FEBS 28639 FEBS Letters 572 (2004) 3-7

Plant stress proteins of the thaumatin-like family discovered in animals

Anna Brandazza^a, Sergio Angeli^a, Mariella Tegoni^b, Christian Cambillau^b, Paolo Pelosi^{a,*}

^aDipartimento di Chimica e Biotecnologie Agrarie, University of Pisa, via S. Michele, 4, 56124 Pisa, Italy ^bArchitecture et Fonction des Macromolecules Biologiques, 31 Chemin J. Aiguier, F-13402 Marseille Cedex 20, France

Received 4 May 2004; revised 2 July 2004; accepted 2 July 2004

Available online 14 July 2004

Edited by Hans Eklund

Abstract Thaumatin-like proteins (TLPs) are polypeptides of about 200 residues synthesized by plants in response to fungal infection. In addition to the exceptionally strong sweet taste exhibited by some members, they are also reported to be endowed with endo-β-1,3-glucanase activity and α-amylase inhibiting properties. However, the detailed mechanism of their antifungal action is not completely understood. So far, TLPs have only been described in plants, with several members of the family expressed in the same species. Here, for the first time in animals, we report the identification of two genes encoding members of the thaumatin-like proteins family in the desert locust Schistocerca gregaria and show their expression in different parts of the body. Southern blot and Western blot experiments revealed the presence of orthologous genes and their expression products in the related species Locusta migratoria. A search through the available genomes yielded similar sequences in the nematode Caenorhabditis but not in Drosophila and other insects. A three-dimensional model of S. gregaria TLP suggests a glucanase function. As in plants, TLPs could play a defense role in insects against pathogens.

© 2004 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

Keywords: Thaumatin; Stress protein; Glucanase; Schistocerca gregaria; Locusta migratoria; Caenorhabditis elegans; Caenorhabditis briggsae

1. Introduction

Among the various polypeptides produced by plants in response to different kinds of stress, thaumatin-like proteins represent a homogeneous family, investigated to a great extent in many species. Their name is derived from the first member isolated from the seeds of the plant Thaumatococcus danielli and studied for its exceptionally potent sweet taste [1]. Thaumatin is a polypeptide of 207 residues, with an unusually high isoelectric point close to 12. It is about one hundred thousand times sweeter than sucrose on a molar basis. The structure of thaumatin has been resolved [2,3], but the reasons for its exceptional sweet taste are not clear. Only several years after the discovery of thaumatin, proteins with similar amino acid sequences have been identified in all the plant species examined [4–6]. The genome of Arabidopsis thaliana contains 27 different annotated genes, whose translation products share between 20 and over 90% of their amino acids. Most of these thaumatinlike proteins (TLPs), however, are not endowed with sweet taste, and not all of them present the high isoelectric point of thaumatin. Their synthesis is induced by the presence of pathogenic molds and fungi [7] and they are generally referred to as pathogenesis-related proteins 5 (PR-5) [4]. Their antifungal activity "in vitro" has been reported in some cases, but little is known on the specific mechanism. Their action against pathogenic microorganisms has been related in some cases to endo- β -1,3-glucanase activity [8,9] and α -amylase inhibiting properties [10]. Moreover, PR-5 proteins have been reported to present membrane-permeabilizing activity [11].

Thaumatin-like proteins so far have been identified only in plants. However, based on genome information, similar proteins were predicted to exist in the nematode *C. elegans* [12]. Here, for the first time in animal species, we report the cloning of two genes encoding members of the thaumatin-like proteins family, in the desert locust *Schistocerca gregaria*. The presence of the expressed protein in different parts of the body was confirmed by Western blot experiments. Southern blot and Western blot revealed orthologous genes and their expression products in the related species *Locusta migratoria*.

2. Materials and methods

2.1. Tau1 cDNA fragment isolation

Total RNA was extracted from the antennae and muscle of *S. gregaria*, using the TrizolTM reagent kit (GIBCO BRL), along with the manufacturer's protocol.

The double stranded cDNA fragments pool was obtained with a subtractive approach by a Representational Difference Analysis (RDA) of cDNA according to the published procedure [13]. Briefly: for each sample (antennae = Tester and muscle = Driver) an amplified cDNA was synthesized and digested with DpnI (Invitrogen). Only the Tester was then ligated with specific linkers and mixed in a ratio of 1:100 with the Driver. The obtained solution, after denaturation and renaturation, was digested with the Exo-Mung nuclease (Invitrogen) to digest the protruding ends in the duplex sequences and reamplified using the same linkers as primers. The obtained product was subjected to the same procedure for two more times and the final amplified DNA fragments cloned into a pGem-T vector system (Promega) for sequencing. One of the clones was identified as a Taumatin-like gene (Tau1) by Blast analysis in the Databank.

2.2. RACE PCR

Based on the sequence of the fragment obtained by RDA analysis, sets of specific primers were synthesized and used for the amplification of the 5' and 3' cDNA ends (RACE).

The 5' RACE was performed using a kit (Invitrogen; 5' RACE system, version 2.0) following the manufacturer's protocol. Briefly Total RNA was reverse transcribed using a specific primer (Tau7 = 5'-GATCCTGTAGCCTGTGTT). The first strand was then 3'-tailed using a Terminal deoxynucleotidyl transferase and dCTP. The

^{*} Corresponding author. Fax: +39-050-571564. *E-mail address:* ppelosi@agr.unipi.it (P. Pelosi).

product was used for the second strand synthesis and amplification with a Poly(G) adapter-primer and a second specific primer (Tau5 = GTCGGATTTGTCGTCGTATG). To increase the specificity of the reaction, a second nested PCR was performed using the Adapter primer and a third specific primer (Tau6 = 5'-GCAGTACTGGTCGGTGTGAA). Following this procedure, we identified two different clones named, respectively, Tau1 and Tau2. For the 3' RACE total RNA was reverse transcribed using an oligodT primer and amplified with the same oligodT and Tau3 (5'-GTCGCTTGCAAGAGTGCGTGC) for the Tau1 clone and Tau2-1 (5'-TGCGGCAACAGGCTGTACTG) for the clone Tau2 respectively. The 5' and 3' fragments were then subcloned into a pGem-T vector system (Promega) for sequencing.

2.3. Southern blot

Genomic DNA was isolated from the legs of *S. gregaria* and *L. migratoria* as follows: the tissues were crushed with liquid nitrogen in lysis buffer (100 mM NaCl, 10 mM Tris–Cl pH 8, 25 mM EDTA, pH 8, 0.5% SDS) and digested overnight at 65 °C with Proteinase K. The proteinase digestion was followed by several extraction with phenol-chloroform–isoamyl alcohol and precipitation with cold ethanol. 10 μg of each genomic DNA was digested with *Eco*RI, separated on a 0.8% agarose gel and transferred onto a Nytran Super Charge nylon membrane (Schleicher & Schuell). For the filter hybridization and detection, we followed the DIG system procedure (Roche). Briefly, samples were hybridized at 42 °C overnight in 50% formamide, 5× SSC 0.1% *N*-lauroyl-sarcosine, 0.02% SDS, and 2% Blocking Reagent with a RNA probe labelled with DIG RNA Labelling Kit (Roche), following the manufacturer's procedure. Washing was performed in 0.2× SSC, 0.1% SDS at 42 °C.

2.4. PCR from different tissues

Total RNA was extracted from different tissues of *S. gregaria* as described above. cDNA was prepared from total RNA by reverse transcription, using an oligo-dT and the ThermoscriptTM RT-PCR System (Invitrogen), along with the manufacturer's protocol in a 20 μl total volume.

Aliquots of 1 µl of crude cDNA were amplified in a Bio-Rad Gene CyclerTM thermocycler, using 5 units of *Thermus aquaticus* DNA polymerase (Amersham Pharmacia Biotech), and Tau4 (5'-GAC-GTCAGCTTCGTTGATGGT TAC) and Tau7 as primers.

2.5. Bacterial expression of locust TLP-Sgre1 C-terminal fragment

The DNA encoding the last 124 residues of TLP of *S. gregaria* (TLP-*Sgre1*) was obtained by PCR from the pGEM plasmid containing the appropriate sequence, using the following primers: Tau4-Nde (AT-CATATGGACGTCAGCTTGGTTGAT) and Tau7-Eco (TAGAAT-TCTCAGCCGAAGGTGATCCTGTAGCCTGT). Sequences were designed to include a *Nde1* restriction site with the ATG codon in the *5'* primer and an *Eco*RI restriction site in the 3' primer. Amplified DNA was digested and cloned into pET-5b (Novagen), previously linearized with the same enzymes. The resulting plasmid, pET-Tau1c, was sequenced and shown to encode the C-terminus fragment of TLP-*Sgre1*.

For protein expression, *E. coli* BL21 (DE3)pLys cells were transformed with the pET-Sg-TAU-1c plasmid. Cultures were induced with 0.4 mM isopropyl thio-β-D-galactopyranoside for 3 h at 37 °C, the cells were harvested by centrifugation, resuspended in 50 mM Tris–HCl, pH 7.4, 300 mM NaCl, 1 mM PMSF, 2 mM EDTA (sonication buffer) and lysed by sonication. The recombinant TLP-Sgre1 c was present at this stage entirely in the pellet as inclusion bodies with yields of about 20 mg per litre of culture. After repetitive washings with 0.1% Triton-X 100 in sonication buffer, the protein was solubilized in 3 M urea and used directly for immunization.

2.6. Preparation of antisera

Antisera were obtained by injecting an adult rabbit subcutaneously with 500 μg of recombinant protein purified as described above, followed by two additional injections of 300 μg after 15 and 30 days. The protein was emulsified with an equal volume of Freund's complete adjuvant for the first injection and incomplete adjuvant for further injections. Animals were bled 15 days after the last injection and the serum was used without further purification. Rabbits were individually housed in large cages, at constant temperature, and all operations were performed according to ethical guidelines in order to minimize pain and discomfort to animals.

2.7. Western blot analysis

After electrophoretic separation under denaturing conditions, proteins were electroblotted on a nitrocellulose membrane, by the procedure of Kyhse-Andersen [14]. After treatment with 0.2% dried skimmed milk and 0.05% Tween 20 in PBS for 2 h, the membrane was incubated with the crude antiserum against the protein at a dilution of 1:500 and then with goat anti-rabbit IgG-horseradish peroxidase conjugate (dilution 1:1000). Immunoreacting bands were detected by treatment with 4-chloro-1-naphthol/hydrogen peroxide.

2.8. Three-dimensional model

Thaumatin was taken from the PDB entry 1RQW [15] and the to-bacco thaumatin-like, anti-fungal protein from entry 1AUN [11]. The butyl-maltose, docked into TLP-Sgre 1, was taken from entry 1BYH. The molecular modelling programme Turbo-Frodo [16] has been used to perform the homology modelling, which was straightforward with the exception of three areas corresponding to long insertions (Fig. 4A): area 2 within residues 145–145, area 3 between residues 174 and 177 and areas 4 and 5, between residues 192 and 207. The first two zones were modelled extended, whereas the two last ones, constrained by disulfide bridges, were modelled as coil—coil. Attention was paid to maximize the interactions with the protein core. Due to the positional uncertainity in these insertions, only topology should be taken into account. Fig. 4 was prepared with the PyMOL Molecular Graphics System [17].

3. Results and discussion

While searching for odorant-binding proteins in a pool of cDNA fragments of the desert locust S. gregaria, we obtained short sequences bearing significant similarity to plants TLPs. Adopting a RACE approach at both ends, we sequenced two complete genes (Fig. 1). They contain, respectively, 894 and 886 bp, with open-reading frames of 738 and 723 bp, encoding proteins of 246 and 241 amino acids. The N-terminal regions of both proteins present putative signal peptides of 19 amino acids [18]. A search through the so far available genomes of animal species produced orthologs of these proteins only in the nematodes C. elegans and Caenorhabditis briggsae. In the first species, several genes have been annotated as "thaumatin-like" with a total of seven different deduced amino acid sequences Accession No. NP_507263, NP 500751, NP_500748, NP_500747, NP_502362, NP_502361, and NP 502360). In the second species three genes, annotated as "hypothetical protein" (GenBank Accession No. CAE59849, CAE65915, CAE72818) show significant similarity to TLPs. To confirm that the genes encoding TLPs did not originate from plant contaminations of the locust samples, we have demonstrated their presence in the genomes of S. gregaria and L. migratoria by Southern blot experiments, using a DIG-labelled RNA probe on preparations of genomic DNA digested with EcoRI (Fig. 2A). In both species, the probe clearly recognizes a single band of about 5 kb, suggesting that the two genes are closely located in the genome.

Using primers corresponding to amino acids 175–181 and 239–246 of the full-length mature protein TLP-Sgre1, we were able to amplify products of about 200 bp, as expected, from both cDNA and genomic DNA of S. gregaria, while primers corresponding to positions 122–130 and 239–246 of the same sequence produced the expected amplification product from cDNA (380 bp), but no bands from genomic DNA. This fact indicates that the TLP gene of S. gregaria contains introns, as in the cases of plants and C. elegans TLPs. PCR experiments performed with the latter primers on cDNA obtained from different parts of the body of S. gregaria yielded amplification

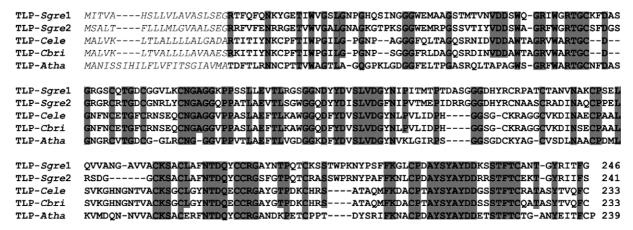


Fig. 1. Complete amino acid sequences of thaumatin-like proteins (TLPs) of *S. gregaria* (TLP-*Sgre1* GenBank Accession No.: AY512591; TLP-*Sgre2* GenBank Accession No.: AY512592), aligned with representative TLPs of *C. elegans* (TLP-*Cele* GenBank Accession No.: NP_502362), *C. briggsae* (TLP-*Cbri* GenBank Accession No.: CAE_59849) and *A. thaliana* (TLP-*Atal* GenBank Accession No.: NP_177641). Conserved residues in all the sequences or at least in the animal sequences are highlighted. Predicted signal peptides are in italics.

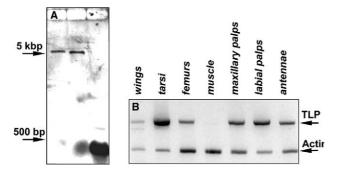


Fig. 2. Panel A. Southern blot experiment showing the presence of a TLP gene in the genomes of *S. gregaria* (Sg) and *L. migratoria* (Lm). Genomic DNA, prepared from leg muscle, was digested with *EcoRI*, blotted on nitrocellulose membrane and, after electrophoretic separation, hybridized with an RNA probe of 400 bp corresponding to 3′ region of the coding sequence of TLP-*Sgre1* and labelled with digoxigenin. Bands were detected using the Roche DIG-Luminescent Detection Kit. 0.5 ng of the DNA fragment corresponding to the probe was used as a positive control (C). Panel B. PCR experiments performed on cDNA extracted from different parts of the body of *S. gregaria*. The primers used corresponded to positions 122–130 and 239–246 of the full-length TLP-*Sgre1*. An amplification band of the expected size (370 bp) was obtained in all the samples except muscle. Primers corresponding to an actin fragment of 170 bp were also included as control.

products of the expected size in all the samples, with the exception of muscle, indicating that the expression of this protein is ubiquitous and probably related to the presence of cuticle (Fig. 2B).

To verify the presence of the gene products of TLP genes discovered in the two species of locusts, we have expressed a fragment of the protein TLP-Sgre1, corresponding to the last 124 residues. The recombinant polypeptide was entirely present in the pellet. After repetitive washings, it was extracted with 3 M urea at a satisfactory degree of purity. This solution was directly used for immunizing a rabbit. The polyclonal antiserum so obtained stained, in Western blot experiments, bands of about 28–29 kDa in legs extracts of S. gregaria and L. migratoria (Fig. 3). Extracts of three plasmid species, Eurycantha calcarata, Sipyloidea sipylus and Carausius morosus,

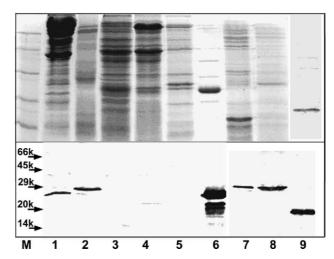


Fig. 3. SDS-PAGE (upper panel) and Western blot (lower panel) of different insect extracts and control proteins. Electrophoretic bands were immunostained with a crude antiserum obtained against a fragment of the thaumatin-like protein (TLP) of *S. gregaria*. M: molecular weight markers (visible on the gel and indicated by arrows on the membrane) were, from the top: bovine serum albumin (66 k), ovalbumin (45 k), carbonic anhydrase (29 k), trypsin inhibitor (20 k) and α-lactalbumin (14 k). 1–5, Legs crude extracts of *L. migratoria*, *S. gregaria*, *Eurycantha calcarata*, *Sipyloidea sipylus* and *Carausius morosus*; 6, TLP purified from kiwi fruits; 7, thorax crude extract of *S. gregaria*; 8, abdomen crude extract of *S. gregaria*; 9, recombinant fragment of *S. gregaria* TLP utilized for the production of antibodies.

failed to cross-react with the antiserum, while a sample of TLP purified from kiwi fruits, as reported in the literature [19], was heavily stained despite the not very high degree of similarity of the TLP-Sgre1 fragment with plant TLPs (around 50%). Thorax and abdomen of *S. gregaria*, as well as other parts of the body not shown in the figure, also express TLP.

The model of TLP-Sgre1 is derived from the structure of tobacco PR5d TLP protein (1AUN, Ref. [11]). It presents 5 significant insertions compared to thaumatin (Fig. 4B–D). These insertions are observed between all insect TLPs and all plant TLPs, while only two insertions are observed with *C. elegans* proteins, at positions ~80 and ~205. Besides the 5

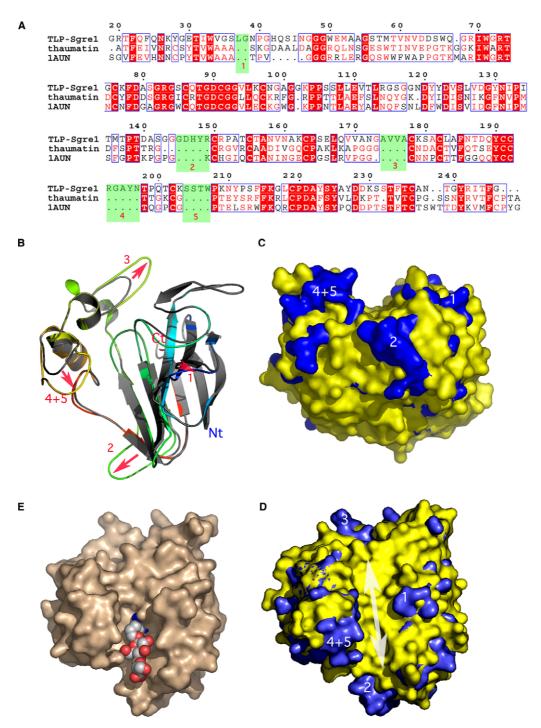


Fig. 4. Model of the 3D structure of TLP-Sgre1. (A) Sequence alignment of TLP-Sgre1 with thaumatin and TLp from tobacco. The 5 major insertions are highlighted in green and numbered 1–5 (red). (B) Ribbon view of TLP-Sgre1 (rainbow colouring) on top of tobacco TLP (gray). The N and C termini are identified (Nt, Ct). The 5 major insertions are indicated. (C) and (D) Molecular surfaces of TLP-Sgre1 (blue) on top of tobacco TLP (yellow). The bottom of the canyon is identified by an arrow. (E) Model of a disaccharide (spheres) docked in TLP-Sgre1 canyon (beige molecular surface). The catalytic acidic residue is colored in blue.

large insertions, several insertion/deletion of one residue are observed. The other parts of the sequences are remarkably similar. Thaumatin and TLPs all present an elongated canyon spanning the whole protein body (Fig. 4C) [15]. In insect TLPs, all the insertions are located near the canyon: looking at Fig. 4C and D, insertion 1 is on the right rim, insertions 2 and 3 are located south and north of the rim, respectively. The

double insertion 4 and 5 is on the south part of left rim. The role of the canyon in PR5d has been proposed to be linked to an antifungal role through the interaction of its negatively charged residues with a receptor of the MAPK transduction pathway [15]. Although the function of this canyon is not documented yet, it is largely admitted that cavities/canyons/groove are favorable places for ligand/substrates binding and

processing. Some authors have demonstrated that β -1,3-glucanase activity is associated with plant TLPs [8,9]. The canyon, present in TLPgre1 and in plant TLPs might also be an ideal place to perform glycan hydrolysis. The shape and size of the canyon is comparable to that of a β -1,4-1,3-glucanase (1BYH). In these enzymes, hydrolysis activity is catalyzed by two acidic residues in close vicinity [20]. Furthermore, the cavity accomodating the polysaccharide often contains hydrophobic residues which interact with the hydrophobic faces of the sugars. These two requirements are also fulfilled in plant TLPs and in our model of insect TLPs: TLP-Sgre1 residues Glu 110 and Asp 123 are conserved in plant TLPs (Fig. 4A), and the south part of the canyon is lined by tyrosines 121, 148 and 223. These observations prompted us to dock a disaccharide into TLP-Sgre1 (Fig. 4E). This experiment confirms that there is enough room to accomodate a polysaccharide in the south part of the canyon.

Homology modelling and loop insertion yielded a topological model of an insect TLP, closely related with the 3D structures of plant TLPs. The canyon spanning the structures of proteins of both families has a favorable size and residue content (acidic/hydrophobic) to fulfill the requirements of glucanase activity. Such an activity has been proposed for plant TLPs. The substitutions observed between plant and insect TLPs all line the canyon. These observations are compatible with a possible role of insect TLPs in sugar binding and/or processing with a specificity different compared to plant TLPs. However, not all plant TLPs show glucanase activity. In a publication by Grenier et al. [8], only 6 of 12 tested TLPs hydrolyzed β-1,3-glucans. Instead, some TLPs are reported to exert their anti-fungal activities by permeabilizing cell membranes. Indeed, enzymatic experiments remain to be done with protein expressed as recombinant to confirm or not the functional hypothesis proposed from the model.

The presence of thaumatin-like proteins in insects, reported here for the first time in animals, extends the use of such defense mechanism at least to some insect species. However, it is possible that members of a protein family will be recruited for different biological functions during the course of evolution of such distantly related groups of organisms. In this respect, it is important to observe that genes encoding TLPs have not been detected in the *Drosophila melanogaster* nor in other insects' genomes. It seems, therefore, that the production of TLPs could represent one of the tools that insects use to control

pathological infections, but such defense mechanism has been conserved only in some species of insects. Detailed information on this aspect will be of basic importance in projects using natural enemies for the control of locusts and other pests in agriculture. We are currently investigating the expression of the protein during the life cycle of locusts and in relationship to their pathological conditions. Finally, the occurrence of TLPs in some insect and nematode species poses the question of their distribution across the diversity of living organisms and their conservation during evolution.

References

- [1] van der Wel, H. and Loeve, K. (1972) Eur. J. Biochem. 31, 221–225.
- [2] Ogata, C.M., Gordon, P.F., de Vos, A.M. and Kim, S.H. (1992) J. Mol. Biol. 228, 893–908.
- [3] Kaneko, R. and Kitabatake, N. (2001) Chem. Senses 26, 167-177.
- [4] Cornelissen, B.J., Hooft van Huijsduijnen, R.A. and Bol, J.F. (1986) Nature 321, 531–532.
- [5] Van Loon, L.C., Gerritsen, Y.A.M. and Ritter, C.E. (1987) Plant Mol. Biol. 9, 593–609.
- [6] Breiteneder, H. and Ebner, C. (2000) J. Allergy Clin. Immunol. 106, 27–36.
- [7] Hu, X. and Reddy, A.S. (1997) Plant Mol. Biol. 34, 949-959.
- [8] Grenier, J., Potvin, C., Trudel, J. and Asselin, A. (1999) Plant J. 19, 473–480.
- [9] Menu-Bouaouiche, L., Vriet, C., Peumans, W.J., Barre, A., Van Damme, E.J. and Rouge, P. (2003) Biochimie 85, 123–131.
- [10] Franco, O.L., Rigden, D.J., Melo, F.R. and Grossi-De-Sa, M.F. (2002) Eur. J. Biochem. 269, 397–412.
- [11] Koiwa, H., Kato, H., Nakatsu, T., Oda, J., Yamada, Y. and Sato, F. (1999) J. Mol. Biol. 286, 1137–1145.
- [12] Kitajima, S. and Sato, F. (1999) J. Biochem. (Tokyo) 125, 1-8.
- [13] Hubank, M. and Schatz, D.G. (1994) Nucleic Acids Res. 22, 5640–5648.
- [14] Kyhse-Andersen, J. (1984) Biochem. Biophys. Methods 10, 203– 209.
- [15] Debreczeni, J.E., Bunkoczi, G., Ma, Q., Blaser, H. and Sheldrick, G.M. (2003) Acta Crystallogr. D: Biol. Crystallogr. 59, 688–696.
- [16] Roussel, A. and Cambillau, C. (1991) Silicon Graphics Geometry Partners Directory 81, 77–78.
- [17] DeLano, W.L. The PyMOL Molecular Graphics System. DeLano Scientific LLC, San Carlos, CA, USA. Available from: http:// www.pymol.org.
- [18] Nielsen, H., Engelbrecht, J., Brunak, S. and von Heijne, G. (1997) Protein Eng. 10, 1–6.
- [19] Wang, H. and Ng, T.B. (2002) Phytochemistry 61, 1-6.
- [20] Keitel, T., Simon, O., Borriss, R.and and Heinemann, U. (1993) Proc. Natl. Acad. Sci. USA 90, 5287–5291.