Communication

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 267, No. 7, Issue of March 5, pp. 4300–4303, 1992
© 1992 by The American Society for Biochemistry and Molecular Biology, Inc.
Printed in U.S.A.

Pancreatic Lipases: Evolutionary Intermediates in a Positional Change of Catalytic Carboxylates?*

(Received for publication, December 3, 1991)

Joseph D. Schrag‡, Fritz K. Winkler§, and Miroslaw Cygler‡¶

From the ‡Biotechnology Research Institute, National Research Council of Canada, Montréal, Québec H4P 2R2, Canada and the §Pharma Division Research Technologies, F. Hoffmann-La Roche Ltd., CH-4002 Basel, Switzerland

Comparison of the fold of lipases from *Geotrichum candidum* and from human pancreas identified a high degree of similarity which was not expected on the basis of their amino acid sequences. Although both enzymes utilize a serine protease-like catalytic triad, they differ in the topological position of the acid. We speculate that these proteins are evolutionarily related and that the pancreatic lipase is an evolutionary intermediate in the pathway of migration of the catalytic acid to a new position within the fold.

The lipases from the fungus *Geotrichum candidum* (GCL)[1] and human pancreas (HPL) have been thought to belong to different structural classes due to a lack of sequence homology beyond the G*X*S*X*G fragment which encompasses the active site serine and is also common to serine proteases, esterases, and other lipases (1). The three-dimensional structures of these two enzymes (2, 3) revealed unexpectedly that GCL and the catalytic, N-terminal domain of HPL share a very similar topology of their central β-sheet and the same location of two (Ser, His) of the three residues comprising the catalytic triad (Fig. 1). The position of the acidic residue is different in the two proteins. In GCL the catalytic Glu-354 is located in a loop after strand seven while in HPL Asp-176 comes from a loop following strand six (Fig. 1). Could these proteins be evolutionarily related despite undetectable sequence homology and if so, is there any evidence in support of the migration of the acidic member of the triad during the evolution?

GCL and four other hydrolytic enzymes, whose three-dimensional structures have been determined recently by x-ray crystallography, show remarkable similarities in their topologies and the geometrical conservation of their central β-sheet.[2] This common topology can be described as $+1+2-1x+2x(+1x)_3$ in Richardson's terminology (5) (Fig. 1). Other enzymes of this fold family include dienelactone hydrolase from *Pseudomonas* (6, 7), wheat serine carboxypeptidase II (8), haloalkane dehalogenase from *Xanthobacter autotrophicus* (9), and *Torpedo californica* acetylcholinesterase (10). All of them utilize a triad arrangement of nucleophile-histidine-acid, reminiscent of that in serine proteases, for the hydrolytic reaction (11). The nucleophile is located at the end of the fifth strand of the β-sheet in a tight "nucleophile elbow" bend[2] which is followed by an α-helix. In all cases the nucleophile is forced into an unfavorable conformation with (+60°,−110°) ϕ,ψ angles. This, however, projects the nucleophile side chain away from the polypeptide backbone and makes it readily accessible for the histidine and the substrate. The histidine residue comes from a loop that follows strand eight of the β-sheet. Finally, the acid comes from a reverse turn following strand seven. The handedness of the triad is opposite to those of the serine proteases (2, 10, 12).[2]

Despite little sequence homology between these enzymes (except for GCL and acetylcholinesterase (2, 13, 14) which share 25% identity), Ollis *et al.*[2] proposed on the basis of the conservation of topology and the three-dimensional structure of the central mixed β-sheet, and the conservation of the topological position of the catalytic triad residues that all these proteins diverged from a common ancestor. The extended, variable length loops on the C-terminal side of the β-sheet provide a way to adapt these proteins to bind a wide range of substrates such as dienelactone, 1-haloalkanes, polypeptide C-termini, acetylcholine, or triacylglycerols.[2]

Although the topological position of the triad residues is conserved in these enzymes, the nature of the nucleophile and the acid was not. The diversity in the identity of the nucleophile, which can be either Cys (6, 7), Asp (9), or Ser (2, 3,

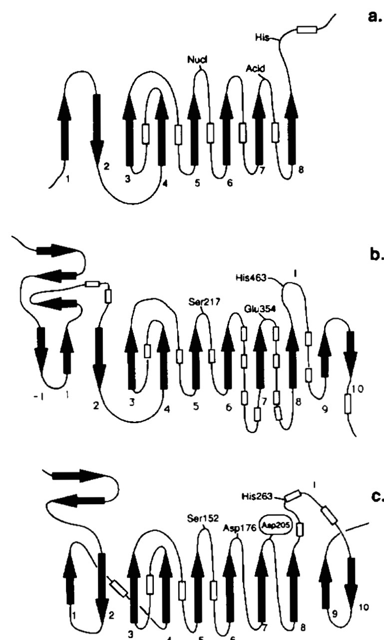

FIG. 1. **Topology diagrams.** *a*, the fold[2] common to dienelactone hydrolase (6, 7), wheat serine carboxypeptidase II (8), haloalkane dehalogenase (9), acetylcholinesterase (10), and *G. candidum* lipase (2). *b*, *G. candidum* lipase; *c*, human pancreatic lipase.

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ To whom correspondence should be addressed.

[1] The abbreviations used are: GCL, *G. candidum* lipase; HPL, human pancreas lipase.

[2] Ollis, D., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., Harel, M., Remington, S. J., Silman, I., Schrag, J., Sussman, J. L., Verschueren, K. H. G., and Goldman, A. (1992) *Protein Eng.*, in press

This is an Open Access article under the CC BY license.

10), reflects the diversity of reactions catalyzed by the different enzymes. The identity of the acidic group has also diverged as the carboxylate is contributed by either Asp (6–9) or Glu (2, 10). The only residue absolutely conserved is the histidine.[2]

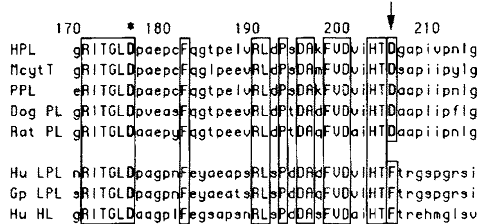

FIG. 2. **Aligned sequences of mammalian pancreatic and lipoprotein lipases within the region corresponding to residues 170–215 of the human pancreatic lipases (the numbering corresponds to that of porcine pancreatic lipase).** The active site Asp is marked with a *star*, while the position equivalent to Asp-205 in HPL is marked with an *arrow*. *HPL*, human pancreatic lipase (3); *McytT*, mouse cytotoxic T-lymphocyte lipase (19); *PPL*, porcine pancreatic lipase (20); *Dog PL*, dog pancreatic lipase (21); *Rat PL*, rat pancreatic lipase (GenBank accession no. M58369); *Hu LPL*, human lipoprotein lipase (23); *Gp LPL*, guinea pig lipoprotein lipase (4); *Hu HL*, human hepatic lipase (22).

A number of lipases and esterases show significant sequence homology to GCL and most likely are members of the same fold family of enzymes. Although the three-dimensional structures have been determined for only two of these enzymes (GCL and *Torpedo* acetylcholinesterase), sequence comparisons strongly suggest that the catalytic Ser, His, and Asp/Glu are in identical topological locations in all of these enzymes (2, 15).

The topology of the N-terminal domain of HPL, although not identical, shows striking similarities to the common topology of the above mentioned proteins[2] (Fig. 1). In Richardson's terminology the portion of HPL which corresponds to this fold is described as $-1{+}3x{-}1x{+}2x(+1x)_3$. Despite the connectivity of the first two strands of the β-sheet being different, the direction of the corresponding strands is the same as in the other enzymes. The similarities between GCL and HPL extend beyond that common fold portion and include a ninth strand. In both enzymes, the loop between the eighth and ninth strands contains the active site histidine and forms a rarely observed left-handed crossover connection. As we mentioned above, the other and most significant topo-

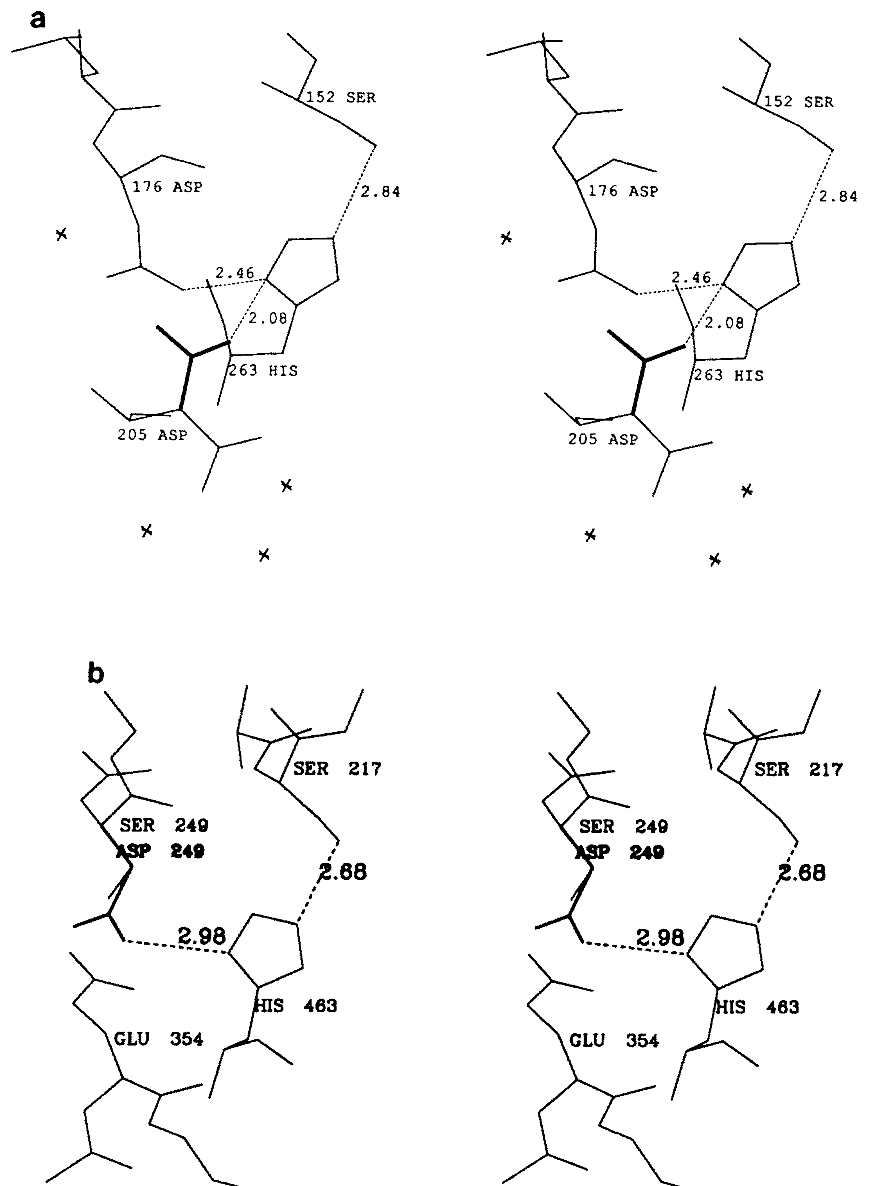

FIG. 3. *a*, model of the catalytic triad Ser-152...His-263...Asp-176 in HPL (*thin lines*). The rotated Asp-205 side chain (*thick lines*) is close to a position to make a good hydrogen bond contact (*dotted line*) to His-263. With small adjustments in the main chain conformation the hydrogen bond geometry can easily be improved. The *crosses* mark the position of nearby bound water molecules. *b*, model of the active site region in GCL (*thin lines*) with Ser-249 replaced by aspartate (*thick lines*) and rotated into position of contact with His-463. Hydrogen bonds are shown as *dotted lines*.

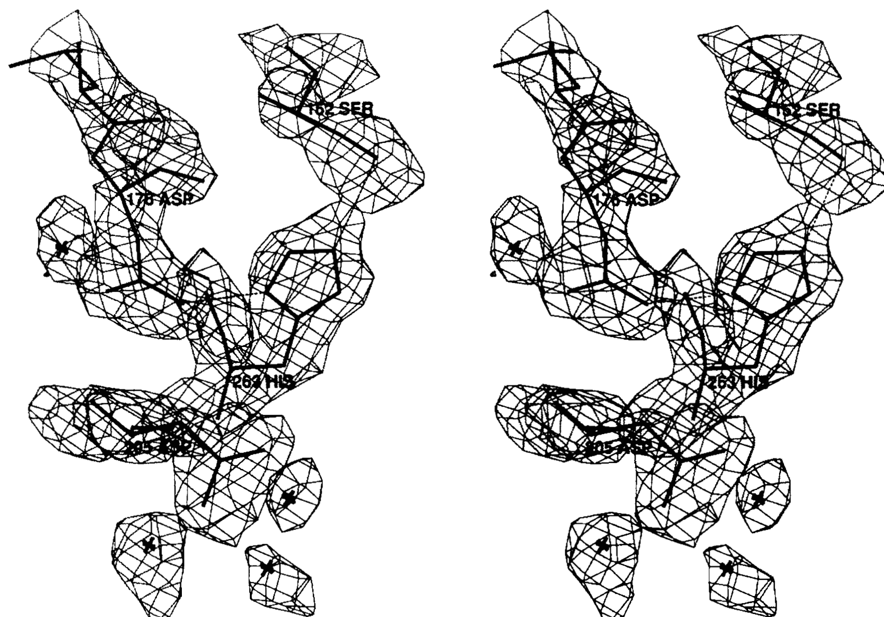

FIG. 4. **Experimental electron density ($2F_{obs} - F_{calc}$) in the region of the catalytic triad of HPL shown in the same view as in Fig. 2*a*.** The hydrogen bonds within the triad are shown as *dotted lines*.

logical departure from the common fold is the position of the catalytic triad acid, which in HPL follows strand six rather than strand seven as in the other enzymes. Examination of the GCL and HPL models reveals that the residue in HPL which corresponds to the triad acid Glu-354 of GCL is an aspartic acid (Asp-205). On the other hand, the residue in GCL that is in the same structural position as the triad participant of HPL, Asp-176, is Ser-249.

HPL shows strong sequence homology to other mammalian pancreatic lipases and also to lipoprotein lipases, suggesting similar three-dimensional structures for these enzymes. The position of Ser-152, His-263, and Asp-176 is conserved in all these sequences, while Asp-205 is conserved only in pancreatic lipases (Fig. 2) and is replaced by phenylalanine in lipoprotein lipases.

Investigation of the three-dimensional model of HPL shows that indeed Asp-205 is in the proximity of His-263 and that, by a simple rotation around torsion angles $\chi 1$ and $\chi 2$, the carboxylate group can be moved into a hydrogen bonding position to this histidine (Fig. 3*a*). Similarly, from the inspection of the GCL model it is apparent that if Ser-249 was replaced by an aspartic acid, its side chain could form a hydrogen bond with His-463 (Fig. 3*b*).

The electron density map of HPL shows clearly that it is the Asp-176 that forms the crucial hydrogen bond to His-263, while Asp-205 is involved in interactions with ordered water molecules (Fig. 4). In conjunction with similar experimental evidence in GCL (2) and other enzymes sharing this fold (6–10), this shows that the fold is flexible enough to accommodate two alternative positions for the acid without disturbing the spatial requirements of the catalytic triad.

The above arguments would strongly suggest that pancreatic and lipoprotein lipases are related to the above mentioned enzymes and would seem to indicate that the pancreatic lipases represent an evolutionary intermediate in the migration of the catalytic acid. The direction of the migration is presumed to have been from strand seven to strand six, rather than the opposite direction. Enzymes with the acid on strand seven include bacterial, fungal, plant, and fish enzymes. Since these life forms predate the existence of mammals one would expect that these enzymes are predecessors to, rather than derivatives of, the mammalian forms.

The common structural motif has been well conserved, but the identities of the nucleophiles and acidic groups have not.[2] The mammalian pancreatic, hepatic, and lipoprotein lipases apparently provide an example of divergence in the location of one of the catalytically essential residues. This fold is thus a very stable scaffold which allows the assembly of a variety of residues into catalytic triads, allows modification of substrate binding loops, resulting in enzymes capable of hydrolysis of wide range of substrates, and also allows alternate arrangements of the residues involved in the catalysis.

Sequence comparisons also suggest that this family of proteins has diverged even further to include proteins which have lost their catalytic hydrolase function. The C-terminal domain of thyroglobulin shows significant sequence homology to GCL and the cholinesterases and has been proposed to be involved in membrane binding (16). The lipoprotein and hepatic lipases show significant homology with vitellogenin (17, 18). The vitellogenins are involved in lipid binding and transport but have no hydrolase activity. This fold family thus appears to be not only a diverse family of hydrolases but a diverse family of proteins with varied functions, and it is very likely that new members of this folding family will frequently be identified.

REFERENCES

1. Brenner, S. (1988) *Nature* **334,** 528–530
2. Schrag, J. D., Li, Y., Wu, S. & Cygler, M. (1991) *Nature* **351,** 761–764
3. Winkler, F. K., D'Arcy, A. & Hunziker, W. (1990) *Nature* **343,** 771–774
4. Enerbäck, S., Semb, H., Bengtsson-Olivecrona, G., Carlsson, P., Hermansson, M.-L., Olivecrona, T. & Bjursell, G. (1987) *Gene* (*Amst.*) **58,** 1–12
5. Richardson, J. S. (1981) *Adv. Protein Chem.* **34,** 167–339
6. Pathak, D., Ngai, K. L. & Ollis, D. (1988) *J. Mol. Biol.* **204,** 435–445
7. Pathak, D. & Ollis, D. (1990) *J. Mol. Biol.* **214,** 497–525
8. Liao, D. & Remington, S. J. (1990) *J. Biol. Chem.* **265,** 6528–6531
9. Franken, S. M., Rozeboom, H. J., Kalk, K. H. & Dijkstra, B. W. (1991) *EMBO J.* **10,** 1297–1302
10. Sussman, J. L., Harel, M., Frolov, F., Oefner, C., Goldman, A., Toker, L. & Silman, I. (1991) *Science* **253,** 872–879
11. Blow, D. M., Birktoft, J. J. & Hartley, B. S. (1969) *Nature* **221,** 337–340
12. Gubernator, K., Müller, K. & Winkler, F. K. (1991) in *GBF Monographs* (Alberghina, L., Schmid, R. D., and Verger, R., eds) Vol. 16, pp. 9–16, VCH Verlagsgesellschaft mbH, Wein-

heim, Germany

13. Shimada, Y., Sugihara, A., Iizumi, T. & Tominaga, Y. (1990) *J. Biochem. (Tokyo)* **107,** 703–707
14. Slabas, R., Windust, J. & Sidebottom, C. M. (1990) *Biochem. J.* **269,** 279–280
15. Sikorav, J.-L., Krejci, E. & Massoulié, J. (1987) *EMBO J.* **6,** 1865–1873
16. Swillens, S., Ludgate, M., Mercken, L., Dumont, J. E. & Vassart, G. (1986) *Biochem. Biophys. Res. Commun.* **137,** 142–148
17. Persson, B., Bengtsson-Olivecrona, G., Enerbäck, S., Olivecrona, T. & Jörnwall, H. (1989) *Eur. J. Biochem.* **179,** 39–45
18. Bownes, M., Shirras, A., Blair, M., Collins, J. & Coulson, A. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85,** 1554–1557
19. Grusby, M. S., Nabavi, N., Wong, H., Dick, R. F., Bluestone, J. A., Schotz, M. C. & Glimcher, L. H. (1990) *Cell* **60,** 451–459
20. DeCaro, J., Boudouard, M., Bonicel, J., Guidoni, A., Desnuelle, P. & Rovery, M. (1981) *Biochim. Biophys. Acta* **671,** 129–138
21. Mickel, F. S., Weidenbach, F., Swarovsky, B., LaForge, K. S. & Scheele, G. A. (1989) *J. Biol. Chem.* **264,** 12895–12901
22. Datta, S., Luo, C.-C., Li, W.-H., Van Tuinen, P., Ledbetter, D. H., Brown, M. A., Chen, S.-W., Liu, S.-W. & Chan, L. (1988) *J. Biol. Chem.* **263,** 1107–1110
23. Wion, K. L., Kirchgessner, T. G., Lusis, A. J., Schotz, M. C. & Lawn, R. (1987) *Science* **235,** 1638–1641