Correlation between carbohydrate-binding specificity and amino acid sequence of carbohydrate-binding regions of *Cytisus*-type anti-H(O) lectins

Yukiko Konami, Kazuo Yamamoto, Toshiaki Osawa* and Taturo Irimura

Division of Chemical Toxicology and Immunochemistry, Faculty of Pharmaceutical Sciences, University of Tokyo, Hongo, Bunkyo-ku,
Tokyo 113, Japan

Received 23 March 1992; revised version received 20 April 1992

A carbohydrate-binding peptide of the di-N-acetylchitobiose-binding Cytisus sessilifolius anti-H(O) lectin I (CSA-I) was isolated from the endoproteinase Asp-N digest of CSA-I by affinity chromatography on a column of N-acetyl-D-glucosamine oligomer-Sepharose (GlcNAc oligomer-Sepharose). The amino acid sequence of the carbohydrate-binding peptide of CSA-I was determined to be DTYFGKTYNPW using a gas-phase protein sequencer. This sequence corresponds to the sequence from Asp-129 to Trp-139 based on the primary structure of CSA-I, and shows a high degree of homology to those of the putative carbohydrate-binding peptide of the Laburnum alpinum lectin I (LAA-I) (DTYFGKAYNPW) and of the Ulex europaeus lectin II (UEA-II) (DSYFGKTYNPW). The binding of these three anti-H(O) lectins is known to be inhibited by di-N-acetylchitobiose but not by 1-fucose. These results strongly suggest that there is a good correlation between the carbohydrate-binding specificity and the amino acid sequence of the carbohydrate-binding regions of di-N-acetylchitobiose-binding lectins.

Cytisus sessilifolius anti-H(O) lectin; Amino acid sequence; Carbohydrate-binding peptide

1. INTRODUCTION

Lectins are widely used as tools for the study of the carbohydrate constituents of cell surfaces and glycoproteins. They are also widely used for the isolation of glycoproteins by affinity chromatography. Leguminous lectins resemble each other in their physicochemical properties regardless of their carbohydrate-binding specificity. Several anti-H(O) lectins have already been isolated and characterized for their carbohydrate-binding specificity [1–19]. On the basis of the inhibition assays using simple sugars as inhibitors, these anti-H(O) lectins were known to be subdivided into two groups. The first group is the so-called eel serum type anti-H(O) lectins including *Ulex europaeus* lectin 1 (UEA-I) and *Lotus tetragonolobus* lectin (LTA). Their binding is in-

Abbreviations: CSA-1, Cytisus sessilifolius lectin I; GleNAc oligomer-Sepharose, N-acetyl-D-glucosamine oligomer-Sepharose; LAA-I, Laburnum alpinum lectin I; UEA-II, Ulex europaeus lectin II; UEA-I, Ulex europaeus lectin I; LTA, Lotus tetragonolobus lectin; HPLC, high-performance liquid chromatography; TFA, trifluoroacetic acid.

*Present address: Yakult Central Institute for Microbiological Research, Yaho, Kunitachi-shi, Tokyo 186, Japan.

Correspondence address: T. Irimura, Division of Chemical Toxicology and Immunochemistry, Faculty of Pharmaceutical Sciences, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan. Fax: (81) (3) 3815-9344.

hibited by L-fucose. The second group is the so-called Cytisus-type anti-H(O) lectins including Ulex europaeus lectin II (UEA-II), Laburnum alpinum lectin I (LAA-I), and Cytisus sessilifolius lectin I (CSA-I). Their binding is inhibited by di-N-acetylchitobiose.

In previous papers [19-21] we have determined the primary structures of LTA [19], UEA-I [20], UEA-II [20] and LAA-I [21], and compared them with those of several other leguminous lectins. Extensive homology was found throughout the stretch of the peptides. Furthermore, carbohydrate-binding peptides have been isolated from LTA, UEA-I, UEA-II and LAA-I by affinity chromatography after treatment of the lectins with endoproteinase Asp-N or Lys-C [22]. The peptides having an affinity for the specific sugars were retarded on the affinity columns of GlcNAc oligomer-Sepharose for LAA-I and UEA-II, and Fuc-gel for LTA and UEA-I. The homology found in the amino acid sequences among these retarded peptides indicates the presence of a carbohydrate-binding region of homologous leguminous lectins. However, there was not enough available data to draw a conclusion on the relationship between the sequence of this region and the carbohydrate-binding specificity.

In this study, we determined the complete amino acid sequence of the di-N-acetylchitobiose-binding anti-H(O) CSA-I using a protein sequencer. After digestion of CSA-I with two kinds of endoproteinase, Lys-C or Asp-N, the resulting peptides were purified by reversed-

phase high-performance liquid chromatography (HPLC) and then subjected to sequence analysis.

A carbohydrate-binding peptide was also isolated from the Asp-N digest of CSA-I by affinity chromatography on a column of GlcNAc oligomer-Sepharose and the amino acid sequence of this peptide was determined using a protein sequencer. More extensive homology was detected in the sequences of the carbohydrate-binding peptide of CSA-I (DTYFGKTYNPW), LAA-I (DTYFGKAYNPW), and UEA-II (DSYFGKTYNPW) than for the homology between isolectins having different carbohydrate-binding specificity. These results strongly suggest that there is a close relationship between the carbohydrate-binding specificity and the sequences of the carbohydrate-binding peptide of the di-N-acetylchitobiose-binding (Cytisus-type) lectins such as CSA-I, LAA-I and UEA-II.

2. MATERIALS AND METHODS

2.1. Materials

The seeds of Cytisus sessilifolius were obtained from Service plus S.A. (Venissieux cedex, France). Sepharose 4B was purchased from Pharmacia (Uppsala, Sweden). C_4 and $C_{18}\mu$ Bondaspheres (100 Å) for reversed-phase chromatography were obtained from Waters (Burling-

ton, MA, USA). Endoproteinase Lys-C (Lysobacter enzymogenes) and Asp-N (Pseudomonas fragi) were purchased from Boehringer GmbH (Mannheim, Germany).

2.2. Preparation of specific affinity adsorbent, GlcNAc oligomer-Sepharose

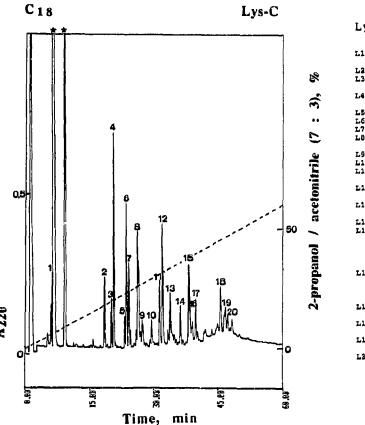
Chitin oligosaccharides were prepared by the limited hydrolysis of chitin according to the method of Rupley [23]. A mixture of tetra-N-acetylchitotetraose, penta-N-acetylchitopentaose, and hexa-N-acetylchitohexaose (1:1:1; w/w) prepared as already mentioned was coupled to amino-Sepharose 4B according to the method of Baues and Gray [24] (GlcNAc oligomer-Sepharose).

2.3. Purification of CSA-I

CSA-1 was isolated and purified by affinity chromatography by the previously reported method [18]. The affinity-purified CSA-I was further purified by reversed-phase HPLC on a column of C₄ using a linear gradient (0–100%) of 2-propanol/acetonitrile (7:3) in distilled water containing 0.1% trifluoroacetic acid (TFA) in 60 min at a flow rate of 1 ml/min.

2.4. Digestion of CSA-I with endoproteinase Lys-C or Asp-N

The purified CSA-I (0.5 mg in 150 μ l of 50 mM phosphate buffer, pH 8.0, 37°C) was digested with 5 μ g of Lys-C or with 2 μ g of Asp-N for 18 h. The peptide fragments obtained were separated by reversed-phase HPLC on a column of C_{18} using a linear gradient (0–60%) of 2-propanol/acetonitrile (7:3) in distilled water containing 0.1% TFA in 60 min at a flow rate of 1 ml/min. Elution profiles were monitored by the absorbance at 220 nm. The peptide fragments were manually collected.



Lys-C fragments

```
Thr-Val-Lys
           Ser-Ile-Lys
His-Ile-Cly-Val-Asp-Val-Asn-Sor-Ile-Lys
Pho-Gly-Leu-Pho-Asn-Sor-Ser-Asp-Asn-Lys
Trp-Asp-Trp-Asp
L2
L3
            Asn-Gly-Glu-Val-Ala-Asn-Val-Val-Ile-Thr-Tyr-Arg-Ala-Pro-Thr-
            Thr-Tyr-Asn-Pro-Trp-Asp-Pro-Asp-Phe-Lys
Thr-Gly-Arg-Val-Ala-Ser-Phe-Glu-Thr-Ser-Phe-Ser-Phe
Lou-Asn-Asp-Hig-Leu-Ser-Phe-Asn-Phe-Asp-Lys
L5
L6
L7
L8
           Trp-Asp-Trp-Arg-Asn-Gly-Glu-Val-Ala-Asn-Val-Val-Ile-Thr-Tyr-
-Arg-Ala-Pro-Thr-Lys
            Ser-Trp-Tyr-Pha-Thr-Ser-Asn-Leu-Glu-Ala-Asn-Pro-Ala
Leu-Ala-Pro-Ala-Asn-Ser-Gln-Tle-Pro-Ser-Gly-Ser-Ser
L10
            Sor-Sor-Asn-Glm-Tio-Tio-Ala-Val-Glu-Phe-Asp-Thr-Tyr-Phe-Gly-
L12
            Pho-Val-Pro-Asn-Qin-Asn-Asn-Ile-Leu-Pho-Gin-Gly-Glu-Ala-Sor-
L13
            Pho-Pho-Leu-Ala-Pro-Ala-Asn-Ser-Gln-Ila-Pro-Ser-Gly-Ser-Ser-
           -Ala-Gly-Leu
Ala-Ile-Leu-Pro-Glu-Trp-Val-Ser-Val-Gly-Pho
           -Asn-Ser-Ser-Asp-Asn-Lys
Ser-Leu-Thr-Val-Ser-Leu-Ser-Tyr-Pro-Ser-Asp-Gln-Thr-Ser-Asn-
           -Tie-Val-Thr-Ala-Ser-Val-Asp-Leu-Lys
Ser-Ala-Gly-Val-Gly-Asn-Ala-Ala-Glu-Phe-Glu-Thr-His-Asp-Val-
L16
            -Lou-Ser-Trp-Tyr-Phe-Thr-Ser-Asn
Val-Ala-Ser-Phe-Glu-Thr-Ser-Phe-Ser-Phe-Val-Val-Lys-Asp-Glu-
           -Pro-Glu-Lya
            L17
1.18
L19
            Ala-Leu-Tyr-Ala-Ala-Pro-Val-His-Tie-Trp-Asp-Sor-Thr-Thr-Gly-
L20
             Pro-Ala-Thr-Arg-Ser-Ile-Gly-Arg-Ala-Leu-Tyr-Ala-Ala-Pro-Val-
           -His-Ile-Tro-Asp-Ser-Thr-Thr-Gly-Arg
```

Fig. 1. Reversed-phase HPLC of Lys-C digest of purified CSA-I on a column of C₁₈. The residues, which were difficult to identify, are written in italics. Two peaks arising from dithiothreitol used for the reduction of CSA-I are marked with an asterisk. Minor peptides in the fragments are marked with a triangle.

2.5. Isolation of a carbohydrate-binding peptide of CSA-I

The purified CSA-I was digested with endoproteinase Asp-N and a mixture of all the obtained peptide fragments was applied to a column of GlcNAc oligomer-Sepharose (5 ml). The column was washed with 10 mM Tris/HCl, pH 6.8, containing 0.15 M NaCl, 1 mM CaCl₂ and 1 mM MnCl₂ at a flow rate of 1.5 ml/h. Each 0.5 ml was collected and analyzed by reversed-phase HPLC on a column of C₁₈ with a linear gradient (0–60%) of 2-propanol/acetonitrile (7:3) in distilled water containing 0.1% TFA in 30 min at a flow rate of 1 ml/min. The elution was monitored by the absorbance at 220 nm.

2.6. Amino acid sequence analysis

The amino acid sequence analyses of the peptide fragments fractionated by HPLC were performed with a 6600 ProSequencer solid-phase protein sequencer (MilliGen/Biosearch, Burlington, MA, USA) and a PSQ-1 gas-phase sequencer (Shimazu, Kyoto, Japan).

3. RESULTS AND DISCUSSION

3.1. Determination of the primary structure of CSA-I

Purification and sequencing of the peptides, obtained after digestion of CSA-I with endoproteinase Lys-C (Fig. 1) or Asp-N (Fig. 2), provided enough overlapping sequences to obtain the complete amino acid sequence of CSA-I (Fig. 3). Some sequenced peptide fragments (L8 and L15) were aligned by the use of the structural

homology with several anti-H(O) ieguminous lectins shown in Fig. 4. The structure of CSA-I contains 244 amino acid residues. The molecular weight of CSA-I calculated on the basis of the primary sequence is 26,575.02. This value is in good agreement with the apparent molecular weight previously determined by SDS-PAGE (27,000) [18]. As shown in Fig. 4, the complete amino acid sequence of CSA-I was compared with those of LAA-I [21], UEA-II [20] and UEA-I [20]. CSA-I, LAA-I and UEA-II are all di-N-acetylchitobiosebinding anti-H(O) lectins. The overall identity value between CSA-I and LAA-I is 86.1% and that between CSA-I and UEA-II is 87.7%. Although UEA-I and UEA-II are isolectins isolated from the seeds of the same species, Ulex europaeus, the overall positional identity value between UEA-I and UEA-II was \$2.0% [20]. On the other hand, the overall identity values between CSA-I and LAA-I (86.1%), and between CSA-I and UEA-I (87.7%) are strikingly high.

3.2. Affinity chromatography of the Asp-N digest of CSA-I

The Asp-N digest of CSA-I was fractionated on a

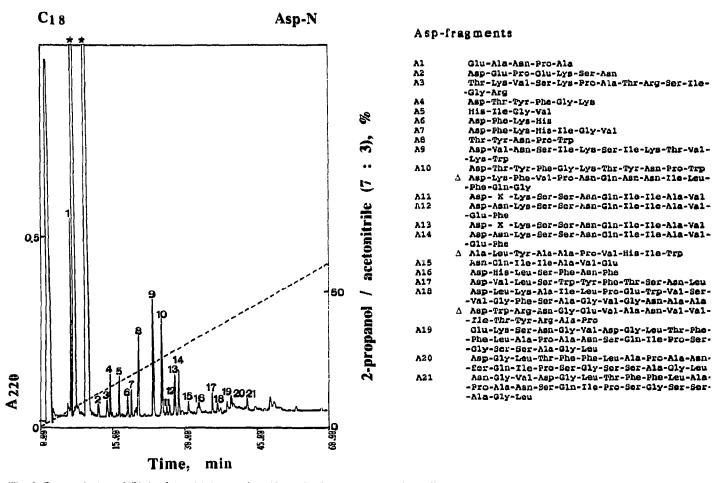


Fig. 2. Reversed-phase HPLC of Asp-N digest of purified CSA-I on a column of C_{18} . The residues, which were difficult to identify, are written in italics. Two peaks arising from dithiothreitol used for the reduction of CSA-I are marked with an asterisk. Minor peptides in the fragments are marked with a triangle.

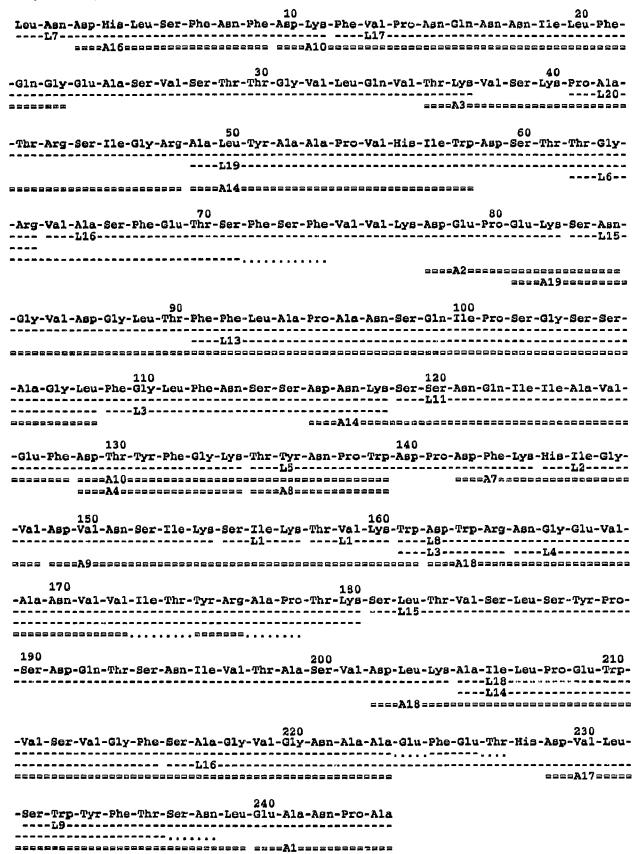


Fig. 3. Complete amino acid sequence of CSA-I. Primary structure in terms of the peptides obtained after digestion with Lys-C (L) and Asp-N (A). Residues were identified by automated sequence analysis of the Lys-C fragments (--L.--) or the Asp-N fragments (==A==). The numbering starts from the N terminus. Dots show residues not clearly determined at this position.

CSA-I LAA-I UEA-II UEA-I	10 20 30 L N D H L S F N F D R F V P N Q N N I L F Q G E A S V S T T G	V -
CSA-I LAA-I UEA-II UEA-I		T
CSA-I LAA-I UEA-II UEA-I	70 80 90 TGRVASFETSFSFVVKDEP.EKSNGVDGLT	0 7 F 4 -
CSA-I LAA-I UEA-II UEA-I	100 110 120 FLAPANSQIPSGSSAGLFGLFNSS.DNKSSNC) I - -) T
CSA-I LAA-I UEA-II UEA-I	130 140 150 I A V E F D T Y F G K T Y N P W D P D F K H I G V D V N S I K S	5 T
CSA-I LAA-I UEA-II UEA-I	160 170 180 K T V K W D W R N G E V A N V V I T Y R A P T R S L T V S I	: S
CSA-I LAA-I UBA-II UBA-I		3 V - T
LAA-I UEA-II	220 230 240 G N A A E F E T H D V L S W Y F T S N L E A N . P A	

Fig. 4. Comparison of complete amino acid sequence of CSA-1 with those of LAA-1, UEA-11 and UEA-1 and their carbohydrate-binding peptide sequences. The sequences were aligned to maximize homology. (-) denotes sequence identity with that of CSA-1. Dots represent gaps introduced for maximal alignment. The sequences of the putative carbohydrate-binding peptides of CSA-1, LAA-1, UEA-11 and UEA-1 are underlined.

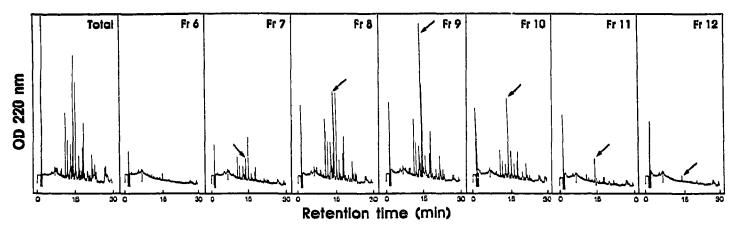


Fig. 5. Elution profiles on reversed-phase HPLC on a column of C₁₈ of fractions obtained by affinity chromatography of Asp-N digest of CSA-I. The purified CSA-I was digested with endoproteinase Asp-N at 37°C for 18 h. The reaction mixtures were applied to a column of GlcNAc oligomer-Sepharose as described in section 2. After the column had been washed with Tris-HCl, pH 6.8, containing 0.15 M NaCl, 1 mM CaCl₂ and 1 mM MnCl₂, each fraction was analyzed by reversed-phase HPLC on a column of C₁₈ as described in section 2. Arrows indicate the peptide of CSA-I retained on the affinity column.

column of GlcNAc oligomer-Sepharose. Each fraction was separately subjected to reversed-phase HPLC on a column of C_{18} . Fig. 5 shows the elution profiles on HPLC of the affinity fractions obtained from CSA-I. In this figure, the elution profile of the original mixture of the Asp-N digest of CSA-I is shown in the left panel (shown as Total). Comparison of the elution profiles of the affinity fractions with this original profile clearly shows that most of the Asp-N digest fragments of CSA-I were recovered in affinity fraction 8. The peptide fragment indicated by the arrows in Fig. 5 appeared to be retained on the GlcNAc oligomer-Sepharose column, because this fragment is a major component of fraction 9 and is also clearly observed in fraction 10, suggesting that it specifically interacts with the N-acetyl-D-glucosamine oligomer. The peptide retained on the column was purified by reversed-phase HPLC on C₁₈ and the amino acid sequence of this peptide was determined using a gas-phase protein sequencer. On the basis of the complete amino acid sequence of CSA-I shown in Fig. 3, it appears that the carbohydrate-binding peptide, DTYFGKTYNPW, from CSA-I corresponds to the sequence from Asp-129 to Trp-139.

3.3. Comparison of the sequence of the carbohydraiebinding peptide of CSA-I with those of LAA-I, UEA-II and UEA-I

The sequence of the peptide of CSA-I (underlined in Fig. 4), which was retained on the GlcNAc oligomer-Sepharose column, was compared with those of LAA-I, UEA-II and UEA-I (also underlined in Fig. 4), and this sequence apparently corresponds to a relatively conserved region of the anti-H(O) leguminous lectins. This region has already been demonstrated to represent carbohydrate-binding sites of LTA, UEA-I, UEA-II and LAA-I [22].

The amino acid sequence of the putative carbohydrate-binding peptide of CSA-I was DTYFGK-TYNPW and this amino acid sequence was compared with those of LAA-I (DTYFGKAYNPW) [22], UEA-II (DSYFGKTYNPW) [22] and UEA-I (DTIGSP-VNFW) [22]. The sequence of the carbohydrate-binding peptide of CSA-I was almost identical with those of LAA-I and UEA-II. The sequence of the carbohydratebinding peptide of UEA-I, an isolectin of UEA-II having L-fucose-binding specificity, was different from those of the three di-N-acetylchitobiose-binding lectins. In view of the differences in the species of the seeds from which CSA-I, LAA-I and UEA-II were isolated, the observed homology of these three lectins seems to be striking. The fact that these three kinds of anti-H(O) lectins have the same di-N-acetylchitobiose- binding specificity and no L-fucose-binding specificity is worthy of remark. The finding provides us with important clues as to the elucidation of the relationship between the carbohydrate-binding specificity and the structure of the carbohydrate-binding region of the leguminous plant seeds. Because the structural homology of these three di-N-acetylchitobiose-binding lectins is extremely high throughout the stretch of these polypeptides, these three lectins may consist of the products from a group of closely related genes.

REFERENCES

- [1] Matsumoto, I. and Osawa, T. (1969) Biochim. Biophys. Acta 194, 180-189
- [2] Matsumoto, I. and Osawa, T. (1972) Biochem. Biophys. Res. Commun. 46, 1810-1815.
- [3] Horejsi, V. and Kocourek, J. (1974) Biochim. Biophys. Acta 336, 329-337.
- [4] Frost, R.G., Reitherman, R.W., Miller, A.L. and O'Brien, J.S. (1975) Anal. Biochem. 69, 170-179.
- [5] Allen, H.J. and Johnson, E.A.Z. (1977) Carbohyd. Res. 58, 253-265.

- [6] Filka, K., Coupek, J. and Kocourek, J. (1978) Biochim. Biophys. Acta 539, 518-528.
- [7] Gurtler, L.G. (1978) Biochim. Biophys. Acta 544, 593-604.
- [8] Pereira, M.E.A., Kisailus, E.C., Gruezo, F.G. and Kabat, E.A. (1978) Arch. Biochem. Biophys. 185, 108-115.
- [9] Yariv, J., Kalb, A.J. and Katchalski, E. (1967) Nature 215, 890– 891.
- [10] Kalb, A.J. (1968) Biochim. Biophys. Acta 168, 532-536.
- [11] Blumberg, S., Hildesheim, J., Yariv, J. and Wilson, K.J. (1972) Biochim. Biophys. Acta 264, 171-176.
- [12] Pereira, M.E.A. and Kabat, E.A. (1974) Biochemistry 13, 3184–3192.
- [13] Matsumoto, I. and Osawa, T. (1970) Arch. Biochem. Biophys. 140, 484-491.
- [14] Pereira, M.E.A., Gruezo, F. and Kabat, E.A. (1979) Arch. Biochem. Biophys. 194, 511-525.
- [15] Horejsi, V. (1979) Biochim. Biophys. Acta 577, 389-393.

- [16] Konami, Y., Tsuji, T., Matsumoto, I. and Osawa, T. (1981) Hoppe-Seyler's Z. Physiol. Chem. 362, 983-989.
- [17] Konami, Y., Yamamoto, K., Tsuji, T., Matsumoto, 1. and Osawa, T. (1983) Hoppe-Seyler's Z. Physiol. Chem. 364, 397-407.
- [18] Konami, Y., Yamamoto, K. and Osawa, T. (1991) Biol. Chem. Hoppe-Seyler 372, 103-111.
- [19] Konami, Y., Yamamoto, K. and Osawa, T. (1990) FEBS Lett. 268, 281-286.
- [20] Konami, Y., Yamamoto, K. and Osawa, T. (1991) J. Biochem. 109, 650-658.
- [21] Konami, Y., Yamamoto, K., Toyoshima, S. and Osawa, T. (1991) FEBS Lett. 286, 33-38.
- [22] Yamamoto, K., Konami, Y., Osawa, T. and Irimura, T. (1992) J. Biochem. 111, 436-439.
- [23] Rupley, J.A. (1964) Biochim. Biophys. Acta 83, 245-255.
- [24] Baues, R.J. and Gray, G.R. (1977) J. Biol. Chem. 252, 57-60.