by this universally present water molecule is presumably activated through the abstraction of a proton by the His48. A formal mechanism for PLA_2 catalysis has been advanced that is based on the stereochemistry described here (*2*) and in a parallel study on the inhibited bee-venom enzyme (*7*).

REFERENCES AND NOTES

1. J. Chang, J. H. Musser, H. McGregor, *Biochem. Pharmacol.* **36**, 2429 (1987); A. Achari *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **52**, 441 (1987); B. Samuelsson, *Drug* **33**, 2 (1987); ———, S.-E. Duhlen, J. Å. Lindgren, C. A. Rouzer, C. N. Serhan, *Science* **237**, 1171 (1987).
2. D. L. Scott *et al.*, *Science* **250**, 1541 (1990).
3. T. L. Hazlett and E. A. Dennis, *Biochim. Biophys. Acta* **961**, 22 (1988); A. G. Tomasselli *et al.*, *J. Biol. Chem.* **264**, 10041 (1989); W. Cho, A. G. Tomasselli, R. L. Heinrikson, F. J. Kezdy, *ibid.* **263**, 11237 (1988).
4. B. W. Dijkstra, K. H. Kalk, W. G. J. Hol, J. Drenth, *J. Mol. Biol.* **147**, 97 (1981).
5. D. L. Scott, A. Achari, S. Brunie, P. B. Sigler, in preparation.
6. R. Renetseder, S. Brunie, B. W. Dijkstra, J. Drenth, P. B. Sigler, *J. Biol. Chem.* **260**, 11627 (1985). The root mean square of a collection of N values x_i is $[(\Sigma x_i^2)/N]^{1/2}$. Here x_i is the distance between the ith pair of a set of N corresponding atoms after superimposition of two molecules or molecular segments. The program used was originally developed by A. Lesk.
7. D. L. Scott, Z. Otwinowski, M. H. Gelb, P. B. Sigler, *Science* **250**, 1563 (1990).
8. C. C. Yang, K. King, T. P. Sun, *Toxicon* **19**, 141 (1980).
9. M. Elder, P. Hitchcock, R. Mason, G. G. Shipley, *Proc. R. Soc. London Ser. A* **354**, 157 (1977); R. H. Pearson and I. Pascher, *Nature* **281**, 499 (1979); H. Hauser, I. Pascher, R. H. Pearson, S. Sundell, *Biochim. Biophys. Acta* **650**, 21 (1981).
10. C. J. van den Bergh, A. J. Slotboom, H. M. Verheij, G. H. de Haas, *J. Cell Biochem.* **39**, 379 (1989).
11. Although there is some overlap of the interfacial binding surface described by Dijkstra *et al.* [B. W. Dijkstra, J. Drenth, K. W. Kalk, *Nature* **289**, 604 (1981)] and that described here, the direction of the substrate's approach to the catalytic surface implicit in the description of Dijkstra *et al.* is nearly perpendicular to that suggested by the orientation of the bound substrate analogue.
12. J. P. Abita, M. Lazdunski, P. P. M. Bonsen, W. A. Pieterson, G. H. de Haas, *Eur. J. Biochem.* **30**, 37 (1972): M. A. Pieterson, J. J. Volwerk, G. H. de Haas, *Biochemistry* **13**, 1455 (1974).
13. M. K. Jain and W. L. C. Vaz, *Biochim. Biophys. Acta* **905**, 1 (1987); R. D. Ludescher *et al.*, *Biochemistry* **27**, 6618 (1988).
14. H. Einspahr and C. E. Bugg, in *Calcium Binding Proteins and Calcium Function*, R. H. Wasserman *et al.*, Eds. (Elsevier North-Holland, New York, 1977), pp. 13–19; A. Persechini, N. D. Moncrief, R. H. Kretsinger, *Trends Neurosci.* **12**, 462 (1989); M. N. G. James, *Annu. Rev. Biochem.* **58**, 951 (1989).
15. Calcium ions are identified by the refinement assigning a compact density of 18 electrons to the site with a temperature factor consistent with those of the surrounding atoms. Attempts to assign a water molecule to the same density caused the refinement to simulate an even more compact higher density by imposing a physically unrealistic negative temperature factor. There is no way to confidently distinguish a calcium from the isoelectronic potassium ion by x-ray crystallography; however, no potassium salts were used in the preparation or purification of the proteins, $diC_8(2Ph)PE$, or any of the crystal-stabilizing supernatant solutions.
16. H. M. Verheij *et al.*, *Biochemistry* **19**, 743 (1980).
17. The coefficients of the Fourier sum are $(2F_o - F_c)\exp(i\phi_c)$, where F_o is the observed amplitude and F_c and ϕ_c are the amplitude and phase, respectively, of the structure factors calculated from the refined model.
18. W. Yuan *et al.*, *Biochemistry* **29**, 6082 (1990).
19. R. A. Crowther, in *The Molecular Replacement Method*, M. G. Rossman, Ed. (Gordon and Breach, New York, 1972, pp. 173–178; R. A. Crowther and D. M. Blow, *Acta Crystallogr.* **23**, 544 (1967).
20. P. M. D. Fitzgerald, *J. Appl. Crystallogr.* **21**, 273 (1988).
21. A. T. Brünger, J. Kuriyan, M. Karplus, *Science* **235**, 458 (1987).
22. W. A. Hendrickson and J. H. Konnert, *Biomolecular Structure, Function, Conformation and Evolution*, R. Srinivasan, Ed. (Pergamon, Oxford, 1981), vol. 1, pp. 43–57.
23. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

24. The research at Yale was supported by NIH grant GM22324 and by the Howard Hughes Medical Institute; the research at the University of Washington was supported by NIH grant HL 36235. S.P.W. is a fellow of the Arthritis Foundation, and D.L.S. is a postgraduate fellow at Yale.

4 June 1990; accepted 27 September 1990

Crystal Structure of Bee-Venom Phospholipase A_2 in a Complex with a Transition-State Analogue

DAVID L. SCOTT, ZBYSZEK OTWINOWSKI, MICHAEL H. GELB, PAUL B. SIGLER

The 2.0 angstroms crystal structure of a complex containing bee-venom phospholipase A_2 (PLA_2) and a phosphonate transition-state analogue was solved by multiple isomorphous replacement. The electron-density map is sufficiently detailed to visualize the proximal sugars of the enzyme's N-linked carbohydrate and a single molecule of the transition-state analogue bound to its active center. Although bee-venom PLA_2 does not belong to the large homologous Class I/II family that encompasses most other well-studied PLA_2s, there is segmental sequence similarity and conservation of many functional substructures. Comparison of the bee-venom enzyme with other phospholipase structures provides compelling evidence for a common catalytic mechanism.

PHOSPHOLIPASES A_2 (PLA_2, EC 3.1.1.4) specifically hydrolyze the 2-ester bond of *L*-glycerophospholipids. These enzymes have been purified from a variety of sources including mammalian pancreas, reptile and insect venoms, and synovial fluid. Since certain cellular forms of the enzyme may catalyze the release of arachidonate and thereby precipitate the inflammatory cascade, modulation of PLA_2 activity is of great pharmacological interest.

The amino acid sequences of the large homologous family of PLA_2s have been divided into two closely related structural classes, Class I (pancreatic juice and elapid venom) and Class II enzymes (crotalid and viper venoms) (*1*). The chemically determined sequence of the PLA_2 from the honeybee (*Apis mellifera*) (*2*) indicated that this enzyme was structurally distinct from the Class I/II superfamily. The amino acid sequence recently deduced from a cDNA clone differs from the chemically determined one, but both suggest that segments containing residues involved in calcium binding and catalysis are conserved (*3, 4*). Bee-venom PLA_2, which is the principal allergen of bee venom, also differs from most other PLA_2s in that it contains an asparagine-linked oligosaccharide whose effect on enzymatic activity and allergenicity is currently under study (*5*). With the possible exception of its behavior toward aggregated substrates, the activity of the bee-venom enzyme is similar to that of other PLA_2s (*6*).

Crystal structures are available for representative Class I and Class II enzymes (*7, 8*). We report the crystal structure of bee-venom PLA_2 in a complex with a transition-state analogue, $diC_8(2Ph)PE$ (Table 1 and Fig. 1). As expected, the conserved sequence segments of the bee-venom PLA_2 preserve the functional substructures found in Class I/II enzymes but are arranged within a different overall architecture (Fig. 2). The interaction of the inhibitor with the bee-venom enzyme shows how the rate-limiting formation of a putative tetrahedral intermediate is fostered by the enzyme's catalytic components and defines the mechanism by which the hydrophobic alkyl moieties of the phospholipid participate in productive-mode binding.

In Fig. 3 the position of $diC_8(2Ph)PE$ is

D. L. Scott, Z. Otwinowski, P. B. Sigler, Department of Molecular Biophysics and Biochemistry and the Howard Hughes Medical Institute, Yale University, New Haven, CT 06511.
M. H. Gelb, Departments of Chemistry and Biochemistry, University of Washington, Seattle, WA 98195.

shown in relation to the Cα backbone of bee-venom PLA_2. The interaction of the transition-state analogue with the catalytic residues is detailed in Fig. 4. His34 corresponds to the active site His48 of Class I/II PLA_2s and, like its counterpart in the large homologous family, immediately precedes the calcium-binding Asp residue in the sequence (*7, 9*). Asp35 is functionally identical to the Asp49 of other PLA_2s. Its carboxylate contributes two of the equatorial ligands to a hepta-coordinated pentagonal bipyramidal calcium-binding cage. Similarly, Asp64 is the counterpart of Asp99, which interacts with His34(48) and presumably neutralizes the positive charge on His34(48) created by bond cleavage during catalysis. In the catalytic network of the Class I/II family there are two absolutely conserved Tyr residues, Tyr52 and Tyr73, whose phenolic hydroxyls interact with the carboxylate oxygen of Asp99. The bee-venom enzyme contributes only one tyrosine, Tyr87, which arises from a nonanalogous position.

In the Class I/II homologous family, the first turn of the amino-terminal α helix is

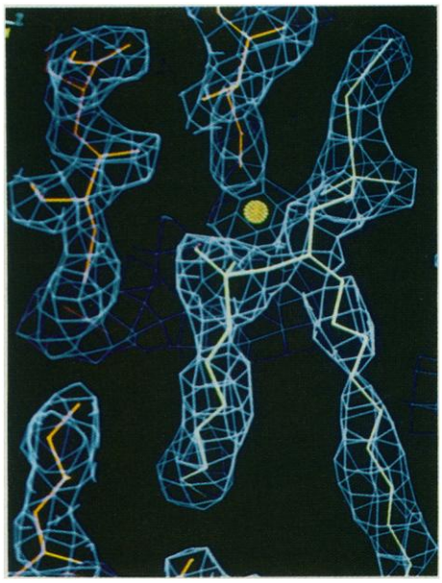

Fig. 1. The electron density of the transition-state analogue, *L*-1-*O*-octyl-2-heptylphosphonyl-*sn*-glycero-3-phosphoethanolamine [diC_8(2Ph)PE]. A short-chain member of a class of strong PLA_2 inhibitors studied by Gelb and co-workers (*17*), in which a phosphonate replaces the *sn-2* ester. This analogue was designed to emulate the transition state formed in the hydrolysis of 1,2-dioctanoyl-*sn*-3-phosphoethanolamine and is therefore designated as diC_8(2Ph)PE. The conformation shown here is that seen in the complex formed with the PLA_2 from bee venom as electron density from a $(2F_o - F_c)$ map (*18*). Yellow ball: calcium ion.

Table 1. Multiple isomorphous replacement (MIR) of bee-venom PLA_2 complexed with diC_8(Ph)PE. Purified bee-venom PLA_2 was purchased from Boehringer, exhaustively dialyzed against distilled water, and then lyophilized. Large (0.4 mm by 0.4 mm by 0.3 mm) single crystals suitable for diffraction work were obtained in 3 days by plating 20-μl droplets containing 10 mg/ml protein, 0.1 M tris, 3 mM diC_8(2Ph)PE, 10 mM $CaCl_2$, 30% saturated NaCl, pH 7.6, onto depression slides in sealed boxes containing 10 ml of 60% saturated NaCl, 0.2 M tris, pH 7.6. The crystals were of space group $I4_122$, $a = b = 89.5$ Å, and $c = 132.5$ Å, with one molecule in the asymmetric unit. Heavy-atom derivatives were prepared by conventional soaking techniques. Crystals were stabilized in 50% saturated NaCl, 0.1 M tris, 10 mM $CaCl_2$, 1 mM diC_8(2Ph)PE, pH 7.2, with the appropriate concentration of heavy atoms. Typical soaking times were less than 1 day. The replacement of calcium ion with barium provided a derivative in which the position of the barium was easily found in a difference Patterson map. The resulting single isomorphous replacement phases were used to solve multiple-site derivatives. Heavy-atom binding sites and occupancies were refined with a modified version of the program PHARE (*15*) that (i) used anisotropic temperature factors to better resolve closely spaced or clustered sites and (ii) reduced the bias contributed by a particular heavy atom to its own refinement. The final MIR map permitted tracing of the main chain and most side chains. Solvent flattening within iterative cycles of phase extension improved the map quality without appreciably changing the original interpretation (*16*). A difference map calculated from the Bijvoet differences of the native intensities and the MIR phases confirmed the location of all five disulfide bridges and the calcium ion. Refinement converged readily to an *R* factor of 0.198 for data in the resolution range 8.3 to 2.0 Å ($F > 1.5\ \sigma_F$). On average, bond lengths, bond angle distances, and planarity deviated from ideal values by less than 0.021, 0.043, and 0.025 Å, respectively. The carbohydrate was not refined.

Structure parameter	Resolution (Å)								
	10.67	8.00	6.40	5.33	4.57	4.00	3.56	3.20	Overall
Figure of merit*	0.93	0.95	0.91	0.89	0.86	0.78	0.69	0.55	0.73
Phasing power of derivative†									
$BaSO_4$	1.09	0.95	1.13	1.18	1.12	0.91	0.84	0.85	0.93
$HgCl_2$	5.10	4.91	3.68	3.63	2.84	2.03	1.50	1.09	2.11
$TlCl_3$	5.04	4.95	3.94	4.42	4.07	2.63	2.22	1.77	3.20
Ethyl-HgCl	0.63	0.69	0.84	0.65	0.63	0.50	0.43	0.40	0.55

*Figure of merit is a measure of the relative reliability of a phase based on the consistency of the MIR analysis. The maximum value is 1.0; the minimum value is 0.0. †The phasing power of a derivative equals the calculated strength of the heavy-atom contribution divided by the estimate of error in the analysis.

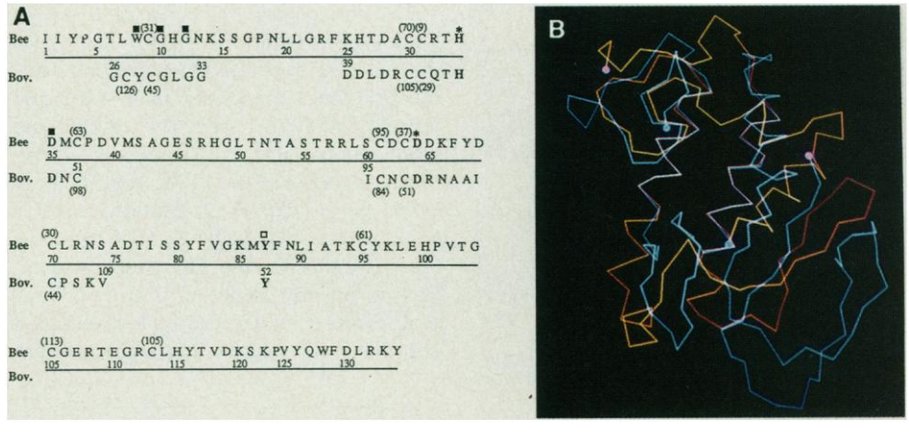

Fig. 2. Comparison of bee-venom PLA_2 with Class I/II enzymes. (**A**) Sequence of bee-venom PLA_2 deduced from a cDNA sequence (*3, 19*). Segments of the Class I/II enzymes (here bovine pancreatic) (*7*) that are structurally "homologous" are included below the respective bee-venom segments (*20*). Numbers in parentheses above or below Cys residues indicate their disulfide partner as seen in the crystal structure; this arrangement differs from those suggested by sequence analysis (*3*). Boldface highlights functionally critical residues: (*), catalytic; (■), calcium binding; and (□), analogue of the invariant supporting Tyr52 of the Class I/II superfamily. (**B**) The Cα trace of bee-venom PLA_2 (red) superimposed upon that of the uninhibited form of the bovine pancreatic PLA_2 (blue) (*7*). Homologous segments of the backbone were superimposed en bloc by a least squares fit of corresponding Cα atoms (*10*). The homology was limited to the segments shown in (A). Similarity of the three-dimensional backbone is limited to the calcium-binding loop and the two long antiparallel helices. Residues 9 to 13 of the bee enzyme correspond to the conserved residues (28 to 32) of the calcium-binding loop of Class I/II phospholipases. Bee-venom PLA_2s antiparallel α-helices (residues 25 to 37 and residues 61 to 74) contain the conserved catalytic residues of the homologous core. Like the Class I/II structures, these helices are held together by disulfide bridges at both ends.

stabilized by a hydrogen-bond network that includes the terminal amino group, the nearly invariant Gln4, elements of residues 69 to 71, and a water-mediated link to the catalytic residues. In the bee-enzyme structure, there is no contact (direct or indirect) between the amino terminus and the catalytic residues. Asp64(99) is connected, however, by a similar network of water-mediated hydrogen bonds to the backbone nitrogen of Thr59, which occupies a position that is stereochemically analogous to the Class I/II amino terminus. Conservation of this network suggests that it is important in stabilizing the active site geometry.

Like the pancreatic enzymes, the transformation of bee-venom PLA_2 from a proenzyme to its activated form also involves the creation of a new amino terminus. Although the amino terminus of bee-venom PLA_2 does not form an α helix and is not connected to the active site, an amino-terminal extension of the enzyme (prepro- or proenzyme) could interfere with interfacial binding by disrupting the hydrogen bond between Ile1 and His11 that helps stabilize the calcium-binding loop or by directly blocking access to the hydrophobic channel or by both processes.

The conformation of bee-venom's calcium-binding loop is strikingly similar to that described for Class I/II enzymes and conserves the general calcium-binding sequence (X-Cys-Gly-X-Gly) found in all catalytically active PLA_2s (Fig. 2). Gly10 and Gly12, which contribute their carbonyl oxygens to bee venom's calcium ion cage, are universally conserved components of PLA_2 calcium-binding domains, whereas Trp8 replaces the conserved Tyr of Class I/II enzymes (*10*).

The bound calcium ion is hepta-coordinated by a pentagonal bipyramidal cage of O atoms. One of the equatorial oxygens of the calcium-binding cage is provided by the phosphonate, which we interpret to simulate the stabilizing effect of the calcium ion on the oxyanion of the tetrahedral transition state. (The oxyanion is further stabilized by a hydrogen bond to N10.) An axial oxygen is provided by the *sn-3* phosphate. These bound substrate analogue oxygens occupy the same sites as two conserved water oxygens that contribute to the ligation cage of uncomplexed Class I/II enzymes.

Figure 3 shows the hydrophobic channel that encloses the roughly parallel alkane side chains of the *sn-1* and *sn-2* substituents and extends approximately 9 Å from the catalytic site (His34 Nδ1) to its opening on the enzyme's surface just below the amino terminus. Arg57, which contributes its proximal methylene groups to the hydrophobic channel, forms a flexible wall that is locked into place by a hydrogen bond between the guanidino group and the *sn-3* phosphate. The analogue's *sn-2* substituent is sharply bent at the tetrahedral phosphonate in a manner reminiscent of crystalline phospholipids (Fig. 1) (*11*). However, unlike the crystal structures of phospholipids, the *sn-3* phosphate and its amino alcohol do not fold back over the *sn-2* esters but project away to form highly polar interactions. As the alkane side chains are well ordered, this channel then provides a high-resolution view of the interaction between the fatty-acid chains of a phospholipid and a protein surface that is designed to face them. These interactions, as

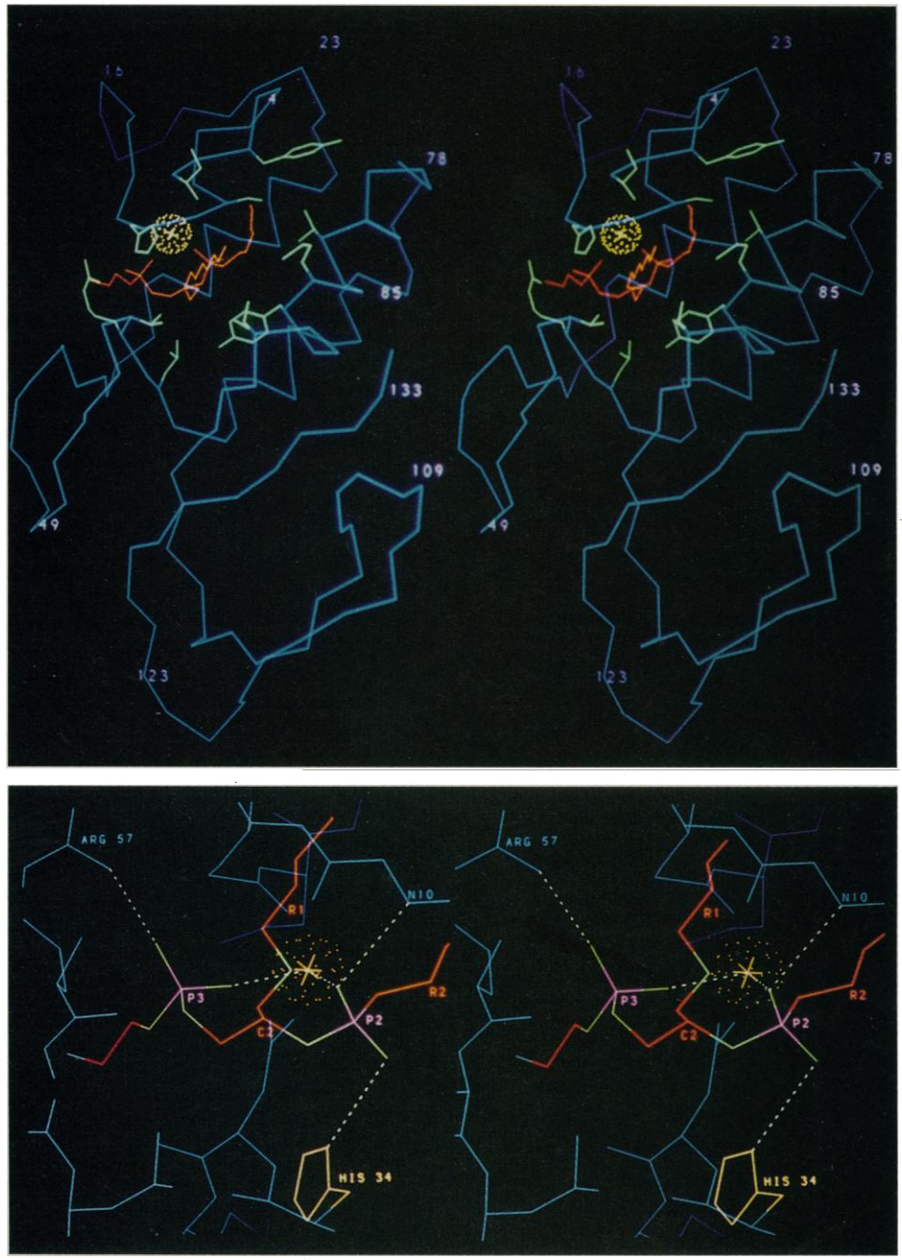

Fig. 3. (**Top**) Complex of $diC_8(2Ph)PE$ with the PLA_2 from bee venom. The Cα trace highlighting the calcium ion cofactor (yellow sphere), the $diC_8(2Ph)PE$ inhibitor (red), and side chains of the hydrophobic channel and the proposed interfacial binding surface (green) (Ile1, Tyr3, Cys9, His11, Thr56, Arg57, Leu59, Val83, Met86, Tyr87, and Ile91). **Fig. 4.** (**Bottom**) Interaction of $diC_8(2Ph)PE$ with the active site of bee-venom PLA_2. Protein components are colored blue except for the gold side chain of the active site His34. The calcium ion is represented by the yellow sphere. In $diC_8(2Ph)PE$, the P atoms of the *sn-2* phosphonate (P2) and *sn-3* phosphate (P3) are magenta, O green, N cyan, and C red. The termini of the *sn-1* (R1) and *sn-2* (R2) alkyl substituents are clipped off in this view. Arg57 completes the hydrophobic channel by a stabilizing contact with the *sn-3* phosphate. The nonbridging oxygen of the *sn-3* phosphate that is coordinated by the principal calcium ion is also hydrogen bonded to N12 of the calcium-binding loop (bond not shown).

well as the calcium-ligation scheme are effectively identical to those seen in the crystal structure of the corresponding *N. n. atra* complex (*12, 13*).

The distal end of the calcium-binding loop bears bee venom's single oligosaccharide, which is N-glycosylated to the side chain of Asn13. Although the inhibitor's *sn-3* amino alcohol makes contact with the proximal sugar, the carbohydrate moiety must be refined before the nature of this contact can be inferred. Electron density for only the proximal three sugars can be found; the more distal saccharides extend into bulk solvent and are apparently disordered. The function of this carbohydrate in enzymatic action (if any) is unclear. Non-glycosylated bee-venom PLA_2 occurs naturally as a minor variant but retains normal activity (*14*).

When compared with its Class I/II relatives, the bee-venom crystal structure provides a clear indication of those components that are essential to PLA_2 function. In a companion article, we interpret these findings in a unifying structural model for PLA_2 catalysis (*13*).

REFERENCES AND NOTES

1. R. L. Heinrikson, E. T. Krueger, P. S. Keim, *J. Biol. Chem.* **252**, 491 (1977).
2. R. A. Shipolini, G. L. Callewaert, R. C. Cottrell, C. A. Vernon, *Eur. J. Biochem.* **48**, 465 (1974).
3. K. Kuchler, M. Gmachi, M. J. Sippl, G. Kreil, *ibid.* **184**, 249 (1989).
4. J. M. Maragnore, R. A. Poorman, R. L. Heinrikson, *J. Protein Chem.* **6**, 173 (1987).
5. A. Weber, H. Schroder, K. Ihalberg, L. Marz, *Allergy* **42**, 464 (1987).
6. G. Lin, J. Noel, W. Loffredo, H. Z. Stable, M. Tsai, *J. Biol. Chem.* **263**, 13208 (1988).
7. B. W. Dijkstra, K. H. Kalk, W. G. J. Hol, J. Drenth, *J. Mol. Biol.* **147**, 97 (1981).
8. S. Brunie, J. Bolin, D. Gewirth, P. B. Sigler, *J. Biol. Chem.* **260**, 9742 (1985).
9. E. A. M. Fleer, H. M. Verheij, G. H. de Haas, *Eur. J. Biochem.* **82**, 261 (1978). When designating a functionally important residue, the first number refers to the position in the bee-venom sequence. The second number is the corresponding sequence number in the Class I/II superfamily from the common numbering system of Renetseder *et al.* (*10*).
10. R. Renetseder, S. Brunie, B. W. Dijkstra, J. Drenth, P. B. Sigler, *J. Biol. Chem.* **260**, 11627 (1985).
11. M. Elder, P. Hitchcock, R. Mason, G. G. Shipley, *Proc. R. Soc. London Ser A* **354**, 157 (1977); R. H. Pearson and I. Pascher, *Nature* **281**, 499 (1979); H. Hauser, I. Pascher, R. H. Pearson, S. Sundell, *Biochim. Biophys. Acta* **650**, 21 (1981).
12. S. White, D. L. Scott, Z. Otwinowski, M. H. Gelb, P. B. Sigler, *Science* **250**, 1560 (1990).
13. D. L. Scott *et al.*, *ibid.*, p. 1541.
14. L. Marz, C. Mollay, G. Kreil, G. Zelger, *Insect Biochem.* **11**, 685 (1981).
15. J. E. Ladner and G. Bricogne, version 9A2 of PHARE kindly provided by the authors (1974); Z. Otwinowski, manuscript in preparation.
16. B.-C. Wang, *Methods Enzymol.* **115**, 90 (1985).
17. W. Yuan and M. H. Gelb, *J. Am. Chem. Soc.* **110**, 2665 (1988); W. Yuan, D. M. Quinn, P. B. Sigler, M. H. Gelb, *Biochemistry* **29**, 6082 (1990).
18. The coefficients of the Fourier sum are $(2F_o - F_c)\exp(i\phi_c)$ where F_o is the observed amplitude and F_c and ϕ_c are the amplitude and phase, respectively, of the structure factors calculated from the refined model.
19. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
20. The segmental correspondence of sequence and structure that relates the bee-venom PLA_2 to the canonical homologous family suggests the evolutionary process of "exon shuffling" [W. Gilbert, *Cold Spring Harbor Symp. Quant. Biol.* **52**, 901 (1987)]. If exon shuffling is operative it is a more subtle and complicated process than normally envisioned. For example, the second exon [J. J. Seilhamer *et al.*, *J. Biol. Chem.* **264**, 5335 (1986)] includes the amino-terminal helix and calcium-binding loop. The bee-venom enzyme appears to have split the first exon. It shows no amino-terminal helix and a perfectly conserved calcium-binding loop. The α-helical segment (residues 76 to 93) that replaces the function of the amino-terminal helix has the opposite polarity and an altogether different sequence.

21. We thank L. Marz for his help early in this project in supplying glycosylated and unglycosylated enzyme and G. Kreil for access to the bee-venom PLA_2 sequence prior to publication. The research at Yale was supported by NIH grant no. GM22324 and by the Howard Hughes Medical Institute; the research at the University of Washington was supported by NIH grant no. HL36235. D.L.S. is a postgraduate fellow at Yale and a graduate student pro forma at the University of Chicago.

4 June 1990; accepted 27 September 1990

Bacterial Origin of a Chloroplast Intron: Conserved Self-Splicing Group I Introns in Cyanobacteria

MING-QUN XU, SCOTT D. KATHE, HEIDI GOODRICH-BLAIR, SANDRA A. NIERZWICKI-BAUER, DAVID A. SHUB*

A self-splicing group I intron has been found in the gene for a leucine transfer RNA in two species of *Anabaena*, a filamentous nitrogen-fixing cyanobacterium. The intron is similar to one that is found at the identical position in the same transfer RNA gene of chloroplasts of land plants. Because cyanobacteria were the progenitors of chloroplasts, it is likely that group I introns predated the endosymbiotic association of these eubacteria with eukaryotic cells.

THE DISCOVERY OF INTRONS IN EUkaryotic genes has stimulated considerable discussion of their origin and evolution. One hypothesis assumes a relatively recent, eukaryotic origin for introns. According to this view, introns are mobile elements that can be transferred between genes, between nuclear and organelle genomes, and horizontally between organisms (*1*). A second hypothesis proposes that introns arose very early, perhaps in the "progenote" ancestor of all living organisms (*2*). According to this scheme, rapidly dividing microorganisms have lost their introns as a result of selection for rapid DNA replication. On the other hand, eukaryotes, under reduced pressure to streamline their genomes, have retained their introns.

Only circumstantial support has been found for either of these views. The demonstration that the group I intron of *Tetrahymena* is able to act as a true enzyme (*3*) added support for the antiquity of introns: self-splicing introns could have been present in the earliest genes of a precellular "RNA world" (*4*). Furthermore, it has been suggested that group I, group II, and nuclear mRNA introns—all of which splice through a series of transesterification reactions—may be evolutionarily related (*5*). However, no trace could be found in contemporary bacteria of these once ubiquitous, primordial introns. Rather, the only prokaryotic introns were in tRNA and ribosomal RNA (rRNA) genes of archaebacteria. The splicing mechanism of these introns is entirely different, being catalyzed by protein enzymes that seem to recognize structural features of the precursor RNA, and is similar to that of introns of eukaryotic nuclear tRNA genes (*6*).

The discovery of group I self-splicing introns in bacteriophages of both Gram-negative and Gram-positive eubacteria (*7, 8*) did not resolve this controversy, because the origin and evolution of viruses are themselves unclear. A convincing argument has been made that the large, tailed, DNA phages are ancient, originating in the earliest eubacteria. However, viruses are genetic mosaics and any particular virus might be a relatively recent assemblage of genes from different sources (*9*). Indeed, some introns (including two in phage T4) encode proteins that mobilize the transfer of the intron to "homing" sites in DNA (*10*), providing further support for the spread of introns through horizontal gene transfer.

In considering a likely source of introns in contemporary eubacteria, we have been impressed by the abundance of introns in

M.-Q. Xu, H. Goodrich-Blair, D. A. Shub, Department of Biological Sciences and Center for Molecular Genetics, State University of New York at Albany, Albany, NY 12222.
S. D. Kathe and S. A. Nierzwicki-Bauer, Plant Research Group, Biology Department, Rensselaer Polytechnic Institute, Troy, NY 12181.

*To whom correspondence should be addressed.