Plant Enzyme Structure. Explaining Substrate Specificity and the Evolution of Function

Maria Hrmova and Geoffrey B. Fincher*

Department of Plant Science, University of Adelaide, Waite Campus, Glen Osmond,
South Australia 5064, Australia

THREE-DIMENSIONAL STRUCTURES OF PLANT ENZYMES

Progress in defining the three-dimensional (3D) structures of plant enzymes has been generally slow, but in the last 5 years momentum has picked up considerably (Fig. 1). By the beginning of 2000 about 140 individual plant protein structures were known, of which 37 related to individual plant enzymes.

Most 3D structural data have been generated by x-ray crystallography. The first and very often the limiting step in this procedure is the production of enzyme crystals. However, if high-quality crystals can be obtained, solving the 3D structure can be greatly facilitated by the recent advent of more powerful x-ray generators, such as synchrotrons coupled with multiwavelength anomalous diffraction, increased computing power, and the use of molecular cloning for rapid determination of amino acid sequences.

But where has all this led us? The emerging conclusion is that both prokaryotic and eukaryotic proteins are comprised of an unexpectedly small number of protein folds, which can be combined, adapted, and fine-tuned to achieve the diverse and quite specific functions mediated by the very large number of proteins that operate at the cellular level. For example, domains that mediate protein-protein interactions are conserved in plant and animal proteins that range in function from regulators of transcription and cytoskeleton organization, to proteins that form K^+ channels across membranes (1). Thus, a relatively small number of structural elements has been conserved, but these are used over and over again in the diversification of protein function during evolution. This is also the case for plant enzymes, in which 3D structure ultimately defines substrate specificity and therefore function.

MOLECULAR MODELING

As more 3D structures are solved by x-ray crystallography and NMR, it is becoming apparent that proteins with 25% to 30% sequence identity over 100 or more amino acid residues are likely to have similar 3D conformations. If the 3D structure of one such protein is known, the structure of the other can be deduced by homology modeling (9). Structures obtained by modeling will be less reliable than those determined experimentally by x-ray crystallography, but can nevertheless provide valuable information on enzyme fold and function.

For example, Harvey et al. (3) used homology modeling to examine the 3D structures of enzymes in the family 3 group of glycoside hydrolases. The only member of the family for which a 3D structure was available was the β-glucan exohydrolase from barley (12). However, this single structure could be used to build reliable models of most of the 100 or so members that constitute this family of enzymes, even though sequence identities of members were as low as 22%. The modeling also revealed that enzymes could be constructed through circular permutations of domains; the sequence of protein domains can be altered without affecting the final 3D structure of the entire enzyme. As a result of the modeling exercise, numerous research groups working on family 3 enzymes from bacteria, fungi, and higher plants have been provided with structural clues as to substrate specificity and catalytic mechanisms of their enzymes of interest. Further, biological function can be predicted with a higher degree of confidence.

In another seminal example, homology modeling allowed the 3D structure of the bean storage protein phaseolin to be linked with the structures of the large group of enzymic and nonenzymic proteins that constitutes the cupin superfamily (2). Our understanding of the evolution of plant enzymes is likely to be greatly enhanced by these structural "connections."

Finally, automated 3D structural modeling programs can be used to rapidly identify unknown proteins and enzymes in high-throughput genomics programs. In this procedure, a "structure" constructed by modeling the amino acid sequence of an unknown protein is compared with actual 3D structures in the databases, using increasingly powerful computers and analytical algorithms. The method has been applied with considerable success in yeast genome projects for the identification of genes encoding unknown proteins (10).

* Corresponding author; e-mail geoffrey.fincher@adelaide.edu.au; fax 61–8–8303–7109.
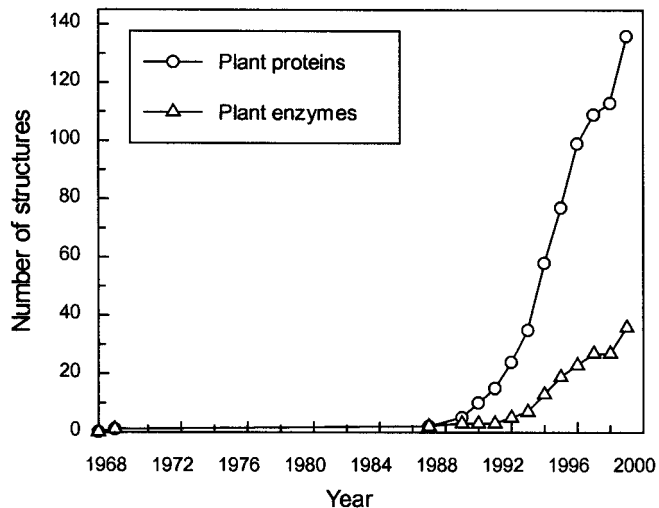
Figure 1. The accumulation of 3D structures of plant proteins and enzymes in the publicly accessible macromolecular structural databases, as of January 2000.

SUBSTRATE SPECIFICITY OF PLANT ENZYMES

The fundamental factors that determine substrate specificity of enzymes are conformational and chemical complementarity between the substrate and its binding site on the enzyme. Thus, the binding site usually consists of a cleft, tunnel, funnel, or other depression on the enzyme's surface. Only those substrates that have complementary shapes will fit into the binding site. Perhaps most intriguing from an evolutionary viewpoint is the precise alignment, or chemical complementarity, of interactive amino acid side-chains on the enzyme surface with corresponding groups on the substrate.

Some general rules of substrate specificity are emerging, in particular for groups of enzymes with common action patterns. For example, an endohydrolase usually has a substrate binding groove or depression that extends across its surface, whether it is a polysaccharide, nucleic acid, or polypeptide endohydrolase (Fig. 2A). Catalytic amino acid residues are located in the substrate-binding cleft. As a result, the endohydrolase can essentially bind anywhere along the polymeric substrate and hydrolyze internal linkages.

In contrast, an exohydrolase needs to align its substrate such that terminal linkages are juxtaposed to catalytic residues. This is usually achieved through a dead-end tunnel, slot, or funnel in the enzyme (Fig. 2, B and C). Specificity can be adjusted from "tight" or "loose" depending on the dimensions of the tunnel or on the geometry of the substrate-binding site. A deep, narrow funnel of the kind observed for barley β-glucosidases severely limits the shape of potential substrates that will fit into the enzyme. In this case, relatively straight $(1\rightarrow4)$-β-oligoglucosides can be threaded all the way to the bottom of the substrate-binding funnel, where catalysis occurs and the non-reducing terminal glucosyl residue is released by hydrolytic action (Fig. 2B). Non-substrate but related molecules such as $(1\rightarrow3)$-β-glucans do not have the correct shape to fit right into the substrate-binding funnel. These non-substrates therefore cannot be brought into contact with catalytic amino acid residues at the bottom of the funnel and are not hydrolyzed by the enzyme.
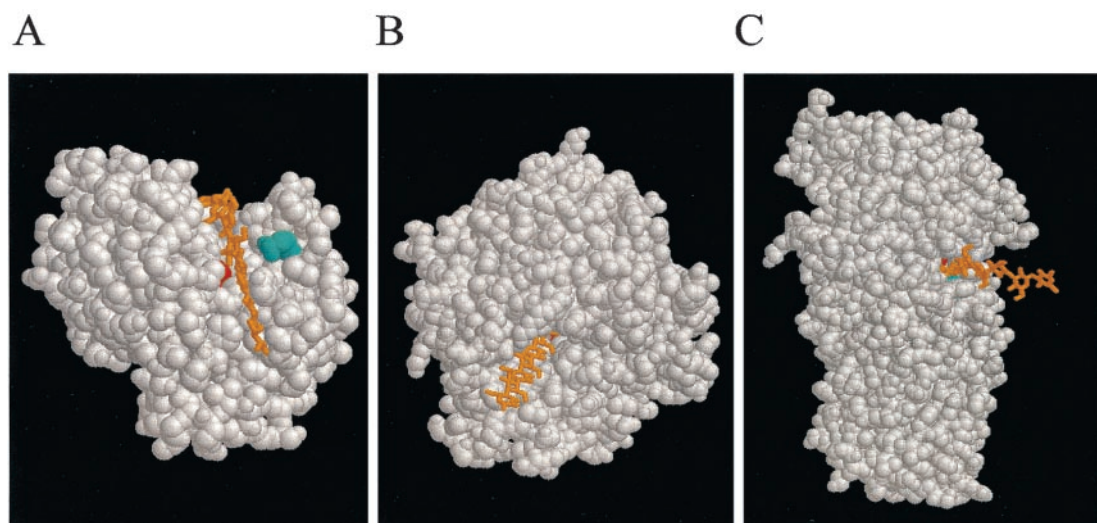
Figure 2. A, Structure of a barley $(1\rightarrow3,1\rightarrow4)$-$\beta$-glucan endohydrolase (11) showing oligosaccharide substrate bound into a cleft that extends across the surface of the enzyme. The catalytic amino acid residues are colored red (nucleophile) and cyan (acid/base). B, A model of the barley β-glucosidase (5) showing a straight, linear $(1\rightarrow4)$-β-oligoglucoside substrate extending to the bottom of a dead-end funnel, where catalytic amino acid residues are located. C, The substrate-binding region of a broad specificity barley β-glucan exohydrolase (12) takes the form of a coin slot that accommodates only two glucosyl residues. Catalytic amino acid residues (colored) are located between the two bound glucosyl residues.

The tight specificity of the barley β-glucosidase may be contrasted with the barley β-glucan exohydrolase. The latter enzyme hydrolyses an extremely broad range of substrates, and its relatively "loose" specificity may again be explained by reference to the 3D structure of the enzyme. The binding site of the β-glucan exohydrolase is much shorter than the binding site of the β-glucosidase; it consists of a shallow "coin slot" into which only two glucosyl residues of the substrate can fit (Fig. 2C). Because the substrate-binding slot is so shallow, most β-glucan substrates can penetrate to the bottom of the slot, irrespective of their conformation, because the majority of the polysaccharide substrate remains "outside" the enzyme. This tolerance of a wide range of substrate shapes is reflected in the broad substrate specificity of the enzyme.

ENZYME INHIBITORS

Higher plants synthesize a range of enzyme inhibitors that function by binding into the active site of the target enzyme, thereby preventing the approach of the natural substrate. As with substrate binding, inhibitor binding requires elements of shape and chemical complementarity between the inhibitor and the enzyme. The 3D structures of a number of enzyme-inhibitor complexes are now solved and provide detailed information on inhibitor action and specificity. In the case of plant α-amylase inhibitors, the 3D structures provide the molecular detail to explain why inhibitors specifically inhibit exogenous α-amylases from pathogenic microorganisms or from insects that attack the plant, but have no effect on endogenous plant α-amylases (6). The structures similarly can demonstrate why inhibitors might only inhibit specific isoforms of an endogenous plant enzyme during particular phases of growth and development. Thus, the barley α-amylase/subtilisin inhibits barley α-amylase 2 in a process that has been linked with the inhibition of precocious germination of grain, but has no effect on barley α-amylase 1, which shares 80% sequence identity with the α-amylase 2 (8).

EVOLUTION OF ENZYME ACTIVITY AND SPECIFICITY

The evolution of enzymic activity and specificity can follow two very different routes (7). First, a protein without catalytic activity but with some well-developed binding capacity for a particular metabolite might accumulate mutations until catalysis occurs. As mentioned earlier, an evolutionary link between enzymic and nonenzymic proteins in the cupin superfamily has been suggested by 3D structural studies and molecular modeling (2). The cupin domain consists of two conserved motifs, each of about 20 amino acids in length and connected by a linker peptide of variable length, which form small β-barrels that are particularly stable. Cupin domains are found in the superfamily both in nonenzymic proteins such as plant storage proteins and in enzymic proteins such as wheat germin, an oxalate oxidase. Cupin proteins of both prokaryotes and eukaryotes are characterized by their small size and their resistance to heat denaturation and proteolytic hydrolysis. These properties are consistent with functional associations between the cupin proteins and plant resistance to biotic and abiotic stresses.

The second evolutionary route to enzyme activity and specificity occurs when an enzyme capable of performing the required catalysis undergoes mutational changes that result in altered substrate specificity. An example here was provided by x-ray crystallography of barley (1→3)-β-glucanases and (1→3,1→4)-β-glucanases, which showed that the 3D folds of the two enzymes are essentially identical (11). The differences in substrate specificity can be attributed to changes in a few amino acid residues lining the substrate-binding cleft, rather than to any large-scale alteration of protein shape. It has been concluded that the pathogenesis-related (1→3)-β-glucanases were recruited to generate the (1→3,1→4)-β-glucanases, which specifically hydrolyze plant cell wall (1→3,1→4)-β-glucans in the graminaceous monocotyledons during wall growth and development (4). Thus, the (1→3)-β-glucanases can degrade the (1→3)- and (1→3,1→6)-β-glucans found in fungal cell walls, consistent with their function in plant-pathogen interactions, whereas the (1→3,1→4)-β-glucanases have evolved to hydrolyze the structurally distinct (1→3,1→4)-β-glucans of plant cell walls.

FUTURE APPLICATIONS

What new concepts have developed from 3D structural studies of plant enzymes and how might these contribute to our future understanding of plant physiology? We can confidently predict that 3D structural analyses of the type used to define enzyme-substrate interactions and the mechanisms of enzymatic catalysis will be extended more broadly into studies on plant cell biology and that molecular modeling will continue to play an important role in these studies. Protein/ligand interactions, other that those of the enzyme/substrate type, will be described in 3D detail. For example, protein inhibitor/enzyme binding, docking of phytohormones with their receptors, the action of specific transporter proteins, and the binding of transcription factors to specific nucleotide sequence motifs could be accurately defined through 3D studies.

Protein/protein interactions that encompass such fundamental processes as signal transduction and plant-pathogen interactions could also be understood

through x-ray crystallography. Developing technologies will need to address problems associated with obtaining 3D structures of membrane-bound proteins and in describing real-time changes in protein conformations during their interactions with other molecules. Time-resolved crystallography, neutron crystallography, and electron cryo-crystallography offer considerable promise in these areas.

If these types of protein/protein and protein/ligand interactions in plants can be described in precise molecular and 3D structural terms, we will place ourselves in a strong position to truly understand the central processes of cell biology. Further, this understanding will present opportunities to enhance plant productivity and end-product quality through rational modification (4) or directed molecular evolution (13) of existing enzymes and de novo design of catalytic proteins. Additional applications in plant production could include the use of specific enzyme inhibitors to regulate or manipulate processes of growth and development.

LITERATURE CITED

1. **Arivand L, Koonin EV** (1999) J Mol Biol **285:** 1353–1361
2. **Dunwell JM, Khuri S, Gane PJ** (2000) Microbiol Mol Biol Rev **64:** 153–179
3. **Harvey AJ, Hrmova M, de Gori R, Varghese JN, Fincher GB** (2000) Proteins: Struct Funct Genet (in press)
4. **Høj PB, Fincher GB** (1995) Plant J **7:** 367–379
5. **Hrmova M, MacGregor EA, Biely P, Stewart RJ, Fincher GB** (1998) J Biol Chem **273:** 11134–11143
6. **Pereira PJ, Lozanov V, Patthy A, Huber R, Bode W, Pongor S, Strobl S** (1999) Structure **7:** 1079–1088
7. **Petsko GA, Kenyon GL, Gerlt JA, Ringe D, Kozarich JW** (1993) Trends Biochem Sci **18:** 372–376
8. **Rodenburg KW, Vallée F, Juge N, Aghajari N, Guo XJ, Haser R, Svensson B** (2000) Eur J Biochem **267:** 1019–1029
9. **Sali A, Blundell TL** (1993) J Mol Biol **234:** 779–815
10. **Sanchez R, Sali A** (1998) Proc Natl Acad Sci USA **95:** 13597–13602
11. **Varghese JN, Garrett TPJ, Colman PM, Chen L, Høj PB, Fincher GB** (1994) Proc Natl Acad Sci USA **91:** 2785–2789
12. **Varghese JN, Hrmova M, Fincher GB** (1999) Structure **7:** 179–190
13. **Zhang JH, Dawes G, Stemmer WPC** (1997) Proc Natl Acad Sci USA **94:** 4504–4509