Volume 222, number 1, 83–88 FEB 05100 September 1987

Amino acid sequence of a 32-residue region around the thiol ester site in duck ovostatin

Hideaki Nagase and Keith Brew*

Department of Biochemistry, University of Kansas Medical Center, Kansas City, KS 66103 and **Department of Biochemistry, University of Miami School of Medicine, Miami, FL 33101, USA*

Received 9 July 1987

To obtain the amino acid sequence at the thiol ester site in duck ovostatin for comparisons with other proteins, the native ovostatin was labeled with $^{14}CH_3NH_2$ at the reactive thiol ester site. The modified protein was reduced, carboxymethylated, and digested with trypsin. ^{14}C-labeled peptides isolated by gel filtration with Sephadex G-50, ion-exchange chromatography on DEAE-cellulose and HPLC were subjected to automated sequence analysis, and the stretch of 32 amino acid residues containing the $^{14}CH_3NH_2$-binding site were determined. A comparison of this sequence with the corresponding sequences in α_2-macroglobulin, and complement components C3 and C4 revealed 72, 31 and 34% homology, respectively. The results indicate that ovostatin is a close relative to plasma α-macroglobulins and may share a common ancestor with C3 and C4.

Ovostatin; α_2-Macroglobulin; Thiol ester; Amino acid sequence

1. INTRODUCTION

Ovostatin (ovomacroglobulin) is a tetrameric glycoprotein proteinase inhibitor synthesized specifically in avian [1–3] and reptilian [4] oviduct. The native molecule (M_r 780000) consists of four identical subunits (M_r 195000) which are linked in pairs by disulfide bonds [3]. The binding of a proteinase to ovostatin is initiated by proteolysis of a specific region of a subunit. This event triggers a conformational change in ovostatin resulting in the entrapment of the enzyme molecule without blocking the active site [5]. Thus, inhibition of proteinases is observed only when high-M_r substrates are used. These observations indicate that the quaternary structure and the mechanism of proteinase inhibition by ovostatin are similar to those of the plasma glycoprotein, α_2-macroglobulin (α_2M) [5,6]. Indeed, the NH_2-terminal sequences of ovostatin and α_2M are sufficiently similar (about 36% identical in their first 14 residues) to suggest that the structural genes for ovostatin and α_2M may have evolved from a common ancestor [3,7,8]. In spite of their similarities, evolutionary divergence has resulted in differences in specificities and chemical properties. Chicken ovostatin inhibits only metalloproteinases stoichiometrically [5] and duck ovostatin inhibits metalloproteinases and serine proteinases but not cysteine proteinases [7], whereas α_2M can bind and inhibit almost all endopeptidases from four major classes of proteinases [6]. In addition, the most striking difference found between chicken ovostatin and α_2M is that chicken ovostatin lacks a reactive thiol ester [5]. As a result, chicken ovostatin, unlike α_2M, does not form a covalent linkage with the entrapped proteinase. Although the example of chicken ovostatin demonstrates that the thiol ester moiety is not essential for proteinases inhibition by this family of proteins, thiol esters have been found in all α-macroglobulins in

Correspondence address: H. Nagase, Department of Biochemistry, University of Kansas Medical Center, 39th and Rainbow Blvd, Kansas City, KS 66103, USA

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/87/$3.50

plasma from various species reported so far [3,9–13].

More recently, however, we have reported that such a thiol ester is present in duck ovostatin [7]. Thiol esters have also been found in complement components C3 [14–16] and C4 [17,18], and similarities in sequence around this region in α_2M and C3 and C4 have suggested that they share a common evolutionary origin [19,20].

In order to locate an internal thiol ester in duck ovostatin and to examine its structural relationship with α_2M, C3 and C4, we have sequenced 32 amino acid residues around the thiol ester site in duck ovostatin. The sequence analysis displays a high degree of similarity with the corresponding sequences of α_2M and to a lesser extent with those of C3 and C4. The results further support the hypothesis that parts of the structural genes for ovostatin, α_2M, C3 and C4, have arisen from a common ancestor.

2. MATERIALS AND METHODS

2.1. *Materials*

Duck ovostatin was purified as described [7]. $^{14}CH_3NH_2 \cdot HCl$ (spec. act. 56 mCi/mmol) was from Amersham. $CH_3NH_2 \cdot HCl$, dithiothreitol, iodoacetic acid and TPCK-treated trypsin were from Sigma. Sephadex G-50 (superfine) and DE-52 were from Pharmacia and Whatman, respectively.

2.2. *$^{14}CH_3NH_2$ labeling of duck ovostatin*

Labeling of duck ovostatin with $^{14}CH_3NH_2$ was first carried out by reacting 225 μl ovostatin (10.6 mg/ml) with 90 μl of 175 mM $^{14}CH_3NH_2$ (10 mCi/ml) in 0.1 M boric acid/NaOH, pH 8.6, containing 0.2 M NaCl. The final concentration of $^{14}CH_3NH_2$ was 50 mM, conditions under which the labeling reaction proceeds efficiently. After a 4 h reaction at 37°C, 1 ml duck ovostatin (10.6 mg/ml) was added to the mixture and it was further incubated at 37°C for 4 h. The sample was then mixed with 24 ml duck ovostatin (10.6 mg/ml) in 0.1 M boric acid/NaOH, pH 8.6, containing 0.2 M NaCl, incubated at 37°C for 1 h. The sample was then treated with 0.2 M CH_3NH_2 for 16 h at room temperature in order to obtain a chemically uniform product.

2.3. *Tryptic digestion of duck ovostatin*

$^{14}CH_3NH_2$-labeled ovostatin was dialyzed against 50 mM NH_4HCO_3, lyophilized, dissolved in 35 ml of 0.1 M Tris-HCl, pH 8.6/8 M urea, reduced with 10 mM dithiothreitol at 37°C for 4 h and alkylated with 30 mM iodoacetic acid. Reduced, carboxymethylated $^{14}CH_3NH_2$-labeled ovostatin was dialyzed against 50 mM NH_4HCO_3, lyophilized, and digested with 5 mg TPCK-trypsin in 20 ml of 50 mM NH_4HCO_3 containing 1 mM $CaCl_2$ for 2 h at 37°C, then with an additional 5 mg TPCK-trypsin for 2 h. Insoluble peptides generated were separated by centrifugation. More than 90% of the radioactivity was recovered in the supernatant. The tryptic fragments containing radioactivity were purified by a series of chromatographic techniques (see section 3). CNBr cleavage was carried out with a 2% (w/v) solution of CNBr in 70% formic acid for 18 h at 20°C.

2.4. *Amino acid analysis*

Peptides were hydrolyzed with 6 N HCl at 110°C for 24 h and amino acid compositions determined by conversion to PTC derivatives and analysis by HPLC. Hydrolysis and conversion was carried out with a Waters Pico Tag workstation, and analysis with a Waters Pico Tag HPLC column, following the procedures suggested by the manufacturer.

2.5. *Sequence analysis*

Automatic sequence analyses were performed with an Applied Biosystems 470A protein sequencer using standard programs. PTH-amino acids were identified and quantified by reverse-phase HPLC using a Hewlett Packard 1090 liquid chromatograph with an Applied Biosystems narrow bore (2.1 × 220 mm) PTH-C18 cartridge system. HPLC solvents were supplied by Applied Biosystems, and elution was carried out following their recommendations.

3. RESULTS AND DISCUSSION

3.1. *Isolation of $^{14}CH_3NH_2$-labeled peptides*

^{14}C-labeled peptides generated by TPCK-treated trypsin digestion were separated first with a column of Sephadex G-50 superfine (4.5 × 55 cm), equilibrated and eluted with 50 mM NH_4HCO_3. Three radioactive peaks were detected but only the

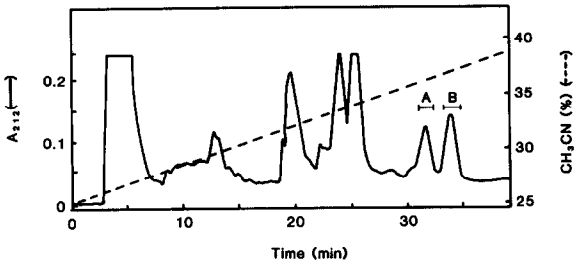

Fig.1. Separation of two ^{14}C-labeled peptides by reverse-phase HPLC. ^{14}C-labeled peptides obtained after DE-52 chromatography were dissolved in 2 ml of 6 M guanidine·HCl/15% formic acid, and 400 μl of the sample was applied to a reverse-phase C-18 column (Vydac) equilibrated with 0.1% trifluoroacetic acid. The peptides were eluted with a linear gradient at a flow rate of 1 ml/min. Each peak of A_{212} was collected and measured for radioactivity. Peaks A and B had 40 and 60% of the total radioactivity, respectively.

fractions containing the major peak were pooled, lyophilized, and subjected to further purification. The lyophilized material from the Sephadex G-50 column was dissolved in 10 mM NH_4HCO_3 and applied to a column of DE-52 (2 × 20 cm) equilibrated with the same solvent. The ^{14}C-labeled peptides bound to DE-52 and were eluted with a gradient of NH_4HCO_3. The majority of the ^{14}C label appeared in one major peak. The pool from DE-52 was lyophilized and dissolved in 2 ml of 6 M guanidine·HCl/15% formic acid. The fraction was further purified by HPLC (Bio Rad MAPS system) with a reverse-phase column (Vydac C18 protein and peptide column) using a gradient of 25–60% acetonitrile containing 0.1% trifluoroacetic acid (fig.1). The ^{14}C label was found in two peaks (peak A with 40% and peak B with 60% of the total radioactivity).

Table 1

Sequence of $^{14}CH_3NH_2$-labeled peptides of duck ovostatin

Cycle	Peptide A		Peptide B		Peptide A-3		
	Residue	Yield	Residue	Yield	Residue	Yield	cpm
1	Ser	112	Ala	1195	Pro	232	64
2	Val	318	Ser	380	Phe	234	40
3	Val	490	Phe	435	Gly	173	262
4	Gly	270	Ser	190	Cm-Cys	98	202
5	Asp	203	Val	490	Gly	148	240
6	Ile	324	Val	385	Glu	127	208
7	Met	230	Gly	260	$^{14}CH_3NH$-Glu	103	1375
8	Gly	202	Asp	180	Asn	46	1050
9	Thr	68			Met	N.D.	180
10	Ser	65					
11	Met	238					
12	Gln	195					
13	Asn	156					
14	Leu	290					
15	His	62					
16	Gln	124					
17	Leu	177					
18	Leu	246					
19	Gln	104					
20	Met	108					
21	Pro	99					
22	Phe	104					
23	Gly	81					
24	Cm-Cys	32					

Yields at each cycle are given in pmol. N.D., not determined

Table 2

Amino acid composition of peptides A, B, and A-3

	Peptide A	Peptide B	Peptide A-3
Asp	3.44 (3)	6.13 (6)	0.88 (1)
Glu	5.90 (5)	9.80 (10)	2.02 (2)
Cm-Cys	N.I. (1)	1.07 (1)	0.72 (1)
Ser	2.5 (3)	5.4 (5)	
Gly	4.07 (4)	6.9 (7)	2.30 (2)
His	0.97 (1)	1.6 (2)	
Arg			
Thr	1.7 (1)	2.55 (3)	
Ala		2.28 (2)	
Pro	0.87 (1)	1.75 (2)	0.89 (1)
Tyr	N.D.	N.D.	
Val	2.32 (2)	3.43 (3)	
Met	3.9 (4)	4.15 (4)	N.I. (1)
Ile	1.28 (1)	2.08 (2)	
Leu	2.74 (3)	5.45 (5)	
Phe	1.47 (2)	3.70 (4)	0.94 (1)
Lys		0.60 (1)	

N.I., not identified; N.D., not determined

3.2. *Sequence analyses of peptide A and peptide B*

From four separate runs of HPLC as shown in fig.1, the pools of peptides A and B were combined separately, dried under vacuum, and one-fifth of each peptide was used for analyses of sequence (table 1) and amino acid composition (table 2). The sequence obtained for the first 24 residues of peptide A appeared to correspond to residues 926–949 in human α_2M [8]. Cm-Cys identified in cycle 24 was possibly the Cys residue which is generated from the cleavage of the thiol ester bond between Cys and Glx analogously to α_2M [21–23]. One SH group is generated from each ovostatin subunit after treating the native duck ovostatin with CH_3NH_2 [7]. Peptide B showed an identical sequence to that of peptide A, but with three additional residues (Ala-Ser-Phe) at the amino-terminus of peptide A (table 1). Peptide B was probably generated by normal tryptic cleavage of duck ovostatin since in α_2M, the residue preceding this region is Arg [8], and peptide A was generated by cleavage at the Phe–Ser bond as a result of residual chymotryptic activity in the trypsin preparation. Indeed, the amino acid composition of peptide B revealed one residue of Lys, and peptide A appeared to be considerably shorter than peptide B (table 2).

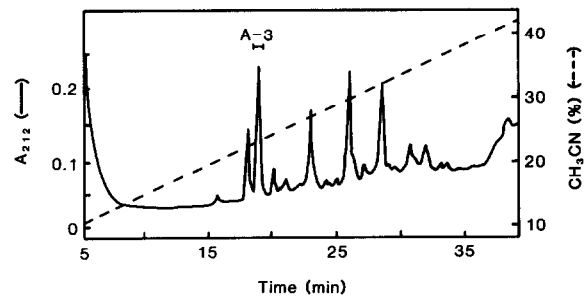

Fig.2. Separation of CNBr fragments of peptide A by reverse-phase HPLC. Peptide A was reacted with 200 μl of a 2% (w/v) solution of CNBr in 70% formic acid at 23°C for 18 h. After drying in vacuo over NaOH pellets, the products were dissolved in 6 M guanidine·HCl and applied to a reverse-phase C18 column (Vydac) attached to an HPLC system (Bio-Rad MAPS). The peptides were eluted with the linear gradient of acetonitrile at a flow rate of 1 ml/min. Each peak of A_{212} was collected and measured for radioactivity. ^{14}C label was found in peak A-3.

To determine a more extended sequence around the thiol ester site of duck ovostatin, the remaining peptide A was cleaved with CNBr and the fragments generated were separated by HPLC with a reverse-phase C-18 column (fig.2). ^{14}C label was only found in peptide A-3. Sequence analysis of A-3 showed that the peptide was generated by cleavage of the Met-Pro between residues 20 and 21 of peptide A (table 1). After each Edman degradation cycle of peptide A-3, the amount of ^{14}C label recovered in the PTH derivative was measured to identify the site of $^{14}CH_3NH_2$ incorporation. Radioactivity was recovered in cycle 7 and to a lesser extent in cycle 8 (table 1), indicating that $^{14}CH_3NH_2$ was covalently linked to residue 7 of peptide A-3. Identification of this residue as Glx was deduced from the amino acid composition of peptide A-3 (table 2) and by the presence of a PTH derivative that eluted at a distinct position from those of normal amino acids together with some PTH-Glu resulting from the cleavage of the methylamine moiety during the conversion of anilinothiazoline to a PTH derivative.

3.3. *Comparison of the sequences around the thiol ester sites of ovostatin, α_2M, C3 and C4*

At present, it appears that reactive inter-chain β-cysteinyl-γ-glutamyl thiol esters are uniquely found in complement components C3 and C4 and

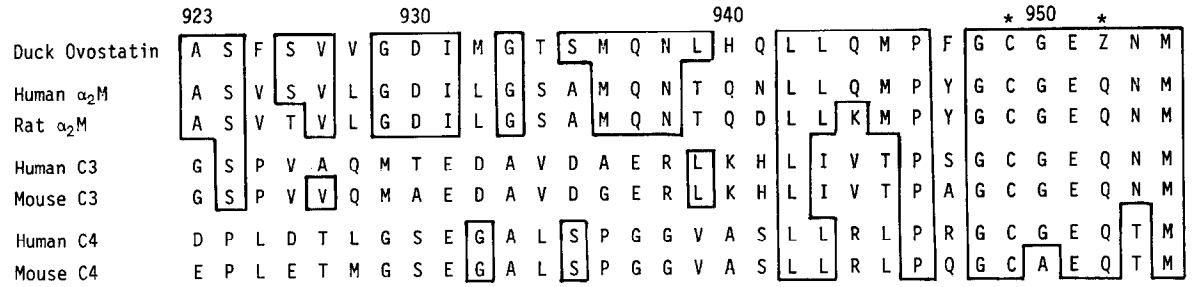

Fig.3. Sequence around the thiol ester sites of duck ovostatin, human [8] and rat [32] α_2M, human [33] and mouse [34] C3 and human [35] and mouse [36] C4. The box denotes residues identical to those in duck ovostatin. The numbers correspond to the residues in human α_2M [8]. Residues indicated by asterisks form thiol ester bonds in the native molecules.

the members of the α_2M family including pregnancy zone protein [24], a 360-kDa α-macroglobulin homologue from fish plasma [13], murinoglobulin [25], rat plasma α_1-inhibitor III [26] and α_2M-like inhibitor from horseshoe crab [27].

Although the functions of the complement factors and proteinase inhibitors are different, the presence of a thiol ester in the proteins is reflected in some common functional properties [19,28]: (i) the proteinase-binding capacity of most α-macroglobulins and the hemolytic activities of C3 and C4 are lost upon incubation with small nucleophilic amines, where these nucleophiles react to form covalent bonds with activated residues; (ii) upon reacting with proteinases, thiol ester groups in α_2M become activated and allow α_2M to bind to an entrapped proteinase covalently while C3 and C4 are coupled to either IgG or to erythrocyte membranes upon activation; (iii) on reacting with either proteinases or nucleophilic amines, free SH groups are generated; and (iv) during heat denaturation characteristic autolytic fragments are produced. We have previously reported that the above effects are not observed with chicken ovostatin because it lacks a thiol ester [5]. The studies reported here demonstrate at the sequence level that in duck ovostatin the thiol ester is conserved, indicating that ovostatins are close relatives of the α_2M group. In the comparison of the 32 residues around the methylamine-labeling site of duck ovostatin with the equivalent region of human α_2M, C3 and C4 (fig.3), 23 residues were found to be identical with those of human α_2M (72% homology) but only 10 and 11 residues identical to those of C3 (31% homology) and C4 (34% homology). Since the sizes of the autolytic fragments (M_r 123000 and M_r 60000) are similar to those produced from α_2M [7,30], and the sequence around the thiol ester displays strong homology between the two molecules, the thiol ester group in duck ovostatin appears to be located at an equivalent site in the sequence to that of α_2M (about 2/3 away from the NH_2-terminus). CH_3NH_2 sensitivity has been also reported for crocodile ovostatin [29].

The presence of an α_2M-like molecule containing a CH_3NH_2-sensitive site in horseshoe crab indicates that the proteins of this family arose at least 550 million years ago [27,31] and have been conserved during the course of evolution. Furthermore, the sequence around this region is highly conserved even in C3 and C4 (fig.3). This may reflect the requirement of a specific sequence and conformation for the formation of the thiol ester group during protein folding although the functions of these proteins are different from those of the proteins in the α_2M family. It is, however, enigmatic that chicken ovostatin lacks a thiol ester in spite of its close structural similarity to the duck protein as reflected in their identical amino-terminal 14-residue sequences [7] and their immunological partial cross-reactivity. In contrast, only 36% homology is found in the NH_2-termini of ovostatin and α_2M [7] and they are immunologically distinct [1,3]. Yet, there is 72% homology between duck ovostatin and human α_2M in the region containing the thiol ester, and all proteins in the α_2M family described in the literature, except

chicken ovostatin, have CH_3NH_2-sensitive sites, presumably reflecting the presence of thiol esters. To understand the evolutionary and functional relationship of this family of proteins, we are currently investigating the sequence in chicken ovostatin which corresponds to the thiol ester sites of other members of this family using molecular cloning techniques.

ACKNOWLEDGEMENTS

This work was supported by National Institutes of Health grants HD 18796 and GM 21363 and a Biomedical Research Support Grant from the University of Kansas Medical Center.

REFERENCES

[1] Miller, H.T. and Feeney, R.E. (1966) Biochemistry 6, 952–958.
[2] Kitamoto, T., Nakashima, M. and Ikai, A. (1982) J. Biochem. (Tokyo) 92, 1679–1682.
[3] Nagase, H., Harris, E.D. jr, Woessner, J.F. jr and Brew, K. (1983) J. Biol. Chem. 258, 7481–7489.
[4] Ikai, A., Kitamoto, T. and Nishigai, M. (1983) J. Biochem. (Tokyo) 93, 121–127.
[5] Nagase, H. and Harris, E.D. jr (1983) J. Biol. Chem. 258, 7490–7498.
[6] Starkey, P.M. and Barrett, A.J. (1977) in: Proteinases in Mammalian Cells and Tissues (Barrett, A.J. ed.) pp.663–696, Elsevier/North-Holland, Amsterdam, New York.
[7] Nagase, H., Harris, E.D. jr and Brew, K. (1986) J. Biol. Chem. 261, 1421–1426.
[8] Sottrup-Jensen, L., Stepanik, T.M., Kristensen, T.E., Wierzbick, D.M., Jones, C.M., Lønblad, P.B., Magnusson, S. and Peterson, T.E. (1984) J. Biol. Chem. 259, 8318–8327.
[9] Barrett, A.J. (1981) Methods Enzymol. 80, 737–754.
[10] Björk, I., Lindblom, T. and Lindahl, P. (1985) Biochemistry 24, 2653–2660.
[11] Gonias, S.G., Balber, A.E., Hubbard, W.J. and Pizzo, S.V. (1983) Biochem. J. 209, 99–105.
[12] Feldman, S.R. and Pizzo, S.V. (1985) Biochemistry 24, 2569–2575.
[13] Starkey, P.M. and Barrett, A.J. (1982) Biochem. J. 205, 105–115.
[14] Howard, J.B. (1980) J. Biol. Chem. 255, 7082–7084.
[15] Tack, B.F., Harrison, R.A., Jonatova, J., Thomas, M.G. and Prahl, J.W. (1980) Proc. Natl. Acad. Sci. USA 77, 5764–5768.
[16] Pangburn, M.K. and Müller-Eberhard, H.J. (1980) J. Exp. Med. 152, 1102–1114.
[17] Campbell, R.D., Gagnon, J. and Porter, R.R. (1981) Biochem. J. 199, 359–370.
[18] Harrison, R.A., Thomas, M.L. and Tack, B.F. (1981) Proc. Natl. Acad. Sci. USA 78, 7388–7392.
[19] Sim, R.B. and Sim, E. (1981) Biochem. J. 193, 129–141.
[20] Sottrup-Jensen, L., Stepanik, T.M., Kristensen, T., Lønblad, P.B., Jones, C.M., Wierzbicki, D.M., Magnusson, S., Domdey, H., Wetsel, R.A., Lundwall, A., Tack, B.F. and Fey, G.H. (1985) Proc. Natl. Acad. Sci. USA 82, 9–13.
[21] Howard, J.B. (1981) Proc. Natl. Acad. Sci. USA 78, 2235–2239.
[22] Sottrup-Jensen, L., Hansen, H.F., Mortensen, S.B., Peterson, T.E. and Magnusson, S. (1981) FEBS Lett. 123, 145–148.
[23] Salvesen, G.S., Sayers, C.A. and Barrett, A.J. (1981) Biochem. J. 195, 453–461.
[24] Sand, O., Folkersen, J., Westergaard, J.G. and Sottrup-Jensen, L. (1985) J. Biol. Chem. 260, 15723–15735.
[25] Saito, A. and Sinohara, H. (1985) J. Biol. Chem. 260, 775–781.
[26] Esnard, F., Gutman, N., Moujahed, E. and Gauthier, F. (1985) FEBS Lett. 182, 125–129.
[27] Quigley, J.P. and Armstrong, P.B. (1983) J. Biol. Chem. 258, 7903–7906.
[28] Tack, B.F. (1983) Springer Semin. Immunopathol. 6, 259–282.
[29] Arakawa, H., Osada, T. and Ikai, A. (1986) Arch. Biochem. Biophys. 244, 447–453.
[30] Harpel, P., Hayes, M.B. and Hugli, T.E. (1979) J. Biol. Chem. 254, 8669–8678.
[31] Sepkoski, J.J. (1978) Paleobiology 4, 223–251.
[32] Gehring, M.R., Shiels, B.R., Northemann, W., DeBruihn, M.H.L., Kan, C.-C., Chain, A.C., Noonan, D.J. and Fey, G.H. (1987) J. Biol. Chem. 262, 446–454.
[33] DeBruihn, M.H.L. and Fey, G.H. (1985) Proc. Natl. Acad. Sci. USA 82, 708–712.
[34] Wetsel, R.A., Lundwall, Å., Davidson, F., Gibson, T., Tack, B.F. and Fey, G.H. (1984) J. Biol. Chem. 259, 13857–13862.
[35] Belt, K.T., Carroll, M.C. and Porter, R.R. (1984) Cell 36, 907–914.
[36] Nonaka, M., Nakayama, K., Yeul, Y.D. and Takahashi, M. (1985) J. Biol. Chem. 260, 10936–10943.