Fast Directed Evolution of Non-Immunoglobulin Proteins by Somatic Hypermutation in Immune Cells

Christian Heinis* and Kai Johnsson[a]

Directed evolution is a widely used technique for modifying or enhancing protein performance. The process of directed protein evolution begins with the creation of a library of mutated genes. Gene products that show improvement with respect to the desired property or set of properties are identified by selection or screening. In this article, we highlight a recently published innovative methodology for the efficient mutagenesis of proteins that has the potential to accelerate the directed evolution of many proteins.

The number of possible protein variants that can be generated by mutation is very large and increases rapidly with the size of the protein and the number of amino acids that are allowed to mutate simultaneously. If, for example, five amino acid substitutions are introduced randomly into a protein of 200 amino acids, the number of theoretically possible mutants exceeds 10^{18}. This number is far beyond the size of a large gene library comprising 10^{9}–10^{10} mutants that can be synthesized and transformed into cells with state-of-the-art recombinant DNA techniques. As most mutations are deleterious, the chance of identifying improved proteins in libraries containing large numbers of mutations and being smaller than the libraries' theoretical size, is very small. For this reason, molecular biologists usually restrict themselves to theoretical library sizes that can be practically generated and analyzed. This can be achieved by limiting the number of mutations that are simultaneously introduced into a protein or by targeting mutations to a small region of the protein. As these restrictions in library design only allow a small fraction of the possible protein variants, it is important that the process of mutagenesis and screening is iteratively repeated to sample a larger sequence space. However, the repetitive generation of protein libraries and subsequent screening is labour-intensive and is therefore limited to one or a maximum of a few cycles.

In order to speed up the process of molecular evolution, researchers have developed in vivo mutagenesis techniques that avoid the process of laborious library construction. In the early days of molecular evolution, organisms harbouring a target gene were exposed to UV light or to chemical mutagens to introduce mutations.[1] Later, *E. coli* strains that exhibit increased mutation rates were employed to mutagenize genes of interest. An *E. coli* strain named XL-Red was constructed with mutations in three independent DNA-repair pathways, thereby exhibiting a spontaneous mutation frequency of about 0.5 mutations per 1000 nucleotides of DNA.[2] A disadvantage of the mutator strain is that mutations are not restricted to the sequence of interest. Widespread mutagenesis in the genome can obscure phenotypic expression of mutations in the target gene. To target evolution to specific genes, a highly error-prone DNA polymerase I has been generated that shows high levels of mutagenesis of a target sequence encoded in a Pol-I-dependent plasmid (8.1×10^{-4} mutations per base pair) and little mutation of the chromosome.[3]

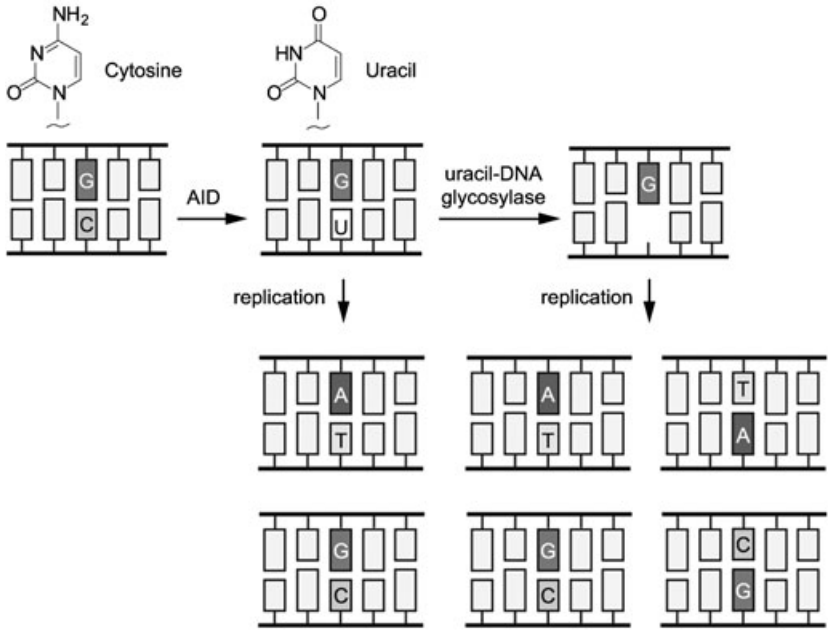
A powerful new strategy for targeted in vivo mutagenesis in immune cells has been published independently by two research groups and is highlighted here. Clifford L. Wang in the group of Matthias Wabl and the team of Roger Tsien have employed the process of somatic hypermutation (SHM) in B lymphocytes for the directed evolution of autofluorescent proteins.[4,5]

The process of somatic hypermutation is used in nature by the immune system to diversify the antibody repertoire, thereby allowing the organism to cope with the immense variability of pathogenic agents and organisms.[6] Upon antigenic stimulation of B lymphocytes, the variable (V) regions of Ig loci mutate at a rate of 10^{-5}–10^{-3} mutations per base pair per generation; this is about 10^{6} times higher than that in the rest of the genome.[7,8] This somatic hypermutation requires activation-induced cytidine deaminase (AID), an enzyme expressed exclusively in activated B cells of the germinal centre that converts deoxycytidine to deoxyuridine.[9] It is thought that replication of the dU/dG template leads to transition mutations. In an alternative process, uracil DNA glycosylase (UNG) removes uracil generated by AID deamination, and mutations are presumably generated from replication of the abasic site (Scheme 1). The hypermutatable segment in the Ig genes encompasses a 2 kb DNA stretch that includes the rearranged V(D)J segment and thus defines antigen receptor specificity. Whether SHM is locus-specific for the IgV gene or whether other regions of the genome are also mutated by the mechanism of SHM is under debate.[10]

In the first work discussed here, the research team of Wang et al. have explored whether hypermutation in activated immune cells can be employed to restore the fluorescence intensity of an enhanced green fluorescent protein (EGFP) mutant with 1000-fold reduced

[a] *Dr. C. Heinis, Prof. K. Johnsson*
Institute of Chemical Sciences and Engineering
École Polytechnique Fédérale de Lausanne (EPFL)
1015 Lausanne (Switzerland)
Fax: (+41) 21-693-9365
E-mail: christian.heinis@epfl.ch

  DOI: 10.1002/cbic.200500011



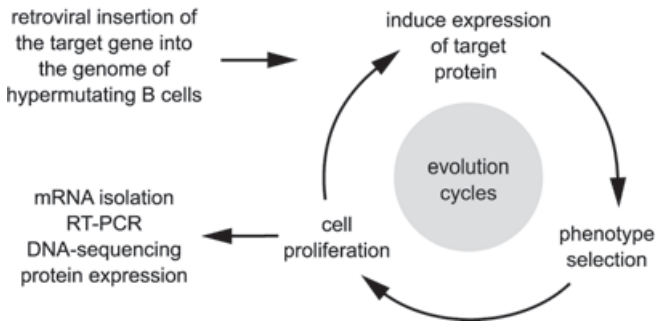
Scheme 1. *Somatic hypermutation (SHM) uses activation-induced cytidine deaminase (AID) for the deamination of cytidine. It is believed that subsequent replication leads to transition mutations. Alternatively, the uridine base is excised by uridine-DNA glycosylase, and transition and transversion mutations are thought to be introduced by replication of the apyrimidinic site.*

fluorescence (Thr66 of the chromophore was replaced by isoleucine).[4] They introduced their target gene by retroviral infection into the cell line 18–81, which displays constitutive hypermutation of the Ig V genes during culture. It has previously been shown that hypermutation does not require Ig; this suggests that non-Ig genes can also be hypermutated.[11,12] Nearby enhancers and other sequences from the Ig intron are expected to direct hypermutation to the V regions.[13] Wang et al. reported that, after five days of expression, fluorescent cells were detected and, after 32 days, several cells with varying fluorescence intensities were present; this indicated that their genes were mutated. Indeed, sequencing of EGFP genes from cells selected by fluorescence-activated cell sorting (FACS) revealed that 1–3 bases were mutated per gene. In all the mutants with enhanced fluorescence intensity, the previously mutated amino acid residue 66 either reverted to threonine or mutated to valine, leucine or methionine, or amino acid residue 221 was changed from leucine to valine. It would be intriguing to see whether other beneficial mutations might accumulate in the target gene if the process of hypermutation and FACS selection were repeated iteratively.

In an independent study, the team of Roger Tsien tested whether the monomeric red fluorescent protein (mRFP) with far-red-shifted emissions can be evolved through iterative SHM and FACS (Scheme 2).[5] Tsien and his colleagues expressed mRFP in Burkitt lymphoma Ramos cells, a human B cell line, and analyzed whether the genes of fluorescent cells, which had been enriched in six consecutive rounds of SHM and FACS, were mutated. They found that 12 out of 20 genes were mutated, each having 1–3 mutations. Moreover, the authors present evidence that the mutation rate can be varied by controlling the level of transcription. They found that $>15\%$ of the cells lose their fluorescence when doxycycline, an inducer of transcription, was present for 120 h, whereas $<5\%$ lost fluorescence in the presence of the doxycycline for only 24 h; this indicated that a longer transcription period generates more mutations. This result is in line with findings by other groups that hypermutation rates are dependent on the level of transcription, and it provides a simple but efficient mechanism for controlling the mutagenesis rate.[14]

The authors next subjected cells harbouring the mRFP gene to 23 rounds of iterated SHM and FACS in order to accumulate beneficial mutations. In each round, cells with a higher emission wavelength and a fluorescence intensity above a fixed threshold were enriched. The authors reported several mutants with improved fluorescence properties, including a mutant with an emission maximum of 649 nm, 37 nm longer than the parental mRFP and 12 nm beyond the furthest red-emitter available so far.[15] The excitation maximum of this mutant was unchanged (590 nm), while the fluorescence quantum yield was slightly reduced (from 0.25 to 0.1). Another noteworthy mutant was selected in round 10 and had the largest excitation wavelength (598 nm), 8 nm shifted towards a longer wavelength. In addition to shifts of excitation and emission wavelengths, all of the reported mutants were more resistant to photobleaching than the parental mRFP.

In order to analyze the evolutionary pathway, Tsien's group sequenced mu-



Scheme 2. *Directed evolution of proteins by iterative SHM of the target gene in activated B cells and selection of cells harbouring an improved phenotype.*

tants with improved fluorescence properties selected after 10, 14, 21 and 23 cycles of mutagenesis and selection. They found that mutations (including silent mutations) were generated in each round and that beneficial mutations (such as V16E and I161M) were preserved from round to round. One position (amino acid residue 65) was mutated to cysteine in round 10 and subsequently to isoleucine in round 14; this indicated that SHM allows continuous exploration of sequence space. The finding that beneficial mutations accumulated in each round emphasizes the power of iterative evolutionary cycles. The authors report that attempts to further increase the emission wavelength of the mutants by saturation mutagenesis of the amino acid positions identified by SHM failed. This suggests that iterative SHM had efficiently explored a very large sequence space and that the best amino acid residues in these positions had been identified. Another indication for the large sequence space that was sampled by SHM is the numerous mutations found in the long terminal repeat sequence (LTR), a region of the retroviral vector that was used by the authors to introduce the mRFP gene into the Ramos genome and that was not biased by phenotypic selection. In one mutant with improved fluorescence, the sequenced 117 bp stretch at the 3′ end of the LTR contained 15 mutations, including an insertion, whereas two clones that were randomly picked from round 2 possessed only one mutation in this region.

In order to test whether the gene of interest was inserted into a Ig V locus, Tsien and colleagues determined the integration loci of the target gene in the Ramos genome. In round 2, the mRFP genes were found in chromosomes that do not contain Ig genes (mainly chromosomes 5, 16, 18 and 20), whereas mutants from round 23 were all found at a single integration site at the Ig heavy-chain locus in chromosome 14. These findings suggest that exogenous genes are mutated at many loci but that desired properties are more likely to be generated in loci with higher mutation frequencies.

In summary, the successful isolation of mRFP mutants with improved fluorescence properties and the recovery of the fluorescence of a EGFP mutant by the two research teams demonstrate that SHM is a powerful tool for evolving other proteins than antibodies. In principle, the technique need not be limited to fluorescent proteins as long as there is a screening or selection assay available to identify cells harbouring proteins with the desired activity. The ability to rapidly repeat mutagenesis and selection cycles will certainly attract other bioscientists to use SHM in B lymphocytes for protein evolution.

Keywords: directed evolution · fluorescence · mutagenesis · proteins · somatic hypermutation

[1] A. W. Nguyen, P. S. Daugherty, *Methods Mol. Biol. (Totowa NJ U.S.)* **2003**, *231*, 39.
[2] A. Greener, M. Callahan, B. Jerpseth, *Mol. Biotechnol.* **1997**, *7*, 189.
[3] M. Camps, J. Naukkarinen, B. P. Johnson, L. A. Loeb, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 9727.
[4] C. L. Wang, D. C. Yang, M. Wabl, *Protein Eng. Des. Sel.* **2004**, *17*, 659.
[5] L. Wang, W. C. Jackson, P. A. Steinbach, R. Y. Tsien, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 16745.
[6] Z. Li, C. J. Woo, M. D. Iglesias-Ussel, D. Ronai, M. D. Scharff, *Genes Dev.* **2004**, *18*, 1.
[7] M. Wabl, P. D. Burrows, A. von Gabain, C. Steinberg, *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 479.
[8] K. Rajewsky, I. Forster, A. Cumano, *Science* **1987**, *238*, 1088.
[9] M. Muramatsu, V. S. Sankaranand, S. Anant, M. Sugai, K. Kinoshita, N. O. Davidson, T. Honjo, *J. Biol. Chem.* **1999**, *274*, 18470.
[10] C. L. Wang, R. A. Harper, M. Wabl, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7352.
[11] M. S. Gordon, C. M. Kanegai, J. R. Doerr, R. Wall, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 4126.
[12] H. M. Shen, A. Peters, B. Baron, X. Zhu, U. Storb, *Science* **1998**, *280*, 1750.
[13] A. G. Betz, C. Milstein, A. González-Fernández, R. Pannell, T. Larson, M. S. Neuberger, *Cell* **1994**, *77*, 239.
[14] J. Bachl, C. Carlson, V. Gray-Schopfer, M. Dessing, C. Olsson, *J. Immunol.* **2001**, *166*, 5051.
[15] A. F. Fradkov, V. V. Verkhusha, D. B. Staroverov, M. E. Bulina, Y. G. Yanushevich, V. I. Martynov, S. Lukyanov, K. A. Lukyanov, *Biochem. J.* **2002**, *368*, 17.

Received: January 12, 2005
Published online on March 16, 2005

