Inositol phospholipid-specific phospholipase C: Complete cDNA and protein sequences and sequence homology to tyrosine kinase-related oncogene products

PANN-GHILL SUH, SUNG HO RYU, KYUNG HO MOON, HAE WON SUH, AND SUE GOO RHEE*

Laboratory of Biochemistry, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892

Communicated by Earl R. Stadtman, April 1, 1988 (received for review March 1, 1988)

ABSTRACT Antibodies against an inositol phospholipid-specific phospholipase C purified from bovine brain were used to screen rat brain Agt11 expression cDNA libraries. Complete sequences of three cDNA inserts yielded a cumulative sequence of 5106 base pairs. The deduced protein had 1289 amino acids with a calculated molecular weight of 148,431. The determination of an open reading frame was aided by the amino acid sequences of 21 tryptic peptides isolated from bovine brain phospholipase C. Only 9 residues of a total of 140 amino acid residues determined for the bovine enzyme were different from those deduced from the rat cDNA. Two regions of phospholipase C (amino acid residues 555-598 and 668-705) exhibited significant amino acid similarities to the products of various tyrosine kinase-related oncogenes (yes, src, fgr, abl, fps, fes, and cck). The homologous domain was located in the region that is not essential for the protein-tyrosine kinase activity but is likely to be involved in an interaction with cellular components that modulate kinase function. Therefore, this unexpected similarity raises the possibility that the 148-kDa phospholipase C and cytoplasmic tyrosine kinases are modulated by common cellular component(s).

Inositol phospholipid-specific phospholipase C has attracted great attention in studies on the signal transduction mechanism since it hydrolyzes phosphatidylinositol 4,5-bisphosphate and generates two second messenger molecules, inositol 1,4,5-trisphosphate and diacylglycerol, when receptors are occupied by calcium-mobilizing hormones and neurotransmitters (reviewed in refs. 1–5). We purified three types of phospholipase C from bovine brain (6, 7). All are single-polypeptide enzymes and the molecular weights, measured under denaturing conditions, were 150,000, 145,000, and 85,000 for phospholipase C types I, II, and III, respectively. Rabbit polyclonal and murine monoclonal antibodies prepared against the three enzymes cross-reacted with their corresponding enzymes but not with the other two enzymes. Nevertheless, these enzymes are apparently similar in their catalytic properties: they are specific for inositol phospholipids, in that phosphatidylinositol 4,5-bisphosphate is the preferred substrate for all three enzymes. In view of the similar catalytic properties, the total immunological dissimilarity of the three enzymes is surprising. It would seem, therefore, that a full understanding of the structure of these isoenzymes might provide important insights into the molecular mechanism of signal transduction. In this paper, we report the cloning and sequencing of phospholipase C cDNA.

MATERIALS AND METHODS

Peptide Purification. The M, 145,000 phospholipase C type II (PLC-II) was purified from bovine brain by the procedure described (6, 7), and 2 mg of purified PLC-II was digested in 1 ml of 0.1 M Tris-HCl (pH 8.0) with 80 μg of trypsin. The resultant peptide mixture was injected into a HPLC AX 300 anion-exchange column (Synchronap, 4.1 × 250 mm) and eluted with a linear gradient of 0–1 M KCl in 50 mM Tris-HCl (pH 7.6). The eluent was divided into seven fractions, one fraction containing peptides unbound to the AX 300 column and six fractions eluted at other salt concentrations. The peptides in each fraction were further purified on a C18 reverse-phase column (Vydac 4.1 × 250 mm) with a linear gradient of 0–50% (vol/vol) CH3CN in 0.1% trifluoroacetic acid developed for 70 min. A total of 21 peptides was isolated, and their partial amino acid sequences were determined with the automated Edman degradation method in the laboratory of either Alan Smith (University of California, Davis) or Joseph Leykam (Michigan State University, East Lansing).

Other Methods. Poly(A) RNA was isolated from rat brain by the guanidine thiocyanate/cesium chloride method (8) and purified by oligo(dT)-cellulose chromatography. A rat brain cDNA library was constructed in the expression vector Agt11 by use of a protocol similar to that in ref. 9. Double-stranded cDNA was synthesized by using oligo(dT)-primed poly(A) mRNA and was size-fractionated on a Sepharose CL-4B column. DNA fragments >2 kilobase were ligated to EcoRI-cleaved Agt11. DNA was sequenced by the dideoxy chain-termination method (10) with bacteriophage T7 DNA polymerase (11). When bands on the sequencing gel were ambiguous and compressed due to the presence of G+C-rich regions, reaction mixtures containing dITP (instead of dGTP) were also included (11). Random sonicated DNA fragments were produced from concatemeric cDNA inserts (12) and cloned into the Sma I site of M13mp10. Each nucleotide was read an average of eight times and at least twice in both directions. Sequence data were assembled and analyzed by the use of the MicroGenie software (Beckman).

RESULTS AND DISCUSSION

We have prepared (6) 23 monoclonal antibodies against PLC-II. A mixture of 7 monoclonal antibodies, directed to different epitopes on PLC-II, was used to screen a rat brain Agt11 cDNA library (Clontech, Palo Alto, CA) as described by Young and Davis (13). Three immunologically positive clones were isolated from 1.2 × 108 transformants. The longest insert pPLC-II-3 (Fig. 1) [1.82 kilobase pairs (kb); nucleotides 2077–3898] was then completely sequenced by the randomly sonicated DNA sequencing strategy. The sequencing results revealed that pPLC-II-3

Abbreviation: PLC-II, phospholipase C type II.

*To whom reprint requests should be addressed: National Institutes of Health, 9000 Rockville Pike, Building 3, Room 203, Bethesda, MD 20892.

The sequence reported in this paper is being deposited in the EMBL/GenBank data base (IntelliGenetics, Mountain View, CA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J03806).
Fig. 1. Schematic representation of phospholipase C cDNA clones. Overlapping clones and the cumulative complete cDNA structure are given. Thick lines and boxes indicate the untranslated regions and coding regions, respectively. Several restriction endonuclease cleavage sites are shown.

 carries in the same reading frame coding sequences for seven trypic peptides (peptides 12, 13, 14, 16, 17, 19, and 20; see Fig. 2), isolated from bovine brain PLC-II. In addition, amino acid sequences homologous to three trypic peptides of bovine PLC-II (peptides 15, 18, and 21; Fig. 2) were found in pPLC-II-3. When the same library was rescreened by plaque hybridization with pPLC-II-3 as probe, no cDNA longer than the probe was isolated. Therefore, a new library was constructed with the Agt11 vector system using EcoRI linkers and rat brain poly(A) RNA.

Upon screening the new library of 5 × 10^6 independent clones with pPLC-II-3 as probe, one clone, pPLC-II-68 that had a 4778-nucleotide insertion (nucleotides 341–5118 in Fig. 2), was isolated. Complete sequence analysis revealed that the insert in pPLC-II-68 contains the polyadenylation signal (14) AATAAA 20 nucleotides upstream from the poly(A) tract and sequences coding for 20 PLC-II peptides (14 identical and 6 closely matched peptides). However, this clone was not long enough (when a 3'-untranslated region of 1143 base pairs was excluded) to code the entire PLC-II

FIG. 2. (Figure continues on the opposite page.)

Nucleotide sequence of rat brain phospholipase C and deduced amino acid sequence. Nucleotide residues are numbered in the 5' → 3' direction, beginning with the first residue of the ATG triplet encoding the initiating methionine, and the nucleotides on the 5' side of residue 1 are indicated by negative numbers. Nucleotide residue numbers are at the right. Deduced amino acid residues are numbered beginning with the initiating methionine and the amino acid residues are numbered at the left. The regions identifying partial amino acid sequences of bovine brain PLC-II peptides are underlined and sequentially numbered. Amino acid residues in parentheses differ between rat and bovine sequences.
molecule that is estimated to have a molecular weight of 145,000. In addition, the first ATG triplet (nucleotides 394–396) in the 5' region of pPLC-II-68 appeared to be unsuitable as the initiation codon. To obtain additional cDNA clones upstream of pPLC-II-68, a primer-extension method was employed to prepare another cDNA library. The 17-mer oligonucleotide whose sequence is complementary to the sequence (nucleotides 718–734 in Fig. 2) located near the 5' end of the pPLC-II-68 was synthesized and used as a primer. The cDNA synthesized was ligated to the pgi11 vector with EcoRI linkers (CGGAATTCCG). The new library was screened with a 5'-most Xho I fragment (nucleotides 341–1635; 1.3 kb) prepared from pPLC-II-68. Six clones with an insert >0.7 kb were isolated from 2 x 10⁶ transfectants. The EcoRI-digested phage DNA from all six clones exhibited two DNA bands in addition to phage DNA, alluding to the presence of an internal EcoRI site in the full-length cDNA encoding PLC-II. When the two EcoRI fragments isolated from a clone pPLC-II-23 (nucleotides 94 to 686; 0.78 kb) were sequenced, a sequence for EcoRI site (TGGAGTCCG, nucleotides 338–347 in Fig. 2), which is different from the sequence of EcoRI linker, was located. The 3'-EcoRI fragment of pPLC-II-23 contained an open reading frame that included sequences of two bovine PLC-II peptides; one sequence exactly matches that of peptide 2 and the other imperfectly matches the sequence of peptide 3 in Fig. 2. The same reading frame of the 5' EcoRI fragment of pPLC-II-23 also contained an amino acid sequence that was homologous to that of bovine PLC-II peptide (Fig. 2, peptide 1) in which glutamine is conservatively changed to asparagine. Furthermore, the 5' EcoRI fragment of pPLC-II-23 and the pPLC-II-68 recognized the same mRNA band of ~5.3 kb when poly(A) mRNA from rat brain was subjected to RNA gel blot analysis (Fig. 3).

The cumulative cDNA sequence of pPLC-II-3, -68, and 23 is 5106 base pairs long [excluding poly(A) tract] as shown in Fig. 2 and revealed an open reading frame that starts with a potential initiation codon. The first ATG triplet is flanked by sequences that fulfill the Kozak criteria for initiation codons (15) and lies within a 100-nucleotide G+C-rich sequence (86%). A translational termination codon (TAG) occurs in-frame after codon 1290 specifying a leucine. The open reading frame initiated by this ATG codon encodes a 1289-amino acid polypeptide (excluding the initiating methionine). The calculated molecular weight of 148,431 is close to that observed for the purified bovine brain PLC-II (6, 16). Among a total of 21 tryptic peptides sequenced for bovine PLC-II, 14 sequences could be matched exactly and the remaining 7 sequences could not perfectly be closely matched to various parts of the deduced amino acid sequence. The mismatching amino acids are indicated inside parentheses underneath the rat brain PLC-II sequence. The fact that the N-terminal amino acids of the 7 imperfectly matching peptides are preceded by either lysine or arginine is in accord with the catalytic specificity of trypsin. In all, the sequences of 140 amino acid residues were established from the 21 PLC-II peptides. Among these, only 9 amino acid residues differed between the bovine and rat enzymes. This excellent sequence identity is consistent with the fact that only 1 out of 23 monoclonal antibodies against bovine brain PLC-II could not recognize rat brain PLC-II.

Fig. 3. RNA gel blot analysis of phospholipase C mRNA. Polyadenylated RNA (5 μg per lane) of rat whole brain was electrophoresed in 1% agarose gels containing 2.2 M formaldehyde and blotted to a nylon filter (Nytran; 0.45-μm pore size; Schleicher & Schuell). The filters were hybridized to either the 32P-labeled pPLC-II-68 (lane A) or the 32P-labeled 5' EcoRI fragment of pPLC-II-23 (lane B) and then subjected to autoradiography. Molecular size standards (in kb) are a Bethesda Research Laboratories "RNA ladder."
Fig. 5. Schematic representation of two regions of sequence identity for cytoplasmic protein-tyrosine kinases. Regions A and B represent noncatalytic and catalytic domains, respectively.

(P.-G.S., S.H.R., W. C. Choi, and S.G.R., unpublished results). The hydrophaticity profile (data not shown) of PLC-II amino acid sequence does not show significant hydrophobic segments shared with other transmembrane proteins. This is consistent with the fact that PLC-II exists mainly in cytosolic fraction (6).

A search of the GenBank3 (Beckman MicroGenie software) found that there are two narrow spans of amino acid sequences in PLC-II that are similar to tyrosine kinase-related oncogene products. As shown in Fig. 4, two regions of PLC-II (amino acid residues 555–598 and 668–705) exhibit a significant amount of sequence identity to the products of the avian sarcoma virus Y-73 v-ray, the Rous sarcoma virus v-src, the Drosophila melanogaster c-src, the Gardener-Rasheed feline sarcoma virus v-fgr, the Abelson murine leukemia virus v-abl, the Caenorhabditis elegans c-abl, the Fujinami sarcoma virus v-fps, the human c-fes, and the mouse FcK. These cytoplasmic oncogene proteins possess two contiguous regions of sequence identity, a C-terminal catalytic domain of 260 residues (domain B designated in Fig. 5) with homology to other tyrosine kinases that are related to transmembrane growth factor receptors and a noncatalytic domain (domain A designated in Fig. 5) that lies immediately on the N-terminal side of the catalytic domain and is absent from kinases that span the plasma membrane (26). The region homologous to phospholipase C is located in domain A. A series of experimental data suggests that domain A is not essential for catalytic activity but may have some important role in directing the cellular actions of the kinase domain. The catalytic domain that lacks domain A sequences could be released from p60v-src (27), p130ps-a-fps (28), and p90ps-a-fps (27) by limited proteolysis as C-terminal 29- to 30-kDa fragments that retain full enzymatic activity. Nevertheless, mutations that infringe on domain A largely abolish the transforming activity of p60v-src (29), p130ps-a-fps (26), and p120ps-a-fps (30). For example, deletion of p60v-src amino acids 149–169 (31) or insertion of dipeptide between Glu-831 and Leu-832 of p130ps-a-fps (26) impaired transforming activity with no obvious effect on tyrosine kinase.

At the present time it is not clear how the noncatalytic domain A affects the ability of tyrosine kinase to transform a host cell. One possibility is, as proposed by Sadowski et al. (26), that the noncatalytic domain directs specific interactions of tyrosine kinases with cellular components that mediate tyrosine kinase function. This proposal is generally consistent with the observation that p60v-src, p130ps-a-fps, and p90ps-a-fps form complexes with two host proteins, a 90-kDa heat shock protein and a 50-kDa tyrosine kinase substrate (32–34). In addition, a sequence encompassing residues 155–160 with domain A of p60v-src was shown to be involved in the complex formation (35). The fact that phospholipase C contains a region homologous to domain A allows speculation that phospholipase C and nonreceptor tyrosine kinases include common regulatory domains that respond to the same or similar regulatory effectors.

We are grateful to Dr. Shiko Kimura for valuable advice during this work.