Structure, Vol. 12, 677–687, April, 2004, ©2004 Elsevier Science Ltd. All rights reserved. DOI 10.1016/j.str.2004.03.005

Aspergillus niger Protein EstA Defines a New Class of Fungal Esterases within the α/β Hydrolase Fold Superfamily of Proteins

Yves Bourne,[2,*] Alinda A. Hasper,[4] Henri Chahinian,[3] Marianick Juin,[1] Leo H. de Graaff,[4] and Pascale Marchot[1,*]

[1]Ingénierie des Protéines
CNRS UMR-6560
Institut Fédératif de Recherche Jean Roche
Université de la Méditerranée
Faculté de Médecine Secteur Nord
F-13916 Marseille Cedex 20

[2]Architecture et Fonction des Macromolécules Biologiques
CNRS UMR-6098

[3]Laboratoire de Lipolyse Enzymatique
CNRS UPR-9025
31 Chemin Joseph Aiguier
F-13402 Marseille Cedex 20
France

[4]Fungal Genomics Section
Laboratory of Microbiology
Wageningen University
Dreijenlaan 2
NL-6703 HA Wageningen
The Netherlands

Summary

From the fungus *Aspergillus niger*, we identified a new gene encoding protein EstA, a member of the α/β-hydrolase fold superfamily but of unknown substrate specificity. EstA was overexpressed and its crystal structure was solved by molecular replacement using a lipase-acetylcholinesterase chimera template. The 2.1 Å resolution structure of EstA reveals a canonical Ser/Glu/His catalytic triad located in a small pocket at the bottom of a large solvent-accessible, bowl-shaped cavity. Potential substrates selected by manual docking procedures were assayed for EstA activity. Consistent with the pocket geometry, preference for hydrolysis of short acyl/propyl chain substrates was found. Identification of close homologs from the genome of other fungi, of which some are broad host-range pathogens, defines EstA as the first member of a novel class of fungal esterases within the superfamily. Hence the structure of EstA constitutes a lead template in the design of new antifungal agents directed toward its pathogenic homologs.

Introduction

The saprophytic fungus, *Aspergillus niger*, is widely distributed in nature where it grows aerobically on organic matter and is found in soil, litter, and compost. It is also found on decaying plant material where it degrades polysaccharides from plant cell walls into monomeric compounds. Since these polysaccharides mainly consist of cellulose and xylan and form a complex matrix, degradation requires a spectrum of polysaccharidases, e.g., pectinases, cellulases, and hemicellulases (de Vries et al., 2002). The products, which are mainly monosaccharides such as D-glucose or D-xylose, are metabolized by the fungus. In *A. niger*, the transcriptional regulator XlnR controls expression of not only extracellular cellulolytic and xylanolytic enzymes that degrade the cellulose and xylan main chains (van Peij et al., 1998a; van Peij et al., 1998b) but also accessory enzymes such as acetyl xylan and feruoyl esterases that modify the side groups of the polysaccharide backbone (van Peij et al., 1998b). By controlling expression of the intracellular D-xylose reductase, XlnR is also involved in D-xylose metabolism (Hasper et al., 2000).

A. niger is considered as a nonpathogenic fungus, although patients with a history of severe illness or immunosuppressive treatment can favor development of the fungus. In contrast, the related *A. fumigatus* species is an important and opportunistic human pathogen that has been the subject of numerous epidemiological studies. It is the most frequent cause of deadly airborne fungal infections in developed countries known as invasive pulmonary aspergillosis, a mycosis that is fatal in immunocompromised patients (Khan et al., 2003). Fungal pneumonias including aspergillosis, cryptococcosis, candidiasis, coccidioidomycosis, histoplasmosis, and blastomycosis are of the major causes of morbidity and mortality among immunosuppressed hosts. The increasing incidence and severity of invasive mycoses have led to new antifungal strategies aimed at substituting amphotericin B, which is no longer the gold standard for a variety of fungal infections. Approval of voriconazole and caspofungin, which display in vitro and in vivo activities against a variety of molds, recently introduced new treatment options (Gupta and Tomas, 2003; Kartsonis et al., 2003).

In the course of our studies aimed at identifying the protein factors that are under control of the *A. niger* XlnR transcriptional activator, we have found a new gene, *estA*, encoding a putative, 538 residue EstA protein of unknown biological function. Sequence alignments showed that EstA shares the α/β-hydrolase fold common to many lipases and esterases, and suggested the presence of a functional catalytic triad. To get insight into the EstA biological function, we overexpressed, purified, and crystallized EstA. The 2.1 Å resolution crystal structure of EstA confirms its belonging to the α/β-hydrolase fold superfamily of proteins and reveals a Ser210/Glu338/His440 catalytic triad located at the bottom of a large, bowl-shaped cavity with full solvent accessibility. The EstA substrate specificity was investigated both in silico and in solution. Moreover, close homologs of EstA were identified from the unfinished genomes of other fungus species, of which some are broad host-range pathogens. Our structural, biochemical, and computational data show that EstA has a marked specificity toward short acyl chain substrates and identify it as the lead member of a new family of

*Correspondence: marchot.p@jean-roche.univ-mrs.fr (P.M.), yves@afmb.cnrs-mrs.fr (Y.B.)

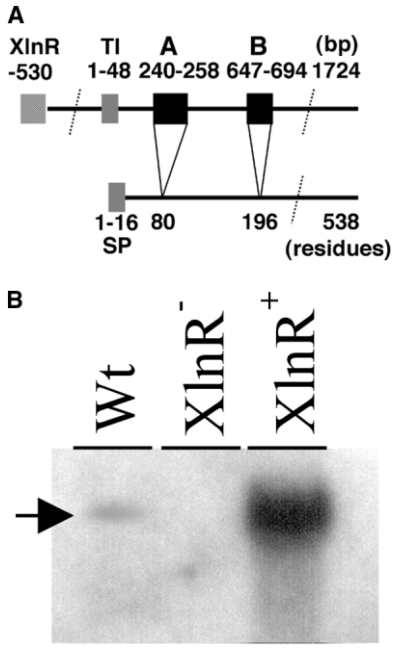

Figure 1. *EstA* Gene Structure and Transcription Pattern

(A) Schematic diagram of the *estA* gene showing the putative XlnR binding site in the promoter, the translation initiation site (TI), the signal peptide (SP) and its encoding sequence, and the two introns, A (59 bp) and B (48 bp).

(B) Northern blot analysis of EstA transcription in the *A. niger* wild-type strain (Wt; the arrow points to the weak band), a XlnR loss-of-function strain ($XlnR^-$), and a *xlnR* multiple copy strain ($XlnR^+$).

fungal esterases within the α/β-hydrolase fold superfamily of proteins.

Results and Discussion

Expression Analysis of the *estA* Gene and Biochemical Characterization of EstA

The *estA* coding region contains two introns within the first 700 bp of the gene (Figure 1). Northern blot analysis of *estA* expression in mycelia of a wild-type *A. niger* strain, a XlnR loss-of-function strain, and a *xlnR* multiple copy strain, all grown on D-xylose, confirmed that *estA* is regulated by the XlnR transcriptional activator. One putative XlnR binding site, with sequence 5′-GGCTAA-3′, was found in the promoter region of *estA* in reverse orientation, 525 bp upstream the *estA* translation start codon.

By placing the *estA* gene under control of the strong endo-xylanase *xlnA* promoter from *A. tubingensis*, we successfully overexpressed a secreted form of EstA in the *A. niger* NW128 strain. The protein was then purified to homogeneity using a limited number of differential ammonium sulfate precipitations and anion exchanges.

The *estA* gene encodes a 538 residue polypeptidic chain with a predicted mass of 58.6 kDa. A PSORT II search for subcellular targeting signals identified an N-terminal signal peptide encompassing the first 16 residues (von Heijne, 1986). This was confirmed by automated Edman sequencing of purified EstA, which unambiguously identified Leu^{17} as the N-terminal residue of the processed protein, yielding a protein of 522 residues with a theoretical mass of ca. 56.9 kDa and theoretical pI of 4.73. Yet, gel filtration and SDS-PAGE analyses, which showed a single peak and a single band, both indicated an apparent mass of ca. 80 kDa, i.e, 40% higher than the calculated mass. Native-PAGE resulted in a diffuse albeit single band, and isoelectric focusing showed three to four intense and two weak bands within the pI 4.6–5.0 range. These observations (data not shown), along with the presence of six consensus sites for Asn-linked glycans in the sequence (cf. below), suggested that EstA is a monomeric, highly glycosylated protein with slight heterogeneity in either the peptidic chain or the glycan moieties, arising from posttranslational modifications.

EstA Is a Member of the α/β-Hydrolase Fold Family of Proteins

BLAST searches using the EstA sequence as a template unambiguously identified EstA as a new member of the α/β hydrolase fold family (Cygler et al., 1993), of which several members have been crystallized. However, attempts to solve the structure of EstA by molecular replacement using, as a template, either of the available α/β-hydrolase protein structures failed. The modular organization of the EstA sequence was therefore analyzed using hydrophobic cluster analysis (HCA) (Callebaut et al., 1997) and compared with HCA plots of the selected highest hits. The highest modular homologies were found with *Torpedo californica* acetylcholinesterase (AChE) (27% sequence identity) and *Geotricum candidum* lipases 1–2 (32%–33% sequence identity), two enzymes of the type B carboxylesterase/lipase subfamily within the α/β-hydrolase fold family, but they were restricted to distinct regions of the HCA plots. This observation prompted us to design a chimera template by connecting the N-terminal two-thirds of the crystalline *G. candidum* lipase molecule to the C-terminal third of the crystalline mouse AChE molecule (cf. Experimental Procedures). This chimera template led to a successful molecular replacement procedure. The final structure has a crystallographic R factor value of 15.3% (R_{free}, 18.0%) in the 30–2.1 Å resolution range, and good stereochemistry (Table 1). Inspection of the final electron density maps allowed us to identify intron A (Figure 1), which had not been predicted in the initial *estA* sequence because of compression artifacts but was then confirmed by cDNA resequencing, and to correct a few uncertainties in the deduced EstA sequence.

As predicted from sequence alignments, EstA shares the canonical α/β hydrolase fold found in esterases and lipases: the EstA molecule consists of a 12-stranded central β sheet surrounded by 13 α helices, with overall dimensions of $55 \times 53 \times 43$ Å (Figure 2). In the crystal, EstA retains the monomeric state observed in solution. Total absence of electron density ahead of the well-ordered residue Ser^{22} suggests a highly disordered N-terminal Leu^{17}-Ala^{21} pentapeptide. The six Cys residues in the sequence are arranged as three disulfide bridges linking residues 85–101, 263–272, and 402–526, the later anchoring the C-terminal helix α_{10} to the protein core. This provides EstA with the same pattern of disul-

Table 1. Data Collection and Refinement Statistics

Data collection[a]	
Beamline (ESRF)	ID14-EH2
Wavelength (Å)	0.933
Resolution range (Å)	30–2.1
Total observations	1,368,385
Unique reflections	104,730
Multiplicity	3.9
Completeness (%)	100 (100)
I/σ(I)	7.6 (2.0)
R_{sym}[b]	6.3 (36.5)
Refinement[c]	
R factor/R_{free} (%)	15.3/18.0
Rmsd[d]	
Bonds (Å)/angles (°)	0.013/1.28
Chiral volume (Å^3)	0.089
Mean B factors[e] (Å)	
Main/side chains	36.1/37.2
Solvent/carbohydrate	46.9/67.6
Cofactors/ion	42.4/22.0
Rmsd on B factors (Å^2)	
Main/side chains	0.57/1.46

[a] Values in parentheses are those for the last shell.
[b] $R_{sym} = |I - <I>|/\Sigma I$, where I is an individual reflection measurement and <I> is the mean intensity for symmetry-related reflections.
[c] R factor $= \Sigma||Fo| - |Fc||/\Sigma|Fo|$, where Fo and Fc are observed and calculated structure factors, respectively. R_{free} is calculated for 2% of randomly selected reflections excluded from refinement.
[d] Root-mean-square deviation from ideal values.
[e] Calculated from 7935 protein atoms, 1149 water and 4 ethylene glycol molecules, 2 sulfate and 1 chloride ions, and 20 glycan moieties.

fide bridges as found in the cholinesterases. A large solvent-accessible bowl-shaped cavity, with dimensions 20 × 23 × 15 Å, hollows one face of the molecule and contains the active site pocket at its base (cf. below). Out of the six potential Asn-type glycosylation sites, five, at positions 72, 138, 259, 345 and 511, are occupied, with a well-ordered complex-type biantennary $GlcNAc_2$-$Man_3GlcNAc_2$ oligosaccharide being linked to Asn^{138}; this is fully consistent with the 23 kDa difference between the EstA calculated and observed masses. The five glycan chains are all located on the same face of the EstA molecule, where they surround the entrance of the active site cavity with Asn^{345} only 20 Å away from the catalytic Ser^{210}, while the nonoccupied Asn^{362} is located on the other face. This suggests that these glycans contribute an active role to the EstA biological function, perhaps in recognizing a yet unidentified partner.

The Structure of EstA Implies an Esterase or a Lipase

At the bottom of the large bowl-shaped cavity, the active site triad, made of residues Ser^{210}, Glu^{338}, and His^{440}, is located in a small pocket 7 Å deep, 15 Å long, and 8 Å wide (Figure 2). EstA presents the so-called nucleophile elbow and the GVSAGA signature motif that characterize members of the α/β hydrolase fold family, and it shares with AChE and the para-nitrobenzyl esterase a rare Glu instead of Asp as the relaying acidic residue in the triad. An ethylene glycol molecule, arising from the cryoprotectant solution used for flash-cooling of the crystals, is bound to the catalytic Ser^{210} where it could mimic part

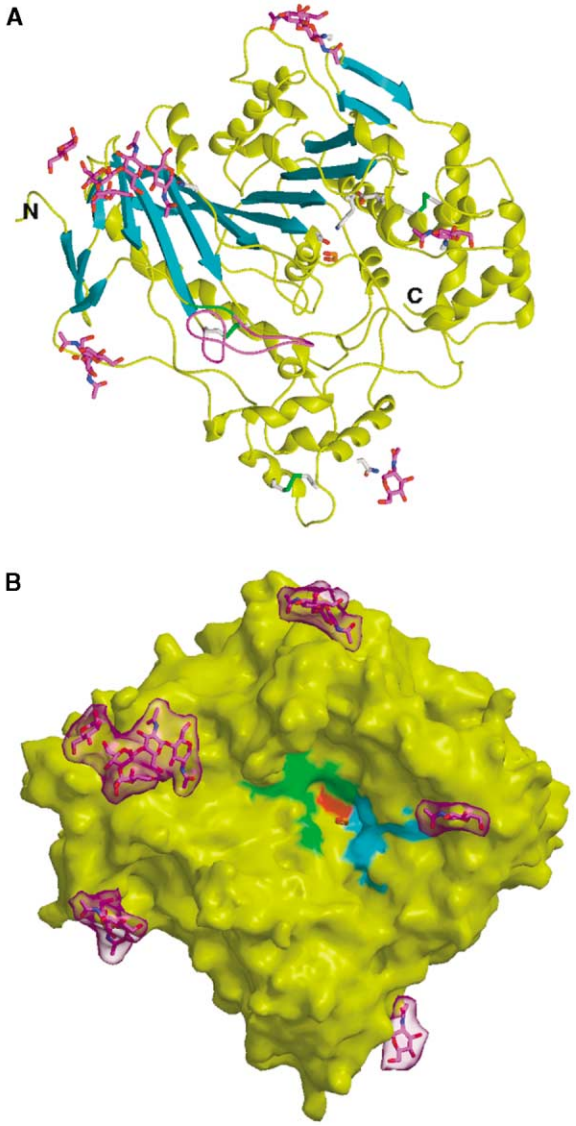

Figure 2. Overall View of the EstA Structure

(A) Ribbon diagram showing the central 12-stranded β sheet in blue and surrounding α helices in yellow. The Asn-linked glycan moieties are displayed in magenta and loop Cys^{85}-Cys^{101} in green. The catalytic triad residues and the ethylene glycol molecule bound to the catalytic Ser^{210} are displayed in white and orange, respectively, at the center of the EstA molecule.

(B) Molecular surface of EstA showing the bowl-shaped active site cavity with the catalytic Ser in red, the aromatic cluster forming the acyl/propyl pocked in blue, and the region containing the conserved Asn^{134} in green. The glycan moieties are shown in magenta under a transparency surface.

of a natural substrate: one of its oxygen atoms occupies the oxyanion hole and establishes hydrogen bonds with the Gly^{127}, Gly^{128}, and Ala^{211} backbone nitrogen atoms (Figure 3).

The active site pocket appears to be made of two subsites: an acyl pocket subsite that contains the catalytic machinery and is lined by an aromatic cluster made of residues Phe^{240}, Trp^{301}, Phe^{342}, and Phe^{400}, and a second subsite made of residues that are mostly polar, as

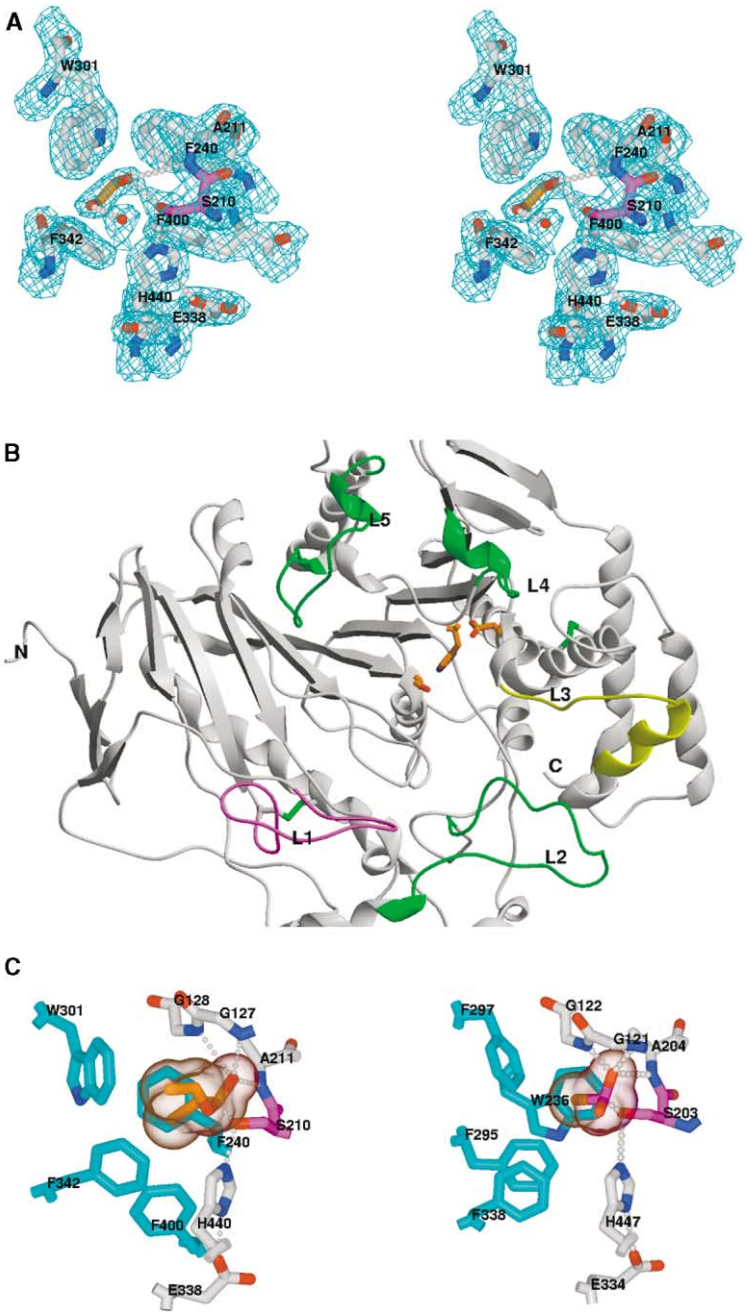

Figure 3. The EstA Active Site Pocket and Catalytic Machinery

(A) Stereoview of the 2.1 Å resolution final 2Fo-Fc electron density map, contoured at 1 σ, showing the ethylene glycol molecule bound to the catalytic Ser^{210}.

(B) Close-up view of the EstA active site pocket with loop Cys^{85}-Cys^{101} (L1) displayed in magenta, loop Phe^{342}-Asn^{359} (L3) in yellow, and loops Thr^{286}-Trp^{301} (L2), Asp^{428}-Ala^{433} (L4), and Phe^{448}-Ala^{468} (L5) in green.

(C) Model of (left) a vinyl propionate in the tetrahedral conformation, bound to the EstA catalytic Ser^{210}, and (right) a soman molecule docked to the catalytic Ser^{203} of mouse AChE (accession code 1MAA), viewed in the same orientation. The aromatic cluster forming the acyl/propyl pocket in EstA and the acyl pocket in AChE are shown in blue.

could be the EstA natural substrate (Figures 2 and 3). The full solvent accessibility of the EstA active site and the lack of a hydrophobic patch at the periphery of this active site pocket largely suggest that EstA functions as true esterase rather than a lipase.

Analysis of electrostatic potentials at the EstA molecular surface revealed a markedly positive potential for the face where the active site cavity opens and a negative potential for the opposite face (data not shown). This overall anisotropic distribution confers EstA a dipole moment of about one-half that of mouse AChE with a vector oriented along the central axis of the active site cavity. Moreover, the negative surface potential appears to form an annular motif that largely surrounds the outer entrance of the active site cavity. Should, in fact, some of the EstA five glycan chains be sialylated, the EstA dipole moment would be even more marked. These observations suggest not only that an electrostatic gradient could guide positively charged substrates or other cations toward the active site, as found for AChE (Ripoll et al., 1993) but also that EstA may belong to the atypical class of electrotactins (Botti et al., 1998), which in the α/β-hydrolase fold family are characterized by a specific surface charge distribution and adhesion properties. The latter would be consistent with the above hypothesized recognition role of the EstA glycans.

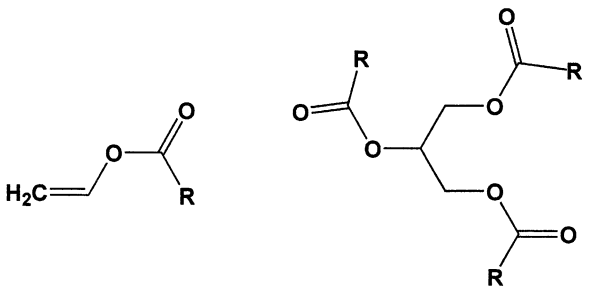

Figure 4. Chemical Structures of the Substrates Hydrolyzed by EstA

The vinyl esters (left) and triacylglycerols (right) differ by the length of the acyl chain, with R = CH_3 (acetate/triacetin), CH_2-CH_3 (proponiate/tripropionin), and $(CH_2)_2$-CH_3 (butyrate/tributyrin).

Substrate Specificity Establishes EstA as an Esterase

To gain insight into the biological function of EstA and establish it as an esterase or a lipase, EstA was assayed on selected substrates (see chemical structures in Figure 4; Table 2). Whereas esterases display maximal activity against solutions of short chain vinyl esters and triacylglycerols at concentrations far below the solubility limit, lipases display maximal activity against emulsified substrates (Chahinian et al., 2002). EstA displays no detectable lipase activity on trioctanoin emulsions or olive oil nor esterase activity on cinnamoyl, feruolyl, or pNP substrates. In contrast, it hydrolyzes vinyl esters and triacylglycerols efficiently, with kinetic parameters consistent with the behavior of an esterase and a substrate specificity similar to that of AChE. Unlike pig liver esterase that clearly prefers vinyl esters to glycerol substrates, EstA has no preference for the nature of the alcohol moiety; instead, it has marked preference for short chain acyl moieties with orders acetate $>$ propionate $>>$ butyrate and triacetin $>>$ tripropionin $>$ tributyrin. Optimal activities were found in the pH 5.0–5.5 range, consistent with the pH requirement for fungal metabolism. These data indicate that the EstA catalytic site cannot accommodate substrates with a vinyl/acyl chain longer than two carbon atoms, and suggest that the Phe^{342} and Phe^{400} side chains confer a hydrophobic environment to the catalytic His^{440}, as often seen for other hydrolases with activities in the low pH range (Mehler et al., 2002).

In the EstA catalytic site we have manually docked a vinyl ester propionate in its tetrahedral configuration so as to mimic the substrate transition state (Figure 3). This model largely supports the EstA specificity for short acyl chains and suggests that its inability to hydrolyze the widely used, small pNP-acetate or other related substrates results from steric hindrance between the bulky phenyl ring and protruding side chains of residues located in the active site region. Overall, our combined biochemical and structural data unambiguously indicate that EstA is a true esterase.

Structural Comparisons

A DALI search for close structural homologs of EstA within a nonredundant set of protein structures from the PDB revealed top-ranked hits (Z score values >35) for the *G. candidum* lipase (Schrag and Cygler, 1993), AChE (Sussman et al., 1991), and para-nitrobenzyl esterase structures (Spiller et al., 1999) (boxed in Figure 6), but only low hits for structures of other fungal esterases such as acetylxylan esterases from *Penicillium purpugenum* (Ghosh et al., 1999) and *Trichoderma reesei* (Hakulinen et al., 2000) and cutinase from *Fusarium solani* (Martinez et al., 1992), although they all belong to the α/β hydrolase fold family.

The root-mean-square deviations between the EstA structure and those of the three top-ranked homologs are 1.42, 1.43, and 1.53 Å for 401, 372, and 395 Cα atoms, respectively. The major structural differences between the four proteins are confined to the shape and dimensions of the bowl-shaped active site cavity, which in EstA is delineated by five loops: loops Cys^{85}-Cys^{101} (L1), Thr^{286}-Trp^{301} (L2), Phe^{342}-Asn^{359} (L3), Asp^{428}-Ala^{433} (L4), and Phe^{448}-Ala^{468} (L5) (Figure 5). In fact, the wide-open cavity in EstA results from the drastic shortening of loop L1, which connects strands b_3 and β_2 on one side of the cavity. Hence EstA lacks the large structural element, i.e., the two-helical lid in the lipases and the homologous long Ω-loop in AChE, which restricts assess to the catalytic site. In addition, the distinctive positionings for the tips of the four other loops, L2–L5, that surround the active site cavity, also contribute a solvent-accessible catalytic triad in EstA. The aromatic patch of four hydrophobic side chains that lines the EstA active site closely resembles the patch made of Phe^{295}, Phe^{297}, Trp^{236}, and Phe^{338} (mouse AChE numbering) that lines the acyl pocket in AChE. This again argues for a short alkyl chain ligand of EstA (Figure 3).

Table 2. Comparative Kinetic Parmeters of *A. niger* EstA, Pig Liver Esterase, and Eel Acetylcholinesterase for Selected Substrates

	Vinyl Esters[a,b]									Triacylglycerols[a,b]								
	Acetate			Propionate			Butyrate			Triacetin			Tripropionin			Tributyrin		
Enzyme	k_{cat}	K_m	k_{cat}/K_m	k_{cat}	K_m	k_{cat}/K_m	k_{cat}	K_m	K_{cat}/K_m	k_{cat}	K_m	k_{cat}/K_m	k_{cat}	K_m	k_{cat}/K_m	k_{cat}	k_{cat}/K_m	k_{cat}/K_m
EstA	400	2.7	148	160	2.9	55	25	ND	—	265	15	18	60	ND	—	20	ND	—
Esterase	315	4.0	79	295	3.0	98	465	2	232	60	3	20	50	0.3	167	70	ND	—
AChE	1300	10.0	130	280	28.5	10	0	0	—	600	18	33	0	0	—	0	—	—

No lipase activity on olive oil and trioctanoin or esterase activity on cinnamoyl, feruolyl, pNP-acetate, -palmitate, and -stearate were detected.

[a] cf. chemical structures in Figure 4.

[b] The k_{cat} and K_m values are expressed in s^{-1} and mM, respectively. ND, not determined.

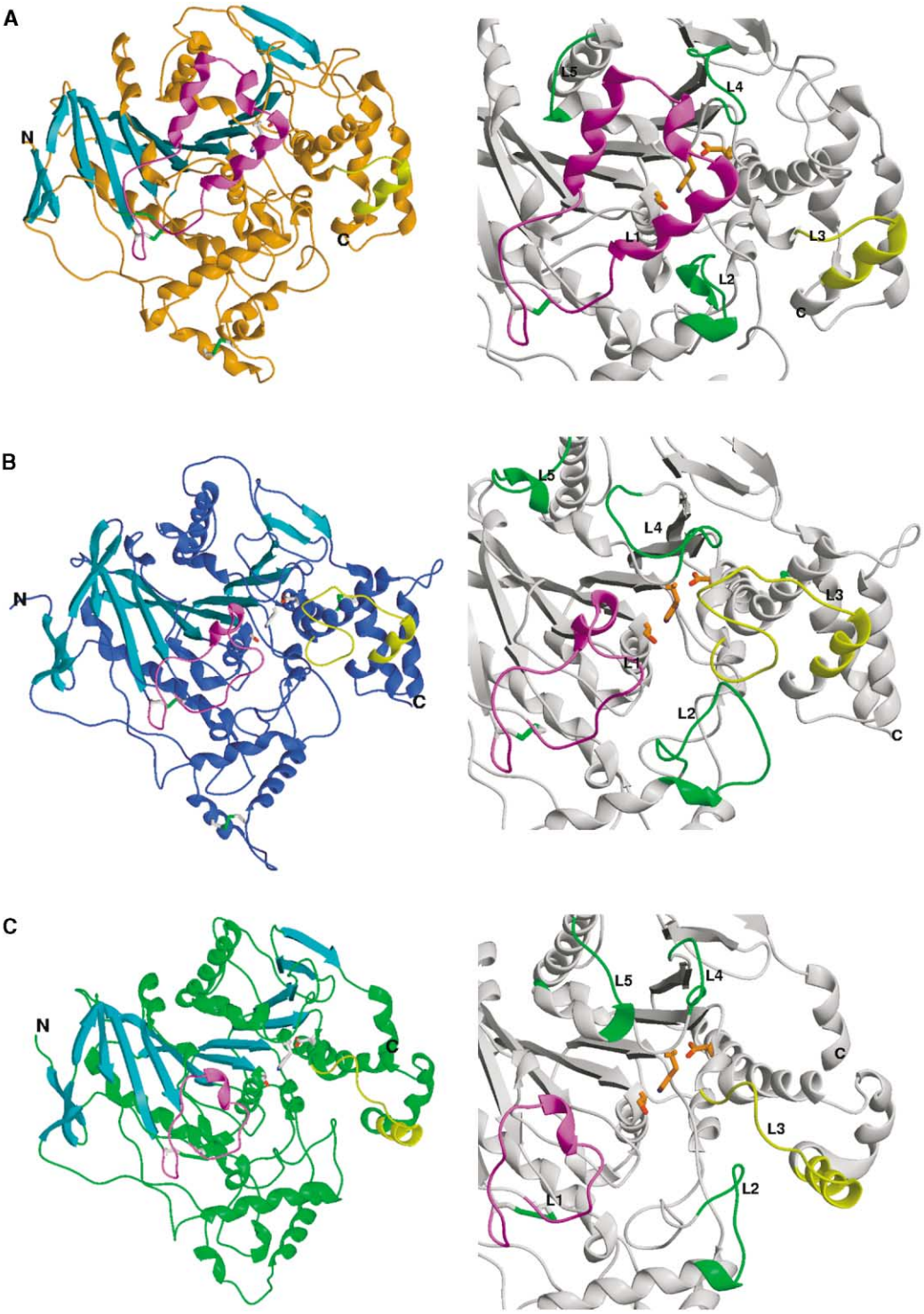

Figure 5. Structural Comparisons

Ribbon diagrams (left) of (A) the *G. candidum* lipase (accession code 1THG), (B) mouse AChE (1MAA), and (C) para-nitrobenzoyl esterase (1C7I), viewed in the same orientation as EstA in Figure 2; and close-up views (right) of the active site pockets in these three molecules, viewed as EstA in Figure 3B. The central β sheet is displayed in cyan with catalytic triad residues in white (left) or orange (right). The lipase large lid motif (in [A]), AChE Ω-loop (in [B]), and homologous esterase loop (in [C]), which occlude access to the active site in these three molecules and correspond to EstA loop L1, are shown in magenta. The surface loop region that connects helix $\alpha^1_{7,8}$ to helix $\alpha^2_{7,8}$ and corresponds to EstA loop L3 is shown in yellow. The surrounding three loops that correspond to EstA loops L2, L4, and L5 are colored in green.

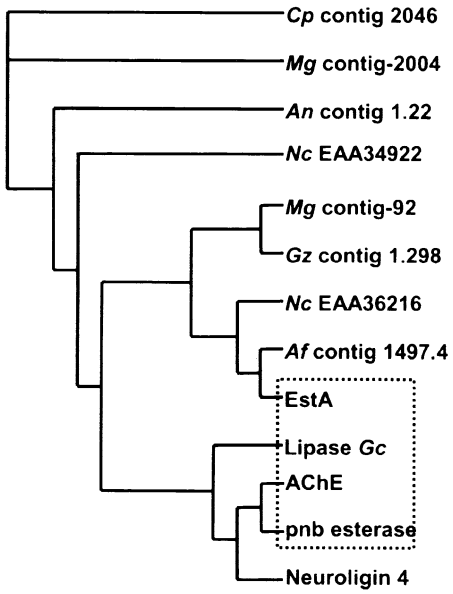

Figure 6. Phylogeny of EstA and Homologous Proteins

Dendrogram of EstA and its sequence-based homologs; the three EstA structural homologs are boxed; the neuroligins are adhesion molecules that belong to the electrotactin class within the α/β-hydrolase fold family of proteins. The branching order and distance score were calculated with TREEVIEW (Page, 1996).

EstA Homologs in Pathogen Fungus Species

BLAST searches using the EstA sequence against a database of unfinished eukaryotic genomes revealed six close homologs (e values in the e^{-100} range) in species *A. nidulans* (contig_1.22), *A. fumigatus* (contig_1497.4), *Coccidioides posadasii* (contig_2046), *Gibberella zeae* (contig_1.298), and *Magnaporthe grisea* (contig_2.92 and contig_2.2004) (Figures 6 and 7). With the exception of *A. nidulans*, all are pathogen species: *A. fumigatus* and *C. posadasii* are respiratory pathogens respectively responsible for invasive pulmonary aspergillosis (Khan et al., 2003) and recurrent epidemics of San Joaquin Valley fever (coccidioidomycosis) in desert regions of the southwestern United States (Pappagianis, 1994); *G. zeae* is a broad host-range pathogen that infects many crops, including wheat and barley, and causes head blight and rot diseases throughout the world (Miller, 2001); and *M. grisea* is the causal agent of rice blast disease, one of the most devastating threats to food security worldwide (Howard and Valent, 1996). A similar search against a database of the proteins predicted from the recently finished *Neurospora crassa* genome identified two additional EstA homologs, EAA36216 and EAA34922 (Galagan et al., 2003). The hypothetical *M. grisea* contig_2004 and *N. crassa* EAA36216 protein sequences contain, before strand b_1, an additional N-terminal region of ~80–95 residues with a predicted transmembrane helix, while that of *N. crassa* EAA34922 is extended by ~55 residues after helix α_{10} in the C-terminal region. However, alignment analysis undoubtedly reveals that all eight hypothetical proteins are close homologs of EstA, with sequence identities in the 45%–50% range.

Topographical analysis of the invariant residues indicates that most of these residues are likely to be involved in salt bridge formation and/or stabilization of the protein core (Figure 7). The pattern of disulfide bridges and the glycosylation site at Asn^{138} (EstA numbering) are well conserved. The catalytic triad and the aromatic patch of residues that characterize the acyl/propyl pocket in EstA are also well conserved, except in *N. crassa* EAA34922 (cf. below). In the sequences of four homologs, the lack of one of the four aromatic residues that line the acyl/propyl pocket, corresponding to Phe^{240} in EstA, seems to be compensated by a Thr-Arg substitution at the neighboring position (residue 399 in EstA), resulting in minor distortions of the shape of the pocket. In contrast, the pattern of residues conservation within the active site does not reveal any particular biological role except for the conserved EstA Asn^{134} that may be involved in substrate recognition. Overall, these proteins exemplify a new structural family of fungal esterases that presumably evolved from a common ancestor and diverged from the other structural homologs of the α/β hydrolase fold family. These fungal esterases, which are closely related in sequence, may also share common catalytic properties and, in turn, similar biological functions.

Only *N. crassa* protein EAA34922 appears to be lacking the hydrolase function due to a Gly substitution to the catalytic Glu residue (Figure 7). This feature is reminiscent of the electrotactins and Ndr proteins, which share the α/β hydrolase fold but respectively lack the catalytic nucleophile residue, and the catalytic nucleophile and His residues (Botti et al., 1998; Shaw et al., 2002). Therefore, members of this atypical class of α/β hydrolase proteins that have gained heterologous recognition properties can also be identified in nonpathogenic fungi, where they may have important biological functions.

In summary, in the *A. niger* genome we have identified a new *estA* gene controlled by the XlnR transcriptional activator and encoding a novel member of the α/β-hydrolase fold family of proteins. To get insights into the EstA biological function, we have undertaken a multipronged approach combining structural, biochemical, and bioinformatics analyses. We show that EstA is the first member of a novel class of fungal esterases that also contains predicted homologs from other fungus species, of which some are broad host-range pathogens. Hence the structure of EstA constitutes a lead template in the design of new antifungal agents directed toward these pathogenic homologs.

Experimental Procedures

Cloning, Gene Structure, and Expression Analysis of the *estA* Gene

Screening of a differential library (Hasper et al., 2000), constructed by subtracting the cDNA of a D-xylose induced XlnR loss-of-function strain from the cDNA of a D-xylose induced *xlnR* multiple copy strain (van Peij et al., 1998a), resulted in isolation of three RsaI fragments, *3R22*, *3R24*, and *3R59*, belonging to the same gene. The complete gene, designated *estA*, was isolated using RsaI fragment *3R22* to probe an *A. niger* genomic library. A 3.8 kb BamHI fragment was obtained from a positive λ-phage and cloned in pUC19, resulting in pIM4420. Propagation of plasmids and λ-phages used *E. coli* strains DH5α and LE392. Sequencing reactions were performed using the Thermo-Sequenase fluorescent-labeled primer cycle sequencing kit

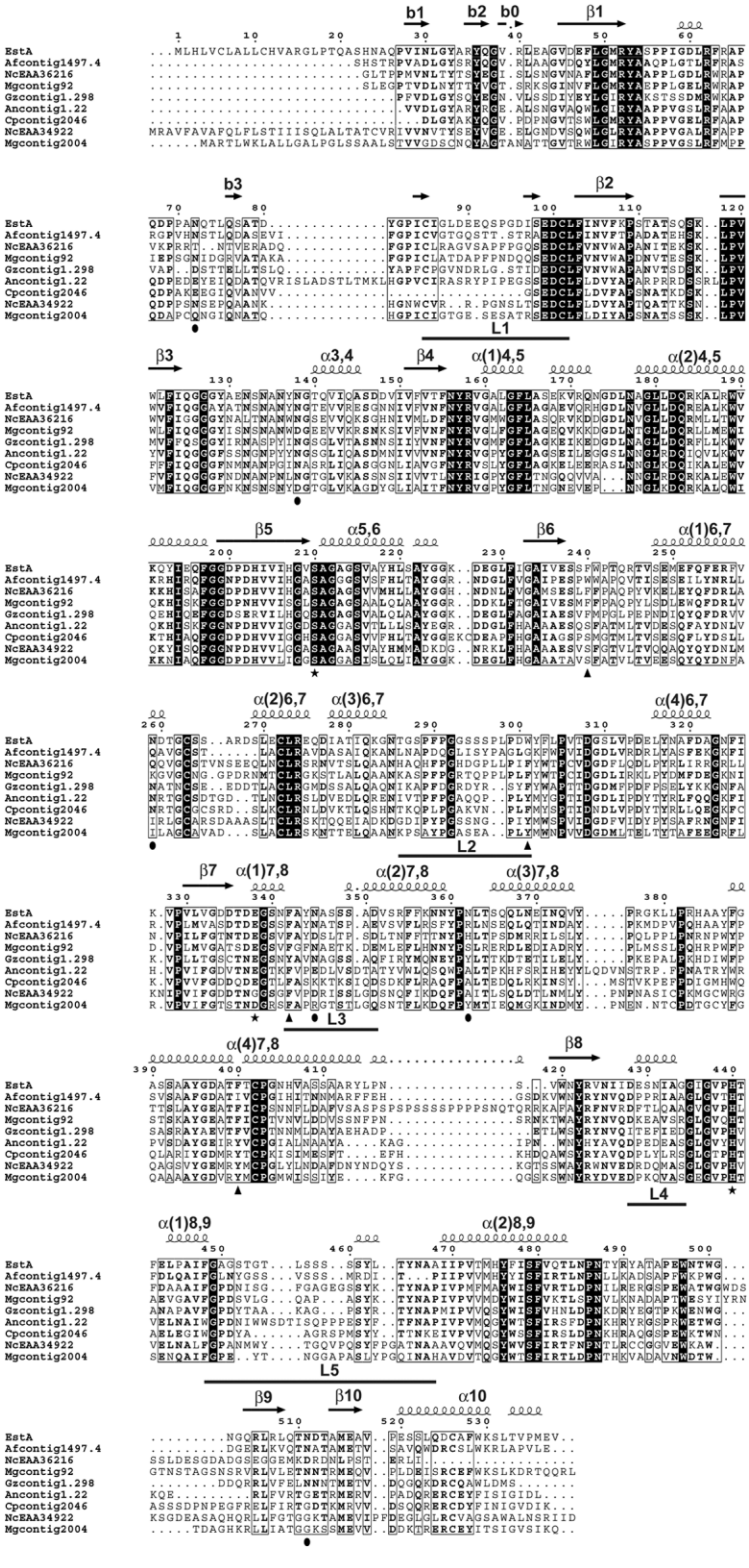

Figure 7. Sequence Conservation between EstA and Its Homologs from Other Fungus Species

Alignment of the EstA primary sequence with those of homologous proteins predicted from the unfinished genomes of *A. fumigatus*, *A. nidulans*, *F. graminearum* (teleomorph *G. zeae*), and *C. posadasii* and from the genome of *N. crassa* genome. The catalytic triad residues are indicated by stars, the putative Asn-linked glycosylation sites by circles, and residues in the aromatic cluster that forms the acyl/propyl pocket by triangles. The EstA secondary structure elements are indicated above the sequence and numbered according to (Cygler et al., 1993).

(Amersham Biosciences) and universal sequencing primers, and using the Cy5 Autoread sequencing kit (Amersham Biosciences) and gene-specific oligonucleotides. The sequences were analyzed on an ALF*express* sequencer (Amersham Biosciences). The *estA* nucleotide sequence has been submitted to the GenBank database, accession number AY456379.

Southern and Northern Blot Analysis

Fungal genomic DNA was isolated from grinded mycelium as previously described (de Graaff et al., 1988). Southern blot analysis was performed according to standard methods (Sambrook et al., 1989) using the VacuGene XL Vacuum blotting system (Amersham Biosciences). For Northern blot analysis, RNA was separated on a

1.6% (w/v) agarose gel in 10 mM sodium phosphate (pH 7.0) and, after capillary blotting to Hybond-N (Amersham Biosciences), hybridized in 50% (v/v) formamide, 10% (w/v) dextran sulfate, 6xSSC, 0.2% (w/v) ficoll, 0.2% (w/v) polyvinylpyrrolidone, 0.2% (w/v) bovine serum albumin, 0.1% (w/v) SDS, and 100 μg ml^{-1} single-stranded herring sperm DNA (Life Technologies). The Hybond membrane was washed at 68°C with 3 mM trisodium citrate, 30 mM NaCl, 0.1% SDS.

Expression and Purification of Protein EstA

The *estA* coding region along with a 3′ noncoding flanking region were fused to the *A. tubingensis xlnA* gene promoter (de Graaff et al., 1994) at its translation start codon using splicing by overlap extension (SOE), resulting in pIM4421. *A. niger* strain NW128 (Witteveen et al., 1997) was transformed into NW128::pIM4421-15 as previously described (Kusters-van Someren et al., 1991). The latter was precultured overnight using 2 liters of Minimal Medium (Pontecorvo et al., 1953) containing 100 mM D-fructose as a carbon source. The mycelium was transferred to 1.4 liters of Minimal Medium containing 50 mM D-xylose for 16 hr (Hasper et al., 2000); it was then removed by filtration and pH of the filtrate was adjusted to a value of 6. The filtrate was subjected to three steps of differential precipitation using ammonium sulfate at 50%, 75%, and 90% saturation. After each step, the precipitated proteins were collected by centrifugation (10 krpm; 20 min) and either resuspended in 10 mM piperazine, pH 6.0, (first two steps) or dialyzed against 10 mM sodium phosphate, pH 6.0 (overnight, 4°C) (third step). The EstA fractions were selected by SDS-PAGE (Coomassie brilliant blue staining) and further purified by successive anion-exchanges on Source 30Q and MonoQ (Amersham Biosciences) using gradients of 0–400 mM NaCl in 10 mM piperazine, pH 6.0, and intermediate desalting by dialysis against 10 mM sodium phosphate, pH 6.0 (overnight, 4°C). This procedure yielded pure EstA, as assessed by SDS- and native-PAGE analyses (cf. below). EstA was dialyzed against 1 mM 2-(*N*-morpholino) ethanesulfonic acid, pH 6.5, 50 mM NaCl, 0.01% (w/v) NaN_3 (crystallization buffer) and concentrated to ca. 10 mg ml^{-1} by ultrafiltration (Centricon Units, Amicon); it was stored on ice.

Biochemical Characterization of Purified EstA

SDS-PAGE and native-PAGE on homogenous 7.5% or 12% gels and isoelectric focusing on pI 4–6.5 gels used a PhastSystem apparatus and silver nitrate staining according to the manufacturer's instructions (Amersham Biosciences). The SDS-PAGE samples were boiled (5 min) in the presence of 2.5% (w/v) SDS with (reducing conditions) and without (nonreducing conditions) 5% (v/v) β-mercaptoethanol. Gel filtration was performed on Superdex-200 in the crystallization buffer using a FPLC apparatus and calibration markers from the LMW and HMW kits (Amersham Biosciences). Automated Edman analysis of the N-terminal ten residues was carried out on a gas-phase microsequencer (Applied Biosystems 476A).

Crystallization and Data Collection

Crystallization was achieved at 4°C by vapor diffusion using hanging drops (1–2 μl) and a protein-to-well solution ratio of 1:1 with PEG-6K 15% (v/v) in 100 mM ammonium sulfate, pH 6.0, as the well solution. The crystals were flash cooled in the nitrogen gas stream (100 K) after successive short soaks with 5%, 10%, and 20% ethylene glycol in the well solution as cryoprotectant. The crystals belonged to the hexagonal space group $P6_3$ with unit cell dimensions a = b = 167.5 Å and c = 112.8 Å, giving a Vm value of 2.86 Å^3/Da (56.5% solvent) for two 80 kDa molecules in the asymmetric unit (Matthews, 1968). Oscillation images were integrated with DENZO (Otwinowski and Minor, 1997), data were scaled and merged with SCALA (CCP4, 1994), and amplitude structure factors were generated with TRUNCATE (CCP4, 1994).

Structure Determination and Refinement

Sequence homology searches within the PDB database and within the ESTHER database for the α/β-hydrolase fold family of proteins (http://bioweb.ensam.inra.fr/ESTHER/general?what=index) (Hotelier et al., 2004) were carried out using PSI- and PHI-BLAST and BLASTP programs, respectively. The EstA structure was solved by the molecular replacement method with AMoRe (Navaza, 1994) using, as search model, the coordinates from a chimera template designed by connecting the N-terminal residues 1–395 of the crystalline *G. candidum* lipase (accession code: 1THG [Schrag and Cygler, 1993]) to the C-terminal residues 384–543 of the crystalline mouse acetylcholinesterase (AChE) (accession code: 1MAA [Bourne et al., 1999]). This procedure yielded a correlation coefficient of 19% and an R factor value of 49.7% in the 15–4 Å resolution range. 1005 residues out of 1028 (98% of the polypeptide chain) were built without any manual intervention and refined to R factor and R_{free} values of 20.3% and 28.5%, respectively, using ARP/wARP (Perrakis et al., 1999). Final refinement stages of the ARP/wARP model were then performed with REFMAC (Murshudov et al., 1997) using data between 30 and 2.1 Å; the resulting SigmaA-weighted 2Fo-Fc and Fo-Fc electron density maps were used to identify errors in the initial sequence and to correct the model with the graphics program TURBO-FRODO (Roussel and Cambillau, 1989).

The final EstA structure comprises residues Ser^{22}/Gln^{26} to Val^{538} for each of the 2 molecules in the asymmetric unit, 2 incomplete biantennary N-linked glycan chains, $GlcNAc_2Man_3GlcNAc_2$, 2 $GlcNAc_2$ disaccharides, and 2 GlcNAc moieties, 1149 solvent molecules, 5 ethylene glycol molecules, and 2 sulfate and 1 chloride ions. The average root-mean-square deviation between the two molecules in the asymmetric unit is 0.15 Å for 513 Cα atoms. The stereochemistry of the structure was analyzed with PROCHECK (Laskowski et al., 1993); with the exception of the putative catalytic Ser^{210}, no residues were found in the disallowed regions of the Ramachandran plot.

Electrostatic surface potentials were calculated with GRASP (Nicholls et al., 1991) using a salt concentration of 100 mM, inner and outer dielectric constants of 2 and 80, respectively, and an average charge on His residues of 0; comparative calculations on mouse AChE (Bourne et al., 1999) used the same parameters. Figures were generated with ESPript (Gouet et al., 1999), SPOCK (Christopher, 1998), and Raster3D (Merritt and Bacon, 1997).

Enzyme Assays

Esterase activities on acyl *p*-nitrophenyl (pNP) substrates (0.25 mM in aqueous solutions) were carried out in 25 mM citrate buffer, pH 4.5, or in 25 mM phosphate buffer, pH 6.0; the reaction was stopped and the pH brought to 9–9.5 with Na_2CO_3, and the released phenol was quantified by spectrophotometry (λ = 405 nm). For palmitate and stearate esters, not soluble in water, stock solutions (25 mM) were in acetonitrile. The lipase activities and the cinnamoyl and feruolyl esterase activities were assayed as described by Sommer et al. (1997) and Donaghy et al. (1998), respectively. Hydrolysis of carboxylic ester solutions and emulsions was followed potentiometrically with a pH stat (TTT 80 Radiometer, Copenhagen, DM), at 25°C and in 30 ml of 2.5 mM Tris, pH 5.0, 0.1 M NaCl, and using EstA concentrations in the 0.1–0.5 μg ml^{-1} range (1–6 nM, based on a 80 kDa mass). Standard conditions were used for measuring enzyme activities at increasing concentrations of vinyl esters and triacylglycerols (Chahinian et al., 2000). No detergent or emulsifier was added to measure activity against oversaturated solutions of the carboxylic esters. Released fatty acids were titrated with 0.1 M NaOH. Correction was made for partial dissociation, at pH 5.0 and 25°C, of acetic, propionic, or butyric acid (respective pKa values: 4.75, 4.87, and 4.87). The catalytic constant (k_{cat}) values were calculated from the maximal velocities (V_m) values recorded at optimal substrate concentration, using masses of 80 kDa for EstA and of 60 and 80 kDa for pig liver esterase and *Electrophorus electricus* AChE (Sigma-Aldrich). The Michaelis constant (K_m) values were determined graphically from double reciprocal Lineweaver-Burk plots of the kinetic data.

Bioinformatics

Predicted sequences of the EstA homologs were identified using BLAST searches within the Fungal Genome Initiative (FGI) database of the Whitehead Institute/Massachusetts Institute of Technology (MIT) Center for Genome Research (*N. crassa, M. grisea, A. nidulans*, and *F. graminearum*) or from the eukaryotic genome database of the National Center for Biotechnology Information (NCBI) (*A. fumigatus* and *C. posadasi*). The deduced protein sequences were aligned using MULTALIGN (Corpet, 1988).

Acknowledgments

We are grateful to Bernard Henrissat (AFMB) who initiated this project; Ester Dekkers (Laboratory of Microbiology) for assistance in expression and purification; the ID14-EH2 staff of the European Synchrotron Radiation Facility (ESRF) for expert assistance during data collection; Pascal Mansuelle and Ludovic Renault (Laboratoire d'Ingénierie des Protéines) for N-terminal sequencing and calculation of surface potentials, respectively; Robert Verger (Laboratoire de Lipolyse Enzymatique, Marseille), Louis Sarda (Laboratoire de Biochimie, Marseille), Palmer Taylor (Dept. of Pharmacology, UCSD, La Jolla, CA) and Arnaud Chatonnet (Dept. de Physiologie Animale, INRA, Montpellier) for fruitful discussions.

Received: December 11, 2003
Revised: January 22, 2004
Accepted: January 23, 2004
Published: April 6, 2004

References

Botti, S.A., Felder, C.E., Sussman, J.L., and Silman, I. (1998). Electrotactins: a class of adhesion proteins with conserved electrostatic and structural motifs. Protein Eng. *11*, 415–420.

Bourne, Y., Taylor, P., Bougis, P.E., and Marchot, P. (1999). Crystal structure of mouse acetylcholinesterase. A peripheral site-occluding loop in a tetrameric assembly. J. Biol. Chem. *274*, 2963–2970.

Callebaut, I., Labesse, G., Durand, P., Poupon, A., Canard, L., Chomilier, J., Henrissat, B., and Mornon, J.P. (1997). Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. Cell. Mol. Life Sci. *53*, 621–645.

CCP4 (Collaborative Computational Program 4) (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. *50*, 760–763.

Chahinian, H., Nini, L., Boitard, E., Dubes, J.P., Sarda, L., and Comeau, L.C. (2000). Kinetic properties of *Penicillium cyclopium* lipases studied with vinyl esters. Lipids *35*, 919–925.

Chahinian, H., Nini, L., Boitard, E., Dubes, J.P., Comeau, L.C., and Sarda, L. (2002). Distinction between esterases and lipases: a kinetic study with vinyl esters and TAG. Lipids *37*, 653–662.

Christopher, J.A. (1998). SPOCK: The Structural Properties Observation and Calculation Kit Program Manual (College Station, TX: The Center for Macromolecular Design, Texas A&M University).

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. *16*, 10881–10890.

Cygler, M., Schrag, J.D., Sussman, J.L., Harel, M., Silman, I., Gentry, M.K., and Doctor, B.P. (1993). Relationship between sequence conservation and three-dimensional structure in a large family of esterases, lipases, and related proteins. Protein Sci. *2*, 366–382.

de Graaff, L., van den Broek, H., and Visser, J. (1988). Isolation and transformation of the pyruvate kinase gene of *Aspergillus nidulans*. Curr. Genet. *13*, 315–321.

de Graaff, L.H., van den Broeck, H.C., van Ooijen, A.J., and Visser, J. (1994). Regulation of the xylanase-encoding *xlnA* gene of *Aspergillus tubigensis*. Mol. Microbiol. *12*, 479–490.

de Vries, R.P., Benen, J.A.E., de Graaff, L.H., and Visser, J. (2002). Plant cell wall degrading enzymes produced by *Aspergillus niger*. In Mycota, Industrial Applications, H.D. Osiewacz, ed. (Berlin Heidelberg: Sringer Verlag), pp. 263–279.

Donaghy, J., Kelly, P.F., and McKay, A.M. (1998). Detection of ferulic acid esterase production by *Bacillus* spp., and *lactobacilli*. Appl. Microbiol. Biotechnol. *50*, 257–260.

Galagan, J.E., Calvo, S.E., Borkovich, K.A., Selker, E.U., Read, N.D., Jaffe, D., FitzHugh, W., Ma, L.J., Smirnov, S., Purcell, S., et al. (2003). The genome sequence of the filamentous fungus *Neurospora crassa*. Nature *422*, 859–868.

Ghosh, D., Erman, M., Sawicki, M., Lala, P., Weeks, D.R., Li, N., Pangborn, W., Thiel, D.J., Jornvall, H., Gutierrez, R., et al. (1999). Determination of a protein structure by iodination: the structure of iodinated acetylxylan esterase. Acta Crystallogr. D Biol. Crystallogr. *55*, 779–784.

Gouet, P., Courcelle, E., Stuart, D.I., and Metoz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics *15*, 305–308.

Gupta, A.K., and Tomas, E. (2003). New antifungal agents. Dermatol. Clin. *21*, 565–576.

Hakulinen, N., Tenkanen, M., and Rouvinen, J. (2000). Three-dimensional structure of the catalytic core of acetylxylan esterase from *Trichoderma reesei*: insights into the deacetylation mechanism. J. Struct. Biol. *132*, 180–190.

Hasper, A.A., Visser, J., and de Graaff, L.H. (2000). The *Aspergillus niger* transcriptional activator XlnR, which is involved in the degradation of the polysaccharides xylan and cellulose, also regulates D-xylose reductase gene expression. Mol. Microbiol. *36*, 193–200.

Hotelier, T., Renault, L., Cousin, X., Marchot, P., and Chatonnet, A. (2004). ESTHER, the database of the α/β-hydrolase fold superfamily of proteins. Nucleic Acids Res. *32*, D145–D147.

Howard, R.J., and Valent, B. (1996). Breaking and entering: host penetration by the fungal rice blast pathogen *Magnaporthe grisea*. Annu. Rev. Microbiol. *50*, 491–512.

Kartsonis, N.A., Nielsen, J., and Douglas, C.M. (2003). Caspofungin: the first in a new class of antifungal agents. Drug Resist. Updat. *6*, 197–218.

Khan, A.N., Jones, C., and Macdonald, S. (2003). Bronchopulmonary aspergillosis: a review. Curr. Probl. Diagn. Radiol. *32*, 156–168.

Kusters-van Someren, M.A., Harmsen, J.A., Kester, H.C., and Visser, J. (1991). Structure of the *Aspergillus niger pelA* gene and its expression in *Aspergillus niger* and *Aspergillus nidulans*. Curr. Genet. *20*, 293–299.

Laskowski, R., MacArthur, M., Moss, D., and Thornton, J. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. *26*, 283–291.

Martinez, C., De Geus, P., Lauwereys, M., Matthyssens, G., and Cambillau, C. (1992). *Fusarium solani* cutinase is a lipolytic enzyme with a catalytic serine accessible to solvent. Nature *356*, 615–618.

Matthews, B.W. (1968). Solvent content of protein crystals. J. Mol. Biol. *33*, 491–497.

Mehler, E.L., Fuxreiter, M., Simon, I., and Garcia-Moreno, E.B. (2002). The role of hydrophobic microenvironments in modulating pKa shifts in proteins. Proteins *48*, 283–292.

Merritt, E., and Bacon, D. (1997). Raster3D: photorealistic molecular graphics. Methods Enzymol. *277*, 505–524.

Miller, J.D. (2001). Factors that affect the occurrence of fumonisin. Environ. Health Perspect. Suppl. 2 *109*, 321–324.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. *53*, 240–255.

Navaza, J. (1994). AMoRe: an automated package for molecular replacement. Acta Crystallogr. A *50*, 157–163.

Nicholls, A., Sharp, K.A., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins *11*, 281–296.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Page, R.D. (1996). TreeView: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. *12*, 357–358.

Pappagianis, D. (1994). Marked increase in cases of coccidioidomycosis in California: 1991, 1992, and 1993. Clin. Infect. Dis. Suppl. 1 *19*, S14–S18.

Perrakis, A., Morris, R., and Lamzin, V.S. (1999). Automated protein model building combined with iterative structure refinement. Nat. Struct. Biol. *6*, 458–463.

Pontecorvo, G., Roper, J.A., Hemmons, L.M., Macdonald, K.D., and Bufton, A.W. (1953). The genetics of *Aspergillus nidulans*. Adv. Genet. *5*, 141–238.

Ripoll, D.R., Faerman, C.H., Axelsen, P.H., Silman, I., and Sussman, J.L. (1993). An electrostatic mechanism for substrate guidance down

the aromatic gorge of acetylcholinesterase. Proc. Natl. Acad. Sci. USA *90*, 5128–5132.

Roussel, A., and Cambillau, C. (1989). TURBO-FRODO (Mountain View, California: Silicon Graphics).

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Molecular cloning: A Laboratory Manual, Second Edition (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Schrag, J.D., and Cygler, M. (1993). 1.8 Å refined structure of the lipase from *Geotrichum candidum*. J. Mol. Biol. *230*, 575–591.

Shaw, E., McCue, L.A., Lawrence, C.E., and Dordick, J.S. (2002). Identification of a novel class in the alpha/beta hydrolase fold superfamily: the N-myc differentiation-related proteins. Proteins *47*, 163–168.

Sommer, P., Bormann, C., and Gotz, F. (1997). Genetic and biochemical characterization of a new extracellular lipase from *Streptomyces cinnamomeus*. Appl. Environ. Microbiol. *63*, 3553–3560.

Spiller, B., Gershenson, A., Arnold, F.H., and Stevens, R.C. (1999). A structural view of evolutionary divergence. Proc. Natl. Acad. Sci. USA *96*, 12305–12310.

Sussman, J.L., Harel, M., Frolow, F., Oefner, C., Goldman, A., Toker, L., and Silman, I. (1991). Atomic structure of acetylcholinesterase from *Torpedo californica*: a prototypic acetylcholine-binding protein. Science *253*, 872–879.

van Peij, N.N., Gielkens, M.M., de Vries, R.P., Visser, J., and de Graaff, L.H. (1998a). The transcriptional activator XlnR regulates both xylanolytic and endoglucanase gene expression in *Aspergillus niger*. Appl. Environ. Microbiol. *64*, 3615–3619.

van Peij, N.N., Visser, J., and de Graaff, L.H. (1998b). Isolation and analysis of xlnR, encoding a transcriptional activator co-ordinating xylanolytic expression in Aspergillus niger. Mol. Microbiol. *27*, 131–142.

von Heijne, G. (1986). A new method for predicting signal sequence cleavage sites. Nucleic Acids Res. *14*, 4683–4690.

Witteveen, C.F.B., van de Vondervoort, P.J.I., van den Broeck, H.C., van Engelenburg, F.A.C., de Graaff, L.H., Hillebrand, M.H.B.C., Schaap, P.J., and Visser, J. (1997). Induction of glucose oxidase, catalase, and lactonase in *Aspergillus niger*. Curr. Genet. *24*, 408–416.

Accession Numbers

The atomic coordinates and structure factors of the EstA structure have been deposited with the Protein Data Bank, accession code 1UKC.