Directed evolution of ampicillin-resistant activity from a functionally unrelated DNA fragment: A laboratory model of molecular evolution

Takato Yano and Hiroyuki Kagamiyama*

Department of Biochemistry, Osaka Medical College, Takatsuki, Osaka 569-8686, Japan

To establish an experimental system to directly observe molecular evolution, a DNA fragment that confers ampicillin resistance on *Escherichia coli* was cloned from an archaeal genomic DNA. The activity of this clone was enhanced by 50 rounds of directed evolution by using DNA shuffling. Analysis of the evolved DNA fragments shows that two genetic regions have coevolved: One region, which has no obvious ORF, is essential for the activity, whereas the other, which appears to encode a protein, is not essential but enhances the activity of the former region. Analysis of the evolutionary intermediates shows that negative mutations are effectively removed while beneficial mutations accumulate and illustrates how a protein has evolved over the course of the evolution experiments. Although the mechanism of the activity remains unclear, the evolved DNA fragments also confer resistance to other drugs that inhibit bacterial cell-wall synthesis. The present system would serve as an experimental model to study evolutionary dynamics in the laboratory and provide the concept of screening natural libraries to obtain starting materials for directed evolution.

Molecular evolution has been studied mainly through theoretical or comparative analysis. Some important questions, however, cannot be fully answered unless we observe directly the process of molecular evolution. For example, how rapidly do new functions of genes evolve? Do the functions keep improving or reach their upper limits and, if in the latter case, what determines the limits? Are there any particular mutations that change significantly the subsequent evolutionary courses? How reproducible are these features of evolution under different conditions? As a first step in answering these questions, we set out to establish an experimental system designed to allow us to monitor the details of molecular evolution.

Directed evolution, consisting of multiple rounds of mutagenesis, screening, and amplification of selected mutants, has been used to modify the properties of existing proteins. Although there have been some impressive successes in protein engineering (1–5), these experiments, using only several rounds of screening and selection at best, simulate the very initial phase of a terminal branching of an evolutionary tree of genes. To achieve more drastic molecular evolution *in vitro*, two criteria were applied to our directed evolution experiment. First, the evolved function is unrelated to, not just a slight modification of, the original biological function of the gene. Second, the number of rounds of screening and selection is high enough to yield a statistically reliable evolutionary trajectory.

We devised the following experimental system to satisfy these criteria: an expression library of *Pyrococcus furiosus* genes is screened for a gene that endows *Escherichia coli* with ampicillin-resistant (amp^R) activity. After the amp^R activity is enhanced by directed evolution, evolutionary intermediates as well as finally evolved mutants are analyzed. *P. furiosus* is a hyperthermophilic archaeon isolated from a marine hydrothermal vent (6). Archaeal cell wall is chemically different from the bacterial peptidoglycan, and therefore archaea are not susceptible to common antibiotics directed against cell-wall synthesis, including β-lac- tam antibiotics such as ampicillin (7). Understandably, no β-lactamase activity has been detected in archaea (8). Thus, the original biological function of the *P. furiosus* gene, even though it shows amp^R activity when expressed in *E. coli*, should be unrelated to peptidoglycan metabolism or inactivation of β-lactam antibiotics. Although the number of rounds of screening and selection we can achieve cannot be predicted, this model system might allow us to simulate molecular evolution in the laboratory.

Materials and Methods

Cloning of amp^R Gene Fragments. *P. furiosus* chromosomal DNA was partially digested with *Sau*3AI, and 1.5- to 3.5-kb fragments were gel purified and ligated with *Bam*HI-digested and BAP-treated pHSG398 (Takara Shuzo, Kyoto) to transform *E. coli* JM109. About 1.5×10^5 transformants were screened on LB agar plates containing 3.5 μg/ml ampicillin and 1 mM isopropylthio-β-D-galactoside (IPTG). Many false-positive colonies appeared at this low ampicillin concentration. All of the colonies were scraped off with sterile water, and a mixture of plasmids was prepared. JM109 cells were retransformed with this plasmid mix and screened on the same plates, and a plasmid mix was prepared again. This procedure was repeated four times in total to "enrich" plasmids with a low amp^R activity. Agarose-gel electrophoresis of the final plasmid mix showed several distinct bands on the background smear of DNA. Eight colonies were picked from the transformants of the final plasmid mix and cultured separately, and the plasmids were prepared. The plasmids contained 3 kinds of insert DNAs (1.8, 2.5, and 2.6 kb). Restriction analysis showed that these DNA fragments cover the same genomic region, and the 1.8-kb DNA was sequenced.

Directed Evolution. DNA shuffling was done as described (1). Fragments of 100 to 300 bp were used to reassemble the 1.2-kb DNA. The rate of mutations per 1 DNA shuffling was 0.3%. In the first to fourteenth rounds, *Kpn*I and *Hin*dIII sites were incorporated in the 5′ and 3′ primers used for PCR, respectively. The shuffled fragment was digested with these enzymes, gel purified, and ligated with pHSG398. This puts the coding region of the shuffled fragment downstream of the *lacZ* promoter of the plasmid. *E. coli* JM109 cells were transformed with the ligated DNA by electroporation. A library of 1×10^6–3×10^6 transformants was screened on eight 9-cm LB agar plates containing an appropriate concentration of ampicillin and 1 mM IPTG

BIOCHEMISTRY

(except for the 14th round). After an incubation at 37°C for 19–24 h, about 100 largest colonies were picked up and spotted on an LB agar plate containing 30 μg/ml chloramphenicol, and the plate was incubated at 37°C for 16–19 h. The cells were scraped off of the spotted plate with sterile water, and a mixture of plasmids was prepared. This screening was done twice for each round; thus, about 200 colonies were selected from a library of 2×10^6–6×10^6. The plasmid mix obtained was used for the next round of DNA shuffling and also to retransform JM109 cells to determine the ampicillin concentration for the next screening. In the 15th to 50th rounds, *Aat*II and *Pst*I sites were incorporated at the 5′ and 3′ ends of the shuffled fragment, respectively, and the fragment was ligated into pBR322 at these restriction sites. This construction puts the shuffled fragment in place of the amino-terminal 60% of the β-lactamase gene of pBR322. The 5′ region of the shuffled fragment would contain a promoter sequence, because this plasmid shows significant amp^R activity even with no promoters in the adjacent region of the vector. The screening procedures were basically the same as those with pHSG398 except that colonies were spotted on plates containing 12.5 μg/ml tetracycline instead of chloramphenicol. The primer pairs used for DNA shuffling are as follows: 5K·3, 5′-GACT-GTACGGGTACCGAACTCCTTTCTATAGACAAAGC-GC-3′ and d3H, 5′-GACTGTACGAAGCTTCCTCCTATC-CTCGTAACGCTGTACC-3′; and 5A, 5′-GACTGTACGGA-CGTCGAACTCCTTTCTATAGACAAAGCGC-3′ and 3P, 5′-GACTGTACGCTGCAGCCTCCTATCCTCGTAACGC-TGTACC-3′.

Plasmid Constructions. The full-length, coding, and 3′ regions of AR50–3 were amplified by PCR by using primer pairs 5K·3 and d3H, 5K·3 and 3X·2, and 5X·2 and d3H, respectively. 3X·2 is 5′-GACTGTACGTCTAGAGTTCCATAGTGCATCGG-TATATCGT-3′, and 5X·2 is 5′-GACTGTACGTCTAGATT-TCTCACCATTAGCTAGGATCCAG-3′. A *Xba*I site is incorporated into 3X·2 or 5X·2 primer. The amplified fragments were ligated to pHSG398 at the corresponding restriction sites to construct pHSG-AR50–3, pHSG-AR50–3/cd, and pHSG-AR50–3/3′. The 3′ region of AR0 was also amplified by using primers; 5X·6, 5′-GACTGTACGTCTAGATTTCCCAC-CAATAGCTAGGGATCCA-3′ and d3H. The amplified 3′ fragments of AR0 and AR50–3 were ligated to pHSG-AR50–3/cd to construct pHSG-AR50–3/cd + AR0/3′ and pHSG-AR50–3/cd + AR50–3/3′, respectively. Tyr-95 was mutated to a stop codon by two-step PCR by using the following primers: 5K·3; rY95st, 5′-CAGCTGGGCCTTAGCTCATTC-3′; fY95st, 5′-GAATGAGCTAAGGCCCAGCTG-3′; and 3X·2.

Results and Discussion

P. furiosus Genomic DNA Fragments with amp^R Activity.
By screening *E. coli* cells transformed with a library of plasmids carrying *P. furiosus* chromosomal DNA fragments, three kinds of plasmids with inserts of different lengths were obtained that had very low but significant amp^R activity. Restriction analysis showed that all three DNA fragments cover the same genomic region, and the nucleotide sequence of the shortest fragment (1.8 kb) was determined. The fragment contained an ORF encoding a 226-aa protein. Database search identified a class of highly conserved archaeal proteins of unknown biological function. Eukaryotic and bacterial proteins were also identified, most of which showed homology with a particular domain of the ORF. Interestingly, among the list was the zinc-binding domain of metallo-β-lactamases (9, 10). Surprisingly, this ORF was not sufficient to confer ampicillin resistance, and further experiments confirmed that a 1.2-kb fragment, which includes a downstream region as well as the ORF, was necessary.
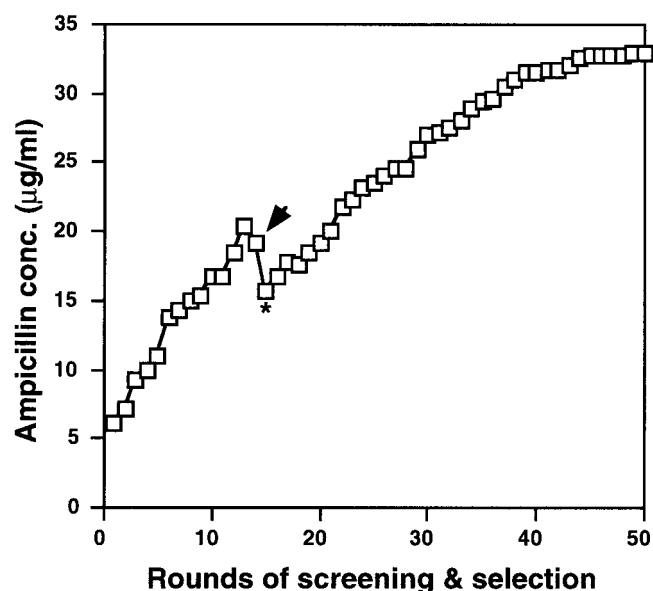

Fig. 1. Trajectory for amp^R activity during 50 rounds of directed evolution. Each point represents the ampicillin concentration at which the screening was done. To set the ampicillin concentration for a given screening, JM109 cells were transformed with a plasmid mix obtained from the previous round of screening, and the cell growth was examined at various concentrations of ampicillin. At every round, the screening was done at as high an ampicillin concentration as possible. At the 14th round (arrow head), IPTG was omitted from the medium. At the 15th round (asterisk), the vectors were changed from a very high-copy plasmid, pHSG398, to a moderately high-copy plasmid, pBR322.

Enhancement of amp^R Activity by Directed Evolution.
Having isolated a starting archaeal DNA fragment, this fragment was subjected to 50 rounds of directed evolution (Fig. 1). The directed evolution experiment was done as follows: the 1.2-kb fragment was treated with DNA shuffling (1) to introduce random mutations (and also recombinations after the second round). Recombination is a key feature of DNA shuffling because it facilitates the accumulation of functionally positive mutations while eliminating negative mutations (3). The shuffled DNA fragment was ligated to a plasmid vector, and the resulting construct was introduced into *E. coli* JM109 cells by using electroporation. At each round of screening, a library of 2×10^6–6×10^6 transformants was screened at an appropriate ampicillin concentration, and a mixture of plasmids was prepared from the about 200 largest colonies. The 1.2-kb fragment was amplified from the plasmid mix by using PCR, and a subsequent round of DNA shuffling was done to introduce further mutations and recombination events among the selected mutant genes. These procedures were repeated at increasing concentrations of ampicillin. At the 14th round, the inducer IPTG was omitted from the selection medium, and at the 15th round, the vector pHSG398 (about 500 copies/cell) was replaced with pBR322 [about 20 copies/cell (11)]. These procedures increased selection pressure by decreasing the expression level of the gene and thus enabled us to do further rounds of directed evolution. The evolutionary trajectory thus obtained was not linear but hyperbolic, as is most evident in the 15th to 50th rounds. That is, the evolution decelerated gradually as it proceeded.

Analysis of the Evolved DNA Fragments.
To evaluate the result of evolution, 5 clones that formed the largest colonies at 31 μg/ml ampicillin were selected from colonies retransformed with a plasmid mix obtained at the 50th round, and their entire evolved

Fig. 2. Mutations introduced into an evolved DNA fragment after 50 rounds of directed evolution. The sequence of the original ''preevolved'' DNA, AR0, is aligned with that of AR50–3. PCR primers used for DNA shuffling anneal to sequences just outside the region shown here. The mutated nucleotide positions are color shaded: missense mutations in the truncated coding region are in blue, and other mutations are in red. Asterisks indicate mutations found in all five AR50 mutants. Start and stop codons are underlined. The nucleotide position, at which the coding and 3′ regions of AR50–3 are separated and the coding region of AR50–3 was linked to the 3′ region of AR0 in Fig. 3, is shown by an arrowhead.

regions were sequenced. The nucleotide sequence of one of those AR50 mutants is compared with that of the original DNA fragment, AR0, in Fig. 2. Each AR50 mutant has 107–129 mutations, including one or two single-base deletions. All five AR50 mutants have a nonsense mutation at nucleotide position (np) 673, resulting in a 42-aa deletion at the carboxy terminus of the ORF. In the truncated coding region, each AR50 mutant has 22–24 missense mutations, 19 of which are found in all of the 5 mutants, whereas only 6 of 20–32 silent mutations are such ''conserved'' mutations. This finding indicates that this ORF is translated into a protein and is important for amp^R activity. The distribution of mutations in the 3′ downstream region is clearly different from that in the coding region and suggests that this region is nonfunctioning. In the 834- to 1,127-np region, how-
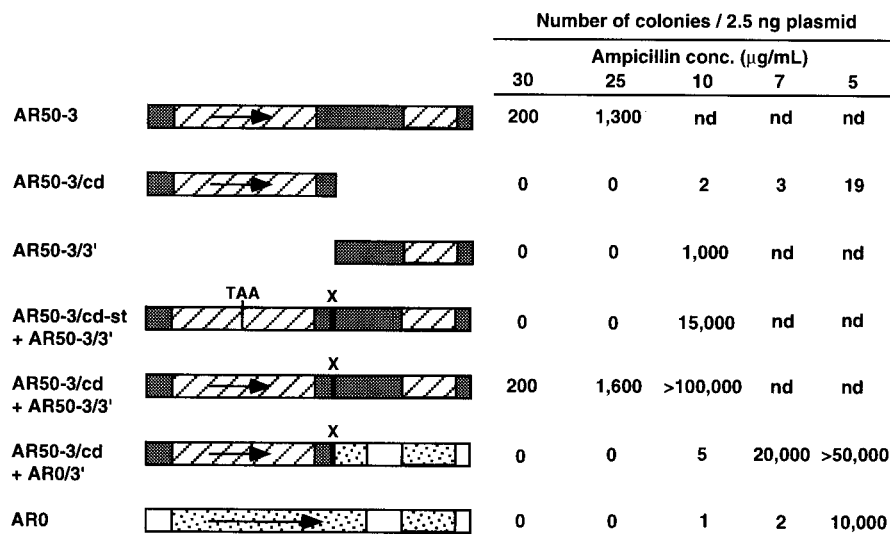
		Number of colonies / 2.5 ng plasmid				
		Ampicillin conc. (μg/mL)				
		30	25	10	7	5
AR50-3		200	1,300	nd	nd	nd
AR50-3/cd		0	0	2	3	19
AR50-3/3'		0	0	1,000	nd	nd
AR50-3/cd-st + AR50-3/3'		0	0	15,000	nd	nd
AR50-3/cd + AR50-3/3'		200	1,600	>100,000	nd	nd
AR50-3/cd + AR0/3'		0	0	5	20,000	>50,000
AR0		0	0	1	2	10,000

Fig. 3. amp^R activity of various regions of the DNA fragment. The regions shown here were put downstream of the *lacZ* promoter of pHSG398, which is located on the left side of the figure. The coding regions and the 3′ regions where many conserved mutations were found are boxed: those of AR50–3 are hatched, and those of AR0 are dotted. Other regions derived from AR50–3 are shaded. Arrows indicate the direction of ORFs; X, *Xba*I sites incorporated during the plasmid constructions; TAA, the position of an inframe stop codon, Tyr-95 (TAC) to stop (TAA). amp^R activity is shown by the number of colonies obtained when *E. coli* JM109 cells were transformed with 2.5 ng of each plasmid and screened on an LB agar plate containing the ampicillin concentration indicated (nd, not determined). Colony numbers varied ± 30% between experiments. Addition of 0.7 mM IPTG increased amp^R activity of pHSG-AR50–3/3′ (to 11,000 colonies at 10 μg/ml ampicillin). The results of pHSG-AR50–3/cd were almost the same as those of a negative control by using pHSG398.
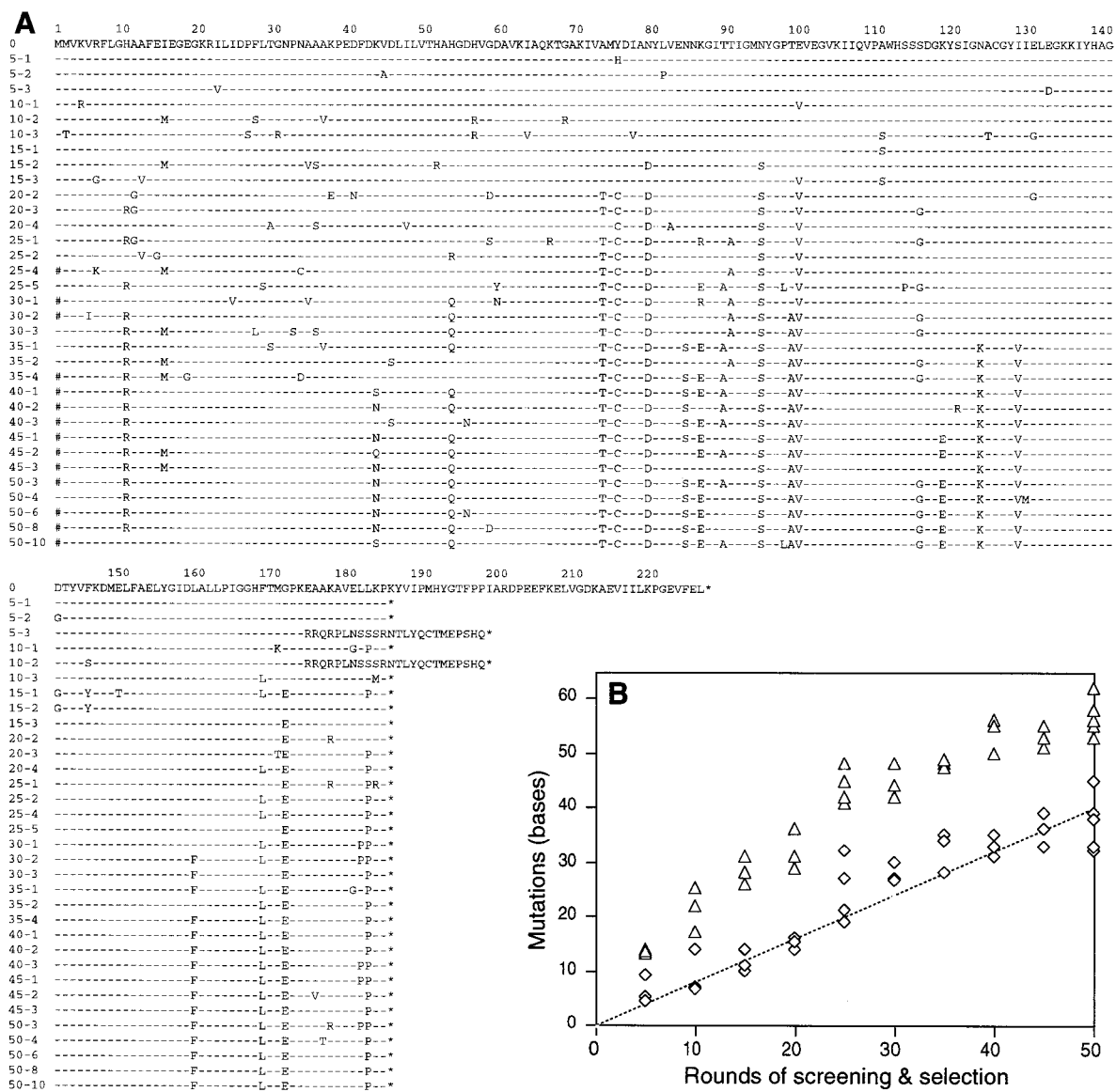
Fig. 4. Analysis of evolutionary intermediates. Entire evolved DNA fragments were sequenced for three clones at every five rounds. Four clones were sequenced for AR25 mutants. (*A*) Predicted amino acid substitutions in the coding region. Asterisks are stop codons and #, mutations in the original initiation methionine. (*B*) Each point represents the number of mutations found in a mutant. Mutations in the essential 3′ region (676–1,175 np) are triangles. The sum of silent mutations in the truncated coding region (121–675 np) and mutations in the 5′ region (1–120 np) is used as a molecular clock (diamonds). A line with a slope of 0.8 mutations/round is tentatively overlaid (see text).

ever, 33 of 41–44 mutations are conserved in the 5 AR50 mutants. Such a high ratio of conserved mutations indicates that this region was also under selection pressure during the evolution experiments and plays some important role in the amp^R activity. This is consistent with the earlier finding that the 3′ downstream region of AR0 could not be deleted without disrupting amp^R activity.

Two Regions Have Coevolved. To determine whether the 3′ region is really necessary, a series of plasmids was constructed by using AR50–3, and their amp^R activities were examined (Fig. 3). First, the coding and 3′ regions were introduced separately into *E. coli* cells. The coding region shows no detectable amp^R activity, although the production of the truncated protein was confirmed by SDS/PAGE of the crude lysate (data not shown). DNA sequencing showed that one of the putative zinc-binding residues, His-53, was mutated to glutamine in all of the 5 AR50

mutants, and that in AR50–6 another ligand residue, Asp-55, was mutated to asparagine. Thus, it seems that the homology with metallo-β-lactamases is spurious. Surprisingly, however, the 3′ region of AR50–3 shows significant amp^R activity, although the activity is lower than that of full-length AR50–3. Moreover, the coding region of AR50–3 is more potent than that of AR0 in enhancing the amp^R activity of the 3′ region of AR0. These findings clearly indicate that two genetic regions have coevolved: the 3′ region is essential for the amp^R activity, whereas the coding region enhances the activity of the 3′ region. To confirm that the coding region needs to be translated into a protein to exert its enhancing activity, a stop codon was introduced into the middle of the AR50–3 ORF. The activity of this construct is, as expected, lower than that of AR50–3 with its intact coding region. What is the product of the 3′ region? Addition of IPTG increased the amp^R activity of the 3′ region when the 3′ region was placed downstream of a *lacZ* promoter, which means this

region needs to be transcribed (in the same direction as the coding region). The product could be a peptide or an RNA, although the distribution pattern of mutations in this region does not look like that of a peptide-coding region. The exact activity of this gene fragment also remains unclear: so far, neither β-lactamase, kinase, nucleotidyltransferase, nor acetyltransferase activity for ampicillin has been detected. *E. coli* cells carrying pHSG398-AR50–3 showed resistance to other drugs directed against different steps of bacterial cell-wall synthesis, cefalexin (one of cephalosporins), fosfomycin, and D-cycloserine (up to 50, 10, and 20 μg/ml, respectively), but no resistance to inhibitors of bacterial protein synthesis, tetracycline and kanamycin. This gene fragment might function by some novel mechanism.

How Did the Evolution Proceed? To analyze the course of evolution, the 1.2-kb fragment was characterized for 3 clones (about 1.5% of each population) at every 5 rounds. A close look at the coding region offers an overview of how the protein evolved (Fig. 4*A*). The introduction of a premature stop codon was one of the first, probably positive, mutations acquired by the protein. Interestingly, another strategy to truncate the protein was tried in an early phase of the evolution: a single base deletion in the codon of Met-170 caused a frame shift and a 29-aa deletion. Although only a small portion of each population was examined, more than 40% of the 184 amino acid residues was mutated at least once. Most of the mutations disappeared soon after their introduction, indicating that these are neutral or negative mutations. Some mutations fixed very quickly once they had appeared in the population; other mutations, for example Ser-115→Gly and Phe-168→Leu, took longer to get fixed; and another group of mutations, for example Ile-15→Met and Thr-89→Ala, appeared relatively early in the evolution but never became widespread in the population. These might be highly positive, moderately positive, and slightly positive mutations, respectively. As a whole, it appears that the variety of mutations in a population has been decreasing as the evolution proceeds; in other words, the protein is converging to that with a particular set of mutations after having tried many possibilities, and it is likely that most of the mutations found in AR50 mutants are positive mutations. The mutations in the essential 3′ region show similar convergence (not shown), and the number of mutations vs. rounds of screening and selection plot gives a hyperbolic appearance (Fig. 4*B*). These features are basically similar to, although not as obvious as, those of the coding region, indicating these two regions have evolved in a similar fashion. More detailed descriptions about individual mutations must await further analysis of the gene products, which is now in progress.

It turned out that the evolved DNA does not contain a segment that is functionally neutral and long enough to be used as a molecular clock. So the sum of silent mutations in the 552-bp coding region and mutations in the 120-bp 5′ upstream region was plotted in Fig. 4*B*. A caveat is that these mutations might not be totally neutral because silent mutations change codon usage, and therefore could affect translation efficiency. Moreover, the 5′ region might contain a promoter sequence. In Fig. 4*B*, a line with a slope of 0.8 mutations/round is tentatively overlaid to visualize linearity. The probability that a point mutation introduced into the coding region is a silent one is 0.22 when the amino acid composition is taken into consideration. The mutation rate of DNA shuffling is 0.3% in this experiment. Thus, the rate of accumulation of silent mutations is 0.36 mutations/round ($552 \times 0.3/100 \times 0.22$), and the mutation rate in the 5′ region is 0.36 mutations/round ($120 \times 0.3/100$). If these mutations are assumed to be totally neutral, the calculated slope is 0.72 mutations/round. Because recombinations would increase the slope, the calculated value is well in accordance with the observed value of 0.8 mutations/round. This means that mutations are steadily being introduced while the evolution is approaching a saturation.

Genes Run out of Positive Mutations. The apparent saturation phenomenon in molecular evolution would be explained by four possibilities: (*i*) The effect of positive mutations is counteracted by the gradual accumulation of negative mutations; (*ii*) the rate of the accumulation of positive mutations decreases as evolution proceeds, that is, the gene runs out of positive mutations; (*iii*) the contribution of each newly introduced positive mutation to the total effect decreases as mutations accumulate; (*iv*) some biological factor sets an upper limit to the activity regardless of the evolutionary capacity of the gene itself. In the fourth case, evolved genes should still have various combinations of positive mutations (have not yet been converged) when the activity approaches a saturation. The data in Fig. 4 conform most reasonably to the second explanation. When a gene that has already evolved a specialized function in an "ecosystem" of genes in the chromosome of an organism is duplicated or introduced into another organism, the gene gains an opportunity to explore evolution again toward a new "niche." Such events have been occurring repeatedly in the real world, as revealed by the presence of many gene families (12). Our experimental system enables us to directly observe how molecular evolution, or "speciation" of genes, proceeds in such a case and would complement theoretical or comparative studies on molecular evolution.

1. Stemmer, W. P. C. (1994) *Nature (London)* **370,** 389–391.
2. Moore, J. C. & Arnold, F. H. (1996) *Nat. Biotechnol.* **14,** 458–467.
3. Zhang, J.-H., Dawes, G. & Stemmer, W. P. C. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 4504–4509.
4. Yano, T., Oue, S. & Kagamiyama, H. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 5511–5515.
5. Hoseki, J., Yano, T., Koyama, Y., Kuramitsu, S. & Kagamiyama. H. (1999). *J. Biochem.* **126,** 951–956.
6. Fiala, G. & Stetter, K. O. (1986) *Arch. Microbiol.* **145,** 56–61.
7. Kandler, O. & König, H. (1998) *Cell. Mol. Life Sci.* **54,** 305–308.
8. Martin, H. H. & König, H. (1996) *Microb. Drug Resist.* **2,** 269–272.
9. Carfi, A., Pares, S., Dúee, E., Galleni, M., Duez, C., Frère, J. M. & Dideberg, O. (1995) *EMBO J.* **14,** 4914–4921.
10. Concha, N. O., Rasmussen, B. A., Bush, K. & Herzberg, O. (1996) *Structure (Cambridge, U.K.)* **4,** 823–836.
11. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
12. Henikoff, S., Greene, E. A., Pietrokovski, S., Bork, P., Attwood, T. K. & Hood, L. (1997) *Science* **278,** 609–614.

BIOCHEMISTRY