Many pathogen recognition genes, such as plant PR-genes, undergo rapid adaptive evolution, providing evidence that these genes play a critical role in plant-pathogen coevolution. Surprisingly, whether rapid adaptive evolution also occurs in genes encoding other kinds of plant defense proteins is unknown. Unlike recognition proteins, plant chitinases attack pathogens directly, conferring disease resistance by degrading chitin, a component of fungal cell walls. Here, we show that nonsynonymous substitution rates in plant class I chitinase often exceed synonymous rates in the plant genus *Arabidopsis* (Cruciferae) and in other dicots, indicating a success of adaptively driven amino acid replacements. We identify individual residues that are likely subject to positive selection by using codon substitution models and determine the location of these residues on the three-dimensional structure of class I chitinase. In contrast to primate lysozymes and plant class III chitinases, structural and functional relatives of class I chitinase, the adaptive replacements of class I chitinase occur disproportionately in the active site cleft. This highly unusual pattern of replacements suggests that fungi directly defend against chitinolytic activity through enzymatic inhibition or other forms of chemical resistance and identifies target residues for manipulating chitinolytic activity. These data also provide empirical evidence that plant defense proteins not involved in pathogen recognition also evolve in a manner consistent with rapid coevolutionary interactions.

Host-pathogen interactions are an important force shaping organismal diversity, yet little is known about the evolution of genes responsible for resistance in the host or virulence in the pathogen. Most genetic studies of plant-pathogen coevolution focus on gene-for-gene interactions that arise from pathogen-detection proteins deployed by the plant host (1–4). These proteins, encoded by *R*-genes, detect pathogen-produced elicitors, causing induced transcription of pathogenesis response (PR) proteins that can confer local or systemic resistance. *R*-genes exist in large multigene families that undergo rapid adaptive diversification, apparently in response to intense selection for new resistance specificities (5–7), supporting the hypothesis that these genes are principle targets of selection in the coevolution of hosts and pathogens. Nevertheless, plants have evolved a complex array of chemical and enzymatic defenses, both constitutive and inducible, that are not involved in pathogen detection but whose effectiveness influences pathogenesis and disease resistance (8–10). The genes underlying these defenses comprise a substantial portion of the host genome. Based on genomic sequencing it is estimated that some 14% of the 21,000 genes in *Arabidopsis* are directly related to defense (11). Many of these ~3,000 genes are not involved in pathogen detection but their products do interact directly with pathogen gene or protein products. Among the PR proteins, for example, are chitinases and endoglucanases that attack and degrade the cell walls of pathogens, and which pathogens counterattack with inhibitors (12). Such systems of antagonistically interacting proteins provide the opportunity for molecular coevolution of individual systems of attack and resistance.

Currently, the magnitude of pathogen-imposed selection is not known for any PR gene, nor have the molecular adaptations composing the response to selection been identified for any plant defense gene. Under the neutral theory of molecular evolution, most replacement substitutions are not expected to have adaptive consequences, even in genic regions where neutral evolution can be rejected. For this reason, identifying which replacements are responsible for enzymatic adaptation has proven difficult. Recently, the availability of x-ray crystallographic structures has revolutionized the understanding of protein anatomy, and with it, the relationship between structure and function at the molecular level. By providing insight into which residues act in substrate binding and catalysis, crystal structure analysis provides a new tool for identifying adaptive amino acid substitutions by virtue of their structural and functional consequences (13).

Here we combine tests of neutrality with crystal structures to analyze selection on a PR protein, class I chitinase.

Members of the chitinase (EC 3.2.1.14) gene family are found in all plants, which express them inducibly as PR proteins and constitutively in tissues vulnerable to pathogen attack (14, 15). Several lines of evidence indicate that chitinases play a direct role in plant defense by attacking chitin, a β-1,4-linked polymer of N-acetyl-D-glucosamine and a major component of fungal cell walls: (i) Purified chitinases can inhibit hyphal growth *in vitro* (15), (ii) constitutive overexpression of certain chitinases in transgenic plants can confer increased pathogen resistance *in vivo* (16–18), and (iii) chitinolytic breakdown products induce the production of defense compounds (phytoalexins) and systemic acquired resistance (19). These antifungal properties are greatly enhanced in the presence of β-1,3-endoglucanase (15), another PR protein that attacks the glucan matrix in which chitin is embedded. Nevertheless, chitinases often fail to confer resistance to certain pathogens in both *in vivo* and *in vitro* experiments, even in instances where the absence of antifungal effects cannot be attributed to failure to detect and respond to the invading pathogen (19–23). Although the mechanisms protecting fungal cell walls from chitinolytic activity remain obscure, inhibitors of plant chitinases have been isolated from several bacteria (24, 25).

The variable effectiveness of specific chitinases against different pathogens and the existence of microbial chitinase inhibitors suggest the hypothesis that chitinases may coevolve with fungi in response to variation in pathogen defenses against chitinolytic activity. One prediction of this hypothesis is that chitinase genes will be subject to positive selection, and that high rates of adaptive evolution will indicate high intensity or frequency of selective episodes. Moreover, adaptive substitutions should affect enzyme...
function, an effect that may be detectable by structural analyses. To test these predictions we applied neutral rate tests, capable of distinguishing selectively favored from selectively neutral substitution patterns, to a set of 22 DNA sequences encoding class I chitinase (sometimes called basic chitinase) from 13 species of Arabis (Cruciferae), a group that includes the nearest relatives of the model plant Arabidopsis thaliana (26). We chose class I chitinase because in A. thaliana this class is encoded by a single gene, is expressed in a manner consistent with a defensive role (14), inhibits hyphal growth of certain fungal species but not others (20), and in tobacco degrades chitin more rapidly than do other classes of chitinase (19).

Materials and Methods

Sequence Data. The coding regions corresponding to 23 mature class I chitinases (Fig. 1) were sequenced in 14 accessions of Arabis spp. as follows. PCR primers based on conserved regions of A. thaliana and Brassica napus class I chitinase (GenBank accession nos. M38240 and M95835) were used to amplify partial genomic sequences of two genes from Arabis fecunda genomic DNA extracts. PCR products were cloned (TA cloning kit, Stratagene) and sequenced. Each gene was amplified with gene-specific primers and direct-sequenced from an additional 60 Arabis fecunda individuals, demonstrating nonallelicism of the two loci in this species. Total RNA was extracted (RNeasy kit, Qiagen, Chatsworth, CA) from A. fecunda, A. drummondii, and A. lemmonii, and 3′ rapid amplification of cDNA ends was used to amplify the carboxyl terminal region of five genes, verifying their transcription in these species. These products were cloned (pGEM-T Vector kit, Promega) and sequenced. Sets of conserved primers within the 5′ and 3′ signal and target peptides were designed to amplify regions corresponding to the mature protein from the remaining species. These products were either gel-purified, cloned, and sequenced or they were direct-sequenced. Primer sequences are available on request. Sequencing was performed on both strands of several independent clones for each gene by using Big Dye Terminators (Applied Biosystems) analyzed on a 377XL DNA sequencer (Applied Biosystems). The number of genes for each species included in the study are: one Arabis alpina; one A. blepharophylla; one A. drummondii; two A. fecunda; one A. glabra; two A. gymnisoniana; one A. holboellii from Colorado; one A. holboellii from Greenland; two A. lemmonii; three A. lignifera; one A. hyalii; three A. microphylla; three A. parishii; and one Halimolobos perplexa [the Arabis clade includes H. perplexa (26)].

Sequence Analysis. Sequences of the single chitinase intron, available for 20 of 23 genes, allowed grouping of genes into three sets distinguished by numerous indels. The first two sets (denoted 1 and 2 in Fig. 2) are from North American Arabis that have n = 7 chromosomes and are paralogous to each other. A third set corresponds to the A. alpina group (denoted 3 in Fig. 2). Coding region sequences corresponding to Fig. 1 were aligned and positions with gaps were excluded from the analyses of nonsynonymous substitution rates (Ks), synonymous substitution rates (Ks), and phylogenetic analyses based on coding region.

Chitinase gene phylogenies were estimated by using parsimony and DNA and protein-maximum likelihood methods implemented in PHYLIP (27) and MOLPHY (28), using B. napus as an outgroup (26). Bootstrap analysis indicated that phylogenies based on coding sequence were reliable only at their roots. Further resolution was attained by using intron as well as coding sequences, but it was not possible to align all introns. Therefore, majority-rule consensus parsimony trees were estimated for each of the three intron groups, and resulting cladograms were rooted by using one gene from each other group. Rooted trees for each of the three sets then were attached at the appropriate location in the consensus phylogeny based on coding sequence. The resulting “combined” phylogeny (Fig. 2) was consistent with phylogenies estimated based on coding sequence alone.

Ks and Ks were estimated for all pairwise comparisons of 22 mature protein sequences by using the program L9S (29, 30). A. glabra was omitted from this and other analyses because part of its active site is deleted. Significance tests for Ks > Ks were evaluated for a small subset of comparisons, using a one-tailed t test with df = infinite (29). Because t is biased toward greater significance when the number of substitutions is small (31), we adopted t > 2.2 (P < 0.014) as an appropriate significance threshold. Simulation models of sequence evolution in lysozyme, which is homologous to class I chitinase, about 1/3 the length of chitinase and has similar substitution rates (32), suggest even when biased this value of t corresponds to P < 0.05 (31). Because codon bias can affect Ks, a measure of codon bias, the effective number of codons (33), was estimated for each of the 22 sequences by using the CODONS program (34).

Maximum-likelihood models of codon substitution (35) for 19 sequences were implemented in CODEML 1.4, a program in the PAML package (36). These models address the question of whether Ks > Ks over the entire set of sequences, while taking into account the phylogenetically correlated structure of the data. CODEML fits a
models average synonymous and nonsynonymous substitutions over all possible ancestral sequences at each interior node, weighted by their likelihoods of occurrence (38). The number of replacements at each site in the protein was estimated by parsimony (53) applied to phylogenies estimated by using DNA and protein maximum-likelihood methods. Graphics were implemented by mapping the number of replacements onto structural models by using INSIGHTII (BioSym Technologies, San Diego). Phylogenies estimated by using different methods yielded similar results. Class I chitinase is modeled by using the crystal structure of barley class II chitinase (39) (Protein Data Bank code 1CNS) with bound hexa-N-acetyl-d-glucosamine (40). The two enzymes are very similar, although class II enzymes lack an N-terminal chitin-binding domain (15). Class I sequences were truncated just 3’ of the proline-rich hinge connecting the chitin-binding domain to the chitinolytic domain (Fig. 1).

For comparative purposes, similar structural analyses were performed on sequence alignments for 12 Solanaceae class I chitinases, 22 primate lysozymes, and 31 plant class III chitinas. Human lysozyme with bound tetra-acetyl-chitotetraose (41) (Protein Data Bank code 1LZR) was used to model primate lysozymes. Hvevamine with bound asallosamin (42) (Protein Data Bank code 1LLO) was used to model class III chitinase. Protein accession numbers are: Solanaceae class I, AF043247–48, A21091, S44869, U02605–7, X07130, X64518, X51599, X16939, Z15140, and class III, 3452147, 1076250, 116328, 167538, 167540, 116330, 486991, 116332, 116337, 1362073, 1362074, 1705812, 1835899, 2072742, 2293066, 2317270, 2457017, 2696227, 2696231, 116327, 2342433, 2342435, 2342439, 2342443, 2342447, 2342451, 2342453, 2342457, 2342459, 2342461, and 320555. For primate lysozyme see ref. 32.

Results

One common neutral rate test compares the number of synonymous (silent) substitutions per synonymous site (Ks) with the number of nonsynonymous (amino acid replacing) substitutions per nonsynonymous site (Kas). In the absence of selection on codon usage, Kas represents the rate of neutral substitution. In most proteins, no nonsynonymous substitutions are deleterious, resulting in purifying selection and Kas << Ks. A rate of amino acid replacement significantly greater than the rate of neutral evolution, Kas > Ks, provides unambiguous evidence of adaptive sequence evolution (31, 32, 35, 38). We found Kas > Ks in 85 of 231 pairwise sequence comparisons, including several of the highest Kas/Ks ratios yet reported (Fig. 3A).

Eleven of these comparisons—all those with Kas/Ks > 3—were significantly greater than 1 (t test, P < 0.01, Fig. 3B). There was no evidence of selection for codon usage (effective number of codons is 53–59). Ratios varied greatly among species—A. blepharophylla, both A. holboellii accessions, A. gunnisoniana, A. lemmontii, and A. lignifera each contained genes that exhibited Kas > Ks in nearly all pairwise comparisons. Comparisons with particularly high Kas/Ks included A. blepharophylla vs. A. holboellii Colorado (Kas/Ks = 9.2, 21 nonsynonymous substitutions (N) vs. only one synonymous substitution (S)). A. holboellii Colorado vs. A gnnisoniana 2a (Kas/Ks = 7.2, N = 15, S = 1) and A gnnisoniana 2a vs. A. lemmontii 2 (N = 8, S = 0). In contrast, genes from A. parishii and A. fecunda exhibited only two comparisons with Kas > Ks, Kas > Ks was found on branches throughout the gene phylogeny (Fig. 2). Genes with significant comparisons are somewhat clustered, but clusters are found in three different clades (Fig. 2).

In the group of Arabis with n = 7 chromosomes we take advantage of the two distinct gene chloroplasts, distinguished by intron sequences (Fig. 2), to compare the frequency of elevated Kas/Ks within intron groups (i.e., orthologues) to those between intron groups (i.e., paralogues). Taking Kas/Ks > 1.9 as an arbitrary cutoff,
we find no difference in the frequency of positively selected pairs in paralogous vs. orthologous comparisons (15 cases of 84 comparisons, and seven cases of 88 comparisons, respectively), nor were intraspecific and interspecific comparisons for the \( n = 7 \) group disproportionate (one case in 12 vs. 21 cases in 159).

Our analysis of \( K_s/K_a \) provides strong evidence for positive selection, but further interpretation is difficult given that (i) many of the comparisons were not significant and (ii) the lineages are phylogenetically correlated. These problems can be circumvented by using maximum-likelihood substitution models to test the hypothesis that positive selection is operating across the entire data set (35). We compared a null model in which substitutions are assumed to be either neutral or deleterious with a positive selection model that includes a third category for advantageous substitutions. For all phylogenies tested, the positive selection model provided a better fit than the null model (\( \chi^2 \) range: 24–57, \( df = 2, P < 0.0001 \)). The positive selection model estimated that substitutions at 64–77% of the amino acid residues are deleterious and 5–14% of substitutions are advantageous, with \( K_s/K_a = 6.8 \) for the latter. Bayesian analysis of the models (35) identified 15 sites that are likely (\( P < 0.05 \)) to sustain advantageous amino acid replacements (Fig. 1). Seven of these sites involve only one alternative residue, but are identified as positively selected because that residue must evolve multiple times in the majority of the tested phylogenies.

Mapping the observed replacements onto the x-ray crystallographic structure of chitinase reveals that numerous replacements occur in the active site cleft (Fig. 4 \textit{Upper Left}). Indeed, the mean number of replacements per site inside the cleft [defined as those residues within 0.6 nm of bound substrate (40)] approximately equals the mean number of replacements per site outside it (rate inside: rate outside = 0.28: 0.19, \( \chi^2 = 1.63, P > 0.05 \), assuming 59 cleft positions vs. 227 outside positions, as shown in Fig. 1), with positively selected sites more than twice as likely to reside within the cleft (0.11: 0.04, \( \chi^2 = 3.94, P < 0.05 \); Fig. 1). Physically adjacent to the two rigidly conserved catalytic residues Glu-141 and Glu-163, sites 191 and 164 sustain multiple advantageous replacements, as do sites 284 and 288, directly across the cleft from Glu-163 (Figs. 1 and 4 \textit{Upper Left}). A similar pattern of amino acid replacements is observed for the class I chitinas of the Solanaceae (0.51: 0.55, \( \chi^2 = 0.11, P > 0.05 \); Fig. 4 \textit{Lower Left}) and other dicots (results not shown). Neutral rate tests for the Solanaceae data reveal one instance in which \( K_a > K_s \) (Solanum tuberosum GenBank accession nos. AF043247–48; \( K_a/K_s = 4.32, P < 0.01, t \) test). These results suggest that rapid adaptive evolution of the active site is a general phenomenon in the class I chitinas of dicotyledonous plants.

\textbf{Discussion}

Molecular evolution of class I chitinas is driven by selection for advantageous mutations, causing an excess of amino acid replacements in the active site and substrate binding cleft. We observed a significant excess of amino acid replacements, compared with neutral expectations, in 11 sequence comparisons in the genus \textit{Arabis} and in one pair of sequences from \textit{S. tuberosum}. Elevated \( K_a/K_s \) ratios have not been observed in ADH1 and chalcone synthase gene sequences from some of the same plant accessions (M. Koch and T.M.-O., unpublished work), indicating our results are not attributable to a genomewide elevation of selection pressures.

The unusual pattern of replacements in class I chitinas, and compared with lysozyme and class III chitinas, is not simply a consequence of constraints imposed by differences in form and function. Although highly divergent, class I chitinas are not unusual in this respect; active site residues in most enzymes are highly conserved. The unusual pattern of replacements in class I chitinas, as compared with lysozyme and class III chitinas, is not simply a consequence of constraints imposed by differences in form and function. Although highly divergent, class I chitinas and lysozymes are homologous, sharing similar structures and catalytic mechanisms (46). Murein, the substrate of lysozyme, differs from chitin only by a lactyl side chain that embellishes the C3 position of alternate N-acetyl-D-glucosamine residues in the polymer; many

\textbf{Fig. 3.} (A) Distribution of 231 \( K_a/K_s \) ratios into intervals, where \( K_a \) and \( K_s \) are the number of nucleotide substitutions that cause amino acid replacements or are silent, respectively, divided by the number of possible substitutions of each type. \( K_a/K_s > 1 \) indicates comparisons where amino acid substitutions are more frequent than predicted by neutral substitution rates. (B) \( K_a \) vs. \( K_s \). Ratios > 1 occur above the line \( K_a = K_s \). ▲ denote significant comparisons.
chitinases, including hevamine (Fig. 4 Lower Right), display lysozyme activity and vice versa (42). Nor are the differences from lysozyme merely a consequence of strong positive selection. Recent analyses demonstrate adaptive protein evolution in primate lysozymes, featuring $K_a/K_s$ ratios similar to those found here (32). Rather, the unusual pattern of adaptive amino acid replacements in the active site of plant class I chitinase is a direct consequence of the mode of natural selection operating at the molecular level.

The situation is reminiscent of the evolutionary consequences of therapeutic use of inhibitors on HIV protease where mutations in the active site reduce inhibitor binding by, for example, introducing a polar group close to a hydrophobic methyl on the inhibitor or removing a hydrophobic contact (47). Such mutations need not be chemically “radical” — for example two chemically “conservative” replacements, V82T and I84V, together reduce binding of Indinavir by a factor of 70, rendering the inhibitor ineffective in cell culture assays. Replacements Q164V and M191I in chitinase may have similar effects as they are very likely to contact any inhibitor bound in the active site cleft. Such mutations are also likely to reduce activity by disrupting substrate binding. Selected replacements at sites 167 and 169 in the mobile loop that includes the catalytic E163 may be compensatory, improving catalysis without necessarily affecting the newly acquired resistance. Indeed, in HIV protease a similarly disposed mutation in a flap that guards the entrance to the active site is believed to compensate for a loss in catalytic efficiency by affecting the dynamic behavior of the flap and hence accessibility to the active site (48).

The presence of positively selected substitutions within the active site cleft indicates rapid adaptive functional modification of class I chitinase. We suggest the observed substitutions are favored as a response to inhibition attributable to carbohydrate or protein inhibitors (12, 24, 25) or to resistance mediated by physical modification of the cell wall (49). Plants are likely to encounter novel resistance or inhibition mechanisms in the face of geographic variation in pathogen species, and chitinases are known to differ in their ability to inhibit different fungi (20–23) and in their substrate specificities (19). Although the precise selective agent remains unknown, the notion that amino acid replacements in class I chitinase may be a response to chitinolytic inhibition is supported by our finding that amino acid replace-
ments in class III chitinase/lysozyme contact the bacterial inhibitor allosamidin (Fig. 3). We predict that enzymes involved in antagonistic enzyme-inhibitor interactions will be especially prone to positive molecular evolution. Interestingly, several candidate systems involve cell wall-attacking enzymes deployed by plants and their pathogens (12, 50).

More generally, rapid adaptive evolution in cell wall-attacking enzymes such as chitinase implies that the arms races and other coevolutionary interactions between plants and pathogens are not solely races for information among detection proteins, but also involve proteins that directly attack pathogens. Changes in these proteins thus may account for more variation in disease resistance than commonly realized. However, the ecological interactions in which novel forms of chitinase are advantageous remain to be elucidated. R-genes were named for their involvement in gene-for-gene resistance against strains of particular pathogen species, and all those characterized to date function to detect invading pathogens. In contrast our results indicate that PR proteins might not be involved in similar gene-for-gene interactions with a single pathogen species, but perhaps respond to geographic or temporal changes in the prevalence of different pathogen species.

We thank I. Dew, J. Neuhaus, B. Schaal, H. Stotz, J. Thompson, Z. Yang, and three anonymous reviewers for thoughtful comments on the manuscript. We also thank J. Daey, D. Pedersen, B. Stranger, M. Tobler, D. Schaadrauch, S. Ripoll, J. Koch, J. Kroymann, and H. Stotz for technical assistance; P. Connors, M. Marler, and B. Roy for seed collections; C.-B. Stewart for providing aligned lysozyme sequences; and M. Koch for access to unpublished data. This work was supported by the Max Planck Institute for Chemical Ecology and National Science Foundations Grant BIR-9626079 to J.G.B. and DEB-9527725 to T.M.-O.