International Journal of Biological, Life and Agricultural Sciences
ISSN: 2415-6612
Vol:6, No:5, 2012

Evolutionary Origin of the αC Helix in Integrins

B. Chouhan, A. Denesyuk, J. Heino, M. S. Johnson, K. Denessiouk

***Abstract*—**Integrins are a large family of multidomain α/β cell signaling receptors. Some integrins contain an additional inserted I domain, whose earliest expression appears to be with the chordates, since they are observed in the urochordates *Ciona intestinalis* (vase tunicate) and *Halocynthia roretzi* (sea pineapple), but not in integrins of earlier diverging species. The domain's presence is viewed as a hallmark of integrins of higher metazoans, however in vertebrates, there are clearly three structurally-different classes: integrins without I domains, and two groups of integrins with I domains but separable by the presence or absence of an additional αC helix. For example, the αI domains in *collagen-binding integrins* from Osteichthyes (bony fish) and all higher vertebrates contain the specific αC helix, whereas the αI domains in *non-collagen binding integrins* from vertebrates and the αI domains from earlier diverging urochordate integrins, i.e. tunicates, do not. Unfortunately, within the early chordates, there is an evolutionary gap due to extinctions between the tunicates and cartilaginous fish. This, coupled with a knowledge gap due to the lack of complete genomic data from surviving species, means that the origin of collagen-binding αC-containing αI domains remains unknown. Here, we analyzed two available genomes from *Callorhinchus milii* (ghost shark/elephant shark; Chondrichthyes – cartilaginous fish) and *Petromyzon marinus* (sea lamprey; Agnathostomata), and several available Expression Sequence Tags from two Chondrichthyes species: *Raja erinacea* (little skate) and *Squalus acanthias* (dogfish shark); and *Eptatretus burgeri* (inshore hagfish; Agnathostomata), which evolutionary reside between the urochordates and osteichthyes. In *P. marinus*, we observed several fragments coding for the αC-containing αI domain, allowing us to shed more light on the evolution of the collagen-binding integrins.

***Keywords*—**Integrin αI domain, integrin evolution, collagen binding, structure, αC helix

I. Introduction

INTEGRINS are a large family of bi-directionally signaling, heterodimeric trans-membrane receptors involved in cell-cell and cell-ExtraCellular Matrix (cell-ECM) interactions [1]. The extracellular domains of integrins interact with their ligands, while the cytoplasmic domains communicate with signaling molecules inside the cell [2]. These interactions are central to the regulation of cell migration, phagocytosis, cell growth, immune system function and development [3]. In addition, integrins are also involved in a number of health issues, including tumor progression [4] and recognition of pathogens [5], especially picornaviruses e.g. echovirus 1 [6,7].

B. Chouhan is with the Structural Bioinformatics Laboratory, Department of Biosciences (Biochemistry), Åbo Akademi University, Turku, Finland, and Turku Centre for Biotechnology, University of Turku and Åbo Akademi University, Turku, Finland (e-mail: bchouhan@abo.fi).

A. Denesyuk is with the Structural Bioinformatics Laboratory, Department of Biosciences (Biochemistry), Åbo Akademi University, Turku, Finland (e-mail: adenesyu@abo.fi).

J. Heino is with the Department of Biochemistry and Food Chemistry, University of Turku, Turku, Finland (e-mail: jyrki.heino@utu.fi).

M. S. Johnson is with the Structural Bioinformatics Laboratory, Department of Biosciences (Biochemistry), Åbo Akademi University, Turku, Finland (e-mail: johnson4@abo.fi).

K. Denessiouk is with the Turku Centre for Biotechnology, University of Turku and Åbo Akademi University, Turku, Finland and Structural Bioinformatics Laboratory, Department of Biosciences (Biochemistry), Åbo Akademi University, Turku, Finland (e-mail: kdenessi@abo.fi).

Integrins are composed of two subunits, α and β. Humans express 18 different integrin α subunits and 8 different β subunits, which are known to form 24 α/β heterodimeric combinations [8]. The α subunits can be further subdivided into those that contain an additional inserted domain, the αI domain, and those without the αI domain [9]. The Rossmann fold structure of the αI domain buds out of the β-propeller domain [10] in the "head" region of the α subunit; ligands are primarily recognized via interactions with the Metal Ion Dependent Adhesion Site (MIDAS) [11]. I domains from different integrins have unique expression patterns and carry out different functions. For example, In humans, I domains of α1β1, α2β1, α10β1 and α11β1 generally function as collagen receptors and grant structural integrity to cells and tissues, whereas the I domains of αXβ2, αDβ2, αMβ2, αLβ2 and αEβ7 have roles in the immune system and are required for interactions of leukocytes with endothelial cells and other cell types and matrix structures [12]. A key structural feature that distinguished I domains of the collagen receptors from those of the immune system receptors is the presence of the αC helix located towards the carboxy-terminus of I domains of the collagen receptors. The αC helix has a critical role in the conformational change between the "closed" and "open" forms of the I domain required for ligand binding. Our earlier studies of integrin evolution showed that I domains are present in the urochordates *C. intestinalis* and *H. roretzi*, whereas the αC helix is absent (Fig. 1) [13, 14].

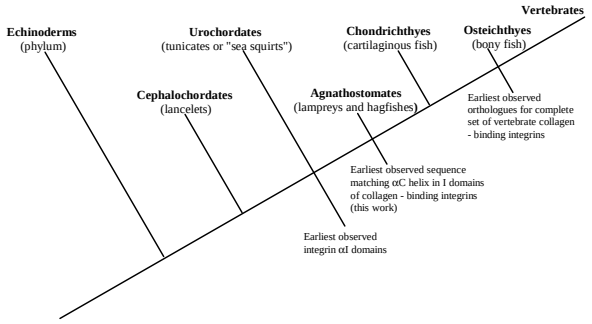

Fig. 1 Evolution of the integrin αI domains in the chordates

The absence of the αC helix in integrins of early chordates suggests that the integrins were either incapable of collagen binding or that other mechanisms were employed to bind collagens. Since orthologues of mammalian collagen-binding αC-containing αI domains are observed in sequences of bony fish and in all more recently diverging species of vertebrates, we can bracket the likely origin of the αC helix containing integrins to after the appearance of the urochordates and before the appearance of the bony fish (Fig. 1) [13]. Unfortunately, many species diverging after the urochordates that once existed are now extinct [15], and thus, it may be impossible to establish the first appearance of either the human-type integrins present in vertebrates today or the origin

of the αC helix typifying the vertebrate collagen receptor I domains. To make things even more uncertain at this time, there is currently a lack of genomic data from the intermediate chordate species that do exist. In the case of the cartilaginous fish, while there are many extant species of sharks and rays, genomic sequence data is spotty at best. Nonetheless, we are fortunate to have available two representative sequenced genomes from the intermediate chordates: *Petromyzon marinus* (the sea lamprey, 5.9x coverage) and *Callorhinchus milii* (the elephant shark, 1.4x coverage). Moreover, Expression Sequence Tags (ESTs) exist for *Eptatretus burgeri* (inshore hagfish), *Raja erinacea* (little skate), and *Squalus acanthias* (dogfish shark) and analysis of these genomic data can give insight into the evolutionary blueprint of the mammalian type integrin αI domains [16]. In our analysis of these data, we could identify "fragments" coding for the αI domain, including αI domains containing the "key" αC helix found in the vertebrate-type collagen binding integrins. Here, we present our analysis of these genomes and discuss their implications for the evolution of the αI–domain containing integrins.

II. Methods

The genome assemblies of the *P. marinus* and *C. milii* were obtained from http://genome.wustl.edu/genomes/view/petromyzon_marinus and http://esharkgenome.imcb.a-star.edu.sg/resources.html, respectively. The sequences of I-domains from nine human integrin α subunits (sequence data from NCBI: http://www.ncbi.nlm.nih.gov/) were used to perform full-scale local tBLASTn [17] searches (using the service at NCBI) of the above genomes. In addition to the genome assemblies of *P. marinus* and the *C. milii*, we used the tBLASTn service to analyze several other incomplete genomes and Expression Sequence Tags (ESTs) from organisms that arose between the divergence of the urochordates and osteichthyes, such as the *R. erinacea*, *S. acanthias* and *E. burgeri*. Sequences from *S. acanthias* were retrieved from NCBI and those of *C. intestinalis* were taken from the supplementary material of Ewan [18]. The Sea Lamprey sequences have been obtained by making local tBLASTn searches in its genome assembly and tBLASTn searches against Sea Lamprey scaffolds at the ENSEMBL database (http://www.ensembl.org/Petromyzon_marinus/Info/Index).
Sequences were aligned using T-COFFEE (http://www.tcoffee.org/) and adjusted manually. Secondary structure prediction employed Jpred [19], GOR [20], Porter [21], Prof_seq [22] and PSIPRED [23].

III. Results

Using the sequences of I domains from the nine human α subunits that contain I domains, we made a comprehensive analysis of all available genome assemblies and ESTs from species that arose after the divergence of the urochordates and before the appearance of osteichthyes, including the genomes of *P. marinus* (sea lamprey), *C. milii* (elephant shark), *R. erinacea* (little skate), *S. acanthias* (dogfish shark) and *E. burgeri* (inshore hagfish). In the sea lamprey we were able to identify three full-length integrin αI domain sequences and four short fragments; in the hagfish one sequence fragment; the search of the skate/shark data did not reveal unambiguously identifiable integrin I domain sequences. The sequences from lamprey and hagfish matched the signature motif of the collagen-receptor αI domain. The three full-length sequences from the sea lamprey, Pma_f1, Pma_f2 and Pma_f3, and the fragment from the inshore hagfish, Ebu_f, are shown in Fig. 2. For comparison, we included two of the four human collagen-binding αI domain sequences containing the αC-helix: α1 and α11; two of the five human leukocyte-specific αI domain sequences lacking the αC helix: αM and αD; and all eight known αI domain sequences from the urochordate sea squirt *C. intestinalis*. The alignment of the 16 sequences (Fig. 2) shows extensive areas of sequence similarity that span the entire length of the αI domain, including secondary structure elements (shown in transparent square boxes in Fig. 2) found in the human α1I domain X-ray structure [24, 25]. The areas of high sequence conservation include conserved residues of the MIDAS motif from all 16 sequences (columns highlighted in black in Fig. 2). The I domains of four human collagen-binding α subunits contain the αC helix, represented by α1 and α11 in the alignment, whereas the five I domains of the human leukocyte-specific integrin α subunits, represented by I domains from αM and αD in the alignment, do not contain the αC helix. The eight sequences from the urochordate *C. intestinalis* also do not contain the αC helix. In contrast, all three of the identified full-length integrin αI domain sequences (Pma_f1, Pma_f2 and Pma_f3 in Fig. 2) from the *P. marinus* contain sequences matching the αC helix present in the human collagen-binding integrins. The secondary structure prediction made for the *P. marinus* sequences, using five different computer-based prediction methods, suggests that the sequences are compatible with formation of an alpha helix in the region where the αC helix would be located in the I domain (Fig. 3). The hagfish is a close relative of the lamprey, both belonging to the super-class Agnathostomata. The alignment in Fig. 2 contains an incomplete sequence fragment from the inshore hagfish (*E. burgeri*; Ebu_f), which matches the signature motif of the collagen-receptor αI domain, and shares strong sequence similarity with the other αI domain sequences, including residues of the highly conserved MIDAS motif. The available sequence terminates just prior to the αC helix present in the collagen-binding I domains and presumably also present within the sequences of the sea lamprey according to the sequence analysis. Without further sequence data spanning the remainder of the presumptive I-domain region, it is not possible to confirm whether the inshore hagfish contains the signature feature i.e. the αC helix of the collagen-binding I domains found in the vertebrates.

#	Name:	Abbr.:	Code:	#	Name:	Abbr.:	Code:
1	Human α1	Hsa_a1	NP_852478 (NCBI)	9	Sea Squirt α1	Cin_a1	ci0100131118
2	Human α11	Hsa_a11	NP_001004439.1 (NCBI)	10	Sea Squirt α2	Cin_a2	ci0100149446
3	Human αM	Hsa_aM	NP_000623.2 (NCBI)	11	Sea Squirt α3	Cin_a3	ci0100130596
4	Human αD	Hsa_aD	NP_005344.2 (NCBI)	12	Sea Squirt α4	Cin_a4	ci0100130838
5	Sea Lamprey f1	Pma_f1	Contig 16236.5, Contig 16236.4, Contig 15215.1 or Scaffold GL479139	13	Sea Squirt α5	Cin_a5	ci0100152002, ci0100152017
6	Sea Lamprey f2	Pma_f2	Contig 42254.2, Contig 24528.1, Contig 36235.1 or Scaffold GL477642	14	Sea Squirt α6	Cin_a6	ci0100131399
7	Sea Lamprey f3	Pma_f3	Contig 1144.7 or Scaffold 501125	15	Sea Squirt α7	Cin_a7	ci0100152615
8	Inshore Hagfish f	Ebu_f	BJ655520.1 (NCBI)	16	Sea Squirt α8	Cin_a8	ci0100130149

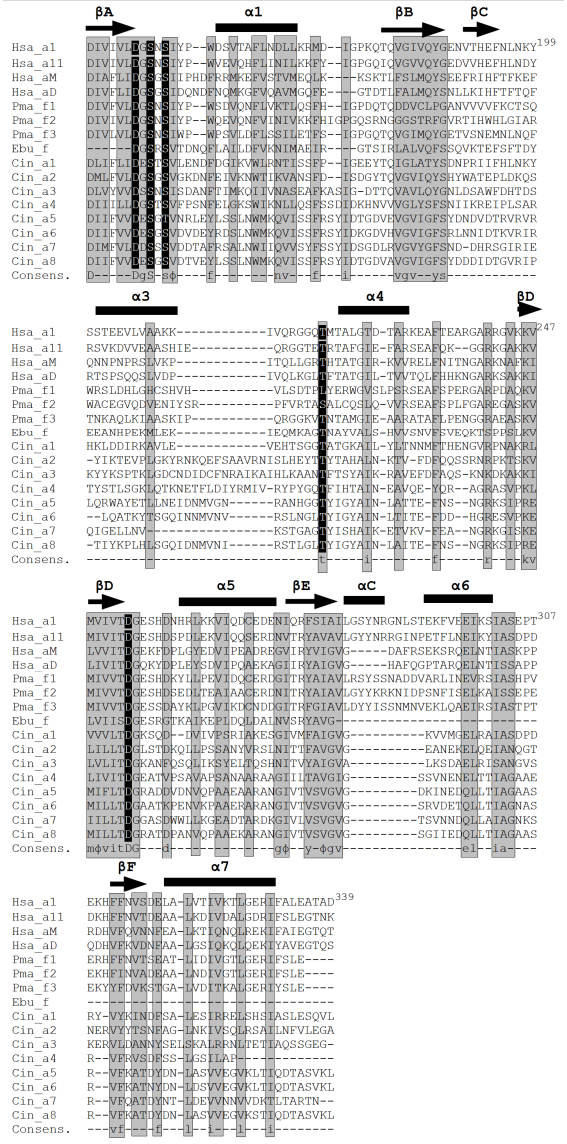

Fig. 2 Sequence alignment of integrin αI domains. Amino acid numbering and secondary structure elements are from the human α1 I domain (PDB code: 1PT6 [24, 25]). Sequence similarity highlighted in grey. MIDAS residues highlighted as black columns

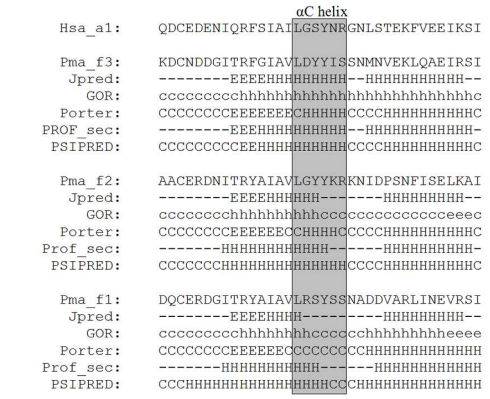

Fig. 3 Secondary structure prediction with five different methods predicts helical structure in all three identified sequences from *P. marinus* within the region that corresponds to the αC helix in the collagen-binding αI domains, here represented by the sequence and structure from human α1 I domain (PDB code: 1PT6 [24, 25])

IV. Conclusion

Early in evolution, prior to the origin of vertebrates, organisms used other means to bind collagens. By the time bony vertebrates arose on the Earth, they had developed specific mechanisms for collagen binding using a set of collagen receptors. Bony fish, for example, have orthologues, indeed multiple isoforms, of each of the human collagen-binding integrins. A key structural feature of the collagen binding function of the human α1, α2, α10 and α11 integrin subunits is the presence of the inserted I domain and its characteristic αC helix, which is absent in the human leuckocyte-binding integrins that also have inserted I domains. The αC helix is defined as a major determinant for collagen binding, where a collagen molecule binds into the groove created by this helix and coordinates the metal ion in the MIDAS site as shown by the human α2 I domain structure in complex with a collagen-like triple-helical peptide [12].

It has been shown, however, on several occasions that for some collagen-binding integrins, modifications or even complete removal of the αC helix in the αI domain may not prevent collagen binding. For example, Kamata et al. [26] observed that deletion of the entire αC helix did not significantly affect collagen binding to the α2I domain of integrin α2β1, while the amino acids of the MIDAS motif and the bound metal ion played crucial roles in the recognition of collagen. It was also shown for the αI domains of the tunicate *C. intestinalis*, which lack the αC helix, that they not only could bind collagen IX, but they could do so in a metal-independent manner [27]. These observations suggest that collagen binding by integrins may have preceded the specialization of integrins into a collagen receptor group as seen throughout the vertebrates, and typified by the presence of the αC helix. It is likely then that the helix serves additional functions in addition to participation in binding collagens. Indeed, the function of integrins is inherently dynamic and in response to ligand binding or a lack thereof integrins change conformation and hence alter their functions in both binding

molecules via the ectodomain and via their cytoplasmic tails. The structure of the αC helix differs depending on the binding state and it is possible that these changes are communicated through to the conformational and functional state of the αβ integrin as a whole.

Notwithstanding the functional role of the αC helix and whether or not it is absolutely required for collagen binding, it is an observed marker that distinguishes the two functionally-distinct sets of I domains found throughout the vertebrates. Where present, the integrins function as collagen receptors. Here, we have sought to identify more closely the origin of the αC helix in chordate evolution. The present data, while limited by the available species and genomes that have been sequenced, as well as the quality of the genomic data, do identify the presence of integin-type I domains in the lamprey *P. marinus* having all of the hallmarks of a vertebrate collagen receptor, including the MIDAS motif, and most telling, the presence of sequence matching the diagnostic αC helix. Integrin I domains are expected to be present in cartilaginous fish and the absence of identifiable sequences is unexpected but very likely reflects the current state of sequencing for these species.

Acknowledgment

This work was supported in part by grants from Stiftelsens för Åbo Akademi Forskningsinstitut, Åbo Akademi University Center of Excellence Program in Cell Stress and Aging, the Sigrid Jusélius Foundation, the Joe, Pentti and Tor Borg Memorial Fund, and the National Doctoral Program in Informational and Structural Biology, and made use of computational infrastructure provided by funding from Biocenter Finland in bioinformatics and Åbo Akademi University.

References

[1] Takada, Y., Ye, X. and Simon, S. (2007) The integrins, *Genome Biol* 8, pp. 215.1-215.9.
[2] Luo, B.H., Carman, C.V. and Springer, T.A. (2007) Structural basis of integrin regulation and signaling, *Annu Rev Immunol* 25, pp. 619-647.
[3] Arnaout, M.A., Goodman, S.L. and Xiong, J.P. (2007) Structure and mechanics of integrin-based cell adhesion, *Curr Opin Cell Biol* 19, pp. 495-507.
[4] Shin, S., Wolgamott, L. and Yoon, S.O. (2012) Integrin trafficking and tumor progression, *Int J Cell Biol* 2012, pp. 516789.1-516789.7.
[5] Ulanova, M., Gravelle, S. and Barnes, R. (2009) The role of epithelial integrin receptors in recognition of pulmonary pathogens, *J Innate Immun* 1, pp. 4-17.
[6] Xing, L., Huhtala, M., Pietiäinen, V., Käpylä, J., Vuorinen, K., Marjomäki, V., Heino, J., Johnson, M. S., Hyypiä, T. and Cheng, R. H. (2004) Structural and functional analysis of integrin α2I domain interaction with echovirus 1, *J Biol Chem* 279, pp. 11632-11638.
[7] Jokinen, J., White, D. J., Salmela, M., Huhtala, M., Käpylä, J., Sipilä, K., Puranen, J. S., Nissinen, L., Kankaanpää, P., Marjomäki, V., Hyypiä, T., Johnson, M. S. and Heino, J. (2010) Molecular mechanism of α2β1 integrin interaction with human echovirus 1, *EMBO J* 29, pp. 196-208.
[8] Shimaoka, M. and Springer, T.A. (2003) Therapeutic antagonists and conformational regulation of integrin function, *Nat Rev Drug Discov* 2, pp. 703–716.
[9] Larson, R.S., Corbi, A.L., Berman, L. and Springer, T. (1989) Primary structure of the leukocyte function-associated molecule-1 alpha subunit: an integrin with an embedded domain defining a protein superfamily, *J Cell Biol* 108, pp. 703–712.
[10] Chouhan, B., Denesyuk, A., Heino, J., Johnson, M. S. and Denessiouk, K. (2011) Conservation of the human-type integrin beta propeller domain in bacteria, *PLoS One* 6, e25069.
[11] Lee, J.O., Rieu, P., Arnaout, M.A. and Liddington, R. (1995) Crystal structure of the A domain from the alpha subunit of integrin CR3 (CD11b/CD18), *Cell* 80, pp. 631-638.
[12] Emsley, J., King, S.L., Bergelson, J.M. and Liddington, R.C. (1997) Crystal structure of the I domain from integrin α2β1, *J Biol Chem* 272, pp. 28512-28517.
[13] Huhtala, M., Heino, J., Casciari, D., de Luice, A. and Johnson M.S. (2005) Integrin evolution: insights from ascidian and teleost fish genomes, *Matrix Biol* 24, pp. 83-95.
[14] Heino, J, Huhtala, M., Kapyla, J. and Johnson M.S. (2009) Evolution of collagen-based adhesion systems, *Int J Biochem Cell Biol* 41, pp. 341-348.
[15] Donoghue, P.C. and Purnell, M.A. (2005) Genome duplication, extinction and vertebrate evolution, *Trends Ecol Evol* 20, pp. 312-319.
[16] Venkatesh, B., Kirkness, E.F., Loh, Y.H., Halpern, A.L., Lee, A.P., Johnson, J., Dandona, N., Viswanathan, L.D., Tay, A., Venter, J.C., Strausberg, R.L. and Brenner, S. (2006) Ancient noncoding elements conserved in the human genome, *Science* 314, p. 1892.
[17] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool, *J Mol Biol* 215, pp. 403-410.
[18] Ewan, R., Huxley-Jones, J., Mould, A.P., Humphries, M.J., Robertson, D.L. and Boot-Handford, R.P. (2005) The integrins of the urochordate *Ciona intestinalis* provide novel insights into the molecular evolution of the vertebrate integrin family, *BMC Evol Biol* 5, pp. 31.1-31.18.
[19] Cole, C., Barber J.D. and Barton, G.J. (2008) The Jpred 3 secondary structure prediction server, *Nucleic Acids Research* 36, Web Server issue W197-W201.
[20] Sen, T.Z., Jernigan, R.L., Garnier, J. and Kloczkowski, A. (2005) GOR V server for protein secondary structure prediction, *Bioinformatics* 21, pp. 2787-2788.
[21] Pollastri, G. and McLysaght, A. (2005) Porter: a new, accurate server for protein secondary structure prediction, *Bioinformatics* 21, pp. 1719-1720.
[22] Rost, B and Sander, C. (1994) Combining evolutionary information and neural networks to predict protein secondary structure, *Proteins* 19, pp. 55-72.
[23] McGuffin, L.J., Bryson, K. and Jones, D.T. (2000) The PSIPRED protein structure prediction server, *Bioinformatics* 16, pp. 404-405.
[24] Salminen, T., Nymalm, Y., Kankare, J., Käpylä, J., Heino, J. and Johnson, M.S. (1999) Large-scale purification and preliminary x-ray analysis of the human integrin α1I domain, *Acta Crystallogr* D55, pp. 1365-1367.
[25] Nymalm, Y., Puranen, J.S., Nyholm, T.K.M., Käpylä, J., Kidron, H., Pentikäinen, O., Airenne, T.T., Heino, J., Slotte, P., Johnson, M.S. and Salminen, T.A. (2004) structural changes in α1 I domain of human integrin α1β1, *J Biol Chem* 279, pp. 7962-7970.
[26] Kamata, T., Liddington, R.C. and Takada, Y. (1999) Interaction between collagen and the α2 I-domain of integrin α2β1. Critical role of conserved residues in the metal ion-dependent adhesion site (MIDAS) region, *J Biol Chem* 274, pp. 32108-32111.
[27] Tulla, M., Huhtala, M., Jäälinoja, J., Käpylä, J., Farndale, R.W., Ala-Kokko, L., Johnson, M.S. and Heino, J. (2007) Analysis of an ascidian integrin provides new insight into early evolution of collagen recognition, *FEBS Lett* 581, pp. 2434-2440.