Protein evolution by hypermutation and selection in the B cell line DT40

Hiroshi Arakawa¹, Hiroaki Kudo¹, Vera Batrak¹, Randolph B. Caldwell¹, Michael A. Rieger², Joachim W. Ellwart³ and Jean-Marie Buerstedde¹,*

¹Institute for Molecular Radiobiology, ²Institute of Stem Cell Research, GSF-National Research Center for Environment and Health, Ingolstaedter Landstrasse 1, D-85764 Neuherberg-Munich and ³Institute of Molecular Immunology, GSF National Research Center for Environment and Health, Marchioninistrasse 25, 81377 Munich, Germany

Received May 24, 2007; Revised July 3, 2007; Accepted July 29, 2007

ABSTRACT
Genome-wide mutations and selection within a population are the basis of natural evolution. A similar process occurs during antibody affinity maturation when immunoglobulin genes are hypermutated and only those B cells which express antibodies of improved antigen-binding specificity are expanded. Protein evolution might be simulated in cell culture, if transgene-specific hypermutation can be combined with the selection of cells carrying beneficial mutations. Here, we describe the optimization of a GFP transgene in the B cell line DT40 by hypermutation and iterative fluorescence activated cell sorting. Artificial evolution in DT40 offers unique advantages and may be easily adapted to other transgenes, if the selection for desirable mutations is feasible.

INTRODUCTION
Natural evolution, based on the selection of beneficial mutations within a population of genetic variants, has created the amazing diversity of life on our planet. While natural selection works on whole genomes, the evolution of individual proteins can be tracked by the analysis of intra-species polymorphisms and inter-species divergence. A fascinating example for the evolution of a single protein by mutation and selection is the affinity maturation of antibodies in B cells (1). Developing B cells activate and diversify their immunoglobulin (Ig) genes by recombination, but only cells encoding antibodies with antigen-binding specificity are stimulated to expand after antigen challenge. These B cells hypermutate their Ig genes and cycles of hypermutation and selection continue until antibodies of sufficiently high antigen-binding affinity emerge.

Recombinant DNA technologies are able to recapitulate the evolution of proteins by mutagenesis and selection in vitro. Most approaches combine the expression of diverse gene libraries in phage, bacteria or translation systems with selection for the encoded proteins (2,3). If the structure-function relationship of a protein is sufficiently understood, site-directed mutagenesis of critical amino acids can be used instead or in addition to random mutagenesis (4). The potential of in vitro protein evolution is illustrated by the example of the green fluorescent protein (GFP) (5) which could be changed into yellow, cyan or blue variants by random and site-directed mutagenesis and selection in vitro (6–8).

Affinity maturation of antibodies can be simulated ex vivo in cultures of hypermutating B cell lines by the enrichment of cells expressing antigen-specific antibodies (9,10). B cell lines could also be used for the optimization of non-immunoglobulin proteins, if the hypermutating activity were directed toward transfected transgenes and cells carrying beneficial mutations were selected. Advantages of this approach are the possibilities to generate an enormous amount of genetic diversity and to select for improved protein variants within a living cell culture. It was recently reported that a gene encoding a red fluorescent protein (FP) was transferred into the hypermutating human B cell line RAMOS, and the selection of cells emitting red-shifted fluorescence yielded the most far red-shifted FP protein known to date (11).

Despite this impressive result, hypermutation is difficult to control in RAMOS, because transgenes usually integrate at random chromosomal sites outside the hypermutating Ig loci. In contrast, transgenes can be easily inserted into the Ig loci of the chicken B cell line DT40 (12). DT40 diversifies its rearranged Ig light chain gene by pseudo V (ψV) gene-templated gene conversion.
rearranged or the unarranged Ig locus of DT40, integration into the rearranged Ig light chain locus was verified by PCR amplification of the V/J intervening sequence of the unarranged locus using primer pairs P4/P5.

**Flow cytometry**

To quantify the appearance of GFP negative cells within proliferating cultures, the AIDR1IgL-eGFP, AID\(^{-/-}\)-IgL-eGFP and ψVAIDR1IgL-eGFP clones were sub-cloned by limited dilution, and 24 subclones of each were analyzed 2 weeks after subcloning by FACS. The preparative FACS sorts to select cells of increased fluorescence were performed using the MoFlo high-speed cell sorter (Cytomation).

**PCR, cloning and sequencing**

The screenings for targeted integration and the amplification of the eGFP transgenes were performed with the expand long template PCR system (Roche) under the following conditions: 2 min of initial incubation at 93°C; 35 cycles consisting of 10 s at 93°C, 30 s at 65°C and 5 min (plus an added 20 s per cycle) at 68°C and a final elongation step of 7 min at 68°C. The primer pair for the amplification of the eGFP gene from sorted cells was P6/P7. The PCR products were digested with HindIII and XbaI, cloned into the pUC119 plasmid vector and sequenced using primers P8 and P9. The mutated eGFP genes were then amplified using Pfu Ultra hotstart polymerase (Stratagene) and primer pair P10/P11, digested by AvrII and cloned into the NheI site of pHypermut2. The orientation and the sequence of the mutant eGFPs in pHypermut2 was verified by sequencing using primers P12 and P13.

**Site-directed mutagenesis**

To combine the codon changes found in different eGFP variants, we designed primers (P14–P26) which included the intended mutations in the center of the primer sequence. In the first step, parts of the eGFP gene were amplified by PCR using at least one mutation-containing primer. In the second step, PCR fragments containing mutations were mixed with other PCR fragments to cover the full eGFP coding sequence. This mixture served as a template for chimeric PCR using the primer pair P10/P11. The full length mutant eGFP PCR fragments weredigested with AvrII and cloned into the NheI site of pHypermut2. The orientation and the sequence of the mutant eGFPs in pHypermut2 were confirmed by sequencing.

**Primers**

- **P1** GGGACTAGTAAAAATGATGCATAACCTTTTG CACA
- **P2** CGATTGAAGAATCTTCCACTCAAATATA CCC
- **P3** CCCACGACTCTAGAGGATCATAATCG GCC
- **P4** TACAAAACCTCCTGCCAGTGCAAGGAGCA GCTGATGGTTTTA

**MATERIALS AND METHODS**

**Cell lines**

Cells were cultured in chicken medium (RPMI-1640 or DMEM/F-12 with 10% fetal bovine serum, 1% chicken serum, 2 mM L-glutamine, 0.1 mM β-mercaptoethanol and penicillin/streptomycin) at 41°C with 5% CO2. The AID expressing DT40 clones, AID R1CL1 and AID R1CL2, used in the study for hypermutation were derived from the AID knockout cell clone DT40 Cre1AID\(^{-/-}\) (16) by stable transfection of cell clone DT40Cre1AID\(^{-/-}\) (16) by stable transfection of a floxed AID----

**Flow cytometry**

To quantify the appearance of GFP negative cells within proliferating cultures, the AIDR1IgL-eGFP, AID\(^{-/-}\)-IgL-eGFP and ψVAIDR1IgL-eGFP clones were sub-cloned by limited dilution, and 24 subclones of each were analyzed 2 weeks after subcloning by FACS. The preparative FACS sorts to select cells of increased fluorescence were performed using the MoFlo high-speed cell sorter (Cytomation).

**PCR, cloning and sequencing**

The screenings for targeted integration and the amplification of the eGFP transgenes were performed with the expand long template PCR system (Roche) under the following conditions: 2 min of initial incubation at 93°C; 35 cycles consisting of 10 s at 93°C, 30 s at 65°C and 5 min (plus an added 20 s per cycle) at 68°C and a final elongation step of 7 min at 68°C. The primer pair for the amplification of the eGFP gene from sorted cells was P6/P7. The PCR products were digested with HindIII and XbaI, cloned into the pUC119 plasmid vector and sequenced using primers P8 and P9. The mutated eGFP genes were then amplified using Pfu Ultra hotstart polymerase (Stratagene) and primer pair P10/P11, digested by AvrII and cloned into the NheI site of pHypermut2. The orientation and the sequence of the mutant eGFPs in pHypermut2 was verified by sequencing using primers P12 and P13.

**Site-directed mutagenesis**

To combine the codon changes found in different eGFP variants, we designed primers (P14–P26) which included the intended mutations in the center of the primer sequence. In the first step, parts of the eGFP gene were amplified by PCR using at least one mutation-containing primer. In the second step, PCR fragments containing mutations were mixed with other PCR fragments to cover the full eGFP coding sequence. This mixture served as a template for chimeric PCR using the primer pair P10/P11. The full length mutant eGFP PCR fragments were digested with AvrII and cloned into the NheI site of pHypermut2. The orientation and the sequence of the mutant eGFPs in pHypermut2 were confirmed by sequencing.

**Primers**

- **P1** GGGACTAGTAAAAATGATGCATAACCTTTTG CACA
- **P2** CGATTGAAGAATCTTCCACTCAAATATA CCC
- **P3** CCCACGACTCTAGAGGATCATAATCG GCC
- **P4** TACAAAACCTCCTGCCAGTGCAAGGAGCA GCTGATGGTTTTA

**MATERIALS AND METHODS**

**Cell lines**

Cells were cultured in chicken medium (RPMI-1640 or DMEM/F-12 with 10% fetal bovine serum, 1% chicken serum, 2 mM L-glutamine, 0.1 mM β-mercaptoethanol and penicillin/streptomycin) at 41°C with 5% CO2. The AID expressing DT40 clones, AID R1CL1 and AID R1CL2, used in the study for hypermutation were derived from the AID knockout cell clone DT40 Cre1AID\(^{-/-}\) (16) by stable transfection of cell clone DT40Cre1AID\(^{-/-}\) (16) by stable transfection of a floxed AID----

**Flow cytometry**

To quantify the appearance of GFP negative cells within proliferating cultures, the AIDR1IgL-eGFP, AID\(^{-/-}\)-IgL-eGFP and ψVAIDR1IgL-eGFP clones were sub-cloned by limited dilution, and 24 subclones of each were analyzed 2 weeks after subcloning by FACS. The preparative FACS sorts to select cells of increased fluorescence were performed using the MoFlo high-speed cell sorter (Cytomation).

**PCR, cloning and sequencing**

The screenings for targeted integration and the amplification of the eGFP transgenes were performed with the expand long template PCR system (Roche) under the following conditions: 2 min of initial incubation at 93°C; 35 cycles consisting of 10 s at 93°C, 30 s at 65°C and 5 min (plus an added 20 s per cycle) at 68°C and a final elongation step of 7 min at 68°C. The primer pair for the amplification of the eGFP gene from sorted cells was P6/P7. The PCR products were digested with HindIII and XbaI, cloned into the pUC119 plasmid vector and sequenced using primers P8 and P9. The mutated eGFP genes were then amplified using Pfu Ultra hotstart polymerase (Stratagene) and primer pair P10/P11, digested by AvrII and cloned into the NheI site of pHypermut2. The orientation and the sequence of the mutant eGFPs in pHypermut2 was verified by sequencing using primers P12 and P13.

**Site-directed mutagenesis**

To combine the codon changes found in different eGFP variants, we designed primers (P14–P26) which included the intended mutations in the center of the primer sequence. In the first step, parts of the eGFP gene were amplified by PCR using at least one mutation-containing primer. In the second step, PCR fragments containing mutations were mixed with other PCR fragments to cover the full eGFP coding sequence. This mixture served as a template for chimeric PCR using the primer pair P10/P11. The full length mutant eGFP PCR fragments were digested with AvrII and cloned into the NheI site of pHypermut2. The orientation and the sequence of the mutant eGFPs in pHypermut2 were confirmed by sequencing.
transgenes, the bsr polyA signal (Figure 1A). The clones contained two floxed integrated the construct into the rearranged cDNA expression cassette (16). Clones which had

Ig light chain the eGFP chain locus were identified by PCR. Transcription of gene of the AID R1 progenitor clone. To remove the Ig light chain locus and the eGFP gene into the rearranged eGFP pHypermut1 vector and transfected into the DT40 variant, light chain locus Ig

Targeted integration of the AIDR1CL1, which is deleted for both alleles of the eGFP marker alone or together with the AID

subcloning by limited dilution (16). In this way, AID

subclones. The AIDR1IgLeGFP1 clones were expanded and cycles of preparative cell sorting and subcloning. Subclones of the AIDR1IgLeGFP1 clone contained cell populations showing lower (3.8%) and higher (0.5%) fluorescence than the dominant GFP positive cell population, whereas subclones of the AID−/−IgLeGFP1 clone consisted only of a homogeneous GFP positive population (Figure 1B). This suggests that the eGFP gene inserted into the rearranged Ig light chain locus is diversified in AIDR1IgLeGFP1 cells giving rise to variants of increased and decreased fluorescent intensity.

RESULTS

Targeted integration of the eGFP gene into the rearranged Ig light chain locus

The eGFP coding sequence was inserted into the pHypermut1 vector and transfected into the DT40 variant, AIDR1CL1, which is deleted for both alleles of the endogenous AID gene and expresses AID as a floxed cDNA expression cassette (16). Clones which had integrated the construct into the rearranged Ig light chain locus were identified by PCR. Transcription of the eGFP gene in these clones is supposed to be driven by the Ig light chain promoter and terminated by the SV40 polyA signal (Figure 1A). The clones contained two floxed transgenes, the bsr marker co-inserted in the rearranged Ig light chain locus and the AID-IRES-gpt gene of the AIDR1 progenitor clone. To remove the bsr marker alone or together with the AID expression cassette, Cre recombinase was induced by tamoxifen before subcloning by limited dilution (16). In this way, AID positive (AIDR1IgLeGFP1) and negative (AID−/−IgLeGFP1) clones could be isolated which had deleted the bsr marker gene from the light chain locus.

Evidence for hypermutation of the eGFP transgene

The AIDR1IgLeGFP1 and AID−/−IgLeGFP1 clones were subcloned by limited dilution and 24 subclones of each were analyzed by FACS for GFP brightness 14 days after subcloning. Subclones of the AIDR1IgLeGFP1 clone contained cell populations showing lower (3.8%) and higher (0.5%) fluorescence than the dominant GFP positive cell population, whereas subclones of the AID−/−IgLeGFP1 clone consisted only of a homogeneous GFP positive population (Figure 1B). This suggests that the eGFP gene inserted into the rearranged Ig light chain locus is diversified in AIDR1IgLeGFP1 cells giving rise to variants of increased and decreased fluorescent intensity.

Mutation activity of artificial evolution system

To analyze the mutation spectrum of the eGFP transgenes, the coding sequence was amplified by PCR from AIDR1IgLeGFP1 and AID−/−IgLeGFP1 cells 6 weeks after subcloning, cloned into a plasmid vector and sequenced. A total of 13 nt changes were found within the 0.7 kb eGFP coding region of 39 sequences from the AIDR1IgLeGFP1 clone (Table 1). Assuming a DT40 doubling time of 10 h, the mutation rate of the eGFP coding region in the rearranged Ig light chain locus of AIDR1IgLeGFP1 cells is approximately $4.7 \times 10^{-6}$ mutations per base pair and division. This mutation frequency is approximately three times lower than the mutation rate of the rearranged VJ segment in the pseudogene deleted $\psi$V AIDR cell line (15) perhaps reflecting the lower incidence of hypermutation hotspots in the eGFP sequence. In both cases, the majority of mutations occurred at C/G base pairs (Table 1). Only 1 nt change which possibly represents a PCR artifact was found in 26 sequences of the AIDIgLeGFP1 clone confirming that the mutation activity of the eGFP transgene is dependent on AID.

Sorting and sequence analysis of cells displaying increased fluorescence

The FACS analysis suggested that a few cells within the AIDR1IgLeGFP1 culture had accumulated mutations which increased either the abundance or the fluorescent intensity of the eGFP protein. To enrich these cells, 10 million AIDR1IgLeGFP1 cells were sorted by preparative FACS using a gate which included only the 0.05% brightest cells (Figure 1C). The cells collected by the FACS sort were expanded and cycles of preparative cell sorting and expansion were repeated twice. Three consecutive sorts were performed independently for three AIDR1IgLeGFP1 subclones.

To trace the rise of mutations responsible for the increased fluorescence, the coding sequences of the eGFP genes were amplified by PCR from cells after the first and the third sort, cloned into a plasmid vector and sequenced. Unexpectedly, few mutations and no obvious mutation signatures were found in the eGFP coding sequence after the first sort (Table 1) despite the fact that the sorted cells displayed increased fluorescence (Figure 1C). We therefore extended the sequence analysis to the region downstream
### Figure 1. Strategy for artificial evolution of eGFP gene.

(A) A physical map of the chicken rearranged Ig light chain locus, the pHypermut1-eGFP targeting construct and the rearranged Ig light chain locus after targeted integration and marker excision is shown. The positions of primers used for the identification of targeted integration events are shown by arrows. (B) FACS profiles of AIDR1IgLeGFP1 and AID−/−IgLeGFP1 clones. The average percentages of events falling into the GFP<sup>high</sup> and GFP<sup>low</sup> gates based on the measurement of 24 subclones are shown. (C) Sorting strategy for cells of increased fluorescence activity.

### Table 1. Mutation profile

| Cell source   | Gene       | Mutations | Number of sequences | Mutations/sequence | Mutations at C | Mutations at G | Mutations at A | Mutations at T | Duplication | Deletion |
|---------------|------------|-----------|---------------------|-------------------|----------------|----------------|----------------|----------------|-------------|----------|
| AID−/−IgLeGFP1 6-week culture | eGFP (723 bp) | 1         | 26                  | 0.04              | 0             | 0              | 1              | 0              | 0          | 0        |
| AIDR1IgLeGFP1 6-week culture | eGFP (723 bp) | 13        | 39                  | 0.33              | 5             | 8              | 0              | 1              | 0          | 0        |
| AIDR1IgLeGFP1 Sort I | eGFP (723 bp)<sup>a</sup> | 15        | 95                  | 0.16              | 8             | 6              | 1              | 0              | 0          | 0        |
| AIDR1IgLeGFP1 Sort I | polyA-L (300 bp)<sup>a</sup> | 88        | 95                  | 0.93              | 1             | 87             | 0              | 0              | 0          | 0        |
| AIDR1IgLeGFP1 Sort III | eGFP (723 bp) | 221       | 85                  | 2.60              | 130           | 93             | 6              | 4              | 0          | 0        |

<sup>a</sup>Their sequences were derived from the different region of the same plasmids.
of the eGFP coding sequence by primer walking. This analysis revealed frequent mutations at the splice donor site of the *Ig light chain* leader intron (88 mutations/95 seq) (Figure 2A, Table 1). Although the role of these mutations downstream of the eGFP polyA signal remains speculative, it is possible that they enhance fluorescence by increasing eGFP mRNA stability.

In contrast, most of the sequences after the third sort showed mutations in the eGFP coding sequence (221 mutations/85 seq) (Figure 2B, Table 1). Comparison of the sequences of each sorted subclone allowed us to reconstruct evolutionary trees showing the stepwise accumulation of eGFP codon changes (Figure 2C). Many of the codon changes were not shared by different subclones and their functional significance remains uncertain. However, a number of codon changes occurred independently in different subclones suggesting that they are related to the increased fluorescence of the sorted cells. Mutated eGFP sequences which were found more than once were named GFP variants (GFPv1–9).

**Development of a new vector for transgene evolution in DT40**

Although we were able to use pHypermut1 for the enhancement of eGFP, this vector is not ideally suited for transgene evolution in DT40, as mutations outside the transgene coding region had apparently been selected for after the first sort. We therefore developed a new targeting vector named pHypermut2 in which the expression of the transgene coding region had apparently been selected for transgene evolution in DT40, as mutations outside the enhancer of eGFP, this vector is not ideally suited for the rearranged locus. The fluorescent intensity of the different transfectants was compared by FACS (Figure 4A). Two independent transfectants of each variant GFP gene were included in the experiment to account for variation among transfectants of the same gene. This analysis revealed that the Emerald transfectant was 1.2 times brighter than the eGFP transfectant, four of the nine GFP variant transfectants (AID/−/IgLGFPv2, AID/−/IgLGFPv5, AID/−/IgLGFPv6 and AID/−/IgLGFPv8) were brighter than the eGFP transfectant. Three variant GFP transfectants (AID/−/IgLGFPv5, AID/−/IgLGFPv6 and AID/−/IgLGFPv8) had even higher fluorescent intensity than the Emerald transfectant (Table 2). Of all variant GFP transfectants, AID/−/IgLGFPv6 had the highest fluorescent intensity (2.5-fold more than the AID/−/IgLGFPv4 and AID/−/IgLGFPv5) had the second highest intensity (2.0-fold brighter than the AID/−/IgLGFPv4) (Figure 4A, Table 2). Little fluorescent variation was observed among transfectants of the same transgene.

To compare the fluorescent intensity of eGFP and GFPv6 in human cells, the pHypermut2 constructs of GFPv6 and eGFP were transiently transduced into NIH3T3 cells transfected with the GFPv6 construct showed on average more fluorescence than cells transfected with the eGFP construct (data not shown).

**Increased fluorescence by the combination of mutations**

As certain combinations of codon changes were not found among the GFP variants, we wondered whether the accumulation of mutations in a single sequence may lead to further enhancement of fluorescent intensity (Figure 4B). At first, we combined Y145F and S202T of GFPv6 with AID/− to generate GFPv10. Because GFPv1 harboring the single Q80E mutation showed less fluorescent intensity than eGFP, we subtracted Q80E from GFPv10 to generate GFPv11. We also added L221I of GFPv8 to GFPv11, thereby generating GFPv12. Finally, we added Q184E and A206T of GFPv2 as well as L141V of GFPv9 to GFPv12 to generate GFPv13. These new variants were cloned into pHypermut2 and inserted into the rearranged *Ig light chain* locus of the clone AID/−/.

Transfectants of all new GFP variants showed higher fluorescent intensity than AID/−/IgLGFPv6 by FACS analysis (Figure 4C, Table 2). AID/−/IgLGFPv10 had 3.0-fold higher fluorescent intensity than AID/−/IgLGFPv1, but AID/−/IgLGFPv11 and AID/−/IgLGFPv12 were even brighter, displaying 3.2- and 3.3-fold more fluorescence respectively than AID/−/IgLGFPv1. AID/−/IgLGFPv13 did not surpass the fluorescent intensity of the other variant isolated GFP variants, as well as the eGFP and the Emerald (8) control sequences, were cloned into pHypermut2. These constructs were then transduced into AID/−/IgLGFPv1. More than 50% of all stable transfectants had integrated the constructs into the rearranged *light chain* locus and these clones were named according to the inserted transgene (AID/−/IgLGFPv1, AID/−/IgLEmerald, AID/−/IgLGFPv13, etc.).

The fluorescent intensity of the different transfectants of the transgene was compared by FACS (Figure 4A). Two independent transfectants of each variant GFP gene were included in the experiment to account for variation among transfectants of the same gene. This analysis revealed that the Emerald transfectant was 1.2 times brighter than the eGFP transfectant, four of the nine GFP variant transfectants (AID/−/IgLGFPv2, AID/−/IgLGFPv5, AID/−/IgLGFPv6 and AID/−/IgLGFPv8) were brighter than the eGFP transfectant. Three variant GFP transfectants (AID/−/IgLGFPv5, AID/−/IgLGFPv6 and AID/−/IgLGFPv8) had even higher fluorescent intensity than the Emerald transfectant (Table 2). Of all variant GFP transfectants, AID/−/IgLGFPv6 had the highest fluorescent intensity (2.5-fold more than the AID/−/IgLGFPv4 and AID/−/IgLGFPv5) had the second highest intensity (2.0-fold brighter than the AID/−/IgLGFPv4) (Figure 4A, Table 2). Little fluorescent variation was observed among transfectants of the same transgene.

To compare the fluorescent intensity of eGFP and GFPv6 in human cells, the pHypermut2 constructs of GFPv6 and eGFP were transiently transduced into the human embryonic kidney cell line HEK293T. By FACS analysis, cells transfected with the GFPv6 construct showed on average more fluorescence than cells transfected with the eGFP construct (data not shown).

**Confirmation of the variant GFP phenotypes**

The phenotype of mutations was best confirmed by retransfection into the genetically stable AID/−/ clone and the analysis of transfectants which have integrated the gene as a single copy into the rearranged locus. The pHypermut2 vector is well suited for this task and all
Figure 2. Mutations downstream and within the eGFP transgene. (A) Mutations identified at the exon/intron border of the Ig light chain leader sequence. The number of times a mutation was found is shown by superscript. (B) Mutations within the eGFP coding sequence. Mutations were mapped below the reference eGFP sequence together with the corresponding amino acid codons. The eGFP sequence which we used had one codon (Val) inserted next to the start codon to yield an optimal translation-termination sequence (Kozak motif). This Val was numbered as codon 1a according to Cramer (26) for the easier comparison to wild-type GFP and previously reported GFP mutants. (C) Pedigrees for the evolution of the eGFP transgenes in culture. The number of times each sequence was found within each subclone is indicated at the right side of the circles. The amino acid changes of each step are shown beside the arrows. Sequences identified in more than one subclone are named v1–v9.

**Figure 2.** Mutations downstream and within the eGFP transgene. (A) Mutations identified at the exon/intron border of the Ig light chain leader sequence. The number of times a mutation was found is shown by superscript. (B) Mutations within the eGFP coding sequence. Mutations were mapped below the reference eGFP sequence together with the corresponding amino acid codons. The eGFP sequence which we used had one codon (Val) inserted next to the start codon to yield an optimal translation-termination sequence (Kozak motif). This Val was numbered as codon 1a according to Cramer (26) for the easier comparison to wild-type GFP and previously reported GFP mutants. (C) Pedigrees for the evolution of the eGFP transgenes in culture. The number of times each sequence was found within each subclone is indicated at the right side of the circles. The amino acid changes of each step are shown beside the arrows. Sequences identified in more than one subclone are named v1–v9.
GFP transfectants. The variants GFPv11 and GFPv12 which offer the most potential for the labeling of vertebrate cells were named GFPnovo1 and GFPnovo2, respectively.

Spectral properties of the new GFP variants

The excitation and the emission maximum of eGFP are at 488 and 509 nm wavelength, respectively (7,8). To determine the spectral profiles of the new variant GFP proteins, the control and variant gene transfectants (AID\(-\)/IgL\(^{\text{GFP}}\), AID\(-\)/IgL\(^{\text{Emerald}}\), AID\(-\)/IgL\(^{\text{GFPv5}}\), AID\(-\)/IgL\(^{\text{GFPv6}}\), AID\(-\)/IgL\(^{\text{GFPnovo1}}\) and the AID\(-\)/IgL\(^{\text{GFPnovo2}}\)) were analyzed using a luminescence spectrometer (Figure 4D). This analysis showed that the fluorescent intensity, but not the excitation and emission spectra had been changed in the variant GFP transfectants. The lack of mutants showing altered excitation or emission properties is most likely due to our selection strategy which only enriched cells for increased fluorescent intensity.

We also examined cells of the eGFP, Emerald and variant GFP transfectants by fluorescence microscopy (Figure 4E). Consistent with the FACS and spectrometer analysis, cells of the GFPv5, GFPv6, GFPnovo1 and GFPnovo2 transfectants produced brighter images than cells of the eGFP and Emerald transfectants, but there was no noticeable change of color.

DISCUSSION

We describe here an artificial evolution system in the B cell line DT40 which offers advantages over alternative approaches and could be of generic value for the optimization of many types of proteins in cell culture. The main advantages are the high transgene-specific mutation rates, the easy confirmation of mutant...
Figure 4. Analysis of variant GFP transfectants. (A) FACS analysis of control and variant GFP transfectants. The average fluorescence of AID London /C0/ IgLeGFP is indicated by a green line for easier comparison. (B) Flow chart of site-directed mutagenesis of GFP variants. (C) Relative fluorescence of the transfectants normalized to the fluorescence of AID /C0/ IgLeGFP. (D) Excitation and emission spectra. (E) Image of single cells by fluorescence microscopy. The image of the same single cells is shown with fluorescence activation (upper row) and without fluorescence activation (lower row).
Table 2. Amino acid changes and brightness of GFP variants

| GFP variant       | Amino acid changes | Number of sequences | Subclone | Relative brightness\(^a\) |
|-------------------|--------------------|---------------------|----------|--------------------------|
| eGFP              | Change from wild-type GFP | –                   | –        | 1.0 ± 0.1                |
| eGFPv1            | Q80E               | 28                  | 1, 2, 3  | 0.7 ± 0.0                |
| eGFPv2            | Q80E, Q184E, A206T | 19                  | 1        | 1.2 ± 0.0                |
| eGFPv3            | Q80E, E124D, L141V, Y237stop | 11               | 3        | 0.3 ± 0.0                |
| eGFPv4            | Q80E, S202T        | 4                   | 1, 2     | 0.9 ± 0.1                |
| eGFPv5            | Q80E, V163A        | 4                   | 1        | 2.0 ± 0.0                |
| eGFPv6            | Q80E, Y145F, S202T | 3                   | 2        | 2.5 ± 0.0                |
| eGFPv7            | Q80E, S208T        | 2                   | 3        | 0.7 ± 0.1                |
| eGFPv8            | Q80E, L221V        | 2                   | 3        | 1.5 ± 0.3                |
| eGFPv9            | Q80E, L141V, S202T | 5                   | 3        | 1.0 ± 0.1                |
| eGFPv10           | Q80E, Y145F, V163A, S202T | –               | –        | 3.0 ± 0.1                |
| eGFPv11 (GFPnovo1)| Y145F, V163A, S202T | –                   | –        | 3.2 ± 0.0                |
| eGFPv12 (GFPnovo2)| Y145F, V163A, S202T, L221V | –               | –        | 3.3 ± 0.0                |
| eGFPv13           | L141V, Y145F, V163A, Q184E, S202T, A206T, L221V | –               | –        | 2.9 ± 0.1                |
| Emerald           | Change from wild-type GFP | S65T, S72A, N149K, M153T, I167T | –        | 1.2                      |

\(^a\)SDS based on the analysis of two clones.

phenotypes and the selection for beneficial mutations within the environment of live vertebrate cells. We have illustrated the utility of the system by generating new GFP variants which display more than 3-fold higher fluorescence activity than the best GFPs currently available for bio-imaging of vertebrate cells.

We have chosen eGFP as a first example for protein evolution in DT40, because the protein has been extensively optimized by random and site-directed mutagenesis and selection of cells displaying brighter fluorescence is straightforward. Only three rounds of FACS sorting during 2 months of culturing were sufficient to isolate mutants of increased fluorescent intensity. Some of the identified mutations are identical to mutations previously reported by others for GFP variants with an altered excitation and/or emission spectrum. For example, the V163A codon substitution is included in the GFP variants Cycle3 (7), T-Sapphire (20), Venus (21–23), ECFP (8) and W1C (8); Y145F is shared by the GFP variants Sapphire (20), p4-3 (8) and EBFP (7,8); L221V is shared by EGFPwe mutants (24). It has been proposed that the V163A substitution facilitates the correct folding and maturation of GFP (6). Other amino acid changes such as S202T and A206T are first described in this study. Although the crystal structure of GFP is resolved (25) it is not easy to explain the increased fluorescence of our variant GFP transfectants without more detailed studies of the encoded mRNAs and proteins. As the configuration of the GFP chromophore has already been optimized, the most likely reason seems to be improved protein maturation or stability. The amino acid substitutions A206T and L221V are localized in a region known to influence GFP dimerization and their effect might be due to the enhancement of dimer formation, but other effects cannot be ruled out. Regardless of the molecular basis, the increased fluorescence conferred by the new GFP variants and in particular GFPnovo1 and GFPnovo2 should be useful in situations where maximal sensitivity is required for bio-imaging.

Compared to alternative artificial evolution techniques, protein optimization in hypermutating B cell lines like DT40 has the advantage that it is performed within the real life environment of a vertebrate cell. This can be important as the properties of many proteins are affected by intra-cellular parameters such as pH and salt concentrations, folding, post-translational modifications and degradation. Although optimization of proteins has been achieved in hypermutating mammalian B cell lines such as Ramos (11) or 18–81 (24), DT40 offers the advantages that transgenes can be inserted as single copies into the Ig light chain locus and AID expression can be controlled. In the future, transgene diversification in DT40 may be accomplished by a combination of hypermutation and gene conversion if homologous conversion donor sequences are inserted upstream of the transgene. It has already been demonstrated that gene conversion can be used for the diversification of transgenes in DT40 (20).

The high efficiency of targeted gene integration in DT40 allows one to standardize the hypermutation of transgenes and the confirmation of mutant phenotypes (Figure 5). We envision that the transgene of interest is first cloned into the pHypermut2 vector and then transfected into a DT40 clone which conditionally expresses AID. Transfectants, which hypermutate the transgene after targeted integration into the rearranged Ig light chain locus, are subjected to iterative cycles of cell selection and expansion. When a cell population of the desired phenotype is isolated, hypermutation is shut off by excision of the AID expression cassette and the cells are subcloned. Transgene sequences from the most promising subclones are amplified by PCR and cloned into the pHypermut2 vector for sequencing. In the end, the mutated transgenes within the pHypermut2 vector are...
transfected back into AID negative cells to verify their phenotype.

In principle, DT40 can be used for the evolution of any coding or non-coding sequence whose transcription is tolerated. However, evolution of transgenes encoding non-fluorescent proteins is only feasible, if rare cells carrying desirable mutations can be enriched from the bulk culture. As FACS sorting is ideal for selection, one might attempt to tie the optimization of non-fluorescent proteins to fluorescent signals. For example, a cell surface receptor expressed on the surface of DT40 might be optimized using a fluorescent coupled ligand. Most likely, the selection strategies will need to be customized for different types of proteins and the outcome of these efforts will determine the success of artificial evolution in DT40.

**ACKNOWLEDGEMENTS**

We are grateful to Claire Brellinger for excellent technical assistance. This work was supported by the Geninteg grant of the FP6 framework and the DGF program ‘Networks in genome expression and maintenance’. Funding to pay the Open Access publication charges for this article was provided by XXX.

*Conflict of interest statement.* None declared.

**REFERENCES**

1. Milstein, C. and Rada, C. (1995) The maturation of the antibody response. In Honjo, T. and Alt, F.W. (eds), *Immunoglobulin Genes*, 2nd edn. Academic Press, London, pp. 57–81.
2. Matsuura, T. and Yomo, T. (2006) In vitro evolution of proteins. *J. Biosci. Bioeng.*, **101**, 449–456.
3. Duñer, P., Jermutus, L. and Minter, R.R. (2006) Harnessing phage and ribosome display for antibody optimisation. *Trends Biotechnol.*, **24**, 523–529.
4. Dwyer, M.A., Looger, L.L. and Hellinga, H.W. (2004) Computational design of a biologically active enzyme. *Science*, **304**, 1967–1971.
5. Shimomura, O. (1979) Structure of the chromophore of *Aequorea* green fluorescent protein. *FEBS Lett.*, **104**, 220–222.
6. Zacharias, D.A. and Tsien, R.Y. (2006) *Methods Biochem. Anal.*, **47**, 83–120.
7. Patterson, G.H., Knobel, S.M., Sharif, W.D., Kain, S.R. and Piston, D.W. (1997) Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. *Biophys. J.*, **73**, 2782–2790.
8. Cubitt, A.B., Woolfenden, L.A. and Heim, R. (1999) Understanding structure-function relationships in the *Aequorea* victoria green fluorescent protein. *Methods Cell Biol.*, **58**, 19–30.
9. Cumbers, S.J., Williams, G.T., Davies, S.L., Grenfell, R.L., Takeda, S., Battista, F.D., Sale, J.E. and Neuberger, M.S. (2002) Generation and iterative affinity maturation of antibodies in vitro using hypermutating B-cell lines. *Nat. Biotechnol.*, **20**, 1129–1134.
10. Seo, H., Masuoka, M., Murofushi, H., Takeda, S., Shibata, T. and Ohta, K. (2005) Rapid generation of specific antibodies by enhanced homologous recombination. *Nat. Biotechnol.*, **23**, 731–735.
11. Wang, L., Jackson, W.C., Steinbach, P.A. and Tsien, R.Y. (2004) Evolution of new nonantibody proteins via iterative somatic hypermutation. *Proc. Natl Acad. Sci. USA*, **101**, 16745–16749.
12. Buerstedde, J.-M. and Takeda, S. (1991) Increased ratio of targeted homologous recombination and iterative affinity maturation of antibodies in vitro using hypermutating B-cell lines. *Cell*, **67**, 179–188.
13. Buerstedde, J.-M., Reynaud, C.A., Humphries, E.H., Olson, W., Evert, D.L. and Weill, J.C. (1990) Light chain gene conversion continues at high rate in an ALV-induced cell line. *EMBO J.*, **9**, 921–927.
14. Sale, J.E., Calandrini, D.M., Takata, M., Takeda, S. and Neuberger, M.S. (2001) Ablation of XRCC2/3 transforms immunoglobulin V gene conversion into somatic hypermutation. *Nature*, **412**, 921–926.
15. Arakawa, H., Saribasak, H. and Buerstedde, J.-M. (2004) Activation-induced cytidine deaminase initiates immunoglobulin gene conversion and hypermutation by a common intermediate. *PLoS Biol.*, **2**, E179.
16. Arakawa, H., Hauschild, J. and Buerstedde, J.-M. (2002) Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion. *Science*, **295**, 1301–1306.
17. Cormack, B.P., Valdivia, R.H. and Falkow, S. (1996) FACS-optimized mutants of the green fluorescent protein (GFP). *Gene*, **173**, 33–38.
18. Zhang, Y., Riester, C., Ayrall, A.M., Soblitzyk, F., Littlewood, T.D. and Reth, M. (1996) Inducible site-directed recombination in mouse embryonic stem cells. *Nucleic Acids Res.*, **24**, 543–548.
19. Arakawa, H., Lodygin, D. and Buerstedde, J.M. (2001) Mutant loxP vectors for selectable marker recycle and conditional knock-outs. *BMC Biotechnol.*, 1, 7.

20. Kanayama, N., Todo, K., Takahashi, S., Magari, M. and Ohmori, H. (2006) Genetic manipulation of an exogenous non-immunoglobulin protein by gene conversion machinery in a chicken B cell line. *Nucleic Acids Res.*, 34, e10.

21. Zapata-Hommer, O. and Griesbeck, O. (2003) Efficiently folding and circularly permuted variants of the Sapphire mutant of GFP. *BMC Biotechnol.*, 3, 5.

22. Nagai, T., Ibata, K., Park, E.S., Kubota, M., Mikoshiba, K. and Miyawaki, A. (2002) A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.*, 20, 87–90.

23. Rekas, A., Alattia, J.R., Nagai, T., Miyawaki, A. and Ikura, M. (2002) Crystal structure of venus, a yellow fluorescent protein with improved maturation and reduced environmental sensitivity. *J. Biol. Chem.*, 277, 50573–50578.

24. Wang, C.L., Yang, D.C. and Wabl, M. (2004) Directed molecular evolution by somatic hypermutation. *Protein Eng. Des. Sel.*, 17, 659–664.

25. Ormo, M., Cubitt, A.B., Kallio, K., Gross, L.A., Tsien, R.Y. and Remington, S.J. (1996) Crystal structure of the Aequorea victoria green fluorescent protein. *Science*, 273, 1392–1395.

26. Crameri, A., Whitehorn, E.A., Tate, E. and Stemmer, W.P. (1996) Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat. Biotechnol.*, 14, 315–319.