

# Progressive Sequence Alignment and Molecular Evolution of the Zn-Containing Alcohol Dehydrogenase Family

Hong-Wei Sun and Bryce V. Plapp

Department of Biochemistry, The University of Iowa, Iowa City, IA 52242, USA

Summary. Sequences of 47 members of the Zncontaining alcohol dehydrogenase (ADH) family were aligned progressively, and an evolutionary tree with detailed branch order and branch lengths was produced. The alignment shows that only 9 amino acid residues (of 374 in the horse liver ADH sequence) are conserved in this family; these include eight Gly and one Val with structural roles. Three residues that bind the catalytic Zn and modulate its electrostatic environment are conserved in 45 members. Asp 223, which determines specificity for NAD, is found in all but the two NADP-dependent enzymes, which have Gly or Ala. Ser or Thr 48, which makes a hydrogen bond to the substrate, is present in 46 members. The four Cys ligands for the structural zinc are conserved except in (-crystallin, the sorbitol dehydrogenases, and two bacterial enzymes. Analysis of the evolutionary tree gives estimates of the times of divergence for different animal ADHs. The human class II ( $\pi$ ) and class III ( $\chi$ ) ADHs probably diverged about 630 million years ago, and the newly identified human ADH6 appeared about 520 million years ago, implying that these classes of enzymes may exist or have existed in all vertebrates. The human class I ADH isoenzymes ( $\alpha$ ,  $\beta$ , and  $\gamma$ ) diverged about 80 million years ago, suggesting that these isoenzymes may exist or have existed in all primates. Analysis of branch lengths shows that these plant ADHs are more conserved than the animal ones and that class III ADHs are more conserved than class I ADHs. The rate of acceptance of point mutations (PAM units) shows that selection pressure has existed for ADHs, implying that these enzymes play definite metabolic roles.

**Key words:** Alcohol dehydrogenase – Molecular evolution – Sequence alignment – Phylogenetic tree – Structure and function of proteins

# Introduction

Alcohol dehydrogenases (ADHs: EC 1.1.1.1) occur in a wide variety of organisms, including animals, plants, yeasts, and bacteria (Brändén et al. 1975). ADHs can be classified, according to the metal ions contained, into three groups: those with zinc, those without any metal ion, and those with iron (Jörnvall et al. 1987). These three groups are respectively represented by horse liver ADH, *Drosophila* ADH, and ADH2 from *Zymomonas mobilis*.

Many Zn-containing ADHs from different species have been characterized; they exist as dimers and tetramers, as represented by the horse and yeast enzymes, respectively. Human ADHs are dimers. They are grouped (Vallee and Bazzone 1983) into three classes: I, II ( $\pi$ ), and III ( $\chi$ ). In addition, a fourth class, human ADH6, has been identified recently (Yasunami et al. 1991). There are three genes for human class I isoenzymes, and the protein subunits coded by these genes are named  $\alpha$ ,  $\beta$ , and  $\gamma$ (Smith et al. 1971). In addition to the Zn-containing ADHs, some other proteins are related to this family as judged by their sequence identities and possible functional and structural similarities. These include liver sorbitol dehydrogenases (Jörnvall et al. 1981; Eklund et al. 1985; Karlsson et al. 1991), Escherichia coli threonine dehydrogenase (Aronson and Somerville 1989), and  $\zeta$ -crystallin from guinea pig lens (Borrás et al. 1989).

Of the Zn-containing ADHs, the enzyme from horse liver is the most extensively studied (Eklund and Brändén 1987). Its three-dimensional structure shows that the enzyme is a dimer of two identical subunits, each containing two zinc atoms and different domains for binding coenzyme and substrate. Other ADHs of this family are also thought to have these features of horse liver ADH (Eklund and Brändén 1987).

Comparisons using sequence alignments are important for studying structure-function and evolutionary relationships among members of the Zncontaining ADH family. Many sequences have been aligned to that of horse liver ADH E isoenzyme. These studies, however, have been limited by the number of sequences available. An evolutionary tree was presented based on the protein sequences of 17 members of this family (Jörnvall et al. 1987), but without detailed information on branch order. Another tree was prepared for 17 animal and plant ADHs based on their DNA sequences (Yokoyama et al. 1990). Both branch lengths and branch order were given, but the information revealed was limited due to the exclusion of many tetrameric members. We have studied 47 members of the Zn-containing ADH family using computer programs that align sequences progressively and produce a phylogenetic tree based on the alignment (Doolittle and Feng 1990; Feng and Doolittle 1990).

# **Materials and Methods**

Sequences included in this study are 42 ADHs from 26 different species, 3 sorbitol dehydrogenases from 3 different species, a threonine dehydrogenase, and  $\zeta$  crystallin. Letter codes used in this study, species names, common names, and references for these enzymes and the protein are listed in Table 1. The Genetics Computer Group sequence analysis software package (GCG, version 7.0, April 1991; Devereux et al. 1984) was used to search two protein sequence data banks, National Biomedial Research Foundation (release 29, June 1991) and SwissProt (release 19, August 1991). PAM stands for the accepted point mutations per 100 residues per 100 million years.

A set of C programs running on a VAX 6410/VMS 5.3 system, including FORMAT, SCORE, PREALIGN, TREE, PAPA3, BLEN, and MULPUB was used to generate the progressive alignment and evolutionary trees (Doolittle and Feng 1990; Feng and Doolittle 1990). The alignment uses the algorithm of Needleman and Wunsch (1970) and the minimum mutation matrix of Dayhoff et al. (1978) for the scoring. Gaps are introduced by comparing the most closely related pair of sequences and are retained by the "once a gap always a gap" rule. The evolutionary tree was based on the progressive alignment. The initial versions of the tree had some negative branch lengths, so modifications were made by switching nearby branches and regrouping some members within a cluster. This was facilitated by using the topology information of a tree generated by PAPA3. The final version of the tree had no negative branch lengths and a low percentage standard deviation.

# Results

# The Alignment

The alignment is given in Fig. 1. It is different from all previous alignments in that it was made progressively rather than pairwise. As a result, it has some gaps or insertions at positions that are different from those assigned previously. For instance, the only insertion for quail ADH appeared before 117 in this alignment instead of before 119 as shown in a previous alignment (Kaiser et al. 1990). For human ADH  $\pi$ , a Ser, a Lys, and two Asn are inserted before 115, 117, and 122 respectively, whereas all four residues were inserted as a single unit before 122 in an alignment made for 17 members of this family (Jörnvall et al. 1987). For barley ADH2, a six-residue gap was previously assigned to position 293-298 (Trick et al. 1988), whereas the current alignment introduced this gap into position 290-295. Perhaps the most striking difference observed is the assignment of a 21-residue deletion for yeast ADHs. This deletion was previously treated as a single gap and assigned to 119–139 (Jörnvall et al. 1987). The present alignment, however, treats this deletion as two gaps and assigns them to 112-127 and 135-139. We have observed that the alignment is dependent on the sequences included, as slightly different assignments for some gaps or insertions were obtained when the number of sequences changed. Nevertheless, the positions for gaps or insertions presented in the current alignment, if different from those assigned previously, should at least provide alternatives for consideration. The ultimate solutions to the position of gaps or insertions will require knowledge of the three-dimensional structures.

With the present alignment, nine residues, eight Gly and one Val, are conserved in all the sequences. These strictly conserved residues can be divided into two clusters. One has four Gly (at positions 66, 71, 77, and 86) and one Val (at 80) and is located in the substrate-binding domain; the other has four Gly (192, 201, 204, and 236) and is located in the coenzyme-binding domain. These 9 residues are among those 22 found strictly conserved when 17 members of this family were aligned (Jörnvall et al. 1987). It is interesting that Val 80, whose conservation was doubted when more sequences became available (Jörnvall et al. 1987), is strictly conserved among all the members of this family. The number of conserved residues increases to 12 when (-crystallin of guinea pig lens is excluded, with the three additional residues being Cys 46, His 67, and Glu 68, which are first or second sphere ligands to the catalytic Zn. (Asp 49, another second sphere ligand to Zn, is conserved among all the enzymes except human ADH6, which has Glu 49.) In addition to

 Table 1. List of enzymes/protein included in this study

Code	Species	Common name	References
Aeu	Alcaligenes eutrophus	A. eutrophus ADH	Jendrossek et al. 1988
Anil	Aspergillus nidulans	A. nidulans ADH1	Gwynne et al. 1987
Ani3	A. nidulans	A. nidulans ADH3	McKnight et al. 1985
Ath	Arabidopsis thaliana	Mouse-ear cress ADH	Chang and Meyerowitz 1986
Cja	Coturnix japonica	Quail ADH	Kaiser et al. 1990
CpoZ	Cavia porcellus	Guinea pig lens crystallin ζ	Borrás et al. 1989
EcaE	Equus caballus	Horse ADHE	Jörnvall 1970
EcaS	E. caballus	Horse ADHS	Park and Plapp 1991
EcaX	E. caballus	Horse ADH $\chi$	Kaiser et al. 1989
EcoT	Escherichia coli	E. coli threonine DH	Aronson and Somerville 1989
Fan	Fragaria ananassa	Strawberry ADH	Wolyn and Jelenkovic 1990
Gga	Gallus gallus	Chicken ADH	Estonius et al. 1990
HsaA	Homo sapiens	Human ADH $\alpha$	Ikuta et al. 1986; von Bahr-Lind- ström et al. 1986
HsaB	H. sapiens	Human ADH $\beta$	Ikuta et al. 1985
HsaC	H. sapiens	Human ADH $\gamma$	Ikuta et al. 1986; Höög et al. 1986
HsaG	H. sapiens	Human sorbitol (glucitol) DH	Karlsson et al. 1991
HsaP	H. sapiens	Human ADH $\pi$	Höög et al. 1987
HsaX	H. sapiens	Human ADH $\chi$	Kaiser et al. 1988
Hsa6	H. sapiens	Human ADH6	Yasunami et al. 1991
Hvu1	Hordeum vulgare	Barley ADH1	Good et al. 1988
Hvu2	H. vulgare	Barley ADH2	Trick et al. 1988
Hvu3	H. vulgare	Barley ADH3	Trick et al. 1988
Kla	Kluyveromyces lactis	K. lactis ADH	Saliola et al. 1990
MacA	Macaca mulatta	Rhesus monkey ADH $\alpha$	Light et al. 1992
MmuA	Mus musculus	Mouse ADHA	Edenberg et al. 1985
MmuX	M. musculus	Mouse ADH $\chi$	Edenberg et al. 1991
OarG	Ovis aries	Sheep sorbitol (glucitol) DH	Karlsson et al. 1991
Osal	Oryza sativa	Rice ADH1	Xie and Wu 1989
Osa2	O. sativa	Rice ADH2	Xie and Wu 1990
Pam	Pennisetum americanum	Pearl millet ADH	Ha et al. 1989
PhaB	Papio hamadrysa	Baboon ADH $\beta$	Trezise et al. 1989
Psa	Pisum sativum	Garden pea ADH	Llewellyn et al. 1987
RnoA	Rattus norvegicus	Rat ADHA	Crabb and Edenberg 1987
RnoG	R. norvegicus	Rat sorbitol (glucitol) DH	Karlsson et al. 1991
RnoX	R. norvegicus	Rat ADH $\chi$	Julià et al. 1988
Rpe	Rana perezi	Frog ADH	Cederlund et al. 1991
Sce1	Saccharomyces cerevisiae	Yeast ADH1	Bennetzen and Hall 1982
Sce2	S. cerevisiae	Yeast ADH2	Russell et al. 1983
Sce3	S. cerevisiae	Yeast ADH3	Young and Pilgrim 1985
Spo	Schizosaccharomyces pombe	S. pombe ADH	Russell and Hall 1983
Stu	Solanum tuberosum	Potato ADH	Matton and Brisson 1990
Tae	Triticum aestivum	Wheat ADH	Mitchell et al. 1989
Tbr	Thermoanaerobium brockii	T. brockii ADH	Peretz and Burstein 1989
Tre	Trifolium repens	White clover ADH	Ellison 1989
Zmal	Zea mays	Maize ADH1	Dennis et al. 1985
Zma2	Z. mays	Maize ADH2	Dennis et al. 1985
Zmo1	Zymomonas mobilis	Z. mobilis ADH1	Keshav et al. 1990

those strictly conserved residues, there are positions where only a few types of amino acid residues are found. These residues, together with those strictly conserved ones, are given in Table 2.

The number of conserved residues increases greatly when only animal and plant ADHs are considered, with the total number of conserved residues being 86. Even more residues are found conserved when plant and animal ADHs are treated separately. As can be seen in Fig. 1 (indicated by a and p), 116

(31%) residues are conserved among 18 animal ADHs and 213 (56%) residues are conserved among 14 plant ADHs. Apparently, these plant ADHs diverged less during evolution than the animal ones. In addition, the plant ADHs seem to have more aromatic residues conserved. Specifically, there are more Phe (14 out of 19 as compared to 6 out of 17 for animal ADHs), Trp (3 out of 3 as compared to 1 out of 3), and Tyr (3 out of 6 as compared to 1 out of 5) conserved among these plant ADHs.

# The Evolutionary Tree

A phylogenetic tree of this family was produced on the basis of the progressive alignment; it is given in Fig. 2. The branch order for human class I ADHs in the current tree is different from that proposed previously, where different methods were employed and a limited number of sequences were included in the analysis (Ikuta et al. 1986; Trezise et al. 1989; Yokoyama et al. 1990). All three possible alternative ways of arrangement for the human class I ADHs were tested and the present order is the best as judged by the criteria for a better tree. In addition, the current order is supported by the fact that the model of the  $\gamma$  isoenzyme is most similar to the x-ray structure of horse ADH E (Eklund et al. 1987).

The 47 members are classified into two large groups. Group I includes all the dimeric ADHs of animals and plants, and group II has all the tetrameric ones including  $\zeta$  crystallin. Group I is further divided into two major clusters, one for 18 animal ADHs and the other for 14 plant ADHs. Among the 18 animal ADHs, human class II enzyme diverged first, followed by class III enzymes and human ADH6, and then the class I ADH from frog diverged from the rest of class I enzymes of other species. The distances from points A, B, and C were given in Table 3 for animal and plant ADHs. They were calculated by adding up the corresponding branch lengths in the evolutionary tree.

In addition to the alignment and the evolutionary tree, the percent identities for all the possible pairs of these 47 enzymes/protein were calculated. The SCORE program produced the percent identities based on pairwise alignment. The TREE program, on the other hand, calculated the percent identities based on the progressive alignment. These two sets of percent identities are given in Table 4; the data show that these two sets differ only for distantly related pairs.

# Discussion

# The Alignment

Sequence alignments are generally valuable as it is unlikely that the three-dimensional protein structures (less than 500 currently) will be determined for most of the known protein sequences (over 20,000 currently). The alignment reveals functionally important residues and usually is the first step in building a reasonable model of the three-dimensional structure. Various sequences of ADHs have been aligned previously (e.g., Jörnvall et al. 1987; Eklund et al. 1990; Xie and Wu 1990). Most of these were produced by aligning the sequence with that of horse liver ADH E, and by applying subjective criteria when a gap or an insertion had to be assigned. In contrast, the alignment presented here, which was used to construct the evolutionary tree, was produced progressively based on objective criteria.

# Requirements for an Enzymatically Functional Member

The alignment of 47 members of the Zn-containing ADH family yielded the current minimal requirements for a functional enzyme of this family. Only a few residues are strictly conserved among all 47 members, probably reflecting functional and structural diversities of the different ADHs. The fact that most of the strictly conserved residues are Gly indicates that they are located at crucial positions where a side chain would disrupt a structure that is required for a functional ADH. A stereoview of the positions of these strictly conserved residues in the three-dimensional structure of horse ADH E is illustrated in Fig. 3. It is interesting that the 9 strictly conserved residues are clustered, rather than distributed over the sequence, with the 5 in the catalytic domain being within a 21-residue fragment and the 4 in the coenzyme-binding domain being within a 45-residue fragment. This suggests that these strictly conserved residues are involved in forming compact cores for the two functional domains.

In addition to the nine strictly conserved residues, four more are conserved if  $\zeta$  crystallin, an NADPH-quinone oxidoreductase (Rao et al. 1992), is not included. These four residues are all related to Zn binding. Cys 46 and His 67 are ligands to the catalytic Zn (Eklund and Brändén 1987). Asp 49 (except human ADH6, which has Glu 49) and Glu 68 are in the second sphere of Zn ligands and have been shown to affect the electrostatic environment near the catalytic Zn for yeast ADH1 (Ganzhorn and Plapp 1988). The third catalytic Zn-binding ligand (at position 174) may also be considered conserved for all but  $\zeta$  crystallin if we assume Glu and Asp could serve as Zn-binding ligands for sorbitol dehydrogenases, threonine dehydrogenase, and ADHs from Thermoanaerobium brockii and Alcaligenes eutrophus. In other zinc enzymes, such as carboxypeptidase (Rees et al. 1981) and thermolysin (Holmes and Matthews 1981), Glu is a ligand. In addition. Asp has recently been identified as a Zn ligand in E. coli alkaline phosphatase (Kim and Wyckoff 1991). Thus, it seems that one major event during the evolution of ADHs is their acquisition of the capability to bind a Zn in their active sites.

There are two other residues that seem to be functionally relevant. One of these is Asp 223, which is conserved in all but  $\zeta$  crystallin and the ADH from *T. brockii* (which use NADP instead of NAD). Asp 223 has been suggested to be important in deter526

	10 20		30	40	50	60	70	80		90 100	110	120	130
					*		* *	* *	*				
	44 444 A 4	a	88888	888	t t t a	a	ttt t aaa aa	tt aa aa	t AA A		A		
Wh e P	P P P PPP PPPP	P P PP				PP PP P PPP	P PPPP PP	PPPPPpp	p ppp	PP PPPPP P 1	P P PP P P PP	PPP PP	P P PPP
Hsab	STAGKVIKCKAAVLWEVKKP	FSIE DV	EVAPPKAY	EVRIKHVAV	GICRTDDHVV	SGNLV SPLPA	ILGHEAAGI	VESVGEGV	TTVKPGDKV	IPLFTPQCGKCR	CKSPEG NYCVKN CKNPES NYCLKN	DL SNPRG	TLODGTRRF
MacA HaaA	STAGKVIKCKAAVLWEVMKP STAGKVIKCKAAVLWELKKP	FSIE DV FSIE EV	/EVAPPKAY /EVAPPRAH	EVRIKNVTV	GICGTDDHVV	SGTMV TPLPV SGTMV TPLPV	ILGHEAAGI	VESVGEGV	TTVEPGDKV	IPLALPQCGKCR1	CKTPER NYCLKN	DV SNPRG	TLODGTSRF
HsaC	STAGKVIKCKAAVLWELKKP	FSIE EV	/EVAPPKAI	EVRIKNVA	GICRSDEHVV	SGNLV TPLPV	ILCHEAAGI	VESVGEGV	TTVKPGDKV	1PLFTPQCGKCR	CKNPES NYCLKN	DL GNPRG	TLQDGTRRF
EcaS EcaE	STAGKVIKCKAAVLWEQKKP STAGKVIKCKAAVLWEEKKP	FSIE EV FSIE EV	/EVAPPKAH /EVAPPKAH	IEVRIKMVA# IEVRIKMVAT	AGICRSDDHVV IGICRSDDHVV	SGTLV APLPV SGTLV TPLPV	IAGHEAAGI IAGHEAAGI	VESIGEGV VESIGEGV	TTVRPGDKV	IPLFIPQCGKCSV IPLFTPQCGKCRV	/CKHPEG NLCLK /CKHPEG NFCLKN	NL SMPRG	TMQDGTSRF
RnoA	STAGKVIKCKAAVLWEPHKP	FTIE DI	EVAPPKA	EVRIKMVAT	GVCRSDDHAV	SGSLF TPLPA	VLGHEGAGI	VESIGEGV	TCVKPGDKV	IPLFSPQCGKCR	CKHPES NLCCQT	KNL TQPKG	ALLDGTSRF
Cja	STAGKVIKCKAAVLWELNKP	FSLE EV	I EVAPPKAP /EVAPPKAP	EVRIKIVA	IGICRSDDHVV	TGALA MPFPV	ILCHEAAGV	VESVGEGV	TLLKPGDAV	IPLFSPQCGECR	CLATES NFCSRS	DL LMPRG DLSSSPTG	TLREGTSRF LMADGTTRF
Gga	STVGKVIKCKAAVLWEANKP	FSLE EV	VEVAPPKA	EVRIKIVAT	IGICRSDDHVV	TGALA MPFPI SCALSDWEFPU	ILCHEAAGV	IESVGEKV	TSLEPGDAV	IPLEVPQCGECRS	CLSTKG NLCIKN	DLSSSPTG	LMADGTTRF
Hsa6	STTGQVIRCKAAILWKPGAP	FSIE EV	EVAPPKA	EVRIKVVA1	IGLCGTENKVL	GSKHLEVLYPT	ILGHEGAGI	VESIGEGV	STVKPGDKV	ITLFLPQCGECTS	CLNSEG NFC1	QFKQSKTQ	LMSDGTSRF
RnoX MmiX	ANQVIRCKAAVAWEAGKP ANOVIRCKAAVAWEAGKP	LSIE EI	I EVAPPQAH I EVAPPKAH	EVRIKIIAT	TAVCHTDAYTL TAVCHTDAYTL	SGADPEGCFPV SGRDPEGCFPV	ILCHEGAGI	VESVGEGV	TKLKAGDTV TKLKAGDTV	IPLYIPQCGECKI IPLYIPOCGECKI	CLNPKT NLCQKI	RV TOCKG	LMPDGTSRF
HsaX	ANEVIKCKAAVAWEACKP	LSIE EI	EVAPPKA	EVRIKIIA	TAVCHTDAYTL	SGADPEGCFPV	ILCHEGAGI	VESVGEGV	TKLKAGDTV	IPLYIPQCGECK	CLNPKTNLCQKI	RV TQGKG	LMPDGTSRF
EcaX HsaP	SAEVIKCKAAVAWEAGKP GTKGKVIKCKAAIAWEAGKP	LCIE EV	VEVAPPKAH VEVAPPKAH	EVRICIA	TAVCHTDAYTL TSLCHTDATVI	DSKFEGLAFPV	ILGHEGAGI IVGHEAAGI	VESVGEGV	TKLKAGDTV TNVKPGDKV	IPLYIPQCGECKI IPLYAPLCRKCKI	CLNPQTNLCQKI CLSPLTNLCGKI	RT TQGKG SNLKSPASDO	LMPDGTSRF
Tae	MATAGKVIECKAAVAWEAGKP	LSIE EV	VEVAPPHAN	EVRVKILY	TALCHTDVYFW	EAKGQTPVFPR	ILCHEAGGI	VESVGEGV	TELVPGDHV	LPVFTGECKDCAL	CKSEES NLCDLL	RINVDRGV	MICDGQSRF
Hvu2	MATAGKVIKCKAAVAWEAGKP MATAGKVIKCKAAVAWEAGKP	LSTE EV	VEVAPPQAN	EVROKILI	TALCHIDVIFW	EANGQTPVFFF	ILGHEAGGI	VESVGEGV	TELVPGDHV	LPVFTGECKBCAP	ICKSEES NLCDLL ICMSEES NLCDLL	RINVDRGV	MICDGQSRF MIDDGQSRF
Zma2	ATAGKVIKCRAAVTWEAGKP	LSIE EV	VEVAPPOAN	EVRIKILY	TALCHTDVYFW	EAKGQTPVFPR	ILGHEAGGI	VESVGEGV	TDVAPGDHV	LPVFTGECKECAL	ICKSEES NMCDLL	RINVDRGV	MIGDGKSRF
Zmal	ATAGKVIKCKAAVAWEAGKP	LSIE EV	VEVAPPOA	EVRVKILF	ISLCHTDVYFW	EAKGQTPVFPR	IFGHEAGGI	LESVGEGV	TDVAPGDHV	LPVFTGECKECA	ICKSAES NMCDLL	RINTDRGV	MIADGKSRF
Pan Hyul	MATAGKVIKCKAAVAVEAGKP MATAGKVIKCKAAVAVEAGKP	LSIE EV	VEVAPPOAN	EVRVKILY EVRVKILF	ISLCHTDVYFW ISLCHTDVYFW	EAKGQTPVFPR EAKGOIPMFPR	IFGHEAGGI IFGHEAGGI	IESVGEGV	TDVAPGDHV	LPVFTGECKECPI LPVFTGECKECPI	ICKSAES NMCDLL	RINTVRGV	MIGDGKSRF
Osal	MATAGKVIKCKAAVAWEAGKP	LSIE EV	VEVA KE	EVRVKILF	ISLCHTDVYFW	EAKGQTPVFPR	IFGHEAGGI	VESVGEGV	TDLAPGDHV	LPVFTGECKECA	ICKSAES NHCDLL	RINTDRGV	MIGDGKSRF
Stu Fan	MSTTVGQVIRCKAAVAWEAGKP MSSTEGKVICCRAAVAWEAGKP	LVME EV	VDVAPPQKI VEVAPPHPI	EVRLKILY:	ISLCHTDVYFW ISLCHTDVYFW	EAKGQNPVFPR	ILGHEAAGI	VESVGEGV VESVGEGV	TELAPGDHV TDLKAGDHV	LPVFTGECKDCA1 LPVFTGECKECD1	ICKSEES NMCSLL ICKSEES NMCDLL	RINTDRGV	MINDGQSRF
Tre	MSNTAGQVIKCRAAVAWEAGKP	LVIE EV	VEVAPPQAC	GEVRLKILFT	TSLCHTDVYFW	EAKGOTPLEPR	IFGHEAGGI	VESVGEGV	THLKPGDHA	LPVFTGECGECPH	ICKSEES NMCNLL	RINTDRGV	MINDNKSRF
Psa Ath	MSNTVGQIIKCRAAVAWEAGKP STTGOIIRCKAAVAWEAGKP	LVIE EV	VEVAPPQAC VEVAPPOKI	GEVRLKILF HEVRIKILF	TSLCHTDVYFW TSLCHTDVYFW	EAKGQTPLFPR	IFGHEAGGI IFGHEAGGI	VESVGEGV	THLEPGDHA	LPVFTGECGECPH	ICKSEES NHCDLL ICQSEES NHCDLL	RINTDRGV	MINDNKSRF MIHDGESRF
Ani3	SVPEVQWAQVVEKAGTP	PVYK Q	VPVPKPGPI	DEILVKMRY	SGVCHTDLHAM	KGDWPLPSKNPL	IGGHEGAGV	VVAKGELVK	DEDFKIGDRA	GIKWLNGSCLSCEN	CHQADE PLC		PHA
Anii Scel	MCIPTNQWAQVAEKVGGP SIPETQKGVIFYESHGK	LEYK DI	I PVPKPGPI	NELLINVKY	SGVCHIDLHAM	HGDWPLPVKLPL	VGGHEGAGI	VVGMGELV	KGWKIGDYA	GIKWLNGSCGECEI	CELGNESNC		PHA
Sce2	SIPETQKAIIFYESNGK	LEHK DI	I PVPKPKP	NELLINVKY	SGVCHTDLHAW	HGDWPLPTKLPL	VGGHEGAGV	VVGMGENV	KGWKIGDYA	GIKWLNGSCMACE	CELGNE SNC		PHA
KIA Sce3	QSTAAIPKTQKGVIFYENKGK	LHYK D	I PVPKPKA I PVPEPKPI	NEILINVKY	SGVCHTDLHAW	HGDWPLPVKLPL	VGGHEGAGV	VVKLGSNV	KGWKVGDLA	GIKWLNGSCHSCEI	CESCHESNC		PDA
Spo	TIPDKQLAAVFHTHGGP	ENVKFE EV	VPVAEPGQI	DEVLVNIKY	TGVCHTDLHAL	QGDWPLPAKMPL NGD FGDFTGP	IGGHEGAGV	VKVGAGV	TRLKIGDRV	SVKWMNSSCGNCEN	CNKAEETIC		PHI
CpoZ	ATGOKLM RAIRVFEFGGP	EVLKVQSD	VAVPIPKD	HQVLIKVHA	CGINPVETYIR	SGTYTRIPLLPY	TPGTDVAGV	VESIGNDV	SAFKKGDRV	SVAWFFRGGGhoe	07300E110		F
OarG	AAAKPENI,SLVVHGPGD	IRLE N	YPIPEPGP VPIPEPCP	NEVLLKMHS	VGICGSDVHYW	CRIGDFVVKKPM	VLGHEASGT	VVKVGSLV	RHLQPGDRV KHIKPGDRV	AIQPGAPRQTDE	CKIGRYNLSPTI	F	F
RnoG	AAPAKGENLSLVVHGPGD	IRLE N	YPIPELCPI	NDVLLKNHS	VGICGSDVHYW	HGRIGDFVVKKPM	VLGHEAAGT	VTKVGPMV	KHLKPGDRV	AIEPGVPREIDE	CKIGRYNLTPSI	F	F
EcoT Thr	MKALS KLKAEEG MKGFANT, SIGK	IWMT DV	VPVPELGHI KEKPAPGPI	NDLLIKIRK FDAIVRPLA	TAICGTDVHIYI VAPCTSDIHT	W DEWSQKTIPVPM VFEGAIGERHNM	VVGHEYVGE ILGHEAVGE	VVGIGQEV VVEVGSEV	KGFKIGDRV	SGEGHITCGHCRN VVPAITPDWRTSE	CRCGRTHLCRNT	L	I AGW
Aeu	MTAMMKAAVFV EPGR	IELA D	KPIPDIGP	NDALVRITT	TTICCTDVH	ILKGEYPVAKGL	TVGHEPVGI	LEKLGSAV	TGYREGQRV	IAGAICPNFNSYA	AQDGVASQDGSY	LMASGQCG	CHGYKATAGW
Aeu	MTANNKAAVFV EPGR 140 150	IELA DI 16	KPIPDIGPI 50	NDALVRITT	TTICGTDVH 180	ILKGEYPVAKGL	TVGHEPVGI 21	LEKLGSAV	TGYREGQRV 220 ;	IAGAICPNFNSYA	AQDGVA SQDGSY 250	LMASGQCG 260	270
Aeu	NTAMMKAAVPV EPGR 140 150	IELA DI 16	KPIPD1GP 50	NDALVRITT	TTICCTDVH 180	ILKGEYPVAKGL 190 200 * #	TVGHEPVGI 2: * * # #	LEKLGSAV	TGYREGQRV	IAGAICPNFNSYA 230 240 * #	AQDGVA SQDGSY 250	LMASCQCG 260	CHGYKATAGW 270
Aeu	нтаникалуру Ерск 140 150	IELA DI 16	KPIPDIGP 50	NDALVRITT	TTICGTDVH 180	ILKGEYPVAKGL 190 200 * t a aa aa aaa	TVGHEPVGI 2: * * # # t t sasaa	LEKLGSAV 10 : # t aaa aaa	TGYREGQRV	IAGAICPNFNSYA 230 240 * t t	AQDGVASQDGSY 250 a a a a	LMASGQCG 260	270 a a
Aeu	нтаникалугу ерск 140 150 а а сод со р р р рррррр ррр	IELA DI 16 aa P.P. PF	KPIPDIGP 50 PP PP	NDALVRITT 170 aaa aa pppp ppp	TTICGTDVH 180 aaa a PP PP	ILKGEYPVAKGL 190 200 * # t a aa aa aaa p p pp ppp	TVGHEPVGI 2: * * # # t t sassa ppppp p	IEKLGSAV 10 # t aaa aaa PPP PPP	TGYREGQRV 220 ; aa a a ; pppp p	IAGAICPNFNSYA 230 240 * # t aa a aa a P PP F	AQDGVASQDGSY 250 a a a a P PP PP P	LMASGQCG 260	270 270 a a p pppp DDDDDUC
Aeu PhaB HsaB	НТАНИКАЛУFV ЕРСК 140 150 а а ала ал р р р рр рррррр ррр тсяскугныFустутSyTVV тсяскугныFустутSyTVV	IELA DI 16 PP PP DEN AVAR DEN AVAR	KPIPDIGP 50 50 PP PP KIDA ASPL KIDA ASPL	NDALVRITT 170 aaa aa pppp ppp EKVCLIGCG EKVCLIGCG	TTICGTDVH 180 aaa a pp pp FSTGYGSAVNV, FSTGYGSAVNV,	ILKGEYPVAKGL 190 200 * t a aa aa aaa p p pp ppp MVTPGSVCAVFGL KVTPGSTCAVFGL	TVGHEPVGI 2: * * t t aaaaa ppppp p GGVGLSAVY GGVGLSAVY	IEKLGSAV 10 # t aaa aaa ppp ppp IGCKAAGAA IGCKAAGAA	TGYREGQRV 220 ; aa a a ; pppp p RIIAVDINKDI RIIAVDINKDI	IAGAICPNFNSYA 230 240 * t aa a aa a p pp r crakakELGATECI crakakELGATECI	AQDGVASQDGSY 250 a a a a p pp pp p p p pp pp pp r N PQDYKKPIQEY N PQDYKKPIQEY	LMASGQCG 260 a a a pp /LKENTDGGV/ /LKENTDGGV/	270 270 a a p pppp DFSFEVIG DFSFEVIG
Aeu PhaB HsaB HacA	HTANHKAAVFV         EFCR           140         150           a         a         aa         aa           p         r	IELA DI 16 P P PP DEN AVAR DEN AVAR DEN AVAR	KPIPDIGP 50 50 PP PP KIDA ASPL KIDA ASPL KIDA ASPH	NDALVRITT 170 aaa aa pppp ppp EKVCLIGCG EKVCLIGCG EKVCLIGCG	TTICGTDVH 180 aaa a pp pp FSTGYGSAVNV, FSTGYGSAVNV, FSTGYGSAVNV, FSTGYGSAVNV,	ILKGEYPVAKGL 190 200 * # t a aa aa aaa p p pp ppp kvyrpcsvcAvFGL kvyrpcstCAvFGL kvyrpcstCAvFGL	TVGHEPVGI 2: * * t t sassas ppppp p GGVGLSAVM GGVGLSAVM GGVGLSAVM	LO # t aaa aaa PPP PPP IGCKAAGAAI IGCKAAGAAI IGCKAAGAAI	TGYREGQRV 220 : pppp p RIIAVDINKDI RIIAVDINKDI RIIAVDINKDI	IAGAICPNPNSYA 230 240 * t aa a aa a p pp t (FAKAKELGATECI (FAKAKELGATECI (FAKAKELGATECI (FAKAKELGATECI	AQDGVASQDGSY 250 a a a ai P PP PP P N PQDYKKPIQEY N PQDYKKPIQEY N PQDYKKPIQEY	LMASCQCG 260 260 260 260 260 260 260 260 260 260	270 270 P PPPP DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG
PhaB HsaB MacA HsaA HsaC	A CARACITY EFCR 140 150 A CARACITAL AND A CARACITY A CARACI	IELA DI 16 P P PP DEN AVAR DEN AVAR DEN AVAR DEN AVAR	KPIPDIGP 50 pp pp KIDA ASPL KIDA ASPL KIDA ASPL KIDA ASPL KIDA ASPL	NDALVRITT 170 aaa aa pppp ppp EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG	180 180 180 PP PP FSTGYGSAVNV, FSTGYGSAVNV, FSTGYGSAVKV, FSTGYGSAVKV,	ILKGEYPVAKGL 190 200 * # t a aa aa aaa p p pp ppp kvrpcsvcavFcL kvrpcstcavFcL kvrpcstcavFgL kvrpcstcavFgL	TVGHEPVGI 2: * * # # t t aaaaa ppppp p GGVGLSAVI GGVGLSAVI GGVGLSAVI GGVGLSAVI	LEKLGSAV # t aaa aaa ppp ppp IGCKAAGAAI IGCKAAGAAI IGCKAAGAAI IGCKAAGAAI	AG A A A A A A A A A A A A A A A A A A	IAGAICPNPNSYA 230 240 * t aa a aa a p pp c (FAKARELGATECI (FAKARELGATECI (FAKARELGATECI (FAKARELGATECI (FAKARELGATECI	AQDGVASQDGSY 250 250 P PP PP P N PQDYKKPIQEY N PQDYKKPIQEY N PQDYKKPIQEY N PQDYKKPIQEY N PQDYKKPIQEY	LMASCQCG 260 260 260 20 20 20 20 20 20 20 20 20 20 20 20 20	270 270 p pppp DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG
Aeu PhaB HsaB MacA HsaA HsaC EcaS	A CARACITY EFCR 140 150 a a a aa aa p p p pp pppppp ppp TCRCKPIHHFUCTSTFSQTTW TCRCKPIHHFLCTSTFSQTTW TCRCKPIHHFUCTSTFSQTTW TCRCKPIHHFUCTSTFSQTTW TCRCKPIHHFUCTSTFSQTTW	IELA DI 16 P P PP DEN AVAR DEN AVAR DEN AVAR DEN AVAR DEN AVAR DEN AVAR	KPIPDIGP 50 pp pp KIDA ASPL KIDA ASPL KIDA ASPL KIDA ASPL KIDA ASPL KIDA ASPL	NDALVRITT 170 aaa aa pppp ppp EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG	180 180 aaa a pp pp PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV,	ILKGEYPVAKGL 190 200 * t a aa aa aa p p pp ppp ukvTPGSVGAVFGL ukvTPGSTCAVFGL ukvTPGSTCAVFGL ukvTPGSTCAVFGL ukvTCGSTCAVFGL ukvTCGSTCAVFGL ukvTCGSTCAVFGL ukvTCGSTCAVFGL ukvTCGSTCAVFGL ukvTCGSTCAVFGL	TVGHEPVGI 2: * * # # t t aaaaa ppppp p GGVGLSAVM GGVGLSAVM GGVGLSAVM GGVGLSAVM GGVGLSAVM GGVGLSVVM GGVGLSVVM GGVGLSVVM	I EKLGSAV # t aaa aaa PPP PPP IGCKAAGAAI IGCKAAGAAI IGCKAAGAAI IGCKAAGAAI IGCKAAGAAI	TGYREGQRV 220 ; PPPP P RILAVDINKDI RILAVDINKDI RILAVDINKDI RILAVDINKDI RILAVDINKDI RILAVDINKDI RILAVDINKDI	IAGAICPNPNSYA 230 240 * t aa a aa p pp r (FAKAKELGATECI (FAKAKELGATECI (FAKAKELGATECI (FAKAKELGATECI (FAKAKELGATECI (FAKAKELGATECI (FAKAKELGATECI (FAKAKELGATECI (FAKAKELGATECI	AQDGVA.SQDGSY 250 p pp pp pt n PQDYKKPTQEN n PQDYKKPTQEN n PQDYKKPTQEN n PQDYKKPTQET n PQDYKKPTQET n PQDYKKPTQET	LMASCQCG 260 260 260 260 260 260 260 260 260 260	270 a a p ppp DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG
Aeu PhaB HsaB MacA HsaA HsaC EcaS EcaE EcaS EcaE	ATAMERAAVFV EFCK 140 150 a a aaa aa p p p pp ppp ppp pp TCRCKPIHHFUCTSTFSQTTW TCRCKPIHHFUCTSTFSQTTW TCRCKPIHHFUCTSTFSQTTW TCRCKPIHHFUCTSTFSQTTW TCRCKPIHHFUCTSTFSQTTW TCRCKPIHHFUCTSTFSQTTW	IELA D 16 16 16 16 16 16 16 16 16 16	KPIPDIGP 50 50 PP PP KIDA ASPL KIDA ASPL KIDA ASPL KIDA ASPL KIDA ASPL KIDA ASPL KIDA ASPL	NDALVRITT 170 aaa aa pppp ppp EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG DKVCLIGCG	180 180 aaa a PP PP PSTGYGSAVNV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV,	ILKGEYPVAKGL 190 200 * t t a aa aaa aaa P P PP PPP NKVTPGSTCAVFGL NKVTPGSTCAVFGL NKVTPGSTCAVFGL NKVTQSSTCAVFGL NKVTQSSTCAVFGL NKVTQSSTCAVFGL	TVGHEPVGI 2: * * # # t t aaaaa ppppp p GGVGLSAVM GGVGLSAVM GGVGLSAVM GGVGLSVVM GGVGLSVVM GGVGLSVVM GGVGLSVVM	I EKLGSAV t aaa aaa ppp ppp IGCKAAGAAI IGCKAAGAAI IGCKAAGAAI IGCKAAGAAI IGCKAAGAAI IGCKAAGAAI IGCKAAGAAI	TGYREGQRV 220 : PPPP P RIIAVDINKDI RIIAVDINKDI RIIAVDINKDI RIIAVDINKDI RIIGVDINKDI RIIGVDINKDI KIIAVDINKDI	IAGAICPNFNSYA 230 240 * t t sa a a a a p pp f craxarElGATECI (rakarELGATECI (rakarELGATECI (rakarELGATECI (rakarELGATECI (rakarELGATECI (rakarELGATECI (rakarELGATCI)	AQDGVA.SQDGSY 250 250 0 P P P P P 0 P P P P 0 PQDYKKPIQET 0 PQDYKKPIQET 0 PQDYKKPIQET 0 PQDYKKPIQET 0 PQDYKKPIQET 0 PQDYKKPIQET	LMASCQCG 260 260 27 27 27 27 27 27 27 27 27 27 27 27 27	270 a a pppp Drsfruig Drsfruig Drsfruig Drsfruig Drsfruig Drsfruig Drsfruig Drsfruig Drsfruig Drsfruig Drsfruig
PhaB HsaB HsaB HsaA HsaC EcaS EcaE RnoA Hua	ATAMERAAVFV EFCR 140 150 a a aaa aa p p p pp ppppp pp 1 1CRCKFIHEFUCTSTFSQTTVV 1CRCKFIHEFUCTSTFSQTTVV 1CRCKFIHEFUCTSTFSQTTVV 1CRCKFIHEFUCTSTFSQTTVV 1CRCKFIHEFUCTSTFSQTTVV SCRCKFIHEFUCTSTFSQTTVV SCRCKFIHEFUCTSTFSQTTVV SCRCKFIHEFUCTSTFSQTTVV	IELA DI 16 16 P P PF DEN AVAN DEN AVAN DEN AVAN DEN AVAN DEN AVAN DEN AVAN DEN AVAN DEN SVAN DEN SVAN DEN SVAN DEN SVAN	KPIPDIGP 50 50 FP PP KIDA ASPL KIDA ASPL KIDA ASPL KIDA ASPL KIDA ASPL KIDA ASPL KIDA ASPL KIDA ASPL KIDA ASPL	ADALVRITT 170 aaa aa pppp ppp EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG DKVCLIGCG DKVCLIGCG DKVCLIGCG	180 180 180 PP PP PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV,	ILKGEYPVAKGL 190 200 * t t a aa aaa aaa P P PP PPP NKVTPGSTGAVFGL NKVTPGSTGAVFGL NKVTQGSTGAVFGL NKVTQGSTGAVFGL NKVTQGSTGAVFGL NKVTQGSTGAVFGL NKVTQGSTGAVFGL NKVTQGSTGAVFGL NKVTQGSTGAVFGL	TVGHEPVGI 2: * * # # t t aaaaa pPPPP P GGVGLSAVI GGVGLSAVI GGVGLSAVI GGVGLSAVI GGVGLSVVI GGVGLSVVI GGVGLSVVI GGVGLSVVI GGVGLSVVI GGVGLSVVI	IEKLGSAV # t aaa aaa ppp ppp GCKAAGAA IGCKAAGAA IGCKAAGAA IGCKAAGAA IGCKAAGAA IGCKAAGAA	TGYREGQRV 220 : PPPP P RIIAVDINKDI RIIAVDINKDI RIIAVDINKDI RIIAVDINKDI RIIGVDINKDI RIIGVDINKDI RIIGVDINKDI RIIAVDINKDI RIIAVDINKDI	IAGAICPNFNSYA 230 240 * t t sa a a a p pp f craxarElGATECI (raxarElGATECI (raxarElGATECI (raxarElGATECI (raxarElGATECI (raxarElGATECI (raxarElGATECI (raxarElGATECI (raxarElGATECI (raxarElGATECI)	AQDGVA.SQDGSY 250 250 0 P0 P0 P1 0 P007KKF1025 0 P007KKF1025 0 P007KKF1025 0 P007KKF1025 0 P007KKF1025 0 P007KKF1025 0 P007KKF1025 0 P007KKF1025 0 P007KKF1025 0 P007KKF1025	LMASCQCG 260 260 260 27 27 27 27 27 27 27 27 27 27 27 27 27	270 a a p pppp DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG
PhaB HsaB HsaB MacA HsaC EcaS EcaE RnoA MuuA Cja Gga	ATAMRKAAVFV EFCR 140 150 a a aaa aa p p p p pp pppp pp 1 1CGKKPIHHFUCTSTFSQTTW 1CGKKPIHHFLOSTFSQTTW 1CGKKPIHHFLOSTFSQTTW 1CGKKPIHHFLOSTFSQTTW 1CGKKPIHHFLSTSTFSQTTW 3CGKKPIHHFLSTSTFSQTTW 3CKKKAIHHFLSTSTFSQTTW 1CKKKAIHHFUCTSTFTETTW	IELA DI 16 16 P P PF DEN AVAR DEN AVAR DEN AVAR DEN AVAR DEN AVAR DEN AVAR DEN AVAR DEN AVAR DEN AVAR HET AAAA HET AAAA	KPIPDIGP 50 50 50 50 50 50 50 50 50 50 50 50 50	ADALURITT 170 aaa aa pppp ppp EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG DKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG	180 180 180 PP PP PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVLQT,	ILKGEYPVAKCI 190 200 * # t t a aa aa aa p p pppp pp pp pp pp ppp p	TVGHEPVGI 2: * * t t saaaa ppppp p GGVGLSAVI GGVGLSAVI GGVGLSAVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI	LEKLGSAV # t aaa aaa ppp ppp iggCkaaGaai iggCkaaGaa	TGYREGQRV 220 ; pppp p RILAVDINKDI RILAVDINKDI RILAVDINKDI RILGVDINKDI RILGVDINKDI RILGVDINKDI RILGVDINKDI RILAVDINKDI RILAVDINKDI	IAGAICPNFNSYA 230 240 * t t RAAXELGATEOI (FAKAKELGATEOI (FAKAKELGATEOI (FAKAKELGATEOI (FAKAKELGATEOI (FAKAKELGATEOI (FAKAKELGATEOI (FAKAKELGATEOI (FAKAKELGATEOI (FAKAKELGATEOI (FAKAKELGATEOI (FAKAKELGATEOI	AQDGVA.SQDGSY 250 250 0 P0 P0 P1 0 P007KKP1025 0 P007KF105 0 P007 0 P007KF105 0 P007 0	LMASCQCG 260 260 Pp JLKENTDGGVI JLKENTDGGVI JLKENTDGGVI JLKENTDGGVI JLKENTDGGVI JLLENSNGGVI JLLENTGGVI JLLENTGGGVI JLLENTGGGVI	270 a a p pppp DTSSFEVIG DTSSFE DTSSFEVIG DTSSFEV
Aeu PhaB HsaB MacA HsaA HsaA HsaA HsaA EcaS EcaE RnoA MuuA Cja Gga Rpe Haa6	ATAMRKAAVFV EFCR 140 150 a a aaa aa p p p p pp pppp pp 1 TCRCKFIHHFUCTSTFSQTTW TCRCKFIHHFUCSTFSQTTW TCRCKFIHHFUCSTFSQTTW TCRCKFIHHFUCTSTFSQTTW CCRCKFIHHFUCTSTFSQTTW SCRCKFIHHFUCTSTFSQTTW SCRCKFIHHFUCTSTFFSQTTW TCRCKFIHHFUCTSTFFSTTWTW TCRCKSIHHFUCTSTFFTETTW TCRCKSIHHFIGTSTFFTETTW TCRCKSIHHFIGTSTFFTETTW	IELA DI 16 16 P P PF DEN AVAR DEN AVAR EN AVAR EN AVAR	KPIPDIGP 50 pp pp KIDA ASPL KIDA ASPL KIDA ASPL KIDA ASPL KIDA ASPL KIDA ASPL KIDA ASPL KIDA AAPL KIDA AAPL KIDS AAPL KIDS AAPL KIDS AAPL	ADALURITT 170 aaa aa pppp ppp EXVCLIGCG EXVCLIGCG EXVCLIGCG EXVCLIGCG EXVCLIGCG EXVCLIGCG EXVCLIGCG EXVCLIGCG EXVCLIGCG	180 180 180 PP PP PSTGYGSAVW, PSTGYGSAVW, PSTGYGSAVW, PSTGYGSAVW, PSTGYGSAVW, PSTGYGSAVW, PSTGYGSAVW, PSTGYGSAVW, PSTGYGAVUQT, PSTGYGAVUQT, PSTGYGAVUQT,	ILKGEYPVAKCI           190         200           *         #           a         aa aa aaa           p         p           p         p           VKVTPCSVCAVFCL         WKVTPCSTCAVFCL           WKVTPCSTCAVFCL         WKVTPCSTCAVFCL           WKVTPCSTCAVFCL         WKVTPCSTCAVFCL           WKVTPCSTCAVFCL         WKVTQCSTCAVFCL           WKVTQCSTCAVFCL         WKVTPCSTCAVFCL           WKVTPCSTCAVFCL         WKVFQCSTCAVFCL           WKVFQCSTCAVFCL         WKVFQCSTCAVFCL           WKVFQCSTCAVFCL         WKVFQCSTCAVFCL           WKVFQCSTCAVFCL         WKVFQCSTCAVFCL	TVGREPVGI 2: * * # # t t t amaaa ppppp p GGVGLSAVI GGVGLSAVI GGVGLSAVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI	LEKLGSAV 4 t aaa aaa 100 100 100 100 100 100 100	AGYREGQRV 220 ; pppp p kIIAVDINKDI KIIAVDINKDI KIIAVDINKDI KIIAVDINKDI KIIAVDINKDI KIIAVDINKDI KIIAVDINKDI KIIAVDINKDI KIIAVDINKDI KIIAVDINKDI KIIAVDINKDI KIIAVDINKDI KIIAVDINKDI KIIAVDINKDI	IAGAICPNFNSYA 230 240 * t t RAAXELGATECI (FAKAKELGATECI (FAKAKELGATECI (FAKAKELGATECI (FAKAKELGATECI (FAKAKELGATECI (FAKAKELGATECI (FAKAKELGATECI (FAKAKELGATECI (FAKAKELGATECI (FKKAGELGATECI (FKKAGELGATECI	40000000000000000000000000000000000000	LMASGQGG 260 260 260 21LEMTDGGV JLEMTDGGV JLEMTDGGV JLEMTDGGV JLEMTDGGV JLEMTDGGV JLEMTDGGV JLEMTDGGV JLEMTDGGV JLEMTDGGV JLEMTDGGV	270 270 P PPPP DFSFEVIG
Aeu PhaB HsaB MacA HsaA HsaA EcaS EcaE RnoA MmuA Gga Rpe Hsa6 RnoX	HTANHKAAVFV EFCR 140 150 a a aaa aa p p p p pp pppp pp 1 1CGKCFHHFVGTSTFSQTTVV 1CGKCFHHFLGVSTFSQTTVV 1CGKCFHHFLGTSTFSQTTVV 1CGKCFHHFLGTSTFSQTTVV 1CGKCFHHFLGTSTFSQTTVV 1CGKCFHHFLGTSTFSQTTVV 1CKKCKAIHHFLGTSTFSQTTVV 1CKKCKAIHHFLGTSTFFSQTTVV 1CKKCKAIHHFLGTSTFFSTFTVTVV 1CKKCKAIHFVGTSTFFETTVV	IELA DI IELA DI IEL	KPIPDIGP 50 50 KIDA ASPL KIDA AAPL KIDA AAPL KIDA AAPL KIDA SAPL KIDA SAPL KIDA SAPL	ADALURITT 170 aaa aa pppp ppp EXVCLIGCC EXVCLIGCC EXVCLIGCC EXVCLIGCC EXVCLIGCC EXVCLIGCC EXVCLIGCC EXVCLIGCC EXVCLIGCC EXVCLIGCC EXVCLIGCC EXVCLIGCC EXVCLIGCC EXVCLIGCC EXVCLIGCC EXVCLIGCC EXVCLIGCC	180 180 PP PP PSTGTSGAVNV, PSTGTSGAVNV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGAVLQT, PSTGYGAVLQT, PSTGYGAAUNT, PSTGYGAAUNT,	ILKGEYPVAKGL 190 200 * # t t a aa aa aaa p p pppp ppp ppp ppp	TYGREPVGI 2: * * # # t t t anaaa ppppp p GGVCLSAVI GGVCLSAVI GGVCLSVI GGVCLSVI GGVCLSVI GGVCLSVI GGVCLSVI GGVCLSVI GGVCLSVI GGVCLSVI GGVCLSVI GGVCLSVI	10 # t aaa aaa PPD PPP IGCKAAGAA IGCKAAGAA IGCKAAGAA IGCKAAGAA IGCKAAGAA IGCKAAGAA IGCKAAGAA IGCKAAGAA IGCKAAGAA IGCKAAGAA	AL A A A A A A A A A A A A A A A A A A	IAGAICPNFNSYA 230 240 * t t RAAXELGATECI (FRAXELGATECI (FRAXELGATECI (FRAXELGATECI (FRAXELGATECI (FRAXELGATECI (FRAXELGATECI (FRAXELGATECI (FRAXELGATECI (FRAXELGATECI (FRAXELGATECI (FRAXELGATECI (FRAXELGATECI (FRAXELGATECI (FRAXELGATECI (FRAXELGATECI (FRAXELGATECI (FRAXELGATECI (FRAXELGATECI)	40000000000000000000000000000000000000	LMASCQCG 260 260 260 260 200 260 200 200	270 270 P PPPP DFSFEVIG DFSFECIG
Aeu PhaB HsaB HsaB HsaC EcaS EcaS EcaS EcaS RnoA MuuA Cja Gga Rpe Hsa6 RnoX MuuX	HTANHKAAVFV EFCK 140 150 a a caa ca p p p pp ppppp pp 1 1CGKKPIHHFUGTSTFSQTTVV 1CGKKPIHHFLGTSTFSQTTVV 1CGKKPIHHFLGTSTFSQTTVV 1CGKKPIHHFLGTSTFSQTTVV 1CGKKPIHHFLGTSTFSQTTVV 1CGKKAIHHFLGTSTFSQTTVV 1CKKKAIHHFUGTSTFFETTVV 1CKKKSIHHFUGTSTFFETTVV 1CKKKSIHHFUGTSTFSETTVV 1CKKKSIHHFUGTSTFSETTVV 1CKKKSIHHFUGTSTFSETTVV 1CKKKSIHHFUGTSTFSETTVV 1CKKKSIHHFKGTSTFSETTVV 1CKKKSIH	IELA DI 16 16 P P PP PP DEN AVAN DEN AV	KPIPDIGP 50 pp pp KIDA ASPLI KIDA ASPLI	aaa aa pppp ppp ExvCllGCG ExvCllGCG ExvCllGCG ExvClLGCG ExvClLGCG ExvClLGCG ExvClLGCG ExvClLGCG ExvClLGCG ExvClLGCG ExvClLGCG ExvClLGCG ExvClLGCG ExvClLGCG ExvClLGCG ExvClLGCG	180 180 180 PP PP PSTGTGSAVNV, PSTGYGSAVNV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGAVQT, PSTGYGAVQT, ISTGYGAAVNT, ISTGYGAAVNT, ISTGYGAAVNT,	ILKGEYPVAKCI 190 200 * # t t a as as as as p p ppps kvTPcSTCAVFGL kvTPcSTCAVFGL kvTPcSTCAVFGL kvTpcSTCAVFGL kvTpcSTCAVFGL kvPcSTC	TYGREPVGI 2: * * # # t t sasaa ppppp p GGVGLSAVI GGVGLSAVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI	1 EKLGSAV 10 # t aaa aaa PPP PPP IGCKAAGAA IGCKAAGAA IGCKAAGAA IGCKAAGAA IGCKAAGAA IGCKAAGAA IGCKAAGAA IGCKAAGAA IGCKAAGAA IGCKAAGAA IGCKAAGAA	AL A A A A A A A A A A A A A A A A A A	IAGAICPNFNSYA 230 240 * t t saaaa ppp f craakelcateci	250 250 250 250 250 250 250 250 250 250	LMASCQCG 260 260 260 27, 200 200 200 200 200 200 200 200 200 200	270 270 P PPPP DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFECIG DFSFECIG
Aeu PhaB HsaB HsaB HsaC EcaE RnoA HsaC Cja Gga Rpe Hsa6 RnoX HsaX EcaX	HTANHKAAVFV EFCK 140 150 a a caa ca p p p pp ppppp pp 1 TCRKKFIHHFUCTSTFSQTTVV TCRKKFIHHFUCTSTFSQTTVV TCRKKFIHHFUCTSTFSQTTVV TCRKKFIHHFUCTSTFSQTTVV TCRKKFIHHFUCTSTFSQTTVV TCRKKFIHHFUCTSTFSQTTVV TCKKKSIHHFUCTSTFFSTTVV TCKKKSIHHFUCTSTFFSTTVV TCKKKSIHHFUCTSTFFSTTVV TCKKKSIHHFUCTSTFSETTVV TCKKKSIHHFUCTSTFSETTVV TCKKKSIHHFUCTSTFSETTVV TCKKKSIHHFUCTSTFSETTVV TCKKKSIHHFUCTSTFSETTVV TCKKKSIH	IELA DI IGI IGI IGI IGI IGI IGI IGI IGI IGI I	KPIPDIGP 50 pp PP KIDA ASPLI KIDA SAPLI KIDA SAPLI KID	aaa aa pppp ppp EkvCLIGCG EkvCLIGCG EkvCLIGCG EkvCLIGCG EkvCLIGCG EkvCLIGCG EkvCLIGCG EkvCLIGCG EkvCLIGCG EkvCLIGCG EkvCLIGCG EkvCLIGCG EkvCLIGCG EkvCLIGCG EkvCLIGCG EkvCLIGCG EkvCLIGCG	180 180 180 PP PP PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGAVLQT, PSTGYGAVLQT, PSTGYGAAVNT, ISTGYGAAVNT, VSTGYGAAVNT,	ILKGEYPVAKGI 190 200 * t t a as as as as p p pppp kvTPcSTCAVFGL kvTPcSTCAVFGL kvTPcSTCAVFGL kvTPcSTCAVFGL kvTpcSTCAVFGL kvTpcSTCAVFGL kvPcSTCAV	TYGREPVGI 2: * * # # t t sasaa ppppp p GGVGLSAVI GGVGLSAVI GGVGLSAVI GGVGLSVVI GGVGLSVVI GGVGLSVVI GGVGLSVVI GGVGLSVVI GGVGLSVVI GGVGLSVVI GGVGLSVVI GGVGLSVVI GGVGLSVVI GGVGLSVVI GGVGLAVII GGVGLAVII GGVGLAVII	10 # t aaa aaa ppp ppp 16CKAAGAA 16CKA	AL A A A A A A A A A A A A A A A A A A	IAGAICPNFNSYA 230 240 * t t sa a aa p pp t c reacher created c	250 250 250 250 250 250 250 250 250 250	LMASCQCG 260 260 260 260 260 260 260 260	270 270 270 270 270 270 270 270
PhaB HsaB HsaC HsaC HsaC HsaC EcaS EcaS EcaE Roa Rac Rac Rac Rac Rac Rac Rac Rac Rac Ra	HTANHKAAVFV EFCK 140 150 a a aaa aa p p p p pp pppp pp p TCRGKPIHHFUGTSTFSQTTW TCRGKPIHHFUGTSTFSQTTW TCRGKPIHHFUGTSTFSQTTW TCRGKPIHHFUGTSTFSQTTW TCRGKPIHHFUGTSTFSQTTW TCRGKPIHHFUGTSTFSQTTW TCRGKSINHFIGTSTFSTTWTU TCKGKSINHFIGTSTFSTTWTU TCKGKSINHFUGTSTFSETTW TCKGKSINHFWGTSTFSETTW TCKGKFUHFWGTSTFSETTW TCKGKFUHFWGTSTFSETTW TCKGKFUHFWGTSTFSETTW TCKGKFUHFWGTSTFSETTW	IELA DI IELA DI IELA DI IELA P P PP PP PP PP PP PP AVAN DEN AVAN DE	KPIPDIGP 50 50 50 50 50 50 50 50 50 50	ANDALURITT 170 aaa aa pppp ppp EKVCLIGCG	180 180 180 PP PP PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGAVUGT, PSTGYGAVUGT, PSTGYGAVUT, PSTGYGAAVNT, PSTGYGAAVNT, PSTGYGAAVNT, PSTGYGAAVNT, PSTGYGAAVNT,	ILKGEYPVAKGI 190 200 * # t t a aa aa aaa p pp pp pp pp pp pp pp pp pp pp pp state a aa aaaa p pp pp pp pp pp pp pp pp pp pp pp pp pp pp pp pp pp pp pp	TYGREPVGI 2: * * # # t t sasaa ppppp p GGVGLSAVI GGVGLSAVI GGVGLSAVI GGVGLSVI	10 # C aaa aaa ppp ppp 16CKAAGAA 16CKAAGAA 16CKAAGAA 16CKAAGAA 16CKAAGAA 16CKAAGAA 16CKAAGAA 16CKAAGAA 16CKAAGAA 16CKAAGAA 16CKAAGAA 16CKAAGAA 16CKAAGAA	AGYREGQRY 220 : pppp p RIIAVDINKDI RIIAVDINKDI RIIAVDINKDI RIIAVDINKDI RIIGVDINKDI RIIGVDINKDI RIIGVDINKDI RIIGVDINKDI RIIGVDINKDI RIIGUDINKDI RIIGUDINKDI RIIGUDINKDI RIIGUDINKDI RIIGUDINKDI RIIGUDINKDI RIIGUDINKDI RIIGUDINKDI RIIGUDINKDI	IAGAICPNFNSYA           230         240           #         t     <	AQDGVA.SQDGSY 250 a a a a p pp pp pt n PQYKKFIQEN N PQYKKFIQEN N PQYKKFIQEN N PQYKKFIQEN N PQYKKFIQEN N PQYKKFIQEN N PQYFKFIQEN N PQYFKFIQEN N PQYFKFIQEN N PQYFKFIQEN N PQFFKFIQEN N PQFFKFIQEN	LMASCQCG 260 260 260 27. 27. 27. 27. 27. 27. 27. 27. 27. 27.	270 270 270 270 270 270 270 270
Aeu PhaB HsaB MacA HsaC HsaC EcaS EcaS EcaE Roa Roa Roa Roa Roa Roa Roa Roa Roa Roa	ATAMRKAAVFV EFCK 140 150 a a aaa aa p p p p pppppp pp j TCRGKPIHHFUGTSTFSQTTVV TCRGKPIHHFUGTSTFSQTTVV TCRGKPIHHFUGTSTFSQTTVV TCRGKPIHHFUGTSTFSQTTVV TCRGKPIHHFUGTSTFSQTTVV TCRGKPIHHFUGTSTFSQTTVV TCRGKAIHHFJGTSTFSQTTVV TCRGKSIHHFJSTSTFSQTTVV TCKGKSIHHFJSTSTFSQTTVV TCKGKSIHHFJSTSTFSTTVVT TCKGKSIHHFJSTSTFSTVTVV TCKGKSIHHFGTSTFSETVVT TCKGKSIHHFGTSTFSETVVT TCKGKSIHHFGTSTFSETVVT TCKGKSIHHFGTSTFSETVVT TCKGKSIHHFGTSTFSETVVT TCKGKSIHHFGTSTFSETVVT TCKGKSIHHFGTSTFSETVVT TCKGKSIHHFGTSTFSETVVT TCKGKSIHFFGTSTFSETVVT TCKGKSIHFFGTSTFSETVVT TCKGKSIHFFGTSTFSETVVT TCKGKSIHFFGTSTFSETVVT TCKGKSIHFFGTSTFSETVVT TCKGKSIHFFGTSTFSETVVT TCKGKSIHFFGTSTFSETVVT TCKGKSIFFGTSTFSETVVT TCKGKFFGTSFSETVIT TCKGKFFGTSFSETVIT TCKGKFFGTSFSETVIT TCKGKFFGTSFSETVIT TCKGKFFGTSFSETVIT TCKGKFFGTSFSETVIT TCKGKFFGTSFSETVIT TCKGKFFGTSFSETVIT TCKGKFFGTSFSETVIT TCKGKFFGTSFSETVIT TCKGKFFGTSFSETVIT TCKGKFFGTSFSETVIT TCKGKFFGTSFSETVIT TCKGFFGTSFSFFGTSFSETVIT TCKGFFGTSFSFFGTSFSFFGTSFSETVIT TCKGFFGTSFSFGTSFSFFGTSFSFFGTSFSFFGTSFSFFGTSFSFFGTSFSFFGT TCKGFFGFGFGTSFSFFGTSFSFFGTSFSFFGT TCKGFFGFGTSFSFFGTSFFFG	IELA DI IELA DI IELA DI IELA DI IELA TALAN IELA T	KPIPDIGP 50 50 PP PP CIDA ASPL CIDA	ANDALURITT 170 aaa aa pppp ppp EKVCLIGGC EKVCLIGCC EXVCLIGCC	180 180 180 PP PP PFGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGAVGAVT, PSTGYGAAVNT, PSTGYGAAVNT, PSTGYGAAVNT, PSTGYGAAVNT, PSTGYGAALTLAV, STGCGATLAV,	ILKGEYPVAKGI 190 200 * # t t a aa aa aaa p pp pp pp kvTPcStCAVFGL kvTPcStCAVFGL kvTPcStCAVFGL kvTQcStCAVFGL kvTQcStCAVFGL kvTQcStCAVFGL kvEpcStCAVFGL kvEpcStCAVFGL kvEpcStCAVFGL kvEpcStCAVFGL kvEpcStCAVFGL kvEpcStCAVFGL kvEpcStCAVFGL kvEpcStCAVFGL kvEpcStCAVFGL kvEpcStCAVFGL kvEpcStCAVFGL kvEpcStCAVFGL kvFkcStVA1FGL kvFkcStVA1FGL kpFkcStVA1FGL	TYGREPVGI 2: * * # # t t govclsAvy GGVGLSAVy GGVGLSAVy GGVGLSVV GGVGLSVV GGVGLSVV GGVGLSVV GGVGLSVV GGVGLSVV GGVGLSVV GGVGLSVV GGVGLSVV GGVGLSVV GGVGLSVV GGVGLSVV GGVGLSVV GGVGLSVV GGVGLSVV GGVGLSVV GGVGLSVV GGVGLAVI GGVGLAVI GGVGLAVI GGVGLAVI GGVGLAVI	10 # t aaa aaa ppp ppp 16CKAAGAA 16CKA	TGYREGQRY 220 : pppp p RIIAVDINKDI RIIAVDINKDI RIIAVDINKDI RIIAVDINKDI RIIGVDINKDI	IAGAICPNFNSYA           230         240           #         t     <	AQDGVA.SQDGSY 250 a a a a p pp pp pi N PQDYKKPIQE N PQDYKKPIQE N PQDYKKPIQE N PQDYKKPIQE N PQDYKKPIQE N PQDYKKPIQE N PQDYKKPIQE N PQDYKKPIQE N PQDFKKPIQE N PKDFKKPIHE N PKDFKKPIHE N PKDFKKPIE N PQDFSK5IQE S PQDFSK5IQE N PQDFSK5IQE	LMASGQGG 260 260 260 260 260 260 260 260	270 270 270 270 270 270 270 270
PhaB HsaB HsaA HsaC HsaC EcaE EcaE EcaE RnoA MmuX HsaC Gga Gga Rpe Hsa6 RnoE Hsa6 RnoE Hsa5 Hsa2 Hsa2 Hsa2 Hsa2 Hsa2 Hsa3 Hvu3 Hvu3 Hvu3	ATAMRKAAVFV EFCR 140 150 a a aaa aa p p p p ppppp ppp 1 TCRCKPIHHFUCTSTFSQTTW TCRCKPIHHFUCTSTFSQTTW TCRCKPIHHFUCTSTFSQTTW TCRCKPIHHFUCTSTFSQTTW TCRCKPIHHFUCTSTFSQTTW TCRCKPIHHFUCTSTFSQTTW TCRCKPIHHFUCTSTFSQTTW TCRCKSINHFISTSTFSQTTW TCRCKSINHFISTSTFSQTTW TCRCKSINHFISTSTFSTTWTV TCRCKSINHFISTSTFSTTWTV TCRCKSINHFISTSTFSTTWTV TCRCKSINHFISTSTFSTTWTV TCRCKSINHFGTSTFSTTVTV TCRCKSINHFGTSTFSETTVU TCRC	IELA DI IELA DI IELA DI IELA DI IELA DI IELA TANAN DEN AVAR DEN AVAR	KPIPDIGP 50 50 PP PP FILDA ASPL KIDA	ADALURITT 170 aaa aa pppp ppp EKVCLIGCC EKVCLIGCC EKVCLIGCC EKVCLIGCC EKVCLIGCC EKVCLIGCC EKVCLIGCC EKVCLIGCC EKVCLIGCC EKVCLIGCC EKVCLIGCC EKVCLIGCC DKVCLIGCCC	180 180 180 PP PP PSTGYGSAVRV, PSTGYGSAVRV, PSTGYGSAVRV, PSTGYGSAVRV, PSTGYGSAVRV, PSTGYGSAVRV, PSTGYGSAVRV, PSTGYGSAVRV, PSTGYGSAVRI, PSTGYGSAVRI, STGYGAVNT, VSTGYGAAVNT, STGYGAAVNT, STGYGAAVNT, STGIGATLRV, ISTGLGATLRV,	ILKGEYPVAKGI 190 200 * t a aa aaa p pp pp pp kvTPGSTGAVFGI kvTPGSTGAVFGI kvTPGSTGAVFGI kvTQGSTGAVFGI kvTQGSTGAVFGI kvTQGSTGAVFGI kvPGSTGAVFGI kvPGSTGAVFGI kvPEGSTGAVFGI kvP	TYGHEPVGI 2: * * # # t t govclsAvy govclsAvy govclsAvy govclsAvy govclsAvy govclsVy	10 # t t aaa aaa ppp ppp 16CKAAGAA 16	TGYREGQRY 220 ; aa a a ; pPPD p ; RI IA VDINKD RI IA VDINKD RI IA VDINKD RI IA VDINKD RI IG VDINKD	IAGAICPNFNSYA 230 240 * t aa a aa a p pp t f rakakelgateci (rakakelgateci) (rakakelgateci (rakakelgateci)	AQDGVA.SQDGSY 250 a a a a p pp pp pi N PQDYKKPIQE N PQDYKKPIQE N PQDYKKPIQE N PQDYKKPIQE N PQDYKKPIQE N PQDYKKPIQE N PQDYKKPIQE N PQDYKKPIQE N PKDFKKPIME N PKDFKKPIME N PQDFSKPIQE N PQDFSKPIQE N PQDFSKPIQE N PQDFSKPIQE N PQDFSKPIQE N PQDFSKPIQE N PQDFSKPIQE N PQDFSKPIQE N PQDFSKPIQE N PKDHTKPVQE N PKDHTKPVQE N PKDHTKPVQE	LMASGQGG 260 260 260 260 260 260 260 260	A A P PPP DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFECIG DVSFECIG DVSFECIG DVSFECIG DVSFECIG DFSF
PhaB HsaB HsaA HsaC HsaC EcaS EcaS EcaS EcaS EcaS EcaS RnoA MuuX HsaC Gga Gga Hsa6 RnoA Hsa6 RnoA Hsa6 RnoA Hsa2 Hsa2 Sa2 Hvu3 Hvu3 Hvu3 Hvu3 Hvu3 Hvu3 Hvu3 Hvu3	ATAMARKAAVFV EFCR 140 150 a a aaa aa p p p p ppppp ppp 1 TCRCKPIHHFUCTSTFSQTTW TCRCKPIHHFUCTSTFSQTTW TCRCKPIHHFUCTSTFSQTTW TCRCKPIHHFUCTSTFSQTTW TCRCKPIHHFUCTSTFSQTTW TCRCKPIHHFUCTSTFSQTTW TCRCKPIHHFUCTSTFSQTTW TCRCKFIHHFUCTSTFSTTTVU TCKCKSINHFISTSTFSTTWTVU TCKCKSINHFISTSTFSTTWTVU TCKCKSINHFISTSTFSETTWU TCKCKSINHFFISTSTFSETTWU TCKCKSINHFFISTSTFSETTWU TCKCKSINHFFISTSTFSETTWU TCKCKSINHFFISTSTFSETTWU TCKCKSINHFFISTSTFSETTWU TCKCKSINHFFISTSTFSETTWU TCKCKSINHFFISTSTFSETTWU TCKCKSINHFFISTSTFSETTWU TCKCKSINHFFISTSTFSETTWU TCKCKFINHFWCTSTFSETTWU TINCKFIFHFVCTSTFSETTVU TINCKFIFHFVCTSTFSETTVU TINCKFIFHFVCTSTFSETTVU TINCKFIFHFVCTSTFSETTVU TINCKFIFHFVCTSTFSETTVU TINCKFIFHFVCTSTFSETTVU TINCKFIFHFVCTSTFSETTVU TINCKFIFHFVCTSTFSETTVU TINCKFIFHFVCTSTFSETTVU	IELA DI IGUA IGU	KPIPDIGP 50 PP PP KIDA ASPL KIDA ASPL K	ADALURITT 170 aaa aa pppp ppp EKVCLIGGC EKVCLIGCC EXVCLIGCCC	180 180 180 PP PP PFGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGAVNT, 1STGYGAVNT, 1STGYGANT, 1STGYGANT, 1STGYGATLNV, 1STGIGATLNV, 1STGIGATLNV, 1STGIGATLNV, 1STGIGATLNV, 1STGIGATLNV,	ILKGEYPVAKGL 190 200 * # t a aa aaa p pp ppp kvTPGSTGAVFGL kvTPGSTGAVFGL kvTPGSTGAVFGL kvTQSSTGAVFGL kvTQSSTGAVFGL kvTQSSTGAVFGL kvTQSSTGAVFGL kvTQSSTGAVFGL kvEPGSTGAVFGL kvEPGSTGAVFGL kvEPGSTGAVFGL kvEPGSTGAVFGL kvEPGSTGAVFGL kvEPGSTGAVFGL kvEPGSTGAVFGL kvEPGSTGAVFGL kvEPGSTGAVFGL kvEPGSTGAVFGL kvEPGSTGAVFGL kvEPGSTGAVFGL kvEPGSTGAVFGL kvEPGSTGAVFGL kvEPGSTAVFGL kvEPGSTAVFGL kvEPGSTAVFGL kvEPGSTAVFGL kvEPGSTAVFGL kvEPGSTAVFGL	TYGHEPVGI 2: * * # # t t goodsav goods	12KLGSAV 10 # t t aaa aaa PPD PPD IGCKAAGAA IGCKAAAGAA IGCKAAGAA IGCKAA IGCKAAGAA IGCKAAGAA IGCKAAGAA IGCKA	TGYREGQRY 220 : pppp p RIIAVDINKDI RIIAVDINKDI RIIAVDINKDI RIIAVDINKDI RIIAVDINKDI RIIGVDINKDI	IAGAICPNFNSYA 230 240 * t t RANAKELGATECI (FAKAKELGATECI (FAKAKELGATECI (FAKAKELGATECI (FAKAKELGATECI (FAKAKELGATECI (FAKAKELGATECI (FAKAKEGATECI (FAKAKEFGATECI (FAKAKEFGATECI (FAKAKEFGATECI (FAKAKEFGATECI (FAKAKEFGATECI (FAKAKEFGATECI (FAKAKEFGATECI (FAKAKEFGATECI (FAKAKEFGATECI (FAKAKEFGATECI (FAKAKEFGATECI (FAKAKEFGATECI (FAKAKEFGATECI (FAKAKEFGATECI (FAKAKEFGCTEF) (FEQAKKFGCTEF) (FEQAKKFGCTEF)	AQDGVA.SQDGSY 250 a a a a P PP PP P1 N PQDYKKPIQET N PQDYKKPIQET N PQDYKKPIQET N PQDYKKPIQET N PQDYKKPIQET N PQDYKKPIQET N PQDYKKPIQET N PQDYKKPIQET N PKDFKKPIHET N PKDFKKPIHET N PKDFKKPIHET N PKDFKKPIHET N PKDFKKPIHET N PKDFKKPIHET N PKDFKKPIET N PKDFKKPIET N PKDFKKPIGET N PKDHKPYQET N PKDHKPYQET N PKDHKPYQET N PKDHKPYQET	LMASGQGG 260 260 260 270 270 270 270 270 270 270 270 270 27	A A P PPP DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFECIG DFSF
Aeu PhaB HsaB HsaC HsaC HsaC EcaS EcaS EcaS EcaS EcaS EcaS EcaS RnoA MuuX HsaC Hsa6 RnoA Hsa6 RnoA Hsa6 RnoA Hsa6 RnoA Hsa6 RnoA Hsa2 Cja Gga Hsa6 Hsa8 Hsa7 Cja Cja Cja Cja Cja Cja Cja Cja Cja Cja	ATAMERAAVSV EFCK 140 150 a a a aa aa p p p p pp ppp ppp pp 1CRCK2FIHFFUSTSTSQTTW 1CRCK2FIHFFUSTSTSQTTW 1CRCK2FIHFFUSTSTSQTTW 1CRCK2FIHFFUSTSTSQTTW 1CRCK2FIHFFUSTSTSQTTW 1CRCK2FIHFFUSTSTSQTTW 1CRCK2FIHFFUSTSTSSQTTW 1CRCK2FIHFFUSTSTSSTTVTV 1CRCK2FIHFFUSTSTSSTTVTV 1CRCK2FIHFFUSTSTSSTVTV 1CRCK2FIHFFUSTSTSSTVTV 1CRCK2FIHFFUSTSTSSTVTV 1CRCK2FIHFFUSTSTSSTVTV 1CRCK2FIHFFUSTSTSSTVTV 1CRCK2FIHFFUSTSTSSTVTVI 1CRCK2FIHFFUSTSTSSTVTVI 1CRCK2FIHFFUSTSTSSTVTVI 1CRCK2FIHFFUSTSTSSTVTVI 1CRCK2FIHFFUSTSTSSTVTVI 1CRCK2FIFFUSTSSTVTVI 1CRCK2FIFFUSTSTSSTVTVI 1CRCK2FIFFUSTSTSSTVTVI 1CRCK2FIFFUSTSTSSTVTVI 1CRCK2FIFFUSTSTSSTVTVI 1CRCK2FIFFUSTSSTVTVI 1CRCK2FIFFUSTSTSSTVTVI 1CRCK2FIFFUSTSTSSTVTVI 1CRCK2FIFFUSTSTSSTVTVI 1CRCK2FIFFUSTSTSSTVTVI 1CRCK2FIFFUSTSSVTVI 1CRCK2FIFFUSTSSVTVI 1CRCK2FIFFUSTSSVTVI 1CRCK2FIFFUSTSSVTVI 1CRCK2FIFFUSTSSVTVI 1CRCK2FIFFUSTSSVTVI 1CRCK2FIFFUSTSSVTVI 1CRCK2FIFFUSTSSVTVI 1CRCK2FIFFUSTSSVTVI 1CRCK2FIFFUSTSSVTVI 1CRCK2FIFFUSTSVTVI 1CRCK2FIFFUSTSVTVI 1CRCK2FIFF	IELA DI IELA DI IEL	KPIPDIGP 50 50 FALL FALL FALL FALL FALL FALL FALL FAL	NDALURITT 170 aaa aa pppp ppp Exvclicccc Exvcliccc Exvcliccc Exvcliccc Exvcliccc	180 180 180 PP PP PFGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGAVAVT, ISTGGAAVNT, ISTGYGAAVNT, ISTGYGAAVNT, ISTGGAAVNT, ISTGIGATLNV, ISTGLGATLNV, ISTGLGATLNV, ISTGLGATLNV, ISTGLGATLNV, ISTGLGATLNV, ISTGLGATLNV, ISTGLGASINV,	ILKGEYPVAKGL 190 200 * t t aaaaa p pp ppp UKVTPGSVGAVFGL UKVTPGSVGAVFGL UKVTPGSTGAVFGL UKVTPGSTGAVFGL UKVTQGSTGAVFGL UKVTQGSTGAVFGL UKVTQGSTGAVFGL UKVTQGSTGAVFGL UKVEPGSTGAVFGL UKVEPGSTGAVFGL UKVEPGSTGAVFGL UKVEPGSTGAVFGL UKVEPGSTGAVFGL UKVEPGSTGAVFGL UKVEPGSTGAVFGL UKVEPGSTGAVFGL UKVEPGSTGAVFGL UKVEPGSTGAVFGL UKPKGSTVAIFGL UKPKGSTVAIFGL UKPKGSTVAIFGL UKPKGSTVAIFGL	2:	12 KLGSAV 10 # t t aaa aba PPD PPD IGCKAAGAA IGCKAAAGAA IGCKAAGAA IGCKAAGAA IGCKAAGAA IGCKAAGAA IGCKAAGAA I	TGYREGQRY 220 : aa a a ; pPPD p RIIAVDINKDI RIIAVDINKDI RIIAVDINKDI RIIAVDINKDI RIIAVDINKDI RIIAVDINKDI RIIGVDI RIIGVDI RII	IAGAICPNFNSYA 230 240 * t t aa a aa a p pp c frakakelgated frakakelgated frakakelgated frakakelgated frakakelgated frakakelgated frakakelgated frakakelgated frakakelgated frakakegated frafakegated fra	AQDGVA.SQDGSY 250 a a a a p pp pp p1 N PQDYKKPIQET N PQDYKKPIQET N PQDYKKPIQET N PQDYKKPIQET N PQDYKKPIQET N PQDYKKPIQET N PQDYKKPIQET N PQDYKKPIQET N PQDYKKPIQET N PKDFKKPIHET N PKDFKKPIHET N PKDFKKPIHET N PKDFKKPIET N PKDFKKPIET N PKDFKKPIET N PKDFKKPIET N PKDHTKPVQET N PKDHTKPVQET N PKDHTKPVQET N PKDHTKPVQET N PKDHTKPVQET N PKDHTKPVQET N PKDHTKPVQET N PKDHTKPVQET	LMASGQGG 260 260 260 260 260 260 260 260	A A P PPP DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFECIG DYSFECIG DYSFECIG DFSF
Aeu PhaB HsaB HsaC HsaC HsaC EcaS EcaS EcaS EcaS EcaS EcaS EcaS RnoA HsaG RnoA Hsa6 RnoA Hsa6 RnoA Hsa6 RnoA Hsa6 RnoA Hsa6 RnoA Hsa6 RnoA Hsa6 Cja Gga Rpe Hsa6 RnoA Hsa7 Cja Cja Cja Cja Cja Cja Cja Cja Cja Cja	ATAMERAAVSV EFCR 140 150 a a a aa aa p p p p pp pppp pp try 1000000000000000000000000000000000000	IELA DI IELA DI IELA DI IELA DI IELA P P PP PP PP PP PP PP AVAR DEN AVAR DEN AVAR DEN AVAR DEN AVAR DEN AVAR DEN AVAR DEN AVAR DEN AVAR DEN SVAR ADI SVAR AD	KPIPDIGP 50 50 FALL FALL FALL FALL FALL FALL FALL FAL	NDALVRITT 170 aaa aa pppp ppp ExvCliGCG EXVCLIGCG EXVCLIGCG EXVCLIGCG EXVCLIGCG EXVCLIGCG EXVCLIGCG EXVCLIGCG EXVCLIGCG EXVCLIGCG EXVCLIGCG DKVCLIGCGC	180 180 180 180 PP pp PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGAVUQT, PSTGYGAVUQT, PSTGYGAVUQT, ISTGGAAVNT, ISTGA	ILKGEYPVAKGL ILKGEYPVAKGL * t a aa aa aaa p pp ppp kvTPGSVGAVFGL kvTPGSTGAVFGL kvTPGSTGAVFGL kvTPGSTGAVFGL kvTQGSTGAVFGL kvTQGSTGAVFGL kvEPGSTGAVFGL kvEPGSTGAVFGL kvEPGSTGAVFGL kvEPGSTGAVFGL kvEPGSTGAVFGL kvEPGSTGAVFGL kvEPGSTGAVFGL kvEPGSTGAVFGL kvEPGSTGAVFGL kvEPKGSTVAIFGL kPKKGSTVAIFGL kPKKGSTVAIFGL kPKKGSTVAIFGL kPKKGSTVAIFGL kPKKGSTVAIFGL kPKGSTVAIFGL	TYGREPVGI 2: * * # # t t t t asaaa ppppp p GGVGLSAVI GGVGLSAVI GGVGLSVVI GGVGLSVVI GGVGLSVVI GGVGLSVVI GGVGLSVVI GGVGLSVVI GGVGLSVVI GGVGLSVVI GGVGLSVVI GGVGLSVVI GGVGLSVVI GGVGLSVVI GGVGLAVII GGVGLAVII GGVGLAVII GGVGLAVII GGVGLAVII GAVGLAANI GAVGLAANI GAVGLAANI GAVGLAANI GAVGLAANI GAVGLAANI GAVGLAANI GAVGLAANI GAVGLAANI	10 # t aga aga ppp ppp IGCKAAGAA IGCKAAA IGCKAA IGCKAA IGCKAAA IGCKAAA IGCKAAA IGCKAA IGCKAAA IGCKAA IGCKAAA	AG A A A A A A A A A A A A A A A A A A	IAGAICPNFNSYA 230 240 * t t s c c c c c c c c c c c c c c c c c	AQDGVA.SQDGSY 250 A A A AA P PP PP P1 N PQPYKKPIQET N PQPFKSIQET N PQPFKSIQET N PQPFKSIQET N PQPFKSIQET N PQPFKFIQET N PQDFKFIQET N PKDHTKPYQET N PKDHTKPYQET	LMASCQCG 260 260 260 260 260 260 260 260	270 270 P PPP DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFECIG
Aeu PhaB HsaB MacA HsaC EcaS EcaS EcaS EcaS EcaS EcaS EcaS RnoA MmuX HsaC Gga Gga Rpe Hsa6 RnoX Hsa6 RnoX Hsa7 Hsa8 Hsa6 RnoA Hsa6 RnoA MmuX Hsa7 Cja Gga Gga Rpe Hsa6 RnoA Hsa6 RnoA Hsa6 RnoA Hsa7 Cja Gga Cja Gga Rpe Hsa6 RnoA Hsa6 RnoA Hsa6 RnoA Hsa7 Cja Gga Rpe Hsa6 RnoA Hsa7 Cja Gga Rpe Hsa6 RnoA Hsa7 Cja Gga Rpe Hsa6 RnoA Hsa7 Cja Gga RnoA Hsa7 Cja Cja Cja Cja Cja Cja Cja Sca2 Hsa6 RnoA Hsa7 Hsa7 Hsa6 RnoA Hsa6 RnoA Hsa7 Sca2 Hsa6 RnoA Hsa7 Sca2 Hsa6 RnoA Hsa7 Sca2 Hsa6 RnoA Hsa7 Sca2 Hsa6 RnoA Hsa7 Sca2 Hsa6 RnoA Hsa7 Hsa7 Hsa7 Hsa7 Hsa7 Hsa6 RnoA Hsa7 Sca2 Hsa6 RnoA Hsa7 Sca2 Hsa6 RnoA Hsa7 Sca2 Hsa7 Hsa7 Hsa7 Hsa7 Hsa7 Hsa7 Hsa7 Hsa7	ATAMERAAVSV EFCR 140 150 a a a aa aa p p p p pp pppp pp 1 1CRGKPIHEFUGTSTSQTTW 1CRGKPIHEFUGTSTSQTTW 1CRGKPIHEFUGTSTSQTTW 1CRGKPIHEFUGTSTSQTTW 1CRGKPIHEFUGTSTSQTTW 1CRGKPIHEFUGTSTSQTTW 1CRGKPIHEFUGTSTSQTTW 1CRGKPIHEFUGTSTSTCTTVW 1CRGKPIHEFUGTSTSTCTTVW 1CRGKPIHEFUGTSTSTETTVW 1CRGKPIHEFUGTSTSSETTWI 1CRGKPIHEFUGTSTSSETTWI 1CRGKPIHEFUGTSTSSETTWI 1LRGKPIPHFUGTSTSSETTWI 1LRGKP	IELA DI IELA DI 16 16 16 16 16 16 16 16 16 16	KPIPDIGP 50 50 50 50 50 50 50 50 50 50 50 50 50	NDALVRITT 170 aaa aa pppp ppp ExvCLIGCG EXVCLICCG EXVCLICCG	180 180 180 PP PP PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGAVLQT, PSTGYGAVLQT, PSTGYGAVNT, 1STGYGAVNT, 1STGYGAVNT, 1STGYGAVNT, 1STGYGAVNT, 1STGYGAVNT, 1STGYGAVNT, 1STGIGATLNV, 1STGIGATLNV, 1STGIGASINV, 1STGIGASINV, 1STGIGASINV, 1STGIGASINV, 1STGIGASINV, 1STGIGASINV, 1STGIGASINV, 1STGIGATLNV, 1STGIGASINV, 1STGIGASINV, 1STGIGATLNV, 1STGIGASINV, 1STGIGASINV, 1STGIGASINV, 1STGIGASINV, 1STGIGATLNV, 1STGIGASINV, 1	ILKGEYPVAKGI 190 200 * * t t a aa aa aaa p p pppp pppt kvTPcStCAVFGI kvTPcStCAVFGI kvTPcStCAVFGI kvTPcStCAVFGI kvTPcStCAVFGI kvTPcStCAVFGI kvEPcSTCAVFGI kvEPcSTCAVFGI kvEPcSTCAVFGI kvEPcSTCAVFGI kvEPcSTCAVFGI kvEPcSTCAVFGI kvEPcSTCAVFGI kvEPcSTCAVFGI kvEPcSTVAVFGI kVEPcSTVAVFGI kPKcSTVAIFGI kPFKcSTVAIFGI kPFKcSTVAIFGI kPFKSTVAIFGI kPFKSTVAIFGI	2: 2: 2: 2: 2: 2: 2: 2: 2: 2:	10 # t ada ada PPP PPP IGCKAAGAA IGCKAA I	AL A A A A A A A A A A A A A A A A A A	IAGAICPNFNSY IAGAICPNFNSY 230 240 * t t t t c p PP 5 p FP 5 p FP 5 p PP 5 p PAXAELATEOI (FAKAKELAATEOI (FAKAKELAATEOI (FAKAKELAATEOI (FAKAKELAATEOI (FAKAKELAATEOI (FAKAKELAATEOI (FAKAKEFCATEOI (FAKAKEFCATEOI (FAKAKEFCATEOI (FAKAKEFCATEOI (FAKAKEFCATEOI (FAKAKEFCATEOI (FAKAKEFCATEOI (FAKAKEFCATEOI (FAKAKEFCATEOI (FAKAKEFCATEOI (FAKAKEFCCTEFI REEAKKFCCTEFI	AQDGVA.SQDGSY 250 A A A AA P PP PP P1 N PQPYKKPIQET N PQPSKPIQET N PQPSKPIQET N PQPSKPIQET N PQPSKPIQET N PQPSKPIQET N PQPFKKPIQET N PQDFKKPIQET N PKDHTKPYQET N PKDHTKPYQET	LMASCQCG 260 260 260 260 260 260 260 260	A A P PPP DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFECIG DFSFECIG DFSFECIG DFALDCAG DRAVECTG DRAVECTG DRSV
Aeu PhaB HsaB HsaA HsaC EcaE RhoA HsaC EcaE RhoA Cja Gga Rpe Hsa6 RnoX HsaP Tae Hvu2 Zma2 Osa2 Zma1 Pam Hvu1 Osa1 Stu Fan Tre	ATAMERAAVFV EFCR 140 150 a a aaa aa p p p p pp pppp pp to 1CRCKPTHEFVGTSTFSQTTVV TCRCKPTHEFUGTSTFSQTTVV TCRCKPTHEFUGTSTFSQTTVV TCRCKPTHEFUGTSTFSQTTVV TCRCKPTHEFUGTSTFSQTTVV TCRCKPTHEFUGTSTFSQTTVV TCRCKPTHEFUGTSTFSTTVTVV TCRCKPTHEFUGTSTFSTTVTVV TCRCKSTHEFUGTSTFSTTVTVV TCRCKSTHEFUGTSTFSTTVTVV TCRCKSTHEFUGTSTFSETVU TCRCKSTLEPHCGTSTFSETVU TCRCKSTLEPHCGTSTFSETVU TCRCKSTLEPHCGTSTFSETVU TLRCKFTHFVGTSTFSETVU TLRCKFTHFVGTSTFSETVU TLRCKFTHFVGTSTFSETVU TLRCKFTHFVGTSTFSETVU TLRCKFTHFVGTSTFSETVU TLRCKFTHFVGTSTFSETVU SLRCKFTHFVGTSTFSETVU SLRCKFTUHFVGTSTFSETVF SLRCKFTUHFVGTSTFSETVF SLRCKFTUHFVGTSTFSET	IELA DI IELA DI 16 16 16 16 16 16 16 16 16 16	APP PP 50 50 50 50 50 50 50 50 50 50	aaa aa pppp ppp Ekvclicge	180 180 180 PP PP PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGAVNT, ISTGGATINV, ISTGIGATINV, ISTGIGATINV, ISTGIGATINV, ISTGIGATINV, ISTGIGATINV, ISTGIGATINV, ISTGIGATINV, ISTGIGATINV, ISTGIGATINV, ISTGIGATINV,	ILKGEYPVAKGI 190 200 * * t t a aa aa aaa p p pppp twyrpcstcavfcl twyrpcstravfcl twprcstvaffcl	2:	10 # t aga aga ppp ppp 10 10 10 20 20 20 20 20 20 20 20 20 2	AG A A A A A A A A A A A A A A A A A A	IAGAICPNFNSYA 230 240 * t t t c p PP 5 p FC p PP 5 p FC p PP 5 p P	AQDGVA.SQDGSY 250 a a a a P PP PP PI N PQDYKKPIQE N PQDYKKPIQE N PQDYKKPIQE N PQDYKKPIQE N PQDYKKPIQE N PQDYKKPIQE N PQDYKKPIQE N PQDYKKPIQE N PQDYKKPIQE N PKDFKYPIE N PKDFKYPIE N PQDFKSIQE N PQDFSKFIQE N PQDFSKFIQE N PQDFSKFIQE N PQDFSKFIQE N PQDFSKFIQE N PKDHKPYQE N PKDHKPYQE	LMASCQCG 260 260 260 260 260 260 260 260	A A P PPP DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFEVIG DFSFECIG DFSF
Aeu PhaB HsaB HsaC HsaC EcaE RnoA HsaC Cja Cja Cja Cga Rpe HsaC RnoX HsaP HsaC HsaC Sa Rpe HsaC Sa Sa Sa HsaC Sa Sa Sa Sa Sa Sa Sa Sa Sa Sa Sa Sa Sa	ATAMERAAVSV EFCR 140 150 a a aaa aa p p p pp ppppp pp 1 1CGKCFHHFVGTSTFSQTTW 1CGKCFHHFLGSTFSQTTW 1CGKCFHHFLGSTFSQTTW 1CGKCFHHFLGSTFSQTTW 1CGKCFHHFLGSTFSQTTW 1CGKCFHHFLGSTFSQTTW 1CGKCFHHFLGSTFSQTTW 1CGKCFHHFUGTSTFSQTTW 1CKKCFHHFVGTSTFSTTTTTTW 1CKKCFHFFUGTSTFSETTW 1CKKCFHFUGTSTFSETTW 1CKKCFHFFUGTSTFSETTW 1CKKCFHFUGTSFSETTW 1CKKCFHFUGTSTFSETTW 1	IELA DI IELA DI 16 16 16 16 16 16 16 16 16 16	KPIPDIGP 50 50 50 50 50 50 50 50 50 50	aaa aa pppp ppp Eevening Eeven	180 180 180 PP PP PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGAVQT, TSTGYGAVQT, ISTGGATUNV, ISTGGATUNV, ISTGIGATUNV,	ILKGEYPVAKGI 190 200 * # t a aa aa aaa p p pppp twytpcsycavpci twythcsycavpci twytpcsycavpci twytpcs	TYGREPVGI 2: * * # # # # t t asasa ppppp p GGVGLSAVI GGVGLSAVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLAVI GGVGLAVI GGVGLAVI GGVGLAVI GAVGLAANI	10 # t aaa aaa ppp ppp 16CKAAGAA 16CKA	AL A A A A A A A A A A A A A A A A A A	IAGAICPNFNSYA 230 240 * t t t r p Pp E provint for the second se	250 250 250 250 250 250 250 250 250 250	LMASCQCG 260 260 260 260 260 260 260 260	270 270 270 270 270 270 270 270
Aeu PhaB HsaB HsaA HsaC EcaE RnoA HsaC Cja Gga Rpe HsaC KnoX HsaP Hsu3 EcaX HsaP Hsu3 Zma1 Pam Hvu1 Zma1 Pam Hvu1 Stu Stu Stu Sta Sta Sta Sta Sta Sta Sta Sta Sta Sta	Ido         Ido           Introduction         Ido           Introduction         Ido           Introduction         Introduction           Introduction         <	IELA DI           161           16           16           16           16           16           16           17           18	KPIPDIGP 50 50 50 50 50 50 50 50 50 50 50 50 50	ARA ARA CALLOR AND ALVRITT 170 aaa aa pppp ppp ERVCLIGCG ERVC	180 180 180 PP PP PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGAVQT, ISTGGAAUNT, ISTGYGAAUNT, ISTGYGAAUNT, ISTGYGAAUNT, ISTGIGATLNV, ISTGIGATLN	ILKGEYPVAKGI 190 200 * * * * a as as as as p p ppps kvtrpcsvcavfcl kvtrpcsvcavfcl kvtrpcstcavfcl kvtrpcstcavfcl kvtpcstcavfcl kvpkcstvaffcl kv	TYGREPVGI 2: * * * * * * * * * * * * * * *	12 KLGSAV 10 # t aaa aaa 10 CKAAGAA 10 CKAAGAA	AL A A A A A A A A A A A A A A A A A A	IAGAICPNFNSYA 230 240 * t t c ppp c rhankelanteo: rpankelanteo: rpankelanteo: rpankelanteo: rpankelanteo: rpankelanteo: rpankelanteo: rpankelanteo: rpankelanteo: rpankelanteo: rpankelanteo: rpankelanteo: rpankera	250 250 250 250 250 250 250 250 250 250	LMASCQCG 260 260 260 260 260 260 260 260	270 270 270 270 270 270 270 270
Aeu PhaB HsaB HsaA HsaC EcaE RnoA HsaC Cja Gga Rpe HsaC HsaC HsaC HsaC HsaC Hsa2 ScaE RnoX HsaC Hsa2 ScaE RnoX Hsa2 ScaE RnoX Hsa2 ScaE RnoX Hsa2 ScaE ScaE RnoX Hsa2 ScaE ScaE Hsa2 ScaE RnoX Hsa2 ScaE ScaE Hsa2 ScaE ScaE Hsa2 ScaE ScaE Hsa2 ScaE ScaE Hsa2 ScaE ScaE Hsa2 ScaE ScaE Hsa2 Hsa2 ScaE ScaE Hsa2 Hsa2 Hsa2 Hsa2 Hsa2 Hsa2 Hsa2 Hsa2	Ido         Ido           ICRCKINHFHVCISTSFSQITVV         ICRCKINHFLGISTSFSQITVV           ICRCKINHFLGISTSFSQITVV         ICRCKINHFLGISTSFSQITVV           ICRCKINHFLGISTSFSQITVV         ICRCKINHFLGISTSFSQITVV           ICRCKINHFLGISTSFSQITVV         ICRCKINHFUSTSFFSQITVV           ICRCKINHFUSTSFFSQITVV         ICRCKINHFUSTSFFSTVV           ICRCKINFFVFGTSFSTVI         ICRCKINFFVFGTSFSETVI           ICRCKSVFHFWISTSFSETVV         ICRCKINFFFGTSFSETVI           ICRCKSVFHFWISTSFSETVI         IDRCKINFFGTSFSETVI           ICRCKSVFHFWISTSFSETVI         ISINCKINFFGTSFSETVI           ISINCKINFFFGTSFSETVI         ISINCKINFFGTSFSETVI           ISINCKINFFGTSFSETVI         ISINCKINFFGTSFSETVI           ISINCKINFFGTSFSETVI         ISINCKINFFGTSFSETVI           ISINCKINFFGTSFSETVI         ISINCKINFFGTSFSETVI           ISINCKINFFGTSFSETVI         ISINCKINFFGTSFSETVI           ISINCKINFFGTSFSET	IELA DI IELA DI IELA DI IELA DI IELA DI IELA TARA P P PP PP PP	KPIPDIGP 50 50 50 50 50 50 50 50 50 50 50 50 50	ARA	180 180 180 PP PP PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGAVLQT, PSTGYGAVLQT, PSTGYGAVLT, ISTGIGATINV, LSTGIATINV, LSTGIATINV, LSTGIATINV, LSTGIATINV, LSTGIATINV, LSTGIATINV, LSTGIATINV, LSTGIATINV, LSTGIATINV, LSTGIATINV, LSTGIATINV, LSTGIATINV, LSTGIATINV, LSTGIATINV, LSTGIATINV, LSTGIATINV, LSTGIATINV, LSTGIATINV,	ILKGEYPVAKGI 190 200 * * t t a as as as as p pp ppp kvtrpcsvcavfcl kvtrpcsvcavfcl kvtrpcstcavfcl kvtrpcstcavfcl kvtrpcstcavfcl kvtrpcstcavfcl kvtpcstvavfcl kvpkcstvavfcl	TYGREPVGI 2: * * * * * * * * * * * * * * *	10 # t aaa aaa ppp ppp 10CKAAGAAA 10CKAAGAA 10CKAAGAA 10CKAAGAA 10CKAAGAA 10CKAAGAA 10CK	AL A A A A A A A A A A A A A A A A A A	IAGAICPNFNSYA 230 240 * t t c p PP 5 p FC p FC	AQDGVA.SQDGSY 250 a a a a p pp pp pi N PQDYKKPIQEN N PQDFKKPIQEN N PQDFKKPIQEN N PQDFKKPIQEN N PQDFKKPIQEN N PQDFKKPIQEN N PKDHTKPVQEN N PKDHTKPVQEN N PKDHTKPVQEN N PKDHTKPVQEN N PKDHTKPVQEN N PKDHKPVQEN N PKDKPVQEN N PKDHKPVQEN N PKDKPVQEN N PKDKPVQEN N PKDHKPVQEN	LMASCQCG 260 260 260 260 260 260 260 260	270 270 270 270 270 270 270 270
Aeu PhaB HsaB HsaA HsaA HsaC EcaE RnoA HsaC EcaE RnoA HsaC BroA HsaC HsaC HsaC HsaC Hsa2 Hsa2 Hsa2 Zma2 Zma2 Zma1 Rom Hsa1 Sca1 Scc1 Scc2	Ido         Ido           ICRCKINHFHFUCTSTSSQTTW         ICRCKINHFIGTSTSSQTTW           ICRCKINHFIGUSTSFSQTTW         ICRCKINHFIGTSTSSQTTW           ICRCKINHFIGTSTSSQTTW         ICRCKINHFIGTSTSSQTTW           ICRCKINHFIGTSTSSTQUTW         ICRCKINHFIGTSTSSQTTW           ICRCKINHFIGTSTSSTSTQUTW         ICRCKINHFIGTSTSSTTTETTW           ICRCKINHFIGTSTSSTTTETTW         ICRCKINHFIGTSTSSTTVI           ICRCKINHFIGTSTSSTSTTTTTW         ICRCKINHFIGTSTSSTSTTTTTW           ICRCKINHFIGTSTSSTSTTY         ICRCKINHFIGTSTSSTTY           ICRCKINHFIGTSTSSTSTTSSTTVI         ISINGKINHFIGTSTSSTSTY           ISINGKINHFIGTSTSSTSTY         ISINGKINHFIGTSTSSTSTY           ISINGKINHFIGTSTSSTSTY         ISINGKINHFIGTSTSSTSTY           ISINGKINHFIGTSTSSTSTY         ISINGKINHFIGTSTSSTSTY           ISINGKINHFIGTSTSSTY         ISINGKINHFIGTSTSSTYN           ISINGKINHFIGTSTSSTSTY         ISINGKINHFIGTSTSSTYN	IELA DI           IELA TARA           IELA VAR           IELA	KPIPDIGP 50 50 PP PP 51 52 53 54 55 55 55 55 55 55 55 55 55 55 55 55	ARA AND ALVRITT 170 aaa aa pppp ppp EEVCLIGCG EEVC	180 180 180 PP PP PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVVV, PSTGYGSAVVV, PSTGYGSAVVV, PSTGYGSAVVV, PSTGYGSAVVV, PSTGYGSAVVV, PSTGYGSAVVV, PSTGYGAVU, ISTGIGAU,	ILKGEYPVAKGI 190 200 * * * a aa aa aaa P P PP PPP WYTPGSTGAVFGL KVTPGSTGAVFGL KVTPGSTGAVFGL KVTPGSTGAVFGL KVTPGSTGAVFGL KVTPGSTGAVFGL KVTPGSTGAVFGL KVEPGSTGAVFGL KVEPGSTGAVFGL KVEPGSTGAVFGL KVEPGSTGAVFGL KVEPGSTGAVFGL KVEPGSTGAVFGL KVEPGSTAVFGL KVEFGSTAVFGL KVEFGSTAVFGL KVEFGSTAVFGL KVEFGSTAVFGL KV	TYGREPVGI 2: * * * * * * * * * * * * * * *	10 # t aaa aaa ppp ppp 10CKAAGAA 10CKA	AL A A A A A A A A A A A A A A A A A A	IAGAICPNFNSYA 230 240 * t t s ppg pr ppg t c prakakelaateoi (prakakelaateoi (prakakelaateoi (prakakelaateoi (prakakelaateoi (prakakelaateoi (prakakelaateoi (prakakelaateoi (prakakelaateoi (prakakelaateoi (prakakelaateoi (prakakefaateoi (prakakefaateoi (prakefaateoi (p	250 250 250 250 250 250 250 250	LMASCQCG 260 260 260 260 260 260 260 260	270 270 270 270 270 270 270 270
Aeu PhaB HsaB HsaA HsaA HsaC EcaE EcaE RnoA HsaC Cja Gga HsaC HsaC HsaC HsaC HsaC HsaC HsaC Hsa	A A A ANY ECC 140 150 A A A A A A A P P P P PP PPPP PPP T TCRKETHHFUCTSTSSQTTW TCRKETHHFUCTSTSSQTTW TCRKETHHFUCTSTSSQTTW TCRKETHHFUCTSTSSQTTW TCRKETHHFUCTSTSSQTTW TCRKETHHFUCTSTSSQTTW TCRKETHHFUCTSTSSQTTW TCRKETHHFUCTSTSSQTTW TCRKETHHFUCTSTSSQTTW TCRKETHHFUCTSTSSTTVTT TCKKETHHFUCTSTSSTTVTT TCKKETHHFUCTSTSSTTVTT TCKKETHHFUCTSTSSTTVTT TCKKETHHFUCTSTSSTTVTT TCKKETHHFUCTSTSSTTVTT TCKKETHHFUCTSTSSTTVTT TCKKETHFUCTSTSSTTVTT TCKKETHFUCTSTSSTTVTT TCKKETHFUCTSTSSTTVTT TCKKETHFUCTSTSSTTVTT TCKKETHFUCTSTSSTTVTT TCKKETHFUCTSTSSTTVTT TCKKETHFUCTSTSSTTVTT TCKKETHFUCTSTSSTTVTT TCKKETHFUCTSTSSTTVTT TCKKETTHFUCTSTSSTTVTT TCKKETTHFUCTSTSSTTVTT TCKKETTHFUCTSTSSTTVTT TCKKETTHFUCTSTSSTTVTT TCKKETTHFUCTSTSSTTVTT TCKKETTHFUCTSTSSTTVTT TCKKETTHFUCTSTSSTTVTT TCKKETTHFUCTSTSSTTVTT TCKKETTHFUCTSTSSTTVTT TCKKETTHFUCTSTSSTTVTT TCKKETTHFUCTSTSSTTVTT TCKKETTHFUCTSTSSTTVTT TCKKETTHFUCTSTSSTTVTT TCKKETTHFUCTSTSSTTVTT TCKKETTHFUCTSTSSTTVTT TCKKETTHFUCTSTSSTTVTT TCKKETTHFUCTSTSSTTVTT TCKKETTHFUCTSTSSTTVTT TCKKETTHFUCTSTSSTTVTTT TCKKETTHFUCTSTSSTTVTTT TCKKETTHFUCTSTSSTTVTTT TCKKETTHFUCTSTSSTTVTTT TCKKETTHFUCTSTSSTTVTTT TCKKETTHFUCTSTSSTTVTTT TCKKETTHFUCTSTSSTTVTTT TCKKETTHFUCTSTSSTTVTTT TCKKETTHFUCTSTSSTTVTTT TCKKETTHFUCTSTSSTTVTTT TCKKETTHFUCTSTSSTTVTTT TCKKETTHFUCTSTSSTTVTTT TCKKETTHFUCTSTSSTTVTTT TCKKETTHFUCTSTSSTTVTTT TCKKETTHFUCTSTSSTTVTTT TCKKETTAFFUCTSTSSTTVTTTT TCKKETTAFFUCTSTSSTTVTTT TC	IELA DI IELA DI IELA DI IELA DI IELA DI IELA IELA DI IELA TANA DEN AVAR DEN	KPIPDIGP 50 50 50 50 50 50 50 50 50 50	ARA ARA ARA	180 180 180 180 PP PP PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGAVQT, PSTGYGAVQT, PSTGYGAVQT, PSTGYGAVQT, ISTGIGATINV, ISTGIATINV, ISTGIATINV, ISTGIGATINV, ISTGIATINV, ISTGIGATINV, ISTGIGATINV, ISTGIGATINV, ISTGIGATINV, ISTGIGATINV, ISTGIGATINV, ISTGIGATINV, ISTGIATINV, ISTGIATINV, ISTGIATINV, ISTGIATINV,	ILKGEYPVAKGI ILKGEYPVAKGI 190 200 * * t t a aa aa aaa P pP pPp PP pP PP pP PP pP PP pP PP pP PP pP PP pP PP pP PP pP PP PP pP PP PP PP PP PP PP PP PP PP PP PP PP PP PP P	TYGREPVGI 2: * * * * * * * * * * * * * * *	10 C C C C C C C C C C C C C	TGYREGQRY 220 220 220 220 220 220 220 220 220 22	IAGAICPNFNSY/ 230 240 * t t c p PF F PF F	AQDGVA.SQDGSY 250 a a a a p pp pp p1 N PQDYKKPIQEN N PQDFKKPIQEN N PQDFKKPIQEN N PQDFKKPIQEN N PQDFKKPIQEN N PQDFKKPIQEN N PKDHKPYQEN N PKDKPYQEN N PKDKPYQEN N PKDKPYQEN N PKDKPYQEN N PKDKPYQEN N PKDKPYQEN N PKDK	LMASCQCG 260 260 260 260 260 260 260 260	270 270 270 270 270 270 270 270
Aeu PhaB HsaB HsaA HsaA HsaC EcaE EcaE RnoA HsaC Cja Gga Rpe Hsa6 KnoX Hsa7 Hsa2 Hsa8 Hsa8 Hsa8 Hsa8 Hsa8 Hsa8 Hsa8 Hsa8	A A A AAA AYY EFCH 140 150 A A A AAA AYY EFCH P P P P PP PPPP PPP T TCRCKPIHHFUCTSTFSQTTW TCRCKPIHHFUCTSTFSQTTW TCRCKPIHHFUCTSTFSQTTW TCRCKPIHHFUCTSTFSQTTW TCRCKPIHHFUCTSTFSQTTW TCRCKPIHHFUCTSTFSQTTW TCRCKPIHHFUCTSTFSQTTW TCRCKAIHHFUCTSTFSQTTW TCRCKAIHHFUCTSTFSQTTW TCRCKAIHHFUCTSTFSTTVTU TCRCKAIHHFUCTSTFSTTVTU TCRCKAIHHFUCTSTFSTTVTU TCRCKAIHHFUCTSTFSTTVTU TCRCKAIHHFUCTSTFSTTVTU TCRCKAIHHFUCTSTFSTTVTU TCRCKAIHHFUCTSTFSTTVTU TCRCKAIHHFUCTSTFSTTVTU TCRCKAIHHFUCTSTFSTTVTU TCRCKAIHHFUCTSTFSTTVTU TCRCKTILHYMCTSTFSETTVI TIGKCFIHFUCTSTFSETTVI TIGKCFIHFUCTSTFSETTVI SIGKCFIYHFUCTSTFSETTVI SIGKCFIYHFUCTSTFSETTVI SIGKCFIYHFUCTSTFSETTVI SIGKCFIYHFUCTSTFSETTVI SIGKCFIYHFUCTSTFSETTVI SIGKCFIYHFUCTSTFSETTVI SIGKCFIYHFUCTSTFSETTVI SIGKCFIYHFUCTSTFSETTVI SIGKCFIYHFUCTSTFSETTVI SIGCGYIHFUCTSTFSETTVI SIGGGYIHFUCTSTFSETTVI SIGGGYIHFUCTSTFSETTVI SIGGGYIHFUCTSTFSETTVI SIGGGYIHFUCTSTFSETTVI SIGGGYIHFUCTSTFSETTVI SIGGGYIHFUCTSTFSETTVI SIGGGYIHFUCTSTFSETTVI SIGGGYIHFUCTSTFSETTVI SIGGGYIHFUCTSTFSETTVI SIGGGYIHFUCTSTFSETTVI SIGGGYIHFUCTSTFSETTVI SIGGGYIHFUCTSTFSETTVI SIGGGYIHFUCTSTFSETTVI SIGGGYIHFUCTSTFSETTVI SIGGGYIHFUCTSTFSETTVI SIGGGYIHFUCTSTFSETTVI SIGGGYIHFUCTSTFSETTVI SIGGGYIHFUCTSTFSETVI SIGGGYIHI SIGGGYIHFUCTSTFSETTVI SIGGGYIHI S	IELA DI IELA DI IELA DI IELA DI IELA P P PP PPD PPD PPD AVAN DEN A	KPIPDIGP 50 50 50 50 50 50 50 50 50 50	NDALURITT 170 aaa aa pppp ppp EKVCLIGGC EKVCLIGCC EVVCLIGCCC EVVCLIGCC EVVCLIGCC EVVCLIGCC	180 180 180 180 180 180 180 180	ILKGEYPVAKGI ILKGEYPVAKGI 190 200 * * t t a aa aa aaa p p pp pp pp pp pp pp twyrposycavreil kwyrposycavreil kwyrposycavreil kwyrposycavreil kwyrosystavreil kwyrosystavreil kwyrosystavreil kwprosycavreil kwprosycavreil kwprosycavreil kwprosycavreil kwprosycavreil kwprosycavreil kwprosycavreil kwprosycavreil kwprosycavreil kwprosycavreil kwprosycavreil kwprosycavreil kwprosycavreil kwprosyvarie	TYGREPVGI 2: * * * * * * * * * * * * * * *	10 10 10 10 10 10 10 10 10 10	TGYREGQRY 220 220 220 220 220 220 220 220 220 22	IAGAICPNFNSY/ 230 240 * t t c ppg ppg ppg ppg ppg ppg ppg	AQDGVA.SQDGSY 250 a a a a P PP DP DP DP N PQDYKKPIQE N PKDKYPIE N PKDKYNIPIHE N PKDHKPYQE N PKDHKFYQE N PKDHFYQE N PKDHFYQE N PKDHKFYQE N PKDHKFYQE N PKDHKFYQE N PKDHKFYQE N PKDHKFYQE N PKDHKFYQE N PKDHKFYQE N PKDHFYQE N PKDHFYQE N PKDHFYQE N PKDHKFYQE N PKDKFYQE N PKDKFYQE	LMASCQCG 260 260 260 260 260 260 260 260	270 270 270 270 270 270 270 270
Aeu PhaB HsaB HsaA HsaC HsaC EcaE EcaE RnoA HsaC Cja Gga Rpe Hsa6 RnoA HsuX Hsa7 Hsa7 Hsu3 Hvu3 Hvu3 Hvu3 Hvu3 Hvu3 Hvu3 Hvu3 Hv	A CALL AND	IELA DI           IELA DI           16           16           16           17           18           18           19           19           19           10           10           11           11           11           11           11           12           12           12           12           12           12           12           12           12           13           14           14           15           15           15           15           15           15           15           15           15           15           15           15           15           16           15           15           15           15           15           15           15           15 <t< td=""><td>KPIPDIGP 50 50 50 50 50 50 50 50 50 50</td><td>NDALURITT 170 asa as pppp ppp EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG DKVCLIGCGC</td><td>180 180 180 180 PP PP PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGAVGAV, PSTGYGAVGAV, PSTGYGAVT, ISTGGATINV, ISTGGATINV, ISTGLGA</td><td>ILKGEYPVAKGI ILKGEYPVAKGI 190 200 * # t a aa aa aaa p pp ppp pp pp pp pp twyTPcStCAVFGL twyTPcStCAVFGL twyTPcStCAVFGL twyTpcStCAVFGL twyTpcStCAVFGL twyTpcStCAVFGL twyTpcStCAVFGL twyTpcStCAVFGL twyPcStCAVFGL twyPcStCAVFGL twPECSTCAVFGL twPECSTCAVFGL twPECSTCAVFGL twPECSTCAVFGL twPECSTCAVFGL twPKCSTVATFGL twPKC</td><td>TYGREPVGI TYGREPVGI 2: * * * * * * t t t t c c GVGLSAV; GGVGLSAV; GGVGLSAV; GGVGLSAV; GGVGLSVI; GGUSSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGGUSSLSV; GGGUSSLSV; GGUSSLS</td><td>10 10 10 10 10 10 10 10 10 10</td><td>TGYREGQRY 220 : 220 : 20</td><td>IAGAICPNFNSY IAGAICPNFNSY 230 240 * t t r ppg ppg ppg ppg ppg ppg ppg</td><td>AQDGVA.SQDGSY 250 a a a a P P2 DP p1 N PQDYKKPIQE N PQDFKKPIQE N PKDHKPYQE N PKDHKPYQE N</td><td>LMASCQCG 260 260 260 260 260 260 260 260</td><td>270 270 270 270 270 270 270 270</td></t<>	KPIPDIGP 50 50 50 50 50 50 50 50 50 50	NDALURITT 170 asa as pppp ppp EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG DKVCLIGCGC	180 180 180 180 PP PP PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGAVGAV, PSTGYGAVGAV, PSTGYGAVT, ISTGGATINV, ISTGGATINV, ISTGLGA	ILKGEYPVAKGI ILKGEYPVAKGI 190 200 * # t a aa aa aaa p pp ppp pp pp pp pp twyTPcStCAVFGL twyTPcStCAVFGL twyTPcStCAVFGL twyTpcStCAVFGL twyTpcStCAVFGL twyTpcStCAVFGL twyTpcStCAVFGL twyTpcStCAVFGL twyPcStCAVFGL twyPcStCAVFGL twPECSTCAVFGL twPECSTCAVFGL twPECSTCAVFGL twPECSTCAVFGL twPECSTCAVFGL twPKCSTVATFGL twPKC	TYGREPVGI TYGREPVGI 2: * * * * * * t t t t c c GVGLSAV; GGVGLSAV; GGVGLSAV; GGVGLSAV; GGVGLSVI; GGUSSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGLSLSV; GGGGUSSLSV; GGGUSSLSV; GGUSSLS	10 10 10 10 10 10 10 10 10 10	TGYREGQRY 220 : 220 : 20	IAGAICPNFNSY IAGAICPNFNSY 230 240 * t t r ppg ppg ppg ppg ppg ppg ppg	AQDGVA.SQDGSY 250 a a a a P P2 DP p1 N PQDYKKPIQE N PQDFKKPIQE N PKDHKPYQE N	LMASCQCG 260 260 260 260 260 260 260 260	270 270 270 270 270 270 270 270
Aeu PhaB HsaB HsaC HsaC HsaC EcaE EcaE RnoA HsaC Gga Gga Rpe Hsa6 RnoA Hsa6 RnoA Hsa7 Hsa7 Hsa2 Hsa2 Usa2 Usa2 Usa2 Usa2 Usa2 Usa2 Usa2 U	Ido         150           Ido         1550           Ido         1500           Ido         1500           Ido         1500           Ido         1500           Ido         1500           Ido         1500	IELA DI IELA DI 16 16 16 16 16 16 16 16 16 16	KPIPDIGP 50 50 50 50 50 50 50 50 50 50	NDALURITT 170 aaa aa pppp ppp EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG EKVCLIGCG DKVCLIGCGC	180 180 180 180 PP PP PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKI, PSTGYGAVLQT, PSTGYGAVLQT, PSTGYGAVLQT, PSTGYGAVNT, 1STGYGAVNT, 1STGYGAVNT, 1STGYGAVNT, 1STGYGAVNT, 1STGYGAVNT, 1STGYGAVNT, 1STGYGAVNT, 1STGYGAVNT, 1STGYGAVNT, 1STGIGATLNV, ISTGIGATLNV, 1STGIGATLNV,	ILKGEYPVAKGI ILKGEYPVAKGI * * * a aa aa aa p p pp pp pp pp kvTPcStCAVFGI kvTPcStCAVFGI kvTPcStCAVFGI kvTPcStCAVFGI kvTPcStCAVFGI kvTPcStCAVFGI kvTPcStCAVFGI kvPcSStCAVFGI kvPcSStCAVFGI kvPcSStCAVFGI kvPcSStCAVFGI kvPcSStCAVFGI kvPcSStCAVFGI kvPcSStCAVFGI kvPcSStCAVFGI kvPcSStCAVFGI kvPcSStCAVFGI kvPkcSTVAVFGI kvPkcSTVAVFGI kvPkcSSTVAFGI kvPkcSSTVAFGI kvPkcSSTVAFGI kvPkcSSTVAFGI kvPkcSSVAFGI kvPkcSVAFGI kvPkcSVAFGI kvPkcSVAFGI kvPkcSVAFGI kvPkcSVAFGI kvPkcSVAFGI kvPkcSVAFGI kvPkcSVAFGI kvPkcSVAFGI kvPkcSVAFGI kvPkcSVAFGI kvPkcSVAFGI kvPkcSVAFGI kvPkcSVAFGI kvPkcSVAFGI kvPkcSVAFGI kvPkcSVAFGI kvPkcSVAFGI kvPkcSVAFGI k	TYGREPVGI 2: * * * * * * * * t t t t t asaaa ppppp p GGVGLSAVI GGVGLSAVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLSVI GGVGLAVII GGVGLAVII GGVGLAVII GGVGLAVII GGVGLAVII GGVGLAVII GGVGLAVII GGVGLAVII GGVGLAVII GGVGLAVII GGVGLAVII GAVGLAANI GGUGLSIAV GGGLSIAVI GGGLSIAVI GGUGLSIAVI GGVGLSIAVI GAVGLAANI GGUGLSINV GGUGLSINV GGUGLSINV GGUGLSINV GGVANI GAVI	10 10 10 10 10 10 10 10 10 10	TGYREGQRY 220 220 220 220 220 220 220 220 220 22	IAGAICPNFNSY IAGAICPNFNSY 230 240 * t t c r r r r r r r r r r r r r	AQDGVA.SQDGSY 250 250 250 250 250 250 250 250	LMASCQCG 260 260 260 260 260 260 260 260	270 270 270 270 270 270 270 270
Aeu PhaB HsaB HsaC EcaE RnoA HsaC EcaE RnoA Gga Rpe HsaC Gga Rpe HsaC RnoX MuxX EcaX HsaX HsaX EcaX HsaX Hsa2 Zaa1 Stu Fan Tre Psa Atia Ania Ania Ania Stu Sce2 Xia Sce3 Spo CaC Sce3 Spo CaC Sce3 Sce3 Spo CaC Sce3	A A A AA AA P P P P PP PPPPP PPP I TCRCKFIHHFUGTSTSQTTVV TCRCKFIHHFUGTSTSQTTVV TCRCKFIHHFUGTSTSQTTVV TCRCKFIHHFUGTSTSQTTVV TCRCKFIHHFUGTSTSQTTVV TCRCKFIHHFUGTSTSQTTVV TCRCKFIHHFUGTSTSQTTVV TCRCKFIHHFUGTSTSQTTVV TCRCKFIHHFUGTSTSSQTTVV TCRCKFIHHFUGTSTSSTTVTVV TCRCKFIHHFUGTSTSSTTVTVV TCRCKFIHHFUGTSTSSTTVTVV TCRCKFIHHFUGTSTSSTTVTVV TCRCKFIHHFUGTSTSSTTVTVV TCRCKFIHHFUGTSTSSTTVTVV TCRCKFIHFUGTSTSSTTVTVV TCRCKFIHFUGTSTSSTTVTVV TCRCKFIHFUGTSTSSTTVTVV TCRCKFIHFUGTSTSSTTVTVV TCRCKFIHFUGTSTSSTTVTVV TCRCKFIHFUGTSTSSTTVTVV TCRCKFIHFUGTSTSSTTVTVV TCRCKFIHFUGTSTSSTTVVV TCRCKFIHFUGTSTSSTTVVV TCRCKFIHFUGTSTSSTTVVV TINCKFIFFFUGTSTSSTVTVV TINCKFIFFFUGTSTSSTVVV SINCKFITHFUGTSTSSTVVV SINCKFITHFUGTSTSSTVVVV SINCKFITHFUGTSTSSTVVVV SINCKFITHFUGTSTSSTVVVV SINCKFITHFUGTSTSSTVVVV SINCKFITHFUGTSTSSTVVVV SINCKFITHFUGTSTSSTVVVV SINCKFITHFUGTSTSSTVVVV SINCKFITHFUGTSTSSTVVVV SINCKFITHFUGTSTSSTVVVVXVVVVVVVVVVVVVVVVVVVVVVVVVVVV	IELA DI IELA DI IELA DI IELA DI IELA DI IELA DI IELA P P PP PP PP	KPIPDIGP 50 50 50 50 50 50 50 50 50 50	NDALURITT 170 aaa aa pppp ppp EEVCLIGCGC	180 180 180 180 PP PP PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVNV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGAVUT, PSTGYGAVUT, PSTGYGAVT, ISTGGATINV, ISTGIGAT	ILKGEYPVAKGI ILKGEYPVAKGI 190 200 * * * * a aa aa aaa p p pp pp pp pp twtrpcstcavfcl kwtrpcstcavfcl kwtrpcstcavfcl kwtrpcstcavfcl kwtrpcstcavfcl kwtrpcstcavfcl kwtrpcstcavfcl kwtrpcstcavfcl kwtpcstavfl kwtpcstavfl kwtpcstvavfl	TYGREPVGI 2: * * * * * * * * * * * * * * *	10 * 10 * 10 10 10 10 10 10 10 10 10 10	TGYREGQRY 220 220 220 220 220 220 220 220 220 22	IAGAICPNFNSY IAGAICPNFNSY 230 240 * t t c r r r r r r r r r r r r r	AQDGVA.SQDGSY 250 a a a a P PP PP PI N PQDYKKPIQE N PQDFKKIQE N PQDFKKIQE N PQDFKKIQE N PQDFKKIQE N PQDFKKIQE N PQDFKKIQE N PQDFKKIQE N PQDFKKIQE N PQDFKKIQE N PKDHKPYQE N P	LMASCQCG 260 260 260 260 260 260 260 260	270 270 270 270 270 270 270 270
Aeu PhaB HsaA HsaC EcaE RhoA HsaC Cja Cga Rpe HsaC Cja Cga Rpe HsaC RhoX HsaP Hsa2 Cas2 Cas2 NuuX HsaP Hsu3 Cas2 Cas2 Sco3 Sco3 Sco3 Cas2 Cas2 Cas2 Cas2 Cas2 Cas2 Cas3 Cas3 Cas3 Cas3 Cas3 Cas3 Cas3 Cas3	Ido         Ido           ICRCKINHFVGTSTSQTTW         ICRCKINFFVGTSTSQTTW           ICRCKINFFIGTSTSQTTW         ICRCKINFFGTSTSQTTW           ICRCKINFFGTSTSTQTTW         ICRCKINFFGTSTSTQTTW           ICRCKINFFGTSTSTPTTTW         ICRCKINFFGTSTSTPTTW           ICRCKINFFGTSTSTSTTTTTW         ICRCKINFFGTSTSTSTTTW           ICRCKINFFGTSTSTSTTTTTW         ICRCKINFFGTSTSTSTTTW           ICRCKINFFGTSTSTSTTTTTW         ICRCKINFFGTSTSSTTVI           ICRCKINFFGTSTSTSTTTTTW         ICRCKINFFGTSTSSTYTU           ICRCKINFFGTSTSTSTTTTT         ICRCKINFFGTSTSSTYTU           ICRCKINFFGTSTSSTYTU         IICRCFINFFGTSTSSTYTU           ICRCKINFFGTSTSSTYTU         ISINGKINFFGTSTSSTYTU           ICRCKINFFGTSTSSTYTU         ISINGKINFFGTSTSSTYTU           ICRCKINFFGTSTSSTYTU         ISINGKINFFGTSTSSTYTU           ICRCKINFFGTSTSSTYTU         ICRCKINFFGTSTSSTYTU           ICRCKINFFGTSTSSTYTU	IELA DI IELA DI IELA DI IELA DI IELA DI IELA DI IELA P P PP PP PP	KPIPDIGP 50 50 50 50 50 50 50 50 50 50	ARA	180 180 180 180 180 180 PP PP PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGSAVKV, PSTGYGAVT, ISTGGATINV, ISTGGATINV, ISTGIGATINV	ILKGEYPVAKCI ILKGEYPVAKCI 190 200 * * * * a aa aa aaa p p pp pp pp pp pp trick ktvTpcstCaVFCL ktvTpcstCaVFCL ktvTpcstCaVFCL ktvTpcstCaVFCL ktvTpcstCaVFCL ktvTpcstCaVFCL ktvEp	TYGREPVGI 2: * * * * * * * * * * * * * * *	10 10 10 10 10 10 10 10 10 10	TGYREGQRY 220 220 220 220 220 220 220 220 220 22	IAGAICPNFNSY IAGAICPNFNSY 230 240 * t t t r r r r r r r r r r r r r	AQDGVA.SQDGSY 250 a a a a p pp pp pi N PQDYKKPIQE N PQDFKSIQE S PQDFSKSIQE S PQDFSKSIQE S PQDFSKSIQE N PQDFKSIQE N PQDFKSIQE N PQDFKSIQE N PQDFKSIQE N PQDFKSIQE N PQDFKSIQE N PKDHTKPVQE N PKDE AG QUKCTPHE N PKDE M	LMASCQCG 260 260 260 260 260 260 260 260	270 270 270 270 270 270 270 270

Fig. 1. Progressive sequence alignment of 47 members of the Zn-containing ADH family. a: Conserved in all animal ADHs (18); p: conserved in all plant ADHs (14); t: conserved in all tetrameric ADHs (14); #: conserved in all but  $\zeta$  crystallin; \*: conserved in all 47 members. For enzyme abbreviations, see Table 1.

	200	290	500	510	520	550	540	330	500		5/0
	a	a a		aa	a aa aa	8888	а	а	а	а	
		PPP P	р в раба	PPPPP	PP PP PP	P PP	PP PP	PPPP P	P PP PPP	P PP	PP
PhaB	RLDTMMASLLCCHE/	ACGTSVIVGVE	PDSQNLSINPML	LLTGRTWK	GAV YGGFKSI	EGIPKLV	ADFMAKKFSI	DALITHVLP	FEKINEGFD	LLRSGKSI	RTVLTF
HsaB	RLDTMMASLLCCHE	ACGTSVIVGVE	PASQNLSINPML	LLTGRTWK	GAV YGGFKSI	C EGIPKLV	ADFMAKKFSI	DALITHVLP	FEKINEGFD	LLHSGKSI	RTVLTF
MacA	RLDTMMASLL CCHE/	ACGTSVIVGVE	PDSQNLSINPML	LLTGRTWK	GAV YGGFKSI	C EDIPKLVA	ADFMAKKFSI	DALITHVLP	FEKINEGFD	LLRSGKSI	RTILTF
HsaA	RLDTMMASLL CCHE/	ACGTSVIVGVE	PDSQNLSMNPML	LLTGRTWK	GAI LGGFKS	C ECVPKLV	ADFMAKKFSI	DALITHVLP	FEKINEGFD	LLHSGKSI	RTILMF
HsaC	RLDTMMASLL CCHE/	ACGTSVIVGVÉ	PDSQNLSINPML	LLTGRTWK	GAI FGGFKSI	C ESVPKLVA	ADFMAKKFSI	DALITNILP	FEKINEGFD	LLRSGESI	RTVLTF
EcaS	RLDTMVAALS CCQE	AYGVSVIVGVE	PDSQNLSMNPML	LLSGRTWK	GAI FGGFKS	C DSVPKLVA	ADF MAKKFAI	DPLITHVLP	FEKINEGFD	LLRSGKSI	RTILTF
EcaE	RLDTMVTALS CCQE	AYGVSVIVGVI	PDSQNLSMNPML	LLSGRTWK	GAI FGGFKS	C DSVPKLV	ADF MAKKFAI	DPLITHVLP	FEKINEGFD	LLRSGESI	RTILTF
RnoA	RLDTMTSALL SCHS/	ACGVSVIVGVE	PSAQSLSVNPMS	LLLGRTWK	GAI FGGFKSI	C DAVPKLV	ADFMAKKFPI	EPLITHVLP	FEKINEAFD	LLRAGKSI	RTVLTF
MmuA	RLDTMTSALL SCHA	ACGVSVVVGVI	PPNAQNLSMNPML	LLLGRTWK	GAI FGGFKS	C DSVPKLV	ADFMAKKFPI	DPLITHVLP	FEKINEAFD	LLRSCKSI	RTVLTF
Cja	RIETMTEALA SCHY	VIVGVSVIVGVI	PAAQKISFDPML	IFSGRTWK	GSV FGGWKSI	C DAVPKLV	ADY MKKKFVI	DPLITHTLP	FTKINEGFD	LLRTGKSI	RTVLVL
Gga	RIETMTAALA SCHN	VYGVSVIVGVI	PAAQKISFDPML	IFSGRTWK	GSV FGGWKSI	C DAVPKLV	ADY MKKKFVI	DPLITHTLP	FTKINEGFD	LLRTGKSI	RSVLVL
Rpe	NTTVMTSALS SSHF	GCGKTVIVGLA	PSSAVMSFDPLL	ILTGRILT	GAV FGGWKSI	C DDVPKLVI	RDY LNKKFDF	DPLITHYMP	FEKINEGFE	LLRNGKSI	RTILTF
Hsa6	NLDVLAAALA SCNE	SYGVCVVVGVI	PASVQLKISGQL	FFSGRSLK	GSV FGGWKSI	<b>QHIPKLV</b>	ADYMAEKLNI	DPLITHTLN	LDKINEAVE	MLKTGKW	
RnoX	NVKVMRSALE AAHKO	GWGVSVVVGVA	ASGEEISTRPFQ	LVTGRTWK	GTA FGGWKSV	/ ESVPKLV:	SEYMSKKIKV	DEFVTGNLS	FDQINKAFD	LMHSGNSI	RTVLKL
MmuX	NVKVMRSALE AAHK	GWGVSVVVGVA	ASGEEISTRPFQ	LVTGRTWK	GTA FGGWKS	/ ESVPKLV:	SEYMSKKIKV	DEFVTGNLS	FDQINQAFD	LMHSGDSI	RTVLKM
HsaX	NVKVMRAALE ACHK	GWGVSVVVGV	ASGEEIATRPFQ	LVTGRTWK	GTA FGGWKS	/ ESVPKLV	SEYMSKKIKV	DEFVTHNLS	FDEINKAFE	LMHSGKSI	RTVVKI
EcaX	NVKVMRAALE ACHK	GWGVSVVVGVA	ASGEEIATRPFQ	LVTGRTWK	GTA FGGWKS	/ ESIPKLV:	SEYMSKKIKV	DEFVTHSLS	FDQINEAFE	LMHAGKSI	RTVVKL
HsaP	GSETMKAALD CTTA	GWGSCTFIGVA	AGSKGLTIFPEE	LIIGRTIN	GTF FGGWKS	/ DSIPKLV	IDY KNKKFNI	DALVTHTLP	FDKISEAFD	LMNQGKSI	RTVLIF
Tae	HVDAMIAAFE CVHD	GWGVAVLVGVI	PHKEAVFKTYPMN	FLNERTLK	GTF FGNYKPI	TDLPEVV	EMYMRKELEI	EKFITHSVP	FSQINTAFD	LMLKGEGL	RCIMRMDQ
Hvu3	HIDAMIATFE CVHD	GWGVAVLVGVI	PHKEAVFKTHPMN	FLNEKTLK	GTF FGNYKP	TDLPEVV	EMYMRKELDI	EKFITHSVP	FSQINTAFD	LMLKGEGL	RCITRTDQ
Hvu2	NADAMI SAFE CVHD	GWGVA	HKEAVFKTHPMN	FLNERTLR	GTF FGNYKPI	TGLPGVV	DMYMRKELEI	DKFITHSLP	FSQINTAFD	LMLRGEGL	RCVIRSEE
Zma2	NVNAMI SAFE CVHD	GWGVAVLVGVI	PHKDDQFKTHPMN	FLSEKTLK	GTF FGNYKPI	TDLPNVV	EMYMKKELEI	EKFITHSVP	FSEINTAFD	LMLKGEGL	RCIMRMED
Osa2	NINAMISCFE CVHD	GWGVAVLVGV	KDDVFKTHPMN	FLNEKTLK	GTLIFGNYKPI	TDLPNVV	ELYMKKELEI	EKFIT SVP	FSEINTAFD	LMLKGESL	RCVMSMYE
Zmal	NINAMIQAFE CVHD	GWGVAVLVGVI	PHKDAEFKTHPMN	FLNERTLK	GTF FGNYKPI	R TDLPNVV	ELYMKKELEV	/EKFITHSVP	FAEINKAFD	LMAKGEGI	RCIIRMEN
Pam	NINAMIQAFE CVHD	GWGVAVLVGVI	PHKDAEFKTHPMN	FLNERTLK	GTF FGNFKPI	R TOLPNVV	ELYMKKELEV	/EKFITHSVP	FSEINKAFD	LMAKGEGI	RCIIRMEN
Hvul	NVNAMIQAFE CVHD	GWGVAVLVGVI	PHKDAEFKTHPMN	FLNERTLK	GTF FGNFKPI	R TDLPNVV	EMYMKKELEV	/EKFITHSVP	FSEINTAFD	LMAKGEGI	RCIIRMDN
Osal	NINAMIQAFE CVHD	GWGIAVLVGVI	PHKDAEFKTHPMN	FLNERTLK	GTF FGNYKPI	R TDLPNVV	ELYMKKELEV	/EKF1THSVP	FSEINTAFD	LMHKGEAI	RCIIRMEN
Stu	HIDAMISAFE CVHD	GWGVAVLVGVI	PHKEAVFKTHPMN	FLNERTLK	GTF FGNYKPI	R SDIPSVV	EKYMNKELEI	EKFITHTLP	FAEINKAFD	LMLKGEGL	RCIITMED
Fan	NIQAMI PAFE CVHD	GWGVAVLVGVI	PHKDAVFTTHPMN	FLNERTLK	GTF FGNYKPI	R TDIPSVV	EKYMNKELEV	/DKFITHQLP	FSQINKAFD	YMLKGEGI	RCIITMEE
Tre	SIQAMISAFE CVHD	GWGVAVLVGVI	PKKDDAFKTHPMN	FLNERTLK	GTF YGNYKPI	R TDLPNVV	EQYMKGELEI	EKFITHSIP	FSEINKAFD	YMLKGES I	RCIIRMEE
Psa	SIQAMISAFE CVHD	GWGVAYLVGVI	SKDDAFKTHPMN	FLNERTLK	GTF YGNYKPI	R TDLPNVV	EKYMKGELEI	LEKFITHTVP	FSEINKAFD	YMLKGESI	RCIIKHEE
Ath	SVQAMIQAFE CVHD	GWGVAVLVGVI	SKDDAFKTHPMN	FLNERTLK	GTF FGNYKP	K TDIPGVV	EKYMNKELEI	LEKFITHTVP	FSEINKAFD	YMLKGESI	RCIITMGA
Ani3	AEKPFQQATE YVRS	H GSVVAIGLE	PANAFLKAPVFTT	VVRMINIK	GSY VGN	R QDGVEAL	DF FARGLI	<b>APFKKAPL</b>	QDLPQIFE	LMGQGKIA	GRYVLEIPE
Anil	SEKPFQQATE YVRS	R GTIVAIGLE	PPDAYLKAPVINT	VVRMITIK	GSY VGN	R QDGVEAL	DF FARGLI	KAPFKTAPL	KDLPKIYE	LMEQGRIA	GRYVLEMPE
Scel	SEAAIEASTR YVRA	N GTTVLVGMI	PAGAKCCSDVFNQ	VVKSISIV	GSY VGN	R ADTREAL	DF FARGLV	CSPIKVVGL	STLPEIYE	KMEKGQIV	GRYVVDTSK
Sce2	SEAALEASTR YCRAI	N GTVVLVGLI	AGAKCSSDVFNH	VVKSISIV	GSY VGN	R ADTREAL	DF FARGLV	<b>(SPIKVV</b> GL	SSLPEIYE	KMEKGQIA	GRYVVDTSK
Kla	SEFAIEQSTN YVRSI	N GTVVLVGLI	RDAKCKSDVFNQ	VVKSISIV	GSY VGN	R ADTREAI	DF FSRGLV	CAPIHVVGL	SELPSIYE	KMEKGAIV	GRYVVDTSK
Sce3	SEAAISLSTE YVRP	C GTVVLVGLE	PANAYVKSEVFSH	VVKSINIK	GSY VGN	R ADTREAL	DF FSRGLI	KSPIKIVGL	SELPKVYD	LMEKGKIL	GRYVVDTSK
Spo	SPKSYEQAAG FARPO	G STMVTVSMI	PAGAKLGADIFWL	TVKMLKIC	GSH VGN	R IDSIEAL	EYVSRGLV	<b>(PYYKVQFF</b>	STLPDVYR	LMHENKIA	GRIVLDLSK
Zmo1	AKSAFNSAVE AIRA	G GRVVAVGLI	PPE KMDLSIPRL	VLDGIEVL	GSL VGTI	R EDLKEAF	QF AAEGKVI	(PKVTKRKV	EEINQIFD	EMEHGKFT	GRMVVDFTHH
CpoZ	NVN LSNDLK LLSC	G GRVIIVGCE	RGSIEINPRDTM	AKESTIS	GVSLFSSTKE	EFQQFASTI	QAG MELGW	VKPVIGSQYP	LEKASQAHE	NIIHSSGTV	GKTVLLM
OarG	VETSIQAGIY ATHS	G GTLVLVGLC	SEM TSVPLVH	AATREVDIK	GVF RY	C NTWPMAI	SMLASKSVN	VKPLVTHRFP	LEKALEAFE	TSKKGLGL	KVMIKCDPSDQNI
HsaG	AEASIQAGIY ATRS	G GTLVLVGLC	SSEM TSVPLLH	AATREVDIK	GVF RY	C NTWPVAI	SMLASKSVN	VKPLVTHRFP	LEKALEAFE	TFKKGLGL	KIMLKCDPSDQNI
RnoG	AESSVQDGIY ATHS	G GTLVVVGMO	SPEM INLPLVH	AAVREVDIK	GVF RY	C NTWPMAV	SMLASKTLN	VKPLVTHRFP	LEKAVEAFE	TAKKGLGL	KVMIKCDPNDQNI
EcoT	APPAFRTMLD TMNHO	G GRIAMLGI	PPSD MSIDWTK	VIFKGLFIK	GIY GREM	F ETW YKM	AAL IQSGLDI	LSPIITHRFS	IDDFQKGFD	AMRSGQSG	KVILSWD
Tbr	NADIMATAVK IVKP	G GTIANVNYH	GEGEVLPVPRLEW	GCGMAHKTIK	GGL CPGGI	R LRMERLI	DLV FYKRVDI	PSKLVTHVFR	GFDNIEKAFM	LMKD KP	KDLIKPVVILA
Aeu	TQATFEQSLR VLKP	G GTLSSLGVY	SSD LTIPLSAR	AAGLGDHKIN	TAL CPGGI	C ERMRRLI	NVI ESGRVDI	LGALVTHQYR	LDDIVAAYD	LFAN QR	DGVLKIAIKPH

Fig. 1. Continued

280

300

200

310

320

330

350

3/10

360

370

mining the coenzyme specificity for ADHs in general (Ohlsson et al. 1974; Brändén et al. 1975; Eklund et al. 1984), and data supporting this role have been obtained for yeast ADH1 (Fan et al. 1991). The second functionally relevant residue is Ser 48 or Thr 48, which is conserved for all the members except  $\zeta$  crystallin. Ser 48 is hydrogen-bonded to the hydroxyl group of the alcohol bound to the catalytic Zn and could function in a proton relay system to facilitate removal of the proton from the alcohol (Eklund et al. 1982). Substitution of Thr 48 with Ala in yeast ADH1 inactivates the enzyme (Plapp et al. 1991).

It appears, therefore, that the minimal requirements for alcohol dehydrogenases of this family include the following: several Gly residues at certain positions that are required to form a basic folded structure; residues that are necessary to bind the catalytic Zn and to modulate its electrostatic environment; an Asp that determines the specificity for NAD; and a Ser or Thr that facilitates proton removal from the substrate.

# Conservation of Ligands for the Noncatalytic Zinc

It is interesting that the four residues responsible for binding the noncatalytic Zn, including cysteines 97, 100, 103, and 111, were conserved among all the members except the three sorbitol dehydrogenases and the ADHs from A. eutrophus and T. brockii. Sorbitol dehydrogenase contains one zinc atom per subunit (Jeffery et al. 1984). The two bacterial ADHs probably also lack the noncatalytic Zn. A structural role was proposed for the noncatalytic Zn of ADHs (Drum et al. 1969; Brändén et al. 1975). In contrast to the catalytic Zn ligands, which originate from different parts of the protein chain, the noncatalytic Zn ligands are close to each other, with the four cysteines being within a 14-residue fragment. This organization also supports a structural role for the noncatalytic Zn because it has been found for some other proteins that catalytic metal atoms have ligands from different parts of a protein, whereas structural metal atoms have ligands from the same

No.	Residue	Exceptions	No.	Residue	Exceptions
31	Р	I(Aeu), L(RnoG, EcoT, Zmo1), Del(Osa1)	192	G	
35	Е	V(Fan), D(RnoG, EcoT, Tbr, Aeu), Q(Ani1, CpoZ)	196	A	C(Spo), L(HsaG, OarG, RnoG, EcoT, CpoZ)
46	С	N(CpoZ)	197	V/I	
47	G/H/R	P(CpoZ)	199	G	A(Aeu)
48	S/T	V(CpoZ)	200	L/A	I(Tbr), Q(Aeu)
49	D	E(Hsa6, CpoZ)	201	G	
51	H/Y	Y(HsaP), K(Hsa6)	202	G/A/P	
62	Р	N(Tbr), G(Asu, Zmo1)	203	V/L/I	
66	G		204	G	
67	Н	T(CpoZ)	207	V/A	N(OarG), T(HsaG, RnoG)
68	E	D(CpoZ)	212	K/R	T(Sce3)
71	G		216	A/G	N(Zmo1), Del(CpoZ)
73	v	I(Gga, Zma1, Pam & Aeu)	218	R	K(Zmo1, CpoZ), T(Aeu), Q(HsaG, OarG, RnoG), N(EcoT)
77	G		219	I/V	T(Ani3)
80	v		220	I/V/L	
86	G		221	G/A	V(HsaG, OarG, RnoG), I(EcoT)
87	D	Q(Aeu)	222	V/I	T(OarG, HsaG, EcoT, CpoZ)
89	V/A		223	D	G(Tbr), A(CpoZ)
97	С	D(Tbr), N(Aeu), R(HsaG, OarG, RnoG), Del(CpoZ)	228	K/R	C(Tbr), Q(Zmo1), G(CpoZ)
100	С	T(Tbr), S(Aeu), D(HsaG, OarG, RnoG), Del(CpoZ)	236	G	
103	С	V(Tbr), A(Aeu), Del(CpoZ)	261	G	K(HsaG, OarG, RnoG)
111	С	S(Tbr, OarG, HsaG), D(Aeu), Del(CpoZ), T(RnoG)	287	G	S(Spo)
144	S/G		292	v	I(HspA, Ani1, Ani3), L(Aeu, EcoT) Del(Hvu2)
146	F	M(Zmol), Q(Aeu), L(HsaG, OarG, RnoG), Y(CpoZ)	293	G	S(Spo), N(Tbr), Del(Hvu2)
159	K/H/R	I(Spo), P(Aue)	294	V/L	Del(Hvu2), M(Sce1, Spo, RnoG), C(CpoZ), I(EcoT), Y(Tbr)
160	I/V/L		359	F/Y	V(Hsa6), H(CpoZ)
169	V/A	G(HsaG, OarG, RnoG, CpoZ)	365	G	N(Spo), Del(Tbr, Aeu), S(CpoZ)
174	С	D(Tbr, Aeu, EcoT), E(HsaG, OarG, RnoG), I(CpoZ)	369	R/K	Del(Hsa6), D(Aeu)
178	Т	S(Osa2), V(HsaG, OarG, RnoG), N(EcoT)			

Table 2. Conserved and consensus residues among members of the Zn-coating ADH family

The numbers refer to the numbering system for horse liver ADH E. For enzyme abbreviations, see Table 1. Del: deletion

part of a protein (Matthews et al. 1974; Monaco et al. 1978; Vallee and Auld 1990; Kim and Wyckoff 1991). Whether the second Zn has a structural role remains to be explored, but the conservation of these residues indicates that they may be necessary at least for the eukaryotic ADHs. Thus, it seems that another major event during the evolution of ADHs was the acquisition of a capability for binding the noncatalytic Zn.

## Other Conserved Amino Acids

His or Tyr is found at position 51 except for human class II ADH and ADH6, which have Thr and Lys at this position, respectively. Thus, there is His in all the class I and tetrameric enzymes and Tyr in all the rest, including all the plant ADHs and class III ADHs. Although both His 51 and Tyr 51 can form a hydrogen bond to the 2'-hydroxyl group of the nicotinamide ribose of NAD (Eklund et al. 1990), they must function differently. His 51 appears to act as a base catalyst for alcohol oxidation through the proton relay system (Eklund et al. 1982; Ehrig et al. 1991; Park 1991; Plapp et al. 1991). A role for Tyr 51, however, is more difficult to assign. The Tyr may bind the coenzyme but might not be required for base catalysis under the physiological conditions where the enzyme functions. It may be relevant that Tyr is also aligned at position 51 in  $\zeta$  crystallin.

There are several other well-conserved positions. Asp is conserved at position 87 in all but the ADH of *A. eutrophus*, where it is Gln. Gly is conserved at position 199 except for the ADH from *A. eutrophus*, which has Ala. Lys or Arg is conserved at position 212 in all but ADH3 from *Saccharomyces cerevisiae*, which has Thr. Gly is found at position 261 for all but the sorbitol dehydrogenases, which have Lys. Gly is located at position 287 for all except for the ADH of *Schizosaccharomyces pombe*, which has Ser. Arg or Lys is conserved at position 369 except for *A. eutrophus* ADH, which has Asp. Although the structural or functional implications of these observations remain to be elucidated, they suggest directions for future studies.



# **Fig. 2.** Phylogenetic tree for 47 members of the Zn-containing ADH family. Percent standard deviation was 5.00 based on the progressive alignment. For enzyme abbreviations, see Table 1.

# Assignment of Insertions or Deletions

Some potentially important residues were not aligned at the same positions as in previous reports. Leu was aligned at position 93 for yeast ADHs, instead of Trp. Whether Trp or Leu was assigned to the position was controlled by the insertion of one Gly. Trp would be at position 93 if a Gly were inserted between residues 95 and 96 instead of between 89 and 90. (The insertion produces a gap between these residues in the other sequences.) Residue 93 has been shown to determine the size of the active site of horse liver ADH (Eklund et al. 1982) and has been assumed to restrict the activities of yeast ADHs toward larger substrates due to the bulky side chain of Trp (Brändén et al. 1975). More importantly, the assignment of Trp to position 93 in yeast ADHs is supported by experimental data obtained from mutant enzymes of yeast ADH1 in which Trp was replaced by Ala or Phe (unpublished data from this laboratory; Creaser et al. 1990). Apparently, the assignment of Leu for Trp at position 93 produced a "better but less reliable" local alignment. Inserting a Gly between 89 and 90 gives a local alignment with better similarity because Ile is then aligned at position 90 where Ile is also found in all the animal ADHs. The alternative assignments for the positions of gaps or insertions should emphasize that the alternatives remain hypothetical until the threedimensional structures are determined.

# The Evolutionary Tree

Another major result of this study is the evolutionary tree for 47 members of the Zn-containing ADH

 Table 3. Distances for animal and plant ADHs since their divergence

	Aı	nimal		Pla	ant
	From A	From C	-	From A	From B
Pha	26.87	17.37	Tae	26.86	7.87
HsaB	26.58	17.08	Hvu3	26.84	7.85
MacA	28.28	18.78	Hvu2	31.87	12.88
HsaA	28.90	18.78	Zma2	26.24	7.25
HsaC	26.59	17.09	Osa2	30.47	11.48
EcaS	27.12	17.62	Zmal	26.40	7.41
EcaE	26.59	17.09	Pam	26.45	7.46
RnoA	27.35	17.85	Hvul	27.27	8.28
MmuA	27.75	18.25	Osa1	27.71	8.72
Cia	27.01	17.51	Stu	26.06	7.07
Gga	26.31	16.81	Fan	28.01	9.02
Rpe	28.85	19.35	Tre	27.58	8.59
Hsa6	32.03	22.53	Psa	26.67	7.68
RnoX	23.07	13.57	Ath	24.16	5.17
MmuX	23.13	13.62			
HsaX	23.35	13.85			
EcaX	22.75	13.25			
HsaP	27.55	18.05			
Average	27 ± 2	17 ± 2		27 ± 2	8 ± 2

Distances were calculated by adding up the corresponding branch lengths shown in the evolutionary tree, which has a % standard deviation of 5.00. Branch points A, B, and C are as indicated in the tree. For abbreviations of enzymes, see Table 1

family. This tree was generated by the matrix-based method (Feng and Doolittle 1990), and was supported by an analysis using the nearest-neighbor procedure (Doolittle and Feng 1990). Like many other evolutionary trees for related sequences, this tree was produced assuming a common ancestor for all the members included. In general, there are two features that are important for an evolutionary tree: its topology or branch order, which shows how related members are grouped and have diverged from each other, and its distances or branch lengths, which should be proportional to the true evolutionary distances.

# Topology of the Tree

The topology of the tree for this family shows several features. First, the two clusters and four subclusters have distinct structural or species differences. All the dimeric ADHs were clustered and further subclustered into two groups for animal ADHs and plant ones; all the tetrameric ones were clustered and further subclustered into two groups, with one including enzymes having two zinc atoms per subunit (represented by yeast ADHs), and the other including enzymes having one zinc atom per subunit (represented by the sorbitol dehydrogenases). It has been shown that evolutionary trees based on threedimensional structures of proteins are almost identical to trees based on primary structures (Johnson et al. 1990). Thus, it is reasonable to assume that members of a cluster are more similar than nonmembers in their three-dimensional structures. In this regard, it is important to establish the threedimensional structure for at least one member from each subcluster. The three-dimensional structures of horse ADH E and the human ADH  $\beta$ 1 (Hurley et al. 1991) are the only ones known for this family.

Secondly, the topology of the tree shows that among all the animal ADHs included, the three classes of human ADHs diverged before speciation occurred for most animal ADHs. The divergence did not happen in humans, but instead in an an-



**Fig. 3.** Stereoview for locations of nine strictly conserved residues in the three-dimensional structure of horse alcohol dehydrogenase E isoenzyme. NAD<sup>+</sup> is bound in the cleft between the two domains of the subunit.

	o	٩	0 N	5	ន	8 9	8 K	3 2	ŝ	ស្ល	5	ß	5	ស្ត	ß	æ	ŝ	ß	ន	5	3 8	ន	ន	2	8	នខ	<b>ং</b> ন	8	ខ្ល	1	8 2	2 2	2	ន	ន	<b>4</b> 8	18	1 2	ន	ន	ខ្ល	5 6	. 1
ut	N	E	o <del>-</del>	5	5	8. 9		2 23	8	5	ģ	8	8	8 2	5	8	8	8	នេះ	s s	3 5	2 2	2	2	2	888	3 10	5	5	8 9	8 9	2 8	8	 	នេះ	5 5		5	g		8 1	2	8
ime	s	٩	0	ç	8		8 5			ŝ	ģ	ģ	5	8 8	2	ò	Ø.	8	5	ю. :	t g	Я КО	ģ	æ	g i	ιΩ μ	0.00	9	8	g ;	8 z		2	8	2	8 9			9	ž	g		. 24
ulign	-	<b>c</b> .		ы Б	<u>4</u>	90 9	n n p u	, 0 , 0		101 102	5	2	2	<u>4</u> 10	9	e g	yo yo	9 9	<u>4</u>	in i in i	9. 9 9. 9	t st	4	4	<u>4</u>	ω. Ω.	n n		ŝ	4	ο οι	)	Ω.	SI SI	SI I	n in	) 10 ) 10		Ω.	ю.		999 999	2
ise a	•	c	- 0		90 :	en e	ы. р. 92	, (v) 2 (2)	. 04 . 99.	ы 10	ŝ	2	5	ຍຄ	çe.	er ge	ge ge	2	en i			1 1	5	2	4	<u>4</u>	8 99 8 99	4	4	2 2	* 5		4	g	ຕ ຕ	5 S			ģ		ø :	4 0 4 5	
uirw	6	o	ه ص	10	5	90 9	n n	, (4 ) (9		4	4	se Se	2	0.0	5	4	Ω.	<b>X</b>	4	Ω Ω	אים אים	. 0	0 0	ŝ	10 10	5 5	t st	4	ŝ	***	t a		ŝ	1	ຕ ເຊິ່	0.5	) (0.		Ű	Q	10 1	9 L	N
e pa	- 	_	- ··	رم وي	0. 9	60 G	ы с р с	, 0 , 0	ເດ. ເຊ	4	ŝ	2	ο φ	0.0 0.0	4	4	ŝ	4	4 1	0. 0 03 0	NØ	1 01 1 01	ы С	4	<b>4</b>	ev e st e	2 02 2 02	4	51 10	5 10 10		-	Ω.	N N	01   E	n a	) 4   ~		Ξ	ю Þ	og e	20	20
n th		o	- 	1 1	2	n e h e	N 0 - 4	2 10	. 00	5.	2	2	2	0 0 0 0	20 20	4	4	er A	ດ ຕ່	0. 10 10 1		. n	ณ ณ	N D	ດ ຕໍ	ณ เ ณ เ	N (N N 17	ୁ ବ୍ୟ	2 2	010	v + 2.00	) <u>–</u>	ev ex	ณ		N 44 20 02	, w	4	eo go	ю Ю	ω. φ	00	2
fror	~	0	e -	. N	00 ·	in in Kong	n a pa	, 0 9	1 01 1 10	5 5	4	∾ 92	2	ດ ທ ອ ທ	ດ ເຊ	2 10	4	4	ດ ເຊິ່	0 0	N 0 ⊏ 0	1 01 1 01	ณ ณ	2	4	01 C	ч сч N <del>г.</del>	ີ ຕ	cv cv	ດ ( ດ (	3 G	· •	N N	сч Т	N	ο α	. 6	e v	~	9 9	ທີ່ ທີ່	n n N Q	5
<i>i</i> ere				v v	0	0 1 10 1	0 4 0 4	1 (N 1 (Q	0 0	4	0 0	0 0	2	0 0 20 0	- 4 - 0	ດ ຕ	4	0 0	4 4	ດ ( ()	4 C.	) 4 1 (V	4	4	4	מינ מיני	N N N N	01 10	N D	4 7	+ √ + 4	- cv - co	2	2	0	α	0 0	8	2	5	രം	4 0 0 0	8
al w	~	Ļ	0 /	ິ ເພ	ດເ ທີ່	4. () ()	50 0 et -c	1 4 1 (1	i Qi . D	N D	ດ ຕ	N 10	4	ແພ	i ini i ini	ŝ	4	N LO	ณ์ ณ	4 ( () ()	N O N O	1 0 1 0	N C	N N	(1) ( (1) (	01 0 4 4	N N t N	01 01	N N	01 7	- 4	i oi i io	CN CN	N N	N I	ہ ہ د	0	0 0	4	N D	থে । বা	0 N	4
gon		- -		in in	4	ດ ( (2) (	n n pa	9 9 9	9 9 9	5	4	5	4	N N B L	4	4	4	4	ດ ຕ	9 ( 19 (	ы с 9 с	າ ດ ວ ຕ	20 20	8 9	4	40	າ າ າ	୍	ณ ณ	20	- 4	1 0. 1 60	8	œ		5 0 - 0	0   0	0 0	22	4	40	אס אי מים	2
e diz	- -	- o	a (7	4	4 I	n i Di l	0.5	t un	1 N 1 N	4	4	0 10	4	0.4	ເ ເ	2 10	0 0	с С	ດ ຕ່	4 ( N (	NO	101 101	2 2	2	5 -	0.0 0.0	2 01	୍ ରା	CN CN	N 1	. 9	1 64 2 60	0	80	80 I	n n pg	. 4	4	cv go	ŝ	তা হাং	N N N N	4
/ the	-	 60		<u>4</u>	<u>4</u>	en o St i	N O N S	1 01 F 10	i Ni i Ni	2	2	<u>⊀</u>	8	2010	0 10	4	4	4	0 0	0 1	n o t c	. 0 . 0	ເນ ເກ	N N	ରା - ରା -		а 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	~		0, 0 0, 0	<u>v</u> a	. N	æ	8	60 ·		0 0 0	- <del>.</del>	~	N N	0 0	N N 9 <del>4</del>	8
elow		-	_	0 0	8	0. 0 20 1	~ ~			9	0 0	2	0 0	60 4 01 01	~	0 0	80	ດ ຍ	ດ ຄ	~ 1	יימ רימ	1 04 1 1-	0 0	8	8	~ c	N N 0 80	0 0	0 0	00	N (7. 0	, ø	0 0	2	~ ~	, , , , , , , , , , , , , ,	. ev . +-	0	сл СЛ	4	~ ~ ~	א הא הא מי	8
ğ		- "	<b>6</b> 0	2	2	0 0 1-1	0 0 - 1		ାର ଜୁନ୍ତ	5 0	2	с Q	0 0	~ ~ ~	່ ຕ ເຕ	0 0	0 0	2 10	2	01 0 N 0	יי שימ	1 01 1 02	4	6 6	8	0 0	и м т	່ດ ອ	0 0	~ ~	να	9 9 9	0 0	0 10	ເນ ຜ່	4 4 N N	- 14 - 14	0 0	0	ດ ຜ	~ (	0 N 0 N	7
Lho	-		~ `` E	. 4	4	۰ ۲ د د	•••	) Q	. 4	4	4	4	0 4	8 F 4 4	- 4 - 10	с С	4 10	4	0 4			, <del>.</del>		8 6	8	4,	о ю - ю	4	4	~ ,	~ ¤	) 4   ()	2	5	o i	0 0 0 4	· 4	0 0	∾ ø	2	0 0 N 0	20 M	~
nt. J		_	-	7 5	20	ທ ຄ	ы а ма	0 - N	010 ~	ъ В	7 5	2	4	44	ം ന	a) N	N N	ы N	4	к. о	0 a 7 a	) @ ) @	8 0	N	6		0 ~ 0 00	8 9	¢	۱ ح د	- a	. A	(i) (i)	4	64 I 10 I	61 6 0 10	i ni i no	61 60	6) 6)	در ۵۵	60 G	N N 0 0	0
ume			-	8 4	4	•	ю к 4 4	2 K0 F 4	. 4	4	<b>6</b>	6 4	4	9 4 4 4	- iO 	- 10	ю 	ю o	4	~ (	0 0 7 0	യെ പെ	ଷ ରା	۲- 80	8		0 FC N 00	. 0	9	et (	, 0 , 0	1 (N) - 10	0 0	4	ດ ທີ່	N N N N	0		۵ ۵	00 00	ດ ເ ອ ເ	50 50 10 50	∾ 0
aligr	-	Ű	~ ~	4	4	чч о (	e -	5 4 5 N	· 4	4	4	4	4	9 10 4 4	- 4 - 10	4	0 0	4	00   4	~ i	οά ~ 0	) 60 4	÷	8	ю .	60 0 N 0	0 ñ D	6	б в	ο f νο ι	с n о a		ŝ	ณั	4	N N M M	0 0 0	া ব	Ω Ω	а С	ຄ.	ດ ຄ	N.
ive		_	-	. 4	4	ৰ ক	4 <del>-</del>	1 4 . 0	, <del>4</del>	4	4	4	4	44	С	ю Ю	ю ю	ю Ю	₩ ₩	05 ( N 1	n d	່ວ ⊢ຕ	4	89 N	- 00	80 F 10	ະ 	2	к Ю	~ •	המ המ	i Ni Ni Ni	N M	οŭ α	ά Νι	N N - 4	। ত • च	i cù co	N G	δù m	0 0	ง ถุด ถุด	~
ress	-	ī	_	4	4	• •		5 97 5 05	, 4 , 4	4	4	4	4	44	4	с С	4	ы С	9 1		ים הים	) 4 ) 0)	4	8 0	4	со ч	~ 0 0		е е	4,	- a	וא מונ	n N	ດ ຜ	ເນ ຍ	0 0 0 4	00 0	60 60	Ω Ω	N N	60 0 60 0	אות אות	~
orog	- -	Š	~ ~	4	4 : 0 :	0 0	ה ל ה ל	1 41 5 05	. 4	4	ю о	9 4	4	9 0 0	വ ന	ю ю	ര	ю 0	- - - - - -	60 i N 1	50 a 0 c	i cò i cò	ю С	8	¢Ò	ة + دى	88 - 0	80 80	60 00	000	50 50 50 60	יא מי	2	R A	Ri i Ni i	ດ ດີ ແ	1 00 1 00		δi Ø	ivi so	ন ন ল ক	νίν ≺ αα	8
he J	N	Ē	- ··	4	4	0 0	5 7 7 7	, 4 ) 0)	0, 0,	4	<del>ب</del>	8	8	44	- 44 - 10	с С	4	4	80 ·	00 1 0 1	~ a	) 60 ) [~	~	8			מי מיע	~	2 7	0.0	0 0. 0 0	ง ณ ม เม	0 0	4	יוס המ	0 4 0 4	· 4	0 0	8 8	0 0	~ ~ ~	אס וא סכו מ	_
m 1	0	-	а	. р.	8 8	0, 0		, 40 200	, d) , d	8	6 2	6) 4	8	00 10 4 4	4	4	4 0	4 0	60 : 47 :	00 ( 10 (	9 0 9 0	, 4 ) 60	æ	~	6 - 6 -	60 0 10 1	0 00 4 10	. e	4	2 4	0 0 0 0	i N b Op	4	8 0	ເຊ ເດີ	0 4 0 4	· 4	5 5	0 0	2	201	N N - 0	~
e fre	т	>	- 	8	<b>6</b> 4	יידי סיס	99 200	) 00 1 70	. 4	4	4	ф Ф	<b>6</b>	α0 44	- 15	ю 2	4	с С	. e	8 9	5 0 7 0	, თ	4	8	9 9	4 ( 0 0	0 0 9 <del>-</del>	രം	е С	~ <	0 0 9 0	ייים הגיפ	27 20	N 20	ດ : ຜີ	0 4 0 4	י מ פ	90 10	4	8 8	~ ~	אי אי סים	6
wer	•	ø	F	9	0,9	9 9 9 0		יי ס כ	• •	9 0	5	ф Ф	9 9	90	- 41 - 101	0 0	4	4	ф, Ф	00 1 00 1	ο α 	, ¢	4	8 8	8 1 9 1	000 107	0 00 7 N		4	~ u	0 a		2 2	N 20	0. 1 10 1	л л р ч	. 0	0	4	0 0	~ (	N N N N	6
nal	N	ε	– م –	9	9 I	9 9	* • 2 g	່ວ	9	S S	5	v Ø	9 9	9 0	ы С	4	5 C	4	φ. 	ω. Ω		- 4 - 00	4	8	ю. Ю.	ю <u>к</u> е о	0 00 0 <del>10</del>	0	8 0		0 0 2 00	1 (N 1 (Q	4	4	יוס וסו	0 0 0 0	0	4	4	ດ ມ	010	ניס ויס סו ס	
liago	-	đ	0	g	2	89	2 9	2 92	9 9	8	Q	м Ф	e e	9 9 9 9	2	4	<u>4</u>	4	9 9	۳ د	± 9		99 99	Ω Ω		in c	y Q y		g	80 9 80 9	9 9	1 01 1 10	2	cv gg		5 G		5	2	99 100	0 0 00 0	N N	·
he d	I	s	a ۵	8	60	5 2		2 8	2	<u>ت</u>	5	ч 0	ģ	9 9 9 9	2	ມ ຜ	ы М	۵ ۵	ч !	0 0 0	2 9	 	8	ω Ω	9 9	0, 9 0, 9	2 02 2 02	8	8	9 1	2 9	9 64 1 92	5	2	5. I	n n n n	์ ณ เญ	হ	0 0	2	in in In in	, v 2 00	4
ve t	ш	o	a ×	ß	50	2	2 2	2 2	. ອ ເສ	8	ğ	5	g	59	2	5	Ω Ω	Ű,	g :	* *	* *	াৰ	т Т	4	ৰ :	4 9 4 9	2 K2	जू	4	93	re. tg		SO SO	ο Ω		0 51 0 51	- <b>4</b>	01 10	ς. Ω	0 0	0 0 5 0	າ ທີ່ ທີ່	0
abo	r	ø	a×	ي ي	¥ :	2 2	2 2	2 2	. T	0 0	е М	Ω Ω	0 0	00	4	e G	0	ŝ	ຍ ດູ	4 1	ою 0 ю		4	4	0	4 9 6 4	2.4 0.0	5	9 9	4 6	9 m 9 m	· 0	α φ	ω Ω		20 41 20 41	4	9	ŝ	N D	~ ~	ი - ი	-
nes	×	ε	5 X	7	ອ ອ	5 5 5 5	מע	,	. 02	ы В	ш М	Ω Ω	0 0	5 90 5 90	5	0,	g	5	ຍ ເຊິ	4 :	t s	: 22		4	9 9	4 9 10 11	лы; 2ю		ຍ ຍ	0 0	) 00 ) 00	) (N ) (Q	60 60	ເບ ເວ	90 (	р 4 м м	. 0	0 0	ŝ	7	~ ·		4
Val	œ	_ _	• ×	ŝ	21 9	טים פים	2 5		. 4	9 0	Ω Ω	4	φ t	5 9 5 9	ω	Ŀ	400	യ പ	0 0	4 i 10 i	ою ою	) 4 () ()	.st	ມ ມ	47 I	ю «	סור סור	- CO - TT	с С	4 4	, 0 1 0	1 0. 1 60	ດ ທ	ດ ທ່	ovi pori	- 4	- 4 - N	01 10	2 2	8	~ *	ი 	2
ns).	r	ø	a 0	ي م	ຍ : ທີ່	9 9 9 9	ts	ι α ι α	ີ 22	8	<b>1</b>	ເ ເ	e N	ຍ ດ	9	0 00	on Og	on Q	9 9		5 KG 2 12	. w	u y	5	ю. Б. 9	eo e No e	מו קסיק	4	4	מש		• • •	20 20	4	N ( 10) (	N N 0 90	CN GO	ري. مر	N D	0 0	0. 0 00 0	0 M	0
egic	œ	۵	8	5	9 U	2 9 2 9	n a		8	8	ģ	ģ	ģ	u g	5	5	ເວ ເປ	е 	90 ( 90 (	9 9 20 0	99 99	9	9	4	99	w 9	, 4 9 9	8	4	60 g	r ev P op	- 01 - 02	01 02	0	N I Rat		0 0	۰. ح	2 2	ດ. 92	ດເຊ ທີ່ເ	N N N 00	2
r bə	Σ	ε	<b>∍</b> <	2	2 2	2 2	2 2	. ພ	8	8	ñ	9 9	Ψ	8 N 10	7	9 22	0 0	ອ ສໍ	9 99	9 9 9 9	29	. <u>.</u>	9 9	۹ 9	99 9 99 9	9 9 9 9	1 4 2 92	1	۲ 92	9 9 9 9	r ov 2 ge	. 01 1 02	8	99. ·	e e	8 6	0	20 20	0 10	2	01 0 90 1	N 01 N 01	8
lign	œ	c	• •	,~ נא	N 1	5 9	מע	i Q		=	ш М	ŵ	ø	80	3	9 0	0	ຍ 22 :	ດ ດ	20	р (с) 19 (с)	, 4 , 9	8	ч ф	9 I	0 V	. 9	9 9	ч 80	- 0	, c	. 0.	0 0	о 10	ovi e po e	יייי ספים	2	~	ດ ດ	2	0, 10	~ 0 ~ 0	ő
on a	ш	υ	¢Ш	5 1	21	2 1	. 9	, w	5	5	w	Q	ŝ	8 Q 0 U	8	ພ ດູ	ຍ ດ	ພ ເຊ	5	5 9	, 4 1 1	• •	ч 	9	<b>T</b> 1	5 5	- 9 - 9	92 10	9 0	0 9		. 4	e Se	5	e e	N N 8 60	9	∾ ¢	2	∾ ₹	0 0 1- 9	ч ю 0 0	4
sed	ω	υ	ສູດ	20	2 9	2 9 2 9	8 5	g	2	ű	5	ŝ	g 0	8 X	8	м М	0	ສ ສ		5 3			2	5		9 9	5 0 1 1	4	9 92	5 9	2 90		(N) 92	90 I	n i n i		. 4	ST.	3	4	0 0 0-0	. 0	~
(ba	τ	s	<b>م</b> ح	<u>م</u>	2 9	2 2	tg		Ű	Ŀ	22	5	g 2	84	4	2 2	4	2 2	g i	2 9	2 0	9	9	g	99		פים	4	8	5 9	2 9		2		N G	N N S 99	сч gg	دم 99	۲۵ وو	©.	0. C 0. C	אימ אוס ב	4
ities	Σ	Ø	υ <b>∢</b>	2	5 3	4 3	t Q	. w	4	g	8	ğ	ଳ ପ	6 4 0 0		ĝ	. e		o Q	29	29		9 9	Q Q	0 9	9 9 9 9	, 4 , 60	4	89 14	ы с т с	, 0	2	0 9	N 1	N 0 N 9	v 0. 0.00	0 0	0 0	N D	0 0	010	~ 4 0 0	0
lenti	I	s	αO	4	29	2 3	t 0.	ପ୍ଲ	5	2	e g	ພ ດູ	4	0 0 0 0	0 0	е N	ຕ ຕ	ຜູ ຕູ	9 9	29	29	• •	9	ю 5	ю. 9.9	2 9 2 9	, 4 , 0	4	е) 4	- 0	- 04	4	2	2	n c p c	2 N 2 N	2	0 0	2 20	ດ ຍ	ດເດ ທີ່ດີ	ניס ע גרס מ	4
nt ic	r	ø	a 00	4	21		ঁয	. z	4	g	2	ພ ດູ	0 0	8 9 8 9	g	ğ	ອ ອ	ດ ເຊິ່		0 4 5 2		- 92 - 92	9	10 12	5		2 5	8	9 9	0.4 0.4	ი თ ა თ	5	2	N 1	N C P C		2	∾ 80	ດ ຍ	0 ·	ou c op c	0 0 0	2
ercei	۵.	e ·	 ه ص	4	* ( *	~	50 0	2	S Q	9 9	37 6	3 15	ສ ສ	5 9 2 9	ມ ຜູ	ы б	ຍ ຜູ	ພ ( ດູ	י ש בי		, 10 ; 17	4	s g	5		а ю с 9	20 20	6	9	» م ج م	1 03 2 55	0 90	2	01 i D. i	n i pr		7 2	2	ci Q	2	ou ĉ ge d	94 97	4
Å	ø	5		1 7	·~ ,	0 t 5	t 4		ы М	4 2	74 E	۳ ۳	ω N	8 9 9 9 9 9	4 8	8	י تە 1	ι Ψ	ט קינ	2 9	29 29	9 9	8	9 10	оч Q (	טינ קס	, no 2 oo	4	17 5	n v	, 0 20	CN GQ	ŝ	ດ ( ທູງ	ы gr	9 09 5 09	2	Q Q	5	2	w C gg o	v ro g oo	4
e 4.	o		ಪ		، ج	- r t 5	. 4	5	N N	5 7	5	2	5	δõ AA	ő 6	ž	ι Ω Ω	ι Ω	9 ( 9 (	2 9	, υ λ	97 17	9	ŝ	ы. Q	y v	200	7 4	4 71	Ωď	2 E 1 C	4	2	αı Ω	a c a r	, 0 20	2	cv go	د، 92	2	מית קאס	v OV b OD	4
Lab	-					- r . r	i n	۰ ۲	4	5	γ μ	•	Ā	ω~, ~	×	ž	יש אי	، س ب	יי ג	υ u •	, 4	4	1 5	e) N	ю 1 0	4 4	- TE	4	4	» د	, ()	CV.	(N) (T)	(N ) (N 4	м ( 15)		~	N	~	CV I	CN C	40 v 	2
				Q B	Gga 19		Hsa(	Mac	Hsa	Ecal	Ecal	Bro	Ĕ	H Be	Rno.	Ĩ	Hsa.	ECB.	EST 1		Part Part	Hvu	Osa	Zma	Ť	un u	H	Psa	Tre	₽Š	j ja	Аец	Hsa(	o a		See 1	Sce	Kla	Sce	Ani3	Anti Bag	R R	å

cestor that was common to the animals (all vertebrates) whose ADHs have been included in this study. This implies that these three classes of enzymes may exist or have existed, not necessarily in active form in many, if not all, vertebrates. This implication is consistent with an earlier investigation of class III ADH from rat liver (Julià et al. 1988).

Thirdly, the topology of this tree shows that human class I ADHs ( $\alpha$ ,  $\beta$ , and  $\gamma$ ) diverged before human, baboon, and monkey class I ADHs, supporting the proposal that duplications of the gene for primate class I ADH could have predated primate radiation (Trezise et al. 1989). This indicates that there should be isoenzymes of class I ADH in primates other than humans. This implication is supported by evidence for the existence of five major forms of baboon class I ADH (Holmes et al. 1986; Holmes and VandeBerg 1986). It is also consistent with the analysis of chromosomal DNA samples derived from various primates (Yasunami et al. 1990).

Finally, the newly identified human ADH6 (Yasunami et al. 1991) also diverged before speciation occurred for most animal ADHs. This enzyme may be the stomach  $\mu$  or  $\sigma$  ADH (Chen and Yoshida 1991), and its position in the present tree suggests that other vertebrates may have similar enzymes.

# Branch Lengths of the Tree

If the topology reveals order of divergence, the analysis of branch lengths should then give estimates of the dates of divergence, provided that an evolutionary clock is available. Thus, if we assume that animals and plants diverged 1000 million years ago (Carroll 1988), it can be estimated that human ADH  $\pi$ , ADH  $\chi$ , and ADH6 diverged about 640, 630, and 520 million years ago, respectively.1 Therefore, the divergence of three classes of human ADHs occurred about 600 million years ago, which approximates the estimated time when vertebrates diverged from invertebrates (Carroll 1988). This estimated divergence time for the three classes of ADHs is about 400 million years earlier than the estimated times when mammals and birds, respectively, diverged from reptiles. Similarly, the divergence of human class I ADHs can be estimated to have occurred about 80 million years ago, which coincides with the time of divergence for primates and rodents (Carroll 1988). These analyses support the idea that three classes of ADHs may exist or may have existed in many vertebrates, and that isoenzymes of class I ADH may exist or have existed in all primates. More specifically, these analyses imply that the three classes of ADHs may exist or have existed in all mammals, birds, and reptiles. Supporting this implication is the fact that class I ADHs have been identified in mammals, birds, and amphibians, and that class III ADHs also exist in rat, mouse, horse, and human species (Julià et al. 1988; Kaiser et al. 1989; Edenberg et al. 1991).

The proposal that ADHs of three different classes may be ubiquitous in most or all vertebrates has also been made previously, based on the estimated divergence time for the ADH of frog, which was 430 million years ago (Cederlund et al. 1991). Frog ADH is the first class I ADH to have diverged (among those included) as shown in the present tree. Its divergence time can be estimated as being about 430 million years ago. The two estimations are identical; both indicate that the three classes diverged before the radiation of vertebrates. It should be emphasized that estimations of divergence times based on the branch lengths are approximate. This is because the number of sequences included is limited, and more importantly because the evolutionary rate for each ADH does not have to be constant. Nevertheless, it is encouraging to observe that the estimates made in this work are consistent with or supported by several earlier studies.

Further analysis of branch lengths reveals varying degrees of conservation for different groups of this family. The average distances since the beginning of their radiation are 17 and 8 for the animal ADHs and the plant ones, respectively. Thus, it appears that these plant ADHs are more conserved than the animal ones.

# PAM Units

The values of PAM units are 4.4 [(165/378) × 10] for plant ADHs<sup>2</sup> and 6.9 [ $(259/375) \times 10$ ] for animal ones. These were calculated based on the progressive alignment and the assumption that animals and plants diverged 1000 million years ago (Carroll 1988). Normally, these calculated values should be corrected for multiple mutations at a site, but values less than 10 need not be. The PAM values for ADHs are higher than those for histone IV (0.09) and glyceraldehyde 3-phosphate dehydrogenase (2.2), comparable with those for insulin (3.5) and trypsinogen (5.1), and lower than those for lysozyme (10) and hemoglobin  $\alpha$  and  $\beta$  chains (14) (Dayhoff 1976). Similar analysis can be applied to compare ADHs of class III with those of class I. The PAM value for four class III enzymes from four different species (human, horse, mouse, and rat) is 1.1  $[(41/375) \times$ 

<sup>&</sup>lt;sup>1</sup> The average distance for 18 animal ADHs since the class II ( $\pi$ ) enzyme diverged is 17.2: 17.2 × (1000/26.7) = 644

<sup>&</sup>lt;sup>2</sup> The total number of conserved residues is 213 for the plant ADHs. The average sequence length of the plant ADHs is 378.  $[(378-213)/378] \times 100 \times 10^8/10^9 = (165/378) \times 10 = 4.4$ 

10], whereas the PAM value for four class I enzymes from the same four species is 2.4 [(90/375)  $\times$  10]. This indicates that the class III ADHs are about twice as conserved as the class I enzymes. This observation is consistent with the results of a previous study, where it has been found that class III ADHs are less variable as compared to class I enzymes (Kaiser et al. 1989).

Human class I ADHs are the major enzymes responsible for ethanol oxidation in the liver, with class II ADH contributing less than 15% of the metabolism (Li et al. 1977). The human class III ADH is almost inactive on ethanol although it is as active as other human ADHs for alcohols with longer chains (Parés and Vallee 1981). Recently, the class III enzymes have been identified as glutathione-dependent formaldehyde dehydrogenases (Koivusalo and Uotila 1991).

Limited information is available about the physiological roles of the ADHs. Yeast ADH1 catalyzes the terminal step in glycolytic fermentation. Rice and maize ADHs are induced under anaerobic conditions where glycolysis is required (Ricard et al. 1986; Kadowaki et al. 1988; Xie and Wu 1989). Human ADH class I enzymes, which have broad substrate specificities, have been suggested to play a general role in the detoxification of various hydroxylated compounds (Kassam et al. 1989). Although the physiological roles of the ADHs remain to be fully established, this study clearly shows that selection pressures exist for plant and animal ADHs, implying that they have specific and important metabolic roles.

Acknowledgments. We thank Drs. Da Fei Feng and Russell Doolittle for providing the programs, and Zhongwen Chen for adapting these programs on the VAX system. This work was supported by USPHS grant AA06223.

### References

- Aronson BD, Somerville RL (1989) The primary structure of *E. coli* L-threonine dehydrogenase. J Biol Chem 264:5226– 5232
- Bennetzen JL, Hall BD (1982) The primary structure of the *Saccharomyces cerevisiae* gene for alcohol dehydrogenase I. J Biol Chem 257:3018-3025
- Borrás T, Persson B, Jörnvall H (1989) Eye lens ζ crystallin relationships to the family of "long-chain" alcohol/polyol dehydrogenases. Protein trimming and conservation of stable parts. Biochemistry 28:6133–6139
- Brändén CI, Jörnvall H, Eklund H, Furugren B (1975) Alcohol dehydrogenases. In: Boyer PD (ed) The enzymes, ed 3, vol XI. Academic Press, New York, pp 104–190
- Carroll RL (1988) Vertebrate paleontology evolution. Freeman, New York
- Cederlund E, Peralba JM, Parés X, Jörnvall H (1991) Amphibian alcohol dehydrogenase, the major frog liver enzyme. Relationships to other forms and assessment of an early gene

- Chang C, Meyerowitz EM (1986) Molecular cloning and DNA sequence of the *Arabidopsis thaliana* alcohol dehydrogenase gene. Proc Natl Acad Sci USA 83:1408–1412
- Chen CS, Yoshida A (1991) Enzymatic properties of the protein encoded by newly cloned human alcohol dehydrogenase *ADH6* gene. Biochem Biophys Res Commun 181:743–747
- Crabb DW, Edenberg HJ (1987) Complete amino acid sequence of rat liver alcohol dehydrogenase deduced from the cDNA sequence. Gene 48:287-290
- Creaser EH, Murali C, Britt KA (1990) Protein engineering of yeast alcohol dehydrogenases; effects of amino acid changes at positions 93 and 48 of yeast ADH1. Protein Eng 3:523– 526
- Dayhoff MO (1976) Survey of new data and computer methods of analysis. In: Dayhoff MO (ed) Atlas of protein sequence and structure, vol 5, suppl 2. National Biomedical Research Foundation, Washington DC, pp 1–8
- Dayhoff MO, Schwartz RM, Orcutt BC (1978) A model for evolutionary change. In: Dayhoff MO (ed) Atlas of protein sequence and structure, vol 5, suppl 3. National Biomedical Research Foundation, Washington DC, pp 345-358
- Dennis ES, Sachs MM, Gerlach WL, Finnegan ET, Peacock WJ (1985) Molecular analysis of the alcohol dehydrogenase 2 (*Adh*2) gene of maize. Nucleic Acids Res 13:727–743
- Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387-395
- Doolittle RF, Feng DF (1990) Nearest neighbor procedure for relating progressively aligned amino acid sequences. Methods Enzymol 183:659–669
- Drum DE, Li TK, Vallee BL (1969) Considerations in evaluating the zinc content of horse liver alcohol dehydrogenase preparations. Biochemistry 8:3783-3797
- Edenberg HJ, Zhang K, Fong K, Bosron WF, Li TK (1985) Cloning and sequencing of cDNA encoding the complete mouse liver alcohol dehydrogenase. Proc Natl Acad Sci USA 82: 2262–2266
- Edenberg HJ, Brown CJ, Carr LG, Ho WH, Hur MW (1991) Alcohol dehydrogenase gene expression and cloning of the mouse χ-like alcohol dehydrogenase. Adv Exp Med Biol 284: 253–262
- Ehrig T, Hurley TD, Edenberg HJ, Bosron WF (1991) General base catalysis in a glutamine for histidine mutant at position 51 of human liver alcohol dehydrogenase. Biochemistry 30: 1062–1068
- Eklund H, Brändén CI (1987) Alcohol dehydrogenase. In: Jurnak FA, McPherson A (eds) Biological macromolecules and assemblies, vol 3. Active sites of enzymes. John Wiley, New York, pp 73–142
- Eklund H, Plapp BV, Samama JP, Brändén CI (1982) Binding of substrate in a ternary complex of horse liver alcohol dehydrogenase. J Biol Chem 257:14349-14358
- Eklund H, Samama JP, Jones TA (1984) Crystallographic investigations of nicotinamide adenine dinucleotide binding to horse liver alcohol dehydrogenase. Biochemistry 23:5982– 5996
- Eklund H, Horjales E, Jörnvall H, Brändén CI, Jeffery J (1985) Molecular aspects of functional differences between alcohol and sorbitol dehydrogenases. Biochemistry 24:8005–8012
- Eklund H, Horjales E, Vallee BL, Jörnvall H (1987) Computergraphics interpretations of residue exchanges between the  $\alpha$ ,  $\beta$  and  $\gamma$  subunits of human liver alcohol dehydrogenase class I isozymes. Eur J Biochem 167:185–193
- Eklund H, Müller-Wille P, Horjales E, Futer O, Holmquist B, Vallee BL, Höög JO, Kaiser R, Jörnvall H (1990) Comparison of three classes of human liver alcohol dehydrogenase. Eur J Biochem 193:303–310

Ellison NW (1989) EMBL Data Library, S06200

- Estonius M, Karlsson C, Fox EA, Höög JO, Holmquist B, Vallee BL, Davidson WS, Jörnvall H (1990) Avian alcohol dehydrogenase: the chicken liver enzyme. Eur J Biochem 194: 593–602
- Fan F, Lorenzen JA, Plapp BV (1991) An aspartate residue in yeast alcohol dehydrogenase I determines the specificity for coenzyme. Biochemistry 30:6397-6401
- Feng DF, Doolittle RF (1990) Progressive alignment and phylogenetic tree construction of protein sequences. Methods Enzymol 183:375–389
- Ganzhorn AJ, Plapp BV (1988) Carboxyl groups near the active site zinc contribute to catalysis in yeast alcohol dehydrogenase. J Biol Chem 263:5446-5454
- Good AG, Pelchen LE, Crosby WL (1988) Nucleotide sequence of a complete barley alcohol dehydrogenase I cDNA. Nucleic Acids Res 16:7182
- Gwynne DI, Buxton FP, Sibley S, Davies RW, Lockington RA, Scazzocchio C, Sealy-Lewis HM (1987) Comparison of the cis-acting control regions of two coordinately controlled genes involved in ethanol utilization in *Aspergillus nidulans*. Gene 51:205-216
- Ha BD, Buffardd D, Breda C, Esnault R (1989) EMBL Data Library, S06693
- Holmes MA, Matthews BW (1981) Binding of hydroxamic acid inhibitors to crystalline thermolysin suggests a pentacoordinate zinc intermediate in catalysis. Biochemistry 20:6912– 6920
- Holmes RS, Courtney YR, VandeBerg JL (1986) Alcohol dehydrogenase in baboons: tissue distribution, catalytic properties, and variant phenotypes in liver, kidney, stomach, and testis. Alcoholism 10:623–630
- Holmes RS, VandeBerg JL (1986) Ocular NAD-dependent alcohol dehydrogenase and aldehyde dehydrogenase in the baboon. Exp Eye Res 43:383-396
- Höög JO, Hedén LO, Larsson K, Jörnvall H, von Bahr-Lindström H (1986) The  $\gamma_1$  and  $\gamma_2$  subunits of human liver alcohol deydrogenase. cDNA structures, two amino acid replacements, and compatibility with changes in the enzymatic properties. Eur J Biochem 159:215–218
- Höög JO, von Bahr-Lindström H, Hedén LO, Holmquist B, Larsson K, Hempel J, Vallee BL, Jörnvall H (1987) Structure of the class II enzyme of human liver alcohol dehydrogenase: combined cDNA and protein sequence determination of the  $\pi$  subunit. Biochemistry 26:1926–1932
- Hurley TD, Bosron WF, Hamilton JA, Amzel LM (1991) Structure of human  $\beta_1\beta_1$  alcohol dehydrogenase: catalytic effects of non-active-site substitutions. Proc Natl Acad Sci USA 88: 8149–8153
- Ikuta T, Fujiyoshi T, Kurachi K, Yoshida A (1985) Molecular cloning of a full-length cDNA for human alcohol dehydrogenase. Proc Natl Acad Sci USA 82:2703–2707
- Ikuta T, Szeto S, Yoshida A (1986) Three human alcohol dehydrogenase subunits: cDNA structure and molecular evolutionary divergence. Proc Natl Acad Sci USA 83:634–638
- Jeffery J, Chester J, Mills C, Sadler PJ, Jörnvall H (1984) Sorbitol dehydrogenase is a zinc enzyme. EMBO J 3:357-360
- Jendrossek D, Steinbüchel A, Schlegel HG (1988) Alcohol dehydrogenase gene from *Alcaligenes eutrophus*: subcloning, heterologous expression in *E. coli*, sequencing, and location of Tn5 insertions. J Bacteriol 170:5248–5256
- Johnson MS, Sutcliffe MJ, Blundell TL (1990) Molecular anatomy: phyletic relationships derived from three-dimensional structures of proteins. J Mol Evol 30:43-59
- Jörnvall H (1970) Horse liver alcohol dehydrogenase. Eur J Biochem 16:25-40
- Jörnvall H, Persson M, Jeffery J (1981) Alcohol and polyol dehydrogenases are both divided into two protein types, and structural properties cross-relate the different enzyme activities within each type. Proc Natl Acad Sci USA 78:4226–4230

- Jörnvall H, Persson B, Jeffery J (1987) Characteristics of alcohol/polyol dehydrogenases. Eur J Biochem 167:195–201
- Julià P, Parés X, Jörnvall H (1988) Rat liver alcohol dehydrogenase of class III. Eur J Biochem 172:73-83
- Kadowaki KI, Matsuoka M, Murai N, Harada K (1988) Induction of two alcohol dehydrogenase polypeptides in rice roots during anaerobiosis. Plant Science 54:29–36
- Kaiser R, Holmquist B, Hempel J, Vallee BL, Jörnvall H (1988) Class III human liver alcohol dehydrogenase: a novel structural type equidistantly related to the class I and class II enzymes. Biochemistry 27:1132–1140
- Kaiser R, Holmquist B, Vallee BL, Jörnvall H (1989) Characteristics of mammalian class III alcohol dehydrogenases, an enzyme less variable than the traditional liver enzyme of class I. Biochemistry 28:8432–8438
- Kaiser R, Nussrallah B, Dam R, Wagner FW, Jörnvall H (1990) Avian alcohol dehydrogenase. Characterization of the quail enzyme, functional interpretations, and relationships to the different classes of mammalian alcohol dehydrogenase. Biochemistry 29:8365–8371
- Karlsson C, Jörnvall H, Höög JO (1991) Sorbitol dehydrogenase: cDNA coding for the rat enzyme. Eur J Biochem 198: 761–765
- Kassam JP, Tang BK, Kadar D, Kalow W (1989) In vitro studies of human liver ADH variants using a variety of substrates. Drug Metab Dispos 17:567–573
- Keshav KF, Yomano LP, An H, Ingram LO (1990) Cloning of the Zymomonas mobilis structural gene encoding alcohol dehydrogenase I (adhA): sequence comparison and expression in E. coli. J Bacteriol 172:2491–2497
- Kim EE, Wyckoff HW (1991) Reaction mechanism of alkaline phosphatase based on crystal structures. Two-metal ion catalysis. J Mol Biol 218:449-464
- Koivusalo M, Uotila L (1991) Glutathione-dependent formaldehyde dehydrogenase: evidence for the identity with class III alcohol dehydrogenase. Adv Exp Med Biol 284:241–251
- Li TK, Bosron WF, Dafeldecker WP, Lange LG, Vallee BL (1977) Isolation of  $\pi$ -alcohol dehydrogenase of human liver: is it a determinant of alcoholism? Proc Natl Acad Sci USA 74:4378–4381
- Light DR, Dennis MS, Forsythe IJ, Liu CC, Green DW, Kratzer DA, Plapp BV (1992) GenBank, M81807
- Llewellyn DJ, Finnegan EJ, Ellis JG, Dennis ES, Peacock WJ (1987) Structure and expression of an alcohol dehydrogenase 1 gene from *Pisum sativum*. J Mol Biol 195:115–123
- Matthews BW, Weaver LH, Kester WR (1974) The conformation of thermolysin. J Biol Chem 249:8030-8044
- Matton DP, Brisson N (1990) Nucleotide sequence of two potato alcohol dehydrogenase cDNAs. Nucleic Acids Res 18: 3070
- McKnight GL, Kato H, Upshall A, Parker MD, Saari G, O'Hara PJ (1985) Identification and molecular analysis of a third *Aspergillus nidulans* alcohol dehydrogenase gene. EMBO J 4: 2093–2099
- Mitchell LE, Dennis ES, Peacock WJ (1989) Molecular analysis of an alcohol dehydrogenase gene from chromosome 1 of wheat. Genome 32:349-358
- Monaco HL, Crawford JL, Lipscomb WN (1978) Three-dimensional structures of aspartate carbamoyltransferase from *E. coli* and of its complex with cytidine triphosphate. Proc Natl Acad Sci USA 75:5276–5280
- Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 48:443–453
- Ohlsson I, Nordström B, Brändén CI (1974) Structural and functional similarities within the coenzyme binding domains of dehydrogenases. J Mol Biol 89:339–354
- Parés X, Vallee BL (1981) New human liver alcohol dehydrogenase forms with unique kinetic characteristics. Biochem Biophys Res Commun 98:122–130

- Park DH (1991) PhD Thesis, Structure and function of isoenzymes of horse liver alcohol dehydrogenase. The University of Iowa, Iowa City
- Park DH, Plapp BV (1991) Isoenzymes of horse liver alcohol dehydrogenase active on ethanol and steroids. J Biol Chem 266:13296-13302
- Peretz M, Burstein Y (1989) Amino acid sequence of alcohol dehydrogenase from the thermophilic bacteria *Thermoanaerobium brockii*. Biochemistry 28:6549–6555
- Plapp BV, Ganzhorn AJ, Gould RM, Green DW, Jacobi T, Warth E, Kratzer DA (1991) Catalysis by yeast alcohol dehydrogenase. Adv Exp Med Biol 284:241–251
- Rao PV, Krishna CM, Zigler JS (1992) Identification and characterization of the enzymatic activity of ζ-crystallin from guinea pig lens. J Biol Chem 267:96-102
- Rees DC, Lewis M, Honzatko RB, Lipscomb WN, Hardman KD (1981) Zinc environment and cis peptide bonds in carboxypeptidase A at 1.75 Å resolution. Proc Natl Acad Sci USA 78:3408-3412
- Ricard B, Mocquot B, Fournier A, Delseny M, Pradet A (1986) Expression of alcohol dehydrogenase in rice embryos under anoxia. Plant Mol Biol 7:321-329
- Russell DW, Smith M, Williamson VM, Young ET (1983) Nucleotide sequence of the yeast alcohol dehydrogenase II gene.
   J Biol Chem 258:2674–2682
- Russell PR, Hall BD (1983) The primary structure of the alcohol dehydrogenase gene from the fission yeast *Schizosaccharomyces pombe*. J Biol Chem 258:143–149
- Saliola M, Shuster JR, Falcone C (1990) The alcohol dehydrogenase system in the yeast, *Kluyveromyces lactis*. Yeast 6:193– 204
- Smith M, Hopkinson DA, Harris H (1971) Development changes and polymorphism in human alcohol dehydrogenase. Ann Hum Genet 34:251-271
- Trezise AEO, Godfrey EA, Holmes RS, Beacham IR (1989) Cloning and sequencing of cDNA encoding baboon liver alcohol dehydrogenase: evidence for a common ancestral lineage with the human alcohol dehydrogenase  $\beta$  subunit for class I alcohol dehydrogenase gene duplications predating primate radiation. Proc Natl Acad Sci USA 86:5454–5458

- Trick M, Dennis ES, Edwards KJR, Peacock WJ (1988) Molecular analysis of the alcohol dehydrogenase gene family of barley. Plant Mol Biol 11:147–160
- Vallee BL, Auld DS (1990) Active-site zinc ligands and activated H<sub>2</sub>O of zinc enzymes. Proc Natl Acad Sci USA 87:220– 224
- Vallee BL, Bazzone TJ (1983) Isozymes of human liver alcohol dehydrogenase. In: Rattazzi MC, Scandalios JG, Witt GS (eds) Current topics in biological medical research, vol 8. Liss, New York, pp 219–244
- von Bahr-Lindström H, Höög JO, Hedén LO, Kaiser R, Fleetwood L, Larsson K, Lake M, Holmquist B, Holmgren A, Hempel J, Vallee BL, Jörnvall H (1986) cDNA and protein structure for the  $\alpha$  subunit of human liver alcohol dehydrogenase. Biochemistry 25:2465–2470
- Wolyn DJ, Jelenkovic G (1990) Nucleotide sequence of an alcohol dehydrogenase gene in octoploid strawberry. Plant Mol Biol 14:855-857
- Xie Y, Wu R (1989) Rice alcohol dehydrogenase genes: anaerobic induction, organ specific expression and characterization of cDNA clones. Plant Mol Biol 13:53–68
- Xie Y, Wu R (1990) Molecular analysis of an alcohol dehydrogenase-encoding genomic clone (*adh*2) from rice. Gene 87:185–191
- Yasunami M, Chen CS, Yoshida A (1990) Multiplication of the class I alcohol dehydrogenase locus in mammalian evolution. Biochem Genet 28:591-599
- Yasunami M, Chen CS, Yoshida A (1991) A human alcohol dehydrogenase gene (ADH6) encoding an additional class of isozyme. Proc Natl Acad Sci USA 88:7610–7614
- Yokoyama S, Yokoyama R, Kinlaw C, Harry DE (1990) Molecular evolution of the zinc-containing long-chain alcohol dehydrogenase genes. Mol Biol Evol 7:143–154
- Young ET, Pilgrim D (1985) Isolation and DNA sequence of *ADH3*, a nuclear gene encoding the mitochondrial isozyme of alcohol dehydrogenase in *Saccharomyces cerevisiae*. Mol Cell Biol 5:3024–3034

Received November 7, 1991/Revised January 8, 1992