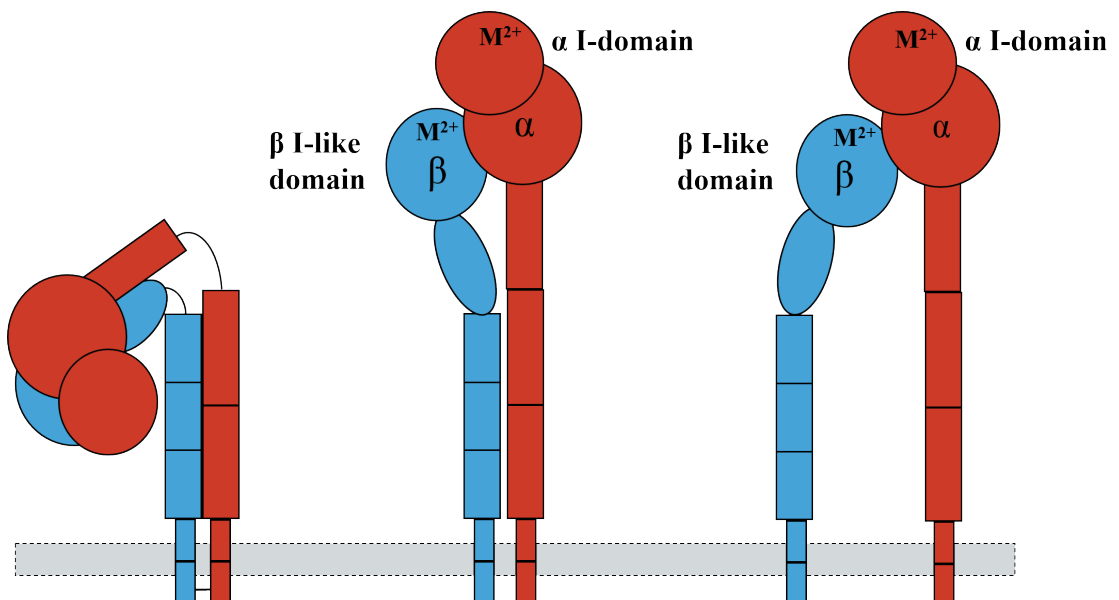


Bhanupratap Singh Chouhan

Integrin evolution: from prokaryotes to the diversification within chordates





Bhanupratap Singh Chouhan

was born on September 04,1985 in Indore, M.P., India. He did his Bachelor of Science in Bioinformatics from Government Autonomous Holkar Science College, Devi Ahilya University in Indore. He then graduated from the University of Turku in 2009 with a Master of Science in Bioinformatics. This Ph.D. thesis work has taken place under the supervision of Professor Mark S. Johnson and Docent Konstantin Denessiouk at the Structural Bioinformatics Laboratory (SBL), Åbo Akademi University, Turku, Finland during 2010-2015.



Integrin evolution: From prokaryotes to the diversification within chordates

Bhanupratap Singh Chouhan

Biochemistry

Faculty of Science and Engineering

Åbo Akademi University

Åbo, Finland

2016

Supervised by:

Professor Mark S. Johnson

Docent Konstantin Denessiouk

Dr. Alexander Denesyuk

Åbo Akademi University

Åbo, Finland

Reviewed by:

Professor Donald Gullberg,

University of Bergen, Norway.

Docent Janne Ravantti,

University of Helsinki, Finland.

Opponent:

Docent Olli Pentikäinen,

University of Jyväskylä, Finland.

ISBN 978-952-12-3448-4

Painosalama Oy – Turku, Finland 2016

Table of contents

1. Review of the literature	1
1.1 Introduction	1
1.2 Integrin bidirectional signaling	6
1.3 Relationship among integrin α and β subunits	12
1.4 N-terminal region: The β -propeller domain	15
1.5 N-terminal region: The vWA domain	17
1.6 Collagens and vertebrate collagen receptors	19
1.7 Structure of collagen-binding and ICAM-binding integrins	21
2. Aims of the study	29
3. Methods	30
3.1 Online databases	30
3.2 Protein sequence analyses	31
3.3 3D modelling and structural analyses	35
4. Results	37
4.1 Conservation of the human integrin-type β -propeller domain in bacteria	37
4.1.1 X-ray structures of α V β 3 and α IIb β 3 and unique structural characteristics	37
4.1.2 β -turns and torsion angles	39
4.1.3 Ca ²⁺ binding motif (DxDxDG)	42
4.1.4 Bacterial sequence dataset	43
4.2 Evolutionary origin of the α C helix in integrins	45
4.3 Early chordate origin of the human-type integrin I-domains	47
4.3.1 Phylogenetic analyses	49
4.3.2 Functional residues shared between human and lamprey α I-domains	51
4.3.3 Sea lamprey α I-domains recognize different mammalian collagen types	55
5. Discussion	57
5.1 Publication I	57
5.1.1 Sequence alignment and phylogenetic tree	57
5.1.2 3D comparative modelling	58
5.1.3 Secondary structure prediction	58
5.1.4 Possible functions	59
5.2 Publication II	59
5.2.1 Genome searches	60
5.2.2 Sequence alignment and secondary structure prediction	60
5.2.3 Importance of the α C helix	61
5.3 Publication III	61
5.3.1 Genome searches	61
5.3.2 Phylogenetic analyses and multivariate plot	62
5.3.3 Structural studies	63
5.3.4 3D comparative modelling	64
6. Conclusions and future directions	65
7. References	67
8. Original publications	85

Original publications

i). Chouhan B, Denesyuk A, Heino J, Johnson MS, Denessiouk K. 2011. Conservation of the human integrin-type β -propeller domain in bacteria. PLoS One 6:e25069.

ii). Chouhan B, Denesyuk A, Heino J, Johnson MS, Denessiouk K. 2012. Evolutionary origin of the α C helix in integrins. WASET. 65:546-549.

iii). Chouhan B, Käpylä J, Denesyuk A, Denessiouk K, Heino J, Johnson MS. 2014. Early chordate origin of the human-type integrin I-domains. PLoS One. 9:e112064.

Contribution of the author

The author has performed multiple sequence alignments, phylogenetic analyses, secondary structure predictions, 3D comparative modelling and structural studies, local genome searches as well as various online database searches during the course of studies that constitute this thesis work. The author also participated in the design, implementation, execution and writing of the publications listed as I-III (original publications) and IV-V (Additional publications).

Additional publications

iv) Johnson MS, Käpylä J, Denessiouk K , Airene TA, Chouhan B, Heino J. 2013. Evolution of cell adhesion to extracellular matrix. In: Keeley W, Mecham RP, editors. Evolution of Extracellular Matrix, Biology of Extracellular Matrix. Springer-Verlag Berlin (Heidelberg). 243-283.

v) Johnson MS, Chouhan BS. Evolution of integrin I domains. 2014. In: Gullberg D, editor. I Domain Integrins (Second Edition). Advances in Experimental Medicine and Biology, Springer (Amsterdam). 819:1-19

Acknowledgements

This thesis work was carried out at the Structural Bioinformatics Laboratory (SBL), Faculty of Science and Engineering, Åbo Akademi University, Finland during 2010-2015. I would like to acknowledge and express my gratitude to all the people who contributed towards the completion of this thesis.

First of all I would like to extend my deepest gratitude to my supervisors, Professor *Mark S. Johnson* and Docent *Konstantin Denessiouk* for granting me the opportunity to conduct my doctoral studies at SBL. Thank you *Mark* and *Konstantin* for your guidance, patience, support and constant encouragement which enabled me to complete this doctoral thesis. I would also like to thank Dr. *Alexander Denesyuk* for his valuable insight, keen supervision and positive feedback. I am very thankful to Professor *Donald Gullberg* and Docent *Janne Ravantti* for taking time out of their busy schedule to read my thesis and providing constructive criticism in order to improve the quality of the text. I am also very thankful to Professor *Jyrki Heino*, Dr. *Jarmo K ypyl * and Dr. *Lari Lehti * for agreeing to be members of my thesis committee and providing valuable ideas regarding this thesis work. I would also like to acknowledge and thank my collaborators/co-authors from Professor *Jyrki Heino*'s laboratory at the University of Turku. I would also like to extend my gratitude to Docent *Olli Pentik inen* for agreeing to be my opponent.

I would also like to thank my seniors and colleagues from SBL: *Lenita Viitanen*, *Mikko Huhtala*, *Mikko Vainio*, *Santeri Puranen*, *Tomi Airenne*, *Outi Salo-Ahen* (Thank you so much Outi for the abstract translations), *Anni Kauko*, *Eva Bligt-Lind n*, *K the Dahlstr m*, *Jukka Lehtonen* and Professor *Tiina Salminen*. Thank you all for the wonderful discussions and coffee room memories. A very special thanks to *Frederik Karlsson* for being very supportive and helpful with all the practical matters throughout my doctoral studies at SBL. I am also very grateful to the secretarial and technical staff at Department of Biochemistry for all their help. I am very thankful to the National Doctoral Programme in Informational and Structural Biology (ISB) for accepting me as a member of this amazing graduate school. Thanks also go out to my colleagues and other members of ISB who made the annual meetings fun and inspirational.

Apart from my wonderful colleagues, I would also like to thank my close friends and partners in crime at SBL (and outside SBL as well): *Nitin 'Marwaadi' Agrawal, Vipin Ranga- 'Dada', Avlokita 'Tikka' Tiwari, Rakesh 'Raakshas' Purushottama, Athresh 'Ornob' Shigaval* and *Parthiban 'Saaree' Marimuthu*. I will always cherish our wonderful time together especially the long weekend lunch sessions and the blunt/crude humour. This Ph.D. journey would not have been as fun without you guys. Thank you also to my band member *Mohit 'MC Elite' Narwal* for the wonderful friendship and entertainment over the years, I will always remember our long 'rap sessions' during the weekend which provided me with a break from the lab work ("Lyrical Prophets"). I owe my warmest gratitude to my dear friends outside my workplace for all the wonderful memories: *Hasan, Aman, Raghu, Yashu, Vinay, Neeraj ji, Tamoghna, Debanga, Rishabh, Ayush, Abhishek, Roshan, Sachin* and the rest of my south-asian friends (Sorry if I missed someone, but you know who you are).

I would like to thank my family for their endless support and constant encouragement which has been a source of great strength. I would especially like to thank my parents for believing in me and letting me chase after my dream. It's been a long journey which could not have been possible without your support. Thank you *Mummy* and *Papa*, I sincerely appreciate all the sacrifices you have made, it means a lot to me. I would also like to thank my loving wife *Neha*, your constant support over the past several months has kept me very focused and inspired me to work harder. A special thanks to all my wonderful cousins and family members especially grandpa (*Nano-sa*) for always being concerned with my well-being and happiness.

I would like to dedicate this thesis to my loved ones who are no longer with me but I always remember them very fondly: my grandma (*nani-sa*), my grandparents (*dado-sa* and *dadi-sa*) and my best friend *Prabhakar Sharma (Prabhu)* who wanted to pursue his own Ph.D. but his untimely demise is something I find difficult to come to terms with, even today so I dedicate this thesis especially to all of you – "you may be gone but you are never over".

The financial support for this thesis work has been absolutely indispensable. Therefore, I wish to acknowledge and thank Åbo Akademi University, ISB, Sigrid Juselius Foundation, Tor, Joe and Pentti Borg Foundation, Magnus Ehrnrooth Foundation and Stiftelsen för Åbo Akademi for the vital financial support.

- Bhanupratap Singh Chouhan,

Åbo, September 2016

Abbreviations

3D	Three-dimensional
β-propeller	Beta Propeller Domain
ADMIDAS	Adjacent to MIDAS site
BLAST	Basic Local Alignment Search Tool
CATH	Class Architecture Topology Homologous superfamily database
CDD	Conserved Domain Database
COGs	Clusters of Orthologous Groups
DDR	Discoidin Domain Receptor
ECM	Extra Cellular Matrix
EDTA	EthyleneDiamineTetraacetic Acid
‘FG-GAP’	Phe-Ala-Gly-Ala-Pro
‘GFOGER’	Gly-Phe-Hyp-Gly-Glu-Arg (Hyp = Hydroxyproline)
‘GLOGEN’	Gly-Leu-Hyp-Gly-Glu-Asn
GOR	Garnier-Osguthorpe-Robson method
GPVI	Glycoprotein VI
I-domain	Inserted domain
I-EGF	Integrin Epidermal growth factor-like domain
JTT	Jones Taylor Thornton method
LAIR	Leukocyte-Associated IG-like Receptor-1
MCMC	Markov Chain Monte Carlo
MIDAS	Metal Ion Dependent Adhesion Site
NCBI	National Center for Biotechnology Information
NMR	Nuclear Magnetic Resonance
PCA	Principle Components Analysis
PDB	Protein Data Bank
PFAM	Protein Families Database
PIR	Protein Information Resources
PRF	Protein Research Foundation
PSI domain	Plexin-Semaphorin-Integrin domain
PSSMs	Position Specific Scoring Matrices
RMSD	Root Mean Square Deviation
RPS-BLAST	Reversed Position Specific BLAST
SCOP	Structural Classification Of Protein database
SMART	Simple Modular Architecture Research Tool
SyMBS	Synergistic Metal Ion Binding Site
ICAM	Inter Cellular Adhesion Molecule
TM	TransMembrane
UniprotKB	Uniprot KnowledgeBase
VCAM	Vascular Cell Adhesion Molecule
VEGF	Vascular Endothelial Growth Factor
vWA	von Willebrand Factor A domain
WAG	Whelan And Goldman method

Abstract

Integrins are a family of large multi-domain cell surface receptors responsible for bidirectional signaling in response to cell-cell, cell-extracellular matrix interactions as well as intracellular interactions. Integrins are implicated in a wide variety of functions such as inflammatory responses, adoptive antigen-specific immunity, tissue remodelling and cell adhesion, proliferation and differentiation. Furthermore, integrins are also known to be associated with a wide variety of diseases and health issues, such as tumour metastasis, immune dysfunction, inflammation, viral infections and osteoporosis – to name a few – making them one of the most complex cell adhesion molecules. Integrins are heterodimers which are composed of an α and a β subunit and humans are known to express 18 α subunits and 8 β subunits which associate non-covalently to form 24 α/β heterodimers out of 144 possible combinations. Orthologues of mammalian integrins are observed throughout vertebrates including the bony fish (osteichthyes), however integrins extracted from early chordates like the tunicates *Ciona intestinalis* or *Halocynthia roretzi* are not direct mammalian orthologues. Even though integrins are observed throughout metazoans, studies have reported that integrins and their signaling machinery are located in unicellular eukaryotes. In addition, the 3D folds of the constituent domains from the integrin α and β subunits have also been detected in bacteria.

The major aims of the research work described in this thesis were to answer the following questions: i) When did the constituent domains from the integrin α and β subunits originate? ii) When did the α I-domains get integrated into the integrin heterodimer and when did the collagen-binding integrin α I-domains originate in the vertebrates? iii) When did the mammalian-type integrin orthologues originate in vertebrates?

In order to address these questions, we analysed the available sequences, genomic data as well as structural data, all of which are discussed in detail in the three studies comprising this thesis. During the course of this thesis: i) We have addressed the origin of pivotal integrin constituent domain like the N-terminal 7-bladed β -propeller domain from the α -subunit by investigating the extent of similarities between the sequences and structures of different integrin domains and similar gene products and protein sequences from bacteria. ii) We have identified characteristic structural features or motifs like the α C helix in order to understand the evolutionary process of collagen-binding integrin α I-domain in vertebrates. iii) Recent advancements in the genome assembly process of organisms like the sea lamprey (agnathostome) and the elephant shark (chondrichthyes) has helped us in understanding the origin and evolution of mammalian-type integrin orthologues. In conclusion, the studies presented in this thesis present novel insights into the evolutionary patterns of the integrins

Key Words: Integrin; molecular evolution; β -propeller; α I-domain; collagen-receptor; α C helix;

Sammanfattning

Integriner är en familj stora multidomän cellytereceptorer ansvariga för dubbelriktad signalering som svar på cell-cell och cell-extracellulär matrix -interaktioner samt intracellulära interaktioner. Integriner är involverade i ett stort antal olika funktioner såsom inflammatoriskt svar, adaptiv antigenspecifik immunitet, vävnadsombildning och celladhesion samt cellproliferation och celldifferentiering. Dessutom är integriner också kända för att vara associerade med en mängd olika sjukdomar och hälsotillstånd, såsom tumörmetastaser, immundysfunktion, inflammation, virusinfektioner och benskörhet, vilket gör dem till en av de mest komplexa celladhesionsmolekylerna. Integriner är heterodimerer sammansatta av en α - och en β -subenhet. Människan är känd för att uttrycka 18 α -subenheter och 8 β -subenheter vilka associerar icke-kovalent bildande 24 α/β -heterodimerer av 144 möjliga kombinationer. Ortologer av däggdjursintegriner observeras hos alla ryggradsdjur inklusive benfiskar (osteichthyes), men integriner utvunna ur tidiga ryggradssträngdjur liksom manteldjur *Ciona intestinalis* eller *Halocynthia roretzi* är inte direkta däggdjursortologer. Även om integriner observeras hos alla flercelliga djur har studier rapporterat att integriner och deras signaleringsmaskineri finns hos encelliga eukaryoter som till exempel *Monosiga brevicollis*. Dessutom har tredimensionella former av de huvuddomänerna från integriners α - och β -subenheter också upptäckts i bakterier.

De viktigaste målsättningarna för forskningsarbetet presenterat i denna avhandling var att besvara följande frågor: i) När uppkom huvuddomänerna i integrinernas α - och β -subenheter? ii) När integrerades α I-domänerna i integrinheterodimeren och när uppkom de kollagenbindande integrin α I-domänerna i ryggradsdjur? iii) När uppkom integrinortologerna av däggdjurstyp hos ryggradsdjur?

För att svara på dessa frågor analyserade vi de tillgängliga sekvenserna, genetisk data samt strukturell data, vilket allt diskuteras i detalj i de tre studier som denna avhandling omfattar. Inom ramen för denna avhandling: i) Har vi studerat ursprunget av en central huvuddomän av integriner, såsom den N-terminala 7-bladiga β -propellerdomänen från α -subenheten, genom att undersöka omfattningen av likheterna mellan sekvenserna och strukturerna av olika integrindomäner och liknande genprodukter och proteinsekvenser hos bakterier. ii) Har vi identifierat de karakteristiska strukturella dragen eller motiven, såsom α C-helix, för att förstå den evolutionära processen av kollagenbindande integrin α I-domänen hos ryggradsdjur. iii) Har nya framsteg i monteringsprocessen av genom hos organismer som havsnejonöga (Agnathostome) och australisk plognos (Chondrichthyes) hjälpt oss att förstå ursprunget och evolutionen av däggdjurstyps integrinortologer. Sammanfattningsvis kan konstateras att de studier som presenteras i denna avhandling presenterar nya insikter i de evolutionära mönstren av integriner.

Nyckelord: integrin; molekylär evolution; β -propeller; α I-domän; kollagenreceptor; α C-helix

1. Review of the literature

1.1 Introduction

Integrins are a large family of bidirectionally signaling, heterodimeric, trans-membrane cell surface receptors that are involved in cell-cell, cell-ECM and even cell-pathogen interactions (Eble and Kühn 1997; Hynes 2002; Legate *et al.*, 2009; Takada *et al.*, 2007). Extracellular integrin domains interact with a wide variety of ligands like collagens, fibronectin and laminins and receptor immunoglobulin fold domains of intracellular adhesion molecules (ICAMs) (Table 1 – Johnson *et al.*, 2009); while the intracellular cytoplasmic domains communicate with the signaling molecules located inside the cell (Luo *et al.*, 2007) and play a pivotal role in the formation of focal adhesions (Legate *et al.*, 2009). These interactions are central to the regulation of cell migration, phagocytosis, cell growth and development (Arnaout *et al.*, 2007). In addition, integrins are also implicated in diseases and health issues like tumor progression (Shin *et al.*, 2012), recognition of pathogens (Ulanova *et al.*, 2009), immune dysfunction (Kishimoto *et al.*, 1987), inflammation (Gahmberg *et al.*, 1998) and osteoporosis (Teitelbaum, 2005).

Integrins are composed of two subunits, α and β subunits (Figure 1). Mammals express at least 18 different integrin α subunits and 8 different β subunits, which are known to form 24 α/β heterodimeric combinations in humans out of 144 possible combinations (Figure 2) (Hynes, 2002; Shimaoka *et al.*, 2003). The reason behind this observed pairing pattern is still not known but it is quite clear from Figure 2 that certain integrin subunits are more promiscuous than others. For example, the subunits $\beta 1$, $\beta 2$ and αV form more heterodimeric associations than rest of the subunits. Interestingly, integrin $\beta 1$ subunit is the most promiscuous subunit as it can form heterodimeric association with 12 integrin α subunits. Table 1 (adapted from the review of Johnson *et al.*, 2009) provides an insight into these receptors and their contributions in the development and sustenance of mammals

The extracellular region consists of ligand binding N-terminal domains followed by the leg or stalk region domains, transmembrane domains and the cytoplasmic domains for both the subunits respectively (Nermut *et al.*, 1988). The first X-ray crystal structure highlighting the ectodomain region of the integrin heterodimer was solved in 2001 by Xiong *et al.* and it clearly showcased the multi-domain structure of the α and β subunit.

Table 1. Mammalian integrins and their functions. Table adopted from 'Integrins during evolution' (Reprinted with permission from Johnson *et al.*, 2009).

Integrin	Ligand	Location of defects in knock-outs and/or main expression sites
$\alpha 7\beta 1$	Laminins	Muscle (Mayer <i>et al.</i> , 1997)
$\alpha 6\beta 4$	Laminins	Skin (hemidesmosomes) (Stepp <i>et al.</i> , 1990)
$\alpha 6\beta 1$	Laminins, ADAMs	Gametes, macrophages, platelets (Georges-Labouesse <i>et al.</i> , 1996)
$\alpha 3\beta 1$	Laminins, (Collagen)	Skin, kidney, lung, cortex (Kreidberg <i>et al.</i> , 1996; DiPersio <i>et al.</i> , 1997)
$\alpha 11\beta 3$	Fibrinogen ('RGD', 'GAKQAGDV'), Fibronectin, vitronectin ('RGD')	Platelets (Pytela <i>et al.</i> , 1986)
$\alpha V\beta 8$	Vitronectin ('RGD')	Vascular development (Müller <i>et al.</i> , 1997; Littlewood and Müller, 2000)
$\alpha V\beta 6$	Fibronectin, TGF- β -LAP ('RGD')	Skin, lung (collagen accumulation) (Munger <i>et al.</i> , 1999)
$\alpha V\beta 5$	Vitronectin ('RGD')	Eye (retinal phagocytosis), bone (osteoclastogenesis) (Nandrot <i>et al.</i> , 2004; Lane <i>et al.</i> , 2005)
$\alpha V\beta 3$	Fibrinogen, fibronectin, vitronectin, tenascin C, osteopontin, bone sialoprotein ('RGD'), MMP-2	Bone (osteoclasts) (Hodivala-Dilke <i>et al.</i> , 1999; McHugh <i>et al.</i> , 2000)
$\alpha V\beta 1$	Fibronectin, vitronectin ('RGD')	In vivo role in tissue fibrosis (Bodary <i>et al.</i> , 1990; Vogel <i>et al.</i> , 1990; Reed <i>et al.</i> , 2015)
$\alpha 8\beta 1$	Fibronectin, vitronectin, tenascin C, osteopontin, nefronectin ('RGD')	Kidney, inner ear (Zhu <i>et al.</i> , 2001)
$\alpha 5\beta 1$	Fibronectin ('RGD')	Embryonic development (blood vessels) (Pytela <i>et al.</i> , 1985; Yang <i>et al.</i> , 1993)
$\alpha 9\beta 1$	Tenascin C, osteopontin, ADAMs, factor XIII, VCAM, VEGF-C, VEGF-D	Lymphangiogenesis (Huang <i>et al.</i> , 2000; Vlahakis <i>et al.</i> , 2005)
$\alpha 4\beta 7$	Fibronectin, VCAM, MadCaM	Peyer's patch (Yang <i>et al.</i> , 1995)
$\alpha 4\beta 1$	Fibronectin, VCAM	Embryonic development (Yang <i>et al.</i> , 1995)
$\alpha 11\beta 1$	Collagens	Periodontal ligament (Popova <i>et al.</i> , 2007)
$\alpha 10\beta 1$	Collagens	Cartilage (Bengtsson <i>et al.</i> , 2005)
$\alpha 2\beta 1$	Collagens, tenascin C, (laminins)	Platelets, epithelium, mast-cells (mesenchymal tissues) (Zutter <i>et al.</i> , 1990; Chen <i>et al.</i> , 2002; Holtkötter <i>et al.</i> , 2002; Senger <i>et al.</i> , 2002; Edelson <i>et al.</i> , 2004; Auger <i>et al.</i> , 2005; Zhang <i>et al.</i> , 2008)
$\alpha 1\beta 1$	Collagens, semaphorin 7A, (laminins)	Mesenchymal tissues (Pozzi <i>et al.</i> , 1998; Gardner <i>et al.</i> , 1999; Ekholm <i>et al.</i> , 2002; Conrad <i>et al.</i> , 2007)
$\alpha D\beta 2$	ICAM, VCAM	Eosinophils (Grayson <i>et al.</i> , 1999; Vieren <i>et al.</i> , 1999)
$\alpha M\beta 2$	ICAM, VCAM, iC3b, factor X, fibrinogen	Leukocytes (phagocytosis) (Altieri <i>et al.</i> , 1990; Diamond <i>et al.</i> , 1990; Elemer and Edgington 1994; Barthel <i>et al.</i> , 2006;)
$\alpha L\beta 2$	ICAM	Leukocytes (recruitment) (Kolanus <i>et al.</i> , 1996)
$\alpha X\beta 2$	Fibrinogen, plasminogen, heparin, iC3b	Leukocytes (Mickletham and Sim, 1985; Loike <i>et al.</i> , 1991; Davis, 1992; Diamond <i>et al.</i> , 1993)
$\alpha E\beta 7$	E-cadherin	Skin, Gut (immune system) (Higgins <i>et al.</i> , 1998)

Figure 1 illustrates the schematic architecture of the non-covalently associated integrin heterodimer. Integrin α subunits (1000-1190 residues) are generally larger than the β subunits (770-800 residues) but there is an exception i.e. integrin $\beta 4$, as it extends to nearly 1800 residues with the additional ~ 1000 residues located within the intracellular C-terminal region. The integrin β subunit ectodomain region consists of eight domains, which includes the β I-like domain, hybrid domain, PSI domain (Plexin, Semaphorin, Integrin), I-EGF modules 1-4 and the β -tail domain. These domains are followed by the transmembrane domain (TM) and the cytoplasmic tail region. The β I-like domain (together with the β -propeller of the α subunit binds extracellular integrin ligands in the absence of the α I-domain) adopts a Rossmann fold and it is inserted in the hybrid domain. The PSI domain is split into two regions (Xiao *et al.*, 2004; Xiong *et al.*, 2004) that are known to be connected by a disulphide bond (Cys13 to Cys435 in integrin $\beta 3$ subunit). The integrin-type epidermal growth factor modules i.e. I-EGF 1-4 domains are cysteine rich regions and each I-EGF domain contains eight cysteines that are bonded together in C1-C5, C2-C4, C3-C6, C7-C8 pattern with the exception of I-EGF1 that lacks the C2-C4 disulphide bond (Beglova *et al.*, 2002; Takagi *et al.*, 2002; Zhu *et al.*, 2008). The β -tail domain consists of an $\alpha+\beta$ fold and it is proposed to play an important role in integrin activation (Arnaout *et al.*, 2005). The interaction between the integrin α subunit and β subunit TM helices results in the resting state of the heterodimer (Adair and Yeager, 2002; Luo *et al.*, 2004; Partridge *et al.*, 2005; Wegener and Campbell, 2008). The cytoplasmic domains (of both α and β subunits) can form interactions with multiple proteins and these domains play a significant role in keeping the dimer in a resting state as well as inside-out activation of integrins (Wegener and Campbell, 2008; Legate *et al.*, 2009).

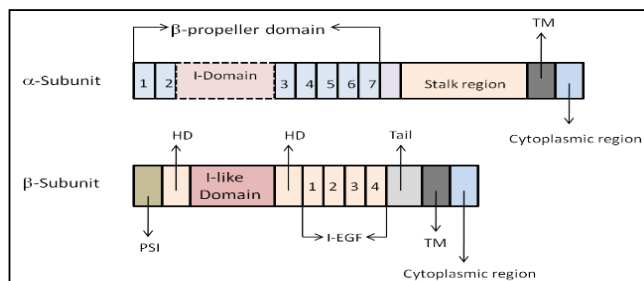


Figure 1. Schematic architecture of a typical integrin α/β heterodimer. In this figure ‘TM’ represents the transmembrane domain, ‘PSI’ represents plexin-semaphorin-integrin domain, ‘HD’ represents the hybrid domain and I-EGF represents Immunoglobulin like folds.

The integrin α subunit consists of four or five extracellular domains and these include the seven bladed β -propeller domain (discussed in detail later in this thesis), which may host an *inserted* I-domain between its second and third repeat, followed by the thigh, calf-1 and calf-2 domains (that constitute the stalk region), a transmembrane domain and a cytoplasmic tail region. The thigh and calf domains adopt a similar structure as they both share a β -sandwich fold (Xiong *et al.*, 2001).

The human integrin α subunits can be subdivided into two major groups: those that contain an additional inserted α I-domain ($\alpha 1$, $\alpha 2$, $\alpha 10$, $\alpha 11$, αD , αX , αL , αM and αE) and those without α I-domain, Figure 2 ($\alpha 4$, $\alpha 9$, $\alpha 6$, $\alpha 7$, $\alpha 3$, αV , $\alpha 5$, $\alpha 8$ and αIIb) (Larson *et al.*, 1989). The α I-domain is about 200 residues long and it is a homolog of the β I-like domain present in all β subunits that adopts a Rossmann fold, which basically consists of six parallel β -strands surrounded by two pairs of α -helices and the α I-domain buds out from a loop located between the second and third repeat of the β -propeller domain, which in turn is located in the *head* region of the integrin α subunit. I-domains have a highly-exposed binding site and can thus recognize bulkier integrin ligands like collagens and ICAMs (Intercellular Adhesion Molecule) through binding of the acidic residue glutamate to its divalent cation (Mg^{2+}) at the binding site called ‘MIDAS’ (Metal Ion Dependent Adhesion Site) (Lee *et al.*, 1995). MIDAS is a characteristic of all the integrin α I-domains and it (MIDAS) is observed all the way back even in the urochordates (Miyazawa *et al.*, 2001; Ewan *et al.*, 2005; Huhtala *et al.*, 2005). MIDAS is also present in all β I-like domains and the site for binding an aspartate of an exposed loop from the protein ligand.

Functionally speaking, integrins with the α I-domains segregate integrins two classes on the basis of their ligand specificity; in the case of the human integrins $\alpha 1\beta 1$, $\alpha 2\beta 1$, $\alpha 10\beta 1$ and $\alpha 11\beta 1$, they are known to bind collagen and contribute to the structural integrity of cells and tissues (Knight *et al.*, 1998, 2000; Zhang *et al.*, 2003). The human integrins $\alpha X\beta 2$, $\alpha D\beta 2$, $\alpha M\beta 2$, $\alpha L\beta 2$ and $\alpha E\beta 7$ are known to play a key role in the interaction of leukocytes with endothelial cells and other matrix structures (Van der Vieren *et al.*, 1996, 1999; Grayson *et al.*, 1999; Garnotel *et al.*, 2000; Noti *et al.*, 2000; Hynes, 2002; Solovjov *et al.*, 2005; Takada *et al.*, 2007). Considerable insight into the structural basis for the function of individual integrin α I-domains has been provided in the variety of structural studies from the past 20 years (Integrin alpha 1 PDB IDs: 1QC5: Rich *et al.*, 1999; 1QCY: Kankare *et al.*, 2003 (Deposition author), Nymalm *et al.*, 2003; 1PT6: Nymalm *et al.*, 2004, Integrin alpha 2 PDB IDs: 1AOX: Emsley *et al.*, 1997, 1DZI: Emsley *et al.*, 2000, 1V7P: Horii *et al.*, 2004; Integrin alpha L PDB IDs: 1LFA: Qu and Leahy, 1996, 1DGQ: Legge *et al.*, 2000, 1CQP:

Kallen et al., 2000, 1MQ8; Shimaoka et al., 2003, 1TOP; Song et al., 2005 3BN3; Zhang et al., 2008; Integrin alpha X PDB ID: Vorup-Jensen et al., 2003; Integrin alpha M PDB IDs: Lee et al., 1995; Xiong et al., 2002 McCleverty and Liddington, 2003). In comparison to integrins without α I-domains where the ligand must have an exposed flexible loop, the insertion of the α I-domain in some integrin α subunits has led to a dramatic shift in the ligand recognition site thereby providing unprecedented access to heavier ligands like collagen fibres bundled into large macroscopic structures and the immunoglobulin-fold ICAM domains.

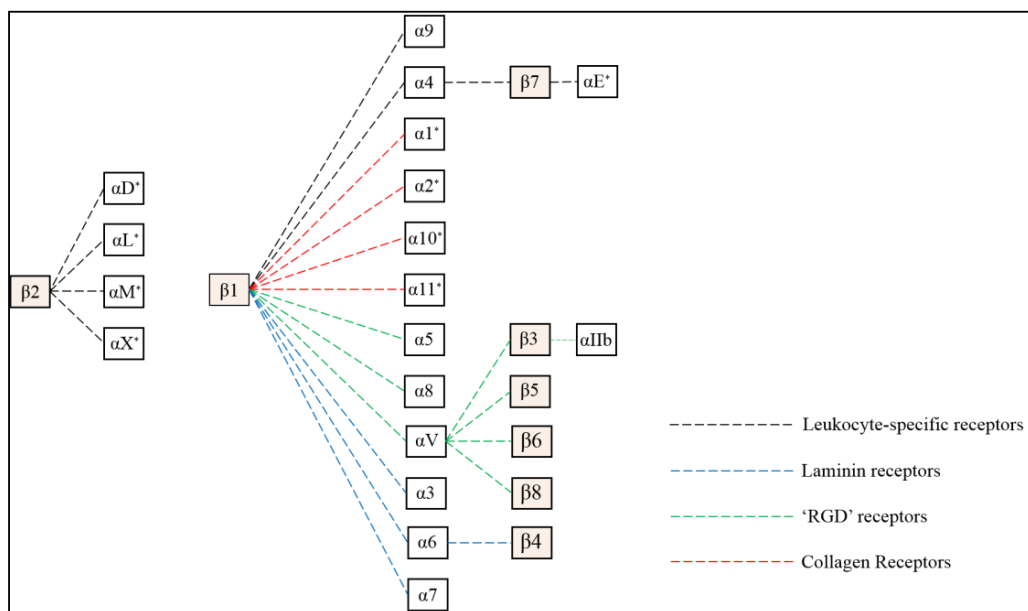


Figure 2. Heterodimeric association pattern of integrin α and β subunits; α subunits marked with an ‘*’ contain the additional I-domain.

All of the earliest diverging integrins do not contain the α I-domain. Consequently, integrins *without* the α I-domain recognize ligands like laminin and fibronectin in a different way, at a narrow cleft at the junction of β -propeller domain (of the integrin α subunit) and the β I-like domain (of the integrin β subunit) (Xiong *et al.*, 2004). The β I-like domain (also known as the β A-domain) and its MIDAS is located in the *head* portion of the integrin β subunit and, in the absence of the α I-domain, it plays a central role in ligand recognition, e.g. of the arginine-glycine-aspartate sequence on the loops of ligands. Extracellular ligand recognition via both of these mechanisms, with or without the α I-domain, results in signal transduction, but integrins are known to be activated via internal cytoplasmic interactions as well.

1.2 Integrin bidirectional signaling

The early structural view of integrins was based primarily on electron microscopy data (Nermut *et al.*, 1988) as well as the analysis of integrin sequences (Nermut *et al.*, 1988; Arnaout, 1990) but implementation of techniques like X-ray crystallography and NMR spectroscopy has led to the solution of multiple three-dimensional integrin structures ranging from the first structures of α I-domains, and later headpieces and ectodomains, the transmembrane (TM) helices and cytoplasmic tail regions, thereby providing a more comprehensive understanding of the structure and function of the integrin heterodimer.

At present, the most widely accepted integrin activation mechanism consists of three distinct conformations as shown in Figure 3 (Arnaout *et al.*, 2005 and Luo *et al.*, 2007). A resting state (or bent state) where the integrin heterodimer exists in an inverted bent V-shape with respect to the location of the membrane, a low-affinity intermediate state (or extended closed state) and an activation state (or extended open state) which is associated with an extended conformation wherein the hybrid domain has swung outward in contrast to the initial resting position and the cytoplasmic tail regions from the α and β subunits do not interact anymore. Several studies have already focused on integrin structural characterisations and activation mechanisms (Vinogradova *et al.*, 2000, 2002, 2004; Li *et al.*, 2001, 2005; Lu *et al.*, 2001; Ulmer *et al.*, 2001; Xiong *et al.*, 2001, 2002, 2004; Beglova *et al.*, 2002; Takagi *et al.*, 2002, 2003; Weljie *et al.*, 2002; Kim *et al.*, 2003; Luo *et al.*, 2003, 2004; Litvinov *et al.*, 2004; Tng *et al.*, 2004; Xiao *et al.*, 2004; Iwasaki *et al.*, 2005; Mould *et al.*, 2005; Shi *et al.*, 2005, 2007; Nishida *et al.*, 2006) but the bent structure is generally seen in structures of the ectodomains. The current integrin activation model clearly suggests that conversion from a low affinity resting state to a high affinity extended state is pivotal for ligand binding and signalling, certain studies have indicated that integrins can bind ligands in a bent state as well (Calzada *et al.*, 2002; Xiong *et al.*, 2002; Chigaev *et al.*, 2003, 2007; Adair *et al.*, 2005; Askari *et al.*, 2010).

Integrins can be activated from the outside as well as from the inside of the cell, which results in a large and reversible conformational change in the heterodimer. When the activation takes place due to interaction between the integrin ectodomain and extracellular ligands (like collagen, ICAM, fibronectin etc.) it is known as 'outside-in' signaling, whereas integrin activation taking place due to signals from within the cell are initiated by interactions between cytosolic proteins like talin and kindlin and the integrin β subunit

cytoplasmic ‘tail’. This is known as ‘inside-out’ signaling. Herein is presented a very brief look at integrin bidirectional signaling.

a) Inside-out signaling: Talin and kindlins bind to the cytoplasmic tail of the integrin β -subunit, which in turn contributes to integrin activation (Tadokoro *et al.*, 2003; Legate and Fässler, 2009; Moser *et al.*, 2009). The tail region of the integrin β -subunit consists of two ‘NPxY’ motifs (where N is asparagine; P is proline; x is any residue and Y is tyrosine) and it has been shown that talin interacts with the membrane proximal ‘NPxY’ motif, while kindlins are known to interact with the membrane distal ‘NPxY’ motif (Calderwood *et al.*, 2002; Calderwood 2004; Banno and Ginsberg 2008; Moser *et al.*, 2009; Calderwood *et al.*, 2013). Talin participates in this interaction through its F3 PTB domain (phospho-tyrosine binding) and this interaction facilitates competitive binding between a lysine from Talin and an arginine from the integrin α -subunit, resulting in the disassociation of a salt bridge linking the α and β subunits, which activates the integrin (Anthis *et al.*, 2009; Lau *et al.*, 2009; Zhu *et al.*, 2009). Integrin activation results from the separation of the cytoplasmic domains that leads to the extension of the extracellular domains causing a switchblade-like extension at the ‘genu’ domain region (Beglova *et al.*, 2002; Takagi *et al.*, 2002)

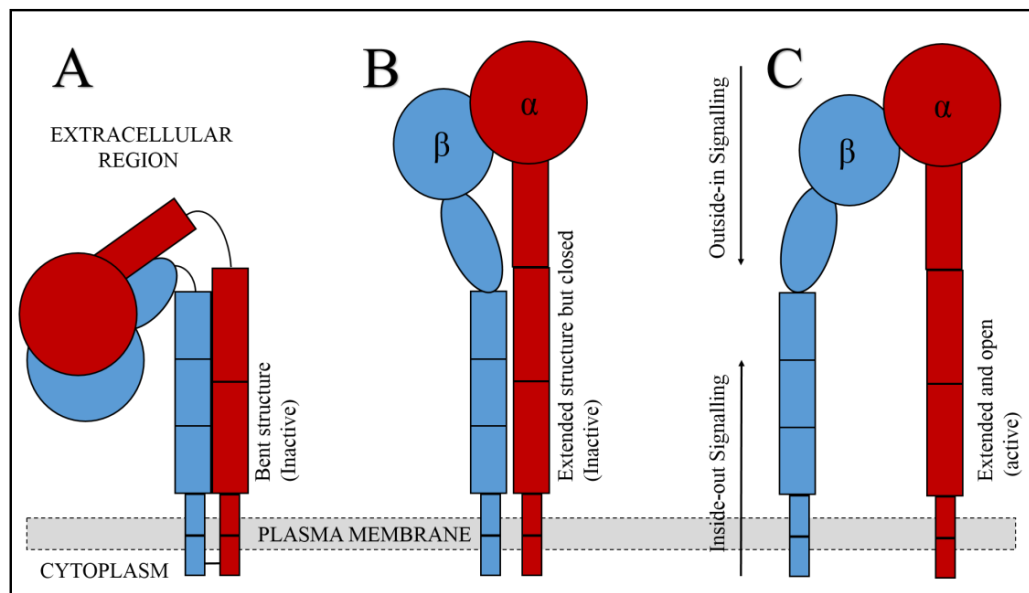


Figure 3. The integrin heterodimer can exist in three conformational states i.e. A) bent conformation where the two subunits are inactive and they are held together by a tight salt bridge; B) extended intermediate closed conformation and C) extended open conformation facilitating bidirectional signaling. The cover illustration for this thesis work includes the pivotal α I-domain.

Kindlins interact with the membrane distal 'NPxY' motif of integrin β -tail region through the FERM (Four-point-one, Ezrin, Radixin, Moesin) domain. Even though kindlins are not directly responsible for integrin activation they do play a pivotal role as co-activators (in partnership with talin). It has been suggested that inhibiting the binding of kindlin and integrins can be disruptive towards the talin-mediated activation of integrins, thereby exhibiting the importance of integrin-kindlin association for proper integrin activation (Montanez *et al.*, 2008; Harburger and Calderwood, 2009; Federico *et al.*, 2013). Furthermore, it has been shown that excessive expression of kindlins (in cell culture systems) does not lead to integrin activation but when kindlins are co-expressed with the talin head region then kindlins can help trigger the activation of integrin α IIb β 3 (Montanez *et al.*, 2008; Moser *et al.*, 2008;). Structural data from an NMR study has further highlighted the fact that kindlins function as integrin co-activators (Bledzka *et al.*, 2012). When it comes to inside-out signaling then the talin mediated integrin activation mechanism has been well studied at the molecular level but the details of the co-operation mechanism between kindlin and talin during the integrin activation process still remains to be elucidated. In conclusion here, it can be said that signals received by other receptors result in the binding of the talin and kindlin to the cytoplasmic tail region of integrins (Watanabe *et al.*, 2008).

b) Outside-in signaling: occurs when integrins bind extracellular ligands and the resulting signal transduction takes place across the bilayer membrane towards the interior of the cell leading to the gene expression, cytoskeletal organization and modulation of the cell cycle (Yamada and Geiger, 1997). This occurs through several signaling pathways associated with outside-in integrin signaling (Ridley *et al.*, 2003; Grashoff *et al.*, 2004; Guo and Giancotti, 2004; Shattil and Newman, 2004). The α I-domain (when present), the β I-like domain and the β -propeller domain play a pivotal role in the recognition and binding of extracellular ligands. The α I-domain and the β I-domain share a few common features, they are both structurally homologous as they both adopt a Rossmann fold (a common feature of all vWA domains), which consists of central β -sheets surrounded by α -helices. Both of these domains bind ligands via direct interaction with a divalent metal ion bound at MIDAS; the β I-like domain also contains two additional binding sites known as SyMBS (Synergistic metal ion-binding site) and ADMIDAS (Adjacent to metal ion-dependent adhesion site), both of which are known to bind a Ca^{2+} cation. In addition, the β -propeller domain plays a pivotal role in integrin function as it participates in ligand recognition either directly (in association with the β I-like domain) or indirectly (through the α I-domain). As the name suggests, the 7-bladed integrin β -propeller domain consist of seven repeats of \sim 60 residues that are arranged

toroidally or radially around a central pore to form seven 'blades' resembling the shape of a propeller wherein each blade is comprised of a four-stand antiparallel β -sheet.

In the case of the *non* I-domain containing integrins, the β I-like domain plays a pivotal role in ligand recognition in conjunction with the β -propeller domain as seen in the X-ray crystal structures of the extracellular region of $\alpha V\beta 3$ in complex with an 'RGD' ligand (PDB ID: 1L5G; Xiong *et al.*, 2002) and $\alpha IIb\beta 3$ headpiece bound to fibrinogen gamma peptide 'LGGAKQRGDV' (PDB ID 2VDO, and 2VDR; Springer *et al.*, 2008). (Figure 5). In the case of the $\alpha V\beta 3$ structure, one of carboxylate oxygen atoms from the aspartate of the 'RGD' tripeptide ligand directly binds to the metal ion, in this case Mn^{2+} at MIDAS and the second carboxylate oxygen hydrogen bonds with the mainchain amide NH from Tyr122 and Asn215. While the guanidinium group of arginine forms two salt bridges, one with Asp218 and the other with Asp150 (both from the β -propeller domain), the glycine residue resides at the interface between the α/β subunits, adds conformational flexibility to the loop. This particular ligand recognition mechanism is relatively strict in comparison to the α I-domain recognition system since the recognition site is positioned at the interface of the α subunit (β -propeller) and β subunit (β I-like domain). One of the shortcomings of this recognition mechanism is the inability to recognize and accommodate bulkier ligands and the recognition sequences e.g. RGD present on ligands need to be on loop regions that can insert into the narrow cleft between the β I-like domain and the β -propeller domain (Xiong *et al.*, 2002). This mechanism is probably further stabilized by the insertion of a positively-charged residue, an arginine, from the β I-domain into the central cavity region of the β -propeller domain. Arg261 from integrin $\beta 3$ is inserted at the core of the β -propeller cavity (from αV) and is surrounded by aromatic side chains e.g. Phe21, Phe159, Tyr224, Phe278, and Tyr406 from the propeller ring residues. Additionally, residues located in the upper ring – Tyr18, Trp93, Tyr221, Tyr275, and Ser403 – interact with residues from the lower ring in order to create a hydrophobic pocket for residues adjacent to Arg261 in the 3_{10} -helix (Xiong *et al.*, 2001).

In the case of the I-domain containing integrin α subunits, the divalent metal ion is located at MIDAS and is coordinated by other residues of the α I-domain and by water molecules. Ligands such as collagens and ICAMs coordinate to the metal cation via a negatively charged residue, glutamate, by displacing a water molecule and the remaining coordination positions are taken up by the highly conserved MIDAS residues such as serine and threonine residues. This complex involving a negatively-charged glutamate of the ligand with the metal ion at MIDAS is observed in the experimental 3D structures of the $\alpha 2$ I-domain in complex with the collagen-like triple-helical ‘GFOGER’ peptide (PDB ID: 1DZI; Emsley *et al.*, 2000), the $\alpha 1$ I-domain in complex with triple-helical ‘GLOGEN’ peptide (PDB ID: 2M32, Chin *et al.*, 2013) and ICAM3 in complex with the αL I-domain (PDB ID: 1T0P; Song *et al.*, 2005). It is proposed that on ligand binding (e.g. to collagens, ICAMs) that the I-domain undergoes a dramatic conformational shift causing the intrinsic ligand, a glutamate

	MIDAS residues	α C Helix	Intrinsic ligand
$\alpha 2$	DVVVVCDENSN SIYP...LTNTF...VTDGESH...ILRFGIAVLG YLNRNALD TKNLIKE...FSIEGT		
$\alpha 1$	DIVIVLDGSGNSIYP...QTMTA...VTDGESH...IQRFSIALLGSYNRGNLSTKEKFVEE...FALEAT		
$\alpha 10$	DVVIVLDGSGNSIYP...ETKTA...VTDGESH...VTRYGIAVLGHYLRQRDPSSFLRE...FGLEGS		
$\alpha 11$	DIVIVLDGSGNSIYP...ETRTA...ITD GESH...VTRYAVAVLGYNRRGINPETFLNE...FSLEGT		
αD	DIVFLIDGSGSIDQ...LTF TA...ITDGQKK...IIRYAIGVG-----HAFQGP TARQE...YAVEGT		
αX	DIVFLIDGSGS ISS...FTYTA...ITDGKKE...IIRYAIGVG-----LAFQNRNSWKE...FAIEGT		
αM	DIAFLIDGSGSIIP...RHTTA...ITDGEKF...VIRYVIGVG-----DAFRSEKSRQE...FAIEGT		
αL	DLVFLFDGSM S LQP...LTNTF...ITDGEAT...IIRYIIGIG-----KHFQTKESQET...YVIEAT		
αE	EIAIILDGSGSIDP...VTKTA...LTDGGIE...VERFAIGVG-----EEFKSARTARE...ISMEGT		
Hra1	DVLFLVDGSGSVGK...TTYTG...LTDGQAK...IATFAVGVG-----EYDISE...FVLEGG		
Cia1	DLIFLIDESTSVLE...GTATG...LTDGKSQ...IVMFAIGVG-----KVMGE...ASLESQ		
Cia2	DMLFVLDGSGSVGK...TTYTA...LTDGLST...ITTFAVGVG-----EANEKE...FVLEGA		
Cia3	DLVYVVDSSNSISD...NTFTS...ITDGKAN...ITVYAIGVA-----LKSDAE...SSGEGQ		
Cia4	DI I I I L D G S T S V F P...QTFIH...ITDGEAT...IILTAVGIG-----SSVNE...-----		
Cia5	DIIFVVDSESGTVNR...GTYIG...LTDGRAD...IVTVSVGVG-----DKINE...VKLEGA		
Cia6	DIIFVVDSESGSVDV...LTYIG...LTDGAAT...IVTVSVGVG-----SRVDE...VKLEGD		
Cia7	DIMFVLDDSSSVDD...GTYIS...LTDGGAS...IVLVS VGVG-----TSVNN...LTARTN		
Cia8	DIIFVVDSESGSVD T...LTYIG...LTDGRAT...IVTVSVGVG-----SGIIE...VKLEGG		

Figure 4. A sequence alignment of the integrin α I-domain region highlighting the ligand-binding divalent metal site MIDAS, the signature α C helix of the collagen receptors and the proposed intrinsic ligand (‘E336’ in the $\alpha 2$ I-domain) present in integrins containing the α I-domain. The human sequences are $\alpha 1$, $\alpha 2$, $\alpha 10$, $\alpha 11$, αD , αX , αL , αM and αE ; while the ascidian sequences are denoted as *Halocynthia Hra1* and *Ciona Cia1-8*.

(‘E336’ in the $\alpha 2$ I-domain) to bind to MIDAS at the β I-domain thereby resulting in outside-in signal transduction (Alonso *et al.*, 2002; Yang *et al.*, 2004; Jokinen *et al.*, 2010; Xie *et al.*, 2010). Some of the features described here are shown in Figure 4; the α I-domain based ligand recognition mechanism is discussed in detail later under ‘Structure of collagen-binding and ICAM-binding integrins’. It is also worth mentioning that there are ligands that

do not adhere to MIDAS but bind in a metal-independent manner e.g. the cholesterol lowering drug lovastatin binding to α L I-domain (Kallen *et al.*, 1999), a snake venom metalloproteinase (Ivaska *et al.*, 1999; Pentikäinen *et al.*, 1999) and echovirus 1 (Bergelson *et al.*, 1994; King *et al.*, 1997; Xing *et al.*, 2004; Jokinen *et al.*, 2010) and human recombinant collagen IX to the $\text{C}\alpha$ 1 I-domain of *C. intestinalis* (Tulla *et al.*, 2007).

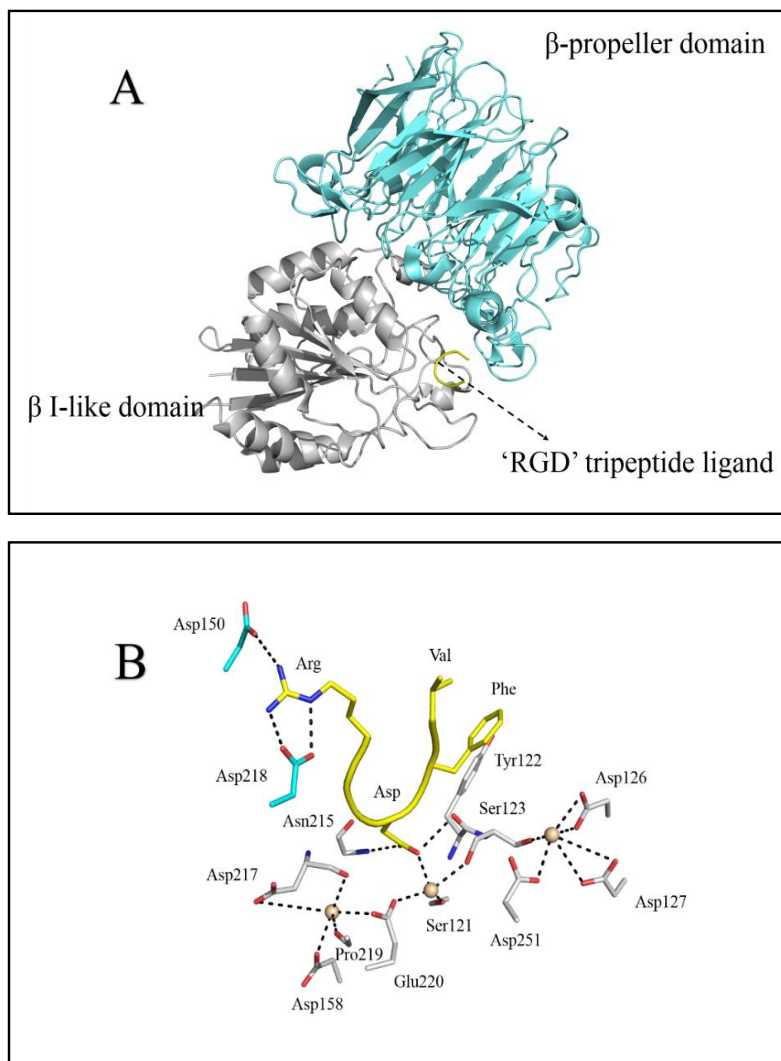


Figure 5. A) Ligand binding at the junction of β -propeller domain (of the integrin α subunit in cyan) and the β I-like domain (of the integrin β subunit in grey). The 'RGD' tri-peptide has been highlighted in yellow (Integrin α V β 3, PDB ID: 1L5G, Xiong *et al.*, 2002). B) Interaction between the ligand 'RGD' (yellow) with integrin residues (β -propeller domain residues in cyan and β I-like domain residues in grey). Here Mn²⁺ ions are highlighted as spheres coordinating residues of MIDAS, ADMIDAS and SyMBS residues.

1.3 Relationship among integrin α and β subunits

Even though several multicellular organisms like fungi and plants do not express integrins (Whittaker and Hynes, 2002; Nichols *et al.*, 2006), they (integrins) probably played a pivotal role in the development and diversification of multicellular animals, the metazoans. The integrin heterodimer has an early origin that most probably predates the first appearance of metazoans (Sebé-Pedrós *et al.*, 2010), being present in a single-cell eukaryote and with most of the constituent domains identifiable in bacterial sequences (Ponting *et al.*, 1999; Johnson *et al.*, 2009). Additionally, the integrin subunits have been detected across the invertebrates (Burke, 1999) and several earliest-diverging animals (Brower *et al.*, 1997; Müller, 1997; Pancer *et al.*, 1997; Reber-Muller, 2001; Nichols *et al.*, 2006; Srivastava *et al.*, 2008; Schierwater *et al.*, 2009; Knack *et al.*, 2008; Srivastava *et al.*, 2010).

Phylogenetic relationships among different integrin subunits have been extensively studied and represented over the past two decades (DeSimone and Hynes, 1988; Hughes 1992; Fleming *et al.*, 1993; Burke 1999; Hynes and Zhao, 2000; Hughes 2001; Johnson and Tuckwell 2003; Ewan *et al.* 2005; Huhtala *et al.* 2005; Takada *et al.*, 2007; Johnson *et al.*, 2009) and they are mostly in agreement with each other. Figures 6 and 7 represent a schematic representation of the phylogenetic distribution among the integrin α and β subunits respectively.

In Figure 6, The integrin α subunits are segregated into two major groups: one group contains the additional I-domain (+ I-domain) while the other group does not (- I-domain). The non I-domain containing group is further segregated based on the *Drosophila melanogaster* integrin α subunits (Hynes and Zhao, 2000; Hughes 2001): the 'PS1' clade (laminin receptor clade), 'PS2' clade (fibronectin receptor clade), 'PS3' clade (invertebrate specific clade) and $\alpha 4/\alpha 9$ clade. This classification scheme is indicative of the 'Position Specific' antigens from *D. melanogaster* that were utilized to define several of the integrin α sequence clusters (Adams *et al.*, 2000). The PS1 clade primarily consists of laminin receptor integrins like $\alpha 3$, $\alpha 6$ and $\alpha 7$ from vertebrates; as well as $\alpha 9$ and $\alpha 10$ from *C. Intestinalis* (Ewan *et al.*, 2005); *ina-1* from *C. elegans* and $\alpha PS1$ from *D. melanogaster* (this is where each 'PS' group gets its naming from). The PS2 clade includes the fibronectin receptor integrins like αIIb , αV , $\alpha 5$ and $\alpha 8$ from vertebrates and $\alpha 11$ from *C. intestinalis*, $\alpha 2$ from *H. roretzi*, *pat-2* from *C. elegans* and $\alpha PS2$ from *D. melanogaster*. The PS3 clade is a special group as it consists of invertebrate sequences exclusively and the integrins $\alpha PS3$, $\alpha PS4$ and $\alpha PS5$ from *D. melanogaster* are known to cluster within this group. The $\alpha 4/\alpha 9$ clade, as the

name suggests is comprised of the $\alpha 4$ and $\alpha 9$ integrins from vertebrates. One common feature shared by the non I-domain containing integrin α -subunits has been the consistency in overall topology of the integrin heterodimer. The subsequent insertion of the α I-domain in some integrins has led to further diversification and specialization of α subunits to cope with the evolution of the chordates. Since vertebrates are characterised by the complexity of the body plan, which includes a closed pressure circulatory system, central nervous system, immune system, cartilage and skeletal system that lends support to higher organisms, the α I-domains were probably a much needed invention in order to accommodate these functions. The I-domain containing integrin α subunits segregate into three distinct clades: collagen-receptor clade, the leukocyte receptor clade and ascidian clade. The collagen receptor clade consists of four subunits: $\alpha 1$, $\alpha 2$, $\alpha 10$ and $\alpha 11$; while the leukocyte surface clade consists of five subunits: αD , αX , αL , αM and αE .

The ascidian (or urochordate) clade is a monophyletic group and it consists of integrin sequences from *C. intestinalis* (Ci $\alpha 1$ -Ci $\alpha 8$) and *H. roretzi* (Hr $\alpha 1$) (Sasakura *et al.*, 2003; Huhtala *et al.*, 2005; Ewan *et al.*, 2005). Although, not much can be said about their functional role at this point of time there have been a few pivotal studies that have certainly attempted to provide an insight. For instance, the Hr $\alpha 1$ sequence from *H. roretzi* is known to adopt a function that is analogous to the vertebrate complement receptor 3 (CR3, Miyazawa *et al.*, 2001). While, Ci $\alpha 1$ I-domain does not bind to collagens I-IV or the 'GFOGER' tripeptide it has been shown to bind strongly to collagen IX, but in a metal-ion/MIDAS independent manner (Tulla *et al.*, 2007). Additionally, it is worth mentioning that several species diverging close to the urochordates are now physically extinct (Donoghue and Purnell, 2005), thereby making it extremely hard to pinpoint the true origin of mammalian integrin orthologues. This problem has been further compounded by a lack of genomic data from the extant species thereby creating a knowledge gap within the evolutionary framework of integrins. Additionally, the α I-domains have not been detected in the lancelet genome (Heino *et al.*, 2008; Putnam *et al.*, 2008). Since genome-wide searches have already revealed that integrins with α I-domains are not observed in the lancelet, or in earlier-diverging invertebrate like the echinoderms (Heino *et al.*, 2008; Johnson *et al.*, 2009), this implies that the insertion of the α I-domain took place after the divergence of deuterostomes and after lancelet.

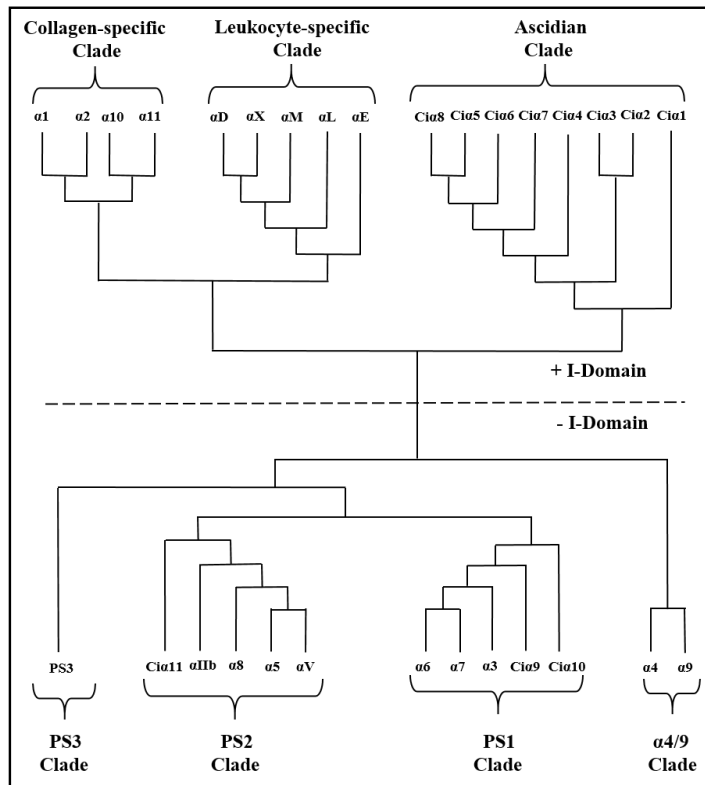


Figure 6. A schematic representation of the phylogenetic distribution among the integrin α subunits, Figure modified from publication V: ‘Evolution of Integrin I Domains’ by Johnson and Chouhan 2014. Sequences indicated as ‘Cia’ belong to *C.intestinalis*; PS3 clade consists of sequences exclusively from invertebrates and other sequences are from higher vertebrates.

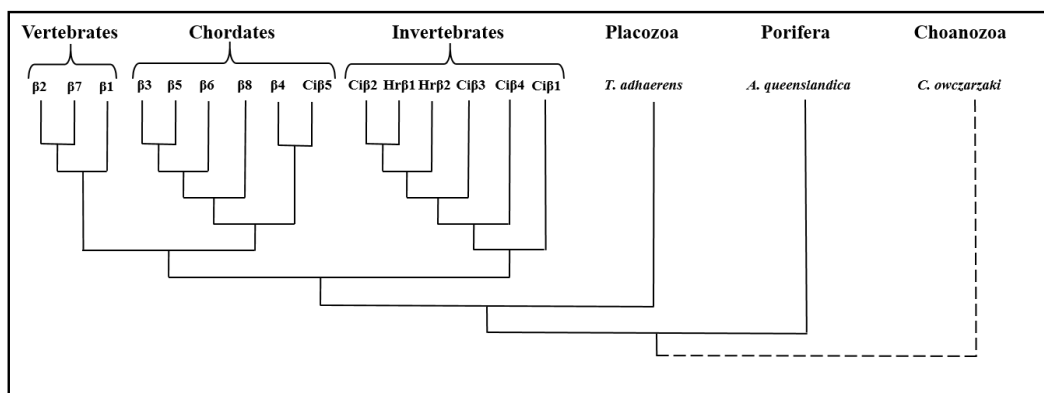


Figure 7. A schematic representation of the phylogenetic distribution among the integrin β subunits, Figure modified from publication V: ‘Evolution of Integrin I Domains’ (Johnson and Chouhan 2014).

Integrin β -subunits cluster into three major clades (Figure 7): the vertebrate clade, the chordate clade and the invertebrate clade. Vertebrate integrin subunits $\beta 1$, $\beta 2$ and $\beta 7$ constitute the vertebrate clade while the vertebrate integrins $\beta 3$, $\beta 6$ and $\beta 8$ along with Ci $\beta 5$ from *C. intestinalis* (the only invertebrate sequence to cluster within the chordate clade) constitute the chordate clade. The invertebrate clade as the name suggests is composed mostly of integrin β subunit sequences from the invertebrates like βPS and βv from *D. Melanogaster* but also including the ascidian sequences from ciona and halocynthia. In addition, integrin β subunit sequences from the placozoan *T. adhaerens*, poriferan *A. queenslandica* and choanozoan *C. owczarzaki* form the outliers to three major clades. Here, the β subunit sequence from the unicellular eukaryote *C. owczarzaki* has been utilized to root the tree.

1.4 N-terminal region: The β -propeller domain

The origin of the integrin constituent domains likely predates the integrin heterodimer itself and integrin constituent domains (like the 7-bladed β -propeller domain or the β I-like domain) have been reported to be a component of different prokaryotic proteins with unknown functions. For instance, in 1999 May and Ponting first reported similarities between bacterial sequences and integrin sequences when a PSI-BLAST run highlighted similarity between the cytoplasmic region from the human integrin $\beta 4$ subunit and a hypothetical protein 'slr1403' from the prokaryote *Synechocystis* sp. PCC680 (cyanobacteria). In addition, May and Ponting also reported that the sequences 'slr1403', 'slr0408' and 'slr1028' contain at least 13 of the repeats found in β -propeller repeats from integrin α subunits. Another study from 2002 highlighted sequence similarity between a clone 'M3G149' from *Gemmata obscuriglobus* and the Ca²⁺-region from integrin αV (Jenkins *et al.*, 2002).

This observation of similarity between bacterial sequences and human integrin sequences was further addressed by our own research group when several prokaryotic sequences were identified that aligned surprisingly well with certain regions from the integrin subunits (Johnson *et al.*, 2009). Some of these examples are as follows: a sequence 'YP 721619' from the cyanobacteria *Trichodesmium erythraeum* showed considerable similarity (over ~450 residues) with N-terminal region of the integrin β -subunit including the β I-like domain, but it lacked similarity with the stalk region domains (e.g. the I-EGF domains) or the TM region domain. Similarly, certain bacterial sequences showed similarity with the N-terminal region of various integrin α -subunits, especially the repeats corresponding to the β -

propeller domain but as expected these sequences did not show any similarity to the stalk region domains either (like the Thigh, Genu, Calf-1 and Calf-2 domains) or the TM region domains. To make things even more complex studies have also reported the presence of 7 (non-integrin type), 8, 6 and 4-bladed β -propeller domains in bacteria (Adindla *et al.*, 2007; Quistgaard *et al.*, 2009). Additionally, individual domains such as I-EGF, Ig and Rossmann folds have been seen in many proteins and in prokaryotes, but the earliest known intact integrin subunits are those reported by Seb e-Pedr s *et al.*, in 2010 in which they highlight the presence of integrin-like sequences in the genomes of unicellular eukaryotes like *C. owczarzaki* and *T. trahens* (Seb e-Pedr s *et al.*, 2010). Here we briefly discuss the two N-terminal regions from the integrin α -subunit and β -subunit i.e. 7-bladed β -propeller domain and β I-like domain.

According to the SCOP (Murzin *et al.*, 1995) and CATH (Sillitoe *et al.*, 2015) databases, the β -propeller fold can exist as 4, 5, 6, 7 or 8-bladed antiparallel β -sheets. The basic repeating unit observed in the β -propeller is a β -meander structural motif which is a series of anti-parallel β -strands linked together by hairpin loops. These anti-parallel β -strands are arranged in a toroidal or radial fashion around a central pore resembling the blades of a propeller. The β -sheets pack tightly in order to give rise to a closed structure regardless of the β -propeller superfamily and fold type (i.e. 4, 5, 6, 7 or 8-bladed). According to one study the β -propeller families display unique characteristics that differentiates them from other repetitive folds (Chaudhuri *et al.*, 2008). These features include immense diversity in β -propeller sequences thereby indicating multiple amplifications based on a single blade at different times, display of a single symmetry i.e. based on the amplification of a single blade and conservation in sequence motifs across the blades of a single β -propeller. Additionally, in many β -propeller structures, including the integrin 7-bladed β -propeller fold, a ‘velcro’ closure arrangement is observed where the N-terminal blade replaces the C-terminal blade of the last repeat (F l p and Jones, 1999).

According to the SCOP database, the integrin 7-bladed β -propeller fold is one of the fourteen protein superfamilies that adopt the same fold. Some of the other notable superfamilies are: the galactose oxidase central domain, nitrous oxide reductase N-terminal domain, WD40 repeat-like domain, clathrin heavy-chain terminal domain, peptidase/esterase ‘gauge’ domain, tricorn protease domain 2, putative isomerase YbhE, oligoxyloglucan reducing end-specific cellobiohydrolase and nucleoporin domain. One of the unique features observed in the integrin 7-bladed β -propeller fold is the presence of a consensus structural repeat known as the ‘cage motif’ (Xiong *et al.*, 2001) or the ‘FG-GAP

motif^o (Springer, 1997; Loftus *et al.*, 1994) with three or four Ca²⁺-ion binding sites located at blades 4-7. Figure 8 below depicts the presence of Ca²⁺-ions in the loops located at the bottom phase of blades 4-7 from the crystal structures of integrins $\alpha V\beta 3$ (PDB ID: 1JV2) and $\alpha IIb\beta 3$ (PDB ID: 2VDR) (Xiong *et al.*, 2002; Springer *et al.*, 2008).

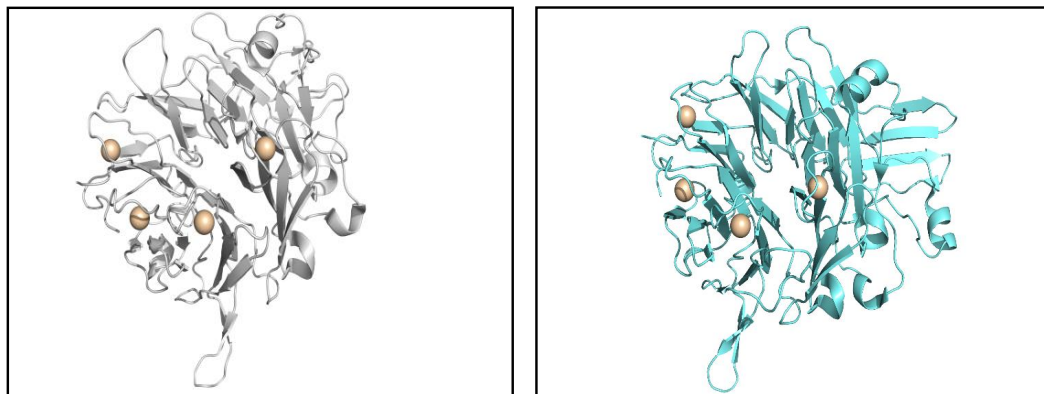


Figure 8. Screenshots of the N-terminal β -propeller domain region from integrins $\alpha V\beta 3$ (left panel, PDB ID: 1JV2) and $\alpha IIb\beta 3$ (right panel, PDB ID: 2VDR) with Ca²⁺ ions highlighted as spheres.

Integrin-like sequences have also been reported in heterokonts (or stramenopiles) which are considered to be a major alternate evolutionary line of eukaryotes containing more than 100,000 known species (including unicellular diatoms). A previous study has reported the presence of integrin-like sequences in the heterokont, a filamentous brown algae, *Ectocarpus siliculosus* (Cock *et al.*, 2010). Some of the examples from the *Ectocarpus* genome are as follows: a sequence has been reported (Genbank/EBI sequence code: CBN77719.1) that shares around 40% sequence identity with the N-terminal region of human integrin αV subunit and its C-terminal region lacks the characteristic integrin ‘KXGFFXR’ motif which is not present in the *Ectocarpus* genome; instead a glycine and alanine-rich low complexity region is observed. Another sequence (EBI sequence ID: CBJ33612.1) has been reported that shares more than 20% sequence identity with the N-terminal region of human integrin $\beta 3$ subunit, including MIDAS region.

1.5 N-terminal region: The vWA domain

The α I-domain and β I-like domain (from α -subunit and β -subunit respectively) adopt the vWA/Rossmann fold and they match with sequences from prokaryotes as well as from *Ectocarpus* (Johnson *et al.*, 2009; Cock *et al.*, 2010). vWA-domain containing proteins have been observed across all domains of life and they are known to be incorporated into proteins

that adopt a wide variety of functions: collagens, complement factors, matrillins, integrins, copines, magnesium chelatase, ion channels etc. (Ponting *et al.*, 1999; Whittaker and Hynes, 2002; Johnson and Tuckwell, 2003). The majority of these proteins having vWA domain are extracellular but the most ancient ones are located in eukaryotes and are intracellular proteins. The vWA domains adopt a classic α/β Rossmann fold where the central parallel β -sheets are flanked by α helices (CDD database, Marchler-Bauer *et al.*, 2015). Studies have shown that the vWA domains have been most represented in proteins that play a pivotal role in the immune system or they are associated with proteins involved in cell-cell and cell-ECM recognition (Colombatti and Bonaldo, 1991; Colombatti *et al.*, 1993; Whittaker and Hynes, 2002). Even though a wide array of knowledge exists in reference to the vWA domains it has not been possible to pinpoint or identify the source of insertion of these domains into the integrin repertoire (Tuckwell, 1999; Johnson and Tuckwell, 2003). But it can be stated that the integrin β -subunits were predicted to contain the vWA domain (Tuckwell, 1999; Ponting *et al.*, 2000), which was later confirmed by X-ray crystal structure study of the $\alpha\beta 3$ integrin heterodimer (Xiong *et al.*, 2001).

In integrins not having an α I-domain, the β I-like domain from the β -subunit plays a pivotal role in binding ligands through its MIDAS where the metal ion at the top crevice coordinates a carboxylate from the ligand sequence e.g. RGD, a feature very commonly observed among most vWA domains. The vWA domains of the integrin β -subunits are considered to be the most ancient among the ones that are involved in cell adhesion (Whittaker and Hynes, 2002). Already from the earliest diverging integrins, their ligands were most likely bound via interactions between the β I-like domain from the β -subunit and the β -propeller domain from the α -subunit (Wimmer *et al.*, 1999). While, the insertion of the vWA domains in the α -subunit is a chordate-specific feature (Heino *et al.*, 2008; Johnson *et al.*, 2009). As discussed earlier this insertion event created three specific groups of I-domain containing integrin α -subunits: urochordate integrin with α I-domains, collagen-binding α I-domains and leukocyte-specific α I-domains; the latter two groups of α I-domains bind their respective ligands in a MIDAS-dependent manner. It is noteworthy that certain vWA domains can bind collagen in a MIDAS-independent manner too (Colombatti *et al.*, 1993; Bienkowska *et al.*, 1997; Huizinga *et al.*, 1997; Romijn *et al.*, 2001; Nishida *et al.*, 2003).

1.6 Collagens and vertebrate collagen receptors

The Extra Cellular Matrix is complex and consists of a large number of biomolecules secreted by cells and the amount and interactions among biomolecules and with cells are tightly regulated for the proper functioning and fate of the cells that the ECM surrounds (Lin and Bissell, 1993). The ECM acts as a support platform that the cells can anchor to and subsequently form specialized tissues, therefore the ECM is essential for the organization, maintenance and remodeling of the body tissues. The vertebrates are characterized by having highly specialized support tissues such as the bone and cartilage of the skeletal system. A family of ECM proteins known as collagens play a significant role in maintaining the structure of those tissues among others; collagens are also involved in functions like cell adhesion, migration, tissue remodelling, differentiation, morphogenesis and wound healing (Myllyharju and Kivirikko, 2004). Collagen molecules are composed of three polypeptide α chains that can assemble to form homotrimers (identical α chains) or heterotrimers (different α chains). The three α chains form a left-handed helix, which is twisted to form a right-handed triple helix and give rise to a rod-shaped coiled-coil structure. This is a common structural feature shared by different collagens. The triple helical sequences are composed of a very basic repetitive 'GXY' motif, where 'G' signifies glycine while 'X' and 'Y' represent any given residue. It has been observed that the residue at position 'X' is often proline, while the residue at 'Y' is hydroxyproline, leading to a frequent repeat of 'GPO' where O represents hydroxyproline (Beck and Brodsky, 1998; Myllyharju and Kivirikko, 2004; Heino *et al.*, 2008).

Collagens in humans are numbered from I to XXVIII, which are further categorized into various subfamilies based on sequence similarities and the complexes they form (Prockop and Kivirikko 1995; Boot-Handford *et al.*, 2003; Boot-Handford and Tuckwell 2003, Myllyharju and Kivirikko, 2004; Ricard-Blum and Ruggiero 2005). Collagen subgroups similar to the human subgroups are found already in early chordates like *C. intestinalis* (Ewan *et al.*, 2005). Some of the different types of collagens expressed in humans are fibril-forming collagens (i.e. collagens I, II, III, V, XI, XXIV and XXVII), beaded-filament forming collagen IV, anchoring fibril collagen VII, network collagens (IV, VIII and X) and TM domain containing α subunit collagens (collagens XIII, XVII, XXII and XXV). Additionally, Fibril Associated Collagens with Interrupted Triple helices or FACITs refer to probably the largest subgroup among the collagens (collagens IX, XII, XIV, XVI, XIX, XX, XXI and XXII). Interestingly, even a primitive organism like the freshwater sponge *E.*

milleri has at least two distinct collagens: one of them is fibril-forming (Exposito and Garrone, 1990) while the other one is non-fibrillar (Exposito *et al.*, 1990). The genome of *M. brevicolis*, a choanoflagellate, encodes at least two genes that consist of the repetitive collagen-like sequence with 'GXY' motif. Furthermore, one laminin-like and several integrin-like genes have also been reported from this same genome (King *et al.*, 2008), suggesting that they could have played an important role in metazoan evolution given that both collagens and integrins function in cell-cell and cell-matrix interactions.

Vertebrates have developed specific mechanisms to recognize and bind different collagen types through various collagen receptors. Besides integrins some other TM collagen receptors are: Discoidin domain receptors (DDR1 and DDR2), Glycoprotein VI (GPVI), Leukocyte-associated IG-like receptor-1 (LAIR-1), Mannose receptor family (MR) and urokinase-type plasminogen activator associated protein (uPARAP) and Endo180 (Heino *et al.*, 2008). All of these receptors clearly belong to different structural groups (Leitinger and Hohenester, 2007) but, just like integrins, they are all multi-domain collagen receptors and exhibit collagen specificity. However, the Mannose receptor family including the Endo180 are endocytosis receptors and they can bind denatured collagen and are thus not classical collagen receptors. Additionally, a recent review also mentions OSCAR and GPR56 as collagen receptors (Zeltz and Gullberg, 2016).

The most abundant and widely distributed collagen-binding integrin subunits are $\alpha 1\beta 1$ (localized on mesenchymal cells, endothelial cells, smooth muscle cells, neurons, fibroblasts etc.) and $\alpha 2\beta 1$ (localized on a variety of epithelial cells, platelets, endothelial cells, keratinocytes, fibroblasts etc.). During embryonic development $\alpha 1\beta 1$ and $\alpha 2\beta 1$ have the broadest tissue expression of the collagen receptors (Barczyk *et al.*, 2010; Leitinger, 2011). Integrin $\alpha 10\beta 1$ appears to be expressed in cartilage, heart, trachea, lung, aorta and spinal chord and plays a critical role in skeletal development (Camper *et al.*, 2001; Lundgren-Åkerlund and Aszödi, 2014). Knockout studies have provided insight into the function of collagen receptor integrins and these studies have reported mild phenotypes where embryonic development was not hampered. For instance, knockout studies have shown that the $\alpha 1$ deficient mice may develop normally (Gardner *et al.*, 1996) while $\alpha 2$ deficient mice display defects in mammary gland branching morphogenesis as well as adhesion of platelets to collagen (Chen *et al.*, 2002, Holtkötter *et al.*, 2002). Meanwhile, $\alpha 10\beta 1$ is more localised on the chondrocytes so it is well worth noticing that $\alpha 10$ knockout mice showcase a defect in the growth plate (Bengtsson *et al.*, 2005) while an $\alpha 10$ truncation in dogs causes a

chondrodysplasia (Kyöstiä *et al.*, 2013). The expression of integrin $\alpha 11\beta 1$ is more prominent in the regions that are rich in interstitial collagen-networks (Tiger *et al.*, 2001) and it is implicated in regulation of periodontal ligament function in the erupting mouse incisor (Popova *et al.*, 2007). It has also been reported that a loss of matrix component binding integrin like αV , $\alpha 3$ and $\alpha 8$ results in more serious defects as compared to a loss of a collagen-receptor integrin (Hynes, 2002).

One of the key areas where the collagen-binding integrins have been studied quite well is collagen-specificity (Kern *et al.*, 1993, Tuckwell, 1995, Tiger *et al.*, 2001 and Tulla *et al.*, 2001). Integrin $\alpha 1$ prefers collagens IV and VI along with the fibril forming collagens, while the $\alpha 10$ subunit shares a similar specificity as $\alpha 1$ but it can also bind collagen II. Meanwhile integrins $\alpha 2$ and $\alpha 11$ preferentially bind the fibril-forming collagens. The two major driving factors behind the success of collagen-binding integrin studies are: firstly, the possibility to crystallize isolated integrin I-domains and their ability to retain specificity towards their respective ligand collagens (Kamata and Takada 1994; Tuckwell *et al.*, 1995), secondly, the synthetic tripeptides (for instance ‘GFOGER’ and integrin $\alpha 1$) have been greatly instrumental in identifying integrin-specific binding sites (Knight *et al.*, 1998, 2000). Additionally, two more integrin binding sites (‘GLOGER’ and ‘GASGER’) were reported for integrins $\alpha 1$ and $\alpha 2$; ‘GFOGER’ has also been reported to be a binding site for integrin $\alpha 11$ (Zhang *et al.*, 2003). There are two important integrin α I-domain structures available from $\alpha 2$ and $\alpha 1$ that showcase binding to ‘GFOGER’ and ‘GLOGEN’ tri-peptides respectively (PDB ID: 1DZI, Emsley *et al.*, 2000; PDB ID: 2M32, Chin *et al.*, 2013).

1.7 Structure of collagen-binding and ICAM-binding integrins

The earliest structural revelations on the integrin heterodimer were from electron microscopy reconstructions, which showed a bent structure (Carrell *et al.*, 1985; Nermut *et al.*, 1988), as well as the analysis of integrin sequences (Nermut *et al.*, 1988; Arnaout, 1990). The I-domain was already a subject of great interest early on (Larson *et al.*, 1989; Arnaout, 1990) and some of the earliest solved integrin crystal structures are that of α I-domain from the α -subunit e.g. αM (PDB ID: 1IDO and 1JLM, Lee *et al.*, 1995a,b) and αL (PDB ID: 1LFA, Qu and Leahy 1995); and α I-domain of the collagen-receptor integrin-type: $\alpha 2$ without (PDB ID: 1A0X, Emsley *et al.*, 1997) and with (PDB ID: 1DZI, Emsley *et al.*, 2000) collagen-like triple-helical ‘GFOGER’ tri-peptide bound.

Subsequently, the first integrin headpiece structures were also solved, for instance: extracellular segment of integrin $\alpha V\beta 3$ (PDB ID:1JV2, Xiong *et al.*, 2001), integrin $\alpha IIb\beta 3$ headpiece segment bound to a fibrinogen peptide (PDB ID:2VDL; Springer *et al.*, 2008), $\alpha 4\beta 7$ headpiece complexed with Fab ACT-1 (PDB ID: 3V4P, Yu *et al.*, 2012), $\alpha 5\beta 1$ integrin headpiece in complex with 'RGD' peptide (PDB ID: 3VI4, Nagae *et al.*, 2012) recently two I-domain containing integrin ectodomain structures were solved, that of $\alpha X\beta 2$ (resolution: 3.5 Å; PDB ID: 3K6S, Xie *et al.*, 2010) and $\alpha L\beta 2$ (resolution: 2.15 Å; PDB ID: 5E6S, Sen and Springer, 2016). In addition, quite a few structures deposited in the PDB repository correspond to the integrin TM and cytoplasmic region. Some of these structures are: NMR structure of $\alpha IIb\beta 3$ cytoplasmic domain (PDB ID: 1M8O, Vinogradova *et al.*, 2002), NMR structure of the cytoplasmic domain of integrin αIIb in DPC micelles (PDB ID: 1S4W, Vinogradova *et al.*, 2004), platelet integrin $\alpha IIb\beta 3$ transmembrane-cytoplasmic heterocomplex (PDB ID: 2KNC, Yang *et al.*, 2009), structures and interaction analyses of the integrin $\alpha M\beta 2$ cytoplasmic tails (PDB ID: Chua *et al.*, 2011), integrin $\alpha IIb\beta 3$ transmembrane complex (PDB ID: 2K9J, Lau *et al.*, 2009) to name a few.

The aforementioned two structures ($\alpha X\beta 2$ and $\alpha L\beta 2$) that include leukocyte ectodomains show the relationship of this inserted domain to the β -propeller from which it buds out of the α subunit, and relationship to the β I-like domain and the β subunit. The insertion of the α I-domain has functioned to substantially enhance the ligand binding capacity of integrins through easy access to more bulky ligands, like collagen fibers and ICAM Ig-fold domains, in comparison to the flexible loops on ligands recognized by non-I domain integrins. As mentioned earlier the α I-domain and the β I-like domain are homologous as they adopt the same fold (Rossmann fold) and they have been categorized as members of the same family (vWA ECM family) but specifically they belong to the vWA ECM protein subfamily along with seven other similar domains (Conserved Domain Database: CDD, Marchler-Bauer *et al.*, 2015). Apart from the ECM subfamily there are at least eighteen different intracellular vWA subfamilies like midasin, copine, complement factors, collagen, trypsin inhibitor and magnesium chelatase to name a few. The vWA domains are mainly associated with proteins involved in cell-cell and cell-ECM recognition as they are key constituent domains of receptor proteins as well as pivotal ECM proteins like collagen (Colombatti and Bonaldo, 1991; Colombatti *et al.*, 1993; Whittaker and Hynes, 2002). One significant feature that is common to both the α I-domain and the β I-like domain is the presence of MIDAS where a divalent cation (like Mg^{2+} or Mn^{2+} but not Ca^{2+} due to its large size) is localized towards the top surface of the domain. Furthermore, the β I-like domain has two additional Ca^{2+} -binding

sites located near MIDAS, they are ADMIDAS (Adjacent to MIDAS) and LIMBS (ligand-associated metal-binding site).

In the structure of $\alpha 2$ I-domain bound to 'GFOGER' tri-peptide (Emsley *et al.*, 2000) the metal ion is coordinated by highly conserved residues: D151 (via a water molecule), S153 (via hydroxyl oxygen) and S155 (via hydroxyl oxygen) located on loop 1, T221 (via hydroxyl oxygen) located on loop 2 and D254 (via a water molecule) and E256 (via a water molecule) located on loop 3. The remaining coordination positions are taken up by water molecules and a negatively charged residue (like glutamate) from the ligand which displaces the water molecule in order to bind to the metal ion. Additionally, phenylalanine from the middle strand (of the ligand GFOGER tripeptide) rests atop the side chains of Q215 and N154 (from the I-domain), phenylalanine from the trailing stand is engaged in hydrophobic interactions with Y157 and L286 (from the I-domain) and arginine is located in an acidic pocket close to E256 and no salt bridge is formed. Furthermore N154 and Y157 from loop 1, H258 from loop 3 form hydrogen bonds with the collagen.

A comparison between the 'GFOGER' tripeptide collagen-bound integrin (PDB ID: 1DZI, Emsley *et al.*, 2000) and unliganded integrin $\alpha 2$ I-domain (PDB ID: 1AOX; Emsley *et al.*, 1997) reveals that movement is observed in the MIDAS loops and $\alpha 1$ helix due to the direct coordination of the metal ion, a 'slinking' motion is observed between the αC helix and the $\alpha 6$ helix resulting in the formation of the collagen binding groove coupled with a large displacement of the $\alpha 7$ helix relaying the signal downwards (secondary structure names for α helices and loops are derived from Emsley *et al.*, 2000). Therefore, in the case of ligand binding at the $\alpha 2$ I-domain, it was observed that the metal ion is displaced about 2.6 Å closer to the T221 in order to establish a direct bond, this movement of the metal ion is closely followed by the MIDAS loop 1 from the I-domain to maintain the direct bonds via S153 and S155. The MIDAS loop 3 is also rearranged and this causes the loss of direct bond between the side chain of D254 and the metal ion and E256 forms a water molecule mediated bond with the metal ion. A shift observed in loop 1 and 3 causes a rearrangement of the αC and $\alpha 7$ helix, where the $\alpha 7$ helix experiences a downward displacement of about 10 Å and this movement in turn results in breaking a salt bridge between E318 from $\alpha 7$ and R288 from the αC helix. This causes the αC helix to unwind and in turn enables the connecting loop at the N-terminus of helix 6 to form an additional turn. Also, R288 shifts closer to MIDAS and forms a water-mediated salt bridge with D254 (Emsley *et al.*, 2000).

These are some of the broad changes that take place with the transition of the $\alpha 2$ I-domain from unliganded to liganded.

A similar movement can also be observed in the case of the αM I-domain where conformational shifts are observed with respect to the change in metal coordination. In one of the crystals (Lee *et al.*, 1995a) of the αM I-domain the metal ion is coordinated by a glutamate from the neighboring I-domain in the crystal lattice thereby completing the coordination sphere of the metal ion and it was proposed that the glutamate behaved as a ‘ligand-mimetic’. A parallel can be drawn between the GFOGER tripeptide bound $\alpha 2$ I-domain and the ‘ligand-mimetic’ αM I-domain structure as MIDAS structure is quite identical and glutamate coordinates the metal ion. While, another structure of the αM I-domain obtained under different conditions which lacked the ‘ligand-mimetic’ has a similar MIDAS motif as the unliganded $\alpha 2$ I-domain (Emsley *et al.*, 1997). A comparison of these two αM I-domain structures (with and without the ‘ligand-mimetic’) with the liganded and unliganded $\alpha 2$ I-domain structure reveals the broad similarities and the conformational changes observed in the I-domains. Both cases detail a movement in the metal coordination and resulting in a bond with a threonine and a loss of bond with an aspartic acid. A similar MIDAS mediated ligand recognition mechanism can also be observed in the recently solved 3D crystal structure of $\alpha 1$ I-domain bound to ‘GLOGEN’₃ (PDB ID: 2M32, Chin *et al.*, 2013) where the glutamate from the collagen-like triple-helical peptides binds at a coordinating position to the divalent metal cation. Similarly, the negatively charged glutamate from ICAM1 (PDB ID: 1MQ8, Shimaoka *et al.*, 2003; PDB ID: 3TCX, Kang *et al.*, unpublished), ICAM3 (PDB ID: 1TOP, Song *et al.*, 2005) and ICAM5 (PDB ID: 3BN3, Zhang *et al.*, 2008) also bind to MIDAS of the αL I-domain.

One of the key diagnostic features of the collagen receptor integrin is the presence of an additional αC helix which is located towards the carboxy-terminus of the I-domain (‘284GYLNR288’ from PDB ID: 1A0X; Emsley *et al.*, 1997, Figure 9). The αC helix can be only observed in the ‘closed’ (inactive) confirmation of the α I-domain as the ‘open’ (active) confirmation forces this helix to unwind and create a groove where the collagen molecule binds and coordinates the divalent metal ion at the MIDAS. Additionally, this αC helix is not present in the I-domains of the leukocyte integrins, I-domains of the tunicates and the β I-like domain from the β -subunit. Apart from the obvious functional role of the αC helix, it serves as a marker that distinguishes between the two functionally-distinct sets of I-domains found throughout the vertebrates. Deletion of the αC helix from the $\alpha 2$ I-domain

does not inhibit binding to collagen I but the affinity is certainly reduced (Käpylä *et al.*, 2000). Recombinant $\text{Ci}\alpha 1$ from the ascidian *C. intestinalis* cannot bind the 'GFOGER' motif from collagen or the fibril-forming collagens. But, it does bind to Collagen IX in a MIDAS and metal ion independent manner (Tulla *et al.*, 2007). These results show that the metal ion dependent and MIDAS mediated adhesion of collagen to integrins is a characteristic of the vertebrates and the evolution of this specialised collagen receptor has played a crucial role in the development of bones, cartilage and the blood vessel system.

Integrins without the α I-domain bind simple signature sequences e.g. "RGD" on solvent exposed loops of their ligands, thereby limiting the ligands that can be recognized. However, with the insertion of the highly solvent-exposed α I-domain integrins have gained access to unprecedented and bulkier ligands and this insertion event has forced the β I-like domain to take on a new responsibility as it binds a conserved 'intrinsic ligand' glutamate (E336 in $\alpha 21$) which in turn aids in stabilizing the integrin active conformation and modulating signal transduction. Structural and experimental data support a mechanism whereby when a ligand binds to MIDAS of the α I-domain then the intrinsic ligand glutamate binds to MIDAS of the β I-like domain as part of the downward integrin signaling mechanism (Alonso *et al.*, 2002; Yang *et al.*, 2004; Xie *et al.*, 2010; Jokinen *et al.*, 2010). But unfortunately, there is no direct structural evidence available to confirm this at this point because the structure is in the inactive bent conformation. Mutational analyses have shown that the mutation of this glutamate can lead to abolishment of integrin activation (Huth *et al.* 2000; Alonso *et al.* 2002). Another supporting observation in favor of this statement is the flexibility displayed by the α I-domain, especially the $\alpha 7$ helix located towards the C-terminal region. The downward displacement of the $\alpha 7$ helix could definitely provide the necessary push for interdomain interaction between the α I-domain and the β I-like domain (Xie *et al.*, 2009).

Figure 9. MIDAS residues coordinating the divalent metal ion (green spheres) in A: Integrin $\alpha 2$ I-domain in complex with ‘GFOGER’ tri-peptide (PDB ID: 1DZI), B: Integrin $\alpha 1$ I-domain in complex with ‘GLOGEN’ tri-peptide (PDB ID: 2M32) and C: Integrin αL I-domain in complex with ICAM-3 (PDB ID: 1TOP). Key water molecules are highlighted as red spheres and MIDAS residues are displayed as sticks in cyan. The I-domain binding glutamate from the three ligands (ICAM, ‘GFOGER’ or ‘GLOGEN’) is shown as Glu11, Glu212 and Glu37 in panels A, B and C respectively (page 27).

Figure 10. A) Representative structures from both known conformations of the $\alpha 2$ I-domain are superposed i.e. the liganded conformation (with ‘GFOGER’ tripeptide in wheatish tint; PDB ID: 1DZI, Emsley et al., 2000) and the unliganded conformation (in grey and the αC helix ‘284GYLNR288’ in green; 1A0X; Emsley et al., 1997) where the αC helix has unwound (RMSD = 0.73 Å). B) A close up shot in panel B. C) Superposition of the unliganded $\alpha 2$ I-domain structure (in grey and the αC helix in green; PDB ID: 1A0X) with the unliganded αL I-domain structure (in lightblue tint; PDB ID: 1ZON; Qu and Leahy, 1996) where the αC helix region is clearly absent (RMSD = 1.24 Å). D) A close up shot in panel D. E) Collagen-like ‘GFOGER’ tripeptide bound to the $\alpha 2$ I-domain through a metal ion coordinating Glu11 (from the tripeptide) located at the top phase of the $\alpha 2$ I-domain and the green highlighted region represents the unwound αC helix (PDB ID: 1DZI). F) ICAM-3 bound to the αL I-domain through a metal ion coordinating Glu37 from (ICAM-3) located at the top phase of the αL I-domain and the green highlighted region represents the absence of an αC helix (PDB ID: 1DZI) (page 28).

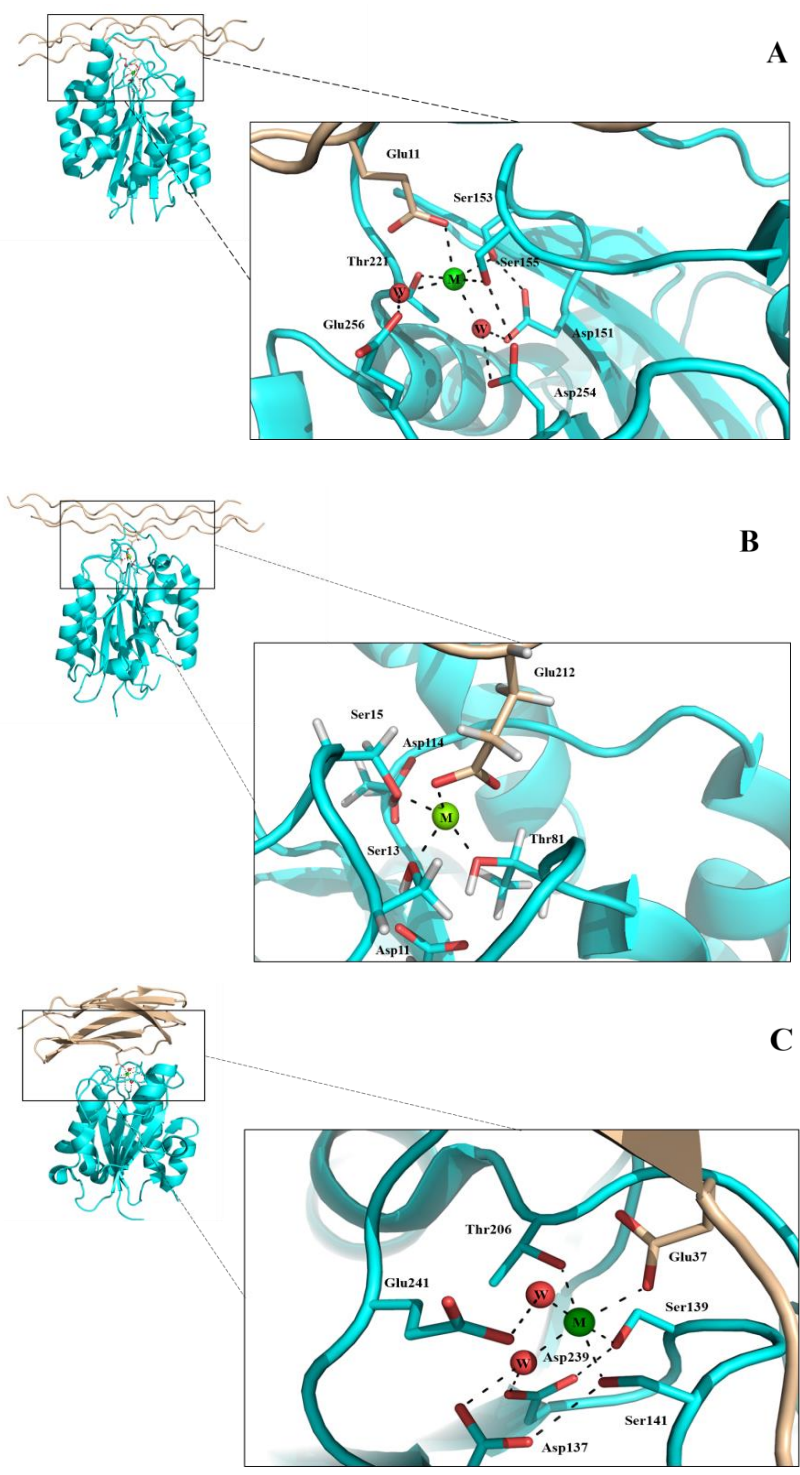
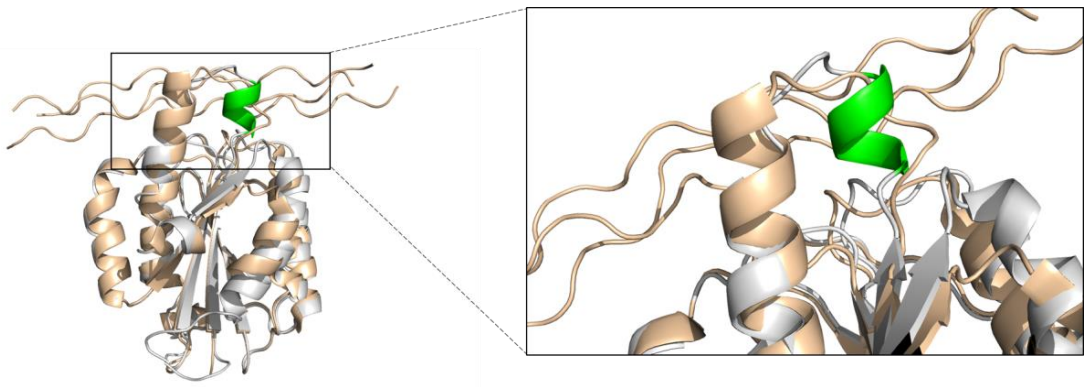
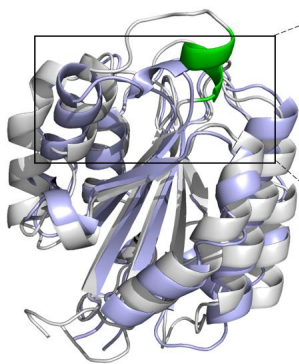


Figure 9.



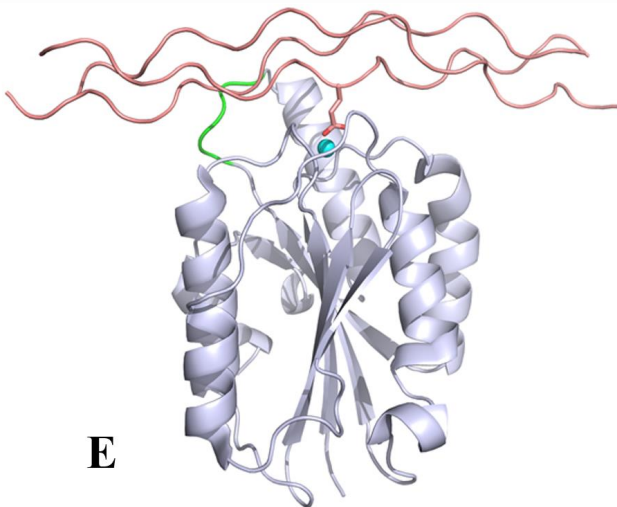
A

B

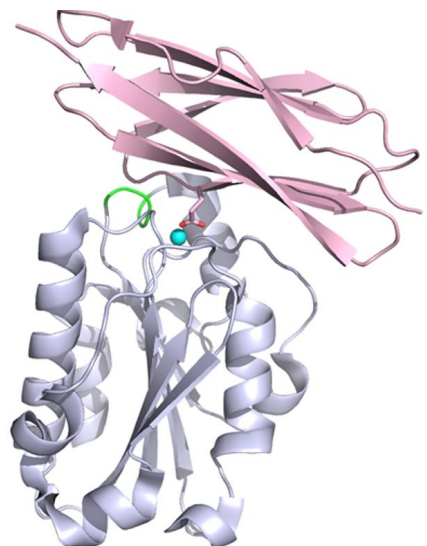


C

D



E



F

Figure 10.

2. Aims of the study

The specific aims of this thesis were:

- i) To determine the possible origin of the key functional integrin domains, especially the vWA-like α I-domain and the 7-bladed β -propeller domain.
- ii) To identify when during evolution that the α I-domain was incorporated into some integrin α subunits along with the possible source of that domain
- iii) To identify how distinct are the integrin α I-domains from the other vWA domains
- iv) To determine the point of origin for the human or mammalian integrin orthologues.

3. Methods

3.1 Online databases

Protein sequences and structures used in this thesis work were obtained from online databases. Protein sequences of interest were extracted from online databases like the NCBI protein database (<http://www.ncbi.nlm.nih.gov/protein>), Uniprot knowledgebase (<http://www.uniprot.org/>) and the Ensembl genome project (<http://www.ensembl.org/>). Elephant shark sequences were downloaded from the Elephant shark genome project (<http://esharkgenome.imcb.a-star.edu.sg/>). The NCBI protein database consists of a vast collection of protein sequences with additional information from sources like RefSeq, Swiss-Prot, GenPept, PIR, PDB, and PRF. The Ensembl project hosts a wide variety of genome databases ranging from early eukaryotic organisms to higher vertebrates and the Uniprot knowledgebase (UniprotKB) contains functional information on protein sequences. Functional information in UniprotKB is derived from a combination of manual annotation from Swiss-Prot as well as automatic annotation from TrEMBL. The emphasis behind every UniprotKB entry is to capture as much annotation information as possible like protein description, functional information, and taxonomic data along with external database references. The BLAST (Basic Local Alignment Search Tool: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>) service at the NCBI web page was utilised to perform searches across various sequence databases in order to identify and create a dataset of homologous sequences.

Protein sequences were also referenced against databases like CDD and Pfam (Protein families database: <http://pfam.xfam.org/>) for validation. CDD is a large collection of well annotated multiple sequence alignment models made publically available as PSSMs (Position Specific Scoring Matrices) which helps in identifying protein domains through RPS-BLAST. CDD also includes external data from sources like NCBI curated domains which contains information derived from experimentally solved 3D structures and it helps in identifying domain boundaries. Furthermore, domain models are incorporated from a variety of different sources like SMART (Simple Modular Architecture Research Tool), PRK (PRotein K(c)lusters), COGs (Clusters of Orthologous Groups of proteins), TIGRFAM (The Institute for Genomic Research's database of protein families) and Pfam (Protein families). The Pfam database is a collection of curated protein families and each protein family is defined by sequence alignments and profile Hidden Markov Model (HMM). The computer program HMMER is implemented to build profile HMMs and subsequently searches are

conducted against a large sequence database based on UniProtKB. It is also noteworthy that Pfam is composed of two major sections: Pfam-A which consists of manually curated entries and it covers majority of the entries present in the database while Pfam-B is composed of automatically generated entries.

All the protein structures were obtained from RCSB Protein Data Bank (Berman *et al.*, 2000). PDB is a key repository that contains experimentally solved protein structures, nucleic acids and complexes which are determined by techniques like X-ray crystallography and Nuclear magnetic resonance. Furthermore, SCOP database (Structural Classification of Proteins: <http://scop.mrc-lmb.cam.ac.uk/scop>) was consulted in order to assign protein families and folds. SCOP database is manually curated in assistance with automated tools and it provides pivotal information in relation to structural and evolutionary relationships shared by proteins with a known structure. The SCOP classification of proteins covers several levels of hierarchy but the key levels are family, superfamily and fold.

3.2 Protein sequence analyses

Multiple sequence alignments were constructed in order to closely assess: the conservation level of key residues among the constituent sequences, inspect the secondary structural elements among the target sequences and provide distance values between all pairs of sequences in order to reconstruct phylogenetic trees. Computer programs like T-COFFEE (Notredame *et al.*, 2000) and ClustalW (Larkin *et al.*, 2007) were utilized for creating alignments which were then manually curated for obvious errors. Some of these refined alignments (unpublished) were also utilised for phylogenetic analyses. Pairwise alignments for 3D-modeling were constructed using using Malign (Johnson and Overington, 1993) from the Bodil package (Lehtonen *et al.*, 2004). Secondary structure prediction (SSP) techniques are aimed at predicting secondary structural elements like α -helices, β -sheets and turns within proteins sequences based on the information derived from the known primary protein sequence. Computer program implemented in SSP are known to perform with a high accuracy therefore we used a combination of various SSP methods namely PHD from PredictProtein (Rost B *et al.*, 2004), PSIPRED (McGuffin LJ *et al.*, 2000), PROF (Ouali and King, 2000), Jpred (Cole *et al.*, 2008), GOR (Sen *et al.*, 2005) and Porter (Pollastri and McLysaght, 2005) in order to predict secondary structural elements for our target sequences and fragments. The Weblogo server (Crooks *et al.*, 2004) was utilised to examine the amino acid frequency at pivotal positions in the β -propeller repeat.

Table 2. (For publication I) Sequences used in the alignment of the β -propeller domain with two representative sequences from human and five representative sequences from different bacterial species.

Protein accession code	Organism	Protein name	PDB ID	UNIPROT	NCBI
2VDR (PDB)	Human	Integrin α IIB	2VDR	-	-
1JV2 (PDB)	Human	Integrin α V	1JV2	-	-
A3VFO	Rhodobacterales bact. HTCC2654	Hyp. Protein	-	A3VFO	ZP_01013484
ASL751	Vibrionales bact. swat 3	Hyp. Protein	-	ASL751	ZP_01816147
Q3JAB2	N. oceani ATCC 19707	Integrin α -like	-	Q3JAB2	YP_343764
A0YNP0	Lyngbya sp. pcc 8106	Hyp. Protein	-	A0YNP0	ZP_01620661
Q31NK2	S. elongatus pcc 7942	Integrin α -like	-	Q31NK2	YP_400354

In publication I, two X-ray crystal structures (α IIB β 3 PDB ID: 2VDR (Springer *et al.*, 2008) and α V β 3 PDB ID: 1JV2 (Xiong *et al.*, 2001) were studied in order to understand the ‘FG-GAP’ (Pfam01839) or the ‘cage’ motif. Upon establishing a consensus repeat pattern within the β -propeller domain, we narrowed down our dataset from 562 bacterial sequences to nine sequences from five different bacterial species which share the seven consensus repeat pattern with the human β -propeller domain. These bacterial sequences were subjected to secondary structure prediction using three different prediction methods (PHD, PSIPRED and PROF) and were subsequently aligned manually against β -propeller domain sequences from α V and α IIB in order to highlight the conservation level of the ‘FG-GAP’/cage motif as well as the secondary structural elements.

In publication II, 16 chordate sequences were aligned across the span of integrin α I-domain region in order to highlight the area of extensive similarity as well as the presence or absence of the characteristic α C helix which is a hallmark of the collagen-binding integrins. In our dataset we included two human collagen-binding integrin α I-domain sequences with

the characteristic α C helix region; α 1 and α 11, two leukocyte surface integrin α I-domain sequences which lack the α C helix; α M and α D, along with all the known α I-domain sequences from *C. intestinalis*. Secondary structure prediction for the sea lamprey sequences

Table 3. (For publication II) Sequences used in the alignment of the I-domain region with 16 representative sequences from different chordate species.

No.	Protein name	Organism	Protein coding
1.	Integrin α 1	Human α 1	NP_852478 (NCBI)
2.	Integrin α 11	Human α 11	NP_001004439.1 (NCBI)
3.	Integrin α M	Human α M	NP_000623.2 (NCBI)
4.	Integrin α D	Human α D	NP_005344.2 (NCBI)
5.	Pma_f1	Sea lamprey f1	Scaffold GL479139
6.	Pma_f2	Sea lamprey f2	Scaffold GL477642
7.	Pma_f3	Sea lamprey f3	Scaffold GL501125
8.	Ebu_f	Hagfish	BJ655520.1 (NCBI)
9.	Cin_ α 1	Sea squirt α 1	ci0100131118
10.	Cin_ α 2	Sea squirt α 2	ci0100149446
11.	Cin_ α 3	Sea squirt α 3	ci0100130596
12.	Cin_ α 4	Sea squirt α 4	ci0100130838
13.	Cin_ α 5	Sea squirt α 5	ci0100152002
14.	Cin_ α 6	Sea squirt α 6	ci0100131399
15.	Cin_ α 7	Sea squirt α 7	ci0100152615
16.	Cin_ α 8	Sea squirt α 8	ci0100130149

(Pma_f1, Pma_f2 and Pma_f3) were conducted with the help of five prediction methods Jpred, GOR, Porter, Prof_seq and PSIPRED (Jones, 1999). X-ray crystal structure of integrin α 1 I-domain (PDB ID: 1PT6 (Nymalm *et al.*, 2004) was used to assign secondary structural elements to the sequence alignment. This multiple sequence alignment clearly

highlights the presence of MIDAS residues as well as the α C helix in the lamprey sequences.

Table 4. (For publication III) Chordate genomes and EST assemblies utilized for the creating three multiple sequence alignment datasets. (Also refer supplementary table1 in publication III). Here ‘-’ indicates classification not available.

Organism	Sequence code used	Scientific name	Subphylum / Superclass / Class / Subclass / Order
Human	Hsa	<i>Homo sapiens</i>	Vertebrata / Tetrapoda / Mammalia / Theria / Primates
Chimpanzee	Ptr	<i>Pan troglodytes</i>	Vertebrata / Tetrapoda / Mammalia / Theria / Primates
Horse	Eca	<i>Equus caballus</i>	Vertebrata / Tetrapoda / Mammalia / Theria / Perissodactyla
Mouse	Mmu	<i>Mus musculus</i>	Vertebrata / Tetrapoda / Mammalia / Theria / Rodentia
Chicken	Gga	<i>Gallus gallus</i>	Vertebrata / Tetrapoda / Aves / - / Galliformes
African clawed frog	Xtr	<i>Xenopus laevis</i>	Vertebrata / Tetrapoda / Amphibia / - / Anura
Green spotted pufferfish	Tni	<i>Tetraodon nigroviridis</i>	Vertebrata / Osteichthyes / Actinopterygii / Neopterygii / Tetraodontiformes
Nile tilapia	Oni	<i>Oreochromis niloticus</i>	Vertebrata / Osteichthyes / Actinopterygii / Neopterygii / Perciformes
Zebrafish	Dre	<i>Danio rerio</i>	Vertebrata / Osteichthyes / Actinopterygii / Neopterygii / Cypriniformes
Common carp	Cca	<i>Cyprinus carpio</i>	Vertebrata / Osteichthyes / Actinopterygii / Neopterygii / Cypriniformes
Elephant shark	Cmi	<i>Callorhynchus milii</i>	Vertebrata / Chondrichthyes / Chondrichthyes / Holocephali / Chimaeriformes
Inshore hagfish	Ebu	<i>Eptatretus burgeri</i>	Vertebrata / - / Myxini / - / Myxiniiformes
Sea lamprey	Pma	<i>Petromyzon marinus</i>	Vertebrata / - / Cephalaspidomorpha / - / Petromyzontiformes
Vase tunicate	Ci	<i>Ciona intestinalis</i>	Tunicata / - / Ascidiacea / - / Enterogona
Sea pineapple	Hro	<i>Halocynthia roretzi</i>	Tunicata / - / Ascidiacea / - / Pleurogona

In publication III, three different multiple sequence alignments datasets (unpublished) were created based on the length of sequence alignment and subjected to phylogenetic analyses (also refer the next section):

- a) 69 sequences with the nearly full length (Pma_f3) sea lamprey integrin α -sequence,
- b) 72 sequences with a coverage of the common region (406-409 residues) including the three lamprey sequences Pma_f1-3 and
- c) 73 sequences with a coverage of the α I-domain region (~200 residues) including the three lamprey sequences Pma_f1-3 and the hagfish fragment (Ebu_f).

Phylogenetic studies: Phylogenetic studies deal with evolutionary relationship among species based on molecular sequence data (like DNA or protein) or morphological data. In the case of molecular sequence data, a well-refined multiple sequence alignment is supplied to a phylogenetic program in order to establish evolutionary relationship among the candidate species from the sequence alignment. This evolutionary relationship can be assessed through various phylogenetic approaches like Neighbour Joining (NJ), Maximum Likelihood (ML), Maximum Parsimony (MP) and Bayesian Method (BM).

Phylogenetic analyses were performed using MEGA (Tamura *et al.*, 2011), MrBayes (Huelsenbeck *et al.*, 2001) and Phylip (Felsenstein, 1989). The NJ phylogenetic test was based on the Jones-Taylor-Thornton (JTT) distance matrix, which was made using MEGA and phylip. The ML phylogenetic test was based on the Whelan and Goldman (WAG, Whelan *et al.*, 2001) matrix, which was selected by ProtTest (Darriba *et al.*, 2011) and MEGA to be the best-fit model based on the Bayesian Information Criterion (BIC) which is implemented for selecting the best statistical model (among a set of finite statistical models) for any given data. Felsenstein's bootstrap method (Felsenstein, 1985) was implemented to validate the tree topology. BM phylogenetic analyses were performed based on the WAG matrix with MCMC analysis. The Bayesian posterior probability was used to assess the confidence level of the tree nodes.

In publication III, three datasets were created (mentioned above) and each dataset was subjected to NJ, ML and BM phylogenetic analyses.

3.3 3D modelling and structural analyses

3D modelling or comparative modelling is a very useful technique for studying proteins that lack an experimentally solved 3D structure. As the name suggests the target protein sequence is modelled to an atomic resolution based on a template structure (which is

actually an experimentally solved 3D structure) belonging to a homologous protein family. Since related proteins share a higher sequence identity along with similar 3D folds, this makes it possible to generate a quality working model. However the lower the sequence identity between the template and the target sequence, the lower the quality of the model. Where the sequence alignment is wrong the resulting model will be wrong too, so much effort is made to evaluate the alignment and estimate levels of reliability.

3D modelling was performed using Modeller (Sali and Blundell, 1993) and Homodge from Bodil (Lehtonen et al., 2004). Modeller generates 3D models by optimally satisfying spatial restraints derived from the alignment, structures related to the target, protein structures in general and expressed as probability density functions (pdfs) for the features restrained. The program also incorporates limited functionality for *ab initio* structure prediction of loop regions of proteins, which are often highly variable even among homologous proteins and therefore difficult to predict by homology modelling. The Homodge program is located in the Bodil package and it helps in generating a 3D model by relying on information from the template structure with minimum alterations. This method is useful especially in case of proteins with high sequence similarity. The generated model does not undergo minimization which makes the process quite fast and quick to assess the quality of the model and prompt any further action like refinement of the model or the sequence alignment. The models constructed with homodge were subjected to energy minimization with the Charmm forcefield (Brooks *et al.*, 1983) from the Discovery Studio package (<http://accelrys.com/>). All of the models were visualized and analysed with Bodil and the side-chain conformations were investigated using the rotamer option. VERTAA from the Bodil package was used to superimpose 3D models on to their respective template structure. In publication III, 3D models for the protein sequences Pma_f1, Pma_f2 and Pma_f3 (from sea lamprey) were created in the ligated and closed conformation respectively. Models for Pma_f1, Pma_f2 and Pma_f3 were created based on the crystal structure of the integrin $\alpha 2$ I-domain (open and ligated form in complex with the collagen-like 'GFOGER' tripeptide, PDB ID: 1DZI (Emsley *et al.*, 2000); Closed form, PDB ID: 1AOX (Emsley *et al.*, 1997). Surf2 program (Prof. Mark S. Johnson, unpublished) was implemented to study the interactions between $\alpha 2$ I-domain and 'GFOGER' tripeptide. In publication I, crystal structures of the extracellular region from integrin $\alpha \text{IIb}\beta 3$ (PDB ID: 2VDR) and $\alpha \text{V}\beta 3$ (PDB ID: 1JV2) were studied using Sybyl (Tripos Associates, Inc., St. Louis, MO) and Bodil. The torsion angles of amino acids (ψ and ϕ) from the β -propeller (for αIIb and αV) domain were calculated using Sybyl.

4. Results

4.1 Conservation of the human integrin-type β -propeller domain in bacteria

Integrins have been extensively studied over the past 30 years and several crystal structures have been solved and deposited in the PDB repository. However, there was a limited amount of information on the prokaryotic origin of individual domains that constitute the integrin heterodimer. The aim of our first study (publication I) was to study the origin of constituent domains from integrin α and β subunits prior to the divergence of multicellular organisms. Certain integrin-type constituent domains can be detected in bacteria and the 7-bladed β -propeller domain (located in the α -subunit) is one such example. The available X-ray structures of integrin extracellular region were studied in order to identify characteristic structural motifs and map them on to a selected set of bacterial sequences to identify the origin of the metazoan-type 7-bladed β -propeller domains in prokaryotes.

4.1.1 X-ray structures of $\alpha V\beta 3$ and $\alpha IIb\beta 3$ and unique structural characteristics

During the course of this project work, ectodomain regions from the available crystal structures i.e. integrins $\alpha V\beta 3$ (PDB ID: 1JV2, 3.10 Å resolution) and $\alpha IIb\beta 3$ (PDB ID: 2VDL, 2.40 Å resolution) were studied. Additionally, an in-depth analysis of the remaining 16 integrin α subunit sequences was done which was helped in identifying structural characteristics that distinguish human integrin-type 7-bladed β -propeller super family from the remaining 13 super families that adopt the 7-bladed β -propeller fold. Furthermore, 32 representative structures from 13 superfamilies were also examined that lead to the identification of four structural characteristics that are uniquely specific to the human integrin-type 7-bladed β -propeller super family and these structural characteristics can be utilized in order to identify sequences that may adopt the same fold. These structural characteristics have been summarized in Figure 11 and described as follows:

i). Presence of the ‘FG-GAP’ motif or the ‘cage’ motif: The ‘FG-GAP’ motif was identified based on a combination of sequence similarities and a structural model of the integrin 7-bladed β -propeller domain. According to this model each propeller domain basically consists of seven repeating blades and each blade consists of four antiparallel β -strands wherein the ‘FG’ (Phe-Gly) pair is located on the first β -strand while the ‘GAP’ (Gly-Ala-Pro) tripeptide is located on the second β -strand (Springer 1997). The cage motif was identified as $\phi\phi G\phi X_{13-20} PX_{2-15} GX_{5-8}$ (ϕ - aromatic residue; G - glycine; X - any residue; P -

proline) consensus sequence in the ectodomain X-ray crystal structure of integrin $\alpha V\beta 3$ (Xiong *et al.*, 2001).

ii). Type I and type II β -turns: are observed in each blade/repeat of the integrin 7-bladed β -propeller domain and they are located on the adjacent loop regions i.e. segment A (β -turn type II) and segment B (β -turn type I) in Figure 11. In depth analysis of 32 representative structures from the remaining 13 superfamilies clearly shows that at least 12 structures do not contain even a single pair of adjacent β -turns in their 7-bladed repeats while the remaining 20 structures contain at least a pair of β -turns in one of the seven blades. Therefore, apart from the integrin 7-bladed β -propeller domain there are no other known representative structures from the 13 superfamilies that contain neighbouring β -turns on adjacent loops in all the repeats of the 7-bladed β -propeller domain.

iii). An intricate H-bonding network: Another integrin specific characteristic is the presence of an intricate H-bonding (hydrogen bonding) network which occurs due to interaction among the two adjacent β -turns (type I and type II). A closer look reveals that five residues from segment A, five residues from segment B and two residues from segment C are involved in creating this elaborate H-bonding network. Furthermore, this characteristic feature is not observed in any of the other representative structures because the β -turns which are pivotal for creating this interaction network are easily located beyond the H-bonding distance.

iv). Presence of a Ca^{2+} binding motif: A Ca^{2+} binding motif 'DxDxDG' (D - aspartate; x - any residue; G - glycine) is located on the opposite end of the β -turns and 'FG-GAP' motif and it grants stability to the β -blade/repeat. This motif spans over two loops (loops two and four) that come together to co-ordinate the divalent cation (Ca^{2+}), although this motif is observed only in four or five blades out of the seven it is a distinguishing characteristic of the integrin-type 7-bladed β -propeller domain. These structural characteristics are discussed in detail below

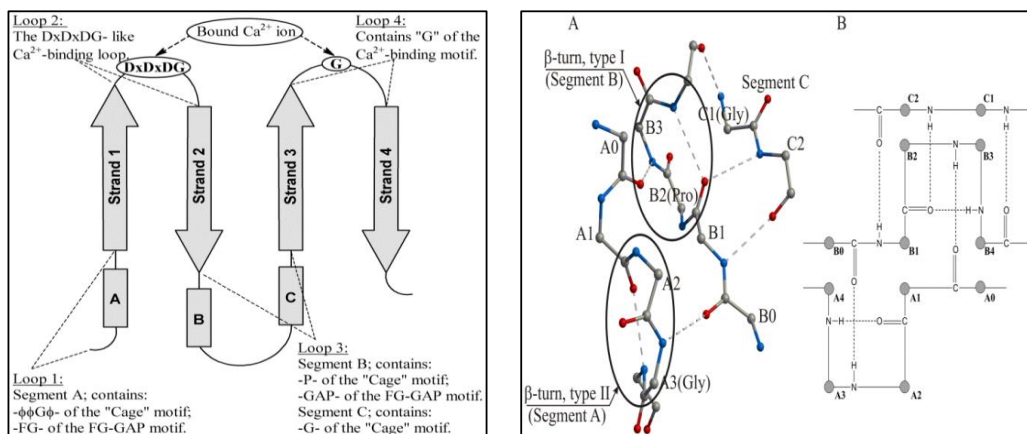


Figure 11. Schematic architecture of a typical β -propeller repeat (left panel), which in turn is comprised of four anti-parallel β -strands (β -strands 1-4) and four connecting loops (Loop1-4). The pivotal A, B and C segments are located on loops 1 and 3 respectively while the Ca^{2+} -binding 'DxDxDG' motif is located on loops 2 and 4. The 'cage' motif and the 'FG-GAP' motif describe the same structural motif; the key residues constituting the cage motif are localised in the segments A, B and C while the key residues that comprise the 'FG-GAP' motif are localised in the segments A and B. An extensive hydrogen bond network (right panel) exists between the segments A, B and C which links five residues from segment A (A0-A4), five residues from segment B (B0-B4) and two residues from segment C (C1,C2). Highly conserved glycine and proline residues at positions A3 and B2 respectively are also clearly highlighted. Figures in the left and right panel are from publication I: Chouhan et al., 2011 reprinted with permission.

4.1.2 β -turns and torsion angles

The β -turns are commonly occurring secondary structural elements that are observed in protein structures and they are classified as coils since they are non-repetitive structures. Longer repetitive structures like the α -helices and β -strands are characterised by the presence of successive residues that have similar torsion angles (Φ and Ψ angles), while non-repetitive structures like β -turn are characterised by different torsion values for each residue. Based on a theoretical conformational analysis β -turns were first described in 1968 where the available conformational freedom for a four residue system was studied which could be stabilised by a hydrogen bond between the CO group of the n residue with the NH group of the $n+3$ residue (Venkatachalam, 1968). It is noteworthy that a β -turn is essentially composed of four residues and based on the torsion angles there are different types of β -

turns: type I, type II, type I', type II', type VIa1, type VIa2, type VIb, type VIII and type IV (Richardson, 1981; Hutchinson and Thornton, 1994).

Among the β -turns, type I and type II are the most commonly occurring and results from publication I have clearly indicated that segment A located on loop 1 consists of a type II β -turn while the segment B on loop 3 consists of type I β -turn. As mentioned earlier the β -turn is essentially composed of four residues and they have been annotated in publication I as A1-A4 for segment A (type II β -turn) and B1-B4 for segment B (type I β -turn). Some distinguishing characteristics of a type II β -turn are: i). torsion angles values $\Phi A2 = -60^\circ$, $\Psi A2 = +120^\circ$, $\Phi A3 = +90^\circ$, $\Psi A3 = 0^\circ$ (A2 and A3 are residue positions in segment A); ii). Glycine occupies the position at 3 (here annotated as A3) and this position is very well conserved; iii). A hydrogen bond exists between the main chain oxygen atom (from the carbonyl group) of the residue at position A1 and the main chain nitrogen atom (from the amino group) at the position A4. Ramachandran plots were prepared (for residue positions A2 and A3 of the β -propeller blades from both the X-ray structures of α IIb and α V) to further substantiate the observations that almost all the residues at position A2 and A3 in the beta propeller blades have torsion angles that correspond to the values of a typical type II β -turn. Furthermore, conservation level of glycine at position A3 (supplementary information provided along with publication I) and the presence of a stabilizing H-bond clearly indicates the presence of a type II β -turn.

As seen in Figure 12, the segment B corresponds to a type I β -turn and the torsion angles are $\Phi B2 = -60^\circ$, $\Psi B2 = -30^\circ$, $\Phi B3 = -90^\circ$, $\Psi B3 = 0^\circ$ (B2 and B3 are residue positions in segment B). A conserved proline residue is located at the position B2 in most of the repeats of the β -propeller domain (supplementary table provided along with publication I). Due to structural and functional constraints imposed on certain positions (like A3 and B2 in type I and type II β -turns) they are taken up by certain specific residues which also results in high level of conservation like glycine and proline (at A3 and B2 respectively). It is noteworthy that these two residues are common to both, the cage motif as well as the 'FG-GAP' motif. However, certain residues do not strictly adhere to the torsion angle values and display deviations, like in the case of type II β -turn residues A2 and A3 from blade 6 of integrin α V subunit are known to deviate from the standard torsion angle values. While blade I from integrin α V subunit has a non-standard conformation with respect to a type I β -turn; $\Phi B2$ P41 = -57° , $\Psi B2$ P41 = 163° , $\Phi B3$ K42 = 64° , $\Psi B3$ K42 = 32° (Figure 12).

In addition to a high level of sequence conservation across the blades of the two β -propeller domains (from human integrin αV and αIIb subunits) 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. The secondary structure elements of segments A, B and C are also identical in all seven blades of the β -propeller domains of integrins αV and αIIb .

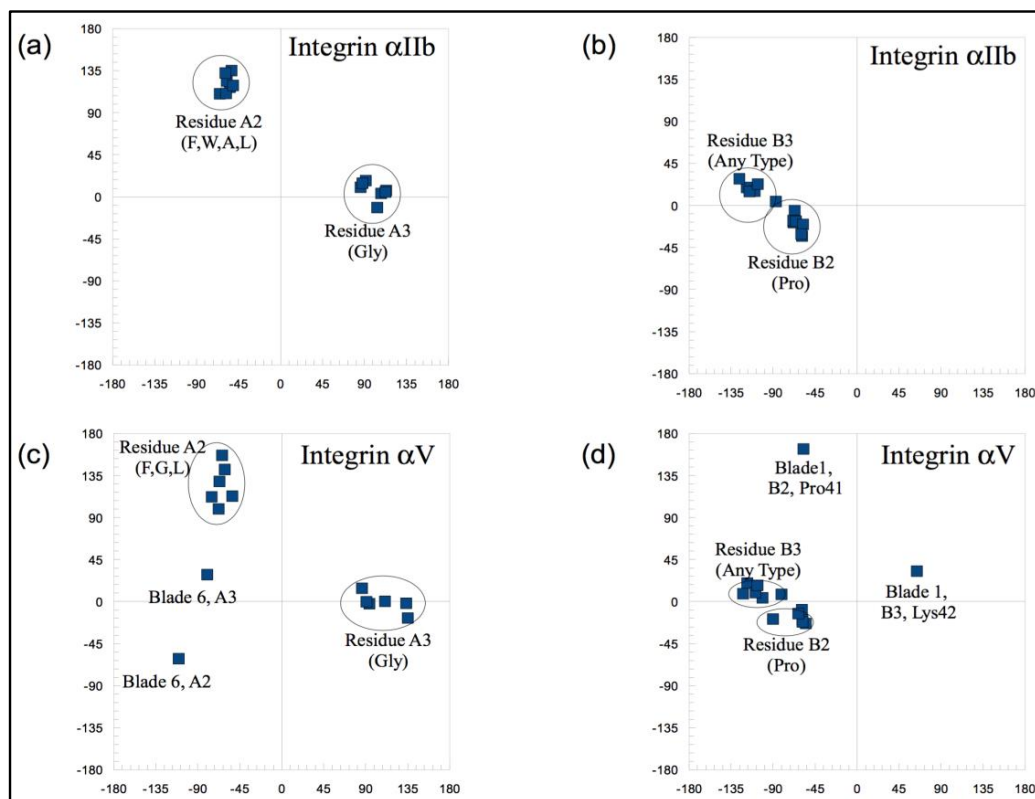


Figure 12. Ramachandran plots A through D highlight the fourteen pairs of torsion angles for residues A2 and A3 from segment A; and residues B2 and B3 from segment B derived from the two integrin structures αV and αIIb . The amino acid composition along with their respective torsion angle values (Ψ and Φ) for second and third amino acid position within the segments A and B correspond to β turn types II (A2 and A3) and I (B2 and B3) respectively. Plots were also prepared for two aforementioned exceptions i.e. blade 6 residues (A2 and A3) from segment A from integrin αV subunit (C) and blade 1 from segment B from integrin αV subunit. Figure from publication I: Chouhan et al., 2011 reprinted with permission.

4.1.3 Ca²⁺ binding motif (DxDxDG)

The divalent Ca²⁺ binding 'DxDxDG' motif is located on loops 2 and 4 ('DxDXD' on loop 2 and 'G' on loop 4) but on the opposite end of 'FG-GAP' repeat/Cage motif and the β -turns. This Ca²⁺ binding motif is not observed in all the blades of the β -propeller domain but it is present in certain repeats (3-4 repeats depending on the integrin α subunit). Interestingly, glycine from the 'DxDxDG' motif which is located on loop 4 coordinates the divalent Ca²⁺ cation through a water molecule (which is bound to the Ca²⁺ cation). It is also worth mentioning here that this 'DxDxDG' motif is also present in quite a few calcium binding proteins that are completely unrelated, for e.g. anthrax protective antigen and human thrombospondin (Rigden and Galperin, 2004). Furthermore, the Ca²⁺ cation is tightly coordinated by residues from the calcium binding loops 2 and 4 (Figure 13). The two motifs: i.e. calcium binding motif (top phase) and the 'FG-GAP' motifs (bottom phase) are responsible for granting stability to the β -propeller domain as they are two key anchor points on the opposite ends of the β -blade.

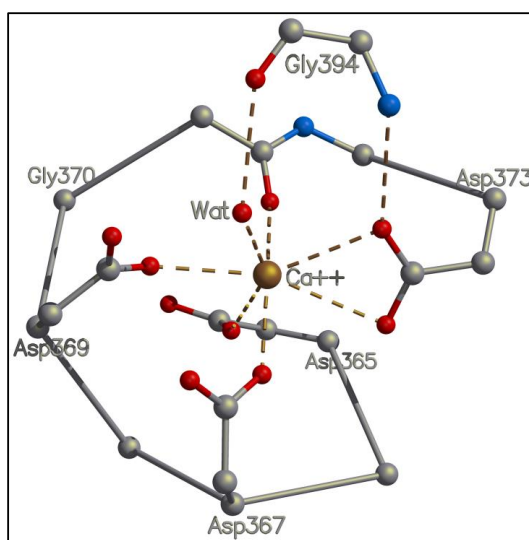


Figure 13. The calcium (Ca²⁺) binding motif 'DxDxDG' present in certain repeats of the human integrin β -propeller domain. A strong network of ionic interactions exists between the divalent cation and the side chains of the conserved residues located between the β -strands 1 and 2; while main chain atoms from the residue located between the β -strands 3 and 4 interact with Ca²⁺ through a conserved water molecule and side chain of an aspartate residue interacts directly. Figure from publication I: Chouhan et al., 2011 reprinted with permission.

4.1.4 Bacterial sequence dataset

In order to identify bacterial sequences that could be most similar to the human β -propeller domain, the Pfam sequence motif with accession code 'pfam01839' was utilized and all the bacterial sequences that showed similarity to at least one copy of the structural motif (which is a characteristic of the human-type 7-bladed β -propeller domain) were extracted. The sequence motif 'pfam01839' in the Pfam database incorporates information on the secondary structural motif 'FG-GAP' as well as the Ca^{2+} binding motif which led to an identification of 1093 sequences (473 eukaryotic sequences and 620 bacterial sequences). These protein sequences were then manually curated to remove any outdated or obsolete sequence records resulting in 464 eukaryotic sequences and 562 bacterial sequences. This dataset was investigated for the presence of at least seven consecutive 'FG-GAP'/cage motif signatures using the 'sequence search' option which resulted in the identification of 229 sequences. Interestingly, some of the sequences displayed a presence of total 14 such signatures which indicates the presence of at least two tandem copies. Subsequently, the sequences that were shorter than the required length to incorporate all the structural elements of the motifs were removed from the dataset. This further reduced our bacterial sequence dataset down to 35 sequences from 21 different bacterial species which consist of seven full-length segments and each of the seven full-length segments carries the Pfam-defined 'FG-GAP'/Cage consensus signature.

These 35 sequences were further examined to identify the presence of 'FG-GAP'/Cage motif in each of the seven repeats which further reduced the dataset down to nine sequences from five different bacterial species. As seen in Figure 14, five out of nine representative bacterial sequences have been aligned with the structural alignment of human integrin αV and αIIb subunits. Furthermore, these five bacterial sequences have been listed in the *materials and methods* section and they were also subjected to secondary structure prediction with the help of PHD, PSIPRED and PROF prediction methods. Our results agree well with the topology and distribution of the β -strands in the available human β -propeller domain structures.

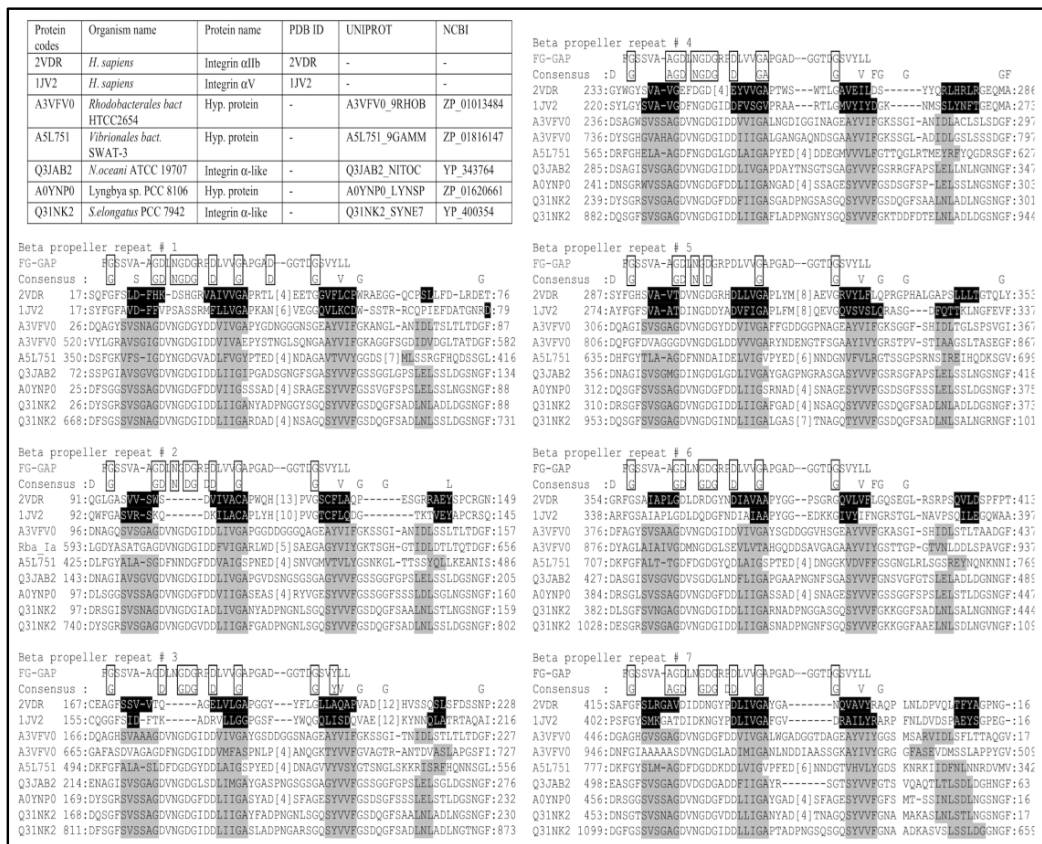


Figure 14. Sequence alignment between the β -propeller domain repeats of the human integrin α IIb and α V subunits with seven bacterial sequences obtained from five different bacterial species. Each bacterial sequence contains seven ‘FG-GAP’ repeats; secondary structural elements for α IIb and α V are highlighted in black while the same are highlighted in grey for the bacterial sequences (based on secondary structure prediction methods). Figure from publication I: Chouhan et al., 2011 reprinted with permission.

The three glycine residues as well as the proline residue (from the ‘FG-GAP’/cage motif) are very well conserved in the bacterial sequences but surprisingly the Ca^{2+} binding motif is also highly conserved in each and every repeat of the bacterial sequences. These results indicate that the bacterial sequences reported in our dataset are similar to the human-type 7-bladed β -propeller domain and they do not share any similarities or features with the remaining 13 families of β -propeller fold proteins. Furthermore, our results also point towards the origin of an ancestral fold that was much more conserved but with evolution this β -propeller fold was adopted for its functions in integrins with the loss of few of its Ca^{2+} binding sites.

4.2 Evolutionary origin of the α C helix in integrins

The nine human integrin α I-domain sequences were utilized to perform comprehensive searches across all the available genomes and EST libraries from organisms that appeared between the divergence of tunicates and the appearance of osteichthyes (Figure 15). Our searches covered the genomes of organisms like *Petromyzon marinus* (sea lamprey), *Callorhynchus milii* (elephant shark), *Leucoraja erinacea* (little skate), *Squalus acanthias* (dogfish shark) and *Eptatretus burgeri* (inshore hagfish).

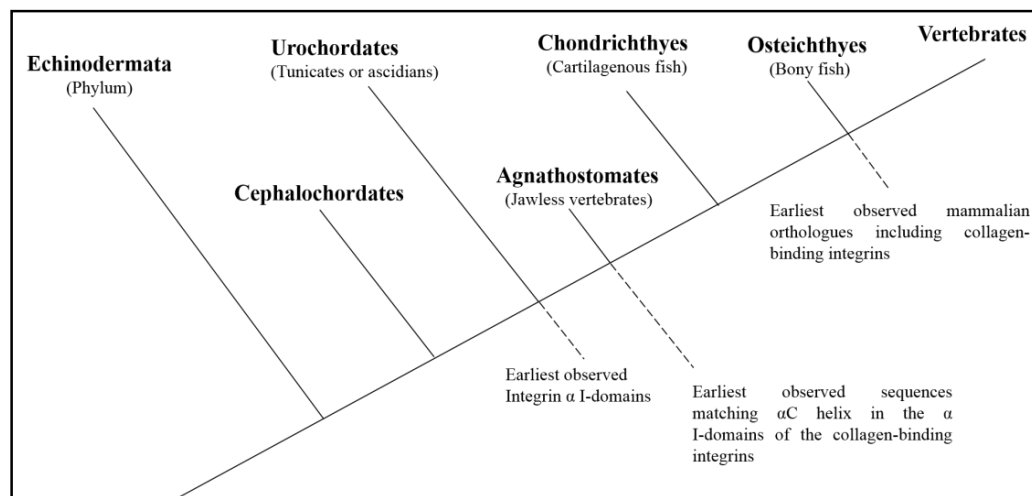


Figure 15. A schematic layout of phylum chordata depicting the evolution of integrin α I-domains. Figure from publication II: Chouhan et al., 2012 reprinted with permission.

During these searches three full length integrin α I-domain sequences from the sea lamprey genome (Pma_f1-f3) were identified, the shark/skate/ray genomic data did not yield any integrin α I-domain sequence and one short EST fragment from the hagfish genome (Ebu_f) was identified. These sequences were aligned with human integrin $\alpha 1$ and $\alpha 11$ (two out of four α C helix containing, collagen-binding human α I-domain sequences) and human integrin αM and αD (two out of five α C helix lacking human leukocyte specific α I-domain sequences) along with all the known α I-domain sequences from the tunicate *Ciona intestinalis*. The resulting dataset consisted of a total of 16 sequences (listed in *materials and methods* section) and the multiple sequence alignment clearly highlights regions of extensive sequence similarity spanning across the entire length of the integrin α I-domain (Figure 17). The secondary structure elements (β strands A-E, α helices C and 1-7) corresponding to the human integrin $\alpha 1$ I-domain (PDB ID: 1PT6) are indicated at the top of the sequence alignment. Additionally, the alignment highlights the conserved MIDAS

residues (which are highlighted as black columns) and regions of high sequence similarity (which are highlighted as grey shade boxes).

It is quite clear from the alignment that apart from the collagen-binding integrin α I-domains ($\alpha 1$ and $\alpha 11$) only the sea lamprey sequences (Pma_f1-3) contain the αC helix region which is a characteristic signature of the collagen-binding integrins while the α I-domain sequences from the leukocyte specific integrins and the tunicates lack this αC helix region. The presence of αC helix region in the lamprey sequences (Pma_f1-f3) was further substantiated by results obtained from the secondary structure prediction computer programs like Jpred (Cole *et al.*, 2008), Gor (Sen *et al.*, 2005), Porter (Pollastri and McLysaght, 2005), PROF (Rost and Sander, 1994) and Psipred (Mcguffin *et al.*, 2000) which agree with the formation of an α helix corresponding to the region where the αC helix is located in the human α I-domains (Figures 16 and 17). The EST fragment (Ebu_f) from inshore hagfish is also shown which terminates just prior to the αC helix region but it does share certain features with the other α I-domains like the presence of MIDAS residues and a strong sequence identity.

Has_a1	:	QDCEDENIQRFSAIALG	SYNRGNLST	TEKFV	EIEKSI
Pma_f3	:	KDCNDDGITRFGIAVLD	YYIS	SNMNVEKLQAE	IRSI
Jpred	:	-----E	EEEEHHHHHHHHH	-----H	HHHHHHHHHH--
Gor	:	ccccccccchhhhhhhhh	hhhhhhhhhhhhhh	hhhhhhhhhhhhhh	hhhhhhhhhhhhc
Porter	:	CCCCCCCCCEEEEE	CHHHHH	CCCCCHHHHHHHHH	HHHHHC
Prof_sec	:	-----E	EEEEHHHHHHHHH	-----H	HHHHHHHHHH--
PSIPRED	:	CCCCCCCCCE	HHHHHHHHHH	CCCCCHHHHHHHHH	HHHHHC
Pma_f2:		AACERDNI	TRYAIAV	LGYYK	RKNIDPSNFISELKAI
Jpred	:	-----E	EEEEHHHHH	-----H	HHHHHHHH--
Gor	:	ccccccccchhhhhhhhh	hhhhhhhhhh	cccccccccccc	cccccccccccc
Porter	:	CCCCCCCCCEEEEE	CHHHHH	CCCCCHHHHHHHHH	HHHHHC
Prof_sec	:	-----H	HHHHHHHHHHH	-----H	HHHHHHHHHH--
PSIPRED	:	CCCCCCC	HHHHHHHHHHH	CCCCCHHHHHHHHH	HHHHHC
Pma_f1:		DQCERDGI	TRYAIAV	LRYS	SNADDVARLINEVRSI
Jpred	:	-----E	EEEEHHH	-----H	HHHHHHHH--
Gor	:	ccccccccchhhhhhhhh	hhhhhhhhhh	cccccccccccc	cccccccccccc
Porter	:	CCCCCCCCCEEEEE	CCCCC	CHHHHHHHHHHH	HHHHHC
Prof_sec	:	-----H	HHHHHHHHHHH	-----H	HHHHHHHHHH--
PSIPRED	:	CCCHHHHHHHHHHH	HHHHHC	CHHHHHHHHHHH	HHHHHC

Figure 16. Secondary structure predictions performed for the three lamprey sequences which are based on the five prediction programs (Jpred, GOR, Porter, Prof and PSIPRED) and these predictions are compared against human integrin $\alpha 1$ I-domain sequence. Figure from publication II: Chouhan *et al.*, 2012 reprinted with permission. The αC helix region (corresponding to the human integrin $\alpha 1$ I-domain) is highlighted in grey.

Therefore, the α C helix definitely serves as a solid marker that can help in distinguishing collagen-binding integrin α I-domains from the remaining two clades (of I-domain) present in the chordates. At the time of this study the amount of available genomic data was quite limited in terms of quality and quantity. The next step was to investigate the lamprey sequences in more detail by conducting binding studies wherein these three sequences are expressed as recombinant fusion proteins and their binding affinities are tested against different types of collagens. Additionally, we were also interested in uncovering the phylogenetic relationship these lamprey sequences share with integrin sequences from the rest of the chordates.

4.3 Early chordate origin of the human-type integrin I-domains

As mentioned earlier searches were conducted across genomic assemblies and EST libraries from organisms that diverged between the appearance of the urochordates and osteichthyes. Apart from the three sea lamprey sequences and the hagfish fragment three very short fragments from the elephant shark survey genome (which was published and available until the end of 2013) were also obtained. Two out of three fragments shared close sequence similarity with human integrin I-domains from α 1 (AAVX01128089.1; 55 residues; 76% identical) and α 2 subunits (AAVX01352230.1; 55 residues; 71% identical). The third fragment showed similarity to the fifth repeat of the β -propeller domain from the human integrin α 2 subunit (AAVX01625876.1; 52 residues; 63% identical).

At the beginning of year 2014 a more detailed version of the elephant shark genome was released and in-depth genomic searches were performed which yielded at least four full-length sequences (orthologous to their respective mammalian integrin counterparts): collagen-binding integrin α 1, α 2, α 10 and leukocyte-specific integrin α E. This development provided the opportunity to create more a diverse dataset in order to perform phylogenetic analyses. Three different sets of phylogenetic trees were constructed based on the length of the sequences alignments as well as the different phylogenetic methods implemented: NJ, ML and Bayesian. Here only one set of phylogenetic trees are discussed i.e. ML trees while rest of the topologies obtained from the NJ and Bayesian methods can be observed in the supplementary information provided for publication III.

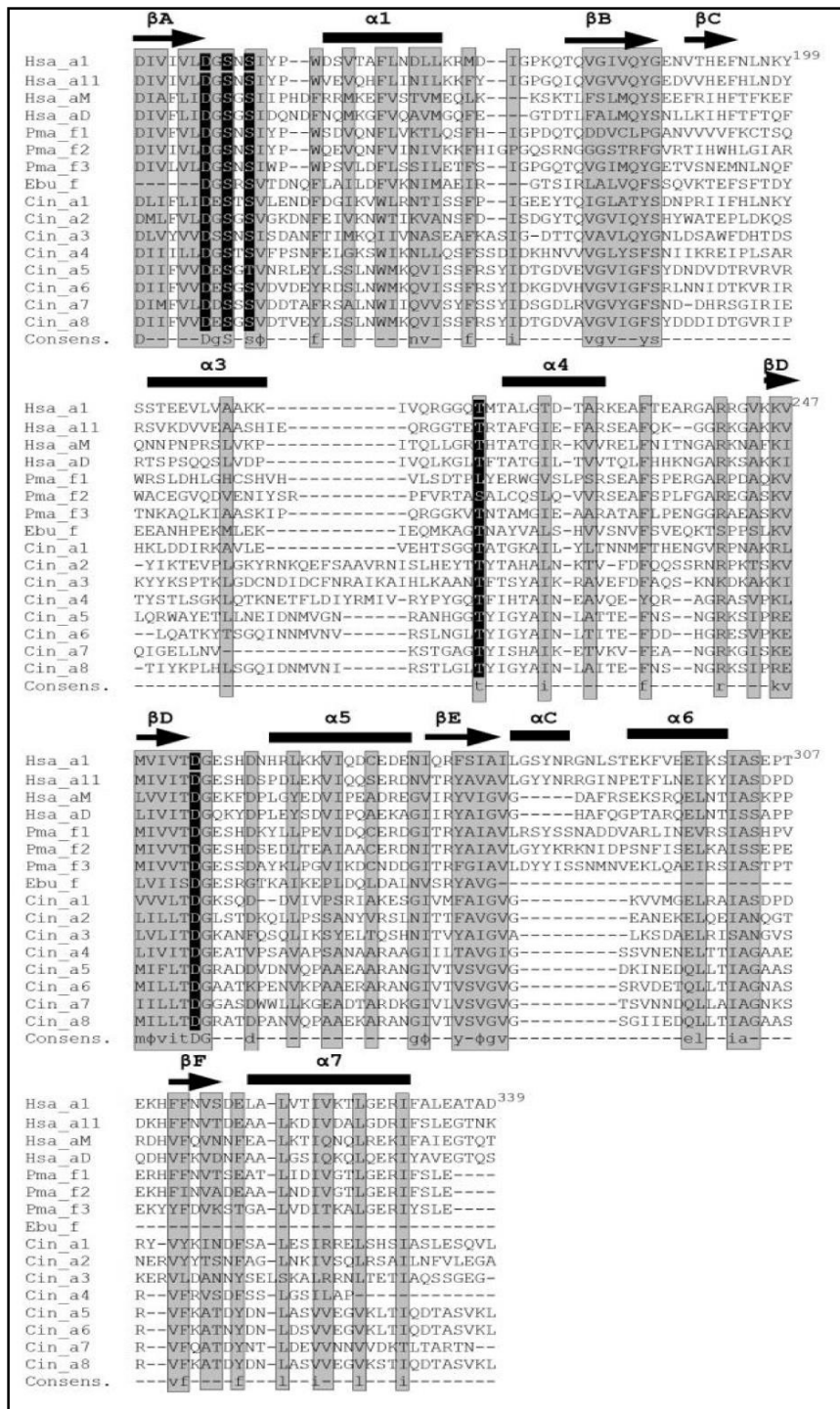


Figure 17. Sequence alignment of the integrin α I-domains and the secondary structure elements are derived from α I-domain; MIDAS residues are highlighted in black while the grey regions indicate similarity. Figure from publication II: Chouhan et al., 2012 reprinted with permission.

4.3.1 Phylogenetic analyses

The three sets of phylogenetic datasets were created based on the following sequence alignments (Figure 18):

- a). A set of 69 chordate sequences that includes a nearly full length sea lamprey sequence (Pma_f3) and four full length elephant shark sequences (integrins $\alpha 1$, $\alpha 2$, $\alpha 11$ and αE).
- b). A set of 72 sequences that have been trimmed down to cover the maximum common area between the three lamprey sequences (Pma_f1-f3).
- c). A set of 73 sequences that span across the integrin α I-domain (~200 residues) including the sea lamprey sequences, elephant shark sequences as well as the incomplete domain fragment from hagfish (Ebu_f).

Phylogenetic trees were inferred based on: NJ method where the pairwise distances between the sequences were calculated based on the JTT matrix, ML method and Bayesian method where the WAG matrix was implemented in order to resolve the phylogenetic relationship among sequences. In addition, 3D PCA multivariate plots were prepared based on the JTT matrix in order to substantiate the observed tree topologies. Most of the clustering patterns are in agreement with previously published trees as the major observed clades are: Tunicate (or Ascidian) clade, Leukocyte specific clade and Collagen-binding clade (DeSimone and Hynes, 1988; Hughes, 1992; Fleming *et al.*, 1993; Burke, 1999; Hughes, 2001; Hynes and Zhao, 2000; Miyazawa *et al.*, 2001; Johnson and Tuckwell, 2003; Ewan *et al.*, 2005; Huhtala *et al.*, 2005; Takada *et al.*, 2007; Johnson *et al.*, 2009). The tunicate clade is an outlier monophyletic group that consists of integrin sequences from the vase tunicate (*C. intestinalis*) and sea pineapple (*H. roretzi*) while the other two remaining clades consist of vertebrate integrin sequences. Some important observations to arise from the phylogenetic trees are mentioned as follows:

Phylogenetic analyses based on full-length integrin sequence alignment: All three methods implicated in deciphering the phylogenetic relationships among the full length integrin sequences essentially agree with each other except in regard to the bootstrap support value between $\alpha 1/\alpha 2$ subunit clustering in the NJ tree which is 52% while in case of ML it is 100% and posterior probability support for Bayesian is also 100%. The nearly full-length sequence Pma_f3 from the sea lamprey genome clearly diverges prior to the integrin $\alpha 10/11$ cluster and in addition, the four sequences extracted from the elephant shark genome ($\alpha 1$, $\alpha 2$, $\alpha 11$ and αE) cluster as near outliers to their respective groups for instance it is quite

clear that they diverge prior to the osteichthyes thereby indicating the presence of vertebrate orthologues in chondrichthyes. It is also noteworthy that the bony fish sequences display the presence of isoforms (like in case of zebrafish - Dre α 11A, Dre α 11B; carp - Cca α L1, Cca α L2 and tilapia – Oni α M-A, Oni α M-B) but one discrepancy that still remains to be addressed is that some bony fish sequences that branch out after α E and α L clusters appear to have diverged prior to the specialization and diversification of α D, α M and α X subunits in the mammals.

Phylogenetic analyses based on aligned common sequence region: In case of largest common sequence region shared among the sea lamprey sequences all three methods essentially are in agreement with each other as the three lamprey sequences diverge prior to the α 10/11 cluster in all the trees. Interestingly the ML and NJ methods produce topologies that are supported by near 100% bootstrap support. Additionally, the posterior probabilities observed at the branches in the Bayesian tree are also nearly 100% especially at the nodes located prior to the divergence of the three lamprey sequences.

Phylogenetic analyses based on alignment of the integrin α I-domain sequences: In this particular case we observed that although the quality of the sequence alignment across the ~200 residue long α I-domain region is quite good, it becomes rather difficult for the phylogenetic programs to distinguish and discriminate among the constituent sequences due to a lack of sufficient similarity differences as compared to a longer sequence alignment. Although, the representative trees here mirror the basic topology from the other trees (derived from longer sequence alignments) but the level of noise is higher and as a result more discrepancies are observed in case of the α I-domain based phylogenetic trees. But the three lamprey sequences do cluster in the collagen-binding clade albeit with certain variations along with poor bootstrap support values and this is also reflected in the 3D multivariate plots.

It is worth noticing that the short EST fragment extracted from the hagfish genome (which terminates prior to the α C helix region) branches out and clusters near the α L I-domains and this pattern is also observed in the multivariate plots. Furthermore, simple reverse BLAST searches have also revealed the α L I-domains to be the closest match for the hagfish EST fragment thereby indicating that an early homologue of the leukocyte specific integrin could exist in the agnathostomes but since the EST fragment is devoid of any additional distinguishing features.

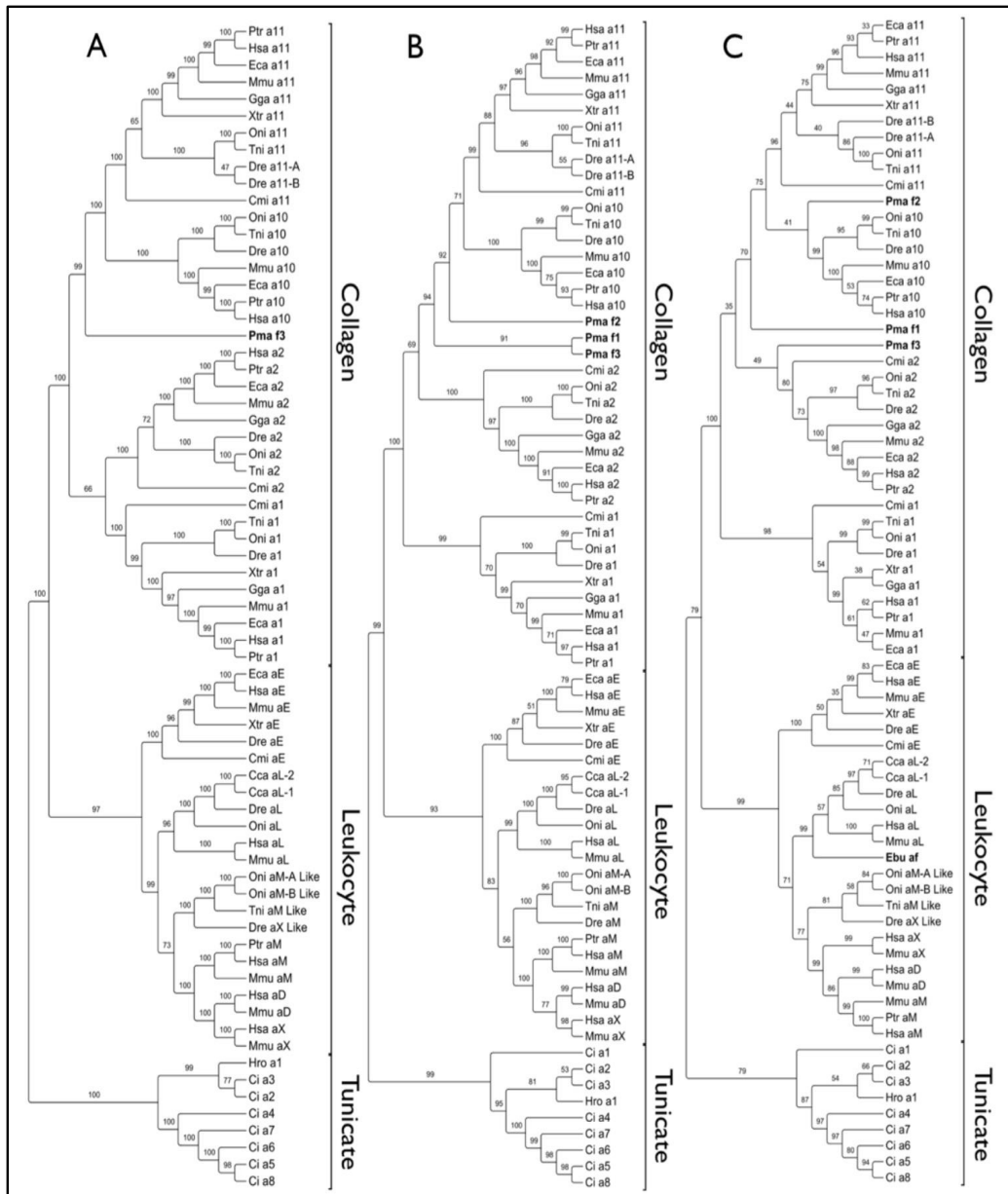


Figure 18. Maximum likelihood phylogenetic analysis of integrin sequences based on: A) full length sequence alignment, B) largest common region sequence alignment between the three lamprey sequences and C) sequence alignment of the α I-domain region. Figure from publication III: Chouhan et al., 2014 reprinted with permission.

4.3.2 Functional residues shared between human and lamprey α I-domains

The available integrin crystal structures which are in complex with their respective ligands were studied in order to identify the functional residues pivotal for the identification of 'GFOGER' and 'GLOGEN' tri-peptides that mimic the collagen tri-peptide. This was

accomplished using the Surf2 program (Prof. Mark S. Johnson, unpublished) which was utilised to study the similarities and differences among the residues that occupy an equivalent position in other human and sea lamprey α I-domain sequences. As discussed earlier in this thesis, the α I-domain clearly provides a large exposed region for the ligands to bind which is mediated through a divalent cation like Mg^{2+} or Co^{2+} and in the case of collagen-binding integrins; the $\alpha 2$ and $\alpha 1$ I-domains recognise and bind the 'GFOGER' and 'GLOGEN' tri-peptides respectively through a glutamate (like 'E11' from the 'GFOGER' tri-peptide). A similar mechanism can also be observed in case of the αL I-domain when it is in complex with ICAM-1 D1 (PDB ID: 3TCX, Kang *et al.*, unpublished), ICAM-3 (PDB ID:1TOP, Song *et al.*, 2004) and ICAM-5 (PDB ID: 3BN3, Zhang *et al.*, 2008) as the recognition and binding process takes place through a glutamate (through E34, E37 and E37 respectively).

Integrin $\alpha 2$ I-domain structure in complex with the 'GFOGER' tri-peptide (PDB ID: 1DZI) was studied with the help of SURF2 program and subsequently a table was prepared wherein all the residues from the I-domain that are located within a vicinity of 4.2 Å of the tri-peptide are displayed and comparisons were made against the other human and agnathostome α I-domain sequences (Table 5). In addition, similar tables were also prepared for $\alpha 1$ I-domain in complex with 'GLOGEN' tri-peptide and αL I-domain in complex with ICAM3 tri-peptide (not shown here, refer article III). All three tables clearly indicate that residues from the sea lamprey α I-domain sequences are more like the collagen receptor α I-domains as compared to the leukocyte specific α I-domain.

3D models were constructed for the lamprey sequences based on the $\alpha 2$ I-domain complex structure (PDB ID: 1DZI) to closely examine the interaction of α I-domains with the 'GFOGER' tri-peptide. The 3D model created for the Pma_f3 α I-domain sequence shows similarity with the $\alpha 2$ I-domain crystal structure as the two sequences (Pma_f3 I-domain and $\alpha 2$ I-domain sequences) share 44% sequence identity and there are only two deletions in the sequences alignment towards the C-terminal region. There are 16 residues from the $\alpha 2$ I-domain which are in the vicinity (4.2 Å) of the 'GFOGER' tri-peptide and two additional residues (total 18) which are a part of MIDAS constitute some of the functionally important interactions. Out of these 18 residues, 12 are identical between the Pma_f3 and the $\alpha 2$ I-domain and 14 residues are identical between the Pma_f3 and the $\alpha 1$ I-domain. One such difference can be observed in the replacement of serine with histidine, while the rest of the 2 residues (out of 3) are conserved that are pivotal for binding R12_B of the 'GFOGER' tri-peptide to $\alpha 2$ I-domain.

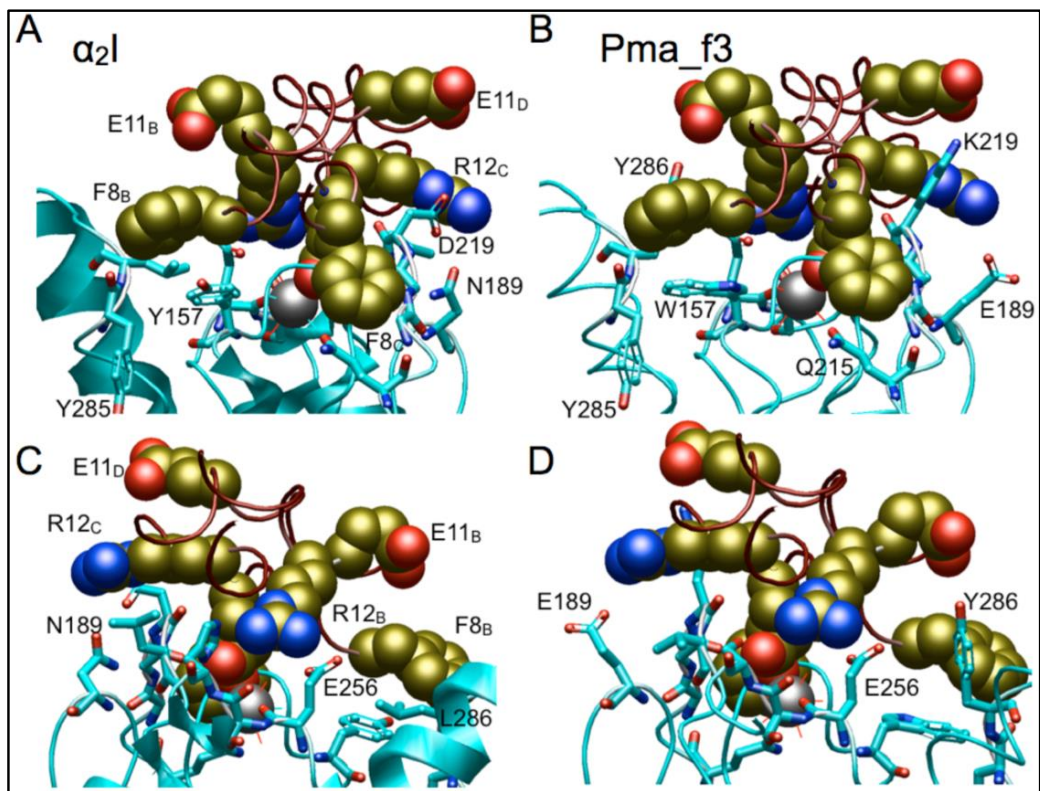


Figure 19. The panels A and C on the left depict the $\alpha 2$ I-domain crystal structure which is in complex with the 'GFOGER' tri-peptide (PDB ID: 1DZI) while the panels B and D on the right depict the lamprey Pma_f3 sequence in complex with same tri-peptide. The $\alpha 2$ I-domain crystal structure and the Pma_f3 3D model were superimposed in order to highlight the relevant residue positions in the Pma_f3 model. Residue side chains from the tri-peptides are shown as CPK model while residue side chains from model as well as the crystal structure are shown as ball and stick model. Figure from publication III: Chouhan et al., 2014 reprinted with permission.

Another example is replacement for the residue equivalent to D219 in the $\alpha 2$ I-domain and in case of Pma_f3 it is a lysine (K219) which can potentially form electrostatic interactions with E11_D (Figure 19). This interaction is particularly important as it helps in determining the collagen subtype preference and in case of human $\alpha 1$ I-domain and $\alpha 10$ I-domain an arginine is present instead of an aspartate. Pma_f1 and Pma_f2 share 9 out of 16 residues in common with human integrin $\alpha 2$ I-domain. But we did observe one discrepancy which remains to be tested and that is the observation that in $\alpha 2$ I-domain there is a threonine (T221) which is required to chelate the divalent cation at MIDAS, however in Pma_f1 there

is no conserved threonine present (even in the close vicinity) and the equivalent residue is uncertain at this point of time

Table 5. Residues from the $\alpha 2$ I-domain structure (PDB ID: 1DZI) located within 4.2 Å of the bound 'GFOGER' tripeptide along with equivalent residues from other human, lamprey α I-domains and the hagfish fragment. Numbered residue indicate sequence numbering within a solved 3D protein structure; PDB ID and resolution of the structure are also mentioned. MIDAS residues (S153, S155 and T221) are indicated in italics and D151 and D254 are not listed here. '*' indicates no equivalent or aligned residue; '?' indicates that residue is not present in the fragment; '†' indicates that alignment is uncertain at that position and '-' indicates that threonine is not present in the sequence nearby. Table from publication III: Chouhan et al., 2014 reprinted with permission.

$\alpha 2$ 1DZI 2.10 Å	S 153	N 154	S 155	Y 157	N 189	Q 215	G 217	G 218	D 219	L 220	T 221	E 256	S 257	H 258	Y 285	L 286
$\alpha 1$ 1PT6 1.87 Å	S 152	N 153	S 154	Y 156	E 188	Q 214	G 216	G 217	R 218	Q 219	T 220	E 255	S 256	H 257	S 284	Y 285
$\alpha 10$	S	N	S	Y	E	R	E	G	R	E	T	E	S	H	H	Y
$\alpha 11$	S	N	S	Y	E	Q	G	G	T	E	T	E	S	H	Y	Y
Pma_f1	S	N	S	Y	A	R	W	G	M	E	R†	E	S	H	S	Y
Pma_f2	S	N	S	Y	F	S	P	F	V	R	T	E	S	H	Y	Y
Pma_f3	S	N	S	W	E	Q	G	G	K	V	T	E	S	S	Y	Y
Ebu_af	S	R	S	T	S	Q	K	A	*	G	T	E	S	D	?	?
αL 3F74 1.70 Å	S 139	M 140	S 141	Q 143	T 175	H 201	L 203	L 204	*	L 205	T 206	E 241	A 242	T 243	*	*
αM 1IDO 1.70 Å	S 141	G 142	S 143	I 146	E 178	Q 204	L 206	G 207	*	R 208	T 209	E 244	K 245	F 246	*	*
αX 1N3Y 1.65 Å	S 140	G 141	S 142	S 144	N 176	Q 202	Q 204	G 205	*	F 206	T 207	K 242	K 243	E 244	*	*
αD	S	G	S	D	N	Q	K	G	*	L	T	Q	K	Y	*	*
αE	S	G	S	D	G	Q	G	S	*	V	T	G	I	F	*	*

4.3.3 Sea lamprey α I-domains recognize different mammalian collagen types

The sea lamprey sequences were synthesized and cloned (by our collaborators at Professor Jyrki Heino's lab) into expression vectors pGEX-2T as recombinant GST fusion proteins in the BL21 tuner strain of *E. coli* bacteria. Despite a minor amount of detected GST the expressed proteins were pure enough to carry out experiments. A solid-phase assay was implemented in order to test the recognition and binding capacity of the sea lamprey α I-domain sequences against different types of collagens. The binding studies were conducted at a fixed concentration of the Pma α I-domains (400 nM) and the results clearly indicate that the Pma α I-domains recognize a wide variety of collagens like: rat collagen I and bovine collagen II (fibrillar collagens), mouse collagen IV (network-forming collagen), and recombinant human collagen IX (FACIT). In general, the Pma α I-domains show highest binding with the rat collagen I while Pma_f3 α I-domain exhibits the best binding with all the tested ligands (Figure 20). The lamprey α I-domains mediate their respective ligands through a metal ion but when all the Pma α I-domains were incubated with EDTA in order to test the binding levels against rat collagen I and they were clearly lower (in comparison to non EDTA collagen I binding levels) indicating the importance of MIDAS and the metal ion as well. Furthermore, the binding affinity of the three lamprey sequences was also tested against the 'GFOGER' tri-peptide and all three Pma α I-domains showed good binding results with Pma_f1 and Pma_f3 α I-domains exhibiting the highest binding.

In addition, comparative binding studies were also conducted against rat collagen I where the binding affinity of Pma_f3 α I-domain in comparison to human collagen receptor α 2 I-domain (wild type human α 2I wt and open conformation mutant human α 2I E318W) was tested. Our results clearly display the lower binding levels for Pma_f3 α I-domain (in comparison to the human orthologous sequences) possibility indicating the presence of higher binding sites on rat collagen I for human α 2I wt or human α 2I E318W (Figure 21).

In conclusion, phylogenetic analyses, 3D comparative modelling as well as experimental testing was implemented in order to highlight the: i) first appearance of features (in the sea lamprey genome) that are a hallmark of the collagen receptor integrins, ii) the three sea lamprey sequences (Pma_f1-3) bind different types of collagens (including the mammalian rat collagen I) in a metal ion dependent manner, iii) vertebrate integrin orthologues exist even in the cartilaginous fish and iv) specialization and diversification of leukocyte integrins (α E, α L, α M, α D and α X) in the bony fish needs to be studied and addressed in more detail at this point of time.

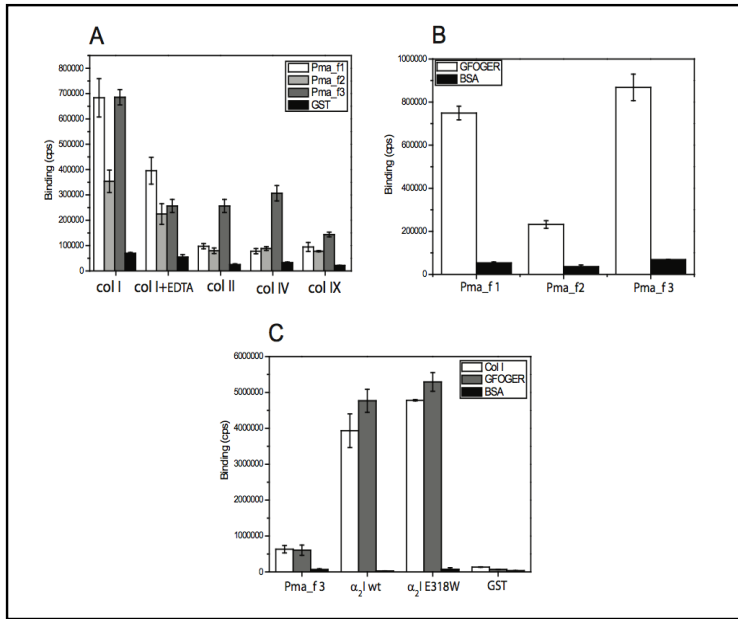


Figure 20. Panel-A: three sea lamprey sequences Pma_f1-f3 recognize different collagen types and exhibit different levels of binding; GST as control for the lamprey sequences. Panel-B: 'GFOGER' tri-peptide binding to the lamprey sequences with BSA as the control. Panel-C:

comparative binding of Pma_f3, wild type human α_2 I-domain and human E318W mutant to rat collagen I and GFOGER tripeptide. Figure from publication III: Chouhan et al., 2014 reprinted with permission.

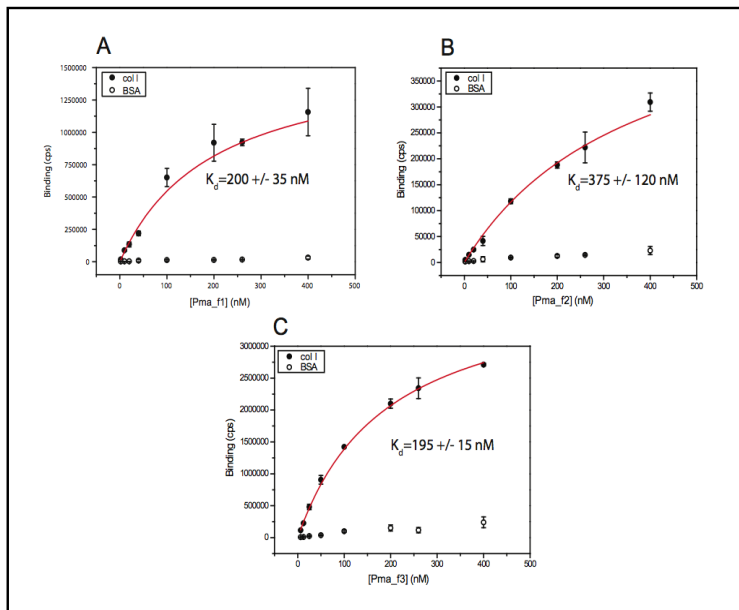


Figure 21. Binding affinities of the three Pma α I-domains against the rat collagen I as a function of the concentration of Pma_f α I. BSA serves as the control in place of collagen I. Figure from publication III: Chouhan et al., 2014 reprinted with permission.

5. Discussion

5.1 Publication I

As discussed earlier, sequence searches performed by our own group led to the identification of different bacterial sequences that aligned surprisingly well with N-terminal domains from the integrin α and β subunits respectively (Johnson *et al.*, 2009). This observation prompted us to study the bacterial sequences further and investigate in more detail whether these sequence were truly of the integrin-type i.e. a member of the thirteen 7-bladed β -propeller superfamilies (SCOP, Murzin *et al.*, 1995) or a novel superfamily adopting the same fold. We performed extensive structural and sequence studies and highlighted different bacterial sequences that share similar structural and sequence characteristics with the integrins.

5.1.1 Sequence alignment and phylogenetic tree

Five sequences from five different bacterial species (of which two sequences contain an additional tandem repeat) were manually aligned against the structural alignment of the β -propeller region from the crystal structures of α V β 3 (PDB ID: 1JV2) and α IIB β 3 (PDB ID: 2VDR) (Figure 14 and Supplementary Figure S1 from the publication I). This alignment clearly shows the presence of integrin specific characteristics like the conservation of key residues i.e. the three glycines and proline from the FG-GAP/cage motif and high level conservation of Ca²⁺-binding repeat in each blade clearly suggesting that the bacterial sequences are similar to the integrin β -propeller superfamily in comparison to the other thirteen 7-bladed β -propeller superfamilies. Additionally, the conservation level of the Ca²⁺-binding repeat in each blade suggests that the bacterial sequences may be quite similar to an ancestral fold of β -propeller domain that evolved later to serve a specific function in the integrin heterodimer thereby making the case of lateral gene transfer less likely.

In order to better understand the evolution of the 7-bladed β -propeller fold containing families, we wanted to construct phylogenetic trees wherein we would include representative sequences from species ranging from prokaryotes all the way through to humans. But, unfortunately since the families are too diverse and the representative sequences from different families lack substantial similar features it was not possible to construct a proper sequence alignment to build these phylogenetic trees.

5.1.2 3D comparative modelling

3D comparative models for the bacterial sequences of interest can be constructed by utilizing the human integrin 7-bladed β -propeller structure as the template but the models would have to be constructed very carefully owing to the key differences observed in the topology of the human β -propeller sequences in comparison to the bacterial sequences. The bacterial sequences contain more conserved features within each repeat and display shorter and more consistent loop regions. Whereas, human sequences have 3 or 4 Ca^{2+} binding sites, the seven bacterial repeats each have a calcium binding site. One possible way to construct a 3D comparative model for the bacterial sequence would be to approach it in a stepwise manner i.e. model one β -repeat at a time and upon completion superpose all the constituent seven β -repeats and model the loop regions to connect them as a single 7-bladed β -propeller domain model furthermore this sort of 3D comparative model can also be utilized to study various interactions and perform some molecular dynamics studies as well. This 3D modelling study has not yet been performed but it is definitely a prospective project work at our lab.

5.1.3 Secondary structure prediction

Secondary structure predictions for the bacterial sequences of interest were carried out in a manner wherein each constituent repeat from the bacterial sequences (seven sets of seven repeats) was submitted to secondary structure prediction programs i.e. PHD from PredictProtein, PSIPRED and PROF. These methods were used in conjunction with each other to improve the accuracy of secondary structure predictions. As it can be seen in the alignment (Figure 14) the predictions align well with the structure of the β -propeller domain from integrins αV and αIIb . The integrin β -propeller fold is known to adopt a ‘Velcro’ fold where the last blade of the last β -repeat i.e. fourth blade of the seventh β -repeat is composed of residues located just adjacent to the first β -repeat of the first blade, thereby locking the structure together. This observation also holds true for all of the bacterial sequences, where a short β -strand was predicted just prior to the first β -repeat of the first strand. Therefore, in the absence of a 3D comparative model, the secondary structure predictions were very insightful as they helped us in studying the bacterial sequences and establishing these sequences to be of the integrin-type β -propeller domain.

5.1.4 Possible functions

At this point of time it is quite difficult to comment on the functions adopted by these β -propeller domains from the bacterial sequences. In the case of the human integrin sequences, the β -propeller domain plays a significant role in recognition and binding of ligands either directly (in conjugation with the β I-like domain) or indirectly (through the α I-domain); also, it is located towards the N-terminal region of the ectodomain of the integrin heterodimer. This is not possible in case of the bacterial sequences we retrieved from the database, since it is very likely that they are not membrane bound since they do not contain a stretch of hydrophobic residues that could indicate the presence of a transmembrane helix region and so they may adopt a functional role in the cytoplasmic region.

Another unique feature of the bacterial sequences is the presence of a well conserved Ca^{2+} -binding motif on all the constituent β -repeats indicating a possible role in calcium signaling (Tisa *et al.*, 1993; Norris *et al.*, 1996; Herbaud *et al.*, 1998; Dominguez, 2004; Zhao *et al.*, 2005). Through the course of evolution some of these Ca^{2+} -binding repeats have been lost and their number has been restricted to three or four depending upon the integrin sequence. Although integrins are not involved in calcium signaling, the three-four Ca^{2+} -binding motifs located on the loop regions between β -strands 1-2 and 3-4 are most probably involved in granting stability to the β -propeller structure. While the loss of Ca^{2+} -binding motifs on the remaining β -repeats may be attributed to acquiring flexibility in order to bind ligands or to host the α I-domain or may be for signal transduction (unlikely). But clearly, this probably could be answered by studying the bacterial and human sequences in more detail either experimentally or through molecular dynamics.

5.2 Publication II

Since the mammalian orthologues have been reported in the sequences of bony fishes and the urochordate integrins (α I-domains) do not contain the characteristic αC helix, it has been difficult to speculate on the origin of the αC containing collagen-binding α I-domains within the evolutionary framework of integrins due to physical extinctions within the early chordates coupled with absence of genomic data from the extant species resulting in a knowledge gap (Donoghue and Purnell, 2005; Huhtala *et al.*, 2005; Jonson *et al.*, 2009). Here we performed extensive genome searches to highlight sequences and fragments from organisms that diverged between urochordates and osteichthyes. Additionally, these sequences were subjected to secondary structure predictions in order to provide more insight into the evolution of collagen-binding integrins.

5.2.1 Genome searches

At the time of this study the available genome assemblies from two chordates *P. marinus* (sea lamprey, 5.9x coverage) and *C. milli* (elephant shark, 1.4x coverage) were downloaded and extensive local tBLASTn searches were performed using the I-domain regions from integrin α subunits. Furthermore, tBLASTn searches were also performed against incomplete genomes and ESTs from organisms like *E. burgeri* (inshore hagfish), *R. erinacea* (little skate), and *S. acanthias* (dogfish shark). The sea lamprey genome searches yielded three full length α I-domains (Pma_f1, Pma_f2 and Pma_f3) and four short fragments, one short fragment from the hagfish genome (Ebu_f) and the *C. milli* genome searches did not yield any unambiguously identifiable integrin α I-domain fragments.

5.2.2 Sequence alignment and secondary structure prediction

The sea lamprey sequences and the hagfish fragment were aligned with representative sequences from α subunits of human collagen-binding integrins ($\alpha 1$ and $\alpha 11$), leukocyte specific integrins (αM and αD) and *C. intestinalis* α I-domains (Cina1- $\alpha 8$) using T-COFFEE (Notredame *et al.*, 2000). The region corresponding to the αC helix in the three lamprey sequences were subjected to secondary structure prediction using PHD from PredictProtein, PSIPRED, PROF, Jpred, GOR and Porter and these methods are in consensus with the formation of an αC helix in the lamprey sequences (Figure 16). The short fragment (Ebu_f) from hagfish genome shows the presence of key MIDAS residues but rest of the sequence data towards the C-terminal region (including the αC helix) is absent, thereby making it difficult to speculate whether this fragment truly contains the signature αC helix or not. However, the sequence data that exists matches best with the immune system integrin sequences that lack the αC helix. The sequence alignment was carried out with the help of the TCOFFEE alignment program and the secondary structure elements from human integrin $\alpha 1$ I-domain were introduced on top of the alignment to guide and improve the quality of the alignment (Figure 17). The sequence identity levels of sea lamprey sequences with human α I-domains are shown in Table 6.

Table 6. Sequence identity of the sea lamprey sequences against the human integrin α I-domain sequences

	Hsa α 1	Hsa α 11	Hsa α M	Hsa α D
Pma_f1	53%	51%	33%	30%
Pma_f2	45%	49%	25%	26%
Pma_f3	44%	53%	26%	28%

5.2.3 Importance of the α C helix

Although the α C helix serves as a distinguishing marker for the collagen-binding integrins, its presence is not absolutely necessary for integrins to bind collagens (Kamata *et al.*, 1999; Tulla *et al.*, 2007). This suggests that integrins could probably bind collagens much prior to their specialization as a collagen receptor group and the α C helix evolved to serve as a marker or it is possible that the evolutionary advantage of possessing an α C helix could have something to do with the conformational change taking place in the activation of α I-domain, where unwinding of the α C helix puts additional pressure on the downward shift of the α 7 helix resulting in the downward transfer of signal to the β I-like domain.

Even though the sequence or the genomic data was limited it provided us with very useful insights and fortunately things began to change at the end of the year 2013 and beginning of the year 2014 as the genome assembly process for these two key genomes (elephant shark and sea lamprey genomes) were gaining momentum. The elephant shark genome was published at the beginning of 2014 (Venkatesh *et al.*, 2014) and the sea lamprey genome (Smith *et al.*, 2013) was also being assembled progressively thereby providing us with a very unique window of opportunity to study integrin sequences from these two genomes in greater detail.

5.3 Publication III

5.3.1 Genome searches

Two of the fragments we recovered from the *C. milli* genome displayed similarity to human integrin α I-domain sequence region from α 1 subunit: fragment AAVX01128089.1; 55 residues; 76% identical and α 2: fragment AAVX01352230.1; 55 residues; 71% identical, respectively. Another fragment showed similarity to the fifth repeat of the β -propeller domain of human integrin α 2 subunit - fragment AAVX01625876.1; 52 residues; 63%

identical. Upon publication of the elephant shark genome we performed extensive searches and recovered at least four full-length integrin α sequences that could be potential human orthologues i.e., $\alpha 1$, $\alpha 2$, $\alpha 11$ (collagen-binding clade) and αE (leukocyte clade). Additionally, our searches also led to the identification of updated version of the sea lamprey fragments: the Pma_f1 was published with two splice variants - ENSPMAP00000003339, 617 amino acids; ENSPMAP00000003342, 582 amino acids, Pma_f2 - ENSPMAP00000008300, 478 amino acids and Pma_f3 - ENSPMAP00000003839, 1099 amino acids. The Pma_f3 fragment is nearly full-length and lacks around 120 residues towards the N-terminus region of the β -propeller domain. Unfortunately, we were not able to find any updates for the hagfish fragment (Ebu_f) and it terminates just prior to the αC helix. Also, the little skate genome was slated to get underway at some point of time but we could not locate any integrins sequences or ESTs from the skate genome. Currently, the skatebase website (<http://skatebase.org>) hosts data downloads for little skate mitochondrion and a little skate genomic contig build, which may be updated in the future.

5.3.2 Phylogenetic analyses and multivariate plot

As discussed earlier in the results section we prepared a total of nine phylogenetic trees; three sets of phylogenetic trees (ML, NJ and Bayesian based) for three sequence datasets based on the length of the alignment and these trees were inferred based on pairwise distances based on either JTT matrix or WAG matrix. In addition, we also prepared 3D multivariate plots (Principle Components Analysis) in order to complement the observed tree topologies. Some key issues observed in the clustering pattern are discussed here, for example placement of the hagfish fragment (Ebu_f) close to the αL cluster in the leukocyte integrin clade. Reverse BLAST searches using the hagfish fragment returns several αL integrin sequences as top hits and this observation is supported by the multivariate plots as they suggest placement of the fragment in vicinity to the αL cluster.

Another issue we observed during our phylogenetic analyses is the clustering pattern for the leukocyte fish integrins, as it seems like the fish cluster that branches out after αE and αL must have diverged prior to the specialization of the human αM , αD and αX cluster. This could imply that the fish leukocytes annotated as αM -like, αD -like and αX -like in the sequence database are precursors to the tetrapod leukocyte integrins. Although it clearly needs to be studied in greater detail, it is worth mentioning here that the genomes of the

lobe-finned fish like the coelacanth and the lungfish will be pivotal in order to gain insight into the diversification of vertebrate leukocyte integrins.

Additionally, phylogenetic trees based on an α I-domain sequence alignment have relatively lower bootstrap values compared to the full-length sequence alignment based phylogenetic trees or the common region sequence alignment based phylogenetic trees. This is due to low similarity differences across the short length of the entire domain (in contrast to the longer sequence alignments). Despite the good quality of sequence alignment behind the α I-domain region tree, there is clearly a certain level of noise (in the tree) but the topology depicts the clustering of the sea lamprey sequences close to the collagen-binding clade despite the low bootstrap scores.

5.3.3 Structural studies

Surf2 computer program (Prof. MS Johnson, unpublished) was used with several X-ray crystal structure complexes in order to identify, tabulate and study the key interaction residues (from α I-domains) involved in recognition and binding of ligands (within 4.2Å distance of the ligand). Furthermore, these key interaction residues were compared with equivalent residues from remaining human integrin α I-domain sequences and sea lamprey sequences (Pma_f1-f3) to highlight the similarities and differences (Refer publication III: Table 2 for $\alpha 2$ I-domain-GFOGER tripeptide complex, Table S2 for $\alpha 1$ I-domain-GLOGEN tripeptide complex and Table S3 for αL I-domain-ICAM3 complex). An interesting observation is the Pma_f1 sequence, where the conserved MIDAS threonine (T221 in case of $\alpha 2$ I-domain) is replaced by an arginine (from sequence pattern 'MER'). It could be said that the glutamate located adjacent to the arginine could possibly take up the function of the threonine but it remains to be experimentally tested. Similar observations were made for the sea lamprey sequences when comparisons were done against the $\alpha 1$ I-domain-GLOGEN interactions (Table S2 in supplementary materials) and αL I-domain-ICAM3 interactions (Table S3 in supplementary materials). These results clearly indicated that the residues from the sea lamprey sequences are more similar to the residues from the collagen-binding integrin α I-domains as compared to the leukocyte integrin α I-domains and the lamprey sequences could potentially bind different collagen-types. In order to test this we performed 3D comparative modelling as well as experimental testing.

5.3.4 3D comparative modelling

The overall quality of a 3D model depends directly on the sequence identity shared between the target sequence with the sequence of a template structure and a good quality sequence alignment. In order to create good quality models for the sea lamprey sequences we considered the $\alpha 2$ I-domain X-ray crystal structure in complex with the collagen-like GFOGER tripeptide. It has a resolution of 2.1 Å and the lamprey sequences share a good level of sequence identity with the template sequence ($\alpha 2$ I-domain sequence). The Pma_f3 lamprey sequence shares 44% sequence similarity with the $\alpha 2$ I-domain sequence and deletions are found at only two position towards the C-terminal region of the domain, while Pma_f2 and Pma_f1 share 42% and 46% sequence identity with $\alpha 2$ I-domain sequence, respectively.

Additionally, the majority of the key functional residues about 10 out of 16 picked by Surf2 and shared between the Pma_f3 sequence and the integrin $\alpha 2$ I-domain sequence are identical which is higher in comparison to the other two models i.e. Pma_f2 and Pma_f1 as they share only 9 and 8 identical residues out of 16 ligand interacting residues. The absence of a threonine residue equivalent to MIDAS coordinating T221 (in $\alpha 2$ I-domain sequence) in Pma_f1 is an issue we were faced with again during the course of 3D modelling but it is not possible to substitute a threonine with an arginine and expect it to occupy the same 3D space as well as function. As discussed earlier the adjacent glutamate may step in to take over the function of the threonine but it remains to be tested.

In any case, the 3D models do provide us with insight into the potential interactions taking place behind the coordination between the sea lamprey sequences and the ligand GFOGER tripeptide. Additionally, the solid phase assay experimental work conducted by o/ur collaborators confirmed the ability of the lamprey sequences (Pma α I-domain sequences) to recognize and bind different collagen-types including collagen I and collagen II (fibrillar collagens), collagen IV (network-forming collagen), collagen IX (FACIT-collagen) and GFOGER tripeptide (for technique see Tulla *et al.*, 2008).

6. Conclusions and future directions

In publication I: we have presented evidence to highlight the origin of the integrin-type 7-bladed β -propeller domain prior to the divergence of multicellular organisms and it appears to be more regular in bacteria than the human integrin domains.

In publication II: we analysed the available genomic data (before the year 2014) from species that arose between the divergence of the urochordates and the osteichthyes in order to identify the presence of Integrin α I-domains. Here, we put forward evidence for the presence of the hallmarks of a vertebrate collagen receptor including the MIDAS motif and the α C helix in the sea lamprey *Petromyzon Marinus*.

In publication III: we have presented evidence to highlight the presence of orthologues of vertebrate integrin α I-domains in the agnathostomes and later diverging species. Advances in the genome assembly process for the sea lamprey genome and the elephant shark genome (at the beginning of the year 2014) helped clarify the evolutionary picture of I-domain containing integrins. Orthologues of the mammalian collagen-binding integrins extend from cartilaginous fish while the leukocyte receptor integrins need to be studied in greater detail. The sea lamprey fragments (Pma_f1, Pma_f2 and Pma_f3) that we identified share similarity with collagen-binding integrins. Additionally, experimental work from our collaborators confirmed that lamprey sequences recognize and bind mammalian collagens at MIDAS in a metal ion dependent manner. Therefore, Integrin α I-domains with vertebrate specific functions arose between the divergence of urochordates and appearance of jawless vertebrates.

Some of the possible future directions are:

3D modelling and structural studies of the β -propeller domain: the β -propeller domain from the integrin α subunit can be studied in greater detail at a structural level. The objective here would be to prepare 3D models for the bacterial 7-bladed β -propeller domain sequences and perform interaction as well as molecular dynamics studies in combination with experimental techniques to extend our understanding of the bacterial sequences. Additionally, phylogenetic studies can also be performed in greater detail to shed more light on the evolutionary pattern of the 7-bladed β -propeller domain from prokaryotes to higher vertebrates.

Evolution of leukocyte-specific integrins in vertebrates: the evolution of leukocyte integrins (α E, α L, α M, α D and α X) in vertebrates. The fish leukocyte clades are not direct mammalian orthologues and the objective here would be to perform a comprehensive phylogenetic analyses which can shed more light on the evolution and specialization of leukocyte integrins in bony fish as well as higher vertebrates.

Insight into the evolution of integrin heterodimerization pattern: in humans there are 24 integrin $\alpha\beta$ heterodimers that are formed from the 18 α subunits and 8 β subunits, although 144 pairs are theoretically possible. The patterning of pair is nonetheless quite complex. The objective here would be to combine details from X-ray structures, observed contacts between domains, sequence variation for individual subunits over species, ligand binding and biological function which would help define the rules for the heterodimer formation with general implications for other cell-surface receptors as well.

7. References

Adair BD, Yeager M. 2002. Three-dimensional model of the human platelet integrin $\alpha\text{IIb}\beta 3$ based on electron cryomicroscopy and X-ray crystallography. *Proc Natl Acad Sci USA*. 99:14059-14064.

Adair BD, Xiong JP, Maddock C, Goodman SL, Arnaout MA, Yeager M. 2005. Three-dimensional EM structure of the ectodomain of integrin $\alpha\text{V}\beta 3$ in a complex with fibronectin. *J Cell Biol*. 168:1109-1118.

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, Chen LX, Brandon RC, Rogers YH, Blazej RG, Champe M, Pfeiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, Gabor GL, Abril JF, Agbayani A, An HJ, Andrews-Pfannkoch C, Baldwin D, Ballew RM, Basu A, Baxendale J, Bayraktaroglu L, Beasley EM, Beeson KY, Benos PV, Berman BP, Bhandari D, Bolshakov S, Borkova D, Botchan MR, Bouck J, Brokstein P, Brottier P, Burtis KC, Busam DA, Butler H, Cadieu E, Center A, Chandra I, Cherry JM, Cawley S, Dahlke C, Davenport LB, Davies P, de Pablos B, Delcher A, Deng Z, Mays AD, Dew I, Dietz SM, Dodson K, Doup LE, Downes M, Dugan-Rocha S, Dunkov BC, Dunn P, Durbin KJ, Evangelista CC, Ferraz C, Ferriera S, Fleischmann W, Fosler C, Gabrielian AE, Garg NS, Gelbart WM, Glasser K, Glodek A, Gong F, Gorrell JH, Gu Z, Guan P, Harris M, Harris NL, Harvey D, Heiman TJ, Hernandez JR, Houck J, Hostin D, Houston KA, Howland TJ, Wei MH, Ibegwam C, Jalali M, Kalush F, Karpen GH, Ke Z, Kennison JA, Ketchum KA, Kimmel BE, Kodira CD, Kraft C, Kravitz S, Kulp D, Lai Z, Lasko P, Lei Y, Levitsky AA, Li J, Li Z, Liang Y, Lin X,

Liu X, Mattei B, McIntosh TC, McLeod MP, McPherson D, Merkulov G, Milshina NV, Mobarry C, Morris J, Moshrefi A, Mount SM, Moy M, Murphy B, Murphy L, Muzny DM, Nelson DL, Nelson DR, Nelson KA, Nixon K, Nusskern DR, Pacleb JM, Palazzolo M, Pittman GS, Pan S, Pollard J, Puri V, Reese MG, Reinert K, Remington K, Saunders RD, Scheeler F, Shen H, Shue BC, Sidén-Kiamos I, Simpson M, Skupski MP, Smith T, Spier E, Spradling AC, Stapleton M, Strong R, Sun E, Svirskas R, Tector C, Turner R, Venter E, Wang AH, Wang X, Wang ZY, Wassarman DA, Weinstock GM, Weissenbach J, Williams SM, Woodage T, Worley KC, Wu D, Yang S, Yao QA, Ye J, Yeh RF, Zaveri JS, Zhan M, Zhang G, Zhao Q, Zheng L, Zheng XH, Zhong FN, Zhong W, Zhou X, Zhu S, Zhu X, Smith HO, Gibbs RA, Myers EW, Rubin GM, Venter JC. 2000. The genome sequence of *Drosophila melanogaster*. *Science*. 287:2185-2195.

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 β -propeller. *Int J Biol Macromol*. 41:454-468.

Alonso JL, Essafi M, Xiong JP, Stehle T, Arnaout MA. 2002. Does the integrin αA domain act as a ligand for its βA domain? *Curr Biol*. 12:340-342.

Altieri DC, Agbanyo FR, Plescia J, Ginsberg MH, Edgington TS, Plow EF. 1990. A unique recognition site mediates the interaction of fibrinogen with the leukocyte integrin Mac-1 (CD11b/CD18). *J Biol Chem*. 265:12119-12122.

Anthis NJ, Wegener KL, Ye F, Kim C, Goult BT, Lowe ED, Vakonakis I, Bate N, Critchley DR, Ginsberg MH, Campbell ID. 2009. The structure of an integrin/talin complex reveals the basis of

- inside-out signal transduction. *EMBO J.* 28:3623-3632.
- Arnout MA. 1990. Structure and function of the leukocyte adhesion molecules CD11b/CD18. *Blood* 75:1037–1050.
- Arnout MA, Mahalingam B, Xiong JP. 2005. Integrin structure, allostery, and bidirectional signaling. *Annu Rev Cell Dev Biol.* 21:381-410.
- Arnout MA, Goodman SL, Xiong JP. 2007. Structure and mechanics of integrin-based cell adhesion. *Curr Opin Cell Biol.* 19:495–507.
- Askari JA, Tynan CJ, Webb SE, Martin-Fernandez ML, Ballestrem C, Humphries MJ. 2010. Focal adhesions are sites of integrin extension. *J Cell Biol.* 188:891-903.
- Auger JM, Kuijpers MJ, Senis YA, Watson SP, Heemskerk JW. 2005. Adhesion of human and mouse platelets to collagen under shear: a unifying model. *FASEB J.* 19:825-827.
- Banno A and Ginsberg MH. 2008. Integrin activation. *Biochem Soc Trans.* 36:229–234.
- Barczyk M, Carracedo S, Gullberg D. 2010. Integrins. *Cell Tissue Res.* 339:269-280.
- Barthel SR, Annis DS, Mosher DF, Johansson MW. 2006. Differential engagement of modules 1 and 4 of vascular cell adhesion molecule-1 (CD106) by integrins $\alpha 4\beta 1$ (CD49d/29) and $\alpha M\beta 2$ (CD11b/18) of eosinophils. *J Biol Chem.* 281:32175-32187.
- Beck K, Brodsky B. 1998. Supercoiled protein motifs: the collagen triple-helix and the alpha-helical coiled coil. *J Struct Biol.* 122:17-29.
- Bergelson JM, St. John NF, Kawaguchi S, Pasqualini R, Berdichevsky F, Hemier ME, Finberg RW. 1994. The I-domain is essential for echovirus-1 interaction with VLA-2. *Cell Adhes. Commun.* 2:455-464.
- Beglova N, Blacklow SC, Takagi J, Springer TA. 2002. Cysteine-rich module structure reveals a fulcrum for integrin rearrangement upon activation. *Nat Struct Biol.* 9:282-287.
- Bengtsson T, Aszodi A, Nicolae C, Hunziker EB, Lundgren-Akerlund E, Fassler R. 2005. Loss of $\alpha 10\beta 1$ integrin expression leads to moderate dysfunction of growth plate chondrocytes. *J Cell Sci.* 118:929–936.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235–242.
- Bienkowska J, Cruz M, Atiemo A, Handin R, Liddington R. 1997. The von willebrand factor A3 domain does not contain a metal ion-dependent adhesion site motif. *J Biol Chem.* 272:25162-25167.
- Bledzka K, Liu J, Xu Z, Perera HD, Yadav SP, Bialkowska K, Qin J, Ma YQ, Plow EF. 2012. Spatial coordination of kindlin-2 with talin head domain in interaction with integrin β cytoplasmic tails. *J Biol Chem.* 287:24585-24594.
- Bodary SC, McLean JW. 1990. The integrin $\beta 1$ subunit associates with the vitronectin receptor αv subunit to form a novel vitronectin receptor in a human embryonic kidney cell line. *J Biol Chem.* 265: 5938-5941.
- Briknarova K, Grishaev A, Banyai L, Tordai H, Patthy L, Llinas M. 1999. The second type II module from human matrix metalloproteinase 2: structure, function and dynamics. *Structure.* 7:1235–1245.

- Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. 1983. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comp Chem.* 4:187–217.
- Burke RD. 1999. Invertebrate integrins: Structure, function and evolution. *Int Rev Cytol.* 191:257-284.
- Bouvard D, Brakebusch C, Gustafsson E, Aszódi A, Bengtsson T, Berna A, Fässler R. 2001. Functional consequences of integrin gene mutations in mice. *Circ. Res.* 89:211-223. (put this in the text somewhere)
- Boot-Handford RP, Tuckwell DS. 2003. Fibrillar collagen: the key to vertebrate evolution? A tale of molecular incest. *Bioessays.* 25:142–151.
- Boot-Handford RP, Tuckwell DS, Plumb DA, Rock CF, Poulosom R. 2003. A novel and highly conserved collagen (pro(α 1)(XXVII)) with a unique expression pattern and unusual molecular characteristics establishes a new clade within the vertebrate fibrillar collagen family. *J Biol Chem.* 278:31067–31077.
- Brower DL, Brower SM, Hayward DC, Ball EE. 1997. Molecular evolution of integrins: genes encoding integrin α subunits from a coral and a sponge. *Proc Natl Acad Sci USA.* 94:9182-9187.
- Calderwood DA, Yan B, de Pereda JM, Alvarez BG, Fujioka Y, Liddington RC, Ginsberg MH. 2002. The phosphotyrosine binding-like domain of talin activates integrins. *J Biol Chem.* 277:21749-21758.
- Calderwood DA. 2004. Talin controls integrin activation. *Biochem Soc Trans.* 32:434-437.
- Calderwood DA, Campbell ID, Critchley DR. 2013. Talins and kindlins: partners in integrin-mediated adhesion. *Nat Rev Mol Cell Biol.* 14:503-517.
- Calzada MJ, Alvarez MV, Gonzalez-Rodriguez J. 2002. Agonist-specific structural rearrangements of integrin α Ib β 3. Confirmation of the bent conformation in platelets at rest and after activation. *J Biol Chem.* 277:39899-39908.
- Camper L, Holmvall K, Wangnerud C, Aszodi A, Lundgren-Akerlund E. 2001. Distribution of the collagen-binding integrin α 10 β 1 during mouse development. *Cell Tissue Res.* 306:107–116.
- Carrell NA, Fitzgerald LA, Steiner B, Erickson HP, Phillips DR. 1985. Structure of human platelet membrane glycoproteins IIb and IIIa as determined by electron microscopy. *J Biol Chem.* 260:1743–1749
- Chaudhuri I, Söding J, Lupas AN. 2008. Evolution of the β -propeller fold. *Proteins.* 71:795-803.
- Chen J, Diacovo TG, Grenache DG, Santoro SA, Zutter MM. 2002. The α (2) integrin subunit-deficient mouse: a multifaceted phenotype including defects of branching morphogenesis and hemostasis. *Am J Pathol.* 161:337-344.
- Chigaev A, Buranda T, Dwyer DC, Prossnitz ER, Sklar LA. 2003. FRET detection of cellular α 4 integrin conformational activation. *Biophys J.* 85:3951-3962.
- Chigaev A, Zwartz G, Graves SW, Dwyer DC, Tsuji H, Foutz TD, Edwards BS, Prossnitz ER, Larson RS, Sklar LA. 2003. α 4 β 1 integrin affinity changes govern cell adhesion. *J Biol Chem.* 278:38174-38182.
- Chigaev A, Waller A, Zwartz GJ, Buranda T, Sklar LA. 2007. Regulation of cell adhesion by affinity and conformational unbending of α 4 β 1 integrin. *J Immunol.* 178:6828-6839.

- Chin YK, Headey SJ, Mohanty B, Patil R, McEwan PA, Swarbrick JD, Mulhern TD, Emsley J, Simpson JS, Scanlon MJ. 2013. The structure of integrin $\alpha 1$ I-domain in complex with a collagen-mimetic peptide. *J Biol Chem.* 288:36796-36809.
- Chouhan B, Denesyuk A, Heino J, Johnson MS, Denessiouk K. 2011. Conservation of the human-type integrin β -propeller domain in bacteria. *PLoS One.* 6:e25069.
- Chouhan B, Denesyuk A, Heino J, Johnson MS, Denessiouk K. 2012. Evolutionary origin of the α C helix in integrins. *WASET.* 65:546-549.
- Chouhan B, Käpylä J, Denesyuk A, Denessiouk K, Heino J, Johnson MS. 2014. Early Chordate Origin of the Vertebrate Integrin α I Domains. *PLoS One.* 9:e112064.
- Chua GL, Tang XY, Amalraj M, Tan SM, Bhattacharjya S. 2011. Structures and interaction analyses of integrin α M β 2 cytoplasmic tails. *J Biol Chem.* 286:43842-43854.
- Cock JM, Sterck L, Rouzé P, Scornet D, Allen AE, Amoutzias G, Anthouard V, Artiguenave F, Aury JM, Badger JH, Beszteri B, Billiau K, Bonnet E, Bothwell JH, Bowler C, Boyen C, Brownlee C, Carrano CJ, Charrier B, Cho GY, Coelho SM, Collén J, Corre E, Da Silva C, Delage L, Delaroque N, Dittami SM, Doulebeau S, Elias M, Farnham G, Gachon CM, Gschloessl B, Heesch S, Jabbari K, Jubin C, Kawai H, Kimura K, Kloareg B, Küpper FC, Lang D, Le Bail A, Leblanc C, Lerouge P, Lohr M, Lopez PJ, Martens C, Maumus F, Michel G, Miranda-Saavedra D, Morales J, Moreau H, Motomura T, Nagasato C, Napoli CA, Nelson DR, Nyvall-Collén P, Peters AF, Pommier C, Potin P, Poulain J, Quesneville H, Read B, Rensing SA, Ritter A, Rousvoal S, Samanta M, Samson G, Schroeder DC, Ségurens B, Strittmatter M, Tonon T, Tregear JW, Valentin K, von Dassow P, Yamagishi T, Van de Peer Y, Wincker P. 2010. The Ectocarpus genome and the independent evolution of multicellularity in brown algae. *Nature.* 465:617-621.
- Cole C, Barber JD and Barton GJ. 2008. The Jpred 3 secondary structure prediction server, *Nucleic Acids Research.* 36(Web Server issue):W197-W201.
- Colombatti A, Bonaldo P. 1991. The superfamily of proteins with von Willebrand factor type A-like domains: one theme common to components of extracellular matrix, hemostasis, cellular adhesion, and defense mechanisms. *Blood.* 77:2305-2315.
- Colombatti A, Bonaldo P, Doliana R. 1993. Type A modules: interacting domains found in several non-fibrillar collagens and in other extracellular matrix proteins. *Matrix.* 13:297-306.
- Conrad C, Boyman O, Tonel G, Tun-Kyi A, Laggner U, de Fougerolles A, Koteliński V, Gardner H, Nestle FO. 2007. $\alpha 1\beta 1$ integrin is crucial for accumulation of epidermal T cells and the development of psoriasis. *Nat Med.* 13:836-842.
- Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: A sequence logo generator. *Genome Res* 14: 1188–1190.
- Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best fit models of protein evolution. *Bioinformatics.* 27:1164-1165.
- Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomaso A, Davidson B, Di Gregorio A, Gelpke M, Goodstein DM, Harafuji N, Hastings KE, Ho I, Hotta K, Huang W, Kawashima T, Lemaire P, Martinez D, Meinertzhagen IA, Nacula S, Nonaka M, Putnam N, Rash S, Saiga H, Satake M, Terry A, Yamada L, Wang HG, Awazu S, Azumi K, Boore J, Branno M, Chin-Bow S, DeSantis R, Doyle S, Francino P, Keys DN, Haga S, Hayashi H, Hino K, Imai KS, Inaba K, Kano S,

- Kobayashi K, Kobayashi M, Lee BI, Makabe KW, Manohar C, Matassi G, Medina M, Mochizuki Y, Mount S, Morishita T, Miura S, Nakayama A, Nishizaka S, Nomoto H, Ohta F, Oishi K, Rigoutsos I, Sano M, Sasaki A, Sasakura Y, Shoguchi E, Shin-i T, Spagnuolo A, Stainier D, Suzuki MM, Tassy O, Takatori N, Tokuoka M, Yagi K, Yoshizaki F, Wada S, Zhang C, Hyatt PD, Larimer F, Detter C, Doggett N, Glavina T, Hawkins T, Richardson P, Lucas S, Kohara Y, Levine M, Satoh N, Rokhsar DS. 2002. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science*. 298:2157-2167.
- DeSimone DW, Hynes RO. 1988. *Xenopus laevis* integrins. Structure and evolutionary divergence of the β subunits. *J Biol Chem*. 163: 5333-5340.
- Diamond MS, Staunton DE, de Fougères AR, Stacker SA, Garcia-Aguilar J, Hibbs ML, Springer TA. 1990. ICAM-1 (CD54): a counter-receptor for Mac-1 (CD11b/CD18). *J Cell Biol*. 111:3129-3139.
- 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.
- DiPersio CM, Hodivala-Dilke KM, Jaenisch R, Kreidberg JA, Hynes RO. 1997. $\alpha 3\beta 1$ Integrin is required for normal development of the epidermal basement membrane. *J Cell Biol*. 137: 729-742.
- Dominguez DC. 2004. Calcium signaling in bacteria. *Mol Microbiol*. 54: 291-297.
- Donoghue PC and Purnell MA. 2005. Genome duplication, extinction and vertebrate evolution. *Trends Ecol Evol*. 20:312-319.
- Ekholm E, Hankenson KD, Uusitalo H, Hiltunen A, Gardner H, Heino J, Penttinen R. 2002. Diminished callus size and cartilage synthesis in $\alpha 1\beta 1$ integrin-deficient mice during bone fracture healing. *Am J Pathol*. 160:1779-1785.
- Eble JA and Kühn K, eds. 1997. Integrin-ligand interactions. Chapman and Hall. New York. 1-273 pp.
- Edelson BT, Li Z, Pappan LK, Zutter MM. 2004. Mast cell-mediated inflammatory responses require the $\alpha 2\beta 1$ integrin. *Blood*. 103:2214-2220.
- Elemer GS, Edgington TS. 1994. Two independent sets of monoclonal antibodies define neoepitopes linked to soluble ligand binding and leukocyte adhesion functions of activated $\alpha M\beta 2$. *Circ Res*. 75:165-171
- Emsley J, King SL, Bergelson JM, Liddington RC. 1997. Crystal structure of the I-domain from integrin $\alpha 2\beta 1$. *J Biol Chem*. 272:28512-28517.
- Emsley J, Knight CG, Farndale RW, Barnes MJ, Liddington RC. 2000. Structural basis of collagen recognition by integrin $\alpha 2\beta 1$. *Cell*. 101:47-56.
- Ewan R, Huxley-Jones J, Mould AP, Humphries MJ, Robertson DL, Boot-Handford RP. 2005. The integrins of the urochordate *Ciona intestinalis* provide novel insights into the molecular evolution of the vertebrate integrin family. *BMC Evol Biol*. 13:5-31.
- Exposito JY, Garrone R. 1990. Characterization of a fibrillar collagen gene in sponges reveals the early evolutionary appearance of two collagen gene families. *Proc Natl Acad Sci USA*. 87:6669-66673.
- Exposito JY, Ouazana R, Garrone R. 1990. Cloning and sequencing of a Porifera partial cDNA coding for a short-chain collagen. *Eur J Biochem*. 190:401-406.
- Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*. 39:783-791.

- Felsenstein J. 1989. PHYLIP - Phylogeny Inference Package. *Cladistics* 5:164-166.
- Federico A, Moretti, Markus Moser, Ruth Lyck, Michael Abadier, Raphael Ruppert, Britta Engelhardt, Reinhard Fässler. 2013. Kindlin-3 regulates integrin activation and adhesion reinforcement of effector T cells. *Proc Natl Acad Sci USA*. 110:17005–17010.
- Fleming JC, Pahl HL, Gonzalez DA, Smith TF, Tenen DG. 1993. Structural analysis of the CD11b gene and phylogenetic analysis of the alpha-integrin gene family demonstrate remarkable conservation of genomic organization and suggest early diversification during evolution. *J Immunol*. 150:480-490.
- Fülöp V, Jones DT. 1999. β -proteins: structural rigidity and functional diversity. *Curr Opin Struct Biol*. 9:715-721.
- Gahmberg CG, Valmu L, Fagerholm S, Kotovuori P, Ihanus E, Tian L, Pessa-Morikawa T. 1998. Leukocyte integrins and inflammation. *Cell Mol Life Sci*. 54:549-555.
- Gardner H, Kreidberg J, Koteliensky V, Jaenisch R. 1996. Deletion of integrin $\alpha 1$ by homologous recombination permits normal murine development but gives rise to a specific deficit in cell adhesion. *Dev Biol*. 175:301-313.
- Gardner H, Broberg A, Pozzi A, Laato M, Heino J. 1999. Absence of integrin $\alpha 1\beta 1$ in the mouse causes loss of feedback regulation of collagen synthesis in normal and wounded dermis. *J Cell Sci*. 112:263-272.
- Garnotel R, Rittie L, Poitevin S, Monboisse J-C, Nguyen P, Potron G, Maquart FX, Randoux A, Gillery P. 2000. Human blood monocytes interact with type-I collagen through $\alpha X\beta 2$ integrin (CD11c-CD18, gp150-95). *J Immunol*. 164:5928-5934.
- Georges-Labouesse E, Messaddeq N, Yehia G, Cadalbert L, Dierich A, Le Meur M. 1996. Absence of integrin $\alpha 6$ leads to epidermolysis bullosa and neonatal death in mice. *Nat Genet*. 13:370-373.
- Grashoff C, Thievensen I, Lorenz K, Ussar S, Fässler R. 2004. Integrin-linked kinase: integrin's mysterious partner. *Curr Opin Cell Biol*. 16:565-571.
- Grayson MH, Van der Vieren M, Sterbinsky SA, Michael Gallatin W, Hoffman PA, Staunton DE, Bochner BS. 1999. $\alpha D\beta 2$ integrin is expressed on human eosinophils and functions as an alternative ligand for vascular cell adhesion molecule 1 (VCAM-1). *J Exp Med*. 188: 2187–2191.
- Guo W, Giancotti FG. 2004. Integrin signaling during tumour progression. *Nat Rev Mol Cell Biol*. 5:816-826.
- Harburger DS and Calderwood DA. 2009. Integrin signaling at a glance. *J Cell Sci*. 122(Pt 2):159-163.
- Heino J, Huhtala M, Kapyla J, Johnson MS. 2008. Evolution of collagen-based adhesion systems. *Int J Biochem Cell Biol*. 41:341-348.
- Herbaud ML, Guiseppi A, Denizot F, Haiech J, Kilhoffer MC. 1998. Calcium signaling in *Bacillus subtilis*. *Biochim Biophys Acta*. 1448:212–226.
- Higgins JM, Mandelbrot DA, Shaw SK, Russell GJ, Murphy EA, Chen YT, Nelson WJ, Parker CM, Brenner MB. 1998. Direct and regulated interaction of integrin $\alpha E\beta 7$ with E-cadherin. *J Cell Biol*. 140:197-210.
- Hodivala-Dilke KM, McHugh KP, Tsakiris DA, Rayburn H, Crowley D, Ullman-Culleré M, Ross FP, Collier BS, Teitelbaum S, Hynes RO. 1990. $\beta 3$ -integrin-deficient mice are a model for Glanzmann thrombasthenia showing placental defects and reduced survival. *J Clin Invest*. 103:229-38.

- Horii K, Okuda D, Morita T, Mizuno H. 2004. Crystal structure of EMS16 in complex with the integrin $\alpha 2$ -I domain. *J Mol Biol.* 341:519-527.
- Holtkötter O, Nieswandt B, Smyth N, Muller W, Hafner M, Schulte V, Krieg T, Eckes B. 2002. Integrin $\alpha 2$ -deficient mice develop normally, are fertile, but display partially defective platelet interaction with collagen. *J Biol Chem.* 277:10789-10794.
- Huang XZ, Wu JF, Ferrando R, Lee JH, Wang YL, Farese RV Jr, Sheppard D. 2000. Fatal bilateral chylothorax in mice lacking the integrin $\alpha 9\beta 1$. *Mol Cell Biol.* 20:5208-5215.
- Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics.* 17:754-755.
- Hughes AL. 1992. Coevolution of the vertebrate α - and β - chain genes. *Mol Biol Evol.* 9:216-234.
- Hughes AL. 2001. Evolution of the integrin α and β protein families. *J Mol Evol.* 52:63-72.
- Huhtala M, Heino J, Casciari D, de Luice A, Johnson MS. 2005. Integrin evolution: insights from ascidian and teleost fish genomes. *Matrix Biol.* 24:83-95.
- Huizinga EG, Martijn van der Plas R, Kroon J, Sixma JJ, Gros P. 1997. Crystal structure of the A3 domain of human von Willebrand factor: implications for collagen binding. *Structure.* 5:1147-1156.
- Hutchinson EG, Thornton JM. 1994. A revised set of potentials for β -turn formation in proteins. *Protein Sci* 3: 2207-2216.
- Huth JR, Olejniczak ET, Mendoza R, Liang H, Harris EA, Lupper ML Jr, Wilson AE, Fesik SW, Staunton DE. 2000. NMR and mutagenesis evidence for an I domain allosteric site that regulates lymphocyte function-associated antigen 1 ligand binding. *Proc Natl Acad Sci USA.* 97:5231-5236.
- Hynes RO, Zhao Q. 2000. The evolution of cell adhesion. *J Cell Biol.* 150:F89-F95.
- Hynes RO. 2002. Integrins: bidirectional, allosteric signaling machines. *Cell* 110:673-687.
- Ivaska J, Käpylä J, Pentikäinen O, Hoffren A.-M, Hermonen J, Huttunen P, Johnson MS, Heino J. 1999. A peptide inhibiting the collagen binding function of integrin $\alpha 2$ I-domain. *J Biol Chem.* 274:3513-3521.
- Iwasaki K, Mitsuoka K, Fujiyoshi Y, Fujisawa Y, Kikuchi M, Sekiguchi K, Yamada T. 2005. Electron tomography reveals diverse conformations of integrin $\alpha IIb\beta 3$ in the active state. *J Struct Biol.* 150:259-267.
- Jenkins C, KedarV, Fuerst JA. 2002. Gene discovery within the planctomycete division of the domain Bacteria using sequence tags from genomic DNA libraries. *Genome Biology.* 3:0031.1-0031.11.
- Johnson MS, Tuckwell D. 2003. Evolution of Integrin I-domains: I domains in Integrins, D Gullberg, ed., Landes Bioscience, Texas, USA, pp.1-26.
- 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.
- Johnson MS, Käpylä J, Dnessiouk K, Airene T, Heino J. 2013. Evolution of cell adhesion to cellular matrix. Keeley F, Mecham RP (eds), In: *Evolution of Extracellular Matrix*, Springer-Verlag Berlin Heidelberg.
- Johnson MS, Overington JP. 1993. A structural basis for the comparison of sequences: An

- evaluation of scoring methodologies. *J Mol Biol.* 233:716-738.
- Jokinen J, White DJ, Salmela M, Huhtala M, Käpylä J, Sipilä K, Puranen JS, Nissinen L, Kankaanpää P, Marjomäki V, Hyypiä T, Johnson MS, Heino J. 2010. Molecular mechanism of $\alpha 2\beta 1$ integrin interaction with human echovirus 1. *EMBO J.* 29:196-208.
- Jones DT. 1999. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292: 195-202.
- Kahn ML. 2004. Platelet-collagen responses: molecular basis and therapeutic promise. *Semin Thromb Hemost.* 30:419-425.
- Kallen J, Welzenbach K, Ramage P, Geyl D, Kriwacki R, Legge G, Cottens S, Weitz-Schmidt G, Hommel U. 1999. Structural basis for LFA-1 inhibition upon lovastatin binding to the CD11a I-domain. *J Mol Biol.* 292:1-9.
- Kamata T, Takada Y. 1994. Direct binding of collagen to the I-domain of integrin $\alpha 2\beta 1$ (VLA-2, CD49b/CD29) in a divalent cation-independent manner. *J Biol Chem.* 269:26006-26010.
- Käpylä J, Ivaska J, Riikonen R, Nykvist P, Pentikäinen O, Johnson M, et al. 2000. Integrin $\alpha 2$ I-domain recognizes type I and type IV collagens by different mechanisms. *J Biol Chem.* 275:3348-3354.
- Kern A, Eble J, Golbik R, Kuhn K. 1993. Interaction of type IV collagen with the isolated integrins $\alpha 1\beta 1$ and $\alpha 2\beta 1$. *Eur J Biochem.* 215:151-159.
- Kim M, Carman CV, Springer TA. 2003. Bidirectional transmembrane signaling by cytoplasmic domain separation in integrins. *Science.* 301:1720-1725.
- King SL, Kamata T, Cunningham JA, Emsley J, Liddington RC, Takada Y, Bergelson JM. 1997. Echovirus 1 interaction with the human very late antigen-2 (integrin $\alpha 2\beta 1$) I domain. Identification of two independent virus contact sites distinct from the metal ion-dependent adhesion site. *J Biol Chem.* 272:28518-28522.
- King N, Westbrook MJ, Young SL, Kuo A, Abedin M, Chapman J, et al., 2008. The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature.* 451:783-788.
- Kishimoto TK, O'Connor K, Lee A, Roberts TM, Springer TA. 1987. Cloning the subunit of the leukocyte adhesion proteins: homology to an extracellular matrix receptor defines a novel supergene family. *Cell.* 48:681-690
- Knack BA, Iguchi A, Shinzato C, Hayward DC, Ball EE, Miller DJ. 2008. Unexpected diversity of cnidarian integrins: expression during coral gastrulation. *BMC Evol Biol.* 8:136.
- Knight CG, Morton LF, Onley DJ, Peachey AR, Messent AJ, Smethurst PA, et al., 1998. Identification in collagen type I of an integrin $\alpha 2\beta 1$ -binding site containing an essential GER sequence. *J Biol Chem.* 273: 33287-33294.
- Knight CG, Morton LF, Peachey AR, Tuckwell DS, Farndale RW, Barnes MJ. 2000. The collagen-binding A-domains of integrins $\alpha 1\beta 1$ and $\alpha 2\beta 1$ recognize the same specific amino acid sequence GFOGER in native (triplehelical) collagens. *J Biol Chem.* 275:35-40.
- Kolanus W1, Nagel W, Schiller B, Zeitlmann L, Godar S, Stockinger H, Seed B. 1996. $\alpha L\beta 2$

- integrin/LFA-1 binding to ICAM-1 induced by cytohesin-1, a cytoplasmic regulatory molecule. *Cell*. 86:233-242
- Kreidberg JA, Donovan MJ, Goldstein SL, Rennke H, Shepherd K, Jones RC, Jaenisch R. 1996. $\alpha 3\beta 1$ integrin has a crucial role in kidney and lung organogenesis. *Development*. 122: 3537-3547.
- Kyöstilä K, Lappalainen AK, Lohi H. 2013. Canine Chondrodysplasia Caused by a Truncating Mutation in Collagen-Binding Integrin Alpha Subunit 10. *PLoS One*. 2013; 8(9): e75621.
- Lane NE, Yao W, Nakamura MC, Humphrey MB, Kimmel D, Huang X, Sheppard D, Ross FP, Teitelbaum SL. 2005. Mice lacking the integrin $\beta 5$ subunit have accelerated osteoclast maturation and increased activity in the estrogen-deficient state. *J Bone Miner Res*. 20: 58-66.
- Larson RS, Corbi AL, Berman L, Springer T. 1989. Primary structure of the leukocyte function-associated molecule-1 α subunit: an integrin with an embedded domain defining a protein superfamily. *J Cell Biol*. 108:703-712.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. ClustalW and ClustalX version 2. *Bioinformatics*. 23:2947-2948.
- Lau TL, Kim C, Ginsberg MH, Ulmer TS. 2009. The structure of the integrin $\alpha \text{IIb}\beta 3$ transmembrane complex explains integrin transmembrane signaling. *EMBO J*. 28:1351-1361.
- Legge GB, Kriwacki RW, Chung J, Hommel U, Ramage P, Case DA, Dyson HJ, Wright PE. 2000. NMR solution structure of the inserted domain of human leukocyte function associated antigen-1. *J Mol Biol*. 295:1251-1264.
- Lehtonen JV, Still DJ, Rantanen VV, Ekholm J, Björklund D, Iftikhar Z, Huhtala M, Repo S, Jussila A, Jaakkola J, Pentikäinen O, Nyrönen T, Salminen T, Gyllenberg M, Johnson MS. 2004. BODIL: a molecular modeling environment for structure-function analysis and drug design. *J Comput Aided Mol Des*. 18:401-419.
- Leitinger B, Hohenester E. 2007. Mammalian collagen receptors. *Matrix Biol*. 26:146-55.
- Leitinger B. 2011. Transmembrane collagen receptors. *Annu Rev Cell Dev Biol*. 27:265-90
- Lee J-O, Rieu P, Arnaout MA, Liddington R. 1995. Crystal structure of the A domain from the α subunit of integrin CR3 (CD11b/CD18). *Cell*. 80:631-638.
- Lee J-O, Rieu P, Arnaout MA, Liddington R. 1995a. Crystal structure of the A domain from the α subunit of integrin CR3. *Cell*. 80:631-638.
- Lee J-O, Bankston LA, Arnaout MA, Liddington RC. 1995b. Two conformations of the integrin A-domain (I-domain): a pathway for activation? *Structure*. 3:1333-1340.
- Legate KR, Wickström SA, Fässler R. 2009. Genetic and cell biological analysis of integrin outside-in signaling. *Genes Dev*. 23:397-418.
- Legate KR, Fässler R. 2009. Mechanisms that regulate adoptor binding to β -integrin cytoplasmic tails. *J Cell Sci*. 122(P2):187-198.
- Li Z, Xi X, Du X. 2001. A mitogen-activated protein kinase-dependent signaling pathway in the activation of platelet integrin $\alpha \text{IIb}\beta 3$. *J Biol Chem*. 2001 276:42226-42232.
- Li Z, Zhang G, Feil R, Han J, Du X. 2005. Sequential activation of p38 and ERK pathways by cGMP-dependent protein kinase leading to

- activation of the platelet integrin α Ib β 3. *Blood*. 107:965-972.
- Lin CQ and Bissell MJ. 1993. Multi-faceted regulation of cell differentiation by extracellular matrix. *FASEB J*. 7:737-743
- Littlewood Evans A, Müller U. 2000. Stereocilia defects in the sensory hair cells of the inner ear in mice deficient in integrin α 8 β 1. *Nat Genet*. 24: 424-428.
- Litvinov RI, Shuman H, Bennett JS, Weisel JW. 2004. Binding strength and activation state of single fibrinogen-integrin pairs on living cells. *Proc Natl Acad Sci USA*. 99:7426-7431.
- Loftus JC, Smith JW, Ginsberg MH. 1994. Integrin-mediated cell adhesion: the extracellular face. *J Biol Chem*. 269:25235-25238.
- Lu C, Ferzly M, Takagi J, Springer TA. 2001. Epitope mapping of antibodies to the C-terminal region of the integrin β 2 subunit reveals regions that become exposed upon receptor activation. *J Immunol*. 166:5629-5637.
- Lundgren-Åkerlund E, Aszòdi A. 2014. Integrin α 10 β 1: a collagen receptor critical in skeletal development. *Adv Exp Med Biol*. 819:61-71.
- Luo BH, Springer TA, Takagi J. 2003. Stabilizing the open conformation of the integrin headpiece with a glycan wedge increases affinity for ligand. *Proc Natl Acad Sci USA*. 100:2403-2408.
- Luo BH, Takagi J, Springer TA. 2004. Locking the β 3 integrin I-like domain into high and low affinity conformations with disulfides. *J Biol Chem*. 279:10215-10221.
- Luo BH, Springer TA, Takagi J. 2004. A specific interface between integrin transmembrane helices and affinity for ligand. *PLoS Biol*. 2:e153.
- Luo BH, Carman CV, Springer TA. 2007. Structural basis of integrin regulation and signaling. *Annu Rev Immunol*. 25: 619–647.
- Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH. 2015. CDD: NCBI's conserved domain database. *Nucleic Acids Res*. 43(Database issue):D222-226.
- Mayer U, Saher G, Fässler R, Bornemann A, Echtermeyer F, von der Mark H, Miosge N, Pöschl E, von der Mark K. 1997. Absence of integrin α 7 causes a novel form of muscular dystrophy. *Nat Genet*. 17: 318-323.
- McCleverty CJ, Liddington RC. 2003. Engineered allosteric mutants of the integrin α M β 2 I-domain: structural and functional studies. *Biochem J*. 372:121-127.
- McHugh KP, HodiVala-Dilke K, Zheng MH, Namba N, Lam J, Novack D, Feng X, Ross FP, Hynes RO, Teitelbaum SL. 2000. Mice lacking β 3 integrins are osteosclerotic because of dysfunctional osteoclasts. *J Clin Invest*. 105:433-40.
- McGuffin LJ, Bryson K and Jones DT. 2000. The PSIPRED protein structure prediction server. *Bioinformatics* 16:404-405.
- Miyazawa S, Azumi K, Nonaka M. 2001. Cloning and characterization of integrin α subunits from the solitary ascidian, *Halocynthia roretzi*. *J Immunol*. 166:1710-1715.
- Montanez E, Ussar S, Schifferer M, Bösl M, Zent R, Moser M, Fässler R. 2008. Kindlin-2 controls bidirectional signaling of integrins. *Genes Dev*. 22:1325-1330.

- Moser M, Legate KR, Zent R, Fässler R. 2009. The tail of integrins, talin and kindlins. *Science*. 324:895-899.
- Moser M, Nieswandt B, Ussar S, Pozgajova M, Fässler R. 2008. Kindlin-3 is essential for integrin activation and platelet aggregation. *Nat Med*. 14:325-330.
- Mould AP, Travis MA, Barton SJ, Hamilton JA, Askari JA, Craig SE, Macdonald PR, Kammerer RA, Buckley PA, Humphries MJ. 2005. Evidence that monoclonal antibodies directed against the integrin β -subunit plexin/semaphorin/integrin domain stimulate function by inducing receptor extension. *J Biol Chem*. 280:4238-4246.
- Myllyharju J, Kivirikko KI. 2004. Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends Genet*. 20:33-43.
- 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.
- Munger JS, Huang X, Kawakatsu H, Griffiths MJ, Dalton SL, Wu J, Pittet JF, Kaminski N, Garat C, Matthay MA, Rifkin DB, Sheppard D. 1999. The integrin α v β 6 binds and activates latent TGF β 1: a mechanism for regulating pulmonary inflammation and fibrosis. *Cell*. 96: 319-328.
- Müller WE. 1997. Origin of metazoan adhesion molecules and adhesion receptors as deduced from cDNA analyses in the marine sponge *Geodia cydonium*: a review. *Cell Tissue Res*. 289:383-395.
- Müller U, Wang D, Denda S, Meneses JJ, Pedersen RA, Reichardt LF. 1997. Integrin α 8 β 1 is critically important for epithelial-mesenchymal interactions during kidney morphogenesis. *Cell* 88: 603-613.
- Nagae M, Re S, Mihara E, Nogi T, Sugita Y, Takagi J. 2012. Crystal structure of α 5 β 1 integrin ectodomain: atomic details of the fibronectin receptor. *J Cell Biol*. 197:131-140.
- Nandrot EF, Kim Y, Brodie SE, Huang X, Sheppard D, Finnemann SC. 2004. Loss of synchronized retinal phagocytosis and age-related blindness in mice lacking α v β 5 integrin. *J Exp Med*. 200: 1539-1545.
- Nermut MV, Green NM, Eason P, Yamada SS, Yamada KM. 1988. Electron microscopy and structural model of human fibronectin receptor. *EMBO* 7:4093-4099.
- Nichols SA, Dirks W, Pearse JS, King N. 2006. Early evolution of animal cell signaling and adhesion genes. *Proc Natl Acad Sci USA*. 103:12451-12456.
- Nishida N, Xie C, Shimaoka M, Cheng Y, Walz T, Springer TA. 2006. Activation of leukocyte β 2 integrins by conversion from bent to extended conformations. *Immunity*. 25:583-594.
- Nishida N, Sumikawa H, Sakakura M, Shimba N, Takahashi H, Terasawa H, Suzuki E, Shimada I. 2003. Collagen-binding mode of vWF-A3 domain determined by a transferred cross-saturation experiment. *Nat Struct Biol*. 10:53-58.
- Noti JD, Johnson AK, Dillon JD. 2000. Structural and functional characterization of the leukocyte integrin gene CD11d. Essential role of Sp1 and Sp3. *J Biol Chem*. 275:8959-8969.
- Norris V, Grant S, Freestone P, Canvin J, Sheikh FN, et al. 1996. Calcium signaling in bacteria. *J Bacteriol*. 178:3677-3682.
- Notredame C, Higgins DG, Heringa J. 2000. T-Coffee: A novel method for multiple sequence alignments. *J Mol Biol*. 302:205-217.

- Nymalm Y, Puranen JS, Nyholm TK, Käpylä J, Kidron H, Pentikäinen OT, Airene TT, Heino J, Slotte JP, Johnson MS, Salminen TA. 2004. Jararhagin-derived RKKH peptides induce structural changes in $\alpha 1$ I-domain of human integrin $\alpha 1\beta 1$. *J Biol Chem.* 279:7962-7970.
- Ouali M, King RD. 2000. Cascaded multiple classifiers for secondary structure prediction. *Protein Sci.* 9: 1162–1176.
- Pancer Z, Kruse M, Müller I, Müller WE. 1997. On the origin of Metazoan adhesion receptors: cloning of integrin α subunit from the sponge *Geodia cydonium*. *Mol Biol Evol.* 14:391-398.
- Partridge AW, Liu S, Kim S, Bowie JU, Ginsberg MH. 2005. Transmembrane domain helix packing stabilizes integrin $\alpha IIb\beta 3$ in the low affinity state. *J Biol Chem.* 280:7294-72300.
- Pentikäinen O, Hoffrén A-M, Ivaska J, Käpylä J, Nyrönen T, Heino J, Johnson MS. 1999. RKKH peptides from the snake venom metalloproteinase of *Bothrops jararaca* bind near the MIDAS site of the human integrin $\alpha 2$ I-domain. *J Biol Chem.* 274:31493-31505.
- Prockop DJ, Kivirikko KI. 1995. Collagens: molecular biology, diseases, and potentials for therapy. *Annu Rev Biochem.* 64:403-434.
- Pollastri G and McLysaght A. 2005. Porter: a new, accurate server for protein secondary structure prediction, *Bioinformatics* 21:1719-1720.
- Ponting CP, Aravind L, Schultz J, Bork P, Koonin EV. 1999. Eukaryotic signaling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer. *J Mol Biol.* 289:729-745.
- Popova SN, Barczyk M, Tiger CF, Beertsen W, Zigrino P, Aszodi A, Miosge N, Forsberg E, Gullberg D. 2007. Alpha11 betal integrin-dependent regulation of periodontal ligament function in the erupting mouse incisor. *Mol Cell Biol.* 27:4306-4316.
- Pozzi A, Wary KK, Giancotti FG, Gardner HA. 1998. Integrin $\alpha 1\beta 1$ mediates a unique collagen-dependent proliferation pathway in vivo. *J Cell Biol.* 142:587-594.
- Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu JK, Benito-Gutierrez EL, Dubchak I, Garcia-Fernandez J, Gibson-Brown JJ, Grigoriev IV, Horton AC, de Jong PJ, Jurka J, Kapitonov VV, Kohara Y, Kuroki Y, Lindquist E, Lucas S, Osoegawa K, Pennacchio LA, Salamov AA, Satou Y, Sauka-Spengler T, Schmutz J, Shin-I T, Toyoda A, Bronner-Fraser M, Fujiyama A, Holland LZ, Holland PW, Satoh N, Rokhsar DS. 2008. The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453:1064-1071.
- Pytela R, Pierschbacher MD, Ginsberg MH, Plow EF, Ruoslahti E. 1986. Platelet membrane glycoprotein IIb/IIIa: member of a family of 'Arg-Gly-Asp' specific adhesion receptors. *Science.* 231: 1559-156.
- Pytela R, Pierschbacher MD, Ruoslahti E. 1985. Identification and isolation of a 140 kd cell surface glycoprotein with properties expected of a fibronectin receptor. *Cell.* 40: 191-198.
- Quistgaard EM, Thirup SS. 2009. Sequence and structural analysis of the Asp-box motif and Asp-box β -propellers; a widespread propeller-type characteristic of the Vps10 domain family and several glycosidehydrolase families. *BMC Struct Biol.* 9:46.
- Qu A, Leahy DJ. 1995. Crystal structure of the I-domain from the CD11a/CD18 (LFA-1, $\alpha L\beta 2$) integrin. *Proc Natl Acad Sci USA.* 92:10277-10281.

- Reber-Muller S, Studer R, Muller P, Yanze N, Schmid V. 2001. Integrin and talin in the jellyfish *Podocoryne carnea*. *Cell Biol Int*. 25:753-769.
- Reed NI, Jo H, Chen C, Tsujino K, Arnold TD, DeGrado WF, Sheppard D. 2015. New information about expression in vivo: The $\alpha\beta 1$ integrin plays a critical in vivo role in tissue fibrosis. *Sci Transl Med*. 7:288ra79
- Ricard-Blum S, Ruggiero F. 2005. The collagen superfamily: from the extracellular matrix to the cell membrane. *Pathol Biol (Paris)*. 53:430-442.
- Ridley AJ, Schwartz MA, Burridge K, Firtel RA, Ginsberg MH, Borisy G, Parsons JT, Horwitz AR. 2003. Cell migration: integrating signals from front to back. *Science*. 302:1704-1709.
- Rich RL, Deivanayagam CC, Owens RT, Carson M, Höök A, Moore D, Symersky J, Yang VW, Narayana SV, Höök M. 1999. Trench-shaped binding sites promote multiple classes of interactions between collagen and the adherence receptors, $\alpha 1\beta 1$ integrin and *Staphylococcus aureus* cna MSCRAMM. *J Biol Chem*. 274:24906-24913.
- Richardson JS. 1981. The anatomy and taxonomy of protein structure. *Adv Protein Chem*. 34:167-339.
- Rigden DJ, Galperin MY. 2004. The DxDxDG motif for calcium binding: multiple structural contexts and implications for evolution. *J Mol Biol*. 343:971-984.
- Romijn RAP, Bouma B, Wuyster W, Gros P, Kroon J, Sixma JJ, Huizinga EG. 2001. Identification of the collagen-binding site of the von Willebrand factor A3-domain. *J Biol Chem*. 276:9985-9991.
- Rost B and Sander C. 1994. Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins*. 19:55-72.
- Rost B, Yachdav G, Liu J. 2004. The PredictProtein server. *Nucleic Acids Res*. 32:W321-W326.
- Sali A, Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*. 234:779-815.
- Sasakura Y, Shoguchi E, Takatori N, Wada S, Meinertzhagen IA, Satou Y, Satoh N. A genomewide survey of developmentally relevant genes in *Ciona intestinalis*. 2003. Genes for cell junctions and extracellular matrix. *Dev Genes Evol*. 213:303-313.
- Schierwater B, Eitel M, Jakob W, Osigus HJ, Hadrys H, Dellaporta SL, Kolokotronis SO, Desalle R. 2009. Concatenated analysis sheds light on early metazoan evolution and fuels a modern “urmetazoon” hypothesis. *PLoS Biol* 7:e20.
- 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 USA*. 107:10142-10147.
- Senger DR, Perruzzi CA, Streit M, Kotliansky VE, de Fougères AR, Detmar M. 2002. The $\alpha 1\beta 1$ and $\alpha 2\beta 1$ integrins provide critical support for vascular endothelial growth factor signaling, endothelial cell migration, and tumor angiogenesis. *Am J Pathol*. 160:195-204.
- Sen TZ, Jernigan RL, Garnier J, Kloczkowski A. 2005. GOR V server for protein secondary structure prediction. *Bioinformatics* 21:2787-2788.
- Shattil SJ, Newman PJ. 2004. Integrins: dynamic scaffolds for adhesion and signaling in platelets. *Blood*. 104:1606-1615.

- Shi M, Sundramurthy K, Liu B, Tan SM, Law SK, Lescar J. 2005. The crystal structure of the plexin-semaphorin-integrin domain/hybrid domain/IEGF1 segment from the human integrin $\beta 2$ subunit at 1.8 Å resolution. *J Biol Chem.* 280:30586-30593.
- Shi M, Foo SY, Tan SM, Mitchell EP, Law SK, Lescar J. 2007. A structural hypothesis for the transition between bent and extended conformations of the leukocyte $\beta 2$ integrins. *J Biol Chem.* 282:30198-30206.
- Shimaoka M, Springer TA. 2003. Therapeutic antagonists and conformational regulation of integrin function. *Nat Rev Drug Discov.* 2:703-716.
- Shimaoka M, Xiao T, Liu JH, Yang Y, Dong Y, Jun CD, McCormack A, Zhang R, Joachimiak A, Takagi J, Wang JH, Springer TA. 2003. Structures of the αL I-domain and its complex with ICAM-1 reveal a shape-shifting pathway for integrin regulation. *Cell.* 112:99-111.
- Shin S, Wolgamott L, Yoon SO. 2012. Integrin trafficking and tumor progression. *Int J Cell Biol.* 516789.
- Siljander, PR, Hamaia S, Peachey AR, Slatter DA, Smethurst PA, Ouwehand WH, Knight CG, Farndale RW. 2004a. Integrin activation state determines selectivity for novel recognition sites in fibrillar collagens. *J Biol Chem.* 279:47763-47772.
- Siljander PR, Munnix IC, Smethurst PA, Deckmyn H, Lindhout T, Ouwehand WH, Farndale RW, Heemskerk JW. 2004b. Platelet receptor interplay regulates collagen-induced thrombus formation in flowing human blood. *Blood.* 103:1333-1341.
- Sillitoe I, Lewis, TE, Cuff AL, Das S, Ashford P, Dawson NL, Furnham N, Laskowski RA, Lee D, Lees J, Lehtinen S, Studer R, Thornton JM, Orengo CA. 2015. CATH: comprehensive structural and functional annotations for genome sequences. *Nucleic Acids Res.* 43(Database issue):D376-381.
- Smith JJ, Kuraku S, Holt C, Sauka-Spengler T, Jiang N, Campbell MS, Yandell MD, Manousaki T, Meyer A, Bloom OE, Morgan JR, Buxbaum JD, Sachidanandam R, Sims C, Garruss AS, Cook M, Krumlauf R, Wiedemann LM, Sower SA, Decatur WA, Hall JA, Amemiya CT, Saha NR, Buckley KM, Rast JP, Das S, Hirano M, McCurley N, Guo P, Rohner N, Tabin CJ, Piccinelli P, Elgar G, Ruffier M, Aken BL, Searle SM, Muffato M, Pignatelli M, Herrero J, Jones M, Brown CT, Chung-Davidson YW, Nanlohy KG, Libants SV, Yeh CY, McCauley DW, Langeland JA, Pancer Z, Fritzsich B, de Jong PJ, Zhu B, Fulton LL, Theising B, Flicek P, Bronner ME, Warren WC, Clifton SW, Wilson RK, Li W. 2013. Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. *Nat Genet.* 45:415-21.
- Solovjov D, Pluskota E, Plow E. 2005. Distinct roles for the α and β subunits in the functions of integrin $\alpha M\beta 2$. *J Biol Chem.* 280:1336-1345.
- Song G, Yang Y, Liu JH, Casanovas JM, Shimaoka M, Springer TA, Wang JH. 2005. An atomic resolution view of ICAM recognition in a complex between the binding domains of ICAM-3 and integrin $\alpha L\beta 2$. *Proc Natl Acad Sci USA.* 102:3366-3371.
- Springer TA, Zhu J, Xiao T. 2008. Structural basis for distinctive recognition of fibrinogen gammaC peptide by the platelet integrin $\alpha IIb\beta 3$. *J Cell Biol.* 182:791-800.

- Srivastava M, Begovic E, Chapman J, Putnam NH, Hellsten U, Kawashima T, Kuo A, Mitros T, Salamov A, Carpenter ML, Signorovitch AY, Moreno MA, Kamm K, Grimwood J, Schmutz J, Shapiro H, Grigoriev IV, Buss LW, Schierwater B, Dellaporta SL, Rokhsar DS. 2008. The Trichoplax genome and nature of placozoans. *Nature*. 454:955-960.
- Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier ME, Mitros T, Richards GS, Conaco C, Dacre M, Hellsten U, Larroux C, Putnam NH, Stanke M, Adamska M, Darling A, Degnan SM, Oakley TH, Plachetzki DC, Zhai Y, Adamski M, Calcino A, Cummins SF, Goodstein DM, Harris C, Jackson DJ, Leys SP, Shu S, Woodcroft BJ, Vervoort M, Kosik KS, Manning G, Degnan BM, Rokhsar DS. 2010. The Amphimedon queenslandica genome and the evolution of animal complexity. *Nature*. 466:720-726.
- Stepp MA, Spurr-Michaud S, Tisdale A, Elwell J, Gipson IK. 1990. $\alpha 6\beta 4$ integrin heterodimer is a component of hemidesmosomes. *Proc Natl Acad Sci U S A* 1990; 87: 8970-8974.
- Tadokoro S, Shattil SJ, Eto K, Tai V, Liddington RC, dePereda JM, Ginsberg MH, Calderwood DA. 2003. Talin binding to integrin beta tails: a final common step in integrin activation. *Science* 302:103-106.
- Takada Y, Ye X, Simon S. 2007. The integrins. *Genome Biol*. 8:215.
- Takagi J, Springer TA. 2002. Integrin activation and structural rearrangement. *Immunol Rev*. 186:141-163.
- Takagi J, Petre BM, Walz T, Springer TA. 2002. Global conformational rearrangements in integrin extracellular domains in outside-in and inside-out signaling. *Cell*. 110:599-511.
- Takagi J, Strokovich K, Springer TA, Walz T. 2003. Structure of integrin $\alpha 5\beta 1$ in complex with fibronectin. *EMBO J*. 22:4607-4615.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 28:2731-2739.
- Teitelbaum SL. 2005. Osteoporosis and integrins. *J Clin Endocrinol Metab*. 90:2466-2468.
- Tiger CF, Fougerousse F, Grundstrom G, Velling T, Gullberg D. 2001. $\alpha 11\beta 1$ integrin is a receptor for interstitial collagens involved in cell migration and collagen reorganization on mesenchymal nonmuscle cells. *Dev Biol*. 237:116-129.
- 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.
- Tng E, Tan SM, Ranganathan S, Cheng M, Law SK. 2004. The integrin $\alpha L\beta 2$ hybrid domain serves as a link for the propagation of activation signal from its stalk regions to the I-like domain. *J Biol Chem*. 279:54334-54339.
- Tuckwell D, Calderwood DA, Green LJ, Humphries MJ. 1995. Integrin $\alpha 2$ I-domain is a binding site for collagens. *J Cell Sci*. 108:1629-1637.
- Tulla M, Pentikainen OT, Viitasalo T, Kapyla J, Impola U, Nykvist P, Nissinen L, Johnson MS, Heino J. 2001. Selective binding of collagen subtypes by integrin $\alpha 11$, $\alpha 21$, and $\alpha 101$ domains. *J Biol Chem*. 276: 48206-48212

- Tulla M, Huhtala M, Jääliñoja J, Käpylä J, Farndale RW, Ala-Kokko L, Johnson MS, Heino J. 2007. Analysis of an ascidian integrin provides new insight into early evolution of collagen recognition. *FEBS Lett.* 581:2434-2440.
- Tulla M, Lahti M, Puranen JS, Brandt AM, Käpylä J, et al. 2008. Effects of conformational activation of integrin $\alpha 11$ and $\alpha 21$ domains on selective recognition of laminin and collagen subtypes. *Exp Cell Res.* 314:1734-1743.
- Ulanova M, Gravelle S, Barnes R. 2009. The Role of Epithelial Integrin Receptors in Recognition of Pulmonary Pathogens. *J Innate Immun.* 1:4-17.
- Ulmer TS, Yaspan B, Ginsberg MH, Campbell ID. 2001. NMR analysis of structure and dynamics of the cytosolic tails of integrin $\alpha 11\beta 3$ in aqueous solution. *Biochemistry.* 40:7498-7508.
- Van der Vieren M, Crowe DT, Hoekstra D, Vazeux R, Hoffman PA, Grayson MH, Bochner BS, Gallatin WM, Staunton DE. 1999. The leukocyte integrin alpha D beta 2 binds VCAM-1: evidence for a binding interface between I domain and VCAM-1. *J Immunol.* 163:1984-1990
- Van der Vieren M, Crowe DT, Hoekstra D, Vazeux R, Hoffman PA, Grayson MH, Bochner BS, Gallatin WM, Staunton DE. 1999. The leukocyte integrin $\alpha D\beta 2$ binds VCAM-1: evidence for a binding interface between I domain and VCAM-1. *J Immunol.* 163:1984-1990.
- Van der Vieren M, Le Trong H, Wood CL, Moore PF, St John T, Staunton DE, Gallatin WM. 1996. A novel leukointegrin, $\alpha D\beta 2$, binds preferentially to ICAM-3. *Immunity.* 3:683-90.
- Venkatachalam CM. 1968. Stereochemical criteria for polypeptides and proteins. Conformation of a system of three linked peptide units. *Biopolymers.* 6:1425-1436.
- Venkatesh B, Lee AP, Ravi V, Maurya AK, Lian MM, Swann JB, Ohta Y, Flajnik MF, Sutoh Y, Kasahara M, Hoon S, Gangu V, Roy SW, Irimia M, Korzh V, Kondrychyn I, Lim ZW, Tay BH, Tohari S, Kong KW, Ho S, Lorente-Galdos B, Quilez J, Marques-Bonet T, Raney BJ, Ingham PW, Tay A, Hillier LW, Minx P, Boehm T, Wilson RK, Brenner S, Warren WC. 2014. Elephant shark genome provides unique insights into gnathostome evolution. *Nature.* 505:174-179.
- Vinogradova O, Haas T, Plow EF, Qin J. 2000. A structural basis for integrin activation by the cytoplasmic tail of the $\alpha 11b$ subunit. *Proc Natl Acad Sci USA.* 97:1450-1455.
- Vinogradova O, Velyvis A, Velyviene A, Hu B, Haas T, Plow E, Qin J. 2002. A structural mechanism of integrin $\alpha 11\beta 3$ "inside-out" activation as regulated by its cytoplasmic face. *Cell.* 10:587-597.
- Vinogradova O, Vaynberg J, Kong X, Haas TA, Plow EF, Qin J. 2004. Membrane-mediated structural transitions at the cytoplasmic face during integrin activation. *Proc Natl Acad Sci USA.* 101:4094-4099.
- Vlahakis NE, Young BA, Atakilit A, Sheppard D. 2005. The lymphangiogenic vascular endothelial growth factors VEGF-C and -D are ligands for the integrin $\alpha 9\beta 1$. *J Biol Chem.* 280:4544-4552.
- Vogel BE, Tarone G, Giancotti FG, Gailit J, Ruoslahti E. 1990. A novel fibronectin receptor with an unexpected subunit composition ($\alpha v\beta 1$). *J Biol Chem.* 265: 5934-5937.
- Vorup-Jensen T, Ostermeier C, Shimaoka M, Hommel U, Springer TA. 2003. Structure and allosteric regulation of the $\alpha X\beta 2$ integrin I-domain. *Proc Natl Acad Sci USA.* 100:1873-1878.

- Watanabe N, Bodin L, Pandey M, Krause M, Coughlin S, Boussiotis VA, Ginsberg MH, Shattil SJ. 2008. Mechanisms and consequences of agonist-induced talin recruitment to platelet integrin α IIb β 3. *J Cell Biol.* 181:1211-1222.
- Wegener KL, Campbell ID. 2008. Transmembrane and cytoplasmic domains in integrin activation and protein-protein interactions. *Mol Membr Biol.* 25:376-387.
- Weljie AM, Hwang PM, Vogel HJ. 2002. Solution structures of the cytoplasmic tail complex from platelet integrin α IIb and β 3 subunits. *Proc Natl Acad Sci USA.* 99:5878-5883.
- Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum likelihood approach. *Mol Biol Evol.* 18:691-699.
- Whittaker CA, Hynes RO. 2002. Distribution and evolution of von Willebrand/integrin A domains: widely dispersed domains with roles in cell adhesion and elsewhere. *Mol Biol Cell.* 13:3369-3387.
- Wimmer W, Blumbach B, Diehl-Seifert B, Koziol C, Batel R, Steffen R, *et al.*, 1999. Increased expression of integrin and receptor tyrosine kinase genes during autograft fusion in the sponge *Geodia cydonium*. *Cell Adhes Commun.* 7:111-24.
- Wu JE, Santoro SA. 1994. Complex patterns of expression suggest extensive roles for the α 2 β 1 integrin in murine development. *Dev Dyn.* 199:292-314.
- Xiao T, Takagi J, Collier BS, Wang JH, Springer TA. 2004. Structural basis for allostery in integrins and binding to fibrinogen-mimetic therapeutics. *Nature.* 432:59-67.
- Xie C, Zhu J, Chen X, Mi L, Nishida N, Springer TA. 2010. Structure of an integrin with α I domain, complement receptor type 4. *EMBO J.* 29:666-679.
- Xing L, Huhtala M, Pietiäinen V, Käpylä J, Vuorinen K, Marjomäki V, Heino J, Johnson MS, Hyypiä T, Cheng RH. 2004. Structural and functional analysis of integrin α 2I domain interaction with echovirus 1. *J Biol Chem.* 279:11632-11638.
- Xiong JP, Stehle T, Diefenbach B, Zhang R, Dunker R, Scott DL, Joachimiak A, Goodman SL, Arnaout MA. 2001. Crystal structure of the extracellular segment of integrin α V β 3. *Science.* 294:339-345.
- Xiong JP, Stehle T, Zhang R, Joachimiak A, Frech M, Goodman SL, Arnaout MA. 2002. Crystal structure of the extracellular segment of integrin α V β 3 in complex with an Arg-Gly-Asp ligand. *Science* 296:151-155
- Xiong JP, Stehle T, Goodman SL, Arnaout MA. 2003. New insights into the structural basis of integrin activation. *Blood* 102:1155-1159.
- 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.
- Yang JT, Rayburn H, Hynes RO. 1993. Embryonic mesodermal defects in α 5 integrin-deficient mice. *Development.* 119:1093-1105.
- Yang W, Shimaoka M, Salas A, Takagi J, Springer TA. 2004. Intersubunit signal transmission in integrins by a receptor-like interaction with a pull spring. *Proc Natl Acad Sci USA.* 101:2906-2911.
- Yang JT, Rayburn H, Hynes RO. 1995. Cell adhesion events mediated by α 4 integrins are

essential in placental and cardiac development. *Development*. 121: 549-560.

Yang J, Ma YQ, Page RC, Misra S, Plow EF, Qin J. 2009. Structure of an integrin α IIb β 3 transmembrane-cytoplasmic heterocomplex provides insight into integrin activation. *Proc Natl Acad Sci USA*. 106:17729-17734.

Yamada KM, Geiger B. 1997. Molecular interactions in cell adhesion complexes. *Curr Opin Cell Biol*. 9:76-85.

Yu Y, Zhu J, Mi LZ, Walz T, Sun H, Chen J, Springer TA. 2012. Structural specializations of α 4 β 7, an integrin that mediates rolling adhesion. *J Cell Biol*. 196:131-146.

Zeltz C, Gullberg D. 2016. The integrin-collagen connection - a glue for tissue repair? *J Cell Sci*. 129:1284.

Zhang WM, Kapyla J, Puranen, JS, Knight CG, Tiger CF, Pentikainen OT, Johnson MS, Farndale RW, Heino J, Gullberg D. 2003. α 11 β 1 integrin recognizes the GFOGER sequence in interstitial collagens. *J Biol Chem*. 278:7270-7277.

Zhang H, Casanovas JM, Jin M, Liu JH, Gahmberg CG, Springer TA, Wang JH. 2008. An unusual allosteric mobility of the C-terminal helix of a high-affinity α L integrin I domain variant bound to ICAM-5. *Mol Cell*. 31:432-437.

Zhang Z, Ramirez NE, Yankeelov TE, Li Z, Ford LE, Qi Y, Pozzi A, Zutter MM. 2008. α 2 β 1 integrin expression in the tumor microenvironment enhances tumor angiogenesis in a tumor cell-specific manner. *Blood* 111:1980-1988.

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 USA*. 102: 5744-5748.

Zhu J, Motejlek K, Wang D, Zang K, Schmidt A, Reichardt LF. 2002. β 8 integrins are required for vascular morphogenesis in mouse embryos. *Development*. 129: 2891-2903.

Zhu J, Luo BH, Xiao T, Zhang C, Nishida N, Springer TA. 2008. Structure of a complete integrin ectodomain in a physiologic resting state and activation and deactivation by applied forces. *Mol Cell*. 32:8498-61.

Zhu J, Luo BH, Barth P, Schonbrun J, Baker D, Springer TA. 2009. The structure of a receptor with two associating transmembrane domains on the cell surface: integrin α IIb β 3. *Mol Cell*. 34: 234-249.

Zutter MM, Santoro SA. 1990. Widespread histologic distribution of the α 2 β 1 integrin cell-surface collagen receptor. *Am J Pathol*. 137:113-120.



ISBN 978-952-12-3448-4