

Similarity between serine hydroxymethyltransferase and other pyridoxal phosphate-dependent enzymes

Stefano Pascarella^a, Verne Schirch^b, Francesco Bossa^{a,*}

^a*Dipartimento di Scienze Biochimiche A Rossi Fanelli, Centro di Biologia, Molecolare del Consiglio Nazionale delle Ricerche, Università La Sapienza, 00185 Roma, Italy*

^b*Department of Biochemistry and Molecular Biophysics, Virginia Commonwealth University, Richmond, VA 23298, USA*

Received 29 July 1993

A structural homology of the pyridoxal-5'-phosphate (PLP)-dependent enzyme serine hydroxymethyltransferase (SHMT) with aspartate aminotransferase (AAT) is proposed. Although the two sequences are very dissimilar, a reasonable alignment was obtained using the profile analysis method. Sequences of AAT and dialkylglycine decarboxylase (DGD), for which crystal structure data are available, have been aligned on the basis of their structure superposition. A profile was then calculated and SHMT sequence aligned to it. Three of the four residues conserved in all aminotransferases (including the PLP-binding lysine) are matched. A profile search with DGD-AAT-SHMT profile is more selective and sensitive than individual sequence profiles for PLP-dependent enzyme detection. Potential homologies with the *eryCI* gene product involved in erythromycin biosynthesis and with amino acid decarboxylases were observed. Homology with AAT will be used as a guideline for planning site-directed mutagenesis experiments on SHMT.

Profile analysis; Site-directed mutagenesis; Pyridoxal-phosphate; Aminotransferase; Serine hydroxymethyltransferase; Dialkylglycine decarboxylase; Structure superposition

1. INTRODUCTION

Pyridoxal-5'-phosphate (PLP) is a versatile cofactor able to catalyze a spectrum of reactions on a variety of amino acid substrates in different structural protein contexts. Two broad groups include those enzymes which break a bond only on the α -carbon of the amino acid substrates (transamination, decarboxylation, β -elimination, or racemization) and enzymes which catalyze elimination or replacement reactions on either β - or γ -carbons. Accordingly, each PLP-dependent enzyme is assigned to a class (transaminase, decarboxylase, etc.) in one of these two groups. Many sequences from these enzymes are known, while only a few spatial conformations are deposited in the Protein Data Bank (PDB) [1]. The most well characterized are the aspartate aminotransferases (AAT) from various sources [2–5] and tryptophan synthase from *Salmonella typhimurium* [6].

Previous observations that several PLP enzymes shared catalytically crucial stereochemical features argued for their evolutionary relationship [7,8]. However, sequence similarity among different classes of PLP en-

zymes is generally weak and the 3-D structures of aspartate aminotransferase and tryptophan synthase showed them not to have a common folding pattern. This has weakened the argument that PLP enzymes are evolutionarily related and share similar structural motifs. Also, standard methods generally fail to find significant alignments among the different classes of PLP-dependent enzymes. In particular, these should match at least some of the residues proven to be critical for activity and conserved in all AATs [9–12], such as glycine-197 (G197 according to the AAT numbering system adopted throughout the paper), which participates in a turn located at domain interface, aspartate-222 (D222), which H-bonds to N1 of PLP, lysine-258 (K258), which forms a Schiff base with PLP, and arginine-386 (R386), which H-bonds with α -carboxylate group of substrate. However, lack of evident sequence similarity between proteins does not necessarily exclude similarity of backbone folding and evolutionary relationship [13]. The most rigorous and reliable way of aligning two sequences is to assign structural equivalences between residues by structure superposition [14,15]. In fact, a functional relationship between fractional sequence similarity of two proteins and root mean square deviation of α -carbons upon optimal superposition of their structures has been proposed by Chothia and Lesk [16]. Recently, it has been shown that dialkylglycine decarboxylase (DGD) from *Pseudomonas cepacia* shares the spatial fold with aspartate aminotransferase [17,18] despite

*Corresponding author. Fax: (39) (6) 444 0062.

Abbreviations: PLP, pyridoxal-5'-phosphate; AAT, aspartate aminotransferases; mCAAT, mitochondrial chicken AAT; DGD, dialkylglycine decarboxylase; SHMT, serine hydroxymethyltransferase; PDB, Protein Data Bank.

their poor sequence identity [10,19]. This observation has again opened the question whether other classes of PLP enzymes may be related to aminotransferases even though they share little sequence similarity.

Our group is interested in the elucidation of the reaction mechanism of serine hydroxymethyltransferase (SHMT), a PLP enzyme involved in one-carbon metabolism and belonging to the group of PLP enzymes which break one of the bonds on the α -carbon of the amino acid substrate [20,21]. Besides its physiological reaction, i.e. the conversion of serine and tetrahydrofolate to glycine and 5,10-methyltetrahydrofolate, SHMT catalyzes to some extent all of the other reactions in this group, such as transamination, decarboxylation, and racemization. Also, it was one of the first enzymes to have its stereochemical properties related to aspartate aminotransferase [22]. The availability of the *glyA* gene coding for SHMT in *E. coli* opens the possibility to investigate the functional role of supposed critical residues via specific mutations. Unfortunately, although 11 sequences of SHMT from various sources are currently known [23-32], no three-dimensional structure has been solved. The current lack of this structural information does not allow a rationally oriented application of site-directed mutagenesis. This report is aimed at detecting sequence similarities between SHMT and other PLP-dependent enzymes, in particular those with available three-dimensional structures and detailed functional data. The results suggest that SHMT has the same folding pattern as AAT and DGD.

2. MATERIALS AND METHODS

Protein sequences and structures have been taken from the SWISS-PROT 24.0 [33] and PDB [1] data banks, respectively. Sequence alignment and data bank search programs were from the Genetic Computer Group sequence analysis software package (GCG) [34], running on a VAX/VMS system. In particular, the profile method [35] as implemented in the GCG routines PROFILEMAKE, PROFILESEARCH, and PROFILEGAP have been used. Secondary structures were assigned to the atomic coordinate set of mitochondrial chicken AAT (mCAAT; PDB code 7AAT) with the program DSSP

[36]. Three-dimensional structures were displayed and analyzed on an INDIGO Silicon Graphics station with the program MIDAS [37]. Secondary structure predictions were calculated with the program PRONET [38] implementing a prediction method based on a neural network [39]. Structure superpositions between DGD and mCAAT fragments were supplied by Michael D. Toney.

3. RESULTS AND DISCUSSION

The AATs are one of the best structurally and functionally characterized groups of PLP-dependent enzymes [3,40]. It is therefore quite natural to consider AAT structures as reference to which one can try to match SHMT sequences. However, standard methods fail to find a satisfactory alignment between SHMT and mCAAT sequences, being able to match only residues D222 and R386, but not residues K258 and G197 as otherwise expected and desirable. On the other hand, sequence alignment between SHMT and DGD obtained with the GCG routine GAP (Fig. 1) matches among other residues also a glycine, an aspartate, a lysine, and an arginine shown in DGD to be structurally equivalent to the mCAAT G197, D222, K258, and R386, respectively [17,18]. This suggested that DGD sequence could serve as a bridge between SHMT and mCAAT for obtaining a credible sequence alignment. Unfortunately, even though mCAAT and DGD share a similar folding, standard sequence alignment methods failed to find a reasonable alignment between these two PLP-enzymes.

On the basis of structure superposition of α -carbons, it was possible to align sequence fragments from DGD and mCAAT accounting for a total of 198 residues and 27 identities (M.D. Toney, personal communication). In order to use this information effectively, we extended as far as possible this structure superposition-based alignment via sequence alignment methods. A profile was calculated from the DGD sequence with the matched mCAAT fragments. The entire mCAAT sequence was then realigned to this profile with the GCG routine PROFILEGAP. The matched mCAAT fragments forced the program to conform the sequence

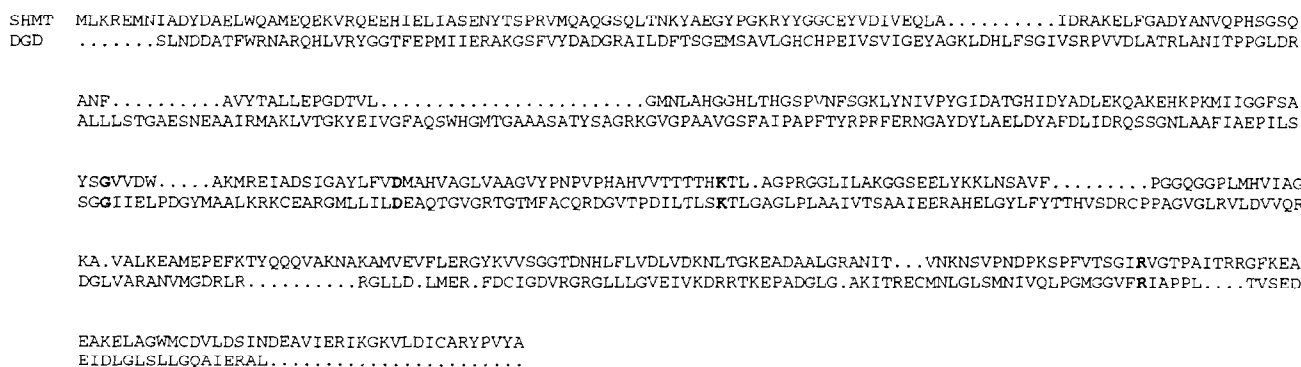


Fig. 1. Alignment between dialkylglycine decarboxylase (DGD) and *E. coli* serine hydroxymethyltransferase (SHMT) sequences obtained with the routine GAP run with standard gap penalties. Residues equivalent to G197, D222, K258 and R386 are boldfaced. Dots represent insertions/deletions introduced by the aligning routine

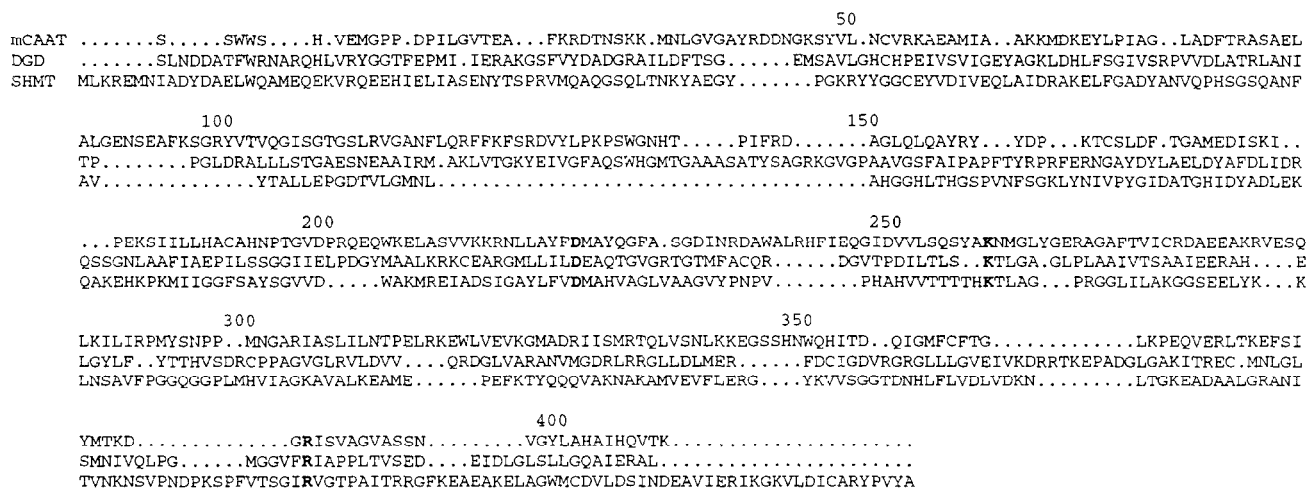


Fig. 2. Alignment among the sequences of dialkylglycine decarboxylase (DGD), *E. coli* serine hydroxymethyltransferase (SHMT) and mitochondrial chicken aspartate aminotransferase (mCAAT), obtained with the routine PROFILEGAP aligning the SHMT sequence to the profile calculated from the DGD-mCAAT alignment which incorporated constraints from structure superposition. Residues conserved in all aminotransferases are boldfaced [9,12]. Sequence positions are numbered according to the AAT system. Dots represent insertions/deletions introduced by the alignment program.

alignment to the structure superposition. The alignment was calculated with gap opening penalties 4.5 and 6.0, and 0.05 for gap elongation penalty. Only minor differences were seen between the two alternatives, and the alignment at 6.0, bearing the least number of gaps, was used. This procedure produced an alignment matching 358 residues with 66 identities. A profile was calculated from this alignment and the *E. coli* SHMT sequence aligned to it with PROFILEGAP. Profiles were calculated using alternatively sequence weights of 1.0 and 4.0 for the DGD, and always 1.0 for mCAAT sequences. Alignments of SHMT to each of these profiles used alternatively gap opening penalties 3.0, 4.5 and 6.0, and 0.05 gap elongation penalty. In Fig. 2, the final alignment among the three sequences obtained at gap opening penalty 6.0 and sequence weights 1.0 is reported. Resulting alignments differed markedly in the N-terminal part, while the central and C-terminal regions had only marginal differences (from position 150). Even though sequence similarity is very low (only 43 invariant residues), some indications would support the hypothesis of structural homology between SHMT and AAT and would justify the assumption of some of the AAT reaction framework features as being shared by SHMT. In fact, three out of the four residues constantly conserved in aminotransferases [10,12] are conserved in all sequences of the proposed alignment: D222, K258 and R386. The residue G197 is not strictly conserved, but it is matched both in DGD and SHMT with a sequence segment containing glycine and other residues compatible with a turn. These three residues and the turn compatible segment are conserved also in all SHMT sequence determined so far [23-32]. Moreover, the same procedure does not align the sequence of tryptophan synthase β subunit, which has a folding different

from that of the AAT family, being unable to match the PLP-binding lysine and the other residues found invariant in the synthase family [41]. Secondary structure predictions were calculated for SHMT and compared with those assigned to mCAAT [36]: a general agreement in the central and C-terminal part (from position 150 to the end; Fig. 2) of the alignment was seen, with 66% of aligned residues predicted to share the same conformation. This region includes mCAAT helices 7, 8, and 9 (positions 170-179, 202-215, 233-246, respectively), which strongly interact to form the foundation of the large domain, helices 14, 15, and 16, which constitutes most of the small domain, and helix 13 (313-344), which connects the two domains. This substructure may represent the most conserved core shared by SHMT and AATs.

To assess the reliability of the proposed DGD-mCAAT-SHMT alignment, a profile was calculated and the SWISSPROT data bank searched with it, using default gap penalties and restricting the search to sequences longer than 100 amino acid residues (with a consequent reduction from 26,000 to 23,000 sequences). In order to evaluate profile sensitivity and selectivity, the results were compared with those of profiles built from the individual sequences and from the DGD-mCAAT alignment. As an evaluation criterion, the following parameters for the different profiles were compared: (a) the scores assigned to some gene-deduced sequences, recently shown to correspond to PLP-dependent enzymes distantly related to aminotransferases [9,11,42]; *cobC* (involved in cobalamine synthesis [43]), *malY* (abolishing endogenous induction of the maltose system [44]), and *cefd* (coding for isopenicillin-*N*-epimerase [45]); (b) the number of consecutive PLP-dependent enzymes in top scoring region of the PROFI-

LESEARCH output. Results of this comparison (Table I) clearly indicate the superior performance of the DGD-mCAAT-SHMT profile. Noteworthy, this profile ranked in position 46 (with score 4.27) the *eryCI* gene product (SWISSPROT code ERBS_SACER) involved in the erythromycin biosynthesis [46]. This entry is followed by other four PLP-enzymes. Although the precise function of *eryCI* gene product is unknown, on the basis of genetic evidence it was suggested that it is involved either as a structural or as a regulatory factor in the synthesis of the deoxyaminosugar desosamine or in its attachment to the macrolide ring. Upon PROFILEGAP alignment, residues D222 and K258 are matched. This would indicate the *eryCI* gene product as a PLP-dependent enzyme, possibly a transaminase, involved in the synthesis of desosamine. Aminotransferases belonging to the fourth subgroup [9,12] appear to be more distantly related (for example serine-pyruvate aminotransferase occurs at *Z* score 3.14). Interestingly, *E. coli* lysine decarboxylase occurs at 2.28. Relation of SHMT structure to that of other PLP enzymes is further supported by the PROFILESEARCH run with the profile calculated from the alignment of eleven SHMT sequences (human and rabbit mitochondrial and cytosolic isoenzymes [23–25], *E. coli* [26], *N. crassa* [27], *S. typhimurium* [28], *C. jejuni* [29], *B. japonicum* [30], *P. sativum* mitochondrial isoenzyme [31], and *H. methylovorum* [32]) obtained with the GCG routine PILEUP. The highest scoring (5.23) non-SHMT sequence is *E. coli* 2-amino-3-ketobutyrate coenzyme A ligase, a PLP-dependent enzyme related to histidinolphosphate aminotransferase, which belongs to the first aminotransferase subgroup [9,12]. Interestingly, *eryCI* gene product is reported at 3.15, while DGD and mitochondrial horse AAT at 2.73 and 2.61, respectively, in the noise region. A weak sequence similarity between the regions encompassing the PLP-binding lysine of decarboxylases and other PLP-dependent enzymes, in particular SHMT, has already been observed ([47] and P. Christen, personal communication). This observation guided us in planning the replacement of residues D222, H225, T253, T254, T255, T256, H257, K258, T230, and R235 (numbering is that of Fig. 2) in *E. coli* SHMT by

site-directed mutagenesis, in order to assess their importance in the enzyme mechanism. Some of the pertinent experiments were already performed with some success [21,48] or are presently underway. The proposed, more rigorous, alignment between SHMT and mCAAT strongly suggests for at least two residues (those equivalent to D222 and R386 in AAT) a relevant role also in SHMT, to be confirmed by appropriate site-directed mutagenesis experiments. In fact, these residues are conserved in all aminotransferases [9,12] and very likely in many other PLP-dependent enzymes (P. Christen, personal communication). The future availability of other three-dimensional structures will provide more precise alignments based on structure superposition and, consequently, more selective and sensitive profiles, which will help in testing the existence of a PLP-enzyme common fold and in the identification of its folding determinants. Careful superposition of DGD and mCAAT 3-D structures will modify and further improve the proposed alignments and hopefully will indicate other SHMT potentially important residues as targets for site-directed mutagenesis studies.

Acknowledgements· We are grateful to 'progetto VAXRMA' for computer support and assistance at the University of Rome La Sapienza. Services provided by the Italian EMBnet node in Bari were essential. We are indebted to Dr. Michael D. Toney and Prof. Johan N. Jansonius for making available structure superposition-based alignments between DGD and mCAAT sequence fragments prior to publication. This work was supported in part by a CNR grant in the framework of 'Progetto speciale bioinformatica' and by a grant from the Commission of the European Communities (Human Capital and Mobility, contract no. ERBCHRXCT930179).

REFERENCES

- [1] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice Jr., M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) *J. Mol. Biol.* 112, 535–542
- [2] Graf-Hausner, U., Wilson, K.J. and Christen, P. (1983) *J. Biol. Chem.* 258, 8813–8826.
- [3] Kirsch, J.F., Eichele, G., Ford, G.C., Vincent, M.G., Jansonius, J.N., Gehring, H. and Christen, P. (1984) *J. Mol. Biol.* 174, 497–525.
- [4] McPhalen, C.A., Vincent, M.G., Picot, D. and Jansonius, J.N. (1992) *J. Mol. Biol.* 227, 197–213.

Table I
Comparison of profile sensitivity and selectivity

Probe profiles	Scores assigned to target sequences ^a			Number of consecutive PLP-dependent enzymes in the profile top scoring region
	<i>cobC</i> (COBC_PSEDE)	<i>malY</i> (MALY_ECOLI)	<i>cefd</i> (CEFD_STRCL)	
DGD	< 2.35	< 2.35	< 2.35	10
SHMT	< 2.35	< 2.35	< 2.35	8
mCAAT	2.60	< 2.33	< 2.33	19
DGD-mCAAT	3.83	< 2.24	2.50	19
DGD-mCAAT-SHMT	5.48	3.64	4.07	45

^a SWISSPROT codes are given in parentheses.

- [5] McPhalen, C.A., Vincent, M.G. and Jansonius, J.N. (1992) *J. Mol. Biol.* 225, 495-517.
- [6] Hyde, C.C., Ahmed, S.A., Padlan, E.A., Miles, E.W. and Davies, D.R. (1988) *J. Biol. Chem.* 263, 17857-17871.
- [7] Dunathan, H.C. (1966) *Proc. Natl. Acad. Sci. USA* 55, 712-716.
- [8] Smith, D.M., Thomas, N.R. and Gani, D. (1991) *Experientia* 47, 1104-1118.
- [9] Mehta, P.K., Hale, T.I. and Christen, P. (1989) *Eur. J. Biochem.* 186, 249-253.
- [10] Mehta, P.K., Hale, T.I. and Christen, P. (1991) in: *Enzymes Dependent on Pyridoxal Phosphate and other Carbonyl Compounds as Cofactors* (Fukui, T., Kagamiyama, H., Soda, K. and Wada, H. eds.) pp. 35-42, Pergamon Press, Oxford.
- [11] Mehta, P.K. and Christen, P. (1993) *Eur. J. Biochem.* 211, 373-376.
- [12] Mehta, P.K., Hale, T.I. and Christen, P. (1993) *Eur. J. Biochem.* 214, 549-561.
- [13] Bashford, D., Chothia, C. and Lesk A.M. (1987) *J. Mol. Biol.* 196, 199-216.
- [14] Rossmann, M.G. and Argos, P. (1975) *J. Biol. Chem.* 250, 7525-7532.
- [15] Taylor, W.R. and Orengo, C.A. (1989) *J. Mol. Biol.* 298, 1-22.
- [16] Chothia, C. and Lesk, A.M. (1986) *EMBO J.* 5, 823-826.
- [17] Toney, M.D., Keller, J.W., Pauptit, R.A., Jaeger, J., Wise, M.K., Sauder, U. and Jansonius, J.N. (1991) *J. Mol. Biol.* 298, 873-875
- [18] Toney, M.D., Hohenester, E., Cowan, S.W. and Jansonius, J.N. (1993) *Science* (in press).
- [19] Keller, J.W., Baurick, K.B., Rutt, G.C., O'Malley, M.V., Sonfrank, N.L., Reynolds, R.A., Ebbesson, L.O.E. and Vajdos, F.F. (1990) *J. Biol. Chem.* 265, 5531-5539.
- [20] Schirch, V. (1982) *Adv. Enzymol.* 53, 83-112.
- [21] Angelaccio, A., Pascarella, S., Fattori, E., Bossa, F., Strong, W. and Schirch, V. (1992) *Biochemistry* 31, 155-162.
- [22] Voet, J.G., Hindenlang, D.M., Blanck, J.J., Ulevitch R.J. and Kallen, R.G. (1973) *J. Biol. Chem.* 248, 841-842.
- [23] Garrow, T., Brenner, A.A., Whitehead, V.M., Chen, X.-N., Duncan, R.G., Korenberg, J.R. and Shane, B. (1993) *J. Biol. Chem.* 268, 11910-11916.
- [24] Martini, F., Angelaccio, S., Pascarella, S., Barra, D., Bossa, F. and Schirch, V. (1987) *J. Biol. Chem.* 262, 5499-5509.
- [25] Martini, F., Maras, B., Tanci, P., Angelaccio, S., Pascarella, S., Barra, D., Bossa F. and Schirch, V. (1989) *J. Biol. Chem.* 264, 8509-8519.
- [26] Plamann, M., Stauffer, L., Urbanowski, M. and Stauffer, G. (1983) *Nucleic Acids Res.* 11, 2065-2075.
- [27] McLung, C.R., Davis, C.R., Page, K.M. and Denome, S.A. (1992) *Mol. Cell. Biol.* 12, 1412-1421.
- [28] Steiert, J.G., Urbanowski, M.L., Stauffer, L.T., Plamann, M.D. and Stauffer, G.V. (1990) *DNA seq.*, 1, 107-113.
- [29] Chan, V.L. and Bingham, H.L. (1991) *Gene* 101, 51-58.
- [30] Rossbach, S. and Hennecke, H. (1991) *Mol. Microbiol.* 5, 39-47.
- [31] Turner, S.R., Ireland, R., Morgan, C. and Rawsthorne, S. (1992) *J. Biol. Chem.* 267, 13528-13534.
- [32] Miyata, A., Yoshida, T., Yamaguchi, K., Yokoyama, C., Tanabe, T., Toh, H., Mitsunaga, T. and Izumi, Y. (1993) *Eur. J. Biochem.* 212, 745-750.
- [33] Baroch, A. and Boeckmann, B. (1991) *Nucleic Acids Res.* 19, 2247-2249.
- [34] Deveraux, J., Haeberli, P. and Smithies, O. (1984) *Nucleic Acids Res.* 12, 387-395.
- [35] Gribksov, M., Luthy, R. and Eisenberg, D. (1990) *Methods Enzymol.* 183, 146-159.
- [36] Kabsch, W. and Sander, C. (1983) *Biopolymers* 22, 2577-2637.
- [37] Ferrin, T.E., Huang, C.C., Jarvis, L.E. and Landgridge, R. (1988) *J. Mol. Graphics* 6, 13-27.
- [38] Pascarella, S. and Bossa, F. (1989) *Comp. Appl. Biosci.* 5, 319-320.
- [39] Holley, H.L. and Karplus, M. (1989) *Proc. Natl. Acad. Sci. USA* 86, 152-156.
- [40] Christen, P. and Metzler, D.E. (1985) *Transaminases*, John Wiley and Sons, New York.
- [41] Bork, P. and Rohde, K. (1990) *Biochem. Biophys. Res. Commun.* 171, 1319-1325.
- [42] Ouzounis, C. and Sander, C. (1993) *FEBS Lett.*, 322, 159-164.
- [43] Crouzet, J., Cauchois, L., Blanche, F., Debussche, L., Thibaut, D., Rouyez, M.-C., Rigault, S., Mayaux, J.-F. and Cameron, B. (1990) *J. Bacteriol.* 172, 5968-5979.
- [44] Reidl, J. and Boos, W. (1991) *J. Bacteriol.* 173, 4862-4876.
- [45] Usui, S. and Yu, C.-A. (1989) *Biochim. Biophys. Acta* 999, 78-85.
- [46] Dhillon, N., Hale, R.S., Cortes, J. and Leadlay, P.F. (1989) *Mol. Microbiol.* 3, 1405-1414.
- [47] Maras, B., Sweeney, G., Barra, D., Bossa, F. and John, R.A. (1992) *Eur. J. Biochem.* 204, 93-98.
- [48] Hopkins, S. and Schirch, V. (1986) *J. Biol. Chem.* 261, 3363-3369.