Evolutionary recruitment of a flavin-dependent monooxygenase for the detoxification of host plant-acquired pyrrolizidine alkaloids in the alkaloid-defended arctiid moth *Tyria jacobaeae*

Claudia Naumann, Thomas Hartmann, and Dietrich Ober*

Institut für Pharmazeutische Biologie der Technischen Universität Braunschweig, Mendelssohnstrasse 1, D-38106 Braunschweig, Germany

Edited by May R. Berenbaum, University of Illinois at Urbana-Champaign, IL, and approved February 26, 2002 (received for review December 17, 2001)

Larvae of *Tyria jacobaeae* feed solely upon the pyrrolizidine alkaloid-containing plant *Senecio jacobaea*. Ingested pyrrolizidine alkaloids (PAs), which are toxic to unspecialized insects and vertebrates, are efficiently *N*-oxidized in the hemolymph of *T. jacobaeae* by senecionine *N*-oxygenase (SNO), a flavin-dependent monooxygenase (FMO) with a high substrate specificity for PAs. Peptide microsequences obtained from purified *T. jacobaeae* SNO were used to clone the corresponding cDNA, which was expressed in active form in *Escherichia coli*. *T. jacobaeae* SNO possesses a signal peptide characteristic of extracellular proteins, and it belongs to a large family of mainly FMO-like sequences of mostly unknown function, including two predicted *Drosophila melanogaster* gene products. The data indicate that the gene for *T. jacobaeae* SNO, highly specific for toxic pyrrolizidine alkaloids, was recruited from a preexisting insect-specific FMO gene family of hitherto unknown function. The enzyme allows the larvae to feed on PA-containing plants and to accumulate predation-deterrent PAs in the hemolymph.

N-oxidation | alkaloid sequestering insects | insect adaptation | chemical defense | gene recruitment

Larvae of the European cinnabar moth *Tyria jacobaeae* are specialized to tansy ragwort (*Senecio jacobaea* L.) as food plant. Pyrrolizidine alkaloids (PAs) present in this plant are sequestered by the larvae and are retained during all stages of metamorphosis (1, 2). PAs are typical plant secondary compounds thought to play an important role as defensive chemicals in the plant and its adapted insect herbivores (3). PAs are hepatotoxic and carcinogenic to vertebrates and mutagenic or genotoxic to insects and even to yeast (4); they also show strong antifeedant activity against various herbivores (5). In species of the genus *Senecio* (Asteraceae), senecionine *N*-oxide is synthesized in the roots as common PA-backbone structure, which subsequently is transported to other organs and tissues where it is chemically modified into the species-specific PA patterns (6). In the gut of mammalian herbivores feeding on PA-containing plants, the PA *N*-oxides are easily reduced, yielding the respective free bases, which are passively absorbed and thus become accessible to the bioactivating microsomal cytochrome P450-dependent monooxygenases (P450) of liver. These enzymes are part of xenobiotic metabolism in vertebrates (7, 8) and transform the tertiary PAs to unstable pyrroles, which are highly reactive alkylating agents. *N*-oxidation was shown to be a second pathway of PA metabolism in mammals, catalyzed by flavin-dependent monooxygenases (FMOs; ref. 9) or by an isoform of P450s (10, 11). The potentially toxic tertiary PAs are oxidized to their *N*-oxides, which no longer serve as substrates for bioactivating P450s (10, 12). In contrast to most mammals, guinea pigs, sheep, and hamsters (9, 13, 14) exhibit a high resistance against the toxic effects of PAs because of high levels of *N*-oxide forming FMO activity and low levels of bioactivating P450 (9).

Herbivorous insects also can cope with xenobiotics by means of P450s, as evidenced by resistance to commercial insecticides (15). However, during evolutionary adaptation to their food plants, herbivorous insects have developed specific mechanisms to tolerate specific plant defense chemicals, often involving P450s. In one example, *Papilio polyxenes*, the black swallowtail, feeds on Apiaceae species containing phototoxic furanocoumarins, which are efficiently detoxified in the insect by a P450 monooxygenase. This enzyme not only shows high substrate specificity for the furanocoumarins occurring in the host plant (16), but it also is specifically induced by these compounds (17) by means of the xanthotoxin-response element, which was identified in the promoter region of this inducible P450 enzyme (18). In another example, *Drosophila mettleri*, one of four *Drosophila* species endemic to the Sonoran desert in the southwestern United States, lives and breeds in soils soaked by rot exudate of the saguaro cactus, which contains toxic isoquinoline alkaloids that accumulate in the soil because of water evaporation (19). *D. mettleri*'s unique isoquinoline alkaloid resistance involves a cytochrome P450 monooxygenase that is specifically induced by isoquinolines from saguaro (20), but not from other cacti (20). Thus, *P. polyxenes* and *D. mettleri* each possess at least one P450 enzyme whose substrate specificity and inducibility evolved in response to specific plant defense compounds ("secondary metabolites") present in the respective host plant.

Tyria jacobaeae larvae do not simply detoxify PAs by *N*-oxidation, but store the nontoxic *N*-oxides as predation deterrents (Fig. 1; 21). The larvae advertise their toxic PAs by a striking yellow-black warning coloration. *N*-oxidation is only one of the different, recently identified strategies insects developed in evolution to safely handle plant-acquired PAs for their own defense (22, 23). The enzyme responsible for the *N*-oxidation of tertiary PAs in *Tyria* is senecionine *N*-oxygenase (SNO), which is a reduced nicotinamide-adenine dinucleotide phosphate (NADPH)-dependent flavin monooxygenase (FMO), not a cytochrome P450 enzyme (21). In contrast to the intensively studied mammalian multisubstrate FMOs (24), the *Tyria* enzyme is strictly substrate specific. It *N*-oxidizes only protoxic PAs, i.e., PAs with structural features

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: PA, pyrrolizidine alkaloid; P450, cytochrome P450-dependent monooxygenase; NADPH, nicotinamide-adenine dinucleotide phosphate; FMO, flavin-dependent monooxygenase; SNO, senecionine *N*-oxygenase; RACE, rapid amplification of cDNA ends.

Data deposition: The sequence reported in this paper has been deposited in the EMBL Nucleotide Sequence Database (accession no. AJ420233).

*To whom reprint requests should be addressed. E-mail: d.ober@tu-bs.de.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

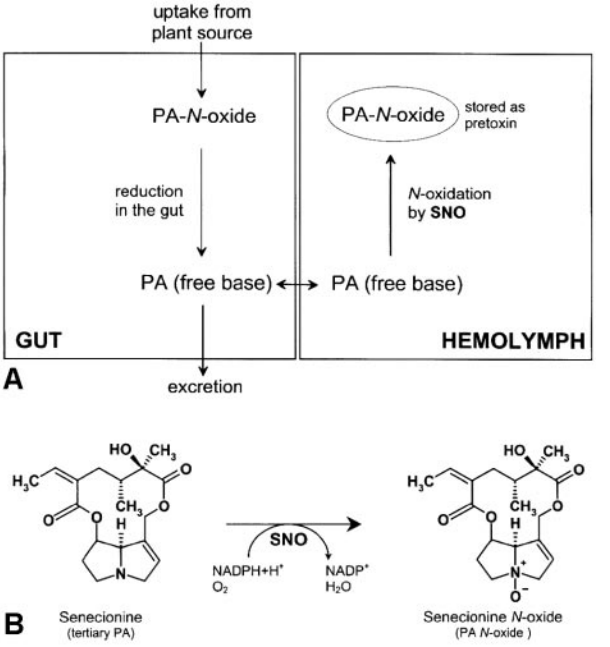

Fig. 1. PA *N*-oxides ingested from plants are reduced in the gut. In their tertiary form as lipophilic free bases, they permeate passively the membrane and are trapped in the hemolymph by efficient re-*N*-oxidation catalyzed by SNO of *T. jacobaeae* (*A*). SNO is an NADPH-dependent monooxygenase (*B*).

required by substrates of the bioactivating cytochrome P450 enzymes, indicating that the major function of SNO is to maintain sequestered protoxic PAs as nontoxic *N*-oxides. No functionally characterized FMO is known from insects (15, 25). To trace the evolutionary origin of SNO, we have obtained peptide sequences from the purified *T. jacobaeae* enzyme, cloned the cDNA, expressed the active protein in *E. coli* and investigated its gene phylogeny. The data indicate that the PA-specific SNO of *T. jacobaeae* was recruited from a preexisting family of FMO-like proteins of hitherto unknown functions.

Materials and Methods

Radiochemicals. [*N*-Methyl-^{3}H]atropine (87 Ci/mmol) was purchased from DuPont. Pyrrolizidine alkaloids applied as enzyme substrates were obtained as described (21).

Purification and Microsequencing of SNO. SNO was purified from hemolymph of *T. jacobaeae* as described (21). The purified enzyme was transferred to a poly(vinylidene difluoride) membrane (Millipore) and was subjected to microsequencing. Peptides of purified SNO obtained by digestion with trypsin were separated by means of HPLC on a Vydac C18 column (4.6 mm × 150 mm) and sequenced according to ref. 26.

RNA Isolation and cDNA Synthesis. Total RNA was isolated by using the RNeasy Mini Kit (Qiagen) from one-week-old larvae of *T. jacobaeae* and of the larvae and adults of *D. melanogaster* frozen in liquid nitrogen. Total RNA (2 μg) was reverse-transcribed by Superscript II (Invitrogen) in a total volume of 50 μl by using an oligo(dT)$_{17}$ primer (0.1 μM; Table 1).

PCR with Degenerate Primers. Degenerate primers (P1–6) were constructed according to peptides resulting from partial sequencing of the SNO protein (Table 1). Amplification of 1 μl of cDNA was performed with *Taq* DNA Polymerase (Qiagen, Chatsworth, CA) in a total volume of 25 μl by using all possible combinations of forward and reverse primers (2.0 μM each), including oligo(dT)$_{17}$ primer (1.0 μM). A touch-down protocol with decreasing annealing temperatures from 60°C to 45°C (0.5°C per cycle, 45°C constant for 20 further cycles) was used for the temperature program . With primers P5 and oligo(dT)$_{17}$, a 554-bp fragment was amplified, which subsequently was electrophoretically purified, subcloned by using the TOPO TA cloning kit (Invitrogen) according to the manufacturer's instructions, and sequenced.

Amplification of the 5′-cDNA End of SNO. For identification of the 5′-end of the cDNA fragment, gene-specific reverse primers

Table 1. Peptides resulting from microsequencing of SNO and primers for cloning of SNO

Peptide		Primer	
N-terminal	**pep1** (S)SASRVTIIGAGYSGLATARYLQDY(S)L		degenerate
	→	P1: 5′-dGTNACNATHATHGGNGCNGG-3′	
	←	P2: 5′-dGCNMGNTAYYTNCARGAYTA-3′	
internal	**pep2** YLQDYGLXYTIFEATP(L)		
	→	P3: 5′-dAYACNATHTTYGARGCNACNCC-3′	
	←	P4: 5′-dGGNGTNGCYTCRAADATNGTRT-3′	
	pep3 VGTDE(D)GIPIYSSN(Y)(K)		
	pep4 HWLEHVQMLR		
	→	P5: 5′-dYTGGYTNGARCAYGTNCARATG-3′	
	←	P6: 5′-dRCATYTGNACRTGYTCNARCCA-3′	
		P7: 5′-dTCT TCC AAC AAA GTC TTT-3′	for 5′-RACE
		P8: 5′-dCAC GGA CGC ATA AAC TGG TGG CAA TAA-3′	
		P9: 5′-dCGC CTG CTT CTT TGC GAA GAG CTT TAA-3′	
		P10: 5′-dGAC CAT ATC GAA TTC TTC AA-3′	
		P11: 5′-dTCC TCA AAC ACA GCA CCG TTT GCA GTG AA-3′	
		P12: 5′-dGTG CAC CAA CTT CGA GGT AAT ATT GGA AA-3′	
	***T. jacobaeae* SNO**	P13: 5′-dATA TAC ATA TGT TTC GGA AAT TTG TGA TAA TG-3′	for full-length cDNA
		P14: 5′-dATA TAC ATA TGA GTT CAG CGT CTC GAG TAT GCA-3′	
		P15: 5′-dATT GGA TCC **TTA** ATC GTC ATA GGG ACA TAC TT-3′	
	***D. melanogaster* CG3006**	P16: 5′-dTAT ATC ATA TGA TGA GCG TTT GCA TAA TTG GG-3′	
		P17: 5′-dTAT ATG ATC AGT TGA GTT TAA TGA ATG TCT CGT C-3′	
	***D. melanogaster* CG3174**	P18: 5′-dTAT ATC ATA TGA TCA AAA CTT CAG TGG ACA AGC-3′	
		P19: 5′-dATA GGA TCC **TCA** TAC AAT AGC TTC TCC GTT CTT C-3′	
		oligo(dT)$_{17}$ primer 5′-dGTCGACTCGAGAATTC(T)$_{17}$-3′	

For the peptides, amino acids marked by an X could not be identified and those in parentheses were uncertain; for nucleotides, the IUB code was used. For primers P13–P19, restriction sites are underlined, and start and stop codons are shown in bold.

(P7-P9 and, thereafter, P10-P12) were designed for two rounds of 5′-rapid amplification of cDNA ends (RACE) by using the 5′ RACE System (Invitrogen) with dC- and dA-tailing according to the manufacturer's instructions.

Generation and Expression of Full-Length cDNA Clones. For amplification of the full-length SNO-cDNA with and without the signal peptide, two forward primers (P13–14) and one reverse primer (P15) were constructed (Table 1). Amplification was performed by using Platinum *Pfx* DNA Polymerase (Invitrogen) in a 100-μl reaction mixture with 10 μl of the oligo(dT)$_{17}$ primed cDNA as template. The resulting 1,388-bp and 1,325-bp fragments were purified by electrophoresis, *Nde*I/*Bam*HI-digested, and ligated into a *Nde*I/*Bam*HI-linearized pET3a vector (Novagen) for expression with the T_7 polymerase system (27). The ligation product was transformed into *E. coli* XL1-Blue cells (Stratagene), and positive clones were selected by PCR amplification by using the primer pair P13/P15 and P14/P15, respectively. Plasmid DNA of a positive clone was purified and used for sequencing and overexpression after transformation in *E. coli* BL21(DE3) (Stratagene) and induction with 0.4 mM isopropyl β-D-thiogalactoside (IPTG). To clone and express the cDNAs coding for the gene products of the *D. melanogaster* genes CG3006 and CG3174 (accession nos. AE003462 and AE003789, respectively), the primer pairs P16–17 and P18–19 were used (Table 1) for PCR amplification with cDNA of *D. melanogaster* larvae and adults as templates. For ligation in a *Nde*I/*Bam*HI-linearized pET3a vector, the 1,266-bp fragment of the CG3006 gene product was digested with *Nde*I/*Bcl*I, and the 1,307-bp fragment of the CG3174 gene product with *Nde*I/*Bam*HI. Further transformation and expression of the resulting plasmids were carried out as described for *T. jacobaeae* SNO.

Sequence Analysis. Analysis of SNO-cDNA was performed by using the SignalP server (28) for prediction of signal peptides and their cleavage sites, the PSORT II server (29) for detection of sorting signals and subcellular localization, and the TMHMM2.0 server (30) for prediction of transmembrane protein helices. The netOglyc server (31) was used to search for *O*-glycosylation sites. For similarity searches, the BLAST (32) algorithm on the GenBank website (http://www.ncbi.nlm.nih.gov) was used. Sequences were aligned by using CLUSTALX (33), and the phylogenetic tree was constructed by using the PROML algorithm of the PHYLIP V.3.6 program package (http://evolution.genetics.washington.edu/phylip.html). RELL bootstrap proportions were calculated with PROTML of the MOLPHY package (34).

Solubilization and Refolding of Expressed Protein from Inclusion Bodies. Inclusion bodies were purified and solubilized by using the protein refolding kit (Novagen), according to the manufacturer's protocol with slight modifications. *E. coli* BL21(DE3) cells expressing SNO were harvested, resuspended in 0.1 culture volume of wash buffer (20 mM Tris·HCl pH 7.5/10 mM EDTA/1% Triton X-100), and broken by sonication. After centrifugation, the pellet was washed twice with 0.1 culture volumes of wash buffer and then solubilized at room temperature in solubilization buffer (0.1 M glycine, pH 11/0.3% *N*-lauroylsarcosine/1.0 mM DTT) at a concentration of 10–20 mg/ml wet weight of the pelleted cell debris, including the inclusion bodies. For refolding, the denatured protein was dialyzed twice (3 h and 18 h, respectively) against 50 sample volumes dialysis buffer (20 mM Tris·HCl, pH 8.5) supplemented with 0.1 mM DTT, twice again (3 h each) against dialysis buffer, and then once again (17 h) against 25 sample volumes dialysis buffer supplemented with 1.0 mM reduced glutathione/0.2 mM oxidized glutathione/1 mM flavinadeninedinucleotide (FAD). All dialysis steps were performed at 4°C. The refolded protein was stored at 4°C and used within 1 week for enzyme assays without further purification steps.

Enzyme Assays. Enzyme activity was determined photometrically at 37°C (substrates: senecionine, seneciphylline, heliotrine, monocrotaline, axillarine, sarracine, dimethylaniline, caffeine, *L*-proline, each 2 mM) or in qualitative tracer assays (substrates: [^{14}C]senecionine, [^{3}H]supinidine [^{14}C]retronecine, [^{14}C]senkirkine, [^{3}H]phalaenopsine, [^{3}H]atropine, each 0.023 μCi/assay and 0.045 μCi/assay for the [^{14}C]- and the [^{3}H]-labeled substrates, respectively). The assays were performed as described (21), but with the modifications that recombinant SNO was assayed directly in the dialysis buffer supplemented with 120 μM NADPH, and that the radioassay was performed in a total volume of 60 μl. Tested substrates for *Drosophila* FMOs were cysteamine, glutathione, and *L*-cysteine (2 mM each) for the photometric assay and [^{14}C]senecionine (0.023 μCi/assay) for the tracer assay.

Results

Identification and Analysis of cDNA Coding for SNO. SNO purified from *T. jacobaeae* shows two distinct bands after SDS/PAGE (21). Both were subjected to microsequencing, resulting in one *N*-terminal peptide of 26 amino acids (pep1). Further, three peptides of 17, 16, and 10 amino acids (pep2 to pep4) were obtained after digestion of the protein with trypsin and sequencing of the fragments (Table 1). Based on the sequence information, degenerate oligonucleotides were synthesized (Table 1) and used for reverse transcription (RT)-PCR with total RNA of one-week-old larvae of *T. jacobaeae* as template. The combination of the oligo(dT)$_{17}$-primer with primer P5 resulted in a 554-bp fragment, which contained the ORF starting with peptide pep4 and ending with the polyA-tail of the cDNA. Because only a truncated cDNA of a total length of 978 bp was obtained with the first set of gene-specific primers for the 5′ RACE technique, a second set of primers was constructed. The ORF of 456 amino acids of the resulting full-length cDNA of 1,701 bp contained the N-terminal peptide as well as all internal peptides identified by microsequencing (Fig. 2). Two of the ambiguous amino acids (the penultimate amino acid of pep1 and the last of pep2) were shown to be glycine and asparagine, respectively, and the amino acid in position 7 of pep1 was shown to be cysteine instead of threonine. The 5′- and 3′-untranslated regions of SNO-cDNA are of 81-bp and 249-bp lengths, respectively. The 3′-untranslated region contains a putative polyadenylation signal AATAAT (35) 31 bp upstream of the polyA tail. Sequence analysis using the SignalP server predicts an N-terminal signal peptide for the vesicular pathway with a proposed cleavage site between amino acids 22 and 23. This prediction is in accordance with the N-terminal peptide resulting from microsequencing starting with amino acid 23. With the PSORT II server, an N-terminal signal peptide is also suggested, and because no ER-retention signal is detectable, an extracellular localization of this protein is predicted. The encoded protein has a predicted subunit size of 52.2 and 49.8 kDa with and without signal peptide, respectively. No putative transmembrane helices were identified. Thus, SNO does not seem to possess a C-terminal membrane anchor as is the case for mammalian multisubstrate FMO (24). The putative binding site for FAD (GAGYSG between residues 10 and 15 of the processed protein without the signal peptide) and NADPH (GAGPSG between residues 193 and 198 of the processed protein without the signal peptide) are highly conserved in sequence motif and position with regard to other known FMOs (24) and with the functionally related bacterial cyclohexanone monooxygenases (36). Asparagine residues being part of the *N*-glycosylation consensus sequence (N-x-S/T-x with x $\neq$ P) are at positions 50, 150, 161, and 229. A putative *O*-glycosylation site was detected at Thr-306. In an alignment of *Tyria* SNO with mammalian FMOs, the potential *N*-glycosylation site at Asn-150 corresponds to the known *N*-glycosylation site at Asn-120 of pig FMO1 (ref. 37, and data not shown).

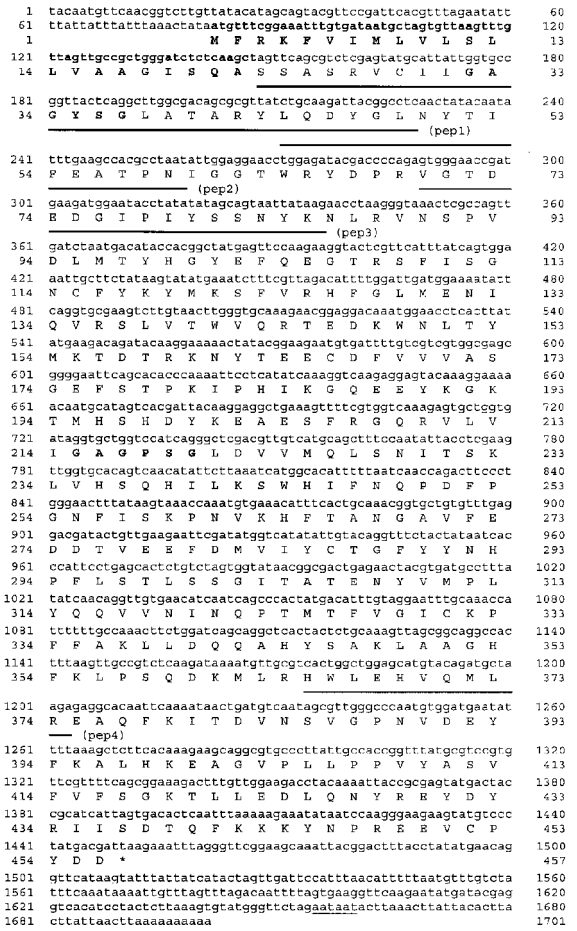

Fig. 2. Nucleotide and deduced amino acid sequences of SNO of *T. jacobaeae*. The N-terminal signal peptide as well as the FAD- and NADPH-binding sites (between amino acids 32–37 and 215–220 of the unprocessed protein, respectively) are shown in bold. The positions of the peptides resulting from microsequencing (pep1-pep4) are indicated by bars. The putative polyadenylation signal in the 3′-untranslated region is underlined.

Expression and Solubilization of Recombinant Proteins. Heterologous expression of the SNO of *T. jacobaeae* in *E. coli* BL21(DE3) resulted in the formation of inclusion bodies, although the native enzyme is a soluble protein (21). Attempts to avoid the formation of inclusion bodies by varying the culture conditions (temperature for growth and induction, concentration of IPTG) were unsuccessful. The inclusion bodies were solubilized under mild denaturing conditions followed by several dialysis steps. The solubilization procedure resulted in a single prominent protein of approximately 50 kDa in the soluble fraction (Fig. 3). Because the solubilized protein remained unstable in solution after transfer to the 10 mM potassium phosphate buffer (pH 7.0), which was used for the biochemical characterization of the native protein (21), enzyme assays were performed in dialysis buffer pH 8.5. At this pH, the native enzyme possesses more than 80% of its activity than at the optimum of pH 7.0. The reduced solubility may be caused by the absence of a putative glycosylation or a phosphorylation, because the calculated isoelectric point of the expressed amino acid sequence of SNO without signal peptide is about 6.9, whereas the native protein showed a pI of 4.9 ± 0.1 (21). Expression of the *Drosophila* cDNAs coding for the CG3006 and CG3174 gene products of 416 and 429 amino acids in length, respectively, also resulted in inclusion bodies. They were solubilized as described above for the SNO.

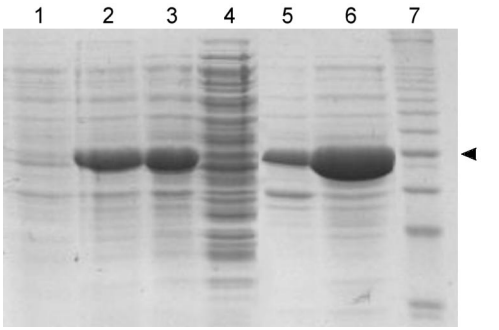

Fig. 3. Expression and solubilization of SNO of *T. jacobaeae* without signal peptide. Coomassie blue stain of an SDS/PAGE on a 12% gel. *E. coli* culture before and after induction with IPTG [lane 1 (7.5 μl) and lane 2 (7.5 μl), respectively], pellet and supernatant after sonification of cells [lane 3 (13 μg) and lane 4 (28 μg), respectively], pellet and supernatant after solubilization of inclusion bodies [lane 5 (3 μg) and lane 6 (18 μg), respectively], and 10-kDa protein ladder (Invitrogen; in lane 7, the 50-kDa band is indicated by an arrowhead).

Substrate Specificity of Native and Recombinant SNO. After expression and refolding of the recombinant SNO, activity was only detectable for the protein without signal peptide, which was used for further study. The recombinant SNO was tested with the substrates listed in Table 2. The substrates cover the five main structural types of PAs (38), as well as some related compounds known to be substrates of mammalian multisubstrate FMO. The data show that the recombinant and the native SNO displays the same high specificity for PAs with the structural features already described for the native enzyme (21); that is, the 1–2 double bond, the esterified allylic hydroxyl group at C9, and a free or esterified hydroxyl group at C7, matching exactly the structural features of PAs found to be responsible for their toxic effects (4). The only observed difference between the native and the recombinant form of SNO is the low specific activity of the latter; i.e., 0.2 to 0.5 nkat/mg in comparison to 77.6 nkat/mg of the native enzyme (21). This low specific activity may be caused by only a small portion of active enzyme after mainly incorrect folding of the recombinant protein or to missing posttranslational modifications.

Table 2. Substrate specificity of recombinant SNO in comparison to native SNO

Substrate	Activity with SNO, %	
	Native	Recombinant
Pyrrolizidine alkaloids		
Senecionine type		
Senecionine	100	100
Seneciphylline	95	93
Senkirkine	n.d.	n.d.
Monocrotaline type		
Monocrotaline	92	94
Axillarine	74	83
Lycopsamine type		
Heliotrine	25	49
Triangularine type		
Sarracine	<1	<10
Phalaenopsine type		
Phalaenopsine	n.d.	n.d.
Necine bases and other substrates		
Retronecine	n.d.	n.d.
Supinidine	n.d.	n.d.
Atropine	n.d.	n.d.
Coffeine	n.d.	n.d.
L-Proline	n.d.	n.d.
Dimethylaniline	n.d.	n.d.

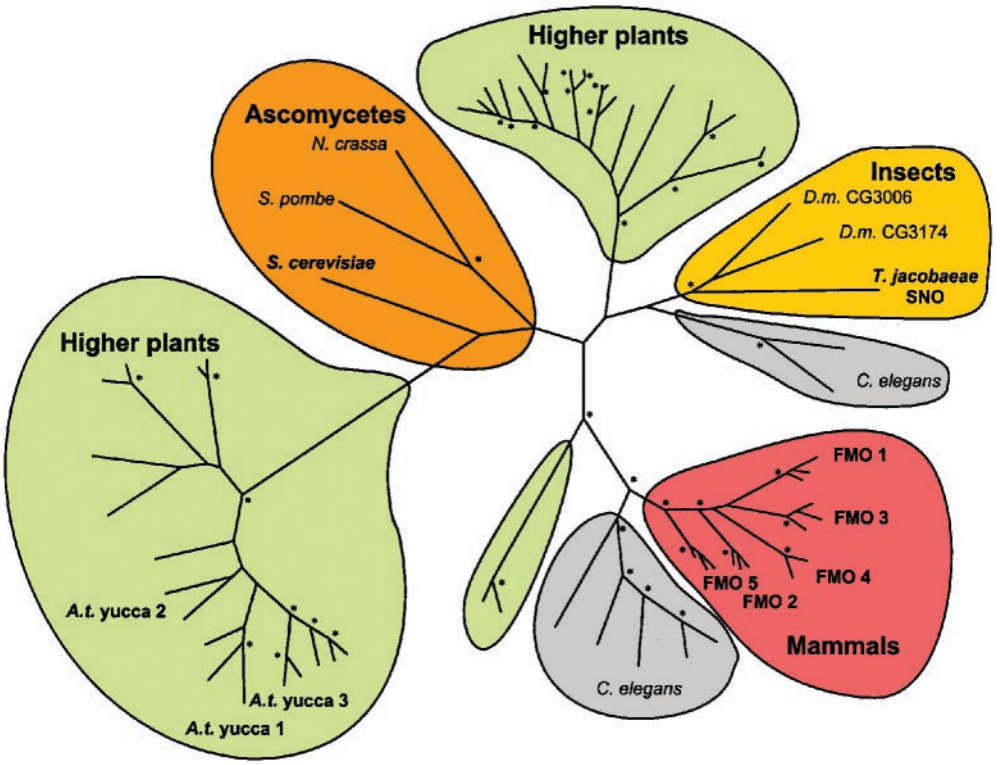

Fig. 4. Phylogenetic analysis of proteins supposed to be related to SNO of *T. jacobaeae*. Sequences resulted from BLAST search (32) with an *e* value of 10^{-9} as the exclusion limit. The tree was constructed by using the PROML algorithm of the PHYLIP V.3.6 program package. Branches with RELL bootstrap proportions ≥ 0.9 are indicated by an asterisk. Sequences functionally characterized are in bold.

Comparative Sequence Analysis of FMOs. By using the amino acid sequence of the SNO of *Tyria jacobaeae* as query sequence, we performed a BLAST search and analyzed all hits as putative homologues that were scored with e values $\leq 10^{-9}$, excluding from the analysis sequences in which no FMO-like motif is detectable. Thus, sequences were excluded which are flavin-dependent but which are NADPH-independent, such as amine oxidases or plant protoporphyrinogen IX oxidases. All putative homologues were checked to eliminate fragments and sequencing artifacts. Those hits that belonged to *Arabidopsis thaliana*, *Caenorhabditis elegans*, and *Homo sapiens* were replaced by the corresponding sequence data of the respective genome projects, as they are available in the databases (http://www.arabidopsis.org, http://www.wormbase.org, http://www.ncbi.nlm.nih.gov/genome/guide/human/). A phylogenetic analysis of eukaryotic FMO sequences thus collected is shown in Fig. 4 (accession numbers are available as supporting information, which is published on the PNAS web site, www.pnas.org).

Most of the sequences that turned up by the BLAST search were of unknown function and emerged mainly from the genome projects of *A. thaliana* and *C. elegans*. Of the sequences with defined function, most belong to the mammalian multisubstrate FMO family, which forms a well defined cluster of sequences and which splits up into five separate clusters representing the five isoforms FMO1 to FMO5. For clarity, only the sequences of human, mouse, and rabbit were considered in this group. In *C. elegans*, seven sequences with unknown function are found, which form two distinct clusters of two and five sequences, respectively. The larger one seems to be related to the mammalian FMOs. Within the ascomycetes, only one FMO-like sequence is found per genome, which was functionally characterized from yeast as responsible for redox regulation (39). In contrast to ascomycetes, plants possess a far greater number of homologues, 29 sequences in the genome of *A. thaliana* alone. The plant sequences form three distinct families of gene products, of which one contains only two uncharacterized sequences. The two other families include both a huge number of sequences fanning out of a few ancestors, showing the increase of homologous sequences in plants caused by gene duplication events as already confirmed for *Arabidopsis*, where duplicated regions encompass about 60% of the genome (40). For three sequences in one of these families, it was suggested that these FMO-like enzymes participate in auxin biosynthesis (41).

The two highest scoring hits resulted from the gene products of the CG3006 and CG3174 genes of *Drosophila melanogaster*, which share 44% amino acid identity and which show 36% and 35% identity to the amino acid sequence of the SNO of *T. jacobaeae*, respectively. Because *D. melanogaster* does not feed on PA-containing plants and has no need to detoxify them, we cloned and expressed both cDNAs of *D. melanogaster*. We were unable to identify an activity and, thus, a physiological role of these gene products of *Drosophila*. The fact that several EST sequences are available in the database for both gene products, and that we were able to clone both cDNAs successfully by RT-PCR, confirms the expression of these genes in *Drosophila*.

Conclusions

The cDNA coding for SNO contains an N-terminal signal peptide characteristic of extracellular proteins. Moreover, the sequence lacks hydrophobic transmembrane anchor motifs, consistent with our biochemical results that SNO is an extracellular soluble enzyme localized in the hemolymph. The solubility of *Tyria* SNO is uncommon, because all mammalian FMOs are microsomal enzymes (24, 42). However, soluble FMOs are probably more widespread than expected, because most FMO-like sequences compared in this study are devoid of transmembrane helices. This observation is also true for yeast FMO, which was shown to be associated with the outer membrane of the endoplasmic reticulum (ER) but which can be solubilized with carbonate or glucose 6-phosphate without destruction of the ER membrane (43). A soluble cytosolic FMO was found in *Trypanosoma cruzi* (44).

Information about the physiological role of most flavin-dependent monooxygenases is sparse. For yeast, which possesses only one FMO-like gene, the encoded enzyme catalyzes the oxi-

dation of biological thiols such as glutathione, cysteine, and cysteamine (45), thus maintaining the optimal redox environment at the cytosolic surface of the ER (39). The production of these oxidizing equivalents is essential for a proper folding of disulfide containing proteins at the ER; yeast FMO is vital to the yeast response to reductive stress (46). A similar role was postulated for the above mentioned *Trypanosoma* FMO (44) as well as for the mammalian FMOs (47). In contrast to the multisubstrate FMOs engaged in xenobiotic metabolism, FMO5 does not accept the substrates typical for the other FMOs, suggesting that FMO5 is not a drug-metabolizing enzyme (48). The phylogenetic tree in Fig. 4 indicates that FMO5 is the closest relative of the common ancestor of all mammalian FMOs. Whether it still possesses the initial function of eukaryotic FMOs, which was postulated to be the redox regulation in the cell, remains speculative.

A role for FMOs in chemical defense, as shown here for SNO of *T. jacobaeae,* also was proposed for marine invertebrates such as sponges and mollusks, in which activation of protoxins might be catalyzed by FMOs (49). FMO-like sequences from genome projects of *C. elegans* and *A. thaliana* await functional characterization. Because eukaryotes like yeast, *Drosophila*, or human possess only one, two, and five FMO-like genes in their genomes, respectively, hitherto unidentified FMO-like sequences from *A. thaliana* may have plant-specific functions. These functions might be involved in plant secondary metabolism or biosynthesis of phytohormones, functions in which monooxygenases of the cytochrome P450 superfamily are involved (50). For the latter, evidence was presented recently, as an FMO-like gene product of *A. thaliana* (YUCCA) seems to catalyze the *N*-oxygenation of tryptamine, which is the rate-limiting step in auxin biosynthesis (41). The involvement of a flavin-dependent monooxygenase is also most probable in seedlings of *Crotalaria scassellatii*, in which a monooxygenase *N*-oxidizes pyrrolizidine alkaloids stored in the seeds with nearly the same substrate specificity as *Tyria* SNO (51).

Insects have systems that are efficient and highly adapted to detoxify plant-derived allelochemicals, in which the cytochrome P450 monooxygenases play a central role (15). Flavin-dependent monooxygenases from insects were not previously known to participate in allelochemical detoxification nor in any other defined metabolic function (15, 25). The genome of *D. melanogaster* contains two predicted FMO-like proteins, which show an unequivocal homology to the FMO of *T. jacobaeae*. Because EST-sequences of these two genes available in the database were extracted from cDNA libraries of embryo and head tissues, the respective proteins are most probably not involved in detoxification processes. Assuming that at least one ortholog of these *Drosophila* genes is or was in earlier times present in the genome of *T. jacobaeae*, it could have a common ancestor with SNO. Of this ancestral gene, the SNO was recruited by a duplication event during adaptation of *T. jacobaeae* to the PA-containing host plant *S. jacobaea*.

We thank Dr. M. Kies and R. Getzlaff (Gesellschaft für Biotechnologishe Forschung, Braunschweig) for microsequencing the protein peptides, M. Brückmann for his help in purification of the SNO, and Dr. W. Martin for discussion. This work was supported by grants from the Deutsche Forschungsgemeinschaft and Fonds der Chemischen Industrie.

1. Aplin, R. T., Benn, M. H. & Rothschild, M. (1968) *Nature (London)* **219,** 747–748.
2. Aplin, R. T. & Rothschild, M. (1972) in *Toxins of Animal and Plant Origin*, eds. De Vries, A. & Kochva, E. (Gordon & Breach, New York), Vol. **2,** pp. 579–595.
3. Hartmann, T. (1999) *Planta* **207,** 483–495.
4. Roeder, E. (1995) *Pharmazie* **50,** 83–98.
5. Boppré, M. (1986) *Naturwissenschaften* **73,** 17–26.
6. Hartmann, T. & Dierich, B. (1998) *Planta* **206,** 443–451.
7. Mattocks, A. R. (1986) *Chemistry and Toxicology of Pyrrolizidine Alkaloids* (Academic, London).
8. Winter, C. K. & Segall, H. J. (1989) in *Toxicants of Plant Origin*, ed. Cheeke, P. R. (CRC, Boca Raton, FL), Vol. I, pp. 23–40.
9. Miranda, C. L., Chung, W., Reed, R. E., Zhao, X., Henderson, M. C., Wang, J. L., Williams, D. E. & Buhler, D. R. (1991) *Biochem. Biophys. Res. Commun.* **178,** 546–552.
10. Williams, D. E., Reed, R. L., Kedzierski, B., Ziegler, D. M. & Buhler, D. R. (1989) *Drug Metab. Dispos.* **17,** 380–386.
11. Miranda, C. L., Reed, R. L., Guengerich, F. P. & Buhler, D. R. (1991) *Carcinogenesis* **12,** 515–520.
12. Williams, D. E., Reed, R. L., Kedzierski, B., Dannan, G. A., Guengerich, F. P. & Buhler, D. R. (1989) *Drug Metab. Dispos.* **17,** 387–392.
13. Huan, J.-Y., Miranda, C. L., Buhler, D. R. & Cheeke, P. L. R. (1998) *Toxicol. Lett.* **99,** 127–137.
14. Huan, J.-Y., Miranda, C. L., Buhler, D. R. & Cheeke, P. R. (1998) *Toxicol. Appl. Pharmacol.* **151,** 229–235.
15. Brattsten, L. B. (1992) in *Herbivores – Their Interactions with Secondary Plant Metabolites*, eds. Rosenthal, G. A. & Berenbaum, M. R. (Academic, San Diego), Vol. 2, pp. 175–242.
16. Ma, R., Cohen, M. B., Berenbaum, M. R. & Schuler, M. A. (1994) *Arch. Biochem. Biophys.* **310,** 332–340.
17. Hung, C. F., Harrison, T. L., Berenbaum, M. R. & Schuler, M. A. (1995) *Insect Mol. Biol.* **4,** 149–160.
18. Prapaipong, H., Berenbaum, M. R. & Schuler, M. A. (1994) *Nucleic Acids Res.* **22,** 3210–3217.
19. Frank, M. R. & Fogleman, J. C. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 11998–12002.
20. Danielson, P. B., Foster, J. L. M., McMahill, M. M., Smith, M. K. & Fogleman, J. C. (1998) *Mol. Gen. Genet.* **259,** 54–59.
21. Lindigkeit, R., Biller, A., Buch, M., Schiebel, H. M., Boppre, M. & Hartmann, T. (1997) *Eur. J. Biochem.* **245,** 626–636.
22. Hartmann, T., Theuring, C., Schmidt, J., Rahier, M. & Pasteels, J. M. (1999) *J. Insect Physiol.* **45,** 1085–1095.
23. Hartmann, T., Theuring, C., Witte, L. & Pasteels, J. M. (2001) *Insect Biochem. Mol. Biol.* **31,** 1041–1056.
24. Cashman, J. R. (1995) *Chem. Res. Toxicol.* **8,** 165–181.
25. Schlenk, D. (1998) *Comp. Biochem. Physiol. C Comp. Pharmacol. Toxicol. Endocrinol.* **121,** 185–195.
26. Weyand, M., Hecht, H., Kiess, M., Liaud, M., Vilter, H. & Schomburg, D. (1999) *J. Mol. Biol.* **293,** 595–611.
27. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990) *Methods Enzymol.* **185,** 60–89.
28. Nielsen, H., Engelbrecht, J., Brunak, S. & Von, H. G. (1997) *Proetin Eng.* **10,** 1–6.
29. Nakai, K. & Horton, P. (1999) *Trends Biochem. Sci.* **24,** 34–36.
30. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. L. (2001) *J. Mol. Biol.* **305,** 567–580.
31. Hansen, J. E., Lund, O., Engelbrecht, J., Bohr, H., Nielsen, J. O., Hansen, J. E. S. & Brunak, S. (1995) *Biochem. J.* **308,** 801–813.
32. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
33. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997) *Nucleic Acids Res.* **25,** 4876–4882.
34. Adachi, J. & Hasegawa, M. (1996) MOLPHY V.2.3 (Institute of Statistical Mathematics, Tokyo).
35. Juretic, N. & Theus, M. (1991) *FEBS Lett.* **290,** 4–8.
36. Sheng, D., Ballou, D. P. & Massey, V. (2001) *Biochemistry* **40,** 11156–11167.
37. Korsmeyer, K. K., Guan, S., Yang, Z. C., Falick, A. M., Ziegler, D. M. & Cashman, J. R. (1998) *Chem. Res. Toxicol.* **11,** 1145–1153.
38. Hartmann, T. & Witte, L. (1995) in *Alkaloids: Chemical and Biological Perspectives*, ed. Pelletier, S. W. (Pergamon, Oxford), Vol. 9, pp. 155–233.
39. Suh, J. K., Poulsen, L. L., Ziegler, D. M. & Robertus, J. D. (2000) *Arch. Biochem. Biophys.* **381,** 317–322.
40. *Arabidopsis* Genome Initiative (2000) *Nature (London)* **408,** 796–815.
41. Zhao, Y., Christensen, S. K., Fankhauser, C., Cashman, J. R., Cohen, J. D., Weigel, D. & Chory, J. (2001) *Science* **291,** 306–309.
42. Ziegler, D. M. (1990) *Trends Pharmacol. Sci.* **11,** 321–324.
43. Suh, J. K., Poulsen, L. L., Ziegler, D. M. & Robertus, J. D. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 2687–2691.
44. Agosin, M. & Ankley, G. T. (1987) *Drug Metab. Dispos.* **15,** 200–203.
45. Suh, J. K., Poulsen, L. L., Ziegler, D. M. & Robertus, J. D. (1996) *Arch. Biochem. Biophys.* **336,** 268–274.
46. Suh, J. K. & Robertus, J. D. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 121–126.
47. Poulsen, L. L. & Ziegler, D. M. (1977) *Arch. Biochem. Biophys.* **183,** 563–570.
48. Overby, L. H., Buckpitt, A. R., Lawton, M. P., Atta, A. A. E., Schulze, J. & Philpot, R. M. (1995) *Arch. Biochem. Biophys.* **327,** 275–284.
49. Kurelec, B., Britvic, S., Krca, S., Muller, W. E. & Zahn, R. K. (1987) *Comp. Biochem. Physiol. C Comp. Pharmacol. Toxicol. Endocrinol.* **86,** 17–22.
50. Durst, F. & O'Keefe, D. P. (1995) *Drug Metabol. Drug Interact.* **12,** 171–187.
51. Chang, A. & Hartmann, T. (1998) *Phytochemistry* **49,** 1859–1866.