An alternative insert of three amino acids is incorporated into collagen XIV in a developmentally regulated fashion

Martin Imhof, Beat Trueb*

M.E. Müller-Institute for Biomechanics, University of Bern, P.O. Box 30, CH-3010 Bern, Switzerland

Received 8 September 1998; received in revised form 12 October 1998

Abstract We have identified a novel splice variant of chicken collagen XIV which contains an insert of three amino acids (Val-Arg-Thr) in the sixth fibronectin type III-like (FNIII) domain. The codons for these amino acids are inserted into the mRNA by skipping of a splice donor site and usage of another donor site 9 bp further downstream in the collagen XIV gene. The percentage of the new splice variant in the total collagen XIV mRNA varies between 22 and 46% in different embryonic tissues. After hatching, however, this percentage increases dramatically and reaches 86% in adult skeletal muscle and 58% in adult gizzard, indicating developmental regulation of this splicing event. Computer modeling suggests that the three extra amino acids cause an increase in the size of a flexible loop connecting two β-strands in the sixth FNIII domain. This increase might affect the exact arrangement of the FNIII domain in the collagen XIV molecule, thereby modulating its interactions with other matrix molecules.

Key words: Alternative splicing; Collagen XIV; Extracellular matrix; Fibronectin type III repeat; FNIII module

1. Introduction

Collagen XIV is a modular protein of the extracellular matrix consisting of two short triple helical segments and three noncollagenous domains. Together with collagens IX, XII, XVI and XIX it has been classified in a subfamily of collagens termed fibril-associated collagens with interrupted triple helices (FACITs, for review see [1–4]). Collagen XIV is expressed in most mesenchymal tissues [5–7] where it appears to interact with other collagens (mainly collagen type VI [8]), glycosaminoglycans and proteoglycans [8–11] and matrix receptors on several cell types [12–14].

The non-collagenous domain at the N-terminus of collagen XIV is extremely large making up nearly 80% of the entire polypeptide. The sequences of the chicken [15–18] and the human [19,20] protein illustrate that this domain is composed of eight fibronectin type III (FNIII) repeats related to fibronectin, two von Willebrand factor A-like (vWFA) domains related to von Willebrand factor and one non-collagenous domain 4 (NC4 domain) related to collagen type IX. At least two splice variants have so far been detected with the human and chicken protein which differ in the length and sequence of their C-terminal non-collagenous domain [16,20]. It is possible that these variations modulate the proposed interaction of collagen XIV with interstitial collagen fibers [1–4].

Here we describe another splice variant of chicken collagen XIV that differs from the common form by the insertion of three amino acids in one of its FNIII domains. Since the expression of this insert is highly regulated during development, we believe that it plays a role in the interaction of collagen XIV with other matrix proteins.

2. Materials and methods

2.1. Genomic DNA cloning

Genomic DNA was isolated from chicken fibroblasts by proteinase K digestion and extensive phenol extraction [21]. A fragment of the type XIV collagen gene was amplified with AmpliTaq polymerase (Perkin Elmer) in the presence of 3 mM $MgCl_2$ using ~1 μg of genomic DNA and a pair of strand specific primers. These primers corresponded to the published cDNA sequence [16] except for one mismatch in each primer that was introduced to facilitate subsequent cloning (upper primer: CAGGAGTCAGAATTCTTGTTATAGATGA, position 2540–2567; lower primer GTATCGGGTACCAGAGGCTGA, position 2709–2729). The fragments were amplified through 35 cycles of 1′ at 95°C, 2′ at 57°C and 2′ at 72°C and ligated into the *Eco*RI/*Kpn*I site of the plasmid pUC19 [22]. The DNA sequence of the insert was determined by the dideoxynucleotide chain-termination method [23] using the enzyme Sequenase 2.0 (USB).

2.2. Analysis of mRNA transcripts

Total RNA was isolated from various embryonic and adult chicken tissues by the guanidinium isothiocyanate method [24] using the RNeasy kit of Qiagen. Poly(A)$^+$ RNA was separated from ribosomal RNA by chromatography on Oligotex beads (Qiagen). The mRNA (~1 μg) was transcribed at 42°C into cDNA with 40 units of AMV reverse transcriptase (Boehringer) following the instructions of the manufacturer. The reaction was primed with the oligonucleotide corresponding to position 2709–2729. The single-stranded cDNA was amplified by PCR using conditions and primers described above except that 35 cycles of 1′ at 95°C, 1′ at 57°C and 1′ at 72°C were performed. The products were resolved on a 3% MetaPhor agarose gel (FMC BioProducts). The band of interest was recovered from the gel with QiaEXII (Qiagen) and cloned into the *Eco*RI/*Kpn*I site of pUC19. The presence or absence of the 9-bp insert was determined by DNA sequencing [23]. At least 36 individual clones from each tissue sample were sequenced. The results were statistically evaluated using the χ^2 test [25].

2.3. Molecular modeling

The three-dimensional structure of the FNIII domain containing or lacking the alternative insert was modeled with the help of a computer using the program Swiss-Model Version 1.1 [26]. The tenth FNIII domain of fibronectin (1TTF) was employed as template [27]. Homology modeling, superposition and energy minimization were done following the tutorial provided with the program.

*Corresponding author. Fax: (41) (31) 632 4999.
E-mail: trueb@mem.unibe.ch

Abbreviations: FNIII domain, fibronectin type III-like domain; NC4 domain, domain related to the non-collagenous domain 4 of collagen IX; RT-PCR, polymerase chain reaction after reverse transcription; vWFA domain, von Willebrand factor A-like domain

The novel nucleotide sequences reported in this paper have been submitted to the GenBank/EMBL database with accession number AJ011841.

PII: S0014-5793(98)01329-5

3. Results

3.1. Identification of an alternatively spliced insert

Previous cloning efforts [15,16] had led to the isolation of 19 different cDNA clones that spanned the entire mRNA of chicken collagen XIV. When the sequences of all these clones were analyzed in detail, a curious variation was observed at position 2673. Compared with the published cDNA sequence, 2 out of 8 clones covering this region possessed an insert of 9 bp (GTAAGAACG). This insert coded for the three amino acids Val-Arg-Thr and occurred in the sixth FNIII repeat of collagen XIV.

To investigate the mechanism by which the 9 bp were inserted into the mRNA, we isolated the corresponding region from the chicken type XIV collagen gene. A PCR with genomic DNA from chicken fibroblasts utilizing two primers that annealed upstream and downstream of the 9-bp insert yielded a genomic fragment of ~2000 bp. Sequencing studies demonstrated that the sixth FNIII repeat was encoded by two separate exons. At the 5′ end, the genomic fragment contained 134 bp of the first exon which were followed, at position 2673 of the cDNA, by an intron harboring the splice donor site GTAAGA (Fig. 1). At the 3′ end, the genomic fragment possessed 56 bp of the second exon which were preceded by the splice acceptor site CCTGACTATCCAG. The intervening sequence between the two exons was 1839 bp. The 9-bp insert was found at the 5′ end of this intron. Skipping of the first splice donor site at position 2673 and usage of an alternative splice donor site 9 bp further downstream (GTCTGT) obviously leads to the extension of the first exon by 9 bp. The three amino acids Val-Arg-Thr are therefore incorporated into collagen XIV by alternative splicing of the RNA.

3.2. Expression of the alternative insert in different tissues and at different developmental stages

To study the occurrence of the 9-bp insert, we analyzed 8 different tissues from 16-day old chicken embryos by RT-PCR. From every tissue sample we obtained a fragment of ~200 bp as predicted from the annealing positions of the two primers used. The yield of this band was dependent on the relative amount of collagen XIV mRNA present in each sample. It was relatively high in all muscular tissues, but relatively low in liver, brain and calvaria (not shown). Since the PCR fragments with and without alternative insert could not be separated from each other on a polyacrylamide gel, they were cloned into the plasmid pUC19. To determine the percentage of fragments with alternative insert, at least 36 clones derived from each tissue were sequenced. These sequencing studies demonstrated that approximately one third of the clones possessed the alternative insert of 9 bp (Fig. 2A). Some differences were noted between the 8 tissues examined, skin containing a relatively high percentage of clones with insert (46%), heart and brain containing a relatively low percentage (22%). A statistical evaluation using the χ^2 test showed that the difference between skin and heart was highly significant ($P < 0.05$). However, the differences between the other tissues (skeletal muscle, gizzard, liver, sternum, calvaria) were too small to allow any predictions. To check the statistical significance in these cases, the sequences of several hundred cDNA clones would have been required.

When the expression of the 9-bp insert was analyzed at five developmental stages, striking differences were observed (Fig. 2B). Skeletal muscle from chicken embryos of day 10 and day 16 expressed the longer splice variant only in a small percentage of all collagen XIV mRNA molecules (25–27%), whereas skeletal muscle from adult chickens of 40 days, 2 years and 8 years expressed it at high level (72–86%, $P < 0.01$). Moreover, the percentage of molecules with insert did not increase in a linear fashion, but rather jumped from a relatively low value during embryogenesis to a high value one month after hatching. An increase in the percentage of the longer splice variant was also observed in the developing gizzard. However, the increase was smaller in this case, embryonic gizzard containing 34% of the longer splice variant, adult gizzard 58% (not shown).

3.3. Computer modeling

Alignment of all FNIII repeats from chicken collagen XIV revealed a relatively weak conservation of the amino acid sequences among the individual repeats (not shown). The length of the modules varied between 88 and 92 residues. Only the sixth module with the alternative insert was distinctly longer (94 residues). The alternative motif Val-Arg-Thr was not found in any of the other repeats.

To see where the three extra amino acids are situated within the sixth FNIII module, we performed protein modeling stud-

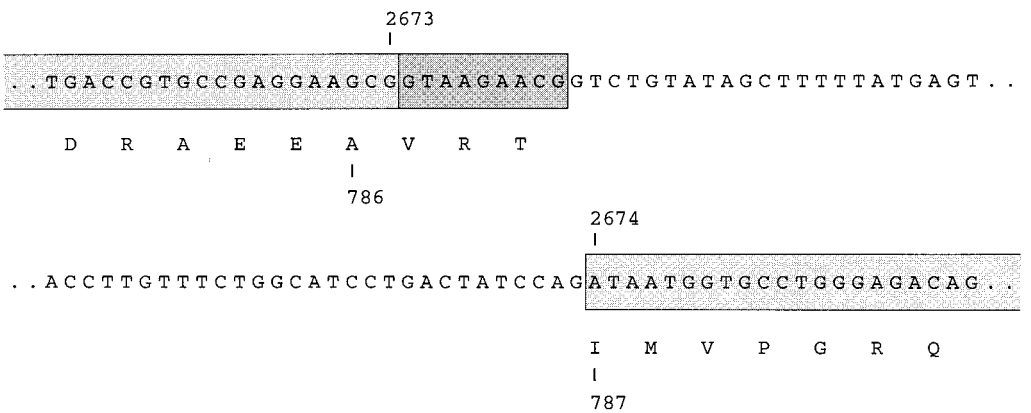

Fig. 1. Structure of the chicken collagen XIV gene around the region coding for the sixth FNIII domain. Exons are boxed, the translated sequence is given by the single letter code. The alternative insert is shaded dark grey. Positions are indicated according to the published cDNA sequence and its deduced amino acid sequence [16].

ies using the Swiss-Model computer program [26]. As template we chose the tenth FNIII repeat from human fibronectin, whose three-dimensional structure has been solved by nuclear magnetic resonance (structure of a single module [27]) as well as by X-ray crystallography (structure of a tandem array of four modules [28]). This structure revealed 7 β-strands termed A, B, C, C′, E, F and G which form two antiparallel β-sheets with three and four strands, respectively. The sixth FNIII repeat of collagen XIV lacking the alternative insert could be superimposed on the published structure without collisions of the amino acid side chains (Fig. 3). In this model, the insertion site for the three amino acids (residue 786) was located within β-strand C′. When the three amino acids Val-Arg-Thr were included in these modeling studies, it became evident that they cannot form a loop in the center of this β-strand. Energy minimization suggested that the three amino acids rather assume the positions of the three preceding amino acids Glu-Glu-Ala, which in turn are displaced towards

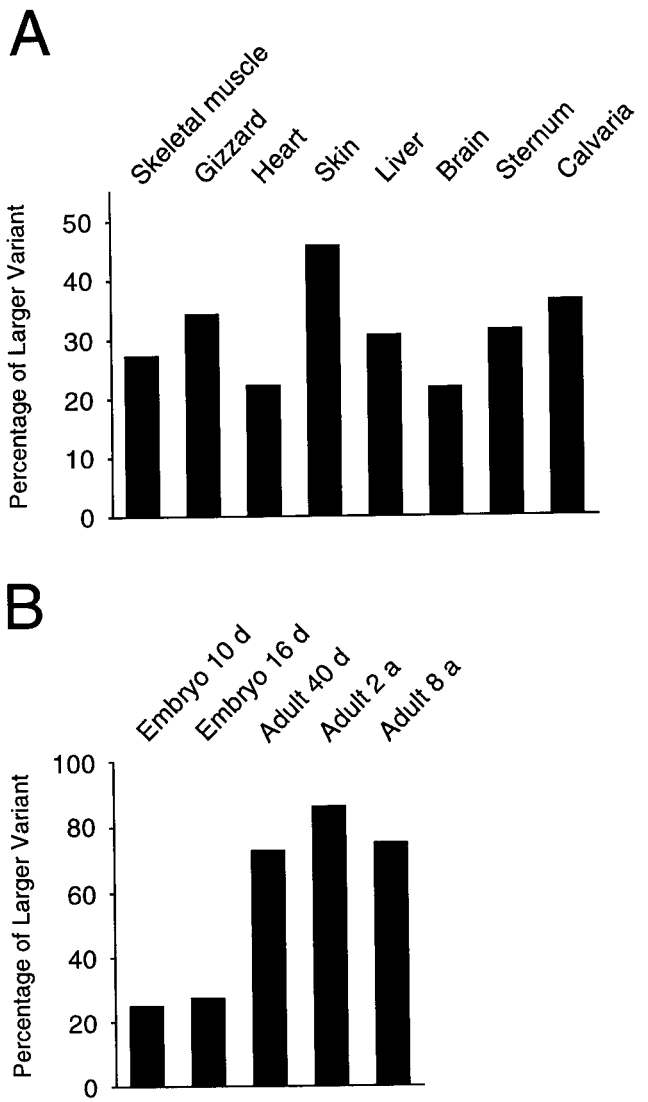

Fig. 2. Percentage of transcripts with alternative insert in the total collagen XIV mRNA population. A: Results obtained with eight different tissues from 16-day old chicken embryos. B: Results obtained with samples from skeletal muscle of five different developmental stages. Each bar represents the percentage calculated from at least 36 individual sequences.

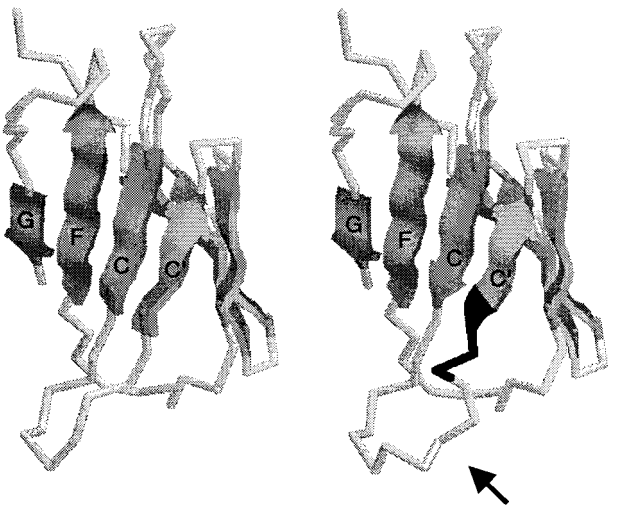

Fig. 3. Structure of the sixth FNIII domain from chicken collagen XIV. The domain lacking the alternative insert is shown on the left, the domain containing the insert on the right. The polypeptide backbone is drawn by a thick line, β-strands are given by flat arrows. The position of the alternative insert is indicated by the black region. Note that the insertion of the three amino acids will increase the size of the loop connecting β-strands C and C′ (black arrow).

the loop connecting β-strands C and C′. It is therefore likely that the incorporation of the alternative insert leads to a remarkable increase in the length of the connecting loop, which might affect the arrangement of the sixth FNIII repeat in the context of the whole collagen XIV molecule.

4. Discussion

Alternative splicing is an important mechanism by which the diversity of proteins, including several proteins of the extracellular matrix, is dramatically increased. Here we describe a novel splice variant of collagen XIV, which contains three extra amino acids in one of its FNIII domains. The codons for the three amino acids are inserted into the final transcript by skipping of a splice donor site and usage of a cryptic donor site 9 bp further downstream. Inclusion of the alternative insert is developmentally regulated. In skeletal muscle, the percentage of transcripts with insert increases from 25% at embryonic stages to 86% at adult stages.

FNIII domains are often encoded by two exons that together encode 90 amino acids [29–32]. The sixth FNIII domain of collagen XIV with the alternative insert is also encoded by two exons, but it is unusual in that it comprises 94 amino acids. No such extended FNIII domain occurs at any other location in collagen XIV, and no similarly extended domain is found in the related proteins fibronectin [29] and collagen XII [33]. It is therefore likely that the 9 bp for the three amino acids were specifically gained at some stage during the molecular evolution of collagen XIV.

The occurrence of the extended FNIII domain is not confined to the chicken molecule. A similar FNIII domain with 94 amino acids is also found at the corresponding position in the human molecule (734–827; note that the position is given incorrectly as 733–827 in [20]). However, the shorter splice variant has not yet been identified in man and it remains to be demonstrated whether human collagen XIV is subject to a similar alternative splicing event or whether the longer variant

is constitutively expressed. There is a potential splice donor site in the human cDNA sequence (GTCATA, position 1777–1782 in GenBank/EMBL database entry M64108) that could be utilized for the creation of a shorter splice variant. However, the sequence of the three extra amino acids is not conserved between the human (Ile-Gly-Thr) and the chicken protein (Val-Arg-Thr).

Since the incorporation of the alternative insert is developmentally regulated and since the length of the insert (although not its sequence) is conserved between avians and mammals we believe that it plays a crucial role. What this role might be is difficult to predict at present as the exact function of collagen XIV is not yet known [1–4,34]. It is conceivable that it is related to changes in the function of collagen XIV during development because the percentage of mRNA molecules with insert increases dramatically after hatching. The insert might therefore modulate an interaction of collagen XIV with another matrix molecule once the tissues of the animal, especially the muscles, are actively used.

It seems unlikely that the alternative insert will directly interact with another matrix molecule because its sequence is not well conserved between chicken and human. Computer modeling rather suggests that the insert induces an increase in the size of a flexible loop which connects two β-strands in the FNIII domain. This increase could affect the arrangement of the neighboring FNIII domains by changing the geometry with which the modules are connected in the entire polypeptide. In this way, the affinity and/or specificity of an adjacent domain for its ligand could be modulated.

In this context it might be of interest to mention a similar situation observed with fibronectin and its cell binding site. This site has been traced down to the three amino acid sequence Arg-Gly-Asp located in the tenth FNIII domain [29]. For a strong and specific interaction with integrins, fibronectin depends on a synergistic binding site which has been mapped to the ninth FNIII domain [35,36]. It has now been demonstrated that the geometry (i.e. the 'tilt' and 'twist' angles [32]) with which the ninth and tenth modules fit together determine both the affinity and specificity of integrin binding. The two modules of fibronectin are arranged in a way that the cell binding site and the synergy region of the neighboring module come sufficiently close together to allow an interaction with a single integrin molecule [28].

These considerations illustrate that it will probably not be possible to elucidate the function of the alternative insert in collagen XIV by performing interaction studies with a single FNIII domain and other matrix molecules. More elaborate experiments with constructs encompassing a tandem array of several FNIII repeats might be required to finally identify a specific role for this alternatively spliced insert.

Acknowledgements: This study was supported by the Swiss National Science Foundation (Grant 31-50571.97) and by the Bernese Cancer Liga. We thank D. Belluoccio, D. Jaquemar and J. Zumbrunn for helpful suggestions.

References

[1] van der Rest, M., Dublet, B. and Champliaud, M.-F. (1990) Biomaterials 11, 28–31.
[2] van der Rest, M. and Bruckner, P. (1993) Curr. Opin. Struct. Biol. 3, 430–436.
[3] Fukai, N., Apte, S.S. and Olsen, B.R. (1994) Methods Enzymol. 245, 3–28.
[4] Brown, J.C. and Timpl, R. (1995) Int. Arch. Allergy Immunol. 107, 484–490.
[5] Keene, D.R., Lunstrum, G.P., Morris, N.P., Stoddard, D.W. and Burgeson, R.E. (1991) J. Cell Biol. 113, 971–978.
[6] Castagnola, P., Tavella, S., Gerecke, D.R., Dublet, B., Gordon, M.K., Seyer, J., Cancedda, R., van der Rest, M. and Olsen, B.R. (1992) Eur. J. Cell Biol. 59, 340–347.
[7] Wälchli, C., Koch, M., Chiquet, M., Odermatt, B.F. and Trueb, B. (1994) J. Cell Sci. 107, 669–681.
[8] Brown, J.C., Mann, K., Wiedemann, H. and Timpl, R. (1993) J. Cell Biol. 120, 557–567.
[9] Font, B., Aubert-Foucher, E., Goldschmidt, D., Eichenberger, D. and van der Rest, M. (1993) J. Biol. Chem. 268, 25015–25018.
[10] Ehnis, T., Dieterich, W., Bauer, M., Kresse, H. and Schuppan, D. (1997) J. Biol. Chem. 20414–20419.
[11] Giry-Lozinguez, C., Aubert-Foucher, E., Penin, F., Deléage, G., Dublet, B. and van der Rest, M. (1998) Matrix Biol. 17, 145–149.
[12] Ehnis, T., Dieterich, W., Bauer, M., Lampe, B. and Schuppan, D. (1996) Exp. Cell Res. 229, 388–397.
[13] Ehnis, T., Dieterich, W., Bauer, M. and Schuppan, D. (1998) Exp. Cell Res. 239, 477–480.
[14] Klein, G., Kibler, C., Schermutzki, F., Brown, J., Müller, C.A. and Timpl, R. (1998) Matrix Biol. 16, 307–317.
[15] Trueb, J. and Trueb, B. (1992) Eur. J. Biochem. 207, 549–557.
[16] Wälchli, C., Trueb, J., Kessler, B., Winterhalter, K.H. and Trueb, B. (1993) Eur. J. Biochem. 212, 483–490.
[17] Gordon, M.K., Castagnola, P., Dublet, B., Linsenmayer, T.F., van der Rest, M., Mayne, R. and Olsen, B.R. (1991) Eur. J. Biochem. 201, 333–338.
[18] Gerecke, D.R., Foley, J.W., Castagnola, P., Gennari, M., Dublet, B., Cancedda, R., Linsenmayer, T.F., van der Rest, M., Olsen, B.R. and Gordon, M.K. (1993) J. Biol. Chem. 268, 12177–12184.
[19] Just, M., Herbst, H., Hummel, M., Dürkop, H., Tripier, D., Stein, H. and Schuppan, D. (1991) J. Biol. Chem. 266, 17326–17332.
[20] Bauer, M., Dieterich, W., Ehnis, T. and Schuppan, D. (1997) Biochim. Biophys. Acta 1354, 183–188.
[21] Kopp, M.U., Winterhalter, K.H. and Trueb, B. (1997) Eur. J. Biochem. 249, 489–496.
[22] Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1987) Current Protocols in Molecular Biology, Greene, New York, NY.
[23] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[24] Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. 162, 156–159.
[25] Sachs, L. (1984) Angewandte Statistik, 6. Auflage, Springer, Berlin.
[26] Peitsch, M.C. (1995) Bio-Technology 13, 658–660.
[27] Main, A.L., Harvey, T.S., Baron, M., Boyd, J. and Campbell, I.D. (1992) Cell 71, 671–678.
[28] Leahy, D.J., Aukhil, I. and Erickson, H.P. (1996) Cell 84, 155–164.
[29] Hynes, R.O. (1990) Fibronectins, Springer, New York, NY.
[30] Campbell, I.D. and Spitzfaden, C. (1994) Structure 2, 333–337.
[31] Potts, J.R. and Campbell, I.D. (1996) Matrix Biol. 15, 313–320.
[32] Bork, P., Downing, A.K., Kieffer, B. and Campbell, I.D. (1996) Q. Rev. Biophys. 29, 119–167.
[33] Yamagata, M., Yamada, K.M., Yamada, S.S., Shinomura, T., Tanaka, H., Nishida, Y., Obara, M. and Kimata, K. (1991) J. Cell Biol. 115, 209–221.
[34] Nishiyama, T., McDonough, A.M., Bruns, R.R. and Burgeson, R.E. (1994) J. Biol. Chem. 269, 28193–28199.
[35] Bowditch, R.D., Hariharan, M., Tominna, E.F., Smith, J.W., Yamada, K.M., Getzoff, E.D. and Ginsberg, M.H. (1994) J. Biol. Chem. 269, 10856–10863.
[36] Aota, S., Nomizu, M. and Yamada, K.M. (1994) J. Biol. Chem. 269, 24756–24761.