Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 5043–5048, May 1996
Biochemistry

Exploring the active site of chorismate mutase by combinatorial mutagenesis and selection: The importance of electrostatic catalysis

(aromatic amino acid biosynthesis/*Bacillus subtilis*/catalytic antibody/metabolic pathway engineering/random oligonucleotide library)

PETER KAST, M. ASIF-ULLAH, NING JIANG, AND DONALD HILVERT†

Departments of Chemistry and Molecular Biology, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, CA 92037

Communicated by Ronald Breslow, Columbia University, New York, NY, January 21, 1996 (received for review October 31, 1995)

ABSTRACT **Chorismate mutase (EC 5.4.99.5) catalyzes the intramolecular rearrangement of chorismate to prephenate. Arg-90 in the active site of the enzyme from *Bacillus subtilis* is in close proximity to the substrate's ether oxygen and may contribute to efficient catalysis by stabilizing the presumed dipolar transition state that would result upon scission of the C—O bond. To test this idea, we have developed a novel complementation system for chorismate mutase activity in *Escherichia coli* by reengineering parts of the aromatic amino acid biosynthetic pathway. The codon for Arg-90 was randomized, alone and in combination with that for Cys-88, and active clones were selected. The results show that a positively charged residue either at position 88 (Lys) or 90 (Arg or Lys) is essential. Our data provide strong support for the hypothesis that the positive charge is required for stabilization of the transition state of the enzymatic chorismate rearrangement. The new selection system, in conjunction with combinatorial mutagenesis, renders the mechanism of the natural enzyme(s) accessible to further exploration and opens avenues for the improvement of first generation catalytic antibodies with chorismate mutase activity.**

Chorismate mutase (CM; EC 5.4.99.5) catalyzes the first committed step in the biosynthesis of the aromatic amino acids Phe and Tyr in bacteria, fungi and higher plants (1). The enzymatic Claisen rearrangement of (−)-chorismate to prephenate (2, 3), like the uncatalyzed thermal reaction (4), proceeds through a chair-like transition state (see Fig. 1*A*). Despite extensive study, however, the origins of the 10^6- to 10^7-fold rate acceleration by CM are poorly understood (5). Crystal structures of CM from *Bacillus subtilis* (BsCM) (6, 7) and *Escherichia coli* (EcCM) (8) reveal extensive shape and charge complementarity between the active sites and a conformationally restricted transition state analog (TSA). Restricting the conformational degrees of freedom of the flexible chorismate molecule and fixing it in a reactive conformation may consequently contribute to lowering the free energy of activation. The crystal structures also suggest that electrostatic and hydrogen-bonding interactions with the transition state species may further augment catalytic efficiency (5, 7, 9). Solvent effects show that the transition state for the uncatalyzed rearrangement is polar in character (10, 11) and kinetic isotope effects indicate that C—O bond cleavage precedes C—C bond formation (11, 12). Residues Arg-90 and Glu-78 in BsCM are positioned to complement a putative C—O dipole electrostatically. Here, we report the development of a novel genetic selection system for CM activity in *E. coli* and its application to probe the importance of the positive charge at position 90.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Strains and Plasmids. The construction of the CM-deficient *E. coli* strain KA12 was described earlier (13). Its genotype is *Δ(srlR-recA)306*::Tn*10*, *Δ(pheA-tyrA-aroF)*, *thi-1*, *endA-1*, *hsdR17*, *Δ(argF-lac)U169*, *supE44*. General cloning and library construction was carried out in strains XL1-Blue (14) and SURE 2 (Stratagene). Plasmids pJX3 (15), pJZ1g (16), and pBSCM2 (17) were obtained from the respective authors. Cloning or assay control vectors were pBluescript I KS(+) (pBLS; Stratagene) (18), pUC19 (19), pACYC184 (20), and pKSS (21).

DNA Manipulations. All nucleic acid manipulations were according to standard procedures (14). Restriction endonucleases and T4 DNA ligase were purchased from New England Biolabs or Boehringer Mannheim. Oligonucleotides were synthesized with a Pharmacia Gene Assembler Plus and purified by polyacrylamide gel electrophoresis. To minimize error frequencies during plasmid constructions, PCRs were carried out for not more than eight cycles at high template concentrations with *pfu* DNA polymerase (Stratagene). All relevant portions of the constructed and selected plasmids were confirmed by sequencing (22) with a kit from United States Biochemical.

Chorismate Mutase Complementation System. The p15A-based plasmid pKIMP-UAUC (see Fig. 1*B*) was constructed by a four-way ligation of three different PCR products and the 2323 bp *Bst*YI–*Cla*I fragment of vector pACYC184. The restriction sites used to fuse the PCR fragments are underlined: PCR 1 (183 bp) was obtained with primers UV5C (CTGG<u>GGTACC</u>GCATGC<u>ATCGAT</u>TAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGTGTGGAATTGTGAGCGGAT) and OMPCL (CCAGTA<u>CGGCCG</u>CGTTTGCTGCGCCTGCTACCAGCAGAGCTGGGACCAGGAGGGACAGTACTTTAACTTTCATGACTGTTTCCTGTGTGAAATTGTTATCC) on template pUC19, PCR 2 (1487 bp) was obtained with primers TYRAC (GCATGA<u>CGGCCG</u>AAGTCATGACTCTGGAAGATCCTCAGAGCCGTTACGGCC) and TYRAN (CCAGTGAATTCGAGCTC<u>GGTACC</u>) on template pJX3, and PCR 3 (1180 bp) was obtained with primers UV5C and PHECN (GTGAAGATCTCGAGCTGGACTAGTAGATAACGCATGGACCAGGCA) on template pJZ1g. Deletion of an internal 72-bp *Bsp*HI fragment from the four-way ligation product yielded pKIMP-UAUC (see Fig. 1*B*). The selection medium (M9c) was based on M9 minimal medium (14) and consisted of Na_2HPO_4 (60 mg/ml)/KH_2PO_4 (30 mg/ml)/NH_4Cl (10 mg/ml)/NaCl (5 mg/ml)/0.2% (wt/vol) D-(+)-Glc/1 mM $MgSO_4$/0.1 mM $CaCl_2$/thiamin-HCl (5 μg/ml)/4-

Abbreviations: CM, chorismate mutase; BsCM, *Bacillus subtilis* CM; EcCM, the CM domain of the bifunctional *Escherichia coli* CM-prephenate dehydratase; TSA, transition state analog; ScCM, *Saccharomyces cerevisiae* CM.

†To whom reprint requests should be addressed.

hydroxybenzoic acid (5 μg/ml)/4-aminobenzoic acid (5 μg/ml)/2,3-dihydroxybenzoic acid (1.6 μg/ml)/L-Trp (20 μg/ml)/sodium ampicillin (100 μg/ml)/chloramphenicol (20 μg/ml), pH 7.0. For Petri dishes, 15 g agar was added per liter. Where required, M9c was supplemented with 20 μg of L-Phe and/or L-Tyr per ml. A more detailed description of this complementation system will be given elsewhere.

Reengineering of *aroH*. Plasmid pKCMT-W (see Fig. 2*B*), which served as the basis for the libraries of randomized active site residues, was constructed by a four-way ligation procedure. Appropriately cut PCR products 4, 5, and 6 were ligated to a 2458-bp *Nsp*I–*Sac*I fragment from pKSS. PCR 4 (296 bp) was obtained (with fusion sites underlined) with primers PTRCS (CGTTCTACATGTTTGACAGCTTATCATCGACT) and PTAH1N (CGAATTCCGCGGATCATCATGGTCTGTT), PCR 5 (232 bp) with AHNCOS (AACAGACCATGGTGATCCGCGGAAT) and WAH3N (CATGTTACCGGTACATACTGCCAGCCGCTAAGCTCGCGTACGGCTTTTGCCGGGAAAACAGCATGCAAATC), and PCR 6 (239 bp) with WAH4S (GTATGTACCGGTAACATGTATGCAGGAAATGGACGTCACAGGCGGTCTTAAGAAGTGCATAAGAGTCATGATGACTGTACAGACAGATGTCCCTCA) and TWAH5N (CGTATTGAGCTCAAAAAAAAGCCCGCTCATTAGGCGGGCTTTATCAAGCGGCCGCACTAGTTTACAATTCAGTATTTTTTGTCAAGCTTAAATCGGGCCTCAATA), all on template pBSCM2.

Construction of *aroH* Libraries. Optimized conditions for the construction of the active site randomized libraries were as follows: PCRs 7 (for library A) and 8 (library B) were carried out with *Eco*RI-linearized template pKCMT-W by using the Stoffel fragment of *AmpliTaq* DNA polymerase (Perkin–Elmer). Denaturation, primer annealing and extension was performed for 20 cycles at 94°C (1 min), at 42°C (2 min), and at 72°C (5 s, with a 6-s extension per cycle), respectively. The 157-bp PCR products 7 and 8 were obtained with oligonucleotides R90AH1S (GGCGGTCTTAAGAAGTGCATA-*NN*(*G*/*C*)GTCATGATGACTGTACAGACAGAT) and TMUTN (ATCAAGCGGCCGCACTAGT) and the primer pair C88R90AH (GGCGGTCTTAAGAAA*NN*(*G*/*T*)ATC-*NN*(*G*/*C*)GTAATGATGACCGTACAGACAGATGTC-CCTCA) and TMUTN, respectively. The concentration of the primers containing the randomized codons (in italics, see also Fig. 2*B*) was limiting in the PCR. The PCR products were extensively purified by agarose gel electrophoresis before and after preparative restriction enzyme digestion with *Afl*II (site underlined) and *Hin*dIII. The resulting 104-bp library fragments were then ligated into a 3057-bp *Afl*II–*Hin*dIII fragment from pKCMT-W previously incubated with *Bsr*GI and gel purified to reduce wild-type *aroH* background.

***In Vivo* Analysis of Active Site Variants.** Purified single colonies of transformants from selective and nonselective minimal media plates were restreaked onto M9c plates with or without Phe and/or Tyr supplementation and incubated at 30°C. Growth of single colonies was recorded 2, 3, 5, and 15 days after seeding and evaluated on an arbitrary relative scale from 0 (no growth) to 5 (wild-type like growth). Such phenotypic characterizations were carried out in duplicate and fully reproducible. Growth rates were also quantified in liquid M9c medium (as above, but Glc was at 0.4%, and ampicillin and chloramphenicol were at 150 μg/ml and 30 μg/ml, respectively). The medium was inoculated 1:100 from a dense preculture and incubated under vigorous shaking at 30°C. A control culture supplemented with Tyr and Phe was run in parallel. $OD_{600\ nm}$ was directly monitored in the culture tubes over a period of 3 weeks and growth rates ($\mu = \Delta \ln OD_{600\ nm}/\Delta t$) and doubling times ($t_d = \ln 2/\mu$) for individual mutants were calculated from the exponential growth phase.

Protein Analysis. Crude protein extracts were obtained from 1.6 ml of a bacterial culture in "terrific broth" (14) (containing 150 μg/ml ampicillin and 20 μg/ml chloramphenicol) grown for 20 h at 37°C. The pelleted cells were resuspended in 400 μl of 50 mM potassium phosphate (pH 7.5)/1 mM dithiothreitol/RNase A and DNase I (each 1 μg/ml; Boehringer Mannheim) and disrupted by sonication at 0°C. Insoluble material was removed by centrifugation at 10,000 × *g* (4°C) for 20 min and the supernatant was stored at −20°C. Protein concentrations were determined using the Bio-Rad protein assay adapted for microtiter plates. Polyacrylamide gel electrophoresis was performed with the Pharmacia PhastSystem (20% High Density PhastGels).

Chorismate Mutase Enzyme Assays. Specific activities *in vitro* were determined by monitoring the initial rate of chorismate disappearance spectrophotometrically at 274 nm at 30°C with 50 μM chorismate in 50 mM potassium phosphate, pH 7.5. For K_m estimations, chorismate concentration was varied between 30 and 500 μM.

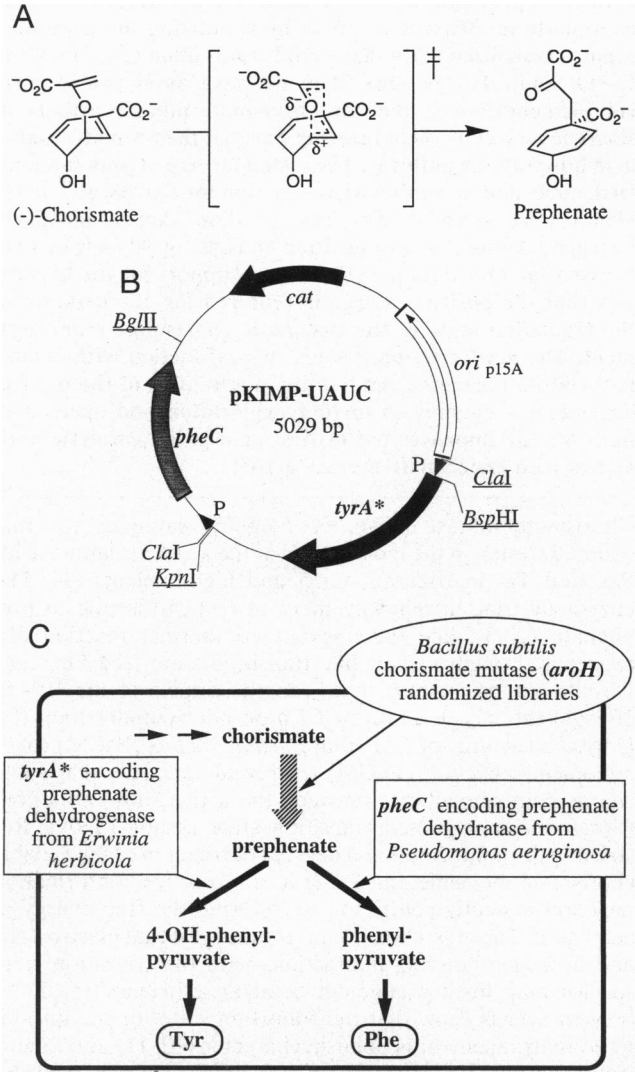

FIG. 1. *In vivo* complementation strategy for the CM reaction. (*A*) Claisen rearrangement catalyzed by CM with the presumed transition state. (*B*) Map of the auxiliary plasmid pKIMP-UAUC with the restriction sites used for its construction (underlined). *cat*, gene for chloramphenicol resistance; P, UV5 *lac* promoter; ori_{p15A}, origin of replication compatible with pMB1 replicons. (*C*) The genes *tyrA** and *pheC* restore the biosynthetic pathways from prephenate to Tyr and Phe in the pKIMP-UAUC-transformed *E. coli* strain KA12 (both endogenous bifunctional CM genes deleted). The hatched arrow represents the selection step for CM activity introduced into the cell via the pMB1 replicon-based *aroH* mutant libraries.

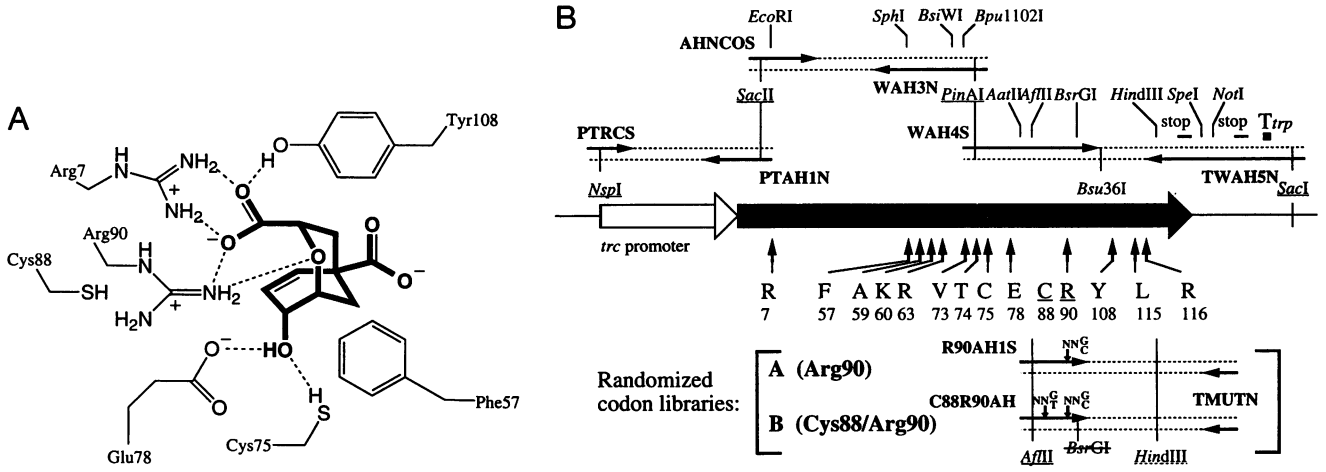

FIG. 2. BsCM active site residues. (*A*) Simplified schematic view of the active site structure of the complex between the enzyme and a TSA (bold lines) according to refs. 6 and 7. (*B*) The *aroH* gene was reengineered from three PCR fragments (fusion sites are underlined) to introduce unique restriction sites suitable for exchanging enzyme residues in contact with the TSA (indicated below the gene with arrows and sequence position). Underlined amino acids Cys-88 and Arg-90 were randomized using the codon formats NN(G/T) and NN(G/C), respectively, where N stands for all four bases. Synthetic oligonucleotides (horizontal arrows) are labeled. The *Bsr*GI site was eliminated in C88R90AH. T*trp*, *trp* terminator.

RESULTS

Engineering of the Chorismate Mutase Complementation System. We designed and established a selection strategy for CM activity in *E. coli* based on the biosynthesis of the essential amino acids Phe and Tyr. The fact that this organism possesses two CMs, fused respectively to prephenate dehydrogenase and prephenate dehydratase, makes metabolic pathway engineering a challenging problem. A simple assay for complementation of Phe auxotrophy based on the spontaneous conversion of prephenate to phenylpyruvate at low pH (23) did not work satisfactorily in our hands. To construct a fully enzymatic and more versatile system, we have devised a complementation assay consisting of three elements (Fig. 1 *B* and *C*): (*i*) the RecA$^-$ *E. coli* host strain KA12 (13) lacking both endogenous CM genes, *tyrA* and *pheA*, and hence prephenate dehydrogenase and prephenate dehydratase activities (23); (*ii*) the auxiliary plasmid pKIMP-UAUC encoding monofunctional forms of prephenate dehydrogenase [from *Erwinia herbicola* (15)] and prephenate dehydratase [from *Pseudomonas aeruginosa* (16)]; and (*iii*) a second, compatible plasmid (pMB1 origin of replication) for expression of *aroH*-encoded variants of the monofunctional BsCM (17).

Control experiments under selective conditions (i.e., in the absence of Phe and/or Tyr in the growth medium) with wild-type *aroH* (plasmid pKCMT-W) and/or empty vectors (pACYC184 and pUC19) demonstrated that all elements of the complementation system were functional and necessary. Because CM is required for the biosynthesis of both Phe and Tyr (Fig. 1*C*), three levels of selection stringency are possible. Complementation of the Phe$^-$ auxotrophy is the least stringent because considerable background growth is observed without CM (unpublished data). In the absence of Tyr, only cells with an active *aroH*-encoded BsCM grow. The most specific and demanding option is the simultaneous complementation of both Phe and Tyr auxotrophies.

Construction of Partially Randomized *aroH* Libraries. First, all residues of BsCM in direct contact with a TSA were identified (Fig. 2) by examining the crystal structure of the complex between BsCM and the TSA (6, 7). These active site residues were then made accessible to mutagenesis by reengineering of the *aroH* gene (Fig. 2*B*). Large fractions of the *aroH* gene were chemically resynthesized, and by making silent mutations, unique restriction sites were introduced that could serve as anchor points for insertion of randomized library fragments. In the design of the oligonucleotides, codons used in highly expressed genes of *E. coli* were favored (according to Version 7 of the program package by Genetics Computer Group).

We have used the complementation assay of Fig. 1 initially to elucidate the importance of the positively charged residue Arg-90 in the substrate binding pocket of BsCM (Fig. 2*A*). An *aroH* gene library (A) was constructed having the codon corresponding to Arg-90 randomized. In a second library (B), codons for Arg-90 and Cys-88 were randomized simultaneously. This combinatorial library allows for greater flexibility in the evolution of other functional active site constellations. Position 88 is not part of the binding site but is in contact with Arg-90 (Fig. 2). During construction of the libraries, care was taken to reduce the potential background of contaminating wild-type plasmid (wild-type codons were recognizable later when sequencing individual clones). To lower the frequency of cells transformed with two different library plasmids and of undesired *in vivo* DNA rearrangements, library ligations were first transformed into *E. coli* strain SURE 2 (having disabled DNA repair and recombination pathways). Subsequently, plasmid pools were isolated and, in the case of the larger library B, additionally digested with *Bsr*GI to eliminate any contaminating wild-type *aroH* (Fig. 2*B*). Diluted aliquots of the pools were transformed into the complementation host strain (KA12/pKIMP-UAUC), and clones able to provide their own Tyr and Phe were selected on suitable minimal medium agar plates. Only 3 of the 111 sequenced clones from library A and none of the 56 analyzed clones from library B contained wild-type codons. The library sizes exceeded the maximal theoretical diversity by at least a factor of 15. Library composition, examined by sequencing plasmid pools and individual clones from nonselective media, was found to be essentially random.

Alternative Active Site Residues Found by Selection. Fig. 3 *A* and *B* compile unambiguously characterized clones from libraries A and B, respectively. All of these clones appeared to be single transformants, and their *aroH* genes were shown by sequencing over the entire inserted PCR fragment to lack additional mutations. The distribution of (inactive) controls indicates good coverage of sequence space, particularly evident in library A.

When only codon 90 was randomized, Arg emerged as indispensable for an active enzyme *in vivo*, since all 44 clones obtained on selective agar plates contained an Arg codon (Fig. 3*A*). From combinatorial library B in which positions 88 and

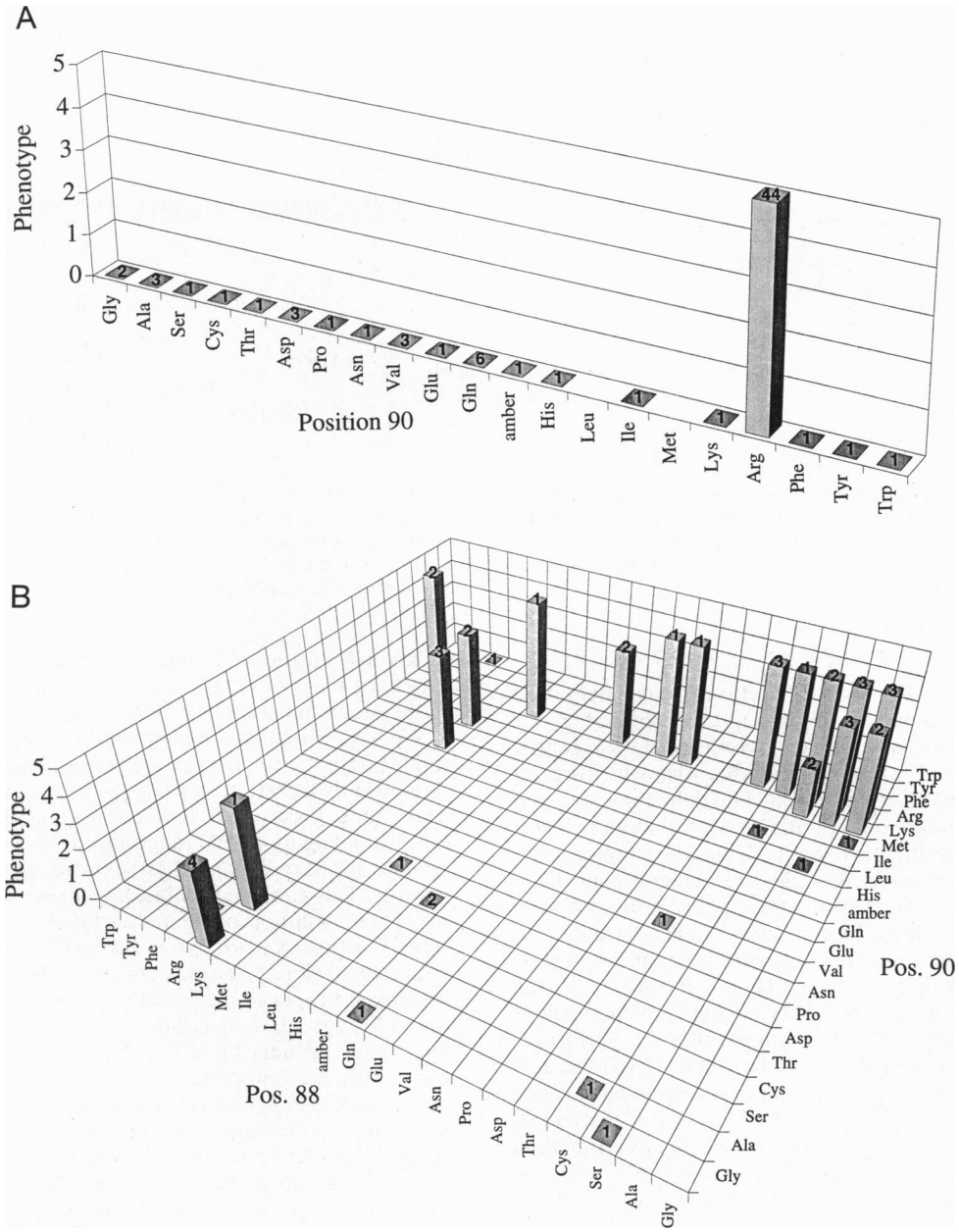

FIG. 3. Growth phenotypes of clones from partially randomized *aroH* libraries. The graphs contain clones picked from selective and nonselective plates. Amino acid residues are ordered according to increasing side chain volume (24). Growth speed (colony size) on selective agar plates is rated from 0 to 5. The number on the tops of columns indicates the frequency of finding a particular mutant. (*A*) Randomizing position 90 (library A). (*B*) Simultaneous randomization of positions 88 and 90 (combinatorial library B).

90 were simultaneously randomized (Fig. 3*B*), we selected many different Arg-90 clones which had wild-type-like activity *in vivo*, irrespective of the residue at position 88. Interestingly, two additional sets of active clones, lacking an Arg residue at position 90, are evident in Fig. 3*B*. Clones with Lys-90 were directly selected if the residue at position 88 was smaller than the wild-type Cys, such as Ser, Ala, and Gly (with phenotypes of 2, 4, and 4, respectively). Alternatively, if position 88 is the basic amino acid Lys, position 90 can be Gly (phenotype 3), Ser, Leu, and Met (phenotypes of 4).

Characterization of a Subset of Selected Active Site Variants. A subset of clones from libraries A and B was examined in more detail. The genotypes and phenotypes of the clones listed in Table 1 were confirmed by retransformation of isolated plasmids into KA12/pKIMP-UAUC and independent sequence and phenotype determination. All clones yielded the expected results with deviations from initial phenotype ratings by at most one unit. Growth doubling times of a representative set of mutants and controls were determined in liquid culture without Phe and Tyr. Although liquid growth conditions may differ inherently from plate assays, the colony scoring system correlates well with the doubling time measurements (Table 1). The liquid culture assays appear to be more sensitive than the plate assays, since mutant Lys-90 (phenotype 0) eventually grew, whereas the vector control did not.

Expression of mutant *aroH* genes was monitored by electrophoresis of a standardized amount of crude protein on SDS/20% polyacrylamide gels stained with Coomassie Blue (data not shown). The CM bands were of roughly equal intensity (within a factor of two) for mutants and wild type. In general, production of the CM variants can thus be considered to be constant, validating comparisons of *in vivo* phenotypes and allowing for direct determination of relative CM activity in crude extracts *in vitro*. As shown in Table 1, CM assays with

Table 1. Properties of a subset of selected clones

CM variant	Phenotype *in vivo**		Relative specific activity *in vitro*,† % of wild type‡
	On plates	t_d, h	
Controls‡			
Wild type	5	2.8	100
Vector	0	No growth	<0.05
Library A			
Lys-90	0	8.1	<0.05
His-90	0	No growth	<0.05
Ala-90	0	NT	<0.05
Ser-90	0	NT	<0.05
Gln-90	0	NT	<0.05
Phe-90	0	NT	<0.05
Library B			
Cys-88/Arg-90	5	NT	180
Tyr-88/Arg-90	5	2.8	<0.05
Gln-88/Arg-90	4	NT	0.06
Gly-88/Lys-90	4	4.3	0.09
Ala-88/Lys-90	4	NT	0.13
Lys-88/Leu-90	4	3.5	0.07
Lys-88/Ser-90	4	NT	0.07
Lys-88/Met-90	4	NT	0.08
Lys-88/Gly-90	3	5.7	<0.05
Ser-88/Lys-90	2	5.7	<0.05
Gly-88/Met-90	0	NT	<0.05
Ala-88/Leu-90	0	NT	<0.05

*The growth behavior under selective conditions was scored on agar plates and in liquid cultures. Doubling times (t_d) were determined in duplicates with standard deviations of <4% between measurements. NT, not tested.

†*In vitro* CM assays were carried out with 150 μg of crude protein extracts or appropriate dilutions. CM content within the cells was estimated to vary by a factor of two. The detection limit of the assay was at $\Delta A_{274nm} = 1.2 (\pm 0.9) \times 10^{-4}$ per min per 150 μg crude extract. Rates below 0.05% of the wild-type value ($\Delta A_{274nm} = 0.55$ per min per 150 μg) were not considered significant.

‡Controls were clones carrying plasmid pKCMT-W (protein identical to Cys-88/Arg-90, but encoded by wild-type codons TGC and AGA) and pBLS (empty vector).

total cellular protein gave a signal to background ratio of more than three orders of magnitude with the wild-type construct pKCMT-W. All mutant clones with a phenotype rating less than 4 had specific activities below the background level. With the exception of the wild-type protein Cys-88/Arg-90 selected from library B, all analyzed CM variants exhibited dramatic drops in *in vitro* activity of 680-fold (Ala-88/Lys-90) or more. Preliminary attempts to determine Michaelis constants with crude extracts of all the mutants in Table 1 with measurable CM activities showed that the K_m is increased in all cases at least three-fold relative to the wild-type extract (apparent K_m of 85 ± 6 μM). While better defined K_m values for these low-activity variants will require purification of individual proteins, this experiment establishes that the specific activities obtained at 50 μM chorismate essentially reflect relative k_{cat}/K_m values for the CM variants.

DISCUSSION

Direct genetic selection is a powerful tool for characterizing the mechanism and altering the properties of enzymes. Applied to libraries of randomized genes, it allows a palette of permissive changes to be evaluated rapidly (25–38). We have demonstrated here that a complex metabolic pathway can be deconvoluted to engineer a selection system. By replacing bifunctional genes with monofunctional versions from three other species (Fig. 1*C*), it became possible to select for CM activity in *E. coli*, the organism of choice for molecular genetics. We have used this system to examine the importance of Arg-90 in the active site of BsCM. Random-codon libraries were created and interesting (active) genes were directly selected. The examination of position 90 alone gives a synopsis of functional alternatives, whereas the relatively small number of clones analyzed from library B (compared to the 400 theoretically possible combinations) provides insight into alternative active site configurations without precluding other functional combinations.

CM activity of individual clones was assessed by three independent methods. First, recording the growth rate on agar plates permitted efficient sorting of clones into five complementation groups. Second, determination of doubling times in liquid cultures paralleled the results obtained on plates. Although this method carries the intrinsic risk of selecting secondary mutants that outgrow the original clone, it is substantially more sensitive than the plate assay. Lastly, CM specific activity was measured directly *in vitro*. All tested mutant proteins exhibited dramatically reduced activities *in vitro*, although some *aroH* variants show strong complementation *in vivo*. This is explained by the high expression level of the *aroH* genes (on a high-copy-number vector and transcribed from the strong, nonrepressed *trc* promoter), probably far in excess of the actual cellular needs for most phenotype 5 clones. Instability of the quaternary structure of the mutant proteins under the conditions of extract preparation and enzyme assays may also account for reduced *in vitro* activity in some cases (e.g., Tyr-88/Arg-90), since the randomized residues are at the interface of two subunits forming the active site in the trimeric BsCM (6, 7).

In the uncatalyzed reaction, C—O bond cleavage substantially precedes the formation of the new C—C bond resulting in a dipolar transition state (10–12, 39). It was proposed (7, 9) that the developing positive charge on the cyclohexadienyl ring may be stabilized by negative polarization of the bottom of the active site through the anionic carboxylate of Glu-78 in the BsCM-catalyzed rearrangement (Figs. 1*A* and 2*A*). The developing negative charge on the ether oxygen could be accommodated by electrostatic interactions with the positively charged guanidinium group of Arg-90 (7, 9). These assignments are now supported by experimental data. Randomization and selection experiments at the active site's negative pole show that a negative charge either at position 78 or 75, although not essential, is clearly needed for full activity *in vivo*, and its loss reduces the enzyme's efficiency by at least a thousand-fold *in vitro* (40). In this report, we show that a positive charge either at position 90 or 88 is essential. Its absolute conservation in active clones may indicate that the oxyanion needs stronger stabilization than the delocalized cation on the cyclohexadiene ring (39, 41). Our results suggest that a dipolar binding site (42) composed of Arg-90 and Glu-78 is a major contributor to the efficacy of CM (43). They thus provide strong but circumstantial evidence that the transition state of the enzymatic Claisen rearrangement, a species previously inaccessible to experiments (12, 39, 44), has a pronounced dipolar character like that of the uncatalyzed reaction (10–12).

The selection experiments demonstrate further that a positive charge at position 90 or 88 alone is not sufficient for activity; steric factors must also be important. Lys-90 in combination with the wild-type Cys-88 was not selected on plates and provides very poor growth in liquid culture. Similarly reduced catalytic activities upon Arg-to-Lys replacements have been reported in other enzymes (45). However, if residue 88 is smaller than Cys (such as Ser, Ala or Gly), Lys-90 becomes a viable alternative on selection plates, possibly due to relieved packing constraints within the protein. Thus, by allowing more flexibility in the combinatorial library B, we were able to select novel active site constellations that fulfill both chemical and steric requirements, but were not available in the more restricted library A. Similar observations emerged from selection experiments in which the role of Glu-78 in BsCM was examined (40). Steric factors were probably also important for

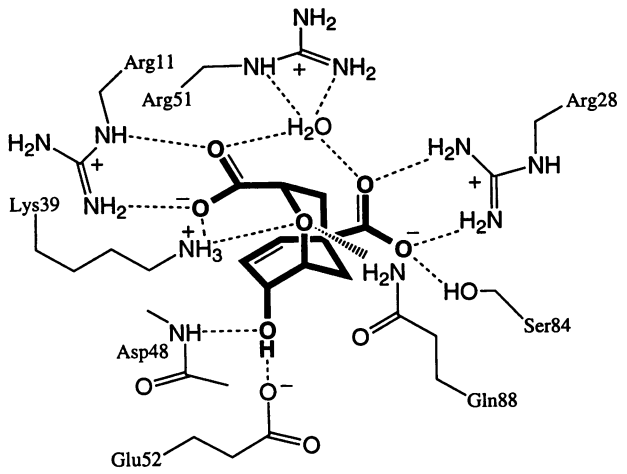

FIG. 4. Interactions between the TSA (bold lines) and active site residues in EcCM (8).

selection of active clones with Lys-88. To provide a positive charge able to interact with the partial negative charge on the substrate's ether oxygen, the ammonium group of Lys-88 must reach into the active site across a nonbulky residue at position 90 (Fig. 2*A*).

Our results also provide insight into the function of other CM catalysts. The structures of the catalytic antibody 1F7 (9), a *Saccharomyces cerevisiae* CM (ScCM) (46), and EcCM (5, 8) (Fig. 4) show active site residues capable of polarizing the ether oxygen of chorismate. In 1F7, however, the corresponding Arg-H95 is not within hydrogen-bonding distance of the TSA, and a negative pole is fully absent in the active site (9); these properties may explain its 10^4-fold lower efficiency (47). EcCM [and the structurally related ScCM, (46)] possesses a pronounced dipolar active site, with Glu-52 [Glu-198] and Lys-39 [Lys-168] providing the negative and positive charge, and it is likely that these residues fulfill roles similar to Glu-78 and Arg-90 in BsCM, respectively. Interestingly, our randomization and selection approach with the *B. subtilis* gene yielded viable solutions which resemble the unrelated EcCM's active site constellation by replacing Arg with Lys as the donor of the positive charge. Analogous experiments with EcCM or ScCM may confirm the importance of charge interactions in these enzymes and elucidate the role of EcCM Gln-88 or ScCM Glu-246 which donate a hydrogen bond to the TSA's ether oxygen (Fig. 4) (46). Randomization of other positions in CM coupled with the powerful selection *in vivo* promises to unravel their functions, while a similar approach with the first-generation catalytic antibody 1F7 (47) may allow the identification of variants with activities that equal or exceed those of the natural enzymes.

We thank Jeffrey D. Hartgerink for initial experiments. We are very grateful to R. A. Jensen for providing us with plasmids pJX3 and pJZ1g and to J. R. Knowles for the gift of plasmid pBSCM2. P.K. acknowledges the Schweizerischer Nationalfonds for a postdoctoral fellowship. This work was supported in part by grants from the Office of Naval Research and the National Institutes of Health (GM38273).

1. Haslam, E. (1993) *Shikimic Acid: Metabolism and Metabolites* (Wiley, New York).
2. Andrews, P. R., Cain, E. N., Rizzardo, E. & Smith, G. D. (1977) *Biochemistry* **16,** 4848–4852.
3. Sogo, S. G., Widlanski, T. S., Hoare, J. H., Grimshaw, C. E., Berchtold, G. A. & Knowles, J. R. (1984) *J. Am. Chem. Soc.* **106,** 2701–2703.
4. Copley, S. D. & Knowles, J. R. (1985) *J. Am. Chem. Soc.* **107,** 5306–5308.
5. Lee, A. Y., Stewart, J. D., Clardy, J. & Ganem, B. (1995) *Chem. Biol.* **2,** 195–203.
6. Chook, Y. M., Ke, H. & Lipscomb, W. N. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 8600–8603.
7. Chook, Y. M., Gray, J. V., Ke, H. & Lipscomb, W. N. (1994) *J. Mol. Biol.* **240,** 476–500.
8. Lee, A. Y., Karplus, P. A., Ganem, B. & Clardy, J. (1995) *J. Am. Chem. Soc.* **117,** 3627–3628.
9. Haynes, M. R., Stura, E. A., Hilvert, D. & Wilson, I. A. (1994) *Science* **263,** 646–652.
10. Copley, S. D. & Knowles, J. R. (1987) *J. Am. Chem. Soc.* **109,** 5008–5013.
11. Gajewski, J. J., Jurayj, J., Kimbrough, D. R., Gande, M. E., Ganem, B. & Carpenter, B. K. (1987) *J. Am. Chem. Soc.* **109,** 1170–1186.
12. Addadi, L., Jaffe, E. K. & Knowles, J. R. (1983) *Biochemistry* **22,** 4494–4501.
13. Kast, P., Asif-Ullah, M. & Hilvert, D. (1996) *Tetrahedron Lett.* **37,** 2691–2694.
14. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
15. Xia, T., Zhao, G., Fischer, R. S. & Jensen, R. A. (1992) *J. Gen. Microbiol.* **138,** 1309–1316.
16. Zhao, G., Xia, T., Fischer, R. S. & Jensen, R. A. (1992) *J. Biol. Chem.* **267,** 2487–2493.
17. Gray, J. V., Golinelli-Pimpaneau, B. & Knowles, J. R. (1990) *Biochemistry* **29,** 376–383.
18. Short, J. M., Fernandez, J. M., Sorge, J. A. & Huse, W. D. (1988) *Nucleic Acids Res.* **16,** 7583–7600.
19. Yanisch-Perron, C., Vieira, J. & Messing, J. (1985) *Gene* **33,** 103–119.
20. Chang, A. C. Y. & Cohen, S. N. (1978) *J. Bacteriol.* **134,** 1141–1156.
21. Kast, P. (1994) *Gene* **138,** 109–114.
22. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
23. Maruya, A., O'Connor, M. J. & Backman, K. (1987) *J. Bacteriol.* **169,** 4852–4853.
24. Chothia, C. (1984) *Annu. Rev. Biochem.* **53,** 537–572.
25. Dube, D. K. & Loeb, L. A. (1989) *Biochemistry* **28,** 5703–5707.
26. Evnin, L. B., Vásquez, J. R. & Craik, C. S. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 6659–6663.
27. Oliphant, A. R. & Struhl, K. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 9094–9098.
28. Reidhaar-Olson, J. F. & Sauer, R. T. (1988) *Science* **241,** 53–57.
29. Liao, H., McKenzie, T. & Hageman, R. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 576–580.
30. Hermes, J. D., Blacklow, S. C. & Knowles, J. R. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 696–700.
31. Dorner, L. F. & Schildkraut, I. (1994) *Nucleic Acids Res.* **22,** 1068–1074.
32. Munir, K. M., French, D. C., Dube, D. K. & Loeb, L. A. (1992) *J. Biol. Chem.* **267,** 6584–6589.
33. Milla, M. E. & Sauer, R. T. (1995) *Biochemistry* **34,** 3344–3351.
34. Dube, D. K., Black, M. E., Munir, K. M. & Loeb, L. A. (1993) *Gene* **137,** 41–47.
35. Munir, K. M., French, D. C. & Loeb, L. A. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 4012–4016.
36. Blacklow, S. C. & Knowles, J. R. (1990) *Biochemistry* **29,** 4099–4108.
37. Kim, S., Ribas de Pouplana, L. & Schimmel, P. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 10046–10050.
38. Gong, J. & Ferreira, G. C. (1995) *Biochemistry* **34,** 1678–1685.
39. Wiest, O. & Houk, K. N. (1994) *J. Org. Chem.* **59,** 7582–7584.
40. Kast, P., Hartgerink, J. D., Asif-Ullah, M. & Hilvert, D. (1996) *J. Am. Chem. Soc.* **118,** 3069–3070.
41. Wiest, O. & Houk, K. N. (1995) *J. Am. Chem. Soc.* **117,** 11628–11639.
42. Smith, P. J. & Wilcox, C. S. (1991) *Tetrahedron* **47,** 2617–2628.
43. Lyne, P. D., Mulholland, A. J. & Richards, W. G. (1995) *J. Am. Chem. Soc.* **117,** 11345–11350.
44. Gray, J. V., Eren, D. & Knowles, J. R. (1990) *Biochemistry* **29,** 8872–8878.
45. Pourmotabbed, T., Dell'Acqua, M., Gerlt, J. A., Stanczyk, S. M. & Bolton, P. H. (1990) *Biochemistry* **29,** 3677–3683.
46. Xue, Y. & Lipscomb, W. N. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 10595–10598.
47. Hilvert, D., Carpenter, S. H., Nared, K. D. & Auditor, M.-T. M. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 4953–4955.