J Mol Evol (2003) 56:162–168
DOI: 10.1007/s00239-002-2389-y

JOURNAL OF MOLECULAR EVOLUTION

Can an Arbitrary Sequence Evolve Towards Acquiring a Biological Function?

Yuuki Hayashi,[1] Hiroshi Sakata,[1] Yoshihide Makino,[1] Itaru Urabe,[1] Tetsuya Yomo[1,2,3,4]

[1] Department of Biotechnology, Graduate School of Engineering, Osaka University, 2-1 Yamada-oka, Suita City, Osaka 565-0871, Japan
[2] Department or Bioinformatic Engineering, Graduate School of Information Science and Technology, Osaka University, 2-1 Yamada-oka, Suita City, Osaka 565-0871, Japan
[3] Intelligent Cooperation and Control Project, PRESTO, JST, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan
[4] Department of Pure and Applied Sciences, University of Tokyo, Komaba, Meguro-ku, Tokyo 153-8902, Japan

Received: 15 April 2002 / Accepted: 6 September 2002

Abstract. To explore the possibility that an arbitrary sequence can evolve towards acquiring functional role when fused with other pre-existing protein modules, we replaced the D2 domain of the fd-tet phage genome with the soluble random polypeptide RP3-42. The replacement yielded an fd-RP defective phage that is six-order magnitude lower infectivity than the wild-type fd-tet phage. The evolvability of RP3-42 was investigated through iterative mutation and selection. Each generation consists of a maximum of ten arbitrarily chosen clones, whereby the clone with highest infectivity was selected to be the parent clone of the generation that followed. The experimental evolution attested that, from an initial single random sequence, there will be selectable variation in a property of interest and that the property in question was able to improve over several generations. fd-7, the clone with highest infectivity at the end of the experimental evolution, showed a 240-fold increase in infectivity as compared to its origin, fd-RP. Analysis by phage ELISA using anti-M13 antibody and anti-T7 antibody revealed that about 37-fold increase in the infectivity of fd-7 was attributed to the changes in the molecular property of the single polypeptide that replaced the D2 domain of the g3p protein. This study therefore exemplifies the process of a random polypeptide generating a functional role in rejuvenating the infectivity of a defective bacteriophage when fused to some preexisting protein modules, indicating that an arbitrary sequence can evolve toward acquiring a functional role. Overall, this study could herald the conception of new perspective regarding primordial polypeptides in the field of molecular evolution.

Key words: Directed evolution — Displayed phage — Domain substitution — g3p protein — Infectivity — Population size — Random polypeptide

Introduction

Functional natural proteins have evolved through iterative processes of mutation and selection. Some novel protein catalysts have been shown to evolve by gene duplication (Jensen 1976), or combination of pre-existing independent protein modules by DNA shuffling (Gilbert 1978; Tonegawa et al. 1978; Go 1981; Barnett et al. 2000). Nevertheless, the function of the duplicated genes or the roles of each of the shuffled pre-existing domains are still associated to their original characteristics. Considering the emergence of a new protein or domain with a functional role, regardless of the mechanism involved, there could be some exceptional cases of evolution where a protein encoded by the duplicated gene or a polypeptide fused as a domain to other pre-existing protein domains may have changed its original role and may elicit a new function. How much then is the possibility that a single polypeptide can evolve into a new protein, or

Correspondence to: T. Yomo; *email:* yomo@bio.eng.osaka-u.ac.jp

domain with a functional role that is totally unrelated to its original feature? To regard an emergent function as "new" or "novel," the amino acid sequence of the duplicated protein or that of the inserted or fused polypeptide must be arbitrary. Otherwise, the evolved function cannot be accredited as "new," and will just be an extension of the original function.

To investigate the possibility that arbitrary sequences can evolve into new proteins or domains with biological functions, several experimental approaches have been applied, in which artificial random sequences were employed as models of the arbitrary sequences. Recently, Keefe and Szostak (2001) reported that roughly one in 10^{11} of random polypeptides containing 80 contiguous random amino acids was an ATP-binding protein. On the contrary, Yamauchi et al. (2002) reported a much higher frequency of finding random sequences that are evolvable towards functional proteins. That is, a primordial esterase was shown to evolve from only ten variants of random polypeptides. Evolving a new functional sequence from any available arbitrary sequence hence seems promising. In this study, we explored the possibility of evolving a random polypeptide towards acquiring a new functional role totally unrelated to its original feature, when it exists as a domain fused with other pre-existing domains.

The ability of a sequence to evolve with regards to its function is intrinsically quantitative (Brookfield 2001) and is defined as the capability to generate heritable, selectable variation (Kirscher and Gerhart 1998). When selectable variation is the issue, the selective property of the members of a large or small population being explored and exploited must be well diversified such that the differences in the property are significant. Can such selectable variation be found when the evolution of a polypeptide is initiated with a single random sequence? If selectable variation is present, will the selective function improve over several generations? To address these issues, we made use of the random polypeptide sequence RP3-42 (Prijambada et al. 1996) and the tetracycline resistant variant of coliphage, fd-tet (Zacher et al. 1980), as the initial materials. In addition, we deliberately limited the population of the arbitrarily chosen mutants for each generation to a maximum of ten, to examine closely the competency of selectable variation.

RP3-42 is a soluble random polypeptide of 139 amino acid residues with no secondary structure (Yamauchi et al. 1998). It was arbitrarily chosen from five of the soluble random polypeptides found in the library prepared previously (Prijambada et al. 1996). The five soluble polypeptides have no homology with known natural proteins in SwissProt database as analyzed by BLAST 2.2.2. Hence, RP3-42 is fit to be an arbitrary sequence representing the 'origin' or initial material of the intended evolutionary study.

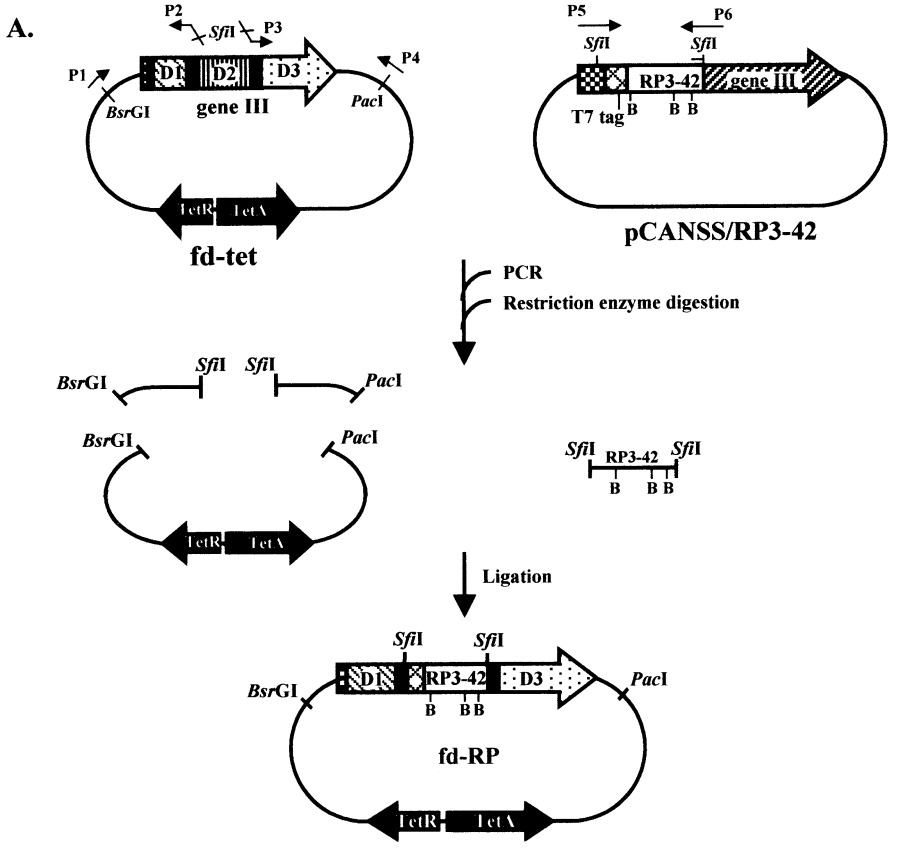
The infectivity of fd-tet was crippled by completely replacing the D2 domain of the g3p minor coat protein (Marvin, 1998) with the random sequence of RP3-42 (Prijambada et al. 1996). The coliphage fd mediates infection of *Escherichia coli* via the minor coat protein g3p (Marvin 1998, Riechmann and Holliger 1997), which consists of three distinct domains connected with flexible glycine-rich linker sequences (Riechmann and Holliger 1997). The D2 domain, wedged between the N-terminal D1 and C-terminal D3 domains, functions for the adsorption of g3p to the tip of the host F-pilus at the initial stage of the infection process (Deng et al. 1999). The phage that is defective in the D2 domain has about seven-order magnitude lower infectivity than the wild-type (Riechmann and Holliger 1997).

Here, we showed that a random polypeptide had generated its functional role in rejuvenating the infectivity of a defective bacteriophage when fused to some preexisting protein domains. We confirmed that from a single random sequence, there will be selectable variation in a property of interest even within a small mutant library, and that the selective property is evolvable through iterative mutation and selection within the time frame of our evolutionary study. Therefore, the results evidently show that a single arbitrary sequence can evolve toward a functional domain. This study hence provides a new prospect for revealing information that could open a new gateway for understanding evolution, especially when its primordial stages are concerned.

Materials and Methods

Materials. The bacterial strains used in this study were *Escherichia coli* K12 strains TG1 [*supE*, *hsd*Δ5, *thi*, Δ(*lac-proAB*) / F′[*traD*36, *ProAB*$^+$, *lac*1^q, *lacD*ΔM15]], JM109 [*recA*1, *endA*1, *gyrA*96, *thi*, *hsdR*17, *supE*44, *relA*1, Δ(*lac-proAB*) / F′[*traD*36, *ProAB*$^+$, *lac*1^q, *lacZ*ΔM15]] and HB101 [*supE*44, *hsdS*20(*r*B$^-$ *m*B$^-$), *recA*13, *ara*-14, *proA*2, *lacY*1, *galK*2, *rpsL*20, *xyl*-5, *mtl*-1, *leuB*6, *thi*-1]. The primers used for the construction of the fd-RP phage genome and for the error-prone PCR were listed in Fig. 1B. Plasmid pCANSS/RP3-42 was prepared previously (Nakashima et al. 2000). Phagemid fd-tet (Zacher et al. 1980) was constructed by subjecting phage vector fUSE1 (Parmley and Smith, 1988) to site-directed mutagenesis (Kunkel 1985; Kunkel et al. 1987) using primer Pfd (Fig. 1B) to eliminate the *Pvu*II site.

Construction of Phage fd-RP. The schematic diagram of the construction of the phage fd-RP is shown in Fig. 1A. Phage fd-RP was derived by replacing the entire D2 domain of g3p of fd-tet with the RP3-42 polypeptide (Prijambada et al. 1996). The *Bsr*GI-*Sfi*I fragment (0.9 kbp) containing the D1 DNA fragments of fd-tet was isolated after amplification by PCR using the RF of fd-tet as a template with P1 and P2 as the primers. The *Pac*I-*Sfi*I fragment (1.8 kbp) containing the D3 DNA fragment of fd-tet was isolated likewise except that P3 and P4 were used as primers. Primers P2 and P3 were used to introduce an *Sfi*I site into the genome of phage fd. The *Sfi*I-*Sfi*I fragment containing the RP3-42 gene (0.5 kbp) was isolated after amplification by PCR using phagemid pCANSS/RP3-42 as a template with P5 and P6 as the primers. The fragments, *Bsr*GI-*Sfi*I and *Pac*I-*Sfi*I, obtained from fd-tet, and the *Sfi*I-



B.

Pfd	**5'GCTAAACAACTTTCAACAGTTTCAGCGGAGTGAGAATAG**
P1	**5'CGACGGCCAGTGAATTGTACACCGTGCATC3'**
P2	**5'ATCACCGGCCTCGTGGGCCTTAGTACCGCCACCCTCAG3'**
P3	**5'ATTTTAGGCCTCTGGGGCCAATGCTGGCGGCGGCTCTGG3'**
P4	**5'CTAGAGGATCTTAATTAATTTTCCCTTAG3'**
P5	**5'CCTTTAGTTGTTCCTTTCTATGCGGCCCACGAGGCCATGGCTAGC3'**
P6	**5'AGTTTAGGCCACAGAGGCCTGGATCGCGAGATCTGTCGACTC3'**
P7	**5'TACTAAGGCCCACGAGGCCATGGCTAGCATGACTGGTGGACAGCAAATGGGT3'**

Fig. 1. Schematic diagram of the construction of fd-RP phage genome (**A**) and the sequences of the primers used for construction and mutagenesis (**B**). (A) B in the RP3-42 fragment indicates *Bam*HI site. The details of the construction are described in Materials and Methods.

*Sfi*I fragment from pCANSS/RP3-42 were simultaneously ligated to the fd-tet vector previously digested with *Bsr*GI and *Pac*I. Ligation of the three fragments, together with the vector, yielded fd-RP, and the structure of the construct was confirmed by DNA sequencing of the *Bsr*GI-*Pac*I region containing the RP3-42 gene.

Derivation of the Phage Clones Constituting Each Generation. The *Sfi*I fragment containing RP3-42 or its derivative sequences were amplified under error-prone PCR conditions as described previously (Arakawa et al. 1996) using primers P6 and P7. The amplified products were digested with *Sfi*I, and the *Sfi*I fragment was cloned into the corresponding region of the fresh fd-RP vector previously digested with *Sfi*I. Fresh fd-RP vector means the shorter vector prepared by first digesting the fd-RP DNA by *Bam*HI (Fig. 1A) and ligating the resultant vector fragment, yielding a vector with a shorter *Sfi*I fragment (700 bp). The fresh vector hence was used to discriminate colonies without insert and to ensure that mutations occurred solely in the *Sfi*I fragment of the fd-RP (Fig. 1A). The resultant derivatives were introduced into *E. coli* JM109 cells by electroporation (Dower et al. 1988), and the cell suspension was plated onto a 2 × YT (Sambrook et al. 1989) agarose medium containing 40 μg/ml tetracycline. Among the tetracycline resistant colonies appearing on the plates incubated at 37°C overnight, 24 colonies were arbitrarily chosen and subjected to colony PCR to confirm the presence of the *Sfi*I fragment insert (1000 bp) in their DNAs. Six to ten colonies out of the positive ones were arbitrarily chosen to constitute a generation. Each of the arbitrarily chosen clones was inoculated into 10 ml 2 × YT medium containing 20 μg/ml tetracycline and the culture was incubated at 37°C overnight. A 500-μl aliquot of the overnight culture was withdrawn and added into equal volume of Luria-Bertani medium (Sambrook et al. 1989) containing 30% glycerol. This mixture was stored at −80°C as phage stocks. The *E. coli* cells from the remaining culture were collected by centrifugation, the RF isolated and purified, and the DNA sequence of the insert analyzed.

The first generation was prepared by subjecting the random sequence in fd-RP to three consecutive rounds of error-prone PCR using primers P6 and P7. Among the variants generated, eight clones were arbitrarily chosen to be the member of the first generation. The clone with the highest infectivity, as evaluated by phage infectivity assay described below, was then selected to be the parent clone for the second generation. The *Sfi*I fragment of the parent clone was then subjected to one round of error-prone PCR using the same primer. Six clones were arbitrarily chosen and these constitute the second generation. Hereafter, each generation was prepared likewise except for the number of arbitrarily chosen clones (see Fig. 3 Legend).

Preparation of the Phage Suspensions of fd-RP and the Arbitrarily Chosen Clones. A 10-μl aliquot of the phage stock described above

was dispensed into 10 ml 2 × YT medium containing 20 μg/ml tetracycline and grown at 37°C overnight. The culture was centrifuged at 6,000 × g for 10 min to remove the bacterial cells, and the supernatant containing the phage particles was filtered through a Dismic 0.45 μm membrane (Toyo Roshi Kaisha, Ltd., Tokyo) to ensure the elimination of remaining bacterial cells, if any. The filtrate containing the phage particles was stored at 4°C, and used as phage suspension for the phage infectivity assay and phage ELISA described below.

Phage Infectivity Assay. The infectivity of a mutant phage was evaluated by the number of tetracycline resistant colonies grown after infecting the *E. coli* host with the phage particles contained in each of the phage suspensions. To determine the infectivity of each clone, the phage particles contained in a 100-μl aliquot of the prepared phage suspension described above were allowed to infect freshly grown *E. coli* TG1 cells of 0.8–0.9 OD_{600} (900 μl) for 40 min on standing at 37°C. The bacteria-phage mixture was spread onto a 2 × YT agarose medium containing 40 μg/ml tetracycline by gently swaying the plate. The tetracycline resistant colonies grown on the plate after incubating overnight at 37°C were counted, and the infectivity of the phage was expressed as the number of colonies per ml phage suspension (cfu/ml). The clone exhibiting the highest infectivity in a generation was chosen to be the parent clone for the next generation, and the value of infectivity of the new parent clone was standardized in reference to the parent clone of the prior generation.

ELISA. A 10 μg/ml of anti-M13 monoclonal antibody (Amersham Pharmacia Biotech) or anti-T7 monoclonal antibody (Novagen) in PBS (100 μl) were immobilized on the wells of the plate (Immuno Module, Nalge Nunc International) at 4°C overnight. Wells were washed with PBS and blocked with 1% BSA in PBS for 1.5 h at 37°C. The phage suspension or its serially diluted one (100 μl each) was then added to each well, and incubated for 1 h at 37°C. The wells were consecutively washed with PBS containing 0.05% tween 20 (PBST) and PBS before the addition of a 100 μl of PBS containing anti-M13 antibody conjugated with horseradish peroxidase and 8% nonfat dry milk. After a further incubation of the plates at 37°C for 1 h, wells were washed with PBST. The color reaction at 30°C for 30 min was initiated by the addition of 100 μl solution containing 0.4 mg/ml 3,3′,5,5′-tetramethylbendizine and 0.015% H_2O_2. The reaction was terminated by the addition of 100 μl of 1 *M* H_2SO_4, and the absorbance at 450 nm was measured with a microplate reader (Molecular Device).

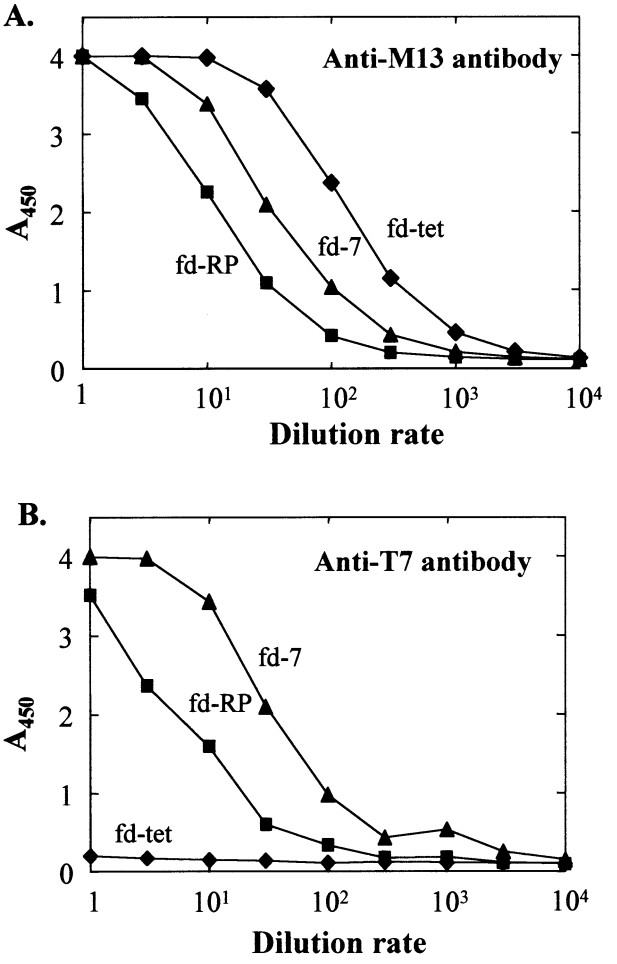


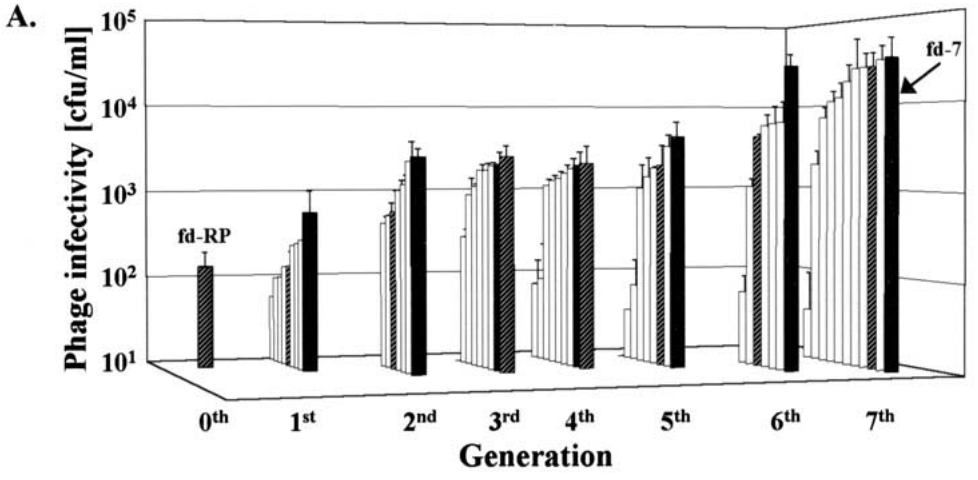
Fig. 2. ELISA of fd-tet, fd-RP, and fd-7 phage suspensions. Serial dilutions of each phage suspension were incubated in wells to react with (**A**) anti-M13 antibody and (**B**) anti-T7 antibody. See details under Materials and Methods. The absorbance at 450 nm is proportional to the amount of phage (A) or the amount of phage and RP3-42 polypeptide and its derivatives (B).

Results and Discussion

Given that the D2 domain of the g3p minor coat protein is required for the adsorption of the phage to the host F-pilus in order to mediate the initial stage of infection (Marvin 1998; Riechmann and Holliger 1997), completely replacing the D2 domain of the fd-tet with an arbitrary sequence of a random polypeptide, RP3-42, will result in a mutant fd phage devoid of the ability to infect *E. coli*. The designed mutant phage was named fd-RP, and was used as the origin for the experimental evolution.

Prior to knowing whether or not the infectivity of fd-RP became defective as a result of the substitution, analyses by phage ELISAs using anti-M13 antibody and anti-T7 antibody were first carried out to know the titre of the phage suspension and the efficacy of expression of the RP3-42 in the fd-RP phage. The presence of the phage particles in the fd-RP phage suspension was verified by phage ELISA using anti-M13 antibody (Fig. 2A). The corresponding dilution rate with 2 A_{450} was estimated by interpolation from the curve drawn with the absorbance of the reaction products on serial dilution of the phage suspension. As compared to the dilution rate of 1.4×10^2 of fd-tet, the fd-RP phage required a dilution rate of 3.7×10^1 (Fig. 2A), indicating that the concentration of phage fd-RP was eleven-times lower than that of fd-tet phage. To confirm whether the fd-RP genome expresses the RP3-42 sequence with T7 epitope tag, phage ELISA assay using anti-T7 antibody was conducted. Figure 2B showed the presence of the T7 epitope tag in fd-RP, implying the expression of the random sequence of RP3-42 in fd-RP phage. In contrast, the wild-type fd-tet did not react with the anti-T7 antibody (Fig. 2B).

The infectivity of the mutant phage was estimated by the number of tetracycline resistant clones, which were derived by the infection of the *E. coli* host with the phage particles contained in each millimeter of the phage suspensions. The infectivity of phage fd-RP



B.

```
fd-RP          1:MASMTGGQQM GRGSRGSILE GPHVGNPPSW GIPKLEKSQA GEPPFWDFCS SSCRNSPVWG SPRLGSSQIG KFPEWDSPNM 80
1st            1:.......... .......... ........T. .V........ .......... .......... .......... .L..R..H.. 80
2nd            1:.......... .......... .S......T. .V........ .......... .......... ...S..P... .L..R..H.. 80
3rd            1:.......... .......... .S......T. .V........ .......... .......... ...S..P... .L..R..H.. 80
4th            1:.......... .......... .S......T. .V........ .......... .......... ...S..P... .L..R..H.. 80
5th            1:.......... .........G .S......T. .V........ .......... ....S..... ...S..P... .L..R..H.. 80
6th            1:.......... .........G .S......I. .V........ .......... ....S..... ...S..P... .L..R..H.. 80
7th (fd-7)     1:.......... ........PG_.S......I. .V.Q...... .......... P...S..... P..S..P... .L..R..H.. 80

fd-RP          81:GDFQLGSKLE TFHVGGIPCW ETSILERSHL GGLPNWENSA AGSTRSGTGL ESTDLAIQAS GANA                 144
1st            81:.......P..G ........R. ......G... V.......P. .......... .......... ....               144
2nd            81:..S...P..G ........R. ......G... V.......P. .......... .......... ....               144
3rd            81:..S...P..G ........R. .....GG... V.......P. .......... .......... ....               144
4th            81:..S...P..G ........R. .....GG... V.......P. .......... .......... ....               144
5th            81:..S...P..G ........R. .....GG... V.......P. .......... .......... ....               144
6th            81:..S...P..G ........R. .....GG... V.P.D...P. .......... .......... ....               144
7th (fd-7)     81:..S...P..G ........R. .....GG... V.P.D.G.P. .......... .......... ....               144
```

Fig. 3. Infectivity of the phage clones in each generation (**A**) and the amino-acid sequence of the clones with highest infectivity in each generation (**B**). (A) The clones are arranged in the order of the lowest to the highest value of infectivity for each generation. Infectivity of each clone was determined in triplicates, except for the first generation. Scale bar indicates standard deviation. Hatched and filled bars represent the parent and the clone with highest infectivity, respectively. The parent clone of each generation was included in the population for comparison. Generations 3 and 5 contain one clone each that has no capability to infect *E. coli* TG1 cells and is indicated with a flat bar. The number of clones arbitrarily chosen to be the members of generations 1, 3, 4, and 5 was 8, while that of generations 2 and 6 was 6, and generation 7 was 10, as indicated in Table 1. fd-RP indicates the origin of the evolutionary study, and clone fd-7 indicates the clone with highest infectivity at the end of the lineage.

in F-pilus bearing bacteria (F^+ cells, TG1) significantly decreased to 1.4×10^2 cfu/ml from 3.0×10^9 cfu/ml of the wild-type fd-tet, indicating that the fd-RP phage suspension is about seven-order magnitude lower in infectivity. Considering the difference in the concentration of the fd-tet and fd-RP phages stated above, fd-RP is a phage of about six-order magnitude lower in infectivity per phage particle. In addition, neither fd-tet nor fd-RP phages infected the F^- cells, HB101. Accordingly, the fd-RP phage possesses a low infectivity, and can be considered as a defective phage.

The evolvability of fd-RP in terms of infectivity was assessed from several cycles of mutation and selection. The first generation was comprised of 9 clones arbitrarily chosen from the variant clones generated by random mutagenesis of the origin, fd-RP having an arbitrary sequence of RP3-42. The infectivity assays of the clones revealed that the first generation comprised a population with selectable variation in view of infectivity (Fig. 3A). The variation was also observed all through the generations within the time frame of the evolutionary study; the diversity in the population of each generation was confirmed by the significant differences in the infectivity among 6–10 arbitrarily chosen clones in each generation (Fig. 3A). Although the third and fourth generations seem to indicate that none of the variant clones within had a higher infectivity than that of the parent clone, the other generations documented the increase in the selectable property attributed to the variant random sequences. As a rule for the experimental evolution, the variant clone with highest infectivity among the arbitrarily chosen one in each generation is to be the parent clone for the next

Table 1. Synonymous and non-synonymous mutations found in the clones in each generation

Generation	Selected clone (clone with highest infectivity)		Non-selected clones[b]		
	No. of synonymous mutations	No. of non-synonymous mutations	No. of clones[c]	No. of synonymous mutations	No. of non-synonymous mutations
1(8)[a]	6	11	7	4.57	8.86
2(6)	1	4	4	1.75	4.25
3(8)	1	1	5	1.40	1.00
4(8)	1	0	5	1.40	1.80
5(8)	0	2	4	0.50	2.75
6(6)	1	3	2	0.50	1.50
7(10)	1	5	7	3.14	4.43
Total average	1.57	3.71		2.29	4.06

[a] The number in parentheses indicates the number of arbitrarily chosen clones comprising the population of each generation.
[b] Non-selected clones pertain to the members of the population excluding the clone exhibiting highest infectivity.
[c] The clones, among the unselected ones, found with full-length DNA sequence of the *Bsr*GI-*Pac*I region containing the RP3-42 gene were used for the estimation of the number of synonymous and non-synonymous mutations.

generation, regardless of the infectivity of the parent clone from which the variant clone with highest infectivity was derived. In addition, it is to be noted that the *Sfi*I fragments containing the variant random sequences were cloned into fresh vector to ensure that the sequence of the mutants varied only in the random sequence region of the fd-RP genome.

To glimpse the extent of development of the infectivity in our evolutionary study, phages fd-tet, fd-RP, and fd-7, were again subjected to infectivity assay. fd-7 is the phage clone exhibiting the highest infectivity in the seventh generation, where the experimental evolution was terminated. Compared with the titre of the origin fd-RP, 1.4×10^2 cfu/ml, in F^+ TG1 bacterial cells, the fd-7 phage was 3.4×10^4 cfu/ml, indicating the fd-7 phage suspension has a 240-fold higher infectivity than the fd-RP phage suspension, an increase by two-order magnitude. In addition, as observed with the fd-tet and fd-RP phages, the fd-7 phage did not infect the F^- host cells, indicating that its infectivity is F-pilus dependent, a characteristic feature of the D2 domain of the wild-type fd-tet phage. As the increase in infectivity of the fd-7 phage suspension is attributed in part to the concentration of phage in the suspension and to the number of polypeptides per phage particle, the phage suspensions were analyzed by phage ELISA using anti-M13 antibody and anti-T7 antibody. For comparison, the corresponding dilution rates of fd-RP and fd-7 phage suspensions giving a color reaction with 2 A_{450} were estimated by interpolation. For the reaction with anti-M13 antibody, the dilution rate of the fd-7 phage was about 3.7×10^1, and that of fd-RP was about 1.3×10^1 (Fig. 2A), inferring a 3-fold higher phage concentration in the fd-7 phage suspension. Therefore, the fd-7 phage has a 80-fold higher infectivity than the fd-RP. Phage ELISA using anti-T7 antibody gave a dilution rate of 3.6×10^1 for fd-7 phage and 5.6×10^0 for fd-RP (Fig. 2B), inferring that a fd-7 phage suspension has a 6.5-fold higher concentration of polypeptides displayed on the phage. Therefore, 37-fold increase in the infectivity of fd-7 was accredited to the evolved polypeptide replacing the D2 domain of the g3p protein.

The numbers of synonymous and non-synonymous mutations found in the selected clones in each generation are listed in Table 1, together with the average numbers of those found in the non-selected clones in each generation. The total average numbers of synonymous and non-synonymous mutations per sequence per generation were 2.17 and 4.0, respectively. These results show that the mutation rates of the selected and non-selected clones are similar, so as the ratio of the synonymous and non-synonymous mutations. Hence, the functional selection during the experimental evolution did not distinctly accelerate nor decelerate the evolutionary rate. If, by any chance, mutants with rare mutation number, i.e., very high or very low, were selected during the evolutionary study, it is then natural to expect an acceleration or deceleration of the evolutionary rate. Therefore, the selected clones are not rare based on the mutation rate of our experimental evolution. Indeed, they were found even in the obligatory small population of six to ten. It should be pointed out that in some cases where non-synonymous mutation rate was low, as observed at the third and fourth generations, the chance of finding advantageous mutants would be slim. Overall, the results show that at the primitive stage of evolution, the function understudy is still far from optimization and mutants with improved function are not rare. In fact, Yamauchi et al. (2002) observed likewise. Accordingly, during the early stage of evolution, functional selection in a small population is possible, but there will be no acceleration or deceleration of the evolutionary rate.

A BLAST 2.2.2 search of the SwissProt amino acid sequences reveals that the selected clone in each of the generations listed in Fig. 3B have no homology with any known polypeptides or proteins. Alignment of the sequences shows that 25 mutations caused the two-order magnitude increase in infectivity at the end of the fd-RP lineage, of which about a 37-fold increase could be attributed to the mutated random polypeptide replacing the D2 domain. In the perspective of protein sequence space, each sequence represents a distinctive amino acid combination, and with the random polypeptide, RP3-42, of 139 residues, the number of possible sequences is 20^{139} (Voigt et al. 2000), of which each sequence has its own fitness. The observance of the increased fitness of the descendants randomly sampled from a maximum of ten sequences in the space implies that, among the possible sequences, those attributing to the infectivity of the phage are not confined only to natural sequences, but rather, constitute significant fractions in the vast sequence space, that is, sequences with a specific function are not very rare. Taking into account that phage genomes frequently undergo mutation and recombination (Lewis et al. 1999), the possibility exists that a defective phage, which is low in its ability to infect certain bacterial cells, could accidentally incorporate an ORF with an arbitrary sequence from the host genome, and the ORF, alone or fused with the original phage proteins, could evolve to be a new functional protein or a domain that could rejuvenate the capacity of a defective phage to infect the host(s).

Here, we have demonstrated that, when fused with other pre-existing protein domains, a random sequence, having no explicit property, was able to evolve towards acquiring a functional role that contributes to the rejuvenation of the infectivity of a defective bacteriophage. The present stage of our experimental evolution might still be in its infancy for us to conclude that every arbitrary sequence is evolvable. However, the study could herald the conception of new perspective regarding the evolution of primordial polypeptides in the field of molecular evolution.

Acknowledgments. This work was supported in part by Grants (10450310,12019244, 12480200,13208016, and 11CE2006) from the Ministry of Education, Science, Sports and Culture, Japan and by a grant from the Rice Genome Project PR-2103, MAFF, Japan.

References

Arakawa T, Jongsareejit B, Tatsumi Y, Tanaka K, Ikeda K, Komatsubara H, et al. (1996) Application of N-terminally truncated DNA polymerase from Thermus thermophilus (delta Tth polymerase) to DNA sequencing and polymerase chain reactions: comparative study of delta Tth and wild-type Tth polymerases. DNA Res 3:87–92

Barnett P, Tabak HF, Hettema EH (2000) Nuclear receptors arose from pre-existing protein modules during evolution. Trends Biochem Sci 25:227–228

Brookfield JF (2001) Evolution: the evolvability enigma. Curr Biol 11:R106–108

Deng LW, Malik P, Perham RN (1999) Interaction of the globular domains of pIII protein of filamentous bacteriophage fd with the F-pilus of *Escherichia coli*. Virology 253:271–277

Dower WJ, Miller JF, Ragsdale CW (1988) High efficiency transformation of *E. coli* by high voltage electroporation. Nucleic Acids Res 16:6127–6145

Gilbert W (1978) Why genes in pieces? Nature, 271:501

Go M (1981) Correlation of DNA exonic regions with protein structural units in haemoglobin. Nature 291:90–92

Jensen RA (1976) Enzyme recruitment in evolution of new function. Annu Rev Microbiol 30:409–425

Keefe AD, Szostak JW (2001) Functional proteins from a random-sequence library. Nature 410:715–718

Kirschner M, Gerhart J (1998) Evolvability. Proc Natl Acad Sci USA 95:8420–8427

Kunkel TA (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc Natl Acad Sci USA 82:488–492

Kunkel TA, Roberts JD, Zakour RA (1987) Rapid and efficient site-specific mutagenesis without phenotypic selection. Methods Enzymol 154:367–382

Marvin DA (1998) Filamentous phage structure, infection and assembly. Curr Opin Struct Biol 8:150–158

Nakashima T, Ishiguro N, Yamaguchi M, Yamauchi A, Shima Y, Nozaki C, Urabe I, Yomo T (2000) Construction and characterization of phage libraries displaying artificial proteins with random sequences. J Biosci Bioeng 90:253–259

Parmley SF, Smith GP (1988) Antibody-selectable filamentous fd phage vectors: affinity purification of target genes. Gene 73:305–318

Prijambada ID, Yomo T, Tanaka F, Kawama T, Yamamoto K, Hasegawa A, et al. (1996) Solubility of artificial proteins with random sequences. FEBS Lett 382:21–25

Riechmann L, Holliger P (1997) The C-terminal domain of TolA is the coreceptor for filamentous phage infection of *E. coli*. Cell 90:351–360

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory Press, New York

Tonegawa S, Maxam AM, Tizard R, Bernard O, Gilbert W (1978) Sequence of a mouse germ-line gene for a variable region of an immunoglobulin light chain. Proc Natl Acad Sci USA 75:1485–1489

Voigt CA, Kauffman S, Wang ZG (2000) Rational evolutionary design: the theory of in vitro protein evolution. Adv Protein Chem 55:79–160

Yamauchi A, Yomo T, Tanaka F, Prijambada ID, Ohhashi S, Yamamoto K, et al. (1998) Characterization of soluble artificial proteins with random sequences. FEBS Lett 421:147–151

Yamauchi A, Nakashima T, Tokuriki N, Hosokawa M, Nogami H, Arioka S, et al. (2002) Evolvability of random polypeptides through functional selection within a small library. Prot Eng 15:619–626

Zacher AN III, Stock CA, Golden JW II, Smith GP (1980) A new filamentous phage cloning vector: fd-tet. Gene 9:127–140