Probing the Role of the C-Terminus of *Bacillus subtilis* Chorismate Mutase by a Novel Random Protein-Termination Strategy†

Marianne Gamper, Donald Hilvert, and Peter Kast*

Laboratorium für Organische Chemie, Swiss Federal Institute of Technology (ETH), ETH-Zentrum, Universitätstrasse 16, CH-8092 Zürich, Switzerland

Received July 17, 2000

ABSTRACT: A novel strategy combining random protein truncation and genetic selection has been developed to identify dispensable C-terminal segments of an enzyme. This approach, which entails the random introduction of premature termination codons, was applied to the last 17 residues of chorismate mutase from *Bacillus subtilis* (BsCM). Although structurally ill-defined, the C-terminus of BsCM has been proposed to cap the active site upon substrate binding and affect catalysis. However, sequence patterns of 178 selected gene variants show that the final 11 residues of the protein can be mutated and even removed without significantly impairing activity in vivo. In fact, none of the randomized residues is absolutely required, but a preference for wild-type Lys111, Ala112, Leu115, and Arg116 is apparent. These residues are part of a C-terminal 3_{10}-helix and provide contacts with the rest of the protein or its ligands. The kinetic parameters of selected enzyme variants show that truncations and mutations do not significantly impair catalytic turnover (k_{cat}) but substantially decrease k_{cat}/K_m. Thus, while the 17 C-terminal residues of BsCM do not participate directly in the chemical rearrangement, they appear to contribute to enzymatic efficiency via uniform binding of the substrate and transition state.

Bacillus subtilis chorismate mutase (BsCM)[1] (*1*) catalyzes the rearrangement of chorismate to prephenate (Figure 1A), the first committed step in the biosynthesis of the aromatic amino acids Phe and Tyr (*2*), by more than a million-fold. The enzyme is a member of the AroH class of chorismate mutases (CMs) (*3*) and has been extensively characterized with regard to structure (*4*, *5*) and function (*1*, *6*−*14*). It is a homotrimeric pseudo-*β*-barrel surrounded by α-helices (*4*, *5*). Three solvent-accessible active sites are located at the subunit interfaces (Figure 1B). BsCM uses an extensive array of interactions to bind and orient the flexible substrate for reaction. In addition, a single cation (Arg90) appears to be essential for catalysis and presumably stabilizes developing negative charge on the substrate's ether oxygen in the transition state (Figure 1A) (*9*). However, the role, if any, of the crystallographically ill-defined C-terminal tail close to the entrance of the substrate binding pocket is unclear.

AroQ enzymes represent a second, more abundant class of CMs (*3*, *15*). Despite similar kinetic parameters (*3*), they have protein architectures entirely different from that of BsCM (*16*). AroQ proteins are all-α-helical, and their active sites are completely buried (*17*, *18*). Furthermore, these catalysts provide important interactions with the substrate's ether oxygen from two critical active site residues (*16*, *19*, *20*). In this report, we have investigated whether burying the active site following substrate binding is also important for BsCM activity (*5*, *7*). In analogy to the AroQ proteins, amino acids from the flexible C-terminal tail of the enzyme

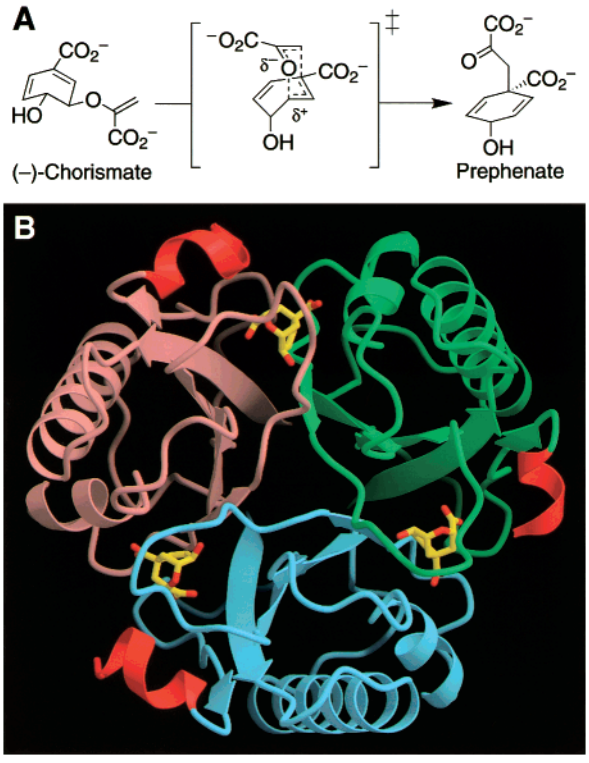

FIGURE 1: CM reaction and structure of WT BsCM. (A) Claisen rearrangement catalyzed by CM with the presumed transition state. (B) Ribbon representation of the BsCM trimer with the crystallographically resolved section of the C-terminal tails shown in red. The last resolved residue in the displayed subunits is either Leu115 or Asp118. A transition state inhibitor is bound to each active site (*4*, *5*).

† This work was supported by the Schweizerischer Nationalfonds.

* To whom correspondence should be addressed. E-mail: kast@org.chem.ethz.ch. Telephone: (+41-1) 632-2908. Fax: (+41-1) 632-1486.

[1] Abbreviations: BsCM, *B. subtilis* chorismate mutase; CM, chorismate mutase; IPTG, isopropyl 1-thio-*β*-D-galactopyranoside; H-AroH, BsCM with an N-terminal His tag; nt, nucleotide(s); PCR, polymerase chain reaction; WT, wild type.

10.1021/bi0016570 CCC: $19.00
Published on Web 10/25/2000

could conceivably cap the active site (Figure 1B) and even provide additional contacts to the substrate in the transition state. Ligand-induced conformational changes observed crystallographically for complexes of BsCM with prephenate or a transition state analogue (*5*), and inferred from Fourier transform infrared (FTIR) spectroscopic studies (*7*), are consistent with a functional role for the C-terminus, but additional information is needed to clarify the nature of its contribution to catalysis.

Site-directed mutagenesis has emerged as an invaluable tool for probing the mechanism of structurally characterized enzymes (*21*). When structural information is lacking, however, appropriate targets for mutational analysis may not be easily recognized. Here, we present a novel random mutagenesis and selection strategy for rapidly analyzing the functional significance of structurally unresolved segments at the C-terminus of an enzyme. We have used this method to assess the role of the 17 C-terminal amino acids of the 127-residue BsCM.

EXPERIMENTAL PROCEDURES

Strains, Plasmids, and General DNA Procedures. Escherichia coli strains XL1-Blue (Stratagene), KA12 (*9*, *10*), and KA13 (*3*, *22*) were used for routine cloning, in vivo complementation, and protein production, respectively. Plasmids pET-22b-pATCH (*3*), pMS15 (*23*), pKET3-W (*11*), pKSS (*24*), pKCMT-R90Q (*9*), and pKIMP-UAUC (*9*) were described previously. Jetquick spin columns (Genomed) were used for DNA purification. Other DNA manipulations were according to standard procedures (*25*). Restriction endonucleases and T4 DNA ligase were from New England Biolabs or Fermentas. Oligonucleotides were synthesized and purified by Microsynth. DNA sequencing was performed on an ABI PRIZM 310 Genetic Analyzer (PE Applied Biosystems).

Reengineering of the aroH Expression System. The *aroH* expression plasmid pMG212H-W was constructed in three consecutive cloning steps (pET-22b-pATCH → pMG208 → pMG209 → pMG212H-W). pMG208 (5050 bp) is pET-22b-pATCH with a deletion of the f1 *ori* (459 bp). It was obtained by ligating the 3764 bp *Sph*I−*Pst*I fragment of pET-22b-pATCH, the 599 bp *Kpn*2I−*Sph*I fragment of pET-22b-pATCH, and the 687 bp *Pst*I−*Kpn*2I PCR fragment generated with primers DELF1-5′ (GTTTTCCGGATTACAATTTCAGGTGGCACTTT) and DELF1-3′ (GTAGTTCGCCAGTTAATAGTTT) on template pET-22b-pATCH. The PCRs were carried out with AmpliTaq Gold DNA Polymerase (PE Applied Biosystems). pMG209 (5950 bp) is pMG208 with a fragment containing the 5′ end of *nahG*, the salicylate-inducible *nahG* promoter, and the entire, divergently transcribed *nahR* gene. To preserve a unique *Nco*I site in pMG209, the *nahGR* region was assembled from two PCRs with pMS15 as a template. The 201 or 927 bp PCR fragments were obtained with primers NAHR-5′ (GTTTGGATCCTTAGATGCGCAAGCCAAGTTT) and DENCO-3′ (CGCAGTTCCATGTGGCCTCGCTT) or DENCO-5′ (CCAAGCGAGGCCACATGGAACTGCG) and NAHR-3′ (GAATTCTAGATCAATCCGTAAACAGGTCAAA), respectively. The two overlapping PCR fragments were then linked by five PCR cycles without additional primers and subsequently amplified by another 25 cycles after addition of outside primers NAHR-5′ and NAHR-3′. After *Xba*I−*Bam*HI digestion, the 1089 bp fragment was ligated with the 4857 bp *Sph*I−*Bgl*II fragment from pMG208 in the presence of adapter LIN8 (CTAGCATG) to join the protruding *Sph*I and *Xba*I ends. pMG212H-W (6203 bp) was constructed by inserting a 414 bp *Nde*I−*Spe*I-digested PCR fragment into the correspondingly cut 5789 bp fragment of pMG209. The PCR product was generated on template pKET3-W with primers 03-NHIS (ACCGATGTCATATGCACCATCATCATCATCATTCTTCTGGTATGATTCGCGGAATTCGCGGAGCAA) and 04-T7TR (CAGCAGCCAACTCAGCTTCCTTTC) and contains an *aroH* gene which encodes WT BsCM with an N-terminal His_6 tag linked by the Ser-Ser-Gly peptide to the second Met in AroH (H-AroH).

Library PCR Templates. The two PCR template plasmids differ only at the *aroH* codon corresponding to Arg90 of BsCM; pKSS-H-Wt contains an Arg (AGA) codon, while pKSS-H-R90Qt specifies Gln (CAG). Both template plasmids (3262 bp each) are composed of the 2927 bp *Eco*RI−*Xba*I fragment from pKSS and a 335 bp *aroH*-containing, *Eco*RI−*Xba*I-digested PCR product. The PCR fragment in pKSS-H-Wt was generated with primers T7PRO2 (TAATACGACTCACTATAGGG) and 05-TEMP (GACTCTCTAGATTACAGAACCACAGCCTTTTCTAAATATACATGTCT) on template pKET3-W. For pKSS-H-R90Qt, primers PMUTS (TTCACACAGGAAACAGACCAT) and 05-TEMP were used on template pKCMT-R90Q.

Construction of aroH Libraries Containing a Partially Randomized 3′ Region. The acceptor plasmid pMG212H-0 (7032 bp) was assembled by replacing the *Eco*RI−*Spe*I *aroH* fragment in pMG212H-W with the 1200 bp *Eco*RI−*Spe*I fragment of pKSS. For construction of libraries, pMG212H-0 was digested with *Eco*RI and *Bam*HI and the 5838 bp acceptor fragment was gel-purified by Geneclean II (Bio101). BsCM's C-terminal 17 residues (Lib1) or 12 residues (Lib2) were randomized by PCR with partially degenerated primer 07-LIB1 (GGACTAGTGGATCCTTAKaaKtcKgtKttKttKgtKaaKgaKaaKtcKggKcgKaaKacKacKgcKttCTCGAGATATACATGTCTGATCTGATCCT) or 08-LIB2 (GGACTAGTGGATCCTTAKaaKtcKgtKttKttKgtKaaKgaKaaKtcKggKcgCAGAACCACAGCCTTTTCTAAAT), respectively. K stands for an equimolar T/G mixture, while lowercase a, t, c, or g indicates a mixture of 60% *aroH* WT nucleotide (nt) with 20% A and 20% T. Lib1 and Lib2 PCRs where carried out with oligonucleotide 06-T3PR (TCGAAATTAACCCTCACTAAAG) as the counter primer on either template pKSS-H-Wt or pKSS-H-R90Qt. Library PCR conditions included 70 ng of template DNA, 50 pmol of each primer, 0.2 mM dNTPs, 2.5 units of HotStarTaq polymerase (Qiagen), and 1.5 mM $MgCl_2$ (only Lib1) in 50 μL of PCR buffer (Qiagen); after 15 min at 95 °C, 25 PCR cycles (1 min at 95 °C, 1 min at 48 °C, and 1 min at 72 °C) and a final extension step (7 min at 72 °C) followed. The purified PCR products were digested with *Eco*RI and *Bam*HI. Purified insert (371 bp) and acceptor (5838 bp) DNA (0.4 pmol each) were ligated overnight at 10 °C, desalted by JETquick purification, eluted in 2 mM Tris-HCl and 0.2 mM EDTA (pH 8.0), and concentrated. One-third was electroporated into Phe and Tyr auxotrophic KA12/pKIMP-UAUC cells (*9*). Aliquots were titrated on LB Amp^{150} agar plates (LB medium, containing 150 μg/mL sodium ampicillin) for library size determination. The remaining cell suspension (90%) was grown overnight (230 rpm at 30 °C) in 100 mL

of LB Amp^{150} medium. Plasmids were purified from a 3 mL culture.

In Vivo Selection of Clones with CM Activity. Ten percent of the isolated plasmid pool was re-electroporated into fresh KA12/pKIMP-UAUC cells to eliminate clones that harbor heterogeneous plasmids (PCR or plasmid purification may result in heteroduplex DNA which can give rise to two different *aroH* genes in one cell during plasmid replication) (*9*). The cells were washed three times in 1× M9 salts (6 mg/mL Na_2HPO_4, 3 mg/mL KH_2PO_4, 1 mg/mL NH_4Cl, and 0.5 mg/mL NaCl) (*26*), and appropriate dilutions were plated onto M9c minimal agar plates (no Phe or Tyr present) or onto nonselective M9c+FY plates (containing Phe and Tyr) and incubated for 4−7 days at 34 °C. M9c minimal medium was based on 1× M9 salts and also contained 0.2% (w/v) D-(+)-Glc, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 5 μg/mL thiamine-HCl, 5 μg/mL 4-hydroxybenzoic acid, 5 μg/mL 4-aminobenzoic acid, 1.6 μg/mL 2,3-dihydroxybenzoic acid, and 20 μg/mL L-Trp. For plates, agar (15 g/L) was added. M9c and M9c+FY plates contained 50 and 100 μg/mL sodium ampicillin, respectively. M9c+FY plates additionally contained 20 μg/mL chloramphenicol, 20 μg/mL L-Phe (F) and 20 μg/mL L-Tyr (Y). Immediately before use, 100 μM IPTG (isopropyl β-D-thiogalactopyranoside) was added to M9c and M9c+FY media.

Single colonies from library platings were restreaked on LB Amp^{150} plates, and plasmid DNA from purified clones was isolated and sequenced using the BigDye Terminator Cycle Sequencing kit (PE Applied Biosystems) with sequencing primer 04-T7TR (CAGCAGCCAACTCAGCTTCCTTTC). The complementation phenotype of isolated plasmids was confirmed after retransformation of KA12/pKIMP-UAUC. Growth of single colonies on M9c and M9c+FY minimal plates was evaluated after 3 days at 34 °C relative to positive (pMG212H-W) and negative (pMG212H-0) control transformants.

Protein Purification. Plasmids carrying interesting *aroH* mutants were transferred to KA13 which carries an IPTG-inducible chromosomal gene for T7 RNA polymerase, enabling high-level gene expression from the T7 promoter on pMG212H plasmids. LB medium (500 mL) containing 150 μg/mL sodium ampicillin was inoculated and shaken at 225 rpm and 37 °C. At an OD_{600} of 0.6, IPTG was added (0.5 mM) and the incubation was continued overnight. After centrifugation, the cell pellet was resuspended in 3 volumes of sonication buffer [50 mM sodium phosphate and 300 mM NaCl (pH 8.0)], and lysozyme was added (1 mg/mL). The cells were ruptured by sonication and treated with RNaseA and DNase I (10 μg/mL each), and insoluble material was removed by centrifugation (10000*g* for 20 min at 4 °C). H-AroH protein variants were purified by affinity chromatography on a matrix containing chelated Ni^{2+} ions (*27*). The supernatant was provided with imidazole (10 mM) and loaded onto a column containing Ni−NTA agarose (Qiagen). The main H-AroH fraction was eluted with 200 mM imidazole and subsequently dialyzed against 20 mM potassium phosphate (pH 7.5). The concentration of the purified enzymes was determined by UV spectroscopy (assuming ϵ_{280} = 8370 M^{-1} cm^{-1}). The values coincide with those determined with the Micro BCA Protein Assay Reagent Kit (Pierce) using bovine serum albumin as the standard. The integrity of purified samples was verified by polyacrylamide gel electrophoresis under denaturing and native conditions using the PhastSystem (20% homogeneous gels, Amersham Pharmacia Biotech).

Chorismate Mutase Assays. Steady-state parameters (k_{cat} and K_m) were derived from the initial rates of chorismate disappearance at 274 nm (ϵ_{274} = 2630 M^{-1} cm^{-1}) or 310 nm (ϵ_{310} = 370 M^{-1} cm^{-1}). Chorismate concentrations varied between 20 μM and 6 mM in assays performed at 30 °C in 50 mM potassium phosphate buffer (pH 7.5). Curve fitting of initial velocity data was carried out with the program KaleidaGraph (Synergy Software).

RESULTS

Targeted Random Termination of Protein Synthesis. A general strategy for mutational and structural analysis of proteins based on random C-terminal truncation, coupled with in vivo selection for desired enzymatic activity, would allow for rapid assessment of the functional importance of C-terminal protein segments. As outlined in Figure 2, we have developed such a strategy to investigate the catalytic role of crystallographically unresolved C-terminal residues in BsCM (Figure 1B). Two large random-library oligonucleotides encompassing BsCM codons 111−127 (Lib1) or 116−127 (Lib2) were designed and synthesized with nucleotide precursors spiked with predetermined levels of other nucleotides (nt). The nature and extent of "contamination" were adjusted to allow for wild-type (WT) amino acids at reasonable frequencies while maximizing the fraction of TAA stops at all mutagenized codons. Of the three possible stop codons, TAA is the preferred choice (*22*, *28*) because translational read-through artifacts with Trp insertion are possible with TGA (*22*), and TAG can be partially suppressed by Gln in our selection strain (*14*). With the codon format and composition described in detail in Figure 3, the probability of encoding a protein with WT sequence is low (2.3×10^{-8} and 2.8×10^{-6}), while at least one prematurely terminating TAA stop codon should occur in 62 and 53% of all Lib1 and Lib2 clones, respectively. The libraries were assembled by PCR, subcloned, and subjected to direct genetic selection for catalytically active variants as shown in Figure 2.

Specific features implemented in the new strategy include (i) the use of acceptor vector and PCR templates devoid of functional WT *aroH* DNA to prevent inadvertent contamination of libraries with WT BsCM sequences (Figures 2 and 3), (ii) a dual promoter system with the weak salicylate promoter (P_{sal}) for in vivo expression during selection experiments and the strong T7 RNA polymerase promoter (P_{T7}) for efficient protein production after transferring the library plasmid into a T7 RNA polymerase expression strain (Figure 2), and (iii) an N-terminal His_6 tag for facile affinity purification of interesting BsCM variants.

Effect of Genetic Selection on Library Composition at the DNA Level. We examined the composition of the gene libraries before application of selective pressure. The libraries were based on template pKSS-H-Wt and typically had between 10^4 and 10^5 members. Experimentally determined nt frequencies at individual positions agreed well with expectations from design (data from 47 unselected clones, not shown). However, insertions or deletions were found in 76 and 31% of the unselected clones from Lib1 and Lib2,

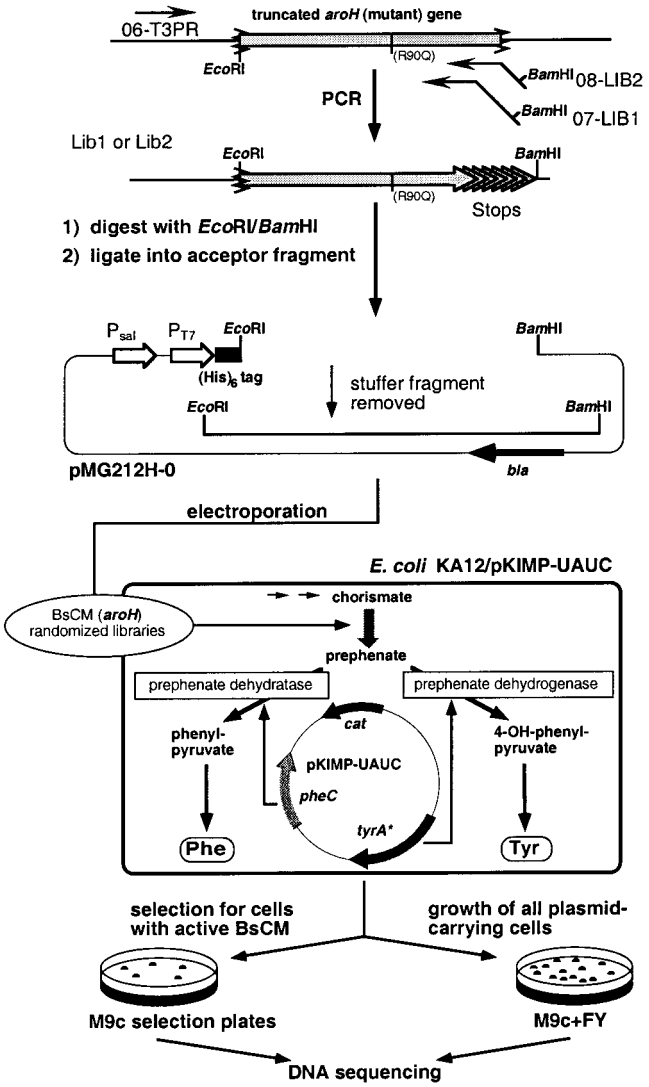

FIGURE 2: Overview of the random protein-termination analysis of the C-terminus of BsCM. Libraries (Lib1 and Lib2) encoding partially randomized C-terminal segments of BsCM were constructed by PCR. As a template, we used an *aroH* fragment of either the WT gene or a mutant encoding the R90Q exchange. Primers 07-LIB1 and 08-LIB2 contain undetermined stretches which are biased toward the WT *aroH* sequence but allow for enhanced premature termination (at TAA stops) within the last 17 and 12 C-terminal codons, respectively. The library PCR products were digested and ligated into the vector fragment of pMG212H-0. The resulting plasmid pools were introduced into KA12/pKIMP-UAUC cells which possess all enzymatic functions necessary for Phe and Tyr biosynthesis except for CM (*9*). Plating of the transformants onto M9c minimal medium agar plates allowed direct selection of *aroH* gene variants encoding catalytically proficient CMs. Library quality was assessed by sequencing clones grown on nonselective M9c+FY plates. P_{sal} and P_{T7} are the promoters used to express *aroH* for in vivo complementation and for subsequent protein isolation, respectively. *bla* and *cat* provide ampicillin and chloramphenicol resistance in the compatible plasmids of the pMG212H type and pKIMP-UAUC, respectively. *pheC* and *tyrA** encode monofunctional forms of prephenate dehydratase and prephenate dehydrogenase, respectively (*9*).

respectively (Table 1). As a consequence of resulting frameshifts, the encoded protein often terminated at a codon other than the intended TAA stop codon and occasionally even extended beyond the position of the natural stop (up to 32% of Lib1 and 12% of Lib2 clones were longer than WT BsCM). We assume that artifacts occurring during commercial chemical synthesis and workup of the long primers (97 nt for Lib1 and 76 nt for Lib2) are largely responsible for these observations (*29*). Investigation of other common cloning artifacts which often complicate construction of large random gene libraries revealed that 73 and 59% of the unselected clones from Lib1 and Lib2, respectively, lacked the *aroH* gene. Moreover, of the 23 full-length plasmids from Lib1, two were double transformants, one had a rearranged N-terminus, and one had a spontaneous point mutation (to a stop codon) within *aroH*.

The power of our direct genetic selection scheme for active CM variants efficiently removed the dysfunctional variants, including all cloning artifacts. On M9c plates, only 1.5% Lib1 and 32% Lib2 clones could grow. None of the selected plasmids had *aroH* deletions, and the fraction of Lib1 clones without frameshifts in *aroH* increased from 24 to 64%. As a result, the percentage of clones terminating at the preferred TAA increased from 50 to 79% (Table 1). It is noteworthy that frameshifts within the first five randomized Lib1 codons (positions 111−115) occur in 43% of the unselected plasmids but were tolerated in only 3% of the selected clones. In contrast, the ensemble of Lib2 clones showed little alteration in the frequency of frameshifts or the distribution of terminating stop codons after application of selection pressure (Table 1). This result is in line with in vivo tests of unselected clones in the assay strain. Only five of 19 representative Lib1 *aroH*-containing plasmids complemented the CM deficiency, while all 26 unselected Lib2 plasmids of the correct size provided growth on selection plates. Thus, no inactive clones were found when the last 12 codons were randomized, but randomization of the last 17 codons reduced the percentage of successful complementers to ~25% of all *aroH*-containing clones. This observation suggests that the five residues additionally randomized in Lib1 may contain critical determinants of activity.

Amino Acid Sequence Patterns of Active Catalysts. We have aligned the protein sequences encoded by 19 unselected and 61 selected Lib1 clones and 26 unselected and 86 selected Lib2 clones. As depicted in panels A and B of Figure 4, unselected clones contained TAA stop codons over the entire randomized segment as specified by library design. However, none of the clones from the selected libraries truncated before position 117 unless they had the leaky TGA stop codon at position 116 (Figure 4A,B). None of the *aroH* genes tested from the unselected libraries which stop upstream of this position complemented the CM deficiency. In contrast, stop codons were tolerated at any position beyond position 116 (Figure 4A,B). Thus, genetic complementation was only possible if the protein possessed at least 116 residues; additional amino acids beyond this length contribute little to in vivo efficacy.

The number of different residues at individual positions compatible with catalytic function is shown in panels C and D of Figure 4. The results are compared with the variability allowed by library design. It is evident that a smaller number of amino acids was tolerated at positions 111, 112, and 115 than would have been possible. Aside from WT Lys, only Asn and Tyr were found at position 111; Ser and Thr were found at position 112 in addition to WT Ala, and Ile and Phe were present at position 115 together with WT Leu. At all other positions, the observed variability in active clones followed from design (or was even higher as a consequence

```
amino acid number      109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128
WT amino acids         L   E   K   A   V   V   L   R   P   D   L   S   L   T   K   N   T   E   L   *
  Randomized in Lib1       ---------------------------------------------------------------------
  Randomized in Lib2                                   -----------------------------------------

Lib1 sequence          CTC GAG aaM gcM gtM gtM ttM cgM ccM gaM ttM tcM ttM acM aaM aaM acM gaM ttM TAA
  Probability for WT       .32 .36 .48 .48 .32 .36 .36 .24 .32 .48 .32 .48 .32 .32 .48 .24 .32
  Probability for TAA      .08 .02 .02 .02 .08 .02 .02 .08 .08 .08 .08 .02 .08 .08 .02 .08 .08

Lib2 sequence          TTA GAA AAG GCT GTG GTT CTG cgM ccM gaM ttM tcM ttM acM aaM aaM acM gaM ttM TAA

aroH PCR template      TTA GAA AAG GCT GTG GTT CTG TAA TCT AGA ...(vector sequence)
                       L   E   K   A   V   V   L   *
```

FIGURE 3: Library design. Codon numbers were assigned as in WT BsCM for all H-AroH variants throughout this paper. Formats and randomized sequence ranges are shown for the coding strands of the two libraries. Lowercase letters represent mixtures consisting of 60% WT nt, 20% A, and 20% T; M positions indicate 50% A and 50% C. The theoretical probabilities for the WT amino acid or a TAA stop codon were calculated for each variable triplet. The PCR template contained a truncated *aroH* gene to avoid preferential hybridization of WT-like library oligonucleotides.

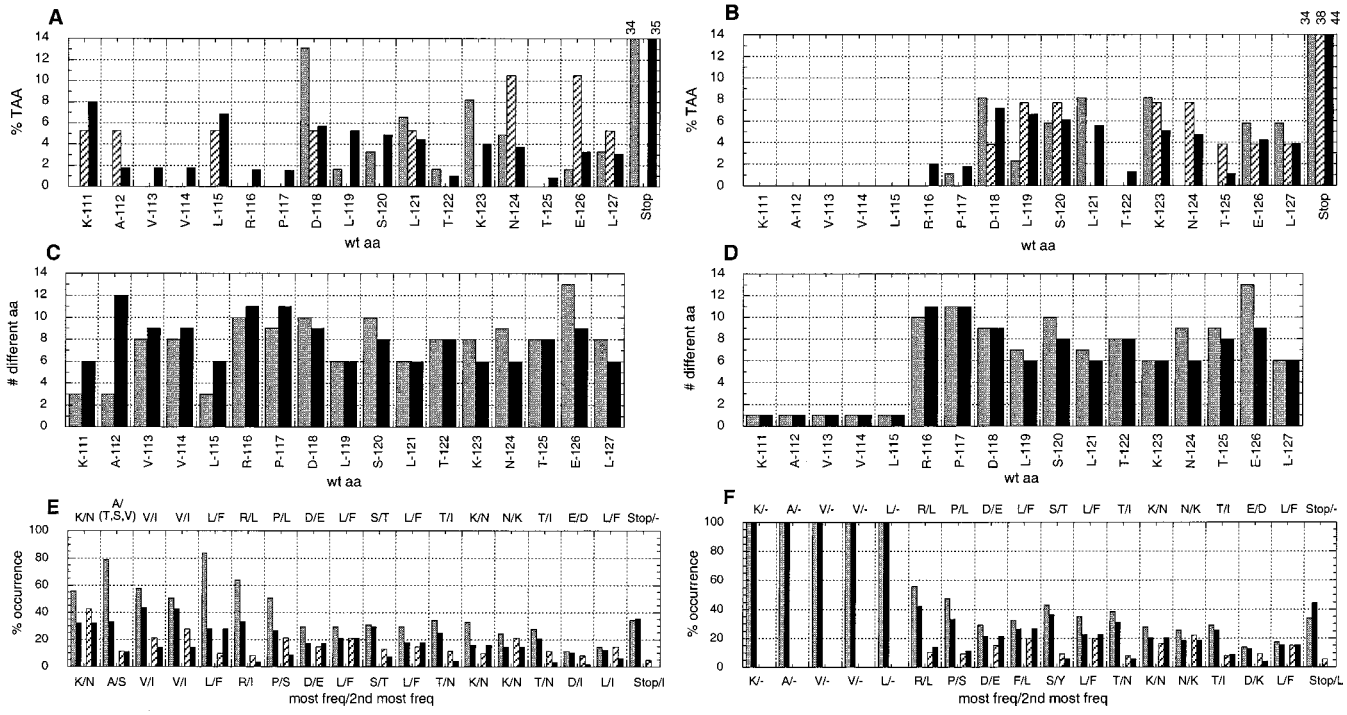

FIGURE 4: Experimentally found and statistically predicted patterns at the C-terminus of BsCM variants. Panels A, C, and E show the sequencing results from selected and unselected Lib1 clones and panels B, D, and F those from Lib2. (A and B) Distribution of terminating TAA stop codons vs the WT BsCM sequence. Column heights indicate the frequency of clones terminating at the corresponding position. Data from selected clones, unselected clones, and predictions based on library design are represented by gray, hatched, and black columns, respectively. (C and D) Experimentally found and allowed amino acid variability at individual BsCM positions. The column heights represent the numbers of different amino acids encoded by active, selected clones (gray) and those theoretically allowed by the random codon format (black). (E and F) Frequencies of the most abundant and second most abundant residues (or terminating stop codons) at individual sequence positions. Active clones with prematurely terminated sequences were included in the calculations at each position. The identities of the most frequent and second most frequent amino acids encoded by active clones are indicated on the bottom axis, and their frequencies are represented by gray and hatched columns, respectively. The adjacent black columns indicate the corresponding calculated frequencies as being derived from library codon formats. The identity of the predicted two most abundant residues is indicated above the panel.

Table 1: Library Analysis on the DNA Level

library	number of clones sequenced	fraction of clones (%)					fraction of terminating stop codons (%)		
		no frameshifts	only deletion(s)	only insertion(s)	deletion(s) and insertion(s)	disallowed nt in randomized DNA	UAA	UAG	UGA
Lib1 unselected	21	24	52	14	10	10	50	40	10
Lib1 selected	61	64	13	23	0	0	79	16	5
Lib2 unselected	26	69	23	8	0	0	85	4	11
Lib2 selected	86	72	13	15	0	1	79	16	5

of frameshifts, particularly at the end of the sequence). Thus, only positions 111 (polar), 112 (small or hydrophilic), and 115 (large and hydrophobic) are moderately conserved.

Although three different residues appear at position 112 in functional clones, WT Ala is strongly favored to an extent (79%) significantly higher than the frequency (33%) calculated for a random distribution. The next most frequent amino acids expected at this position are Thr, Ser, and Val, each at 11% abundance. Experimentally, Ser occurred in 11% and Thr in 10% of the complementing clones, but Val was not found. Analogous data for each of the randomized positions are summarized in panels E and F of Figure 4. It is obvious

Table 2: Catalytic Parameters of Some Purified BsCM Variants

H-AroH variant	highest S concentration used (mM)	C-terminal amino acid sequence[a]	k_{cat}[b] (s^{-1})	K_m[b] (mM)	k_{cat}/K_m[c] (M^{-1} s^{-1})
5-8	3.8	KAVVLR	26 ± 2	9.3 ± 1.0	2.8×10^3
1-3	4.0	NSNVLRP	30 ± 6	15 ± 4	2.0×10^3
V-7	3.9	KAVVLLT	26 ± 10	16 ± 7	1.6×10^3
4-5	5.8	KAVVLRP	22 ± 1	3.9 ± 0.2	5.7×10^3
6-9	3.4	KAVVLRP	24 ± 1	4.1 ± 0.3	5.7×10^3
5-11	5.8	KAVVLRPN	23 ± 1	4.0 ± 0.3	5.6×10^3
WT H-AroH[d]	4.0	KAVVLRPDLSLTKNTEL	47 ± 2	0.081 ± 0.011	5.8×10^5
WT AroH[e]		KAVVLRPDLSLTKNTEL	46 ± 3	0.067 ± 0.005	6.9×10^5

[a] All BsCM variants that are listed terminate with the nonleaky TAA stop codon. [b] The standard deviations are given for the values of k_{cat} and K_m. Large errors occur when, for technical reasons, the highest substrate concentration used in the enzyme assay could not be increased significantly above the deduced K_m. [c] The k_{cat}/K_m ratio was calculated from individual parameters. [d] WT H-AroH was measured under the same conditions as the mutants (310 nm; including high substrate concentrations). [e] Values for WT AroH (BsCM) were determined previously (*10*).

that the positions most biased toward the WT amino acid are 112 and 115, but preferences for WT Lys111, Arg116, and Pro117 are also apparent. At these five positions, the WT amino acid is clearly preferred over the second most abundant one. At other randomized positions, the WT residue appeared roughly at the statistically expected frequency. Thus, residue identity beyond Pro117 is evidently unimportant for CM activity in vivo.

Catalytic Parameters of Tail Variants. All proteins examined in this work had nine additional N-terminal amino acids when compared to BsCM (see Experimental Procedures for details on these H-AroH variants). The added His tag, which projects away from the trimer (perpendicular to and behind the plane of Figure 1B), does not affect the catalytic parameters (Table 2). Twenty variants with interesting C-terminal sequences were isolated, and their catalytic activity was examined. In general, modification of the C-terminal segment led to large increases in K_m of at least an order of magnitude relative to the WT value. The lowest K_m values (≈0.5 mM) were found for mutants that were similar in size to or larger than WT H-AroH. Table 2 lists the properties of a subset of short but active enzyme variants. The five enzymes had k_{cat} values between 22 and 30 s^{-1}, barely a factor of 2 below the WT value. The most efficient short proteins were encoded by clones 4-5, 6-9, and 5-11 (plasmids 4-5 and 6-9 carry different genes for the same protein). All three clones exhibited a K_m of ∼4 mM which is a factor of 50 higher than the WT value. The only difference between clones 5-11 and 4-5 (or 6-9) is an additional Asn (at position 118) which is obviously unimportant for catalysis or substrate binding. However, removal of Pro117, as in clone 5-8, increased the K_m by another factor of 2. Clones 1-3 and V-7 have, besides C-terminal truncations, substitutions of the relatively conserved WT Ala112 or Arg116 and Pro117, causing K_m to increase above 15 mM. Nevertheless, all clones in Table 2 grew essentially at WT rates on selection plates.

Intragenic Complementation of the R90Q Mutation. Mutation of the catalytically essential Arg90 to Gln90 completely inactivates BsCM (*9*). Conceivably, the R90Q defect could be complemented intragenically by a suitable change within the C-terminal segment. In particular, a cationic residue extending into the active site might interact with the ether oxygen of chorismate and thus provide the necessary transition state stabilization. To test this idea with our randomization approach, several libraries were prepared using an *aroH* gene with the R90Q mutation as a template. The fraction of complementing clones ranged from 0 to 4×10^{-5}. We sequenced five active genes obtained from several independent libraries (three with Lib1 and two with Lib2 formats). Interestingly, complementation was due not to a particular C-terminal sequence able to compensate for the missing Arg but to a spontaneous point mutation which probably was introduced during PCR. In all five cases, the mutation reverted Gln (CAG) back to Arg (CGG), albeit via a codon distinct from the WT codon (AGA). This result shows that it is extremely unlikely that any single nt exchange within the entire *aroH* gene can compensate for the loss of Arg90 in BsCM. It also indicates that the probability of intragenic complementation of the R90Q mutation by C-terminal amino acids is very low (if at all possible). More extensive mutagenesis of this segment would be necessary to explore this issue.

DISCUSSION

Structural Interpretation of the Selection Results. Residues 111−115 of BsCM adopt a 3_{10}-helix in the crystal structure (*5*) (Figure 1B). No electron density was observed for the remaining C-terminal residues (116−127) in the unliganded enzyme. However, in complexes with a transition state analogue inhibitor or with prephenate, additional interpretable electron density can be detected in some subunits in the crystals up to residue 120 (in the best case). Residues 116−120 have distinct conformations in different subunits, and some are involved in crystal packing contacts. Structural information for the remaining amino acids is entirely lacking (*5*). Ligand-induced conformational changes as inferred from the ordering of C-terminal residues in the crystals of the complexes (*5*) and from Fourier transform infrared (FTIR) spectroscopic studies (*7*) raise the possibility that the flexible C-terminal tail might play an important role in enzyme function.

Our results now show that none of the last 17 residues is essential for catalysis and that the 11 C-terminal amino acids can even be deleted without destroying CM activity. Ala112 and Leu115 are the most conserved residues in the C-terminal tail. Both amino acid side chains are oriented toward the protein interior and are engaged in intimate, mostly hydrophobic interactions with the core of the protein. Interestingly, a naturally occurring A112V mutant was reported to have little or no activity in vivo or in vitro (*7*). Consistent with

this, variants with Val112, even though allowed by design, were not found among active clones in our experiments. Besides Ala112 and Leu115, the only C-terminal residues observed in the crystal structure to have (minor) contacts with the rest of the protein are Lys111 and Arg116, in accord with the weak sequence conservation at these positions. Of all the positions that were examined, only Leu115 and Arg116 are within 2.5−4.0 Å of any ligand atom. Thus, the pattern of conservation at individual C-terminal tail positions coincides with observable contacts to ligand and the rest of the protein.

Implications for the Catalytic Mechanism. The catalytic parameters of the truncated variants show that the last 11 residues are not involved in the chemical step of catalysis. In fact, the values for k_{cat} in Table 2 are not more than 3-fold lower than the estimated rate constant for the actual chemical rearrangement in WT BsCM (*11*). Moreover, even replacement of the partially conserved residues Lys111, Ala112, and Arg116 does not affect k_{cat}. Only Leu115 was present in all the clones that were tested. Thus, sequestering the transition state from bulk solvent as seen in AroQ mutases (*17*, *18*) is not necessary for efficient catalysis of the rearrangement on the protein (the k_{cat} step), although it appears to contribute to high k_{cat}/K_m values. Residues from the C-terminal tail apparently do not help stabilize the transition state relative to the bound ground state. Moreover, none of the terminal 17 amino acids, whose side chain could potentially provide a polar interaction with the substrate's ether oxygen, is required for activity. This suggests that, unlike AroQ proteins where two residues [Lys and Gln (or Glu)] (*16*−*20*) are used to stabilize developing negative charge at the ether oxygen, members of the AroH family utilize only a single residue (i.e., Arg90 in BsCM) (*5*, *9*, *12*). Attempts to compensate for the loss of this essential cation via suitable mutations in the C-terminal tail failed. From the frequency and nature of the revertants obtained here, it is obvious that restoration of activity, if possible at all, will require either more extensive modification of the C-terminal tail or more than a single nt change elsewhere in the gene.

In so far as K_m provides information about substrate affinity (*11*, *13*), the large increase in this parameter upon truncation supports a role for the C-terminus in ligand binding. Moreover, given the nearly unchanged k_{cat} values, this segment appears to contribute to enzyme efficiency via uniform binding of both the ground state and transition state. The evolutionary conservation of BsCM residues Ala112, Leu115, Arg116, and Asp118 in the three other AroH class sequences (from *Bacillus stearothermophilus*, *Streptomyces coelicolor*, and *Synechocystis* sp. PCC6803) currently available in databases further supports a structural and/or functional role for the C-terminal extension. It was speculated previously (*5*, *7*) that an ordering of the C-terminal residues upon ligand binding may contribute to the binding energy for the reactive, but energetically disfavored and thus rare, pseudodiaxial conformer of chorismate by an induced-fit mechanism. Viscosity-variation experiments have established that WT BsCM is partially diffusion-controlled (*11*). Nevertheless, the "on" rate constants for substrate and product binding are significantly lower than is typical for diffusion-limited enzymes. This can be explained, in part, by the relatively low concentration of the reactive diaxial conformer of chorismate in solution (*11*, *30*). It is also conceivable, however, that a viscosity-sensitive conformational rearrangement of C-terminal residues during ligand binding may slow the enzyme.

Utility of the Novel Random Termination Analysis Approach. In a single experiment, application of our random termination mutagenesis strategy to the 17 C-terminal amino acids of BsCM has yielded a large amount of information about the importance of an entire protein segment and about permissible substitution patterns at individual positions within that segment. Since only viable variants are analyzed, valuable insight is gained without wasting resources on inactive proteins. Furthermore, by optimization of the frequency of the cleanly terminating TAA stop codons within the examined stretch, unambiguous results are obtained; any fragments C-terminal to a TAA stop are not required for activity, irrespective of sequence. Provided a suitable selection or an efficient screening system is available, our strategy can be generally used to probe the importance of C-terminal segments of any protein. The technique could be particularly useful for engineering minimized versions of active proteins (*31*, *32*).

NOTE ADDED IN PROOF

The complete C-terminal tail of BsCM was recently resolved crystallographically (*33*), but the nonphysiological conditions of crystallization (2.2 M ammonium sulfate at pH 3.0) and the prominent intermolecular contacts made by the terminal residues in the crystal suggest that the observed conformation of this protein segment has little relevance to function.

ACKNOWLEDGMENT

We are grateful to Sandra Zangger and to the participants of the 1999 ETH-Praktikum "Biologische Chemie: Gesteuerte Evolution von Proteinen" for their dedicated experimental work. We thank Kai Walter and Corinna Grisostomi for providing chorismate, Marc Schell for plasmid pMS15, and Kinya Hotta for preparing Figure 1.

REFERENCES

1. Gray, J. V., Golinelli-Pimpaneau, B., and Knowles, J. R. (1990) *Biochemistry 29*, 376−383.
2. Haslam, E. (1993) *Shikimic Acid: Metabolism and Metabolites*, John Wiley & Sons, New York.
3. MacBeath, G., Kast, P., and Hilvert, D. (1998) *Biochemistry 37*, 10062−10073.
4. Chook, Y. M., Ke, H., and Lipscomb, W. N. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 8600−8603.
5. Chook, Y. M., Gray, J. V., Ke, H., and Lipscomb, W. N. (1994) *J. Mol. Biol. 240*, 476−500.
6. Gray, J. V., Eren, D., and Knowles, J. R. (1990) *Biochemistry 29*, 8872−8878.
7. Gray, J. V., and Knowles, J. R. (1994) *Biochemistry 33*, 9953−9959.
8. Rajagopalan, J. S., Taylor, K. M., and Jaffe, E. K. (1993) *Biochemistry 32*, 3965−3972.
9. Kast, P., Asif-Ullah, M., Jiang, N., and Hilvert, D. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 5043−5048.
10. Kast, P., Asif-Ullah, M., and Hilvert, D. (1996) *Tetrahedron Lett. 37*, 2691−2694.
11. Mattei, P., Kast, P., and Hilvert, D. (1999) *Eur. J. Biochem. 261*, 25−32.
12. Cload, S. T., Liu, D. R., Pastor, R. M., and Schultz, P. G. (1996) *J. Am. Chem. Soc. 118*, 1787−1788.

13. Gustin, D. J., Mattei, P., Kast, P., Wiest, O., Lee, L., Cleland, W. W., and Hilvert, D. (1999) *J. Am. Chem. Soc. 121*, 1756−1757.
14. Kast, P., Hartgerink, J. D., Asif-Ullah, M., and Hilvert, D. (1996) *J. Am. Chem. Soc. 118*, 3069−3070.
15. Gu, W., Williams, D. S., Aldrich, H. C., Xie, G., Gabriel, D. W., and Jensen, R. A. (1997) *Microb. Comp. Genomics 2*, 141−158.
16. Lee, A. Y., Stewart, J. D., Clardy, J., and Ganem, B. (1995) *Chem. Biol. 2*, 195−203.
17. Lee, A. Y., Karplus, P. A., Ganem, B., and Clardy, J. (1995) *J. Am. Chem. Soc. 117*, 3627−3628.
18. Sträter, N., Schnappauf, G., Braus, G., and Lipscomb, W. N. (1997) *Structure 5*, 1437−1452.
19. Liu, D. R., Cload, S. T., Pastor, R. M., and Schultz, P. G. (1996) *J. Am. Chem. Soc. 118*, 1789−1790.
20. Zhang, S., Kongsaeree, P., Clardy, J., Wilson, D. B., and Ganem, B. (1996) *Bioorg. Med. Chem. 4*, 1015−1020.
21. Gerlt, J. A. (1994) *Curr. Opin. Struct. Biol. 4*, 593−600.
22. MacBeath, G., and Kast, P. (1998) *BioTechniques 24*, 789−794.
23. Schell, M. A., and Wender, P. E. (1986) *J. Bacteriol. 166*, 9−14.
24. Kast, P. (1994) *Gene 138*, 109−114.
25. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
26. Miller, J. H. (1992) *A Short Course in Bacterial Genetics. A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
27. Van Dyke, M. W., Sirito, M., and Sawadogo, M. (1992) *Gene 111*, 99−104.
28. Poole, E. S., Brown, C. M., and Tate, W. P. (1995) *EMBO J. 14*, 151−158.
29. Bartel, D. P., and Szostak, J. W. (1993) *Science 261*, 1411−1418.
30. Copley, S. D., and Knowles, J. R. (1987) *J. Am. Chem. Soc. 109*, 5008−5013.
31. Pessi, A., Bianchi, E., Crameri, A., Venturini, S., Tramontano, A., and Sollazzo, M. (1993) *Nature 362*, 367−369.
32. Braisted, A. C., and Wells, J. A. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 5688−5692.
33. Ladner, J. E., Reddy, P., Davis, A., Tordova, M., Howard, A. J., and Gilliland, G. L. (2000) *Acta Crystallogr. D56*, 673−683.

BI0016570