Homology modeling and identification of serine 160 as nucleophile of the active site in a thermostable carboxylesterase from the archaeon Archaeoglobus fulgidus

G. Manco¹, L. Camardella, F. Febbraio, G. Adamo, V. Carratore and M. Rossi

Istituto di Biochimica delle Proteine ed Enzimologia, Via Marconi 10, 80125 Naples, Italy

¹To whom correspondence should be addressed.
E-mail: manco@dafne.ibpe.na.cnr.it

The hyperthermophilic Archaeon Archaeoglobus fulgidus has a gene (AF1763) which encodes a thermostable carboxylesterase belonging to the hormone-sensitive lipase (HSL)-like group of the esterase/lipase family. Based on secondary structure predictions and a secondary structure-driven multiple sequence alignment with remote homologous proteins of known three-dimensional structure, we previously hypothesized for this enzyme the α/β-hydrolase fold typical of several lipases and esterases and identified Ser160, Asp255 and His285 as the putative members of the catalytic triad. In this paper we report the building of a 3D model for this enzyme based on the structure of the homologous brefeldin A esterase from Bacillus subtilis whose structure has been recently elucidated. The model reveals the topological organization of the fold corroborating our predictions. As regarding the active-site residues, Ser160, Asp255 and His285 are located close each other at hydrogen bond distances. The catalytic role of Ser160 as the nucleophilic member of the triad is demonstrated by the [3H]diisopropylphosphofluoridate (DFP) active-site labeling and sequencing of a radioactive peptide containing the signature sequence GDSAGG.

Keywords: affinity labeling/Archaeoglobus fulgidus/catalytic triad/homology modeling/thermostable carboxylesterase

Introduction

Some studies (Krejei et al., 1991; Hemilä et al., 1994) have suggested that esterases, lipases and cholinesterases belong to a large family of phylogenetically related proteins with representatives in the domains of Eukarya and Bacteria including proteins lacking enzymatic activity. Three subfamilies have been identified: the C group includes cholinesterases from vertebrates and invertebrates, lipases from fungi, a number of esterases and some non-enzymatic proteins; the L group includes lipases from vertebrates and bacteria, lipoprotein lipases, lecithin–cholesterol acyltransferases and related non-enzymatic vitellogenins from flies; the H group was also named hormone-sensitive lipase (HSL) by Hemilä et al. (1994) since these authors reported the cloning and sequencing of a gene from Alicyclobacillus acidocaldarius (formerly Bacillus) which encoded a protein of unknown function (ORF3) homologous to the HSL from human and rat (Langin et al., 1993). In this group were included several esterases and lipases but it is unclear at the moment whether the sequence similarities among these enzymes imply comparable biochemical properties. In addition, some members of the HSL group were identified only as gene sequences and there is no indication of their functionality.

Recently we reported the purification of a new esterase (EST2) from A. acidocaldarius and demonstrated its identity with the aforementioned ORF3. On the basis of a secondary structure-driven multiple sequence alignment we predicted for this enzyme the α/β-hydrolase topological fold common to several esterases and lipases and identified Ser155, Asp252 and His282 as the putative members of the catalytic triad (Manco et al., 1997). The gene was overexpressed in Escherichia coli and the protein was purified and characterized to obtain further insights into the structure–function relationship of this interesting representative of the HSL group (Manco et al., 1998a). The enzyme has been crystallized (De Simone et al., 1999) and resolution of the 3D structure is currently under way. This will permit the comparison with the recently solved structure of brefeldin A esterase (BFAE) from Bacillus subtilis, a mesophilic representative of the HSL group (Wei et al., 1999).

More recently, the ORF AF1763 identified in Archeoglobus fulgidus as a hypothetical esterase (AFEST) (Klenk et al., 1997) has been overexpressed in E. coli, purified and demonstrated by us to be a hyper-thermophilic carboxylesterase (Manco et al., 2000). In this work we exploited a combined approach of affinity labeling and molecular modeling to investigate the structural features of the enzyme.

Materials and methods

Homology modeling

AFEST and BFAE sequence alignment was obtained from a secondary structure-driven multisequence alignment of the HSL group (not shown) using the Clustal W program (Thompson et al., 1994) and refined manually to reduce to the minimum gaps which were not allowed inside the structural elements.

Molecular modeling was performed on a Silicon Graphics O2 workstation using the commercial software packages InsightII (InsightII User Guide, October 1995, Biosym/MSI, San Diego, CA).

The high-resolution X-ray crystal structure of BFAE was used as a template structure. Several 3D models were constructed using the Modeller module (Sali et al., 1995) within InsightII. The methodology is based on satisfaction of spatial restraints that are obtained from an alignment of a target sequence with related 3D structures at high resolution (1.8–2.4 Å), using a conjugate gradient and a molecular dynamics-simulated annealing as optimization procedures. Each model was also opportune minimized using the Discover3 module within InsightII.

The resulting models were verified using the on-line software WHAT IF at http://biotech.ebi.ac.uk:8400/chk/whatif/index.html.
Enzyme labeling with $[^3]$Hdiisoproporphosphofluoridate

Protein expression and purification have been described elsewhere (Manco et al., 2000).

A sample of 350 $\mu$g of pure protein (11 nmol) was labeled by incubation with 140 $\mu$Ci of $[^3]$HDFP (23 nmol; 60×10$^6$ c.p.m.) in 300 $\mu$L of 20 mM Tris–HCl buffer (pH 8.0) at 37°C for 2 h and then overnight at room temperature. The unreacted inhibitor was separated from protein by gel filtration through Sephadex G-25 equilibrated and eluted with 20 mM Tris–HCl (pH 8.0) containing 1 mM MgCl$_2$, 0.5 mM EDTA, 100 mM NaCl. Fractions of 0.5 ml were collected and radioactivity eluting at the void volume was measured in a Packard Tri-Carb 300 liquid scintillation counter. Radioactive fractions were pooled and enzyme activity and radioactivity were measured. The radioactivity and enzyme activity recoveries were 50% and 13%, respectively.

Protein alkylation and digestion

The radioactively labeled protein (175 $\mu$g, corresponding to 5.5 nmol of protein) was lyophilized, dissolved in denaturing buffer containing 0.5 M Tris–HCl (pH 8.0), 2 mM EDTA, 6 M guanidine-HCl and reduced with diithiothreitol (100 nmol) for 3 h at 37°C under nitrogen. 4-Vinylpyridine (9 $\mu$mol) was added and the reaction proceeded for 45 min at room temperature in the dark under nitrogen. The protein was immediately desalted by gel filtration on a PD-10 column (Pharmacia, Uppsala, Sweden) equilibrated in 0.1% trifluoroacetic acid and lyophilized. The lyophilized protein was resuspended in 25 mM Tris–HCl (pH 8.5) containing 10% acetonitrile and 1 mM EDTA and digested with 5 $\mu$g of endoproteinase Lys-C (Boheringer, Mannheim, Germany) for 4 h at 37°C. The peptide mixture was separated by reversed-phase HPLC on a C$_{18}$ Vydac column with a linear gradient of acetonitrile in 0.1% trifluoroacetic acid. The absorbance at 220 nm and $[^3]$H radioactivity were monitored.

Amino acid sequencing

Automated Edman degradation was performed on a Procise Protein Sequencer Model 492, equipped with a 140C Microgradient System from Perkin-Elmer (Applied Biosystems Division). Fractions were collected at each cycle, after HPLC elution of the released phenylthiohydantoin amino acid and aliquots were counted for $[^3]$H radioactivity.

Results and discussion

The pairwise comparison between AFEST and members of the C and L groups of the esterase/lipase family whose 3D structure is known shows a very low level of sequence identity (about 30% or less; data not shown). Nevertheless we hypothesized for AFEST the $\alpha$-$\beta$-hydrolase fold on the basis of secondary structure predictions and a ‘threading’ technique (Manco et al., 1999b), the latter being a powerful means to identify remote relatives of proteins when the sequence identity is close to 25%, a range of sequence identity which is frequently referred to as the ‘twilight zone’ (Chothia and Lesk, 1986; Sander and Schneider, 1991). The recently reported 3D structure of brefeldin A esterase (BFAE) from Bacillus subtilis, a member of the HSL group (Wei et al., 2000), confirmed that the $\alpha$-$\beta$-hydrolase fold is typical of this group of proteins and thereby confirmed our predictions. In the current absence of a 3D structure of AFEST and of the related EST2 from A. acidocaldarius (42% identity; Manco et al., 2000) and in order to plan a site-directed mutagenesis program to approach the study of the basis of thermostability, thermophilicity and the mechanism of action of these enzymes, we constructed a 3D model of AFEST using BFAE of $B$. subtilis (PDB code: 1JKM) as template. We generated a secondary structure-driven sequence alignment between AFEST and BFAE (Figure 1) and submitted this alignment to the modeling procedure described in the Materials and methods section. Figure 2A shows the general folding pattern of the best AFEST model obtained. The superimposed backbone traces of AFEST and 1JKM (not shown) displayed a 0.57 Å r.m.s. deviation on 297 C$\alpha$ atoms, using a 2.75 Å cut-off distance. In a global superposition (308 C$\alpha$ atoms) the r.m.s. deviation was 1.07 Å. The Ramachandran plot (not shown) indicated that most (95%) of residues have $\phi$ and $\psi$ angles in the core and allowed regions. The putative nucleophile Ser160 displayed $\phi$ and $\psi$ angles as expected. Most bond lengths, bond angles and torsion angles evaluated with the WHAT IF program were within the range of values expected for a naturally folded protein (data not shown). The schematic structural model shown in Figure 2A consists of one $\beta$-sheet made up of eight $\beta$-strands ($\beta_1$–$\beta_8$) twisted by ~90°, corresponding to the canonical $\alpha$-$\beta$ fold (Ollis et al., 1992). The identified strands were $\beta_1$ (Glu54–Lys62), $\beta_2$ (Asn65–Tyr73), $\beta_3$ (Val81–Tyr85), $\beta_4$ (Thr112–Asp117), $\beta_5$ (Lys153–Gly158), $\beta_6$ (Gln184–Tyr188), $\beta_7$ (Pro246–Ala253) and $\beta_8$ (Val274–Tyr280). The main differences between AFEST and BFAE are the absence of the first of the two $\alpha$-helices at the N-terminus which represent a sub-domain peculiar to BFAE not observed in the other $\alpha$-$\beta$ hydrolases characterized up to now (Wei et al., 1999) and
the shortening of some loops. The same has been observed in a comparison with the model of the enzyme from *A. acidocaldarius* (data not shown). The importance of these loops in the structure thermostabilization will be a topic of future investigations.

According to the α/β-hydrolase fold (Ollis *et al*., 1992) and our previously made predictions (Manco *et al*., 1998b), Ser160, Asp255 and His285 were located close to each other and connected by hydrogen bonds as in BFAE (Figure 2B), thus representing a classical catalytic triad. The above results agree with site-directed mutagenesis data reported for human HSL (Østerlund *et al*., 1997) and *Moraxella* TA144 lipase (Feller *et al*., 1991) and more recently for *A. acidocaldarius* EST2 (Manco *et al*., 1999). The active-site serine is part of a conserved sequence GXSXG, GDSAGG in the HSL group, which has been found in many other enzymes containing the catalytic triad (Brenner, 1988). The serine is located in a short turn between β-strand 5 and α-helix 5 in a strained conformation with backbone torsion angles φ = 62.8° and ψ = −120.6°, exposing the hydroxyl group to the solvent and to the catalytic histidine. For steric reasons the super-secondary structure strand-turn–helix requires glycines at positions +2 and −2 with respect to the nucleophile Ser160. For the same reason a residue with a short lateral chain, Ala161 in this case, is placed after the serine. The aspartic acid before the serine is substituted by glutamate in members of the C group and has been proposed to be involved in the coordination of water needed for substrate hydrolysis (Cygler *et al*., 1993).

To test further the role of Ser160 as the nucleophilic element of the triad we used a biochemical approach, as follows. The active enzyme was labeled as described previously (Manco *et al*., 1998a) with [3H]DFP which totally inhibits the activity of serine-type enzymes by irreversible phosphorylation of the essential nucleophilic residue; it was made free of unreacted inhibitor by gel filtration on a Sephadex G-25 column and subjected to protease digestion in order to identify the labeled peptide and the reactive serine residue. After endopeptidase Lys-C digestion and HPLC separation of the resulting peptide mixture, most of the radioactivity was found associated with a single peak (Figure 3). Amino acid sequencing revealed that the fraction contained two co-eluting peptides. Taking advantage of the different amounts of the peptides in the mixture, simultaneous sequencing was carried out. Twenty-five cycles of Edman degradation yielded the sequences (i) MLDMPIDP/VYYQLAEYFDLSPK and (ii) IFVGGD(S)AG-

---

**Fig. 3.** Pure AIFEST was labeled with [3H]DFP and digested with Lys-C. Fragments (solid line) detected at 220 nm (left axis) were separated with an acetonitrile gradient in 0.1% TFA. Almost all radioactivity (dotted line) measured as c.p.m. (right axis) co-migrated with the peak eluted at 58 min. The importance of these loops in the structure thermostabilization will be a topic of future investigations.

**Fig. 4.** Radioactivity detected at each cycle during the Edman sequencing. Most of the radioactivity was eluted at the seventh cycle where [3H]DFP-modified Ser160 was expected. The counts indicated represent 200 µl of a total of 8.5 ml collected at each step.

---

GNLAAAVSIMARDSGE and the peptides were identified on the basis of the sequence deduced from DNA. The first peptide corresponded to the amino-terminal region including the N-terminal methionine (positions 1–22). The second peptide corresponded to positions 154–178 of the total sequence. The [3H]DFP-modified serine residue was identified by the appearance of radioactivity in the fractions released following sequential degradation and was localized at the seventh cycle where Ser160 of the 154–178 peptide was expected to be eluted (Figure 4). At this cycle, a very low signal for the serine residue could be detected (data not shown). Since the [3H]DFP-modified serine is expected to be eluted at a different position from that of serine, the small amount of serine detected at this cycle was probably due to partial cleavage of the diisopropyl moiety from the labeled serine residue during the routine of Edman degradation. Serine residues not corresponding to radioactive signals were also detected at the seventeenth and twenty-third cycles corresponding to positions 170 and 176 of the 154–178 peptide and at nineteenth cycle corresponding to position 19 of the 1–22 peptide. This result together with the information resulting from the 3D model demonstrate conclusively that Ser160 is the nucleophilic member of the catalytic triad in AIFEST.

**Conclusion**

This study set up the guidelines for a deeper analysis of the relationship between the structure and function of a *A. fulgidus* esterase. The 3D model will provide a 3D framework for the rational design of site-specific mutations to address the problem of protein thermophilicity, thermostability and substrate specificity before we succeed in the resolution of the 3D structure for this and the related thermophilic subbacterial EST2 from *A. acidocaldarius*. The fact that these two enzymes share many features and are practically colinear in sequence (Manco *et al.*, 2000) and yet are derived from organisms in different domains of life is noteworthy, since it speaks of the probable gene transfer that may account for the similarity. Even more intriguing is the similarity of these enzymes with two members of the HSL group (*Moraxella* and human lipases) showing adaptation to low temperatures (Feller *et al*., 1991; Langin *et al*., 1993). We hope that the accumulation of new data on these and other enzymes in the HSL group will accelerate...
the understanding of the evolutive and structure–function relationships in this interesting group of enzymes.

Acknowledgements

This study was partly supported by a European Union Contract (B104-CT96-0488), by the CNR Target Project on ‘Biotechnology’ and by a grant from Recordati Industria Chimica e Farmaceutica SpA.

References


Received August 2, 1999; revised November 29, 1999; accepted December 8, 1999