Seeking high-priority mutations enabling successful antibody-breeding: systematic analysis of a mutant that gained over 100-fold enhanced affinity

Hiroyuki Oyama, Yuki Kiguchi, Izumi Morita, Chika Yamamoto, Yuka Higashi, Miku Taguchi, Tatsuya Tagawa, Yuri Enami, Yuriko Takamine, Hanako Hasegawa, Atsuko Takeuchi & Norihiro Kobayashi*

"Antibody-breeding" has provided therapeutic/diagnostic antibody mutants with greater performance than native antibodies. Typically, random point mutations are introduced into the VH and VL domains of parent antibodies to generate diverse libraries of single-chain Fv fragments (scFvs), from which evolved mutants are selected. We produced an scFv against estradiol-17β with 11 amino acid substitutions and a >100-fold improved affinity constant ($K_a = 1.19 \times 10^{10} \text{M}^{-1}$) over the parent scFv, enabling immunoassays with >30-fold higher sensitivity. We systematically analyzed contributions of these substitutions to the affinity enhancement. Comparing various partial scFv revertants based on their $K_a$ indicated that a revertant with four substitutions (VH-L100GQ, VL-I29V, -L36M, -S77G) exhibited somewhat higher affinity ($K_a = 1.46 \times 10^{10} \text{M}^{-1}$). Finally, the VH-L100GQ substitution, occurring in VH complementarity-determining region (CDR) 3, was found to be the highest-priority for improving the affinity, and VL-I29V and/or VL-L36M cooperated significantly. These findings encouraged us to reconsider the potential of VH-CDR3-targeting mutagenesis, which has been frequently attempted. The substitution(s) wherein might enable a "high rate of return" in terms of selecting mutants with dramatically enhanced affinities. The "high risk" of generating a tremendous excess of "junk mutants" can be overcome with the efficient selection systems that we developed.

As exemplified by the fact that pioneers in the field received the Nobel Prize in Chemistry in 2018,[1] the “antibody-breeding” approach (i.e., in vitro molecular evolution of antibody molecules) revolutionized the generation of therapeutic and diagnostic antibody agents.[2–5] Standard strategies[2–5] rely upon the introduction of random point mutations or site-directed mutations into the heavy and light chain variable (VH and VL) domains of a parent antibody to generate diverse libraries of mutated antibody fragments, e.g., single-chain Fv fragments (scFvs)[6,7] or Fab fragments. Mutated fragments with improved binding characteristics contained therein are then selected and isolated with the aid of genotype-phenotype linking technology,[7, e.g., phage display[4,9, ribosomal display[10, or yeast cell-surface display[11].

To date, we have produced affinity-matured scFv mutants binding to small biomarkers [e.g., estradiol-17β (E2)[12–14], cotinine[15], cortisol[16], and Δ9-tetrahydrocannabinol (THC)[17] to establish more sensitive immunochemical assay systems. Actually, these mutants enabled 3–100-fold higher affinity in competitive enzyme-linked immunoassays (ELISAs), compared with the corresponding parental scFvs (Fig. 1). Because these scFvs were generated via mutagenesis based on error-prone polymerase chain reaction (PCR) experiments[12–18], the missense mutations (causing amino acid substitutions) were introduced randomly. Consequently, some of them might function as “key mutations” for increasing in the affinity, whereas other mutations might be “junk mutations” that contribute little, nothing, or even decrease the affinity. Our previous results (Fig. 1) indicated the number of substitutions correlated with the extent of affinity improvement; thus, the anti-THC scFv with a single

Kobe Pharmaceutical University, 4-19-1, Motoyama-Kitamachi, Higashinada-ku, Kobe, 658-8558, Japan. *email: no-kobay@kobepharma-u.ac.jp
substitution gained 10-fold higher affinity (based on the calculated equilibrium affinity constant $K_a$), whereas the anti-E2 scFv with 11 substitutions exhibited a $K_a$ that was increased (improved) by $>100$-fold. The scFv mutants against cortisol and cotinine showed intermediate improvement, exhibiting $>30$-fold and $>40$-fold higher $K_a$s as the results of three and five substitutions, respectively. However, a reasonable explanation for such correlations requires evidence that most or many (if not all) of these multiple substitutions participated in increasing the affinity to at least some extent.

Regarding the positions of the substitution(s), a remarkable difference was found between the anti-cortisol scFvs (only in the VL) and the anti-THC (only in the VH) (Fig. 1). In contrast, with the anti-E2 and anti-cotinine scFvs that showed more significant improvements, the substitutions were spread over both the VH and VL domains. The most successful example of *in vitro* affinity maturation for an antibody against a small compound was reported for an scFv against fluorescein derivative ($K_a > 1 \times 10^{12}$ M$^{-1}$), which represented an affinity increase of $>2,000$-fold$^{19}$. This "super" mutant, which is almost beyond native antibodies, was the product of 14 different substitutions, 12 of which were located in the VH domain (Supplementary Fig. S1A).

Recently, some approaches have been developed for improving antibody functions via computational analysis, in order to minimize the trial and error that is an inevitable aspect of the conventional strategies$^{20-23}$. However, using an empirical approach toward more efficient mutagenesis strategies, e.g., targeting more limited regions (hotspots) with more controlled amino acid-substitution patterns is still important. To collect useful information, we sought the highest-priority substitutions for successful affinity maturation, among numerous randomly introduced multiple substitutions. We simultaneously pursued the greatest enhancement in affinity that was achievable with the fewest substitutions.
We selected our affinity-matured anti-E₂ scFv with a $10^{10}$-order $K_a$ and 11 substitutions as the subject of this study. Our novel approach employed here, i.e., systematic analysis using partial scFv revertants (scFvs with some substituted amino acids restored to the original sequence), revealed that the most critical substitution among the 11 was the one from leucine (L) to glutamine (Q) at the VH100g position, which occurred in complementarity-determining region (CDR) 3 of the VH domain (VH-CDR3). This mutation alone resulted in 17-fold enhanced $K_a$ compared with the wild-type scFv (i.e., the parent scFv without any artificial mutations).

Results

Origin, and structural and binding characteristics of target anti-E₂ scFv (scFv#M3rd). The anti-E₂ scFv focused on in this study, named scFv#M3rd(amb) here (this was originally reported as “scFv#m3-a18” (see Fig. 1A)), is an affinity-matured mutant that showed a $10^{10}$-range $K_a$ value against free (i.e., not immobilized) E₂ molecules. This scFv was our “third-generation mutant” that was generated previously in our laboratory after three iterative mutagenesis and selection steps performed on the wild-type scFv (scFv#WT)12–14 (Fig. 2A). scFv#WT was constructed by linking the VH and VL domains of a mouse anti-E₂ antibody (Ab#E4-4)12 via a common linker sequence composed of glycine (G) and serine (S) in the sequence of (GGGGS)$_6$ and attaching a FLAG tag24 at the C-terminus. The VH and VL domains contained 124 and 107 amino acids (Fig. 2A)12, which belonged to subgroups IIID and V25, respectively. In this study, we used the numbering and classifications defined by Kabat et al.25 The VH domain contained 15-residue CDR3, which is substantially longer than the average length of 8.7 residues for VH-CDR3 for mouse antibodies against any antigens26 or 8.50 residues for mouse antibodies against haptens27. Seven amino acids following the residue at position 100 (underlined in Fig. 2A) are defined as the inserted residues in the Kabat-rule, and were named 100a–100g. The $K_a$ values of each scFv, determined by the Scatchard analysis28, are shown together with those of the first- and the second-generation scFv mutants (scFv#M1st and #M2nd, respectively) in Fig. 2A. Previously, we showed that the $K_a$ values of each scFv, determined by the Scatchard analysis28, are shown together with those of the first- and the second-generation scFv mutants (scFv#M1st and #M2nd, respectively) in Fig. 2A.

Figure 2. (A) Summary of the process used for generating affinity-matured scFvs against estradiol-17β ($E_2$)12–14. Three steps of genetic evolution, i.e., scFv#WT (the wild-type scFv combining the VH and VL domains derived from a mouse anti-E₂ antibody)12 → scFv#M1st12–14 → scFv#M2nd13,14 → scFv#M3rd(amb)14, were performed, each of which involved random mutagenesis based on error-prone PCR and phage display-aided selection of improved species. The amino acid sequences of the wild-type VH and VL domains are shown in the purple box. The VH- and VL-CDRs, determined with the Kabat definition25 (H1, H2, H3, L1, L2, and L3), are shown with red and green, respectively. The $K_a$ values of each scFv, determined by the Scatchard analysis28, are shown together with the increasing magnitudes observed with each step. The primary scFv structures are schematically illustrated, where the new amino acid substitution(s) introduced during the first, second, and third mutagenesis steps is represented with a red, dark blue, and purple star(s), respectively. The amino acids before and after each substitution are indicated with the one-letter code. (B,C) Schematic illustration of the primary structures of the scFvs introduced with a reverse mutation(s) ([shown with magenta cross(es)]) for returning upstream (B) from scFv#M3rd(amb) to scFv#M2nd or (C) from scFv#M2nd to scFv#M1st. Two substitutions were simultaneously restored in scFv#R2-3. The downward double arrows (↓↓) mean >10-fold decrease in the affinity compared with the parent scFv before reverse mutation(s) was introduced.

We selected our affinity-matured anti-E₂ scFv with a $10^{10}$-order $K_a$ and 11 substitutions as the subject of this study. Our novel approach employed here, i.e., systematic analysis using partial scFv revertants (scFvs with some substituted amino acids restored to the original sequence), revealed that the most critical substitution among the 11 was the one from leucine (L) to glutamine (Q) at the VH100g position, which occurred in complementarity-determining region (CDR) 3 of the VH domain (VH-CDR3). This mutation alone resulted in 17-fold enhanced $K_a$ compared with the wild-type scFv (i.e., the parent scFv without any artificial mutations).
scFv#M3rd(amb) was specific enough and applicable for use with clinical specimens. The 11 amino acid substitutions in scFv#M3rd(amb) are located both in the V_{H} and the V_{L} domains (five and six substitutions, respectively), and both in CDRs and framework regions (FRs) (five and six substitutions, respectively).

In the scFv#M3rd(amb) gene variant, a T→A transversion introducing a nonsense mutation from TTTG [encoding L] to TG (amber termination codon) occurred in the V_{H}-100g-codon, which encodes the residue near the end of V_{H}-CDR3. Considering that we used Escherichia coli (E. coli) XL1-Blue as host cells, which is an supE suppressor strain, this amber codon was expected to be readthrough and translated as Q, and this was confirmed by liquid chromatography/tandem mass spectrometry (LC/MS/MS) fingerprinting of the affinity-purified scFv#M3rd(amb) protein (Supplementary Fig. S2). As further confirmation, we modified the scFv gene variant replacing the TAG codon at the position with CAG (encoding Q) via oligonucleotide-directed mutagenesis. The scFv gene variant scFv#M3rd(amb) protein (Supplementary Fig. S2). As further confirmation, we modified the scFv#M3rd(Q) protein (Supplementary Fig. S2). As further confirmation, we modified the scFv gene variant replacing the TAG codon at the position with CAG (encoding Q) via oligonucleotide-directed mutagenesis. The scFv#M3rd(Q) protein was expressed in E. coli and the purified protein was confirmed by liquid chromatography/tandem mass spectrometry (LC/MS/MS) fingerprinting of the affinity-purified scFv#M3rd(Q) protein (Supplementary Fig. S2).

Determining the substitutions responsible for enhanced affinity (the first stage). In this study, we employed the “elimination approach” to discover a minimum and essential set of substitutions responsible for the increased E2-binding affinities of scFv#M3rd(amb) or (Q) [denoted as scFv#M3rd(amb/Q), hereafter]. We first evaluated the affinity of partial scFv revertants, where one of the multiple substitutions was restored to the original amino acid. The $K_a$ values were determined by Scatchard analysis, using tritium-labeled E2, taking care to avoid inadequate estimations due to any additional structures like “bridges” that link with signal-groups or proteins necessary for immobilization.

We first analyzed the third mutagenesis step involving eight amino acid substitutions that converted scFv#M1st into #M2nd, resulting in a 2.4-fold higher $K_a$. We produced seven revertants (scFv#R2-1–#R2-7), each with a single restoration, except that scFv#R2-3 contained two restorations at the contiguous V_{H}84 and V_{L}85 positions (Fig. 2C). Considerably lowered affinity was found only for scFv#R2-5 and #R2-7 (i.e., 64-fold and 3.0-fold decreases, respectively), indicating that the methionine (M) at the VL36 and the G at the VL77 were responsible for the higher affinity. We note here that the substitutions excluding VL-L36M, found in scFv#M2nd, should have cooperatively reduced the affinity of its parent scFv, while the VL-L36M substitution maintained 47% of the affinity that observed for scFv#M3rd(amb). These results show that the substitution at the V_{H}50 was critical but the mutation at the V_{L}50 was not critical.

Then, we analyzed the second mutagenesis step involving eight amino acid substitutions that converted scFv#M1st into #M2nd, resulting in a 2.4-fold higher $K_a$. We produced seven revertants (scFv#R2-1–#R2-7), each with a single restoration, except that scFv#R2-3 contained two restorations at the contiguous V_{H}84 and V_{L}85 positions (Fig. 2C). Considerably lowered affinity was found only for scFv#R2-5 and #R2-7 (i.e., 64-fold and 3.0-fold decreases, respectively), indicating that the methionine (M) at the VL36 and the G at the VL77 were responsible for the higher affinity. We note here that the $K_a$ value of scFv#R2-5 (9.3 × $10^{-8}$ M$^{-1}$) was significantly lower than that of its parent scFv, scFv#M1st (2.6 × $10^{-8}$ M$^{-1}$). This finding suggests a situation where the 10 substitutions excluding V_{H}-L36M, found in scFv#M2nd, should have cooperatively reduced the affinity of the parent scFv (i.e., scFv#M1st) significantly (26-fold), but that the effect of the single V_{H}-L36M substitution on increasing the affinity predominated.

The first mutagenesis step provided scFv#M1st with a 3.0-fold higher $K_a$ over scFv#WT. This improvement should be due to a leucine (I) to valine (V) substitution at the V_{L}29 position, which was solely introduced during this step. Therefore, this mutation was estimated to be essentially responsible for the increased affinity of the third-generation mutant, which was confirmed later (see below).

Determining the substitutions responsible for enhanced affinity (the second stage). During the first-stage examination, we identified the following amino acid substitutions that must have been key for increasing the affinity: V_{H}-L100Q (introduced during the 3rd mutagenesis step), V_{L}-L36M and V_{L}-S77G (the 2nd mutagenesis step), and V_{L}-129V (the 1st mutagenesis step). However, substitutions selected in different mutagenesis stages do not always function cooperatively. Thus, we generated a new scFv containing these selected four substitutions, named scFv#4mut (Fig. 3A). This mutant showed an even slightly higher $K_a$ over scFv#WT. This improvement was significant compared to the expected $K_a$ value (27133) (Supplementary Fig. S2). Chemical identity between scFv#M3rd(Q) and #M3rd(amb) was supported by the LC/MS/MS fingerprinting data of the newly generated scFv#M3rd(Q) (Supplementary Fig. S2).
or V177 [G → S], named scFv#R3-1, #R3-2, #R3-3, and #R3-4, respectively, scFv#R3-1 exhibited a significantly (44-fold) decreased $K_a$ compared to scFv#4mut, whereas the decreases of 1.6–3.4-fold were observed with the other three revertants retaining the V H-L100gQ substitution (Fig. 3B). These observations strongly suggested that the VH-L100gQ substitution was most important for the dramatically increased affinity. Comparison of the $K_a$ values of scFv#R3-2, #R3-3, and #R3-4 suggested that the extent of cooperation with VH-L100gQ might be in the order of VL-I29V $\approx$ VL-L36M $>$ VL-S77G. This was further analyzed using three revertants with double restorations (Fig. 3C). Regarding the revertants that retained the VH-L100gQ substitution, the order of the $K_a$ was as follows: scFv#R4-2 (containing VL-L36M) $>$ #R4-1 (VL-I29V) $>$ #R4-3 (VL-S77G), all of which showed 109-order $K_a$ values. However, the other three revertants that lacked the VH-L100gQ substitution showed obviously lower affinity ($K_a = 1.3–4.1 \times 10^8$ M$^{-1}$).

Finally, we directly evaluated the contribution of the V H-L100gQ substitution by determining the $K_a$ of revertant scFv#R5-1 (Fig. 3D). The result ($K_a = 1.5 \times 10^9$ M$^{-1}$) indicated that the single V H-100gQ substitution improved the affinity by 17-fold, suggesting the great potential of performing V H-CDR3-directed mutagenesis anew, which has commonly been performed (often with simultaneous V L-CDR3-randomization), particularly in early antibody-engineering studies$^{31-34}$. Naturally, we became interested in the potential of substituting the other 18 amino acids (beside Q and the original L) at the V H100g position. Thus, we prepared 18 scFv additional mutants by replacing the V H100g amino acid of scFv#M3rd(amb) with one of the remaining 18 amino acids, and compared their $K_a$ values with those of the already-evaluated V H-100gQ and V H-100gL mutants. As shown in Fig. 3E, the Q-substituted mutant [i.e., scFv#M3rd(Q)] exhibited the most improved affinity, and the next highest was the mutant substituted with N (≈14-fold improvement over the V H-100gL mutant), a homolog of Q with an amide group but with one less carbon atom in the side chain. Moderate, but significant (>5-fold) improvements were observed for the G-, glutamic acid (E)-, H-, and arginine (R)-substituted mutants. Amino acids with aliphatic and hydrophobic side chains (I and V) or aromatic rings [phenylalanine (F), tryptophan (W), and tyrosine (Y)] contributed little to the improvement. It was surprising that substitution with S, though well-recognized as a residue (as well Y) that often plays important roles in interactions with antigens$^{35}$, deteriorated the affinity down to undetectable range, as also seen with the cysteine (C)-substituted mutant. We should also note that the Q substitution, which increases the binding affinity most, is not among the top 10 amino acids that frequently appear in the CDR sequences of mouse antibodies against haptens [i.e., Y, S, G, L, N, W, threonine (T), I, alanine (A), and R$^{32}$]: this suggests that it may be difficult to use a prediction-based approach for improving amino acid

**Figure 3.** Comparison of the affinities ($K_a$ values determined by the Scatchard analysis$^{28}$) between (A) scFv#M3rd(amb) and scFv#4mut having four substitutions, and between (B–D) scFv#4mut and its partial revertants retaining (B) three substitutions, (C) two substitutions, and (D) a single substitution. The restored substitutions are shown with magenta crosses. (E) The affinities ($K_a$) of 20 different scFv#M3rd(amb) variants, each of which had a different amino acid at the V H100g position (shown in abscissa with the one-letter code), were compared. ND means “not determined” because of too low affinities. The downward double arrows (↓↓) mean >30-fold decrease in the $K_a$ value compared with scFv#4mut.
sequences in this most diverse CDR. Nonetheless, these data (Fig. 3E) suggested to us the possibility that every position in VH-CDR3 might be substitutable with much more potent amino acids that cause dramatically enhanced affinity, but that it should rarely be achieved via error-prone-PCR-based mutagenesis.

Summary of affinity-maturation results from scFv#WT to scFv#M3rd(amb/Q). These findings enabled us to order the affinity-maturation process by focusing the importance of VH-L100gQ substitution, as shown in Fig. 4. The most improved mutant species with single, double, and triple substitutions were estimated to be scFvR5-1, #R4-2, and #R3-4, respectively, the $K_a$ values of which increased as the numbers of substitutions increased. Comparison of the $K_a$ values between scFvR3-1 and R3-3 (Fig. 3B), and between scFvR4-3 and R4-6 (Fig. 3C), indicates a greater potential of substitution at VH-L100gQ than VL-L36M under the presence of one or two other substitution(s). These observations were compatible with the result where scFvR5-2 with a single VL-L36M substitution exerted the $K_a$ value that was only 2.2-fold greater than scFv#WT and much lower than scFvR5-1. Quadruple substitutions in scFv#4mut were necessary to reach (and even exceed) the affinity of scFv#M3rd(amb/Q), the high-affinity mutant with 11-amino acid substitution. Therefore, assistance of VL-S77G substitution was necessary, although it was not as potent as that of VL-I29V and VL-L36M substitutions. After all, it was shown that, in scFv#4mut, the 4 substitutions functioned additively and showed a 170-fold higher affinity than that of scFv#WT.

Analytical utility and structural aspects of the high-affinity scFvs. Immunoassay sensitivities basically correlate with the affinities of the antibodies used, and usually antibodies with a higher affinity enable immunoassays with higher sensitivity. Indeed, scFv#M3rd(amb), #M3rd(Q), and #4mut (showing $K_a$ values in the $10^9$-range), as well as scFv#R3-4, #R4-1, #R4-2, and #R5-1 (showing $K_a$ values in the $10^9$-range), displayed dramatically enhanced sensitivities in competitive ELISAs, as shown by 8.0–17-fold lower midpoint values (~13–28 pg/assay) than scFv#WT (220 pg/assay) in dose–response curves (Fig. 5). These dose–response curves with improved sensitivities cover a measurable range required for clinical applications.

Protein modeling of the scFv#4mut and #WT, docked with E2, is shown in Fig. 6. Although such in silico approaches offer structural information with more or less limited reliability compared with X-ray crystallography, the modeling of scFv#4mut strongly suggested that none of the 4 substituted residues (Q at the VH100g position, or V, M, and G at the VL29, 36, and 77 positions, respectively) forms direct contacts with the E2 molecule in the immune complex. The Q residue was located near the C-terminus of VH-CDR3 and should function by raising the loop structure of this CDR. Substitution from the original L to Q is estimated to alter the steric conformation of VH-CDR3 loop, but the modeling did not suggest obvious interaction of VH-CDR3-related residues with E2 molecule. Instead, a possibility was shown where, in scFv#4mut, VH-CDR2 and VL-CDR1 might interact with E2. Thus, the hydroxy group of Y residue (at the VL32 in the CDR1) and the carboxy group of E residue (at the VH50 in the CDR2) might anchor E2 molecule via the hydrogen bonds with the hydroxy groups of the A- and D-ring of the steroid skeleton, respectively, resulting in drastic change of the orientation of the E2 molecule in the paratope compared with that in scFv#WT. The V (at the VL29) and/or M (at the VL36) residue(s) introduced in scFv#4mut might trigger such VL-CDR-dependent events.

We should note here that both the introduced M residue and original L residue are unusual amino acids as for the VL36 position: in fact, the residues at the VL35–41, which form the beginning of VL-FR2, are highly conserved in the sequence WYQQK(lysine)P(proline)G found in all VL subgroups. The G residue at the VL77 position is a member of FR3: the original amino acid was S, which frequently appears at this position. The contribution of this G was not significant as that of the V or M residue, but was essential for increasing the affinity to that observed with scFv#3rd(amb/Q). The mechanism whereby the G residue contributed to the affinity, despite its considerably long-distance from the paratope, is of great interest.
Figure 5. Dose–response curves of competitive ELISAs obtained with the scFvs shown in Fig. 4. The vertical bars indicate SD for intra-assay variance (n = 4). The midpoints of the curves (pg/assay) were as follows: scFv#4mut, 12.6 ± 0.12 [mean ± SD (n = 4)]; #M3rd(amb), 15.0 ± 0.76; #M3rd(Q), 16.8 ± 1.57; #R3-4, 14.3; #R4-2, 15.6; #R4-1, 19.0; #R5-1, 27.6; and #WT, 220 (average of determinations in duplicate). In these assays, the scFv concentrations were adjusted to give bound enzyme activities at B0 (the reaction without E2 standard) of approximately 1.0–1.5 absorbance after a 30-min enzyme reaction. The background absorbance (observed without addition of scFvs) was lower than 5.0% of the B0 absorbance. We should note that, in the original article where scFv#M3rd(amb) (denoted as scFv#m3-a18 therein) was generated14, we reported its $K_a$ value as $1.3 \times 10^{10}\text{M}^{-1}$, and the midpoint value in the ELISA using this scFv was determined to be 10.0 ± 1.2 pg/assay. In this study, we re-determined the $K_a$ in triplicate. The midpoint values were also re-determined to perform equal and strict comparisons between the scFvs, because we had to use a newly prepared E2–BSA conjugate to coat the ELISA microplates. Difference in the quality of these conjugates (mainly in the hapten/protein molar ratio) influences on the ELISA sensitivity and often makes it difficult to strictly reproduce previous experimental data.

Discussion

For those of us, working in the fields of analytical and diagnostic chemistry, antibodies with high affinity for target molecules are an essential tool. Because the higher affinity enables more sensitive analytical/diagnostic systems, “antibody-breeding” that generates mutant antibody fragments with improved affinity is an attractive methodology. However, the conventional approach combining random mutagenesis and panning-based selection has very often failed to provide satisfactorily improved mutants, despite much time consuming effort. To overcome such challenges, revolutionary strategies are required for efficiently introducing functional mutations without unnecessarily enlarging the diversity and for reliably selecting rare and improved mutants without overlooking them.

To achieve the former requirement, information is needed for designing “decisive mutations (substitutions)” essential for elevating the antigen-binding affinity. It would also be of great help if it were possible to discover “hot spots” in the V_{H} and/or V_{L} domains of antibodies, where a wide range of mutations introduced thereto significantly improves the affinity. Of course, it is more desirable if fewer mutations introduced in a narrower region facilitate isolating improved species with high probability.

Previously, we summarized the results of studies wherein antibody mutants were generated that bound to small molecules (haptens)15-18. Among the affinity-matured products reported, mutant scFvs or Fabs that showed $K_a$ values greater than $1 \times 10^9\text{M}^{-1}$ (i.e., a standard value required for subpicomole-order analysis), owing to >10-fold enhancement, were selected and their structures are illustrated in Supplementary Fig. S1A. Four out of the six mutants generated, including the scFv with a $K_a$ value greater than $1 \times 10^{10}\text{M}^{-1}$ and the greatest improvement19, involved totally 10 or more substitutions and some of them were in V_{H}-CDR3, similar to our scFv#M3rd(amb). Because two of these four mutants (i.e., anti-fluorescein-biotin and anti-fluorescein antibody fragments) were generated via error-prone PCR15 or related methods, only some of the multiple substitutions introduced might have driven the enhanced affinity. Seeking substitutions that do in fact contribute to the binding affinity is inevitable for developing simpler strategies that depend less on trial and error.

Thus, we analyzed our anti-E2 scFv#M3rd(amb): this mutant had 11 amino acid substitutions and showed an extremely high ($10^{10}\text{-order}$) $K_a$ against free E2 molecules, which was over 100-fold higher than that of the parent antibody (scFv#WT). To explore the hierarchy of these substitutions regarding contribution to the enhanced affinity, we employed a unique approach based on the comparison of the affinities of various partial scFv revertants. Although this case study was performed with a particular antibody mutant, we obtained several suggestive results that exceeded our initial expectations, as summarized below. First, a mutant with only four substitutions showed even higher affinity than scFv#M3rd(amb), i.e., the $K_a$ was $1.46 \times 10^{10}\text{M}^{-1}$, corresponding to a 170-fold improvement compared to scFv#WT. The incorporated amino acids (and positions) were Q (V_{H}100g), V (V_{L}29), M (V_{L}36), and G (V_{L}77). Second, the most influential residue was Q at V_{H}100g (locates in V_{H}-CDR3), which alone improved the affinity by 17-fold. Third, none of the substitutions at V_{H}100g with the remaining 19 different amino acids resulted in an equivalently improved affinity. Fourth, the extent that these substitutions contributed to the enhanced affinity was in the order of $Q \gg M \geq V \gg G$, and these four substitutions seemed to function in an additive manner. Consequently, the mutant with three substitutions (Q/M/V residues) showed 63% of the affinity observed with the mutant with four substitutions, and the mutant with two substitutions, Q/M or Q/V,
exhibited 36% or 29% affinity, respectively. These mutants with Q/M/V, Q/M, and Q/V substitutions demonstrated 107-fold, 62-fold, and 49-fold higher affinity than the parent scFv#WT, maintaining 109-order $K_a$ value.

These findings impressed upon us the importance of identifying a decisive substitution (‘ace’ substitution) like the VH-L100gQ substitution described above. Considering the mutation patterns found in other reported high-affinity mutants as well (Supplementary Fig. S1A), the VH-CDR3 should be focused on as a “hot region” where such an ace substitution might be discovered with high probability. VH-CDR3 often works as the “ace CDR” in antigen recognition36,37, and thus at the dawn of antibody engineering (~1990), “hard randomization”39 was frequently performed for multiple amino acid residues therein (often together with VL-CDR3) by the site-directed introduction of degenerated NNS [i.e., (A/C/G/T)(A/C/G/T)(C/G)] or NNK [i.e., (A/C/G/T)(A/C/G/T)(G/T)] codons encoding any of 20 different amino acids32,33. This was actually a potent strategy for generating new prototype antibody fragments that gained different specificities, but was unlikely to be suitable for improving the affinity while maintaining the original specificity. In fact, our previous affinity-maturated scFvs against cotinine, cortisol, and THC did not contain substitutions in VH-CDR3 (Fig. 1B–D).

Such outcomes should be mainly attributable to the highly diversified structure and nature of VH-CDR3, each with different binding specificity and affinity function to allow for “the best fitting” against only a limited antigen structures. Therefore, the “hardly” mutagenized libraries only rarely generate satisfactorily improved mutants, which, furthermore, should be buried by tremendously large excess of mutants with deteriorated binding performance. The most popular selection systems, i.e., those combining phage display and panning do not always facilitate successful isolation of such rare and hidden species, mainly due to the biased propagation of phage clones displaying antibody fragments and the competition with a large excess of the phage clones displaying antibodies.
with weaker or deteriorated affinities. Moreover, the hard mutagenesis of more than seven amino acid residues, which should produce >20^7 (≈ 1.28 × 10^8) different amino acid sequences, generates libraries too large to deal with using standard experimental conditions. It is no wonder that these difficulties made us gradually avoid the V_H-CDR3-focused strategies for the purpose of affinity-maturation.

However, the present findings suggest a much simpler strategy for mining a possible ace mutation(s) in V_H-CDR3. The parent scFv (scFv#WT) has a V_H-CDR3 composed of 15 amino acids (Fig. 2A), which also correspond to the definition by Chothia et al. We here assumed that 13 of these residues, avoiding the less-variable D and Y residues at the V_H101 and V_H102 positions, are important for generating affinity against E2. Searching a very small scFv library, i.e., a sum of 13 groups of sequences each containing 20 different sequences generated by the hard randomization of one of the 13 positions (therefore, theoretically composed of only 13 × 20 = 260 amino acid sequences), should have enabled discovery of the mutant with the ace VH-L100gQ substitution. Even in the cases where two (or more) ace-equivalent mutations were present in the CDR3 and cooperated together, each of them could be separately discovered. Parallel examinations following hard mutagenesis of two or three serial residues (each generates 12 × 20 = 120, 12 × 20 × 20 = 4,800 or 12 × 20 × 20 × 20 = 88,000 sequences, respectively) might help in finding "ace-mutation motifs". In our laboratory, development of a novel and efficient strategy for discovering affinity-matured scFv mutants, named "colony-array profiling", is now progressing. We believe that this approach should be particularly suitable for examining such small libraries and will be of great help, the results of which will be reported in the near future. After finding such an ace mutation, some cooperating mutations (e.g., corresponding to the V and M residues in this study) should be searched for. Error-prone PCR might be still suitable for this purpose.

One of the motives for undertaking this study was an article published in 1990, where the author sought amino acid substitutions that caused a >200-fold difference in the K_a (determined by the fluorescence quenching method) between two hybridoma-derived antibodies against p-azophenylarsonate. Although 19 amino acid differences were originally observed between these antibodies, the author finally showed that only three amino acid substitutions (two in V_H-CDR2, one in V_H-CDR3) were needed to produce this increased affinity by examining artificial antibodies produced with the aid of synthetic oligonucleotides (Supplementary Fig. S1B). A more striking example was shown by the affinities observed for anti-digoxin antibodies with the integrated CDR3 and VL-CDR1 tend to play important roles in forming the antigen-binding cavity. These results are reasonable because, for antibodies against hapten, V_H-CDR2 and V_L-CDR1 tend to play important roles in forming the antigen-binding cavity, and this speculation was supported by the successful affinity maturation of anti-tacrolimus scFvs due to randomizing VH1-CDR2 and VL-CDR1 with NNS codons. However, even in such cases, the approach for seeking ace substitutions mentioned above can easily applied by extending it to more than one CDR in a parallel manner.

We are greatly interested in identifying the greatest possible increase in the affinity with the fewest substitutions, because this suggests the potential of in vitro affinity maturation of antibodies. In this study, we achieved 170-fold higher affinity by introducing four substitutions (with scFv#4mut), 107-fold higher affinity by introducing two substitutions (with scFv#R4-2), and 62-fold higher affinity by introducing three substitutions (with scFv#R3-4), and these improved scFv mutants showed >10^8 M^−1 order K_a values. Considering the difficulty in performing extensive improvement from already-matured antibodies (e.g., whose K_a values exceed 10^8 M^−1), our present results might be worthy of attention.

Materials and Methods

Buffers. The following buffers were used in this study. PB: 50 mM sodium phosphate buffer (pH 7.3); PBS: PB containing 9.0 g/L NaCl; G-PBS: PBS containing 1.0 g/L gelatin; T-PBS: PBS containing 0.050% (v/v) Tween 20; and PVG-PBS: G-PBS containing 1.0 g/L polyvinyl alcohol with an average polymerization degree of 500.

scFvs. Anti-E2 scFv#WT, scFv#M1st, scFv#M2nd, and scFv#M3rd(amb) (originally named as scFv#E4-412-14, scFv#m1-e712-14, scFv#m2-c413-14, and scFv#m3-a1816, respectively) were prepared as soluble proteins as we described previously. Other scFvs having reverse mutation(s) (i.e., revertants) were produced by expressing the corresponding scFvs genes in E. coli XL1-Blue cells as described previously. These scFvs were constructed by PCR using synthetic oligo-DNAs predesigned to introduce the targeted mutation(s) as shown below, whose nucleotide sequences were confirmed by the standard method. The scFv proteins were obtained as periplasmic extracts from mass-cultured transformants, and used for K_a determinations and ELISAs. For SDS-PAGE analysis, the scFvs were affinity-purified with anti-FLAG M2 antibody-immobilized agarose gel (Sigma–Aldrich).

Preparation of gene fragments encoding the scFv revertants. Among the 21 scFv genes synthesized in this study, typical instances were selected and their preparations are shown below. PCR experiments were performed using an adequate scFv gene (subcloned into the pEXmide 5 vector) with 0.5–50 ng as the template in a buffer (100 μL) containing Ex Taq (TaKaRa–Bio) (0.5 or 2.5 U) or KOD FX DNA polymerase (TOYOBO) (2.5 U), 20 nmol of each dNTP, and a combination of reverse/forward primers (30–100 pmol each), unless otherwise specified. Usually, the following cycling condition was used: 94°C (2 min), then 35 cycles of 98°C for 10 sec, 55°C for 30 sec, and 72°C for 1 min, followed by a hold step at 72°C for 10 min. The nucleotide sequences of the primers are shown in Supplementary Table S1. Every scFv gene fragment synthesized was digested with Nco I and Sal I, ligated with the similarly digested pEXmide 5 vector, and introduced in the E. coli cells by electroporation.
a) **scFv#R1-1.** Three sets of PCRs were performed using the scFv#M3rd(amb) gene\(^{14}\) as the template and one of the following three combinations of primers: (i) R1 and F1, (ii) R2 and F2, or (iii) R3 and E\(_2\)-VL-For-2\(^{12}\) (Supplementary Figure S4A). Using the three kinds of PCR products (i–iii), two overlap-extension PCR steps were performed. First, the products i and ii (each 200 ng) were mixed and subjected to 10 cycles of PCR in a 25-\(\mu\)L buffer solution containing ExTaq polymerase. A portion of the reaction solution (10 \(\mu\)L) was mixed with R1 and F2 primers and re-amplified similarly, but for 15 cycles in a 100-\(\mu\)L buffer solution containing ExTaq polymerase. Second, the resulting product was gel-purified, and a portion (200 ng) was mixed with product iii (200 ng) and subjected to a similar serial two-step amplification to generate the desired gene fragment.

b) **scFv#R2-1.** PCR was performed using the scFv#M2nd gene\(^{13}\) as template with R1 and F3 primers. The product obtained was used as reverse mega-primer (MP) in the next PCR, in combination with the E\(_2\)-VL-For-2 primer\(^{12}\), to generate the desired gene fragment.

c) **scFv#4mut.** PCRs were performed using the scFv#WT gene\(^{12,13}\) as template (Supplementary Figure S4B). The \(V_I\)-portion gene was prepared by amplification with E\(_2\)-VL-Rev and F4 primers. The \(V_I\)-portion gene was prepared as follows. First, PCR was performed using R5 and F5 primers. The product obtained was gel-purified and used as reverse MP1 in the next PCR, in combination with the E\(_2\)-VL-For-2 primer. The product was then used similarly as forward MP2 with the E\(_2\)-VL-Rev primer\(^{12}\) to generate the \(V_I\)-portion gene fragment with \(5\')-end sequence that was complementary to the \(3\')-end sequence of the \(V_I\)-portion gene fragment. These \(V_I\)- and \(V_I\)-portion gene fragments (each 200 ng) were mixed and submitted to overlap-extension PCR as described above (see entry a) to generate the desired gene fragment.

d) **scFv#R3-1.** PCRs were performed using the scFv#WT gene as template. The \(V_I\)-portion gene was prepared as follows. First, PCR was performed using R4 and MP1 primers. The product obtained was gel-purified and used as MP3, which was submitted to overlap-extension PCR with MP2 as described above (see entry a) to generate the gene fragment containing whole the \(V_I\) with a portion of the \(3\')-side of \(V_I\) (extending over \(V_I\)-CDR3). On the other hand, a gene fragment containing the \(V_I\)-portion was amplified using E\(_2\)-VL-Rev and E\(_2\)-VL-For primers. These two gene fragments were combined by the overlap-extension PCR to generate the desired gene fragment.

e) **scFv#R4-3.** PCR was performed using the scFv#WT gene as template with R2 and F5 primers. The product obtained was digested with BamH I and Sal I to generate a gene fragment covering the \(V_I\) with the mutation that substitutes the \(V_{I77}\) residue, whereas the plasmid having scFv#4mut gene was digested with BamH I and Sal I and gel-purified to remove the corresponding gene fragment. The \(V_{I77}\)-mutated gene fragment was ligated into the digested plasmid to construct the desirable gene as the form already incorporated in the plasmid.

f) **scFv#R5-1.** The desired gene fragment was generated by PCR using the scFv#WT gene as template with an MP (the \(V_I\)-portion gene fragment prepared in entry c) and the E\(_2\)-VL-For-2 primer.

**Determination of the scFv \(K_x\) values.**

a) **Scatchard analysis\(^{28}\).** Mixtures of \([1, 2, 6, 7\text{-}^{3}H\text{-}]\text{-E}_{2}\ (3.53 \text{ TBq/mmol}; \text{PerkinElmer}) \ (~250 \text{ Bq})\), varying amounts of standard \(E_{2}\), and a constant amount of each scFv (adjusted to bind to ~50% of the tritium-labeled \(E_{2}\)) were incubated in G-PBS (500 \(\mu\)L) at 4°C for 240 min. The bound (B) and free (F) fractions were separated using a dextran-coated charcoal method, and the radioactivity of the B fraction was measured.

b) **SPR analysis.** The kinetic parameters of selected scFvs, which were affinity-purified with anti-FLAG-M2 agarose (Sigma–Aldrich)\(^{11}\), to the E\(_2\)-bovine serum albumin (BSA) conjugate (prepared according to the previous method\(^{12}\)) were determined with Biacore T200 SPR system (GE Healthcare). Kinetic measurements were carried out at 25°C in G-PBS with a constant flow rate of 30 \(\mu\)L/min. In the measurements, five different concentrations (0.16–10 \(\mu\)g/mL) of the purified scFvs were used. The kinetic evaluation of data was performed using Biacore T200 evaluation software (GE Healthcare).

**ELISA.** The 96-well microplates (#3590; Corning) coated with the E\(_2\)-BSA conjugate (see above) were incubated at 4°C for 240 min with a mixture of \(E_{2}\) standard (50.0 \(\mu\)L/well) and soluble scFv protein (100 \(\mu\)L/well), both diluted in PVG-PBS. The microplates were washed 3 times with T-PBS and probed with an anti-FLAG M2 antibody labeled with peroxidase (POD) (Sigma–Aldrich) diluted in G-PBS (0.20 \(\mu\)g/mL; 100 \(\mu\)L/well)\(^{12-17}\). After incubation at 37°C for 30 min, the microplates were washed similarly and the captured POD activity was determined colorimetrically (490 nm), as described previously\(^{12-17}\). To construct the ELISA dose–response curves, Image j software\(^{48}\) (NIH) was used for curve fitting and determining the reaction parameters. The midpoint (i.e., IC\(_{50}\)) values were derived from a four-parametric logistic equation [log(analyte dose) vs. B/B\(_{0}\)(%)] as the E\(_{50}\) values. The unit “X g/assay” was used in the abscissa, which refers to the total mass (X g) of analyte that was added to each assay chamber (microwell) for the competitive antigen–antibody reactions.

**Data availability**

The data sets generated during the current study are available from the corresponding author upon request.

Received: 30 October 2019; Accepted: 27 February 2020; Published online: 16 March 2020
References

1. Gilmore, E. et al. Monoclonal antibodies: technologies for early discovery and engineering. *Science* **362**, 762–768 (2018).

2. Chothia, C. et al. Conformations of immunoglobulin hypervariable regions. *Nature* **503**, 117–122 (2013).

3. Griffiths, A. D. et al. Antibody engineering toward high-sensitivity high-throughput immunosensing of small molecules. *Anal. Chem.* **89**, 2905–2912 (2017).

4. Smith, G. P. & Petrenko, V. A. Phage display. *Annu. Rev. Biochem.* **60**, 95–120 (1991).

5. Kennedy, P. J., Oliveira, C., Granja, P. L. & Sarmento, B. Monoclonal antibodies: technologies for early discovery and engineering. *Crit. Rev. Biotechnol.* **38**, 394–408 (2018).

6. Chiu, M. L. & Gilliland, G. L. Engineering antibody therapeutics. *Proc. Natl. Acad. Sci. USA* **112**, 7729–7736 (2015).

7. Smith, G. P. & Petrenko, V. A. Phage display. *Chem. Rev.* **97**, 391–410 (1997).

8. Kabat, E. A., Wu, T. T., Perry, H. M., Gottesman, K. S. & Foeller, C. Sequences of Proteins of Immunological Interest. *Proteins: Struct. Funct. Genet.* **25**, 219–227 (1996).

9. Chothia, C. et al. Standard conformations for the canonical structures of immunoglobulins. *Nature* **273**, 89–95 (1978).

10. Derda, R. et al. Crystal structure of an in vitro affinity-matured antibody mutant targeting a small biomarker. *Nat. Chem.* **8**, 1011–1019 (2016).

11. Schildbach, J. F. et al. Improved affinity tag based on the FLAG® peptide for the detection and purification of proteins. *Anal. Chem.* **89**, 1027–1032 (2017).

12. Oliva, B., Bates, P. A., Querol, E., Avilés, F. X. & Sternberg, M. J. E. Automated classification of antibody complementarity determining region 3 of the heavy chain (H3) loops into canonical forms and its application to protein structure prediction. *J. Mol. Biol.* **279**, 1193–1210 (1998).

13. Al-Lazikani, B., Lesk, A. M. & Chothia, C. Standard conformations for the canonical structures of immunoglobulins. *J. Mol. Biol.* **273**, 927–948 (1997).

14. Schildbach, J. F. et al. Altered hapten recognition by two anti-digoxin hybridoma variants due to variable region point mutations. *J. Biol. Chem.* **266**, 4640–4647 (1991).

15. Vargas-Madrazo, E., Lara-Ochoa, F. & Almagro, J. C. Canonical structure repertoire of the antigen-binding site of immunoglobulins suggests strong geometrical restrictions associated to the mechanism of immune recognition. *J. Mol. Biol.* **254**, 497–504 (1995).

16. Smith, G. P., Johnson, G. & Kabat, E. A. Length distribution of CDRH3 in antibodies. *Anal. Chem.* **59**, 1027–1032 (1987).

17. Morita, I. Diversity of phage-displayed libraries of peptides during panning and amplification. *Molecules* **16**, 337–354 (2011).

18. Smith, G. P. & Petrenko, V. A. Phage display. *Crit. Rev. Biotechnol.* **38**, 394–408 (2018).

19. Boder, E. T., Midelfort, K. S. & Wittrup, K. D. Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Biotechnol. J.* **39**, 1–5 (2010).

20. Smith, G. P. & Petrenko, V. A. Phage display. *Chem. Rev.* **97**, 391–410 (1997).

21. Chothia, C. et al. Standard conformations for the canonical structures of immunoglobulins. *Nature* **273**, 89–95 (1978).

22. Sormanni, P., Aprile, F. A. & Vendruscolo, M. Third generation antibody discovery methods: rational design. *Curr. Opin. Struct. Biol.* **21**, 11–15 (2011).

23. Collis, A. V. J., Brouwer, A. P. & Martin, A. C. R. Analysis of the antigen combining site: correlations between length and sequence of the hypervariable loops and the nature of the antigen. *J. Mol. Biol.* **325**, 337–354 (2003).

24. Chothia, C. & Lesk, A. M. The attractions of proteins for small molecules and ions. *Ann. N. Y. Acad. Sci.* **51**, 660–672 (1949).

25. Kabat, E. A. & Foeller, C. Sequences of Proteins of Immunological Interest. *Proteins: Struct. Funct. Genet.* **25**, 219–227 (1996).

26. Smith, G. P., Johnson, G. & Kabat, E. A. Length distribution of CDRH3 in antibodies. *Anal. Chem.* **59**, 1027–1032 (1987).

27. Collis, A. V. J., Brouwer, A. P. & Martin, A. C. R. Analysis of the antigen combining site: correlations between length and sequence of the hypervariable loops and the nature of the antigen. *J. Mol. Biol.* **325**, 337–354 (2003).

28. Sechatard, G. The attractions of proteins for small molecules and ions. *Ann. N. Y. Acad. Sci.* **51**, 660–672 (1949).

29. Hosoda, H., Kobayashi, N. & Nambara, T. Effect of bridge heterologous combination on sensitivity in enzyme immunoassay for 11-deoxycortisol. *Chem. Pharm. Bull.* **49**, 2061–2065 (2001).

30. Hosoda, H., Kobayashi, N., Ishii, N. & Nambara, T. Bridging phenomena in steroid immunoassays. The effect of bridge length on sensitivity in enzyme immunoassay. *Chem. Pharm. Bull.* **31**, 953–958 (1983).

31. delta- and beta-phenylalanine substitutions can increase the affinity of an in vitro affinity-matured antibody mutant targeting a small biomarker. *Chem. Pharm. Bull.* **49**, 2061–2065 (2001).

32. Barbas, C. F. III, Bain, J. D., Hoekstra, D. M. & Lerner, R. A. Semisynthetic combinatorial antibody libraries: a chemical solution to the diversity problem. *Proc. Natl. Acad. Sci. USA* **89**, 4457–4461 (1992).

33. Griffiths, A. D. et al. Isolation of high affinity human antibodies directly from large synthetic repertoires. *EMBO J.* **13**, 3245–3260 (1994).

34. Valjakka, J. et al. Crystal structure of an in vitro affinity-matured human antibody Fab in complex with testosterone. *J. Biol. Chem.* **277**, 44021–44027 (2002).

35. Fellouse, F. A. et al. Molecular recognition by a binary code. *J. Mol. Biol.* **284**, 1113–1162 (2005).

36. Chothia, C. et al. Conformations of immunoglobulin hypervariable regions. *Nature* **347**, 877–883 (1987).

37. Al-Lazikani, B., Lesk, A. M. & Chothia, C. Standard conformations for the canonical structures of immunoglobulins. *J. Mol. Biol.* **273**, 927–948 (1997).

38. Sharon, J. Structural correlates of high antibody affinity: three engineered amino acid substitutions can increase the affinity of an in vitro affinity-matured antibody mutant targeting a small biomarker. *Chem. Pharm. Bull.* **49**, 2061–2065 (2001).

39. Chothia, C. et al. Conformations of immunoglobulin hypervariable regions. *Nature* **347**, 877–883 (1987).

40. Al-Lazikani, B., Lesk, A. M. & Chothia, C. Standard conformations for the canonical structures of immunoglobulins. *J. Mol. Biol.* **273**, 927–948 (1997).

41. Schildbach, J. F. et al. Altered hapten recognition by two anti-digoxin hybridoma variants due to variable region point mutations. *J. Biol. Chem.* **266**, 4640–4647 (1991).

42. Vargas-Madrazo, E., Lara-Ochoa, F. & Almagro, J. C. Canonical structure repertoire of the antigen-binding site of immunoglobulins suggests strong geometrical restrictions associated to the mechanism of immune recognition. *J. Mol. Biol.* **254**, 497–504 (1995).
46. Siegel, R. W., Baugher, W., Rahn, T., Drengler, S. & Tyner, J. Affinity maturation of tacrolimus antibody for improved immunoassay performance. *Clin. Chem.* **54**, 1008–1017 (2008).
47. Jirholt, P., Ohlin, M., Borrebaeck, C. A. K. & Söderlind, E. Exploiting sequence space: shuffling *in vivo* formed complementarity determining regions into a master framework. *Gene* **215**, 471–476 (1998).
48. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to Image: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
49. Oyama, H., Yamaguchi, S., Nakata, S., Niwa, T. & Kobayashi, N. Correction to "Breeding" diagnostic antibodies for higher assay performance; a 230-fold affinity-matured antibody mutant targeting a small biomarker. *Anal. Chem.* **91**, 15325 (2019).
50. Guex, N., Diemand, A. & Peitsch, M. C. Protein modelling for all. *Trends Biochem. Sci.* **24**, 364–367 (1999).
51. Grosdidier, A., Zoete, V. & Michielin, O. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res.* **39**, W270–W277 (2011).
52. Delano, W. *The PyMOL Molecular Graphics System*; DeLano Scientific LLC: Palo Alto, CA, USA, 2008. Available at, http://www.pymol.org.

**Acknowledgements**

We thank Dr. Eskil Söderlind (Avena Partners AB, Sweden) and Professor Carl A.K. Borrebaeck (Lund University, Sweden) for providing the pEXmide 5 vector.

**Author contributions**

N.K. conceived the project; N.K. and H.O. designed the research; H.O., Y. K., I.M., C.Y., Y.H., M.T., T.T., Y.E., Y.T., H.H. and A.T. performed the research and analyzed the data; N.K. and H.O. wrote the paper.

**Competing interests**

The authors declare no competing interests.

**Additional information**

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41598-020-61529-7.

**Correspondence** and requests for materials should be addressed to N.K.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020