ponting CP (1999) Integrin α- and β4-subunit-domain homologues in cyanobacterial proteins. Trends Biochem Sci 24: 12–13.

19. Jenkins C, Kedar V, Fuerst JA (2002) Gene discovery within the planctomycete division of the domain Bacteria using sequence tags from genomic DNA libraries. Genome Biology 3: research0031–research0031.11.

20. Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 247: 536–540.

21. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. (2000) The Protein Data Bank. Nucleic Acids Res 28: 235–242.

22. Xiong JP, Stehle T, Diefenbach B, Zhang R, Dunker R, et al. (2001) Crystal structure of the extracellular segment of integrin αVβ3. Science 294: 339–345.

23. Xiong JP, Stehle T, Goodman SL, Arnaout MA (2004) A novel adaptation of the integrin PSI domain revealed from its crystal structure. J Biol Chem 279: 40252–40254.

24. Xiong JP, Mahalingham B, Alonso JL, Borrelli LA, Rui X, et al. (2009) Crystal structure of the complete integrin alphaVbeta3 ectodomain plus an alpha/beta transmembrane fragment. J Cell Biol 186: 589–600.

25. Xiao T, Takagi J, Coller BS, Wang JH, Springer TA (2004) Structural basis for allostery in integrins and binding to fibrinogen-mimetic therapeutics. Nature 432: 59–67.

26. Springer TA, Zhu J, Xiao T (2008) Structural basis for distinctive recognition of fibrinogen γC peptide by the platelet integrin αIIbβ3. J Cell Biol 182: 791–800.

27. Zhu J, Luo BH, Xiao T, Zhang C, Nishida N, et al. (2008) Structure of a complete integrin ectodomain in a physiologic resting state and activation and deactivation by applied forces. Mol Cell 32: 849–861.

28. Xie C, Zhu J, Chen X, Mi L, Nishida N, et al. (2010) Structure of an integrin with an αI domain, complement receptor type 4. EMBO J 29: 666–679.

29. Adindla S, Inampudi KK, Guruprasad L (2007) Cell surface proteins in archaeal and bacterial genomes comprising "LVIVD", "RIVW" and "LGxL" tandem sequence repeats are predicted to fold as beta-propeller. Int J Biol Macromol 41: 454–468.

30. Quistgaard EM, Thirup SS (2009) Sequence and structural analysis of the Asp-box motif and Asp-box beta-propellers; a widespread propeller-type characteristic of the Vps10 domain family and several glycoside hydrolase families. BMC Struct Biol 9: 46.

31. Corbi AL, Miller LJ, O'Connor K, Larson RS, Springer TA (1987) cDNA cloning and complete primary structure of the α subunit of a leukocyte adhesion glycoprotein, p150,95. EMBO J 6: 4023–4028.

32. Springer TA (1997) Folding of the N-terminal, ligand-binding region of integrin α-subunits into a β-propeller domain. Proc Natl Acad Sci U S A 94: 65–72.

33. Richardson JS (1981) The anatomy and taxonomy of protein structure. Adv Protein Chem 34: 167–339.

34. Hutchinson EG, Thornton JM (1994) A revised set of potentials for β-turn formation in proteins. Protein Sci 3: 2207–2216.

35. Rigden DJ, Galperin MY (2004) The DxDxDG motif for calcium binding: multiple structural contexts and implications for evolution. J Mol Biol 343: 971–984.

36. Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, et al. (2008) The Pfam protein families database. Nucleic Acids Res 36: D281–D288.

37. Rost B, Yachdav G, Liu J (2004) The PredictProtein server. Nucleic Acids Res 32: W321–W326.

38. McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. Bioinformatics 16: 404–405.

39. Ouali M, King RD (2000) Cascaded multiple classifiers for secondary structure prediction. Protein Sci 9: 1162–1176.

40. Emsley J, Knight CG, Farndale RW, Barnes MJ, Liddington RC (2000) Structural basis of collagen recognition by integrin alpha2beta1. Cell 101: 47–56.

41. Norris V, Grant S, Freestone P, Canvin J, Sheikh FN, et al. (1996) Calcium signalling in bacteria. J Bacteriol 178: 3677–3682.

42. Dominguez DC (2004) Calcium signalling in bacteria. Mol Microbiol 54: 291–297.

43. Herbaud ML, Guiseppi A, Denizot F, Haiech J, Kilhoffer MC (1998) Calcium signalling in Bacillus subtilis. Biochim Biophys Acta 1448: 212–226.

44. Tisa LS, Olivera BM, Adler J (1993) Inhibition of Escherichia coli chemotaxis by omega-conotoxin, a calcium ion channel blocker. J Bacteriol 175: 1235–1238.

45. Zhao Y, Shi Y, Zhao W, Huang X, Wang D, et al. (2005) CcbP, a calcium-binding protein from Anabaena sp. PCC 7120, provides evidence that calcium ions regulate heterocyst differentiation. Proc Natl Acad Sci U S A 102: 5744–5748.

46. Cock JM, Sterck L, Rouzé P, Scornet D, Allen AE, et al. (2010) The Ectocarpus genome and the independent evolution of multicellularity in brown algae. Nature 465: 617–621.

47. Lehtonen JV, Still DJ, Rantanen VV, Ekholm J, Björklund D, et al. (2004) BODIL: a molecular modeling environment for structure-function analysis and drug design. J Comput Aided Mol Des 18: 401–419.

48. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: A sequence logo generator. Genome Res 14: 1188–1190.

49. Kraulis PJ (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J Appl Crystallog 24: 946–950.

50. Merritt EA, Bacon DJ (1997) Raster3D: photorealistic molecular graphics. Methods Enzymol 277: 505–524.