BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **224,** 206–211 (1996)
ARTICLE NO. 1008

Expressed Sequence Tags Identify Human Isologs of the ARF-Dependent Phospholipase D

Gérard Ribbes,* Joëlle Henry,† Clotilde Cariven,* Pierre Pontarotti,† Hugues Chap,* and Michel Record*,[1]

**Unité INSERM 326, Phospholipides Membranaires, Signalisation Cellulaire et Lipoproteines, and †d'Immunolologie et de Génétique Humaine, Hôpital Purpan, 31059, Toulouse, France*

Received May 20, 1996

By searching into Expressed Sequence Tags databases (dbEST) using Blast X algorithm software and a plant phospholipase D as template, we have identified a cDNA from human brain (Z45777) which encodes for a protein similar to the amino acid region 743-929 of the human phospholipase D1 (PLD1), and a cDNA from human liver (R93485) which encodes for a protein similar to region 815-932 of PLD1. Sequence comparison between cloned phospholipases showed the presence of 3 conserved amino acid sequences: AFVGGIDLAYGRWD (box A), IIGSANINDRS (box B), and YIYIENQFFI (box C). Phylogenic analysis indicated that the cDNA from brain and liver encoded for human isologs of PLD1. © 1996 Academic Press, Inc.

The programs of systematic genomes sequencing have supplied numerous new cDNA sequences named EST for ''Expressed Sequence Tagged'' into the dbEST data bank.

The new informations accumulated within a year in such a data bank have been multiplied by 2.4 (from 172 175 to 416 419 sequences). Thus, it appears conceivable to characterize an unknown cDNA, i.e. to clone it, using the classical BLAST algorithmic program software, and at least one reference sequence as template (1).

In this *in silico*[2] cloning technique, one has to assume that a minimal sequence must be highly conserved throughout the evolution. Based on all these assumptions, we have undertaken to search for a cDNA of the human homologue of a plant phospholipase D. This enzyme is a phosphodiesterase which hydrolyses phospholipids, and plays an important role in cell signaling (2).

EXPERIMENTAL PROCEDURES

Methods. In step 1, using Blast X software (3), the cDNA of *ricinus communis* phospholipase D (PLDrc) (1) was converted into amino acid sequence and compared with a protein database (Swiss Prot). In step 2, using tBlastN program, a protein of unknown function presenting a high homology with PLDrc was compared with the Expressed Sequence Tags traduced into the 6 reading frames of dbEST database. The general strategy was plotted in Fig. 1. The philogeny between phospholipases D was obtained using ClustalW (4) and Phylip softwares (5). The full sequencing of the human brain clone was performed by automatic sequencing with UP-RP primers using an IBI automatic equipment.

RESULTS

The ''cloning'' of a cDNA from human brain using plant phospholipase D as template. The publication of the first phospholipase D ever cloned experimentally was reported[3] from castor

[1] To whom correspondence should be addressed.
[2] ≪ in silico ≫ referred to the basic material of computer hardware
[3] May 1994.

0006-291X/96 $18.00

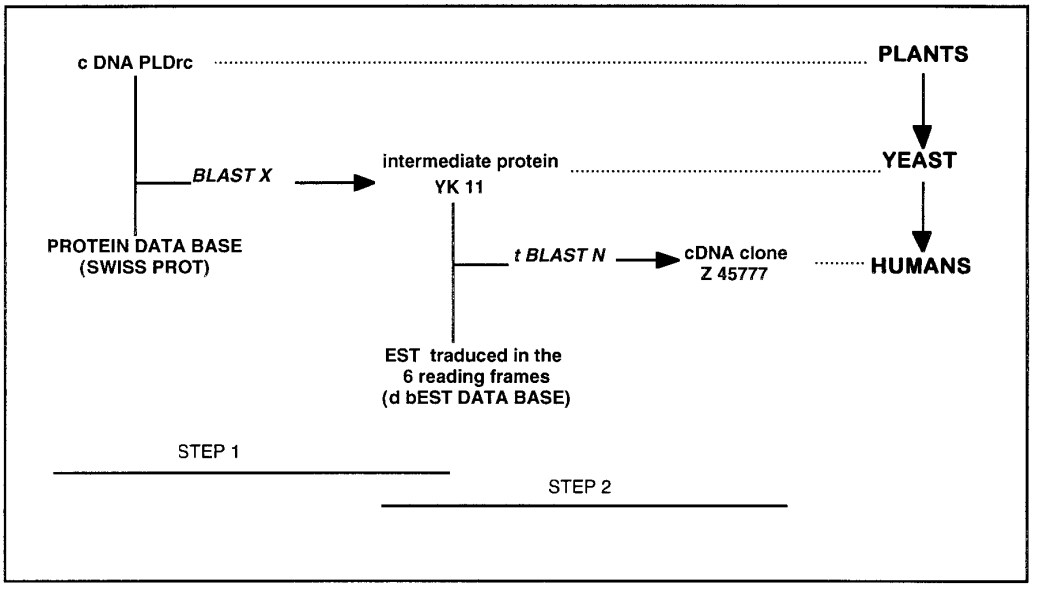


FIG. 1. General strategy for in silico cloning. The BLAST software contains the various options allowing to perform the search. The first step is a cDNA/protein comparison. The cDNA used as template is converted into the proteins corresponding to the six reading frames and compared with protein databases with BLAST X option (Step 1). The second step is a protein / cDNA (EST) comparison. The intermediate protein thus unraveled is compared with the EST traduced into proteins corresponding to the six reading frames, using tBLAST N option. (tBlastN software compare a protein query sequence against a nucleotide sequence database dynamically translated in the six reading frames)

bean (*ricinus communis*) (1). At that time the authors mentioned that this enzyme showed no similarities with sequences of other cloned lipolytic enzymes. Soon after, an hypothetical 195.2 Kd protein from yeast was deposited into databases and called YK11 (Swiss Prot accession number P36126). The corresponding cDNA had been cloned in the course of the sequencing program of chromosome 11 in yeast. When searching for similarity between *ricinus communis* phospholipase D (PLDrc) and other protein(s) we found[4] two conserved regions between PLDrc and YK11 as reported in fig 2. Such a result was obtained by comparing the six frame translation products of the cDNA from PLDrc, against a non redondant protein sequence database, using Blast X software. When sending the amino acid sequence of YK11 in the dbEST databases using the tBlastN algorithmic software, we found also some similarity with a cDNA from human brain (clone accession number Z45777 in dbEST[5]).

At that time no homology of this clone was observed with Genbank release 81 and Swissprot release 28. This cDNA clone was hold by the Genexpress (at Genethon, France) and had been obtained using a normalized cDNA library according to Soares et al (6). It should be note that the homology between YK11 and the protein from brain was located between the two conserved boxes evidenced between PLDrc and YK11 (Fig.2). However brain has been known as a tissue with one of the highest PLD activity (7). We postulated that the human brain cDNA clone could encode for a PLD related protein. When sequencing the full length of this clone, we

[4] March 1995.
[5] April 1995.

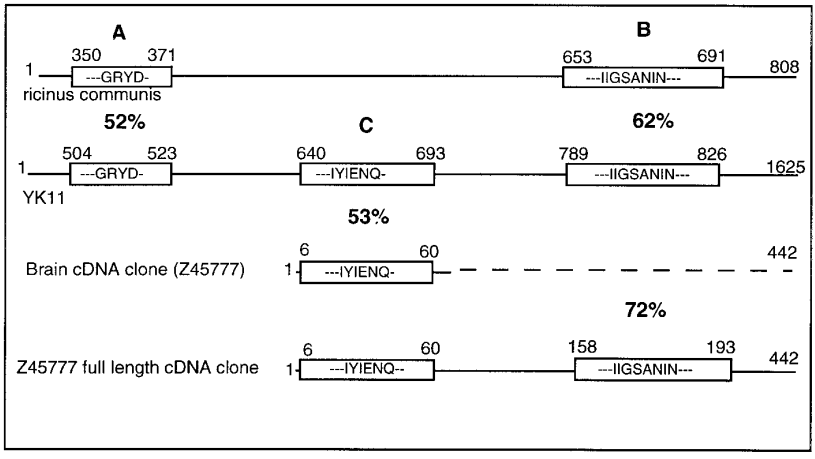


FIG. 2. ≪ Isolation ≫ of a human cDNA clone by computer. The phospholipase D cloned from castor bean (PLDrc) was compared with databases of proteins and the YK11 protein from yeast of unknown function was ''isolated.'' This first sequence comparison allowed to evidence to regions of strong similarity (boxes A and B). The YK11 sequence was then compared with the Expressed Sequence Tags databases, and similarity was found in a region located between boxes A and B, with an unknown protein encoded by a cDNA (Z45777) from human brain (box C). Full length sequencing of that cDNA exhibited the box B previously evidenced in YK11 and PLDrc.

found effectively the same box ''B'' as in YK11 and PLDrc (Fig.2). However the clone was truncated on the 5′ end as compared with YK11 and PLDrc.

The human brain cDNA clone is identical to part of the human PLD cDNA. In the course of the experimental cloning of a full length cDNA for a human PLD from brain, using a labeled probe constructed from the truncated cDNA, the first sequence of a human PLD from Hela cells was reported (8).

The cloning had been realized from the characterization of a protein required for meiosis in yeast and which exhibited a phospholipase D activity. The gene corresponding to this protein was defective in a yeast mutant unable to sporulate (9,10). This protein, called by authors SPO14p, turned out to be identical to YK11. In that respect, SPO14p has recently replaced YK11 in databases.

Our human brain cDNA clone is mostly identical to the sequence 761-1085 of arf-PLD (PLD1) and to sequence 636-826 of SPO14p (Fig.3). Therefore the cDNA clone Z45777 encodes for a phospholipase D from brain and certainly contains the catalytic site of PLD, since its sequence corresponds to a highly conserved domain throughout the evolution. It should be noted that the amino acid sequences: AFVGGIDLAYGRWD, IIGSANINDRS, and YIYIENQFFI, when introduced individually into databases sorted ten proteins only (Table 1), among which six have been experimentally identified as phospholipases D; the others displayed strong similarities with PLD1 allowing to postulate they were also phospholipases D. By means of box B, we have thus ≪ isolated ≫ the clone R93485 from human liver (Table 1). Noteworthy, the box C was not observed first in the PLDrc (Fig.2), because the PLDrc sequence probably contains a sequencing error between the two conserved boxes A and B. When changing of open reading frame, then the YIYIENQFFI sequence signature appears (Fig.3 and Table 1).

Philogeny of phospholipases D. Philogeny between the various phospholipases D we have characterized (Table 1) was reported in Fig 4A. This analysis showed that the cDNAs from human brain and liver are isologs of the PLD1. The philogenyc tree indicated that the group

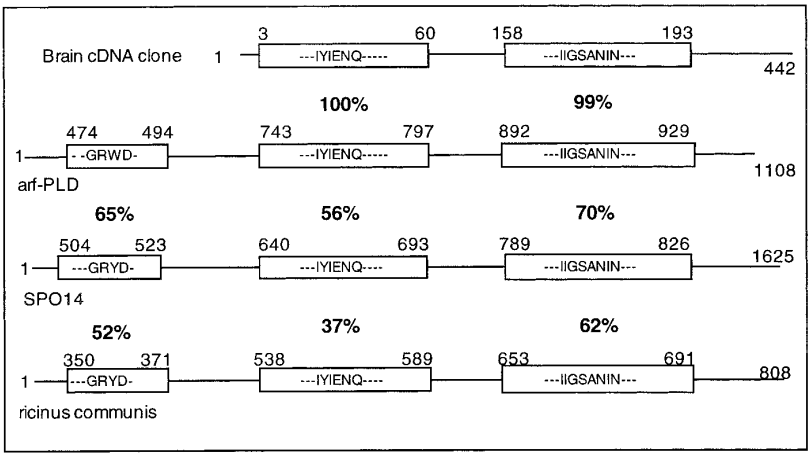


FIG. 3. The Z45777 clone and YK11 belonged to the phospholipase D family. Boxes C and B of the protein encoded by the Z45777 cDNA were also found in the recently cloned phospholipase D (arf-activated or ''PLD1'') from the human strain HeLa cells (8). Boxes A, B, and C in PLD1 were highly conserved in the yeast SPO14p recently evidenced as a phospholipase D (9), and which is identical to the YK11 protein ''isolated'' by computer search in Fig.2. The box C could be evidence in PLDrc by changing of open reading frame.

of yeast and human phospholipases D was distinct from that of plant phospholipases D. Interestingly, plant phospholipase D activities are not regulated by PIP2 or ARF (11). In addition, human phospholipase D appears tightly regulated, since the extend of phosphatidyl-

TABLE 1
''Sequence Signatures'' of Phospholipases D

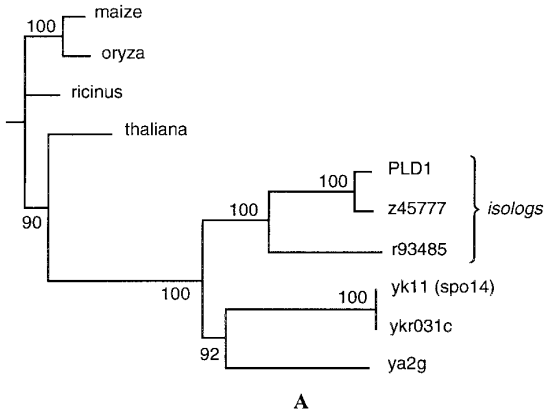
Origin	GB id.	Box A	Box B	Box C
Human				
human brain*	Z45777		IIGSANINDRS	YIYIENQFFI
human PLD1 **(8)**	U38545	AFVGGIDLAYGRWD	IIGSANINDRS	YIYIENQFFI
human liver*	R93485		IIGSANINDRS	
Yeast				
YK11/SPO14 **(9)**	P36126	AFIGGTDLCYGRYD	IIGSANINERS	FIYIENQFFI
YKR031c **(10)**	S38103	AFIGGTDLCYGRYD	IIGSANINERS	FIYIENQFFI
YA2G*	Q09706	TFIGGIDLCFGRYD	VIGSANINERS	FIYIENQFFV
Plants				
Zea mays **(14)**	D73410	SFIGGIDLCDGRYD	IIGSANINQRS	FIYIENQYFL
Oryza sativa **(14)**	D73411	SFVGGLDLCDGRYD	IIGSANINQRS	FIYIENQYFL
Thaliana **(15)**	U36381	SFVGGIDLCDGRYD	IIGSANINQRS	FIYIENQYFL
ricinus **(1)**	L33686	SFVGGLDLCDGRYD	IIGSANINQRS	FIYIENQYFL[a]
Consensus		s**FhGGh**D**LC GR**o**D**	f**IGSANIN**p**RS**	o**IYIENQ**o**F**f

Note. This table was obtained by searching into databases using either the YIYIENQFFI, IIGSANINDRS or AFVGGIDLAYGRWD motives. Only these proteins were sorted. They correspond to established phospholipases D (reference number is indicated in brackets) or putative ones (*).

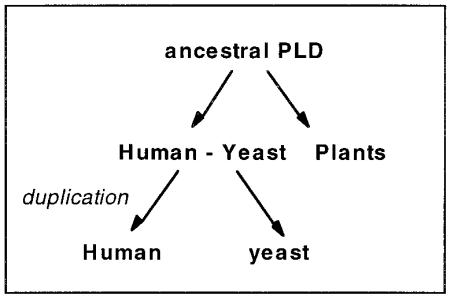
GB id.: abbreviation of GenBank identification number. Boxes A, B, and C refer to fig. 2.

[a] This motif can only obtained by switching between frame 1 to frame 2.

o: aromatic residues, f: aliphatic residues, p: polar residues, s: small residues, h: hydrophobic residues.



A



B

FIG. 4. Philogenyc tree of phospholipases D. In **A,** philogenyc tree obtained by ClustalW and Phylip softwares with the bootstrap values (probability that the filiation be true). Symbols: maize = corn; oryza = rice ; thaliana = arabidopsis thaliana; PLD1= human phospholipase D1; Z45777= human brain protein; R93485= human liver protein; YK11 (spo14)= yeast phospholipase D; YKR031c was identical to YK11; Ya2g was another member of yeast proteins with putative phospholipase D activity. Interspecy phospholipase D philogeny is represented in **B.**

choline is limited to a few percent (12), and also a factor of 50 Kd additional to ARF and PIP2 is required for the full enzyme activation (13).

Therefore the distinct evolution of the group of yeast and human phospholipases D comparatively to plants (Fig.4B), might reflect the apparition of regulatory sequences towards the catalytic activity. On the opposite, since the three boxes A, B and C were highly conserved throughout the evolution, one can consider they belong to the catalytic domain. Multiple sequence alignments of the various proteins was reported in Fig.5.

DISCUSSION

The strategy for ‘‘cloning’’ through computer search appears as a convenient approach, which might avoid useless experimental step during the course of an experimental cloning. By computer searching we found YK11 yeast protein, a human brain and human liver protein (cDNA Z45777 and R93485 clones) whose functions were unknown but which showed strong similarity in some part of their sequence. In the present case our ‘‘ computer strategy ’’ has been experimentally confirmed. The YK11/SPO14p protein has been identified as phospholipase D, and the protein corresponding to Z45777 and R93485 are partial PLD sequences. Without the

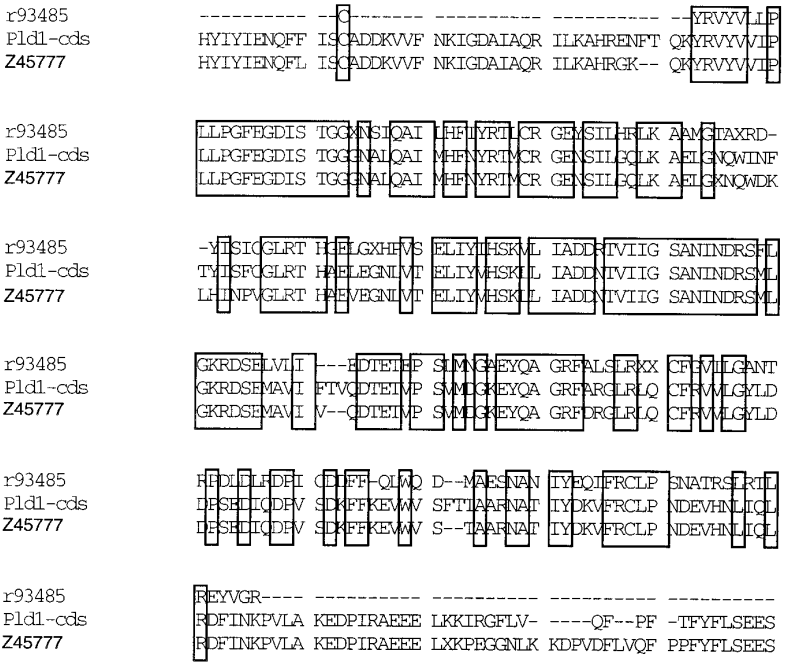


FIG. 5. Multiple alignment of sequences of the predicted protein products from Z45777 and R93485 human isologs of PLD1. Multiple alignments were performed using Gene Work program . Boxes indicated conserved amino acids.

observation of a yeast mutant (9) which was defective for PLD activity the cloning of a human PLD would have taken longer time, requiring for instance purification of the protein. The characterization of cDNA clones from human brain and liver which are similar to part of PLD1 (Fig. 5), constitutes an advanced step for experimental screening of cDNA banks.

REFERENCES

1. Wang, X., and Zheng, L. (1994) *J. Biol. Chem.* **269,** 20312–20317.
2. Exton, J. H. (1990) *J. Biol. Chem.* **265,** 1–4.
3. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.,* **215,** 403–410.
4. Higgins, D. G., Bleasby, A. J., and Fuchs, R. (1991) *Cabios,* **8,** 189–191.
5. Higgins, D. G., and Sharp, P. M. (1989) *Cabios,* **5,** 151–153.
6. Soares, B. M., Bonaldo, M. F., Jelene, P., Su, L., Lawton, L., and Efstratiadis, A. (1994) **91,** 9228–9232.
7. Chalifa, V., Möhn, H., and Liscovitch, M. (1990) *J. Biol. Chem.* **265,** 17512–17519.
8. Hammond, S. M., Altshuller, Y. M., Sung, T. S., Rudge, S. A., Rose, K., Engebrecht, J., Morris, A. M., and Frohman, M. A. (1995) *J. Biol. Chem.* **270,** 29640–29643.
9. Rose, K., Rudge, S. A., Frohman, M., Morris, A. J., and Engebrecht, J. (1995) *J. Biol. Chem.* 12151–12155.
10. Waksman, M., Eli, Y., Liscovitch, M., and Gerst, J. E. (1996) *J. Biol. Chem.* 2361–2364.
11. Brown, H. A., Gutowski, S., Moomaw, C. R., Slaughter, C., and Sternweis, P. (1993) *Cell* **75,** 1137–1144.
12. Tronchere, H., Planat, V., Record, M., Tercé, F., Ribbes, G., and Chap, H. (1995) *J. Biol. Chem.* **270,** 22, 13138–13146.
13. Bourgoin, S., Harbour, D., Desmarais, Y., Takai, Y., and Beaulieu, A. (1995) *J. Biol. Chem.* **270,** 3172–3178.
14. Ueki, J., Morioka, S., Komari, T., and Kumashiro, T. (1995) *Plant Cell Physiol.* **36,** 903–914.
15. Dyer, J. D., Zheng, L., and Wang, X. (1995) *Plant Physiol.* **109,** 1497–1501.