Tiziana Cabras[1]
Federica Iavarone[2]
Davide Pirolli[2]
Maria Cristina De Rosa[3]
Alberto Vitali[3]
Gavino Faa[4]
Massimo Cordaro[5]
Irene Messana[1]
Jörgen Ekström[6]
Massimo Castagnola[2,3]

[1]Dipartimento di Scienze della Vita e dell'Ambiente, Univ. di Cagliari, Cagliari, Italy
[2]Istituto di Biochimica e Biochimica Clinica, Facoltà di Medicina, Università Cattolica, Roma, Italy
[3]Istituto di Chimica del Riconoscimento Molecolare, CNR, Roma, Italy
[4]Dipartimento di Scienze Chirurgiche, Sezione di Patologia, Univ. di Cagliari, Cagliari, Italy
[5]Istituto di Clinica Odontoiatrica, Facoltà di Medicina, Università Cattolica, Roma, Italy
[6]Department of Pharmacology, Institute of Neuroscience and Physiology, Sahlgrenska Academy at the University of Gothenburg, Göteborg, Sweden and Dipartimento di Scienze Biomediche, Univ. di Cagliari, Cagliari, Italy

Received March 21, 2013
Revised April 17, 2013
Accepted April 17, 2013

Research Article

Top-down HPLC–ESI-MS characterization of rat gliadoralin A, a new member of the family of rat submandibular gland glutamine-rich proteins and potential substrate of transglutaminase

During HPLC–ESI-MS/MS analysis of rat submandibular saliva secreted under isoprenaline stimulation, a protein with an experimental $[M+H]^{1+} = 10\,544.24$ m/z was detected (17.5 ± 0.7 min). The MS/MS fragmentation pattern, manually investigated, allowed establishing an internal sequence in agreement with a DNA-derived sequence of an unknown rat protein coded D3Z9M3 (Swiss-Prot). To match the experimental MS/MS fragmentation pattern and protein mass with theoretical data, the removal from the N terminus of the signal peptide and from the C terminus of three amino acid (a.a.) residues (Arg-Ala-Val) and the cyclization of the N-terminal glutamine in pyroglutamic had to be supposed, resulting in a mature protein of 90 a.a. HPLC–ESI-MS/MS of the trypsin digest ensured 100% sequence coverage. For the high glutamine content (34/90 = 37.8%) we propose to name this protein rat gliadoralin A 1–90. Low amounts of five different isoforms were sporadically detected, which did not significantly change their relative amounts after stimulation. Gliadoralin A is substrate for transglutaminase-2, having Lys 60 and different Gln residues as major determinants for enzyme recognition. *In silico* investigation of superior structures evidenced that a small part of the protein adopts an α-helical fold, whereas large segments are unfolded, suggesting an unordered conformation.

Keywords: Glutamine-rich protein / Rat / Saliva / Submandibular / Top-down
DOI 10.1002/jssc.201300312

1 Introduction

Saliva is a bodily fluid responsible for the protection of hard and soft tissues of the oral cavity and attractive for diagnosis of diseases because its collection is economical and noninvasive and it can be performed without the help of healthcare workers [1]. For this reason, many studies with different proteomic platforms have been recently carried out to deeply investigate human salivary proteome generating a list of more than 2400 different protein components [2–5]. Salivary proteins can be divided into two families, one including proteins secreted by salivary glands (parotid, submandibular, sublingual, and minor glands) that account not more than 400 components, but represent more than 90% in weight of the salivary proteome and the second including all the others originating from defoliating epithelial cells and mucosal transudate [5]. In humans the proteins of the first family can be further divided into several classes, such as proline-rich proteins (PRPs; acidic, basic, and basic glycosylated), mucins, α-amylases, S-type cystatins, histatins, statherin, and P-B peptide [6, 7]. The most demanding challenge of the current investigations of the saliva proteome is to establish the multifunctional role of these protein families and comparative studies on saliva of other mammals can provide precious information to accomplish this aim [3]. Indeed the study of the animal kingdom always offers interesting examples of how evolution has generated different structures in order to fulfill similar purposes. For this reason, some years ago our research group started comparative proteomic

Correspondence: Professor Massimo Castagnola, Istituto di Biochimica e di Biochimica Clinica, Facoltà di Medicina, Università Cattolica, Largo Francesco Vito 1, 00168 Roma, Italy
E-mail: massimo.castagnola@icrm.cnr.it or mcastagnola@rm.unicatt.it
Fax: +39-06-3053598 or +39-06-3057612

Abbreviations: **a.a.**, amino acid; **GRPs**, glutamine-rich proteins; **PRPs**, proline-rich proteins; **TFA**, trifluoroacetic acid; **TG-2**, transglutaminase 2; **XIC**, extracted ion current

This article is dedicated to Frantisek Svec on the occasion of his 70th birthday.
Colour Online: See the article online to view Fig. 7 in colour.

studies of the proteins present in whole saliva of some mammals (rat, pig) in order to establish similarities and differences with respect to human saliva and to gather valuable information for better understanding the functions of different families of salivary proteins [8, 9].

To reach these purposes the approach was always focused on a top-down RP-HPLC–ESI-MS platform, often integrated by bottom-up experiments [10]. Top-down proteomics platforms allow the intact proteome to be investigated and further make it possible to elucidate the structure of slightly different isoforms of the same proteins [11]. Recently, attracted by the advantages of this strategy, other groups have also utilized top-down platforms to the study of human saliva [12–15]. In this study, we describe how the application of an integrated top-down/bottom-up MS-based platform to rat submandibular saliva allowed us to determine the entire sequence and some properties of an until now unknown glutamine-rich protein and of its derivatives, that we propose to term, for its high glutamine content, rat gliadoralin A 1–90. Gliadoralin A is a new member of the family of glutamic acid/glutamine-rich secretory proteins from rat submandibular glands [16–20]. We also predicted an atomic model for protein native structure, which might be used as reference to understand its functional properties.

2 Materials and methods

2.1 Reagents and apparatus

All common chemicals and reagents were of analytical grade and were purchased from Merck (Darmstadt, Germany), Sigma Aldrich (St. Louis, MO, USA), and Pierce Biotechnology (Rockford, IL, USA). Low-resolution HPLC–ESI-MS measurements were carried out by a Surveyor HPLC system (ThermoFisher, San Jose, CA, USA) connected by a T splitter to a photodiode-array detector and an LCQ Advantage mass spectrometer (ThermoFisher). The chromatographic column was a Vydac (Hesperia, CA, USA) C8 with 5 μm particle diameter (column dimensions 150 × 2.1 mm). High-resolution HPLC–ESI-MS/MS experiments were carried out by an Ultimate 3000 Micro HPLC apparatus (Dionex, Sunnyvale, CA, USA) equipped with a FLM-3000-Flow manager module and coupled to an LTQ Orbitrap XL apparatus (ThermoFisher). A Zorbax 300 SB-C8 (3.5 μm particle diameter; column dimension 1 mm id × 15 cm) was used as chromatographic column.

2.2 Ethical approval

The animal experiments were performed in accordance with the "Principles of laboratory animal care (NIH publication No. 85-23, revised 1985)" and were approved by the Ethics Committee for Animal Experiments, Göteborg, Sweden. At the end of the experiments, the rats, under pentobarbitone anesthesia, were killed by an intravenous overdose of pentobarbitone.

2.3 Saliva collection

Six female and five male Sprague–Dawley rats (Charles River, Sulzfeld, Germany) were used, weighing (mean ± SEM) 301 ± 7 and 395 ± 6 g, respectively. The animals were maintained on a standard-pelleted diet. They were anaesthetized with pentobarbitone (55 mg/kg intraperitoneal). The body temperature, measured with a rectal probe, was maintained at 38°C by means of a thermostatically controlled blanket. The animals were fitted with a femoral venous polyethylene catheter to serve as conduit for drug administration, and a tracheal cannula. The submandibular ducts were reached externally and exposed under the mylohyoid muscle. A fine polyethylene tube, filled with distilled water, was inserted into each submandibular duct and secured with ligatures. Isoprenaline hydrochloride (Sigma Chemicals, St. Louis, MO, USA) was infused intravenously (20 μg/kg body-weight per minute). The first drop of saliva falling from the tip of the duct-cannula was discarded. Thereafter, saliva from the two submandibular glands was collected in preweighed ice-chilled Eppendorf tubes filled with 250 μL of 0.2% of trifluoroacetic acid (TFA), over two 30 min periods of isoprenaline infusion (0–30 min and 30–60 min). Each tube was immediately weighed and, if necessary, additional TFA was added so as to obtain equal amounts of saliva and TFA. Saliva and TFA were vigorously mixed and centrifugated at 8000 *g* at 4°C for 5 min. The acid supernatant was then removed, frozen, and stored at −80°C until lyophilized and then, transported to the laboratories in Cagliari and Rome, Italy, for the analysis. The saliva secreted was expressed in microliters, assuming the density of saliva to be 1.0 g/mL. The submandibular glands were removed, pressed gently between gauze pads and weighed.

2.4 RP-HPLC–ESI-MS analysis

The following solutions were utilized in low-resolution ESI-MS chromatographic separations: (eluent A) 0.056% aqueous TFA and (eluent B) 0.050% TFA in acetonitrile/water 80/20 v/v. The applied gradient was linear from 0 to 54% in 39 min (linear) and from 54 to 100% in 10 min (linear), at a flow rate of 0.30 mL/min. The T splitter permitted 0.20 mL/min to flow toward the diode array detector and 0.10 mL/min toward the ESI source. The photodiode array detector was set at a wavelength of 214 and 276 nm. During the first 5 min of separation eluate was not addressed to the mass spectrometer to avoid source contamination and instrument damage due to the high salt concentration. Mass spectra were collected every 3 ms in the positive ion mode. MS spray voltage was 5.0 kV and capillary temperature was 255°C.

High-resolution HPLC–ESI-MS/MS experiments were performed by using the following eluents: (A) 0.1% v/v aqueous formic acid and (B) 0.1% v/v formic acid in acetonitrile. The applied gradient was 0–4 min 5% B, 4–40 min from 5 to 55% B (linear), 40–50 min from 55 to 100% B (linear), at a flow rate of 80 μL/min. High-resolution positive MS/MS spectra were collected in full scan using the lock mass for internal

mass calibration (polydimethyl cyclosiloxane, 445.1200 *m/z*) with the resolution of 60 000 and 30 000, respectively, and *m/z* range from 350 to 2000. In data-dependent acquisition mode the three most intense multiply charged ions were selected and fragmented by using collision-induced dissociation (35% normalized collision energy) and spectra were recorded. Alternatively, fragmentation was carried out using the same conditions on selected multiply charged ions corresponding to specific protein masses. Tuning parameters were: capillary temperature 250°C, source voltage 4.0 kV, capillary voltage 36 V, tube lens voltage 150 V.

2.5 Preparative RP-HPLC protein purification

Semipreparative RP-HPLC was utilized in order to partially purify the proteins detected in rat submandibular saliva. Acidic solution from rat submandibular saliva (200 μL or more, when available) were purified on by a Beckman System Gold 125S HPLC system (Beckman, Palo Alto, CA, USA) equipped with a PDA detector settled at 214 and 280 nm. The column was a Vydac C8 (column dimensions 250 × 10 mm, 5 μm). Eluents were those utilized for the low-resolution HPLC–ESI-MS separations. The gradient was from 0 to 54% of B in 39 min (linear) and from 54 to 100% B in 5 min (linear) and the flow rate was 2.8 mL/min. Fractions were collected in concomitance with the peak exit. The content of any purified fraction was checked by using HPLC–ESI-MS procedures described in the previous sections. Fractions containing the protein(s) of interest were pooled and used for enzymatic treatments as described in the following sections without further purification.

2.6 Reaction with transglutaminase 2

Different quantities of gliadoralin A 1–90 preparations were incubated in 50 mM Tris-HCl, 10 mM $CaCl_2$, 1 mM dithiothreitol, and 1 mM EDTA buffer, pH 7.5, with 0.02 EC units/mL of guinea pig liver transglutaminase-2 (TG-2) at 37°C. One EC unit corresponded to the formation of 1.0 μmole of hydroxamate per minute from Na-CBZ-glutaminylglycine and hydroxylamine at pH 6.0 at 37°C. Reaction was stopped by addition of 0.1 M EDTA (final concentration 33 mM), usually after 5 h of incubation. The reaction mixtures were centrifuged and the soluble fraction and the precipitate were separated. Two aliquots of the soluble fraction were directly analyzed by both low and high resolution HPLC–ESI-MS. The soluble fraction and the precipitate obtained after TG-2 incubation were submitted to trypsin digestion.

2.7 Trypsin digestion

Fifty micrograms of the freeze-dried powder of the purified protein, protein mixture, or reaction products after treatment with TG-2 were dissolved in 50 μL of ultrapure H_2O and digested with 0.5 μg of Trypsin by using the kit "Trypsin Singles Proteomic Grade" (Sigma-Aldrich) according to the manufacturer instructions. The reaction was stopped with formic acid (0.1% final concentration) after an overnight incubation, the samples were lyophilized, dissolved in 40 μL of 0.1% aqueous formic acid and submitted to high-resolution HPLC–ESI-MS/MS analysis.

2.8 HPLC–ESI-MS and MS-MS data analysis

Deconvolution of averaged ESI mass spectra was automatically performed by MagTran 1.0 software [21]. The relative abundances of proteins and derivatives were determined by measuring the extracted ion current (XIC) peak area and relating it to 1.0 mL of whole saliva. Under identical experimental conditions this value is linearly proportional to peptide concentration and it can be used with confidence to monitor relative abundances [22]. In the determination of XIC peak area a correct choice of the *m/z* values for protein detection is necessary, avoiding *m/z* potentially overlapping with ESI spectra of other close-eluting proteins in crowded chromatographic elution ranges (see Results). The window for all the *m/z* values chosen was in a range of ±0.5 *m/z*. The percentage error of the measurements was less than 10%. Data obtained by HPLC–ESI-MS/MS LTQ Orbitrap XL apparatus were manually investigated after extraction by the software resident in Xcalibur. The trypsin digest of the intact protein was analyzed by the Proteome Discoverer 1.3 program, based on SEQUEST cluster as search engine (University of Washington, USA, licensed to Thermo Electron Corp., San Jose, CA, USA) against Swiss-Prot *Rattus Norvegicus* proteome (UniprotKB/Swiss-Prot proteine knowledgebase release 2013_01 of 09-gen-13 containing 538849 sequence entries, taxonomical restriction: *Rattus Norvegicus*, 7839 sequence entries). For peptide matching the following limits were used: Xcorr scores greater than 1.5 for singly charged peptide ions and 2.0 and 2.5 for doubly and triply charged ions, respectively, one missed cleavage site. Different searches were carried out allowing the recognition of various post-translational modifications. Precursor mass search tolerance was set to 10 ppm and fragment mass tolerance was set to 1.5 Da.

2.9 *In silico* structural analysis

Secondary structure prediction was done using JPRED [23], SOPMA [24], PSIPRED [25], and PORTER [26]. Analysis of conserved domains was performed using conserved domains database, a resource on NCBI (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml). Tertiary structure models were performed using the I-TASSER approach that combines the methods of threading, *ab initio* modeling, and structural refinement to generate a protein structure [27]. The accuracy of the I-TASSER models is assessed by the C-score, which is a combination of the significance score of threading templates and the structural convergence of the

I-TASSER simulations. Typically, a good predicted model is obtained when the C-score is between −5 and 2. Model 1 had the highest score (C-score = −3.27). As a comparative measure, we also used ROBETTA modeling server, that uses an *ab initio* prediction method to construct model structures of a protein sequence if no related structure exists [28]. The models were validated by PROCHECK [29], VERIFY3D [30], and ProSA-Web (Protein Structure Analysis) available at https://prosa.services.came.sbg.ac.at/prosa.php. Visualization and molecular graphics were done using DiscoveryStudio v 3.5 (Accelrys Inc.) on a HP Z800 workstation.

3 Results

3.1 Determination of the sequence of the mature protein by an integrated solution-based top-down bottom-up HPLC–ESI-MS platform

In the high-resolution HPLC–ESI-MS profile of the acidic soluble fraction of rat submandibular saliva a protein with an experimental $[M+H]^{1+}$ (monoisotopic) = 10 544.24 *m/z* was detected in the elution range 17.5 ± 0.7 min (Fig. 1). Manual inspection of the deconvoluted MS/MS fragmentation spectrum performed on the $[M+10H]^{10+}$ ion (1055.83 *m/z*) allowed us to determine an internal protein sequence rich in glutamine residues as reported in Table 1. The search (BLAST program) using the inverse sequence with an *L* residue matched perfectly with an internal sequence (51–67 residues) of the *Rattus Norvegicus* unknown protein-coded D3Z9M3 in the Swiss-Prot data bank, whose sequence was predicted only on the basis of its DNA sequence (Fig. 2). The theoretical $[M+H]^{1+}$ of the whole D3Z9M3 protein was 12 540.46 *m/z* with a Δmass = 1996.22 *m/z* from the experimental $[M+H]^{1+}$ observed in the TIC profile. The *b* ions of the MS/MS spectrum were in agreement with the theoretical fragmentation (Table 1) only after the removal of the first 16 a.a. residues corresponding to the signal peptide (MLVILLMVVVLALSSA) and the presence of a N-terminal pyroglutamic moiety, both common modifications of salivary proteins [5]. However, the theoretical $[M+H]^{1+}$ of the whole protein was 10 870.44 *m/z*, with a difference of 326.20 *m/z* from the experimental one, suggesting also C-terminal modifications. The perfect agreement between experimental and theoretical values of *y* series and of the $[M+H]^{1+}$ of the whole protein was obtained after removal of the last three a.a. residues (RAV) from the C-terminus (Table 2) providing a mature protein of 90 a.a. residues (Fig. 2). Manual inspection of MS/MS fragmentation spectra carried out on other multiply-charged ions ($[M+9H]^{9+}$ = 1173.15; $[M+11H]^{11+}$ = 959.94; $[M+12H]^{12+}$ = 880.03; $[M+13H]^{13+}$ = 812.41 *m/z*) allowed to confirm the sequence with other sporadic fragments of the *b* and *y* series from the 3rd to the 72nd residues (Table 3). On the basis of the high glutamine content and of its origin we propose to term this protein rat gliadoralin A 1–90.

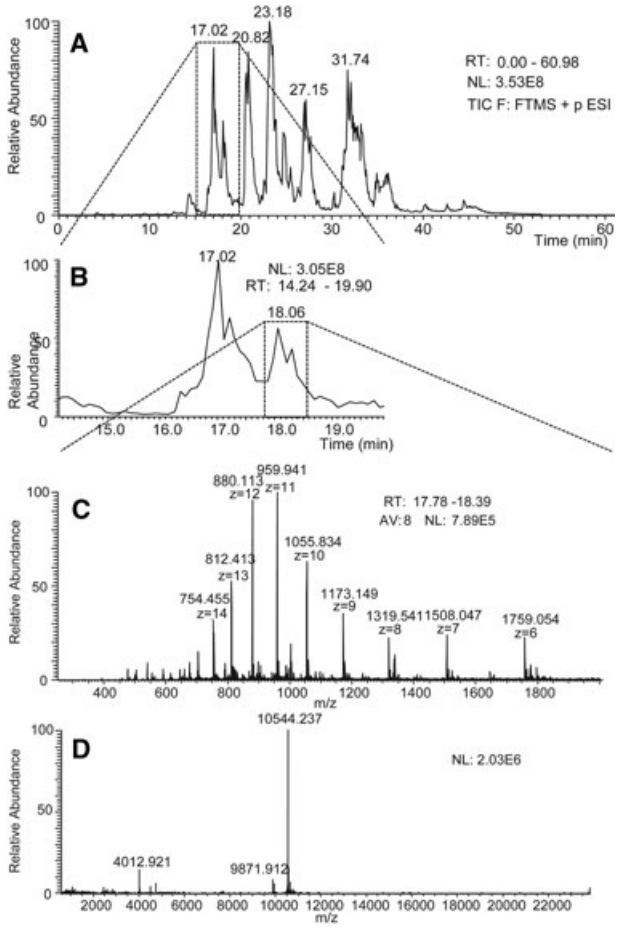

Figure 1. High-resolution RP-HPLC–ESI-MS identification of gliadoralin A 1–90 (and 1–85) in submandibular saliva of rat. (Panel A) Total ion current (TIC) profile. (Panel B) TIC profile in the 14.24–19.90 min elution range. (Panel C) High resolution ESI spectrum in the 17.78–18.39 min elution range. (Panel D) Deconvolution of the ESI spectrum of panel C, where the $[M+H]^{1+}$ of gliadoralin A 1–90 (10 544.24 *m/z*) and 1–85 (9871.91 *m/z*) were recognized. RT: retention time; NL: normalization level; AV: average.

In order to confirm the sequence of the mature protein, it was partially purified by preparative RP-HPLC. Other protein masses detected in the purified fraction were identified as different isoforms of the main component (see the following sections). After trypsin digestion the mixture was analyzed by high-resolution RP-HPLC–ESI-MS/MS in data-dependent analysis. The results ensuring sequence coverage of 100% are reported in Table 3. Even though the DNA translated sequence was of great help for the determination of the primary structure of the mature protein, the results reported in Table 3 show that careful analysis of MS/MS fragmentation of the intact protein and of its trypsin fragments made it possible an almost complete *de novo* sequencing. Several tryptic fragments (Fr. 16–36 and 61–67) with glutamine residue on the N-terminus were detectable also with a cyclic pyroglutamic moiety at the C-terminus. Some fragments (Fr. 20–36, 21–36, 37–55) derived from non specific trypsin cleavages (after

Table 1. Sequence detected and post-translational modifications necessary to ensure matching between theoretical and experimental MS/MS data of the *b* series

	Deconvoluted $[M+H]^{1+}$ exp *b* series	Theor. $[M+H]^{1+}$	Theor. $[M+H]^{1+}$	Theor. $[M+H]^{1+}$
a.a. residue (one letter code)	Intact	Intact from D3Z9M3	Signal peptide removal (−1652.99)	*N*-terminal Pyrogln res (−17.02)
	10 544.24	12 540.46	10 887.47	10 870.44
	5806.71	7476.73	5823.73	5806.71
Q	5678.65	7348.67	5695.67	5678.65
Q	5550.59	7220.61	5567.61	5550.59
E	5421.55	7091.57	5438.57	5421.55
Q	5293.49	6963.51	5310.51	5293.49
S	5206.46	6876.48	5223.48	5206.46
Q	5078.40	6748.42	5095.42	5078.40
Q	4950.34	6620.36	4967.36	4950.34
L/I	4837.25	6507.27	4854.28	4837.25
S	4750.22	6420.24	4767.25	4750.22
Q	4622.16	6292.18	4639.19	4622.16
Q	4494.11	6164.13	4511.13	4494.11
Q	4366.05	6036.07	4383.07	4366.05
Q	4237.98	5908.01	4255.01	4237.98
Q	4109.93	5779.95	4126.95	4109.92
		Δ mass from exp. 1670.01 ± 0.2	Δ mass from exp. 17.02 ± 0.2	

P_{19}, S_{20}, and Q_{55}, respectively), but they were detected in low amounts with respect to the canonical fragments.

3.2 Determination of the isoforms of rat gliadoralin A

As described, above coincident with the chromatographic peak of rat gliadoralin A 1–90 minor relative amounts of other proteins with $[M+H]^{1+}$ 10870.44, 10700.33, 9950.98, 9871.87, and 9278.75 *m/z* (± 0.04 *m/z*) were sporadically detected. Analysis of the MS/MS fragmentation (not reported) of the intact isoforms performed by selected ion monitoring experiments on different multiply charged ions allowed us to establish that the protein with the higher mass is the gliadoralin A isoform with three further residues (RAV) on the C-terminus (Theor. $[M+H]^{1+}$ = 10870.44 *m/z*, gliadoralin A 1–93), the protein with $[M+H]^{1+}$ = 10 700.33 *m/z* is the isoform with a further *R* on the C-terminus (Theor. $[M+H]^{1+}$ = 10 700.33 *m/z*, gliadoralin A 1–91), the isoform with $[M+H]^{1+}$ = 9871.87 *m/z* is the isoform missing five a.a. residues (YQQPR) at the C-terminus (Theor. $[M+H]^{1+}$ = 9871.90 *m/z*, gliadoralin A 1–85), the isoform with $[M+H]^{1+}$ = 9950.98 *m/z* corresponds to gliadoralin A 1–90 missing the first five a.a residues (pyro-QDPNR) at the N-terminus (Theor. $[M+H]^{1+}$ = 9950.98 *m/z*, gliadoralin A 6–90) and finally the isoform with $[M+H]^{1+}$ = 9278.75 *m/z* corresponds to gliadoralin A 1–90 missing the 5 a.a residues at both the N- and C-termini (Theor. $[M+H]^{1+}$ = 9278.74 *m/z*, gliadoralin A 6–85).

3.3 Determination of the secretion of the gliadoralin A isoforms under isoprenaline stimulation

In the absence of the secretagogue, there was no on-going secretion of saliva from the duct-cannulated submandibular glands. In response to the continuous intravenous infusion of isoprenaline (20 μg/kg body-weight per minute) for 60 min, the mean volume of saliva secreted from the two glands was the same or almost the same in female and male rats, despite the fact that the glands of the female rats (mean ± SEM, 196 ± 8 mg, $n = 12$) were much lighter ($P < 0.001$, Student's *t*-test for unpaired values) than those of the male rats (294 ± 13 mg, $n = 10$). The volume secreted from each pair of glands during the first 30 min period of observation was 176 ± 12 μL in the female rats and 179 ± 20 μL in the male rats. The volume secreted during the following 30 min period was less ($P < 0.001$, Student's *t*-test for paired values). It was 131 ± 10 μL in the six female rats and 125 ± 21 μL in the five male rats.

A ***(Sequence from DNA)***

```
1         10         20         30         40         50         60
MLVILLMVVV LALSSAQDPN RDFVVSSQDV RERQPSSQQG TVGGQSQESQ LRDQQQQQSL

          70         80         90        100
QQSQEQQPQP QLQQTKQPQR PVKGQPLPQQ QQQQNQRPRP RYQQPRRAV
```

B ***(Primary structure of the mature protein)***

```
 1        10         20         30         40         50         60
<QDPNRDFVVS SQDVRERQPS SQQGTVGGQS QESQLRDQQQ QQSLQQSQEQ QPQPQLQQTK

         70         80         90
 QPQRPVKGQP LPQQQQQQNQ RPRPRYQQPR

<Q : pyro-glutamine (pyrrolidone carboxylic acid)
```

Figure 2. Primary sequence of the unknown protein coded D3Z9M3 in the Swiss-Prot data bank DNA-derived and primary sequence of the mature protein (gliadoralin A 1–90) experimentally characterized by RP-HPLC–ESI-MS and MS/MS data in rat submandibular saliva.

Table 2. Post-translational modifications necessary to ensure matching between theoretical and experimental MS/MS data of the *y* series and to the $[M+H]^{1+}$ of the intact protein

	Deconvoluted $[M+H]^{1+}$ exp *y* series	Theor. $[M+H]^{1+}$	Theor. $[M+H]^{1+}$
	Intact	Signal peptide removal and *N*-term pyroglu (Table 1)	Removal *C*-term RAV (−326.20)
	10 544.24	10 870.44	10 544.24
	5795.02	6121.23	5795.02
S	5707.99	6034.20	5707.99
L	5594.91	5921.11	5594.90
Q	5466.85	5793.05	5466.85
Q	5338.79	5664.99	5338.79
S	5251.76	5577.96	5251.76
Q	5123.70	5449.90	5123.70
E	4994.64	5320.86	4994.65
Q	4866.60	5192.80	4866.60
Q	4738.54	5064.74	4738.54
		Δ mass from exp. 326.20 ± 0.2	

The relative amounts of the different gliadoralin A isoforms were evaluated in the two sample series (0–30 min and 30–60 min) by RP-HPLC–ESI-MS analysis applying an XIC procedure. Five *m/z* values were selected for each isoform, from the $[M+7H]^{7+}$ to the $[M+11H]^{11+}$ multiply charged ions for gliadoralin A 1–85, 6–85, and 6–90 isoform and from the $[M+8H]^{8+}$ to the $[M+12H]^{12+}$ multiply charged ions for the gliadoralin A 1–90, 1–91, and 1–93 isoforms. The XIC peak area was related to the volume of saliva collected and reported as arbitrary units for mL of saliva. The mean values obtained analyzing the secretion of 11 rats (six females and five males) are reported in Fig. 3. Figure shows that the mean percentage of gliadoralin A 1–90 accounted for about 70%, while the other isoforms for about 5–10% each and that the percentages did not changed significantly upon prolonged isoprenaline stimulation. It is relevant to outline that the values reported in the figure are mean values, while the different minor isoforms were sporadically detectable in the different samples. They were never all simultaneously detected in the same sample. The unique isoform always detectable in high percentage in all samples was gliadoralin A 1–90. Even though gender-related differences were observed for some rat salivary protein [16], no gender-significant difference in the secretion of the gliadoralin A isoforms was observed.

3.4 Determination of the amino acid residues involved in the cross-linking by transglutaminase-2

A purified preparation of gliadoralin A 1–90 containing minor amounts of its 1–93 and 1–85 isoforms were submitted to the action of TG-2 following the procedure described in the Experimental section. The reaction product was divided in two fractions: the soluble fraction and a precipitate (as the major product ≈ 80%) that were separately analyzed.

HPLC–ESI-MS and MS/MS analysis of the soluble fraction showed two main peaks (Fig. 4). In the first eluting peak (18.02 min) the unreacted gliadoralin A 1–90, its isoform, 1–85 and the fr. 6–67 of gliadoralin ($[M+H]^{1+}$ exp. = 7113.55 *m/z*, theor. = 7113.51 *m/z*) were detected as major components (attribution confirmed by MS/MS data). In the second eluting peak (18.51 min) the fr. 1–67 ($[M+H]^{1+}$ exp. = 7706.81 *m/z*, theor. = 7706.77 *m/z*) as major component was detected (attribution confirmed by MS/MS data), but a third minor but very large peak was also evidenced, containing minor amounts of gliadoralin cyclic derivatives generated by TG-2, one with a $[M+H]^{1+}$ exp. = 10 527.23 *m/z* (10 544.25 − 10 527.23 = 17.02) corresponding to the internally cross-linked cyclogliadoralin A 1–90, the other with a $[M+H]^{1+}$ exp. = 9854.91 *m/z* (9871.94 − 9854.91 = 17.03) corresponding to the internally cross-linked cyclogliadoralin A 1–85 and finally one with $[M+H]^{1+}$ exp. = 7689.77 *m/z* (7706.81 − 7689.77 = 17.03) corresponding to the internally cross-linked cyclogliadoralin A 1–67 (Fig. 4). As for the naturally occurring protein, the high resolution RP-HPLC–ESI-MS/MS carried out in data depending analysis provided satisfactory MS/MS for the intact different cyclo-derivatives. It should be considered that in the presence of isobaric heterogeneous cyclo-derivatives, as suggested from the large chromatographic peak (Fig. 4), the interpretation of MS/MS data is puzzling for superimposed fragmentations of different structures. Nonetheless, the MS/MS carried out on the $[M+8H]^{8+}$ ion (1233.37 *m/z*) of the cyclogliadoralin A 1–85 provided interesting results. Gliadoralin A has only two lysine residues: K_{60} and K_{67}. Experimental MS/MS data fit well with the hypothesis of the involvement of K_{60} in the internal cross-link. Fragment y_{24} ($[M+H]^{1+}$ 2889.58_{exp}; 2889.58_{theor} *m/z*) was indeed determinant for the exclusion of K_{67} (Fig. 5A). For the identification of glutamine residue(s) involved in the cross-link, various fragments of the *y* and *b* series restricted the choice to Q_{53}, Q_{55}, Q_{57}, Q_{58}, and Q_{61}. Among these, the presence of the cross-linked MS/MS internal fragments 57–67 ($[M+H]^{1+}$ 1302.66_{exp}; 1302.73_{theor} *m/z*) and 47–60 ($[M+H]^{1+}$ 1632.83_{exp}; 1632.80_{theor} *m/z*) suggested the involvement of one of the two Q_{57} and Q_{58} residues (Fig. 5A).

In the RP-HPLC–MS and MS/MS analysis of the trypsin digest of the proteins present in the chromatographic peak containing the cyclo-derivatives it was possible to recognize a peptide eluting at 15.84 min with $[M+H]^{1+}$ = 3695.86 *m/z* as cross-linked 37–67 peptide (Theor. $[M+H]^{1+}$ = 3695.85 *m/z*) confirming the involvement of K_{60} (Fig. 5B). Different fragments of the *b* and *y* series detected on the MS/MS spectrum ($[M+4H]^{4+}$ ion = 925.22 *m/z*) were excluded the involvement of some glutamine residues, particularly the fragment b_{24} ($[M+H]^{1+}$ 2844.35_{exp}; 2844.35_{theor} *m/z*) which excluded the Q_{61} residue. The detection of the internal fragment corresponding to the cyclo $Q_{57}QTKQPQRP_{65}$ ($[M+H]^{1+}$ 1075.56_{exp}; 1075.56_{theor} *m/z*) confirmed the hypothesis of the

Table 3. MS/MS principal data for the characterization of the sequence of gliadoralin A 1–90

	Seq	Intact Protein	*b* and *y* fragments detectable in the trypsin digest by manual inspection					
1	pyroQ		—	—				
2	D	—	—	—				
3	P	*y*	*y*	*y*				
4	N	—	—	—				
5	R	—	—	*b,y*				
6	D	—	612.27	*b,y*	—			
7	F	*b*		*b,y*	*b,y*			
8	V	—		*b,y*	*b,y*			
9	V	—		*b,y*	*b,y*			
10	S	—		*b,y*	*b,y*			
11	S	—		*b,y*	*b,y*			
12	Q	—		*b*	*b,y*			
13	D	*b*		*b*	*b,y*			
14	V	—		*b,y*	*b,y*			
15	R	—		—	—			
16	E	—	—	1744.83	1151.57			
17	R	—	*b,y*					
18	Q	—	*b,y*	—				
19	P	—	*y*	*y (b,y)*				
20	S	—	*b,y*	*y (b,y)*	—			
21	S	—	*b,y*	*y (b,y)*	—	—		
22	Q	*b*	*b,y*	*b,y*	*y*	—		
23	Q	*b*	*b,y*	*y (b,y)*	*y*	*y*		
24	G	*b,y*	*b,y*	*b,y (b,y)*	*y*	*y*		
25	T	*b,y*	*b,y*	*b,y (b,y)*	*b,y*	*y*		
26	V	*b,y*	*b,y*	*b,y (b,y)*	*b,y*	*b,y*		
27	G	*b,y*	*b,y*	*b,y (b,y)*	*b,y*	*y*		
28	G	*b,y*	*b,y*	*b,y (b,y)*	*y*	*y*		
29	Q	*b,y*	*b,y*	*b,y (b,y)*	*b,y*	*y*		
30	S	*b,y*	*b,y*	*b,y (b,y)*	*b,y*	*b*		
31	Q	*b*	*b,y*	*b,y (b,y)*	*b,y*	*b,y*		
32	E	*b,y*	*b,y*	*b,y (b,y)*	*b,y*	*y*		
33	S	*y*	*y*	*b,y (b,y)*	*b,y*	*b,y*		
34	Q	*b*	*b,y*	*b,y (b,y)*	*b*	*b*		
35	L	*b,y*	*b,y*	*b,y (b,y)*	*b,y*	*b*		
36	R	*b,y*	—	—	—	—		
37	D	*b,y*	2287.11	2001.96	1776.85	1689.82	—	—	—
38	Q	*b*		(1984.94)[a)]			*b*	*Y*	*y*
39	Q	*b*					—	*b,y*	*y*
40	Q	*b,y*					*b*	*b,y*	*y*
41	Q	*b,y*					*b,y*	*b,y*	*y*
42	Q	*b,y*					*y*	*b,y*	*y*
43	S	*b,y*					—	*b,y*	*y*
44	L	*b,y*					*b,y*	*b,y*	*b,y*
45	Q	*b,y*					*b,y*	*b,y*	*b,y*
46	Q	*b,y*					*b,y*	*b,y*	*b,y*
47	S	*b,y*					*b,y*	*b,y*	*y*
48	Q	*b,y*					*b,y*	*b,y*	*y*
49	E	*b,y*					*b,y*	*b,y*	*b,y*
50	Q	*b,y*					*b,y*	*b,y*	*y*
51	Q	*b,y*					*b,y*	*b,y*	*b,y*
52	P	*y*					*y*	*b,y*	*y*
53	Q	*b*					*b*	*b,y*	—
54	P	*y*					*y*	*y*	*y*
55	Q	—					—	*b,y*	—
56	L	*y*					2281.05	*b,y*	—

Table 3. Continued

	Seq	Intact Protein	*b* and *y* fragments detectable in the trypsin digest by manual inspection					
57	Q	*y*					*b,y*	*y*
58	Q	—					*b,y*	*y*
59	T	—					*b*	*y*
60	K	—					—	*y*
61	Q	*y*	—				2879.40	*b,y*
62	P	—	*y (b,y)*					—
63	Q	—	*y (b,y)*					—
64	R	—	*(b,y)*					*b*
65	P	—	*y (b,y)*					—
66	V	—	*b (b,y)*					—
67	K	—	(*y*)					—
68	G	—	852.51	—	—	—		3712.89
69	Q	—	(835.48)	*y*	—	—		
70	P	—		*y*	*b,y*	*y*		
71	L	—		*b,y*	*b,y*	*b,y*		
72	P	*y*		*y*	*y*	*y*		
73	Q	—		*b,y*	*y*	—		
74	Q	—		*b,y*	*y*	*y*		
75	Q	—		*b,y*	*y*	*y*		
76	Q	—		*b,y*	*y*	*y*		
77	Q	—		*b,y*	—	—		
78	Q	—		*y*	—	—		
79	N	—		*b,y*	—	—		
80	Q	—		*b,y*	—	*y*		
81	R	—		—	—	—		
82	P	—		1677.85	—	—		
83	R	—			—	*b*		
84	P	—			1931.00	—		
85	R	—				—		
86	Y	—	—			2184.15		
87	Q	—	*b*					
88	Q	—	*b,y*					
89	P	—	*y*					
90	R	—	—					
		10 544.24	691.35					

a)In parenthesis the mass and the fragments of tryptic peptides detected with *N*-terminal pyro-Glu moiety.

involvement of one of the two Q_{57}, Q_{58} residues (Fig. 5B). Moreover, some peptide masses found in the trypsin digest suggested that other Q residues were involved in the internal cross-link: (a) a peptide with $[M+H]^{1+} = 5696.82$ *m/z* that could correspond to the cross-link between the tryptic peptides 37–67 and 18–36 (theor. $[M+H]^{1+} = 5696.80$ *m/z*); (b) a peptide with $[M+H]^{1+} = 5372.70$ *m/z* that could correspond to the cross-link between the tryptic peptides 37–67 and 68–81 (theor. $[M+H]^{1+} = 5372.68$ *m/z*); (c) traces of a peptide with $[M+H]^{1+} = 4386.21$ *m/z* that could correspond to the cross-link between the tryptic peptides 37–67 and 86–90 (theor. $[M+H]^{1+} = 4386.20$ *m/z*) (Fig. 6). Due to the small amount of these peptides the few MS/MS data obtained were very poor and not suitable for supporting any further structural interpretation.

The cross-linked peptide 61–85 (Theor. $[M+H]^{1+} = 3000.61$ *m/z*) or any other mass corresponding to this peptide cross-linked to other fragments was not detected in the trypsin digest, further excluding the involvement of K_{67}.

The HPLC–ESI-MS analysis of the trypsin digest of the cross-linked precipitate identified the same three masses detected in the trypsin digest of the cyclo-derivatives: $[M+H]^{1+} = 5696.81$ *m/z* (cross-link between peptides 37–67 and 18–36), $[M+H]^{1+} = 5372.70$ *m/z*, (cross-link between the peptides 37–67 and 68–81) and $[M+H]^{1+} = 4386.21$ *m/z* (cross-link between the peptides 37–67 and 86–90). Moreover, traces of a peptide with a $[M+H]^{1+} = 4846.43$ *m/z* that could correspond to the cross-link between the tryptic peptides 37–67 and 6–15 (theor. $[M+H]^{1+} = 4846.41$ *m/z*) were detected (Fig. 6). Also in this case the quality of MS/MS data did not allow us any further structural interpretation.

All these peptides were not detectable in the trypsin digest of the unreacted protein.

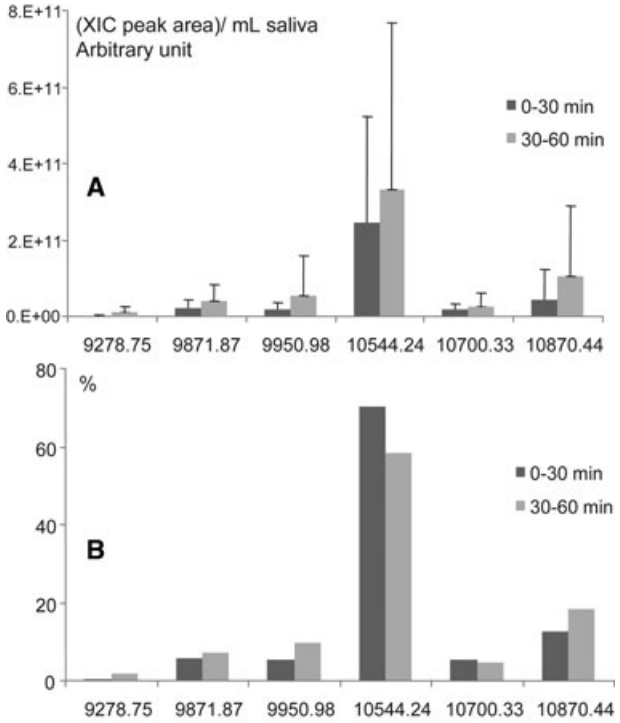

Figure 3. Mean of the values of (XIC) peak area (arbitrary unities) per milliliter (Panel A) and mean percentages (Panel B) of the different isoforms of gliadoralin A detected in rat submandibular saliva ($N = 11$; six females; five males), collected in the 0–30 (black), and 30–60 (gray) min period of isoprenaline stimulation. No significant differences were observed between the two periods of collection.

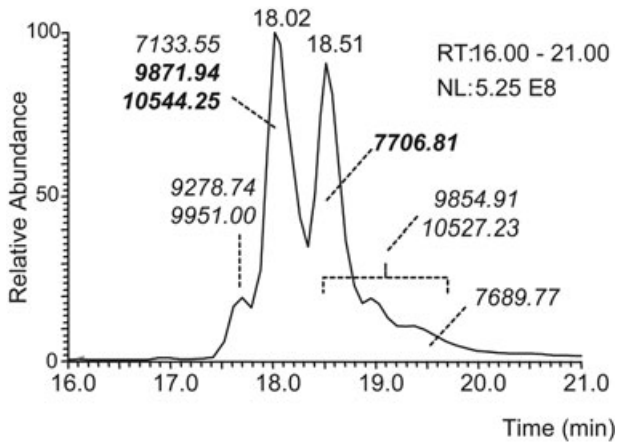

Figure 4. TIC profile of the soluble fraction of the reaction product obtained by the action of (guinea pig) TG-2 on partially purified gliadoralin A 1–90. The main $[M+H]^{1+}$ values detected in the TIC peaks (by deconvolution of the different ESI spectra) are reported.

On the whole, the global manual analysis of the MS and MS/MS data demonstrated that K_{60} is probably the unique lysine recognized as TG-2 substrate, while different glutamines along the entire sequence (but not all) are potential candidates.

3.5 *In silico* determination of secondary and tertiary structures

The purification of the protein for experimental structural studies was impossible for the presence of different isoforms in the same chromatographic peak. *In silico* prediction of 3D protein structures from amino acid sequences is a challenging task in the case of this protein, since, as mentioned above, its primary sequence is novel and it does not show any homology to known structures. We investigated the protein by computational methods, including secondary structure prediction analyses and domain searches. All secondary structure prediction methods resulted in similar secondary structure elements indicating that overall, only a small part of the protein is predicted to adopt an α-helical fold, whereas large segments, particularly of the *N*-terminal region, are proposed to be unfolded. The suggested helical parts converge into three regions from position 32 to 49, from 55 to 58, and from 73 to 78. The prediction of unfolded regions may be most likely due to the numerous glutamine (37.8%), proline (12.2%), serine (8.9%), glutamic acid (3.3%), and lysine (2.2%) residues. These residues often indicate intrinsically unstructured protein domains [29]. Accordingly, the protein is depleted in tryptophan (0.0%), tyrosine (1.1%), phenylalanine (1.1%), cysteine (0.0%), isoleucine (0.0%) compared with the average-folded proteins in the Protein Data Bank [31].

The protein has no conserved domains as identified by either NCBI or other publicly available search engines. Since tertiary structure prediction using template-based modeling proved difficult, results from I-TASSER were compared to those generated using ROBETTA, a *de novo* modeling approach.

The stereochemical quality of the models was assessed by PROCHECK. Ramchandran plots of both models showed that all the residues are located in the core (93.2%) and allowed (6.8%) regions denoting a stable structure. The reliability of the modeled fold was also checked with VERIFY-3D, which evaluates the compatibility of a given residue in a certain 3-D environment. The VERIFY-3D scores of our models are always positive, meaning that the conformation adopted by the residues in the models is compatible with the surrounding environment. Furthermore, the low ProSA *z*-score values obtained (−3.44 for the I-TASSER and −3.41 for the ROBETTA models, respectively) confirm the good quality of the predicted structures.

Interestingly, as shown in Fig. 7, although the predicted 3D models show some differences, the secondary structure largely overlaps, with α-helices present in approximately the same places as for the 2-D model. In particular the resulting I-TASSER and ROBETTA models feature a long α-helix (Q_{31}–Q_{50}) that agrees with the secondary structure prediction. Main differences with the 2-D model are that the β-sheet content is zero, and the fraction of α-helix is increased, especially in the I-TASSER prediction. In the figure the two lysine residues, relevant for TG-2 recognition are evidenced.

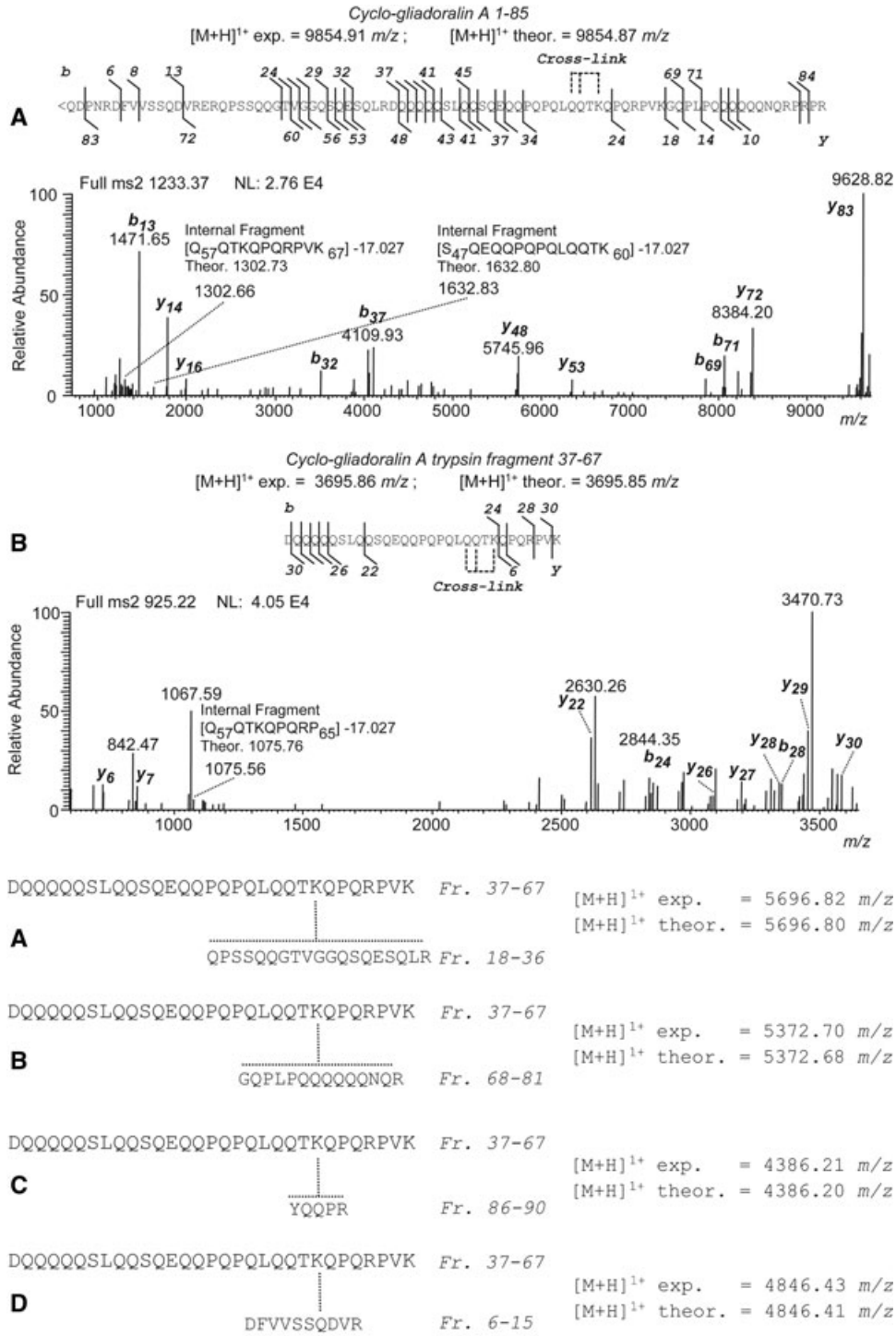

Figure 5. (Panel A) MS/MS fragmentation carried out on the $[M+8H]^{8+}$ ion (1233.37 *m/z*) of the cyclogliadoralin A 1–85. (Panel B) MS/MS fragmentation carried out on the $[M+4H]^{4+}$ ion (925.22 *m/z*) of trypsin fragment 37–67 of one of the cyclo-derivatives of gliadoralin A generated by TG-2. Both MS/MS spectra are in agreement with the cross-link between K_{60} and one of the two Q_{57} or Q_{58} residues.

Figure 6. Cross-links generated by the action of TG-2 on rat gliadoralin A. A, B, and C $[M+1H]^{1+}$ ions were detected in the trypsin digest of cyclogliadoralin A, while all the $[M+1H]^{1+}$ ions (from A to D) were detected in the trypsin digest of the precipitate collected after TG-2 reaction. All these peptides were not detectable in the trypsin digest of the unreacted protein. For the attribution of the possible glutamine residues involved in the cross-link see the Section 4.

4 Discussion

The results of this study at first demonstrated the power of an integrated high resolution HPLC–ESI-MS top-down bottom-up platform for the detailed description of primary structure of a relatively big protein, of its isoforms and its reaction products. Although the DNA sequence has been of great help to facilitate the determination of the primary structure of gliadoralin A 1–90, it is evident from Table 3 that the integrated MS/MS information obtained on the intact protein and on its trypsin digest is enough to solve the puzzle of the *de novo* sequence determination. The utilization of further digest maps by this way can surely complete even more complicated structures. However, a satisfactory, even partial, MS/MS of the

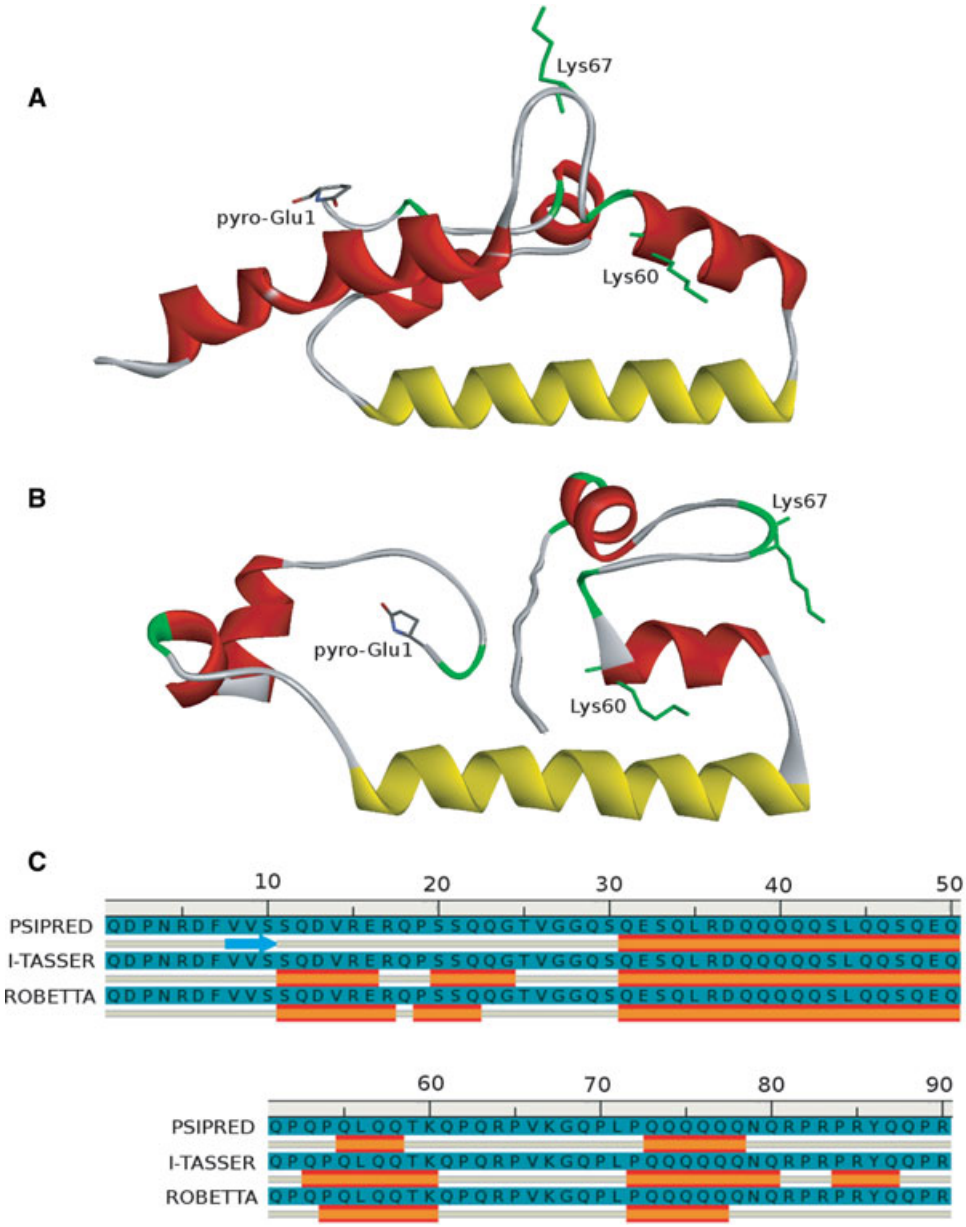

Figure 7. Structural modeling of gliadoralin A 1–90. Ribbon illustrations of 3D model generated by (A) I-TASSER and (B) ROBETTA. Ribbon corresponding to helix 31–50 is colored in yellow, the other helices in red; coils and turns are colored in white and green, respectively. K_{60} and K_{67} residues, potentially involved in enzyme recognition are shown in green color; Pyro-Glu_1 is colored by atom type. (C) Secondary structure prediction using PSIPRED.

intact protein (top-down approach [11]) is the pivotal information for the complete structural protein characterization, with a particular concern for the different isoforms.

Gliadoralin A should be included in the family of (glutamic acid/glutamine/) glx-rich proteins (GRPs) secreted by the rat submandibular gland [16]. This protein family was detected several years ago by cDNA analysis in rat submandibular glands by various groups [16–20]. It is characterized by the presence of conserved tandem-repeats of 23 residues (NQEPPATSGSEEEQQQQEPTQAQ) that are not detectable in gliadoralin A. Gliadoralin A does not share structural similarity with these and other known proteins, except a very low sharing (less than 30% similarity) with the human cornifin A, a protein devoted to the formation of the cross-linked envelope of keratinocytes [32].

The a.a. composition of rat glutamine-rich proteins is reminiscent of that found in most human PRP genes, strengthening the hypothesis of a possible ancestral origin common with other mammals [18]. Table 4 compares the amino acid composition of gliadoralin A 1–90 against that of the rat GRPs until now identified, human acidic (*PRP-2* allele) and basic (*PRB-2 L* allele) PRPs as well as pig PRPs. A general similarity is recognizable between rat and human, while the a.a. content of pig PRPs are somewhat different. Gliadoralin A shows by far the highest glutamine content, while the proline content is in line with that of the other rat GRPs, but significantly lower than that of human and pig PRPs. Moreover, while gliadoralin A and the other rat proteins are probably secreted as entire proteins (except for the sporadic presence of low percentages of other isoforms identified

Table 4. Amino acid percentages in several glutamine-glutamic acid-proline rich proteins of rat, human, and pig

	Gliadoralin A 1–90	Q9JLN8[a] Rat GRP	P08568 Rat GRP	P08462 RAT GRP	P02810[b] Human acidic PRP	P02812[c] Human Basic PRP	Q95JC9 Pig Basic PRP
Ala (A)	—	7.8	8.8	7.9	0.7	1.8	9.2
Arg (R)	10.0	4.3	3.1	3.5	4.0	3.0	5.0
Asn (N)	2.2	5.2	4.8	5.2	0.7	4.8	0.3
Asp (D)	4.4	3.4	4.4	3.5	6.7	1.0	0.8
Cys (C)	—	—	—	—	—	—	—
Gln (Q)	37.8	15.5	15.8	15.7	23.3	15.5	0.8
Glu (E)	3.3	19.0	19.3	19.2	3.3	1.0	1.1
Gly (G)	4.4	5.2	4.8	4.8	20.7	20.5	18.0
His (H)	—	0.4	0.4	0.4	2.0	—	0.2
Ile (I)	—	0.9	1.3	0.9	1.3	0.2	—
Leu (L)	4.4	—	2.6	0.4	2.0	0.5	0.5
Lys (K)	2.2	1.7	2.2	1.7	1.3	6.8	0.9
Met (M)	—	—	—	—	—	—	—
Phe (F)	1.1	2.6	0.9	2.6	0.7	—	0.5
Pro (P)	12.2	14.2	13.6	14.8	27.3	38.5	62.0
Ser (S)	8.9	9.5	10.1	9.2	4.0	6.2	0.9
Thr (T)	2.2	6.0	5.7	5.7	—	—	—
Trp (W)	—	—	—	—	—	—	—
Tyr (Y)	1.1	0.9	—	0.9	—	—	—
Val (V)	5.6	3.4	2.2	3.5	2.0	0.2	—

a) Signal peptide removed.
b) *PRP-2* allele.
c) *PRB2 L* allele.

in this study) human acidic and basic PRPs, and even more pig PRPs are submitted to a complex proprotein cleavage before granule storage under the action of different convertases. The cleavage is particularly relevant for human and pig basic PRPs, where in glandular saliva only smaller peptides deriving from the bigger proprotein are detectable, in humans with dimensions ranging from 90 to 50 a.a. residues, in pigs from 40 to 10 a.a residues [8, 33, 34]. A relevant difference between these proteins is that gliadoralin A, as well the other rat GRPs, are specific products of submandibular gland, human acidic PRPs are secreted by both major glands, preferentially by parotid, while basic PRPs are only secreted by parotid glands, both in humans and in pig [3, 5, 6]. The functional meaning of this regio-anatomical selectivity is amazing and puzzling.

For their high glutamine content rat and human proteins should be good substrate for transglutaminases. Effectively, gliadoralin A 1–90 is a satisfactory substrate for guinea pig TG-2. From the MS/MS data K_{60} (but not K_{67}) seems to be the unique lone-pair donor for the enzyme. The *in silico* structural modeling indicates that the two lysine residues are both well exposed to the solvent, but while K_{67} is included in a flexible loop and the side chain is oriented towards the exterior of the protein, K_{60} has the side chain oriented towards the main α-helix of the protein. The recognition of specific lysine residues seems to be governed only by steric availability, while the spacing and structure of neighboring residues seems to be a crucial factor for the recognition of the targeted glutamine residues. From our MS/MS data many different glutamine residues can be recognized by the enzyme. Q_{57} or Q_{58} are candidates for one of the intramolecular bond generating the cyclo-derivative. In general, proline residues seem to be relevant for the glutamine recognition. A glutamine residue is usually not recognized as a substrate if it occurs between two proline residues or if it has a +1 or +3 flanking proline residues. On the contrary, a glutamine with −1 or +2 flanking proline residues is good potential acceptor of lysine lone-pair, as well as two adjacent glutamine residues even in consecutive reactions [35–38]. According to these rules Q_{18}, Q_{51}, Q_{53}, Q_{61}, Q_{69}, and Q_{88} should not be potential substrates. Further than Q_{57} or Q_{58} residues, MS data from the tryptic digest of the cyclo-derivative and of the precipitate collected after TG-2 reaction, indicated a cross-link involving a Q between 18–36 residues (six Q residues, the most probable being Q_{22} or Q_{23}) and a cross-link involving a Q between 68–81 residues (eight Q residues, the most probable being Q_{83}). Minor recognition sites could be Q_{12}, unique Q residue in the peptide 6–15, and Q_{87} in the 86–90 peptide.

The susceptibility to TG-2 recognition led to the supposition that gliadoralin A can participate, perhaps together with the other GRPs rat submandibular proteins, to the formation of the mucosal protein pellicle, a protein network involving different secreted salivary protein, that is devoted to the reinforcement and protection of oral mucosal epithelial surface and only partially characterized in humans [39–41].

The presence of a signal peptide common to salivary secretory proteins of other mammals strongly suggests that

gliadoralin A is secreted from the granule of salivary (acinar) glands by the secretion pathway typical of salivary proteins such as human statherin or PRPs. However, in future studies it would be of interest to investigate whether gliadoralin is detectable in the bloodstream. Interestingly, a secretory granular localization of GRP of the acinar cells of the rat submandibular glands has been demonstrated [42]. Moreover, isoprenaline, known to induce exocytosis, induce the secretion of GRPs proteins from the rat submandibular gland [43]. Future studies will determine if gliadoralin is secreted further than by isoprenaline stimulation, upon the stimulation of the sympathetic innervations. The unchanged salivary gliadoralin A concentration over time (Fig. 3) concomitant with a decrease in the volume of saliva secreted, should be attributed to a desensitization effect of isoprenaline on the gland cells rather than to a depletion of the gliadoralin A gland content. The removal of the last three C-terminal residues (RAV), bringing to the gliadoralin A 1–90 isoform, is an event probably occurring before the secretion, while the detection of minor amounts of other isoforms suggests that the cleavages deriving from trypsin-like (and chimotrypsin-like) proteinases should occur after the secretion of the principal isoform.

The authors acknowledge the financial support of Cagliari University, Catholic University of Rome, MIUR, Italian National Research Council (CNR), Regione Sardegna and Nando Peretti Foundation, according to their programs of scientific diffusion.

The authors have declared no conflict of interest.

6 References

[1] Tabak, L. A., *J. Dent. Educ.* 2001, *65*, 1335–1339.

[2] Helmerhorst, E. J., Oppenheim, F. G., *J. Dent. Res.* 2007, *86*, 680–693.

[3] Castagnola, M., Cabras, T., Vitali, A., Sanna, M. T., Messana, I., *Trends Biotechnol.* 2011, *29*, 409–418.

[4] Castagnola, M., Cabras, T., Iavarone, F., Fanali, C., Nemolato, S., Peluso, G., Bosello, S. L., Faa, G., Ferraccioli, G., Messana, I., *Expert Rev. Proteomics* 2012, *9*, 33–46.

[5] Ekström, J., Khosravani, N., Castagnola, M., Messana, I., in: Ekberg, O. (Ed.), *Dysphagia, Diagnosis and Treatment*, Springer-Verlag, Heidelberg, Berlin 2012, pp. 19–47.

[6] Messana, I., Inzitari, R., Fanali, C., Cabras, T., Castagnola, M., *J. Sep. Sci.* 2008, *31*, 1948–1963.

[7] Oppenheim, F. G., Salih, E., Siqueira, W. L., Zhang, W., Helmerhorst, E. J., *Ann. N.Y. Acad. Sci.* 2007, *1098*, 22–50.

[8] Fanali, C., Inzitari, R., Cabras, T., Fiorita, A., Scarano, E., Patamia, M., Petruzzelli, R., Bennick, A., Messana, I., Castagnola, M., *J. Sep. Sci.* 2008, *31*, 516–522.

[9] Ekström, J., Murakami, M., Inzitari, R., Khosravani, N., Fanali, C., Cabras, T., Fujita-Yoshigaki J., Sugiya, H., Messana, I., Castagnola, M., *J. Sep. Sci.* 2009, *32*, 2944–2952.

[10] Castagnola, M., Cabras, T., Iavarone, F., Vincenzoni, F., Vitali, A., Pisano, E., Nemolato, S., Scarano, E., Fiorita, A., Vento, G., Tirone, C., Romagnoli, C., Cordaro, M., Paludetti, G., Faa, G., Messana, I., *J. Matern. Fetal Neonatal. Med.* 2012, *25* (Suppl. 5), 27–43.

[11] Messana, I., Cabras, T., Iavarone, F., Vincenzoni, F., Urbani, A., Castagnola, M., *J. Sep. Sci.* 2013, *36*, 128–139.

[12] Whitelegge, J. P., Zabrouskov, V., Halgand, F., Souda, P., Bassilian, S., Yan, W., Wolinsky, L., Loo, J. A., Wong, D. T., Faull, K. F., *Int. J. Mass. Spectrom.* 2007, *268*, 190–197.

[13] Halgand, F., Zabrouskov, V., Bassilian, S., Souda, P., Wong, D. T., Loo, J. A., Faull, K. F., Whitelegge, J. P., *J. Am. Soc. Mass. Spectrom.* 2010, *21*, 868–877.

[14] Halgand, F., Zabrouskov, V., Bassilian, S., Souda, P., Loo, J. A., Faull, K. F., Wong, D. T., Whitelegge, J. P., *Anal. Chem.* 2012, *84*, 4383–4395.

[15] Peng, Y., Chen, X., Sato, T., Rankin, S. A., Tsuji, R. F., Ge, Y., *Anal. Chem.* 2012, *84*, 3339–3346.

[16] Mirels, L., Bedi, G. S., Dickinson, D. P., Gross, K. W., Tabak, L. A., *J. Biol. Chem.* 1987, *262*, 7289–7297.

[17] Mirels, L., Kopec, L., Yagil, C., Dickinson, D. P., Dziejman, M., Tabak, L. A., *Arch. Oral Biol.* 1990, *35*, 1–5.

[18] Rosinski-Chupin, I., Rougeon, F., *J. Biol. Chem.* 1990, *265*, 10709–10713.

[19] Ten Hagen, K. G., Beres, T. M., Szpirer, J., Szpirer, C., Tabak, L. A., *Biochem. J.* 1997, *324*, 177–184.

[20] Sakulsak, N., Wakayama, T., Hipkaeo, W., Iseki, S., *Arch. Oral Biol.* 2007, *52*, 507–517.

[21] Zhang, Z., Marshall, A.G., *J. Am. Soc. Mass Spectrom.* 1998, *9*, 225–233.

[22] Levin, Y., Schwarz, E., Wang, L., Leweke, F. M., Bahn, S., *J. Sep. Sci.* 2007, *30*, 2198–2203.

[23] Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M., Barton, G. J., *Bioinformatics* 1998, *14*, 892–893.

[24] Geourjon, C., Deléage, G., *Comput. Appl. Biosci.* 1995, *11*, 681–684.

[25] McGuffin, L. J., Bryson, K., Jones, D. T., *Bioinformatics* 2000, *16*, 404–405.

[26] Pollastri, G., McLysaght, A., *Bioinformatics* 2005, *21*, 1719–1720.

[27] Zhang, Y., *BMC Bioinformatics* 2008, *9*, 40.

[28] Kim, D. E., Chivian, D., Baker, D., *Nucleic Acids Res.* 2004, *32* (Web Server Issue), W526–W531.

[29] Morris, A. L., MacArthur, M. W., Hutchinson, E. G., Thornton, J. M., *Proteins* 1992, *12*, 345–364.

[30] Luthy, R., Bowie, J. U., Eisenberg, D., *Nature* 1992, *356*, 83–85.

[31] Tompa, P., *Trends Biochem. Sci.* 2002, *27*, 527–533.

[32] Greco, M. A., Lorand, L., Lane, W. S., Baden, H. P., Parameswaran, K. N., Kvedar, J. C., *J. Invest. Dermatol.* 1995, *104*, 204–210.

[33] Patamia, M., Messana, I., Petruzzelli, R., Vitali, A., Inzitari, R., Cabras, T., Fanali, C., Scarano, E., Contucci, A., Galtieri, A., Castagnola, M., *Peptides* 2005, *26*, 1550–1559.

[34] Messana, I., Cabras, T., Pisano, E., Sanna, M. T., Olianas, A., Manconi, B., Pellegrini, M., Paludetti, G., Scarano, E., Fiorita, A., Agostino, S., Contucci, A. M., Calò, L., Picciotti, P. M., Manni, A., Bennick, A., Vitali, A., Fanali, C., Inzitari, R., Castagnola, M., *Mol. Cell. Proteomics* 2008, *7*, 911–926.

[35] Esposito, C., Caputo, I., *FEBS J.* 2004, *272*, 615–631.

[36] Pastor, M. T., Diez, A., Pérez-Payá, E., Abad, C., *FEBS Lett.* 1999, *451*, 231–234.

[37] Piper, J. L., Gray, G. M., Khosla, C., *Biochemistry* 2002, *41*, 386–393.

[38] Parameswaran, K. N., Velasco, P. T., Wilson, J., Lorand, L., *Proc. Natl. Acad. Sci. USA* 1990, *87*, 8472–8475.

[39] Bradway, S. D., Bergey, E. J., Jones, P. C., Levine, M. J., *Biochem. J.* 1989, *261*, 887–896.

[40] Yao, Y., Lamkin, M. S., Oppenheim, F. G., *J. Dent. Res.* 1999, *78*, 1696–1703.

[41] Yao, Y., Lamkin, M. S., Oppenheim, F. G., *J. Dent. Res.* 2000, *79*, 930–938.

[42] Moriera, J. E., Tabak, L. A., Bedi, G. S., Culp, D. J., Hand, A. R., *J. Histochem. Cytochem.* 1989, *37*, 515–528.

[43] Cooper, L. F., Elia, D. M., Tabak, L. A., *J. Biol. Chem.* 1991, *266*, 3532–3539.

