Protein Science (1998), 7:1092–1105. Cambridge University Press. Printed in the USA.

Motifs and structural fold of the cofactor binding site of human glutamate decarboxylase

KUNBIN QU,[1] DAVID L. MARTIN,[2] AND CHARLES E. LAWRENCE[1,3]

[1]Biometrics Laboratory and [2]Laboratory of Nervous System Disorders, Wadsworth Center for Laboratories and Research, New York State Department of Health, Albany, New York 12201
[3]National Center for Biotechnology Information, National Library of Medicine, National Institute of Health, Bethesda, Maryland 20894

(RECEIVED September 23, 1997; ACCEPTED January 6, 1998)

Abstract

The pyridoxal-P binding sites of the two isoforms of human glutamate decarboxylase (GAD65 and GAD67) were modeled by using PROBE (a recently developed algorithm for multiple sequence alignment and database searching) to align the primary sequence of GAD with pyridoxal-P binding proteins of known structure. GAD's cofactor binding site is particularly interesting because GAD activity in the brain is controlled in part by a regulated interconversion of the apo- and holoenzymes. PROBE identified six motifs shared by the two GADs and four proteins of known structure: bacterial ornithine decarboxylase, dialkylglycine decarboxylase, aspartate aminotransferase, and tyrosine phenol-lyase. Five of the motifs corresponded to the α/β elements and loops that form most of the conserved fold of the pyridoxal-P binding cleft of the four enzymes of known structure; the sixth motif corresponded to a helical element of the small domain that closes when the substrate binds. Eight residues that interact with pyridoxal-P and a ninth residue that lies at the interface of the large and small domains were also identified. Eleven additional conserved residues were identified and their functions were evaluated by examining the proteins of known structure. The key residues that interact directly with pyridoxal-P were identical in ornithine decarboxylase and the two GADs, thus allowing us to make a specific structural prediction of the cofactor binding site of GAD. The strong conservation of the cofactor binding site in GAD indicates that the highly regulated transition between apo- and holoGAD is accomplished by modifications in this basic fold rather than through a novel folding pattern.

Keywords: Bayesian statistics; Gibbs sampling; human glutamate decarboxylase; multiple sequence alignment; structural prediction by homology

Glutamate decarboxylase (GAD, EC 4.1.1.15) is the pyridoxal-5′-phosphate (pyridoxal-P) dependent enzyme that synthesizes γ-aminobutyric acid (GABA), the major inhibitory neurotransmitter in vertebrate brain. Adult brain contains two major forms of GAD called GAD65 and GAD67, which are the products of two genes (Erlander et al., 1991; Martin & Rimvall, 1993). Each GAD is composed of two major domains: a C-terminal domain of about 500 amino acids, and a 95–100 amino acid N-terminal domain. The C-terminal domain contains the pyridoxal-P binding site and lengthy segments that have identical sequences in GAD65 and GAD67; the overall sequence identity for this domain of human GAD is 73%. The sequences of the amino terminal domains of GAD65 and GAD67 differ substantially (23% identity). This domain is involved in targeting GAD to membranes and in the formation of heteromultimers of GAD65 and GAD67 (Solimena et al., 1993, 1994; Dirkx et al., 1995; Sheikh & Martin, 1996); it contains apparent palmitoylation sites in GAD65 and GAD67 (Christgau et al., 1992; Martin & Barke, 1997) and phosphorylation sites in GAD65 (Namchuk et al., 1997).

Pyridoxal-P plays a key role in the regulation of GAD activity (reviewed by Martin & Rimvall, 1993). GAD is unusual, if not unique, among pyridoxal-P–dependent enzymes in brain in being present mainly as inactive apoenzyme (GAD without bound pyridoxal-P) (Martin et al., 1991). This apoGAD serves as a reservoir of inactive enzyme that can be converted to active holoGAD when additional GABA synthesis is required. The interconversion between apo- and holoGAD does not result from simple dissociation and association of pyridoxal-P but occurs by a highly regulated cycle of reactions (Porter et al., 1985; Martin & Rimvall, 1993). In this cycle, apoGAD is formed when pyridoxal-P is converted to pyridoxamine-P by a transamination reaction catalyzed by GAD and the pyridoxamine-P dissociates from the enzyme. HoloGAD is formed by a two-step reaction between apoGAD and free pyridoxal-P. Although other decarboxylases can also carry out these reactions, the regulation of the reactions appears to be unique to GAD. With GAD, polyanions such as ATP favor the formation

Reprint requests to: Charles E. Lawrence, P.O. Box 509, Wadsworth Center, Empire State Plaza, Albany, New York 12201; e-mail: lawrence@wadsworth.org.

of apoGAD and stabilize it against thermal inactivation, while inorganic phosphate strongly enhances the activation of apoGAD by pyridoxal-P (Porter & Martin, 1988). Evidently we must understand the interactions of pyridoxal-P with GAD to understand both the catalytic mechanism and the regulation of the enzyme.

Structural motifs are conserved across great evolutionary distances and homology modeling is possible if the conserved motifs can be identified. The structural fold of the pyridoxal-P binding site of some decarboxylases and transaminases has been identified by comparing their X-ray structures (Momany et al., 1995a). Momany et al. also aligned GAD and several other decarboxylases with the sequence of ornithine decarboxylase from *Lactobacillus 30a* by using a combination of conventional alignment algorithms and manual alignment. This alignment identified several segments within GAD that appeared to correspond to the pyridoxal-P binding site. In the work reported here, we have used a recently developed, automated algorithm for multiple sequence alignment and database searching to identify proteins that are related to GAD according to objective statistical criteria and to align the sequence of GAD with proteins of known structure and thereby predict the structural fold of GAD's co-factor binding site. As described below, we searched a nonredundant database of approximately 100,000 protein sequences and recruited 512 related proteins. Our results substantially refine those of Momany et al. by more clearly defining the conserved regions by objective, automated statistical procedures and criteria, by identifying a previously unrecognized conserved motif, and by identifying additional conserved residues. Furthermore, we exploit the objective character of our method to test the validity of the predictions stemming from it.

Results

PROBE (Neuwald et al., 1997), a recently developed multiple-sequence alignment propagation algorithm (Liu & Lawrence, 1996) and database search procedure, has significantly advanced the identification and alignment of subtly related protein sequences. The key to the success of this procedure is the use of Bayesian model selection methods to identify and align only those regions that the data justify as related. Not surprisingly, the conserved motifs in subtly related proteins are predominantly subsequences that interact with other molecules such as ligands.

The structures of eight pyridoxal-P dependent enzymes are available from the Protein Data Bank (PDB): ornithine decarboxylase from *Lactobacillus 30a* (EC 4.1.1.17; PDB accession number 1ORD), dialkyglycine decarboxylase from *Pseudomonas cepacia* (EC 4.1.1.64; 1DGD), aspartate aminotransferase (EC 2.6.1.1; several versions are available in PDB; we used 9AAT, which is the mitochondrial form from chicken heart), tyrosine phenol-lyase from *Citrobacter intermedius* (EC 4.2.1.20; 1TPL), D-amino acid aminotransferase from a *Thermophillic bacillus* (EC 2.6.1.21; 1DAA), tryptophan synthase from *Salmonella typhimurium* (EC 4.2.1.20; 1WSY), alanine racemase from *Bacillus stearothermophilus* (EC 5.1.1.1; 1SFT), and glycogen phosphorylase from rabbit muscle (EC 2.4.1.1; 1PYG). The amino acid sequences of these proteins are only distantly related to each other as well as to the two GADs, as demonstrated by the low BLAST scores and nonsignificant p-values (Table 1). Four other pyridoxal-P–dependent enzymes have been crystallized, but their coordinates are not publicly available. These are ω-amino acid-pyruvate aminotransferase (Ω-AT, EC. 2.6.1.18; Watanabe et al., 1991), phosphoserine aminotransferase (PS-AT, EC. 2.6.1.52; Stark et al., 1991), glutamate-1-semialdehyde aminomutase (AD-MT, EC 5.4.3.8; Hennig, 1997) and cystathionine β-lyase (β-LS, EC 4.4.1.8; Clausen et al., 1996). These proteins also are only distantly related to the two GADS and to most of the seven proteins of known structure in PDB (Table 1); although 1DGD, AD-MT, and Ω-AT are significantly related.

Motif identification and selection

PROBE identified 10 motif models in the final alignment and found 512 proteins that fit the overall model out of 100,000 proteins in a nonredundant sequence database at NCBI. The 10 motifs ranged from 16 to 27 residues in length and included a total of 199 residues. Along with the two GADs, four of the seven pyridoxal-P dependent proteins from PDB were significantly related to the model: specifically, aspartate aminotransferase, dialkylglycine decarboxylase, tyrosine phenol-lyase, and ornithine decarboxylase (Table 2). These enzymes are all members of the α family of pyridoxal-P–dependent proteins (Alexander et al., 1994). The other four pyridoxal-P proteins from PDB (tryptophan synthase, D-amino acid aminotransferase, alanine racemase, and glycogen phosphorylase) did not fit the alignment model, implying that their sequences are significantly distinct.

To assure that a protein and its homologs are not recruiting themselves to the model, we performed a Jackknife test with 1DGD, 9AAT, 1ORD, 1TPL, GAD65, and GAD67. To do this, we removed the query protein, say 1DGD, and all of its homologs from the collection of 512 related proteins, rebuilt the model, and then tested the query using this reduced model. As indicated in Table 2, five of these six proteins are significantly related to the model (p-value < 0.05) and 1TPL is marginally significant (p-value $= 0.062$). These results suggest that the model incorporates characteristics common to all of these protein sequences.

Not all of the 10 motifs were conserved (statistically significant) in every protein. To select motifs worthy of further study, we examined the maximum *a posteriori* probability (MAP) score. MAP scores above zero indicate that the motif is more likely to belong to the model than to the unaligned random background and thus merit further study. Five motifs met this criterion (Table 3). Motif 10 had a marginal MAP score (-0.75) but was included in our extended analysis. Thus, our final set included motifs 2, 3, 4, 5, 6, and 10. The sequence fragments, i.e., motif elements, from the two GADs and the four proteins of known structure that fit the six motifs of our model are listed in Figure 1 along with their corresponding p-values.

Each of the six significant motifs represents a conserved structure in the four proteins of known structure that fit our model. Structural data indicate that the aminotransferases and some pyridoxal-P–dependent decarboxylases are composed of a large pyridoxal-P binding domain and one small domain that closes on the large domain when the substrate binds (McPhalen et al., 1992b). The large domain has a seven-stranded core of α/β folds (Fig. 2). Pyridoxal-P lies in a crevice formed by the connecting loops in the carboxyl edge region of the seven beta strands. The six motifs are shown in Figure 3 for each of the known structures that fit our model (1DGD, 9AAT, 1ORD, and 1TPL). In every one of these structures, our motifs 2 through 6 correspond to six of the seven α/β elements and the loops that form the binding cleft for pyridoxal-P. Only one α/β element of the seven-stranded structure (helix C, β strand 3, and its connecting loop III; see Fig. 2) was not included in our model, possibly because it is too far from the pyridoxal-P binding site to retain widely shared features. Motif 10

Table 1. *GAD and pyridoxal-P–dependent enzymes of known structure are only distantly related as determined by BLAST score and p-value*[a]

Enzyme[b]	1DGD	9AAT	1ORD	1TPL	PS-AT	Ω-AT	AD-MT	GAD65	GAD67	β-LS	1WSY	1DAA	1SFT	1PYG
1DGD	2,197	25 >0.999	40 0.840	39 0.920	41 0.730	178 2e-29	97 6e-27	33 >0.999	32 >0.999	27 >0.999	36 0.998	34 >0.999	30 >0.999	33 >0.999
9AAT		2,191	34 >0.999	35 >0.999	33 >0.999	36 0.998	34 >0.999	33 >0.999	41 0.740	31 0.999	38 0.970	30 >0.999	33 >0.999	30 >0.999
1ORD			3,874	35 >0.999	36 >0.999	36 >0.999	32 >0.999	31 >0.999	38 >0.999	34 >0.999	36 >0.999	32 >0.999	30 >0.999	44 0.019[c]
1TPL				2,374	33 >0.999	50 0.075	37 0.992	30 >0.999	28 0.999	32 >0.999	34 >0.999	31 >0.999	28 >0.999	43 0.530
PS-AT					1,860	27 >0.999	31 >0.999	29 >0.999	31 >0.999	31 >0.999	33 >0.999	32 >0.999	38 0.950	34 >0.999
Ω-AT						2,355	138 7e-20	35 >0.999	32 >0.999	32 >0.999	29 >0.999	36 0.998	30 >0.999	37 0.990
AD-MT							2,236	30 >0.999	31 >0.999	33 >0.999	32 >0.999	29 >0.999	39 0.920	36 0.997
GAD65								3,124	1836 3e-286	34 >0.999	41 0.820	29 >0.999	37 0.997	34 >0.999
GAD67									3,138	29 >0.999	29 >0.999	29 >0.999	38 0.990	31 >0.999
β-LS										2,039	30 >0.999	29 >0.999	34 >0.999	33 >0.999
1WSY											2,050	29 >0.999	29 >0.999	29 >0.999
1DAA												1,484	34 >0.999	43 0.41
1SFT													2,033	36 0.994
1PYG														4,429

[a]A BLAST search was performed by individually searching each sequence in the top row against a database containing only these 14 proteins. Within each box of the table, the first row is the BLAST score, while the second row is the *p*-value.

[b]The enzymes are: 1DGD, dialkyglycine decarboxylase; 9AAT, aspartate aminotransferase; 1ORD, ornithine decarboxylase; 1TPL, tyrosine phenol-lyase; PS-AT, phosphoserine aminotransferase; Ω-AT, ω-amino acid-pyruvate aminotransferase; AD-MT, glutamate-1-semialdehyde aminomutase; GAD65; GAD67; β-LS, cystathionine β-lyase; 1WSY, tryptophan synthase; 1DAA, D-amino acid aminotransferase; 1SFT, alanine racemase; 1PYG, glycogen phosphorylase.

[c]This significant *p*-value is due to the limited data available from the 14 proteins and the relatively longer length of 1PYG and 1ORD, resulting in many more ways of matching them. It is a false positive from BLAST.

belongs to the small domain. Thus, even though the four proteins of known structure are very distantly related (average BLAST pairwise score about 35; Table 1), the motifs identified by PROBE correspond very well to one another and share a common structure.

Motif folding

Five of the six motifs in our model belong to the pyridoxal-P binding domain. The loops in the carboxyl end of the five motifs play very important roles in forming the cleft for the cofactor and substrate. Motif 2 corresponds to the N terminal part of helix A and beta strand 1, and contains residues PH1 and PH3 (Table 4A) that interact with the phosphate of pyridoxal-P. Helix A helps to stabilize the binding of pyridoxal-P through strong dipolar coupling between the positive end of the helix and the negatively charged phosphate group and through interactions between polar side chains and the phosphate group (Momany et al., 1995b). This substructure is observed in all aminotransferases and decarboxylases with known structure. α-Helix dipole coupling is also thought to be important in the binding of phosphate moieties in many other proteins (Hol et al., 1978). Motif 3 corresponds to beta strand 2 and helix B along with connecting loop II and contains a possible proton donor to the cofactor (PD, Table 4A). Beta strand 3, helix C, and loop III, which follow motif 3, do not correspond to any motif in the final model. Motif 4 corresponds to beta strand 4 and the N terminal of helix D connected by loop IV and contains the glycine IN (Table 4A) that is at the interface between the large and small domains. Motif 5 starts with the C terminal part of helix D, followed by beta strand 5, which contains the highly conserved (see Fig. 2) aspartate residue SB (Table 4A) that makes a salt bridge with the pyridine nitrogen atom (N_1) of pyridoxal-P. It ends with loop V, which lies in the reverse direction of loop IV and helps to create the crevice for the coenzyme. Loop V also contains residue OHD and residue AR (Table 4A); residue OHD makes a

Table 2. *Total p-value for GAD and four pyridoxal-P dependent proteins of known structure*[a]

Proteins	$-\log_{10}$ (p-value)
1DGD	37.47
9AAT	18.27
1ORD	5.17
1TPL	1.36
GAD65	14.86
GAD67	15.68
9AAT(Jackknife)	6.59
1DGD(Jackknife)	8.39
1ORD(Jackknife)	2.33
1TPL(Jackknife)	1.21
GAD65(Jackknife)	9.36
GAD67(Jackknife)	11.62

[a] p-Values are 10^{-x}, where x is the value in the table, for example, the p-value for 1TPL after Jackknife is: $10^{-1.21} = 0.062$. p-Values before the Jackknife test were obtained by using the model to scan a database which contained the eight strutural proteins from PDB and the two GADs. p-Values for the Jackknife p-value experiment were generated by rebuilding the model using a database that had been purged of the query protein and its homologs, and then determining whether the query protein fit the reduced model.

hydrogen bond to the –OH of pyridoxal-P, while residue AR sits in front of the pyridoxal ring. Motif 6 corresponds to the C terminal of helix E, followed by strand 6 and beta strand 7. These two beta strands are linked by loop VI containing the highly conserved (see Fig. 2) lysine SF, which forms the Schiff base with the carbonyl group of pyridoxal-P. Strand 7 runs in the reverse direction of strand 6 and is the only anti-parallel strand among the seven strands.

Motif 10 is in the small domain, starting with a beta strand followed by a major section of helix. The beta strand is close to the active site and undergoes a conformational change when the protein binds the substrate, but none of its residues interact directly with the cofactor (McPhalen et al., 1992a).

Conserved residues

In previous work several residues in these structures were found to make contact with the pyridoxal-P (McPhalen et al., 1992a; Antson et al., 1993; Momany et al., 1995b; Toney et al., 1995). These residues were also identified by our model as shown in Table 4A. The degree of conservation in a multiple alignment is often mea-

Table 3. *Motif MAP scores from multiple sequence alignment*[a]

Motif	MAP score
1	−110.97
2	53.81
3	17.12
4	314.45
5	399.18
6	315.71
7	−192.93
8	−114.28
9	−150.44
10	−0.75

[a] The MAP score is the log of the probability ratio of drawing the sequence segment from the model compared with that from the randomly generated sequence with alignment space adjustment taken into account. MAP scores above zero indicate that the motif is more likely to belong to the model than to some unaligned random columns.

```
Protein   P-Value          Motifs

Motif2    (magenta)        ***************
1DGD      4.9:     104  RALLLSTGAESNEAAI    119
9AAT      1.9:     101  VTVQGISGTGSLRVGA    116
1ORD      1.7:     191  TYFVLGGSSNANNTVT    206
1TPL      2.0:      92  HIVPTHQGRGAENLLS    107
GAD65     3.4:     236  DGIFSPGGAISNMYAM    251
GAD67     2.9:     245  DGIFSPGGAISNMYSI    260


Motif3    (orange)         * ***************  **  **
1DGD      4.1:     124  LVTGKYEIVGFAQSWHGMTGAAASAT    149
9AAT      4.3:     122  FFKFSRDVYLPKPSWGNHTPIFRDAG    151
1ORD      4.5:     209  LVSNGDLVLFDRNNHKSVYNSALAMA    234
1TPL      0.3:     114  GQYVAGNMYFTTTRYHQEKNGAVDVD    139
GAD65     2.5:     268  AALPRLIAFTSEHSHFSLKKGAAALG    293
GAD67     2.1:     277  AAVPKLVLFTSEQSHYSIKKAGAALG    302


Motifs4   (olive)          ************************
1DGD      3.2:     203  NLAAFIAEPILSSGGIIELPDGY       225
9AAT      5.8:     183  KSIILLHACAHNPTGVDPRQEQW       205
1ORD      1.8:     280  RPFRLAVIQLGTYDGTIYNAHEV       302
1TPL      1.0:     175  IAYICLAVTVNLAGGQPVSMANM       197
GAD65     1.6:     328  VPFLVSATAGTTVYGAFDPLLAV       350
GAD67     2.1:     337  VPFYVNATAGTTVYGAFDPIQEI       359


Motif5    (aquamarine)     ************************
1DGD      8.8:     231  RKCEARGMLLILDEAQTGVGRT    252
9AAT      5.7:     210  SVVKKRNLLAYFDMAYQGFASG    231
1ORD      1.4:     304  KRIGHLCDYIEFDSAWVGYEQF    325
1TPL      2.8:     202  ELTEAHGIKVFYDATRCVENAY    223
GAD65     9.9:     352  DICKKYKIWMHVDAAWGGGLLM    373
GAD67    11.0:     361  DICEKYNLWLHVDAAWGGGLLM    382


Motif6    (cyan)       *  ****** **** ***************
1DGD      8.0:  259  QRDGVTPDILTLSKTLGAGLPLAAIVTSAAI    289
9AAT      3.2:  245  EQGIDVVLSQSYAKNMGLYGERAGAFTVICR    275
1ORD      1.7:  342  PEDPGIIVVQSVHKQQAGFSQTSQIHKKDSH    372
1TPL      0.3:  244  MFSYADGCTMSGKKDCLVNIGGFLCMNDDEM    274
GAD65     3.5:  383  GVERANSVTWNPHKMMGVPLQCSALLVREEG    413
GAD67     2.8:  392  GIERANSVTWNPHKMMGVLLQCSAILVKEKG    422


Motif10   (violet)       ************************
1DGD      9.5:  406  RIAPPLTVSEDEIDLGLSLLGQAIE    430
9AAT      2.2:  385  GRISVAGVASSNVGYLAHAIHQVTK    410
1ORD      2.4:  540  LFLMTPAETPAKMNNLITQLLQLQR    564
1TPL      2.1:  405  LTIPRRVYTYAHMDVVADGIIKLYQ    429
GAD65     4.3:  559  MVISNPAATHQDIDFLIEEIERLGQ    583
GAD67     4.8:  568  MVISNPAATQSDIDFLIEEIERLGQ    592
```

Fig. 1. Motifs identified from the multiple sequence alignment of pyridoxal-P–dependent proteins. The six motifs are shown for 1DGD, 9AAT, 1ORD, 1TPL, GAD67, and GAD65. The total p-value for each whole protein is listed in Table 2. The p-value for each individual motif is listed in this figure. The color name for each motif is the color displayed for that motif in Figure 3. The asterisks (*) above the amino acid symbols denote the columns selected by the model. The numbers before and after each sequence are the residue numbers of the first and final amino acids of the motif as used in the PDB structural representation. For 1DGD and 9AAT, the number differs from the residue number in the residue table of the PDB file. Our finding set also contained histidine decarboxylase and aromatic L-amino acid decarboxylase, which were aligned by Bu and Tobin (1994) with GAD. The results in this figure are consistent with their results.

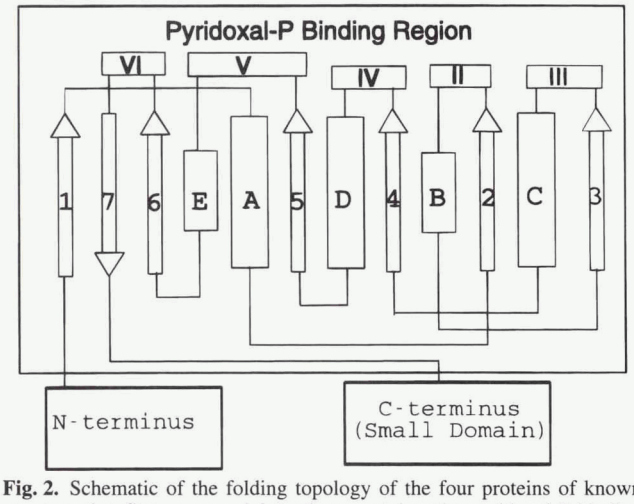

Fig. 2. Schematic of the folding topology of the four proteins of known structure that fit to our model. In these proteins, the pyridoxal-P binding domain is called the large domain. The C-terminal region often contains another domain called the small domain, which closes upon substrate binding. Helices are represented by perpendicular rectangles, strands are represented by flat thick arrows, and loops that form the binding cleft are represented by horizontal rectangles.

sured in information bits (Schneider & Stephens, 1990), and these residues carried high information content in our model (Table 4; Fig. 4). Figure 4 illustrates the six conserved motifs as a stylized bar graph with the single letter amino acid symbols drawn proportional to the degree of conservation of that amino acid. The height of the letters in this graph thus simultaneously indicates the degree of conservation of a position in the sequence and the importance of a particular amino acid at that position.

Because the model represents a large set of very distantly related proteins, there are only two positions that are even close to be completely conserved and thus well represented by a consensus. Even though there are only two highly conserved positions in this superfamily, the availability of a large number of distantly related sequences makes it possible to obtain an alignment because there are many moderately and weakly conserved positions. In a moderately or weakly conserved position, the most likely residue often has a probability of under 50%. As a consequence, the most conserved residue type at some positions may not be present in any particular sequence or even in a small set of sequences from this superfamily. For example, even though threonine is the most conserved amino acid at position 9 of motif 2, it only has a value of 0.68.

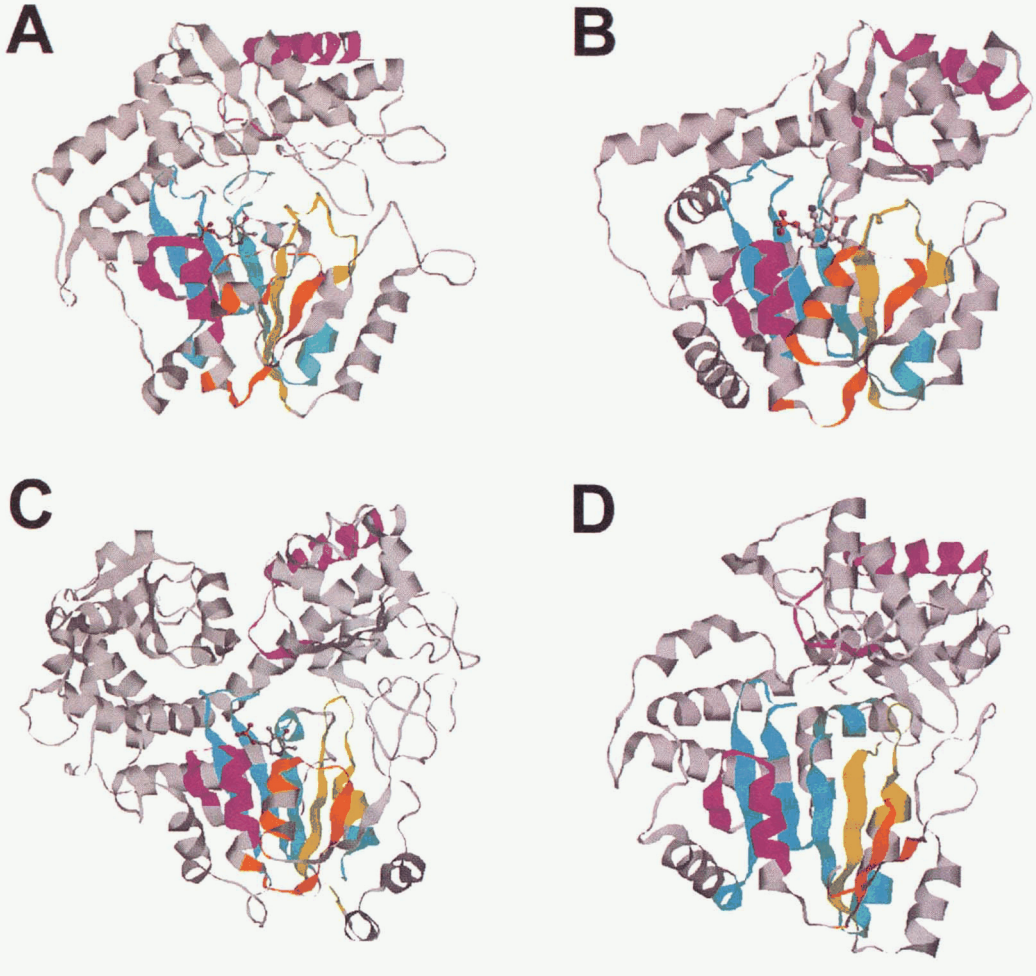

Fig. 3. Structural views showing the positions of our motifs in the four proteins of known structure. Motifs 2–6 and 10 are colored as magenta, orange, olive, aquamarine, cyan, and violet, respectively. **A:** 1DGD. **B:** 9AAT. **C:** 1ORD. Residues 598 to 730 have been removed to provide better view of the pyridoxal-P binding domain. **D:** 1TPL.

Table 4. *Conserved residues in the six pyridoxal-P proteins*

Residue name and its function	Motif	1DGD	9AAT	1ORD	1TPL	GAD65	GAD67	Info bits[g]
A. Conserved residues reported in the literature								
SF (Schiff base with PLP)	6	Lys272[a]	Lys258[a,b,c]	Lys355[c]	Lys257[d]	Lys396	Lys405	3.34 ± 0.51
SB (Salt bridge with N1 of PLP)	5	Asp243[a]	Asp222[a,b,c]	Asp316[c]	Asp214[d]	Asp364	Asp373	3.80 ± 0.51
OHD (H bond to OH of PLP)	5	Gln246[a]	Tyr225[a,b,c]	Trp319[c]	Arg217[d]	Trp367	Trp376	0.92 ± 0.51
PH1	2	Gly111[a]	Gly108[b,c]	Ser198[c]	Gly99[d]	Gly243	Gly252	1.64 ± 0.51
PH2 (H bond to OP of PLP)	6	Thr269[h]	Ser255[b,c]	Ser352[c]	Ser254[d]	Asn393	Ans402	1.68 ± 0.51
PH3	2	Ala112[a]	Thr109[b]	Ser199	Arg100[d]	Ala244	Ala253	1.40 ± 0.51
PD (Cofactor binding, possible proton donor)	3	Trp138[a]	Trp140[b,c]	His223[c]	Tyr128	His282	His291	1.18 ± 0.51
AR (Hydrophobic environment of pyridinium ring)	5	Ala245[a]	Ala224[a,b]	Ala318	Thr216[d]	Ala366	Ala375	1.62 ± 0.51
IN (Interface of large and small domain)	4	Gly217[e,f]	Gly197[e,f]	Gly294[e,f]	Gly189[e,f]	Gly342	Gly351	1.76 ± 0.51
B. Residues with high information bit but without full literature match								
	2	Thr110	Ser107[b]	Gly197	Gln98[d]	Gly242	Gly251	1.52 ± 0.51
	4	Ser214	Asn194[b]	Thr291	Leu186	Thr339	Thr348	1.36 ± 0.51
	6	Ser271	Ala257	His354[c]	Lys256	His395	His404	1.40 ± 0.51
C. Residues with high information bit but not noted in the structural literature								
mtf2α1	2	Ser114	Ser111	Ala201	Ala102	Ser246	Ser255	1.56 ± 0.51
mtf2α2	2	Ala118	Gly115	Val205	Leu106	Ala250	Ser259	1.42 ± 0.51
mtf3β1	3	Ile131	Val133	Val216	Met121	Ala275	Leu284	1.24 ± 0.51
mtf3β2	3	Val132	Try134	Leu217	Tyr122	Phe276	Phe285	1.78 ± 0.51
mtf4β	4	Phe207	Leu187	Leu284	Cys179	Val332	Val341	1.42 ± 0.51
mtf4α	4	Tyr225	Trp205	Val302	Met197	Val350	Ile359	1.56 ± 0.51
mtf5β	5	Gly237	Asn216	Cys310	Gly208	Lys358	Asn367	1.82 ± 0.51
mtf6β1	6	Th264	Val250	Ile347	Asp249	Asn388	Asn397	1.52 ± 0.51
mtf6β2	6	Ala282	Gly268	Gln365	Leu267	Ala406	Ala415	1.44 ± 0.51
mtf10α1	10	Gly421	Leu400	Leu555	Val420	Leu574	Leu583	1.56 ± 0.51
mtf10α2	10	Leu425	Ile404	Leu559	Ile424	Ile578	Ile587	1.88 ± 0.51

[a]Toney et al. (1995).
[b]McPhalen et al. (1992).
[c]Momany et al. (1995)
[d]Antson et al. (1993).
[e]Mehta et al. (1993)
[f]Pascarella et al. (1993)
[g]Information bit is used to estimate the residue uncertainty at each position. The lower the information bit, the bigger the uncertainty of the residue type at that position. The maximum information bit at each position is given by $\log_2(20) = 4.3$ (Schneider & Stephens, 1990). The error uncertainty is due to the limited sample sequence number.
[h]From the structural data, this residue is about 8 Å away from the nearest oxygen atom on the phosphate group, implying a false positive identification.

Each of the five core motifs contains at least one highly conserved residue. The most highly conserved are the aspartate that forms a salt bridge with the pyridinium ion and the lysine that forms a Schiff base of pyridoxal-P, but conserved residues that interact with the phosphate and the ring of pyridoxal-P were also identified in the model. The three-dimensional orientations of these critical residues are shown in Figure 5A, B, and C for 9AAT, 1DGD, and 1ORD. The residues are not illustrated for 1TPL because its structure is available only as the apoenzyme.

Eleven additional conserved residues that have not been previously identified are shown in Table 4C. None of these residues interact directly with pyridoxal-P, but rather are located some distances from it. Furthermore, examination of the 1AMA (from PDB), a co-crystal of aspartate aminotransferase and alpha-methyl aspartate, the substrate analog, shows that none of these residues are in close proximity with the substrate. Thus, these residues do not appear to be involved in either substrate or cofactor interaction; rather they appear to play a role in stabilizing structure.

Two of these residues, mtf10α1 and mtf10α2, are found in motif 10 In all four proteins of known structure, they are located near the center of a large helical segment that lies on the surface of the small domain. In each case these two residues are part of a series of hydrophobic amino acids that form the hydrophobic side of an amphipathic helix. Both face inward and interact with hydrophobic amino acids in the protein's interior. Thus, our model predicts that motif 10 is an amphipathic helix with these two conserved residues on its hydrophobic face. There is experimental evidence supporting this prediction; this motif is located at the carboxyl terminal of both GADs (its C-terminal Q is the third residue from the C-terminal of GAD) and an antiserum against a peptide corresponding to the carboxyl-terminal residues of GAD immunoprecipitate the enzyme (Karlsen et al., 1992).

Potential roles of the remaining residues in Table 4C are perhaps most easily described using a three-dimensional frame of reference. If one rotates the structure so that the pyridoxal-P is at the top and in front of the roughly rectangular plane formed by the core binding motifs, the phosphate group will point to the left, as shown in Figure 3, and the residues in Table 4C can be placed in the following three groups: right (right edge of the rectangle), left, and bottom.

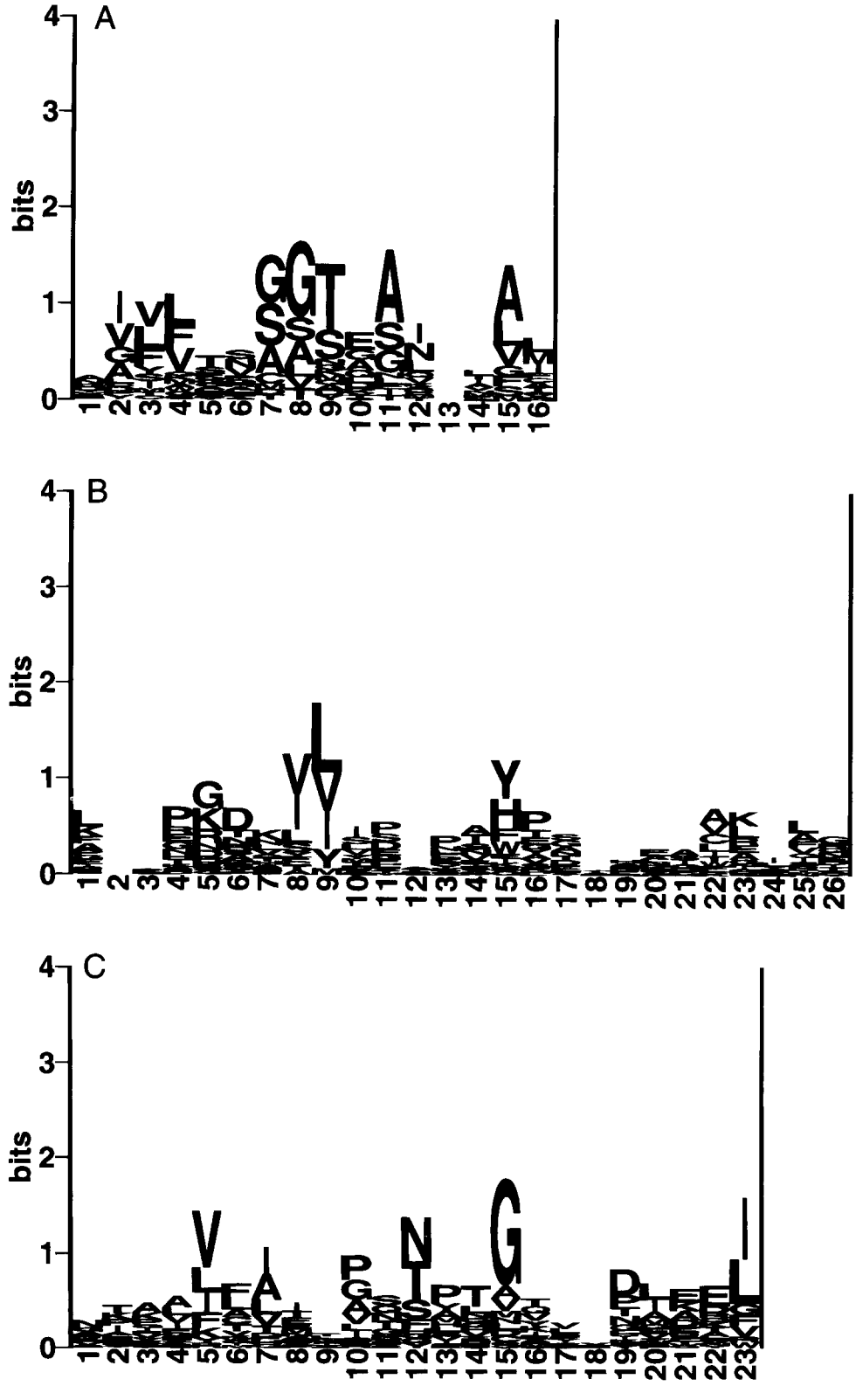

Fig. 4. Sequence logos of the six motifs identified by PROBE. The sequence logos were constructed from a database containing 27 distantly related sequences. The 27 sequences were obtained by purging the original identified 512 sequences at BLAST score of 50, thus decreasing the weight from the closely related sequences. The horizontal axis represents the position of the residue within the motif. The vertical axis represents the amount of information (in bits) that this position holds. The height of the one letter residue symbol at each position is proportional to the information bit of the residue at that position. The uncertainty due to the limited number of the sampling sequences is ±0.51 for all the positions. (**A**) Motif 2; (**B**) motif 3; (**C**) motif 4; (**D**) motif 5; (**E**) motif 6; and (**F**) motif 10. (*Figure continues on facing page.*)

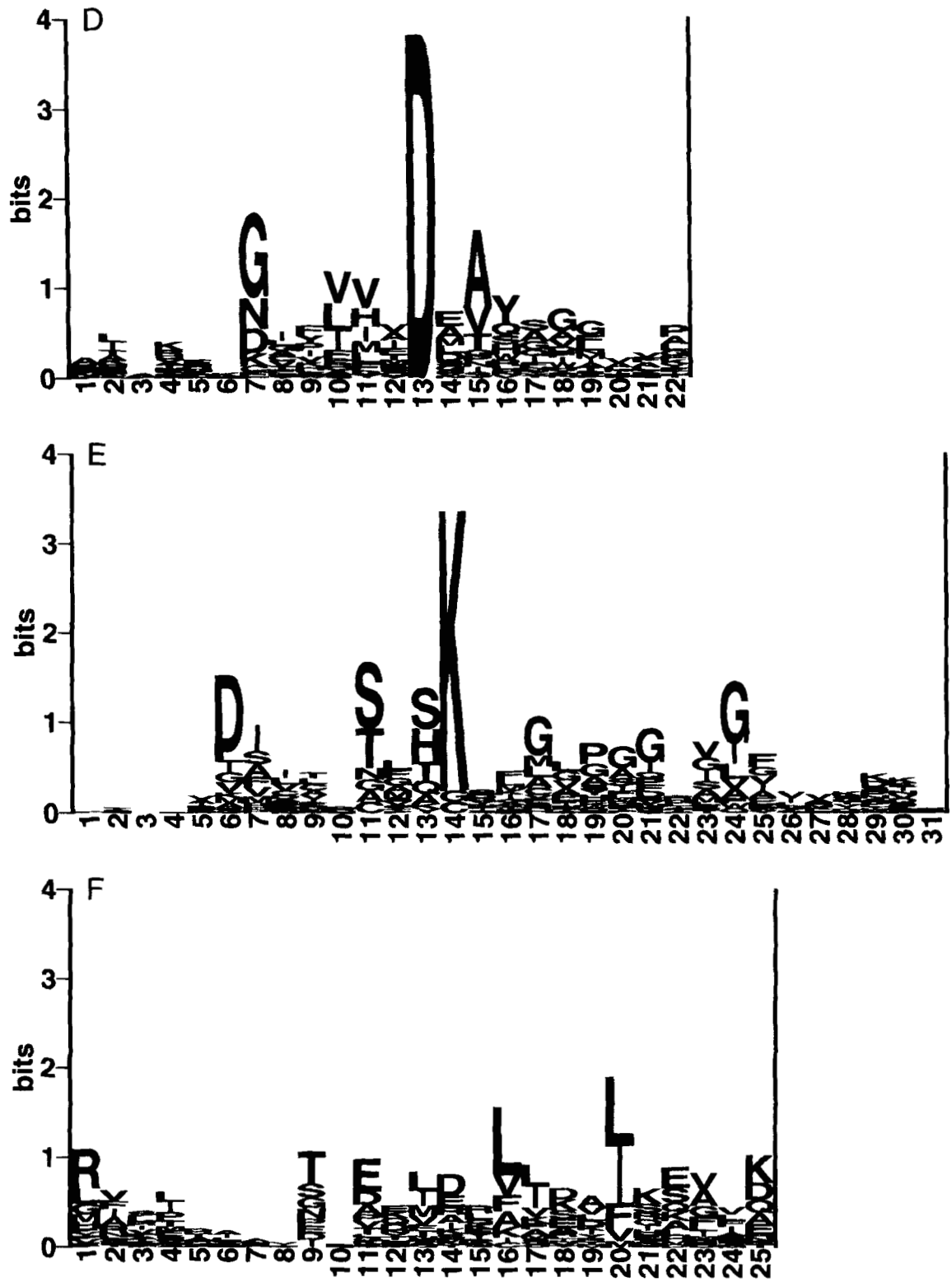

Fig. 4. *Continued.*

The right group includes residues mtf3β1, mtf3β2, mtf4β, and mtf4α. Residues mtf3β1 and mtf3β2 are adjacent in sequence and are located in the middle of beta strand 2. The side chain of Mtf3β1 is at front of the plane and mtf3β2 is pointing behind the plane. Residue mtf4β is located near the middle of the long beta strand 4, and the side chain is at the front of the plane (except of

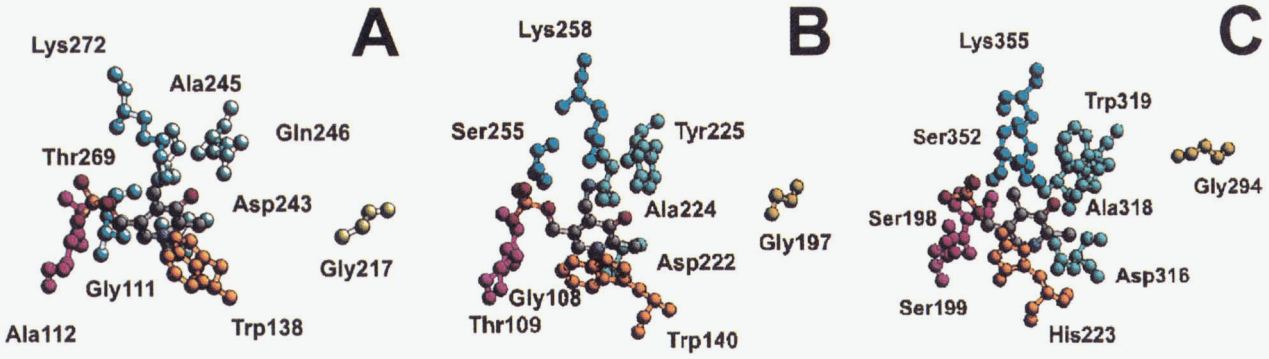

Fig. 5. Relative orientations of the critical conserved residues (listed in Table 4A) to pyridoxal-P. Each residue has the same color as the motif that it resides in (**A**) 1DGD, (**B**) 9AAT, and (**C**) 1ORD.

the one from 1DGD). They are in contact or approximate contact and are surrounded by a pocket of hydrophobic amino acids. They appear to play a role in maintaining the correct structural relationship between beta strands 2 and 4. Residue mtf4α is also on the right edge of the pyridoxal-P α/β core and is behind the β strand plane. The side chain of this mtf4 points into the hydrophobic core consisting of beta strands 4, 5, and helix D, interacting with the hydrophobic residues from strands 4 and 5, and thus may provide further stabilization to the pyridoxal-P α/β core.

The left group includes residues mtf2α1, mtf2α2, and mtf6β2. The first two residues are approximately in the middle of helix A and separated by three residues. The side chain of residue mtf2α1 points toward beta strand 7, in close proximity to residue mtf6β2, which resides in strand 7. These three residues may play a role in anchoring the left flank of the pyridoxal-P binding core, interacting with each other to stabilize this end of the tertiary core structure.

The bottom group consists of residues mtf5β and mtf6β1. Residue mtf5β is located in the loop just before beta strand 5, and residue mtf6β1 is at the N-terminal of beta strand 6. We cannot suggest a more specific structural or functional role for these residues at this time.

Structural superposition and prediction

The superposition of the α-carbon atoms of the five core motifs of 1DGD on each of the other three enzymes is shown in Figure 6. The root-mean-square (RMS) deviations of the fits for 1DGD with 9AAT, 1ORD, and 1TPL were 4.21, 5.00, and 5.17 Å. We also used a recently developed structural neighboring and alignment procedure, VAST, which identifies remote structural homologs (Gibrat et al., 1996) and aligns significantly related structural fragments. The motifs identified here strongly overlap structurally conserved fragments identified by VAST, and the RMS deviations obtained from VAST for these proteins are also in the range of 4 to 5 Å.

To determine whether the identified side chains of GAD can pack around pyridoxal-P in a manner similar to these other enzymes, we modeled the five core motifs of GAD by homology using Xlook (Xlook Version 2.0, 1996; Molecular Applications Group, Palo Alto, California). The four modeled structures of each GAD based on these four pyridoxal-P proteins from PDB (parents) were consistent with their parents and also in good agreement among themselves. A more detailed atomic prediction was achieved by carefully examining the pyridoxal-P binding pocket of the three proteins (1DGD, 9AAT, and 1ORD). Nearly all of the residues at van der Waals contact distances are included in Table 4A. For 1ORD, the five residues indicated to make direct contact with the pyridoxal-P ring are all that make such contact. As Table 4A demonstrates, the aligned residues of GAD are identical to these five residues. Accordingly, we have developed a structural prediction of the pyridoxal-P binding core for GAD65 and GAD67 based on the structure of 1ORD. Figure 7A shows the predicted structures for the pyridoxal-P motif binding core of GAD65. Because the sequence of GAD65 is highly homologous to that of GAD67, the predicted five pyridoxal-P binding core motifs are very similar to that of GAD65 (not shown here). Figure 7B shows the superposition of the pyridoxal-P binding pocket (the Van der Waals interacting residues) from the predicted GAD65 model and 1ORD.

His-354 in 1ORD, which is adjacent to the Schiff base-forming Lys-355, makes a salt bridge with the phosphate of pyridoxal-P. This position also has a high information bit (position 13 of motif 6, see Fig. 2E), and the corresponding residue in the two GADS (His 395–404) appears to interact with the cofactor in the same way (Fig. 7).

The role of the motifs in subunit interactions

Motifs 2, 3, and 6 form part of the interface between the two subunits that make up the basic dimeric structure of 1ORD and 9AAT and the predicted dimeric structure of GAD (Table 5). In most cases, the interface-forming residues within a motif in one subunit interact with residues that are not part of the conserved motifs in the other subunit, although in a few instances, a specific residue in one motif of one subunit interacts with the same residue in the other subunit. The interactions between the conserved motifs and residues in the other subunit constitute only a fraction of the total number of residues at the interface between the subunits. In 1ORD and 9AAT the cofactor itself also forms part of the interface between the subunits, as polar residues in one subunit interact with the cofactor phosphate in the active site of the other subunit, and our GAD model incorporates a similar feature (Table 5; McPhalen et al., 1992a; Momany et al., 1995b). The interface-forming residues in 1DGD were not identified because the structure in PDB contains only one subunit. For 1TPL, only the structure of the apoenzyme is available in PDB, and it appears to be in a much more open conformation than the other proteins.

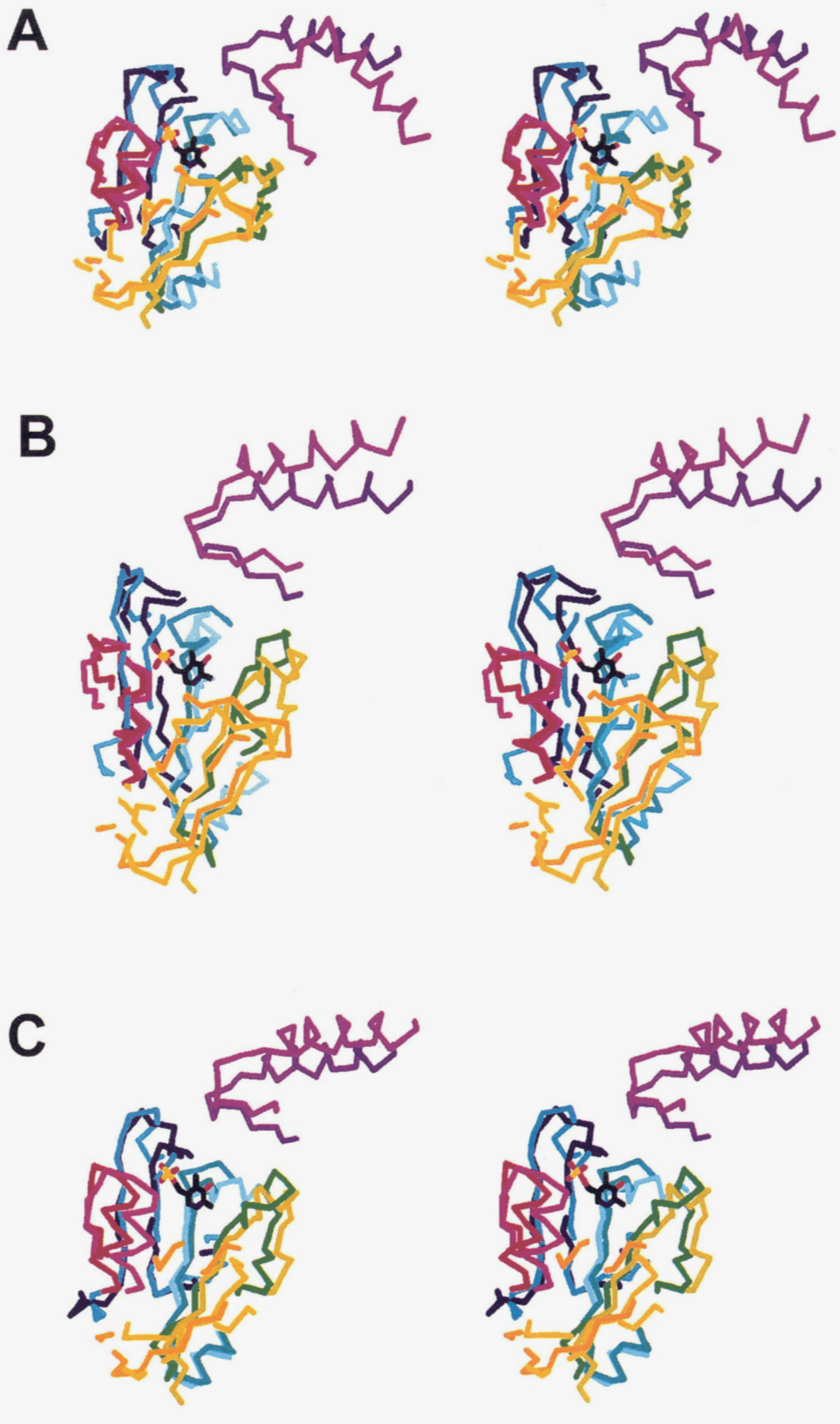

Fig. 6. Stereo views of the superimposition of the five core motifs from the pyridoxal-P–dependent proteins of known structure. The parent structure is 1DGD and the motifs are: 2, magenta; 3, orange; 4, olive; 5, aquamarine; 6, cyan; 10, violet. The target proteins are 9AAT, 1ORD, and 1TPL, respectively. The motifs are: 2, red; 3, yellow; 4, green; 5, turquoise; 6, blue; 10, purple. **A:** 9AAT over 1DGD. The RMS is 4.21 Å. It increases to 5.06 Å when motif 6 is included. **B:** 1ORD over 1DGD. The RMS is 5.00 Å. It changes to 4.81 Å when motif 6 is included. **C:** 1TPL over 1DGD. The RMS is 5.17 Å. It increases to 5.41 Å when motif 6 is included.

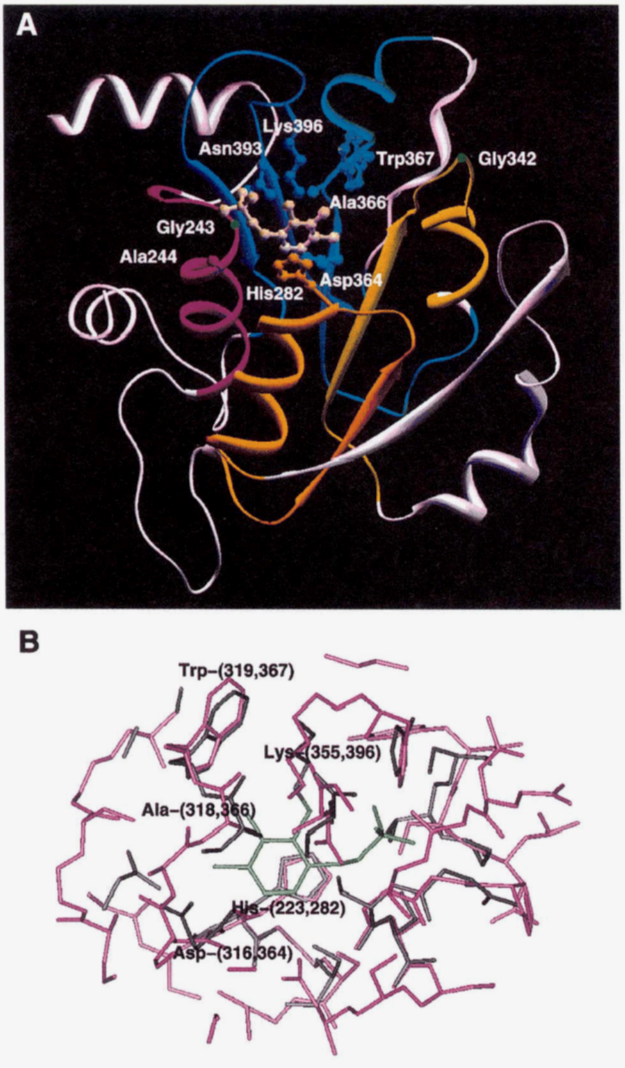

Fig. 7. Predicted fold of pyridoxal-P binding core for GAD65 from Xlook. Because Xlook did not permit the specification of a covalent bond between the Schiff base lysine and the pyridoxal-P ring, we performed an adjustment to demonstrate that there was sufficient flexibility in the lysine side-chain dihedral angles and space in the structure to permit the lysine to closely approach the C4′ of pyridoxal-P ring. **A:** Predicted structural fold of pyridoxal-P binding core for GAD65. The motifs are colored as in Figure 3. The pyridoxal-P is represented in ball-and-stick form and is colored in yellow. The side chains of those residues in Table 4A are also represented in ball-and-stick form except two glycines (Gly342 and Gly243). The backbone positions of the two glycines are represented by two green dots. **B:** Superimposition of the pyridoxal-P binding pocket of the predicted GAD65 structure and 1ORD based on superposition of pyridoxal-P. The binding pocket for 1ORD consists of two shells of residues. The residues in the first shell are those that make direct van der Waal's contact with pyridoxal-P. The second shell consists of those that make direct van der Waals contact with the residues in the first shell. The residues for GAD65 are the those within 8 Å of pyridoxal-P. GAD65 is shown in magenta, 1ORD in black, pyridoxal-P in green. The residues that make direct van der Waal's contact with pyridoxal-P ring residues are text labeled for 1ORD with corresponding aligned residues from GAD65.

Discussion

Our results prompt us to predict that the folding of the core pyridoxal-P binding site in each GAD resembles that of the four pyridoxal-P binding proteins of known structure examined here.

Table 5. *Core motif residues involved in subunit interactions*

	Subunit A		Subunit B	
Enzyme	Motif	Residue	Motif	Residue
1ORD	2	Leu195	—	Gln362
		Ser199	—	Thr395
		Asn200	—	Phe392
		Asn203	—	Leu391
	3	Asn212	—	Tyr629
		His223	—	Thr395
		Lys224	—	Asn390
		Met233	3	Met233
			—	Leu391
			—	Phe392
	6	Ala358	—	Tyr135
			—	Tyr136
		Phe360	—	Phe145
			—	Val153
			—	Phe154
			—	Phe398
		Ser361	—	Asp157
		Gln362	6	Gln362
			2	Leu195
			—	Phe398
	—	Cofactor	—	Ser394
			—	Thr395
			—	Ser396
9AAT	2	Ile106	2	Ile106
			—	Tyr295
		Arg113	—	Pro293
			—	Met294
	3	Pro145	—	Pro293
		Asp149	—	Arg121
	6	Lys258	—	Tyr70
		Leu262	—	Lys68
		Tyr263	—	Tyr70
		Gly265	—	Lys68
		Arg266	—	Asn297
	—	Cofactor	—	Tyr70
			—	Ser296
GAD65	2	Pro241	2	Pro241
			—	Leu435
		Ala244	—	Leu435
			—	Ser436
		Ile245	2	Ile245
			—	Tyr433
		Met248	—	Tyr433
			—	His432
	3	Phe283	—	Lys431
		Lys287	—	His432
			—	Gln428
		Leu292	3	Leu292
			—	Met252
			—	Tyr433
	6	His395	—	Gly174
		Lys396	—	Lys172
			—	Gly174
		Met398	—	His175
		Gly399	—	Asn180
			—	Gln181
		Val400	—	His175
		Leu402	—	Thr173
			—	Thr202
			—	Tyr437
			—	Asp438
			—	Thr439
		Gln403	—	Thr439
			—	Asp438
			—	Tyr437
	—	Cofactor	—	Ser436

Accordingly, in each GAD the core should be composed of an α/β fold of at least six strands with helices on either side, as described by the motifs 2 through 6 in Figures 1 and 2. We further predict that GAD has most, if not all, of the residues that have been cited in the structural literature to be important in this class of pyridoxal-P binding proteins (Table 4A). Thus, the basic features of the pyridoxal-P binding site of each GAD appear to resemble each other and other pyridoxal-P binding proteins. We also predict that GAD has a small domain with a large helix like the four structurally available proteins. This prediction is supported by the presence of motif 10 in the model and by the high information bit of glycine-(342,351) in the GADs, as this residue lies at the interface between the large and small domains in the proteins of known structure. Our overall structural prediction agrees with that of Momany et al., but, as discussed below, refines it in several important ways.

The primary evidence supporting these predictions is our finding that the structural motifs and critical residues of the four pyridoxal-P binding enzymes with known structure are correctly predicted by our alignment methods. Because the method that we use is completely automatic, it provides an unbiased means to test predictions. This set of predictions makes a good set of positive controls because, as shown in Table 1, all of these proteins are as distant from one another in sequence as they are from GAD.

Our results refine those of Momany et al. by focusing attention on those motifs and residues that are conserved according to statistical criteria. In their alignment Momany et al. included sequences corresponding to our motifs 2–7 and part of 8, and also many amino acids that lie between our motifs. We excluded motifs 1, 7, and 8 because of nonsignificant MAP scores, and also numerous other nonsignificant residues. Thus, our motifs 2–6, which form the pyridoxal-P binding site, include 108 significant amino acids, which is fewer than half of the number included from the GAD sequence by Momany et al. We also identified motif 10, which was not included by Momany et al., and identified 11 other conserved residues that appear to play important roles in establishing the core pyridoxal-P binding site. Thus, we expect our findings to substantially improve the ability to design studies of the regulation of cofactor function in GAD.

The functions of most of the highly conserved residues are predicted by our motif models. In GAD, the pyridine ring is sandwiched by a hydrophobic methyl group of alanine and a side chain ring group, both of which are within van der Waals' contact range. This ring stacking is important for enzyme catalysis, because it favors the quinonoid intermediate formation with improved ring stacking (John, 1995). Also, the prediction indicates that lysine-(396,405) for GAD65 and GAD67 form a Schiff base with pyridoxal-P and that the adjacent histidine interacts with the phosphate of pyridoxal-P as it does in 1ORD. Aspartate-(364,373) binds to the pyridine nitrogen of pyridoxal-P in a similar fashion to the four proteins with known structure. The indole hydrogen of tryptophan-(367,376) bends toward the hydroxyl group of pyridine ring to make a hydrogen bond. The glycine-(342,351) (motif 4), which lies at the interface of the large and small domains and is well conserved in other proteins, was also identified. A larger residue does not appear to be permitted at this position in most proteins, because a bulky residue might interfere with the folding of the polypeptide chain or the change from the open to the closed conformation that occurs when the substrate binds (Picot et al., 1991; Mehta et al., 1993).

When analyzing the homology modeling results based on 1ORD, we noticed that the side chain of tryptophan-(352,401), which is near the binding pocket, had two totally different conformational states: one pointed into the pyridoxal-P binding pocket, and the other directed out of the pocket. Both fit well with the rest of the protein and the pyridoxal-P coenzyme. The fact that all animal GAD sequences (except *Caenorhabditis elegans*) in our findings have a tryptophan at this position suggests the speculative possibility that this residue plays a role in GAD's regulated uptake and release of pyridoxal-P.

Our results also have implications for the quaternary structure of GAD. The basic subunit structure of GAD and the other decarboxylases and transaminases in its group is dimeric (Stark et al., 1991; McPhalen et al., 1992a; Momany et al., 1995b; Sheikh & Martin, 1996). In the proteins of known structure, the pyridoxal-P is located deep within each subunit and near the interface with the other subunit. Indeed, residues from one subunit can reach into the cofactor binding site of the other subunit and interact with the phosphate group. Our results strongly suggest that several residues in motifs 2, 3, and 6 form part of the internal face of each subunit and are positioned to interact directly with residues in the other subunit of the enzyme. Thus, the C-terminal domain of GAD, which contains the core pyridoxal-P binding structure, is doubtless responsible for the basic dimeric structure of the enzyme. The specific residues involved at the subunit interface of GAD may differ from those given in Table 5, however, because many of the residues are not part of the conserved motifs and are distant from the cofactor where the proposed structure is probably most accurate. Unrelated interactions between sequences in the N-terminal domain appear to be responsible for the larger heteromeric forms of GAD (Solimena et al., 1993, 1994; Dirkx et al., 1995).

The interactions between the cofactor phosphate and residues in the other subunit are interesting because they may provide a means of allosteric communication between the two active sites. However, the residues that participate in these interactions are not located in the conserved motifs, and we have confidence that this prediction for GAD is much lower than in the predicted interactions of the conserved motifs with the cofactor.

Two other aminotransferases, which are not in the PDB, ω-amino acid pyruvate aminotransferase and phosphoserine aminotransferase, have a structural fold similar to that of 9AAT (Watanabe et al., 1991; Stark et al., 1991). They also fit our model with total p-values of 37.82 and 7.49; although they have very limited sequence homology with 9AAT (Table 1). Structures of two additional pyridoxal-P–dependent proteins, glutamate-1-semialdehyde aminomutase (from α family) and cystathionine β-lyase (from γ family, Alexander et al., 1994), have been determined recently (Clausen et al., 1996; Hennig, 1997). They also fit our model with p-values of 26.43 and 16.87, respectively. When we compared Figure 2 of Hennig and Figure 1 of Clausen with our alignment model, we found that the secondary structures and the relative tertiary orientation of the six identified motifs in AD-MT and β-LS corresponded very well to those of the four proteins from PDB.

The remaining four pyridoxal-P–dependent enzymes in PDB (1WSY, 1DAA, 1SFT, and 1PYG) do not fit our model, and they bear a quite different tertiary fold indeed. They appear to belong to other superfamilies. A detailed classification and analysis will be discussed in another article from our group.

In summary, we find that GAD is a member of a large superfamily of pyridoxal-P binding proteins that includes all the α and γ proteins (Alexander et al., 1994). Our findings indicate that this superfamily contains six conserved motifs that contain an average about 20–30% of the residues in these proteins. These motifs and

conserved residues appear to play important roles in stabilizing of the monomer and dimeric structure and also in cofactor binding.

Methods

Brief summary of the multiple-sequence alignment method: PROBE

Lawrence et al. (1993) developed a Gibbs sampler to detect conserved patterns in multiple sequences that was subsequently extended and fully delineated as a Bayesian Monte Carlo Markov chain process (Liu et al., 1995; Neuwald et al., 1995). This method was further integrated with a hidden Markov Model (HMM) propagation algorithm (Liu & Lawrence, 1996), extensively improved, and combined with a database search procedure to make an improved multiple sequence identification tool: PROBE (Neuwald et al., 1997).

PROBE can start with either a set of sequences or a single sequence. For a set of sequences, PROBE performs a Gibbs alignment using propagation and a genetic algorithm. It creates motif models based on the alignment. These alignment motifs are next used to search the nonredundant database (which contains 103,515 sequences) to extract additional sequences that fit the original motif models created by PROBE. It uses these additional sequences to refine the motif models and then repeats the process until the number of the recruited sequences does not increase significantly. The original set should contain at least five sequences.

When starting with a single sequence, PROBE performs a transitively extended BLAST search to attempt to find an initial set of related sequences to form the basis of the above procedure.

Applying PROBE to our case

Our initial attempt to run PROBE on the set of the six pyridoxal-P binding proteins with known structure failed because these six proteins share very limited sequence similarity. PROBE was then run under the single sequence condition for each of the six sequences. Tyrosine phenol-lyase failed to recruit related sequences. For the other five families, related proteins were identified. Because histidine decarboxylase shares significant sequence identity with GAD (Bu & Tobin, 1994), another PROBE group based on histidine decarboxylase and glutamate decarboxylase was created. All of these sets were combined into an extended set. When PROBE was applied to the extended set, 512 sequences were identified as members of this superfamily. A final fine alignment adjustment using propagation was carried out on this set, which contains 512 protein sequences with 50 sampling iterations.

In addition to the motif locations from the results of the alignment, PROBE provides other useful information. A p-value is assigned to each motif and to the whole protein sequence. The p-value is the probability of finding a motif by chance in a search of random, unrelated sequences of the same length. The p-values listed in Figure 1 and Table 2 are in negative log base 10 units of the original value.

Examination of PROBE output

We used the following procedures to examine the raw output of PROBE: (1) we checked the p-values of each motif to determine whether they are significant or not (p-value < 0.05); (2) we used the MAP score to identify significant motifs. Alignment scores always represent a balance between the benefit of aligning related residues and the penalty of introducing gaps. There is an entropic explosion in the number of alignments with increasing number of gaps, k. Specifically as k increased, the number of alignments grows exponentially. Traditionally, gap penalties have been employed to control this explosion. Here gap penalties are not used. Rather, the entropic explosion is controlled through the inclusion of a factor that is inversely proportional to the number of alignments with k gaps (see Liu et al., 1998, for details). The MAP score represents this balance and works to find significant motifs because it is an approximation of the rigorous Bayesian posterior probability for model selection (Zhu et al., 1997). (3) We tested the predicted model with the Jackknife test. This test is performed to assure that a protein and homologs that may be contained in the model set do not recruit themselves. The Jackknife p-value test was conducted by removing the query sequence and its homologs from the model, creating a new model, and scanning the database of interest with the new motif model to determine whether the query sequence is related to this reduced model.

Detailed structure prediction

Xlook, a software package based on a database search strategy and the energy minimization algorithms of Levitt (1983, 1992), was used for detailed structural prediction. This procedure uses the coordinates from a protein with known structure, the parent, as a basis to derive the predicted coordinates of the target protein. The coordinates of the parent are combined with coordinates of target homologous segments from the structural database, to construct 10 initial models. An average model based on these 10 is derived. Energy minimization is employed to reduce steric overlap and produce the final predicted structure.

Availabilities

The statistical model for the GADs, the 512 proteins recruited by the model and the coordinates predicted for GAD are available at: www.wadsworth.org/res&res/bioinfo. PROBE can be obtained by anomymous ftp at: ncbi.nlm.nih.gov.

Acknowledgments

We thank Michael Palumbo for the position adjustment performed in Figure 7. This work is partially supported by grants MH35664 to D.L.M. from the USPHS/DHHS, and by grants DEFG0296ER62266 from DOE and 5R01HG0125702 from NIH to C.E.L. The authors would also like to express thanks to the Computational Molecular Biology and Statistics Core at Wadsworth Center for Laboratories and Research.

References

Alexander FW, Sandmeier E, Mehta PK, Christen P. 1994. Evolutionary relationships among pyridoxal-5′-phosphate-dependent enzymes Regio-specific α, β and γ families. *Eur J Biochem 219*:953–960.

Antson AA, Demidkina TV, Gollnick P, Dauter Z, Von Tersch RL, Long J, Berezhnoy SN, Phillips RS, Harutyunyan EH, Wilson KS. 1993. Three-dimensional structure of tyrosine phenol-lyase. *Biochemistry 32*:4195–4206.

Bu DF, Tobin AJ. 1994. The Exon-Intron organization of the genes (GAD1 and GAD2) encoding two human glutamate decarboxylase (GAD67 and GAD65) suggests that they derive from a common ancestral GAD. *Genomics 21*:222–228.

Christgau S, Aanstoot HJ, Schierbeck H, Begley K, Tullin S, Hejnaes K, Baekkeskov S. 1992. Membrane anchoring of the autoantigen GAD_{65} to

microvesicles in pancreatic β-cells by palmitoylation in the NH_2-terminal domain. *J Cell Biol Vol 118*:309–320.

Clausen T, Huber R, Laber B, Pohlenz HD, Messerschmidt A. 1996. Crystal structure of the pyridoxal-5′-phosphate dependent cystathionine beta-lyase from *Escherichia coli* at 1.83 Å. *J Mol Biol 262*:202–224.

Dirkx R Jr, Thomas A, Li L, Lernmark A, Sherwin RS, DeCamilli P, Solimena M. 1995. Targeting of the 67-kDa isoform of glutamic acid decarboxylase to intracellular organelles is mediated by its interaction with the NH_2-terminal region of the 65-kDa isoform of glutamic acid decarboxylase. *J Biol Chem 270*:2241–2246.

Erlander MG, Tillakaratne NJK, Feldblum S, Patel N, Tobin AJ. 1991. Two genes encode distinct glutamate decarboxylases. *Neuron 7*:91–100.

Gibrat JF, Madej T, Bryant SH. 1996. Surprising similarities in structure comparison. *Curr Opin Struct Biol 6*:377–385.

Hennig M, Grim B, Contestabile R, John RA, Jansonius JN. 1997. Crystal structure of glutamate-l-semialdehyde aminomutase: An alpha2-dimeric vitamin B6-dependent enzyme with asymmetry in structure and active site reactivity. *Proc Natl Acad Sci USA 94*:4866–4871.

Hol WGJ, van Duijnen PT, Berendsen HJC. 1978. The α-helix dipole and the properties of proteins. *Nature 273*:443–446.

John RA. 1995. Pyridoxal phosphate-dependent enzymes. *Biochim Biophys Acta 1248*:81–96.

Karlsen AE, Hagopian WA, Petersen JS, Boel E, Dyrberg T, Grubin CE, Michelsen BK, Madsen OD, Lernmark Å. 1992. Recombinant glutamic acid decarboxylase (representing the single isoform expressed in human islets) detects IDDM-associated 64,000-M_r autoantibodies. *Diabetes 41*:1355–1359.

Lawrence EC, Altschul SF, Boguski MS, Liu JS, Neuwald AF, Wootton JC. 1993. Detecting subtle sequence signals: A Gibbs sampling strategy for multiple alignment. *Science 262*:208–214.

Levitt M. 1992. Accurate modeling of protein conformation by automatic segment matching. *J Mol Biol 226*:507–533.

Levitt M. 1983. Protein folding by constrained energy minimization and molecular dynamics. *J Mol Biol 170*:723–764.

Liu JS, Lawrence CE. 1996. Statistical models for multiple sequence alignment: Unifications and generalizations. *Proc Am Stat Assoc Stat Comput Sect*:1–8.

Liu JS, Neuwald A, Lawrence CE. 1995. Bayesian models for multiple local sequence alignment and Gibbs sampling strategies. *JASA 90*:1156–1170.

Liu JS, Neuwald A, Lawrence CE. 1998. Markovian structures in biological sequence alignments. *JASA* Forthcoming.

Martin DL, Barke K. 1997. Are GAD_{65} and GAD_{67} associated with specific pools of GABA in brain? *Perspect Dev Biol*. Forthcoming.

Martin DL, Martin SB, Wu SJ, Espina N. 1991. Regulatory properties of brain glutamate decarboxylase (GAD): The apoenzyme of GAD is present principally as the smaller of two molecular forms of GAD in brain. *J Neurosci 11*:2725–2731.

Martin DL, Rimvall K. 1993. Regulation of γ-aminobutyric acid synthesis in the brain. *J Neurochem 60*:395–407.

McPhalen C, Vincent MG, Jansonius JN. 1992a. X-ray structure refinement and comparison of three forms of mitochondrial aspartate aminotransferase. *J Mol Biol 225*:495–517.

McPhalen CA, Vincent MG, Picot D, Jansonius JN, Lesk AM, Chothia C. 1992b. Domain closure in mitochondrial aspartate aminotransferase. *J Mol Biol 227*:197–213.

Mehta PK, Hale TI, Christen P. 1993. Aminotransferases: Demonstration of homology and division into evolutionary subgroups. *Eur J Biochem 214*:549–561.

Momany C, Ernst S, Ghosh R, Chang NL, Hackert ML. 1995b. Crystallographic structure of a PLP-dependent ornithine dearboxylase from *Lactobacillus* 30a to 3.0 Å resolution. *J Mol Biol 252*:643–655.

Momany C, Ghosh R, Hackert ML. 1995a. Structural motifs for pyridoxal-5′-phosphate binding in decarboxylases: An analysis based on the crystal structure of the *Lactobacillus* 30a ornithine decarboxylase. *Protein Sci 4*:849–854.

Namchuk M, Lindsay L, Turck CW, Kanaani J, Baekkeskov S. 1997. Phosphorylation of serine residues 3, 6, 10, and 13 distinguishes membrane anchored from soluble glutamic acid decarboxylase 65 and is restricted to glutamic acid decarboxylase 65alpha. *J Biol Chem 272*:1548–1557.

Neuwald A, Liu JS, Lawrence CE. 1995. Gibbs motif sampling: Detection of bacterial outer membrane protein repeats. *Protein Sci 4*:1618–1632.

Neuwald A, Liu JS, Lipman DJ, Lawrence CE. 1997. Extracting protein alignment models from the sequence database. *Nucleic Acids Res 25*:1665–1677.

Pascarella S, Schirch V, Bossa F. 1993. Similarity between serine hydroxymethyltransferase and other pyridoxal phosphate-dependent enzyme. *FEBS 331*:145–149.

Picot D, Sandmeier E, Thaller C, Vincent MG, Christen P, Jansonius JN. 1991. The open/closed conformational equilibrium of aspartate aminotransferase studies in the crystalline state and with a fluorescent probe in solution. *Eur J Biochem 196*:329–341.

Porter TG, Martin DL. 1988. Stability and activation of glutamate apodecarboxylase from pig brain. *J Neurochem 51*:1886–1891.

Porter TG, Spink DC, Martin SB, Martin DL. 1985. Transaminations catalysed by brain glutamate decarboxylase. *Biochem J 231*:705–712.

Schneider TD, Stephens RM. 1990. Sequence logos: A new way to display consensus sequences. *Nucleic Acids Res 18*:6097–6100.

Sheikh SN, Martin DL. 1996. Heteromers of glutamate decarboxylase isoforms occur in rat cerebellum. *J Neurochem 66*:2082–2090.

Solimena M, Aggujaro D, Muntzel C, Dirkx R, Butler M, DeCamilli P, Hayday A. 1993. Association of GAD-65, but not of GAD-67, with the Golgi complex of transfected Chinese hamster ovary cells mediated by the N-terminal region. *Proc Natl Acad Sci USA 90*:3073–3077.

Solimena M, Dirkx R Jr, Radzynski M, Mundigl O, De Camilli P. 1994. A signal located within amino acids 1–27 of GAD65 is required for its targeting to the Golgi complex region. *J Cell Biol 126*:331–341.

Stark W, Kallen J, Markovic-Housley Z, Fol B, Kania M, Jansonius JN. 1991. The three-dimensional structure of phosphoserine aminotransferase from *Escherichia coli*. In: Fukui T, Kagamiyama H, Soda K, Wada H, eds. *Enzymes dependent on pyridoxal phosphate and other carbonyl compounds as cofactors*. New York: Pergamon Press.

Toney MD, Hohenester E, Keller JW, Jansonius JN. 1995. Structural and mechanistic analysis of two refined crystal structures of the pyridoxal phosphate-dependent enzyme dialkylglycine decarboxylase. *J Mol Biol 245*:151–179.

Watanabe N, Yonaha K, Sakabe K, Sakabe N, Aibara S, Morita Y. 1991. Crystal structure of ω-amino acid: Pyruvate aminotransferase. In: Fukui T, Kagamiyama H, Soda K, Wada H., eds. *Enzymes dependent on pyridoxal phosphate and other carbonyl compounds as cofactors*. New York: Pergamon Press.

Xlook Version 2.0. 1996. Product of Molecular Application Group, Palo Alto.

Zhu J, Liu JS, Lawrence CE. 1997. Bayesian alignment and inference. *ISMB-97 5*:358–368.