Volume 333, number 1,2, 39–43 FEBS 13132 October 1993

The primary structure of carboxypeptidase S1 from *Penicillium janthinellum*

Ib Svendsen[a,*], Theo Hofmann[b], Jim Endrizzi[c], S. James Remington[c], Klaus Breddam[a]

[a]*Carlsberg Laboratory, Department of Chemistry, Gamle Carlsberg Vej 10, DK-2500 Copenhagen, Denmark*
[b]*Department of Biochemistry, University of Toronto, Toronto, Ont. M5S 1A8, Canada*
[c]*University of Oregon, Institute of Molecular Biology, Eugene, OR 97403-1229, USA*

(Received 23 August 1993; revised version received 8 September 1993)

The complete amino acid sequence of carboxypeptidase S1 from *Penicillium janthinellum* has been determined by N-terminal sequencing of the reduced and vinylpyridinated protein and of peptides obtained by cleavage with cyanogen bromide, iodosobenzoic acid, hydroxylamine, endoproteinase LysC, endoproteinase AspN and Glu-specific proteinase from *B. licheniformis*. The enzyme consists of a single peptide chain of 433 amino acid residues and contains 9 half-cystine residues and one glycosylated asparagine residue. A comparison to other carboxypeptidases shows that the enzyme is homologous to carboxypeptidase-Y and carboxypeptidase-MIII from malt. Specificity and binding of substrates are discussed from a three-dimensional model based on the known structure of carboxypeptidase-Y from *Saccharomyces cereviciae* and carboxypeptidase II from wheat.

Carboxypeptidase; Serine protease; Amino acid sequence

1. INTRODUCTION

The serine carboxypeptidases constitute a group of carboxypeptidases that employ the catalytic mechanism of the serine endopeptidases. Two types of primary structures among this group of enzymes have been found [2]: (a) single-chain enzymes containing approximately 420 amino acid residues; and (b) enzymes that are processed into two chains [3] of approximately 260 and 160 amino acid residues. The primary structures of five serine carboxypeptidases have previously been determined, i.e. from group (a) carboxypeptidase-Y from yeast [4,5,6] and carboxypeptidase III from germinated barley [7] and from group (b) carboxypeptidases I and II from germinated barley [8,9] as well as carboxypeptidase W-II from wheat [10]. Comparisons of the sequences show a number of regions which are conserved in all five enzymes. The three-dimensional structure of carboxypeptidase W-II was described in 1990 [11] and recently, the structure of the single chain carboxypeptidase-Y was completed (Endrizzi, Breddam and Remington, in preparation). Both structures suggest that some of the amino acids located in the poorly conserved region around the processing point in the two-chain enzymes are situated in the active site. However, comparison of the two three-dimensional structures suggests that the two sub-classes of serine carboxypeptidases are unrelated within this region and that the previous alignment of the sequences is in error [7]. The two single-chain enzymes carboxypeptidases-Y and MIII, both of which exhibit pronounced preferences for hydrophobic amino acid residues at the P_1 and P_1' positions, are similar in parts of this region. The structure–function studies on carboxypeptidase-Y [12] would be greatly facilitated if primary structures of other single- chain serine carboxypeptidases with different substrate preferences were available. Carboxypeptidase S-1 [13] with a preference for basic amino acid residues at the P_1 and P_1' [14] is one such enzyme and its primary structure is reported here.

*Corresponding author. Fax: (45) (33) 27-4708.

Abbreviations: CPD-S1, carboxypeptidase S1 (from *P. janthinellum*); CPD-Y, carboxypeptidase Y (from yeast); CPD-MIII, carboxypeptidase MIII (from barley malt); CPDW-II, carboxypeptidase II from wheat; CNBr, cyanogen bromide. The binding site notation is that of Schechter and Berger [1].

2. EXPERIMENTAL

Carboxypeptidase S1 was isolated as previously described [14]. Amino acid sequence analysis was performed using an Applied Biosystems gas phase sequencer, model 470A or a pulsed liquid sequencer, model 477A. Reduction, vinylpyridination and cleavages with CNBr, iodosobenzoic acid and hydroxylamine, digestions with EndoLysC protease (Boehringer), Glu-specific enzyme (isolated from *B. licheniformis* [15]) and AspN cleaving enzyme (Boehringer) were carried out as previously described [16]. Separation of peptides was performed either by gel chromatography on BioGel P6, Sephadex G50, superfine, in 30% acetic acid or by high performance liquid chromatography on a Vydac C4 column, using a linear gradient of acetonitrile from 10–60% in 0.1% trifluoroacetic acid.

C-Terminal digestion of CPD-S1 was performed in the following way: 1.1 mg reduced and vinylpyridinated CPD-S1 was dissolved in 100 μl 0.05 M MES, 1 mM EDTA, 0.5% SDS, pH 6.5. 50 μl H_2O was added followed by 1 μl CPD-Y (19 mg/ml). 25 μl aliquots were withdrawn at different times, acidified by addition of 7 μl of 0.6 M HCl and applied directly to the amino acid analyzer.

3. RESULTS AND DISCUSSION

The complete amino acid sequence of CPD-S1 was obtained by N-terminal sequencing of the reduced and vinylpyridinated protein and by sequencing of peptides derived by cleavage with CNBr, hydroxylamine, iodosobenzoic acid, EndoLysC protease, endoproteinase AspN and Glu-specific protease from *B. licheniformis*. Digestion of the vinylpyridinated protein with CPD-Y identified the C-terminal sequence: 1 Thr, 2 Ser and 1 Ile were released; the time course suggested that the single Thr and two Ser preceded Ile. It was not possible to assign the sequence of the first three amino acids, but the results are consistent with the C-terminal sequence -Ile-Ser-Ser-Thr-OH, identified by N-terminal sequencing of the C-terminal EndoLysC peptide. As seen in Fig.

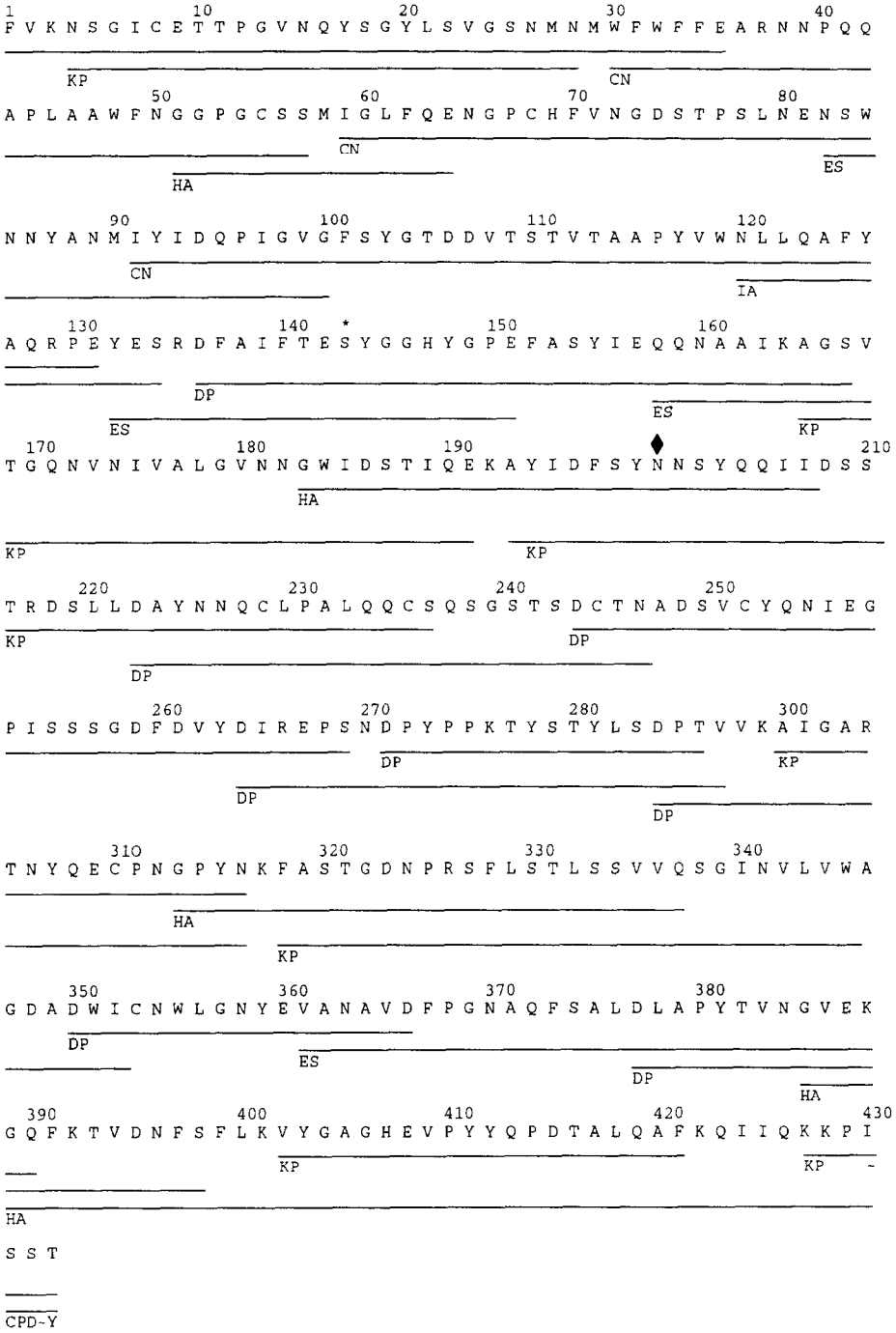

Fig. 1. The amino acid sequence of carboxypeptidase S1 (CPD-S1) from *Penicillium janthinellum* obtained by N-terminal sequencing and sequencing of peptides obtained from cleavage with cyanogen bromide (CN), hydroxylamine (HA), iodosobenzoic acid (IA), Glu-specific protease from *B. licheniformis* (ES), EndoLysC protease (KP), endoproteinase AspN (DP) and carboxypeptidase-Y (CPD-Y). The diamond denotes the glycosylated Asn-200 and the asterisk the Ser-143 in the catalytic site.

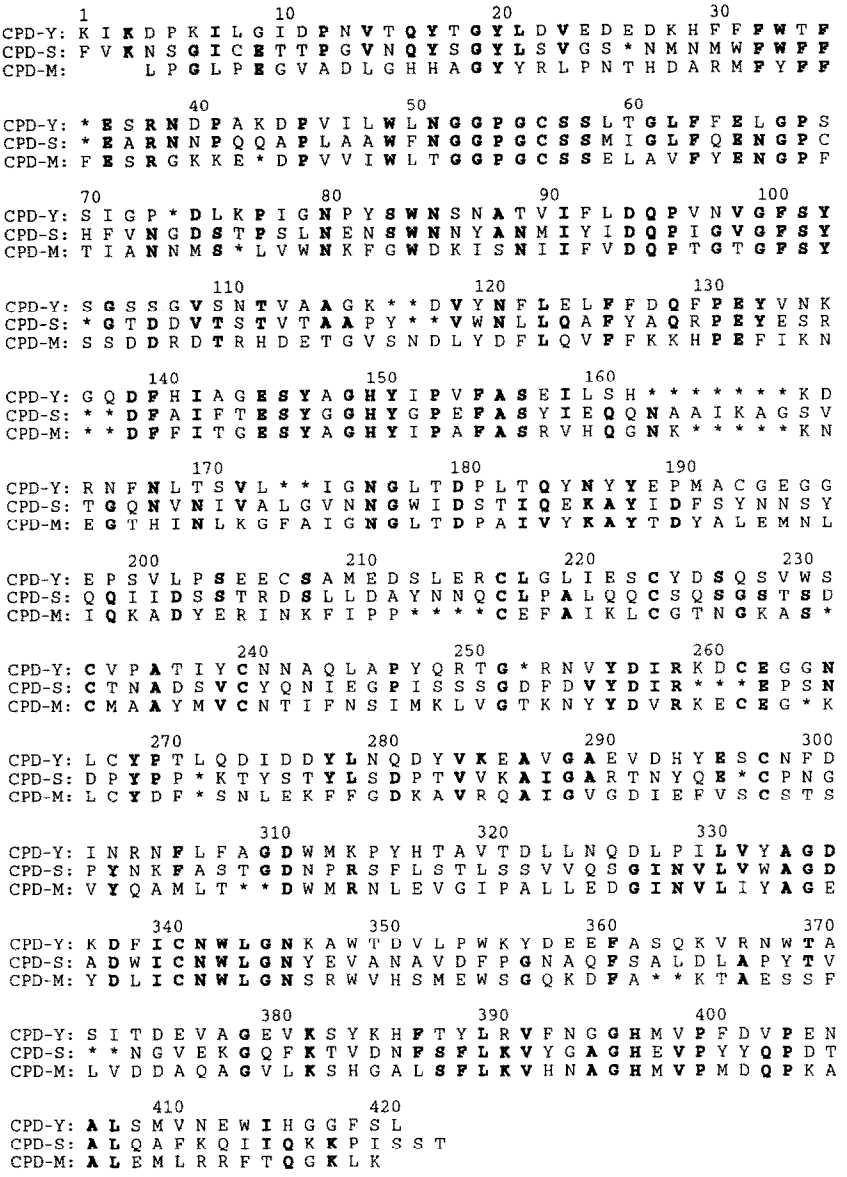

Fig. 2. Comparison of the sequences of the single-chain serine carboxypeptidases CPD-Y, CPD-S1 and CPD-MIII. Identical amino acid residues are in bold letters, * show gaps.

1 the enzyme consists of a single peptide chain of 433 amino acid residues. The enzyme is a glycoprotein with a content of about 3% carbohydrate [14]. There are two potential *N*-glycosylation sites in CPD-S1: positions 200 and 396 (Fig. 1) of which the latter is not glycosylated. Position 200, however, was always empty when sequenced and we therefore assume that this position is glycosylated. Recently, the sequence around the glycosylated Asn in the commercially available carboxypeptidase-P was described [17]. Although this enzyme is also isolated from a *Penicillium janthinellum* strain the sequences around the glycosylation point are not identical.

When the sequences of the five serine carboxypeptidases previously determined are aligned, with due allowance for insertions and deletions, certain regions in the N-terminal and C-terminal portions of the peptide chain are identical in all the enzymes [7]. In contrast, very little identity is observed between residues 180 and 395 [7]. CPD-S1 falls well into this pattern (not shown) but in order to establish whether the single-chain enzymes would exhibit a higher degree of identity among each other, in particular in the middle portion of the peptide chain, the sequence of CPD-S1 was aligned with those of CPD-Y and CPDM-III (Fig. 2). 138 amino acid residues (32%) are in identical positions when CPD-S1 is compared to CPD-Y and 132 (30%) when compared to CPD-MIII. It is seen that identical regions are found evenly distributed throughout the sequence suggesting that the single-chain enzymes are more closely related to each other than to the double-chain enzymes. The essential serine may be assigned to Ser-150 on the basis of the high degree of identity in this region. Zieske et al.

[17] have shown by labelling with DFP that the catalytic site Ser was located in a peptide almost identical to the CPD-S1 sequence Asp-150 to Glu-164.

The global folding pattern of CPD-Y (to be published) and CPDW-II [11] are similar. Both contain a central 11-stranded mixed β-sheet flanked by 15 helices, although there are 18 insertions and deletions (totalling 65 amino acid residues) between the two. Several of these insertions and deletions result in alteration of secondary structure (to be discussed below). The conserved disulfide bridges (to be discussed below) together with the catalytic triad residues give landmarks from which to align the sequences. To further analyze differences between the sequences which may be involved in specificity, a molecular model of the active site region of CPD-S1 was constructed using FRODO on an Evans and Sutherland graphics station.

The three-dimensional structure of CPD-Y identifies the location of the disulfide bridges: 56–298, 193–207, 217–240, 224–233, and 262–268 (Fig. 2). All five disulfides are located on one face of the molecule and surround the active site pit, suggesting that they may be involved in keeping the structural integrity of the extensively hydrophobic active site cavity.

In CPD-MIII, all disulfide bridges are conserved except 193–207. In CPD-S1, 193–207 and 262–268 are missing. A new disulfide bridge has most likely been formed between residues 8 and 69, since these residues of the CPD-Y model can form a disulfide with good geometry without moving the main chain (see Fig. 3). The fairly high degree of conservation of disulfide bridges within the group of single-chain serine carboxypeptidases support the suggestion that they are structurally very similar.

An overlay of CPD-Y and CPDW-II based on α-coordinates reveals 2 conserved disulfides, 56–298 and 217–240. Disulfide 217–240 is shifted 3Å relative to its position in CPD-WII due to displacement of three helices (residues 180–250) which pack against each other at the edge of the active site. Disulfide 262–268 is near the S1′ specificity pocket of both enzymes. In CPD-Y, it positions Tyr-269 on one side of the S1′ subsite. The equivalent disulfide of CPDW-II (246–268) spans the break between the A- and B-chains of CPDW-II and is 10 Å from disulfide 262–268 of CPD-Y, resulting in a more open S1′ subsite in CPDW-II. Residues 260–270 adopt completely different conformations in the two enzymes and are involved in determining substrate specificity in the S1′ subsite.

CPD-Y has two disulfides that CPDW-II lacks, 193–207 and 224–233. In addition to disulfide 217–240, disulfide 224–233 links antiparallel helices 204–227 and 230–251 at the edge of the active site (these helices are several turns longer in CPD-Y relative to CPDW-II). The other CPD-Y disulfide that CPDW-II lacks, 193–207, links helix 204–227 to the C-terminus of helix 180–193.

Interestingly, the free cysteinyl residue (341 in Fig. 2) characteristic of these single chain enzymes, but absent in the double-chain enzymes, is situated in a conserved region. The function of this amino acid residue, which renders these enzymes sensitive to mercurials, has been investigated by site directed mutagenesis [18] and chemical modifications [19]. However, a clear conclusion with regard to its function has not been obtained.

3.1. *Specificity at subsite S'_1*

Binding of free arginine has located residues in CPDW-II involved in binding the P'_1 residue of substrate [20]. The aliphatic portion of the arginine side chain lies in a cleft formed by Tyr-60 and Tyr-239

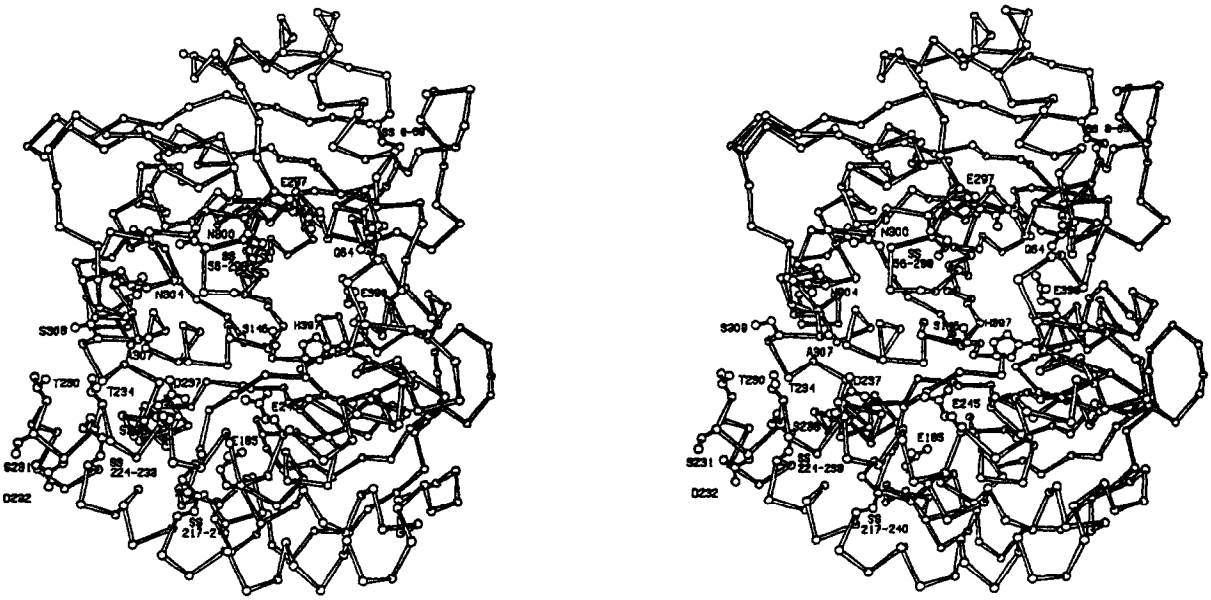

Fig. 3. Molecular model of CPD-S1 based on the α-carbon coordinates of the 2.8 Å crystal structure of CPD-Y. Side chains of residues referred to in the text are shown. As in Fig. 2, amino acid residues are numbered according to the CPD-Y sequence.

(CPDW-II numbering). Thus, Ile-60 and Tyr-256 (CPD-Y numbering, Tyr-256 is topologically equivalent to Tyr-239 of CPDW-II) presumably form a hydrophobic cleft for the aliphatic portion of basic side-chain at S'_1 of CPD-S1. Glu-272 and Glu-398 interact electrostatically with the terminal charge of the arginine side chain. Binding of benzylsuccinate to CPD-Y at pH 4.5 has implicated Met-398 as a crucial residue for binding substrate at S'_1. CPD-S1 has a glutamate in position 398 (and is adjacent to His-297 of the catalytic triad), and thus it is likely that this residue plays a role in specificity determination at S'_1 by interacting electrostatically with Lys and Arg side chains at the P'_1 of substrates. Additionally, Glu-64 of CPDW-II (Phe-64 in CPD-Y) is hydrogen bonded to Glu-398, possibly orienting it for interaction with basic side chains at P'_1. CPD-S1 has Gln at position 64, suggesting that it too interacts with residue 398, orienting it for P'_1 specificity determination (see Fig. 3). Model building suggests that Glu-260, Asp-268 and/or Glu-297 may be able to interact electrostatically with basic side chains at P'_1 of substrate, possibly through bridging water molecules. Glu-260 and Asp-268 are not included in the model since this region (residues 260–270) differs greatly in sequence between the serine carboxypeptidases and adopts completely different conformations in CPD-Y and CPDW-II.

3.2. *Specificity at subsite S_1*

Of the residues that surround the putative S_1 subsite of CPD-Y, Ile-340, Cys-341, and Tyr-147 are conserved in the CPD-S1 sequence. Model building suggests that Ile-340 and Cys-341 form a hydrophobic cleft for binding the aliphatic portion of basic side chains at S_1. Tyr-185 and Leu-245 lie on the periphery of the S_1 subsite of CPD-Y. CPD-S1 has Glu in positions 185 and 245, and model building suggests that they could interact with the side chains of basic residues at the P_1 position of substrates hence accounting for the preference of CPS-S1 for basic P_1 residues. Glu-185 appears to be out of range for direct electrostatic interaction with side chains at P_1, and may influence substrate preference beyond P_1.

3.3. *Extended binding site?*

Analysis of the final model of CPD-Y reveals a solvent-exposed region at the edge of the active site pit which is extremely hydrophobic. This region, made up of residues Val-230, Trp-231, Pro-235, Thr-237, Ile-238, Phe-300, Leu-307, Phe-308, and Trp-312, contributes over 750 Å^2 of contiguous solvent-exposed surface area (using a 1.4 Å probe radius). This region may be involved in recognition of longer hydrophobic peptides, and may be the reason why CPD-Y requires a prosequence for proper folding [21]. Residues 230, 231, 234, 235, 237, 238, 307, and 308 of CPD-Y form 4 'new' turns of helix and represent insertions into CPD-Y relative to CPDW-II. The disulfide linked helices 204–227 and 230–251 and helix 299–307 appear to be conserved between CPD-Y, CPD-S1, and CPDM-III, i.e. they all have the insertions which contribute the 4 new turns of helix relative to CPDW-II. The sequence alignment suggests that Thr-230, Ser-231, Thr-234, Asn-235, Asp-237, Ser-238, Asn-300, Asn-304, Ala-307, Thr-309, and Asn-312 comprise this hydrophobic surface in CPD-S1 (see Fig. 3, Asn-312 and Asn-235 have been left out for clarity) as opposed to those of CPD-Y and CPD-MIII which essentially are exclusively hydrophobic. These differences may be related to their substrate preferences. Like CPD-Y, CPDM-III has an abundance of hydrophobic residues in this region, consistent with its preference for hydrophobic peptide substrates.

The results obtained in the present study will be used for further studies on CPD-Y by means of genetic engineering.

Acknowledgements: The skilled technical assistance of Pia Breddam, Bodil Corneliussen and Lone Sørensen is gratefully acknowledged.

REFERENCES

[1] Schechter, I. and Berger, A. (1967) Biochem. Biophys. Res. Commun. 27, 157–162.
[2] Breddam, K. (1987) Carlsberg Res. Commun. 51, 83–128.
[3] Doan, N.P. and Fincher, G.B. (1988) J. Biol. Chem. 263, 11106–11110.
[4] Martin, B.M., Svendsen, I., Viswanatha, T. and Johansen, J.T. (1982) Carlsberg Res. Commun. 47, 1–14.
[5] Svendsen, I., Martin, B.M., Viswanatha, T. and Johansen, J.T. (1982) Carlsberg Res. Commun. 47, 15–28.
[6] Valls, L.A., Hunter, C.P., Rothmann, J.H. and Stevens, T.H. (1987) Cell 48, 887–897.
[7] Sörensen, S.B., Svendsen, I. and Breddam, K. (1989) Carlsberg Res. Commun. 54, 193–202.
[8] Sörensen, S.B., Breddam, K. and Svendsen, I. (1986) Carlsberg Res. Commun. 51, 475–485.
[9] Sörensen, S.B., Svendsen, I. and Breddam, K. (1987) Carlsberg Res. Commun. 52, 285–296.
[10] Breddam, K., Sörensen, S.B. and Svendsen, I. (1987) Carlsberg Res. Commun. 52, 297–312.
[11] Liao, D.-I. Breddam, K., Sweet, R.M., Bullock, T. and Remington, S.J. (1990) J. Biol. Chem. 265, 6528–6531.
[12] Bech, L. and Breddam, K. (1988) Carlsberg Res. Commun. 53, 381–393.
[13] Hoffmann, T. (1976) in: Methods Enzymol. (L. Lorand, ed.) Academic Press, vol. XLV, pp. 587–599.
[14] Breddam, K. (1988) Carlsberg Res. Commun. 53, 309–320.
[15] Svendsen, I., Jensen, M.R. and Breddam, K. (1991) FEBS Lett.292, 165–167.
[16] Ballance, G.M. and Svendsen, I. (1988) Carlsberg Res. Commun. 53, 411–420.
[17] Zieske, L.P., Hsi, K.-L., Chen, L. and Yuan, P.M. (1992) Arch. Biochem. Biophys. 295, 76–83.
[18] Winther, J.R. and Breddam, K. (1987) Carlsberg Res. Commun. 52, 263–273.
[19] Breddam, K. (1983) Carlsberg Res. Commun. 48, 9–19.
[20] Laio D.-I., Breddam, K., Sweet, R.M., Bullock, T. and Remington, S.J. (1992) Biochemistry 31, 9796–9812.
[21] Winther, J. and Sörensen, P. (1991) Proc. Natl. Acad. Sci. USA 88, 9330–9334.