Amino acid sequence of CAP37, a human neutrophil granule-derived antibacterial and monocyte-specific chemotactic glycoprotein structurally similar to neutrophil elastase

Jan Pohl[1], H. Anne Pereira[2], Nancy M. Martin[2] and John K. Spitznagel[2]

[1]*Microchemical Facility, Winship Cancer Center* and [2]*Department of Microbiology and Immunology, Emory University School of Medicine, Atlanta, GA 30322, USA*

Received 28 August 1990

We report the amino acid sequence of CAP37, a human neutrophil granule protein with antibacterial and monocyte-specific chemotactic activity. CAP37 is a single-chain protein consisting of 222 amino acid residues. It has three *N*-glycosylation sites, at Asn residues 100, 114 and 145. Some species of CAP37 are glycosylated at all three sites; some at Asn-114 alone, others at Asn-114 and Asn-110 or Asn-145. CAP37 has 45% sequence identity to human neutrophil elastase, and 30–37% identity to several other granule serine proteinases. Despite these similarities, CAP37 is not a serine proteinase because the active site residues serine and histidine are replaced.

Cationic granule protein; Serine proteinase; Neutrophil; Monocyte chemotaxis; Antibacterial activity

1. INTRODUCTION

We report here the primary amino acid sequence of a monocyte-specific chemotactic, cationic protein originally isolated from human neutrophil (PMN) granules [1,2] owing to its killing of Gram-negative bacteria. We named it CAP37 (cationic antimicrobial protein of M_r 37 kDa). Cationic proteins from human PMN such as BPI/CAP57/BP [4,5], cathepsin G [6,7], the defensins [8], and CAP37/azurocidin [1,3,9,10] have been regarded as oxygen-independent anti-infective PMN components. However, we recently showed that CAP37 has primary chemotactic action on monocytes. This suggests it has another, perhaps more important role in defensive inflammation, specifically recruiting monocytes as it is released during inflammation and contributing to the influx of monocytes during the second wave of inflammation.

The amino acid sequence of CAP37 was undertaken to deduce any relationships that may exist between its structure and known functions. In this report we present the complete amino acid sequence of CAP37 as determined by protein sequencing.

Correspondence address: H.A. Pereira, Department of Microbiology and Immunology, Emory University School of Medicine, Room 3152 Rollins Research Center, Atlanta, GA 30322, USA

Abbreviations: fMLP, formyl-Met-Leu-Phe; MCP-1, monocyte chemotactic protein-1; PE-CAP37, reduced-pyridylethylated CAP37; PMN, polymorphonuclear leukocyte; PTH, phenylthiohydantoin; TFA, trifluoroacetic acid

2. EXPERIMENTAL

Sequenal grade chemicals for amino acid analysis, protein sequencing and HPLC purification were from Applied Biosystems Inc. Sequencing grade trypsin, endoproteinase Asp-N and *N*-glycosidase-F (EC 3.2.2.18) were from Boehringer Mannheim Biochemicals. Crystalline porcine pepsin was from Sigma Chemicals. All other chemicals were of analytical grade. Reduced-alkylated CAP37 was prepared by reduction of the native CAP37 [2] with 2-mercaptoethanol in 6 M guanidine hydrochloride followed by alkylation of the reduced protein with 4-vinyl pyridine and HPLC-desalting, yielding PE-CAP37 [11].

Cleavage of PE-CAP37 and the native CAP37 with trypsin and endoproteinase Asp-N was done in 0.1 M Tris-HCl buffer, pH 8.0, 1 M guanidine hydrochloride (E/S = 1:20 to 1:100, w/w, at 30°C, 10–20 h). Cleavage of PE-CAP37 with pepsin was done in 0.01% aqueous TFA (E/S = 1:50, 30°C, 0.5 h). The resulting digests were acidified with TFA and the component peptides separated by HPLC. The purified glycopeptides D5 and T9 (50 pmol) were treated with *N*-glycosidase-F (1 pmol) in 50 μl of 0.1 M sodium phosphate buffer, pH 7.2, 1 M guanidine hydrochloride, 10 mM EDTA-Na_2 (30°C, 24 h). At the end of the reaction, the deglycosylated peptides (D5F and T9F) were absorbed (at 25°C, 2 h) onto three pieces (2 × 4 mm) of Immobilon P membrane (Millipore Co.). The membranes were washed several times with water and directly subjected to Edman degradation [12]. Treatment of native CAP37 (1 μg) with *N*-glycosidase-F was performed in 10 μl of 0.1 M sodium phosphate buffer, pH 7.2, 10 mM EDTA-Na_2 (E/S = 1:150, w/w, 30°C, 40 h).

Purification of peptides was done by microbore reverse-phase HPLC on an Applied Biosystems Model 130A Micro Separation System equipped with a microbore LC flow cell and 0.005 inch I.D. tubing. Applied Biosystems Aquapore RP-300 C8 or Aquapore OD-300 C18 silica columns (1 × 250 mm) were equilibrated in 0.1% (v/v) aqueous TFA and developed at flow rates of 50–80 μl/min using a linear gradient (0.5% per min) of acetonitrile/water/TFA (4:1:0.004 v/v/v). Material absorbing at 214 nm was manually collected and stored at −20°C. Rechromatography of some fractions in a second solvent system consisting of 2-propanol/acetonitrile/water/TFA (7:2:1:0.008) was required before sequencing. All samples

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

were sequenced in an Applied Biosystems Model 477A Protein Sequencer with on-line Model 120A PTH Analyzer. Automated Edman degradation [13] was done using the manufacturer's reaction vessel cycle PRO-1, with modifications including the use of a higher reaction temperature (49°C) and a 30% increase in the delivery time of trimethylamine vapors. Cysteine was determined as PTH-*S*-β-(4-pyridylethyl)-Cys. The amino acid composition of certain peptides was determined using a Model 420A Derivatizer/130A PTC Analyzer equipped with an on-line hydrolysis unit assembly (Applied Biosystems).

SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was performed according to the method of Laemmli [14]. All protein samples were solubilized in 0.625 M Tris (pH 6.8), 4% (w/v) SDS, and 1% (v/v) 2-mercaptoethanol at 100°C for 5 min and analyzed on a 12.5% gel. The proteins were visualized by silver staining .

3. RESULTS AND DISCUSSION

The complete sequence of PE-CAP37, as summarized in Fig. 1, was determined by analysis of fragments generated by cleavage with endopeptidases. All endoproteinase Asp-N peptides (D1–D8), representing the full-length sequence, were recovered. Fragments D1 (residues 1–58) and D5 (residues 96–151), which could not be sequenced completely because of their length, were subfragmented by trypsin, yielding peptides D1T1–D1T4 and D5T1–D5T4, respectively. Sequencing of tryptic (T1–T12) and peptic (P1–P11) peptides provided the necessary overlaps and allowed the order of the endoproteinase Asp-N peptides to be determined. The tryptic peptides T12a and T12b, were identified as C-terminal since their C-terminal amino acid residues (glycine and proline, respectively), are incompatible with the cleavage specificity of trypsin. As further proof, two C-terminal peptides, each with a different C-terminus amino acid, were isolated in approximately the same quantities from peptic (peptides P11a,b) and endoproteinase Asp-N (peptides D8a,b) digests. Thus, the mature (active) CAP37 was determined to be a single chain polypeptide, composed of either 221 or 222 residues, and heterogeneous at the carboxy-terminus.

CAP37 contains eight cysteine residues which are found at positions corresponding to those in elastase, complement factor D and chymotrypsin [15,16]. During HPLC purification of the tryptic digest of native CAP37, peptides T3 and T4 co-migrated as an equimolar mixture of both; however, after reduction and alkylation, they migrated separately and were recovered as individual entities. This indicates that, in the native protein, Cys-26 and Cys-42 form a disulfide bridge. Peptide T9, containing Cys-154 and Cys-160, was isolated both from native and reduced-alkylated CAP37 as a single entity indicating that, in CAP37, Cys-154 and Cys-160 are either linked to each other through a disulfide bond or that they are in a free state. The oxidation state and pairing of the remaining Cys residues is currently being examined, and preliminary evidence indicates that the remaining two intramolecular disulfide bonds between Cys-123–181, and Cys-171–196 are also conserved.

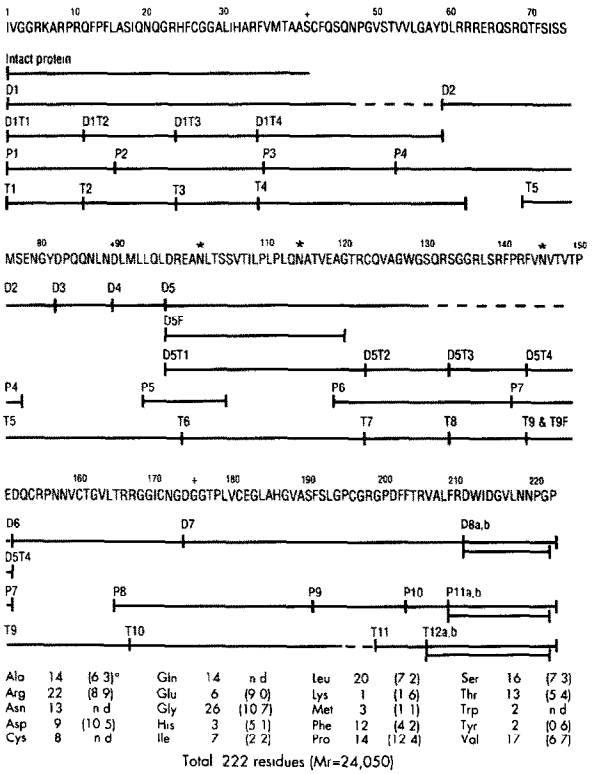

Fig. 1. Amino acid sequence and composition of CAP37. The amino acid sequence of the pyridylethylated protein was determined by automated Edman degradation of intact protein, and of its fragments generated by endoproteinase Asp-N (D), trypsin (T) and pepsin (P). Dashed lines indicate positions in sequencer runs where further sequence information was either not obtainable or unnecessary. The sites of *N*-glycosylation, Asn-100, Asn-114 and Asn-145 (*), and the positions of residues equivalent to the catalytic residues of serine proteinases (+) are indicated.

[a]Values in parentheses represent previously published amino acid composition (% total amino acids) [1].

CAP37 contains three potential *N*-glycosylation sites of the −Asn(CHO)-X-Thr− type. These sites, residues 100, 114 and 145 (Fig. 1), were identified in the following way: (i) on sequencing of fragments D5, D5T1, D5T4, T6 and T9, no PTH-amino acid could be recovered or identified in positions 100, 114 and 145 of the CAP37 sequence, suggesting the possibility of *N*-glycosylation; (ii) sequencing of peptides D5 and T9, after their exposure to *N*-glycosidase-F (peptides D5F and T9F in Fig. 1), yielded PTH-aspartic acid in cycles corresponding to residiues 100, 114 and 145, confirming that these residues are *N*-glycosylated asparagines; (iii) unglycosylated Asn-100 and Asn-145 were found in peptides P5 and P7, respectively, showing that CAP37 also exists as a partially nonglycosylated protein.

The existence of partially glycosylated or unglycosylated species of CAP37 is also evidenced by its electrophoretic and chromatographic behavior, a 37 kDa species as reported by our laboratory [1,2], and a 29 kDa species as observed by others [9,10]. Two different preparations of CAP37 (identity established by a complete match of the first 20 amino acids), one which

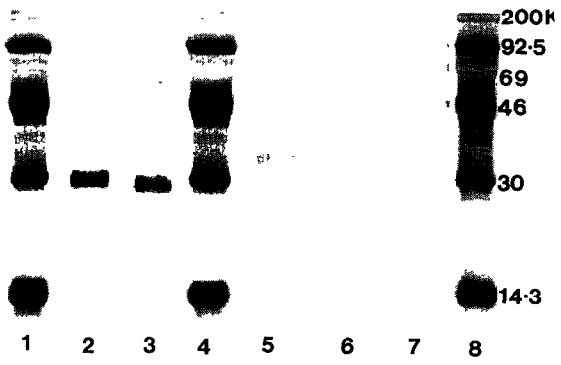

Fig. 2. Silver stained SDS-PAGE comparing native CAP37 with *N*-glycosidase-F-treated CAP37. The following samples, 2 μg 'rainbow' molecular mass markers (Amersham Corp.) in lanes 1, 4 and 8, 1 μg native CAP37 migrating at M_r 31 and M_r 37 kDa in lanes 2 and 5, respectively, *N*-glycosidase-F-treated CAP37 in lanes 3 and 6, and *N*-glycosidase-F in lane 7, were analyzed on 12.5% SDS-PAGE. The M_r of the standards are indicated in the figure. The extra band seen at M_r 35.5 kDa in lanes 3 and 6, is *N*-glycosidase-F.

migrated at approximately 37 kDa, and the other which migrated at approximately 31 kDa on SDS-PAGE (Fig. 2, lanes 2 and 5) were subjected to *N*-glycosidase-F treatment. Following treatment both preparations of CAP37 migraterd at approximately 28 kDa (Fig. 2, lanes 3 and 6) indicating that the carbohydrate moieties may constitute up to 24% of the total mass of the protein.

During the HPLC fractionation of the enzymatic digests of CAP37, glycopeptides D5, D5T1, D5T4 and T9 were each found in several fractions. Sequencing of these fractions showed them to be indistinguishable from each other. We conclude that these fractions represent glycopeptides with the same sequence and differing carbohydrate content.

A search of the GeneBank sequence data library [17] reveals that CAP37 as we suggested [11] has its strongest homology with human neutrophil elastase (45%), and 30–37% homology with human complement factor D, human cathepsin G, rat mast cell proteinases I and II, and the cytotoxic T-lymphocyte serine proteinases. CAP37 and human neutrophil elastase (Fig. 3) have 50% identity between residues 1–95 and residues 123–220. Importantly, these two regions contain all eight strictly conserved cysteine residues of CAP37 and elastase. Also, the disulfide bridges formed by four of these, Cys-26 to Cys-42 and Cys-154 to Cys-160, have their counterpart in elastase [16], suggesting that CAP37, like elastase, has a chymotrypsin-like type of disulfide pairing.

A 27-residue segment of CAP37 defined by residues 96-122 is different from elastase. It contains two *N*-glycosylation sites, Asn-100 and Asn-114, indicating its surface location. The Asn-114 site found glycosylated without exception lacks a counterpart in elastase and other granule serine proteinases suggesting a special role, perhaps in intracellular sorting and targeting. Interestingly, the beginning of this segment coincides with one of the exon/intron junctions of elastase [18], which maps between Gln-93 and Leu-94 of the translated elastase sequence (Fig. 3), and it is also found in the related proteinases, RCMP II and adipsin [19,20]. In other serine proteinases, the exon/intron junctions are located in the surface areas of these proteins and are characterized by high structural divergency [21,22]. It is therefore likely that an exon/intron junction is located in the same region of the CAP37 gene and could be of

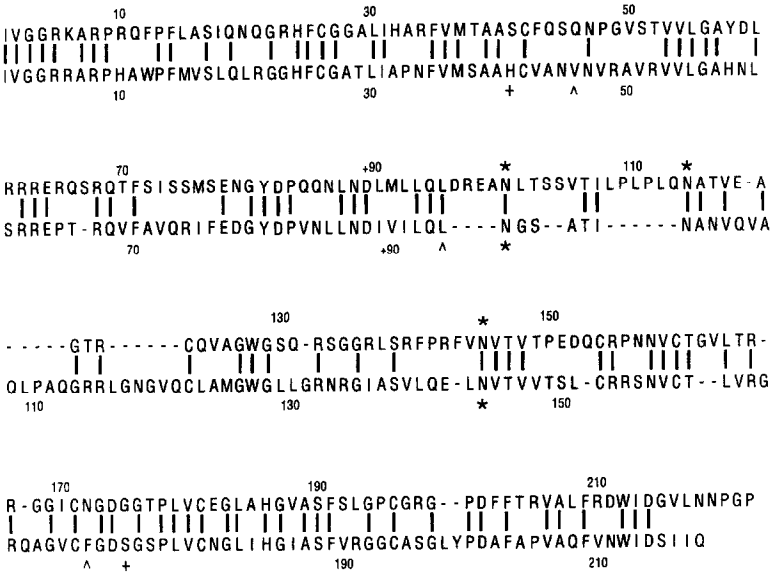

Fig. 3. Alignment of primary structures of CAP37 (top line) and human neutrophil elastase [16]. The identical residues are connected by vertical bars (|); the gaps (−) are inserted in order to obtain maximum sequence identity. Positions of the catalytic residues (+) and of the glycosylation sites (*) in both proteins, and positions of introns (^) of the elastase gene [18] are indicated.

evolutionary importance for the functional differentiation of CAP37.

The most critical sequence differences between CAP37 and elastase involve the catalytic residues. His-41, Asp-88 and Ser-173 in neutrophil elastase, are strictly conserved in all serine proteinases [23]. An equivalent catalytic aspartic acid Asp-89, is found in CAP37; however, the histidine and serine catalytic residues, are substituted by serine-41 and glycine-175. These findings explain the lack of serine proteinase function in CAP37 [11,24].

Significantly, the segments of the CAP37 sequence in the immediate vicinity of Ser-41 and Gly-175 are identical to, or highly conserved with, elastase, and include some of the residues which form the primary specificity pocket, S1. Ile-170, Gly-173, Ala-189 and Asp-201 of CAP37 have their topological counterparts in Val-168, Gly-171, Ala-187 and Asp-201 of elastase. Asn-172 and Ser-191 of CAP37 have their equivalents in Phe-170 and Val-190 of elastase, suggesting that S1 in CAP37 is more polar than in elastase. Asp-174 of CAP37 has its equivalent in Asp-172 of elastase, a residue forming a salt bridge with Ile-1, as found in other trypsin-like proteinases [25,26,27]. One might therefore expect a similar role for Asp-174 and Ile-1 in CAP37. Thus, even though the serine proteinase catalytic apparatus of CAP37 has been abrogated, the overall spatial arrangement of its primary substrate binding cleft seems to be preserved.

Analysis of the gene structure of CAP37 will contribute to the understanding of its evolutionary relationship to mammalian serine proteinases. In the neutrophil elastase, adipsin and RMCPII proteinase genes [18,19,20], the codons for catalytic residues His, Asp and Ser are located on separate exons [18]. Assuming the same exon/intron pattern for CAP37, it is likely that the present structure of CAP37 was formed by evolutionary processess involving excission of its two catalytic residues, followed by gene duplication of these exons and recombination.

Are the structure of CAP37 derived from its amino acid sequence and its bactericidal and chemotactic activity related? Overall the molecule is basic, having 23 positively charged residues as opposed to 15 acidic residues (Fig. 4). The amino terminus up to residue 70 has 11 basic residues, with four of them being present in the first 10 amino acids. The next section of the molecule (amino acids 78 to 118) shows a preponderance of acidic residues (6 acidic, 1 basic) followed by a second basic region ending at amino acid 142, thereafter the COOH terminus shows an alternating pattern of basic and acidic residues.

The basic amino-terminus may enhance electrostatic interaction between CAP37 and the negatively charged lipid A of CAP37 sensitive bacteria such as *Salmonella*. The binding of BPI/CAP57, another cationic protein to *Salmonella*, is dependent on both electrostatic and hydrophobic bindings [28–30]. Recent evidence shows the N-terminal pentapeptides IIGGR of cathepsin G [31], and IVGGR of elastase, the latter of which is identical to the N-terminus of CAP37, to possess antibiotic activity [32].

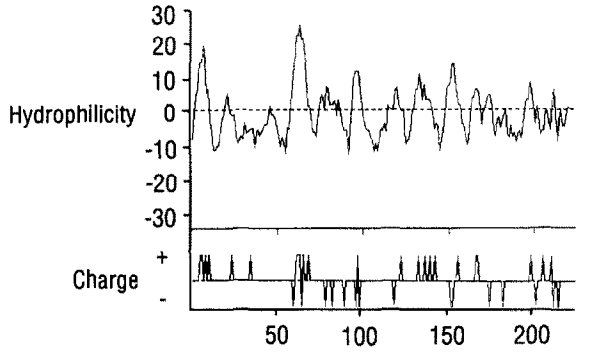

Fig. 4. Plot of hydrophilicity and charge distribution of CAP37. Hydrophilic residues are designated as positive, and hydrophobic residues as negative. Basic residues are represented as positively charged and acidic residues as negative.

The hydrophilicity plot of mature CAP37 protein using the method of Hopp and Woods [33] indicates a protein with alternating hydrophobic and hydrophilic domains, with one extended hydrophobic stretch between amino acids 26 and 58. The three highest points of hydrophilicity are from residues 61 to 66, 5 to 10, and 150 to 156. Proteins such as α-casein, β-casein, and been venom mellitin, which are cationic and contain hydrophobic sections show chemotactic activity. It is believed that these molecules are taken up more avidly into lipid bilayers than other proteins, enhancing their affinity for cell membranes and subsequent chemotactic binding [34]. The same may apply to CAP37, a cationic polypeptide with alternating hydrophobic stretches. Comparison of the CAP37 sequences with known monocyte chemoattractants such as fMLP, MCP-1, and the defensins shows no similarities. The availability of the covalent structure will be helpful in defining these mechanisms as well as to allow synthesis of pharmacologically useful compounds.

Acknowledgements: We would like to thank Linda R. Jones for help with preparation of the manuscript. This work was supported by Public Health Service Grants AI17662, AI26589 and AI28018 from the National Institute of Allergy and Infectious Disease.

REFERENCES

[1] Shafer, W.M., Martin, L.E. and Spitznagel, J.K. (1984) Infect. Immun. 45, 29–35.
[2] Pereira, H.A., Shafer, W.M., Pohl, J., Martin, L.E. and Spitznagel, J.K. (1990) J. Clin. Invest. 85, 1468–1476.
[3] Shafer, W.M., Martin, L.E. and Spitznagel, J.K. (1986) Infect. Immun. 53, 651–655.
[4] Weiss, J., Elsbach, P., Olsson, I. and Odeberg, H. (1978) J. Biol. Chem. 253, 2664–2672.

[5] Pereira, H.A., Spitznagel, J.K. Winton, E.F. Shafer, W.M., Martin, L.E., Guzman, G.S., Pohl, J., Scott, R.W., Marra, M.N. and Kinkade, J.M. Jr (1990) Blood (in press).
[6] Shafer, W.M., Onuka, V. and Hitchcock, P.J. (1986) J. Infect. Dis. 153, 910–917.
[7] Shafer, W.M., Onunka, V.C. and Martin, L.E. (1986) Infect. Immun. 54, 184–188.
[8] Ganz, T., Selsted, M.E., Szklarek, D., Harwig, S.S.L., Daher, K., Bainton, D.F. and Lehrer, R.I. (1985) J. Clin. Invest 76, 1427–1435.
[9] Gabay, J.E., Scott, R.W., Campanelli, D., Griffith, J., Wilde, C., Marra, M.N., Seeger, M. and Nathan, C.F. (1989) Proc. Natl. Acad. Sci. USA 86, 5610–5614.
[10] Campanelli, D., Detmers, P.A., Nathan, C.F. and Gabay, J.E. (1990) J. Clin. Invest. 85, 904–915.
[11] Pereira, H.A., Spitznagel, J.K., Pohl, J., Wilson, D.E., Morgan, J., Palings, I. and Larrick, J.W. (1990) Life Sci. 46, 189–196.
[12] Matsudaira, P. (1987) J. Biol. Chem. 262, 10035–10038.
[13] Hewick, R.M., Hunkapiller, M.W., Hood, L.E. and Dreyer, W.J. (1981) J. Biol. Chem. 256, 7990–7997.
[14] Laemmli, U.K. (1970) Nature (Lond.) 227, 680–685.
[15] Niemann, M.A., Bhown, A.S., Bennett, J.C. and Volanakis, J.E. (1984) Biochemistry 23, 2482–2486.
[16] Sinha, S., Watorek, W., Karr, S. Giles, J., Bode, W. and Travis, J. (1987) Proc. Natl. Acad. Sci. USA 84, 2228–2232.
[17] Pearson, W.R. and Lipman, D.J. (1988) Proc. Natl. Acad. Sci. USA 85, 2444–2448.
[18] Takahashi, H., Nukiwa, T., Yoshimura, K., Quick, C.D., States, D.J., Holmes, M.D., Wang-Peng, J., Knutsen, T. and Crystal, R.G. (1988) J. Biol. Chem. 263, 14739–14747.
[19] Benfey, P.N., Yin, F.H. and Leder, P. (1987) J. Biol. Chem. 262, 5377–5384.
[20] Phillips, M., Djian, P. and Green, H. (1986) J. Biol. Chem. 261, 10821–10827.
[21] Rogers, J. (1985) Nature (Lond.) 315, 458–459.
[22] Craik, C.S., Rutter, W.J. and Fletterick, R. (1983) Science 220, 1125–1129.
[23] Neurath, H. (1984) Science 224, 350–352.
[24] Wilde, C.G., Snable, J.L., Griffith, J.E. and Scott, R.W. (1990) J. Biol. Chem. 265, 2038–2041.
[25] Bode, W., Wei, A.Z., Huber, R., Meyer, E., Travis, J. and Neumann, S. (1986) EMBO J. 5, 2453–2458.
[26] Navia, M.A., McKeever, B.M. Springer, J.P., Lin, T.-Y., Williams, H.R., Fluder, E.M., Dorn, C.P. and Hoogsteen, K. (1989) Proc. Natl. Acad. Sci. USA 86, 7–11.
[27] Remington, S.J. Woodbury, R.G. Reynolds, R.A. Matthews, B.W. and Neurath, H. (1988) Biochemistry 27, 8097–8105.
[28] Ooi, C.E., Weiss, J., Elsbach, P., Frangione, B. and Mannion, B. (1987) J. Biol. Chem. 262, 14891–14894.
[29] Farley, M.M. Shafer, W.M. and Spitznagel, J.K. (1988) Infect. Immun. 56, 1589–1592.
[30] Gray, P.W., Flaggs, G. Leong, S.R., Gumina, R.J., Weiss, J., Ooi, C.E. and Elsbach, P. (1989) J. Biol. Chem. 264, 9505–9509.
[31] Salvesen, G., Farley, D., Shuman, J., Przybyla, A., Reilly, C. and Travis, J. (1987) Biochemistry 26, 2289–2293.
[32] Bangalore, N., Travis, J., Onunka, V.C., Pohl, J. and Shafer, W.M. (1990) J. Biol. Chem. (In press).
[33] Hopp, T.P. and Woods, K.R. (1981) Proc. Natl. Acad. Sci. USA 78, 3824–3828.
[34] Wilkinson, P.C. (1982) Chemotaxis and Inflammation, 2nd edn., Churchill Livingstone, Edinburgh.