Lo, K.-M., Jones, S. S., Hackett, N. R., & Khorana, H. G. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 2285-2289.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Martinez-Carrion, M., Turano, C., Chiancone, E., Bossa, F., Giartosio, A., Riva, F., & Fasella, P. (1967) *J. Biol. Chem. 242*, 2397-2409.
Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499-560.
McPherson, A., Jr. (1976) *J. Biol. Chem. 251*, 6300-6303.
Metzler, C. M., Metzler, D. E., Martin, D. S., Newman, R., Arnone, A., & Rogers, P. (1978) *J. Biol. Chem. 253*, 5251-5254.
Miki, T., Yasukochi, T., Nagatani, H., Furuno, M., Orita, T., Yamada, H., Imoto, T., & Horiuchi, T. (1987) *Protein Eng. 1*, 327-338.
Morino, Y., Tanase, S., Watanabe, T., Kagamiyama, H., & Wada, H. (1977) *J. Biochem. (Tokyo) 28*, 847-852.
Nagashima, F., Tanase, S., Higaki, T., Morino, Y., Murao, S., & Nishino, T. (1986) *J. Biochem. (Tokyo) 99*, 1017-1025.
Obaru, K., Nomiyama, H., Shimada, K., Nagashima, F., & Morino, Y. (1986) *J. Biol. Chem. 261*, 16976-16983.
Okayama, H., & Berg, P. (1982) *Mol. Cell. Biol. 2*, 161-170.
Ouchterlony, O. (1949) *Acta Pathol. Microbiol. Scand. 26*, 507-515.
Ovchinnikov, Y., Egorov, C. A., Aldanova, N. A., Feigina, M. Y., Lipkin, V. M., Abdulaev, N. G., Grishin, E. V., Kiselev, A. P., Modyanov, N. N., Braunstein, A. E., Polyanovsky, O. L., & Nosikov, V. V. (1973) *FEBS Lett. 29*, 31-33.
Rigby, P. W. J., Dieckman, M., Rohdes, C., & Berg, P. (1977) *J. Mol. Biol. 113*, 237-251.
Teranishi, H., Kagamiyama, H., Teranishi, K., Wada, H., Yamano, T., & Morino, Y. (1978) *J. Biol. Chem. 253*, 8842-8847.
Wain-Hobson, S., Nussinov, R., Brown, R. J., & Sussman, J. L. (1981) *Gene 13*, 355-364.
Yoshimoto, T., Murayama, N., Honda, T., Tone, H., & Tsuru, D. (1988) *J. Biochem. (Tokyo) 104*, 93-97.

Inducible (Class 3) Aldehyde Dehydrogenase from Rat Hepatocellular Carcinoma and 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin-Treated Liver: Distant Relationship to the Class 1 and 2 Enzymes from Mammalian Liver Cytosol/Mitochondria†

John Hempel,* Kimberly Harper,‡ and Ronald Lindahl‡
Department of Microbiology, Biochemistry, and Molecular Biology, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, and Biochemistry Program, Department of Biology, University of Alabama, Tuscaloosa, Alabama 35487
Received June 8, 1988; Revised Manuscript Received August 22, 1988

ABSTRACT: Peptides from rat liver aldehyde dehydrogenase (AlDH) induced by 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) treatment match the AlDH structure from HTC rat hepatoma cells (HTC-AlDH) at all positions examined, indicating induction of the same gene product by two independent routes. This 452 amino acid residue, class 3 AlDH structure differs substantially from the 500-residue AlDH structures isolated from normal liver cytosol (class 1) and mitochondria (class 2). Despite a 29.8% identity in 429 overlapping amino acids vs the human class 1 enzyme (27.7% vs class 2), neither the N- nor C-termini coincide, and gaps are introduced to optimize the alignment. Two residues placed in the active site of human liver AlDH by chemical modification, Cys-302 and Glu-268, are conserved in class 3 AlDH as Cys-243 and Glu-209. Cys-243/302 is the only cysteine residue conserved in *all* known AlDH structures. Gly-245 and Gly-250 of class 1/2 AlDHs, fitting the patterns of glycine residues in coenzyme binding fold of other dehydrogenases, are also conserved. Otherwise, Cys-49, Cys-162, and Glu-487, to which functional importance has also been ascribed, are not retained in the class 3 structure. Overall, a high conservation of Gly, Pro, and Trp and similar patterns of predicted secondary structure indicate general conservation of tertiary structure, as noted with other distantly related proteins. Three exon boundaries from the human liver mitochondria AlDH gene directly correspond to the N-terminus of the rat class 3 protein and to two of the gaps in the alignment. Three contiguous matches also occur within the 16-residue C-terminal "extension" of the class 3 structure vs the hypothetical protein sequence obtaining from the 3′ noncoding sequence of the human mitochondrial gene in the absence of the stop codon. Thus, exon addition, splice junction alterations, and stop codon "migration" appear to have occurred during evolution of the class 1 and 2 structures from an ancestor common to class 3.

Liver aldehyde dehydrogenases (AlDHs)[1] display a wide substrate specificity and notably participate in clearance of ethanol-derived acetaldehyde. In human, bovine, equine, and ovine liver, two NAD-dependent enzyme forms, one cytosolic and one mitochondrial, have been described (Eckfeldt & Yonetani, 1976; Greenfield & Pietruszko, 1977; Crow et al., 1974; Kitabatake et al., 1981). At present, the complete primary structures of the human (Hempel et al., 1984, 1985;

† This work was supported by a grant (AA 06985) from the National Institute on Alcohol Abuse and Alcoholism (to J.H. and R.L).
* Address correspondence to this author at the University of Pittsburgh.
‡ University of Alabama.

[1] Abbreviations: HTC, hepatocellular carcinoma; AlDH, aldehyde dehydrogenase (EC 1.2.1.3); TCDD, 2,3,7,8-tetrachlorodibenzo-*p*-dioxin; HPLC, high-performance liquid chromatography; TLCK, N^{α}-tosyllysine chloromethyl ketone; DABITC, 4-(*N*,*N*-dimethylamino)azobenzene 4′-isothiocyanate.

0006-2960/89/0428-1160$01.50/0

Hsu et al., 1988) and horse (von Bahr-Lindström et al., 1984; Johansson et al., 1988) enzyme pairs are known, and data on the bovine pair are being gathered (Guan & Weiner, 1987). The positional identity between homotopic forms (e.g., human cytosolic vs horse cytosolic) is substantially greater (~90–95%) than between the heterotopic forms from the same species (~70% positional identity); thus these liver cytosolic and mitochondrial AlDHs may be considered separate families [Dayhoff et al., 1983; but see also Doolittle (1981)]. On this basis "class 1" and "class 2" designations have been proposed for the respective families (Lindahl et al., 1988). From comparison of the sequences of the human and horse enzyme pairs, together with data from chemical modifications and natural polymorphisms, interpretations have been made concerning active sites residues, coenzyme binding site regions, potential regions of subunit contact, and segments possibly reflecting ancestral internal duplication (Hempel & Jörnvall, 1987).

In the rat, an apparently dissimilar distribution of aldehyde dehydrogenases exists. Two mitochondrial forms and two microsomal forms are noted in normal rat liver (Lindahl & Evces, 1984a, and references cited therein). Additional forms inducible by such xenobiotics as phenobarbital and 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) (Deitrich et al., 1977), as well as during carcinogenesis, have been purified and characterized with regard to basic physicochemical properties (Lindahl & Evces, 1984b). Some properties of the inducible rat liver AlDHs differ from those of the "normal" liver forms while others are typical. The AlDH induced by TCDD treatment and appearing in association with hepatocellular carcinoma (HTC·AlDH) also occurs normally in rat bladder (Lindahl, 1986) and lung and stomach (Koivusalo, 1988). They are dimeric structures with $K_{cat}/K_{m,benzaldehyde} \gg K_{cat}/K_{m,propanal}$ and $K_{cat,NADP^+} > K_{cat,NAD^+}$ but retain sensitivity to disulfiram characteristic of the tetrameric class 1 enzymes (Kitson, 1983; MacKerrell et al., 1985). Therefore, we sought to obtain structural data on these class 3 aldehyde dehydrogenases from the rat, which might offer some functional correlations. The structure of the HTC·AlDH has recently been reported in connection with expression of the cDNA encoding that enzyme (Jones et al., 1988). Here the enzyme induced by TCDD treatment is concluded to be identical with that structure. Moreover, we now report interpretations from comparison of this extensively divergent, rat class 3 AlDH primary structure with the human/horse AlDH structures.

Experimental Procedures

TCDD-Inducible AlDH. Rat liver AlDH induced by TCDD was purified (Lindahl & Evecs, 1984a) and S-alkylated with iodo[^{14}C]acetic acid in the presence of 6 M guanidine hydrochloride (Hempel et al., 1985). After dialysis vs 100 mM ammonium bicarbonate the ^{14}C-carboxymethylated protein was digested with trypsin (1:50 w/w) and prefractionated over Sephadex G-75 (1.5 × 190 cm) in 30% acetic acid. The eluate was monitored for absorbance at 280 nm and for radioactivity. Pooled fractions were rechromatographed over 3.9 × 300 mm C-18 μBondapak (Waters) or Ultropac (LKB) columns in aqueous 0.1% trifluoroacetic acid with linear gradients of acetonitrile. The eluates were monitored at 214 and 280 nm; aliquots from peaks of UV absorbance were analyzed for amino acid content after hydrolysis (24 h, 110 °C, 6 N HCl, 0.5% phenol in vacuo) with a Beckman Model 6300 amino acid analyzer. Material from selected peaks was submitted either to manual analysis, using the DABITC double-coupling method with residue identification by two-dimensional thin-layer chromatography (Chang et al., 1978; von Bahr-Lindström et al., 1982), or to automated Edman degradation in the presence of precycled polybrene, in an upated Beckman 890 M protein sequenator. PTH amino acids were identified by HPLC (Hawke et al., 1982). Some larger peptides were further digested with the Glu-specific protease from *Staphylococcus aureus* (Miles) or chymotrypsin (Sigma, TLCK treated). All peptides are named by an alternate letter–number combination in which the letters denote the type of cleavage: T, trypsin; S, staphylococcal Glu-specific protease; C, chymotrypsin. The numbering of peptides is consecutive according to their position in the sequence.

HTC·AlDH. The primary structure of HTC·AlDH is taken from Jones et al. (1988) as deduced from cDNA and supported by CNBr and tryptic peptide data.

The structure was compared, by using the Lipman–Pearson algorithm (Lipman & Pearson, 1985), to all entries in the May, 1988, version of the National Biomedical Protein Identification Resource Database (George et al., 1986), including both human aldehyde dehydrogenases.

Secondary structural character was assessed by using the α-helical, β-sheet, and reverse-turn potentials of Chou and Fasman (1978). Helical and β-sheet calculations were based on successive six- and five-residue segments, respectively. No segments of shorter length or any including α-helical or β-sheet terminating tetrapeptides were predicted as α or β, despite some with a sufficiently high numerical potential. For prediction of reverse turns, $p_t > 1.5 \times 10^{-4}$ was used as the cutoff, to avoid overprediction.

Results

TCDD-Inducible AlDH. Peptides from rat liver AlDH induced by TCDD treatment were isolated after Sephadex and reverse-phase chromatographies. All sequences obtained were identical with segments from the HTC enzyme as indicated in Figure 1. Compositional data for these petides are listed in supplemental Table 1 (see paragraph at end of paper regarding supplementary material), with the exception of T20, which showed low Gly and high Glx values due to contamination from low levels of other peptides. Due to limiting amounts of material and unambiguous sequencer results from T20 over 29 cycles, the remaining material was cleaved with chymotrypsin, from which T20C1 was obtained. The composition and phenylthiohydantoin derivatives of this peptide and the composition of T20C2 (not shown) support the remainder of the T20 sequence shown in Figure 1. Thus, all data equate the hepatoma-associated enzyme with the TCDD-induced enzyme. The HTC/TCDD structure represents the first example of a class 3 AlDH (Lindahl et al., 1988).

Alignment of HTC/TCDD AlDH with Human Liver Cytosolic AlDH. The greatest degree of positional identity to this class 3 enzyme, revealed by a database search using the Lipman–Pearson algorithm (Lipman & Pearson, 1985), was against human liver class 1 aldehyde dehydrogenase. Two gaps totaling 7 positions introduced into the class 1 structure and nine gaps totaling 15 positions inserted into the class 3 structure optimize the alignment. By so doing, and excluding the residues of one structure corresponding to a gap in the other, a positional identity of 29.8% is achieved within the 429 aligned residue pairs, providing an initial alignment score of 120 (optimized: 414). These values are 26.2 and 59.7 standard derivations above the mean obtained from a comparison of the class 3 sequence with 20 randomly generated sequences of the same composition as the human class 1 structure. These identities occur over the entirety of the alignment, even though the N-terminal 56 residues of the human class 1 and the C-terminal 16 residues of class 3 AlDH are without apparent

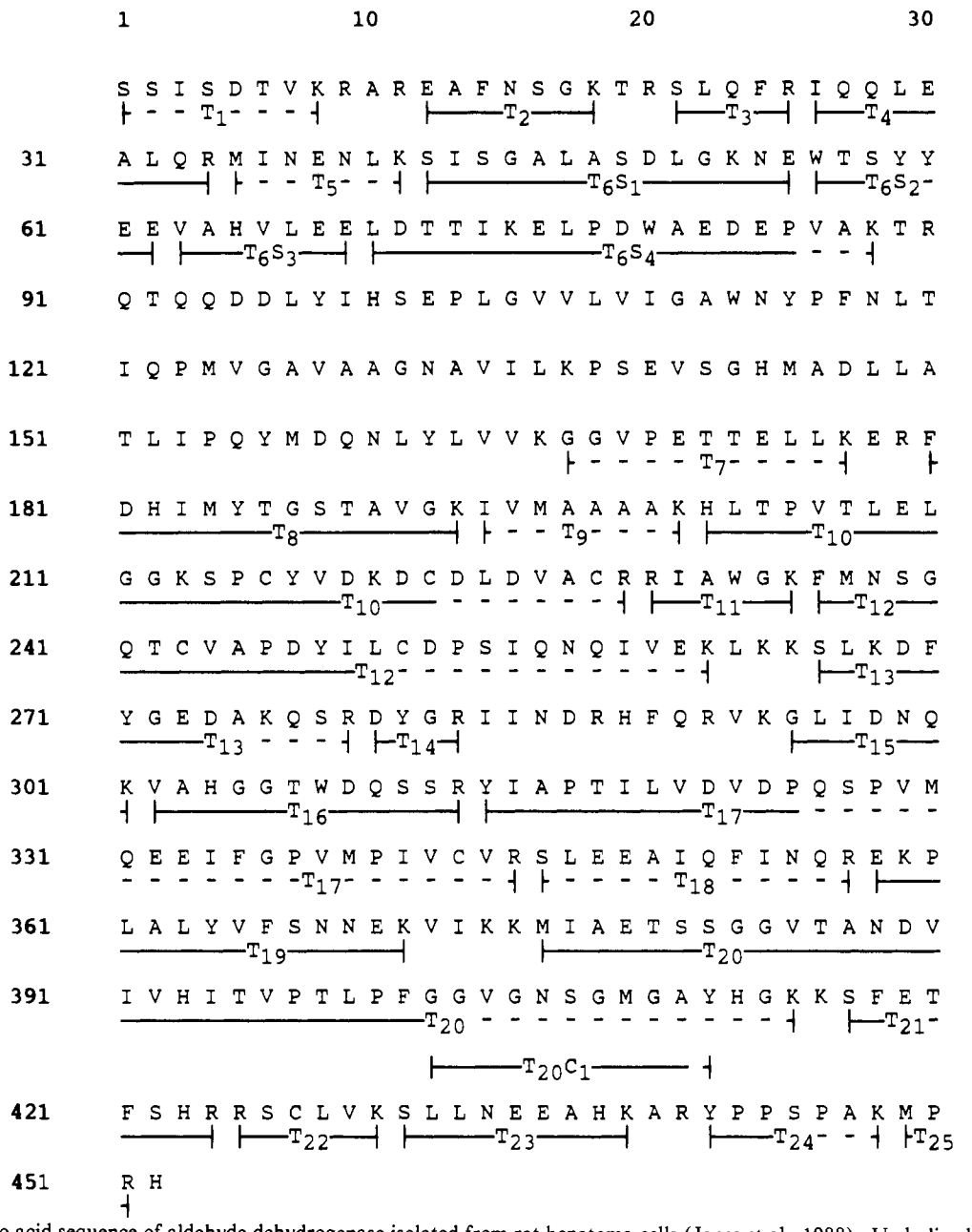

FIGURE 1: Amino acid sequence of aldehyde dehydrogenase isolated from rat hepatoma cells (Jones et al., 1988). Underlined segments denote peptides with identical sequences from TCDD-induced rat liver AlDH, as supported by Edman degradation and compositional analysis (solid underlining) or compositional analysis alone (dashed underlining). Compositions are given in supplemental Table 1.

counterpart in the other structure[2] (Figure 2). Although only two segments of more than four contiguous identical residues are found, single and dipeptide identities are numerous and in register without resort to placement of an excessive number of gaps [gaps/100 residues = 2.6, similar to values of other putatively related protein pairs cited by Doolittle (1981)], and many of the remaining aligned, nonidentical residues are similar pairs. Particular residues from the human structures that are conserved in this alignment include Gly-245, Gly-250, Glu-268, and Cys-302, while a high overall conservation of Gly, Pro, and Trp is also noted (Table I). These conservations are considered further under Discussion. The highest score against a non-AlDH structure was achieved in comparison to the hypothetical Ec-RF2 protein from Epstein–Barr virus (Bankier et al., 1983), 23% positional identity over 91 residues with intitial and optimized alignment scores of 57 and 66, respectively.

Differences in Comparison with Other Class 1/2 AlDHs. The alignment of Figure 2 is maintained also in comparison of the rat class 3 enzyme to the other class 1 or class 2 structures from human and horse liver. The human class 1 structure is marginally higher in number of matching residues, with 128. The horse class 1 enzyme [itself with 91% positional identity with the human class 1 (von Bahr-Lindström et al., 1984)] reveals 11 positions that do not retain the matches of Figure 2 (cf. Table II); offsetting this loss are four "new" matches not seen in Figure 2. Of this total of 15 exchanges, 4 involve deviations from conserved isoleucine residues, and a pronounced loss of Ile in the horse vs human class 1 structures has already been noted (von Bahr-Lindström et al., 1984).

[2] The alignment poses inherent problems concerning positional nomenclature when discussing conserved residues. For instance, Cys-302 of human and equine NAD-dependent AlDHs corresponds to Cys-243 of the rat class 3 structure. Under Discussion, such a conserved residue will be referred to as Cys-243/302, always giving the correct positional number from the class 3 structure first.

Table I: Amino Acid Residue Conservation between Human Class 1 and Rat Class 3 AlDHs[a]

amino acid	% of class 1	% of class 3
Cys	2/11 (18)	2/7 (29)
Asp	4/26 (15)	4/25 (16)
Asn	3/20 (15)	3/18 (17)
Thr	4/28 (14)	4/23 (17)
Ser	6/29 (21)	6/31 (19)
Glu	5/38 (13)	5/29 (17)
Gln	1/18 (5)	1/21 (5)
Pro	8/24 (33)	8/24 (33)
Gly	21/48 (44)	21/31 (68)
Ala	11/38 (29)	11/34 (32)
Val	14/38 (37)	14/37 (38)
Met	0/8 (0)	0/12 (0)
Ile	13/35 (37)	13/29 (45)
Leu	11/37 (30)	11/36 (31)
Tyr	2/16 (13)	2/15 (13)
Phe	6/22 (27)	6/13 (46)
Trp	3/7 (43)	3/4 (75)
Lys	11/38 (29)	11/30 (37)
His	0/8 (0)	0/12 (0)
Arg	3/17 (18)	3/20 (15)
total	128/500 (25.6)	128/452 (28.3)

[a] Denominators reflect total compositions of the respective enzymes and are not adjusted to reflect gaps introduced to optimize the alignment of Figure 2. Thus, total percentages on the last line differ from percentages conserved within the alignment (text) where residues corresponding to gaps are not counted.

A larger number of deviations from the matches in Figure 2 are seen in comparing class 2 AlDHs from human or horse liver to the class 3 structure, although the net difference is again minimal. When the human class 2 sequence is compared with the class 3 structure, 24 of the above 128 identities are lost, while 15 new matches are noted, providing initial and optimized Lipman–Pearson scores of 114 and 364, respectively. With the horse class 2 structure, 27 of the 128 matches are lost and 17 are gained. Many of these alterations are identical with ones detected vs human class 2 (Table II).

At most positions where at least one of the class 1/2 AlDHs is identical with the rat class 3 structure (Table II), the differing residues reflect conservative substitutions. However, at position 216/275, where Cys is conserved only between the class 3 and human class 1 structures, the exchanges Cys ↔ Asn (class 3 vs class 2) and Cys ↔ Phe (class 3 vs horse class 1) are much less common, as indicated by PAM 250 scores of −4 in each case (Dayhoff, 1983). The effect of the exchange for Asn may be negligible, though, since it only serves to strengthen a predicted reverse turn. In the rat structure Cys-216 is predicted as the third residue of a reverse turn of $P_t = 5.3 \times 10^{-4}$ (Ser-Pro-Cys-Tyr); in the human class 1 structure this segment has a $P_t = 2.4 \times 10^{-4}$ (Ser-Pro-Cys-Ile). In both class 2 structures, with Ser-Pro-Asn-Ile segments, $P_t = 3.8 \times 10^{-4}$.

Also, as a result of these changes in the class 1/2 structures relative to Cys-216/275, Cys-243/302 emerges as the only cysteine residue retained in all AlDHs presently known, including the enzyme from *Aspergillus* (Pickett et al., 1987).

Correlations with the Human Class 2 AlDH Gene. Considering the organization of the human class 2 gene (Hsu et al., 1988), several additional correlations may be noted relative to the alignment of Figure 2. First, the N-terminal "extension" of the class 1 and 2 structures corresponds to the coding portion of the first exon and the entirety of the second exon of this gene. Thus, the N-terminus of the class 3 structure corresponds exactly to the beginning of the third exon of the human class 2 gene. Additionally, gaps in the alignment (Figure 2) directly correspond to the boundaries between human mitochondrial AlDH exons 4 and 5 and exons 6 and 7, respectively. Finally, the 13th exon of the human class 2 gene, if not punctuated by a stop codon, would allow deduction of a sequence Lys-Leu-Pro-Pro-Ser corresponding in exact register to the sequence Arg-Tyr-Pro-Pro-Ser (residues 441–445) of the class 3 structure (Figure 3), although this degree of similarity is not maintained at the nucleotide level, nor is it seen in the 3′ noncoding sequence (Hsu et al., 1985) of the human class 1 cDNA.

Table II: Positions in Rat Class 3 Aldehyde Dehydrogenase Where Identity (Underlined) Is Found with at Least One of the Four Class 1/Class 2 Enzymes from Human/Horse Liver

Class 3		Class 1		Class 2		Class 1/2
position	rat	human	horse	human	horse	position
8	Lys	Lys	Ala	Lys		64
36	Ile	Ile	Val	Ile		95
50	Asp	Asn		Asp		109
58	Ser	Ala		Ser		117
64	Ala	Ala	Gly	Asp		123
66	Val	Cys		Val		125
67	Leu	Ile	Leu	Leu		126
121	Ile	Ile	Leu	Ala		176
129	Ala	Ser		Ala		184
138	Pro	Pro		Val		193
151	Thr	Ser	Thr	Asn		206
153	Ile	Ile		Ile	Thr	208
164	Val	Ile		Ile	Val	220
191	Val	Val		Ile	Val	249
193	Lys	Lys		Arg	His	251
201	Lys	Lys		Ser	Arg	259
216	Cys	Cys	Phe	Asn		275
218	Val	Val		Ile		277
224	Leu	Leu		Met		283
225	Asp	Asp	Glu	Asp		284
234	Gly	Gly	Ala	Ala		293
238	Asn	His		Asn		297
244	Val	Ile	Val	Cys	Gly	303
249	Ile	Ile	Leu	Thr		308
254	Ser	Ser		Asp		313
255	Ile	Ile		Ile	Val	314
261	Glu	Arg		Glu		320
268	Lys	Lys		Ser		328
285	Ile	Ile		Val		345
290	Phe	Tyr		Phe		350
293	Val	Ile		Ile	Val	353
295	Gly	Asp		Gly		355
296	Leu	Leu		Tyr		356
308	Trp	Trp		Ala		374
315	Ile	Val	Ile	Ile		381
322	Asp	Asn		Asp		388
339	Met	Gln		Met		405
346	Ser	Ser		Thr		412
347	Leu	Leu		Ile		413
348	Glu	Asp		Glu		414
349	Glu	Asp		Glu		415
351	Ile	Ile		Val		417
362	Ala	Ser	Phe	Ala		428
365	Val	Val	Ser	Val		431
373	Ile	Ile	Leu	Leu		436
377	Ile	Ile		Asn		440
381	Ser	Ser	Ala	Gln		444
390	Val	Val		Val	Ile	453
396	Val	Val		Phe		459
418	Phe	Phe		Leu		482
429	Val	Val		Val	Ile	493
Total differentially identical to Class 3 enzyme		32	25	23	22	

Discussion

TCDD-Inducible AlDH. The equivalence of the AlDH forms associated with rat hepatocellular carcinoma and induced by TCDD is consistent with earlier indications based on kinetic, immunological, and physiochemical data (Lindahl & Evces, 1984a,b).

Alignment with Human and Horse Liver AlDH. Comparison of the class 3 structure against the National Biomedical Research Foundation Database reveals that the class 1 and

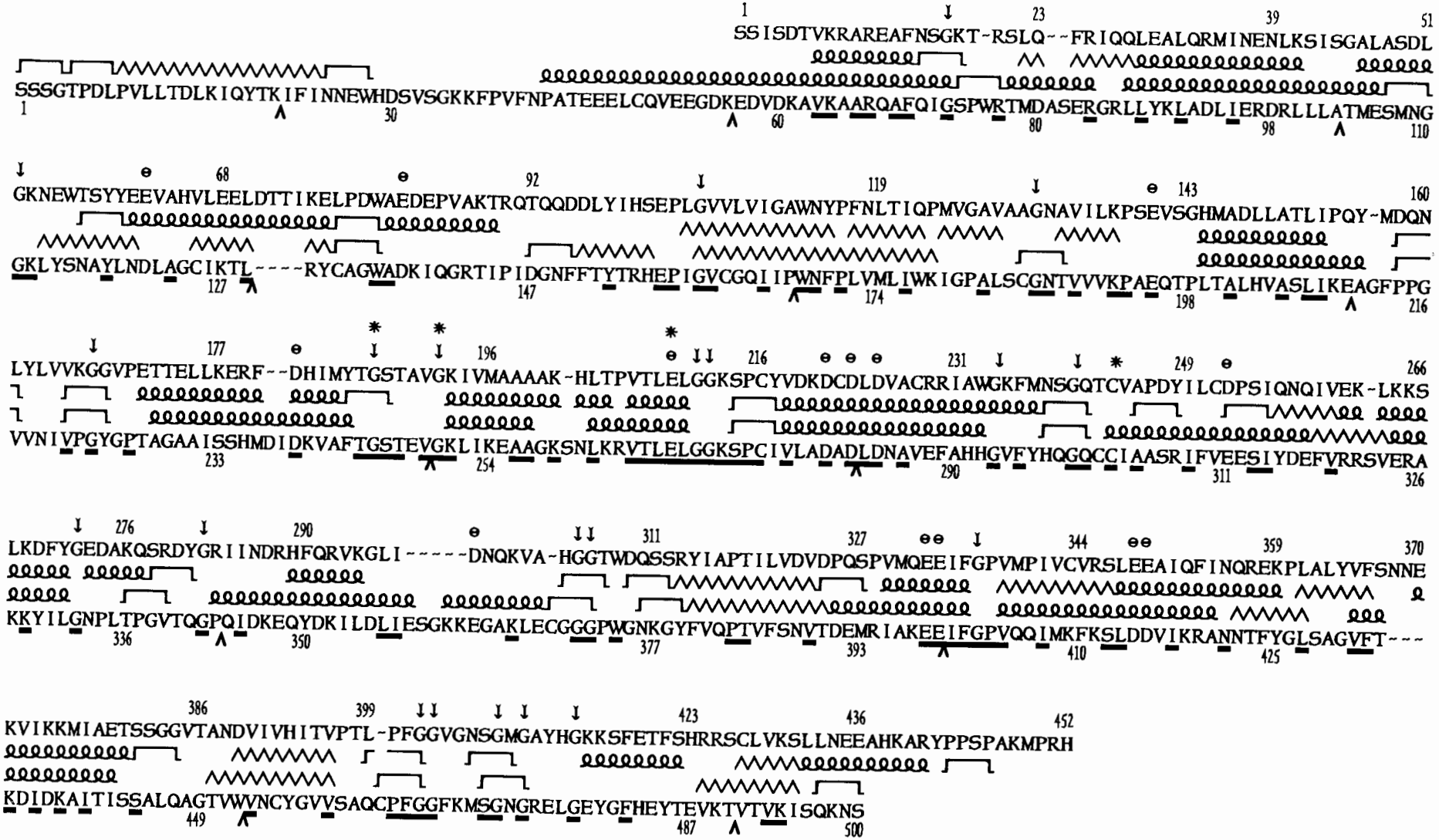

FIGURE 2: Alignment of rat class 3 aldehyde dehydrogenase (upper sequence) with human liver class 1 aldehyde dehydrogenase (lower sequence). The alignment was generated by using the *fastp* algorithm of lipman and Pearson (1985). Below the class 1 sequence, residues conserved between the two structures are indicated by heavy bars, and the corresponding locations of exon boundaries in the class 2 gene sequence (Hsu et al., 1988) are denoted by (∧). Above the class 3 structure, conserved glycine residues (arrows), conserved acidic residues, including Asp ↔ Glu exchanges (Θ), and residues specifically ascribed functional roles (*) are indicated. Predicted (Chou & Fasman, 1978) secondary structures are indicated below and above the class 3 and class 1 structures, respectively. Spirals represent predicted α-helical segments, sawteeth represent regions of predicted β-sheet structure, and footed brackets denote predicted reverse turns.

	437 440 452
Rat Class 3 AlDH	Leu Asn Glu Glu Ala His Lys Ala Arg Tyr | Pro Pro Ser | Pro Ala Lys Met Pro Arg His
	TTGAATGAAGAAGCTCACAAGGCCAGGTAT|CCCCCAAGC|CCAGCCAAGATGCCCCGGCAC
Human Class 2 AlDH	CAGAAGAACTCATAAGAATCATGCAAGCTT|CCTCCCTCA|GCCATTGATGGAAAGTTCAGC
	Gln Lys Asn Ser * Glu Ser Cys Lys Leu | Pro Pro Ser | Ala Ile Asp Gly Lys Phe Ser
	500

FIGURE 3: Comparison of the C-terminus of rat class 3 AlDH with the hypothetical extension of the human class 2 structure, based on the alignment of Figure 2 and cDNA sequences. The known C-terminal protein sequence of the class 3 AlDH is given together with its corresponding cDNA (Jones et al., 1988). Below this the human class 2 cDNA (Hsu et al., 1985) is given with the actual C-terminal tetrapeptide sequence ending with Ser-500. Downstream from the stop codon, the hypothetical protein sequence, which would obtain in the absence of the stop codon (but in register with it), is also given. The resulting amino acid identities are boxed.

2 AlDHs as now known from human and horse liver share a clear but distant relationship. The tetrameric class 1 and 2 enzymes have subunits of 500 amino acid residues with 90–95% positional identity between the homotopic, species variants and ~70% identity between heterotopic, interclass pairs from the same species, all without resort to placement of gaps to maximize the alignment (von Bahr-Lindström et al., 1984; Hempel et al., 1985; Johansson et al., 1988). In constrast, neither the N- nor C-terminus of the 452 residues class 3 AlDH structure occurs at the same position relative to the class 1/2 AlDHs, and the chains attain maximal positional identity (29.8%) only by placement of 11 gaps into the two structures (Figure 2).

Overall, many of the apparent residue replacements are conservative changes. This is particularly noticeable in the case of those positions listed in Table II, where at least one of the human/horse enzymes displays a match to the class 3 structure. The nonmatching residues almost uniformly constitute conservative replacements, viz., Ile ↔ Leu ↔ Val, Thr ↔ Ser, Glu ↔ Asp, and Gly ↔ Ser ↔ Ala (Dayhoff et al., 1983).

Regarding the common ancestor of these three enzymes, the nearly equal degree of relatedness of the class 1/2 AlDHs to this class 3 structure (Table II, Figure 4) indicates that divergence leading to class 3 occurred prior to the divergence of the class 1/class 2 families. This conclusion is further supported from examination of Table II. Of the 51 positions listed where uniformly identical residues are not found between the five mammalian AlDHs considered herein, 15 positions show symmetry in matches of the class 3 structure against only the class 1 enzymes, with identical, nonmatching residues in the class 2 structures (e.g., position 138/193) while the converse applies at 12 positions (e.g., position 50/109).

Correlations with the Human Mitochondrial AlDH Gene. A further striking indication of the common ancestry of the class 1, 2, and 3 AlDHs is the relationship of exon boundaries in the human class 2 AlDH gene to noncorresponding portions of the two structures in Figure 2. At its N-terminus, the class 3 structure lacks all residues corresponding to exons 1 and 2 of the human class 2 enzyme. The actual N-terminus corresponds exactly to the beginning of exon 3, while two of the gaps introduced to the alignment occur at exon boundaries (see Results). Additionally, the correspondence between the C-terminal 16-residue "extension" of the class 3 protein structure and the hypothetical protein sequence, which would obtain downstream from the 3′ coding end of the human class 2 AlDH cDNA in the absence of the stop codon, is also noted (Figure 3). The correlation is suggestive of possible remnants of a common ancestral structure.

Together, the above relationships suggest that cryptic splice sites and stop codons have been actuated by selected nucleotide base replacements in the course of evolving from the class 3 and class 1/2 structures. Thus, effective wholesale loss (or

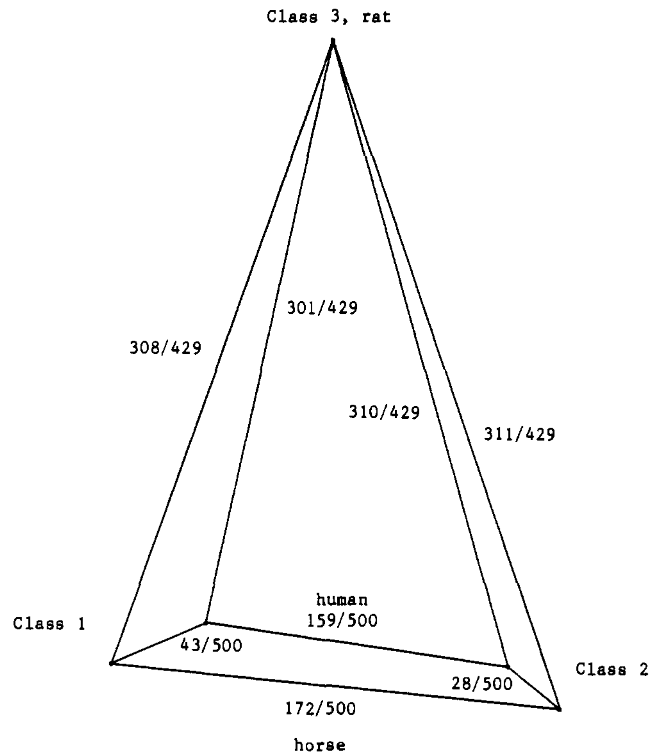

FIGURE 4: Relationships between rat class 3 AlDH and human and horse class 1/2 AlDHs. Values represent the number of differences per number of aligned residues, excluding the 71 residues of the class 1/2 structures and the 23 residues of the class 3 structure that have no counterpart in the opposing structure, as a result of nonaligned N- and C-termini and gaps introduced in optimizing the alignment.

gain) of entire exons at the 5′ end of the common ancestral gene would have obtained, to account for the noncorresponding N-termini just as minor adjustments in the precise positioning of the exon 4/5 and 6/7 boundaries would be revealed through two of the gaps now placed to optimize the alignment of Figure 2, and mutations yielding a stop codon in place of the GCT encoding Ala-437 of the class 3 structure (or vice versa) offer the simplest explanation for the relationship (Figure 3) between the translated cDNA sequences of the two AlDHs.

These observations are consistent with a previous correlation between AlDH protein structure and AlDH gene structure, noted in connection with the extreme variability between the N-terminal 21 amino acids of the cytosolic and mature mitochondrial human liver enzymes, in contrast to the far more highly conserved residues of this pair, from position 22 onward. The exon 1/exon 2 boundary occurs, relative to the protein sequence, between residues 21/22 (Hsu et al., 1988). A further correlation, noted relative to the boundary between class 2 exons 7 and 8, is discussed below in connection with Gly-187/245 and Gly-192/250. Together, all of these observations underscore the importance of exons as units of

evolution (Lonberg & Gilbert, 1985).

Conservation of Previously Identified Specific Residues. Derivatizations of Cys-302 with iodoacetamide (Hempel et al., 1982) and a coenzyme analogue (von Bahr-Lindström et al., 1985), and of Glu-268 with bromoacetophenone (Abriola et al., 1987), have placed each residue in the active site. Recently, reactivity with *N*-ethylmaleimide of Cys-49 and Cys-162 in horse class 2 AlDH has supported a catalytic role for these residues (Tu & Weiner, 1988a,b). The exchange of Glu-487 in human class 2 AlDH for Lys in the low activity mutant form of the enzyme (found in Oriental individuals with a sensitivity to alcohol) has supported an important function for that residue (Yoshida et al., 1984). In addition, but without direct proof, distributions of glycine residues like those found in the coenzyme binding fold of dehydrogenases have been noted in two segments from the human class 1 and 2 structures, at positions 223–229 and 245–250, as well as internally repetitive patterns of amino acids, which may reflect internal duplications, between two pairs of segments at positions 77–80 vs 90–101 and 224–247 vs 425–444 (Hempel et al., 1985).

The alignment shown in Figure 2 supports the importance of Glu-209/268, Cys-243/302, and Gly-187/245 + Gly-192/250 since all are clearly conserved. In contrast, neither Cys-49, Cys-162, Glu-487, two of the three glycine residues at positions 223–229, or either internally repetitive pattern is found in the class 3 structures. Despite the importance of cysteine and of acidic residue(s) indicated from various studies (Deitrich, 1967; Kitson, 1981; Bennett et al., 1982; Dunn & Buckley, 1985, and references cited therein), which might suggest the likelihood of high overall conservation of these residues, the conservation of Glu-209/268 and Cys-243/302 is given added emphasis from Table I, showing the low overall conservation of Cys, Glu, and Asp. Including Glu ↔ Asp substitutions with the total of Asp and Glu strictly conserved gives 14/64 acidic residues conserved (22%) in the class 1 or 14/54 (26%) in the class 3 structures; all acidic residue conservations are noted in Figure 2. The conservation of Cys-243/302 is yet further emphasized in that it is the only cysteine residue retained in all AlDH structures presently known (cf. Tables I and II, exchanges at position 216/275).

The significance of the Gly-187/245 and Gly-192/250 conservations is given added support from the human mitochondrial AlDH genomic data, since the boundary between exons 7 and 8 occurs in the human and horse enzymes between Gly-245 and Gly-250. Between the known segments composing the nucleotide binding folds of alcohol dehyrogenases and pyruvate kinase a distant relationship was earlier detected at the primary and tertiary protein structural level, and exon boundaries were found to correspond to positions within the mononucleotide binding fold (Brändén et al., 1984; Lonberg & Gilbert, 1985) where the glycine-rich region is found in alcohol and other dehydrogenases (Rossmann et al., 1975; Wierenga & Hol, 1983). Thus, the location of an exon boundary between Val-249 and Gly-250 (human class 2 gene) appears to lend further indirect support to identification of the segment containing Gly-187/245 and Gly-192/250 as presumptively part of the nucleotide binding domain.

Conservation in Flanking Segments. Conservtion of Gly-187/245 + Gly-192/250, Glu-209/268, and Cys-243/302, all located within ~10% of the linear sequence, directs notation of consensus sequences flanking these residues. Glu-209/268 is the fourth residue of a 10-residue segment, Val-Thr-Leu-Glu-Leu-Gly-Gly-Lys-Ser-Pro, which is invariant in all AlDH structures thus far known and which is indicated to terminate in a reverse turn (see Results). Gly-187/245 and Gly-192/250 are included in a Thr-Gly-Ser-Thr-X-Val/Ile-Gly sequence. This consensus is followed by a basic and then an alkyl-branched residue in all five mammalian AlDHs currently known, but not in the *Aspergillus* enzyme (Pickett et al., 1987). Finally, Cys-243/302, apparently the only essential cysteine residue in AlDHs, is flanked within a consensus Gly-Gln-X-Cys-X-Ala sequence. Interestingly, in the human class 2 and *Aspergillus* structures the residues at *both* X positions are Cys, and in all human and horse AlDHs presently known the first X is Cys (Johansson et al., 1988). The rat class 3 AlDH structure is thus the first AlDH to be reported without any Cys-Cys sequences.

Conservation of Secondary Structure. Despite the limitations of assessing secondary structure on the basis of primary structure, a striking indication of largely similar seconary structure between subunits of the tetrameric class 1 and dimeric class 3 AlDHs emerges when the predictions are combined with the *fastp* alignment (Figure 2). Some segments with few or no identities are predicted as both helical or both β-sheet, and at least half of the predicted reverse turns are seen as closely, if not exactly, coinciding. It is particularly interesting that three of these coinciding reverse-turn predictions do not include any matching amino acids, and one further includes only one identity within the tetrad. The alignment of reverse turns suggests conserved bends in the two chains. The prevalence of Pro, Gly, and Trp at the second, third, and fourth positions, respectively, of reverse turns (Chou & Fasman, 1978) may partially explain why they are among the most highly conserved residues of the alignment (Table I). Of the three retained tryptophan residues, one (Trp-80/135) is at the fourth position of a predicted turn, and another (Trp-308/274) occurs between consecutive conserved turns. Six of the conserved glycine residues are within apparently conserved predicted turns, as are two proline residues. Otherwise, Gly and Pro frequently serve to terminate conserved predicted helices and β-sheets and to maintain chain bending points. The overrepresentation of Gly among residues conserved between distantly related proteins has been noted before as relating to maintenance of tertiary structure (Jörnvall, 1984; Blundell, 1986; & Jörnvall et al., 1987). The greater general conservation of nonpolar vs polar residues (Table I) probably reflects greater selective pressure against change in the interior of proteins. Thus, on the basis of the general correspondence of predicted helical, sheet, and turn regions between the two structures, and the relationship of Gly, Pro, and Trp (specifically) and nonpolar residues (generally) to this correspondence, it seems likely that the tertiary structures formed by the corresponding portions of the two structures (Figure 2) are similar.

Summary

In summary, evidence of the relatedness of the class 1/2 AlDHs of human and horse liver and the inducible class 3 AlDH of rat liver is seen at the primary, predicted secondary, and gene structural levels. The primary structural relationships support the functional significance of some (but not all) amino acids previously indicated by other studies.

Acknowledgments

We are indebted to Tracy Dunn, University of Wisconsin, for supplying livers of TCDD-treated rats. We are grateful for the generous and enthusiastic help from Roger Hendrix in preparing Figures 2 and 3.

Supplementary Material Available

Amino acid compositions of 28 peptides from TCDD-in-

duced rat AlDH (2 pages). Ordering information is given on any current masthead page.

Registry No. TCDD, 1746-01-6; AlDH, 9028-86-8.

REFERENCES

Abriola, D. P., Fields, R., Stein, S., MacKerell, A. D., Jr., & Pietruszko, R. (1987) *Biochemistry 26*, 5679–5684.

Bankier, A. T., Deininger, P. L., Farrell, P. J., & Barrell, B. G. (1983) *Mol. Biol. Med. 1*, 21–45.

Bennett, A. F., Buckley, P. D., & Blackwell, L. F. (1982) *Biochemistry 21*, 4407–4413.

Blundell, T. (1986) *Chem. Scr. 26B*, 213–219.

Brändén, C.-I., Eklund, H., Cambillau, C., & Pryor, A. J. (1984) *EMBO J. 3*, 1307–1310.

Chang, J.-Y., Brauer, D., & Wittmann-Liebold, B. (1978) *FEBS Lett. 93*, 205–214.

Chou, P., & Fasman, G. (1978) *Annu. Rev. Biochem. 47*, 251–276.

Crow, K. E., Kitson, T. M., MacGibbon, A. K. H., & Batt, R. D. (1974) *Biochim. Biophys. Acta 350*, 121–128.

Dayhoff, M. O., Barker, W. C., & Hunt, L. T. (1983) *Methods Enzymol. 91*, 524–545.

Deitrich, R. A. (1967) *Arch. Biochem. Biophys. 119*, 253–263.

Deitrich, R. A., Bludeau, P., Stock, T., & Roper, M. (1977) *J. Biol. Chem. 252*, 6169–6179.

Doolittle, R. F. (1981) *Science 214*, 149–159.

Dunn, M. F., & Buckley, P. D. (1985) *Prog. Clin. Biol. Res. 174*, 15–27.

Eckfeldt, J. H., & Yonetani, T. (1976) *Arch. Biochem. Biophys. 175*, 717–722.

George, D. G., Barker, W. C., & Hunt, L. T. (1986) *Nucleic Acids Res. 14*, 11–15.

Greenfield, N. J., & Pietruszko, R. (1977) *Biochim. Biophys. Acta, 483*, 35–45.

Guan, K., & Weiner, H. (1987) *Prog. Clin. Biol. Res. 232*, 15–24.

Hawke, D., Yuan, P.-M., & Shively, J. E. (1982) *Anal. Biochem. 120*, 302–311.

Hempel, J., & Jörnvall, J. (1987) *Prog. Clin. Biol. Res. 232*, 1–14.

Hempel, J., Pietruszko, R., Fietzek, P., & Jörnvall, H. (1982) *Biochemistry 21*, 6834–6838.

Hempel, J., von Bahr-Lindström, H., & Jörnvall, H. (1984) *Eur. J. Biochem. 141*, 21–35.

Hempel, J., Kaiser, R., & Jörnvall, H. (1985) *Eur. J. Biochem. 153*, 13–28.

Hsu, L. C., Tani, K., Fugiyoshi, T., Kurachi, K., & Yoshida, A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 3771–3775.

Hsu, L. C., Bendel, R. E., & Yoshida, A. (1988) *Genomics 2*, 57–65.

Johansson, J., von Bahr-Lindström, H., Jeck, R., Woenckhaus, C., & Jörnvall, H. (1988) *Eur. J. Biochem. 172*, 527–533.

Jones, D. E., Jr., Brennan, M. D., Hempel, J., & Lindahl, R. (1988) *Proc. Natl. Acad. U.S.A. 85*, 1782–1786.

Jörnvall, H. (1984) in *Evolution and Tumor Pathology of the Neuroendocrine System* (Falkmer, S., Håkanson, R., & Sundler, F., Eds.) pp 165–180, Elsevier Publishers, Amsterdam.

Jörnvall, J., Persson, B., & Jeffery, J. (1987) *Eur. J. Biochem. 167*, 195–201.

Kitabatake, N., Sasaki, R., & Chiba, H. (1981) *J. Biochem. (Tokyo) 89*, 1223–1229.

Kitson, T. M. (1981) *Biochem. J. 199*, 255–258.

Kitson, T. M. (1983) *Biochem. J. 213*, 551–554.

Koivusalo, M., Aarnio, M., & Rautoma, P. (1988) in *Enzymology and Molecular Biology of Carbonyl Metabolism* (Weiner, H., & Flynn, T. G., Eds.) Liss (in press).

Lindahl, R. (1986) *Cancer Res. 46*, 2502–2506.

Lindahl, R., & Evces, S. (1984a) *J. Biol. Chem. 259*, 11986–11990.

Lindahl, R., & Evces, S. (1984b) *J. Biol. Chem. 259*, 11991–11996.

Lindahl, R., Hempel, J., Holmes, R., & Weiner, H. (1988) in *Enzymology and Molecular Biology of Carbonyl Metabolism* (Weiner, H., & Flynn, T. G., Eds.) Liss (in press).

Lipman, D. J., & Pearson, W. R. (1985) *Science 227*, 1435–1441.

Lonberg, N., & Gilbert, W. (1985) *Cell 40*, 81–90.

MacKerrell, A. D., Jr., Vallari, R. C., & Pietruszko, R. (1985) *FEBS Lett. 179*, 77–81.

Pickett, M., Gwynne, D. I., Buxton, F. P., Elliott, R., Davies, R. W., Lockington, R. A., Scazzocchio, C., & Sealy-Lewis, H. M. (1987) *Gene 51*, 217–226.

Rossmann, M. G., Liljas, A., Brändén, C.-I., & Banaszak, L. J. (1975) *Enzymes (3rd Ed.) 7*, 61–102.

Tu, G.-C., & Weiner, H. (1988a) *J. Biol. Chem. 263*, 1212–1217.

Tu, G.-C., & Weiner, H. (1988b) *J. Biol. Chem. 263*, 1218–1222.

von Bahr-Lindström, H., Hempel, J., & Jörnvall, H. (1982) *J. Protein Chem. 1*, 257–262.

von Bahr-Lindström, H., Hempel, J., & Jörnvall, H. (1984) *Eur. J. Biochem. 141*, 31–42.

von Bahr-Lindström, H., Jeck, R., Woenckhaus, C., Sohn, S., Hempel, J., & Jörnvall, H. (1985) *Biochemistry 24*, 5847–5851.

Wierenga, R. K., & Hol, W. G. J. (1983) *Nature 302*, 842–844.

Yoshida, A., Huang, I.-Y., & Ikawa, M. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 258–261.