

# Nucleotide exchange and excision technology (NExT) DNA shuffling: a robust method for DNA fragmentation and directed evolution

Kristian M. Müller\*, Sabine C. Stebel, Susanne Knall, Gregor Zipf<sup>1</sup>,  
Hubert S. Bernauer<sup>1</sup> and Katja M. Arndt

Institut für Biologie III, Universität Freiburg, Schänzlestraße 1, 79104 Freiburg, Germany and  
<sup>1</sup>ATG:Biosynthetics, Freiburg, Germany

Received March 30, 2005; Revised and Accepted July 8, 2005

## ABSTRACT

**DNA shuffling is widely used for optimizing complex properties contained within DNA and proteins. Demonstrated here is the amplification of a gene library by PCR using uridine triphosphate (dUTP) as a fragmentation defining exchange nucleotide with thymidine, together with the three other nucleotides. The incorporated uracil bases were excised using uracil-DNA-glycosylase and the DNA backbone subsequently cleaved with piperidine. These end-point reactions required no adjustments. Polyacrylamide urea gels demonstrated adjustable fragmentation size over a wide range. The oligonucleotide pool was reassembled by internal primer extension to full length with a proofreading polymerase to improve yield over *Taq*. We present a computer program that accurately predicts the fragmentation pattern and yields all possible fragment sequences with their respective likelihood of occurrence, taking the guesswork out of the fragmentation. The technique has been demonstrated by shuffling chloramphenicol acetyltransferase gene libraries. A 33% dUTP PCR resulted in shuffled clones with an average parental fragment size of 86 bases even without employment of a fragment size separation, and revealed a low mutation rate (0.1%). NExT DNA fragmentation is rational, easily executed and reproducible, making it superior to other techniques. Additionally, NExT could feasibly be applied to several other nucleotide analogs.**

## INTRODUCTION

Since the first reports of hybrid gene synthesis (1), PCR-based gene cross-overs (2) and PCR-based gene synthesis (3), were

published, the idea of directing evolution (4) initialized the development of various methods for the shuffling of gene libraries (5), which permit homologous recombination *in vitro*. To date, however, all of these methods have not been without disadvantage or difficulty. In the well-established protocol of Stemmer, DNase is used to fragment DNA requiring careful optimization of the digest conditions, e.g. time, temperature, amount of nuclease and DNA (4,6). Other methods such as the staggered extension process (7) and random-priming (8) are limited by the DNA composition, and matters are complicated further by the lack of controllability of the range of fragment sizes generated. Methods such as RACHITT (9) also require DNase digests and are even more labor intensive. The race for the best method is still on. Simple comparisons (9) can be helpful but need to be taken with caution since the gene length, the homology of the shuffled gene libraries and the intended cross-over rate would have to be taken into account. Besides the homology-dependent methods, which are related to the presented data, homology-independent methods have also been developed based on DNA fragment fusion [e.g. thio-ITCHY (10) and SHIPREC (11)].

We have devised a new method that, we are confident, is both rational and robust. Nucleotide Exchange and Excision Technology (NExT) DNA shuffling is based on the random incorporation of 'exchange nucleotides'. The occurrence and position of these exchange nucleotides in the DNA will dictate the subsequent fragmentation pattern without the need for further adjustment. We used highly homologous libraries with a few members to be able to analyze our fragmentation and shuffling results in detail. The key advantages of our method are (i) calculable experimental setup aided by a computer program, (ii) reproducible end-point reactions without adjustments, (iii) no gel purification required, (iv) efficient reassembly with a proofreading polymerase, (v) gene recombination including very short fragments of only a few bases, (vi) low error rate and (vii) practically no contamination with unshuffled clones.

\*To whom correspondence should be addressed. Tel: +49 761 203 2748; Fax: +49 761 203 2745; Email: kristian@biologie.uni-freiburg.de

## METHODS

### Cloning steps

Genes used in the NEXt DNA shuffling procedures were the 657 bp chloramphenicol acetyl transferase I (CAT) wild-type gene (CATwt, SwissProt: P00483, PDB: 1NOC:B) and variants coding for an N-terminally 10 amino acid truncated (the Met start codon was added again), C-terminally 9 amino acid truncated or double-truncated CAT (CAT\_Nd10, CAT\_Cd9, CAT\_Nd10\_Cd9) or C-terminally 26 amino acid truncated CAT (CAT\_Cd26). For shuffling and error-prone PCR genes were amplified using the primers Pr-N-shuffle (5'-ATT-TCTAGATAACGAGGGCAA-3') and Pr-C-shuffle (5'-ACT-TCACAGGTCAAGCTTTC-3') for the wild-type and N-terminally truncated genes, Pr-N-shuffle and Pr-Cdx-shuffle (5'-CTTCACAGGTCAAGCTTATCA-3') for the C-terminally truncated and for the double-truncated genes. Priming sites were located shortly before and after the gene adding in total 45 nt to the genes and contained the restriction sites XbaI and HindIII for cloning into the vector pLisc-SAFH11 (12) thus replacing part of the original plasmid. Plasmids were transformed by electroporation in *Escherichia coli* strain RV308 (ATCC No. 31608) using standard methods (13). Mutated variants of the various clones were obtained by error-prone PCR using 2.5 U *Taq* polymerase in a 50  $\mu$ l reaction supplemented with vendor-supplied buffer (Genaxxon, GeneCraft, Amersham) and with 7 mM MgCl<sub>2</sub>, 0.5 mM MnCl<sub>2</sub>, 0.4 mM each dNTP and 50 ng template. The PCR protocol was as follows: 1 cycle of 94°C, 3 min; 30 cycles of 92°C, 1 min; 60°C, 1 min; 72°C, 2 min; 1 cycle of 72°C, 7 min.

### Uridine exchange PCR

The uridine versus thymidine exchange PCR mixture contained 50 ng template (0.017 pmol of a 4340 bp plasmid), 25 pmol of each primer (see above), 0.2 mM of dATP, dGTP and dCTP each, a 0.2 mM mixture of dUTP:dTTP in various ratios, 5 U *Taq* DNA polymerase (GeneCraft, Amersham, Genaxxon), and 5  $\mu$ l 10 $\times$  PCR buffer containing 160 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 670 mM Tris-HCl, pH 8.8 (at 25°C), 15 mM MgCl<sub>2</sub>, 0.1% Tween-20 for the reactions shown in the ethidium bromide stained gels and 5  $\mu$ l 10 $\times$  PCR buffer containing 100 mM Tris-HCl, pH 9.0 (at 25°C), 500 mM NaCl, 15 mM MgCl<sub>2</sub>, 1% Triton X-100 for the reactions shown in the autoradiographed gel. The volume was adjusted to 50  $\mu$ l with H<sub>2</sub>O. Before adding to the reaction, the 100 mM nucleotide stock solutions (Peqlab, Germany) were diluted in water to 10 mM for dATP, dGTP and dCTP, and to 1 mM for dUTP and dTTP. For radioactive experiments, 0.5  $\mu$ l of a 3.3  $\mu$ M [<sup>32</sup>P]dCTP solution or 0.5  $\mu$ Ci, respectively, were added. The cyclor program was as follows: 1 cycle of 94°C, 1 min; 25 cycles of 92°C, 30 s; 62°C, 20 s; 72°C, 2 min; final incubation 72°C, 4 min. To obtain sufficient product, four 50  $\mu$ l reactions were combined, separated on a 1% agarose gel, purified using one column of a PCR clean-up kit (Amersham GFX Gel Band Purification Kit) and eluted with 50  $\mu$ l of 10 mM Tris, pH 8.0. For radioactive experiments, the clean-up kit was used without the gel step. The concentration of the PCR product was determined by taking the baseline corrected 260 nm value of an absorption spectrum from 220 to

350 nm of a 1:30 diluted 5  $\mu$ l aliquot in a 140  $\mu$ l microcuvette. Product yield for 200  $\mu$ l of PCR was 10–17  $\mu$ g.

### Enzymatic digest and chemical cleavage

About 15  $\mu$ g in 45  $\mu$ l (minimal 7  $\mu$ g) of the purified PCR product were supplemented with 6  $\mu$ l supplied uracil-DNA-glycosylase (UDG) 10 $\times$  buffer and 2 U *E.coli* UDG (Peqlab, Germany), adjusted to 60  $\mu$ l with water and digested for 1 h at 37°C. The DNA was cleaved by adding piperidine (Sigma) to a final concentration of 10% (v/v) and heated for 30 min at 90°C in a thermocycler with heated lid. Piperidine is toxic and should be handled in a hood. Alternatively, piperidine was replaced by a 5 M NaOH stock solution added at 10% (v/v) to the cleavage reaction.

### Fragment purification

Fragments were purified directly from the piperidine or NaOH cleavage using the QiaexII kit (Qiagen) according to the manufacturer manual. The capture buffer included was added and neutralized (~20  $\mu$ l of 3 M Na-acetate, pH 5.3). After two washing steps, fragments were extracted two times with 25  $\mu$ l of 10 mM Tris pH 8.0 and pooled. Two centrifugation steps with transfer to a fresh tube ensured that oligonucleotides were not contaminated with matrix. Note that Qiagen recommends this kit for fragments longer than 40 bp. For the extraction of fragments from polyacrylamide urea gels, the excised slices were crushed and incubated either with 1 ml water or diffusion buffer containing 0.5 M ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA, 0.1% SDS, pH 8.0 in a thermomixer (Eppendorf) at 37°C, 1000 r.p.m. overnight. The water extracted oligonucleotides were precipitated by adding sodium acetate, MgCl<sub>2</sub> and 2-propanol. The diffusion-buffer-extracted fragments were purified with the QiaexII kit as described above. Initially, fragments were quantified by mixing with SYBR Green II (Molecular Probes) and measuring fluorescence emission intensity relative to a 60 bp oligonucleotide calibration curve.

### Denaturing polyacrylamide urea gel

Gels were composed of 6.7 M urea, 11.3% polyacrylamide/bisacrylamide (37.5:1), 1 $\times$  TBE, ammoniumperoxodisulfate and TEMED (13). Gels (10 cm  $\times$  8 cm  $\times$  1 mm) were prepared freshly, as older gels did not run properly, and electrophoresed in a Hoefer mighty small basic unit (Amersham) heated to 56°C with an attached temperature-controlled water bath. Before loading, the cleaved DNA was concentrated to 7  $\mu$ l in a speed-vac in order to evaporate the piperidine, supplemented with 25  $\mu$ l of deionized formamide and heated to 80°C for 3 min in a thermocycler; 9  $\mu$ l of the sample were loaded on the gel. For the radioactive experiments, 7  $\mu$ l of the DNA-formamide sample were additionally supplemented with 3  $\mu$ l of 60% sucrose solution, which improved loading, and 7  $\mu$ l H<sub>2</sub>O and then 15  $\mu$ l were loaded. Oligonucleotides of 20, 38, 48, 58, 65 or 68 nt, as well as a 100 bp ladder (New England Biolabs) with added bromophenol blue dye served as visible length standard. For radioactive experiments, the oligonucleotides were kinased with [ $\gamma$ -<sup>32</sup>P]ATP and purified by size exclusion. After heating and a 10 min pre-run at 100 V, the gel was loaded and run at 170 V until the dye was 1–2 cm from the bottom of the gel. The gel was stained in 30 ml,

1.2 µg/ml ethidium bromide for 5 min. Note that with longer incubation times the smaller fragments start to elude from the gel (therefore SYBR Green II was not suited for staining) and exposure to UV light bleaches the gels.

### Gene reassembly and amplification

For the reassembly ~2 µg of the purified DNA fragments (typically 20 µl) were mixed with 4 µl of a mix of 10 mM of each dATP, dTTP, dCTP and dGTP (800 µM final), 4 U Vent DNA Polymerase (NEB) with 1–4 µl of 25 mM MgSO<sub>4</sub> and 5 µl supplied 10× buffer. For the experiments with *Taq*, 1 or 4 µl of the dNTP mix were tested (without noticeable differences in yield) and 5 U of the enzyme were used. The volume was adjusted to 50 µl. In case of the Vent polymerase, even <1 µg of DNA was sufficient. Cycles for the reassembly were as follows (Eppendorf mastercycler): 1 cycle of 94°C, 3 min; 36 cycles of 92°C, 30 s; 30°C, 60 s + 1°C per cycle (cooling ramp 1°C/s); 72°C, 1 min + 4 s per cycle; final incubation at 72°C, 3 min. Ten microliters of the reassembly product (this volume was chosen to ensure diversity) were amplified using a standard PCR reaction (25 pmol primer, 0.2 mM dNTPs, 25 cycles, 40 s elongation time) with the appropriate primers listed in cloning steps. Amplified genes were cloned via the *Xba*I and *Hind*III restriction sites, and plasmids prepared from *E. coli* grown on plates without selection pressure were named, e.g. pNd10\_Cd9\_control# and equivalents. Clones were sequenced using the Big Dye termination kit 1.1 or 3.0 (Applied Biosystems) and analyzed in an ABI Prism sequencer.

## RESULTS AND DISCUSSION

### Test-libraries

The NExT procedure was developed and tested by increasing the functionality of truncated mutants of chloramphenicol acetyl transferase I (CAT), which mediates resistance against the antibiotic chloramphenicol. Directed evolution was independently applied to four sets of variants truncated at the genetic level. The first library of CAT mutants was shortened by ten amino acids at the N-terminus (CAT\_Nd10) while maintaining the start methionine, the second library by 9 amino acids at the C-terminus (CAT\_Cd9), the third library by 26 amino acids at the C-terminus (CAT\_Cd26), and the fourth library was truncated at both ends by 10 and 9 amino acids (CAT\_Nd10\_Cd9), respectively. Besides testing the NExT method, these experiments were set up to elucidate the structure–function relations of this thermostable enzyme. We also wanted to test the applicability of our structure perturbation strategy (14) to improve the thermostability of already thermostable enzymes. Detailed data for the NExT shuffling were obtained using test ‘libraries’ with three to six members selected from error-prone PCR diversification steps (15) containing 14–49 mutations within the genes of 627 (CAT\_Nd10, CAT\_Cd9) to 579 (CAT\_Cd26) bp in length. The small number of library members with a manageable set of mutations ensured that almost all mutations found in recombined clones could be unambiguously traced to parental segments. The biochemical and biophysical characterization of improved CAT enzymes will be published

elsewhere (S.C. Stebel, manuscript in preparation). For example, the melting temperature of the truncated CAT\_Nd10 variant was increased by 24°C as detected by circular dichroism measurements.

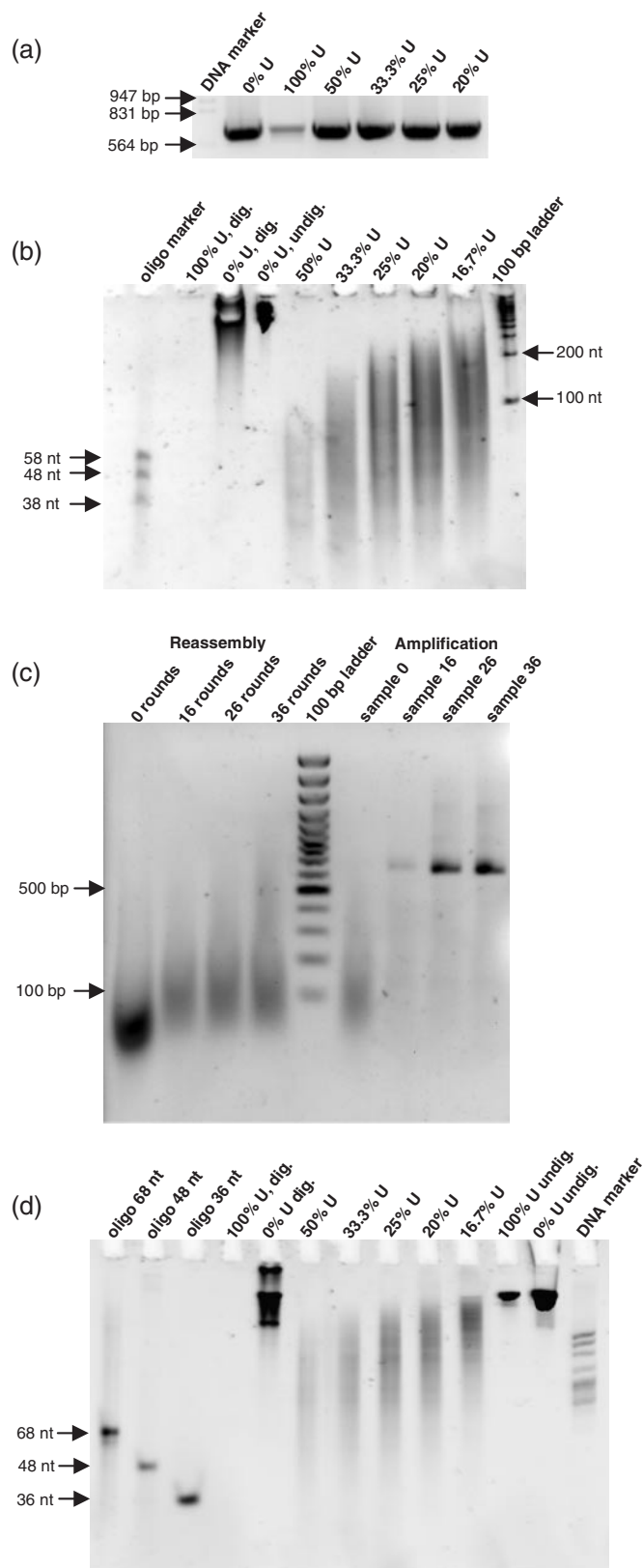
### Preferred NExT implementation

In the following, we first demonstrate the NExT DNA shuffling in the preferred protocol and later give data on variations of this technique. Uridine was chosen as exchange nucleotide because dUTP is known to be incorporated into the DNA by various polymerases (16). A NExT shuffling procedure was carried out according to the following protocol: first, a PCR reaction with *Taq* polymerase amplified the gene pool and additionally incorporated uridine. Various ratios of dUTP:dTTP could be used to obtain the optimal fragmentation (Figure 1a). No apparent difference in the amount of PCR product was observed when using dUTP fractions of up to 50% within the dUTP and dTTP mixture. Only the PCR sample containing uridine alone yielded about a quarter of the product. Second, the PCR product was agarose gel-purified to separate it from any non-uracil containing template. Third, the exchange nucleotide was cleaved out by incubating with uracil-DNA-glycosylase (UDG) (17). This enzyme attacks double-stranded as well as single-stranded DNA using a hydrolytic mechanism to remove the uracil moiety by a nucleophilic attack at the C1' position (18). Fourth, piperidine was used to split the backbone positions where a uracil had been cleaved out by UDG. Piperidine is a well-established cleaving agent for chemical sequencing (19). As in our case, the base moiety is already eliminated at this step we propose that piperidine mainly promotes two base-catalyzed beta-eliminations of the phosphates. This is supported by our observation that replacing piperidine with NaOH resulted in an almost identical fragment distribution (data not shown). The result of such a cleavage reaction was analyzed with high resolution on denaturing polyacrylamide urea gels for dUTP fractions ranging from 100 to 0% (Figure 1b) and quantified by image analysis (Figure 2a). Several length distributions with defined maxima were easily obtained including pools of very small and large fragments, which is optimal for shuffling short genes or long gene clusters, respectively. Such initial tests determined the optimal dUTP fraction for the given gene. Once established, the chosen dUTP fraction of 33.3% for the test libraries was found to be highly reproducible. The gel analysis step was no longer required and omitted for subsequent experiments.

Resulting fragments were cleaned simply and quickly direct from the cleavage reaction solution using a silica-based resin. The full-length gene was reassembled from the fragments in an internal primer extension procedure with increasing annealing temperatures using a proofreading DNA polymerase such as Vent (Figure 1c). In an internal primer extension, also named ‘recursive’ PCR (3), the fragments serve each other as primers and thus get longer with each cycle of the PCR reaction, until full-length products are achieved. As a final step, products of the assembly reaction were amplified with a standard PCR reaction, cloned and sequenced. While establishing the method, the assembly reaction was monitored by agarose gel electrophoresis (Figure 1c). The assembly process was stopped after a serial increase in the number of PCR cycles

and products obtained at these points were subjected to the amplification PCR. Despite the harsh chemical cleavage conditions with piperidine or NaOH, the assembly worked very efficiently, and the use of a proofreading polymerase

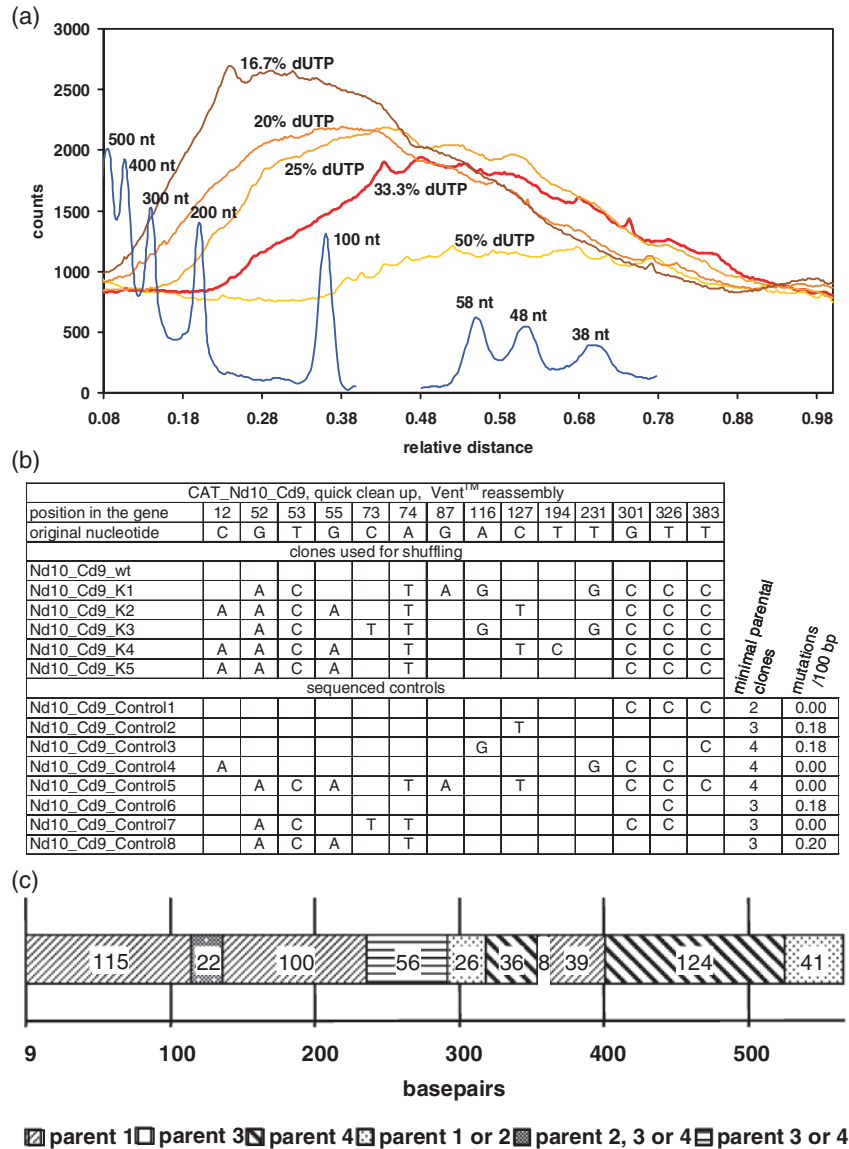
further improved the yield. In particular, the gel purification step for a defined fragment size range was omitted and yet no full-length product could be amplified without several cycles of the reassembly process demonstrating a very efficient fragmentation.



### Evaluation of NExT shuffling

The NExT DNA shuffling procedure described so far has been applied to the directed evolution of a 600 bp long CAT gene truncated at both ends (CAT\_Nd10\_Cd9). In the course of these experiments, a defined library of six clones with different mutation patterns between nucleotides 12 and 383 was shuffled based on a 33.3% uridine exchange PCR. Eight shuffled clones taken from control plates without selection pressure were sequenced (Figure 2b). The unique mutation pattern of these clones showed that all clones tested were derived from at least two (e.g. clone 1) to four (e.g. clone 4) parental clones. Within the 372 bp stretch amenable to analysis, this resulted in one cross-over per 93 to 186 bp with a mean fragment length of 114 bp. Sequencing also determined the error level of this procedure. Within 4425 bases sequenced, four alterations were found (one A to G and one T to C transition, a 1 bp insertion and a 1 bp deletion) giving a mutation rate of 0.09%. This is remarkably lower than an error rate of 0.7% reported previously for DNase shuffling (6). As detailed below, this is not a unique feature of using Vent polymerase. As our fragment distribution and cross-over rate were comparable to previous experiments, we are inclined to attribute the previously reported error rates more to the DNase digest and to the UV damage due to gel visualization rather than the fragment size and the polymerase. A low mutation rate is

**Figure 1.** Analysis of the NExT DNA shuffling technology. (a) 1% agarose gel showing the uracil-PCR products of CAT\_Nd10 clones obtained with different amounts of uridine in the reactions. For the PCR program, an extended elongation time of 2 min was chosen based on a test series showing that the yield was significantly improved compared to shorter times (data not shown). %U was calculated by  $c(\text{dUTP})/[c(\text{dUTP}) + c(\text{dTTP})] \times 100$ . (b) Polyacrylamide urea gel stained with ethidium bromide showing UDG/piperidine digests of CAT\_Nd10 PCR products obtained with various dUTP:dTTP ratios (1:0, 0:1, 1:1, 1:2, 1:3, 1:4, 1:5) to determine an optimal ratio. Digests between 1 and 3 h yielded equivalent results, indicating a selective and consistent reaction. From left to right: lane 1, oligonucleotides with 58, 48 and 36 bases as size marker; lane 2, 100% dUTP PCR digested; lane 3, 0% dUTP digested; lane 4, 0% dUTP undigested; lane 5, 50% dUTP digested; lane 6, 33.3% dUTP digested; lane 7, 25% dUTP digested; lane 8, 20% dUTP digested; lane 9, 16.7% dUTP digested; lane 10, 100 bp DNA ladder. Note that residual amounts of piperidine contribute to slightly distorted lanes. (c) 1% agarose gel of CAT\_Nd10\_Cd9 gene fragment libraries from DNA containing 33.3% U showing the reassembly process with Vent DNA polymerase and the amplification of reassembled genes with *Taq* polymerase. Lane 1, fragments without reassembly PCR; lane 2, fragments after 16 cycles of reassembly; lane 3, fragments after 26 cycles of reassembly; lane 4, fragments after 36 cycles of reassembly; lane 5, 100 bp DNA ladder; lane 6, amplification PCR of fragments without reassembly; lane 7, amplification PCR of fragments subjected to 16 reassembly cycles; lane 8, amplification PCR of fragments subjected to 26 reassembly cycles; lane 9, amplification PCR of fragments subjected to 36 reassembly cycles. (d) Polyacrylamide urea gel with UDG/T4 endonuclease V digests of CAT wild-type PCR products containing various dUTP:dTTP ratios to analyze enzymatic fragmentation. Lanes 1–3, oligonucleotides with 68, 48 and 36 bases; lanes 4–10, digests of PCR products obtained with 100%, 0%, 50%, 33.3%, 25%, 20% and 16.7% dUTP; lanes 11–12, PCR products without digest obtained with 100% and 0% dUTP; lane 13, pBR322/HpaII DNA marker. Note that the migration behavior of DNA without uracil incorporation is influenced by the digestion with UDG/piperidine or UDG/T4 endonuclease V. A small fraction of the cleavage might be attributed to this treatment.



**Figure 2.** Quantification of fragmentation size range and analysis of shuffling results. (a) Lane density plot of lane 1 and lanes 5–10 of Figure 1b detailing the fragment sizes based on the fraction of dUTP used. For the shuffling of all CAT variants, a uracil-PCR containing 33.3% dUTP was used producing fragments ranging from about 30 to 200 bases in length (thick red line). The image was acquired with a FluorS Multiimager and the plot generated using the Quantity One software (Bio-Rad). For clarity of the plot, the signal of the 100 bp ladder was shifted by  $-750$  counts and the signal of the oligonucleotides by  $-500$  counts. (b) Sequencing results of a NEXt DNA shuffling experiment with a CAT\_Nd10\_Cd9 gene mutant library with quick clean-up of fragments and reassembly using a proofreading polymerase. About 500–571 bases per clone were sequenced. The test shuffling was prepared with a 33.3% uracil-PCR containing 26 ng (52%) truncated CAT wild-type fragments and 4.8 ng (9.6%) fragments of each mutant. The bottom panel lists the sequences of clones obtained without selection pressure focusing on the shuffled mutations, the minimal number of parental clones as can be deduced from the mutation patterns and the frequency of additionally introduced mutations not listed in the table. On average, the 372 bp segment analyzed is composed of 3.25 parental clones. Owing to the excess of wild-type, which was added for backcrossing, the real number of parental clones is likely to be higher than the minimal value listed. (c) Schematic representation of a clone obtained from a NEXt DNA shuffling experiment with four equally mixed parental clones of CAT\_Cd26 with up to 49 mutations between bases 9 and 575. The composition assuming a minimal number of parental clones is shown by boxes shaded according to originating parent clone. The length of the fragment is given in the box. Cross-over positions were calculated as midpoints between two parent defining mutations. In this experiment, four clones were sequenced and an overall mean fragment length of 86 bases was detected. The clone shown displays a mean fragment length of 57 bases.

particularly important when shuffling of longer DNAs is envisioned as this will avoid dilution of the gene pool with dysfunctional or undesired molecules. Using a proofreading polymerase for the amplification of the gene assembly could further lower the error rate. In another experiment, four parental truncated CAT genes (CAT\_Cd26) containing a total of 49 mutations spread from bases at position 9–575 were shuffled and five clones sequenced. A detectable mean fragment

length of 86 bases was found, including fragments down to only 8 bases (Figure 2c). The mean fragment length in this experiment is smaller than in the previous one, as more mutations result in a better detection of the fragment length. The short fragments can be explained by the possibility that the QiaexII kit used purified significant amounts of short fragments or, more likely, by efficient priming with frequent strand switching for each PCR cycle. In general,

gene assembly is a complex process mainly but not only determined by fragment length.

A complete directed evolution series based on error-prone PCR and NExT DNA shuffling was applied to improve the enzymatic activity of truncated CAT\_Nd10, which grew on plates up to only 25 µg/ml chloramphenicol in the presence of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG), and CAT\_Cd9, which failed to grow at all. After optimization, several clones of both libraries grew even at 400 µg/ml chloramphenicol/IPTG (higher concentrations were not tested), demonstrating the efficacy of this technique. In addition, the preferred method was applied to TEM-1 β-lactamase using a dUTP fraction of 30%. The fragmentation and assembly worked in the first experiment only ensuring sufficient material to start with but without any prior or intermediate tests or analysis steps (data not shown).

### A computer program to predict NExT fragmentation

Since the dUTP incorporation and the resulting fragmentation are based on deducible principles, a computer program was developed and named NExTProg (20) (<http://www.molbiotech.uni-freiburg.de/next>, <http://www.ATG-biosynthetics.com>). This permits the prediction of the NExT fragmentation pattern of double-stranded DNA allowing the researcher to tailor the dUTP:dTTP ratio without the need for experiments. The program was designed to read a DNA sequence file and dUTP:dTTP values, and calculate all possible fragments, their likelihood of occurrence and relative distribution. The complementary strand for a given DNA is automatically generated and taken into account. The program displays the result in a bar chart and allows the export of all calculated data as tabulated lists for further use (Figure 3a). When upper and lower ranges for the fragment size are set (e.g. due to gel purification), the program calculates the potential loss of material and adjusts the relative likelihood of the individual fragments.

The underlying mathematics is as follows: let the probability that a thymidine in a given DNA sequence is replaced by uridine be ' $p$ '. A fragment between two given thymidines is generated if both are replaced by uridine, which equals the probability  $p \times p$ . However, this fragments can only exist if all  $n$ -number thymidines between the two uridines are not replaced resulting in the overall probability of  $p \times p \times (1 - p)^n$ . For fragments including one or both ends of the DNA, one or both  $p$  of the  $p \times p$  probability are set to 1. Note that the sum for all possible patterns of the uridine incorporation in the gene is 1, but the sum of the probabilities for the fragments is larger as each pattern results in several fragments. Thus, for comparing fragmentation results, we normalized the fragment probabilities by dividing through the sum of all values. Our calculation approach is distinguished from previous calculations (21), as we take into account the fact that both ends of a fragment need to be generated and we do not face the problems of partial sequence preferences and uncertain digest conditions, which severely hamper the calculation of DNase digests.

For a gene with  $x$  number of thymidines, there are  $\sum_{i=1}^{x+1} i$  possible fragments [which is according to Gauß  $n \times (n + 1)/2$  for even  $n$  values with  $n = x + 1$ ]. Thus, for a typical 1000 bp gene with, for example, 251 thymidines and 241 adenines

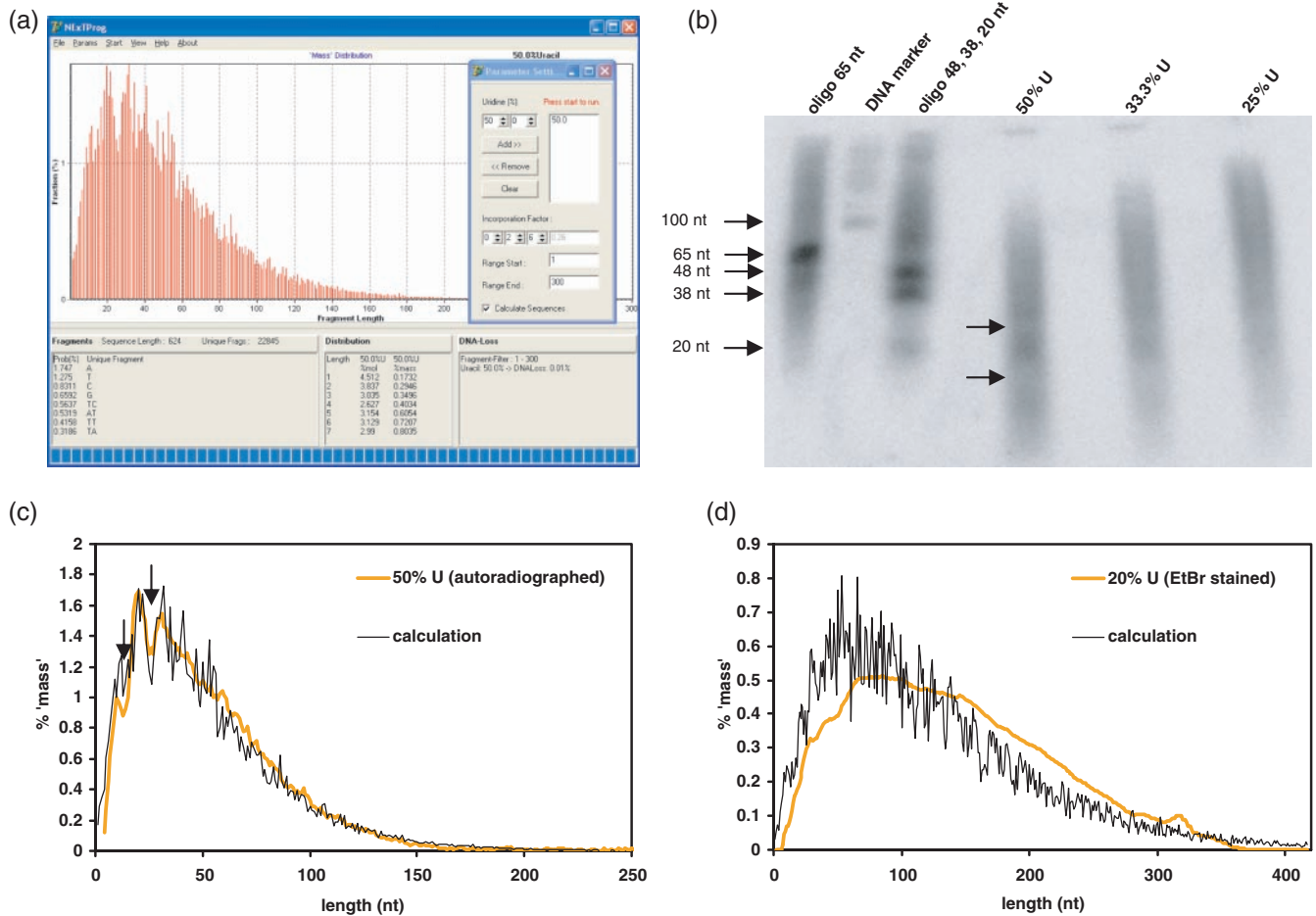
that are calculated as thymidines in the opposite strand, the program calculates and generates in a few seconds  $31\,878 + 29\,403 = 61\,281$  fragments with their probabilities and individual sequences. This formula gives the theoretical maximal number as also fragments of zero length (i.e. two neighboring Ts) are calculated, which are not taken into account by the software. As most users are probably interested in an overview, the program pools all fragments of the same length, sums up their probabilities and gives the distribution as percentage of the sum of all probabilities versus the length (shown as %mol). To reflect the visualization on electrophoreses gels, the 'mass' distribution is calculated by multiplying the probability of a certain fragment length with its length (shown as %mass). These values are normalized and represent the percent bases, which we termed 'mass' as defined by length in base pairs. For the fragment sequence, output fragments of identical sequence are listed only once with their summed probabilities, and all fragments are sorted by decreasing likelihood of occurrence.

### Determination of the incorporation rate of dUTP versus dTTP

Before one can compare the measured and calculated values, there is one important point to consider: the incorporation rate of uridine versus thymidine by the polymerase in the exchange PCR. This value might not only depend on the ratio of dUTP:dTTP but also on the type of polymerase used, the absolute concentration of nucleotides and the buffer. Thus, this value needs to be set when using the program.

To provide a default value for the uridine incorporation rate, a quantitative analysis of a fragmentation experiment with widely used conditions [standard *Taq* polymerase (22), 0.2 mM dNTPs] were performed. DNA was radioactively labeled with [ $^{32}$ P]dCTP in the uridine incorporating PCR. Denaturing polyacrylamide gels were run as described and autoradiographed, thus avoiding signal distortions due to inefficient staining of short fragments (Figure 3b). The calculation of '%mass' already implies that longer oligonucleotides contain more radioactive [ $^{32}$ P]dCTP per molecule. The relative migration for the fragment smear and the radioactively labeled markers were taken from line density plots. The relative distance of the markers were fitted with the equation  $relative\ distance = a \times \ln(\text{length in bp}) + b$  with the variables  $a$  and  $b$ , and the relative distance of the fragment smear converted to nucleotide length using the reverse equation. This resulted in an intensity signal versus a continuous length distribution, which was made discrete to integer numbers by combining rounded nucleotide length values and by averaging the respective intensity values. This distribution was then normalized by dividing each intensity value through the sum of all values and is shown in percent in Figure 3c.

This experiment was compared with program calculations to determine the relative uridine versus thymidine incorporation value. For the calculations, incorporation values from 0.2. to 0.3 in 0.01 increments were used. Setting a value of 0.26 resulted in the best agreement of measurement and prediction based on a mean root square analysis of the fragment size range between 10 and 150 bases where the length calculation is probably reliable and also between 4 and 200 bases (Figure 3c). In case of using the range between



**Figure 3.** Comparison of calculated NEXt fragmentations with radioactively labeled and ethidium bromide stained fragmentation experiments. (a) Graphical front end of NEXtProg 1.0 (20). This program reads DNA sequences and calculates all possible fragments. The program exports lists of fragment length versus normalized fraction of molecules or mass as defined by number of nucleotides, respectively. The sequences of all fragments can be generated, whereby identical sequences are combined, and exported for subsequent assembly calculations. (b) Denaturing PAGE of radioactively labeled DNA samples. From left to right: lanes 1–3, marker oligonucleotides kinased with [ $^{32}$ P]ATP; lanes 4–6, fragments of a gene (CAT\_Cd26, 624 Bp) based on the indicated amount of uridine and [ $^{32}$ P]CTP in the exchange PCR (each lane contained 0.3  $\mu$ Ci). The gel was autoradiographed with a phosphor screen (Kodak) and read with a phosphorimager (Biorad Fx). Note the inhomogeneities in the lower third of the lanes which correspond to sequence-specific peaks in the fragmentation. (c) Measured and calculated fragment distributions used to determine the incorporation rate of uridine versus thymidine in the exchange PCR. The orange line represents a line density plot of the radioactive 50% U lane in panel b, which was converted from relative migration distance to nucleotide length based on the marker nucleotides, set to integer numbers by averaging the respective values and normalized. The black line represents the calculation of NEXtProg for the fragment 'mass' distribution for the same gene with 50% uridine and an incorporation rate of 0.26, which provided the lowest root mean square deviation. (d) The orange line is the line density plot of a 20% U reaction (Figures 1b and 2a) stained with ethidium bromide, which was converted to fragment length and normalized. The black line is the calculation of NEXtProg with 20% uridine and an incorporation rate of 0.26. Note that the staining of short single-stranded oligonucleotides with ethidium bromide is inefficient and consequently longer fragments are overrepresented in the normalized plot.

the markers (20 and 100 bases), a factor of 0.25 scored marginally better. It is remarkable how well these plots overlay and peaks and dips within the fragment smear (indicated by arrows in Figure 3b and c) can be explained. Importantly, the scaling of the y-axis falls into place without further adjustment. We are very confident that the uridine incorporation rate has been determined and not a factor accounting for incomplete fragmentation, because of the near-perfect agreement of calculation and experiment under many experimental conditions tested for fragmentation. Measurements of the radioactive 33.3% uridine fragmentation and calculations agreed equally well. The radioactive 25% dUTP fragmentation produced significant amounts of fragments above the 100 nt marker, and thus showed some deviations due to scaling problems, since

fragment size determination was less accurate in the upper part of the gel. However, using the same incorporation rate value of 0.26 and comparing the results with the ethidium-bromide-stained fragmentation results, which were more accurate for longer fragments due to the available markers, demonstrated a good agreement (Figure 3d) considering the length dependency of the gel staining. Note that the different buffers used for the PCR had no significant effect.

#### Variation of NEXt shuffling—endonuclease V versus piperidine

Our preferred form of the NEXt DNA shuffling method emerged from several tests and experiments. In the following section, variations of the method are described.

As an alternative to piperidine, other endonucleases such as endonuclease IV (23), exonuclease III (24) or T4-endonuclease V (25) could be used to cleave the DNA backbone (17). We tested the latter but discontinued its use even before data were published by Miyazaki (26). Even in combination with UDG, fragmentation was not as efficient as with piperidine. This is demonstrated by, for example, the cleavage of the 50% U PCR product by piperidine in Figure 1b, which has its center of mass within the 38–58 nt marker range, in comparison with the analogous endonuclease V cleavage in Figure 1d, which has its center of mass above the 68 nt marker. Even more problematic was the high error rate inherent to this procedure. Applying an UDG and T4 endonuclease V fragmentation with gel extraction, and a reassembly to a CAT wild-type gene and sequencing six clones with a total of 3930 bases gave a mutation rate of 1.75%. We propose two reasons for this finding. First, the DNA backbone is cleaved by the T4 endonuclease V with its lyase activity, which catalyzes a  $\beta$ -elimination reaction leaving a 3' unsaturated aldehyde (4-hydroxy-2-pentenal) attached at the phosphate group (17,27). The further chemical reaction leading to the free phosphate group is unlikely to be complete and such generated fragments are an unfavorable starting point for a polymerase. Second, the size distribution of fragments hints an incomplete backbone cleavage at sites with a cleaved uracil, so that gapped templates could lead to erroneous nucleotide incorporation. Further experiments could solve these problems, but were not carried out as the piperidine cleavage worked well and was much more cost-effective.

#### Variation of NExT shuffling—*Taq* versus Vent

One important variation of the NExT DNA shuffling tested was the use of the proofreading polymerase Vent instead of *Taq* for the fragment reassembly procedure. *Taq* has been the enzyme of choice for most published fragment reassembly procedures. The use of a proofreading polymerase has been reported (28) for the main purpose to reduce the error rate. However, as shown below the polymerase was not the main factor determining the error rate.

The 3'→5' proofreading exonuclease activity of Vent DNA polymerase relies on removal of mismatched nucleotides only from the 3' terminus of the priming strand, until polymerization can be initiated from an annealed end. Thus, point mutations in nucleotides even in close proximity to the 3' end will pose no problem (29). For comparison, two reassembly procedures were run in parallel with the same fragment pool using either *Taq* or Vent, respectively. Subsequent amplification PCR reactions with *Taq* resulted in one band each in an agarose gel. Quantification by image analysis revealed a 35 times higher yield for the procedure with Vent. Interestingly, sequencing of 6988 bases of the *Taq*-based procedure revealed only five additional mutations (0.075%). Thus, within the statistical limitations based on available sequences, *Taq* and Vent provided the same low error rate for DNA reassembly. The low error rate for *Taq* is in agreement with other reports (30). The significantly improved yield can be explained by several traits of proofreading polymerases such as Vent. The strand displacement activity might help in the presence of many hybridization reactions, and a half-life of ~8 h at 95°C compared with 1.6 h half-life of *Taq* DNA polymerase ensures

fitness during the long reassembly reaction (31). Another difficulty is that *Taq* adds additional dATPs to the 3' hydroxyl terminus (32), a possible hindrance in the reassembly reaction. Hence, the benefits of *Taq* are limited to templates difficult to amplify with proofreading polymerases.

#### Variation of NExT shuffling—gel purification versus quick clean-up

As a further variation of the fragmentation and assembly procedure, it was tested to what extent a high resolution urea gel purification of the fragments compares to the quick clean-up with a one-step silica matrix. Gel purification to ensure that no undesired long fragments get involved in the reassembly reaction could aid the cross-over frequency. For the fragment extraction from urea gels either water or a diffusion buffer were used. In both cases, fragments remained in the gel as seen by staining of previously extracted gel material. To elucidate the cross-over rate with gel extraction, a shuffling experiment with three parental clones (CAT\_Nd10 mutants) was set up. Two of the clones contained one and one clone contained two distinct mutations within a stretch of 100 bp. Sequencing of eight shuffled clones revealed that six clones had one cross-over within the 100 bp stretch. A total of 3851 bases were sequenced and 12 errors were found equaling a mutation rate of 0.31%. Thus, the cross-over rate was in the range of the shuffling procedure using the quick clean-up. The error rate is significantly higher which might be explained by UV damage due to visualization even on the weak 366 nm light source used and/or chemical modifications due to the gel. For the preferred NExT DNA shuffling method, we omitted the gel purification step, because of the additional work without significant benefit, loss of material and the higher error rate.

#### Variation of NExT shuffling—dUTP and alternative analogs

This study focused solely upon uracil as exchange nucleotide, however the technique could equally be applied to the incorporation of several other analogs. One such example, 8-oxo-guanine, can be cleaved out by 8-oxoguanine DNA glycosylase (formamidopyrimidine-DNA glycosylase, Fpg) (33). This base could prove useful in AT-rich regions where DNA cleavage by UDG is too frequent or in GC-rich genes where thymidines are seldom found. Alternatively, a combination of several exchange nucleotides together could generate fragments of the desired range. In our case, the use of additional analogs was not necessary. Furthermore, nucleotides other than dUTP are significantly more costly. Nonetheless, our software is easily adapted to other nucleotides.

If required, incorporation of exchange nucleotides into the primers of the incorporating PCR should ensure that these regions of the gene library could also be shuffled in both strands.

NExT is based on rational and predefined dUTP:dTTP ratios and is well suited for the shuffling of short genes and large gene assemblies. Owing to the robustness and simplicity of NExT shuffling, even those with little prior experience in this area should be able to apply this technique.



## ACKNOWLEDGEMENTS

We thank Jody Mason for critically reading the manuscript. Funding to pay the Open Access publication charges for this article was provided by Institut für Biologie III, Universität Freiburg. H.S.B. was supported by AiF/ProInno, grant KU 0326701 KAJ1.

*Conflict of interest statement.* None declared.

## REFERENCES

- Sung, W.L., Zahab, D.M., Yao, F.L., Wu, R. and Narang, S.A. (1986) Simultaneous synthesis of human-, mouse- and chimeric epidermal growth factor genes via 'hybrid gene synthesis' approach. *Nucleic Acids Res.*, **14**, 6159–6168.
- Meyerhans, A., Vartanian, J.P. and Wain-Hobson, S. (1990) DNA recombination during PCR. *Nucleic Acids Res.*, **18**, 1687–1691.
- Prodromou, C. and Pearl, L.H. (1992) Recursive PCR: a novel technique for total gene synthesis. *Protein Eng.*, **5**, 827–829.
- Stemmer, W.P. (1994) Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature*, **370**, 389–391.
- Neylon, C. (2004) Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. *Nucleic Acids Res.*, **32**, 1448–1459.
- Stemmer, W.P. (1994) DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. *Proc. Natl Acad. Sci. USA*, **91**, 10747–10751.
- Zhao, H., Giver, L., Shao, Z., Affholter, J.A. and Arnold, F.H. (1998) Molecular evolution by staggered extension process (StEP) *in vitro* recombination. *Nat. Biotechnol.*, **16**, 258–261.
- Shao, Z., Zhao, H., Giver, L. and Arnold, F.H. (1998) Random-priming *in vitro* recombination: an effective tool for directed evolution. *Nucleic Acids Res.*, **26**, 681–683.
- Coco, W.M., Levinson, W.E., Crist, M.J., Hektor, H.J., Darzins, A., Pienkos, P.T., Squires, C.H. and Monticello, D.J. (2001) DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nat. Biotechnol.*, **19**, 354–359.
- Lutz, S., Ostermeier, M. and Benkovic, S.J. (2001) Rapid generation of incremental truncation libraries for protein engineering using alpha-phosphothioate nucleotides. *Nucleic Acids Res.*, **29**, E16.
- Sieber, V., Martinez, C.A. and Arnold, F.H. (2001) Libraries of hybrid proteins from distantly related sequences. *Nat. Biotechnol.*, **19**, 456–460.
- Arndt, K.M., Müller, K.M. and Plückthun, A. (2001) Helix-stabilized Fv (hsFv) antibody fragments: substituting the constant domains of a Fab fragment for a heterodimeric coiled-coil domain. *J. Mol. Biol.*, **312**, 221–228.
- Sambrook, J. and Russel, D.W. (2001) *Molecular Cloning: A Laboratory Manual*. 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
- Hecky, J. and Müller, K.M. (2005) Structural perturbation and compensation by directed evolution at physiological temperature leads to thermostabilization of beta-lactamase. *Biochemistry*, in press.
- Cadwell, R.C. and Joyce, G.F. (1992) Randomization of genes by PCR mutagenesis. *PCR Methods Appl.*, **2**, 28–33.
- Patel, P.H. and Loeb, L.A. (2000) Multiple amino acid substitutions allow DNA polymerases to synthesize RNA. *J. Biol. Chem.*, **275**, 40266–40272.
- Friedberg, E.C., Walker, G.C. and Siede, W. (1995) *DNA Repair and Mutagenesis*. American Society of Microbiology, Washington.
- Drohat, A.C., Jagadeesh, J., Ferguson, E. and Stivers, J.T. (1999) Role of electrophilic and general base catalysis in the mechanism of *Escherichia coli* uracil DNA glycosylase. *Biochemistry*, **38**, 11866–11875.
- Maxam, A.M. and Gilbert, W. (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.*, **65**, 499–560.
- Müller, K.M. and Zipf, G. (2004) NExTProg 1.0; download available at <http://www.molbiotech.uni-freiburg.de/next> or <http://www.ATG-biosynthetics.com>.
- Moore, G.L. and Maranas, C.D. (2000) Modeling DNA mutation and recombination for directed evolution experiments. *J. Theor. Biol.*, **205**, 483–503.
- Kaledin, A.S., Sliusarenko, A.G. and Gorodetskii, S.I. (1980) Isolation and properties of DNA polymerase from extreme thermophilic bacteria *Thermus aquaticus* YT-1. *Biokhimiia*, **45**, 644–651.
- Ljungquist, S. (1977) A new endonuclease from *Escherichia coli* acting at apurinic sites in DNA. *J. Biol. Chem.*, **252**, 2808–2814.
- Richardson, C.C. and Kornberg, A. (1964) A deoxyribonucleic acid phosphatase-exonuclease from *Escherichia coli*. I. Purification of the enzyme and characterization of the phosphatase activity. *J. Biol. Chem.*, **239**, 242–250.
- Dodson, M.L., Schrock, R.D., III and Lloyd, R.S. (1993) Evidence for an imino intermediate in the T4 endonuclease V reaction. *Biochemistry*, **32**, 8284–8290.
- Miyazaki, K. (2002) Random DNA fragmentation with endonuclease V: application to DNA shuffling. *Nucleic Acids Res.*, **30**, e139.
- Levin, J.D. and Demple, B. (1990) Analysis of class II (hydrolytic) and class I (beta-lyase) apurinic/apyrimidinic endonucleases with a synthetic DNA substrate. *Nucleic Acids Res.*, **18**, 5069–5075.
- Zhao, H. and Arnold, F.H. (1997) Optimization of DNA shuffling for high fidelity recombination. *Nucleic Acids Res.*, **25**, 1307–1308.
- Mattila, P., Korpela, J., Tenkanen, T. and Pitkanen, K. (1991) Fidelity of DNA synthesis by the *Thermococcus litoralis* DNA polymerase—an extremely heat stable enzyme with proofreading activity. *Nucleic Acids Res.*, **19**, 4967–4973.
- Lorimer, I.A. and Pastan, I. (1995) Random recombination of antibody single chain Fv sequences after fragmentation with DNaseI in the presence of Mn<sup>2+</sup>. *Nucleic Acids Res.*, **23**, 3067–3068.
- Kong, H., Kucera, R.B. and Jack, W.E. (1993) Characterization of a DNA polymerase from the hyperthermophile archaea *Thermococcus litoralis*. Vent DNA polymerase, steady state kinetics, thermal stability, processivity, strand displacement, and exonuclease activities. *J. Biol. Chem.*, **268**, 1965–1975.
- Clark, J.M. (1988) Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res.*, **16**, 9677–9686.
- Boiteux, S., O'Connor, T.R. and Laval, J. (1987) Formamidopyrimidine-DNA glycosylase of *Escherichia coli*: cloning and sequencing of the fpg structural gene and overproduction of the protein. *EMBO J.*, **6**, 3177–3183.