Pergamon

Bioorganic & Medicinal Chemistry 7 (1999) 375–379

BIOORGANIC & MEDICINAL CHEMISTRY

Conformational Studies on the Specific Cleavage Site of Type I Collagen (α-1) Fragment (157–192) By Cathepsins K and L by Proton NMR Spectroscopy

Atsuko Y. Nosaka,* Kenji Kanaori,† Naoki Teno, Hiroko Togame, Tetsuya Inaoka, Michihiro Takai and Toshio Kokubo

International Research Laboratories, Ciba–Geigy Japan Ltd., PO Box 1, Takarazuka 665-8666, Japan

Received 14 August 1998; accepted 13 October 1998

Abstract—Cathepsins K and L are cysteine proteinases which are considered to play an important role in bone resorption. Type I collagen is the most abundant component of the extracellular matrix of bone and regarded as an endogenous substrate for the cysteine proteinases in osteoclastic bone resorption. We have synthesized a fragment of Type I collagen (α-1) (157–192) as a substrate for the cathepsins and found that cathepsins K and L cleave the fragment at different specific sites. The major cleavage sites for cathepsin K were Met159-Gly160, Ser162-Gly163 and Arg165-Gly166, while those for cathepsin L were Gly166-Leu167 and Gln180-Gly181. The structure of the fragment was analyzed in aqueous solution by circular dichroism and proton NMR spectroscopy and the difference in the molecular recognition of collagen by cathepsins K and L was discussed from the structural aspect.

Introduction

Bone formation and resorption are well balanced between osteoclast and osteoblast cells in the process of bone remodeling. Imbalances caused by a net increase in osteoclast activity result in excessive bone resorption and lead to pathological conditions such as osteoarthritis and osteoporosis. Osteoclasts are multinuclear cells that solubilize hydroxyapatite and degrade protein matrix which mainly consists of Type I collagen fibers, through secretion of proteolytic enzymes into an extracellular, low pH compartment on the bone surface. It has been reported that cysteine proteinase inhibitors effectively reduce bone resorption, both in vitro and in vivo.[1,2] Cathepsin L, first isolated from rat liver lysosome[3] was found to be considerably active among the lysosomal cysteine proteases with regard to the ability to hydrolyze azocasein,[4] elastin,[5] and collagen.[6] From previous studies on substrate preference, inhibitor preference and immunoreactivity, cathepsin L was suggested to be responsible for the degradation of proteinaceous matrix in the osteoclastic bone resorption process.[7–10] Cathepsin K is also a recently discovered cysteine proteinase, initially cloned from rabbit osteoclasts,[11] that is selectively and highly expressed in human osteoclasts and osteoclastomas.[11–17] Cathepsin K is highly active against Type I collagen and elastin, at pH values where cathepsin L is at best minimally active.[7] Gelb et al. have recently linked defects in the gene encoding cathepsin K to pycnodysostosis, a disease characterized by skeletal defects such as dense, brittle bones, short stature and poor bone remodeling.[18] Thus, current data suggest that osteoclastic bone resorption is mostly linked to the activity of cathepsins L and K.[13,19] Hence the design for potent inhibitors against the cathepsins would make important contributions in arresting the pathological processes, such as osteoarthritis, osteoporosis, and multiple myeloma osteoclastomas. In order to design efficacious and specific therapeutics, it is essential to obtain the information about their proteolytic specificity based on the molecular recognition.

The bone matrix is approximately 90% Type I collagen and it is regarded as an endogenous substrate for proteinases, e.g. collageneses and cathepsins, in osteoclastic bone resorption.[20–22]

In this study, we have synthesized a short fragment of Type I collagen [Type I collagen(α-1) fragment (157–192)] which consists of 36 amino acid residues as a substrate for cathepsins K and L and found that the fragment is specifically cleaved at several sites differently by cathepsins K and L. The solution structure of the fragment was analyzed by proton NMR spectroscopy

Key words: Cathepsin; collagen; NMR; temperature coefficient; CD.
*Corresponding author. Tel: +81-258-47-9315; fax: +81-258-47-9315.
†Present address: Department of Polymer Science and Engineering, Kyoto Institute of Technology, Kyoto 606-8585, Japan.

0968-0896/99/$ - see front matter
PII: S0968-0896(98)00227-2

and the difference in the molecular recognition of the collagen fragment by the cathepsins was discussed from the structural aspect to provide clues for understanding the substrate specificities of these cysteine proteinases.

Results

Identification of the cleavage sites of the fragment

As shown in Figure 1, the hydrolytic cleavage of the collagen fragment with cathepsin L was found to be specific. Two major cleavage sites were identified to be Gly166-Leu167 and Gln180-Gly181 and relative rates of the cleavage were 74 and 100, respectively.

On the other hand, specificity of the cleavage by cathepsin K was different from that by cathepsin L. Three major cleavage sites were Met159-Gly160, Ser162-Gly163, and Arg165-Gly166. The relative rates of the cleavage were 86, 100 and 89, respectively. To our knowledge this is the first characterization of substrate specificity of cathepsin K. Thus the results clearly indicated that the both cathepsins which share high homology, do show different substrate specificity.

Circular dichroism (CD) measurements

CD spectrum of the fragment in water at pH 3.0 (Fig. 2) shows no discernible features between 240 and 200 nm, suggesting that the peptide is present under this condition in a predominantly random conformation.

NMR assignments of the proton resonances

The resonance assignment of the proton NMR signals of the fragment in acetic acid solution was performed using the conventional sequential assignment procedure at pH 3.0 and 298 K. The identification of amino acid spin systems was first established by means of direct (DQF-COSY),[23,24] and relayed through-bond connectivities (HOHAHA),[25,26] followed by sequential resonance assignments using through-space NOE connectivities.[27,28] However, some of the NH-CαH peaks were not clearly separated in the fingerprint region of DQF-COSY, and sequential NOE cross-peaks of ROESY spectra (typically, NH(i)-NH($i+1$), CαH(i)-NH($i+1$) and CαH(i)-NH($i+1$)). Figure 3 shows the finger print region of ROESY[27,28] spectrum of the fragment. The unambiguous assignment was achieved for the backbone protons (Hα and NH). However, for the side chain protons the complete assignment was difficult

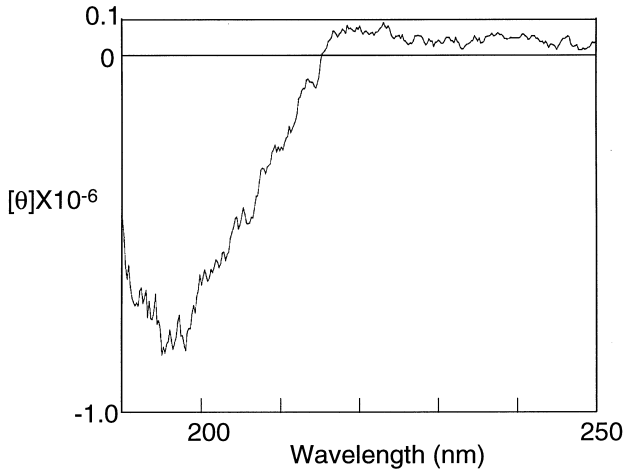

Figure 2. CD spectrum of Type I collagen (α-1) fragment (157–192). Sample concentration was 80 μM in 38 mM acetic acid solution (pH 3.0). The ellipticity units/mole peptide residue (Θ) are given in deg·cm^2·dmol^{-1}.

because the fragment contains 12 glycine, 6 proline, and 6 hydroxy proline residues, and those proton NMR signals severely overlap each other particularly in the sequence regions of Hyp-Gly-X. No medium nor long-range NOEs were observed in the ROESY spectra besides intra-residual and sequential NOEs. Thus, the NOE results also indicate no rigid secondary conformation of the fragment in aqueous solution.

Temperature coefficient

Temperature dependence of amide-proton chemical shifts ($\Delta\delta_{NH}/\Delta T$) has been shown to correlate with the presence of intramolecular hydrogen bonds and can give a measure of the protection of the amide protons from the solvent.[29] Lowered temperature coefficients are usually taken to indicate some degree of intramolecular hydrogen bonding, generally as part of an element of secondary structure such as a turn, helix, or β-hairpin.[30] In Figure 4 the temperature coefficients of the amide proton chemical shifts of Type I collagen (α-1) fragment (157–192) measured in aqueous solution were plotted as a function of the residue number. The temperature dependence of the amide protons was high for most parts of the molecule (between 6 and 10 ppb/K), indicating that there were no hydrogen bonds present. However, one clearly outstanding value of only 3.5 ppb/K was obtained for NH of residue 181, indicating the existence of a hydrogen bond. Although present CD and NOE results indicate that the major population of

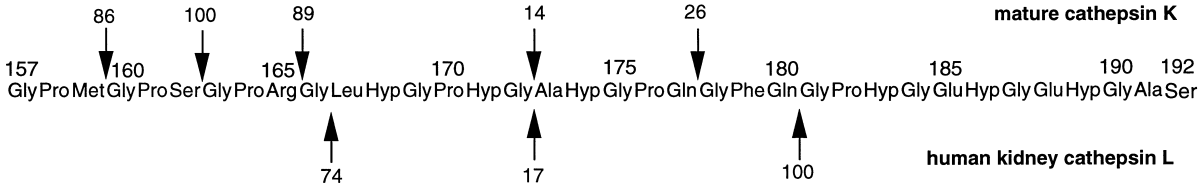

Figure 1. A sequence of Type I collagen (α-1) fragment (157–192) with cleavage sites and relative digestion rates by cathepsin K (upper) and cathepsin L (lower).

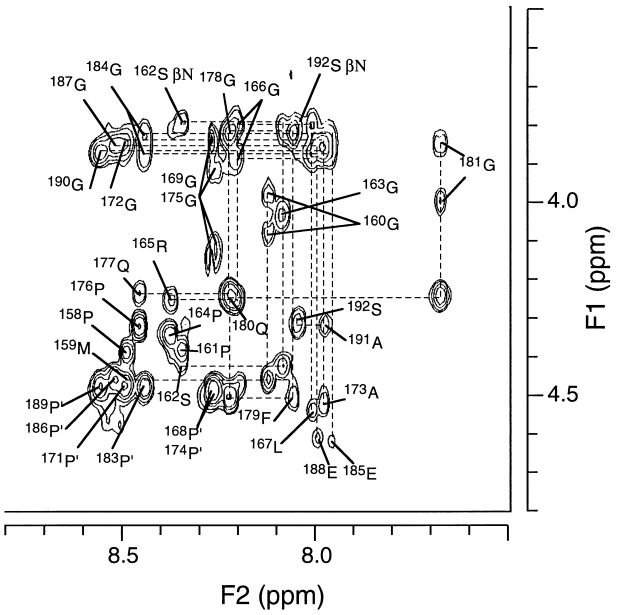

Figure 3. NH-CαH region of ROESY spectrum (mixing time 200 ms) of Type I collagen (α-1) fragment (157–192) recorded at pH 3.0 and 300 K. Intra-residue NH-CαH cross-peaks are labeled by residue name and number. Cross peaks labeled by P or P′ indicate sequential conectivities between α proton of the proline or hydroxy proline residue and NH proton of the following residue. Dashed lines indicate other sequential connectivities.

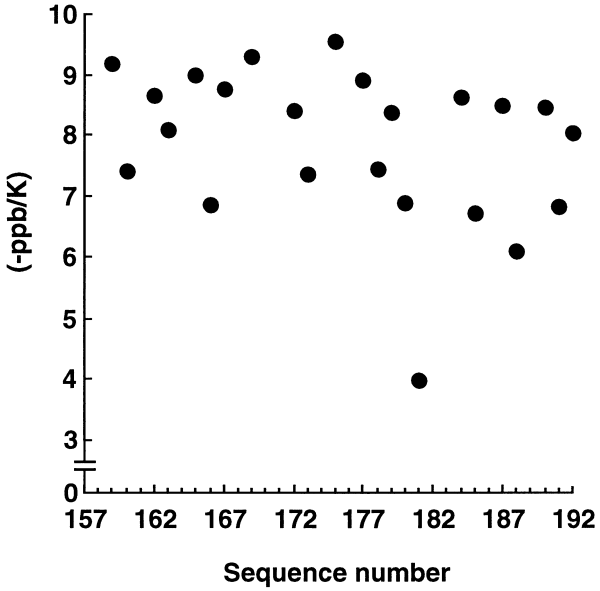

Figure 4. Amide temperature coefficients of Type I collagen (α-1) fragment (157–192) derived as (-ppb/K) values versus amino acid sequence of the peptide in aqueous solution. The temperature coefficients were determined from a series of HOHAHA spectra recorded at from 280 to 320 K by every 10 degrees.

the Type I collagen (α-1) fragment (157–192) takes a random structure in aqueous solution, this small temperature coefficient value for Gly181 suggests that this part of the peptide is involved in an ordered structure, less accessible to the solvent. Since Gly and Pro residues are known to readily take a β-turn, the residues around Gly181 are possibly involved in a turn structure. Although medium range NOEs were not observed for the residues around Gly181, NOE cross peaks might be weakened by the molecular motion and by the heterogeneity of the local conformations. It should be noted that the residue 181 coincides exactly with the cleavage site of the fragment by cathepsin L.

Discussion

It has been demonstrated that Type I collagen is critical for bone development.[31] The helical trimeric molecules of Type I collagen are comprised of two α1(I) chains and one α1(II) chain. The present cleavage experiment indicates that cathepsins K and L have the activity of digestion of the non-helical collagen strand of the Type I collagen (α-1) fragment (157–192), which can be regarded as the denatured form of the intact collagen monomers. Although both the cathepsins share high homology (41%), the preference of the cleavage site of the Type I collagen fragment significantly differs between cathepsins K and L. Cathepsin L cleaves the fragment at Gly166-Leu167 and Gln180-Gly181, whereas cathepsin K preferentially cleaves the fragment at the three sites: Met159-Gly160, Ser162-Gly163 and Arg165-Gly166.

The α chains of Type I collagen are constructed from repeating Gly-X-Y triplets, where X and Y are most frequently Pro and Hyp, respectively, although any other residue can appear in these positions. Cathepsin L recognized non-consensus regions of the peptide, such as Gly166-Leu167 and Gln180-Gly181. Taking account of the small temperature coefficient value of 3.5 ppb/K for Gly181 and the high digestion rate (100) at Gln180-Gly181 by cathepsin L, cathepsin L may recognize a certain structural factor of the substrate like a turn structure. The other cleavage site (Gly166-Leu167) also showed a relatively low temperature coefficient (ca. 6 ppb/K) which might also be indicative of relatively restricted structure as compared to the other parts of the peptide. On the other hand, cathepsin K recognized only the specific consensus sequence of the Type I collagen (α-1) fragment (157–192). The peptide contains six -Gly-Pro-Z-Gly- sequences as shown in Figure 1: 157–160 (z; Met), 160–163 (z; Ser), 163–166 (z; Arg), 169–172 (z; Hyp), 175–178 (z; Gln) and 181–184 (z; Hyp). Cathepsin K recognized and exerted the former three quartet residues (z; Met, Ser, or Arg) and the fifth quartet residue (z; Gln) but not the other two (z; Hyp). Actually, the mapping study of the cleavage sites for the various collagen fragments demonstrated that cathepsin K preferably cleaved the various collagen fragments at the peptide bond between Z and Gly of -Gly-Pro-Z-Gly-, where Z is Met, Arg, Ser or Gln, while cathepsin L cleaved them mostly in non-consensus region (The data will be published elsewhere). Thus present results indicate that cathepsin K cleaves Type I collagen (α-1) fragment (157–192) by recognizing the specific consensus residues in the extended strand while the recognition by cathepsin L involves both the conformational preference and the specific residues other than the consensus sequence.

Cathepsins K and L degrade Type I collagen.[13,32] It is, however, unclear whether the cleavage occurs in the intact helical region or in unraveled α monomers. At the ends of the triple helical domain, short non-helical peptides, namely the telopeptides, having a non-repeating sequence and spanning from 9 to 25 residues, are located. It was reported that the primary cleavage by cathepsin K seems to occur in the telopeptide region. Following cleavage may occur within the α monomers. On the other hand, cathepsin L is considered to cleave Type I collagen only in the telopeptide region.[13] In addition to the collagenase activity, cathepsins K and L display a powerful gelatinase activity. At 0.1 nM concentration of the enzyme, denatured collagen is totally degraded within 30 min by cathepsin K.[13] The monomers of the Type I collagen contain the consensus sequence more than 20%. This high number of cleavage sites would result in a stronger collagenolytic and gelatinase activities exhibited by cathepsin K than cathepsin L.

The difference in substrate specificities between cathepsins K and L implies different functions of both the cathepsins in bone remodeling and the possibility to develop selective inhibitors against cathepsins K and L, leading to the osteoporosis therapeutics. Further investigation on the interaction of both the proteinase with various length and sequence of the substrate would provide more unequivocal information about the cleavage processes. Assignment of a novel cleavage site would also provide further insights into the mechanisms of collagen turnover in development and disease.

Conclusion

Osteoclastic bone resorption is currently considered to be correlated with the activity of two cysteine proteinases, cathepsins K and L. Type I collagen is regarded as an endogenous substrate for both the proteinases. We synthesized a short fragment of Type I collagen [Type I collagen (α-1) fragment (157–192)] which consists of 36 amino acid residues as a model substrate. The distinct difference in the cleavage sites of the fragment by cathepsins K and L implies that the both cathepsins which share high homology, do show different substrate specificity. Cathepsin K cleaves Type I collagen (α-1) fragment (157–192) by recognizing the specific consensus residues in the extended strand, namely -Gly-Pro-Z-Gly-, where Z is Met, Arg, Ser or Gln, whereas the recognition by cathepsin L seems to involve both the conformational preference and the specific residues other than the consensus sequence. Thus the substrate specificities strongly suggest different functions of both cathepsins in the process of bone remodeling.

Experimental

Peptide synthesis

Type I collagen (α-1) fragment (157–192) was synthesized by a solid phase fragment assembly.[33–36] The fragment was divided into 5 sub-fragments, which were each synthesized by the Fmoc method. To the C-terminus the sub-fragment synthesized on the HMPB-BHA-resin[37] was successively coupled with other 4 sub-fragments to give the 36 residues long peptide.

Digestion of the peptide

Type I collagen (α-1) fragment (157–192) was subjected to digestion with cathepsins L and K. The purity of the enzymes is both more than 95% (SDS-PAGE). Cathepsins L and K show the hydrolytic activity between pH 3.5 and 5.5, and between pH 4.5 and 6.0, respectively. Optimal pH was about 5.0 and 5.5, for cathepsins L and K, respectively.

The fragment (0.1 mM) was hydrolyzed with 3.3 μg/mL of the mature form of human kidney cathepsin L at 300 K in 100 mM citrate buffer (pH 5.5) containing 1 mM EDTA and 3.5 mM 2-mercaptoethanol. Products of the proteolytic cleavage were analyzed by reverse phase HPLC after 1, 2, 4 and 24 h digestion. All well-separated peaks on HPLC were collected from the reaction mixture after 24 h digestion and subjected to MALDI-TOF MS and amino acid sequencing to identify cleavage sites. Hydrolysis of the fragment by cathepsin K was performed by using the mature form of human cathepsin K. We examined the hydrolysis with 0.1 mM of the peptide substrate and 0.06 μg/mL of cathepsin K at 300 K in 100 mM phosphate buffer (pH 6.0) containing 1 mM EDTA, 100 mM NaCl and 2 mM dithiothreitol. The reactions were monitored by reverse phase HPLC over 24 h. The cleavage sites were determined by sequencing and mass analysis of each hydrolyzed fragment which was isolated by HPLC. Addition of a cysteine proteinase inhibitor, egg white cystatin, completely blocked the hydrolysis, and addition of a high concentration (80 mM) of cysteine accelerated the reaction considerably, confirming that the cleavage is actually catalyzed by the cysteine proteinase.

The area of each fragment on HPLC increased linearly against time. The cleavage rates were determined with the slope up to 2 h and the relative rates to that at the maximum cleavage site (100) were shown in Figure 1.

CD spectroscopy

The fragment was dissolved in 38 mM acetic acid solution to a final concentration of 80 μM. The CD spectra were recorded at ambient temperature in a quartz cell of path length 0.01 cm on a JASCO J-600 spectropolarimeter at pH 3.0. All CD spectra were reported in terms of ellipticity units per mole of peptide residue. The mean residue weight was calculated from the amino acid composition.

Proton NMR spectroscopy

All proton NMR spectra were recorded on a Bruker AMX-600 spectrometer. The derived data were processed on a Bruker X-32 computer with UXNMR software. For resonance assignments, double-quantum

filtered shift correlated spectroscopy (DQF-COSY),[23,24] and homonuclear Hartmann–Hahn spectroscopy (HOHAHA),[25,26] nuclear Overhauser enhancement (NOESY) and rotating frame nuclear Overhauser effect spectroscopy (ROESY),[27,28] were run according to the time proportional phase incrementation method (TPPI).[38] ROESY spectra were recorded with mixing times of 200 and 350 ms, and HOHAHA experiments were acquired with mixing time of 60 ms. The data size was 2048 complex points for f_2 and 512 points for f_1 and repetition delay was 1.8 s. The f_1 and f_2 data were apodized with the squared sinebell function and then Fourier transformed. The water OH resonance was suppressed by irradiation during the relaxation delay. 3-(trimethylsilyl)propionate (TSP)-d_4 was used as an internal standard of chemical shift. The fragment was dissolved in 90% H_2O/10% D_2O (v/v) containing 50 mM of CD_3COOD (pH was adjusted by adding ND_4OD to 3.0). All the NMR measurements were performed at pH 3.0 to avoid signal broadening of the NH protons by the exchange with solvent protons. We assumed the structure of the fragment is the same as that at physiological pH. The final concentration of the peptide was 5 mg/200 μL (7.5 mM). After dissolution of the peptide in D_2O solution in a sample tube, the sample was inserted into the NMR magnet and thermally equilibrated for 30 min. The temperature coefficients of the amide proton chemical shifts were measured with a series of HOHAHA spectra. The amide proton chemical shifts of each residue were monitored by changing temperature from 280 to 320 K, by every 10 degrees.

References and Notes

1. Vaes, G. *J. Cell Biol*. **1968**, *39*, 676.
2. Baron, R.; Neff, L.; Louvard, D.; Courtoy, P. J. *J. Cell Biol*. **1985**, *101*, 2210.
3. Kirschke, H.; Langer, J.; Wiederanders, B.; Ansorge, S.; Bohley, P. *Eur. J. Biochem*. **1977**, *74*, 293.
4. Barrett, A. J.; Kirschke, H. *Methods Enzymol*. **1981**, *80*, 535.
5. Mason, R. W.; Johnson, D. A.; Barret, A. J.; Chapman, H. A. *Biochem. J*. **1982**, *226*, 233.
6. Kirschke, H.; Kembhavi, A. A.; Bohley, P.; Barrett, A. J. *Biochem. J*. **1982**, *201*, 367.
7. Koga, H.; Yamada, H.; Nishimura, Y.; Kato, K.; Imoto, T. *J. Biochem*. **1990**, *108*, 976.
8. Kageshita, T.; Yoshii, A.; Kimura, T.; Maruo, K.; Ono, T.; Himeno, M.; Nishimura, Y. *Arch. Dermatol. Res*. **1995**, *287*, 266.
9. Kakegawa, H.; Nikawa, T.; Tagami, K.; Kamioka, H.; Sumitani, K.; Kawata, T.; DrobnicKosorok, M.; Lenarcic, B.; Turk, V.; Katsunuma, N. *FEBS Lett*. **1993**, *321*, 247.
10. Delaisse, J. M.; Ledent, P.; Eckhout, Y.; Vaes, G. *Cysteine Proteinases and Their Inhibitors*, Walter de Gruyter: Berlin, 1986; pp. 259–267.
11. Tezuka, K.; Tezuka, Y.; Maejima, A.; Sato, T.; Nemoto, K.; Kamioka, H.; Hakeda, Y.; Kumegawa, M. *J. Biol. Chem*. **1994**, *269*, 1106.
12. Brömme, D.; Okamoto, K. *Biol. Chemi. Hoppe-Seyler*. **1995**, *376*, 379.
13. Brömme, D.; Okamoto, K.; Wang, B. B.; Biroc, S. *J. Biol. Chem*. **1996**, *271*, 2126.
14. Drake, F. H.; Dodds, R. A.; James, I. E.; Connor, J. R.; Debouck, C.; Richardson, S.; Lee-Rykaczewski, E.; Coleman, L.; Rieman, C.; Barthlow, R.; Hastings, G.; Gowen, M. *J. Biol. Chem*. **1996**, *271*, 12511.
15. Bossard, M. J.; Tomaszek, T. A.; Thompson, S. K.; Amegadzie, B. Y.; Hanning, C. R.; Jones, C.; Kurdyla, J. T.; McNulty, D. E.; Drake, F. H.; Gowen, M.; Levy, M. A. *J. Biol. Chem*. **1996**, *271*, 12517.
16. Inaoka, T.; Bilbe, G.; Ishibashi, O.; Tezuka, K.; Kumegawa, M.; Kokubo, T. *Biochem. Biophys. Res. Commun*. **1995**, *206*, 89.
17. Shi, G.-P.; Chapman, H. A.; Bhairi, S. M.; DeLeeuw, C.; Reddy, V. Y.; Weiss, S. J. *FEBS Lett*. **1995**, *357*, 129.
18. Gelb, B. D.; Shi, G. P.; Chapman, H. A.; Desnick, R. J. *Science* **1996**, *274*, 1236.
19. Delaisse, J. M.; Boyde, A.; Maconnachie, E.; Ali, N. N.; Sear, C. H. J.; Eeckhout, Y.; Vaes, G.; Jones, S. J. *Bone* **1987**, *8*, 305.
20. Burleigh, M. C.; Barrett, A. J.; Lazarus, G. S. *Biochem. J*. **1974**, *137*, 387.
21. Evans, P.; Etherington, D. J. *Eur. J. Biochem*. **1978**, *83*, 87.
22. Maciewicz, R. A.; Etherington, D. J. *Biochem. Soc. Trans*. **1985**, *13*, 1169.
23. Piantini, U.; Sørensen, O. W.; Ernst, R. R. *J. Am Chem. Soc*. **1982**, *104*, 6800.
24. Rance, M.; Sørensen, O. W.; Bodenhausen, G.; Wagner, G.; Ernst, R. R.; Wüthrich, K. *Biochem. Biophys. Res. Commun*. **1983**, *117*, 479.
25. Braunschweiler, L.; Ernst, R. R. *J. Magn. Reson*. **1981**, *53*, 521.
26. Davis, D. G.; Bax, A. *J. Am. Chem. Soc*. **1985**, *107*, 2820.
27. Jeener, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R. *J. Chem. Phys*. **1979**, *71*, 4546.
28. Macura, S.; Huang, Y.; Suter, D.; Ernst, R. R. *J. Magn. Reson*. **1981**, *43*, 259.
29. Dyson, H. J.; Rance, M.; Houghten, R. A.; Wright, P. E. *J. Mol. Biol*. **1988**, *201*, 161.
30. Rose, G. D.; Gierasch, L. M.; Smith, J. A. *Adv. Protein Chem*. **1985**, *37*, 1.
31. Liu, X.; Wu, H.; Byrne, M.; Jeffrey, J.; Krane, S.; Jaenisch, R. *J. Cell Biol*. **1995**, *130*, 227.
32. Sanesige, S.; Mano, H.; Tezuka, K. I.; Kakudo, S.; Mori, Y.; Honda, Y.; Itabashi, A.; Yamada, T.; Miyata, K.; Hakeda, Y.; Ishii, J.; Kumegawa, M. *Biochem. J*. **1995**, *309*, 721.
33. Barlos K.; Gatos, D.; Schäfer, W. *Angew. Chem. Int. Ed. Engl*. **1991**, *30*, 590.
34. Lloyd-Williams, P.; Albericio, F.; Giralt, E. *Tetrahedron* **1993**, *49*, 11065.
35. Benz, H. *Synthesis* **1994**, *337*.
36. Athanassopoulos, P.; Barlos, K.; Gatos, D.; Hatzi, O.; Tzavara, C. *Tetrahedron Lett*. **1995**, *36*, 5645.
37. Riniker, B.; Flörsheimer, A.; Fretz, H.; Sieber, P.; Kamber, B. *Tetrahedron* **1993**, *49*, 9307.
38. Marion, D.; Wüthrich, K. Biochem. *Biophys. Res. Commun*. **1983**, *113*, 967.