Unexpected Divergence of Enzyme Function and Sequence: "*N*-Acylamino Acid Racemase" Is *o*-Succinylbenzoate Synthase†

David R. J. Palmer,‡ James B. Garrett,‡ V. Sharma,§ R. Meganathan,§ Patricia C. Babbitt,|| and John A. Gerlt*,‡

Department of Biochemistry, University of Illinois, Urbana, Illinois 61801, Department of Biological Sciences, Northern Illinois University, DeKalb, Illinois 60115, and Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, California 94143

Received January 21, 1999; Revised Manuscript Received February 16, 1999

ABSTRACT: A protein identified as "*N*-acylamino acid racemase" from *Amycolaptosis* sp. is an inefficient enzyme ($k_{cat}/K_m = 3.7 \times 10^2$ M^{-1} s^{-1}). Its sequence is 43% identical to that of an unidentified protein encoded by the *Bacillus subtilis* genome. Both proteins efficiently catalyze the *o*-succinylbenzoate synthase reaction in menaquinone biosynthesis ($k_{cat}/K_m = 2.5 \times 10^5$ and 7.5×10^5 M^{-1} s^{-1}, respectively), suggesting that this is their "correct" metabolic function. Their membership in the mechanistically diverse enolase superfamily provides an explanation for the catalytic promiscuity of the protein from *Amycolaptosis*. The adventitious promiscuity may provide an example of a protein poised for evolution of a new enzymatic function in the enolase superfamily. This study demonstrates that the correct assignment of function to new proteins in functional and structural genomics may require an understanding of the metabolism of the organism.

The number of enzymes greatly exceeds the number of protein folds (*1*, *2*). As a result, the members of many groups of enzymes related by >50% pairwise sequence identity are known to catalyze the same reaction on structurally similar substrates (enzyme families). However, as the amount of sequence and structural information increases, the members of many divergently related groups of enzymes that are related by <50% pairwise sequence identity (superfamilies) have been discovered to catalyze different reactions on structurally diverse substrates. The latter, designated mechanistically diverse superfamilies (*3*, *4*), offer the potential for defining structure−function relationships in a variety of reactions that are important in biochemistry.[1] The members of mechanistically diverse superfamilies catalyze different overall reactions that are related by a common partial reaction in which a chemically similar intermediate is generated and significantly stabilized by conserved functional groups and elements of secondary and tertiary structure. These superfamilies include the enolase, crotonase (enoyl CoA hydratase), vicinal oxygen chelate, *N*-acetylneuraminate lyase, and amidohydrolase superfamilies.

The various members of the enolase superfamily catalyze reactions that are initiated by abstraction of the α-proton of a carboxylate anion substrate to generate a stabilized enolate anion intermediate (*6*): racemization catalyzed by mandelate racemase and "*N*-acylamino acid racemase"; dehydration catalyzed by enolase, D-galactonate dehydratase, D-glucarate dehydratase, and *o*-succinylbenzoate synthase; cycloisomerization catalyzed by muconate lactonizing enzyme; and deamination catalyzed by β-methylaspartate ammonia lyase. Each member is known to require a divalent metal ion for catalysis, with the ligands located in conserved positions in the sequences and structures; the enolate anions are generated in the first coordination sphere of the divalent metal so these can be stabilized by potentially large electrostatic interactions. However, the identities of the functional groups that generate the enolate anion or direct it to product differ, depending upon the reaction that is catalyzed. On the basis of the functional groups involved in abstraction of the α-proton, the enolase superfamily can be divided into three subgroups: mandelate racemase (MR,[2] one His and one Lys), muconate lactonizing enzyme (MLE, two Lys), and enolase (one Lys).

We have rationalized how different reactions can be catalyzed by homologous active site architectures in this superfamily (*3*, *6*). Structures are available for enolase (*7*), MR (*8*), MLE (*9*, *10*), D-glucarate dehydratase (*11*), and D-galactonate dehydratase.[3] In each enzyme, the active site is located at the interface between (1) one or more loops in the N-terminal $\alpha_4\beta_3$ domain that sequesters the substrate from solvent and (2) the C-terminal end of the $(\beta/\alpha)_8$ barrel

† This research was supported by NIH Grants GM-40570 and GM-52594 to J.A.G. and GM-50262 to R.M. and a PhRMA Foundation Faculty Development Award to P.C.B.

* To whom correspondence should be addressed. E-mail: j-gerlt@uiuc.edu.

‡ University of Illinois.

§ Northern Illinois University.

|| University of California.

[1] The formation of an extensive array of hydroxylated, unsaturated, and epoxidized fatty acids in plants is accomplished by members of a family of enzymes (sequence identities of >80%). This diversity of function has been attributed to differences in the access of the substrate to a reactive oxygen species generated by a non-heme iron center (*5*).

[2] Abbreviations: MLE, muconate lactonizing enzyme; MR, mandelate racemase; NAAAR, *N*-acylamino acid racemase; OSBS, *o*-succinylbenzoate synthase; SHCHC, 2-hydroxy-6-succinyl-2,4-cyclohexadiene carboxylate.

[3] J. G. Clifton, S. J. Wieczorek, J. A. Gerlt, and G. A. Petsko, unpublished observations.

10.1021/bi990140p CCC: $18.00
Published on Web 03/16/1999

domain. Each metal ligand and acid or base catalyst is located at the C-terminus of a separate β-strand in the barrel domain. In these positions, the functional groups surround the substrate and intermediate and can be altered independently to generate different overall reactions while retaining the shared partial reaction, abstraction of an α-proton which generates an enolic intermediate. The N-terminal domain largely determines substrate specificity via hydrophobic, hydrogen-bonding, and/or electrostatic interactions with the distal portions of the substrate. This design likely simplified the evolution of new reactions since catalysis and specificity determinants are located on separate domains that can be varied separately.

As the sequences of additional genomes become available,[4] we identify new members of the enolase superfamily. However, given the observation that approximately 40% of the open reading frames (orfs) in sequenced genomes have unknown functions, it is not surprising that we have not been able to assign functions to all new members of the superfamily; their sequences are not very similar to those of previously characterized members of the superfamily. However, in some cases, we have been able to use membership in a particular subgroup to both assign new functions to enzymes catalyzing known reactions [D-glucarate dehydratase is also an L-idarate dehydratase and a D-glucarate/L-idarate epimerase (*12*−*14*)] and deduce the functions of new, unknown members [D-galactonate dehydratase (*15*) and L-rhamnonate dehydratase[5]]. We note that assignment of function to new proteins is not restricted to the enolase superfamily and is a major focus of functional and structural genomics (*16*, *17*).

We describe the assignment of the "correct" function [*o*-succinylbenzoate synthase (OSBS)] to a member of the enolase superfamily that first was described as catalyzing a different reaction [*N*-acylamino acid racemase (NAAAR)]. As would be expected for membership in this superfamily, both are initiated by abstraction of the α-proton of a carboxylate substrate, the shared partial reaction in the superfamily. This correction of function was made possible by the deposition of a new sequence encoding a protein with unknown function in the databases. The ability of a member of the superfamily to catalyze two reactions with significantly different proficiencies may provide an example of how new catalytic activities evolve within the enolase superfamily.

MATERIALS AND METHODS

Pfu DNA polymerase was purchased from Stratagene. Restriction enzymes, T4 DNA ligase, and T7 polynucleotide kinase were purchased from Gibco BRL. Oligonucleotide primers were synthesized and DNA sequencing was performed by the Genetic Engineering Facility (University of Illinois Biotechnology Center, Urbana, IL). All other reagents were of the highest-quality grade commercially available and purchased from Sigma or Aldrich.

An FPLC system (Pharmacia Biotech) was used for purification of proteins. Spectrophotometric assays were performed with a Perkin-Elmer Lambda 14 spectrophotometer.

Isolation of NAAAR from Amycolaptosis. The gene encoding NAAAR was amplified from *Amycolaptosis* sp. T-1-60 chromosomal DNA (IFO 15079, Institute for Fermentation, Osaka, Japan) by PCR and inserted into the pET17b expression vector. The resulting plasmid was used to transform *Escherichia coli* strain BL21(DE3), and the protein was expressed at 37 °C without induction by IPTG. The cells were lysed, and the NAAAR was purified to electrophoretic homogeneity by chromatographies on DEAE-Sephacel, Resource-Q, and Sephacel S-200.

Isolation of YtfD (OSBS) from Bacillus subtilis. The gene encoding YtfD was amplified from *B. subtilis* chromosomal DNA by PCR and inserted into the pET15b expression vector. The resulting plasmid was used to transform *E. coli* strain BL21(DE3), and the protein was expressed at 37 °C without induction by IPTG. The cells were lysed, and N-terminal His-tagged YtfD was purified to electrophoretic homogeneity from the soluble fraction with a column of Chelating Sepharose Fast-Flow charged with Ni^{2+}.

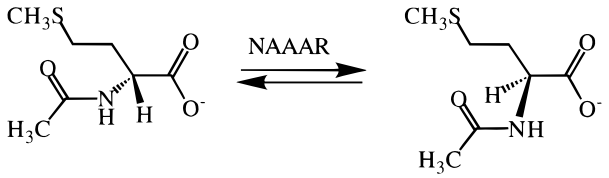
Isolation of OSBS from E. coli. The *men*C gene was amplified from *E. coli* MG1655 chromosomal DNA by PCR and inserted into a modified pET15b expression vector. The OSBS was purified as described for the OSBS from *B. subtilis*.

Assay of Enzymatic Activities. *N*-Acylamino acid racemase activity was measured at 25 °C with a polarimeter using enantiomers of *N*-acetylmethionine, 50 mM Tris-HCl (pH 8), and 0.1 mM $MnCl_2$.

OSBS activity was measured at 25 °C using SHCHC, 50 mM Tris-HCl (pH 8), and 0.1 mM $MnCl_2$ by quantitating the decrease in absorbance of 2-hydroxy-6-succinyl-2,4-cyclohexadiene carboxylate (SHCHC) ($\Delta\epsilon = -4000\ M^{-1}\ cm^{-1}$ at 293 nm). SHCHC was prepared (*18*) from chorismate (*19*) by the coupled actions of isochorismate mutase (encoded by the *ent*C gene from *E. coli*) and SHCHC synthase (encoded by the *men*D gene from *E. coli*). SHCHC was purified by anion-exchange chromatography on a column of Dowex 1x8-400 eluted with a gradient of 0 to 0.3 M LiCl; the bislithium salt of SHCHC was precipitated by addition of 10:1 acetone/methanol.

RESULTS AND DISCUSSION

Discovery of N-Acylamino Acid Racemase. The Takeda Chemical Co. was seeking a commercially viable catalyst for the racemization of *N*-acylamino acids (*20*−*23*) so that it might be able to convert a racemic mixture of substrates to enantiomerically pure amino acids in the presence of either a D- or L-acylase. A protein designated "*N*-acylamino acid racemase" (NAAAR) was identified in a species of *Amycolaptosis* (a Gram-positive Streptomycete) by a microbial screening program (49 000 strains and isolates).



NAAAR was reported to catalyze racemization of a range of *N*-acylamino acids, with *N*-acetyl-L-methionine being the best substrate. Derivatives of charged amino acids were not

[4] At www.tigr.org/tdb/mdb/mdb.html.

[5] B. K. Hubbard, J. Delli, and J. A. Gerlt, unpublished observations.

Table 1: Kinetic Constants for the OSBS and NAAAR Reactions

	OSBS		NAAAR[a]	
protein	k_{cat} (s^{-1})	k_{cat}/K_m (M^{-1} s^{-1})	k_{cat} (s^{-1})	k_{cat}/K_m (M^{-1} s^{-1})
NAAAR from *Amycolaptosis*	120 ± 10	$(2.5 \pm 0.6) \times 10^5$	11.7 ± 0.6	$(3.7 \pm 0.7) \times 10^2$
YtfD from *B. subtilis*	150 ± 10	$(7.5 \pm 1.0) \times 10^5$	$<2 \times 10^{-3}$	$<5 \times 10^{-2}$
OSBS from *E. coli*	19 ± 1	$(1.6 \pm 0.3) \times 10^6$	$<2 \times 10^{-3}$	$<5 \times 10^{-2}$

[a] *N*-Acetyl *S*-methionine as the substrate.

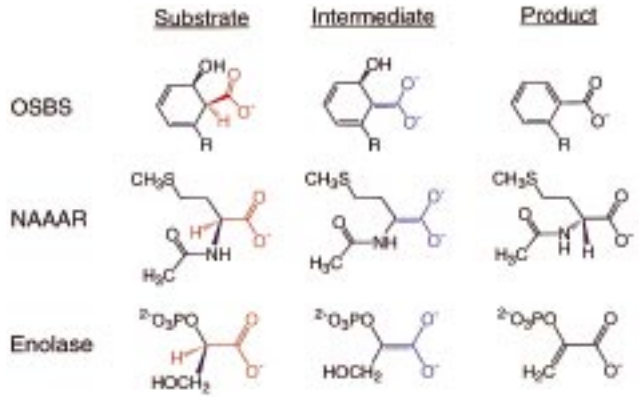


FIGURE 1: Comparison of the substrates, enolic intermediates, and products in the NAAAR-, OSBS-, and enolase-catalyzed reactions. The common substrate structure, α-proton of a carboxylate group, is highlighted in red; the common enolic intermediate structure is highlighted in blue.

substrates, but NAAAR was tolerant of the identity of the acyl group, with L-Ala-L-Met also being a substrate. Although the biosynthetic pathways for arginine and lysine include *N*-acetylated and *N*-succinylated intermediates, respectively, no step involves racemization of an *N*-acylamino acid (*24*), thereby raising the question of the physiological role of NAAAR in the metabolism of *Amycolaptosis*.

On the basis of its similarity to the structurally characterized MLE (29% sequence identity), we predicted (*6*) that NAAAR uses two Lys acid or base catalysts: one to abstract the α-proton of the substrate to generate an enolic intermediate and the second to deliver a solvent-derived proton to the opposite face of the intermediate to generate the product (Figure 1). Because MR uses one His and one Lys as acid or base catalysts (*25*), we began to investigate the NAAAR-catalyzed reaction so we might understand the mechanistic consequences of the diverged functional group.

Characterization of NAAAR. We amplified the gene for NAAAR from chromosomal DNA, expressed the protein in *E. coli*, and purified it to homogeneity. We confirmed that NAAAR catalyzes the reported racemization reaction (*21*) and determined that racemization is accompanied by the incorporation of solvent-derived deuterium, in support of the two-base mechanism we proposed on the basis of the relationship of its sequence to that of both MLE and MR.[6] Our studies further disclosed that (1) NAAAR is an inefficient enzyme as assessed by the value of its k_{cat}/K_m (Table 1) and (2) its reaction is subject to substrate inhibition at concentrations exceeding 50 mM. Both of these kinetic "problems" could restrict the use of NAAAR in the desired commercial process. More notably, both the low value of k_{cat}/K_m relative to that expected for an enzyme-catalyzed reaction (*26*) and the uncertain importance of racemization of *N*-acylamino acids in microbial physiology caused concern about the identity of the reaction catalyzed by this protein.

Sequence Relationships of NAAAR with Members of the Enolase Superfamily. When our studies of NAAAR were initiated, the databases contained no proteins other than MLEs whose sequences were >25% identical with that of NAAAR. However, after completion of the *B. subtilis* genome project (*27*), we found a protein, YtfD, using the BLASTP algorithm (*28*) whose sequence is 43% identical to that of NAAAR. No function was assigned to YtfD, but the database annotation reported that YtfD is "similar to *N*-acylamino acid racemase," "similar to muconate cycloisomerase of *Pseudomonas putida*," and "cotranscribed with *men*BE operon genes."[7]

The *men* operon encodes enzymes that catalyze steps in the anabolic pathway for menaquinone (*29*), essential in electron transfer chains in eubacteria (Figure 2A). When the *men* operons in *B. subtilis* (*27*) and *E. coli* (*29*, *30*) were compared (Figure 2B), each was found to contain six genes. The functions of the proteins encoded by all six genes in *E. coli* are known (Figure 2); in *B. subtilis*, the proteins encoded by two genes, YtxM and YtfD, were reported in the database annotations to have unknown functions. The sequences of MenH in *E. coli*[8] and YtxM are 28% identical, so we assigned the MenH function to YtxM.

The sequence of the remaining protein, YtfD, is 16% identical to that of *o*-succinylbenzoate synthase (OSBS or MenC) in *E. coli*, the other protein encoded by the *E. coli* operon that did not have an assigned counterpart in *B. subtilis*. This low level of identity undoubtedly dissuaded the annotators of the *B. subtilis* genome from proposing that YtfD is OSBS. However, both YtfD and the OSBS from *E. coli* are members of the enolase superfamily (22 and 19% identical, respectively, with that of MLE), and the active sites of each are predicted to contain two Lys acid or base catalysts. As a result, we hypothesized that YtfD is the OSBS in *B. subtilis*. Given the 43% identity relating YtfD and NAAAR, we also hypothesized that NAAAR is the OSBS in *Amycolaptosis*.

Reactions Catalyzed by YtfD, NAAAR, and the OSBS from E. coli. To test this hypothesis, we first determined that the genes encoding YtfD and NAAAR complement a *men*C mutant of *E. coli* (JRG862) that is otherwise unable to grow anaerobically on glycerol minimal medium with fumarate as the terminal electron acceptor (*31*). We then measured the OSBS activities of purified YtfD, NAAAR, and OSBS from *E. coli*; the kinetic constants are included in Table 1.

The value of k_{cat}/K_m for the reaction catalyzed by the OSBS from *E. coli* (1.1×10^6 M^{-1} s^{-1}) is within the range expected

[6] B. K. Hubbard, J. B. Garrett, and J. A. Gerlt, unpublished observations.

[7] PIR locus I39598.

[8] R. Meganathan, unpublished observations.

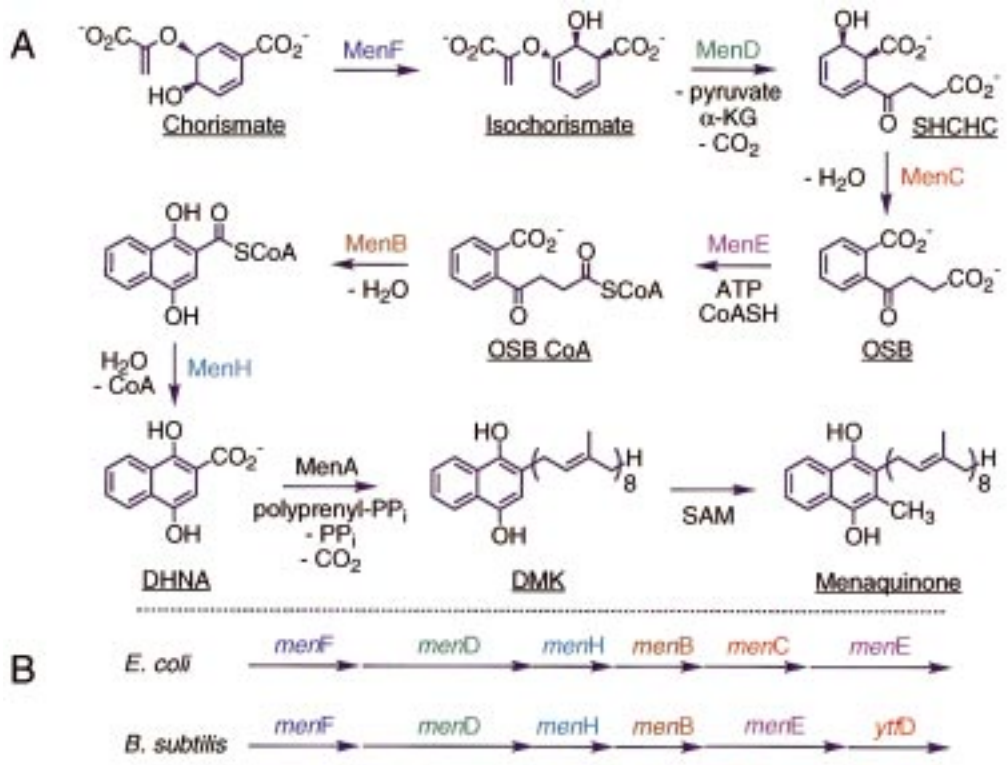


FIGURE 2: (A) Menaquinone anabolic pathway (*29*) and (B) the menaquinone operon in *E. coli* (top) and the menaquinone operon in *B. subtilis* (bottom).

for an efficient enzyme (*26*).[9] YtfD catalyzes the OSBS reaction with the value of its k_{cat}/K_m (7.5×10^5 M^{-1} s^{-1}) being comparable to that of the OSBS from *E. coli*, confirming that YtfD is the OSBS in *B. subtilis*. NAAAR also catalyzes the OSBS reaction, with the value of its k_{cat}/K_m (2.5×10^5 M^{-1} s^{-1}) being similar to those measured for YtfD and OSBS from *E. coli* and exceeding the value measured for the NAAAR reaction by a factor of 700 (Table 1).

The NAAAR and OSBS reactions catalyzed by the protein from *Amycolaptosis* are (1) inhibited competitively by salicylhydroxamate, with similar values for the K_is describing the inhibition (1.5 ± 0.5 μM for the NAAAR reaction and 0.5 ± 0.05 μM for the OSBS reaction), and (2) impaired significantly by mutagenesis of the predicted active site Lys residues (K163R and K263R). Both observations are consistent with the same active site catalyzing both reactions. We conclude that the metabolic function of the protein from *Amycolaptosis* is the OSBS reaction.

We assayed the OSBSs from *B. subtilis* and *E. coli* for NAAAR activity (Table 1), and neither protein has detectable activity. Hence, the NAAAR activity by which the OSBS from *Amycolaptosis* was first isolated is not a general property of OSBSs.

Sequence Relationships of OSBSs. We previously noted that the *Synechocystis* genome encodes an unknown member of the MLE group (SynORF); i.e., its active site is predicted to contain two Lys acid or base catalysts. We noted that SynORF was most homologous to OSBS from *E. coli* and MLE II from *P. putida*; however, we felt that the low levels of sequence identity with known members of the MLE group (20% identical with that of the OSBS and 18% identical with that of the MLE II) prevented correct prediction of the reaction that SynORF catalyzes. However, considering the low levels of sequence identity relating the OSBSs from *Amycolaptosis*, *B. subtilis*, and *E. coli*, we predicted and determined that SynORF is an OSBS (data not shown). Using our assignments of function and those made by others for the OSBSs from *E. coli* (*21*), *Haemophilus influenzae*,[10] *Staphylococcus aureus*,[11] and *Mycobacterium tuberculosis*,[12] we aligned the sequences of the seven known OSBSs (Figure 3) and computed their pairwise identities (Table 2). Given the divergence of the sequences, the alignment cannot be absolutely correct, although the most conserved elements (e.g., ligands for the divalent metal ion and two Lys acid or base catalysts) reliably predict membership in the superfamily. Only 15 amino acid residues are conserved in these seven sequences, and on the basis of the structure of MLE (*9*, *10*), six are predicted to be located in the active sites and function as metal ion ligands or acid, base, or electrophilic catalysts.

With others (*32*), we noted that enzymes that mediate cofactor-dependent reactions can be related by low levels of sequence identity; benzoylformate decarboxylase catalyzes the same reaction as pyruvate decarboxylase, but the positions and identities of the acid or base catalysts are not conserved. However, the low values of the pairwise identities

[9] As isolated, YtfD, NAAAR, and OSBS from *E. coli* each require a divalent metal ion for activity, as expected for membership in the enolase superfamily (*6*). The kinetic constants were measured with 0.1 mM $MnCl_2$, a concentration observed to saturate each enzymatic activity. $MgCl_2$ also activates these enzymes, with the values of k_{cat}/K_m similar (within a factor of 3) to those measured with $MnCl_2$; the concentrations of $MgCl_2$ required for saturation range from 5 to 20 mM.

[10] SWISS-P locus MENC_HAEIN, accession number P44961.
[11] Genbank locus SAU51132, accession number U51132.
[12] EMBL locus MTY25D10, accession number Z95558.4.

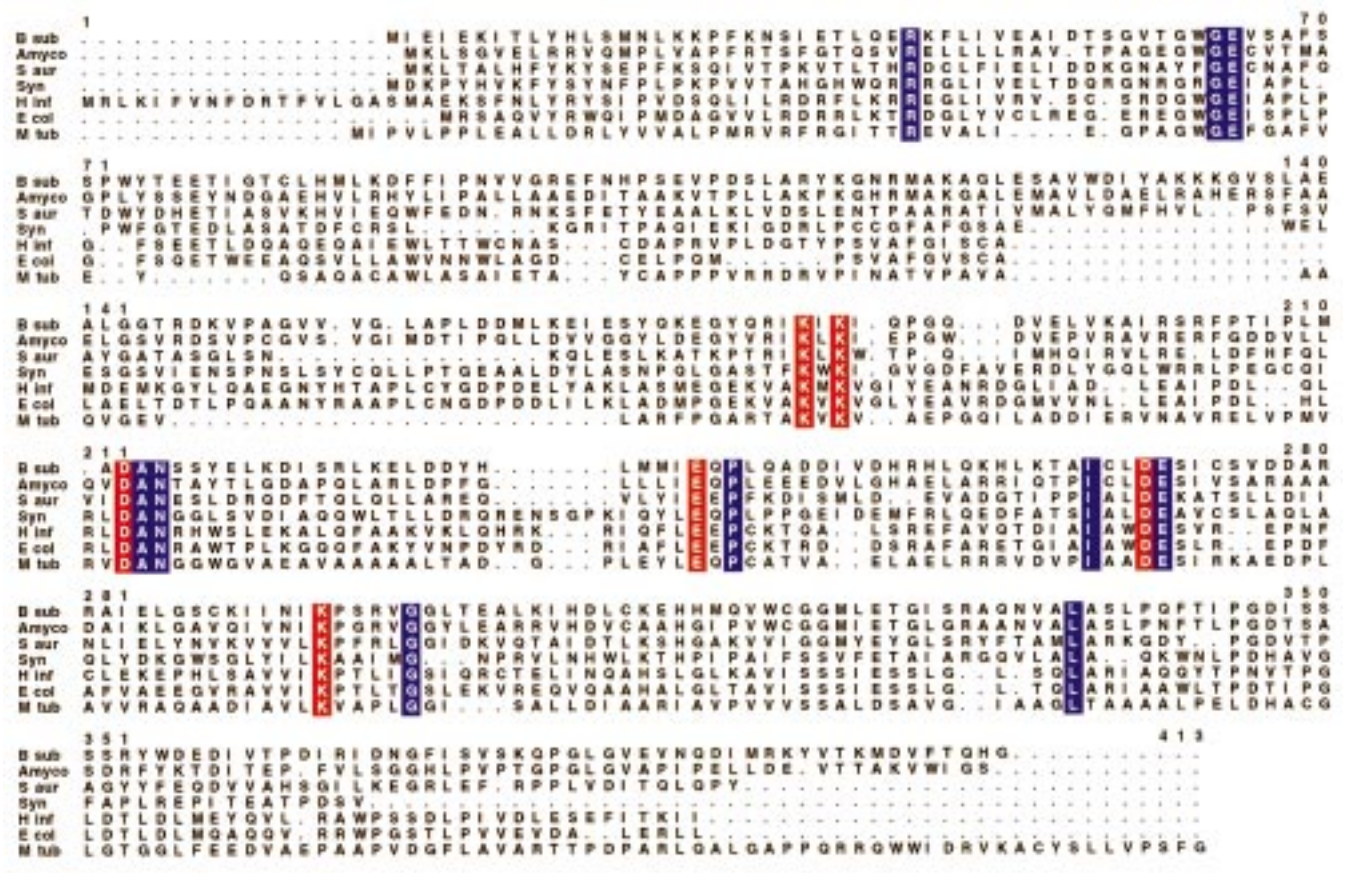


FIGURE 3: Sequence alignment of the known OSBSs. The conserved residues highlighted in red are the homologues of the active site residues in MLE (*9*, *10*); the conserved residues highlighted in blue are located elsewhere in the structure.

Table 2: Pairwise Sequence Identities Relating the Orthologous OSBSs

	Amycolaptosis	*B. subtilis*	*S. aureus*	*Synechocystis*	*H. influenzae*	*E. coli*	*M. tuberculosis*
Amycolaptosis	100	43	25	23	20	19	19
B. subtilis		100	27	22	15	15	14
S. aureus			100	23	20	19	19
Synechocystis				100	22	20	20
H. influenzae					100	57	26
E. coli						100	25
M. tuberculosis							100

observed for the OSBSs is unusual for cofactor-independent orthologues (divergently related enzymes that catalyze the same reaction in different organisms). For example, when MenB, MenD, MenE, and MenF from *E. coli* (Figure 2) are used as queries in the gapped-BLASTP algorithm (*28*), orthologues are "mined" simultaneously in the *H. influenzae*, *B. subtilis*, *M. tuberculosis*, and *Synechocystis* genomes. Using the same algorithm, the orthologues of OSBS (MenC) from *E. coli* are found only by iterations using as queries the most distantly related orthologues identified in preceding searches. The iterative PSI-BLAST algorithm mines all of the OSBSs, although the sequences having the most similar final "scores"/"E values" are not OSBSs but MLEs; after that of the OSBS from *B. subtilis*, the sequence of OSBS from *Amycolaptosis* is more identical to that of several MLEs and D-galactonate dehydratase (*15*) than to those of the other OSBSs.

Comparison with the Enolase-Catalyzed Reaction. The OSBS reaction is a dehydration in which the α-proton of a carboxylate substrate likely is abstracted by a basic catalyst (one Lys) and then the *β*-hydroxyl group is eliminated, presumably by assistance of an acidic catalyst (the second Lys). Enolases catalyze a dehydration reaction by an acid- or base-catalyzed mechanism (Figure 1). The sequences of the enolases encoded by the five genomes that encode OSBSs are aligned in Figure 4; these display pairwise sequence identities in excess of 55% (Table 3).

The contrast between the relatively high level of similarity among enolases with the markedly lower identities relating the OSBSs requires explanation. Our hypothesis is that the greater divergence of the OSBSs is permitted by relaxed requirements for a conserved active site geometry. The dehydration reaction catalyzed by OSBS is very exergonic as a result of aromatization. According to the Hammond postulate, the transition state for the reaction will resemble the substrate, suggesting little scission of the bond to the α-proton. As expected, the substrate for OSBS has been reported to undergo nonenzymatic dehydration under mild conditions (*18*), at least when compared to malate (*33*) and, by analogy, 2-phosphoglycerate, suggesting that OSBSs are not as proficient as catalysts (*34*) as enolases. Perhaps, the low proficiency that characterizes efficient catalysis of the

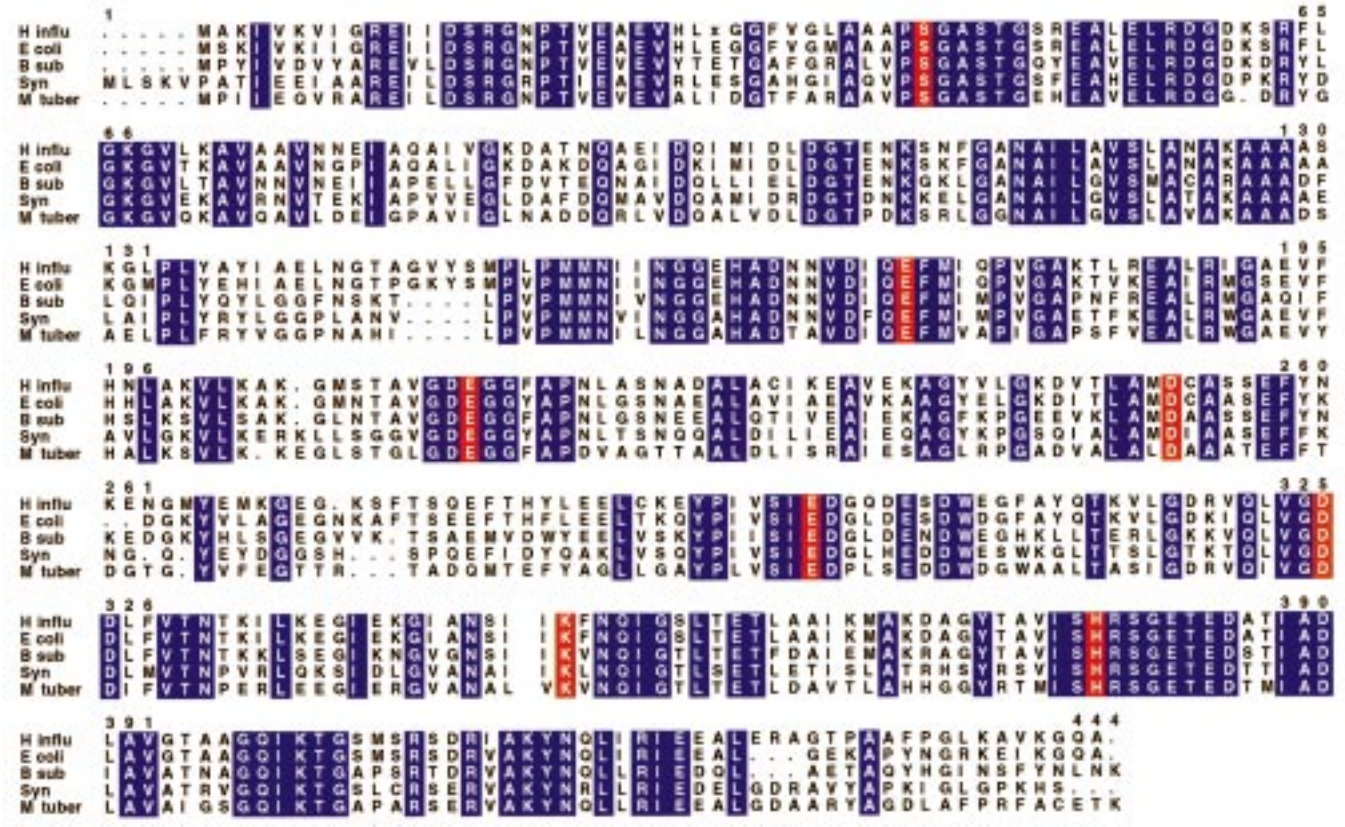


FIGURE 4: Sequence alignment of the enolases encoded by genomes that are known to encode OSBSs. The conserved residues highlighted in red are the homologues of the active site residues in yeast enolase (*7*); the conserved residues highlighted in blue are located elsewhere in the structure.

Table 3: Pairwise Sequence Identities Relating the Orthologous Enolases

	H. influenzae	*E. coli*	*B. subtilis*	*Synechocystis*	*M. tuberculosis*
H. influenzae	100	86	64	58	56
E. coli		100	64	59	55
B. subtilis			100	60	58
Synechocystis				100	59
M. tuberculosis					100

OSBS reaction allowed the sequences of orthologues to diverge. And, perhaps, the three-dimensional structures and positions of the acid or base groups also have diverged within the confines of the $(\beta/\alpha)_8$ barrel domain without compromising the catalytic proficiency. Structural studies of the OSBSs from *Amycolaptosis*, *E. coli*, and *B. subtilis* are underway to explore this possibility.

Evolution of New Enzymes in the Enolase Superfamily. Both the NAAAR- and OSBS-catalyzed reactions are initiated by abstraction of the α-proton of a carboxylate substrate, the common partial reaction in the superfamily (Figure 1). However, the ability of the OSBS from *Amycolaptosis* to catalyze the much more demanding racemization reaction of *N*-acylamino acids is unexpected; i.e., the proficiency of OSBS for catalysis of racemization is much greater than that for catalysis of dehydration. Our hypothesis is that the ability of this OSBS to catalyze the NAAAR reaction is catalytic promiscuity that has no impact on the organism in its metabolic niche.

We have proposed that mechanistically diverse superfamilies arise from a process in which the gene encoding an enzyme catalyzing a low level of the desired new reaction is duplicated (*3*, *4*, *6*). Selection then allows alterations in both substrate specificity and the mechanisms for diverting the stabilized enolate anion intermediate to the new product. Accordingly, this and other OSBSs may represent facile starting points for the evolution of new functions in the enolase superfamily. Alternatively, we will not be surprised if unknown members of the MLE subgroup function as racemases of *N*-acylated amino acids.

Implications for Structural and Functional Genomics. The initial identification of OSBS from *Amycolaptosis* as an (inefficient) NAAAR should signal caution to the fields of structural and functional genomics; neither the observation of a function nor the detection of significant sequence identity or structural homology is sufficient for correct assignment of the actual biological function. Given the increasing number of enzyme superfamilies whose members stabilize similar intermediates and/or transition states but catalyze different overall reactions (*3*, *4*), we expect that incorrect assignments of functions to new proteins discovered in genome sequencing projects will be made unless the measured or predicted activity is placed in the context of the metabolism of the organism.

The discovery that orthologues containing stringently conserved active site groups can otherwise diverge signifi-

cantly in sequence emphasizes the complexities for prediction of function for members of even well-characterized superfamilies. The example reported here calls into question the common practice of tentative assignment of function for unknown proteins based on pairwise sequence comparisons with known proteins. Instead, a more rigorous analysis of relationships within entire superfamilies may be required to generate good hypotheses for the functional identities of unknown orfs.

REFERENCES

1. Dorit, R. L., Schoenbach, L., and Gilbert, W. (1990) *Science 250*, 1377−1382.
2. Chothia, C. (1992) *Nature 357*, 543−544.
3. Babbitt, P. C., and Gerlt, J. A. (1997) *J. Biol. Chem. 272*, 30591−30594.
4. Gerlt, J. A., and Babbitt, P. C. (1998) *Curr. Opin. Chem. Biol. 2*, 607−612.
5. Broun, P., Shanklin, J., Whittle, E., and Somerville, C. (1998) *Science 282*, 1315−1317.
6. Babbitt, P. C., Hasson, M., Wedekind, J. E., Palmer, D. R. J., Barrett, W. C., Reed, G. H., Rayment, I., Ringe, D., Kenyon, G. L., and Gerlt, J. A. (1996) *Biochemistry 35*, 16489−16501.
7. Larsen, T. M., Wedekind, J. E., Rayment, I., and Reed, G. H. (1996) *Biochemistry 35*, 4349−4358.
8. Landro, J. A., Gerlt, J. A., Kozarich, J. W., Koo, C. W., Shah, V. J., Kenyon, G. L., Neidhart, D. J., Fujita, S., and Petsko, G. A. (1994) *Biochemistry 33*, 635−643.
9. Helin, S., Kahn, P. C., Guha, B. L., Mallows, D. G., and Goldman, A. (1995) *J. Mol. Biol. 254*, 918−941.
10. Hasson, M. S., Schlichting, I., Moulai, J., Taylor, K., Barrett, W., Kenyon, G. L., Babbitt, P. C., Gerlt, J. A., Petsko, G. A., and Ringe, D. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 10396−10401.
11. Gulick, A. M., Palmer, D. R. J., Babbitt, P. C., Gerlt, J. A., and Rayment, I. (1998) *Biochemistry 37*, 14358−14368.
12. Palmer, D. R. J., and Gerlt, J. A. (1996) *J. Am. Chem. Soc. 118*, 10323−10324.
13. Palmer, D. R. J., Hubbard, B. K., and Gerlt, J. A. (1998) *Biochemistry 37*, 14350−14357.
14. Hubbard, B. K., Koch, M., Palmer, D. R. J., Babbitt, P. C., and Gerlt, J. A. (1998) *Biochemistry 37*, 14369−14375.
15. Babbitt, P. C., Mrachko, G. T., Hasson, M. S., Huisman, G. W., Kolter, R., Ringe, D., Petsko, G. A., Kenyon, G. L., and Gerlt, J. A. (1995) *Science 267*, 1159−1161.
16. Terwilliger, T. C., Waldo, G., Peat, T. S., Newman, J. M., Chu, K., and Berendzen, K. J. (1998) *Protein Sci. 7*, 1851−1856.
17. Zarembinski, T. I., Hung, L. W., Mueller-Dieckman, H. J., Kim, K. K., Yokota, H., Kim, R., and Kim, S. H. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 15189−15193.
18. Popp, J. L., Berliner, C., and Bentley, R. (1989) *Anal. Biochem. 178*, 306−310.
19. Rieger, C. E., and Turnbull, J. L. (1996) *Prep. Biochem. Biotechnol. 26*, 67−76.
20. Tokuyama, S., Hatano, K., and Takahashi, T. (1994) *Biosci. Biotechnol. Biochem. 58*, 24−27.
21. Tokuyama, S., and Hatano, K. (1995) *Appl. Microbiol. Biotechnol. 42*, 853−859.
22. Tokuyama, S., and Hatano, K. (1995) *Appl. Microbiol. Biotechnol. 42*, 884−889.
23. Tokuyama, S., and Hatano, K. (1996) *Appl. Microbiol. Biotechnol. 44*, 774−777.
24. Voet, D., and Voet, J. G. (1995) *Biochemistry*, 2nd ed., Chapter 24, Wiley, New York.
25. Kenyon, G. L., Gerlt, J. A., Petsko, G. A., and Kozarich, J. W. (1995) *Acc. Chem. Res. 28*, 178−186.
26. Albery, W. J., and Knowles, J. R. (1976) *Biochemistry 15*, 5631−5640.
27. Kunst, F., et al. (1997) *Nature 390*, 249−256.
28. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res. 25*, 3389−3402.
29. Meganathan, R. (1996) in *Escherichia coli and Salmonella: Cellular and Molecular Biology* (Neidhart, F. C., Curtis, R., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umbarger, H. E., Eds.) pp 642−656, ASM Press, Washington, DC.
30. Blattner, F. R., et al. (1997) *Science 277*, 1453−1474.
31. Guest, J. R. (1977) *J. Bacteriol. 130*, 1038−1046.
32. Hasson, M. S., Muscate, A., McLeish, M. J., Polovnikova, L. S., Gerlt, J. A., Kenyon, G. L., Petsko, G. A., and Ringe, D. (1998) *Biochemistry 37*, 9918−9930.
33. Bearne, S. L., and Wolfenden, R. (1995) *J. Am. Chem. Soc. 117*, 9588−9589.
34. Radzicka, A., and Wolfenden, R. (1995) *Science 267*, 90−93.

BI990140P