

## Deletion of a Single Amino Acid Residue from Different 4-Coumarate:CoA Ligases from Soybean Results in the Generation of New Substrate Specificities\*

Received for publication, March 19, 2002, and in revised form, October 23, 2002  
Published, JBC Papers in Press, November 5, 2002, DOI 10.1074/jbc.M202632200

Christian Lindermayr, Judith Fliegmann, and Jürgen Ebel‡

From the Department Biologie II/Botanik, Ludwig-Maximilians Universität, Menzinger Str. 67, D-80638 München, Germany

**Plant 4-coumarate:coenzyme A ligases, acyl-CoA ligases, peptide synthetases, and firefly luciferases are grouped in one family of AMP-binding proteins. These enzymes do not only use a common reaction mechanism for the activation of carboxylate substrates but are also very likely marked by a similar functional architecture. In soybean, four 4-coumarate:CoA ligases have been described that display different substrate utilization profiles. One of these (*Gm4CL1*) represented an isoform that was able to convert highly ring-substituted cinnamic acids. Using computer-based predictions of the conformation of *Gm4CL1*, a peptide motif was identified and experimentally verified to exert a critical influence on the selectivity toward differently ring-substituted cinnamate substrates. Furthermore, one unique amino acid residue present in the other isoenzymes of soybean was shown to be responsible for the incapability to accommodate highly substituted substrates. The deletion of this residue conferred the ability to activate sinapate and, in one case, also 3,4-dimethoxy cinnamate and was accompanied by a significantly better affinity for ferulate. The engineering of the substrate specificity of the critical enzymes that activate the common precursors of a variety of phenylpropanoid-derived secondary metabolites may offer a convenient tool for the generation of transgenic plants with desirably modified metabolite profiles.**

The well regulated secondary metabolism of plants can be conceived as an adaptive network of specialized enzymes, responding to diverse environmental and internal stimuli. This network ensures an efficient partitioning of resources for the benefit of the plant. The phenylpropanoid pathway of plants represents a particular example for this concept. This pathway uses phenylalanine-derived intermediates for the biosynthesis of compounds used as UV protectants, defense chemicals, signaling compounds, and flower pigments, as well as for the building units of one of the most important structural polymers, lignin. One essential early step in the biosynthesis of phenylpropanoids is the activation of differently ring-substituted cinnamic acids to the corresponding coenzyme A thioesters. This reaction is catalyzed by 4-coumarate:CoA ligase

(4CL,<sup>1</sup> EC 6.2.1.12), an enzyme that employs a two-step reaction mechanism related to other adenylate-forming enzymes (1) such as fatty acyl-CoA synthetases, peptide synthetases, and luciferases. In the first step, ATP forms an adenylate intermediate with the carboxyl moiety of the respective substrate from which, in the second step, the activated group is transferred to an acceptor, mostly the sulfhydryl group of an enzyme-bound cofactor, followed by the concomitant release of AMP. Crystal structures are available for some members of the protein family, allowing the structural modeling of plant 4CLs (2, 3).

Plant 4CLs have been characterized from a wide range of species, showing different isoform distribution patterns. Some plants exhibit only a single form, whereas others contain multiple isoforms. From those containing multiple isoforms, these can exhibit similar or highly divergent substrate specificities in discriminating differently ring-substituted cinnamic acids. For example, in parsley (*Petroselinum crispum*) (4), potato (*Solanum tuberosum*) (5), and loblolly pine (*Pinus taeda*) (6), the multiple 4CL isogenes encode identical or very similar proteins. Conversely, soybean (*Glycine max*) (7, 8), tobacco (*Nicotiana tabacum*) (9), aspen (*Populus tremuloides*) (10), hybrid poplar (*Populus trichocarpa* x *P. deltoides*) (11), and *Arabidopsis thaliana* (12) contain structurally divergent 4CL isoforms. Due to the pronounced differences at the structural, enzymatic, and regulatory levels, the four soybean isozymes deserve special attention (7, 8). The two genes encoding isoforms 3 and 4 are very closely related and could not be distinguished with respect to the enzymatic properties of the encoded proteins. These isoforms were strongly up-regulated at the transcriptional level by elicitation or infection (7, 8). In contrast, the structurally diverse isoform 1 was demonstrated to be down-regulated after elicitation, whereas isoform 2 showed nearly no response (8). In addition to this rather unusual behavior, isoform 1 displayed an interesting biochemical feature: it represented the first 4CL for which a cDNA was isolated to date, where the encoded enzyme was able to convert sinapate to its respective CoA ester (8), a step potentially of critical importance for the synthesis of syringyl-type lignin.

Protein sequence alignments of 4CLs revealed the existence of conserved peptide motifs (5, 7, 11, 12). Box I, described as a putative nucleotide-binding motif, has been used as a signature for the superfamily of adenylate-forming enzymes (13, 14), whereas the absolute conservation of the box II motif seemed to be restricted to 4CLs (5). The box I and II motifs have been the subject of recent investigations using mutant 4CL isoforms

\* This work was supported by the Deutsche Forschungsgemeinschaft (Grant Eb 62/11–3), the Fonds der Chemischen Industrie, and a fellowship of the state of Bavaria (to C. L.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed. Fax: 49-89-1782274; E-mail: j.ebel@botanik.biologie.uni-muenchen.de.

<sup>1</sup> The abbreviations used are: 4CL, 4-coumarate:CoA ligase; 3,4-DMC, 3,4-dimethoxy cinnamate; 4-coumarate, 4-hydroxy cinnamate; caffeate, 3,4-dihydroxy cinnamate; ferulate, 3-methoxy, 4-hydroxy cinnamate; sinapate, 3,5-dimethoxy, 4-hydroxy cinnamate.

from *Arabidopsis* (2, 3, 15). Interestingly, the modification of highly conserved residues included in each of these boxes, which have been postulated to be essential for the enzymatic reaction, did not result in the total loss of activity but showed rather subtle changes of the kinetic parameters of the conversion of caffeate to the respective CoA ester (2). Putative substrate-binding domains within the amino acid sequence have been identified by domain-swapping approaches (15) and have been shown to include specificity-conferring residues that were predicted to be in direct contact with the ring substituent of the cinnamic acid substrate, ferulate (3).

In our earlier report, we described the first cloned 4CL isoenzyme capable of converting sinapate to its thiol ester (8). We now aimed to pinpoint the major difference(s) in the structures of the soybean isozymes that are responsible for the different substrate specificities. In a first attempt, domain swapping between one isoform using a narrow substrate range and one isoform with a broad substrate range was used to confine the regions important for the binding of the acid substrate. Second, structural comparisons have been conducted using 4CL protein alignments in conjunction with crystal structures derived from enzymes utilizing similar enzymatic mechanisms. This allowed the deduction of regions or single amino acid residues with a suspected importance in the enzymatic reaction toward the cinnamic acid substrate. Site-directed mutagenesis was then used to generate modified proteins, which were assayed for altered substrate specificities.

Intriguingly, the deletion of one single amino acid residue from *Gm4CL2* as well as from *Gm4CL3*, which is absent in wild-type *Gm4CL1*, resulted in the generation of new specificities and allowed the mutant enzymes to convert sinapate. The possibility of converting 4CL isoforms into sinapoyl-synthesizing enzymes may lead to valuable transgenic plants, where both lignin production and composition may be customized by using inducible promoters in conjunction with recombinant 4CL proteins expressing the substrate specificity of choice.

#### EXPERIMENTAL PROCEDURES

**Structural Analysis**—Amino acid sequences were aligned and modeled using SWISS-Model ([www.expasy.ch](http://www.expasy.ch)). The crystal structure of luciferase (Protein Data Bank codes 1LCI, 1BA3) (16) was used as a template for the prediction of the putative conformations of *Gm4CL1*, *Gm4CL3*, as well as *Gm4CL3ΔVal*<sup>367</sup>.

**4CL Hybrids**—The plasmids pQE-30/*Gm4CL1* and pQE-31/*Gm4CL3*, described in (8), were used for the generation of hybrids between the soybean 4CL isoforms 1 and 3. Restriction sites available at identical positions in both cDNAs or generated by PCR at the respective positions were used to assemble different portions of the cDNAs according to Fig. 1A and to yield the chimeric proteins depicted in Fig. 1B.

**Site-specific Mutagenesis**—The modification of single nucleotide residues was performed using proof-reading PCR according to Refs. 17 and 18. Briefly, for each mutation, a pair of oligonucleotides was synthesized including the desired alterations surrounded by at least eight nucleotides of the original sequence. The size of the primers was adjusted to yield a melting temperature of 75–77 °C by using the following formula:  $T_m = 81.5 + 0.41 \times GC (\%) - 675/\text{number of bases} - \text{sequence deviation} (\%)$ . For the amplification, 10 ng of column-purified plasmid DNA (Genomed, Bad Oeynhausen, Germany) was used in a total volume of 15  $\mu\text{l}$ , including 1  $\mu\text{M}$  of each primer, 200 mM dNTPs, and one unit of *Pfu* polymerase (Promega/Serva). In total, 18 cycles were conducted, consisting each of 35 s at 94 °C, 45 s at 55 °C, and 12 min at 72 °C, preceded by a melting step at 94 °C for 2 min and followed by a final extension step at 72 °C for 10 min. Subsequently, the parental template DNA was digested with *DpnI* (Promega/Serva), and the amplified plasmids were purified and transformed into *Escherichia coli* DH5 $\alpha$  (19). The mutations were verified by sequencing.

**Expression in *E. coli* and Enzyme Purification**—Heterologous expression, enzymatic assays, and calculations of isoenzyme specificities were performed as described (8). Isolation of recombinant proteins was achieved by affinity dye chromatography combined with immobilized metal chelate affinity chromatography. Briefly, crude bacterial extracts

were applied to reactive blue CL-6B (Sigma) in 50 mM Tris/HCl, pH 8, 14 mM  $\beta$ -mercaptoethanol, and 30% (v/v) glycerol (buffer A), the column was washed with buffer A including 0.6 M NaCl, and proteins were eluted by addition of 2 M NaCl. The eluate was applied to nickel-nitrilotriacetic acid agarose (Qiagen), equilibrated previously in buffer A including 2 M NaCl. After washing (buffer A, 1 M NaCl, 10 mM imidazole), the bound proteins were eluted by raising the imidazole concentration to 200 mM.

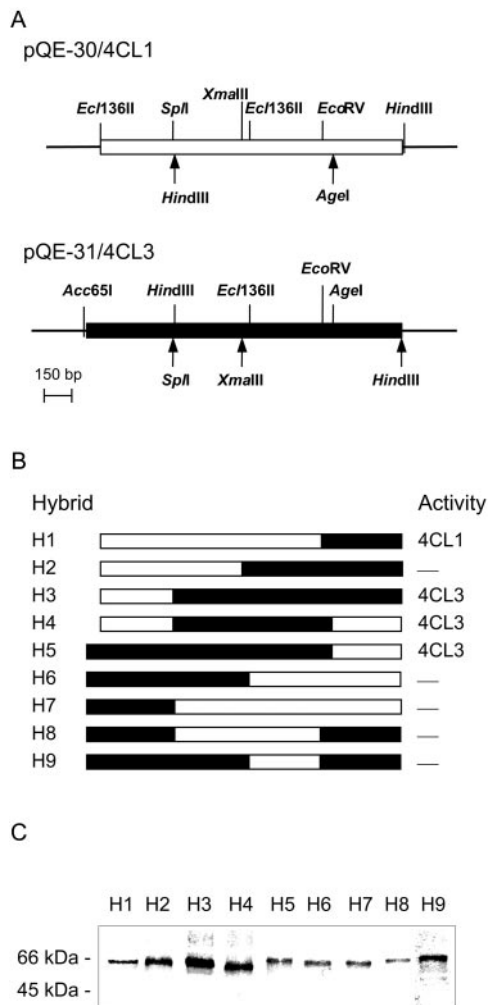
**Antisera and Bacterial Strains**—An antiserum raised against parsley 4CL (20) was used in conjunction with anti-rabbit IgG from goat coupled to alkaline phosphatase (Sigma) to verify the expression from different constructs according to Ref. 8. *E. coli* DH5 $\alpha$  was used for the propagation of recombinant plasmids, and the strain SG13009 (Qiagen) was used for the expression of proteins.

#### RESULTS

**Domain Swapping**—The plant 4CLs belong to a family of AMP-binding enzymes utilizing a common two-step activation mechanism for carboxylate substrates as diverse as fatty, acetic, amino, or cinnamic acids or chlorobenzoate and luciferin (1, 13, 16, 21). PheA, the phenylalanine-activating subunit of gramicidin S synthetase 1 from *Bacillus brevis*, which employs the same reaction mechanism, has been crystallized in a ternary complex with the substrates AMP and phenylalanine (22). Luciferase from the firefly *Photinus pyralis*, another member of this enzyme family from which a structure is available, showed a structure similar to PheA, which allowed the presumptive assignment of the amino acid residues implicated in the binding of the substrates when compared with PheA (16). Since firefly luciferase and plant 4CLs share 31–33% identical amino acid residues, it was then possible to use the structure of the firefly luciferase as a template to model the hypothetical three-dimensional conformation of *Gm4CL1*. All these proteins fold into a larger N-terminal domain and a smaller C-terminal domain with the substrate-binding pocket being located near the transition of the N-terminal domain to the C-terminal domain. This putative substrate-binding region located to the central part of *Gm4CL1*.

We tested this prediction by using a domain-swapping approach exploiting the two soybean 4CL isoforms, *Gm4CL1* and *Gm4CL3*, which showed the most diverse substrate specificities (8). Following the assembly of different hybrids from *Gm4CL1* and *Gm4CL3* cDNAs (Fig. 1A), the derived cDNAs were cloned into a bacterial expression vector (pQE30 or pQE31). As summarized in Fig. 1B, nine chimeric proteins have been generated, from which five were shown to be inactive. The hybrid proteins H1 and H5, which each contained the highly conserved C-terminal domain of the other form, showed no differences in substrate specificities (Table I), excluding an impact of this protein region on the utilization of the cinnamic acid substrate. The replacement of the central region of *Gm4CL1* with the respective portion of *Gm4CL3* (H3, H4) resulted in a significant alteration of the substrate specificity. The hybrid proteins H3 and H4 accepted only the same restricted substrate range of cinnamic acid substrates as isoform 3 (Table I). The failure to generate active hybrid proteins in the reverse combination of isoforms 1 and 3 (H7, H8) may have been due to the N-terminal extension present in *Gm4CL3* as opposed to *Gm4CL1*, which may have resulted in a structural perturbation that was not tolerated by an isoform 1-type enzyme. The hybrid forms H2 and H6, which were merged in the middle of the central region, as well as H9, which contained only the central part of *Gm4CL1* in the backbone of *Gm4CL3*, all yielded inactive enzymes despite a positive expression of all of these recombinant proteins (Fig. 1C).

**Alanine Scanning**—The involvement of the central domain of 4CL in the recognition of cinnamic acid substrates was surveyed in more detail. The sequential superposition of the central protein regions of PheA, which have been shown to build



**FIG. 1. Construction of chimeric 4CL proteins.** A, schematic representation of the coding regions of *Gm4CL1* and *Gm4CL3*, as present in the respective expression vectors. Internal restriction sites are labeled above the bars representing the open reading frames. New sites generated by PCR are marked with arrows below the bars. B, scheme of hybrid proteins generated after fusion of different portions of the open reading frames from *Gm4CL1* (open boxes) and *Gm4CL3* (black boxes). The type of substrate specificity of each chimeric protein is given to the right, as deduced from the data shown in Table I. —, no enzyme activity detectable. The fusion points for the various chimeric proteins are as follows: Hybrid H1: *Gm4CL1*<sub>M1-G400</sub>::*Gm4CL3*<sub>Y425-P570</sub>; H2: *Gm4CL1*<sub>M1-A256</sub>::*Gm4CL3*<sub>A280-P570</sub>; H3: *Gm4CL1*<sub>M1-A131</sub>::*Gm4CL3*<sub>Y156-P570</sub>; H4: *Gm4CL1*<sub>M1-A131</sub>::*Gm4CL3*<sub>Y156-T444</sub>::*Gm4CL1*<sub>G421-N546</sub>; H5: *Gm4CL3*<sub>M1-T444</sub>::*Gm4CL1*<sub>G421-N546</sub>; H6: *Gm4CL3*<sub>M1-E294</sub>::*Gm4CL1*<sub>L272-N546</sub>; H7: *Gm4CL3*<sub>M1-K159</sub>::*Gm4CL1*<sub>I136-N546</sub>; H8: *Gm4CL3*<sub>M1-K159</sub>::*Gm4CL1*<sub>I136-G400</sub>::*Gm4CL3*<sub>Y425-P570</sub>; H9: *Gm4CL3*<sub>M1-E294</sub>::*Gm4CL1*<sub>L272-G400</sub>::*Gm4CL3*<sub>Y425-P570</sub>. In the above list of chimeric proteins, the subscript single-letter codes represent the amino acids, and the subscript numbers represent the position numbers. C, detection of the recombinant chimeric *Gm4CL* proteins generated in *E. coli* by Western blotting. Crude protein extracts (10  $\mu$ g of protein each) were separated on 10% SDS-polyacrylamide gels and transferred onto nitrocellulose filters. For immunodetection, antiserum raised against parsley 4CL (20) combined with goat anti-rabbit IgG conjugated to alkaline phosphatase was used. The relative molecular masses of protein standards are shown on the left.

part of the active site cavity, in the first step to luciferase and then to a computer-generated model of *Gm4CL1*, led to the selection of a protein region of 4CL presumably involved in the binding of the cinnamic acid substrate (*Gm4CL1*<sup>331–348</sup>, Fig. 2A). In addition, in the vicinity of the selected amino acid sequence, a tripeptide (VPP), which was strictly conserved between 4CLs (Fig. 2A, positions 284–286 in *Gm4CL1*), was observed. The residues Ala<sup>236</sup>, Ile<sup>330</sup>, Cys<sup>331</sup>, and Ala<sup>322</sup>, Ala<sup>301</sup>, Thr<sup>278</sup> of PheA have been found to line both sides of the

specificity pocket for the phenylalanine side chain, completed by Trp<sup>239</sup> at the bottom of the cavity (22). From these, four residues were located in or close to the highly conserved peptides shown in the restricted alignment of PheA with the soybean and *Arabidopsis* 4CLs (Fig. 2A). Accordingly, these residues represented promising candidates for a functional role in 4CLs as well.

Both regions of *Gm4CL1* highlighted in Fig. 2A, therefore, were chosen for an alanine-scanning mutagenesis approach to attribute functional importance to specific amino acid side chains by replacing consecutive residues with alanine. The substrate specificities of the recombinant proteins were determined and compared with the wild-type *Gm4CL1* enzyme (Fig. 2B). One replacement (T331A) did not have any influence on enzyme activity, whereas eight of the mutations displayed major effects resulting in nearly complete loss of conversion of cinnamic acid substrates (G333A, Y336A, G337A, M338A, T339A, E340A, G342A, L344A, M348A) despite the positive expression of all of the recombinant proteins (Fig. 2C). Strikingly, the modification of residues 284–286, as well as of residue 334, resulted in larger activities with the highly substituted substrate sinapate when compared with the respective reference values obtained with 4-coumaric acid. With the exception of the mutants P285A and Q334A, this alteration of the enzyme activity ratio was also observed with the artificial substrate 3,4-dimethoxy cinnamate (3,4-DMC).

**Deletion of a Single Residue**—The alignment of soybean 4CL1 with other plant isoforms for which the substrate range was specified revealed one crucial difference (Fig. 2A and data not shown): the absence of a single valine residue (lacking between Pro<sup>343</sup> and Leu<sup>344</sup> of *Gm4CL1*) in the region, which was shown to build the phenylalanine-binding pocket in PheA (22). According to the designation of functional side chain residues, as identified by the alanine-scanning mutagenesis, this part of the putative binding pocket showed only a moderate influence on the selectivity for differently substituted cinnamic acids (Fig. 2B). Furthermore, *Gm4CL1* represents the only isoform that is able to activate sinapate and 3,4-DMC. We therefore tested the impact of the valine deletion by engineering it into the two other isoforms (*Gm4CL2* and 3) of soybean, which have been shown to display intermediate and narrow substrate utilization ranges, respectively (8). The mutant enzymes were indeed able to convert sinapate, although the affinities were shown to be rather low ( $K_m$  values of 1208 and 323  $\mu$ M, respectively, Table II) when compared with *Gm4CL1* ( $K_m$  value of 4.7  $\mu$ M). *Gm4CL2dVal*<sup>345</sup> did not show any further significant alteration of the catalytic efficiency, in contrast to *Gm4CL3dVal*<sup>367</sup>. The latter mutant form did not only gain the capacity to convert sinapate but also the capacity to utilize the artificial substrate 3,4-DMC, showing a slightly better relative conversion rate in comparison with sinapate. Moreover, the efficiency of activating ferulic acid was greatly enhanced when compared with the wild-type 4CL3 enzyme (Table II). In the case of *Gm4CL3*, the deletion of one single residue resulted in a considerable alteration of the functionality of the isoenzyme, albeit accompanied by a decrease of the efficiency for 4-coumarate, the highly preferred substrate of the wild-type isoform, as well as for caffeate.

## DISCUSSION

The soybean 4-coumarate:CoA ligase 1 (1) (*Gm4CL1*) cDNA (8) represents a valuable tool for investigations on the active site of a plant 4CL isoform, which is capable of converting the highly ring-substituted cinnamic acids, sinapate and 3,4-DMC, to the respective CoA thioesters. The aim of our present study was to elucidate the critical difference, inherited by *Gm4CL1*, that allows this enzyme to accommodate a great range of dif-

TABLE I  
Substrate specificities of chimeric 4CL enzymes generated by domain swaps between *Gm4CL1* and *Gm4CL3*

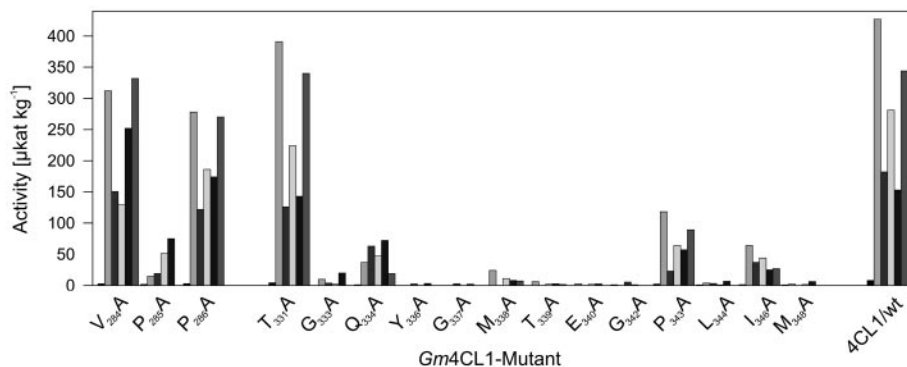
The recombinant 4CLs were expressed in *E. coli*, and specific enzyme activities were determined as described previously (8). The cinnamic acid substrates were used at the following concentrations in the assays: cinnamate, 2 mM; 4-coumarate, caffeate, sinapate, and 3,4-DMC, 0.5 mM; ferulate, 0.1 mM.

	4CL1	4CL3	H1	H3	H4	H5
	$\mu\text{kat kg}^{-1}$					
Cinnamate	11	24	3.2	240	183	13
4-Coumarate	439	101	147	2170	900	60
Caffeate	186	70	36	540	162	20
Ferulate	330	0	54	2.6	0	0
Sinapate	157	0	31	0	0	0
3,4-DMC	358	0	93	0	0	0

A

<i>Gm4CL1</i>	284	V P P	286	331	T F G Q G Y G M T E A G P - L A I S M	348
<i>Gm4CL2</i>	285	V P P	287	332	K L G Q G Y G M T E A G P V L T M S L	350
<i>Gm4CL3</i>	307	V P P	309	354	V L G Q G Y G M T E A G P V L S M C L	372
<i>At4CL1</i>	302	V P P	304	349	K L G Q G Y G M T E A G P V L A M S L	367
<i>At4CL2</i>	295	V P P	297	342	K L G Q G Y G M T E A G P V L A M S L	360
<i>At4CL3</i>	305	V P P	307	352	I L G Q G Y G M T E A G P V L S M S L	370
PheA	279	L P P	281	318	T Y I N A Y G P T E T T I C - A T T W	335

B



C

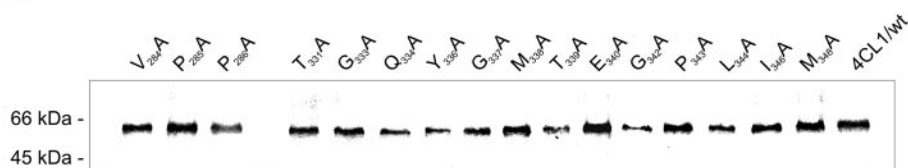


FIG. 2. Alanine-scanning mutagenesis of *Gm4CL1*. A, partial amino acid alignment of soybean (*Glycine max*, *Gm*) and *A. thaliana* (*At*) 4CL isoforms 1, 2, and 3, respectively, in comparison with PheA, the phenylalanine-activating subunit of gramicidin S synthetase 1 from *B. brevis*. The denoted numbering is according to each single polypeptide, as found in the databases. B, enzyme activity profiles of different recombinant *Gm4CL1* proteins generated by site-directed mutagenesis in the protein regions depicted in panel A in comparison with the wild-type (*wt*) enzyme activity of *Gm4CL1* (*4CL1/wt*). The recombinant proteins were expressed in *E. coli*, and the specific activities were measured in crude extracts as described previously by using 2 mM cinnamate or 0.5 mM of all other cinnamic acid substrates, respectively (8). The conversion rates for the cinnamic acids, as catalyzed by the recombinant proteins, are each depicted in the order cinnamate, 4-coumarate, caffeate, ferulate, sinapate, and 3,4-DMC (left to right), respectively, and are denoted by different shading of the columns. C, detection of the single-residue mutant proteins and wild-type *Gm4CL1* generated in *E. coli* by Western blotting, performed as described in the legend for Fig. 1. The relative molecular masses of protein standards are shown on the left.

ferently substituted substrates. The analyses were provoked and facilitated by the availability of two other soybean 4CL cDNAs, coding for *Gm4CL2* and *Gm4CL3*, which are distinguished from *Gm4CL1* by an intermediate and narrow substrate utilization range, respectively (8). By successively pin-

pointing one region of the primary sequence of the enzyme contributing to the active site, a single amino acid residue was identified that, in *Gm4CL2* and *Gm4CL3*, prohibited the conversion of the highly substituted substrates.

One further important prerequisite for the design of our

TABLE II  
Substrate specificities of deletion mutants of the soybean 4CL2 and 4CL3 isoenzymes in comparison to wildtype 4CL1

The recombinant 4CLs were expressed in *E. coli*, purified to homogeneity, and enzyme kinetics were measured as previously described (8). 3,4-DMC, 3,4-dimethoxy cinnamate; dVal<sup>345</sup>, *Gm4CL2* lacking valine at position 345; dVal<sup>367</sup>, *Gm4CL3* lacking valine at position 367; NC, no conversion; wt, wildtype.

	$K_m$		rel $V_{max}$		rel $V_{max} K_m^{-1}$	
	$\mu M$		% of 4-coumarate		$\mu M$	
<i>Gm4CL1</i>						
Cinnamate	4400		8		0.002	
4-Coumarate	50		100		2.00	
Caffeate	81		37		0.46	
Ferulate	4.0		72		18.00	
Sinapate	4.7		32		6.81	
3,4-DMC	49		86		1.76	
	wt	dVal <sup>345</sup>	wt	dVal <sup>345</sup>	wt	dVal <sup>345</sup>
<i>Gm4CL2</i>						
Cinnamate	1700	1900	72	46	0.04	0.02
4-Coumarate	42	49	100	100	2.38	2.04
Caffeate	13	46	37	61	2.85	1.33
Ferulate	140	298	71	100	0.51	0.34
Sinapate	NC	1208		36		0.03
3,4-DMC	NC	NC				
	wt	dVal <sup>367</sup>	wt	dVal <sup>367</sup>	wt	dVal <sup>367</sup>
<i>Gm4CL3</i>						
Cinnamate	1100	361	45	37	0.04	0.10
4-Coumarate	9	15	100	100	11.12	6.67
Caffeate	50	136	74	53	1.48	0.39
Ferulate	3100	14	25	10	0.01	0.71
Sinapate	NC	323		3		0.01
3,4-DMC	NC	311		32		0.10

experiments was structural information describing proteins executing similar enzymatic reactions. 4CLs belong to the superfamily of AMP-binding enzymes, which also includes PheA, the adenylation domain of gramicidin S synthetase, and the firefly luciferase (13, 16, 21, 22). The structures of these proteins have been determined, in the case of PheA even as a ternary complex in the presence of the substrates AMP and phenylalanine (16, 22). Based on the amino acid sequence alignment of *Gm4CL1* and the unliganded luciferase, a structural model was created, which then was used to superpose PheA. In the predicted conformation of *Gm4CL1*, the putative substrate-binding cavity located to the central amino acid sequence including the amino acid residues 284 to 348 of ligase 1. This prognosis was verified by domain-swapping experiments (Fig. 1). The hybrid H4 revealed a conversion of the broad substrate range to the restricted one, as typified by *Gm4CL3*, after the substitution of a sequence of just 288 amino acid residues in the center of the *Gm4CL1* protein with the respective region of *Gm4CL3*. The difficulties in generating active hybrid proteins experienced with the majority of the constructs prevented more in-depth studies of the localization of the regions that may be important for the determination of the substrate range. This may have been due to the higher divergence in similarity and size of the soybean isoforms 1 and 3 (58% identity (8)) as compared with the *Arabidopsis* forms 1 and 2 (86% identity (12)), for which a larger number of active hybrid enzymes have been generated recently (15). The analyses of the substrate recognition profiles of the chimeric proteins from *Arabidopsis* pointed to the same region in the central part of the enzymes, as was found for soybean. Moreover, the dissection of this inner part of the *Arabidopsis* 4CLs resulted in the designation of two substrate-binding domains (sbd I and sbd II), which independently were able to confer a change in the substrate range (15). Included in these domains are the regions that were analyzed by alanine-scanning mutagenesis in the present study (Fig. 2A). The sequence of *Gm4CL1*<sup>284–286</sup> corresponds to a highly conserved tripeptide in the alignment of the

amino acid sequences of a large number of the adenylation-forming enzymes (data not shown) and was therefore chosen for closer inspection. According to the structural data, this tripeptide (Leu<sup>279</sup>-Pro<sup>280</sup>-Pro<sup>281</sup> in PheA) is in close proximity to the phenolic side chain of the aromatic acid substrate (22). Interestingly, the exchange of each of the respective positions in *Gm4CL1* with alanine resulted in an increase of the activity with more highly ring-substituted cinnamic acids with the remarkable exception of the loss of 3,4-DMC activation by the mutation P285A (Fig. 2B). This artificial substrate contains an *O*-methyl group in *para*-position, which may bear a steric hindrance in preventing a productive enzyme-substrate complex in the modified substrate groove. Although the participation of the tripeptide in substrate binding cannot be firmly deduced without the elucidation of the real structure of a 4CL in the presence of the acid substrate, our data clearly support the computational model of *Gm4CL1*, which places this peptide at one end of the substrate-binding pocket. The different ring substitutions of the cinnamic acids implicate different spatial requirements for the substrates, which may or may not fit into the groove built by this stretch of amino acid residues, depending on their respective side chains. This conclusion was corroborated by the recent description of a substrate specificity-conferring amino acid in 4CL2 from *Arabidopsis* (3). The position Met<sup>293</sup> of *At4CL2*, which was speculated to correspond to Thr<sup>278</sup> of PheA, was shown to influence the spatial arrangement of the substrate-binding pocket in such a way that only the exchange with a smaller residue allowed the utilization of ferulate (3). Regardless of minor displacements of the PheA amino acid sequence in relation to different plant 4CL protein alignments used in the aforementioned publication (3), published recently elsewhere (15), or presented in our study (Fig. 2A), the accumulated results indicate that *in silico* biology can be applied to approach the questions posed in these studies.

According to the alignment of Fig. 2A, the second selected region of *Gm4CL1* (positions 331 to 348) corresponds to a small amino acid motif of PheA, which includes residues contributing

to the fixation of both substrates (phenylalanine and AMP) by both main chain carbonyl oxygens (Ala<sup>322</sup>, Gly<sup>324</sup>, Ile<sup>330</sup>) as well as side chain interactions (Asn<sup>321</sup>, Tyr<sup>323</sup>, Thr<sup>326</sup>, Glu<sup>327</sup>) (22). The importance of this central part of the polypeptide chain of *Gm4CL1* is reflected by the loss of activity observed for 9 out of 13 single mutant enzymes (Fig. 2B). Interestingly, the exchange at Q334A did not result in a total loss of activity despite the importance of the respective functionally analogous residue Asn<sup>321</sup> of PheA. Here, the side chain was proposed to interact both with the exocyclic nitrogen of adenine as well as with N1 of the purine ring through a well ordered water molecule (22). Moreover, the substrate specificity of the mutant Q334A displayed a reversal of the pattern when compared with that of the wild-type *Gm4CL1* enzyme: caffeate and sinapate were preferably converted, in contrast to coumarate, ferulate, and 3,4-DMC. This result could indicate a major influence on cinnamic acid substrate discrimination rather than on AMP binding.

The most conspicuous difference of the central part of the polypeptide chain between *Gm4CL1* and all other known plant 4CL proteins was the absence of a conserved valine residue between the positions Pro<sup>343</sup> and Leu<sup>344</sup> of *Gm4CL1* (Fig. 2A and data not shown). We substantiated the important contribution of this amino acid to the substrate specificity by engineering its absence into two other 4CL isoforms of soybean that were known to accept only a subset of the ring-substituted cinnamic acids as substrates (8). The single amino acid deletions had a profound impact on the range of accommodated substrates, generating new substrate specificities for *Gm4CL2* and *Gm4CL3* (Table II). The mutation of ligase 3, which efficiently converted only 4-coumarate and caffeate, to a form that not only accepted two additional substrates (sinapate and 3,4-DMC) but also showed a drastic improvement of the affinity for another (ferulate), illustrated the relevance of this amino acid for substrate discrimination. Computer-based modeling suggested that the deletion of the single valine residue caused a displacement of the respective peptide loop of *Gm4CL3* and concomitantly resulted in a steric rearrangement of the orientation of the neighboring leucine side chain. Interestingly, the exchange at L344A of *Gm4CL1* resulted in the total loss of enzyme activity (Fig. 2B). Furthermore, recent evidence indicated the participation of additional neighboring regions of the polypeptide chain, which allowed the mutant *Arabidopsis* enzymes to convert ferulate (3, 15). Taken together, these findings suggest that spatial restrictions of the binding groove rather than specific interactions between the substrate and amino acid residues of the polypeptide chain may be decisive in determining the substrate specificity of distinct 4CL isoenzymes.

In summary, the comparison of enzymatically highly divergent isoforms of 4CL from soybean enabled the prediction of functionally important amino acid residues. Active mutant en-

zymes with single amino acid exchanges revealed some of the structural principles responsible for the ability to utilize differently substituted cinnamic acids to a different extent. The principles detected for 4CL isoforms from soybean as well as from other sources (2, 3) help to understand the catalytic action of this important enzyme of the general phenylpropanoid pathway of plants. They could be useful also for the design of novel 4CL specificities. For example, the manipulation of the lignin composition of trees used for paper production is an economically important trait since a higher content of syringyl monolignols improves the commercial lignin degradation. Although recent discoveries concerning alternative pathways for the synthesis of monolignols provide interesting targets for genetically modified crop plants (23–25), the utilization of 4CL to preferentially activate a range of cinnamic acid substrates leading to highly substituted lignin building units may also constitute a valuable tool.

*Acknowledgments*—We thank K. Hahlbrock (Köln, Germany) for providing the *P. crispum* 4CL antiserum and A. Mithöfer for critically reading the manuscript and valuable discussions.

#### REFERENCES

- Knobloch, K.-H., and Hahlbrock, K. (1975) *Eur. J. Biochem.* **52**, 311–320
- Stuible, H.-P., Büttner, D., Ehlting, J., Hahlbrock, K., and Kombrink, E. (2000) *FEBS Lett.* **467**, 117–122
- Stuible, H.-P., and Kombrink, E. (2001) *J. Biol. Chem.* **276**, 26893–26897
- Douglas, C. J., Hoffmann, H., Schulz, W., and Hahlbrock, K. (1987) *EMBO J.* **6**, 1189–1195
- Becker-André, M., Schulze-Lefert, P., and Hahlbrock, K. (1991) *J. Biol. Chem.* **266**, 8551–8559
- Zhang, X. H., and Chiang, V. L. (1997) *Plant Physiol.* **113**, 65–74
- Uhlmann, A., and Ebel, J. (1993) *Plant Physiol.* **102**, 1147–1156
- Lindermayr, C., Möllers, B., Fliegmann, J., Uhlmann, A., Lottspeich, F., Meimberg, H., and Ebel, J. (2002) *Eur. J. Biochem.* **269**, 1304–1315
- Lee, D., and Douglas, C. J. (1996) *Plant Physiol.* **112**, 193–205
- Hu, W.-J., Kawaoka, A., Tsai, C.-J., Lung, J., Osakabe, K., Ebinuma, H., and Chiang, V. L. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 5407–5412
- Allina, S. M., Pri-Hadash, A., Theilmann, D. A., Ellis, B. E., and Douglas, C. J. (1998) *Plant Physiol.* **116**, 743–754
- Ehlting, J., Büttner, D., Wang, Q., Douglas, C. J., Somssich I. E., and Kombrink, E. (1999) *Plant J.* **19**, 9–20
- Babbitt, P. C., Kenyon, G. L., Martin, B. M., Charest, H., Slyvestre, M., Scholten, J. D., Chang, K.-H., Liang, P.-H., and Dunaway-Mariano, D. (1992) *Biochemistry* **31**, 5594–5604
- Hofmann, K., Bucher, P., Falque, L., and Bairoch, A. (1999) *Nucleic Acids Res.* **27**, 215–219
- Ehlting, J., Shin, J. J. K., and Douglas, C. J. (2001) *Plant J.* **27**, 455–465
- Conti, E., Franks, N. P., and Brick, P. (1996) *Structure* **4**, 287–298
- Nelson, M., and McClelland, M. (1992) *Methods Enzymol.* **216**, 279–303
- Papworth, C., Braman, J., and Wright, D. A. (1996) *Strategies* **9**, 3–4
- Hanahan, D. (1983) *J. Mol. Biol.* **5**, 557–580
- Ragg, H., Kuhn, D. N., and Hahlbrock, K. (1981) *J. Biol. Chem.* **256**, 10061–10065
- Fulda, M., Heinz, E., and Wolter, F. P. (1994) *Mol. Gen. Genet.* **242**, 241–249
- Conti, E., Stachelhaus, T., Marahiel, M. A., and Brick, P. (1997) *EMBO J.* **16**, 4174–4183
- Humphreys, J. M., Hemm, M. R., and Chapple, C. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96**, 10045–10050
- Li, L., Popko, J. L., Umezawa, T., and Chiang, V. L. (2000) *J. Biol. Chem.* **275**, 6537–6545
- Guo, D., Chen, F., Inoue, K., Blount, J. W., and Dixon, R. A. (2001) *Plant Cell* **13**, 73–88