

Evolution of an Arbitrary Sequence in Solubility

Yoichiro Ito,¹ Toshihiro Kawama,¹ Itaru Urabe,¹ Tetsuya Yomo^{1–4}

¹ Department of Biotechnology, Graduate School of Engineering, Osaka University, 2-1 Yamada-oka, Suita City, Osaka 565-0871, Japan

² Department of 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 Yamada-oka, Suita City, Osaka 565-0871, Japan

⁴ Department of Pure and Applied Sciences, University of Tokyo, Komaba, Meguro-ku, Tokyo, 153-8902, Japan

Received: 20 May 2003 / Accepted: 21 August 2003

Abstract. We have investigated the evolvability of an insoluble random polypeptide, RP3-34, to a soluble form through iterative mutation and selection with the aid of the green fluorescent protein (GFP) folding reporter. To assess the solubility of the polypeptides in the selected clones of each generation, the polypeptide genes were detached from the GFP fusions and expressed with a His₆ tag. The solubility of the variant random polypeptides increased in each generation within the scope of the evolutionary process, and the polypeptides assumed a soluble form from the fourth generation. Analysis of the synonymous and nonsynonymous mutations found in the deduced amino acid sequence of the selected polypeptides revealed that selection had accelerated the evolutionary rate. The solubility and hydrophobicity of the polypeptides and the 25 arbitrarily chosen random polypeptides found in a previously prepared library were determined, analyzed, and interpreted from the landscape on the protein sequence space. This study showed the evolvability of an insoluble arbitrary sequence toward a soluble one, hence, it provides a new perspective on the field of artificial evolution.

Key words: Artificial random polypeptide — GFP-based evolution — Hydrophobicity — Landscape — Protein sequence space — Protein solubility

Introduction

Solubility is one important aspect for the biological functions of protein in either the soluble or insoluble state. Protein solubility is generally attributed to the hydrophobicity of its sequence. The soluble proteins were distinguished from the multihelical membrane proteins based on hydrophobicity but not from the single-helical membrane proteins (Yanagihara et al. 1986). Based on the index of Kyte and Doolittle (1982), the average hydropathy indexes of the analyzed natural proteins in the NBRF database fell across the narrow range of -1.5 to $+1.5$ (Yanagihara et al. 1986), far narrower than the maximum range of -4.5 (for Arg) to $+4.5$ (for Ile). Restriction of the hydrophobicity of natural proteins to within a narrow range could have been dictated by the needs of a variety of amino acid residues to assume unique conformations with specific functions. Moreover, the intershifting of protein solubility between the soluble and the insoluble forms could have been controlled within that narrow range of hydrophobicity. This suggests that solubility may change due to small changes in the hydrophobicity of the protein brought about by mutation of the amino acid resi-

Correspondence to: Tetsuya Yomo, Department of Biotechnology, Graduate School of Engineering, Osaka University, 2-1 Yamada-oka, Suita City, Osaka 565-0871, Japan; email: yomo@bio.eng.osaka-u.ac.jp

dues. Protein solubility as one of the elements contributing to its function, hence, may have worked as a selection pressure during the process of protein evolution.

Recently, we have demonstrated that artificial polypeptides with random sequences of about 140 amino acid residues (Priambada et al. 1996; Yamauchi et al. 1998) have the capacity to evolve toward acquiring biological functions such as an esterase activity (Yamauchi et al. 2002) and phage infectivity (Hayashi et al. 2003), where the latter emerged from an arbitrarily chosen soluble random polypeptide. The evolvability of a soluble arbitrary sequence hence permits room to accommodate the possibility of an evolutionary route initiating from any of the insoluble sequences.

In this work, we used an insoluble random polypeptide, RP3-34, as the initial sequence for the intended green fluorescent protein (GFP)-based evolutionary study. RP3-34 is composed of 149 amino acid residues and has no homology with any known natural proteins in the SwissProt database as analyzed by BLAST 2.2.2. It was arbitrarily chosen from 20 insoluble random polypeptides found in a previously prepared library (Priambada et al. 1996). Here, we show that an insoluble arbitrary sequence can evolve to a soluble form through iterative mutation and selection, which is based on the fluorescence emitted by the GFP folding reporter (Waldo et al. 1999). In addition, the study was extended to the analysis of the hydrophobicity in relation to solubility of the polypeptides and of the 25 random polypeptides obtained previously (Priambada et al. 1996). Interpretation of the data by means of the landscapes on the protein sequence space is presented.

Materials and Methods

Bacterial Strains, Plasmids, and GFP Mutants

Escherichia coli strains used in this study were DH5 α (DE3) and KP3998 (Miki et al. 1987). *E. coli* DH5 α (DE3) was prepared by infecting *E. coli* strain DH5 α with λ DE3 phage using the λ DE3 Lysogenization Kit (Novagen). The *E. coli* KP 3998 was a generous gift from Dr. Takeyoshi Miki (Kyushu University). A library of hybrid plasmids containing genes encoding the random polypeptides in the multicloning site of pEOR was prepared previously (Priambada et al. 1996). A plasmid pET21aSH (Yamauchi et al. 2002) was used for expressing random polypeptides with a C-terminal His₆ tag, while pETHLGT1, constructed as described below, was used for expressing random polypeptides fused with GFPuv5, a GFP variant. The GFPuv5 gene was prepared by mutating the GFPuv4 gene (Ito et al. 1999) to replace Ile-167 with Thr and to eliminate the *Bam*HI and *Nde*I sites without changing the amino acid sequence. When GFPuv5 was expressed in *E. coli* DH5 α , the whole-cell fluorescence was about 1.2 times brighter than that of GFPuv4, the mutant with the highest fluorescence in a previous work (Ito et al. 1999).

Construction of pETHLGT1

The oligonucleotide sequence 5'-GGATCCCAGGGCCTCTGGGGCCGCACACCACCACCACCACCACGGCGGT-3' (underscores indicate *Bam*HI and *Sfi*I sites, respectively, and italic characters represent the linker sequence coding for the amino acid sequence of AGGAAHHHHHHGG) followed by the GFP gene was prepared by PCR and inserted into the *Bam*HI/*Eco*RI sites of pET21a(+) (Novagen). The *Nhe*I/*Sfi*I fragment of the resultant plasmid was replaced with a T1 terminator DNA fragment obtained by PCR with plasmid pPROTet.E133 Vector (Clontech) as a template, and oligomers 5'-TCTGCAGCTAGCAGAGGCATCAAATAAAAC-3' and 5'-TGCTGAGGCCACAGAGGCCTCTAGGGCGGCGGATT-3' (underscores indicate *Nhe*I and *Sfi*I sites, respectively) as the primers, yielding plasmid pETHLGT1, on which *Nhe*I/*Sfi*I sites become accessible for the insertion of target polypeptide genes. The T1 terminator was inserted in front of the GFP-coding region as a transcriptional stop to avoid transcriptional leakage from the T7 promoter. The fluorescence intensity of GFP fused with a target polypeptide was used as the index for solubility of the polypeptide (Waldo et al. 1999).

Random Mutagenesis and Selection

The artificial evolution was initiated with an arbitrarily chosen insoluble random polypeptide, RP3-34, fused with GFP Error-prone PCR was applied for the mutagenesis of the gene of the parent polypeptide for each generation with primers 5'-CTCAGCCATATGGCTAGCATGACTGGTGGACAGCAAATGGGT-3' and 5'-AGTTTAGGCCACAGAGGCCTGATCGGATCTGTCGACTC-3' (underscores show *Nhe*I and *Sfi*I sites, respectively). The 1st to 5th mutageneses were performed with Δ Tth DNA polymerase as described by Arakawa et al. (1996), while the 6th to 10th were done with GeneMorph Mutagenesis Kit (Stratagene), following the manufacturer's protocol. The PCR products thus obtained were separated by agarose gel electrophoresis, and the DNA fragments corresponding to about 500 bp were isolated with Gel Extraction Kit (Qiagen). The fragments were then digested by *Nhe*I/*Sfi*I and ligated with *Nhe*I/*Sfi*I-digested pETHLGT1, and the resultant hybrid plasmids were used to transform *E. coli* DH5 α (DE3) cells. The resultant transformants, comprising the mutant library of each generation, were grown at 37°C for 19 h on LB plates containing 75 μ g/ml ampicillin.

The selection process consists of the following three screening steps. The first screening involved the selection of about 30 colonies emitting high green fluorescence on eye view under the fluorescent light from the approximately 2000 transformants obtained above. The second screening involved estimating the whole-cell fluorescence of the selected transformants using a Hitachi F-2000 spectrofluorometer ($\lambda_{\text{ex}} = 488$ nm, $\lambda_{\text{em}} = 510$ nm, with both $\Delta\lambda = 10$ nm). Each of the selected transformants was grown at 37°C in an LB medium with ampicillin (75 μ g/ml) to an OD₆₆₀ = 0.3 before the addition of 1 mM isopropylthiogalactoside (IPTG). After the 3-h induction, the cells were harvested by centrifugation and resuspended in phosphate-buffered saline (PBS), such that the cell density was about OD₆₆₀ = 0.2. The fluorescence of the cell suspension was then measured by spectrofluorometry, after which five to eight clones with high whole-cell fluorescence were selected. The nucleotide sequences of the random polypeptide genes of these clones were analyzed, and those clones containing genes without any mutations were again selected and termed as semiselected clones hereafter. In the last screening process, the *Nhe*I/*Sfi*I fragments containing the variant random polypeptides of the semiselected clones were recloned

independently into a fresh pETHLGT1 previously digested with *NheI/SfiI* to ensure that the sequence of the mutants varies only in the random polypeptide gene of the hybrid plasmid. The whole-cell fluorescence of three independent colonies resulting from each transformation of the *E. coli* DH5 α (DE3) containing each of the hybrid plasmids were measured as described above. The clone with the highest average value was then selected as the parent for the next generation.

Expression of Variant Random Polypeptides and Solubility Assay

The *NheI/SfiI* fragment containing a variant random polypeptide gene in pETHLGT1 carried by a selected clone was isolated and recloned into pET21aSH previously digested with *NheI/SfiI*, and the resultant hybrid plasmid was used to transform *E. coli* DH5 α (DE3) cells. The variant polypeptide was expressed in the cell by IPTG induction as described above. The 25 random polypeptides arbitrarily chosen in our previous work from a library of random polypeptides (Priambada et al. 1996) were expressed in *E. coli* KP3998 cells after each gene was recloned to a pEOR vector as described by Priambada et al. (1996) with the slight modification that IPTG induction for the expression of the 25 random polypeptides was carried out for 2 h.

Solubility of the polypeptides was determined by SDS-PAGE throughout and the fraction soluble was estimated from the amount of the target polypeptide in the soluble fraction (D_s) over the total amount of expressed target polypeptide (D_T) (Waldo et al. 1999). The total amount of the expressed target polypeptide was estimated by summing the amount of the polypeptide in the soluble (D_s) and insoluble (D_i) fractions. The soluble fraction comprised the supernatant obtained from the centrifuged, sonicated sample of resuspended cell pellets collected from a 3-ml culture, while the pellet collected comprised the insoluble fraction. Both fractions were denatured by boiling with SDS sample dye containing mercaptoethanol before being subjected to SDS-PAGE on a 15% gel. The proteins bands were visualized by Coomassie brilliant blue R250 staining.

Results and Discussion

The experimental evolution was initiated with an arbitrarily chosen insoluble random polypeptide of 149 amino acid residues, RP3-34, fused with GFP. From the mutant library of about 2000 clones of the first generation prepared by random mutagenesis of the RP3-34 gene, we selected the clone with the highest fluorescence through the three-step screening process described under Materials and Methods and used it as the parent clone for the second generation. The same mutation and selection cycle was carried out for the succeeding generations. The increase in the whole-cell fluorescence of the selected clone is evident in each generation (Fig. 1A). Nevertheless, the amount of the expressed variant random polypeptide in each of the selected clones varies by an average of 108 ± 8.4 , as measured by densitometry of the intensity of the target polypeptide band of the whole cell of each clone on SDS-PAGE gel. Furthermore, it should be noted that the cell concentration of each clone was adjusted to the same OD₆₆₀

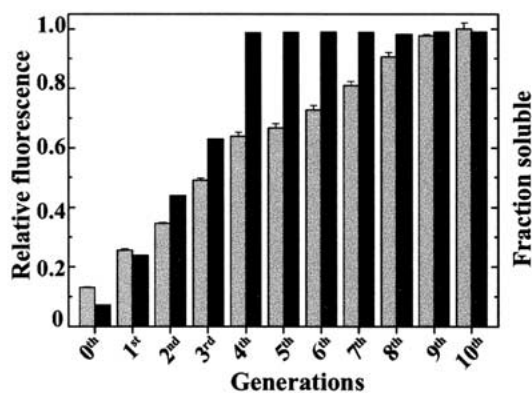
before the SDS-PAGE analysis on the expression of the polypeptide. These results indicate that the increase in whole-cell fluorescence can represent the increase in the level of fluorescence of the fused GFP molecule in the cell brought about by the directed evolution.

To assess the solubility of the target polypeptides from the selected clones of the experimental evolution, the polypeptide genes of the GFP fusion were isolated and cloned to the pET21aSH expression vector. The expressed His₆-tagged polypeptides of the 0th to the 10th generations were named RP3-34H, ITP1-1, ITP2-1, ITP3-1, ITP4-1, ITP5-1, ITP6-1, ITP7-1, ITP8-1, ITP9-1, and ITP10-1, respectively, and the respective deduced amino acid sequences are listed in Fig. 1B. An increase in the solubility of the variant polypeptides was apparent in the evolutionary process (Fig. 1A), in good agreement with the increase in the fluorescence intensity of the corresponding GFP fusions. From the fourth generation, all the variant polypeptides in the selected clones were soluble ones. These results clearly show that an arbitrary sequence of an insoluble polypeptide can evolve toward a soluble polypeptide. Because a soluble arbitrary sequence has the capacity to evolve and acquire a new function (Hayashi et al. 2003), the possibility exists that an insoluble arbitrary sequence can evolve in the same way, channeling through the routes of insoluble to soluble sequences.

The numbers of synonymous and nonsynonymous mutations in all the selected clones and the semiselected clones (see Materials and Methods) are listed in Table 1. The ratios of the total average numbers of nonsynonymous to synonymous mutations were 6.6 for the selected clones and 2.7 for the semiselected clones. However, our previous work showed that the ratio for the selected clones was 2.4 and was similar to that for the nonselected clones (1.8), indicating that the functional selection during the evolutionary process did not distinctly accelerate or decelerate the evolutionary rate (Hayashi et al. 2003). On the contrary, the high ratio of selected clones compared to those of the semiselected and nonselected clones suggests that the selection imposed in this study had accelerated the evolutionary rate. The difference in the resultant ratios in the two studies may be due to the fact that the best clone of each generation in this study was selected from a library of approximately 2000, a population far greater than that accessed in the previous study, which involved the selection of the best clone from a very small library of 6 to 10. This implies that at the primitive stage of evolution, the selection of the best clone in each generation from a larger population will drive the acceleration of the evolutionary rate.

The study was also extended to the analysis of the hydrophobicity of polypeptides. The plot of the

A.



B.

RP3-34H	1:	MASMTGGQQMGRGSRGCSNLGDPQFGDIPVWKTPIILGVLQACKPPSWEPLQLELQDTPSLENSQNGVHPWURLPKM	75
ITP1-1	1: P	75
ITP2-1	1: G P G A R	75
ITP3-1	1: G SP V G A . R R	75
ITP4-1	1: V SP V G APR R	75
ITP5-1	1: V SP V G APR R	75
ITP6-1	1: V SPE V G APR H . R	75
ITP7-1	1: VE SPE V G APR H . R S	75
ITP8-1	1: IE SPE V G APR H . R S	75
ITP9-1	1: IE SPE V G APR H . R S	75
ITP10-1	1: IE SPD V G APR H . R S	75
RP3-34H	76:	EYLNQGGPLPNWIIQVGS LPCWGAPRLEALQVGG LPSWEVSNLETPCSWIHEIWIYRSRVD RSRDQASGA AHHHHHHH	149
ITP1-1	76: R	149
ITP2-1	76: K R	149
ITP3-1	76: K A P RG T	149
ITP4-1	76: K A A RG . T TR	149
ITP5-1	76: D K . P . A A D RG . T TR	149
ITP6-1	76: D K . P . A A D RG . T TR	149
ITP7-1	76: D K . P . A A C D RG . T TR	149
ITP8-1	76:	N D K . P . A A C D RG . T TR	149
ITP9-1	76:	N D T K . P . A A C D RG . T TR	149
ITP10-1	76:	N D T K . P . A A C D RG . T TR	149

Fig. 1. GFP-based evolution of an insoluble random polypeptide, RP3-34. **A** Relative whole-cell fluorescence of selected clones expressing GFP-fused polypeptide (gray bars) and solubility of the polypeptides detached from each of the selected GFP fusions (black bars). Solubility of the His₆-tagged polypeptides was estimated as described under Materials and Methods. The density of the corresponding protein bands of the soluble (supernatant) and

insoluble (precipitate) fractions, obtained from the sonicated 3-ml culture, on a scanned image of a 15% SDS-PAGE gel using a scanner in the transmission mode (ScanMaker 8700, MICROTEK), was analyzed by NIH Image (<http://rsp.info.nih.gov/nih-image/>).

B Deduced amino acid sequences of the selected polypeptides.

hydrophobicity of the selected polypeptides, calculated from the deduced amino acid composition (Fig. 1B), against the solubility (Fig. 2; triangles) shows that solubility monotonically increases with a decrease in hydrophobicity. To know whether such a clear correlation between the solubility and the hydrophobicity applies to all polypeptides in the global protein sequence space, we conducted the same analysis on 25 arbitrarily chosen polypeptides previously obtained from a library of random polypeptides (Priambada et al. 1996). The results showed that although the two parameters roughly correlate, there were many cases where a polypeptide with a higher solubility was more hydrophobic (Fig. 2, circles),

depicting a rugged solubility landscape on the protein sequence space.

Let the protein sequence space be drawn in terms of a landscape with the horizontal as the hydrophobicity arranged in the order of lowest to highest and the vertical as solubility (Fig. 3), and envision the results stated above as a landscape on the protein sequence space. The data obtained from the 25 random polypeptides then suggest a global landscape, as shown in Fig. 3, based on the fact that these polypeptides were arbitrarily chosen from a large library of random sequences. The landscape is a rugged terrain with its global slope of higher solubility with lower hydrophobicity representing the rough correlation of the

Table 1. Synonymous and nonsynonymous mutations found in the selected and semiselected clones in each generation

Generation	Selected clone ^a		Semiselected clones ^b		
	No. synonymous mutations	No. nonsynonymous mutations	No. clones	Average No. synonymous mutations	Average No. nonsynonymous mutations
1	2	5	4	2	4.5
2	0	3	5	3	6.2
3	0	7	7	2.7	9.1
4	1	6	6	1.3	4.5
5	1	3	7	1.9	3.1
6	0	2	6	0.2	1.8
7	0	3	5	0.4	1.6
8	0	2	5	0	1.2
9	1	1	3	0.7	1.3
10	0	1	3	0.7	0.3
Total average	0.5	3.3		1.4	3.8

^a Selected clone pertains to the clone with the highest whole-cell fluorescence in each generation and was the parent clone of the following generation.

^b Semiselected clones pertain to clones selected in the second screening during the selection process (see Materials and Methods for details).

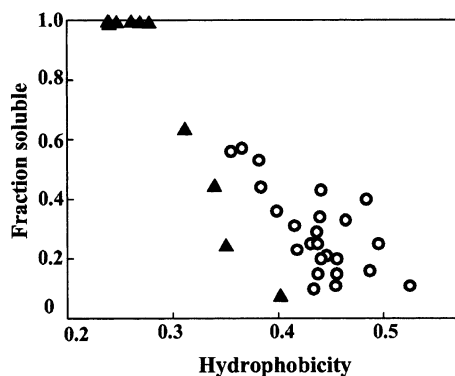


Fig. 2. Relationship between the solubility and the hydrophobicity of the polypeptides. Triangles denote His₆-tagged polypeptides detached from each of the selected GFP fusions and circles denote the arbitrary chosen 25 random polypeptides in our previous work (Priambada et al. 1996). Densitometry was analyzed by NIH image for the His₆-tagged polypeptides and by a densitograph system equipped with Lan&Spot Analyzer (ATTO Corp., Japan) for the 25 random polypeptides. The hydrophobicity was calculated based on the Fauchere and Pliska (1983) hydrophobicity index.

two parameters (Fig. 2; circles) and its ridges and valleys representing the many exceptional cases, such as those polypeptides with various solubilities but a similar hydrophobicity level (Fig. 2; circles). If sequences are then to be randomly sampled (Fig. 3; yellow stars) from such a rugged landscape, there is no doubt that similar relationship between the solubility and hydrophobicity will be observed. When we envisage on such a rugged global landscape the course of an evolution involving many consecutive selection steps on a local sequence space, the clear correlation between solubility and hydrophobicity of the selected clones (Fig. 2; triangles) in our experimental evolution may depict an evolutionary course that appears to be forced on a ridge of the landscape (Fig. 3; red arrows). That is, the imposed selection pressure could have tailored such evolutionary route of the selected poly-

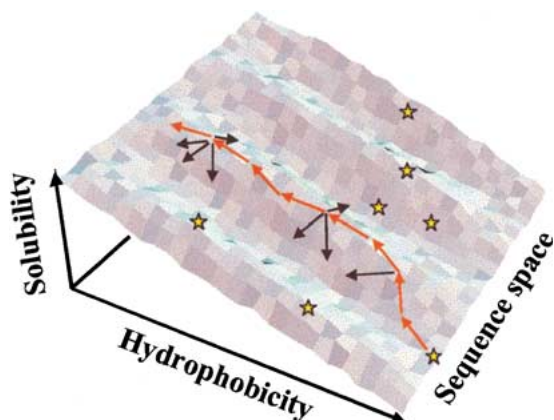


Fig. 3. An imaginary schematic landscape of the protein sequence space based on the hydrophobicity and solubility of the polypeptides. The colored block denotes the rugged landscape, comprising a global slope with ridges and valleys. Yellow stars indicate the locations of arbitrarily sampled polypeptides on the rugged global landscape. Red arrows denote the route taken by an insoluble arbitrarily chosen polypeptide in its evolution toward a soluble form, as in the case of RP3-34. Black arrows indicate possible outcomes of a local search on points within the evolutionary route under selection. See text for details.

peptides, which may correspond to the adaptive walk on a Mt. Fuji-type landscape (Aita and Husimi 2000). The observed monotonous increase in the property even on a rugged landscape guarantees the evolvability of the polypeptides.

Here we used GFP as a reporter for protein solubility. By using difference reporters, i.e., chloramphenicol acetyltransferase (Maxwell et al. 1999) and β -galactosidase (Wigley et al. 2001), we expect to observe different evolutionary routes. Therefore, it is interesting to see whether the routes lie along a ridge on the protein sequence space landscape. In addition, it is also of interest to analyze other landscapes drawn using factors other than hydrophobicity that are

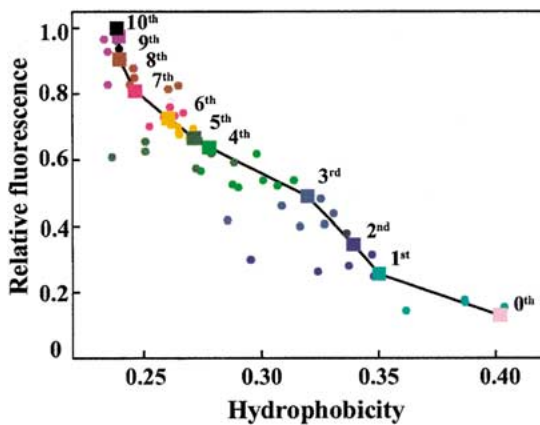


Fig. 4. Relationship between the whole-cell fluorescence of the selected and semiselected clones and the hydrophobicity of the polypeptides in the corresponding clones in each generation of the GFP-based evolution of the insoluble RP3-34. Squares denote the polypeptides of the selected clones, while circles denote those of the semiselected clones. Each color corresponds to clones of each generation indicated in Fig. 1A.

reported to affect protein solubility (Wilkinson and Harrison 1991). Furthermore, as the solubility of a natural protein is closely correlated to protein folding, such evolution in solubility may lead a random polypeptide to have a folded structure.

Although the local landscape of the selected polypeptides was smooth, a local search at any point along the selected route could reflect rugged terrain, as we expected there to be mutant sequences that possess a different relationship between their solubility and hydrophobicity, i.e., one could be less hydrophobic than the selected polypeptides (Fig. 3; black arrows). Such ruggedness in the local landscape is consistent with the results of the partial local search (second screening) of high-fluorescence clones at each generation. The plot of the fluorescence intensity of the semiselected and selected clones in the evolutionary process against the hydrophobicity of their corresponding polypeptides indicates that the polypeptide in the GFP fusion expressed in the clone with the highest fluorescence intensity in each generation is not always the least hydrophobic, particularly, those in the second, third, fourth, fifth, and ninth generations (Fig. 4).

The hydrophobicity of the understudied polypeptides was in the range of 0.2–0.6 (Fig. 2). A recalculation of the hydrophobicity values of these polypeptides using the index of Kyte and Doolittle (1982) yielded a range of -1.0 to -0.2 , which fell within that of the analyzed natural proteins found in the NBRF database: -0.8 to $+0.1$ for single-helical membrane proteins and -1.5 to $+0.5$ for soluble proteins (Yanagihara et al. 1986). The existence of both soluble and insoluble polypeptides in such a narrow hydrophobicity range suggests that any minor change in amino acid composition may cause

the polypeptides to approach the brink of a new form in terms of solubility. Hence, there is a possibility that the intershifting between the soluble and the insoluble forms will be observed in any artificial evolution under the selection pressure of a property other than solubility. In such cases, it will be effective to combine the use of solubility selection with an additional selection pressure to achieve efficient evolution.

We showed here that interpretation of the evolutionary process via the landscapes on the protein sequence space has provided relevant information on the evolvability of polypeptides even in a rugged landscape. Overall, we have demonstrated that an insoluble arbitrary sequence can evolve and become soluble. As soluble arbitrary sequences were proved evolvable toward acquiring new functions (Yamauchi et al. 2002; Hayashi et al. 2003), any insoluble sequence can also evolve likewise by first taking the routes from the insoluble to soluble sequences. This study, hence, provides a new perspective in the field of artificial evolution.

Acknowledgments. This work was supported in part by Grant 11CE2006 and a grant from The 21st Century Center of Excellence Program of the Ministry of Education, Culture, Sports, Science and Technology, Japan, and by a grant from the Rice Genome Project PR-2103, MAFF, Japan.

References

- Aita T, Husimi Y (2000) Adaptive walks by the fittest among finite random mutants on a Mt. Fuji-type fitness landscape II: Effect of small non-additivity. *J Math Biol* 41:207–231
- Arakawa T, Jongsareejit B, Tatsumi Y, et al. (1996) Application of N-terminally truncated DNA polymerase from *Thermus thermophilus* (delta Tth polymerase) to DNA sequencing and polymerase chain reaction: comparative study of delta Tth and wild-type Tth polymerase. *DNA Res* 3:87–92
- Fauchere JL, Pliska V (1983) Hydrophobic parameter π of amino-acid side chains from the partitioning of N-acetyl amino acid amides. *Eur J Med Chem* 18:369–375
- Hayashi Y, Sakata H, Makino Y, Urabe I, Yomo T (2003) Can an arbitrary sequence evolve towards acquiring a biological function? *J Mol Evol* 56:162–168
- Ito Y, Suzuki M, Husimi Y (1999) A novel mutant of green fluorescent protein with enhanced sensitivity for microanalysis at 488 nm excitation. *Biochem Biophys Res Commun* 264: 556–560
- Kyte J, Doolittle RF (1982) A simple method for displaying the hydrophobic character of a protein. *J Mol Biol* 157:105–132
- Maxwell KL, Mittermaier AK, Forman-Kay JD, Davidson AR (1999) A simple *in vivo* assay for increased protein solubility. *Protein Sci*, pp 1908–1911
- Miki T, Yasukochi T, Nagatani H, et al. (1987) Construction of a plasmid vector for the regulatable high level expression of eukariotic genes in *Escherichia coli*: An application to over production of chicken lysozyme. *Protein Eng* 1:327–332
- Prijambada ID, Yomo T, Tanaka F, et al. (1996) Solubility of artificial proteins with random sequences. *FEBS Lett* 382:21–25

- Waldo GS, Standish BM, Berendzen J, Terwilliger TC (1999) Rapid protein-folding assay using green fluorescent protein. *Nature Biotechnol* 17:691–695
- Wigley WC, Stidharm RD, Smith NM, Hunt JF, Thomas PJ (2001) Protein solubility and folding monitored in vivo by structural complementation of a genetic marker protein. *Nature Biotechnol.* pp 131–136
- Wilkinson DL, Harrison RG (1991) Predicting the solubility of recombinant protein in *Escherichia coli*. *Biotechnology* 9:443–448
- Yamauchi A, Yomo T, Tanaka F, et al. (1998) Characterization of soluble artificial proteins with random sequences. *FEBS Lett* 421:147–151
- Yamauchi A, Nakashima T, Tokuriki N, et al. (2002) Evolvability of random polypeptides through functional selection within a small library. *Protein Eng* 15:619–626
- Yanagihara N, Suwa M, Mitaku S (1989) A theoretical method for distinguishing between soluble and membrane proteins. *Biophys Chem* 34:69–77