Protein Engineering vol.15 no.7 pp.619–626, 2002

Evolvability of random polypeptides through functional selection within a small library

Asao Yamauchi[1], Toshihiro Nakashima[1,2], Nobuhiko Tokuriki[1], Masato Hosokawa[1], Hideki Nogami[1], Shingo Arioka[1], Itaru Urabe[1] and Tetsuya Yomo[1,3,4,5]

[1]Department of Biotechnology, Graduate School of Engineering, Osaka University, 2-1, Yamada-oka, Suita Shi, Osaka, 565-0871, [2]The Chemo-Sero-Therapeutic Research Institute, Kawabe, Kyokushi, Kikuchi, Kumamoto 869-1298, [3]PRESTO, JST, 2-1 Yamadaoka, Suita, Osaka 565-0871 and [4]Department of Pure and Applied Sciences, University of Tokyo, Komaba, Meguro-ku, Tokyo 153-8902, Japan

[5]To whom correspondence should be addressed.
E-mail: yomo@bio.eng.osaka-u.ac.jp

A directed evolution with phage-displayed random polypeptides of about 140 amino acid residues was followed until the sixth generation under a selection based on affinity to a transition state analog for an esterase reaction. The experimental design deliberately limits the observation to only 10 clones per generation. The first generation consists of three soluble random polypeptides and seven arbitrarily chosen clones from a previously constructed library. The clone showing the highest affinity in a generation was selected and subjected to random mutagenesis to generate variants for the next generation. Even within only 10 arbitrarily chosen polypeptides in each of the generations, there are enough variants in accord to capacity of binding affinity. In addition, the binding capacity of the selected polypeptides showed a gradual continuous increase over the generation. Furthermore, the purified selected random polypeptides exhibited a gradual but significant increase in esterase activity. The ease of the functional development within a small sequence variety implies that enzyme evolution is prompted even within a small population of random polypeptides.

Keywords: esterase activity/phage display/protein evolution/random protein

Introduction

Proteins have evolved basically by iterating between diversification through mutation and selection in accord with function. Because the diversification process creates an enormous variety of sequences, how many of the variants should be exploited to allow the observation of evolution, especially in the case of primordial polypeptides, in an attempt to understand the initial stage of protein evolution?

Although nobody would ever provide the explicit route taken by natural proteins, experimental molecular evolution has imparted several insights towards the understanding of protein evolution (Matsuura *et al.*, 1999; Arnold *et al.*, 2001; Kashiwagi *et al.*, 2001). We have previously shown that polypeptides with random sequences have a very slim chance of having such unique properties (Prijambada *et al.*, 1996; Yamauchi *et al.*, 1998). Recently, Keefe and Szostak showed roughly one in 10^{11} random polypeptides containing 80 contiguous random amino acids had ATP-binding ability (Keefe and Szostak, 2001). Consecutively, it is possible to take new functional proteins directly from a large library (approximately 10^{13}) of random polypeptides (Keefe and Szostak, 2001; Wilson *et al.*, 2001) and further improve a property by directed evolution (Keefe and Szostak, 2001).

How many random sequences need to be searched to permit the observance of evolving polypeptides towards acquiring higher functions? Is it feasible to decipher evolution from observing only a few variants per generation out of the vast number of available sequences, especially at the primitive stage? To address these issues, we deliberately limit our evolutionary studies on populations of arbitrarily chosen mutant random polypeptides to a maximum of 10 for each generation, although a phage display system with selection using transition state analogs (TSAs) has the capability of yielding highly diversified mutants of $10^{8–9}$ rendering it an efficient tool for generating a protein catalyst (Patten *et al.*, 1996; Fujii *et al.*, 1998; Forrer *et al.*, 1999). Here, we show that even examining only 10 arbitrarily chosen polypeptides with random sequences per generation could guarantee the observance of evolution, in which there is a continuous gradual increase in a function of the polypeptides via iterating mutation and selection.

Materials and methods

Bacterial strains, phage, plasmids and chemicals

Escherichia coli strains used in this study were TG1 [K12, Δ(*lac-pro*) *supE thi hsdD5*, F′ *traD36 proAB lacI*^{q} *lacZ* Δ*M15*] (Maniatis *et al.*, 1982) and BL21(DE3) (Novagen). Three phage clones displaying random polypeptides RP3-04, RP3-42 and RP3-45 were prepared as described previously (Nakashima *et al.*, 2000). Phage library PL1 and phagemid pCANSS were made available previously (Nakashima *et al.*, 2000). Helper phage M13KO7 was purchased from Amersham Pharmacia Biotech. PET21aSH vector for expressing C-terminal His_6-tagged random polypeptides is a derivative of pET21a(+) (Novagen), in which the *Hin*dIII/*Xho*I cloning site was replaced with a *Hin*dIII/*Xho*I dsDNA fragment containing an *Sfi*I site. The *Hin*dIII/*Xho*I dsDNA fragment was prepared by annealing 5′-AGCTTGGCCTCTGGGGCCGCACACCACCACCACC ACCACTAAC-3′ with 5′-TCGAGTTAGTGGTGGTGGTGG-TGGTGTGCGGCCCCAGAGGCCA-3′. The TSA, *p*-nitrophenyl hydrogen 4-(hydroxycarbonyl)butylphosphonate (Ohkubo *et al.*, 1994), was kindly synthesized by Dojindo (Kumamoto, Japan).

Preparation of phage particles

In the case of the 10 random polypeptides per generation, each arbitrarily chosen phagemid clone harboring a variant gene of the polypeptide was grown in 2× YT (Maniatis *et al.*, 1982) containing 100 μg/ml ampicillin and 2% glucose to an OD_{600} of 0.7–1.0, and the phage was rescued by adding the helper phage M13KO7 [2×10^{10} plaque-forming units (pfu)/ml]. After standing for 30 min at 30°C and shaking at the same temper-

ature for another 30 min, the bacterial culture containing phage particles was centrifuged. The pellets rinsed with 2×YT containing 100 μg/ml ampicillin and 50 μg/ml kanamycin were suspended in the same medium and incubated at 30°C for 18 h. The cells were removed by centrifugation at 4°C for 15 min and the supernatant filtered by a Dismic-25 cs disposable syringe unit (0.45 μm, ADVANTEC MFS). The phage particles in the filtrate were precipitated by incubating at 4°C overnight after the addition of one fifth volume of 20% poly(ethylene glycol) 6000 (Wako Pure Chemical Industries) in 2.5 M NaCl. The phage particles collected by centrifugation at 10 000 *g* for 20 min were resuspended in 50 mM Tris–HCl (pH 7.0) containing 150 mM NaCl. The phage suspension was adjusted to a titer of 10^{13} transducing units (t.u.)/ml as estimated by the number of ampicillin resistant colonies (Marks *et al.*, 1991).

For the initial library of arbitrarily chosen 1000 clones, the preparation of phage particles is essentially the same as described above, except for the preparation of the library, from which the clones were derived. The library was prepared by consecutive error-prone and standard polymerase chain reaction (PCR) of the random polypeptide gene of the clone with highest binding affinity in the first generation of the 10-membered library. The mixture of the variant mutant genes so obtained was used to transform *E.coli* TG1, and from the transformants, 1000 clones were arbitrarily chosen, mixed, and used as initial library for the preparation of phage particles.

In vitro selection system

Each of the random polypeptides displayed on the surface of M13 phage as fusion proteins with the pIII coat protein were panned with the TSA (Figure 1a) of an esterase reaction (Ohkubo *et al.*, 1994). The TSA plate was prepared as follows: The 96-well NH_2 microtiter plates (CovaLink^{TM} NH_2 Module, NalgeNunc International) were coated with the TSA using the carbodiimide coupling method by the addition of 50 μl of 33 μM TSA in 0.1 M MES (pH 4.7) containing 0.15 M NaCl and 50 μl of 26 mM 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC). The plates were incubated overnight at 20°C before the plate was washed with 0.3 ml of water three times and blocked with glycine by incubating the plate for 1.5 h at 20°C after the addition of 0.3 ml of 1 M glycine. The well was then washed with 0.3 ml of 50 mM Tris–HCl (pH 7.0) containing 150 mM NaCl three times before ready for use. Non-TSA plates were prepared with the same protocol except that 33 μM glycine was coated onto the wells of the microtiter plates instead of 33 μM TSA.

For the 10 random polypeptides per generation, the level of binding affinity of every clone was evaluated by the number of phages bound to the TSA plate as eluted with the TSA-containing buffer. The procedure is as follows: phages of ~10^{10} t.u. suspended in 100 μl of 50 mM Tris–HCl (pH 7.0) containing 150 mM NaCl and 0.5% skim milk were incubated for 2 h at 15°C in a single well of the TSA plate for the

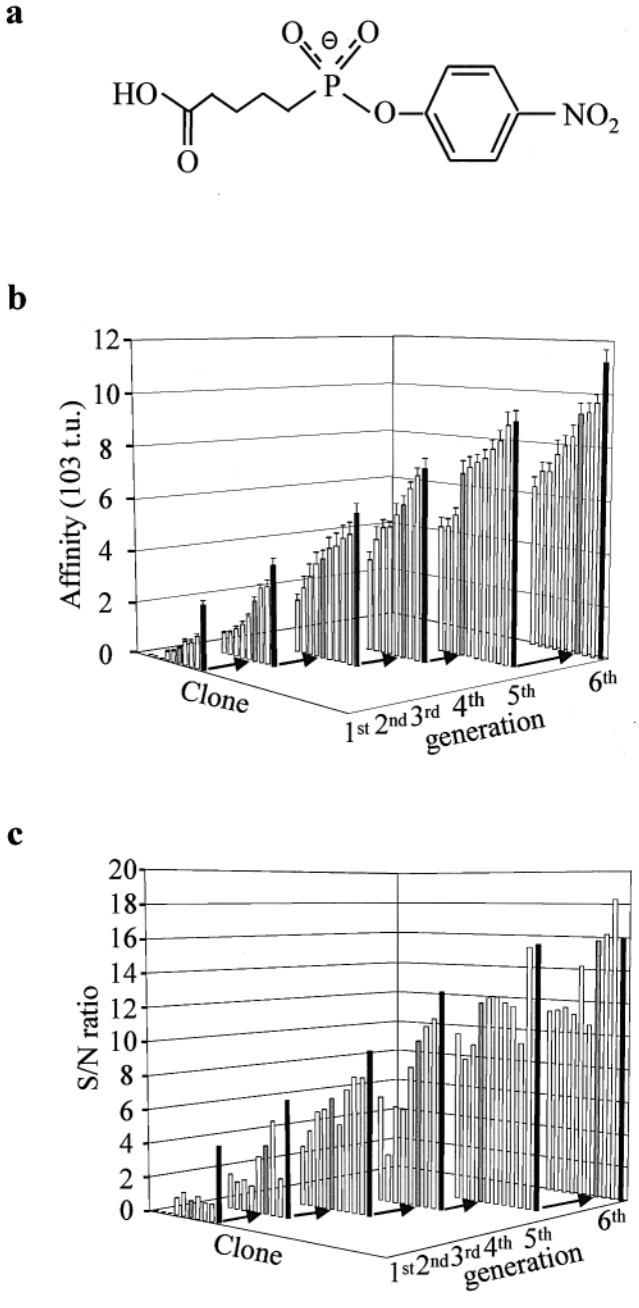

Fig. 1. Binding affinity and specificity of each clone in the small phage libraries displaying random polypeptides towards the TSA for an esterase reaction. (**a**) Structure of TSA used for the selection. (**b**) Binding affinity is expressed as the number of eluted phage bound to the TSA (see Materials and methods). The values are arranged in the order of the lowest to the highest value for each generation. (**c**) S/N ratio calculated as the ratio of the numbers of phages eluted using TSA-containing and non-TSA buffers. The clones are arranged in the same order as in (b). Binding affinity and S/N ratio were determined in triplicate for each clone. The black bars indicate the phage clones showing the highest value in each generation. The gray bar in the first generation shows the RP3-42-displaying clone, and those in the second to sixth generations show the parent clone of each generation. For each generation, about 10 clones were arbitrarily chosen from the generated mutant library after confirming that the genes isolated from the clones do not have deletion mutations and have intact T7- and E-epitope tag sequences (Nakashima *et al.*, 2000). Including the parent clone, the number of clones in the library of first to third generations was 10 and 12 for the fourth to the sixth generations. In the second and fifth generations, it was found that two of the clones had the same amino acid sequences but different nucleotide sequences. In the sixth generation, three clones were observed likewise. In the fourth generation, there were three clones with the same amino acid sequences, of which one had different nucleotide sequences, and two clones had the same amino acid sequences but different nucleotide sequences. Note that the amino acid sequences of the latter two clones are different from the set of former three clones. The results from the clones with different amino acid sequences in each generation are shown in (b) and (c). Scale bar in (b) indicates SD. The averages and SD of S are shown in (b), while the averages and SD of N of the clones with highest binding affinity in first, second, third, fourth, fifth and sixth generations were $5.5\times10^2 \pm 56$, $5.6\times10^2 \pm 20$, $5.9\times10^2 \pm 24$, $5.7\times10^2 \pm 43$, $5.8\times10^2 \pm 37$ and $6.7\times10^2 \pm 59$, respectively.

phages to bind with the TSA. The well was then washed 16 times by consecutive gentle addition of 200 µl of 150 mM Tris–HCl (pH 7.0) containing 150 mM NaCl and aspiration of 200 µl of the solution, leaving 100 µl of the solution to prevent the well from drying. The plate was incubated at 15°C for 2 h after the 16th wash, after which washing was again performed five times using the same procedures. The bound phages were released by incubating the plate at 15°C for 2 h following consecutive addition of 200 µl of 250 µM TSA in the same buffer and aspiration of 200 µl of the solution. The number of phages eluted in the solution, expressed as t.u., was measured by the number of ampicillin resistant colonies (Marks *et al.*, 1991).

Basically, the same selection system described above was applied to phage particles prepared from an initial mixture of 1000 clones. Phage particles were eluted after each of the 12 panning rounds, and the eluted mixed phages were used to infect *E.coli* TG1. All transfectants of a generation were collected and used for the preparation of phage particles.

Random mutagenesis and construction of the phage library

A random polypeptide gene on the phagemid, prepared from *E.coli* TG1 cells infected with a selected phage, was amplified by error-prone PCR as described previously (Arakawa *et al.*, 1996) using the *Thermus thermophilus* DNA polymerase and the primers, 5′-CCGCCTTTGAGTGAGCTGAT-3′ and 5′-TCGTCACCAGTACAAACCACAACG-3′. The mutated products were isolated and further amplified by standard PCR as described previously (Nakashima *et al.*, 2000). For the first to fourth generations, the forward primer used for standard PCR was 5′-ATCCTCGCAACTGCGGCCCACGTGGCCAT-GGCTAGCATGACTGGTGGACAGCAAATGGGT-3′ and the reverse primer was 5′-AGTTTAGGCCCCAGAGGCTAAT-CGCAGTCTGTCGACTC-3′ to allow screening for the *Bgl*I restriction site. For the fifth and sixth generations, primers used were basically the same except that the 19th base of the reverse primer was changed to C instead of T to generate *Sfi*I restriction site for screening purposes. The change of base on the primer did not affect the amino acid sequences. The *Bgl*I or *Sfi*I fragment from the PCR products was ligated into pCANSS, and *E.coli* TG1 cells were transfected with the ligated DNA as described previously (Nakashima *et al.*, 2000). The phage library was prepared from the phagemid library as described previously (Nakashima *et al.*, 2000).

Expression and purification of His-tagged random polypeptides

A polypeptide gene on the phagemid carried by a selected phage was recloned into the *Nhe*I/*Sfi*I site of the newly constructed expression vector, pET21aSH, described above. The *E.coli* BL21(DE3) cells harboring the hybrid plasmid grown at 37°C on LB medium (Maniatis *et al.*, 1982) containing 50 µg/ml ampicillin to an OD_{600} of 0.6–1.0 was further incubated for 5 h at 30°C after the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) (final concentration, 1 mM). Cells harvested after IPTG induction were disrupted in 50 mM potassium phosphate (pH 8.0) containing 8 M urea, 0.3 M NaCl and 10 mM imidazole, and the cell suspension centrifuged. The supernatant was applied to a Ni-NTA Superflow (Qiagen) column equilibrated with the same buffer used for disrupting the cells. The column was washed with 50 mM potassium phosphate (pH 8.0) containing 8 M urea, 0.3 M NaCl and 20 mM imidazole, and the polypeptide was eluted with an imidazole gradient of 20–80 mM. The fractions showing a single band on SDS–polyacrylamide gel electrophoresis (SDS–PAGE) were pooled, and the pooled fractions transferred to Spectra/Por® dialyzing membrane (MWCO: 3500; Spectrum Laboratories) and concentrated by surrounding the tube with poly(ethylene glycol) 20 000 (Wako Pure Chemical Industries). After incubating at room temperature for 6 h, the tube was gently rinsed with distilled water and dialyzed against 62.5 mM sodium acetate, pH 5.0. Dialysis was first carried out at room temperature with two buffer changes, each at 2-h intervals, to completely remove the urea from the polypeptide solution. The polypeptide solution was further dialyzed at 4°C with three buffer changes at consecutive 2-, 12- and 1-h intervals. The dialyzate was sterilized using a Dismic-25 cs disposable syringe unit (0.45 µm) and the purified polypeptide was stored at 4°C.

All purified polypeptides were confirmed homogeneous by SDS–PAGE, and the amino acid compositions coincided well with those deduced from the nucleotide sequences. The molar absorption coefficients of the polypeptides at 280 nm, determined from the amino acid composition and A_{280} value of the purified polypeptide solution as previously described (Suga *et al.*, 1996), were as follows: RP3-42H, 37 000 M^{-1} cm^{-1}; YSLP1-1, 63 000 M^{-1} cm^{-1}; YSLP3-1, 67 000 M^{-1} cm^{-1}; and YSLP6-1, 65 000 M^{-1} cm^{-1}.

Esterase activity

The hydrolysis rate of *p*-nitrophenyl acetate was measured at 37°C. Reactions were initiated by mixing 20 µl of 2.5 mM *p*-nitrophenyl acetate in 1.25% methanol with 80 µl of a polypeptide solution containing 62.5 mM sodium acetate (pH 5.0). After incubating for 1, 2, 3 or 4 h, the reaction mixture was chilled on ice and the pH was increased to 7.2 by adding 25 µl of 250 mM potassium phosphate (pH 8.0). The precipitate formed after the pH shift of the reaction mixture was separated by centrifugation, and the absorbance of the supernatant at 400 nm was measured. The 0 h sample was prepared in the same manner, except that the substrate and polypeptide were mixed on ice without incubation at 37°C. To examine the effect of the TSA, different concentrations of the TSA (0.47, 2.35 and 4.65 mM) were added into the substrate solution used for the reaction. The concentration of *p*-nitrophenol released by the reaction was calculated using a molar absorption coefficient of 10 400 M^{-1} cm^{-1}, which was determined experimentally at pH 7.2.

Results and discussion

The experimental evolution was initiated with 10 random polypeptides of about 140 amino acid residues displayed on the surface of a filamentous bacteriophage as fusion proteins with the pIII coat protein (Nakashima *et al.*, 2000). Selection from a phage-displayed library of random polypeptide variants does not, in principle, require any structural information regarding the parent protein of interest (Dunn, 1996). Taking into account that a correlation exists between catalytic activity and affinity to respective TSA (Baca *et al.*, 1997) and that the initial materials are in their primordial states (Prijambada *et al.*, 1996; Yamauchi *et al*., 1998), panning conditions with the TSA (Figure 1a) for an esterase reaction were optimized to allow detection of a minute binding affinity, if present.

The first generation comprised of RP3-04-, RP3-42- and RP3-45-displaying phage (Nakashima *et al.*, 2000) and seven clones displaying different random polypeptides arbitrarily chosen from a large library (PL1) constructed previously

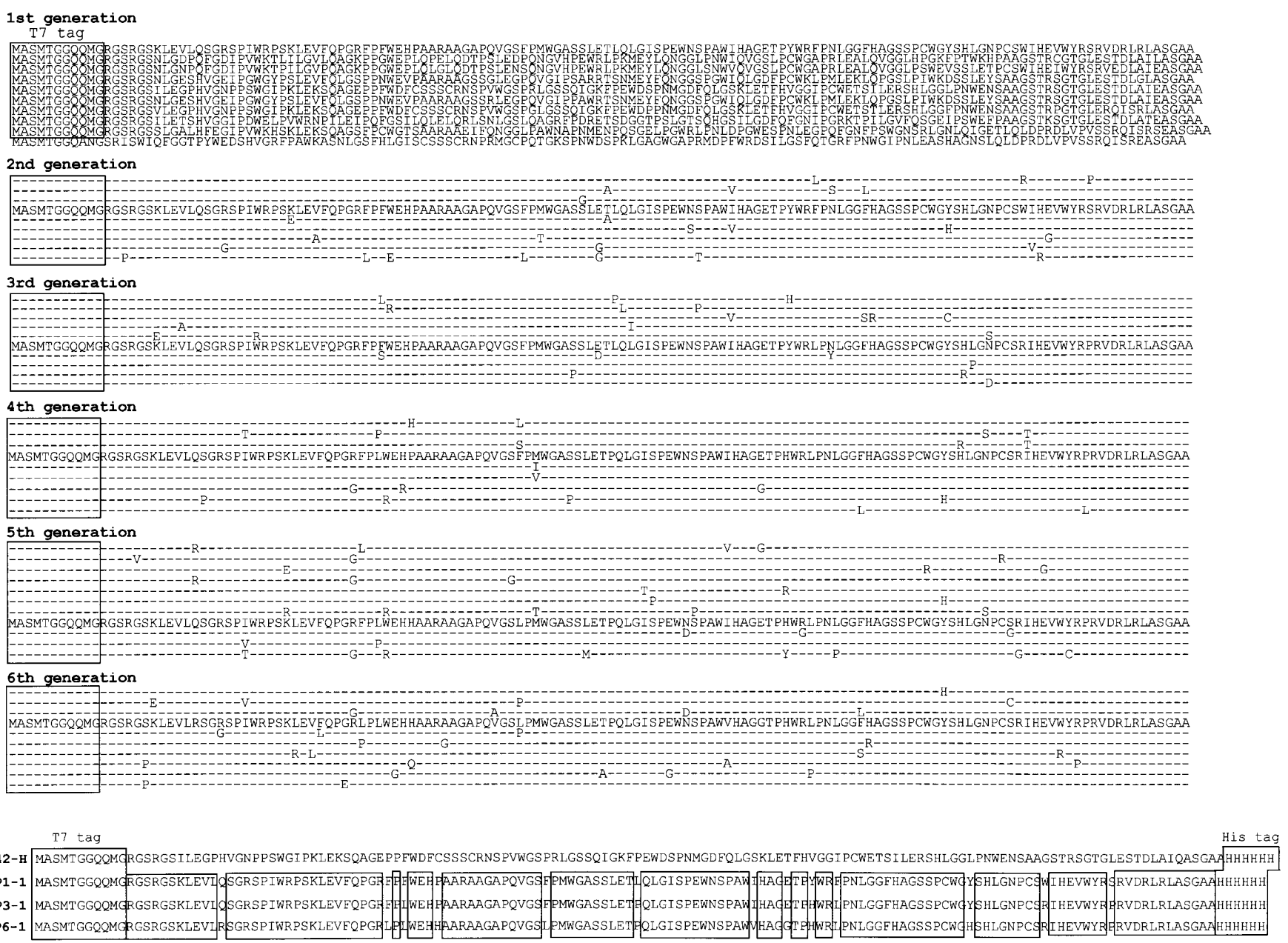

Fig. 2. Deduced amino acid sequences of random polypeptides and their derivatives. (**a**) Sequences displayed by all the phage clones in each generation shown in Figure 1b. Sequences in each generation are arranged following the order of the highest binding affinity value to the lowest of the clones shown in Figure 1b. Except for the parent clone of the second to the sixth generation, dashed lines indicate the amino acids that are the same in all the sequences. The T7 tag sequence, the first 11 amino acids for the T7 gene 10 protein, is boxed. The T7 tag sequence of the clone in the first generation with lowest binding affinity was mutated. (**b**) The sequences of RP3-42H, YSLP1-1, YSLP3-1 and YSLP6-1. Amino acids that are the same in two or three sequences are boxed.

(Nakashima, *et al.*, 2000). RP3-04, RP3-42 and RP3-45 are soluble random polypeptides with no secondary structure propensity but are as compact as a molten globule (Yamauchi *et al.*, 1998). In addition, all the displayed random polypeptides have no homology with known natural proteins in the SwissProt database as analyzed by BLAST 2.2.1 (Figure 2a). Hence, the initial material of the evolutionary study could be assumed to be like any primitive sequence in view of evolvability regarding affinity to the TSA. When the 10 phage clones of the first generation were assayed for binding affinity, significant differences in binding ability to the TSA were observed (Figure 1b), implying that even from a small library of 10 random polypeptides, experimentally detectable variety could be found. Hence, sampling of as few as 10 from a huge number of possible sequences permits the observation of an evolving population.

The clone with the highest binding affinity in the first generation was subjected to random mutagenesis generating varieties for the second generation. Among the point mutants randomly generated, eight clones were arbitrarily chosen and together with the parental clone were subjected to the binding affinity assay as in the first generation. The nine clones, which comprised the second generation, showed significant differences in their binding affinity, where three clones even had stronger binding affinity than the parent clone (Figure 1b). These results mark the implementation of Darwinian evolution in the system. The same mutation and selection cycle was carried out for the succeeding generations, and a similar distribution pattern in the binding affinity was observed among the members of the population in each of the generations (Figure 1b). The differences in the binding affinity, though they may appear small, are significant such that the accumulation of the point mutations among the selected polypeptides in each generation brought about a continuous increase in the binding affinity over the generations. Such a continuous increase could not have been observed if the differences in the binding affinity were due to experimental noise, which could increasingly hinder an adaptive walk (Levitan and Kauffmann, 1995). Furthermore, the fact that $>30\%$ of the clones in each of the mutant libraries showed an enhanced binding affinity, the polypeptides in each generation are considered a primitive populace (Trakulnaleamsai *et al.*, 1995; Matsuura *et al.*, 1998). Hence, the property has the allowance for improvement through the process of Darwinian evolution as theoretically proposed by Matsuura *et al.* (Matsuura *et al.*, 1998).

The deduced amino acid sequences of the polypeptide region of all the members in each generation are listed in Figure 2a, and no homology was detected between the selected random polypeptides and the natural proteins with known amino acid sequences. As shown in Figure 2a, even with only a few mutations accumulated, 13 sites, from the initial stage to the sixth generation, the acquired binding capacity of the mutant YSLP6-1 in the sixth generation is significantly high as compared to those in the initial stage of the experimental evolution. The easy ascent towards acquiring a new property, the binding affinity, by a small number of mutations in the random polypeptides, indicates that a significant fraction of amino acid sequences other than the known natural proteins have a certain function, at least in the binding activity.

Will an increase in variety make any difference in the observation of protein evolution? About 1000 clones were arbitrarily chosen from the mutant polypeptides generated from the selected clone in the first generation, and were subjected to several panning rounds with the TSA. The binding affinity of the clones increased gradually on each panning round but becomes saturated on the 11th round, as shown by no further increase in the binding affinity of the population (Figure 3). Out of the 10 clones sampled from the final population after the panning rounds, nine were found to have the same amino acid sequence (two amino acid substitutions, W46R and L75P, from the parent sequence of the second generation) and one has the same sequence as that of the third clone in the second generation (see Figure 2b). In addition, the binding affinity of the major clone (4.2×10^3 t.u.) was found to be similar to the clone with highest binding affinity (3.8×10^3 t.u.) in the second generation. These results infer that a large variety is not always effective for obtaining a good mutant.

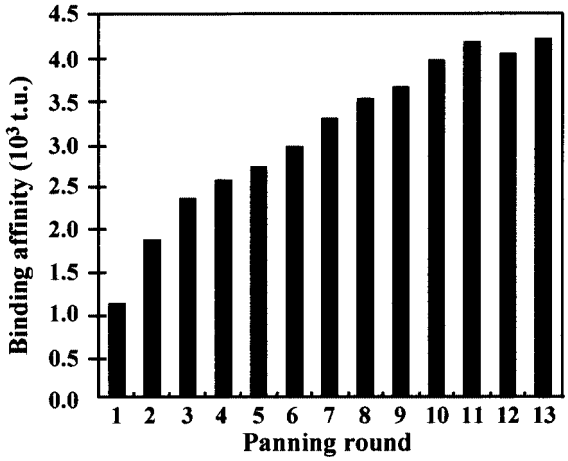

Fig. 3. Changes in the binding affinity of mixture of clones from an initial library of about 1000 clones as a function of panning round.

In our experimental evolution, the phage most highly eluted from the TSA plate by the TSA-containing buffer convenes the clone with highest binding affinity in each generation. However, the possibility remains that the phage eluted by the TSA-containing buffer may include non-specifically bound phages. Hence, the ratio of the numbers of phages eluted with the TSA-containing and non-TSA buffers, represented as the S/N ratio, were calculated (Figure 1c). The results indicate that the S/N ratio, an indication of the specificity of the TSA binding, increases with the increase in the binding affinity. It was also confirmed that when panning was done on the non-TSA plate, the number of eluted clones eliciting the highest binding affinity in each generation, using either the TSA-containing or non-TSA buffer, is almost the same and is at ground level (data not shown).

Considering that the TSA is a transition state analog for an esterase reaction, the esterase activity of the polypeptides may have increased through the process of evolution, as a number of studies have demonstrated that catalytic efficiency usually increases as a function of increased affinity for a respective TSA (Philipps *et al.*, 1992; Angeles *et al.*, 1993; Stewart and Benkovic, 1995). Note that the binding affinity of the random polypeptides was estimated by the titer of the M13 phage displaying the polypeptide as a fusion protein with the pIII coat protein. Therefore, to assess the kinetic parameters for the esterase activities of the selected (advantageous) polypeptides in different generations accurately, the polypeptides were detached from the fused protein displayed on the phage prior to purification. The polypeptide genes in the phagemids from RP3-42-displaying phage clones, the clones showing the high-

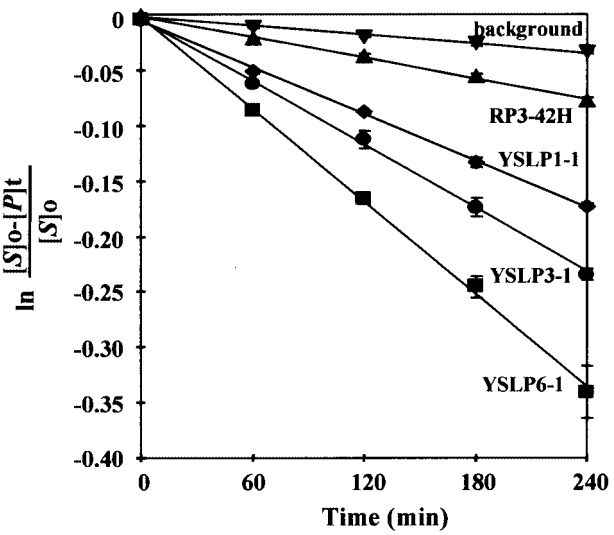

Fig. 4. Course of the hydrolysis of *p*-nitrophenyl acetate in the presence and absence of the polypeptides selected from the libraries of the first, third and sixth generations. Esterase activity was measured as described in Materials and methods. Polypeptide concentrations ([*E*]) were 22, 19, 18 and 20 μM, for RP3-42H, YSLP1-1, YSLP3-1 and YSLP6-1, respectively. ▲, RP3-42H; ◆, YSLP1-1; ●, YSLP3-1; ■, YSLP6-1; ▼, background.

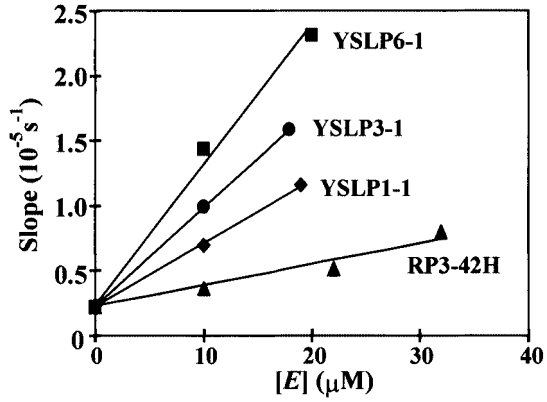

Fig. 5. Effect of polypeptide concentration on the slope of the linear lines for the hydrolysis of *p*-nitrophenyl acetate shown in Figure 4. ▲, RP3-42H; ◆, YSLP1-1; ●, YSLP3-1; ■, YSLP6-1.

est binding affinity toward TSA in the first, third, and sixth generations, were independently recloned to an expression vector as described in Materials and methods. The expressed polypeptides with a His_6 tag were then purified and named RP3-42H, YSLP1-1, YSLP3-1 and YSLP6-1, respectively. The deduced amino acid sequences are shown in Figure 2b. The esterase activity of the purified polypeptides was measured using *p*-nitrophenyl acetate as a substrate (Figure 4). In addition, it was examined whether the TSA inhibits the activity competitively.

In the presence of a competitive inhibitor, *I*, the rate of hydrolysis of an ester, *S*, is expressed as

$$-d[S] / dt = k_{cat}[E][S] / \{(1 + [I]/K_I)K_m + [S]\} + k_B[S] \quad (1)$$

where K_I is the inhibition constant, k_B is the rate constant for ester hydrolysis in the absence of a polypeptide, *E*, i.e. the rate constant of the background reaction. Here, K_I is equivalent to the ratio of the dissociation rate constant to the association rate constant, and hence K_I represents the dissociation constant, K_d, of TSA. When the activity of a polypeptide is at a primitive level, the Michaelis constant, K_m, must be much larger than the mM order. Hence, applying this to our experimental conditions where the initial concentration of ester, $[S]_0$, is 0.5 mM, $(1 + [I] / K_I)K_m >> [S]$. In such a case, Equation (1) is simplified to

$$-d[S] / dt = [k_{cat}[E] / \{(1 + [I] / K_I)K_m\} + k_B][S] \quad (2)$$

Integration of equation (2) gives:

$$\ln\{([S]_0 - [P]_t) / [S]_0\} = -[k_{cat}[E] / \{(1 + [-I] / K_I)K_m\} + k_B]t \quad (3)$$

where *t* is the reaction time, and $[P]_t$ is the concentration of *p*-nitrophenol released by the reaction at time *t*. In the absence of the inhibitor, Equation (3) is simplified to

$$\ln\{([S]_0 - [P]_t) / [S]_0\} = -(k_{cat}[E] / K_m + k_B)t \quad (4)$$

Figure 4 shows that ester hydrolysis catalyzed by the polypeptides followed a linear relationship expected from Equation (4), in agreement with large K_m values. Figure 5 shows the effect of polypeptide concentration on the slope of the linear lines for the esterase hydrolysis, as shown in Figure 4. The linear lines for RP3-42H, YSLP1-1, YSLP3-1 and YSLP6-1 intersect at the *y*-axis**.** The intersecting points show a k_B value of 2.3×10^{-6} s^{-1} for the background reaction and the slope of each line gives k_{cat} / K_m values of 0.16, 0.49, 0.75 and 1.07 M^{-1} s^{-1} for RP3-42H, YSLP1-1, YSLP3-1 and YSLP6-1, respectively (Table I). These results clearly show that the polypeptides are evolving towards acquiring higher catalytic functions. A negative control (Yamauchi *et al.*, 1998) ruled out the possibility that the observed esterase activity is due to contamination of enzyme(s) present in the host cell.

Is the increase in activity over the generations a consequence of sequence differences? Figure 2a seems to indicate a tendency of the advantageous clones in incorporating Arg or His into several positions. The amino acid sequences were analyzed in depth to determine whether the two residues have any influence on the increase in the activity of the advantageous clones. The number of Arg and His with reference to the increase and decrease in number in all sequences of the six generations was counted. The clone in the sixth generations that has highest binding affinity showed only two additional Arg residues. The probability of the average number of Arg being incorporated

Table I. Kinetic parameters for the esterase activity of RP3-42H, YSLP1-1, YSLP3-1 and YSLP6-1

Protein	Observed k_{cat} / K_m (M^{-1} s^{-1}) ± SD	Calculated k_{cat} / K_m (M^{-1} s^{-1})	K_I (mM)
RP3-42H	0.16 ± 0.02	0.06	ND
YSL1-1	0.49 ± 0.01	0.07	26.67
YSL3-1	0.75 ± 0.01	0.08	2.24
YSLP6-1	1.07 ± 0.06	0.09	0.58

The observed k_{cat} / K_m was estimated from the data shown in Figure 5 using Equation 4 (see text). The calculated k_{cat} / K_m value was obtained assuming that the activity is the sum of the individual activities of the amino acid residues contained in each peptide. Esterase activities of various amino acids and their derivatives were measured. From the results, only the following k_{cat} / K_m values were used for calculations: His, 6.48×10^{-3} M^{-1} s^{-1}; Cys, 0.79×10^{-3} M^{-1} s^{-1}; Lys, 0.12×10^{-3} M^{-1} s^{-1}; and the N-terminal amino group, 0.28×10^{-3} M^{-1} s^{-1}. Amino acids yielding insignificant k_{cat} / K_m values were neglected. K_I for TSA was calculated using the slope of the linear lines in Figure 6b and the observed k_{cat} / K_m. $k_B = 2.3\times10^{-6}$ s^{-1}. ND indicates that inhibition by 0.93 mM TSA was not observed under the assay conditions employed.

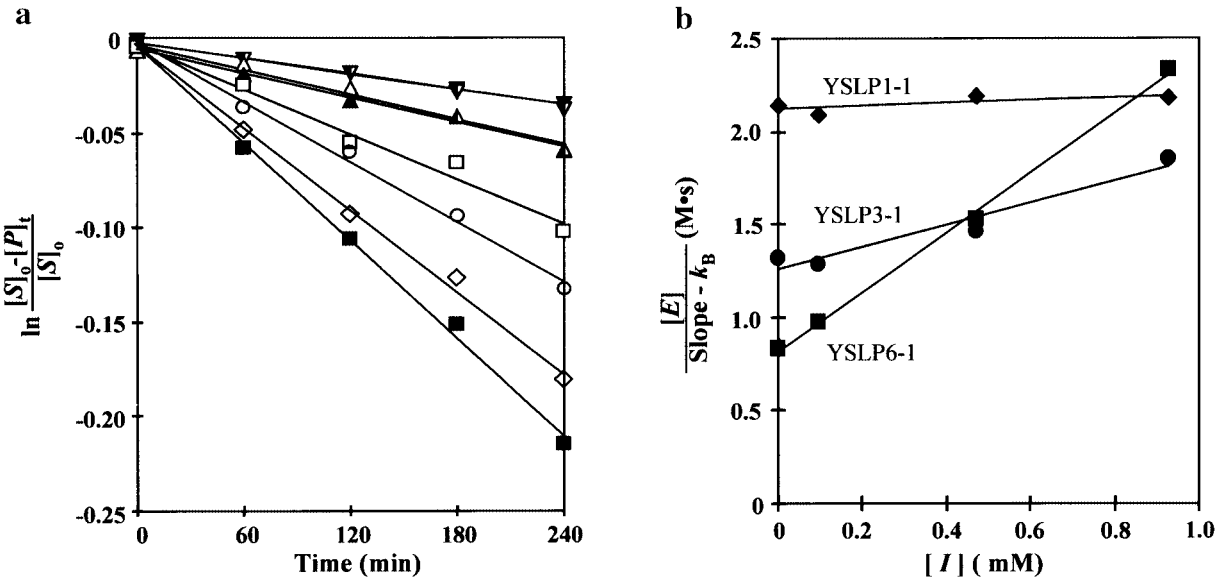

Fig. 6. Effect of TSA on the esterase activity of polypeptides. Esterase activity was measured as described in Materials and methods in the presence and absence of the TSA. Polypeptide concentration ([*E*]) was 10 μM. (**a**) Course of the hydrolysis of *p*-nitrophenyl acetate in the presence (open symbol) or absence (filled symbol) of TSA. ▼, background; {▽}, background (0.93 mM TSA); ▲, RP3-42H; △, RP3-42H (0.93 mM TSA); ■, YSLP6-1; ◇, YSLP6-1 (0.094 mM TSA); ○, YSLP6-1 (0.47 mM TSA); □, YSLP6-1 (0.93 mM TSA). (**b**) Values for the ordinate were calculated from the slope of the linear lines as shown in (a), [*E*] value of 10 μM, and the k_B value of 2.3×10^{-6} s^{-1} given from Figure 5. ◆, YSLP1-1; ●, YSLP3-1; ■, YSLP6-1.

in any of the sequences in the sixth generation is 1 ± 1.47, and the level of significance in reference to all non-selected members within the six generations is within the standard deviation (SD). On the other hand, there are three additional His residues in the clone showing the highest binding affinity in the sixth generation. The probability of the average number of His being incorporated in any of the sequences is –0.75 ± 1.104. Therefore, there is a possibility for the His residue to play a role in the observed esterase activity. However, the increase in the observed activity is much larger than the calculated value, which is based on the assumption that the activity is the sum of the individual activities of amino acid residues contained in each polypeptide (Table I). The gradual increase in the activity predicted from the calculated values hence cannot justify the faster rate of increase observed experimentally. Accordingly, the increase in the activity is not due to the change in the amino acid composition but due to the change in the sequence.

Given that transition-state analogs serve as respective inhibitors to enzyme-catalyzed reactions and that catalytic efficiency is a function of binding affinity (Philipps *et al.*, 1992; Angeles *et al.*, 1993), inhibition of esterase activity by the addition of TSA was investigated. While TSA showed no influence on the activity of RP3-42H, an increase in inhibitory effect was observed towards reactions catalyzed by YSLP6-1 upon an increase in the concentration of TSA (Figure 6a). In addition, it was confirmed that the TSA addition does not affect the background reaction (Figure 6a). Applying equation (3) to the slopes of YSLP6-1 in Figure 6a, [*E*] / (slope – k_B) which corresponds to $\{(1 + [I] / K_I)K_m\} / k_{cat}$ is plotted against the concentration of the inhibitor. From the slope of the linear line of YSLP6-1 in Figure 6b, a K_I value was calculated using k_{cat} / K_m values in Table I. Furthermore, the inhibition constant K_I, which corresponds to the dissociation constant, K_d, between the polypeptides and the TSA, decreased significantly on generations, guaranteeing that there is an increase in the esterase activity of the selected polypeptides over the generations even when the polypeptides were detached from the phage. In addition, these results implied that the affinity assay, by panning to the TSA as a selection system, is reliable.

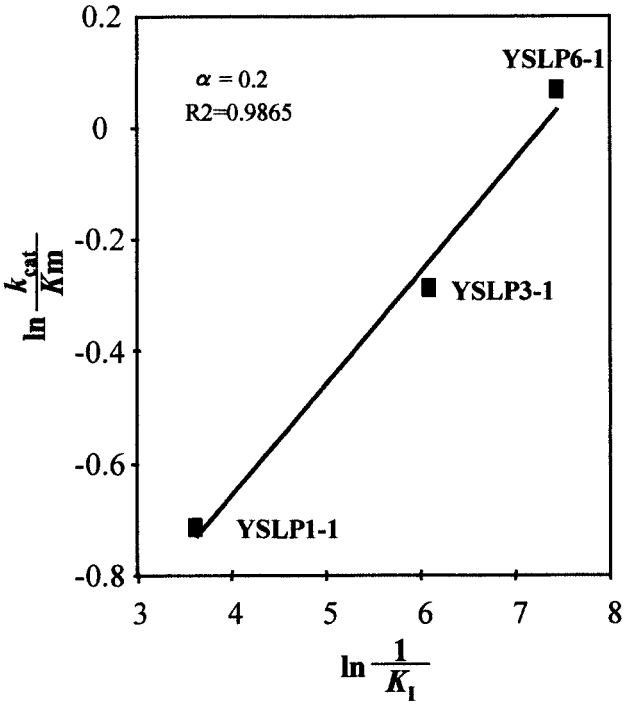

Fig. 7. Plot of TSA $1/K_I$ values versus k_{cat}/K_m for random polypeptides YSLP1-1, YSLP3-1 and YSLP6-1. Data represented in this figure were from Table I.

According to the transition state theory (Stewart and Benkovic, 1995), the more stabilized the transient state, the more enhanced is the catalytic activity of the polypeptide.

Hence, one may expect that the polypeptides with lower free energy change in binding to the TSA have a more stabilized transition state as compared with the substrate, and accordingly have lower activation free energy of the catalytic reaction (Stewart and Benkovic, 1995). Figure 7 shows a good correlation between the binding energy and the activation free energy of the polypeptides, in agreement with results obtained with an antibody-catalyzed reaction (Angeles *et al.*, 1993).

Here, we evidently showed that the evolution of random polypeptides with respect to esterase activity is being prompted even within a small population of polypeptides with arbitrary sequences. The ease of the polypeptide evolution with a small number of arbitrary sequences indicates that a significant fraction of all possible sequences may have functions, at least binding activity in correlation with catalytic activity. It is noteworthy to consider that as long as an assay system is available for a function of interest, the observation of a primitive polypeptide acquiring the function in the course of evolution is of no doubt possible. Furthermore, regardless of what the initial material was, be it a random polypeptide, or a new protein derived by DNA shuffling or combinations of pre-existing independent protein modules (Gilbert, 1978; Tonegawa, *et al.*, 1978; Go, 1981; Barnett *et al.*, 2000), a small variety of amino acid sequences are sufficient to bring about Darwinian evolution. Although the observed k_{cat} / K_m of the esterase activity of the random polypeptides at the primordial stage is still about three orders lower than the catalytic antibodies (Charbonnier *et al*., 1997; Nakashima, *et al*., 2000), a continuous increase is expected as its evolution progresses. In addition, this study provides promise for the observation of the gradual evolutionary process from the random polypeptide with highly flexible structure to a uniquely folded structure by functional selection as theoretically proposed (Saito *et al.*, 1997; Yomo *et al.*, 1999), and for obtaining intermediate polypeptides. The availability of intermediate polypeptides evolving to functional proteins at different stages of the evolutionary process, especially where initial stages are concerned, will expand the horizon for exploring the structure–function relationship of proteins.

Acknowledgements

The authors thank Dr Takashi Sagawa (Kyoto University) for his technical support in experiments with the TSA analog. 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

Angeles,T.S., Smith,R.G., Darsley,M.J., Sugasawara,R., Sanchez,R.I., Kenten,J., Schultz,P.G. and Martin,M.T. (1993) *Biochemistry*, **32**, 12128–12135.

Arakawa,T. *et al.* (1996) *DNA Res.*, **3**, 87–92.

Arnold,F.H., Wintrode,P.L., Miyazaki,K. and Gershenson,A. (2001) *Trends Biochem. Sci.*, **26**, 100–106.

Baca,M., Scanlan,T.S., Stephenson,R.C. and Wells,J.A. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 10063–10068.

Barnett,P., Tabak,H.F. and Hettema,E.H. (2000) *Trends Biochem. Sci.*, **25**, 227–228.

Charbonnier,J.B., Golinelli-Pimpaneau,B., Gigant,B., Tawfik,D.S., Chap,R., Schindler,D.G., Kim,S.H., Green,B.S., Eshhar,Z. and Knossow,M. (1997) *Science*, **275**, 1140–1142

Dunn,I.S. (1996) *Curr. Opin. Biotechnol.*, **7**, 547–553.

Forrer,P., Jung,S. and Pluckthun,A. (1999) *Curr. Opin. Struct. Biol.*, **9**, 514–520.

Fujii,I., Fukuyama,S., Iwabuchi,Y. and Tanimura,R. (1998) *Nature Biotechnol.*, **16**, 463–467.

Gilbert,W. (1978) *Nature*, **271**, 501.

Go,M. (1981) *Nature*, **291**, 90–92.

Kashiwagi,A., Noumachi,W., Katsuno,M., Alam,M.T, Urabe,I. and Yomo,T. (2001) *J. Mol. Evol.*, **52**, 502–509.

Keefe,A.D. and Szostak,J.W. (2001) *Nature*, **410**, 715–718.

Levitan,B. and Kauffmann,S. (1995) *Mol. Diversity*, **1**, 53–68.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*, 1st edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Marks,J.D., Hoogenboom,H.R., Bonnert,T.P., McCafferty,J., Griffiths,A.D. and Winter,G. (1991) *J. Mol. Biol.*, **222**, 581–597.

Matsuura,T., Yomo,T., Trakulnaleamsai,S., Ohashi,Y., Yamamoto,K. and Urabe,I. (1998) *Protein Eng.*, **11**, 789–795.

Matsuura,T., Miyai,K., Trakulnaleamsai,S., Yomo,T., Shima,Y., Miki,S., Yamamoto,K. and Urabe,I. (1999) *Nature Biotechnol.*, **17**, 58–61.

Nakashima,T., Ishiguro,N., Yamaguchi,M., Yamauchi,A., Shima,Y., Nozaki,C., Urabe,I. and Yomo,T. (2000) *J. Biosci. Bioeng.*, **90**, 253–259.

Ohkubo,K., Urata,Y., Seri,K., Ishida,H., Sagawa,T., Nakashima,T. and Imagawa,Y. (1994) *J. Mol. Catal.*, **90**, 355–365.

Patten,P.A., Gray,N.S., Yang,P.L., Marks,C.B., Wedemayer,G.J., Boniface,J.J., Stevens,R.C. and Schultz,P.G. (1996) *Science*, **271**, 1086–1091.

Phillips,M.A., Kaplan,A.P., Rutter,W.J. and Bartlett,P.A. (1992) *Biochemistry*, **31**, 959–963.

Prijambada,I.D., Yomo,T., Tanaka,F., Kawama,T., Yamamoto,K., Hasegawa.A., Shima.Y., Negoro,S. and Urabe,I. (1996) *FEBS Lett.*, **382**, 21–25.

Saito,S., Sasai,M. and Yomo,T. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 11324–11328.

Stewart,J.D. and Benkovic,S.J. (1995) *Nature*, **375**, 388–391.

Suga,Y., Yomo,T. and Urabe,I. (1996) *J. Ferment. Bioeng.*, **81**, 259–261.

Tonegawa,S., Maxam,A.M., Tizard,R., Bernard,O. and Gilbert W. (1978) *Proc. Natl Acad. Sci. USA*, **75**, 1485–1489.

Trakulnaleamsai,S., Yomo,T., Yoshikawa,M., Aihara,S. and Urabe,I. (1995) *J. Biosci. Bioeng.*, **79**, 107–118.

Wilson,D.S., Keefe,A.D. and Szostak,J.W. (2001) *Proc. Natl Acad. Sci. USA*, **98**, 3750–3755.

Yamauchi,A. *et al.* (1998) *FEBS Lett.*, **421**, 147–151.

Yomo,T., Saito,S. and Sasai,M. (1999) *Nature Struct. Biol.*, **6**, 743–746.

Received October 10, 2001; revised March 19, 2002; accepted March 23, 2002