Protein Engineering, Design & Selection vol. 25 no. 10 pp. 657–668, 2012
Published online June 12, 2012 doi:10.1093/protein/gzs034

Generation and characterization of new highly thermostable and processive M-MuLV reverse transcriptase variants

Aurimas Baranauskas[1†], Sigitas Paliksa[1†], Gediminas Alzbutas[1], Mindaugas Vaitkevicius[1], Judita Lubiene[1], Virginija Letukiene[1], Sigitas Burinskas[1], Giedrius Sasnauskas[2] and Remigijus Skirgaila[1,3]

[1]Thermo Fisher Scientific, Graiciuno 8, LT-02241 Vilnius, Lithuania and [2]Institute of Biotechnology, Vilnius University, Graiciuno 8, LT-02241 Vilnius, Lithuania

[3]To whom correspondence should be addressed.
E-mail: remigijus.skirgaila@thermofisher.com

Received January 6, 2012; revised April 20, 2012; accepted May 14, 2012

Edited by Philip Holliger

***In vitro* synthesis of cDNA is one of the most important techniques in present molecular biology. Faithful synthesis of long cDNA on highly structured RNA templates requires thermostable and processive reverse transcriptases. In a recent attempt to increase the thermostability of the wt Moloney Murine leukemia virus reverse transcriptase (M-MuLV RT), we have employed the compartmentalized ribosome display (CRD) evolution *in vitro* technique and identified a large set of previously unknown mutations that enabled cDNA synthesis at elevated temperatures. In this study, we have characterized a group of the M-MuLV RT variants (28 novel amino acid positions, 84 point mutants) carrying the individual mutations. The performance of point mutants (thermal inactivation rate, substrate-binding affinity and processivity) correlated remarkably well with the mutation selection frequency in the CRD experiment. By combining the best-performing mutations D200N, L603W, T330P, L139P and E607K, we have generated highly processive and thermostable multiply-mutated M-MuLV RT variants. The processivity of the best-performing multiple mutant increased to 1500 nt (65-fold improvement in comparison to the wt enzyme), and the maximum temperature of the full-length 7.5-kb cDNA synthesis was raised to 62°C (17° higher in comparison with the wt enzyme).**

Keywords: *Moloney Murine leukemia virus reverse transcriptase/processive reverse transcriptase/thermostable reverse transcriptase*

Introduction

Reverse transcriptases (RTs) are the key enzymes in the life cycle of retroviruses that convert viral single-stranded RNA into the double-stranded DNA. RTs possess three enzymatic activities: the RNA-dependent DNA polymerase, the DNA-dependent DNA polymerase and RNase H, which degrades RNA strand in the RNA–DNA hybrid. The synthesis of cDNA is probably the second most important technique in present molecular biology after the polymerase chain reaction and its modifications. RTs are extensively used to generate cDNA libraries for cloning, end-point and quantitative RT-polymerase chain reaction (PCR), RACE technique, microarray analysis, RNA amplification and other applications (Sambrook *et al.*, 2001).

The major challenges in cDNA synthesis reactions are the secondary RNA structures that may slow down or even halt the RT (Wu *et al.*, 1996; Suo and Johnson, 1998; Lanciault and Champoux, 2006), and the degradation of template RNA by the RNase H activity of wt RTs that leads to truncated cDNA. The RNA secondary structures are destabilized at higher temperatures. Unfortunately, high temperatures may also lead to the RNA degradation due to Mg^{2+} ions in the reaction buffer (Gerard *et al.*, 2002a). Therefore, the optimal temperature for quantitative synthesis of full-length cDNA is 50–55°C (temperatures >55°C can be used on heavily structured RNA targets, given the quantitative synthesis of full-length cDNA is not essential for the experiment).

Another issue for cDNA synthesis at increased temperatures is the lack of an efficient thermostable RT. This limitation can be overcome either by thermal stabilization of mesophilic RTs, or by utilization and improvement of the promiscuous RT activity of thermophilic DNA-dependent DNA polymerases. A nice example of the latter approach is the work of Sauter and Marx, who employed high-throughput screening to identify the M1 mutant of the thermostable Klentaq DNA polymerase with 40-fold increased RT activity (Sauter and Marx, 2006). The corresponding Taq DNA polymerase mutant could perform cDNA synthesis at 63–68°C and subsequent DNA amplification in a single tube; however, the efficiency and sensitivity of cDNA synthesis reaction, as well as possibility to synthesize long cDNA, remain unclear (Kranaster *et al.*, 2010). Another example is the Tth DNA-dependent DNA polymerase, which has an intrinsic RT activity in the presence of Mn^{2+} ions, and is currently used in some commercial kits. However, despite the recent progress in cDNA synthesis using thermostable DNA-dependent DNA polymerases, the majority of applications still rely on the mesophilic enzymes Moloney Murine leukemia virus reverse transcriptase (M-MuLV RT) and avian myeloblastosis virus enzyme (AMV RT). The M-MuLV RT is a monomeric protein containing the N-terminal polymerase domain and the C-terminal RNase H domain (Georgiadis *et al.*, 1995; Das and Georgiadis, 2004; Lim *et al.*, 2006; Coté and Roth, 2008). The wild-type M-MuLV RT typically is used at 37–42°C and is rapidly inactivated at higher temperatures. The heterodimeric AMV RT has a higher temperature optimum (45–50°C, usable up to 60°C), but is less popular than

[†]Both authors contributed to work equally.

M-MuLV RT due to the lower cDNA yield and higher RNase H activity (Ståhlberg *et al.*, 2004).

The limited temperature range of the wt M-MuLV and AMV RTs has prompted numerous attempts to improve their thermostability and the efficiency of cDNA synthesis. First, the RNase H active site of both M-MuLV and AMV RTs was switched off by site-specific mutations (Kotewicz *et al.*, 1988; Gerard *et al.*, 2002b). The enzyme variants without the RNase H activity (H^-) perform better in the full-length cDNA synthesis, are compatible with the RACE technique, and are also more thermostable in the presence of a template-primer (Gerard *et al.*, 1997, 2002b; Pinto and Lindblad, 2010). The latter finding was attributed to tighter binding of RTs to primer-template substrates that are unaltered by the RNase H activity (Gerard *et al.*, 2002b).

Owing to the simple monomeric structure, M-MuLV RT also became an ideal candidate for further improvements by rational design (Kotewicz *et al.*, 1988; Gerard *et al.*, 2002b; Yasukawa *et al.*, 2010) and high-throughput screening (Potter *et al.*, 2002; Arezi and Hogrefe, 2009). As a result, many M-MuLV RT mutants with improved thermostability, fidelity, substrate affinity and/or reduced terminal deoxynucleotidyl-transferase activity are currently available as commercial enzymes (Superscript II, Superscript III, AffinityScript, RevertAid Premium, Maxima RT). Unfortunately, most of this research is referenced in patents and patent applications only (Chen *et al.*, 2004; Hogrefe *et al.*, 2006). Just a few mutants disclosed in the patents were further biochemically characterized and described in the scientific literature (Gerard *et al.*, 2002b; Arezi and Hogrefe, 2009). In a recent study Arezi and Hogrefe (2009) employed a high-throughput screening approach to identify five mutations (E69K, E302R, W313F, L435G, N454K) that collectively increase the half-life of M-MuLV RT at 55°C from <5 min to ~30 min in the presence of the template-primer. The pentuple M-MuLV RT mutant M5 was able to synthesize the full-length 9-kb cDNA over a broad range of temperatures (37–55°C) and was more tolerant to the common quantitative RT-PCR inhibitors (Arezi *et al.*, 2010). The increase in the thermostability was attributed primarily to the tighter substrate binding (Arezi and Hogrefe, 2009). Noteworthy, both the RNase H^+ and the RNase H^- variants of the M5 mutant displayed a comparable thermostability and ability to synthesize full-length cDNA, arguing that the RNase H^- phenotype is not obligatory for the efficient cDNA synthesis and heat resistance of RTs. Moreover, the preserved RNase H activity may be advantageous on the GC-rich RNA templates, as degradation of the RNA template in the RNA–DNA hybrid by the RNase H domain may facilitate synthesis of the second DNA strand; this would eliminate the need for the optional RNase H treatment step that is recommended by many suppliers of the H^- M-MuLV RTs.

In an attempt to increase the thermostability of the wt M-MuLV RT, we have developed the compartmentalized ribosome display (CRD) evolution *in vitro* technique (Janulaitis *et al.*, 2009; Skirgaila *et al.*, manuscript in preparation). The CRD technique is essentially the combination of two previously developed selection methods, ribosome display (RD) (Zahnd *et al.*, 2007), and *in vitro* compartmentalization (IVC) (Miller *et al.*, 2006). Initially, the ternary mRNA-ribosome–protein complex is generated by *in vitro* translation of mRNA encoding the RT. As in the RD method, this establishes a non-covalent linkage between the enzyme (RT) and the genetic material (mRNA). The ternary complexes placed in the reverse transcription reaction mixture are then emulsified yielding compartments containing no more than one ternary complex per compartment. Upon the increase in reaction temperature, the ternary complexes dissociate, and the liberated RT starts cDNA synthesis from its own mRNA. Only mRNA, which encodes active RT variants capable of full-length cDNA synthesis at the desired temperature, is amplified and transferred to the next selection round. Experimental details of the CRD technique will be described elsewhere (Skirgaila *et al.*, manuscript in preparation). By employing the CRD method, we have identified a number of previously unknown M-MuLV RT mutations that enable cDNA synthesis at elevated temperatures (Janulaitis *et al.*, 2009). In this study, we report the thermal inactivation rate, substrate-binding affinity and the processivity of the M-MuLV RT mutants carrying the individual substitutions. We show that the performance of individual mutations correlates remarkably well with their selection frequency in the CRD experiment. By combining the best-performing mutations, we have constructed highly processive multiply-mutated M-MuLV RT mutants capable of the full-length 7.5-kb cDNA synthesis at the temperatures up to 62°C.

Materials and methods

Enzymes and oligonucleotides

The T4 polynucleotide kinase (PNK), proteinase K and RiboLock RNase inhibitor were supplied by Fermentas. The radiolabeled nucleotides were from PerkinElmer. All oligonucleotides were purchased from Metabion. The protein expression, purification, polymerase activity measurements, determination of the DNA synthesis rate and the modeling of the full-length enzyme–substrate complex are described in the 'Supplementary data' section.

Thermal inactivation rates

The stability of RT mutants was determined by measuring the residual activity of the enzymes at 37°C after the preincubation at 50°C in the presence of the template–primer substrate. The reaction mixture containing RT (1.85 ng/ 10 μl or ~2.4 nM) and poly(rA)/oligo$(dT)_{18}$ substrate (6 μM) in the RT reaction buffer [50 mM Tris–HCl (pH 8.3 at 25°C), 50 mM KCl, 4 mM $MgCl_2$, 10 mM DTT] was incubated at 50°C. The aliquots of 10 μl were taken after 0, 8, 15, 30, 60, 120, 240 and 300 min and mixed with an equal volume of additional reaction mixture (RT buffer supplemented with 6 μM poly rA/oligo dT_{18}, 0.5 mM dTTP and 0.365 μM 0.555 TBq/mmol [^{3}H]-dTTP) preincubated at 37°C. The reaction was carried out for 20 min at 37°C and stopped by adding 5 μl of ice-cold 100 mM ethylenediaminetetraacetic acid (EDTA). The activity units were determined by measuring the [^{3}H]-dTMP incorporation into a polynucleotide fraction adsorbed on DE–81 filter paper (Whatman). The dependencies of the residual activity on the preincubation time were plotted using Prism 4 program (GraphPad software Inc.). The thermal inactivation rates k_{inact} were determined by fitting single exponentials to the experimental data. The thermal inactivation half-life values $t_{1/2}$ were calculated from the ratio $\ln(2)/k_{inact}$.

RNA/DNA substrate-binding studies

The RNA–DNA hybrid substrate was prepared by annealing the RNA oligonucleotide LA-237 (sequence 5′-CGGGAU ACCGUCCAGCGACAUUCUUCCUCGGUACAUAAUCUC CUUUGG-3′) with a 1.2-fold excess of the complementary DNA oligonucleotide JV-08 (5′-CCAAAGGAGATTATG TACCGAGGAAGAATGTCGCTG-3′) in 50 mM Tris–HCl (pH 8.3), 10 mM EDTA and 20 mM KCl. The RNA strand prior to annealing was 5′-labeled using T4 PNK and [γ-^{32}P]-ATP.

The RTs were diluted with the dilution buffer [10 mM Tris–HCl (pH 8.3), 20 mM NaCl, 1 mM DTT, 0.2% Triton X-100], mixed with 2 pM of the radiolabeled primer–template substrate in 20 μl of the binding buffer [50 mM Tris–HCl (pH 8.3), 50 mM KCl, 10 mM DTT and 10% glycerol], and incubated for 30 min at room temperature. Free and protein-bound RNA–DNA heteroduplexes were separated on the 8% non-denaturing polyacrylamide gels using 40 mM Tris–acetate (pH 8.3) supplemented with 1 mM EDTA as the running buffer. The electrophoresis was performed for 2 h at 10 V/cm at room temperature. Radiolabeled bands were detected using Typhoon Trio imager (GE Healthcare) and quantified using TotalLab software. The equilibrium dissociation constants (K_D) were determined by fitting equation (1) to the experimental data:

$$[E \cdot S] = \{K_D + [E]_0 + [S]_0 - [(K_D + [E]_0 + [S]_0)^2 - 4 \cdot [E]_0 \cdot [S]_0]^{0.5}\}/2 \quad (1)$$

where $[E]_0$ and $[S]_0$ are total enzyme and template–primer concentrations, respectively, and [E·S] is the concentration of the enzyme–substrate complex.

Synthesis of the 1.1- and 7.5-kb cDNA

The RNA/DNA substrate was prepared by mixing 1 μg of 1.1 kb (or 7.5 kb) target RNA (sequences are provided in Supplementary Table S1), 1 μl of 100 μM oligo(dT)$_{18}$ (or 0.1 μl of 100 μM Eco31Irev 5′-CACGCTCACCGGCTC CAGAT-3′ oligonucleotide in the case of the 7.5-kb cDNA synthesis), 2 μl of 10 mM solutions of each dNTPs and nuclease-free water (final volume 13.5 μl). The mixture was heated at 65°C for 5 min and placed in a cold rack (+4°C). Then the reaction mixture was supplemented with RT buffer [final composition: 50 mM Tris–HCl (pH 8.3), 50 mM KCl, 10 mM DTT, 4 mM $MgCl_2$] and RiboLock RNase Inhibitor to 1 U/μl. RTs were diluted in the dilution buffer [10 mM Tris–HCl (pH 8.3), 20 mM NaCl, 1 mM DTT, 0.2% Triton X-100] to 100 U/μl concentration. 2 μl (200 U) of each enzyme were added to 20 μl of the reaction mixture. Reactions were performed in the Mastercycler gradient PCR machine (Eppendorf) under a wide range of temperatures (37–69°C) for 1 h. After the reactions were completed, 4 μl of 6× alkaline electrophoresis loading buffer (180 mM NaOH, 6 mM EDTA, 18% Ficoll 400, 0.05% bromcresol green) were added, and the reaction products were incubated at 80°C for 10 min, chilled on ice and loaded onto 1% alkaline agarose gel. The gel was prepared in 30 mM NaCl, 2 mM EDTA (pH 7.5) buffer and then equilibrated in electrophoresis buffer (30 mM NaOH, 2 mM EDTA) overnight at 4°C. Electrophoresis was performed at 3 V/cm for 1 h, the gel was neutralized for 30 min in 300 ml of 0.5 M Tris–HCl buffer (pH 7.5), stained in the 0.5 μg/ml ethidium bromide solution for 30 min and visualized under UV light.

Processivity measurements

The substrate for processivity measurements was prepared by annealing 100 nM 5′-FAM-labeled primer FAMECO (5′-CACGCTCACCGGCTCCAGATTTATCAGCAATAAAC CAGCCAGCCGGAAGGGCCGA-3′) with the synthetic 7.5-kb RNA (200 nM) in the annealing buffer [50 mM Tris-HCl (pH 8.3), 40 mM KCl and 1 mM EDTA]. Prior to the reactions the enzyme (10 nM) was preincubated with the substrate (2 nM) in 10 μl of reaction buffer [50 mM Tris–HCl (pH 8.3), 50 mM KCl and 10 mM DTT] supplemented with 5 mM of each dNTP (final concentration 1 mM) for 30 min at room temperature to allow the formation of protein–substrate complexes. The reactions were started by adding 40 μl of the reaction buffer [50 mM Tris–HCl (pH 8.3), 50 mM KCl, 10 mM DTT and 5 mM $MgCl_2$ (final Mg^{2+} concentration 4 mM)] supplemented with a trap [8 mg/ml poly(A) RNA (GE Healthcare) and 0.32 mg/ml oligo(dT)$_{18}$ (Fermentas)], incubated at 37°C for 30, 40, 50 or 60 min and stopped by adding EDTA to a final concentration of 100 mM. The control experiments were performed by adding enzyme to the reaction mixture containing both the substrate and trap and incubating the mixture at 37°C for 60 min. The synthesized cDNA was released from the RNA–cDNA–RT complex by treating the reaction mixture with proteinase K (0.5 mg/ml) in the presence of 3 M urea. After 3 h at 37°C proteinase K was inactivated by heating at 65°C for 30 min, followed by phenol-chloroform extraction. The reaction products were ethanol precipitated and dissolved in 10 μl of nuclease-free water. One microliter of reaction products was mixed with 0.24 μl of GeneScan 500ROX size standard (Applied Biosystems) and 9 μl of formamide. DNA fragments were fractionated using a sequence analyzer (Applied Biosystems) and analyzed with ABI PRISM GeneMapper Software version 3.0. The peaks with a signal level above the background were integrated to obtain the intensity at each position, and product length for each position was determined by applying the amplified fragment length polymorphism analysis pattern. The processivity was calculated using the equation $P = \sum(L_n \cdot I_n)/\sum(I_n)$, where L_n is the length and I_n is the intensity of each analyzed cDNA fragment.

Due to the limited length of DNA fragments in the GeneScan 500ROX size standard and the properties of the sequencing gel, the above procedure is only suitable for the analysis of reaction products up to 500 nucleotides in length. As multiply-mutated RTs synthesize much longer cDNA fragments during a single binding event, their processivity was determined using an alternative method. The DNA primer Eco31Irev was 5′-labeled using T4 PNK and [γ-^{33}P]ATP and annealed to the 7.5 kb synthetic RNA yielding the radiolabeled primer–template substrate. After the reactions (performed as above), the products were analyzed by alkaline agarose gel electrophoresis along with the 5′-^{33}P phosphorylated DNA size markers. The DNA fragments in dried gels were visualized by phosphorimager. The approximate fragment lengths and intensities required for processivity determination were quantified using TotalLab software.

Data analysis

Nonlinear regression was performed using GraphPad Prism 4 software (GraphPad Software Inc.). The results from repeated experiments are presented as the average value ± one standard deviation.

Results

The combination of RD (Mattheakis *et al.*, 1994) and IVC (Tawfik and Griffiths, 1998) techniques was employed for *in vitro* evolution and selection of M-MuLV RT variants capable of cDNA synthesis at elevated reaction temperatures (Janulaitis *et al.*, 2009; Skirgaila *et al.*, manuscript in preparation). During this selection, we have identified a number of previously uncharacterized mutations potentially implicated in the increase of RT thermostability. To elucidate the contribution of individual mutations on the RT properties, we have subjected 28 most frequently selected amino acid positions [D200 (30 clones), L603 (18), T330 (15), L139 (14), Q221 (6), T287 (6), I49 (5), N479 (5), H594 (5), F625 (5), P65 (4), H126 (4), L333Q (4), A502 (4), E607 (4), K658 (4), P130, Q237, Y344, M428, D449, A644,N649, L671 (3 clones each), M39, M66, Q91, W388 (2 clones each)] to mutational and biochemical analysis. This set does not include the frequently mutated, but previously characterized amino acid positions D524 (31), D653 (23), D583 (21), H204 (7) and H638 (18) (Potter *et al.*, 2002; Chen *et al.*, 2004), and the C-terminal amino acids (679–680), which were mutated due to the amplification and cloning procedures used in the evolution *in vitro* setup.

Preliminary screen of RT mutants for increased thermostability

At 28 selected positions of the wt Mu-MLV RT, we first introduced mutations that were most frequently identified in the original evolution *in vitro* experiment (Janulaitis *et al.*, 2009; Skirgaila *et al.*, manuscript in preparation) – D200N (25 of 30 mutations at this position), L603W (17/18), T330P (15/15), L139P (14/14), Q221R (6/6), T287A (6/6), I49V (4/5), N479D (5/5), H594R (4/5), F625S (3/5), P65S (4/4), H1 26R (2/4), L333Q (3/4), A502V (4/4), E607K(2/4), K658R (3/4), P130S (3/3), Q237R (3/3), Y344H (3/3), M428L (3/3), D449G (2/3), A644V (2/3), N649S (3/3), L671P (3/3), M3 9V (1/2), M66L (2/2), Q91R (1/2) and W388R (2/2). Mutant enzymes were constructed, expressed and affinity purified as described in Supplementary Methods.

The thermostability of the RT point mutants was evaluated by measuring the RT activities on a synthetic poly(rA):oligo$(dT)_{18}$ substrate at 37°C and 50°C, and calculating the 50°/37°C activity ratio (expressed in percentage). The best-performing mutants (D200N, L603W, T330P, Q221R, N479D, H594R, E607K, P130S, D449G, M66L and W388R) exhibited a substantially higher 50°/37°C activity ratio (≥70%, Fig. 1) than the wt M-MuLV RT (~50%). The wt RT and all mutant enzymes were also tested in a reverse transcription reaction for the ability to synthesize 1.1-kb cDNA at elevated temperatures. The reactions were performed on a synthetic RNA substrate in a temperature gradient using a PCR instrument, and the highest temperature of the 1.1-kb cDNA synthesis was determined by analyzing the reaction products on an alkaline agarose gel. The maximum temperature determined for the wt RT was 48°C (data not shown). Despite the significantly increased 50°/37°C activity ratio of many point mutants, only four RT variants were able to accomplish full-length 1.1-kb cDNA synthesis at higher temperatures than the wt enzyme: L603W (≤53°C), and D200N, Q221R, E607K (≤50°C) (Fig. 1).

Partial site-saturation mutagenesis of selected amino acid positions

The *in vitro* evolution of the M-MuLV RT was performed on the mutant library generated by error-prone PCR (Janulaitis *et al.*, 2009; Skirgaila *et al.*, manuscript in preparation). As this method has certain bias and restrictions, only a limited number (four to six) of mutations per amino acid position could be analyzed. Thus, the most frequent mutation identified in the original selection at a certain position may be suboptimal in terms of enzyme thermostability. To verify the performance of the most recurrent amino acid substitutions, we performed partial site-saturation mutagenesis at the most

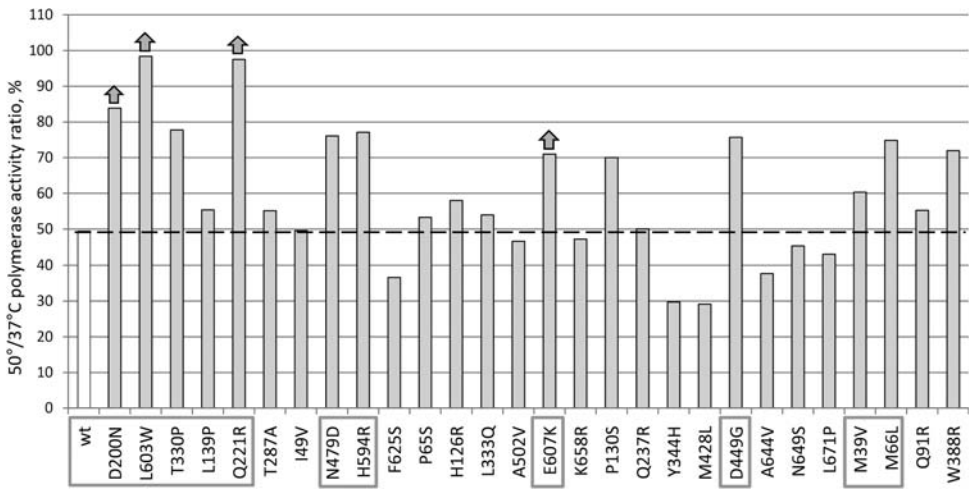

Fig. 1. Activity of the M-MuLV RT point mutants. The 50°/37°C activity ratio (expressed in percentage) of the wt enzyme (white column) and point mutants (gray columns) was determined as described in Supplementary Methods. The average values of at least two independent measurements are given. Gray arrows above the columns indicate RT variants that synthesize 1.1-kb cDNA at ≥50°C. Amino acid positions selected for partial site-saturation mutagenesis are denoted by gray rectangles.

frequently (≥6) selected amino acid positions D200, L603, T330, L139, Q221 and T287. This group was also supplemented with N479, H594, E607, D449, M39 and M66 amino acids, as their mutations resulted in a higher 50°/37°C activity ratio (Fig. 1). Typically, the set of mutations at each amino acid position included all the mutations selected during the original *in vitro* evolution experiment (e.g. D200N, D200A, D200G), similar amino acids (e.g. D200Q, D200E), amino acids with the opposite properties (e.g. negatively/positively charged—D200K, D200R, D200H, polar/non-polar—D200V) and completely different amino acids (D200W, D200P). The best-performing amino acid positions were mutagenized more extensively (D200 – 11 variants; L603 – 8 variants) than the others (M39 – 4 variants, M66 – 3 variants). All the mutants were constructed, expressed, affinity purified and examined for the 50°/37°C activity ratio and for the maximum 1.1-kb cDNA synthesis temperature as described above (Supplementary Fig. S1).

As anticipated, replacements by similar amino acids in most cases yielded similar phenotypes. For example, the L603 mutants L603F and L603Y had a comparable 50°/37°C activity ratio (∼100%) to the L603W mutant (the most frequent variant identified during the selection), and were also able to synthesize 1.1-kb cDNA at 50°C. The L603M mutant, identified in the original selection only once, also had a high 50°/37°C activity ratio (∼80%), but all other replacements (L603P, L603I, L603V and L603G) had a lower 50°/37°C activity ratio (<50%) than the wt enzyme (50%, Supplementary Fig. S1). Thus, only aromatic or large hydrophobic side chains at the 603 position increase the thermostability of the RT. Surprisingly, nearly all substitutions at the D200 position, with the exception of the completely inactive D200P variant, resulted in an increased 50°/37°C activity ratio (Supplementary Fig. S1). Notable examples are the replacement of aspartate residue by a large tryptophan residue (D200W, 50°/37°C activity ratio ∼100%) and the replacement of aspartate by a chemically similar glutamate (D200E, 50°/37°C activity ratio ∼80%). The screen also revealed several other mutants with increased 50°C/37°C activity ratio (≥70%): Q221K, Y; E607A, G; H594K, Q; D449E, N, A, V; M39N; M66V, I. However, only 9 mutations at five positions of 84 single amino acid mutants screened enabled synthesis of 1.1-kb cDNA at >48°C (D200N, D200H, L603W, L603F, L603Y, Q221R, E607K, H594K and H594Q), and only in one case (position H594) the partial site-saturation mutagenesis revealed mutants (H594K and H594Q) that synthesized 1.1-kb cDNA at a higher temperature than the most frequently selected variant H594R (Supplementary Fig. S1).

Characterization of M-MuLV RT point mutants

Thermal inactivation at 50°C. To understand the mechanism behind the beneficial effect of individual mutations, we performed detailed biochemical characterization of the wt enzyme and the best-performing point mutants D200N, L603W, T330P, L139P, Q221R, T287A, N479D, H594K, E607K, D449G, M66L and W388R. All proteins were expressed with an N-terminal His-tag and purified using affinity chromatography and P11 ion-exchange column as described in Supplementary Methods.

Despite the relatively high activity of some mutants at 50°C (Fig. 1), their thermostability in the absence of template-primer was quite low. The residual polymerase activity after 5-min preincubation at 50°C in the absence of template-primer (≤21%) was comparable to the residual activity of the wt enzyme (13%, data not shown). However, as the template-primer is always present in the reverse transcription reactions, a more relevant parameter for the RT thermostability is the half-life ($t_{1/2}$) of the enzyme in the presence of the RNA–DNA substrate. To this end we have measured the thermal inactivation rates of the wt RT and purified point mutants at 50°C in the presence of the template-primer (see Materials and Methods for details). The $t_{1/2}$ values obtained for the mutants D200N (∼180 min), L603W (∼150 min), L139P (∼100 min) and T330P (∼90 min) exceeded the $t_{1/2}$ of the wt enzyme (40 min) up to 4-fold (Fig. 2A). The $t_{1/2}$ values of the other mutants fell within the range of 45–76 min. Intriguingly, a decreased thermal inactivation rate at 50°C of some mutants did not result in a higher 50°/37°C activity ratio. For example, despite the longer $t_{1/2}$ of the L139P mutant, its 50°/37°C activity ratio (∼60%) was only slightly higher in comparison with the wt enzyme (∼50%, Fig. 1). On the contrary, the $t_{1/2}$ of the Q221R mutant (∼50 min) was comparable to the wt enzyme, but the 50°/37°C activity ratio was significantly higher (∼94%, Fig. 1).

Substrate-binding affinity. It was demonstrated that RT mutants may improve thermostability through tighter binding to the template–primer substrate (Gerard *et al.*, 2002b; Arezi and Hogrefe, 2009). Therefore, we have determined the equilibrium dissociation constants (K_D) for the interaction of the wt and mutant RTs with the RNA–DNA substrate using the electrophoretic mobility shift assay (EMSA, see Materials and Methods for details). The values of equilibrium binding constants (K_D) determined for RNase H^- and RNase H^+ RTs vary from 1 nM to 19 nM depending on the experimental assay (EMSA, filter binding, surface plasmon resonance, pre-steady state analysis) and the substrate used for the binding measurements (DNA/DNA, RNA/DNA, RNA/RNA and RNA) (Schultz *et al.*, 1999; Gerard *et al.*, 2002b; Wu *et al.*, 2005; Ndongwe *et al.*, 2012). In our case the determined K_D for wt M-MuLV RT (1.5 nM, Fig. 2B) was in good agreement with the value reported for the RNase H^- M-MuLV RT by Gerard *et al.* (2002b) (4.3 nM, RNA/DNA substrate, filter-binding assay). The substrate-binding affinity of the D200N, L603W, L139P, N479D, H594K, E607K and D449G mutants was slightly (2- to 5-fold) increased in comparison with the wt enzyme (Fig. 2B). This may account for the slower thermal inactivation (Fig. 2A) and higher 50°/37°C activity ratio (Fig. 1) of these RT variants.

Processivity. An increased substrate-binding affinity may also affect the enzyme's processivity, which is an essential parameter for the efficient synthesis of long cDNA molecules. We have tested the processivity of the wt and mutant RTs using an FAM-labeled DNA primer annealed to a long RNA transcript in the presence of a substrate trap. The elongated primer was analyzed on a sequencing machine along with the 35–500 nt ROX-labeled DNA ladder and quantified as described in Materials and Methods. As the secondary structures of the RNA template may cause pausing of the enzyme, the processivity of RTs was highly dependent on the RNA sequence (data not shown). The processivity of the wt enzyme determined on the 7.5-kb RNA template (∼24 nucleotides, Fig. 2C) was in good agreement with the

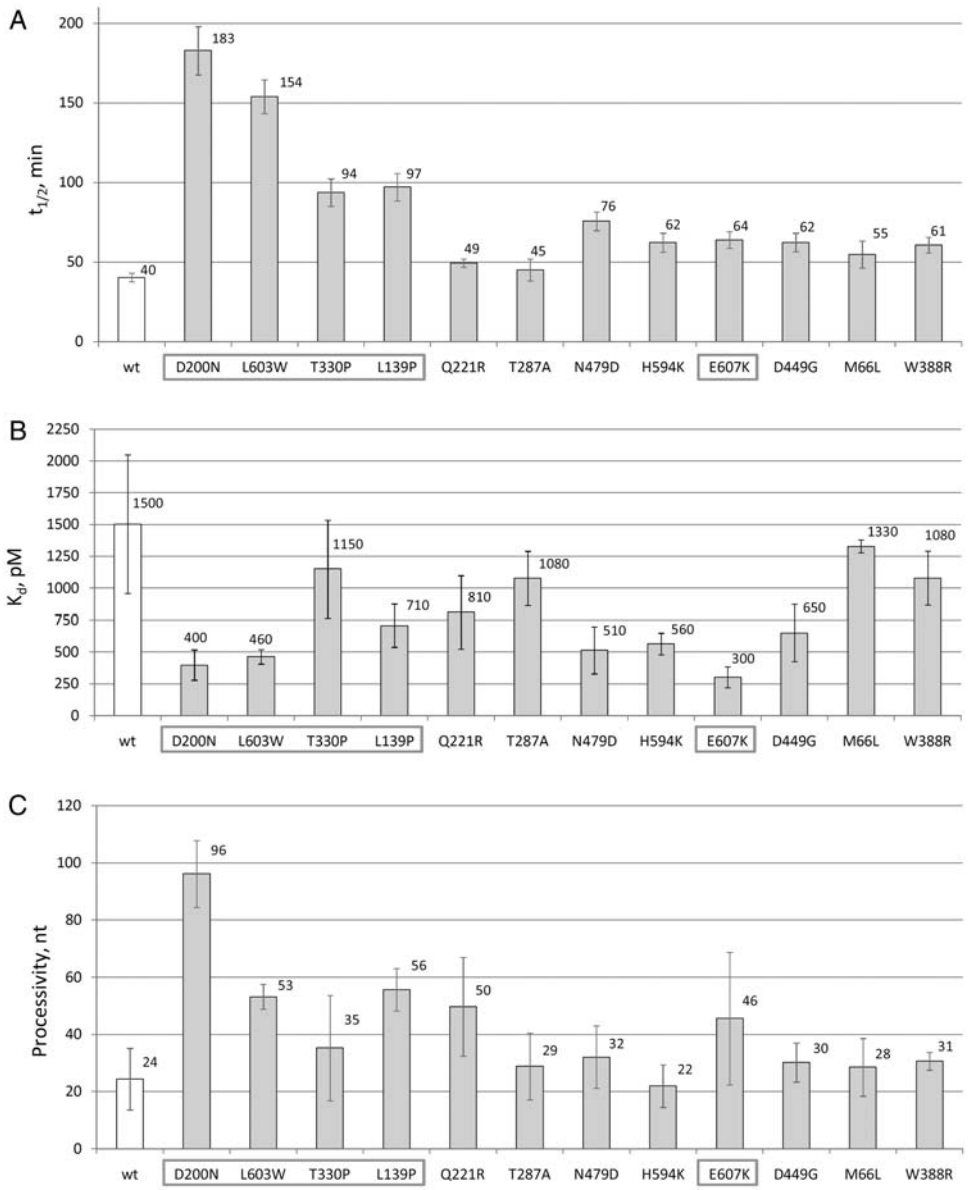

Fig. 2. Biochemical characterization of the M-MuLV RT point mutants. (**A**) Thermal inactivation half-life ($t_{1/2}$, minutes) of point mutants at 50°C. The inactivation rates were measured in the presence of the template–primer substrate as described in Materials and Methods. (**B**) Substrate-binding affinity. The equilibrium dissociation constants (K_D) of the RT – primer–template complexes were measured using the EMSA. (**C**). Enzyme processivity (in nucleotides) at 37°C. The average values of at least three (A) or two (B and C) independent measurements ± 1 SD are given. Mutations selected for the construction of multiply-mutated RTs are denoted by gray rectangles.

previously reported values of 20–40 nucleotides (Gerard *et al.*, 2002b). The stretches of cDNA synthesized by the E607K, Q221R, L139P, L603W and D200N mutants during a single polymerization event were 2- to 4-fold longer in comparison with the wt enzyme (processivity 46–96 nt, Fig. 2C). Increased processivity of these mutants correlates with tighter substrate binding (Fig. 2B). However, mutants N479D, H594K and D449G, despite their improved affinity to the substrate (Fig. 2B), displayed comparable processivity to the wt enzyme (22–32 nt, Fig. 2C).

Design of M-MuLV RT multiple mutants

An ideal RT should quickly and efficiently synthesize full-length cDNA at 50°C or even higher temperatures. None of the point mutants met the above requirement, although our analysis revealed a group of previously uncharacterized mutations that increased the enzyme thermostability, affinity to the substrate and the processivity. As beneficial mutations may act synergistically, we have attempted to generate superior RT variants by combining the individual mutations. For this rational design approach, we have selected five substitutions: D200N and L603W [the best-performing mutations, most frequently found in the original selection (Janulaitis *et al.*, 2009; Skirgaila *et al.*, manuscript in preparation)]; T330P and L139P [the third and the fourth most frequent mutations (Janulaitis *et al.*, 2009; Skirgaila *et al.*, manuscript in preparation)], and E607K (one of the few mutations that enabled the synthesis of 1.1-kb cDNA at elevated temperature, Fig. 1). Despite a similar ability, the Q221R mutation was omitted due to extremely low stability of the Q221R mutant in the absence of the substrate (complete inactivation after 5 min at 50°C, data not shown). A stepwise addition of

point mutations resulted in the following multiple mutants: (i) double mutant mut2 (D200N; L603W); (ii) triple mutant mut3 (D200N; L603W; T330P); (iii) quadruple mutant mut4 (D200N; L603W; T330P; E607K); and (iv) pentuple mutant mut5 (D200N; L603W; T330P; E607K; L139P). To assess the performance of the rational design approach, we have also characterized the multiple mutant mut4_S identified during the *in vitro* evolution procedure that contains three of five mutations selected for the construction of mut2-mut5 mutants (L139P; D200N; E607K) and an additional W388R mutation. All the multiple mutants were expressed with an N-terminal His-tag and purified using the affinity chromatography and P11 ion-exchange column as described in Supplementary Methods. The determined specific activities of the multiple mutants at 37°C were slightly higher in comparison with the wt enzyme (~20% increase in the case of mut2, and 50–60% increase in the case of mut3-mut5 and mut4_S, data not shown).

Characterization of M-MuLV RT multiple mutants

Thermal inactivation at 50°C. The thermal inactivation of the multiply-mutated variants was analyzed in the presence of the substrate at 50°C as described above. The determined half-life ($t_{1/2}$) values for the mut2–mut5 multiple mutants are provided in Fig. 3A. The combination of the beneficial D200N and L603W mutations in the double mutant mut2 increased the half-life of the enzyme to ~360 min (Fig. 3A). The addition of the T330P mutation in the triple mutant mut3 further increased the half-life to ~430 min. The E607K mutation slightly decreased the half-life of the quadruple mutant mut4 ($t_{1/2} \approx 360$ min), but the fifth replacement L139P yielded the most thermostable RT variant mut5, with a 12-fold higher $t_{1/2}$ value (~500 min) than the wt enzyme (~40 min, Fig. 3A) and a 1.5-fold higher $t_{1/2}$ value than the multiple mutant mut4_S indentified in the evolution *in vitro* experiment (~310 min, Fig. S2A). In the absence of the template–primer substrate, the increase in thermostability was less prominent: the residual activity after the 5-min-long preincubation at 50°C exceeded the residual activity of the wt enzyme by only ~2-fold (data not shown).

Substrate-binding affinity. EMSA revealed that the double mutant mut2 binds the primer–template with comparable affinity ($K_D \approx 500$ pM) to the mutants bearing the individual D200N and L603W mutations (Fig. 3B). However, the addition of the T330P mutation in the triple mutant mut3 increased the affinity ~3-fold ($K_D \approx 150$ pM); further addition of the E607K and L139P mutations in the multiple mutants mut4 and mut5 resulted in even higher DNA-binding affinities (Fig. 3B). The K_D of the best-performing variant mut5 (~30 pM) was 50-fold higher in comparison with the wt enzyme (~1500 pM, Fig. 3B) and was similar to the K_D of the multiple mutant mut4_S identified in the original selection (~36 pM, Supplementary Fig. S2B).

Processivity. The tighter substrate binding of the RT point mutants led to the concomitant increase in the enzyme processivity (Fig. 2B and C). The same applies to the multiply-mutated RTs (Fig. 3C). As the stretches of cDNA synthesized by the mut2–mut5 and mut4_S RT variants exceeded the length of the fluorescently labeled marker used in the sequencing machine (see below), we have measured their processivity using a radiolabeled DNA/RNA substrate as described in Materials and Methods. The combination of the beneficial D200N and L603W mutations acted cooperatively in the mut2 double mutant, increasing its processivity to 613 nt; a stepwise addition of the T330P, E607K and L139P mutations in the mut3, mut4 and mut5 variants led to the processivity values of ~1000–1500 nt, up to 64-fold higher in comparison with the wt enzyme (Fig. 3C). The processivity of the multiple mutant mut4_S identified in the selection experiment was ~1000 nt (Supplementary Fig. S2C) and fell within the processivity range (~600–1500 nt) of the rationally designed mutants mut2-mut5. As expected, RTs processivity correlated with increased enzymes affinity for the substrate (Fig. 3B and C), suggesting that a tighter substrate binding is the major mechanism behind the higher thermostability and processivity of the multiply-mutated RTs.

DNA synthesis rate. The above experiments confirmed that we had successfully generated RT variants with increased thermostability, substrate-binding affinity and processivity. To test whether these mutations also affected the DNA synthesis rate, we have determined the nucleotide incorporation rates for the wt and mutant RTs using 10 nM enzyme, saturating concentration of the 1.3-kb RNA transcript with an annealed primer (200 nM), and the standard concentration of all dNTPs (1 mM each, see Supplementary Methods for details). The observed nucleotide incorporation rate in this experimental setup is determined by many processes, including enzyme association with the template-primer, DNA elongation, enzyme stalling due to the secondary structures of the template RNA, and enzyme dissociation from the RNA-DNA hybrid upon the incomplete or complete cDNA synthesis. We have performed the initial reaction rate measurements with the wt enzyme, single (D200N and L603W) and multiple mutants at 37°C and 50°C. The initial nucleotide incorporation rate (corresponding to the reaction V_{max}) of all RT variants at 50°C was about 1.5- to 2-fold higher than at 37°C (Supplementary Fig. S3). For example, the determined rates of the wt RT were 4.4 s^{-1} and 6.5 s^{-1} at 37°C and 50°C, respectively. The improvement of the DNA synthesis rate observed for all RT variants at 50°C may be attributed to the increase in the chemical reaction rate and reduced stability of the RNA secondary structures. The reaction rates of multiple mutants decreased in the line mut2 > mut3 > mut4 > mut5 (Supplementary Fig. S3), showing the negative correlation with the affinity to substrate (mut2 < mut3 < mut4 < mut5; Fig. 3B). This relationship also holds for the D200N mutant (both substrate-binding affinity and DNA synthesis rate were comparable to the mut2 double mutant), but fails for the L603W mutant (the binding affinity was similar to mut2, but DNA synthesis rate was 1.5-fold slower), the wt enzyme (both the binding affinities and the reaction rates were lower in comparison to the mut2-mut5 mutants), and the multiple mutant mut4_S (the binding affinity was similar to the mut5 pentuple mutant, but the DNA synthesis rate was 1.5-fold higher, Fig. 3B, Supplementary Figs S2B and S3). The lack of a consistent relationship between the binding affinity and the initial nucleotide incorporation rate suggests a multiplex relationship between these two parameters in RT reactions. Indeed, a tighter binding of RT to the primer–template may both increase the average

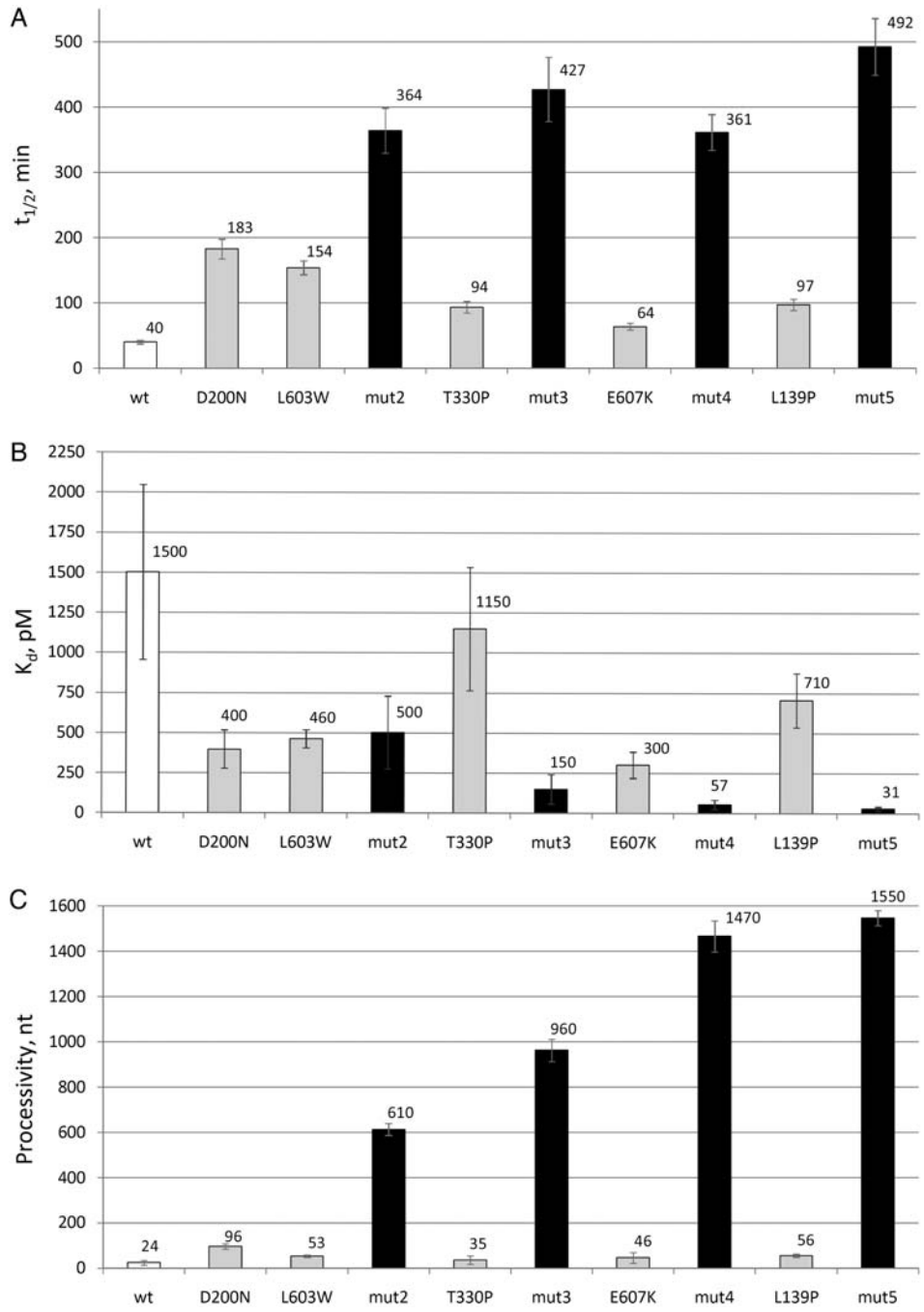

Fig. 3. Biochemical characterization of the multiple M-MuLV RT mutants mut2, mut3, mut4 and mut5 constructed by a rational design. (**A**) Thermal inactivation half-life ($t_{1/2}$, minutes) at 50°C. The inactivation rates were measured in the presence of the template–primer substrate. (**B**) Affinity to substrate. The equilibrium dissociation constants of the RT—primer–template complexes (K_D, pM) were measured using the EMSA. (**C**) Enzyme processivity (in nucleotides) at 37°C. The average values of at least three (A) or two (B and C) independent measurements ± 1 SD are given. For clarity, all graphs contain data for the multiple mutants (black columns), the wt enzyme (white column) and point mutants used for the rational design of the multiple mutants mut2-mut5 (gray columns).

nucleotide incorporation rate due to faster association of the enzyme with the substrate and a reduced number of dissociation/association events necessary to produce a full-length product, but it may also have an opposite effect due to slower enzyme dissociation after a full-length cDNA is made.

The highest temperature of the 7.5-kb cDNA synthesis. To compare the performance of the wt M-MuLV RT and the newly generated variants in the real-life applications, we have determined the highest temperature suitable for the synthesis of 7.5-kb cDNA. The reactions with the wt enzyme, single mutants used for the generation of the multiple mutants, and all the multiply-mutated variants were carried out at 37–69°C for 1 h using a 7.5-kb-long RNA transcript as a template (see Materials and Methods for details). The wt RT could accomplish synthesis of the full-length 7.5-kb cDNA only at $\leq$45°C (Fig. 4A and B). In a similar setup, the wt enzyme synthesized 1.1 kb at the temperatures of up to 48°C (data not shown), indicating that the synthesis of a

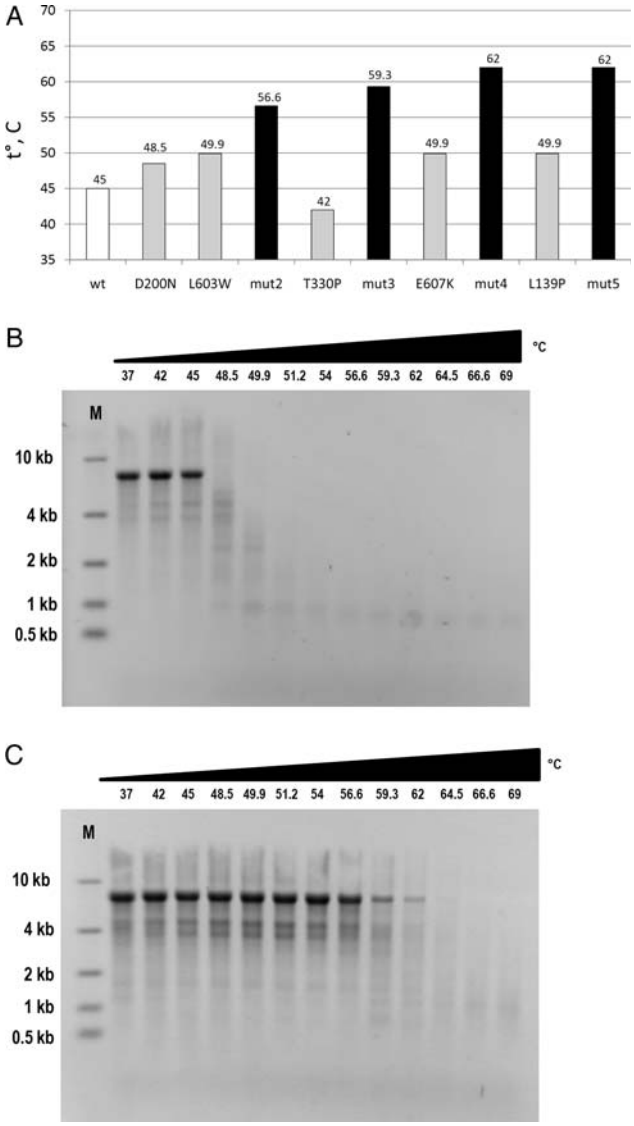

Fig. 4. Synthesis of 7.5-kb cDNA. The reverse transcription reactions were performed on a 7.5-kb RNA template at various temperatures. The yield of the full-length cDNA was assessed by analyzing the reaction products by alkaline agarose electrophoresis. (**A**) The highest 7.5-kb cDNA synthesis temperatures of the wt RT (white column), the multiple mutants mut2–mut5 (black columns), and the point mutants used for the rational design of the mut2–mut5 multiple mutants (gray columns). (**B and C**) 7.5-kb cDNA synthesis experiments performed with the wt enzyme (**B**) and the best-performing multiple mutant mut5 (**C**). The samples were analyzed by alkaline agarose electrophoresis. FastRuler High Range DNA Ladder (Thermo Fisher Scientific; 10 000, 4000, 2000, 1000 and 500 bp) was used as a reference.

longer cDNA is more reliant on the (thermo)stability of the enzyme. Most single amino acid mutants (D200N, L603W, E607K and L139P) could synthesize 7.5-kb cDNA at higher temperatures than the wt enzyme ($\leq$48–50°C, Fig. 4A and Supplementary Fig. S4A). The multiple mutants mut2, mut3, mut4 and mut5 showed a further increase in the maximum synthesis temperature (56°C, 59°C, 62°C and 62°C, respectively, Fig. 4A), while the multiple mutant mut4_S could synthesize 7.5-kb cDNA at $\leq$59°C (Supplementary Fig. S4B). Thus, the best-performing multiple mutant mut5 was able to accomplish synthesis of the full-length 7.5-kb cDNA at 17° higher temperature than the wt enzyme. The temperature range for efficient cDNA synthesis by the mut5 mutant is 37–59°C (Fig. 4C).

Fidelity of cDNA synthesis. Application of the newly generated RT variants *in vitro* would be possible only if their DNA synthesis fidelity is not compromised. The error rate reported for wt M-MuLV RT is in the range of 1 error per 15 000–27 000 nucleotides synthesized (Potter *et al.*, 2003; Arezi and Hogrefe, 2007). Preliminary assessment of the mut5 variant fidelity (see Supplementary Materials and Methods for details) confirmed that its fidelity is comparable to the wt enzyme (<1 error per 10 000 nucleotides, data not shown). This is not surprising, as the CRD technique employed for RT selection involves full-length cDNA synthesis from mRNA encoding the target enzyme. The fidelity of nucleotide incorporation therefore is required to secure the population of selected RTs during the repeated cycles of selection.

Discussion

We have previously applied the combination of RD and IVC methods to identify a number of multiple M-MuLV RT mutants capable of cDNA synthesis at elevated temperatures. The selected variants carried a number of previously unknown substitutions potentially implicated in the increase of enzyme thermostability (Janulaitis *et al.*, 2009; Skirgaila *et al.*, manuscript in preparation). However, the impact of individual mutations on the ability of M-MuLV RT to synthesize cDNA at elevated temperatures remained unknown. In this report, we constructed and characterized RT point mutants that carried substitutions at a number of previously uncharacterized amino acid positions, and by combining the best-performing mutations we were able to generate multiple RT mutants with dramatically increased substrate-binding affinity, processivity and thermostability.

In the first set of experiments, we constructed 28 M-MuLV RT point mutants that carried previously uncharacterized substitutions most frequently selected in the evolution *in vitro* experiment. We determined their activity ratio at 50°/37°C and the ability to synthesize full-length cDNA at elevated temperatures (Fig. 1). By giving the preference to the cDNA synthesis ability, we have selected 12 amino acid positions for partial site-saturation mutagenesis (56 additional point mutants at 12 selected positions).

Twelve best-performing point mutants were subjected to a more detailed biochemical characterization. Heat inactivation experiments in the presence of a substrate-primer revealed that the life time of the point mutants D200N, L603W, T330P and L139P at 50°C increased 2.5- to 4-fold in comparison with the wt enzyme (Fig. 2A). The higher thermostability in all cases was accompanied by the 1.5- to 5-fold improved substrate binding (decreased K_D values, Fig. 2B) and increased processivity (Fig. 2C). However, the thermal inactivation rates in the absence of template–primer remained largely unchanged. This suggests that all newly identified mutations increase enzyme thermostability primarily through tighter substrate binding; a similar mechanism for the RT stabilization was reported previously (Gerard *et al.*, 2002b; Arezi and Hogrefe, 2009).

To understand the structural mechanism behind the improved performance of the D200N, L603W, L139P and T330P point mutants, we used the full-length homology

model of the M-MuLV RT (Fig. 5A; see Supplementary Methods for details). The aspartate 200 resides on an α-helix that is close to the RT catalytic center (Fig. 5B): the catalytic aspartates D224 and D225, located on a loop connecting two anti-parallel β-strands, are only 7–9 Å away from the D200 residue. Therefore, mutations at the 200th position may cause structural rearrangements of the catalytic and neighboring residues that in turn may alter the substrate-binding affinity and the rate of catalysis. Surprisingly, all tested substitutions at the 200th position, both drastic (D200W, D200G) and conservative (D200E, D200N), had a positive effect on the 50°/37°C activity ratio, the only substitution not tolerated being D200P (Supplementary Fig. S1). Moreover, the best-performing variant in terms of cDNA synthesis, D200N, also displayed significantly improved thermostability, substrate-binding affinity, processivity and DNA synthesis rate (Fig. 2 and Supplementary Fig. S3).

The L139 amino acid is located at the core of the palm domain. It forms a hydrophobic cluster with the residues from several distinct sequence regions, including the active site structural motive β-strand—loop—β-strand (Fig. 5B). Large and polar substitutions of the L139 drastically diminished the 50°/37°C activity ratio of mutant enzymes (Supplementary Fig. S1); this may be attributed to the disruption of the hydrophobic cluster. On the contrary, the L139P mutation had a positive effect on the thermostability, substrate-binding affinity and processivity of the enzyme (Fig. 2). Presumably, a proline residue at the 139 position stabilizes the hydrophobic cluster (I218, L220, L273) through increased backbone rigidity (Sakaguchi *et al.*, 2007); this in turn may improve the template–primer binding at the active site. Another example of structure stabilization by proline is the T330P mutant. The T330 residue is located at the terminus of a small helical fragment of the thumb domain (the structure of this fragment was modeled using I-Tasser server (Roy *et al.*, 2010), as it was absent in the available crystallographic structures). As the primer–template typically interacts with the palm subdomain of polymerases between the fingers and the thumb subdomains of the enzyme, this M-MuLV RT helix presumably is located in close proximity to the substrate, and the neighboring residue of T330, the lysine K329, may form direct contact to the phosphodiester backbone (Fig. 5C). As the coil-helix boundary positions serve as the starting points of protein unfolding (Fierz *et al.*, 2009), the backbone rigidity provided

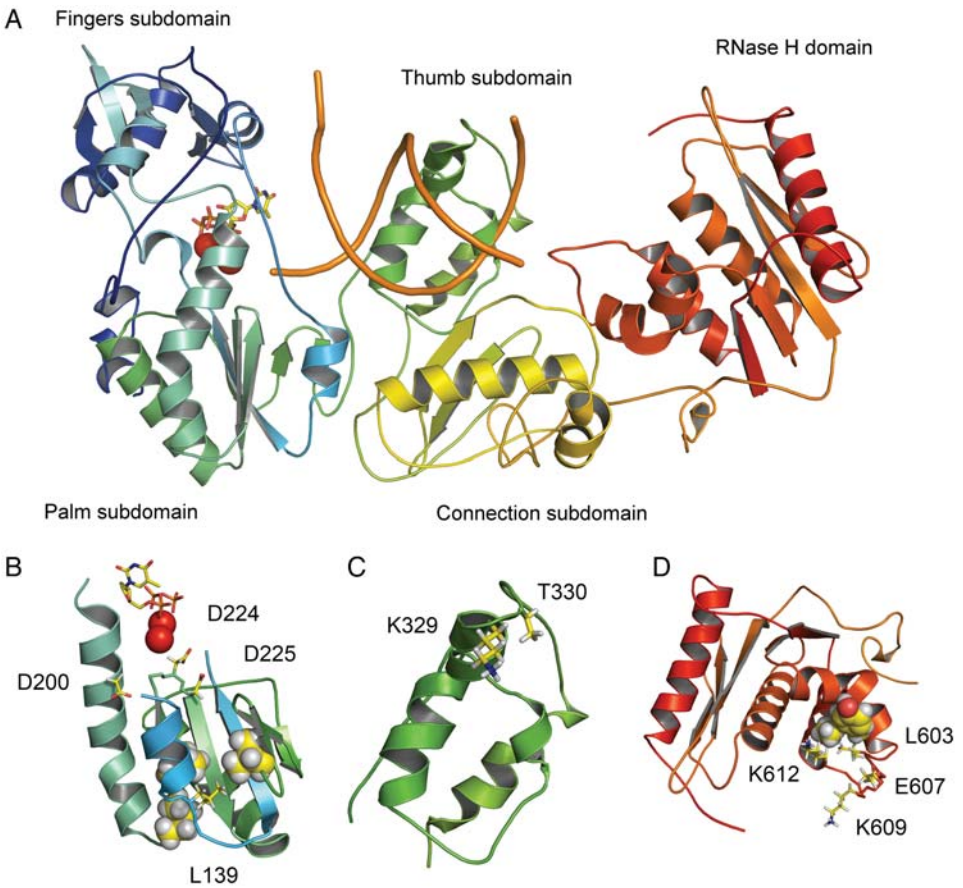

Fig. 5. Structural context of the newly identified thermostable mutations. (**A**) The model of the full-length M-MuLV RT (see Supplementary Methods for details) colored in a rainbow mode ranging from blue (N-terminus) to red (C-terminus). The polymerase finger and palm subdomains are colored in blue and green, thumb subdomain is colored in light-green, connection subdomain—yellow and RNase H domain—red and orange. Amino acid numbering is the same as in PDB ID 2fjw (Goodwin *et al.*, 2006). The magnesium ions (red spheres), incoming dTTP (sticks), and part of the DNA–DNA substrate (cartoon) are taken from the aligned structure of HIV RT [PDB ID 1RTD (Huang *et al.*, 1998)]. (**B**) The D200 amino acid (sticks) is located on the H-helix in the vicinity of the polymerase active site (the active site residues D224 and D225 are shown as sticks). The L139 residue (sticks) forms a hydrophobic cluster with the aliphatic amino acids I218, L220 and L273 shown as spheres. (**C**) The T330 residue (sticks) is located at the terminus of a small helical fragment of the thumb subdomain. The neighboring residue K329 (sticks) may form direct contact with the phosphodiester backbone of the substrate. (**D**) The RNase H domain residue L603 (sticks) forms a hydrophobic cluster with the I617 and Y598 residues (spheres) and together with K609, K612 (Lim *et al.*, 2006) and E607 residues (sticks) is located on the putative substrate-binding loop.

by the T330P substitution should stabilize the DNA-contacting helical motif, again resulting in a stabilized protein and enhanced primer–template binding.

A similar stabilization mechanism can be proposed for the L603W mutant. Like T330, the L603 residue is positioned at the helix-unstructured loop boundary at the terminus of the so-called 'C- helix' in the RNase H domain (Lim *et al.*, 2006). According to the model of the full-length RNase H domain, L603 forms a hydrophobic cluster together with the I617 and Y598 residues (Fig. 5D), and is located on the same loop as the putative substrate phosphate backbone interacting lysines K609 and K612 (Lim *et al.*, 2006). All large hydrophobic/aromatic substitutions of L603 improved the RT 50°/37°C activity ratio (Fig. S1). Presumably, large side chains (e.g., indole ring in the case of the L603W mutant) stabilize the hydrophobic cluster. This in turn may stabilize the neighboring C-helix and the loop with the DNA/RNA contacting residues, resulting in the enhanced substrate binding, protein stability and processivity of the mutant enzymes (Fig. 2). Our model of the substrate-bound full-length M-MuLV RT also accounts for the effect of the E607K mutation (5-fold improved substrate binding, Fig. 2B). Indeed, an extra positive charge introduced into the putative DNA-interacting loop (Fig. 5D) should facilitate binding of the negatively charged RNA–DNA heteroduplex.

As discussed above, the best-performing mutations D200N, L603W, T330P and L139P are distributed in different structural regions of the enzyme, suggesting that their beneficial effects might sum up in multiple enzyme mutants. By combining the four above mutations and the E607K mutation, selected due to the strongest impact on the substrate-binding affinity (Fig. 2B), we have produced RT variants mut2, mut3, mut4 and mut5. As expected, a stepwise addition of beneficial mutations had an incremental effect on the thermostability, substrate-binding affinity and processivity of the multiple mutants. The best-performing RT variant was the pentuple mutant mut5 having all five beneficial mutations. In comparison with the wt enzyme, it showed a ~12-fold longer life time at 50° ($t_{1/2}$ ~500 min vs. ~40 min, Fig. 3A), a ~50-fold higher substrate-binding affinity (K_D ~30 pM vs. ~1500 pM, Fig. 3B) and a ~65-fold higher processivity (~1500 nt vs. ~24 nt, Fig. 3C). Moreover, the processivity of all multiple mutants (600–1500 nt per single polymerization event) exceeded the processivity of AMV RT [~400 nt (Gerard *et al.*, 2002b)] making them ideal tools for the long cDNA synthesis. Indeed, the 7.5-kb cDNA synthesis experiments confirmed the exceptional ability of the multiply-mutated variants to synthesize cDNA at increased reaction temperatures (Fig. 4A). The best-performing multiple mutant mut5 synthesized full-length 7.5-kb cDNA even at 62°C, i.e. at 17° higher temperature than the wt enzyme. Such extreme cDNA synthesis temperatures might prove useful for the amplification of relatively short, but highly structured RNA in experiments where quantitative conversion of RNA to cDNA is not required.

Intriguingly, all thermostable multiple mutants constructed and analyzed in this study were RNase H proficient (data not shown). This challenges the traditional belief that the efficient synthesis of long cDNA at elevated temperatures requires RNase H^- phenotype. Moreover, in the evolution *in vitro* experiment (Janulaitis *et al.*, 2009; Skirgaila *et al.*, manuscript in preparation), the RNase H inactivating mutations (positions D524, D653 and D583), were selected with a comparable frequency (31, 23 and 21 clones, respectively) to the mutations at the D200, L603, T330 and L139 positions characterized here (30, 18, 15 and 14 clones, respectively), also suggesting that under the employed selection conditions [increased cDNA synthesis temperature (Janulaitis *et al.*, 2009; Skirgaila *et al.*, manuscript in preparation)], improved substrate-binding achieved via the D200, L603, T330 and L139 substitutions is as advantageous as the loss of the RNase H activity. Presumably, the RNase H and polymerase domains of the M-MuLV RT enzyme do not act concertedly, but instead compete for the primer–template substrate. At any given time the RT acts in either of two modes, primer extension/cDNA synthesis or RNase H/template degradation. These two activities in the wt M-MuLV RT are balanced and optimized for the conversion of retroviral RNA into the double-stranded DNA. However, for the in vitro applications, where highly efficient cDNA synthesis is required, the RT can be converted into a dominant cDNA polymerase in two alternative ways: either via inactivation of the RNase H activity (Gerard *et al.*, 2002b), or via enhancement of the DNA polymerase activity or the substrate-binding affinity [ref. Arezi and Hogrefe (2009) and this study].

In the original evolution *in vitro* experiment (Janulaitis *et al.*, 2009; Skirgaila *et al.*, manuscript in preparation), we have identified a large set of multiple M-MuLV RT mutants capable of cDNA synthesis at elevated temperatures. However, none of them carried the full set of the best-performing mutations D200N, L603W, T330P and L139P identified in the current study. The closest match to the rational design mutants mut3–mut5 was the multiple mutant mut4_S, which carried three of five mutations selected for the rational design (L139P, D200N and E607K), and only a single additional mutation W388R. The determined parameters of the mutant mut4_S ($t_{1/2}$ at 50°C ~300 min, processivity ~1000 bp, K_D ~36 pM and the maximum 7.5-kb cDNA synthesis temperature ~59°C) are comparable to the rational design mutant mut4, but are inferior to the best mutant mut5. This implies that the rational design approach used in our study, which involves the selection and combining of the best-performing individual mutations, may yield a better end-product than screening of multiple mutants isolated directly in the evolution *in vitro* experiment.

Taken together, our results demonstrate that CRD technique (Janulaitis *et al.*, 2009; Skirgaila *et al.*, manuscript in preparation) is a valuable addition to other polymerase *in vitro* evolution schemes such as polymerase genetic complementation assay (Sweasy and Loeb, 1993), compartmentalized self-replication (Ghadessy *et al.*, 2001) or phage display-based polymerase screening (Xia *et al.*, 2002; Henry and Romesberg, 2005). The distinguishing feature of CRD methodology is the possibility to perform screening of active RT completely *in vitro* under a broad range of conditions (high temperature, different buffers) and at a single-molecule level.

By selecting and combining the best-performing mutations identified by CRD, we were able to produce a set of highly thermostable and processive M-MuLV RT variants. A similar CRD-based setup could be applied for the improvement of the other nucleic acid enzymes.

Supplementary data

Supplementary data are available at *PEDS* online.

Acknowledgements

The authors thank Dr R. Rimšelienė and Dr D. Šikšnienė for the help with protein engineering, construction and analysis, and Dr R. Leipuvienė and Dr Š. Zigmantas for the help with manuscript preparation.

Funding

This work was supported by Thermo Fisher Scientific, Graiciuno 8, LT-02241 Vilnius, Lithuania.

References

Arezi,B. and Hogrefe,H. (2009) *Nucleic Acids Res.*, **37**, 473–481.
Arezi,B. and Hogrefe,H.H. (2007) *Anal. Biochem.*, **360**, 84–91.
Arezi,B., McCarthy,M. and Hogrefe,H. (2010) *Anal Biochem.*, **400**, 301–303.
Chen,L., Setterquist,R. and Latham,G. (2004) US Patent Application, US20050232934.
Coté,M.L. and Roth,M.J. (2008) *Virus Res.*, **134**, 186–202.
Das,D. and Georgiadis,M.M. (2004) *Structure*, **12**, 819–829.
Fierz,B., Reiner,A. and Kiefhaber,T. (2009) *Proc. Natl. Acad. Sci. U S A*, **106**, 1057–1062.
Georgiadis,M.M., Jessen,S.M., Ogata,C.M., Telesnitsky,A., Goff,S.P. and Hendrickson,W.A. (1995) *Structure*, **3**, 879–892.
Gerard,G.F., Fox,D.K., Nathan,M. and D'Alessio,J.M. (1997) *Mol. Biotechnol.*, **8**, 61–77.
Gerard,G.F., Collins,S. and Smith,M.D. (2002a) *Biotechniques*, **33**, 984, 986. 988 passim.
Gerard,G.F., Potter,R.J., Smith,M.D., Rosenthal,K., Dhariwal,G., Lee,J. and Chatterjee,D.K. (2002b) *Nucleic Acids Res.*, **30**, 3118–3129.
Ghadessy,F.J., Ong,J.L. and Holliger,P. (2001) *Proc. Natl. Acad. Sci. U S A*, **98**, 4552–4557.
Goodwin,K.D., Lewis,M.A., Tanious,F.A., Tidwell,R.R., Wilson,W.D., Georgiadis,M.M. and Long,E.C. (2006) *J. Am. Chem. Soc.*, **128**, 7846–7854.
Henry,A.A. and Romesberg,F.E. (2005) *Curr. Opin. Biotechnol.*, **16**, 370–377.
Hogrefe,H., Arezi,B. and Xing,W. (2006) International Patent Application, WO2007022045.
Huang,H., Chopra,R., Verdine,G.L. and Harrison,S.C. (1998) *Science (80-).*, **282**, 1669–1675.
Janulaitis,A., Skirgaila,R. and Siksniene,D. (2009) International Patent Application, WO2009125006.
Kotewicz,M.L., Sampson,C.M., D'Alessio,J.M. and Gerard,G.F. (1988) *Nucleic Acids Res.*, **16**, 265–277.
Kranaster,R., Drum,M., Engel,N., Weidmann,M., Hufert,F.T. and Marx,A. (2010) *Biotechnol. J.*, **5**, 224–231.
Lanciault,C. and Champoux,J.J. (2006) *J. Virol.*, **80**, 2483–2494.
Lim,D., Gregorio,G.G., Bingman,C., Martinez-Hackert,E., Hendrickson,W.A. and Goff,S.P. (2006) *J. Virol.*, **80**, 8379–8389.
Mattheakis,L.C., Bhatt,R.R. and Dower,W.J. (1994) *Proc. Natl. Acad. Sci. U S A*, **91**, 9022–9026.
Miller,O.J., Bernath,K., Agresti,J.J., *et al.* (2006) *Nat. Methods.*, **3**, 561–570.
Ndongwe,T.P., Adedeji,A.O., Michailidis,E., *et al.* (2012) *Nucleic Acids Res.*, **40**, 345–359.
Pinto,F.L. and Lindblad,P. (2010) *Anal Biochem.*, **397**, 227–232.
Potter,J., Zheng,W. and Lee,J. (2003) *Focus (Invitrogen newsletter).*, **25**, 19–24.
Potter,R.J., Lee,J., Smith,M.D., Dhariwal,G., Gerard,G.F. and Rosenthal,K. (2002) WO2004024749.
Roy,A., Kucukural,A. and Zhang,Y. (2010) *Nat. Protoc.*, **5**, 725–738.
Sakaguchi,M., Matsuzaki,M., Niimiya,K., Seino,J., Sugahara,Y. and Kawakita,M. (2007) *J. Biochem.*, **141**, 213–220.
Sambrook,J., Russell,D.W. and Laboratory,C.S.H. (2001) *Molecular cloning: a laboratory manual/Joseph Sambrook, David W. Russell*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
Sauter,K.B.M. and Marx,A. (2006) *Angew Chem. Int. Ed. Engl.*, **45**, 7633–7635.
Schultz,S.J., Zhang,M., Kelleher,C.D. and Champoux,J.J. (1999) *J. Biol. Chem.*, **274**, 34547–34555.
Ståhlberg,A., Kubista,M. and Pfaffl,M. (2004) *Clin. Chem.*, **50**, 1678–1680.
Suo,Z. and Johnson,K.A. (1998) *J. Biol. Chem.*, **273**, 27259–27267.
Sweasy,J.B. and Loeb,L.A. (1993) *Proc. Natl. Acad. Sci. U S A*, **90**, 4626–4630.
Tawfik,D.S. and Griffiths,A.D. (1998) *Nat. Biotechnol.*, **16**, 652–656.
Wu,L., Huang,M., Zhao,J. and Yang,M. (2005) *Acta Biochim. Biophys. Sin. (Shanghai).*, **37**, 634–642.
Wu,W., Henderson,L.E., Copeland,T.D., Gorelick,R.J., Bosche,W.J., Rein,A. and Levin,J.G. (1996) *J. Virol.*, **70**, 7132–7142.
Xia,G., Chen,L., Sera,T., Fa,M., Schultz,P.G. and Romesberg,F.E. (2002) *Proc. Natl. Acad. Sci. U S A*, **99**, 6597–6602.
Yasukawa,K., Mizuno,M., Konishi,A. and Inouye,K. (2010) *J. Biotechnol.*, **150**, 299–306.
Zahnd,C., Amstutz,P. and Plückthun,A. (2007) *Nat. Methods*, **4**, 269–279.