Volume 329, number 3, 336–340 FEBS 12816 August 1993

Complete amino acid sequence of puroindoline, a new basic and cystine-rich protein with a unique tryptophan-rich domain, isolated from wheat endosperm by Triton X-114 phase partitioning

Jean-Erik Blochet[a], Catherine Chevalier[b], Eric Forest[c], Eva Pebay-Peyroula[c], Marie-Françoise Gautier[d], Philippe Joudrier[d], Michel Pézolet[e], Didier Marion[a,*]

[a]*Laboratoire de Biochimie et Technologie des Protéines, INRA, 44026 Nantes Cédex 03, France*
[b]*Institut de Biologie, INSERM U211, 44035 Nantes Cédex 01, France*
[c]*Institut de Biologie Structurale, 38047 Grenoble Cédex, France*
[d]*Laboratoire de Technologie des Céréales, INRA, 34060 Montpellier Cédex 01, France*
[e]*CERSIM, Département de Chimie, Université Laval, Québec G1K 7P4, Canada*

Received 25 January 1993; revised version received 7 June 1993

A new basic protein has been isolated from wheat endosperm by Triton X-114 phase partitioning. It contains five disulfide bridges and is composed of equal amounts of a polypeptide chain of 115 amino acid residues and of the same chain with a C-terminus dipeptide extension. The most striking sequence feature is the presence of a unique tryptophan-rich domain so that this protein isolated from wheat seeds has been named puroindoline. The similar phase partitioning behavior in Triton X-114 of this basic cystine-rich protein and of purothionins suggests that puroindoline may also be a membranotoxin that might play a role in the defense mechanism of plants against microbial pathogens.

Puroindoline; Thionin; Triton X-114; Basic protein; Cystine-rich protein; Tryptophan

1. INTRODUCTION

During the hydrated part of their life, i.e. maturation and germination, seeds are very sensitive to viruses, fungi and bacteria. Many proteins involved in the microbial defense mechanism of plants have been identified and some of them display, in vitro, an effective antimicrobial activity [1,2]. Thionins were the first antimicrobial proteins to be isolated from wheat endosperm half a century ago and they have been the subject of several studies [3,4]. These low molecular weight basic and cystine-rich proteins are toxic for various microorganisms and animal cells and this cytotoxicity is most probably exerted through their interaction with membranes and especially with membrane lipids (see [3,4] for reviews).

These biological properties of thionins on membranes were also observed for some protein and peptide toxins found in the venom of different animals [5,6]. However, these toxins generally exhibit an important polymorphism [7,8] while thionins are so far the only wheat endosperm membranotoxins to have been identified. The non-specific lipid transfer proteins (nsLTP), that are other abundant basic and cystine-rich proteins of wheat endosperm [9,10], are known to interact with cell membranes [9,11]. It has been shown that the radish seed nsLTP has, in vitro, an antifungal activity but its efficiency is very low compared to that of thionins [12]. Recently, by using the TX114 phase partitioning method generally used to isolate proteins that are tightly bound to membrane lipids [13], we have isolated in the detergent-rich phase both thionins and other unknown basic and cystine-rich proteins [14] that might have a similar antimicrobial activity as that of thionins in perturbing membrane integrity. In this paper, we report on the purification and on the primary structure of the most abundant protein of these new basic proteins.

*Corresponding author.

Abbreviations: nsLTP, non-specific lipid transfer protein; HPS, hydrophobic protein of soja; TX114, Triton X-114; EDTA, ethylene-diamine-tetraacetic acid; RP-HPLC, reverse-phase high-performance liquid chromatography; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis.

2. MATERIALS AND METHODS

2.1. *Materials*

Triticum aestivum (var. Camp Rémy) seeds were kindly provided by Dr. Doussinault from the Station d'Amélioration des Plantes (Rennes, Le Rheu, France). Endosperm flour was prepared using a Brabender milling device. Sequencing grade endoproteases were from Boehringer Mannheim (Peypin, France) and carboxypeptidase Y from Sigma (France). Organic solvents were either HPLC (Merck, France) or sequencing grades (Pierce, Netherlands). All other chemicals were of analytical grade. Sephadex G75 was from Pharmacia (France) and the Neobar CS cation-exchange column was from Dyno Particles (Nor-

Published by Elsevier Science Publishers B.V.

way). RP-HPLC columns were from the Merck or from Société Française de Chromatographie (France). FSTPSCAN and CLUSTAL programs provided by the PC-GENE software (Intelligenetics, USA) were used for the determination of sequence homologies in the SWISS PROT data bank.

2.2. *Extraction and purification of wheat endosperm proteins*

TX114 phase partitioning was used on 1 kg wheat endosperm flour according to a modification of a procedure previously described [14]. Soluble proteins were extracted at 4°C by stirring 1 kg wheat endosperm flour with 10 l of Tris-HCl 100 mM pH 7.8 containing 5 mM EDTA and 0.1 M KCl (Tris-KCl buffer). After centrifugation at 5,000 × *g* the residue was extracted at 4°C with 5 l of Tris-KCl buffer containing 4% TX114 (w/w). After centrifugation at 5,000 × *g*, the supernatant was heated to 30°C for 1 h and centrifuged 15 min at 5,000 × *g*. The upper detergent poor phase was discarded and the same volume of fresh Tris-KCl buffer containing 0.06% TX114 was added. The solution was mixed for 1 h at 4°C and phase partitioning procedure was repeated. Finally, the upper phase was discarded and the lower detergent-rich phase was removed and the proteins were precipitated overnight with 5 l of diethylether/ethanol 1:3 (v/v) at −20°C. After centrifugation at 2,000 × *g* the protein pellet was reextracted at −20°C three times with 1 l of diethylether-ethanol and finally with 1 l of diethylether. After centrifugation at 2,000 × *g*, the protein pellet was dried overnight under reduced pressure.

The dry white protein powder obtained after extraction was dispersed in 0.05 M acetic acid (5% w/v) and solid NaCl was added to the supernatant up to 20 M. After centrifugation at 15,000 × *g*, the supernatant containing purothionins was discarded and the pellet was dialyzed overnight against 0.05 M acetic acid and fractionated on a column (10 cm × 100 cm) packed with Sephadex G75 (Pharmacia, France) as previously described [9,14]. The eluted fractions were analyzed by SDS-PAGE according to the procedure of Laemmli [15] and those containing polypeptides with apparent molecular weight below 20 kDa were pooled and freeze-dried. 50 mg of this protein fraction were dissolved in 0.05 M ammonium acetate, pH 5.5, and applied to a cation exchange CS Neobar HPLC column (1.3 × 3 cm). Proteins were eluted with a concentration gradient of ammonium acetate, pH 5.5 containing 20% acetonitrile, from 0.05 to 0.7 M. The flow rate was 3 ml/min.

2.3. *Determination of the amino acid sequence*

Reduction and alkylation with 4-vinylpyridine and subsequent enzymatic hydrolysis by trypsin and endoproteinase Lys-C were carried out as previously described [9]. The pyridylethylated protein was digested by chymotrypsin under the same conditions used for trypsin hydrolysis. The protein was hydrolyzed by endoproteinase Asp-N according to the procedure suggested by the manufacturer. Cyanogen bromide cleavage on the alkylated protein dissolved in an aqueous solution of 70% formic acid was carried out under nitrogen for 24 h. Fractionation of peptides, amino acid analysis and N-terminal amino acid sequencing were previously described [9].

2.4. *Electrospray mass spectrometry*

Precise molecular weight determination of proteins were obtained using a Perkin-Elmer Sciex APIII triple quadrupole mass spectrometer equipped with a nebuliser-assisted electrospray (ionspray) source [16]. The protein was introduced into the ion source of the spectrometer as a 40 μM aqueous solution containing 10% formic acid.

3. RESULTS AND DISCUSSION

The ion-exchange chromatography elution profile of the low molecular weight protein fraction collected after size exclusion chromatography is presented in Fig. 1. A main basic protein is eluted between 0.38 and 0.44 M ammonium acetate pH 5.5. Sequencing indicates that its purity is above 95% and SDS-PAGE shows a single band (Fig. 1) corresponding to a protein with an apparent molecular weight of 16 kDa. After dialysis, about 150 mg of freeze-dried protein were obtained from 1 kg of endosperm flour. This yield is twice that generally obtained for purothionins [17].

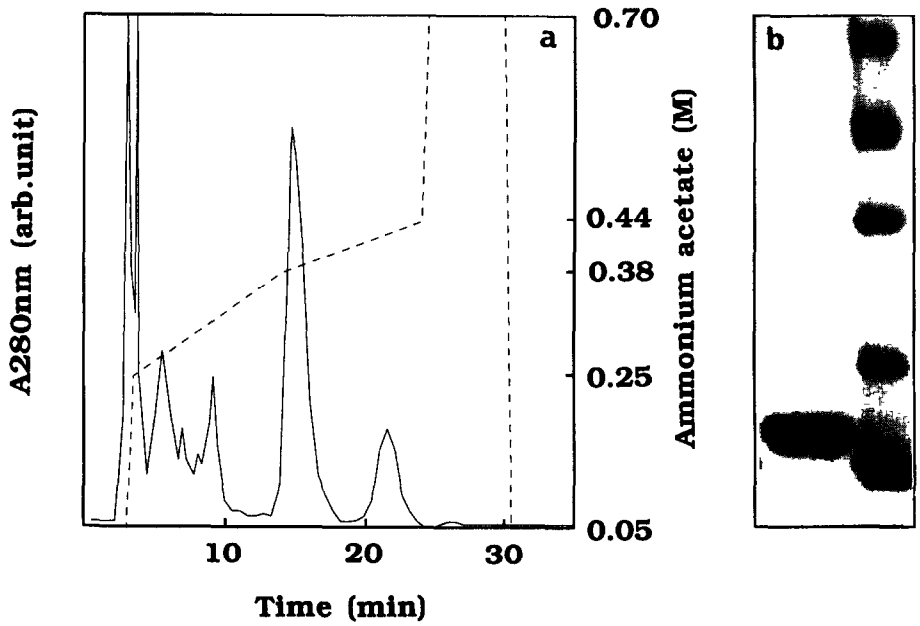

Fig. 1. (a) Cation-exchange chromatography of the low molecular weight proteins obtained after TX114 phase partitioning and size exclusion chromatography. Proteins were eluted using an ammonium acetate gradient from 0.05 M to 0.7 M containing 20% (v/v) acetonitrile (flow rate: 3 ml/min). (b) SDS-PAGE of the main protein eluting at about 0.38 M (left side) and standard proteins from bottom to top 14.4 (α-lactalbumin), 20.1 (trypsin inhibitor), 30 (carbonic anhydrase), 43 (ovalbumin) and 67 kDa (albumin) (right side).

The complete amino acid sequence obtained by automated sequence analysis by Edman degradation of the whole protein and of peptides obtained by enzymatic hydrolysis and cyanogen bromide cleavage is shown in Fig. 2. The carboxypeptidase digestion resulted in the simultaneous release of threonine, isoleucine and glycine during the first minutes that could be due to an heterogeneity at the C-terminus. The sequence and the heterogeneity at the C-terminus was confirmed by electrospray mass spectrometry where two main peaks corresponding to proteins with molecular weight of 12,749 and 12,918 were obtained for the native protein with a mass precision of 2 (Fig. 3). The lower molecular weight is in good agreement with the determined 115 amino acid sequence that ends with a threonine residue while the higher molecular weight peak could only correspond to the same protein with an Ile-Gly dipeptide extension after the C-terminal threonine. A more precise analysis of the mass spectrum of the protein allowed to detect minor fragments corresponding to the two main polypeptide chains having lost the N-terminal Asp residue (peaks at 12,634 and 12,803), and a polypeptide chain of 118 amino acid residues corresponding to an Ile-Gly-Tyr tripeptide C-terminal extension (Fig. 3). This heterogeneity of the protein chain length was not suppressed by extracting the protein in the presence of a mixture of protease inhibitors (results not shown), suggesting that this phenomenon is a normal pos-synthetic event and does not occur during extraction. Such a heterogeneity was observed at the N-terminal end of the hydrophobic protein of soybean (HPS) which also belongs to the family of cystine-rich proteins [18,19]. It was recently shown that the homologous hydrophobic protein of maize is synthesized as a precursor of higher molecular weight with an extension at the N-terminus [20] and that thionins are synthesized as a precursor of higher molecular weight with an acidic polypeptide at the C-terminus of the mature protein [3,4]. This suggests that the protein isolated by TX114 phase partitioning might be synthesized with polypeptide extensions at the N- and especially at the C-terminus, and that the post-translational cleavages were not quite specific, as for HPS. In this regard, it is interesting to note that carboxypeptidases are involved in the maturation of barley amylases [21].

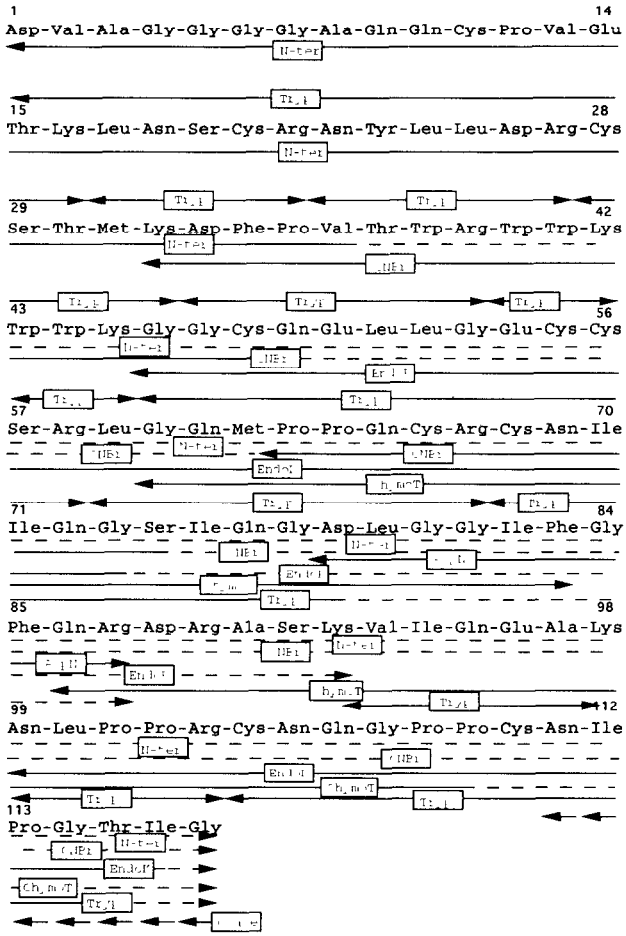

Fig. 2. The amino acid sequence of puroindoline as determined by automatic sequencing of peptides obtained by enzymatic hydrolysis and cyanogen bromide cleavage of the pyridylethylated protein. N-ter, N-terminal sequence; Tryp, tryptic peptide; chymoT, chymotryptic peptide; EndoK, endoproteinase Lys-C peptide; AspN, endoproteinase Asp-N peptide; CNBr, cyanogen bromide peptide; Cxase, carboxypeptidase Y amino acid release. Dashed lines represent unsequenced peptides residues.

The new protein, extracted by non-ionic detergent has a mean molecular weight of 12.8 kDa and is a basic protein with a calculated isoelectric point of 11. The fact that the molecular weight was overestimated by SDS-PAGE has already been observed for other basic proteins [22]. The cysteine content is identical to that found for the exogenous α-amylase inhibitors where the Cys-Cys and Cys-Arg-Cys motifs are also found [23–26], suggesting that all cysteines are involved in disulfide bonds. This is further confirmed by the fact that the cysteine residues could not be alkylated without reduction of the protein (results not shown). These Cys-Cys and Cys-X-Cys sequence features are also found in lipid transfer proteins [9,10]. However, the distribution of the other cysteine residues differs from that observed for exogeneous α-amylase inhibitors, and compared to the wheat nsLTP, the new protein has two additional cysteine residues in the N-terminal domain (Fig. 4). Finally, no significant identity or similarity is found with the sequences of other wheat endosperm proteins. Among wheat endosperm proteins, only those extracted with TX114 are able to interact tightly with membrane lipids. The new protein displays some common cysteine features with purothionins but differs markedly by its molecular weight and amino acid sequence.

The most surprising structural feature of this protein is the presence of an amphiphilic domain from residues 38 to 45 containing 5 tryptophan and 3 basic residues. From the SWISS-PROT data bank, such a tryptophan rich domain is unique for a protein. However, a tetrade-

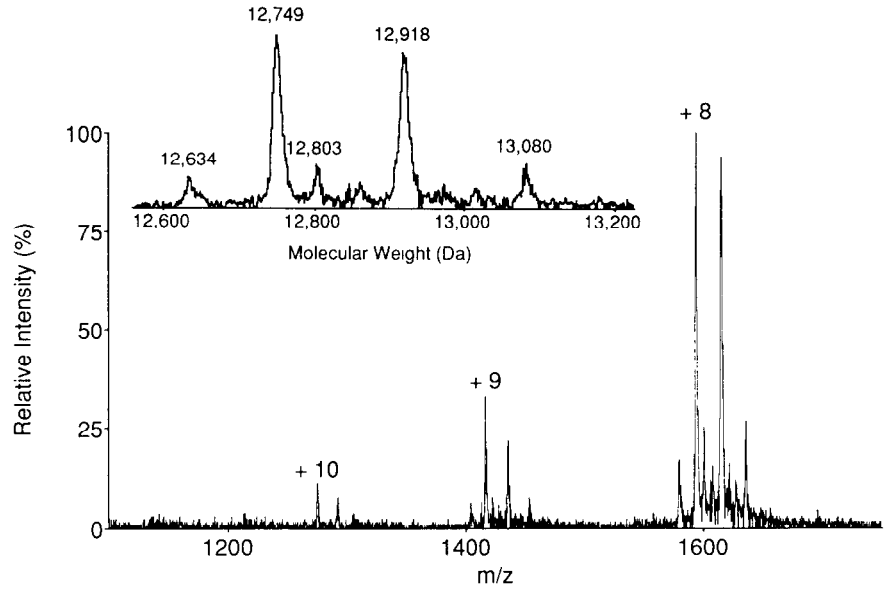

Fig 3. Electrospray mass spectrum of 20 pmol of protein. The spectrum is the resulting sum of seven 7s scans in the mass-to-charge ratio range 1,100–1,750. The insert shows the reconstructed molecular mass profile (as obtained by using a deconvolution algorithm program).

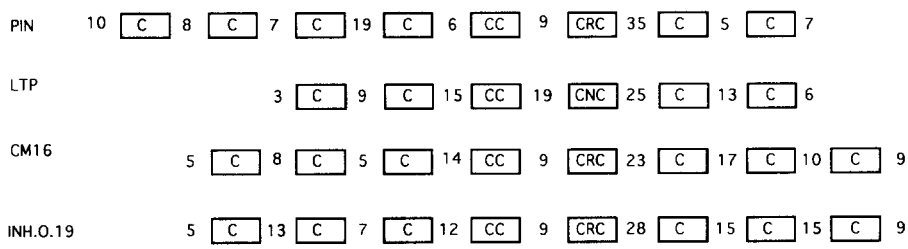

Fig. 4. Cysteine alignments of different wheat cystine-rich proteins. Puroindoline (PIN), lipid transfer protein (LTP) [9,10]; a subunit of the exogeneous α-amylase tetrameric inhibitor (CM16) [24,25]; a dimeric exogeneous α-amylase inhibitor (INH 0.19) [23]. Number between cysteine refers to the amino acid length of inserted peptides. Letters in boxes are cysteine (C), arginine (R) and asparagine (N).

capeptide, indolicidin, has been recently isolated from bovine neutrophils and exhibits a similar tryptophan-rich sequence [27]. In vitro, this peptide has antibacterial and antifungal activities. Tryptophan is generally involved in the 'toxic site' of different animal toxins which perturb the membrane function [28]. Thus, the wheat protein isolated in the current study has been named puroindoline in regard to the presence of this unique tryptophan rich domain. The similar TX114 phase partitioning behavior of puroindoline and purothionins suggests that the new basic protein can strongly bind polar lipids and, therefore, membranes. Like purothionins, this putative membranotoxin might exhibit antibacterial and antifungal properties and might, thus, contribute to the defense mechanism of plants against predators.

Acknowledgements: The authors are grateful to Jean-Pierre Compoint for his technical assistance for the purification of the protein. This work was supported by grants from Institut National de la Recherche Agronomique (INRA) (AIP 91/4749) and from the European Economic Community through ECLAIR (AGRE 0052). M.P. received a sabbatical fellowship from INRA and J.E.B. was supported by BSN and INRA.

REFERENCES

[1] Collinge, D.B. and Slusarenko, A.J. (1987) Plant Mol. Biol. 9, 389–390.
[2] Dixon, R.A. and Harrison, M.J. (1990) Adv. Gen 28, 165–234.
[3] Garcia-Olmedo, F., Rodriguez-Palenzuela, P., Hernandez-Lucas, C., Ponz, F., Marana, C., Carmona, M.J., Lopez-Fando, J., Fernandez, J.A. and Carbonero, P. (1989) Oxford Surv. Plant Mol. Cell Biol. 6, 31–60.
[4] Bohlmann, H. and Apel, K. (1991) Annu. Rev. Plant Physiol. Plant Mol. Biol. 42, 227–240.
[5] Dufourcq, J. and Faucon, J.F. (1977) Biochim. Biophys. Acta 467, 1–11.
[6] Ménez, A., Gatineau, E., Roumestand, C., Harvey, A.L., Mouawad, L., Gilquin, B. and Toma, F. (1990) Biochimie 72, 575–578.
[7] Endo, T. and Tamiya, N. (1987) Pharmacol. Ther. 34, 403–451.
[8] Olivera, B.M., Rivier, J., Scott, J.K., Hillyard, D.R. and Cruz, L.J. (1991) J. Biol. Chem. 25, 22067–22070.

[9] Désormeaux, A., Blochet, J.E., Pézolet, M. and Marion, D. (1992) Biochim. Biophys. Acta 1121, 137–152.
[10] Dieryck, W., Gautier, M.F., Lullien, V. and Joudrier, P. (1992) Plant Mol. Biol. 19, 707–709.
[11] Kader, J.C. (1990) in: Subcellular Biochemistry (Hilderson, H.J. Ed.) pp. 69–111, Plenum, New York.
[12] Terras, F.R.G., Goderis, I.J., Van Leuven, F., Vanderleyden, J., Cammue, B.P.A. and Broekaert, W.F. (1992) Plant Physiol. 100, 1055–1058.
[13] Bordier, C. (1981) J. Biol. Chem. 25, 1604–1607.
[14] Blochet, J.E., Kaboulou, A., Compoint, J.P. and Marion, D. (1991) in: Gluten Proteins (Bushuk, W. and Tkachuk, R. Eds.) pp. 314–325, American Association of Cereal Chemists, Minnesota, St-Paul, USA.
[15] Laemmli, U.K. (1970) Nature 227, 680–685.
[16] Fenn, J.B., Mann, M., Meng, C., Wong, S.F. and Whitehouse, C.M. (1989) Science 246, 64–71.
[17] Mak, A.S. and Jones, B.L. (1976) J. Sci. Food Agric. 27, 205–213.
[18] Odani, S., Takehiko, K., Ono, T., Seto, Y. and Tanaka, T. (1987) Eur. J. Biochem. 162, 485–491.
[19] Baud, F., Pebay-Peyroula, E., Cohen-Addad, C., Odani, S. and Lehmann, M.S. (1993) J. Mol. Biol., in press.
[20] Josè-Estanyol, M., Ruiz-Avila, L. and Puigdomenech, P. (1992) Plant Cell 4, 413–423.
[21] Sogaard, M., Olsen, F.L. and Svensson, B. (1991) Proc. Natl. Acad. Sci. USA 88, 8140–8144.
[22] See, Y.P. and Jackowski, G. (1989) in: Protein Structure: a Practical Approach (Creighton, T.E. Ed.) pp. 1–21, IRL Press, Oxford, UK.
[23] Maeda, K., Kakabayashi, S. and Matsubara, H. (1985) Biochim. Biophys. Acta 828, 213–221.
[24] Garcia-Maroto, F., Marana, C., Mena, M., Garcia-Olmedo, F. and Carbonero, P. (1990) Plant Mol. Biol. 14, 845–853.
[25] Gautier, M.F., Alary, R. and Joudrier, P. (1990) Plant Mol. Biol. 14, 313–322.
[26] Gautier, M.F., Alary, R., Lullien, V. and Joudrier, P. (1991) Plant Mol. Biol. 16, 333–334.
[27] Selsted, M.E., Levy, J.N., Van Abel, R.J., Cullor, J.S., Bontems, R.J. and Barany, G. (1991) in: Peptides: Chemistry and Biology (Smith, J.A. and Rivier, J.E. Eds.) pp. 905–907, ESCOM Science Publishers, Leiden, The Netherlands.
[28] Gatineau, E., Toma, F., Montenay-Garestier, T., Takechi, M., Fromageot, P. and Ménez, A. (1987) Biochemistry 26, 8046–8055.