Structural and Phylogenetic Analyses of the GP42 Transglutaminase from *Phytophthora sojae* Reveal an Evolutionary Relationship between Oomycetes and Marine Vibrio Bacteria*\(^5\)

**Background:** The *Phytophthora sojae* GP42 transglutaminase induces defense responses in parsley and potato.

**Results:** GP42 folds into a novel structure that has arisen through convergent evolution.

**Conclusion:** GP42 has unique structural and enzymatic properties distinct from mammalian or bacterial transglutaminases.

**Significance:** This offers a basis to engineer durable broad spectrum resistance in plants with the design and use of GP42 inhibitors.

Transglutaminases (TGases) are ubiquitous enzymes that catalyze selective cross-linking between protein-bound glutamine and lysine residues; the resulting isopeptide bond confers high resistance to proteolysis. *Phytophthora sojae*, a pathogen of soybean, secretes a Ca\(^{2+}\)-dependent TGase (GP42) that is activating defense responses in both host and non-host plants. A GP42 fragment of 13 amino acids, termed Pep-13, was shown to be absolutely indispensable for both TGase and elicitor activity. GP42 does not share significant primary sequence similarity with known TGases from mammals or bacteria. This suggests that GP42 has evolved novel structural and catalytic features to support enzymatic activity. We have solved the crystal structure of the catalytically inactive point mutant GP42 (C290S) at 2.95 Å resolution and identified residues involved in catalysis by mutational analysis. The protein comprises three domains that assemble into an elongated structure. Although GP42 has no structural homolog, its core region displays significant similarity to the catalytic core of the Mac-1 cysteine protease from Group A *Streptococcus*, a member of the papain-like superfamily of cysteine proteases. Proteins that are taxonomically related to GP42 are only present in plant pathogenic oomycetes belonging to the order of the Peronosporales (*e.g.*, *Phytophthora*, *Hyaloperonospora*, and *Pythium* spp.) and in marine *Vibrio* bacteria. This suggests that a lateral gene transfer event may have occurred between bacteria and oomycetes. Our results offer a basis to design and use highly specific inhibitors of the GP42-like TGase family that may impair the growth of important oomycete and bacterial pathogens.

Transglutaminases (TGases)\(^4\) ((R)-glutamyl-peptide:amine-\(\gamma\)-glutamyltransferases, EC 2.3.2.13) catalyze an acyl transfer reaction between peptide-bound glutamine residues and primary amines, including the \(\epsilon\)-amino group of peptide-bound lysine residues. TGase activity has been implicated in a multitude of physiological activities in animals, plants, and bacteria (1–3). Human TGases such as the blood coagulation factor XIIIa or the tissue TGase 2 have been structurally characterized (4, 5). Despite the low degree of homology at the amino acid sequence level, it appears that human TGases share a core structural fold with the papain-like cysteine proteases (1). The reaction mechanism of TGases and cysteine proteases is based on a catalytic triad formed by Cys, His, and Asp residues and proceeds via an acyl-enzyme intermediate in which the active site cysteine is covalently bound to the ligand (1). This is a striking example of the conservation in evolutionary distant taxa of a tertiary structure in proteins mechanistically related but with highly diverging primary sequences. Bioinformatics analyses performed by Makarova et al. (6) demonstrated a statistically significant sequence similarity between mammalian TGases and a group of microbial proteins of which the only functionally characterized member is a protease (7). Subsequently, it was suggested that animal TGases have evolved from...
ancestral cysteine proteases, but to date, there is a lack of evidence to support such a model of divergent evolution.

*Phytophthora* species are pathogens that infect various plants, leading to major economic and environmental damage. *Phytophthora sojae* in particular causes root and stem rot of soybeans. Immunolocalization assays demonstrated that the GP42 TGase is associated with hyphal cell walls of *P. sojae* growing *in planta* (8). We have previously shown that extracellular TGase activity is detectable in at least 10 different *Phytophthora* spp. and that proteins serologically related to GP42 are expressed in all these species (9). The GP42 TGase is similar in both its biochemical and its enzymatic characteristics to mammalian TGases (9), but its biological function is yet to be elucidated. Notably, GP42 serves as a microbe-associated molecular pattern that triggers plant immunity in a plant pattern recognition receptor-dependent manner (9–11). A stretch of 13 amino acids (Pep-13: VWNQPVRGFKVYE), located in the C-terminal region of GP42, has been shown to be sufficient for receptor-mediated perception and plant immunogenic activity (11).

To provide a platform for understanding the molecular mechanism of the GP42-catalyzed transamidation reaction and to allow for its comparison with other known enzymes, we have determined the crystal structure of a catalytically inactive form of GP42 in which the active site cysteine residue is replaced with a serine. The core region of GP42 superposes well with the canonical domain characteristic of proteins belonging to the papain-like protein superfamily of cysteine proteases. However, the GP42 sequences surrounding the catalytic core fold into domains that are novel and have not been observed to date in other proteins. Moreover, the arrangement of the catalytic triad in GP42 differs from that seen in cysteine proteases. Hence, the structural analysis clearly suggests that GP42 has arisen through convergent evolution. GP42-related sequences are only present in closely related oomycetes belonging to the order of the Peronosporales and in a small group of marine *Vibrio* bacteria that can infect numerous fish species and marine invertebrates (12). We show that the TGase-like protein from *Vibrio harveyi* displays significant TGase activity *in vitro*. We therefore conclude that TGases from plant pathogenic oomycetes were acquired from bacteria by horizontal gene transfer, probably conferring a selective advantage to Peronosporales over other oomycetes.

**EXPERIMENTAL PROCEDURES**

**Recombinant Protein Expression and Purification**—The cDNA sequence encoding the mature *P. sojae* GP42 (GenBank accession AAA67875.1, amino acids 163–529) was cloned into the pPIC9K vector (Invitrogen) and subjected to site-directed mutagenesis using the GeneEditor in vitro site-directed mutagenesis system (Promega) to replace Cys-290 with serine. The expression of recombinant GP42 (C290S) was performed in *Pichia pastoris* GS115 according to the multi-copy *Pichia* GS115 method. 450 μl of reservoir solution (0.1 M K$_2$HPO$_4$/KH$_2$PO$_4$, pH 7.9, 5 mM DTT, 5 mM CaCl$_2$, and 0.1 M KCl) were mixed with 450 μl of reservoir solution: 0.1 M CAPS, pH 6.0, 10 mM DTT, 5 mM CaCl$_2$, 10 μg/μl N,N′-dimethylecasein, 1 μM [2,3-3H]putrescine (60–120 Ci/mmol; Hartmann Analytic), 10 mM putrescine. Protease activity was determined using the ENZCHEK peptidase/protease assay kit (Invitrogen) following the instructions provided by the manufacturer.

**Crystallography**—Recombinant GP42 (C290S) crystals grew with the sitting drop vapor diffusion method. 450 μl of GP42 (C290S) at a concentration of 2–3 mg/ml in 0.1 M K$_2$HPO$_4$/KH$_2$PO$_4$, pH 7.9, 5 mM DTT, 5 mM CaCl$_2$, and 0.1 M KCl were mixed with 450 μl of reservoir solution: 0.1 M CAPS, pH 11, 0.2 M Li$_2$SO$_4$, 1.6 M (NH$_4$)$_2$SO$_4$. For cryoprotection, crystals were transferred into 3 mM sodium malonate (pH 7.4) before freezing them in liquid nitrogen. Heavy atom soaks were carried out with 1 mM [Ta$_6$Br$_{12}$]$^{2-}$ and 2 Br$^-$; (Jena Biosciences) dissolved in 3 mM sodium malonate (pH 7.4). The crystals turned green and were transferred back into 3 M...
sodium malonate (pH 7.4) prior to flash-freezing. The crystals belong to space group P6\(_{2}2\) and contain two molecules per asymmetric unit (see Table 1). Both native and derivative datasets were collected at beamline PXIII of the Swiss Light Source in Villigen, Switzerland. Diffraction data were processed with XDS (14).

Crystal Structure Determination and Structural Refinement—The structure was solved by single isomorphous replacement with anomalous scattering phasing using autoSHARP (15). As the phasing power was poor beyond 6 Å resolution, phase extension to the resolution of the native dataset was performed. To further improve the initial phases, the two-fold non-crystallographic symmetry and the high solvent content of the crystals (78%) were exploited. Solvent flattening was performed with RESOLVE (Los Alamos National Laboratory). The resulting electron density map was of good quality, allowing manual model building using Coot (16). Crystallographic refinement was carried out with Refmac5 (CCP4) (17). PyMOL (35) was used to create the structure figures. The topology diagram was created using topdraw (CCP4) (18).

Phylogenetic Analysis—To determine the extent and distribution of TGase-like sequences among organisms, known and predicted proteins within the National Center for Biotechnology Information (NCBI) non-redundant databases were searched by BLASTP using the P. sojae GP42 sequence as a query. Predicted proteins from whole genome sequencing of *Hyaloperonospora arabidopsis* (VBI Microbial Database Version 5.0), *Phytophthora infestans* (available through the Broad Institute of MIT and Harvard), *Phytophthora sojae* and *Phytophthora ramorum* (www.jgi.doe.gov), and from *Pythium ultimum* (Pythium Genome Database) were retrieved and searched using BLASTP. Sequences were edited to remove redundant and fragmentary proteins, and low scoring sequences (expected value > 10\(^{-5}\)). Protein sequence alignments were made using CLUSTALW, MUSCLE, and T-COFFEE, and phylogenetic inferences and bootstrap values were calculated by maximum likelihood, neighbor joining, or the unweighted pair group method with arithmetic means, using the computer software PHYLIP and MEGA. Jalview was used to further analyze the alignments made with CLUSTALW. N- and C-terminal sequences that did not correspond to the mature sequence of *P. sojae* GP42 (amino acids 163–529) were removed. The alignment was colored according to percentage of sequence identity (19).

**RESULTS AND DISCUSSION**

Overall Structure of GP42—GP42 from *P. sojae* consists of an N-terminal prodomain region (amino acids 1–162) that is cleaved off during enzyme maturation and a region that comprises the TGase (amino acids 163–529). The TGase was recombinantly expressed in *P. pastoris* and purified from the culture medium using a two-step chromatography protocol (see “Experimental Procedures”). We solved the crystal structure of the catalytically inactive point mutant (C290S) at a resolution of 2.95 Å by single isomorphous replacement with anomalous scattering (Table 1). The structure of GP42 reveals a highly intertwined fold. Its overall appearance resembles a sea horse, with a head, a body, and a curved tail (Fig. 1, A and B). To facilitate discussion of its structural elements, GP42 can be divided into three domains: the head (domain I), body (domain II), and tail (domain III). The head, depicted in yellow, consists of a small \(\beta\)-sheet formed by two \(\beta\)-strands (\(\beta2\) and \(\beta3\)) that faces a short \(3_{10}\) helix. The body (shown in cyan) is compact, globular, and predominantly composed of \(\alpha\)-helices (\(\alpha2\), \(\alpha3\), \(\alpha4\), \(\alpha5\), \(\alpha8\), and \(\alpha9\)). The long \(\alpha5\) helix forms the core of the domain, whereas the other helices are arranged circularly around it. The body domain also contains a short \(3_{10}\) helix following \(\alpha8\) and a small \(\beta\)-sheet (\(\beta4\) and \(\beta5\)) between \(\alpha4\) and \(\alpha5\). The tail domain (depicted in navy) has an elongated, curved shape and features elaborate loops. The core of the tail domain is primarily formed by a long and highly twisted antiparallel \(\beta\)-sheet (\(\beta\)-strands \(\beta1\), \(\beta8\), \(\beta9\), \(\beta10\), \(\beta11\)) that wraps tightly around helix \(\alpha6\). Loops connecting the \(\beta\)-strands in some cases feature helices (\(\alpha7\), \(\alpha8\), and a \(3_{10}\) helix). The tail domain is connected to the body domain via a small four-stranded antiparallel \(\beta\)-sheet (\(\beta6\), \(\beta7\), \(\beta8\), and \(\beta12\)). An additional link to the body domain is provided by the N-terminal helix \(\alpha1\) and strand \(\beta1\). Although both are part of the tail domain, the chain enters the body domain with residues following \(\beta1\). GP42 contains seven cysteine residues. Intramolecular disulfide bridges between Cys-256 and Cys-271 and between Cys-262 and Cys-282 stabilize the body domain, whereas a third bridge links Cys-305 with Cys-527 in the tail domain. The presence of disulfide bridges is quite unusual because they can interfere with conformational changes upon activation in mammalian TGases (1, 20). The seventh cysteine, Cys-290, is the active-site nucleophile for TGase activity. Replacement of Cys-290 with serine was necessary for crystallization as active TGases are able to catalyze *in vitro* disordered intermolecular cross-linking, resulting in insoluble aggregates (5, 21, 22). As replacement of cysteine with serine is unlikely to affect the fold of the protein, we expect that both side chains will be in equivalent positions in GP42. For

| **Table 1** X-ray data collection and structural refinement |
|---------------------------------|-----------------|-----------------|
| **Data collection**             | **Native**      | **Ta\(_{6}Br_{12}\) derivative** |
| Resolution (Å)                  | 30–2.95 (3.03–2.95) | 50–5.50 (5.64–5.50) |
| Space group                     | P6\(_{2}2\)     | P6\(_{2}2\)     |
| Cell dimensions                 | a and c (Å)     | 195.5, 137.3    |
|                                | \(\alpha\), \(\beta\), and \(\gamma\) | 90, 90, 120     |
| \(\Delta R_{	ext{work}}\)      | 9.0 (58.3)      | 10.2 (27.2)     |
| \(\Delta R_{\text{free}}\)     | 1.0             | 1.255           |
| Completeness (%)                | 96.4 (98.2)     | 99.8 (100)      |
| Redundancy                      | 3.7             | 6.5             |
| **Refinement**                  |                 |                 |
| \(R_{	ext{work}}\)/\(R_{	ext{free}}\) (%) | 21.5/23.5      |                 |
| Figure of merit B-factors       | 0.85            |                 |
| Chain A (Å\(^2\))              | 47.0            |                 |
| Chain B (Å\(^2\))              | 47.0            |                 |
| No. of atoms                    |                 |                 |
| Protein                         | 5297            |                 |
| CAPS                            | 28              |                 |
| r.m.s.d.                        |                 |                 |
| Bond lengths (Å)                | 0.01            |                 |
| Bond angles (°)                 | 0.9             |                 |
| Ramachandran plot\(^a\)        |                 |                 |
| Favored (%)                     | 656 (96.5)      |                 |
| Allowed (%)                     | 24 (3.5)        |                 |

\(^a\) Data collection was at Beamline PXIII (SLS, Villigen, Switzerland). Values in parentheses are for highest resolution shell.

\(^b\) Ramachandran plot generated with Rampage (CCP4) (34).
simplicity, we will from here on refer to the residue at position 290 as Cys-290.

The amino acid sequence corresponding to Pep-13 is located on β-strands β8 and 8′, near Cys-290 (Fig. 1, A and B). Belonging to two different sheets, these two strands are separated by a single proline residue (Pro-396) that causes the chain to bend at this position. One face of β8/8′ is fully exposed to solvent, rendering many of the residues of Pep-13 accessible to putative ligands. Moreover, because Pep-13 is part of β-strands 8 and 8′, many of its residues are conformationally restrained. The surface exposure of the Pep-13 sequence within GP42 is one feature that logically explains why this sequence constitutes an immunogenic motif that triggers plant immunity through activation of plant pattern recognition receptors.

Architecture of the Active Site—Inspection of the structure reveals that Cys-290 is located at the interface between the body and tail domain, at the N terminus of helix αå. One side of Cys-290 is covered by the loop connecting β-strands β8′ and
β9, but the other side is exposed to the surface and poised to interact with its substrate. The reaction mechanism of all known TGases is based on a Cys-His-Asp triad or, less frequently, a Cys-His dyad (1). Nucleophilic attack on the substrate is performed by the sulfhydryl group of the cysteine following activation by a thiolate-imidazolium ion pair involving the histidine side chain. Six histidines are present in the sequence of GP42, but only one of these, His-291, is in close proximity to Cys-290. The remaining histidines have distances of at least 17 Å from the Cys-290 sulfhydryl group and thus cannot participate in catalysis. His-291 is located immediately next in sequence to Cys-290, and it is part of the same αhelix.

Such an unusual arrangement of two neighboring amino acids forming a catalytic dyad or triad has so far not been described in the literature for any cysteine protease or TGase. Moreover, the distance between the imino nitrogen atom of His-291 and the sulfhydryl group of Cys-290 is, with 4.8 Å in chain A and 5.0 Å in chain B, rather large, and it is therefore unclear whether Cys-290 can be directly activated by the His-291 imidazole group. However, the His-291 side chain does form a salt bridge with Asp-328 (3.0 Å distance) located on the α5 helix, and therefore Cys-290, His-291, and Asp-328 could constitute a functional catalytic triad (Fig. 1C). Different rotamers of His-291 would bring its side chain closer to Cys-290. Such rotamers are not observed in the crystal structure but may occur in solution.

It is also conceivable that the main function of His-291 is not to lower the pKₐ value of Cys-290 but to support catalysis by stabilizing an intermediate. Such a mechanism was shown for FabH, an enzyme that catalyzes a transacylase reaction in the biosynthesis of fatty acids (23). The active site of this enzyme is composed of Cys-112, His-244, and Asn-274. The distance between Cys-112 and His-244 is, with 3.6 and 4.1 Å in different protomers, also larger than a hydrogen bond. Thus, deprotonation of Cys-112 would not be supported by interaction with His-244. Instead, the pKₐ of Cys-112 is lowered due to its localization at the N terminus of an α-helix and the resulting dipole effect (24). The GP42 residue Cys-290 is also located at the N terminus of an α-helix (α4), and a similar dipole effect could increase its acidity. It is therefore possible that the His-291-Asp-328 pair stabilizes a reaction intermediate by hydrogen bonding.

Mapping the electrostatic potential of GP42 onto its surface reveals that most of the protein is either electropositive or neutral (Fig. 1D). Strikingly, however, a strong negative potential delineates a groove adjacent to the active site. This region features several acidic residues projecting from different β-strands and loops. The presence of a highly negatively charged region in close proximity to the active site suggests that the groove may interact with a positively charged substrate. Although the identity of GP42 substrates is not yet known, it is likely that the enzyme catalyzes the cross-linking between peptides that contain basic amino acids.

Mutational Analyses of Residues Cys-290, His-291, and Asp-328—To determine the function of residues Cys-290, His-291, and Asp-328, which constitute the catalytic triad in GP42, point mutants of all three residues were generated, and the culture filtrates of P. pastoris expressing the recombinant proteins were tested for enzymatic activity (Table 2, supplemental Fig. S1). The advantage to using P. pastoris for protein expression is that there is no detectable TGase activity in the culture filtrate of the untransformed strain. In a standard assay, based on the incorporation of [3H]putrescine into N,N'-dimethylcasein, the specific TGase activity of GP42 was 30.15 units/mg. This value was about 2 orders of magnitude above the value obtained for the guinea pig liver TGase (0.117 units/mg). As expected, the C290S mutant had no detectable activity. We also generated single mutants H291A, H291F, H291Y, and D328N as well as the double mutant H291A/D328N. The D328N mutation is a conservative exchange unlikely to cause structural changes. In line with our hypothesis that Asp-328 is part of the catalytic triad of GP42, the D328N mutant had only 2.6% of wild-type GP42 activity (Table 2). To determine how size and polarity of His-291 affects catalytic activity, it was first replaced with Phe or Tyr. The His-291 side chain is solvent-exposed, and the two substitutions are unlikely to produce steric clashes. Both mutants exhibit negligible activity (Table 2), suggesting that neither Tyr-291 nor Phe-291 is able to functionally replace His-291. However, the slightly larger side chains of Phe-291 and Tyr-291 may also make access of the substrate to the active site more difficult, thus affecting activity. To test this possibility, we mutated His-291 to Ala and found that the H291A mutant retained 19.1% of wild-type activity (Table 2). An explanation for this unexpected finding could be that due to the small size of the Ala side chain, the substrate can enter the active site more easily, perhaps facilitating cleavage in the absence of the proton-abstracting histidine side chain. A second possibility is that one or two water molecules could enter the active site in the H291A mutant, linking Cys-290 with Asp-328 and thus helping to deprotonate Cys-290 through a proton relay system. The additive effect of H291A and D328N in the double mutant, which has an activity of 0.3% of the wild type (Table 2), clearly demonstrates the crucial role of the two residues in catalysis. Finally, we also mutated Asn-394 to Ala (N394A) because this residue forms direct hydrogen bonds with Ser-290 (Fig. 1C) as well as with Trp-496. It is likely that Asn-394 also makes a similar interaction with Cys-290 in the wild-type protein, although the hydrogen bond would be somewhat weaker. Replacement with Ala reduced the activity to 0.02% (Table 2), indicating that Asn-394 may be responsible for the correct positioning of Cys-290.

GP42 Exhibits Structural Homology to the Cysteine Protease Mac-1 from Group A Streptococcus—A search for tertiary structure homologs using the DALI server did not reveal any

| TABLE 2 |
|------------|
| **TGase activity of GP42 point mutants** |
| GP42 | TGase activity | Relative TGase activity |
| Wild type | 3.0 ± 1.80 | 100 |
| C290S | ND* | ND* |
| D328N | 0.077 ± 0.001 | 2.6 |
| W393F | 22.70 ± 3.72 | 75.3 |
| N394A | 0.007 ± 0.0001 | 0.02 |
| H291A/D328N | 0.088 ± 0.001 | 0.3 |
| H291A | 5.76 ± 0.58 | 19.1 |
| H291F | ND* | ND* |
| H291Y | 0.068 ± 0.001 | 0.2 |

* ND, not detectable
Structure of a Phytophthora Transglutaminase

FIGURE 2. The structure of the catalytic core domain is conserved between the P. sojae GP42 and the bacterial Mac-1 cysteine protease. Structurally conserved regions in GP42 (A) and Mac-1 (B) are shown in pink. Alignment of the Ca atoms using lsqkab resulted in an r.m.s.d. value of 1.4 Å over 77 aligned residues (out of 367 GP42 residues). C, close-up view of the superposition of GP42 and Mac-1, centered at their active sites. The postulated catalytic triads of GP42 (Ser-290, His-291, and Asp-328) and Mac-1 (Ala-94, His-262, and Asp-284) and the GP42 residues Trp-393 and Tyr-443 that superpose with the catalytic residues His-262 and Asp-284 of Mac-1 are shown as stick representations and are highlighted in green (GP42) and yellow (Mac-1).

significant (based on Z-scores >6.0) homology of GP42 to other proteins when the individual domains are considered. However, searches with the complete structure of GP42 yielded modest hits with bacterial cysteine proteases, indicating the presence of some structural similarity. The highest score of Z = 6.4 was obtained with the IgG-degrading cysteine protease Mac-1 from Group A Streptococcus. Inspection of the crystal structure of the inactive (C94A) point mutant of Mac-1 (25) reveals that its core structure indeed possesses some similarity with GP42 (Fig. 2, A and B). Alignment of the Ca atoms of both structures using lsqkab (26) confirms this by yielding a root mean square deviation (r.m.s.d.) value of 1.4 Å for 77 superposed residues out of 367 GP42 residues. Regions of GP42 that can be superposed onto Mac-1 include elements of both the body and the tail domains. Most noticeably, a region comprising the active-site cysteine of GP42, including helix α4 and the two neighboring β-strands β8/β9, superposes well onto the corresponding helix and β-strands that form the catalytic core in Mac-1. These structural elements are shared by all members of the papain-like protein superfamily, including the papain-like cysteine proteases and peptide: N-glycanase or N-acetyl transferases. Interestingly, the superposition described above brings the two catalytically active cysteine residues (Cys-290 in GP42 and Cys-94 in Mac-1) in almost perfect alignment, with an r.m.s.d. value of 0.5 Å (Fig. 2C). However, the remaining residues of the catalytic triad of Mac-1 (His-262 and Asp-284) have no equivalent residues in GP42. Although His-262 and Asp-284 of Mac-1 align well with Trp-393 (r.m.s.d. = 1.0 Å) and Tyr-443 (r.m.s.d. = 2.0 Å) of GP42, neither Trp-393 nor Tyr-443 could function in a catalytic triad (Fig. 2C). The residues that most likely participate in the catalytic triad of GP42, His-291 and Asp-328, face into the opposite direction. Thus, although the central cysteine and key secondary structure elements are structurally conserved in the two proteins, the catalytic triads have apparently evolved differently.

Previous studies have shown that replacing Trp-393 and Pro-396 within the Pep-13 motif with alanines compromised the ability of the protein to induce defense responses in plants and also affected TGase activity to a similar extent (2–6% of wild-type activity) (9, 11). Interestingly, the structural comparison shows that Trp-393 aligns with the catalytically important His-262 of the bacterial Mac-1 protease (Fig. 2C). The W393F mutant retained 75% activity, indicating that the tryptophan is not directly involved in catalysis. Trp-393 may, however, play a structural role or be required for substrate binding. It was proposed for the human TGase 2 and red sea bream tissue TGase that highly conserved tryptophans close to the active site regulate substrate entry to the active site or are involved in substrate binding (27–29). We therefore think it likely that a rearrangement of the catalytic triad occurred during evolution of the Phytophthora TGases to accommodate aromatic residues (such as Trp-393 and Tyr-443) that were required for specificity. It is tempting to speculate that such a change was accompanied by a switch from hydrolytic to transamidating reactions.

Phylogenic Analysis of TGase Sequences Delineates Bacterial and Oomycete Sequences—Using the GP42 sequence as a query, we performed BLAST searches of public databases and sets of predicted proteins resulting from whole genome sequencing of selected organisms (as described under “Experimental Procedures”). More than 100 sequences were obtained, but these were reduced to a core set of 61 sequences after removing redundant, fragmentary, and low scoring hits (expected value >10−5) (supplemental Fig. S2). This core set of 61 TGase-like sequences originated from a limited number of species, as shown in supplemental Table S1. Remarkably, all seven of the prokaryotic TGase-like sequences were from marine bacteria, whereas all 54 of the eukaryotic TGases were from plant pathogenic oomycetes belonging to the order of the Peronosporales.

To determine the relationships among the 61 TGase-like sequences retrieved from databases searches, we performed alignments and phylogenetic analysis using diverse programs. The TGase-like proteins consistently separate into bacterial and eukaryotic groups regardless of the methods used to make the alignments or construct the phylogeny. The branch separation of bacterial from oomycete TGase-like proteins always returns the maximum bootstrap value (equal to the number of replications), whereas other branches within the two groups vary in their placement and bootstrap values depending on the algorithm (Fig. 3A). The oomycete TGase-like proteins form a monophyletic group that is well separated from the bacterial proteins. Within the oomycete branch, TGase-like proteins from different species of Phytophthora, Hyaloperonospora, and...
Pythium are variously interspersed (orthologous) and clustered (paralogous) among each other. Our expectation was to identify TGase-like sequences in marine oomycetes that are more likely to share the same habitat as compared with the *Vibrio* species, but there is no evidence of TGase-like sequences in the genome of *Saprolegnia* and *Aphanomyces* species. Our analysis therefore suggests that a single ancestral gene gave rise to all Peronosporales TGases. This hypothesis is further supported by the analysis of surface residue conservation among all members of the oomycete TGase-like proteins (Fig. 3B and supplemental Fig. S3). Amino acid residues that were identical in >75% of the sequences are predominantly concentrated at the active site and in the surrounding area, which is probably involved in substrate binding and recognition. The residues forming the catalytic triad and the first six amino acids of the Pep-13 fragment (VWNQPV), which are located in close proximity to the active site, are present in >90% of the analyzed sequences. The less conserved residues in the C-terminal portion of Pep-13 are not as relevant as Trp-393 and Pro-396 for elicitor activity. The heptamer VWNQPVR was found to be completely inactive, demonstrating that the C-terminal part is important for plant perception and induction of immune responses (11). The higher degree of variability might be a strategy of the pathogen to evade plant perception. Whether such TGase homologs with a degenerated Pep-13 motif still trigger plant immune responses and display TGase activity requires further experiments.

The high structural similarity among the oomycete TGase-like proteins suggests that they share similar substrates and a similar catalytic mechanism. The presence of functionally identical TGases, highly conserved within the order of the Peronosporales, is an important criterion of genuine microbial-associated molecular patterns that are by definition present in a broad range of microorganisms.

*The Oomycete TGases Originate from Marine Vibrio Species*—The phylogenetic analysis has shown that the bacterial TGase-like sequences also formed a monophyletic group, and within this, a subgroup of TGase-like proteins from *Vibrio* species clustered together. In fact, most (six out of seven) of the bacterial sequences originated from Vibrionales, the exception being one

[FIGURE 3. Structurally related GP42 TGases are conserved in all plant pathogenic oomycetes. A, phylogenetic analysis of TGase protein sequences. Shown is a hypothetical radial tree resulting from comparison of 61 TGase protein sequences. The unrooted, maximum likelihood tree was produced from a MUSCLE alignment. Bootstrap values from 100 replicates are shown for major branches. The scale bar represents 20% weighted sequence divergence. B, surface representation of conserved amino residues within the oomycete TGases. The program Jalview was used to identify residues that are identical in >75% (highlighted in blue) of the 54 GP42-related protein sequences from *Phytophthora*, *Pythium*, and *Hyaloperonospora*. These highly conserved residues include the active-site residues Cys-290, His-291, and Asp-328 (highlighted in red) and the N-terminal part of the Pep-13 motif (highlighted in yellow).]
predicted protein from *Hahella chejuensis*. Alignment of the amino acid sequences of GP42 and its closest homolog from *V. harveyi* (VhTGase) revealed the presence of a highly conserved core sequence (31% identity; 45% similarity over 336 amino acids; Fig. 4A). In particular, the positions of the residues involved in the catalytic reaction (Cys-His-Asp) are marked with asterisks. B, surface representation of conserved amino residues between the *P. sojae* GP42 and bacterial TGases. The program Jalview was used to identify residues that are identical in >75% (highlighted in cyan) of the six GP42-related protein sequences from *Vibrio* spp. The conserved residues include the active-site residues Cys-290 and His-291 (highlighted in red) and the N-terminal part of the Pep-13 motif (highlighted in yellow). Asp-328 (highlighted in red), which is postulated to be part of the catalytic triad in GP42, is not conserved between the oomycete and the *Vibrio* sequences.

FIGURE 4. A GP42-related TGase is present in marine bacteria. A, sequence alignment of the *P. sojae* GP42 and *V. harveyi* TGase. Identical (black) and similar (gray) residues are highlighted. Based on the crystal structure of GP42 and the structure prediction of VhTGase, conserved secondary elements are indicated by cylinders (α-helices) and arrows (β-strands). The positions of the residues involved in the catalytic reaction (Cys-His-Asp) are marked with asterisks. B, surface representation of conserved amino residues between the *P. sojae* GP42 and bacterial TGases. The program Jalview was used to identify residues that are identical in >75% (highlighted in cyan) of the six GP42-related protein sequences from *Vibrio* spp. The conserved residues include the active-site residues Cys-290 and His-291 (highlighted in red) and the N-terminal part of the Pep-13 motif (highlighted in yellow). Asp-328 (highlighted in red), which is postulated to be part of the catalytic triad in GP42, is not conserved between the oomycete and the *Vibrio* sequences.
proteins is very similar to the pattern obtained for the oomycete TGases alone (Fig. 4B and supplemental Fig. S6). High sequence similarity in regions adjacent to the catalytic residues and Pep-13 motif favors the hypothesis that the Vibrionales and oomycete TGase-like proteins arise from a common ancestor rather than from convergent evolution. There is a high probability that a lateral gene transfer event with Vibrio spp. as the bacterial source occurred prior to speciation and radiation of the Peronosporales. Combined with the absence of TGase-related sequences in animals, plants, fungi, and other bacteria, it is unlikely that a common ancestral protein has been selectively retained or lost through evolution because this would have required a high number of independent gene losses. Interestingly, homologs of the necrosis- and ethylene-inducing peptide 1 (Nep1)-like proteins (NLPs), which represent a major class of toxin-like virulence factors in Peronosporales, are also present in the genomes of several Vibrionales (30). Thus, a comprehensive search of the P. sojae genome for bacterium-derived genes might reveal additional acquisitions and help to document at which frequency gene transfer occurred. Although the biological function of GP42 has not yet been elucidated, the protein is thought to play a crucial role for pathogen lifestyle or in the infection process on plants. This hypothesis is strongly supported by several lines of evidence that allowed us to consider the Pep-13 motif as a genuine microbe-associated molecular pattern (9–11). We conclude that the GP42-related TGase family probably confers an important gain of pathogenicity to all the Peronosporales that may have supported colonization of land plants. Several virulence-associated proteases have been identified from V. harveyi strains (31, 32). Whether the TGase-like proteins from marine Vibrio species are playing a crucial role for pathogenicity on crustacean and mollusks also needs to be demonstrated.

Acknowledgments—We thank Dr. Leighton Pritchard for helpful discussions about the evolutionary aspect of TGases, and we are grateful to the staff at beamline PXIII at the Swiss Light Source (SLS), Villigen, Switzerland, for technical support.

REFERENCES

1. Lorand, L., and Graham, R. M. (2003) Nat. Rev. Mol. Cell Biol. 4, 140–156
2. Serafini-Fracassini, D., and Del Duca, S. (2008) Ann. Bot. 102, 145–152
3. Yokoyama, K., Nio, N., and Kikuchi, Y. (2004) Appl. Microbiol. Biotechnol. 64, 447–454
4. Liu, S., Cerione, R. A., and Clardy, J. (2002) Proc. Natl. Acad. Sci. U.S.A. 99, 2743–2747
5. Yee, V. C., Pedersen, L. C., Le Trong, I., Bishop, P. D., Stenkamp, R. E., and Teller, D. C. (1994) Proc. Natl. Acad. Sci. U.S.A. 91, 7296–7300
6. Makarova, K. S., Aravind, L., and Koonin, E. V. (1999) Protein Sci. 8, 1714–1719
7. Pfister, P., Wasserfallen, A., Stettler, R., and Leisinger, T. (1998) Mol. Microbiol. 30, 233–244
8. Hahlbrock, K., Scheid, D., Logemann, E., Nürnberger, T., Parniske, M., Reindold, S., Sacks, W. R., and Schmelzer, E. (1995) Proc. Natl. Acad. Sci. U.S.A. 92, 4150–4157
9. Brunner, F., Rosahl, S., Lee, J., Rudd, J. J., Geiler, C., Kauppinen, S., Rasmussen, G., Scheil, D., and Nürnberger, T. (2002) EMBO J. 21, 6681–6688
10. Halim, V. A., Hunger, A., Macioszek, V., Landgraf, P., Nürnberger, T., Scheel, D., and Rosahl, S. (2004) Physiol. Mol. Plant Pathol. 64, 311–318
11. Nürnberger, T., Nennstiel, D., Jabs, T., Sacks, W. R., Hahlbrock, K., and Scheil, D. (1994) Cell 78, 449–460
12. Austin, B., and Zhang, X. H. (2006) Structure of a Phytophthora Transglutaminase
13. Lorand, L. (2002) Proc. Natl. Acad. Sci. U.S.A. 99, 979–991
14. Murthy, S. N., Iismaa, S., Begg, G., Freymann, D. M., Graham, R. M., and Lorand, L. (2002) Proc. Natl. Acad. Sci. U.S.A. 99, 2738–2742
15. Pinkas, D. M., Strop, P., Brunger, A. T., and Kossel, C. (2007) PLoS Biol. 5, e327
16. Ottmann, C., Luberacki, B., Künfer, I., Koch, W., Brummer, F., Weyand, M., Mattinen, L., Pirhonen, M., Anderluh, G., Seitz, H. U., Nürnberger, T., and Oecking, C. (2009) Proc. Natl. Acad. Sci. U.S.A. 106, 10359–10364
17. Cheng, S., Hu, Y. H., Zhang, M., and Sun, L. (2010) Vaccine 28, 2716–2721
18. Liu, P. C., Lee, K. K., Tu, C. C., and Chen, S. N. (1997) J. Biol. Chem. 272, 443–446
19. Agniswamy, J., Nagiec, M. J., Liu, M., Schuck, P., Musser, J. M., and Sun, P. D. (2006) Structure 14, 225–235
20. Kabsch, W. (1976) Acta Crystallogr. D. 33, 225–235
21. The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC, New York

Structure of a Phytophthora Transglutaminase