DNA AND CELL BIOLOGY
Volume 14, Number 1, 1995
Mary Ann Liebert, Inc., Publishers
Pp. 61–71

Gene Evolution of Epoxide Hydrolases and Recommended Nomenclature

JEFFREY K. BEETHAM,[1,2] DAVID GRANT,[1] MICHAEL ARAND,[3] JOAN GARBARINO,[4] TOMOHIRO KIYOSUE,[5] FRANCK PINOT,[1] FRANZ OESCH,[3] WILLIAM R. BELKNAP,[4] KAZUO SHINOZAKI,[5] and BRUCE D. HAMMOCK[1]

ABSTRACT

We have analyzed amino acid sequence relationships among soluble and microsomal epoxide hydrolases, haloacid dehalogenases, and a haloalkane dehalogenase. The amino-terminal residues (1–229) of mammalian soluble epoxide hydrolase are homologous to a haloacid dehalogenase. The carboxy-terminal residues (230–554) of mammalian soluble epoxide hydrolase are homologous to haloalkane dehalogenase, to plant soluble epoxide hydrolase, and to microsomal epoxide hydrolase. The shared identity between the haloacid and haloalkane dehalogenases does not indicate relatedness between these two types of dehalogenases. The amino-terminal and carboxy-terminal homologies of mammalian soluble epoxide hydrolase to the respective dehalogenases suggests that this epoxide hydrolase, but not the soluble epoxide hydrolase of plant or the microsomal epoxide hydrolase, derives from a gene fusion. The homology of microsomal to soluble epoxide hydrolase suggests they derive from a gene duplication, probably of an ancestral bacterial (epoxide) hydrolase gene. Based on homology to haloalkane dehalogenase, the catalytic residues for the soluble and microsomal epoxide hydrolases are predicted. A nomenclature system based on divergent molecular evolution is proposed for these epoxide hydrolases.

INTRODUCTION

HYDROLYSIS IS A CENTRAL PROCESS of many metabolic pathways. Hydrolytic enzymes catalyze this process and include many of the lipases, esterases, proteases, epoxide hydrolases, and dehalogenases. These enzymes function in metabolizing both endogenous and xenobiotic-derived compounds.

The epoxide hydrolases (EH, E.C.3.3.2.3) add water to epoxides to form the corresponding diol. The epoxide hydrolases are grouped by characteristics of enzymatic activity and biochemical separation. In mammals, these groups include the soluble epoxide hydrolases (sEH, also referred to as cytosolic EH), microsomal epoxide hydrolases (mEH), leukotriene A_4 hydrolase (LTA_4H), cholesterol epoxide hydrolase (Watabe *et al.*, 1981), and hepoxilin hydrolase (Pace-Asciak and Lee, 1989). The LTA_4H functions to produce hormone derivatives of arachidonic acid (Samuelsson, 1983), as may the sEH; mammalian sEHs hydrolyze epoxides of arachidonic acids with regio and enantiomeric selectivity (Oliw *et al.*, 1982; Chacos *et al.*, 1983; Zeldin *et al.*, 1993), forming and/or degrading oxylipins that may have biological roles. The sEH is also implicated in decreasing deleterious effects of oxidative stress due to endogenous and xenobiotic-derived compounds (El-Tantawy and Hammock, 1980; Kramer *et al.*, 1991). A principal function of the mEH is in detoxifying xenobiotics (Oesch, 1973).

Epoxide hydrolases are less understood in nonmammalian species. As discussed below, several EHs have been described from plants where they may have some of the same roles attributed to mammalian EHs in addition to having involvement in production of cutin (complex polymers de-

Departments of Entomology and Environmental Toxicology[1] and Biochemistry and Biophysics,[2] 303 Briggs Hall, University of California, Davis CA 95616.
[3]Institute of Toxicology, University of Mainz, Mainz, Germany.
[4]United States Department of Agriculture, Agricultural Research Service, Western Regional Research Center, Albany, CA.
[5]Laboratory of Plant Molecular Biology. The Institute of Physical and Chemical Research (RIKEN). Tsukuba Life Science Center. Tsukuba, Japan.

rived largely from hydroxylated fatty acids and found on the external surfaces of many plants) and phytoalexins (defensive compounds of diverse structures that are induced by exposure to foreign compounds, organisms, and/or by damage) (Blee and Schuber, 1992; Kiyosue *et al.*, 1994; Stapleton *et al.*, 1994). In insects, an EH with activity against octane oxide has been purified (Mullin and Wilkinson, 1980) and appears distinct from other insect EHs that degrade certain cyclodiene insecticides and insect juvenile hormone. The insect EHs that degrade the juvenile hormones may be important regulators of insect development and reproduction (Hammock, 1985; Halarnkar and Schooley, 1990; Harshman *et al.*, 1991).

We have been investigating the biological functions of the EHs, especially of sEH. Most of our present understanding of the sEHs derives from biochemical and physiological analysis. An evolutionary analysis of the sEHs based on molecular aspects may provide a framework within which biochemical and physiologic data of analogous and/or homologous EH activities from varied species may be compared. The recent isolation of cDNAs encoding sEHs of mammals and plants permits a molecular evolutionary analysis of these enzymes.

The first cDNAs encoding sEHs of mammals (human, mouse, and rat) were reported in 1993 (Table 1). More recently, the cDNAs for two plant sEHs (*Solanum tuberosum* and *Arabidopsis thaliana*) have been isolated (Table 1). While the recombinant sEH of plants has been shown to have mammalian-sEH like activities, the plant enzyme is only two-thirds the size of mammalian sEH and has relatively low overall identity. The evolutionary relatedness of these enzymes has not been established.

This paper describes the results of comparing mammalian sEH sequences to other EHs as well as to sequences on file with nucleic acid and protein databases. We propose that the mammalian sEH is composed of two evolutionarily distinct regions. The amino-terminal region is homologous to bacterial haloacid dehalogenase (HAD1; Table 1). The carboxy-terminal region is homologous to sEH of plant, mEH, and bacterial haloalkane dehalogenase (HLD1; Table 1).

METHODS

Protein sequences were analyzed with FASTA, PILEUP, GAP, and BESTFIT programs on the University of Wisconsin Genetic Computer Group (GCG) program version 7.0 (Devereux, 1989; Devereux *et al.*, 1984), and with BLAST (Altschul *et al.*, 1990) utilizing the BLOSUM62 matrix (Henikoff and Henikoff, 1992) on the National Center for Biotechnology Information (NCBI) e-mail server (blast@ncbi.nlm.nih.gov). Protein sequences were screened against database sequence libraries utilizing FASTA and BLAST. Statistical significance of FASTA search products was evaluated using RDF2 (Pearson and Lipman, 1988) on the FLAT e-mail server at Gunma University, Japan (flatnetserv@smlab.eg.gunma-u.ac.jp). Sequence pairs were aligned using GAP and BESTFIT with a gap weight of 3.0 and length weight of 0.10. Simultaneous alignment of greater than two sequences was by PILEUP.

RESULTS

Comparison of human sEH (hsEH) to all sEHs with FASTA yielded optimized scores, all in excess of 400 (Table 2), and FASTA alignments over essentially all (>95%) of the target sequences. The corresponding BLAST scores for the sEHs were greater than 140 with probabilities of less than $2.1e^{-26}$. The percent identity shared between the sEHs varies from 30 to 92% (Table 2). Figure 1 portrays an

TABLE 1. DESCRIPTION OF SEQUENCES USED

Enzyme name	*Organism*	*Abrv.*	*Acc. #*	*Reference*
Soluble epoxide hydrolase	Human	hsEH	L05779	(Beetham *et al.*, 1993)
Soluble epoxide hydrolase	Mouse	msEH	L05781	(Grant *et al.*, 1993)
Soluble epoxide hydrolase	Rat	rsEH	X65083	(Knehr *et al.*, 1993)
Soluble epoxide hydrolase	Potato (*Solanum tuberosum*)	psEH	U02497	(Stapleton *et al.*, 1994)
Soluble epoxide hydrolase	Mouse-eared cress (*Arabidopsis thaliana*)	asEH	D16628	(Kiyosue *et al.*, 1994)
Microsomal epoxide hydrolase	Human	hmEH	J03518	(Skoda *et al.*, 1988)
Microsomal epoxide hydrolase	Rat	rmEH	A26732	(Falany *et al.*, 1987)
Microsomal epoxide hydrolase	Rabbit	bmEH	S04342	(Hassett *et al.*, 1989)
Leukotriene A4 hydrolase	Human	LTA_4H	J02959	(Funk *et al.*, 1987)
Haloacid dehalogenase	*Pseudomonas* sp.	HAD1	P24069	(Schneider *et al.*, 1991)
Haloacid dehalogenase	*Xanthobacter autotrophicus*	HAD2	M81691	(Ploeg *et al.*, 1991)
Haloacid dehalogenase	Bacterium	HAD3	JQ0932	(Not published)
Haloacid dehalogenase	*Pseudomonas* sp.	HAD4	P24070	(Schneider *et al.*, 1991)
Haloacid dehalogenase	*Pseudomonas cepacia*	HAD5	S29096	(Murdiyatmo *et al.*, 1992)
Haloacid dehalogenase	*Pseudomonas putida*	HAD6	A44830	(Jones *et al.*, 1992)
Haloalkane dehalogenase	*Xanthobacter autotrophicus*	HLD1	P22643	(Janssen *et al.*, 1989)

Accession number (Acc. #) corresponds to the unique DNA sequence identifier used by the sequence database. The deduced protein sequences of these DNA sequences were use for analysis of homology. (ABRV. = abbreviation)

TABLE 2. MATRIX OF PERCENT IDENTITIES AND FASTA SCORES

	PERCENT IDENTITY										
	soluble EH					microsomal EH					
	1	2	3	4	5	6	7	8	9	10	11
(length)	hsEH	rsEH	msEH	asEH	psEH	rmEH	hmEH	bmEH	HAD1	HLD1	LTA_4H
1 (554)		73	73	37	30	22	21	23	24	24	20
2 (554)	2276		92	34	30	19	22	21	23	25	16
3 (554)	2282	2674		34	30	22	20	21	20	21	16
4 (323)	494	453	434		57	26	25	22	17	24	22
5 (321)	445	443	422	1059		22	24	19	15	19	22
6 (455)	95	72	86	97	92		84	83	18	20	19
7 (455)	78	79	41	85	73	2183		83	20	22	19
8 (455)	82	79	81	88	72	2165	2166		22	23	19
9 (227)	103	103	105	36	36	38	33	32		24	19
10 (310)	201	218	190	168	126	96	86	87	31		19
11 (611)	38	30	46	39	50	60	46	61	43	48	
	OPTIMIZED FASTA SCORE										

Percent identities were determined by comparing protein sequence pairs utilizing GAP. Identities were also equivalent to values calculated by BESTFIT.

alignment of all sEHs. These data clearly indicate that the sEHs are homologous, but that the sEHs of plant lack the amino-terminal region of mammalian sEH.

Utilizing the FASTA and BLAST programs to query the sequence databases with the amino or carboxy terminus of hsEH, and amino-terminal homology to haloacid dehalogenase HAD1 and a carboxy-terminal homology to haloalkane dehalogenase HLD1 were indicated. The search scores, corresponding probabilities, and residues aligned are shown in Table 3. The FASTA probability factors for both amino- and carboxy-terminal identities between bsEH, and the dehalogenases were greater than 10 standard deviation units, strongly indicating evolutionary relatedness; factors greater than 3 are considered as possibly significant, greater than 6 as probably significant, and greater than 10 as significant (Lipman and Pearson, 1985).

The percent identities and FASTA scores of HAD1 and HLD1 compared to all sEHs are shown in Table 2. The FASTA scores for HAD1 compared to mammalian sEH are all greater than 100, but compared to plant sEH they are less than 40; these scores indicate significant identity of HAD1 to mammalian sEH but not to the plant enzymes. The FASTA scores for HLD1 compared to all sEHs are greater than 126. Based on these identities, alignments of the human sEH with the HAD1 and HLD1 are shown (Fig. 2). Three catalytic residues of HLD are conserved in all sEHs based on this alignment ("*" residues, Fig. 2b). The FASTA score of 31 for comparison of HLD1 to HAD1 (Table 2) does not indicate a significant similarity.

By sequence analysis the mEHs are homologous to the sEHs and HLD1. The percent shared identities and FASTA scores are shown in Table 2. Although the percent identities of mEH's to sEH's are low (19–25%), the local similarities determined in the FASTA analysis are on the cusp between possibly and probably significant (Table 3). Further support for homology is that in aligning the mEHs with sEHs and HLD1, the three catalytic residues of HLD1 are conserved, but the acidic Asp of HLD1 is replaced with Glu in mEH (Fig. 3a). It should be noted that a partial sequence identity between the mEH and HLD1 over 78 residues was noted on cloning the HLD1 (Janssen *et al.*, 1989).

The overlap of the linear sequences of these EHs and dehalogenases is shown (Fig. 4a). The evolutionary relationships of these enzymes were deduced by analyzing the identities within the multiple alignment (Fig. 3, difference matrix) and within the paired alignment of hsEH with HAD1 (Fig. 2a). A cladogram depicting these relationships is shown (Fig. 4b).

DISCUSSION

The overall sequence identities of the sEHs are quite varied, from 30% to 92% (Table 2). However, analysis with BLAST and FASTA strongly supports the hypothesis that these enzymes are homologous. Biochemical characteristics of the recombinant enzymes, such as similar activities and inhibition by sEH-selective compounds, provide biological support of common ancestry (see sEH references in Table 1). Similarly, the homology of the amino terminus of mammlian sEH to HAD1 and the carboxy terminus of mammalian sEH to plant sEH and microsomal EH and HLD1 is supported by sequence identity, conservation of catalytic residues, and similar enzymatic activity (hydrolysis of substrates without cofactors).

The sequence data do not support the existence of an evolutionary relationship between LTA_4H and the sEHs, mEHs, or dehalogenases (Table 2). Furthermore, the enzymatic activity of LTA_4H requires a metal cofactor whereas the sEHs, mEHs, and dehalogenases described here do not. These data suggest that at least two types of unrelated epoxide hydrolases have evolved from separate ancestral genes.

HAD1 and HLD1 share no significant sequence identity that would indicate homology between these dehalogenases (Table 2). The orthologs of these two enzymes can be found together in bacteria such as *Xanthobacter autotrophicus* allowing the dehalogenation of 1,2-dihalogenated alkanes (Ploeg *et al.*, 1991). This observation suggests the mammalian sEH is the result of a fusion between the genes ancestral to HAD1 and HLD1. With this assumption, a scheme of

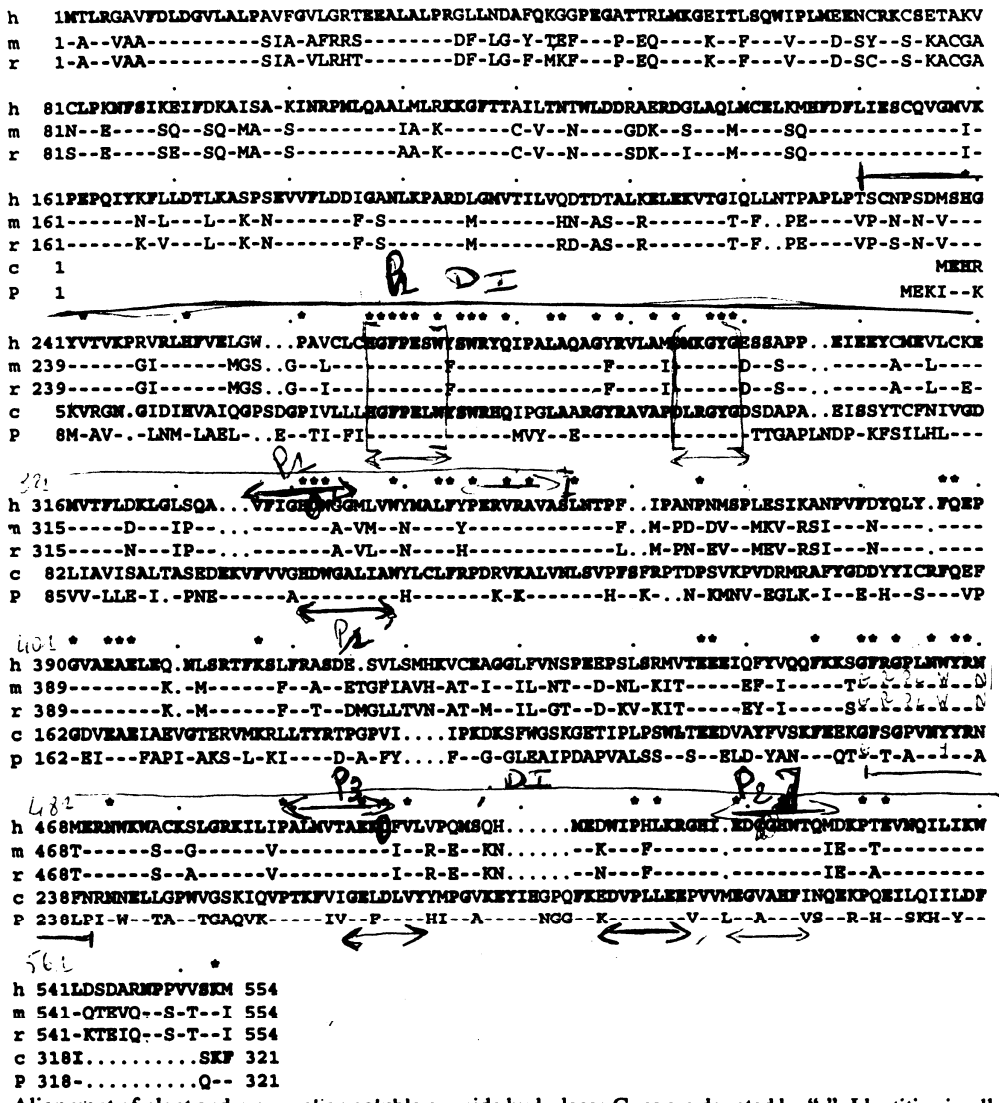

FIG. 1. Alignment of plant and mammalian soluble epoxide hydrolases Gaps are denoted by ".". Identities in all plant and mammalian sEHs are indicated by "*". Bold residues of hsEH are conserved in all mammalian sEHs (also indicated as dashes "-" in msEH and rsEH sequences). Bold residues of asEH are conserved in psEH (also indicated as dashes "-" in psEH). h, hsEH; m, msEH; r, rsEH; c, asEH; p, psEH.

relatedness can be established for the sEHs, mEHs, haloalkane dehalogenases, and haloacid dehalogenases (Fig. 4b).

Amino-terminal analysis

The homology between the amino terminus of mammalian sEH and HAD1 is intriguing and may indicate that mammalian sEH contains two catalytic domains. As noted earlier, although the plant lacks the mammalian sEH amino-terminal region, the mammalian and plant sEHs share characteristics of activities and inhibition. What (if any) is the function of the amino terminus of mammalian sEH? It has been suggested without experimental support that the mammalian amino-terminal region functions to allow dimerization of the sEH (Knehr *et al.*, 1993); the native mammalian sEHs studied exist as homodimers. The homology to HAD1 suggests that at least a part of the amino-terminal region may possess a function or catalytic activity analogous to the corresponding region of HAD1.

Although no catalytic mechanism has yet been established for the haloacid dehalogenases, the subclass of dehalogenases to which HAD1 belongs is characterized in part by an insensitivity to thiol-modifying reagents. Furthermore, site-directed mutation of HAD1 and the closely related HAD5 (both are members of the same subclass of haloacid dehalogenases that convert L-2-chloropropionate to D-lactate, but do not convert D-2-chloropropionate) identified residues critical to activity. Substitution in HAD1 of Asp10 to Ala10 resulted in complete loss of activity (Schneider *et al.*, 1993);

TABLE 3. STATISTICAL SIGNIFICANCE (*P* AND *z* VALUES) OF PROTEIN SIMILARITY SCORES

		Blast results				*FASTA results.* *Residues aligned*			
Query	*Target*	*hsEH residue*	*Segment identity*	*Score*	P	*Query*	*Target*	*Segment identity*	Z *(SD)*
hsEH	HAD1	148–204	30%	79	0.01	112–229	107–219	28%	10.5
		112–124	61%	45	0.7				
hsEH	HLD1	264–358	31%	146	$3.9e^{-12}$	246–402	32–193	25%	21.4
hsEH	rmEH					211–382	99–274	23%	6.0
HLD1	rmEH					15–144	110–246	23%	5.8

Protein and DNA databases were searched using the BLAST program with query sequences consisting of hsEH residues 1–264 or 252–554, and using the FASTA program with query sequences consisting of hsEH residues 1–230 or 190–554. The BLAST results include what appears to be an insignificant alignment ($P = 0.7$). Unlike FASTA, BLAST does not allow gaps in constructing an alignment but will report low scoring sequence pairs when a comparison also results in the report of a proximal and significantly high scoring sequence pair. The highest scoring dehalogenase target sequence for each FASTA search, and the sequence of rmEH, were compared against the appropriate sEH query segment for statistical significance (Z) utilizing FASTA RDF-2. The FASTA RDF2 comparison of HLD1 to HAD1 and rmEH utilized the entire sequence of the proteins. The λ and κ values for the BLAST searches were 0.322 nats/unit and 0.137 for the carboxy-terminal search, and 0.322 nats/unit and 0.139 for the amino-terminal search. *P*, Probability; SD, standard deviation units.

Asp-10 is conserved in HAD-1 through HAD6 and in mammalian sEH as well. Similarly, residues Arg-42 and His-20 are critical to the activity of HAD5 (Murdiyatmo *et al.*, 1992). Arg-42 is conserved in the HADs we examined (Arg-41 in HAD1; Fig. 2a) and in the mammalian sEHs corresponds to a conserved Lys. However, His-20 is not conserved in all HADs examined (Asn-19 in HAD1, and Gln in HAD2) or in sEH (note though the retention of charge in HAD1 and HAD2). The region of greatest identity between the HADs and sEHs corresponds to hsEH residues 112–229, but no function has yet been ascribed to the corresponding region of the HAD's.

Carboxy-terminal analysis

HLD1 is a member of the α/β hydrolase fold enzyme family, a group of evolutionarily related enzymes that include other hydrolases such as acetylcholine esterase, lipases, and carboxypeptidase II (Ollis *et al.*, 1992). The α/β fold enzymes have a two-domain organization. Domain I consists of an α/β sheet that forms a catalytic pocket containing the nucleophile–acid–His catalytic triad and is structurally continuous. However, in terms of linear protein sequence, Domain I is split by Domain II. Domain II in HLD1 sits like a lid over the catalytic cleft (Franken *et al.*, 1991) positioning residues critical to substrate binding and specificity (Ollis *et al.*, 1992). Much of the structural variance between α/β enzymes occurs in Domain II and is evidenced by the varied lengths of Domain II sequences (Ollis *et al.*, 1992).

Domain I of HLD1 spans residues 20–155 and 230–310, and Domain II spans 156–229 (Franken *et al.*, 1991). By the alignment of hsEH and HLD1 (Fig. 2b), the equivalent residues of hsEH Domain I are 230–365 and 473–545, and for Domain II are 366–472 (Fig. 2b, "↓"). Alignment of all sEH with all mEH and HLD1 indicates the deduced Domain I of mEH spans residues 112–245 and 374–452, Domain II spans residues 256–373 (shown in part in Fig. 3a).

Establishing the deduced domains of the sEH and mEH may be important both for predicting catalytic residues and noncatalytic residues that interact with substrate. For example, if residues critical to substrate binding are to be found in Domain II of the sEHs as they are in other α/β enzymes, a conserved basic residue might function in stabilizing or positioning the acid moiety of acidic substrates. Longer-chain fatty acids are excellent substrates for the mammalian sEHs. There are three basic residues conserved in all sEHs that could serve this stabilizing function: Arg-466, Lys-406, and Lys-375 (which is Arg in some sEHs). Preliminary experiments in one author's laboratory (Hammock) indicate that chemical modification of Lys-406 inactivates the enzyme.

The homology of the carboxyl terminus of sEH to HLD1, especially the shared identity with the catalytic residues of HLD1, is important for predicting a catalytic mechanism for the sEH. The mechanism for the HLD1 was established by X-ray crystallographic analysis of a HLD1-substrate intermediate (Verschueren *et al.*, 1993). The HLD1 mechanism involves the nucleophilic attack of Asp-123 on the halogen-substituted terminal carbon of the substrate, forming a covalently bound ester intermediate. The His-289–Asp-260 pair activate a water molecule that hydrolyzes the ester intermediate to release the product. By analogy, the alignment of hsEH with HLD1 suggests that hsEH Asp-334 is the nucleophile that attacks the epoxide-carbon of the substrate, and Asp-495–His-523 are the water-activating pair. This mechanism has recently been suggested for the mammalian sEH and mEH (Arand *et al.*, 1994; Lacourciere and Armstrong, 1994) and recent data provide experimental support for this hypothesis (Lacourciere and Armstrong, 1993; Hammock *et al.*, 1994).

The multiple alignments between plant and animal sEH, mEH, and HLD1 presented here support a predicted mechanism for the EHs analogous to that of HLD1. The alignment suggests that the catalytic residues of mEH are the nucleophile Asp-226 and the water-activating pair Glu-404–His-431 (Fig. 3a). While the assignment of acidic water-activa-

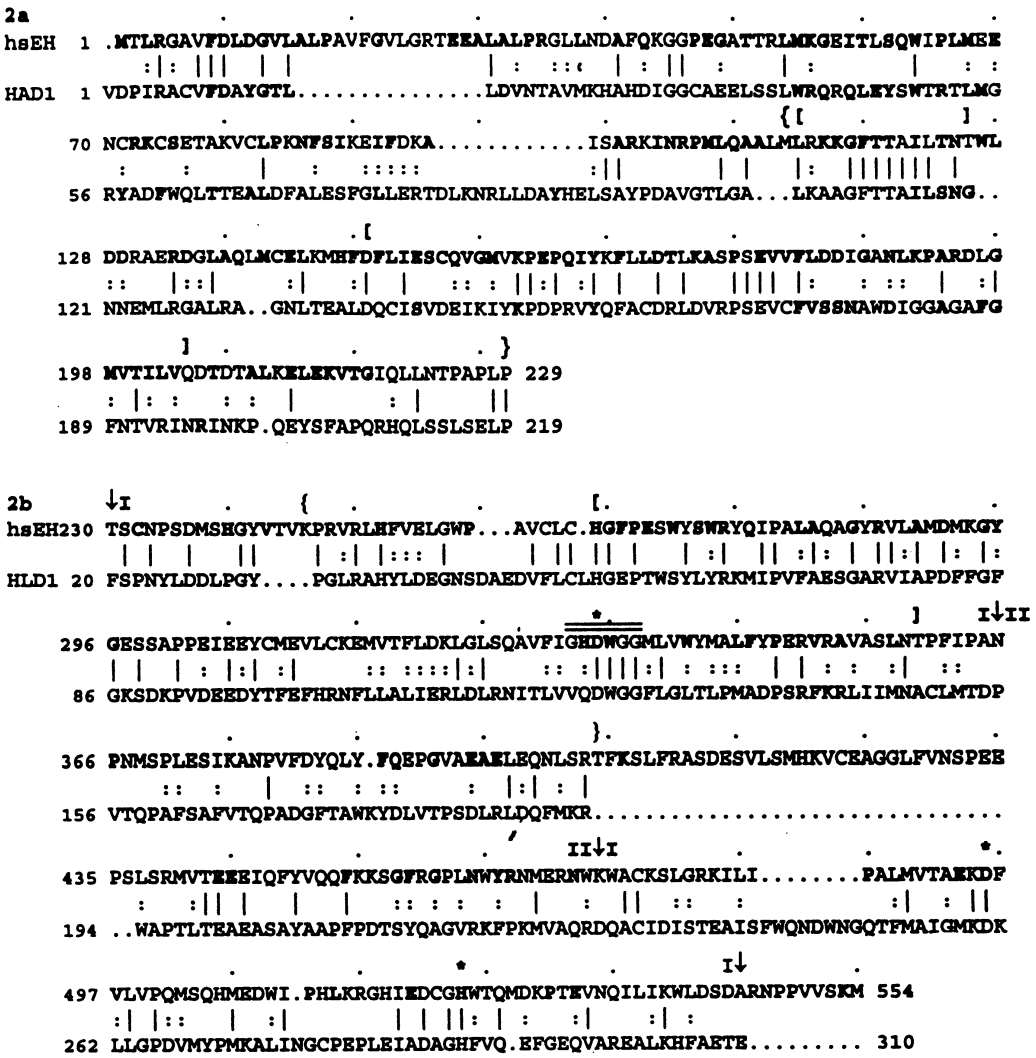

FIG. 2. Alignment of hsEH with two dehalogenases Regions aligned by FASTA or BLASTP analysis are bounded by "{}" and "[]", respectively. The alignment shown of regions outside these boundaries may not be valid. Identities and similarities are denoted by "|" and ":". Alignment utilized GAP and is equivalent to that produced with BESTFIT. Gaps are denoted by ".". a. Alignment of hsEH amino-terminal region to HAD1. Bold residues of hsEH are conserved with sEH of rat and mouse. Bold residues in HAD1 are conserved with five other haloacid dehalogenases (HAD2-6). b. Alignment of hsEH carboxy-terminal region to HLD1. Bold residues of hsEH are conserved with sEH of rat, mouse, arabidopsis, and potato. The catalytic residues of HLD1 are denoted by "*". Domain I and II boundaries of HLD1 are denoted by "↓" and "I" or "II". The consensus nucleophilic region of HLD1 (small,X,nucleophile,X,small,small) is denoted by =.

tion to Asp-352 was previously suggested by ourselves and others (Arand *et al.*, 1994; Lacourciere and Armstrong, 1994), Asp-352 does not show the high conservation observed with Glu-404. In addition, the positions of Asp-352 and Glu-404 relative to His-431 support assignment of water-activation function to the Glu-404 residue. In HLD1 and in a closely related α/β enzyme (dienelactone hydrolase; Ollis *et al.*, 1992), the separation of the catalytic acid-His pair is 28–30 residues. A similar spacing is observed in the five sEHs (27–34 residues) and three mEHs (26 residues). This contrasts with an Asp-352–His-431 spacing of 78 residues.

Finally, the domain organization of the α/β enzymes also supports catalytic assignment to Glu-404 rather than Asp-352. In superimposing the domains of HLD1 upon the sEH and mEH, Glu-404 is in Domain I while Asp-352 is in Domain II; all catalytic residues of α/β enzyme family members are within Domain I. We can attempt to reposition Asp-352 by extending mEH Domain I to include this residue. In so doing, the total residues of Domain I increase from 217 to 239. This length is as compared to a Domain I length of 209 in mammalian sEH, 213 in plant sEH, and 217 in HLD1. However, the major difference between sequence lengths of α/β enzymes is due to variance of lengths in Domain II, not

```
          ii i     .     ^  .        .          i^i       .        .        .
rmEH 148 HGWPGSFYEFYKIIPLLTDPKSHGLSDEHVFEVICPSIPGYGYSEASSKKGLNSVATARI...FYKLMTR
hmEH 148 HGWPGSFYEFYKIIPLLTDPKNHGLSDEHVFEVICPSIPGYGFSEASSKKGFNSVATARI...FYKLMLR
bmEH 148 HGWPGSFFEFYKIIPLLTDPKSHGLSDEHIFEVICPSIPGYGFSQASSKKGFNSVSTARI...FYKLMLR
rsEH 263 HGFPESWFSWRYQIPALAQA.........GFRVLAIDMKGYGDSSSPP..EIEEYAMELLCEEMVTFLNK
msEH 262 HGFPESWFSWRYQIPALAQA.........GFRVLAIDMKGYGDSSSPP..EIEEYAMELLCKEMVTFLDK
hsEH 264 HGFPESWYSWRYQIPALAQA.........GYRVLAMDMKGYGESSAPP..EIEEYCMEVLCKEMVTFLDK
asEH  30 HGFPELWYSWRHQIPGLAAR.........GYRAVAPDLRGYGDSDAPA..EISSYTCFNIVGDLIAVISA
psEH  31 HGFPELWYSWRHQMVYLAER.........GYRAVAPDLRGYGDTTGAPLNDPSKFSILHLVGDVVALLEA
HLD1  54 HGEPTWSYLYRKMIPVFAES.........GARVIAPDFFGFGKSDKPV..DEEDYTFEFHRNFLLALIER

                    .    ii    .        i. ^     ^  .                        ^  .
rmEH 215 LGFQK...FYIQGGDWGSLICTNMAQMVPNHVKGL.HLNMAFISR...255//374 YKENLGQGIMVHK
hmEH 215 LGFQE...FYIQGGDWGSLICTNMAQLVPSHVKGL.HLNMALVLS...255//374 YKENLGQGWMTQK
bmEH 215 LGFQE...FYIQGGDWGALVCTNMAQLVPSHVKGL.HLNMALILR...255//374 YKENLGQGFMAHK
rsEH 322 LGIPQA...VFIGHDWAGVLVWNMALFHPERVRAVASLNTPL..MPPN364//473 KWSCKALGRKIL.
msEH 321 LGIPQA...VFIGHDWAGVMVWNMALFYPERVRAVASLNTPF..MPPD363//472 KWSCKGLGRKIL.
hsEH 323 LGLSQA...VFIGHDWGGMLVWYMALFYPERVRAVASLNTPF..IPAN365//473 KWACKSLGRKIL.
asEH  89 LTASEDEKVFVVGHDWGALIAWYLCLFRPDRVKALVNLSVPFSFRPTD136//243 ELLGPWVGSKIQ.
psEH  92 I.APNEEKVFVVAHDWGALIAWHLCLFRPDKVKALVNLSVHFSKR..N136//243 ELTAPWTGAQVK.
HLD1 113 LDLRN...ITLVVQDWGGFLGLTLPMADPSRFKRLIIMNACLMTDPVT157//230 DQACIDISTEAIS

                 . ^      .s      .        . ^ ^    .     sis .  ^     .   s
rmEH 387 HEGMKVFVPTGFSAFPSELLHAPEKW.....VKVVKYPKLISYSYMERGGHFAAFEEPKLLAQDIRKFVSLA452
hmEH 387 HERMKVYVPTGFSAFPFELLHTPEKW.....VRFFKYPKLISYSYMVRGGHFAAFEEPELLAQDIRKFLSVL452
bmEH 387 HERLKVHVPTGFAAFPCEIMHVPEKW.....VRTTKYPQLISYSYMPRGGHFAAFEEPELLARDICKFVGLV452
rsEH 485 .......VPALMVTAEKDIVLRPEMSKN.....MENNWIPFLKRGHI.EDCGHWTQIEKPAEVNQILIKWLKTE544
msEH 484 .......VPALMVTAEKDIVLRPEMSKN.....MEKKWIPFLKRGHI.EDCGHWTQIEKPTEVNQILIKWLQTE543
hsEH 485 .......IPALMVTAEKDFVLVPQMSQH.....MEDDWIPHLKRGHI.EDCGHWTQMDKPTEVNQILIKWLDSD544
asEH 255 .......VPTKFVIGELDLVYYMPGVKEYIHGPQFKEEDVPLLEEPVVMEGVAHFINQEKPQEILQIILDFI...318
psEH 255 .......VPTKFIVGEFDLVYHIPGAKEYIHNGGFKKKDVPLLEEVVVLEGAAHFVSQERPHEISKHIYDFI...318
HLD1 243 FWQNDWNGQTFMAIGMKDKLLGPDVM.....YPMKAALINGCPEPLEIADAGHFVQEFGEQVAREALKHFAETE310
```

a

NUMBER OF DIFFERENCES

		1 rmEH	2 hmEH	3 bmEH	4 rsEH	5 msEH	6 hsEH	7 asEH	8 psEH	9 HLD1
1	rmEH		20	29	123	121	125	116	123	132
2	hmEH	0.123		24	122	122	126	118	127	131
3	bmEH	0.178	0.147		123	123	128	122	128	131
4	rsEH	0.755	0.748	0.755		8	30	103	114	120
5	msEH	0.742	0.748	0.755	0.049		27	102	111	122
6	hsEH	0.767	0.773	0.785	0.184	0.166		99	114	119
7	asEH	0.712	0.724	0.748	0.632	0.626	0.607		55	119
8	psEH	0.759	0.784	0.790	0.704	0.685	0.704	0.340		125
9	HLD1	0.810	0.804	0.804	0.736	0.748	0.730	0.730	0.772	

b DIFFERENCE RATIO

FIG. 3. Multiple alignment and difference matrix of soluble and microsomal epoxide hydrolases and HLD1. a. Complete sequences were aligned by PILEUP and the region corresponding to HLD1 Domain I (excepting the 34 most amino-terminal residues) is shown. As noted in text, the linear sequence of Domain I is interrupted by Domain II; this is indicated by "//". Gaps are denoted by ".". Identities among all sequences are indicated by "i", among all epoxide hydrolases but not HLD1 by "^". Conserved similar residues (D = E; K = R = H; Q = N; F = W) are denoted by "s". Catalytic (and deduced catalytic) residues are in large, bold print. Positions in the alignment at which there were gaps in any of the sequences were omitted in calculating the difference matrix and cladogram. b: Difference matrix from the alignment of a. Residues not considered in the matrix are as noted in a.

domain I. A mEH Domain I length of 217 residues is closer to that of the sEH's and HLD1 than is a length of 239.

Another region showing high conservation of residues between mEH, sEH, and HLD1 is within Domain I at hsEH residues 264–277. In HLD1 and other α/β enzymes this structurally conserved region is putatively involved in forming an oxyanion hole that stabilizes an enzyme–substrate tetrahedral intermediate (Ollis *et al.*, 1992; Verschueren *et al.*, 1993). In the sEHs and mEHs, it may function similarly as well as functioning to polarize the epoxide in a push pull mechanism of initial attack of Asp-334. In this model, the helix dipole occurring within the conserved region (hsEH residues 264–277) is positioned near the catalytic pocket, stabilizing a tetrahedral anion-intermediate consisting of the enzyme covalently linked to an oxyanion substrate intermediate. Such an oxyanion is a possible early

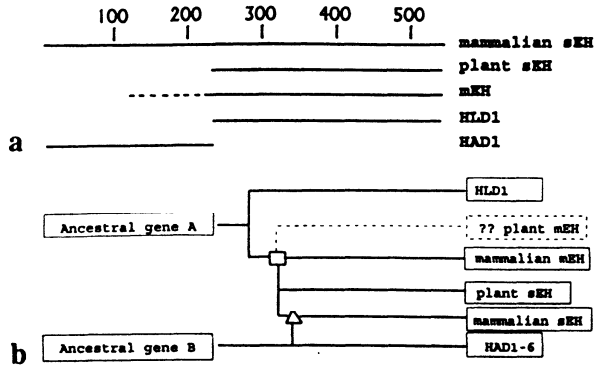

FIG. 4. Relatedness of the soluble and microsomal epoxide hydrolases and the dehalogenases. a. The linear sequence of mammalian sEH (~554 residues) is aligned with other epoxide hydrolases and dehalogenases. Solid lines represent regions of proposed homologies between the proteins. The dashed line for the amino terminus of microsomal EH indicates there is no homology of this region to the other proteins represented. b. Cladogram depicting one possible scheme of the evolutionary relatedness of these enzymes. The open triangle indicates a gene fusion, the open square a gene duplication. The dashed line and box denoting the plant mEH represents our speculation that the ortholog of the mammalian mEH will be found in plant, a discovery that would further support the placement of the duplication event prior to the plant and mammal sEH split. Indeed, a plant-like enzyme with mammalian mEH-like biochemical properties has been reported (Pinot *et al.*, 1992).

intermediate formed as the enzyme reacts with the substrate (Fig. 5). Additionally, the conserved region may form a hydrogen bond to the ester carbonyl of the tetrahedral intermediate, as is indicated to occur with HLD1 (Verschueren *et al.*, 1993).

NOMENCLATURE

We have proposed evolutionary relationships between soluble epoxide hydrolases of mammals and plants, microsomal epoxide hydrolase of mammals, and two dehalogenases. Given the distribution of sEH across the species from plant to mammal, it seems likely that sEH will be found in many or most eukaryotes, and perhaps prokaryotes as well. The homologs of mammalian microsomal EH are also likely to be found in many nonmammalian species. There have been scattered reports of EH activities in other species but for which protein sequences have not been established. These include the cholesterol EH (Watabe *et al.*, 1981), hypoxilin hydrolase (Pace-Asciak and Lee, 1989), plant EHs such as are found in soybean (Blee and Schuber, 1992), fungal EHs (Chen *et al.*, 1993), and insect EHs (Hammock, 1985; Halarnkar and Schooley, 1990; Harshman *et al.*, 1991). What will be their mechanistic and evolutionary relatedness to those EHs already sequenced?

The large number of expected epoxide hydrolases suggests that a nomenclature system encompassing all epoxide hydrolases that are evolutionarily related would facilitate investigations of these enzymes. One has only to look at the examples of other enzyme families such as the cytochromes P450 (Nelson *et al.*, 1993), the glutathione transferases (Buetler and Eaton, 1992; Mannervik *et al.*, 1992), or the UDP glucuronosyltransferases (Burchell *et al.*, 1991) to see the utility of a molecular-based nomenclature system to inter- and intra-species comparisons of analogous or homologous enzymes. The early establishment of a system would clearly decrease the difficulty of implementation. In the case of the cytochromes P450, there were fewer than one dozen cDNA sequences established when a nomenclature committee was formed to recommend a molecular-based classification system (Nebert *et al.*, 1991). During the 2 years from committee formation to publication of the system, the number of P450 sequences established had increased to 65 (Nebert *et al.*, 1987). We fell that present number of EHs whos sequences

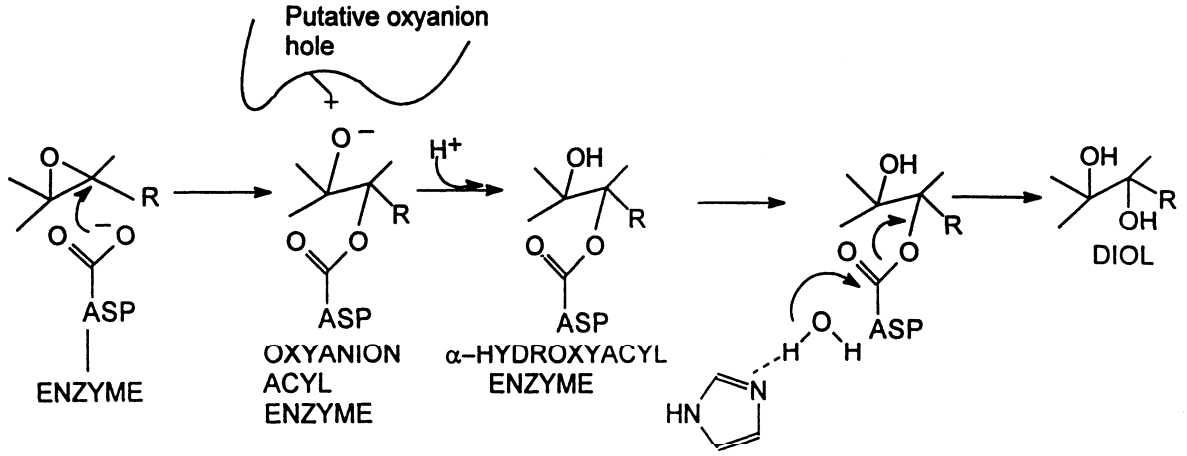

FIG. 5. Putative oxyanion intermediate. Possible catalytic scheme of the soluble and microsomal epoxide hydrolases depicting an oxyanion intermediate that may be stabilized by an oxyanion hole formed by the enzyme.

are described (eight soluble and microsomal EHs from six eukaryote species) is adequate to establish the framework of an open-ended nomenclature system for these enzymes.

Historically, the EHs have been named by species of origin, substrate, and/or characteristics of fractionation. Although these characteristics are important to understanding enzyme function, they cannot be used to directly infer evolutionary relatedness between enzymes and may be ambiguous when used for nomenclature. Substrates may vary for orthologous proteins. Characteristics of fractionation may also vary; the sEH of mammals was named for its predominant fractionation to the 100,000 × *g* supernatant, but the sEH of rodent is in peroxisomal, cytosolic, and (in trace amounts) microsomal fractions. A molecular classification system for epoxide hydrolases would be less arbitrary than the historical *ad hoc* naming system, and may be much more consistent.

What we propose is a superfamily designated by the root name "HYL" and consisting of all enzymes that are evolutionarily related to the α/β hydrolase fold family. Division of the superfamily into families, subfamilies and genes would be by protein sequence identities. The families would be signified by an arabic number; any member of one family would have less than 40% identity to any member of another family. The HYL group would contain many families, but with respect to the EHs there would be three families: those containing the mEH, the mammalian sEH, and the plant sEH. Each family would divide into subfamilies signified by a lowercase letter; any member of one subfamily would have greater than 55% identity to members of the same subfamily. The gene number would be designated by an arabic number (preceded by a dash for mouse to bring this nomenclature in line with accepted mouse nomenclature).

We suggest the gene, transcript, cDNA, and enzyme all utilize the same name. The gene and its cDNA would be italicized, *e.g.*, *HYL2a1* (*Hyl2a-1* for mouse). The mRNA or protein would be written without italics, *e.g.*, HYL2a1 for mouse or any other species.

If there is no evidence of multiple subfamilies or genes within a family, no subfamily or gene designation need be specified. In the case of the present α/β-related EHs there are no subfamilies yet identified. Of course if authors wish to adopt this system but utilize trivial names within a text, they could reference the trivial names to the "formal" names.

Based on the difference matrix of Fig. 3b we can establish the molecular nomenclature names for the EHs. A summary of the family/subfamily boundaries and classification tree for these EHs is given in Fig. 6. As shown, the mEHs would be family HYL1, mammalian sEH HYL2, and plant sEH HYL3.

The nomenclature system described here for those epoxide hydrolases that are related to the α/β hydrolases is similar to the nomenclature systems for the P450s, the glutathione transferases, and the UDP glucuronosyltransferases noted earlier. We are suggesting that percent identity serves as a reasonable and measurable characteristic that can reflect both evolutionary and functional aspects of a protein. An important objective of this nomenclature system is that is should agree with the biological data. Thus, percent identity should be considered a guideline to naming, not an absolute, inflexible criterion. If data exist that argue for a protein being grouped other than where percent identity would indicate, then the data should be considered in whole. For example, we have proposed 40% identity as the minimum identity between family members. The mammalian sEHs are also characterized as deriving from a gene fusion event. In the possible event that a protein encoded by a fused gene homologous to the mammalian sEHs is described that does not share at least 40% identity to mammalian sEHs, perhaps the hypothetical protein should be included in the sEH HYL2 family. With this in mind, while this system may be utilized immediately, it can also be considered a framework to a finalized system to be produced by a nomenclature committee that should be formed and meet on some regular basis.

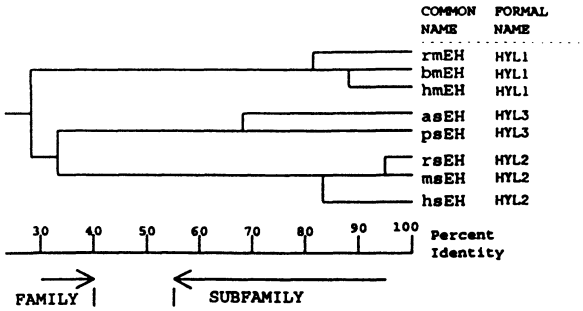

FIG. 6. Classification tree for soluble and microsomal epoxide hydrolases. Percent identities are derived from a difference matrix (Fig. 3).

ACKNOWLEDGMENTS

This publication was supported in part by grants ES02710 from the National Institutes of Health (NIEHS), U.S. EPA (R819658) Center for Ecological Health Research at the University of California at Davis, SFB 302 from the German Research Council, 91-37302-6185 from the U.S. Department of Agriculture, ES05707 from NIEHS Center for Environmental Health Research/National Institutes of Health (NIH), and by special coordination funds from the Science and Technology Agency (STA) of Japan. J.K.B. was supported in part by NIH Biotechnology Training Grant GM08343 and T.K. by an STA fellowship.

REFERENCES

ALTSCHUL, S.F., GISH, W., MILLER, W., MEYERS, E.W., and LIPMAN, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. **215**, 403–410.

ARAND, M., GRANT, D.F., BEETHAM, J.K., FRIEDBERG, T., OESCH, F., and HAMMOCK, B.D. (1994). Sequence similarity of mammalian epoxide hydrolases to the bacterial haloalkane dehalogenase and other related proteins: Implication for the potential catalytic mechanism of enzymatic epoxide hydrolase. FEBS Lett. **338**, 251–256.

BEETHAM, J.K., TIAN, T., and HAMMOCK, B.D. (1993). cDNA cloning and expression of a soluble epoxide hydrolase from human liver. Arch Biochem. Biophys. **305**, 197–201.