Volume 284, number 2, 161–164 FEBS 09823 June 1991

ADONIS 001457939100533T

The complete amino acid sequence of ribonuclease from the seeds of bitter gourd (*Momordica charantia*)

Hiroyuki Ide, Makoto Kimura, Mariko Arai and Gunki Funatsu

Laboratory of Protein Chemistry and Engineering, Faculty of Agriculture, Kyushu University, Fukuoka 812, Japan

Received 28 January 1991; revised version received 17 April 1991

The complete amino acid sequence of ribonuclease (RNase MC) from the seeds of bitter gourd (*Momordica charantia*) has been determined. This has been achieved by the sequence analysis of peptides derived by enzymatic digestion with trypsin, lysylendopeptidase, and chymotrypsin, as well as by chemical cleavage with cyanogen bromide. The protein contains 191 amino acid residues and has a calculated molecular mass of 21259 Da. Comparison of this sequence with sequences of the fungal RNases, RNase T2, and RNase Rh, revealed that there are highly conserved residues at positions 32–38 (TXHGLWP) and 81–92 (FWXHEWXKHGTC). Furthermore, the sequence of RNase MC was found to be homologous to those of *Nicotiana alata* *S*-glycoproteins involved in self-incompatibility sharing 41% identical residues.

Ribonuclease; *Momordica charantia*; Amino acid sequence; Sequence comparison

1. INTRODUCTION

A large number of ribonucleases (RNase) with different specificities have been isolated from a variety of organisms and were investigated extensively from both structural and functional points of view. Particularly, RNases with a lower molecular mass, such as RNase T1 and RNase A, are well understood in terms of a model system of nucleic acid/protein interaction [1–3]. In addition, some RNases have become invaluable tools in molecular biology research because of their restricted substrate specificity.

Although a vast amount of information has been accumulated on microbial and mammalian RNases, very few studies of the plant RNases have thus far been pursued. Recently, it was found that *S*-glycoproteins involved in self-incompatibility in *Nicotiana alata* are related to RNase T2 from *Aspergillus oryzae* [4]. Subsequently, McClure et al. demonstrated that *S*-glycoproteins in fact contain RNase activity [5] and that the enzyme exhibited its activity in pollen to hydrolyze rRNA, thereby arresting pollen tubule growth [6]. In addition, Nürnberger et al. found that RNase was induced upon phosphate starvation, and proposed that RNase functions to scavenge exogenous phosphate [7].

Correspondence address: G. Funatsu, Laboratory of Protein Chemistry and Engineering, Faculty of Agriculture, Kyushu University, Fukuoka 812, Japan

Abbreviations: ribonuclease, RNase; RNase MC, ribonuclease from *Momordica charantia*; *S*-CM-RNase MC, *S*-carboxymethylated RNase MC; CNBr, cyanogen bromide; RP-HPLC, reverse-phase high-performance liquid chromatography; DABITC/PITC, 4-*N*,*N*-dimethylaminoazobenzene-4′-isothiocyanate/phenylisothiocyanate

In the course of the studies on protein-synthesis inhibitory proteins, Watanabe et al. isolated a 19 kDa protein with RNase activity from the seeds of sponge gourd (*Luffa cylindrica*) [8], and also we isolated RNase MC, with a similar molecular mass and isoelectric point, in higher yield from the seeds of bitter gourd (*Momordica charantia*). Irie et al. showed, using dinucleoside monophosphates, that RNase MC has a remarkably high specificity toward CpU, ApU, and UpU (unpublished result). As an initial step toward understanding the structure/function relationships of RNases of plant origin, we have determined the primary structure of RNase MC and compared its sequence with those of other RNases.

2. MATERIALS AND METHODS

RNase MC was prepared from a low-pH extract of the seeds by gel-filtration on Sephadex G-75 and CM-cellulose column chromatography in 10 mM phosphate buffer, pH 6.5, followed by S-Sepharose column chromatography in 10 mM phosphate buffer, pH 6.0. RNase MC was eluted at around 0.25 M of NaCl from a CM-cellulose column with a linear gradient of NaCl. The protein, purified further by S-Sepharose column chromatography, migrated as a single component on SDS-PAGE (the details will be described elsewhere).

Succinylation and carboxymethylation of RNase MC were done by the procedures of Habeeb et al. [9] and Crestifield et al. [10], respectively.

Five mg of RNase MC was digested with 0.1 mg of trypsin in water adjusted to pH 8.0 with 0.1% NH_4OH at 37°C for 4 h. Tryptic digestion of the *N*-succinylated *S*-CM-RNase MC was performed by the same procedure. Five mg of RNase MC was digested with 0.1 mg of lysylendopeptidase in 10 mM Tris-HCl buffer, pH 9.0, containing 4 M urea, at 35°C for 4 h. Digestion of *S*-CM-RNase MC with chymotrypsin was carried out by as described above for tryptic digestion. 200 μg of RNase MC was digested by carboxypeptidase Y in 50 mM phosphate buffer, pH 6.5, containing 0.5% SDS for 5 min at

37°C using an enzyme/protein molar ratio of 1:50. CNBr cleavage of RNase MC was done as described by Steers et al. [11].

The peptides derived from enzymatic digestions as well as CNBr cleavage were separated by RP-HPLC on a C4 column of YMC-Gel (4.6 × 250 mm) using an acetonitrile gradient in 5 mM phosphate buffer, pH 6.0. In case of a second chromatography run, peptides were purified on YMC-Gel (4.6 × 150 mm) using 0.1% aqueous trifluoroacetic acid as the solvent.

Amino acid analysis was performed, after hydrolysis for 24 h at 110°C with 5.7 N HCl containing 0.05% 2-mercaptoethanol, on a Hitachi HPLC system (Type 655A). Sequence determination of the isolated peptides was carried out by manual Edman degradation employing the DABITC/PITC double-coupling procedure [12]. All sequence reagents were of the highest purity available and the other reagents were of analytical grade.

3. RESULTS AND DISCUSSION

3.1. *Sequence determination*

RNase MC, purified as described in Materials and Methods, was subjected to amino acid analysis and sugar analysis. The amino acid composition derived from protein hydrolysate was $Asp_{18.20}$, $Thr_{14.60}$, $Ser_{15.42}$, $Glu_{16.04}$, $Pro_{10.13}$, $Gly_{15.43}$, $Ala_{13.46}$, $Val_{8.28}$, $Met_{0.88}$, $Ile_{7.83}$, $Leu_{13.25}$, $Tyr_{2.97}$, $Phe_{15.22}$, $Lys_{10.05}$, $His_{5.05}$, $Arg_{9.16}$, $1/2Cys_{8.30}$. The sugar analysis with PAS-staining indicated the absence of carbohydrate chains in RNase MC. The N-terminal sequence of RNase MC up to position 9 was determined to be Phe-Asp-Ser-Phe-Trp-Phe-Val-Gln-Gln- by the DABITC/PITC double-coupling method. Carboxypeptidase Y digestion of RNase MC using 1:50 molar ratio of enzyme/substrate released the following equivalents of amino acids: Ile, 0.91; Phe 1.73. The complete amino acid sequence of RNase MC was derived from analyses of peptides obtained by cleavages with trypsin, lysylendopeptidases, chymotrypsin, and CNBr.

RNase MC was first digested with trypsin and the resulting peptides were isolated by several runs of RP-HPLC using a YMC-Gel C4 column. Twelve peptides were obtained in pure form and subjected to sequence analysis by manual Edman degradation using DABITC as a coupling reagent. As summarized in Fig. 1, the

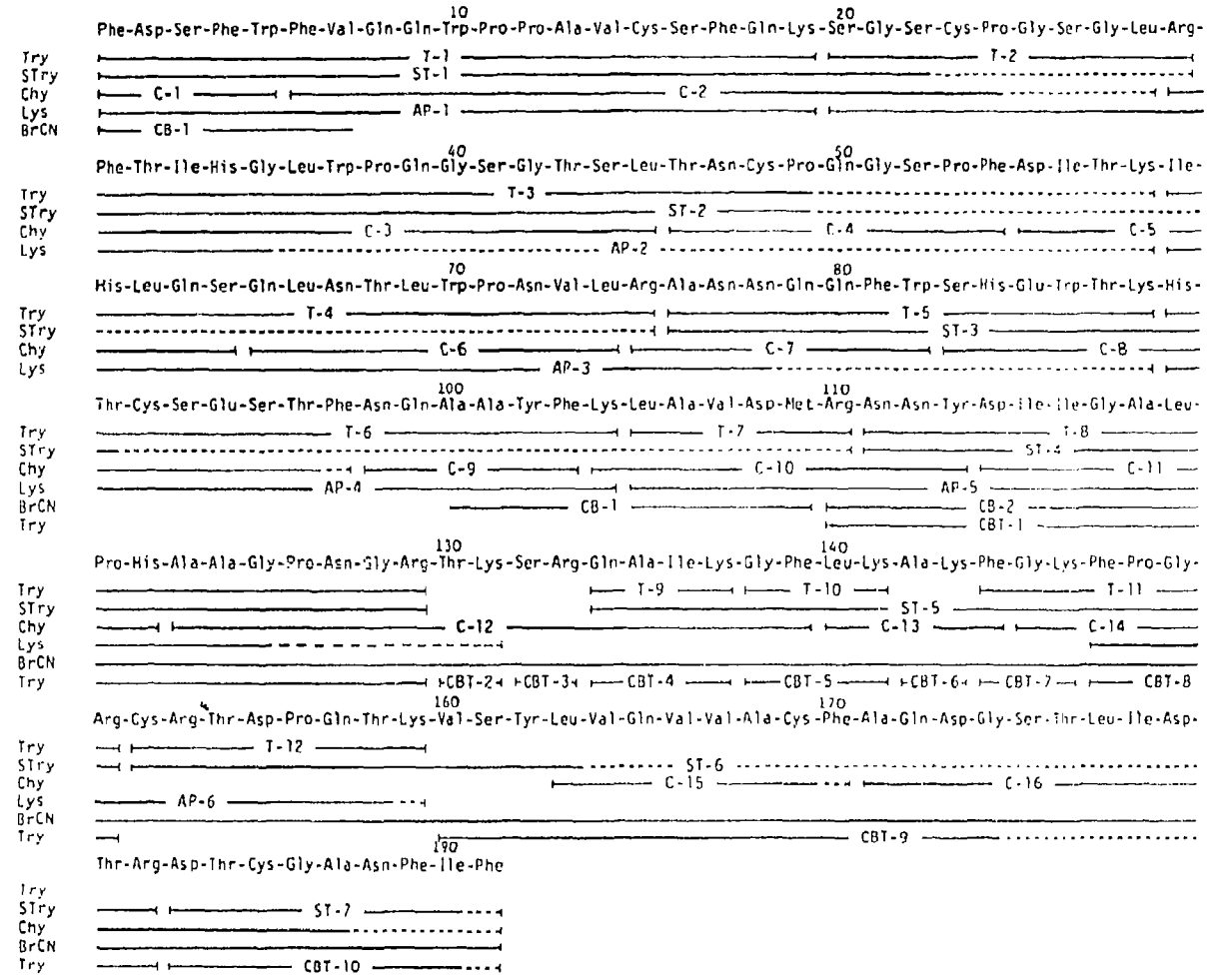

Fig. 1. The amino acid sequence of RNase MC from the seeds of *Momordica charantia*. Try and STry indicate peptides derived from cleavage with trypsin of the intact and *N*-succinylated RNase MC, respectively. Lys, Chy, and BrCN indicate peptides derived from cleavages with lysylendopeptidase, chymotrypsin and CNBr of *S*-CM-RNase MC, respectively. Sequence data on individual peptides are indicated as follows: (——) sequenced by the DABITC/PITC method; (– – –) not identified by sequencing.

complete sequences of all peptides except T-3 were determined. The alignment of tryptic peptides was then obtained from amino acid sequences of selective arginyl and lysyl cleavage products from RNase MC.

Tryptic digestion of the *N*-succinylated *S*-CM-RNase MC yielded 7 peptides which were separated by RP-HPLC. The amino acid sequences of these peptides were determined as summarized in Fig. 1. In this analysis, the C-terminus of RNase MC was deduced as Phe-Ile-Phe on the basis of sequence data and amino acid composition of peptide ST-7, together with the result obtained from carboxypeptidase Y digestion of the protein. *S*-CM-RNase MC was cleaved by lysylendopeptidase and the resulting peptides were resolved by RP-HPLC. Six peptides were obtained in pure form and subjected to sequence analysis. Although sequence information thus determined could overlap most of the tryptic peptides, two parts of the middle regions (positions 50–58 and 130–133) and the C-terminal region (position 164–191) were still not established. These gaps were filled by sequencing the peptides obtained by chymotryptic digestion and CNBr cleavage of the protein.

Chymotryptic digestion of the protein yielded sixteen peptides which were separated by RP-HPLC and the resulting peptides were sequenced. Sequence of peptides C-4, C-5 and C-12 could establish the remainder of the sequence, and those of peptides C-10, C-15 and C-16 could provide alignments T-7 to T-8, and ST-6 to ST-7. For determination of the sequence of the C-terminal region, 5 mg of the protein was cleaved with CNBr in 70% formic acid for 20 h. As predicted from the presence of one methionine residue, two fragments (CB-1 and CB-2) were obtained in yields of 70% and 75%, respectively. The C-terminal fragment CB-2 was further digested with 20 μg of trypsin and the digest was separated by RP-HPLC. Sequence analysis of peptide CB-T-9 could provide the overlap between peptides C-15 and C-16. By the combination of all results, the complete amino acid sequence of RNase MC was unambiguously determined as given in Fig. 1.

The protein consists of 191 amino acid residues and the amino acid composition derived from the sequence is Asp_8, Asn_{10}, Thr_{15}, Ser_{16}, Glu_2, Gln_{14}, Pro_{11}, Gly_{16}, Ala_{13}, Val_8, Met_1, Ile_8, Leu_{13}, Tyr_3, Phe_{15}, Lys_{10}, His_5, Arg_9, Trp_6, $1/2Cys_8$. This is in agreement with the result obtained from the amino acid analysis of the protein. The protein sequence thus determined includes no potential Asn-linked glycosylation sites (Asn-Xaa-Thr/Ser), which is consistent with the finding that RNase MC is insensitive to PAS-staining. From the sequence, the molecular weight of the protein was calculated to be 21 259 Da.

3.2. *Sequence comparison*

Comparison of the amino acid sequence of RNase MC with those of other RNases revealed that RNase MC shows homology to RNase T2 from *Aspergillus oryzae* [13] and to RNase Rh from *Rhizopus niveus* [14], which are base non-specific RNases from fungi, as shown in Fig. 2. Although the substrate specificity of RNase MC is different from those of fungal RNases and the level of sequence identity is relatively low, they share two segments of conserved sequence, TXHGLWP and FWXHEWXKHGTC, at positions 32–38 and 81–92 in the sequence of RNase MC, respectively. Recently, His-34 (His-53 in RNase T2) and His-89 (His-115) in RNase T2) were identified as the active site residues in RNase T2 [4] and also Glu-85 (Glu-105 in RNase Rh) in RNase Rh (M. Irie, personal communication). Interestingly, these three residues are conserved in RNase MC, suggesting that they may be involved in

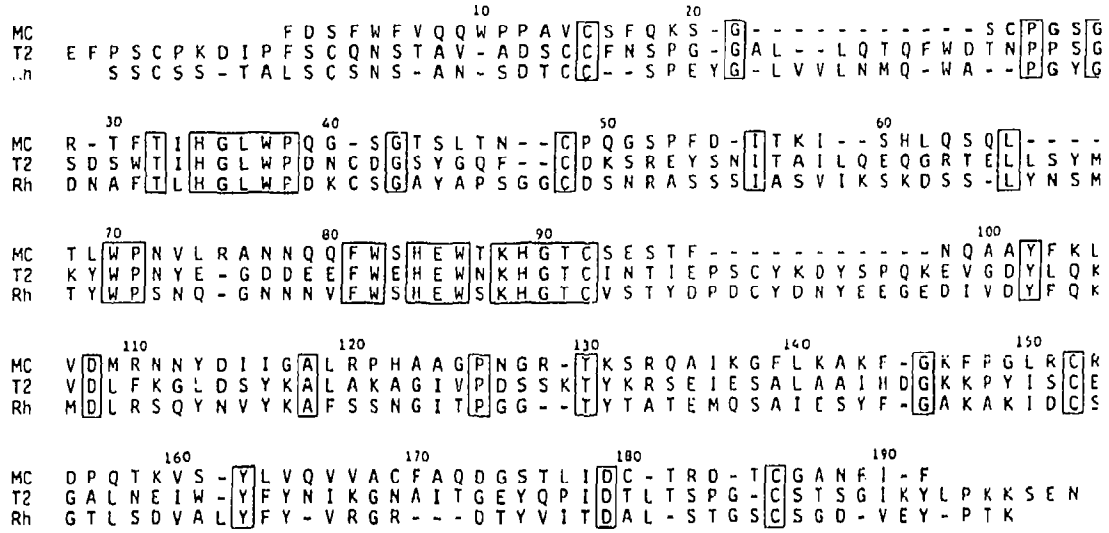

Fig. 2. Comparison of the amino acid sequence of RNase MC with those of RNase T2 and RNase Rh from *Aspergillus oryzae* [13] and *Rhizopus niveus* [14], respectively. Identical residues are enclosed in boxes.

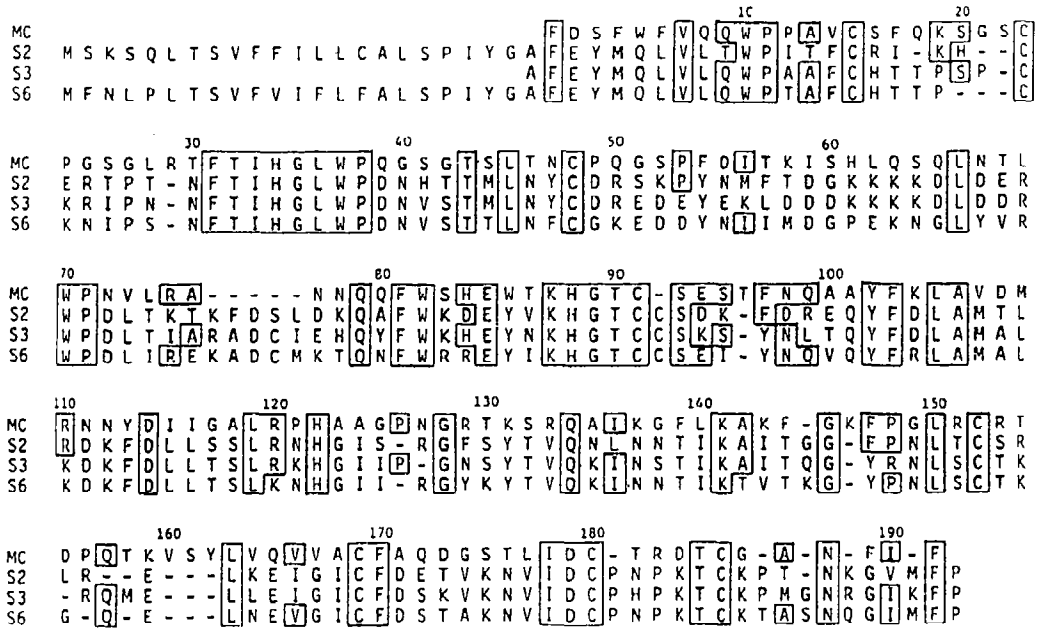

Fig. 3. Comparison of the amino acid sequence of RNase MC with those of *Nicotiana alata* S-glycoproteins S2, S3, and S6 [15,16]. The residues identical with RNase MC are boxed.

the catalytic mechanism and/or in the formation of the active site in RNase MC.

Recently, McClure et al. found that *Nicotiana alata* S-glycoproteins involved in self-incompatibility are RNase [5]. Hence, the sequence of RNase MC was also compared with those of the S-glycoproteins S2, S3 and S6, as shown in Fig. 3. This comparison shows that the sequence of RNase MC can be readily aligned with those of S-glycoproteins by making some insertions and deletions and that RNase MC is about equally homologous to all S-glycoproteins, sharing 41% identical residues; this value is higher than those obtained by comparison with fungal RNases. The putative active sites around the His residues described above in RNase MC are again highly conserved in all S-glycoproteins, having 8 and 5 consecutive identical residues. In addition to the two conserved segments, there are also other individual positions in the sequence which are completely conserved in all molecules, as shown in Fig. 3. In particular, all 8 cysteine residues in RNase MC are totally conserved in all S-glycoproteins, whereas 3 cysteine residues, Cys-23, Cys-169, and Cys-180, are substituted in the sequences of the fungal RNases. This finding would suggest that the RNase MC and S-glycoproteins share structural features and that they have diverged from a common ancestral protein.

REFERENCES

[1] Heinemann, U. and Saenger, W. (1982) Nature 299, 27–31.

[2] Wlodawer, A., Bott, R. and Sjolin, L. (1982) J. Biol. Chem. 257, 1325–1332.

[3] Arni, R., Heinemann, U., Tokuoka, R. and Saenger, W. (1988) J. Biol. Chem. 263, 15358–15368.

[4] Kawata, Y., Sakiyama, F., Hayashi, F. and Kyogoku, Y. (1990) Eur. J. Biochem. 187, 255–262.

[5] McClure, B.A., Haring, V., Ebert, P.R., Anderson, M.A., Simpson, R.J., Sakiyama, F. and Clarke, A.E. (1989) Nature 342, 955–957.

[6] McClure, B.A., Gray, J.E., Anderson, M.A. and Clarke, A.E. (1990) Nature 347, 757–760.

[7] Nürnberger, T., Abel, S., Jost, W. and Glund, K. (1990) Plant Physiol. 92, 970–976.

[8] Watanabe, K., Minami, Y. and Funatsu, G. (1990) Agric. Biol. Chem. 54, 2085–2092.

[9] Habeeb, A.F.S.A., Cassidy, H.G. and Singer, S.J. (1958) Biochim. Biophys. Acta 29, 587–593.

[10] Crestfield, A.M., Moore, S. and Stein, W.H. (1963) J. Biol. Chem. 238, 622–627.

[11] Steers, E., Craven, G.R., Anfinsen, C.B. and Bethune, J.L. (1965) J. Biol. Chem. 240, 2478–2484.

[12] Chang, J.Y., Brauer, D. and Wittmann-Liebold, B. (1978) FEBS Lett. 93, 205–214.

[13] Kawata, Y., Sakiyama, F. and Tamaoki, H. (1988) Eur. J. Biochem. 176, 683–697.

[14] Horiuchi, H., Yanai, K., Takagi, M., Yano, K., Wakabayashi, E., Sanda, A., Mine, S., Ohgi, K. and Irie, M. (1988) J. Biochem. 103, 408–418.

[15] Anderson, M.A., Cornish, E.C., Mau, S.-L., Williams, E.G., Hoggart, R., Atkinson, A., Bonig, I., Grego, B., Simpson, R., Roche, P.J., Haley, J.D., Penschow, J.D., Niall, H.D., Tregear, G.W., Coghlan, J.P., Crawford, R.J. and Clarke, A.E. (1986) Nature 321, 38–44.

[16] Anderson, M.A., McFadden, G.I., Bernatzky, R., Atkinson, A., Orpin, T., Dedman, H., Tregear, G., Fernley, R. and Clarke, A.E. (1989) Plant Cell 1, 483–492.