The Chemistry of the Reaction Determines the Invariant Amino Acids

During the Evolution and Divergence of

Orotidine 5'-monophosphate Decarboxylase*

Thomas W. Traut and Brenda R. S. Temple

Department of Biochemistry and Biophysics,

University of North Carolina School of Medicine,
Chapel Hill, NC 27599-7260, USA
Abstract  Orotidine 5’-phosphate decarboxylase has the largest rate enhancement for any known enzyme. For an average protein of 270 amino acids from more than 80 species, only eight amino acids are invariant, and seven of these correspond to ligand-binding residues in the crystal structures of the enzyme from four species. It appears that the chemistry required for catalysis determines the invariant residues for this enzyme structure. A motif of three invariant amino acids at the catalytic site (DXKXXD) is also found in the enzyme hexulose-phosphate synthase. Although the core of OMP decarboxylase is conserved, it has undergone a variety of changes in subunit size or fusion to other protein domains, such as orotate phosphoribosyltransferase, during evolution in different kingdoms. The phylogeny of OMP decarboxylase shows a unique subgroup distinct from the three kingdoms of life. The enzyme subunit size almost doubles from Archaea (average Mr of 24.5 kDa) to certain fungi (average Mr of 41.7 kDa). These observed changes in subunit size are produced by insertions at 12 sites, largely in loops and on the exterior of the core protein. The consensus for all sequences has a minimal size of <20 kDa.

Key words: Molecular evolution; chemistry determines invariance; OMP decarboxylase; hexulose-phosphate synthase
Orotidine 5'-monophosphate (OMP) decarboxylase (EC 4.1.1.23) catalyzes the synthesis of UMP in the final reaction of the \textit{de novo} pathway for the biosynthesis of pyrimidine nucleotides (Scheme 1). Since all other pyrimidine nucleotides can be derived from UMP, this enzyme appears essential to all organisms (1), and it has now been sequenced from over 80 species. We present an analysis of this diverse set of sequences that demonstrates that even while the subunit size has almost doubled in some species, a core set of important amino acids has remained invariant or highly conserved.

This apparent structural flexibility is of interest since this enzyme has the currently greatest observed rate enhancement, defined as $k_{\text{cat}}/k_{\text{non}}$, by a factor of $10^{17}$ (2). The remarkable catalytic proficiency of OMP decarboxylase results from the exceptional stability of the carboxyl group of OMP, since the uncatalyzed decarboxylation of OMP has a $t_{1/2}$ of 78 million years. This then makes it more significant that this enzyme has no cofactors (3), and makes interpretation of the catalytic mechanism a challenge.

Four high quality crystal structures of OMP decarboxylase were recently produced: one from Archaea (\textit{M. thermoautotrophicum}; (8)), two from the Eubacteria \textit{B. subtilis} (9) and \textit{E. coli} (10), and one from a eukaryote, \textit{S. cerevisiae} (11). These enzymes were bound to the product UMP (9) or to tight binding transition state analogs (8,10,11). This enables a comparison of the amino acid residues in these structures shown to be necessary for ligand binding with the amino acids that are found to be invariant or highly conserved in all available sequences. Consistent with the demands dictated by the chemistry for this proficient reaction, the amino acid residues that have been
maintained as invariant across all kingdoms of life are almost completely those that are
essential for catalysis.

Furthermore, the recent abundance of OMP decarboxylase sequence data has made
it practical to define the phylogeny of this essential enzyme in comparison to earlier
phylogenetic analyses. Work by Woese and colleagues led to the widely used ribosomal
RNA sequences as a reference standard (4,5), since this molecule is an essential and
presumably original component of the oldest organisms. Efforts have also been made to
use proteins for this purpose. Since no single protein had been sequenced in enough
organisms, Doolittle et al. (6) chose to use 57 different proteins, for each of which
sequences existed from at least four species. Their results show that phylogenetic
relationships are not too dissimilar from relationships based on RNA, but their time
scale for important divergence points was significantly different than for trees based on
RNA studies.

An element of uncertainty has entered the recently developed paradigm for a tree
of life with the findings that genes may be transferred laterally between species.
Comparison of the complete genomes of Methanoccocus jannashii, Escherichia coli,
Synechocystis 6803, and Saccharomyces cerevisiae suggested that lateral transfer may have
been on a large scale (7). The authors detected a pattern for two classes of genes:
informational (maintenance and expression of DNA, and signaling), and operational
(enzymes of general metabolism). For yeast, informational genes were most closely
related to Methanoccocus jannashii (Archaea), while operational genes were closer to
Escherichia coli (Bacteria).

If transfer of genes between very different species is at all extensive, then no single
gene (molecule) may serve as a unique reference standard to define any tree of life. It
may then be necessary to have more such reference standards. Proteins that have an
essential function, such as OMP decarboxylase, are good candidates for this role. The
enzyme activity has been widely detected (1), and this implies that new sequences will continue to become available to help phylogenetic analyses.
EXPERIMENTAL PROCEDURES

Sequence Analysis: The amino acid sequences for all defined OMP decarboxylases were obtained from standard databases. A majority of sequences is available in Swiss-Prot; a few sequences were found by BLAST searches of Genbank, and then translated. These latter sequences could be verified as true OMP decarboxylase sequences by the location of a critical motif, DxKxxDlxxT, where the capital letters represent invariant residues, ‘x’ is any residue, and the two italic letters are almost invariant - being present in all but two or three of 82 sequences (see Fig. 1).

Sequences were initially aligned with the program Pileup of the University of Wisconsin Computer Group (GCG9). Since nine of these sequences are contained in a bifunctional UMP synthase, the OMP decarboxylase portion was identified by comparison to the consensus for all the monofunctional OMP decarboxylase sequences. For phylogenetic analyses a final alignment was obtained with the program CLUSTAL X (12,13). An evolutionary tree was constructed using the PHYLIP fitch program (14) from evolutionary distances calculated with the PHYLIP protdist program.
RESULTS

Global Sequence Alignment and Identification of Signature Sequences

Figure 1 presents 13 representative sequences selected from an alignment of 82 OMP decarboxylase plus eleven hexulose-phosphate synthase sequences. The first eleven sequences are OMP decarboxylases. Because most of the signature motif (DxKxxDlxxT; evident at amino acid residues 145 - 154) was also located in sequences for hexulose-phosphate synthase, two sequences for this additional enzyme complete the listing in Figure 1. An effort was made to use sequences representative of the taxonomic subgroups, while also including the range for subunit size and diversity. This alignment also illustrates that there are 11 consensus sequence segments, which define the secondary structure elements of the \( \alpha/\beta \) barrel core structure (8-11). Inserts of varying sizes occur throughout the sequence, and always at positions where there are loops in the protein structures.

In the key signature sequence at residues 145 - 154, the D, K, and D are absolutely invariant, suggesting that they were likely to be involved in the catalytic mechanism. Also invariant are an aspartate at position 59, a lysine at position 110, a glycine at 399, a glutamine at 422, and an arginine at 446. The position numbers are based on the alignment in Figure 1, but all the individual enzymes are actually smaller, as shown next to their species name. Since alignment programs allow some gaps to optimize the alignment, the result shown in Figure 1 was visually adjusted to have Glutamine422 become invariant, aided by the information from the four crystal structures. These are yeast OMP decarboxylase bound to barbiturate monophosphate (BMP) (11), the \textit{Bacillus subtilis} enzyme bound to UMP (9), and the enzymes from \textit{Methanobacterium thermoautotrophicum} and \textit{Escherichia coli} bound to 6-azauridine monophosphate (8,10).

All four OMP decarboxylase proteins show an \( \alpha/\beta \) barrel core structure, with 8 central \( \beta \) strands surrounded by 9 helices (8-11). In each of these structures 9 to 11
amino acids make important hydrogen bonds to the ligand at the catalytic site; the nine consensus residues are indicated in Fig. 1 by symbols for hydrogen bonding (closed or open circles). Up to six other amino acids bind to a few of these amino acids that bind the nucleotide ligand directly, and help to stabilize them. Only one of these is at a consensus position (Asp145). Thus, of these ten consensus residues shown to be necessary in the four crystal structures, seven of them are at an invariant position in Fig. 1 (shown in bold), and the other three are at very conserved residues (in italic). The invariant amino acids tend to be at the ends of β strands, or in loops. Only Thr154 and Arg446 are in a helix. These results show the importance of the overall core α/β barrel structure, as well as the few essential amino acids therein.

While nine of the ten residues shown by the four crystal structures to participate in binding were predicted in the sequence alignment, each of the structures also shows one or more additional amino acids that have a secondary role at the catalytic site, but these are not at the same position in the sequence alignment. This limited variety may be due to actual variations in the separate structures, and/or the somewhat different nucleotide ligands being bound.

The usefulness and limits of such an alignment analysis become apparent. Figure 2 depicts the yeast protein structure, and the specific residues involved in binding to the substrate analog BMP (11). Of seven amino acids whose side chains bind directly to BMP, five are invariant for more than 80 different species (Asp59, Lys147, Asp150, Gln422, Arg446) and one is highly conserved (Thr154). But, while five additional amino acids make side chain contacts to stabilize other amino acids that bind BMP, only two of these are invariant (Lys110, Asp145) while the other three are not even conserved (#57, #176, #311).

Of the many loops in the crystal structures, three are specifically identified in Fig. 1, and are known to participate at the catalytic site. OMP decarboxylases in the four protein structures, as well as in man, or mouse are normally dimeric (1). An important
structural feature that could not be anticipated from the sequence alignment is that the catalytic site is formed by segments of the two adjoining subunits in the four enzyme structures. Thus, Loop A on one subunit containing Asp150 - Thr154 connects across the dimer interface to complete the catalytic site made by residues Asp145 and Lys147 plus other amino acids shown in Figure 3. Therefore, one half of the signature motif (Asp145 X Lys147 X X Asp150 Ile151 X X Thr154) contributes to the catalytic site of one subunit in the functional dimer, while the second half of this motif contributes to the catalytic site of the adjoining subunit. Given the close spacing of these key residues in the primary sequence, it was not anticipated that they could contribute to separate catalytic sites on different subunits. However, this significant result for the active site architecture corroborates earlier kinetic studies which showed that only the dimer form of OMP decarboxylase was catalytically competent (15). Thus the recently obtained crystal structures (8-11) are entirely consistent with these earlier kinetic studies.

Two loops (Loops B and C, Fig. 4) tend to be poorly defined in the apoenzyme, but clearly close over the catalytic pocket when a ligand is bound, and each loop contains a residue needed for binding (Fig. 1). Loops A and B appear to be constant in size, while Loop C is clearly variable, as defined in Table I. It is evident that this loop has a minimal size in Archaea, but is much larger in other species. Since the structure from M. thermoautotrophicum (Archaea) is the only structure showing water molecules at the active site (8), it appears that this loop functions to exclude water where this loop is large enough (11).

Figure 1 also shows three highly conserved amino acids that are not found to be involved in binding a ligand, and therefore may only be needed for part of the structure: Ile151, Pro398 and Gly399. Thus, of 12 amino acids identified in the sequence alignment, nine of these are shown to be involved in the active site of the crystal structures. And of the nine amino acids shown in the structures to be specifically
involved in binding a ligand, seven of these were apparent in the sequence alignment presented in Figure 1.

**Phylogeny Based on OMP Decarboxylase**

A phylogenetic analysis of the OMP decarboxylase sequences (Fig. 4) is in general fairly consistent with earlier evolutionary trees. Four large clusters are evident. Three of these correspond to the well established taxonomic kingdoms Archaea, Eubacteria, and Eukaryota. However, a quite distinct fourth group is evident; designated as “Mixed”, this group consists mainly of four mycobacteria, two members of the Thermus/Deinococcus group and one myxobacterium. More surprisingly, included in this group is one multicellular eukaryote, *Trypansosoma cruzi*. Since all other eukaryotes group well together, and at quite a distance from the mycobacterial subgroup, it may well be that the inclusion of *T. cruzi* in this subgroup is an example of lateral gene transfer.

Among the eukaryotes, the multicellular species *C. elegans* and *D. melanogaster* are farther outliers from the main cluster than the simple slime mold *D. discoidium*. Also of interest is that the many fungi in this data set show as much or more diversity than all the multicellular eukaryotes.

In examining the subunit sizes of OMP decarboxylases from the different species, it was evident that the Archaea had a distinctly smaller protein than the eukaryotes. When all proteins were analyzed for their size versus evolutionary distance from *A. pernix* (Archaea), it became evident that the subunit size varied almost two-fold from Archaea to the Pyrenomycetes fungi, and there appeared to be a modest correlation of increasing protein size with evolutionary distance from *A. pernix* (Fig. 5A). Since no unique functions or benefits have been described for the larger proteins, the data were reanalyzed by subtracting for each protein sequence the variable extensions at the
N- and C-termini (Fig. 1, Inserts 1, 2, and 12), as well as the single large insert of the Pyrenomycetes fungi, an example of which is Insert 7 for *N. crassa* at residues 195 - 299 (Fig. 1). The replot for this set of truncated sequences now shows much less variation in size (Fig. 5B). The horizontal line suggests an average subunit size for the core domain of about 228 amino acids, with a variation of about 25 amino acids from this mean. This variation represents the remaining smaller inserts in the different species (see Table I). If only those residues for all species that are consistently present at any given position were used for this plot (the minimal consensus), the data set would reduce to about 180 amino acids. This would then be the smallest core domain, with an $M_r <20$ kDa.

A detailed analysis of all 12 insert positions for the major phylogenetic groups is listed in Table I. Insert numbers correspond to those in Fig. 1. Inserts 1 and 2, where they occur in the four structures, are separate helices, while Insert 12 extends the final helix in these proteins (Fig. 3). Insert 2 is the immediate N-terminal addition found in almost all sequences. Since a few species have no Insert 2, and since for Eubacteria Insert 2 is fairly small, it is interpreted here as an early extension at the N-terminus. Only the eukaryotes have an additional N-terminal extension, Insert 1, which is generally in the 16-23 amino acid range (see Table I). The average size for most inserts is quite small. This reflects the fact that for a given subgroup, only a few member species actually have an insert at that position. The inclusion of standard deviation values in Table I shows that these values are often zero, indicating no variation in insert size, and this suggests that the function of many loops at which inserts occur may place a constraint on the size of the insert at a given position.

The analysis of insert variations and their effect on subunit size is summarized in Table II. Thus the average subunit for all species contains 270 amino acids. Eliminating the N- and C-terminal inserts produces a core subunit of 237 amino acids. Such an average core protein still contains many small inserts; by comparison, for Archaea alone
the core size would only be 207 amino acids. If no inserts of any size are included, then
the value for the minimal consensus sequence is ≥178 amino acids.

An additional level of complexity in the evolution of OMP decarboxylase is that
the gene for this enzyme became fused with a second gene in all multicellular
eukaryotes studied (Fig. 6). The second gene always codes for orotate phosphoribosyl-
transferase (orotate PRTase), the enzyme that immediately precedes OMP decarboxylase
in the pathway for the de novo synthesis of UMP, so that the fused gene now codes for a
bifunctional protein designated as UMP synthase (1). Since the arrangement of the
fused genes, or of their protein domains, is commonly with the orotate PRTase
preceding OMP decarboxylase, and since it is found in the slime mold D. discoideum, this
fusion presumably occurred once at the beginning of the metazoan expansion, and has
been stably maintained thereafter.

While T. cruzi also has a bifunctional UMP synthase, the domains are linked in the
reverse order. The ready ability of the OMP decarboxylase to fuse with an orotate
PRTase domain for a stable bifunctional UMP synthase is evident in the four protein
structures. Each one shows that both the N-terminus and C-terminus extend side-by-
side on one surface of the protein. These termini are evident at the upper left of the
subunit structure in Fig. 3, or at the lower corners in Fig. 4. Therefore linkage at either of
these termini with orotate PRTase leads to a fairly similar overall structure for the
bifunctional UMP synthase.

This unusual pattern is again consistent with the origin of the T. cruzi OMP
decarboxylase and UMP synthase being completely separate from all other eukaryotes.
Since it has been proposed above that the acquisition of the OMP decarboxylase gene
was by lateral transfer for this parasite, then the unusual reverse fusion with orotate
PRTase (Fig. 6) may represent an independent event for the evolution of UMP
synthases.
An additional bifunctional protein containing OMP decarboxylase is suggested for *D. radiodurans* by the linkage of an open reading frame sequence that is 5' to and contiguous with the coding sequence for OMP decarboxylase. The sequence of this putative protein domain, containing 315 amino acids, does not show significant identity with any of the orotate PRTases, and it remains unidentified.

**Implications for Hexulose-phosphate Synthase**

A BLAST search of GenBank with a bacterial OMP decarboxylase sequence identified the sequence for 3-hexulose-6-phosphate synthase (EC 4.1.2.X) from 11 species. This enzyme catalyzes the reversible aldol condensation shown in Scheme 2, by which methanophile bacteria are able to incorporate one carbon compounds for biosynthesis of hexoses, which in turn are utilized for other pathways.

![Scheme 2](image_url)

It is very likely that the two enzymes diverged from a common ancestor, since of the eight invariant amino acids in OMP decarboxylase five are also found in hexulose-phosphate synthase at the corresponding position (last two sequence entries, Fig. 1). This enzyme from various methanophile bacteria is always an oligomer, though it may be a dimer (16,17), a tetramer (18), or a hexamer (19). Since this enzyme also has the same unique motif found at the junction of the two subunits of OMP decarboxylase in forming the active site, then it would be consistent with the observed oligomer forms of hexulose-phosphate synthase for it to have a similar architecture for the catalytic site.
Kinetic studies have emphasized that for hexulose-phosphate synthase divalent metals appear to be essential for activity, with Mg^{2+} and Mn^{2+} being equally effective, while other divalent cations had modest or no benefit (16,18,19). However, it was noted that this enzyme was somewhat unstable, and that the presence of metals improved stability for long term storage at -60 °C or short exposure to 60 °C (16,18,19). Thus, it is not clear if the metals are essential for catalysis, or simply function in stabilizing the protein. The latter has been demonstrated as a benefit of metals in preparations with yeast OMP decarboxylase, although this enzyme does not require metals for activity (3,20).
DISCUSSION

The results of such an extensive sequence analysis for a single gene support our emerging understanding of the conservation of protein structure during evolution. OMP decarboxylases from three kingdoms have now been defined as having an α/β barrel structure. By comparing those amino acids that are invariant or highly conserved across more than 80 species to the amino acids shown to be functional in the crystal structures, we see that through evolutionary divergence only the most structurally and functionally essential amino acids are conserved, and that amino acids identified by such an alignment are highly significant and most likely to be involved directly in catalysis.

However, the validity of the above judgment is highly dependent on the size of the data base. Two papers analyzed OMP decarboxylase sequences available earlier, and found 48 invariant residues for a set of 17 sequences (21), which became reduced to 10 invariant residues for a set of 20 sequences (22). It must again be noted that on average such sequences code for 270 amino acids. In the present analysis only eight of these amino acids remain as invariant for 82 sequences, and seven of these invariant residues participate in binding a transition state analog. Such a result is consistent with the hypothesis that invariant residues are chiefly specified by their essential role in the actual chemical reaction.

What may be essential for this enzyme is the core α/β barrel structure, which is most likely very similar in all species. It is worth noting that inserts that occur at different positions of the sequence, and in different species, almost always occur in loops between elements of secondary structure (see Fig. 3). This suggests that such inserts are easily tolerated if they are on the outside of the core structure, and do not alter this core structure or sterically impede access of the ligands to the catalytic site. Thus the single 100 amino acid insert shown for *N. crassa* (Insert 7) may be viewed as a
small domain attached to the side of the main α/β barrel core. Since at least seven
species of fungi have retained this large insert, it may have a function yet to be
discovered.

The segments of consensus sequence generally have well defined boundaries (Fig. 1),
and therefore defining the size of an insert between two such consensus segments is
then not difficult. In a few cases, such boundaries are not as well defined, and we
therefore set the boundary so as to accommodate the minimum sequence for any given
species within such a segment. This may clearly influence the Insert size shown for
some species.

Only the mycobacteria (Mixed group, Table I) have evolved inserts at each of the
12 positions. All other species are lacking inserts at two or more of the 12 positions. No
functions are currently known for any such insertions, and they may simply reflect a
stochastic process that increases the size of transcribed RNA by various processes. The
great variation in insert sizes and location suggests that such alterations with time of an
ancestral protein core may be normal events that may confer no benefit, but are
tolerated in the absence of any deficit. While inserts may occur anywhere in the protein
sequence, many of these may cause a harmful change in structure, and are therefore not
maintained or observed.

The increase in size with divergence distance from Archaea is produced by up to
12 inserts. About half of this size increase is produced by the 3 inserts that extend the
N- and C-termini, and half of the increase is by all other inserts combined to the core of
the protein. Inspection of the values for standard deviations in Table I suggests two
possible patterns for such alterations in the protein’s size. For the majority of inserts, the
standard deviation is quite small. This would be consistent with a single insertion
event, at a given position, and that divergence of species from that event occurred with
minimal changes at that site. Where the standard deviation is quite large, as for Insert 2
with Archaea, this represents an N-terminal extension that may have occurred
separately in the different species (by alteration of the start codon), and therefore shows considerable diversity in the size of this extension.

In addition to its ability to handle smaller inserts or additions to the core structure, the core domain itself is easily joined to at least two other protein domains (Fig. 6). With the demonstration that *T. cruzi* can form UMP synthase in the alternate configuration (23) (Fig. 6), it becomes evident that in such gene fusions there must be adequate linker DNA to code for the connecting polypeptide between the two domains. This assumption is based on the fact that the OMP decarboxylase from yeast is only functional as a dimer. Kinetic studies obtained the same result for this enzyme activity in the mammalian UMP synthase, which adds further evidence that only the dimer is functional (15,24). Furthermore, the recent crystal structure (25) of the bacterial orotate phosphoribosyltransferase presented a comparable architecture of an active dimer, since this enzyme also requires a loop from one subunit to form a part of the catalytic site of the adjacent subunit. Therefore, in the evolution of the bifunctional UMP synthase, the linkage of the two catalytic domains must be sufficiently extensive or flexible to permit each of the domains to align with its cognate domain in the adjacent UMP synthase subunit so that both the orotate phosphoribosyltransferase domains and the OMP decarboxylase domains can form functional catalytic sites across respective dimer interfaces.

That this fusion of the same two genes occurred at least twice also suggests that some benefit is associated with the coupling of these two protein domains. Although the two domains of UMP synthase catalyze sequential metabolic steps, evidence for the strict channeling of the common metabolite (OMP) between the two domains was not observed (26). However, with the ability to separately clone and express the two domains for the human UMP synthase, it was shown that the bifunctional protein has much greater stability than either of the independent catalytic domains (27), and such a benefit may explain two different gene fusion events to produce UMP synthase. This
latter finding also implies that some interaction must occur between the two different domains in UMP synthase to provide this observed stability.
Footnotes

* Supported by a grant from the Medical Alumni Association Endowment Fund of the University of North Carolina.

1 The abbreviations used are: BMP, barbiturate 5'-monophosphate; OMP, orotidine 5'-monophosphate
REFERENCES

1. Traut, T. W., and Jones, M. E. (1996) Prog. Nuc. Acid Res. Mol. Biol. 53, 1-78
2. Radzicka, A., and Wolfenden, R. (1995) Science 267, 90-93
3. Miller, B. G., Smiley, J. A., Short, S. A., and Wolfenden, R. (1999) J. Biol. Chem. 274, 23841-23843
4. Woese, C. R., and Fox, G. E. (1977) Proc. Natl. Acad. Sci. USA 74, 5088-5090
5. Woese, C. R., Gutell, R., Gupta, R., and Noller, H. F. (1983) Microbiol. Rev. 47, 621-669
6. Doolittle, R., Feng, D.-F., Tsang, S., Cho, G., and Little, E. (1996) Science 271, 470-476
7. Rivera, M. C., Jain, R., Moore, J. E., and Lake, J. A. (1998) Proc. Natl. Acad. Sci. USA 95, 6239-6244
8. Wu, M., Mo, Y., Gao, J., and Pai, E. (2000) Proc. Natl. Acad. Sci. USA 97, 2017-2022
9. Appleby, T., Kinsland, C., Begley, T., and Ealick, S. (2000) Proc. Natl. Acad. Sci. USA 97, 2005-2010
10. Harris, P., Poulsen, J.-C. N., Jensen, K. F., and Larsen, S. (2000) Biochemistry 39, (in press)
11. Miller, B. G., Hassell, A. M., Wolfenden, R., Milburn, M. V., and Short, S. A. (2000) Proc. Natl. Acad. Sci. USA 97, 2011-2016
12. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) Nuc. Acids Res. 22, 4673-4680
13. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) Nuc. Acids Res. 24, 4876-4882
14. Felsenstein, J. (1989) Cladistics 5, 164-166
15. Traut, T. W., and Payne, R. C. (1980) Biochemistry 19, 6068-6074
16. Kato, N. (1990) Meth. Enzymol. 188, 397-401
17. Yanase, K., Ikeyama, K., Mitsui, R., Ra, S., Kita, K., Sakai, Y., and Kato, N. (1996) *FEMS Microbiol. Lett.* **135**, 201-205

18. Müller, R. H., and Babel, W. (1990) *Meth. Enzymol.* **188**, 401-405

19. Ferenci, T., Strøm, T., and Quayle, J. R. (1974) *Biochem. J.* **144**, 477-486

20. Cui, W., DeWitt, J., Miller, S., and Wu, W. (1999) *Biochem. Biophys. Res. Comm.* **259**, 133-135

21. Radford, A. (1993) *J. Mol. Evol.* **36**, 389-395

22. Kimsey, H. K., and Kaiser, D. (1992) *J. Biol. Chem.* **267**, 819-824

23. Gao, G., Nara, T., Nakajima-Shimada, J., and Aoki, T. (1999) *J. Mol. Biol.* **285**, 149-161

24. Traut, T. W., Payne, R. C., and Jones, M. E. (1980) *Biochemistry* **19**, 6062-6068

25. Scapin, G., Grubmeyer, C., and Sacchettini, J. C. (1994) *Biochemistry* **33**, 1287-1294

26. Traut, T. W. (1989) *Arch. Biochem. Biophys.* **268**, 108-115

27. Yablonski, M. J., Pasek, D. A., Han, B.-D., Jones, M. E., and Traut, T. W. (1996) *J. Biol. Chem.* **271**, 10704-10708

28. Christopher, J. (1998) (Program Manual), The Center for Macromolecular Design, Texas A&M University, College Station, TX.

29. Kraulis, P. (1991) *J. Appl. Cryst.* **24**, 946-950
Table II  Scaling the Core Subunit Size of OMP Decarboxylase

| Group          | Protein | (# aa) |
|----------------|---------|--------|
| All ODCases    | native  | 270    |
| All ODCases    | core    | 237    |
| Archaea        | core    | 207    |
| Consensus Sequence | core  | ≥178   |
Figure Legends

Figure 1. Sequence alignment for OMP decarboxylases and hexulose-phosphate synthases. The first eleven species represent OMP decarboxylase sequences, and these are separated as phylogenetic subgroups; the last 2 species represent hexulose-phosphate synthase sequences. The length of each polypeptide chain is shown next to the species name. Above each set of sequences, the first line shows the type of bond made by the corresponding amino acid in the structures of the 4 proteins:

- H-bond from aa side chain to ligand; ○ H-bond from aa backbone to ligand;
- H-bond from aa side chain to other key aa. Also on this line are shown the insert positions (numbered), for sequence residues beyond the consensus core. On the second line are indicated important amino acids that are invariant (bold) or highly conserved (italic). Residues involved in ligand binding are boxed.

Figure 2. Ribbon diagram of the OMP decarboxylase subunit structure bound to BMP (11), and with invariant plus conserved residues shown. The structure is for the yeast enzyme (PDB 1DQX), but the residue numbers correspond to Figure 1. Figure created with SPOCK (28), and rendered with MOLSCRIPT (29).

Figure 3. Dimer structure of OMP decarboxylase, showing the positions of Loops A and B, and the different inserts (numbered as in Fig. 1). Secondary structures in white or in grey designate the two separate subunits; in black are insertions of variable length. Figure created with SPOCK (28), and rendered with MOLSCRIPT (29).
Figure 4. Phylogeny of OMP decarboxylase. Arrows, and the numbers associated with them, show where in this evolutionary scheme these particular inserts first appeared; Insert numbers are from Figure 1. Inserts are shown here only for a size $\geq 5$ amino acids. The many species shown by abbreviated names include: Acremonium chrysogenum, Acremonium lolii, Aeropyrum pernix, Aquifex aeolicus, Arabidopsis thaliana, Archaeoglobus fulgidus, Aspergillus fumigatus, Aspergillus niger, Aspergillus oryzae, Aspergillus parasiticus, Bacillus caldolyticus, Bacillus subtilis, Bartonella bacilliformis, Bos taurus, Caenorhabditis elegans, Candida albicans, Candida boidinii, Candida glabrata, Candida maltosa, Candida parapsilosis, Candida tropicalis, Candida utilis, Dictyostelium discoidium, Drosophila melanogaster, Emericella nidulans, Endomyces magnusii, Epichloe typhina, Escherichia coli, Haemophilus influenza, Helicobacter pylori, Homo sapiens, Hypocrea jecorina, Kluyveromyces lactis, Kluyveromyces marxianus, Lactobacillus plantarum, Lactococcus lactis, Methanobacterium thermoautotrophicum, Methanococcus jannashii, Mucor circinelloides, Mucor racemosus, Mus musculus, Mycobacterium bovis, Mycobacterium leprae, Mycobacterium smegmatis, Mycobacterium tuberculosis, Myxococcus xanthus, Naegleria gruberi, Neurospora crassa, Nicotiana tabacum, Pachysolen tannophilus, Penicillium chrysogenum, Phycomyces blakesleeanus, Picha angusta, Picha anomala, Picha fabianii, Picha ohmeri, Picha stipitis, Pseudomonas aeruginosa, Pyrococcus abyssi, Pyrococcus horikoshii, Rhizomucor pusillus, Rhizopus niveus, Saccharomyces cerevisiae, Saccharomyces exiguus, Saccharomycopsis fibuligera, Salmonella typhimurium, Schizophyllum commune, Schizosaccharomyces pombe, Sordaria macrospora, Streptococcus pneumoniae, Sulfolobus solfataricus, Synechocystis sp.3, Thermoplasma acidophilum, Thermotoga maritima, Thermus thermophilus, Trichoderma harzianum, Trypanosoma cruzi, Ustilago maydis, Yarrowia lipolytica.
Figure 5. Change in subunit size during the evolution of OMP decarboxylases.
A. Total size of the native protein.  B. Size of the core protein. Proteins from different
taxonomic kingdoms or subgroups: ▼, Archaea; ◊, Eubacteria; □, Multicellular
Eukaryota; ●, Mycobacteria; ▲, Pyrenomycetes; ○, Other Fungi.

Figure 6. Examples of gene fusion of OMP decarboxylase with a second protein domain.
In most multicellular eukaryotes OMP decarboxylase is fused with, and C-terminal to,
orotate phosphoribosyltransferase. In T. cruzi, this linkage is reversed, and in
D. radiodurans fusion is with some unidentified protein.
### Table I. Variable Inserts Attached to the Core of the OMP Decarboxylase Protein

| Species                          | Subunit Size (#aa) | Inserts and Sizes (#aa) | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 | LoopB |
|----------------------------------|--------------------|-------------------------|----|----|----|----|----|----|----|----|----|----|----|----|-------|
|                                  | Native             | Truncated*              | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 |       |
| **Archaea** (n=8)                |                    |                         |    |    |    |    |    |    |    |    |    |    |    |    |       |
| Average                          | 226                | 207                     | 0  | 14.3| 3.0| 5.0| 1.4| 0.6| 2.4| 0.9| 1.6| 6.1| 0  | 4.4| 2.0   |
| Std Dev                          | 12.3               | 2.4                     | 0  | 13.3| 0  | 1.6| 0.7| 0.7| 1.8| 0.4| 1.1| 0.8| 0  | 4.6| 1.2   |
| **Eubacteria** (n=14)            |                    |                         |    |    |    |    |    |    |    |    |    |    |    |    |       |
| Average                          | 233                | 226                     | 0  | 4.4 | 3.0| 5.1| 0.6| 1.0| 5.4| 5.6| 1.2| 7.5| 0  | 2.3| 10.9  |
| Std Dev                          | 10.9               | 8.0                     | 0  | 4.6 | 1.2| 0.6| 1.2| 0  | 3.3| 2.3| 1.1| 1.6| 0  | 2.7| 0.5   |
| **Mixed group** (n=8)            |                    |                         |    |    |    |    |    |    |    |    |    |    |    |    |       |
| Average                          | 266                | 243                     | 0  | 16.4| 12.8|7.4| 6.9| 3.5| 3.0| 6.5| 7.9| 8.5| 1.0| 6.8| 6.3   |
| Std Dev                          | 25.8               | 14.1                    | 0  | 2.7 | 2.3| 1.5| 0.4| 1.1| 0  | 1.4| 3.7| 1.6| 1.6| 3.7| 2.1   |
| **Multi-Cellular Eukaryota** (n=9) |                |                         |    |    |    |    |    |    |    |    |    |    |    |    |       |
| Average                          | 265                | 222                     | 18.6|14.3| 3.0| 5.0| 4.1| 4.0| 5.2| 0.1| 0.2| 3.4| 0  | 8.9| 12.0  |
| Std Dev                          | 3.1                | 3.0                     | 1.9| 0.5 | 0  | 0  | 0.3| 0  | 1.9| 0.3| 0.7| 2.1| 0  | 0.9|       |
| **Pyrenomycetes and Ascomycota Fungi** (n=8) |          |                         |    |    |    |    |    |    |    |    |    |    |    |    |       |
| Average                          | 379                | 333                     | 245| 14.0| 3.0| 5.0| 8.0| 4.0| 97.6|0  | 0  | 10.5| 0 | 9.0| 20.6  |
| Std Dev                          | 12.6               | 9.4                     | 7.5| 2.4 | 0  | 0  | 0  | 0  | 5.0| 0  | 0  | 7.8| 0  | 1.3| 1.9   |
| **Other Fungi** (n=35)           |                    |                         |    |    |    |    |    |    |    |    |    |    |    |    |       |
| Average                          | 271                | 229                     | 16.4|14.1| 3.0| 5.6| 4.6| 4.0| 7.9| 0  | 0  | 6.3| 1.3| 10.0| 12.0  |
| Std Dev                          | 9.1                | 7.3                     | 1.7| 0.3 | 0  | 2.9| 0.5| 0  | 5.1| 0  | 0.2| 2.8| 2.3| 1.3|       |
| **Total** (n=82)                 |                    |                         |    |    |    |    |    |    |    |    |    |    |    |    |       |
| Average                          | 269.5              | 237.0                   | 228.3|    |    |    |    |    |    |    |    |    |    |    |       |
| Std Dev                          | 41.6               | 33.8                    | 12.4|    |    |    |    |    |    |    |    |    |    |    |       |

* Truncation 1 is the size of the protein after deletion of inserts 1, 2, and 13 at N- and C-termini. Truncation 2 is Truncation 1 minus insert 7.*
Figure 4

Eukaryota

Multicellular

Eubacteria

Archaea
Figure 5

A: native protein

B: core protein
Figure 6

| Protein Description       | Domain Arrangement                                      |
|---------------------------|---------------------------------------------------------|
| Most UMP synthases:       | N-[orotate PRTase]OMP decarboxylase-C                   |
| T. cruzi UMP synthase:    | N-[OMP decarboxylase]orotate PRTase-C                   |
| D. radiodurans ODCase:    | N-[ORF unidentified]OMP decarboxylase-C                 |
The chemistry of the reaction determines the invariant amino acids during the evolution and divergence of orotidine 5'-monophosphate decarboxylase
Thomas W. Traut and Brenda R.S. Temple

J. Biol. Chem. published online July 11, 2000

Access the most updated version of this article at doi: 10.1074/jbc.M003468200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts