FOR THE RECORD

A novel zinc-binding motif found in two ubiquitous deaminase families

JONATHAN REIZER,¹ STEWART BUSKIRK,¹ AMOS BAIROCH,² AIALA REIZER,¹
AND MILTON H. SAIER, JR.¹

¹ Department of Biology, University of California at San Diego, La Jolla, California 92093-0116

² Department of Medical Biochemistry, University of Geneva, 1 rue Michel Servet, 1211 Geneva 4, Switzerland

(RECEIVED December 30, 1993; ACCEPTED March 9, 1994)

Abstract: Two families of deaminases, one specific for cytidine, the other for deoxycytidylate, are shown to possess a novel zinc-binding motif, here designated ZBS. We have (1) identified the protein members of these 2 families, (2) carried out sequence analyses that allow specification of this zinc-binding motif, and (3) determined signature sequences that will allow identification of additional members of these families as their sequences become available.

Keywords: allosteric site; aminohydrolase; cytidine; deaminase; deoxycytidylate; evolution; zinc-binding motif; zinc metallo-enzyme

Zinc plays a pivotal role in enzyme structure and function, and several zinc-binding motifs have been identified for specific types of proteins (Berg, 1986, 1990; Vallee & Auld, 1990; Vallee et al., 1991; von Arnim & Deng, 1993). The cytidine and deoxycytidylate deaminases comprise 2 families of zinc-dependent enzymes but, nevertheless, little is known about the zinc-binding site(s) of these deaminases, which were proposed to share a common general mechanism of catalysis (Yang et al., 1992; Moore et al., 1993). We therefore (1) identified the protein members of these 2 families, (2) carried out sequence analyses that allow specification of a novel zinc-binding motif, and (3) determined signature sequences that will allow identification of additional members of these families as their sequences become available.

The proteins included in this study as well as their abbreviations and catalytic functions where known are presented in the legend to Figure 1. These proteins belong to 2 distinct families, as revealed by pairwise sequence comparison scores (see Table 1). Four proteins constitute the currently sequenced cytidine deaminase family (CDF): a human and 2 bacterial cytidine deaminases as well as a *Bacillus cereus* deaminase that acts on the cytosine-containing antibiotic blasticidin-S (Isono, 1991). All the intra-family similarity scores of the CDF protein members were found

to be high (i.e., >15 SD) except for those obtained with cytidine deaminase of *Escherichia coli*, which is most similar to cytidine deaminase of *Bacillus subtilis* (9 SD).

The second family, the deoxycytidylate deaminase family (DCF), includes 12 proteins from diverse sources: 2 bacteriophage, 6 bacteria, 2 yeast, the worm *Caenorhabditis elegans*, and a mammal (see proteins listed in the legend to Figure 1). Each protein of this family exhibits a high similarity score (10–107 SD) to at least 2 other members of this family, but proteins of the

Fig. 1. Average similarity profiles of the aligned protein sequences of the cytidine deaminase family (CDF; left panel) and deoxycytidylate family (DCF; right panel). The average similarity across the entire alignment is shown as a dotted line. Regions of alignment outside those shown in the 2 plots lack significant similarity. The bars labeled AAS and ZBS denote the positions of the putative allosteric or active site and the zinc-binding site, respectively. The 4 proteins that comprise the CDF and the 12 proteins that comprise the DCF, their abbreviations, and their accession numbers in the SWISS-PROT data bank, respectively, are as follows; CDF: cytidine deaminases of *E. coli* (CDD(Eco); P13652), of *B. subtilis* (CDD(Bsu); P19079), and of *Homo sapiens* (CDD(Hsa); P32320); blasticidin-S deaminase of *B. cereus* (BSD(Bce); P33977). DCF: dCMP deaminase of bacteriophage T2 (DCD(T2); P00814), of bacteriophage T4 (DCD(T4); P16006), of *S. cerevisiae* (DCD(Sce); P06773), and of *H. sapiens* (DCD(Hsa); P32321); the functionally unidentified open reading frames of *B. subtilis* (ORF161(Bsu); P21335 and ORF189(Bsu); P32393), of *E. coli* (YfhC(Eco); P30134), of *V. fischeri* (ORF147(Vfi); P33968), and of *C. elegans* (ORF197(Cel); P30648); the RibG protein of *B. subtilis* (RibG(Bsu); P17618) and of *E. coli* (RibG(Eco); P25539); and the RibG homologue of *S. cerevisiae* (YBR12.03(Sce); P33312).

Reprint requests to: Jonathan Reizer, Department of Biology, University of California at San Diego, La Jolla, California 92093-0116; e-mail: reizer@ucsd.edu.

Table 1. Binary comparisons of the amino acid sequences of members of the cytidine deaminase family (CDF) and the deoxycytidylate deaminase family (DCF)^a

	CD0(Bsu) (136)	CD0(Hsa) (145)	CD0(Eco) (329)	ORF189(Bsu) (189)	DCD(Hsa) (178)	DCD(T2) (188)	DCD(T4) (193)	DCD(Sce) (312)	ORF161(Bsu) (161)	YfhC(Eco) (178)	ORF197(Cel) (197)	ORF147(Vfi) (147)	RibC(Bsu) (361)	RibC(Eco) (367)	YBR12.03(Sce) (244)
BSD(Bce) (140)	33(132) [17]	31(110) [15]	27(25) [4]	N.D. ^b	46(11) [2]	25(52) [1]	25(52) [1]	18(33) [0]	57(7) [1]	17(78) [1]	14(37) [1]	N.D.	25(69) [3]	20(30) [1]	30(23) [0.5]
CD0(Bsu) (136)		49(132) [44]	37(49) [9]	24(34) [4]	30(37) [2]	30(20) [3]	30(20) [3]	33(24) [1]	18(55) [1]	N.D.	35(17) [1]	26(43) [2]	N.D.	30(27) [2]	34(30) [0.2]
CD0(Hsa) (145)			39(52) [7]	19(27) [0]	24(70) [4]	33(21) [2]	33(21) [2]	17(29) [0]	N.D.	18(38) [2]	21(24) [1]	24(32) [4]	N.D.	29(14) [1]	56(9) [0.6]
CD0(Eco) (329)				22(82) [0]	67(6) [1]	31(36) [0]	31(36) [0]	21(43) [0]	27(30) [1]	60(5) [1]	22(86) [1]	N.D.	20(58) [2]	30(33) [1]	40(15) [0.7]
ORF189(Bsu) (189)					37(141) [23]	45(69) [27]	42(77) [18]	35(149) [23]	33(49) [6]	31(62) [15]	28(150) [15]	34(147) [26]	29(79) [7]	27(77) [5]	27(11) [0.7]
DCD(Hsa) (178)						44(54) [14]	44(54) [14]	31(146) [22]	31(32) [7]	28(43) [7]	41(144) [40]	41(111) [24]	32(76) [7]	29(70) [4]	18(51) [0.3]
DCD(T2) (188)							94(193) [107]	39(69) [12]	39(69) [11]	28(95) [10]	40(47) [11]	27(130) [14]	43(79) [9]	35(78) [4]	27(11) [0.8]
DCD(T4) (193)								39(54) [14]	40(69) [12]	28(95) [10]	40(47) [11]	27(129) [13]	43(79) [7]	35(78) [4]	27(11) [0.8]
DCD(Sce) (312)									33(30) [5]	39(39) [6]	27(129) [9]	34(121) [16]	28(39) [2]	19(78) [4]	12(83) [0.8]
ORF161(Bsu) (161)										44(154) [46]	23(103) [7]	35(151) [14]	31(157) [14]	29(155) [9]	22(51) [2]
YfhC(Eco) (178)											53(30) [10]	23(103) [7]	31(151) [16]	31(157) [14]	18(33) [0.1]
ORF197(Cel) (197)												41(127) [28]	25(83) [7]	24(116) [6]	21(33) [0.6]
ORF147(Vfi) (147)														37(27) [4]	20(40) [2]
RibC(Bsu) (361)														41(365) [84]	25(183) [16]
RibC(Eco) (367)															24(189) [17]

^a Abbreviations are as indicated in the legend of Figure 1. Values in parentheses below the designation of the protein refer to the number of residues in the intact protein. Values presented in the table that are not in brackets or parentheses represent percent identity for segments having the number of compared residues indicated in parentheses. The FASTA program, using the dipeptide mode (ktup = 2) (Pearson & Lipman, 1988), was used to assess similarities of the indicated proteins. Comparison scores in standard deviations, using the RDF2 program (Pearson & Lipman, 1988) and 150 shuffles, are given in brackets below the values for percent identity. Shaded boxes denote intrafamily homologous members of the CDF (left triangle) and DCF (right triangle).

^b N.D., not detected.

CDF are not demonstrably homologous to proteins of the DCF (i.e., interfamily similarity scores ≤ 4 SD).

The average similarity plots of all currently known proteins of the CDF and DCF (Fig. 1, left and right panels, respectively) reveal regions of striking similarity as well as regions of marked divergence for both families. Regions represented in Figure 1 are the only ones that exhibit significant similarity, and regions that possess an average similarity score of zero are not included in the plot. One region of similarity, designated AAS (allosteric or active site) and indicated by black bars in Figure 1, is possibly an allosteric effector-binding site. Deoxycytidylate deaminases are known to be sensitive to allosteric regulation by dCTP and dTTP (Maley et al., 1990), and the allosteric site in the phage T2 deaminase has been proposed to be localized to the first 125 residues of the protein, a region that includes all of the presently identified AAS sequence (Maley et al., 1983). A second region of conservation, designated ZBS (zinc-binding site) and indicated by shaded bars in Figure 1, follows the AAS region and includes a novel zinc-binding motif common to all proteins of both the DCF and the CDF (see below and Yang et al., 1992; Moore et al., 1993; Weiner et al., 1993).

Figure 2 presents the multiple alignments of the AAS and ZBS regions of the sequences of CDF and DCF members. Based on the aligned ZBS regions, the signature sequence [CH]-A-E-X₍₂₎-[LIVMA]-[LIVM]-X₍₁₇₋₃₃₎-P-C-X₍₂₋₈₎-C-X₍₃₎-[LIVM], valid for both families, was determined (Bairoch, 1993). This signature sequence detects all proteins of the DCF and CDF except for the

yeast protein YBR12.03 (see below). It is selective for these deaminases because it did not recognize any other protein in the SWISS-PROT database (release 27.0) (Bairoch & Boeckmann, 1993). Interfamily divergence and intrafamily conservation of this signature sequence are revealed in Figure 2.

It is noteworthy that a more degenerate and condensed form of this signature sequence, i.e., [CH]-A-E-X₍₂₁₋₃₇₎-P-C-X₍₂₋₈₎-C, which is conserved in all proteins of the DCF and CDF except the yeast protein YBR12.03, contains residues commonly associated with zinc chelation (Vallee & Auld, 1990, and references therein). Furthermore, previous studies on the *E. coli* cytidine deaminase have shown that Zn²⁺ is chelated to the histidyl residue in the signature sequence described above (unpubl. data cited by Yang et al., 1992). Consequently, we suggest that this condensed signature represents a novel, ubiquitous zinc-binding motif.

Support for the conclusion that zinc is coordinated at the ZBS region has come from analyses of the bacterial and mammalian adenosine deaminases, which exhibit a common mechanism of action with the cytidine deaminases (Frick et al., 1989; Kati & Wolfenden, 1989a, 1989b). The crystal structure of the murine adenosine deaminase shows that zinc is coordinated to a histidyl residue at its active site (Wilson et al., 1991), and comparisons between the sequence of this adenosine deaminase and those of proteins of the DCF and CDF show that the only region of commonality corresponds to the first part of the ZBS signature (Yang et al., 1992).

CDF

	AAS
	<div> <div> <div>!</div> <div>!</div> <div>!</div> </div> <div> <div>!</div> <div>!</div> <div>!</div> </div> </div>
CDD (Bsu)	26 VGAALITkdGkvyrGcNIE..nAaysmcn
CDD (Hsa)	37 VGAALITqeGrifkGcNIE..nAcypIgi
BSD (Bce)	32 VGAairTktGeiisavhIE..ayigrvtv
CDD (Eco)	72 mLAqLrgvsGtwyfGaNmEfigAtmqqtv
Consensus	VGAAL-T--G---G-NIE---A-----

DCF

	AAS
	<div> <div>!</div> </div>
ORF161 (Bsu)	26 iGAVlV..ingeIIarahN
YfhC (Eco)	41 VgAVlV..hnnRvIgeGwN
RibG (Bsu)	26 VgAVvV..KdqqIvgmGah
RibG (Eco)	28 VgCvIV..KdqeIvgeGYh
DCD (Hsa)	37 VgAcIVnsenkIvgiGYN
ORF197 (Cel)	72 VgCvIVdKdncIvsvGYN
DCD (T2)	24 VgAVIe..KngRIIstGYN
DCD (T4)	24 VgAVIe..KngRIIstGYN
ORF189 (Bsu)	28 VgAtIV..rdkRmIagGYN
ORF147 (Vfi)	37 VgAVIt..KhnrIvsvGYN
DCD (Sce)	185 VgCvIV..recRvIatGYN
Consensus	VgAVIV-K--RII--GYN

ZBS

	<div> <div>!!</div> <div>!</div> <div>!</div> <div>!!</div> <div>!!</div> </div>
53	CAERTAlfkAVSeGdteFqmlav.....aaDtpgp..vSPCGaCRQvisE
64	CAERTAlfkAVSeGdteFqmlav.....asDmgddfiSPCGaCRQvmrE
59	CAEaiAlgsAVSnGqKdFdtIvavrhpysdevDrsirrvvSPCGmCRelisd
101	hAEqsaIshAwlsGeKalaaItvn.....ytPCGHCRQfmnE
	CAE--AI--AVS--G--K--F--I-----D-----SPCG--CRQ---E

ZBS

	<div> <div>!!!</div> <div>!!</div> <div>!</div> </div>
53	HAEmIvIdeAckalgtwrleGATLYVTLePC.....PmCagavvlSrvekV
68	HAEmAlrgggglvmqnyrlidATLYVTLePC.....vmCagamIhSrIgrV
49	HAEvhaIhmAga....haeGAdiYVTLePCshygktPpCAeIIInSGIKrV
50	HAEvhaIrmAge....kakGATaYVTLePCshhgtrPpCcdalIaaGvarV
84	HAELNAIm..nk..nstdvkGcsmYvaLlPC.....neCAkIIQaGIKeV
117	HAEmNAIi..nk..rcttlhdCTvYVTLePC.....nkCAqmlIQSrvKkV
104	HAELNAIIfAar..ngssieGATmYVTLePC.....PdCAkaIaQSGIKkI
104	HAELNAIIfAar..ngssieGATmYVTLePC.....PdCAkaIaQSGIKkI
70	HAEmNAIIfAar..fgvptdGAeiYVThyPC.....iqCcksIIQaGIKeV
68	HAEmNAIIfAar..dlegcdiWThfPC.....PnCAakIIQSGIKV
233	HAEmNAIIfAar..drvg.qnATLYcdtPC.....ltCsvkIvQGISeV
	HAEmNAI--A-----GAT--YVTLePC-----P-CA--IIQSGIK-V

Fig. 2. Multiple alignment of the regions that comprise the putative allosteric or active site (AAS) and the zinc-binding site (ZBS) in members of the cytidine deaminase family (CDF) and the deoxycytidylate deaminase family (DCF). The consensus sequences of the aligned AAS and ZBS regions are shown below the aligned sequences. The numbers to the left of the sequences shown denote the position of the first residue presented in each indicated protein. Abbreviations are as indicated in the legend of Figure 1.

The second region of conservation, which precedes the ZBS in all of the sequences examined here, is most likely a part of an allosteric effector site.

The DCF includes 6 ORFs with unknown functions. Based on the analyses reported here, the following suggestions are made regarding the potential functions of some of these proteins:

1. The *gef* gene family encodes membrane proteins that induce respiratory arrest, membrane leakiness, and cell death in *E. coli* (Poulsen et al., 1992). We suggest that, like all DCF proteins with known functions, the primary function of YfhC(Eco), which was isolated from a Gef-resistant mutant (Poulsen et al., 1992), is to deaminate a pyrimidine ring.
2. Because ORF161(Bsu) exhibits high similarity to YfhC(Eco) (46 SD), we propose that these 2 ORFs possess the same function. Similarly, the high degree of similarity between ORF197(Cel) and DCD(Hsa) (40 SD) suggests that ORF197(Cel) is a deoxycytidylate deaminase.
3. The *luxG* operon, which is transcribed convergently with ORF147(Vfi), encodes proteins mediating bioluminescence in *Vibrio fischeri*, a process that requires the oxidation of reduced riboflavin phosphate (FMNH₂) (Lee et al., 1993, and references therein). Suggestions that ORF147(Vfi) may be involved in supplying this substrate have been based on the similarity of ORF147 to the RibG proteins (4 and 6 SD) (Lee & Meighen, 1992; Lee et al., 1993). Because ORF147(Vfi) is more similar to the dCMP deaminases (13–26 SD) than to the riboflavin deaminases, we suggest that if ORF147 does catalyze a RibG-like reaction, it evolved along a pathway different from that used by the RibG proteins of *B. subtilis* and *E. coli*.
4. The *Saccharomyces cerevisiae* YBR12.03 protein was reported to be involved in riboflavin biosynthesis (Baur et al., 1993). Although it shows a convincing degree of similar-

ity to the C-terminal half of bacterial RibG proteins (>16 SD), it lacks the N-terminal half, which contains the AAS and ZBS motifs. Inspection of the nucleotide sequence reveals that the apparent absence of the N-terminal region is due neither to a frameshift error nor to intron interruption. We propose that in yeast this riboflavin biosynthetic step is catalyzed by an oligomeric enzyme composed of 2 distinct subunits, one of which corresponds to the missing N-terminal half containing the AAS and ZBS regions.

Acknowledgments

This work was supported by Public Health Service grants 5R01AI 21702 and 2R01AI 14176 from the National Institute of Allergy and Infectious Diseases.

References

- Bairoch A. 1993. The PROSITE dictionary of sites and patterns in proteins: Its current status. *Nucleic Acids Res* 21:3097–3103.
- Bairoch A, Boeckmann B. 1993. The SWISS-PROT protein sequence data bank: Recent developments. *Nucleic Acids Res* 21:3093–3096.
- Baur A, Schaaff-Gerstenschlager I, Boles E, Miosga T, Rose M, Zimmermann FK. 1993. Sequence of a 4.8 kb fragment of *Saccharomyces cerevisiae* chromosome II including three essential open reading frames. *Yeast* 9:289–293.
- Berg JM. 1986. Potential metal-binding domains in nucleic acid binding proteins. *Science* 232:485–487.
- Berg JM. 1990. Zinc fingers and other metal-binding domains. Elements for interactions between macromolecules. *J Biol Chem* 265:6513–6516.
- Frick L, Yang C, Marquez VE, Wolfenden R. 1989. Binding of pyrimidin-2-one ribonucleoside by cytidine deaminase as the transition-state analogue 3,4-dihydrouridine and the contribution of the 4-hydroxyl group to its binding affinity. *Biochemistry* 28:9423–9430.
- Isono K. 1991. Current progress on nucleoside antibiotics. *Pharmacol Ther* 52:269–286.
- Kati WM, Wolfenden R. 1989a. Contribution of a single hydroxyl group to transition-state discrimination by adenosine deaminase: Evidence for an "entropy trap" mechanism. *Biochemistry* 28:7919–7927.
- Kati WM, Wolfenden R. 1989b. Major enhancement of the affinity of an

- enzyme for a transition-state analog by a single hydroxyl group. *Science* 243:1591–1593.
- Lee CY, Meighen EA. 1992. The *lux* genes in *Photobacterium leiognathi* are closely linked with genes corresponding in sequence to riboflavin synthesis genes. *Biochem Biophys Res Commun* 186:690–697.
- Lee CY, Szittner RB, Miyamoto CM, Meighen EA. 1993. The gene convergent to *luxG* in *Vibrio fischeri* codes for a protein related in sequence to *RibG* and deoxycytidylate deaminase. *Biochim Biophys Acta* 143:337–339.
- Maley GF, Duceman BW, Wang A, Martinez J, Maley F. 1990. Cloning, sequence analysis, and expression of the bacteriophage T4 *cd* gene. *J Biol Chem* 265:47–51.
- Maley GF, Guarino DU, Maley F. 1983. Complete amino acid sequence of an allosteric enzyme, T2 bacteriophage deoxycytidylate deaminase. *J Biol Chem* 258:8290–8297.
- Moore JT, Silversmith RE, Maley GF, Maley F. 1993. T4-phage deoxycytidylate deaminase is a metalloprotein containing two zinc atoms per subunit. *J Biol Chem* 268:2288–2291.
- Pearson WR, Lipman DJ. 1988. Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* 85:2444–2448.
- Poulsen LK, Larsen NW, Molin S, Andersson P. 1992. Analysis of an *Escherichia coli* mutant strain resistant to the cell-killing function encoded by the *gef* gene family. *Mol Microbiol* 6:895–905.
- Vallee BL, Auld DS. 1990. Zinc coordination, function, and structure of zinc enzymes and other proteins. *Biochemistry* 29:5647–5659.
- Vallee BL, Coleman JE, Auld DS. 1991. Zinc fingers, zinc clusters, and zinc twists in DNA-binding protein domains. *Proc Natl Acad Sci USA* 88:999–1003.
- von Arnim AG, Deng X. 1993. Ring finger motif of *Arabidopsis thaliana* COP1 defines a new class of zinc-binding domain. *J Biol Chem* 268:19626–19631.
- Weiner KXB, Weiner RS, Maley F, Maley GF. 1993. Primary structure of human deoxycytidylate deaminase and overexpression of its functional protein in *Escherichia coli*. *J Biol Chem* 268:12983–12989.
- Wilson DK, Rudolph FB, Quijcho FA. 1991. Atomic structure of adenosine deaminase complexed with a transition-state analog: Understanding catalysis and immunodeficiency mutations. *Science* 252:1278–1284.
- Yang C, Carlow D, Wolfenden R, Short S. 1992. Cloning and nucleotide sequence of the *Escherichia coli* cytidine deaminase (*ccd*) gene. *Biochemistry* 31:4168–4174.