Crystal Structure of Hydroxynitrile Lyase from *Sorghum bicolor* in Complex with the Inhibitor Benzoic Acid: A Novel Cyanogenic Enzyme[†,‡]

Hanspeter Lauble,*[,§] Burkhard Miehlich,[§] Siegfried Förster,[§] Harald Wajant,[||] and Franz Effenberger*[,§]

Institut für Organische Chemie, Universität Stuttgart, Pfaffenwaldring 55, D-70569 Stuttgart, Germany, and Institut für Zellbiologie und Immunologie, Universität Stuttgart, D-70569 Stuttgart, Germany

Received April 29, 2002; Revised Manuscript Received July 25, 2002

ABSTRACT: The crystal structure of the hydroxynitrile lyase from *Sorghum bicolor* (SbHNL) in complex with the inhibitor benzoic acid has been determined at 2.3 Å resolution and refined to a crystallographic *R*-factor of 16.5%. The SbHNL sequence places the enzyme in the α/β hydrolase family where the active site nucleophile is predicted to be organized in a characteristic pentapeptide motif which is part of the active site strand−turn−helix motif. In SbHNL, however, a unique two-amino acid deletion is next to the putative active site Ser158, removing thereby the putative oxyanion hole-forming Tyr residue. The presented X-ray structure shows that the overall folding pattern of SbHNL is similar to that of the closely related wheat serine carboxypeptidase (CPD-WII); however, the deletion in SbHNL is forcing the putative active site residues away from the expected hydrolase binding site toward a small hydrophobic cleft, which also contains the inhibitor benzoic acid, defining thereby a completely different SbHNL active site architecture where the traditional view of a classic triad is not given any more. Rather, we propose a mechanism involving general base catalysis by the carboxy-terminal Trp270 carboxyl group and proton transfer toward the leaving nitrile group by an active site water molecule. The unexpected interactions of the inhibitor with the new SbHNL active site also reveal the structural basis for the enzyme's limited substrate specificity. The implications of this structure on the evolution of catalysis in the hydroxynitrile lyase superfamily are discussed.

Hydroxynitrile lyases (HNLs)[1] constitute a diverse family of enzymes that catalyze the stereospecific cleavage of a wide range of cyanohydrins into aldehydes or ketones and hydrogen cyanide. The release of toxic HCN (cyanogenesis) as a defense against herbivores is widely distributed in higher plants, including important food plants such as cassava, sorghum, or millet (*1*). Several distinctly different types of HNLs exist in nature, and it has been suggested that this group of enzymes has originated by convergent evolution from different ancestral proteins (*2*). Previous structural work on these enzymes has focused on HNL from *Hevea brasiliensis* (HbHNL) (*3*), *Manihot esculenta* (MeHNL) (*4*), and *Prunus amygdalus* (PaHNL) (*5*), and a general base mechanism closely related to the base-catalyzed chemical addition of HCN to carbonyl compounds is now generally accepted for MeHNL (*6*).

The HNL from *Sorghum bicolor* (SbHNL, EC 4.1.2.11) catalyzes the dissociation of its native substrate (*S*)-4-hydroxymandelonitrile into 4-hydroxybenzaldehyde and HCN (*7*). According to the nucleotide sequence, SbHNL is synthesized as a 510-amino acid polypeptide chain, which is further processed to a functional double-chain form designated subunits α and β (*8*). The active enzyme form is an $\alpha_2\beta_2$ heterotetramer assembled as a dimer of $\alpha\beta$ dimers (*9*). The mature enzyme is N-glycosylated and exists in three different isoforms (SbHNL-I, -II, and -III). The SbHNL amino acid sequence is not similar to that of any other known HNL, but the sequence of SbHNL is 60% identical and 73% similar to that of wheat serine carboxypeptidase (CPD-WII) (Figure 1) (*10*). On the basis of the sequence alignment with CPD-WII, we assume that mature SbHNL starts with Gln1. Residues of CPD-WII identified in crystallographic studies (*11*) to be located in the active site are also conserved in SbHNL, suggesting an active site charge-relay system similar to that observed in serine proteases. The residues that are putative components of the catalytic triad in SbHNL are Ser158 (Ser146 in CPD-WII), His414 (His397 in CPD-WII), and Asp359 (Asp338 in CPD-WII).

Despite these common biochemical properties, however, SbHNL shows a significant difference. The SbHNL sequence reveals a two-amino acid deletion immediately adjacent to the CPD-WII catalytic Ser146 (Ser158 in SbHNL), thereby modifying the characteristic pentapeptide motif G-X-S-X-G/A, often considered the hallmark of serine hydrolase activity. Derewenda et al. (*12*) postulate that this sequence

[†] This work was supported by Grants A03U and A10U from the Bundesministerium für Forschung und Technologie. This work was generously supported by the Fonds der Chemischen Industrie and the Degussa AG.

[‡] The atomic coordinates and observed structure factors have been deposited with the Protein Data Bank (entry 1GXS).

* To whom correspondence should be addressed. Fax: (+49) 711-685-4269. E-mail: PeterLauble@t-online.de.

[§] Institut für Organische Chemie.

[||] Institut für Zellbiologie und Immunologie.

[1] Abbreviations: HNLs, hydroxynitrile lyases; HbHNL, *H. brasiliensis* HNL; MeHNL, *M. esculenta* HNL; SbHNL, *S. bicolor* HNL; SbHNL-II, *S. bicolor* HNL isoenzyme II; CPD-WII, carboxypeptidase from wheat; DA, decanoic acid; rms, root-mean-square; CHA, cyclohexanecarbaldehyde.

10.1021/bi020300o CCC: $22.00
Published on Web 09/13/2002

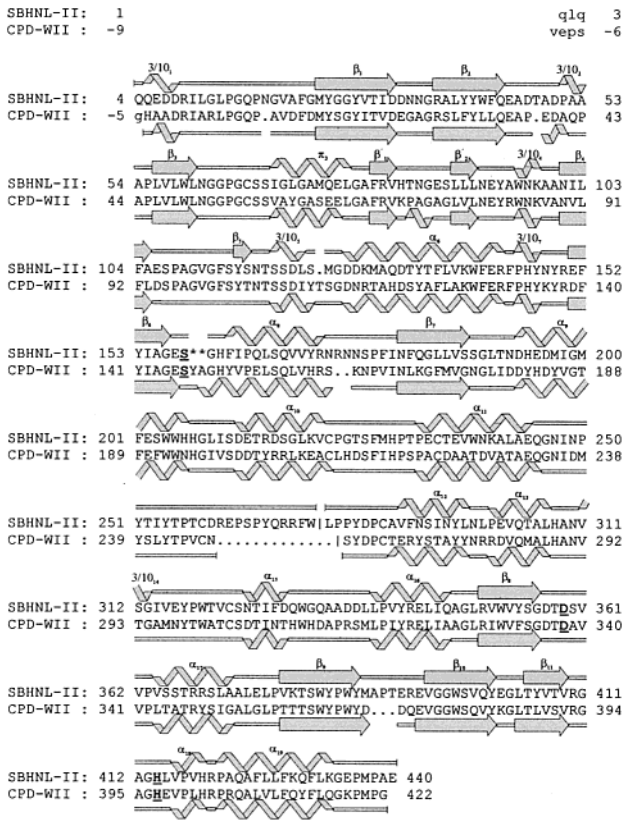

FIGURE 1: Amino acid sequence and secondary structure alignment of SbHNL-II (top line) and CPD-WII (bottom line) performed with the program Insight II (*20*). Only the mature sequences were used, and residues represented in lowercase letters were not modeled. The coordinates of the CPD-WII−arginine complex were used in the alignment, and numbers are given to the residues as in PDB entry 3SC2 (*11*). The SbHNL sequence is corrected for SbHNL isoenzyme II as described in Materials and Methods. The CPD-WII active site residues and the corresponding analogue SbHNL-II residues are shown in bold und underlined. The two asterisks in the SbHNL-II sequence next to Ser158 represent the two-residue deletion in the active site loop structure. The secondary structure elements of both enzymes were assigned with the program PROMOTIF (*27*) and are given above the SbHNL-II sequence and below the CPD-WII sequence.

motif implicates a structural motif of the enzyme which positions the nucleophilic residue in a sharp turn, creating thereby an arrangement of hydrogen bond donors that is supposed to stabilize the transition state of the catalytic reaction. The relative replacement of the primary catalytic serine residue in the SbHNL structure element, which we intend to call the active site loop, compared to the invariant signature motifs observed in all other serine-dependent hydrolases, is highly unusual and suggests that SbHNL may have a completely different active site geometry for the enzymatic retro addition of HCN from cyanohydrins than CPD-WII has for the hydrolysis of an amide bond. This suggestion is particularly interesting since SbHNL exerts no carboxypeptidase activity toward a standard set of potential substrates (*13*), despite the pronounced overall sequence homology with CPD-WII.

We recently crystallized the SbHNL isoenzyme II (SbHNL-II) in the presence of the inhibitor benzoic acid (*14*) and have now determined the three-dimensional structure to 2.3 Å resolution. With the structure presented here, we have examined the relationship of the modified SbHNL-II active site sequence motif to the enzyme's overall three-dimensional structure. We further have analyzed interactions of the active site residues with the inhibitor benzoic acid, from which the enzyme reaction mechanism is inferred. The presence of the inhibitor also allows us to define the factors that are important for the limited substrate specificity.

Table 1: Crystallographic Summary

diffraction data	
space group	*C*2
cell dimensions	$a = 150.7$ Å
	$b = 103.7$ Å
	$c = 90.6$ Å
	$\beta = 101.3°$
resolution range (Å)[a]	17−2.3 (2.38−2.30)
no. of measured reflections	100992
no. of unique reflections	51747
completeness (%)	88.7 (82.9)
$I/\sigma I$	8.5 (5.7)
R_{sym} (I)[b]	0.059 (0.114)
refinement	
resolution range (Å)	8.0−2.3
no. of reflections ($F_o > 2\sigma F_o$)	50480
R_{cryst}[c]	0.165
R_{free}[d]	0.222
no. of protein atoms	6748
no. of ligand atoms	146
no. of solvent atoms	331
average *B*-factor (Å^2)	
protein	17
benzoic acid	27
decanoic acid	25
solvent	26
rmsd from target values	
bond lengths (Å)	0.02
bond angles (deg)	3.1
regions of Ramachandran plot[e] (%)	92.3, 7.6, 0.1

[a] Values in parentheses are for the highest-resolution shells. [b] $R_{sym} = \sum_{hkl}\sum_i I_i - \langle I\rangle / \sum_i I_i$, where I_i is the intensity of the ith reflection (hkl) or a symmetry-related reflection and $\langle I\rangle$ is the scaled mean intensity. [c] $R_{cryst} = \sum|F_o - F_c|/\sum F_o$. [d] $R_{free} = R_{cryst}$ calculated for 10% of the reflections omitted from the refinement. [e] Most favored, additionally allowed, generously allowed.

MATERIALS AND METHODS

Crystallization and Data Collection. Details of crystal growth and data collection have been described elsewhere (*14*). Briefly, the SbHNL-II−benzoic acid complex data set was obtained from a single crystal using a MAR image plate detector and graphite monochromatized Cu Kα radiation produced with an Enraf Nonius FR 571 rotating anode generator. The data set was processed with MOSFLM (*15*). The data were scaled with SCALA and reduced with TRUNCATE (*16*). The crystals belong to space group *C*2 with two $\alpha\beta$ dimers per asymmetric unit with the following cell dimensions: $a = 150.7$ Å, $b = 103.7$ Å, $c = 90.6$ Å, and $\beta = 101.3°$. Data processing statistics are given in Table 1.

Structure Determination. The structure was determined by molecular replacement with XPLOR (*17*), using an $\alpha\beta$ dimer from wheat serine carboxypeptidase [CPD-WII, PDB entry 3SC2 (*11*)] as a search model. Water and covalently attached ligands were removed from the structure, and residues of CPD-WII different from those of SbHNL-II were modeled as Ala. The rotation function and Patterson correlation refinement gave two clear solutions, corresponding to each $\alpha\beta$ dimer of the heterotetramer. The translation function of

each solution gave 11σ and 12σ peaks, respectively. Analysis of the packing of the molecular replacement solution showed no unfavored interactions of the two $\alpha\beta$ subunits and with symmetry-related molecules, respectively.

Structure Refinement. All refinement steps were performed using the XPLOR package (*17*). For calculation of R_{free}, 10% of the reflections were set aside (*18*). Difference Fourier maps were inspected using the program XTALVIEW (*19*). The molecular replacement parameters were optimized by rigid-body refinement using data between 8 and 4 Å. To allow rearrangement of the molecules in the asymmetric unit against each other, each $\alpha\beta$ dimer was treated separately and noncrystallographic symmetry restraints were not applied during refinement. The *R*-factor after rigid-body refinement was 43.3%. Positional refinement of the model obtained from rigid-body refinement using data between 8 and 3 Å resolution reduced the *R*-factor to 29.7%. The $2F_o - F_c$ and $F_o - F_c$ electron density maps were calculated using all data between 17 and 3 Å resolution. Inspection of these maps clearly indicated that the overall structure of the SbHNL-II $\alpha_2\beta_2$ heterotetramer is well-defined, as illustrated by density modifications at positions where there are sequence insertions and deletions and clear extra electron density for the additional 10 residues which extend from the carboxy-terminal helix of the α subunit. During the rebuilding, however, it became clear for both protein molecules in the asymmetric unit that amino acid residues Pro11, Ser408, Pro409, and Ser410 had clear extra electron density, allowing modeling as Leu11, Thr408, Val409 (Val392 in CPD-WII), and Arg410 (Arg393 in CPD-WII), respectively. The $F_o - F_c$ map further revealed clear negative difference density extending from the Cα atom of Val112, from the Cβ atom of Pro79, and from the Cχ atom of Leu203; additionally, the side chains of Pro79 and Val112 come into closer than van der Waals contacts with Gln44 and Gln106, respectively. On the basis of the CPD-WII sequence, we have therefore chosen to model in Ala79, Gly112, and Ser203, respectively. This model was subjected to further rounds of positional and simulated annealing (3000 K) refinement, including all data in the resolution range of 8−2.7 Å. The $2F_o - F_c$ and $F_o - F_c$ electron density maps were calculated, and the model was carefully analyzed. The modified amino acids of both protein molecules fit the electron density very well. Additional density appeared next to Asn117 and Asn310 in both $\alpha\beta$ subunits and was modeled as N-linked acetylglucosamine. This interpretation is in agreement with the glycosylation pattern predicted for SbHNL (*10*) and with that reported for the CPD-WII structure. For both molecules in the asymmetric unit with the addition of data up to 2.7 Å resolution, a segment of electron density unaccounted for appeared in the $2F_o - F_c$ and $F_o - F_c$ electron density maps. This elongated electron density feature is located in the S_1' carboxypeptidase binding subsite and has the shape of a carboxylate or a carboxamide group with a substantial aliphatic tail. Thus, a model for decanoic acid (DA) was built into the density. For lack of a better explanation, we assume that this fatty acid represents a substance that copurifies with SbHNL-II. Subsequent cycles of positional refinement led to an *R* of 22.5% and an R_{free} of 27.8% for all data in the resolution range of 8−2.5 Å. The $2F_o - F_c$ and $F_o - F_c$ maps were calculated and confirm the fit of the DA models to their density. Water molecules were added using the program Insight II (*20*), refined using individual *B*-factors, and checked for their fit into $2F_o - F_c$ electron density maps. Only water molecules with *B*-factors of >50 Å^2 were retained. Residual electron density extends from N117-acetylglucosamine on both molecules in the asymmetric unit, and an additional glucosamine−L-fucose moiety has been incorporated on each side. Models for these carbohydrate linkages were also based on the CPD-WII structure.

This model was subjected to further rounds of positional, simulated annealing (3000 K) and *B*-factor refinement, including all data in the resolution range of 8−2.3 Å. The $2F_o - F_c$ and $F_o - F_c$ electron density maps were calculated, and the model was carefully analyzed. The maps clearly revealed the location and orientation of the inhibitor benzoic acid and also showed an active site water molecule in both subunits. An additional water was found in the active site region of subunit A, while the attempt to locate an analogous water molecule bound to subunit B was unsuccessful. The protein structure along with the water molecules, two decanoic acids, and two benzoic acids were refined against data between 8 and 2.3 Å resolution. The $2F_o - F_c$ maps were calculated and confirm the fit of these models to their density. The refinement converged to an *R* of 16.5% and an R_{free} of 22.2% for all 50 480 reflections in the resolution range of 8−2.3 Å. The final model comprises protein residues α4−α270 and β283−β440 for both dimers in the asymmetric unit, eight carbohydrate residues, two benzoic acid molecules, two decanoic acid molecules, and a total of 331 water molecules. The three remaining N-terminal residues of both α subunits are disordered. Clear electron density modifications have revealed differences with respect to the reported amino acid sequence. On the basis of these features and in agreement with the CPD-WII sequence, we have therefore chosen to model in Leu11, Ala79, Gly112, Ser203, Thr408, Val409, and Arg410. Differences with respect to the SbHNL cDNA-derived sequence information may be attributed to isoenzyme specific residues in SbHNL-II.

In a Ramachandran plot, 92.3% of all residues are in the most favored region, 7.6% in the additionally allowed region, and 0.1% in the generally allowed region. The average *B*-factor for all protein atoms is 17 Å^2, for the carbohydrate moieties 42 Å^2, for the benzoic acid molecules 27 Å^2, for the decanoic acid molecules 25 Å^2, and for all water molecules 26 Å^2. A comparison of the two independently refined $\alpha\beta$ dimers in the asymmetric unit shows root-mean-square (rms) differences of 0.2 Å for backbone atoms and 0.6 Å for all atoms. Details of the refinement statistics are summarized in Table 1. The geometric parameters of the model have been analyzed with PROCHECK (*21*) and are within the expected deviations from ideality calculated from other structures determined to 2.3 Å resolution. These statistics correspond to an estimated error in atomic coordinates of 0.20 Å according to Luzzati (*22*). The final coordinates have been deposited with the Protein Data Bank (entry 1GXS).

RESULTS AND DISCUSSION

Overall Structure. The overall fold and connectivity of $\alpha\beta$ SbHNL-II in complex with benzoic acid are very similar to those of CPD-WII, consistent with a level of sequence homology of 73% between the two enzymes. As this folding

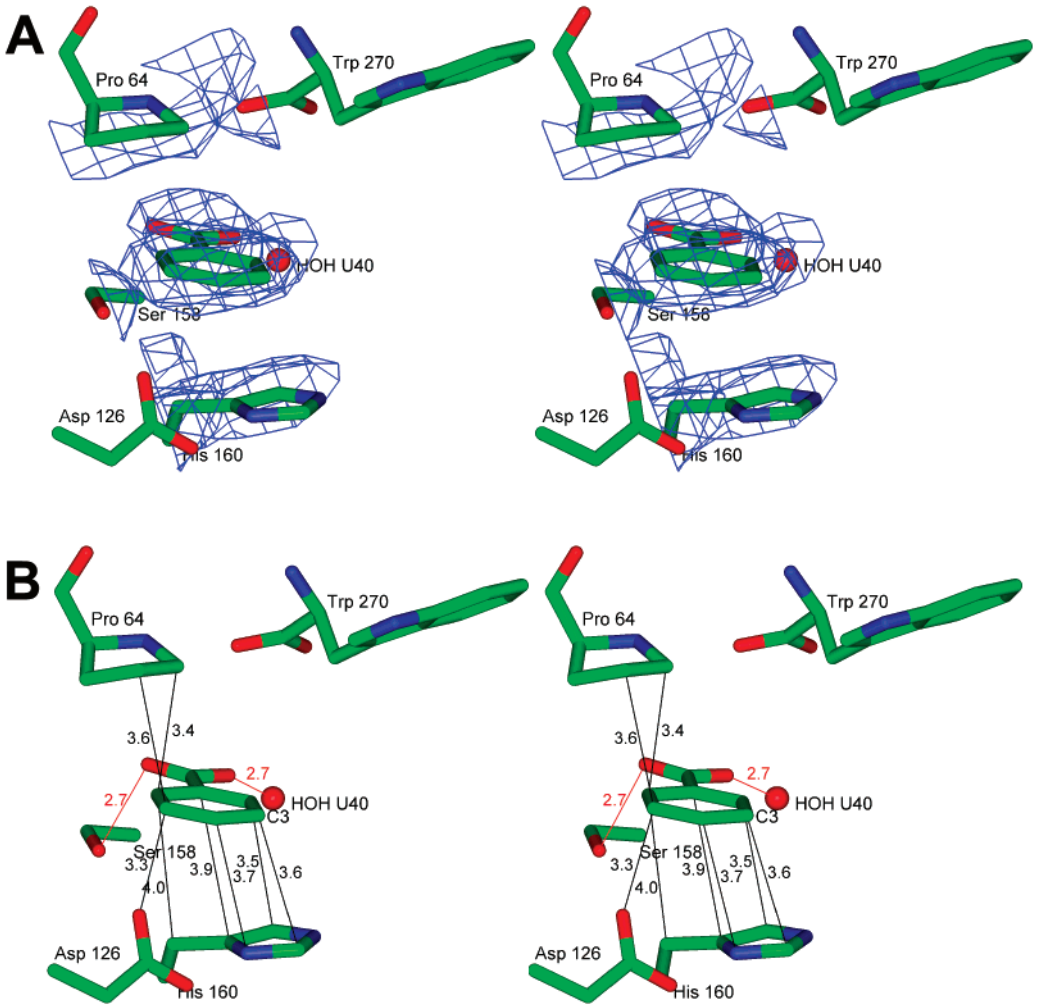

FIGURE 2: Stereoview of the SbHNL-II active site in complex with benzoic acid. Molecule A of the asymmetric unit is shown. (A) Final $2F_o - F_c$ electron density map contoured at 1σ, showing the density of the inhibitor benzoic acid and the neighboring residues. (B) Intermolecular interactions of the inhibitor molecule benzoic acid and selected active site residues. Hydrogen bonds are shown as red lines and van der Waals interactions as black lines. Interatomic distances are given in angstroms.

geometry has been described in detail (*11*), only a brief overview is given here (for notation, see Figure 1). SbHNL-II shows an α/β hydrolase fold consisting of an 11-stranded mixed β sheet (β_1−β_{11}) of which the central six are parallel. Two small antiparallel twisted β strands (β'_1 and β'_2) form an additional short sheet on one side of the core scaffold. This central sheet structure is surrounded by 12 α helices, six short 3_{10} helices, and one π helix. The remaining residues form the loop structures, most significantly the additional 10 amino acids extending from the carboxy terminus of the α subunit in SbHNL-II, which have no homologues in CPD-WII. These residues form a loop toward the "front" of the molecule, thereby occluding part of the SbHNL-II carboxypeptidase cavity. The relevance of this structural arrangement will be discussed below. Superposition of the SbHNL-II−benzoic acid complex structure and the CPD-WII−arginine complex structure (PDB entry 3SC2) by least-squares optimization of 371 equivalent Cα atoms shows an rms difference of 0.8 Å. Major differences of the main chain atoms are located at the SbHNL-II active site loop and will be discussed in detail below.

SbHNL-II Active Site and Inhibitor Binding. The SbHNL-II active site is located between β sheet 6 and α helix 8 and was identified not only by the location of the putative catalytic residue Ser158 but also by the presence of the inhibitor benzoic acid. In each of the two subsites, the conformation of benzoic acid was clearly defined in the final $2F_o - F_c$ electron density map as shown in Figure 2A.

The SbHNL-II active site is constructed as a box which is lined by atoms from Gly63, Pro64, Asp126, Ser158, His160, Phe161, Leu190, Trp270, Trp330, and Ala333. The geometry of the active site is fixed by an intricate network of polar and nonpolar interactions involving all these residues. Ala333 opens the box toward a small tunnel which is formed by Asp193, Ser226, Phe227, Met228, Ala334, and Asp335, leading to the surface of the molecule.

Inside the cavity, one water molecule and the inhibitor benzoic acid are buried. Carboxylic oxygen O1 of benzoic acid forms a hydrogen bond to Ser158 OG (Figure 2B); carboxylic oxygen O2 is hydrogen bonded to the active site water U40, while the phenyl ring lies parallel between His160 and the nonpolar face of Pro64. Most specific for this sandwich mode binding are the π stacking interactions of C2−C4 with the His160 imidazolium group. An additional water molecule is found in the active site region of subunit A and placed between Trp270 carboxylic oxygen O1 and His414 NE2. An attempt to locate an analogous water molecule in the active site region of subunit B, however, was unsuccessful, so this water may not be critical for a

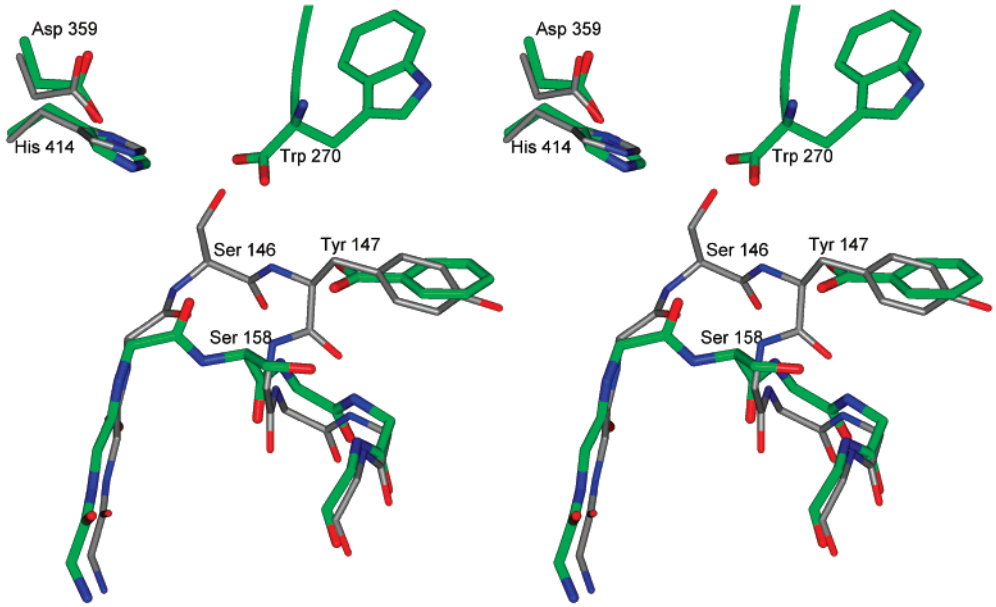

FIGURE 3: Stereoview of the superposition of the active sites of the SbHNL-II−benzoic acid complex (green) and the CPD-WII−arginine complex (gray). The two molecules are superimposed as described in the text. Note how the two-residue deletion in the SbHNL-II active site loop is forcing Ser158 away from its S'_1 carboxypeptidase binding subsite with respect to the corresponding Ser146 in the CPD-WII structure. This figure also shows that benzoic acid in SbHNL-II occupies basically the same space as does the side chain of Tyr147 in CPD-WII.

balanced saturation of the H-bonding properties of the active site.

A surprising aspect of the SbHNL-II inhibitor binding mode is the observation that the His414 residue of the putative catalytic triad is positioned near the carboxyl moiety of benzoic acid, but not close enough for direct hydrogen bonding, and so may contribute only in a small way to the general stabilization of the inhibitor−substrate enzyme complex. A second interesting aspect of this structure reveals the superposition of the SbHNL-II−benzoic acid complex and the CPD-WII−arginine complex showing that the phenyl moiety of the inhibitor fills in SbHNL-II the same space created by deletion of Tyr147 in CPD-WII (see Figure 3).

Comparison of the SbHNL-II Active Site with That of CPD-WII. On the basis of the conserved sequence motifs of SbHNL-II and CPD-WII, it was hypothesized that the SbHNL-II homologues of the CPD-WII catalytic triad residues were able to facilitate a general base mechanism in SbHNL-II reminescent to the cyanogenese mechanism observed in MeHNL (*6*). Figure 3 shows the superposition of the active site of SbHNL-II and CPD-WII, based on the alignment of the corresponding backbone atoms.

While the hydrogen bonding between SbHNL-II His414 and SbHNL-II Asp359 indicates a close match to the corresponding residues His397 and Asp338 of CPD-WII, it is apparent from this figure that the spatial orientation of SbHNL-II Ser158 is completely different from that of the corresponding CPD-WII triad residue Ser146. The calculated rms differences for the SbHNL-II active site loop Cα atoms are 1.2 Å for Glu157, 3.9 Å for Ser158, and 1.4 Å for Gly159. The large differences are due to a dipeptide deletion next to the putative catalytic Ser158, compared to the corresponding CPD-WII β turn residues Glu145, Ser146, Tyr147, Ala148, and Gly149. The consequence of this deletion is a flip of the peptide bond between Glu157 and Ser158 which implies that SbHNL-II Ser158 is oriented away from the corresponding CPD-WII S'_1 subsite. In its new position, Ser158 OG is separated by more than 9 Å from its putative catalytic partner His414 NE2, and is therefore too far away for direct interaction. Surprisingly, the superposition reveals that in the SbHNL-II active site the carboxyl group of the SbHNL-II specific C-terminal Trp270 is spatially homologous to the hydroxyl group of CPD-WII Ser146. As simulation studies also indicate that the Trp270 carboxyl moiety can form a hydrogen bond to the active site water U40, which forms a bridge to the substrate, we suggest that it is important for catalysis. A detailed analysis of the functional role of Trp270 is presented below.

The "Broken" SbHNL-II Carboxypeptidase Binding Site. The second important difference between the SbHNL-II and CPD-WII structures is an extra 10-residue loop (Arg260−Trp270) at the SbHNL-II C-terminus of the α subunit. The loop is anchored to the surface of SbHNL-II through a salt bridge from Arg267 to Asp328, while the peptide bond between Phe269 and Trp270 is held in place by a hydrogen bond between Phe269 O and Tyr251 OH. These interactions place the indol moiety of C-terminal residue Trp270 into the S_1 subsite, present in CPD-WII (*23*). It is apparent that occluding of the SbHNL-II S_1 subsite by Trp270 prevents binding of peptide substrates, and we therefore describe this cavity as a broken binding site. In addition to the conformational rearrangement of the active site loop, the model of SbHNL-II indicates that this enzyme is unlikely to function as a carboxypeptidase, and this is in agreement with experimental data (*13*).

The putative SbHNL-II carboxypeptidase S'_1 subsite is constructed like that of CPD-WII and occupied by an elongated electron density feature which was modeled as decanoic acid (DA) und subsequently refined (see Materials and Methods). The carboxylate group of DA is hydrogen bonded to Glu157 OE1, Asn61 ND2, and Gly62 N. The aliphatic chain is projecting toward the protein surface and

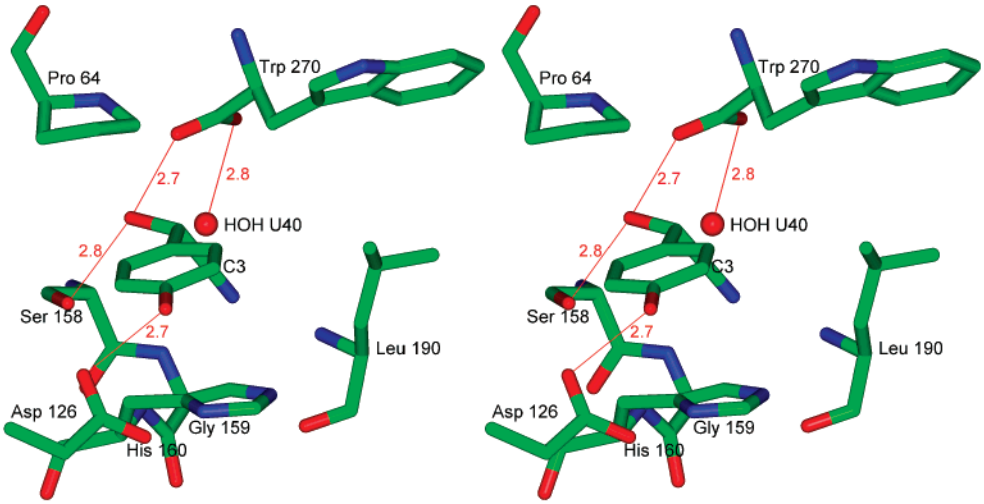

FIGURE 4: Stereoview of the hypothetical complex between (*S*)-4-hydroxybenzaldehyde cyanohydrin and SbHNL-II. Potential hydrogen bonds are shown as red lines. Interatomic distances are given in angstroms.

makes van der Waals contacts with Asn249, Tyr251, Tyr265, and Phe269. Inspection of a SbHNL-II inhibitor free structure (currently underway), prepared from a different charge of sorghum seeds, also shows this density, indicating that it is not an artifact. As there is no direct evidence that this fatty acid represents an endogenous cofactor, we assume that it is copurified with SbHNL-II.

Substrate Specificity of SbHNL-II-Catalyzed Cyanohydrin Formation. In addition to its physiological role, SbHNL-II is an important biocatalyst for the stereoselective synthesis of (*S*)-cyanohydrins (*24*). Experimental studies that were carried out demonstrate that SbHNL-II is able to discriminate in its specificity between aromatic and aliphatic carbonyl substrates; for example, benzaldehyde is converted with high enantiomeric excess into the corresponding (*S*)-benzaldehyde cyanohydrin, while the aliphatic counterpart cyclohexanecarbaldehyde (CHA) is not accepted as a substrate (*24*). The studies also indicate that binding and catalysis of substituted benzaldehydes depend significantly on the relative size and orientation of the substituents.

To understand the structural basis for this unusual specificity profile, a structure-based model of benzaldehyde in the SbHNL-II active site was constructed with the carbonyl group at the benzoate O1 position and the phenyl ring sandwiched between the side chain of His160 and the nonpolar face of Pro64. If benzaldehyde is extended at C4 by a hydroxyl group, this OH is nicely positioned within 2.5 Å of Asp126 OD1. Binding of 4-methoxybenzaldehyde in a similar way, however, would result in a steric clash of the additional methyl group with the main chain and side chain of Trp330. This interpretation is in agreement with experimental studies, showing that 4-methoxybenzaldehyde is not accepted as a substrate (*24*). Model building also shows that introduction of a double bond substituent at C1 of benzaldehyde, e.g., 3-phenylpropenal, would prohibit binding in the SbHNL-II active site due to steric conflicts, and this is also in agreement with experimental data. The model further reveals that the active site volume is widely opened toward the active site tunnel entrance next to C3, thereby facilitating the accommodation of benzaldehydes with large C3 substituents, most significantly demonstrated by the highly stereoselective conversion of 3-phenoxybenzaldehyde.

The substrate preference, however, is not exclusively determined by the size of the substrate as anticipated from the binding studies from aromatic and aliphatic carbonyl compounds. Cyclohexanecarbaldehyde differs from benzaldehyde in the presence of sp^3-hybridized C atoms with additional axial protons, rather than exclusively equatorial protons as in the phenyl ring of benzaldehyde, and is not accepted as a substrate. Additionally, carbonyl substrates with small aliphatic side chains such as butyraldehyde and isobutyraldehyde are also not accepted as substrates. Hypothetical models of SbHNL-II formed with CHA and with butyraldehyde reveal several closer than van der Waals contacts of the axial protons, in particular at C2−C4 of CHA and at C4 of butyraldehyde with Pro64 and His160, respectively. Most of the steric hindrance, however, could be released by some tilting and simultaneous rotation of either substrate relative to the active site cleft. The docking experiments, however, indicate that the fit will not be as good as the fit observed for benzaldehyde with respect to His160. In conclusion, the formation of stacking interactions of aromatic and heteroaromatic substrate side chains with the active site His160 imidazolium group appears to be most critical for inhibitor−substrate binding and helps define the specificity of this HNL.

Implications for the Catalytic Mechanism. The mechanism proposed for cyanogenesis of MeHNL (*6*) and HbHNL (*3*), which involves an active site histidine as a general base for abstracting a proton from the catalytic serine, cannot be applied to SbHNL. From the crystal structure of the SbHNL-II−benzoic acid complex, it can be seen that such a mechanism is not possible since His414 NE2 is 9 Å from Ser158 OG (see above). To establish the geometry of the cyanohydrin substrate and to investigate the possibility of conformational rearrangements of putative catalytic residues, we modeled the structure of the SbHNL-II active site with (*S*)-4-hydroxybenzaldehyde cyanohydrin, including also the active site water molecule U40 by energy minimization, using the INSIGHT II parameters (*20*). The result is shown in Figure 4.

From this simulation, the substrate is predicted to form a hydrogen bond between the cyanohydrin hydroxyl group and Ser158 OG (2.8 Å), while the nitrile group is directed toward

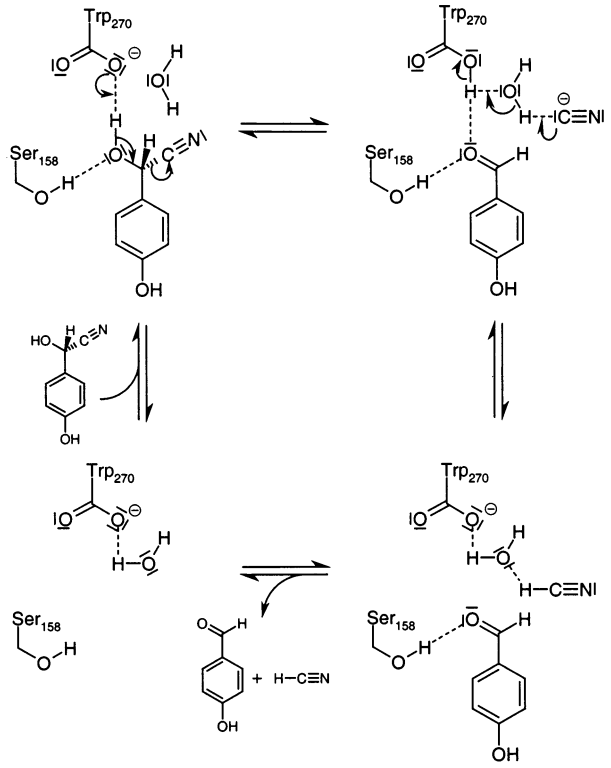

FIGURE 5: Simplified reaction mechanism for SbHNL-II (see the text).

Leu190 N (3.2 Å) and Gly159 N (3.1 Å). The phenyl ring is nicely sandwiched between the side chain of His160 and the nonpolar face of Pro64, and the additional hydroxyl group at C4 is within 2.7 Å of Asp126 OD1 and thereby assists in fixing the substrate with respect to the active site. The docking experiment further reveals that the carboxylate moiety of C-terminal residue Trp270 undergoes a movement to form a good hydrogen bond contact between Trp270 carboxylate O1 and the cyanohydrin hydroxyl group (2.7 Å), while the active site water U40 adjusts to maintain a hydrogen bond contact between the Trp270 carboxylate O2 (2.8 Å) and the substrate nitrile group (3.1 Å). The simulation indicates that there are no unfavorable contacts in the docked complex and only the active site water U40 had to move slightly to accommodate the (*S*)-cyanohydrin substrate in the putative catalytic position. Note that binding of an (*R*)-cyanohydrin substrate in a similar way would be impossible as its nitrile group would clash with the side chain atoms of Trp270.

Given the geometrical constraints expected for suitably positioned catalytic residues, the carboxylate group of C-terminal Trp270 is well situated to serve as the catalytic base in the first step of abstracting a proton from the cyanohydrin hydroxyl group as one of its carboxylate oxygens is 2.7 Å distant. In fact, the pH−rate dependence of SbHNL-II is consistent with the requirement to titrate a carboxylate group with a pK_a of ~4.7 to generate the necessary base (*24*). As there is no other direct interaction between the nitrile group and any protein residue, the active site water U40 is suggested to be the most likely candidate for proton transfer. In conclusion, a simplified reaction mechanism of SbHNL cyanogenesis is presented in Figure 5.

The entering cyanohydrin is hydrogen bonded to Ser158 OG and C-terminal Trp270 O1, which abstracts a proton from the OH group of (*S*)-4-hydroxybenzaldehyde cyanohydrin. A proton transfer from the Trp270 carboxyl group to the leaving nitrile group occurs via active site water U40. Protonation of the cyanide ion results in the products hydrogen cyanide and 4-hydroxybenzaldehyde, thus completing the catalytic mechanism. We have omitted, however, any discussion of whether the proton transfer occurs simultaneously via a substrate−Trp270−active site water ternary complex or alternatively via a charged intermediate.

An argument in favor of the combination of the Trp270 carboxylate moiety and one water molecule being the catalytic machinery is based on the chemical results (see above) showing that the SbHNL-II-catalyzed addition of HCN to carbonyl substrates in the reverse reaction occurs with pronounced *S* stereoselectivity. Due to the geometry of the active site, the cyanide ion can attack exclusively from below the phenyl plane (see Figure 4). Since the catalytic base Trp270 is situated above this plane, the active site water U40 is required to deprotonate HCN.

Evolution of SbHNL Structure and Function. The homology between SbHNL-II and carboxypeptidase CPD-WII shows without doubt that both molecules have evolved and diverged from a common ancestor late in evolution. Indeed, the level of sequence identity between different CPD subtypes of various monocotyledone species is lower than the level of sequence identity between SbHNL-II and CPD-WII (data not shown). Because SbHNL-II already has a broken carboxypeptidase active site (see above), we suggest that the ability to catalyze cyanogenesis was a secondary adaptation of an ancestral CPD enzyme of *S. bicolor*. From this perspective, it is apparent that the evolution of cyanogenic function in SbHNL-II is characterized by both the deletion of Trp147 and Ala148 (necessary for substrate binding) and the α subunit C-terminal extension up to Trp270 (necessary for catalysis) of the ancestral molecule. Recombinant methods should help in analyzing the structural features and the chemical constraints that have been selected by evolution to generate the new catalytic activity in SbHNL-II.

ACKNOWLEDGMENT

We thank A. Baro for her help in preparing the manuscript.

REFERENCES

1. Seigler, D. S. (1991) Cyanide and cyanogenic glycosides, in *Herbivores: Their interactions with secondary plant metabolites* (Rosenthal, G. A., and Berenbaum, M. R., Eds.) pp 35−77, Academic Press, New York.
2. Wajant, H., and Effenberger, F. (1996) *Biol. Chem. 377*, 611−617.
3. Zuegg, J., Gruber, K., Gugganig, M., Wagner, U. G., and Kratky, C. (1999) *Protein Sci. 8*, 1990−2000.
4. Lauble, H., Förster, S., Miehlich, B., Wajant, H., and Effenberger, F. (2001) *Acta Crystallogr. D57*, 194−200.
5. Dreveny, I., Gruber, K., Glieder, A., Thompson, A., and Kratky, C. (2001) *Structure 9*, 803−809.
6. Lauble, H., Miehlich, B., Förster, S., Wajant, H., and Effenberger, F. (2001) *Protein Sci. 10*, 1015−1022.
7. Bove, C., and Conn, E. E. (1961) *J. Biol. Chem. 236*, 201−210.
8. EMBL access number AJ421152.
9. Wajant, H., and Mundry, K.-W. (1993) *Plant Sci. 89*, 127−133.

10. Wajant, H., Mundry, K.-W., and Pfizenmaier, K. (1994) *Plant Mol. Biol. 26*, 735−746.
11. Liao, D., Breddam, K., Sweet, R., Bullock, T., and Remington, S. J. (1992) *Biochemistry 31*, 9796−9812.
12. Derewenda, U., Brzozowski, A. M., Lawson, D. M., and Derewenda, Z. S. (1992) *Biochemistry 31*, 1532−1541.
13. Wajant, H. (2002) unpublished results.
14. Lauble, H., Knödler, S., Schindelin, H., Förster, S., Wajant, H., and Effenberger, F. (1996) *Acta Crystallogr. D52*, 887−889.
15. Leslie, A. G. W. (1992) *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography 26*, 10−14.
16. Collaborative Computational Project Number 4 (1994) *Acta Crystallogr. D50*, 760−763.
17. Brunger, A. T. (1997) *X-PLOR*, version 3.851, Yale University Press, New Haven, CT.
18. Brunger, A. T. (1992) *Nature 355*, 472−475.
19. McRee, M. D. (1999) XtalView-Xfit: A versatile program for manipulating coordinates and electron density, *J. Struct. Biol. 125*, 156−165.
20. *INSIGHT II/DISCOVER User Guide*, release 98.0 (1998) BIOSYM/Molecular Simulations Inc., San Diego.
21. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 282−291.
22. Luzzati, V. (1952) *Acta Crystallogr. 5*, 802−810.
23. Bullock, T. L., Breddam, K., and Remington, S. J. (1996) *J. Mol. Biol. 255*, 714−725.
24. Hörsch, B. (1990) Enzymkatalysierte Synthese von (*R*)-und (*S*)-Cyanhydrinen in organischen Loesungsmitteln und deren saure Hydrolyse zu optisch aktiven 2-Hydroxycarbonsaeuren, Ph.D. Dissertation, University of Stuttgart, Stuttgart, Germany.
25. Babbitt, P. C., Hasson, M. S., Wedekind, J. E., Palmer, D. R. J., Barrett, W. C., Reed, G. J., Rayment, I., Ringe, D., Kenyon, G. L., and Gerlt, J. A. (1996) *Biochemistry 35*, 16489−16501.
26. Babbitt, P. C., and Gerlt, J. A. (1997) *J. Biol. Chem. 272*, 30591−30594.
27. Hutchinson, E. G., and Thornton, J. M. (1996) PROMOTIF: a program to identify and analyze structural motifs in proteins, *Protein Sci. 5*, 212−220.

BI020300O