

ORIGINAL ARTICLE

Open Access



Expression of thermostable MMLV reverse transcriptase in *Escherichia coli* by directed mutation

Marzieh Divbandi¹ , Ahad Yamchi^{2*} , Hadi Razavi Nikoo¹, Abdolvahab Moradi¹ and Alijan Tabarraei^{1,3*}

Abstract

The functionality of Moloney murine leukemia virus reverse transcriptase (MMLV RT) will increase with the improvement of its solubility and thermal stability. Introduce directed mutation at specific positions of the MMLV RT sequence and codon optimization is needed to achieve these properties. The two RT coding sequences with (rRT-K) and without directed mutations (rRT-L) were versatility optimized and expressed to analyze the ribonuclease H (RNase H) inactivity and thermostable polymerase activity. For this purpose, the five-point mutations (438–442aa) and three-point mutations (530, 568, and 659 aa) were done at the RT connection domain and RNase H active site, respectively. High expression levels of rRT-L and rRT-K were obtained in *E. coli* BL21(DE3) and BL21(shuffle) strains, 0.5 mM IPTG concentration at 37 °C, and 8 hours' post-induction condition. Then, recombinant enzymes were purified and verified by Ni-NTA resin and western blotting. Insilico analysis (IUpred 3.0) showed that the directed mutation in the RNase H domain caused the formation of disorder regions or instability in the RNase H domain of rRT-K compared to rRT-L. The modified RT-PCR and the RT-LAMP reactions proved the RNase H inactivity of rRT-K. In addition, increasing of thermostability of rRT-K compared to rRT-L and commercial RT was evaluated by the RT-PCR and RT-LAMP reactions. The results showed that rRT-K could successfully tolerate 60 °C in the two methods. This study revealed that the directed mutations and the versatile sequence optimization can promise to produce thermostable commercial enzymes to decrease non-specific one-step RT-PCR and RT-LAMP products.

Key points

- The versatile sequence optimization to improve the yield of the recombinant enzyme.
- Directed specific mutations can promise to produce thermostable and soluble RT commercial enzymes.
- Predicted hot mutations in the active site of the RNase H domain of recombinant RT without any effect on its structure.

Keywords Reverse transcriptase, Recombinant protein expression, Murine Moloney Leukemia Virus (MMLV), Site-directed mutagenesis, Sequence optimization, Thermostability

*Correspondence:
Ahad Yamchi
yamchi@gau.ac.ir
Alijan Tabarraei
alijant@yahoo.com

¹Department of Microbiology, Faculty of Medicine, Golestan University of Medical Sciences, Gorgan, Iran
²Department of Biotechnology, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Golestan, Iran
³Infectious Disease Research Center, Golestan University of Medical Sciences, Gorgan, Iran

Introduction

Reverse transcriptase (RT) is the key protein of Moloney murine leukemia virus (MMLV), which is the most widely used for cDNA synthesis and RNA amplification due to its robust polymerase activity and high fidelity which is routinely used in RT-PCR, RT-LAMP, and RACE (Hu and Hughes 2012; Nishimura et al. 2015; Pinto and Lindblad 2010; Zajac et al. 2013). RT is a ~75 kDa single polypeptide, as an enzyme that catalyzes DNA from RNA or DNA templates, as well as ribonuclease activity that cleaves the RNA in an RNA/DNA hybrid via the hydrolytic mechanism (Levin et al. 1988; Martín-Alonso et al. 2021). RT has three main functional and structural domains, including the polymerase domain (PD, aa 1–361), connection domain (CD, aa 362–496), and RNase H domain (RD, aa 497–671) (Coté and Roth 2008; Tanese and Goff 1988). Heat stability of RT is necessary to improve the synthesis of cDNA with long lengths and one-step downstream reactions of RT. Researchers have used various approaches to modify the sequence of native RT and improve its properties (Konishi et al. 2014; Mizuno et al. 2010). However, native RT expresses at a low expression level and solubility, making it not preferable for commercial purposes (Das and Georgiadis 2001; Nuryana et al. 2022). The versatile sequence optimization and directed mutations are key points to increase the yield and efficiency of enzymes (Konishi et al. 2014; Mizuno et al. 2010). The studies revealed that incorporating site-directed mutations could be related to increased thermal stability (Konishi et al. 2014; Mizuno et al. 2010; Narukawa et al. 2021a). In this study, we considered two points in designing the RT sequence. The first was to apply five-point mutations (438–442aa) and three-point mutations (530, 568, and 659 aa) at the RT connection domain and RNase H active site, respectively, to increase the thermostability of recombinant RT. The second was versatile optimization including RT coding sequence and redesign of the ribosome binding site (RBS) of the pET9a vector to obtain a high yield of recombinant RT. Then two constructs with (rRT-K) and without (rRT-L) mutations were expressed in *E. coli*. After purification of recombinant enzymes, the successful expression of respected enzymes was verified by SDS-PAGE and western blotting. The thermostable polymerase activity and ribonuclease H (RNase H) inactivity of rRT-K compared to rRT-L was evaluated after expressing in 0.5 mM IPTG concentration at 37 °C, and 8 h post-induction condition. The RT-PCR and RT-LAMP reactions confirmed the thermostable activity of rRT-K compared to rRT-L and commercial enzymes.

Materials and methods

In-silico studies

Designing of site-directed mutagenesis, sequence optimizing, and constructs synthesizing in expression vector

The 2086-bp coding sequence of RT from the MMLV (GenBank Accession No: NC-001501.1) was optimized to increase the expression level of this protein in *E. coli*. The determination of codon adaptation index (CAI), codon pair bias, and GC content related to the coding sequence was done based on the *E. coli* codon usage database (<http://www.kazusa.or.jp/codon/>). In addition, the negative *cis*-sequence elements consisting of mRNA destabilizing, and cryptic regions were removed. Also, the sequence of the RBS region was removed by *Xba*I and *Bam*HI restriction enzymes in the pET9a vector and it was redesigned to avoid secondary structure formation in the RBS site. Also, the full length of CDS was optimized to form weak secondary structures to speed translation by ribosome machinery (Masoudi et al. 2021; Yamchi et al. 2024). Additionally, we have designed two sequences including the RT-L sequence without point mutations and the RT-K sequence which contains five-point mutations (438–442aa) and three-point mutations (530, 568, and 659 aa) at the RT connection domain and RNase H active site, respectively. To predict the effective mutations in the RNase H active site including the inactivation of the RNase H domain while maintaining domain structure, we used the Phyre Investigator tool (<http://www.sbg.bio.ic.ac.uk/phyre2/html/help.cgi?id=help/investigator>). The two graphs consisting of mutation analysis and sequence profile were used to identify the effective mutation and residue preference, respectively (Kelley et al. 2015). The secondary structure of mRNA was optimized and the stability of structures was calculated at the RNA fold web server, Vienna RNA (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>). The final nucleic acid sequences were synthesized by Biomatik Co. (Canada), and cloned into the restricted enzyme sites (*Xba*I and *Bam*HI) of pET9a. The final diagram of the RT-L and the RT-K constructs is presented in (Fig. 1).

Homology modeling and molecular docking of RT-L and RT-K

To predict the 3-dimensional (3D) structure of RT-L and RT-K, their sequence was submitted to the I-TASSER server (<https://zhanglab.ccmb.med.umich.edu/I-TASSER>), which uses a hierarchical approach to protein structure prediction. The best model of RT-L and RT-K was refined by UCSF chimera 1.16 (<https://www.cgl.ucsf.edu/chimera/>) and Swiss-PDB viewer (spdb viewer) software (<https://spdbv.unil.ch/>). The validation of the 3D structure was analyzed by ERRAT (<https://www.doe-mbi.ucla.edu/errata/>) and Rampage servers (<https://www.ram-pageservers.com/>). ERRATA server predicts the structure

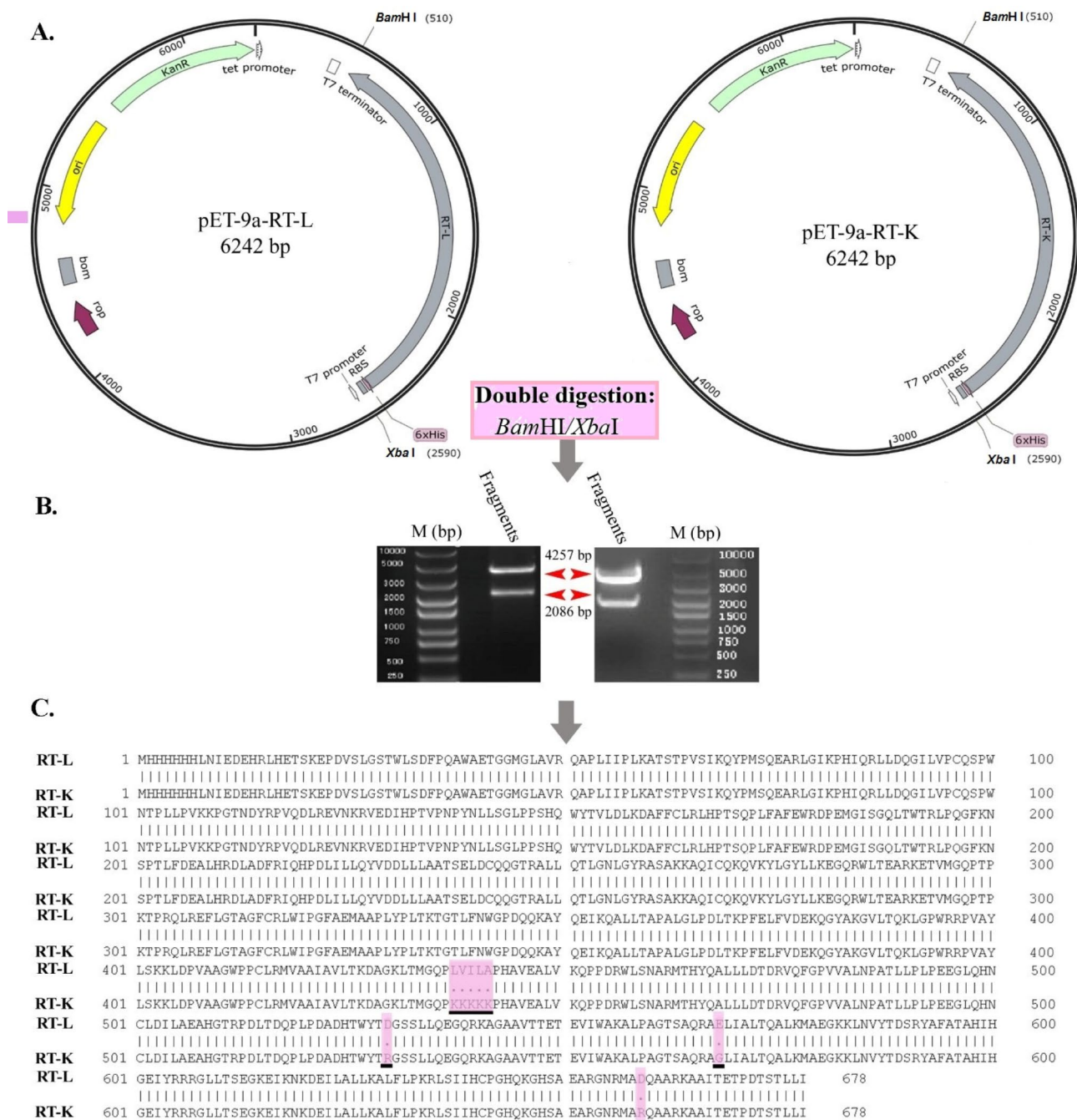


Fig. 1 **A** The constructs including *rRT-L* and *rRT-K* genes based on the pET-9a expression vector. **B** Electrophoretic pattern of digested *rRT-L* and *rRT-K* constructs with *XbaI* and *BamHI*. **C** The alignment of two *rRT-L* and *rRT-K* sequences. Site point-directed mutations were highlighted

accuracy based on the quality factor. The best quality factor should be more than 95%. Blind docking between RT-L and RT-K with RNA ligand was performed using the HDock server (<http://hdock.phys.hust.edu.cn>) to evaluate the binding affinity of the polymerase domain of rRTs. HDock server is based on a hybrid algorithm of template-based modeling and ab initio-free docking. The results were analyzed using the PDBSum (<http://www.ebi.ac.uk/thornton-srv>) server.

Prediction of order and dis-order regions of the recombinant RT-L and RT-K

The order and dis-order regions of the recombinant RT-L and RT-K were predicted by IUpred 3.0 (<https://iupred3.elte.hu/>). The statistical algorithm is used to predict the order and dis-order regions based on a 20×20 matrix, which characterizes the general preference of each pair of amino acids to be in contact as observed in a dataset of globular proteins (Erdős et al. 2021).

In-vitro studies

Escherichia coli strains, growth kinetics, and transformation

Two strains of *E. coli* BL21(DE3) and BL21(shuffle) were used for protein expression. The recombinant plasmids were dissolved in sterile TE buffer (1 M Tris pH8, and 0.5 M EDTA) and then transformed into strains of the *E. coli* competent cells according to standard laboratory protocol (Hanahan et al. 1991). The transformed bacterial cells were cultured overnight in LB broth (supplemented with 100 µg/mL kanamycin) at 37 °C in a shaking incubator (180 g) for 16 h. The fresh LB (containing 100 µg/mL kanamycin) was inoculated and grown to reach an OD_{600nm} of 0.6, Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added at different concentrations (0.1, 0.5, and 1 mM). Additionally, samples were cultured at two different temperatures (20 °C and 37 °C), and ten post-induction times (2–20 h) to optimize the induction conditions.

SDS-PAGE analysis and western blotting

Cell pellets were dissolved in the lysis buffer (150 mM NaCl, 50 mM Tris-Cl pH8.0, 1.0% Triton X-100, and protease inhibitor cocktail (ab271306) and sonicated at 150 Hz amplitude with a 2 mm probe for 10 cycles each consisting of 20s of sonication followed by 10s intervals on ice. The sonication solution was centrifuged at 8500 g, at 4 °C for 15 min. The inclusion body and supernatant of cells were separately collected and the inclusion body part dissolved in 8 M urea. The protein of the inclusion body and supernatant were separately quantified by the Bradford assay. The size and the yield of recombinant RT-L and RT-K proteins were confirmed and compared by coomassie-blue stained 12% SDS-PAGE, respectively (Laemmli 1970). Western blot was done using mouse anti-His (Santa Cruz, USA) with a dilution ratio of 1:2500 and horseradish peroxidase (HRP)-conjugated anti-mouse IgG antibody (Sigma-Aldrich, Germany) with a dilution ratio of 1:4500. Finally, 3,3'-Diaminobenzidine (DAB, Sigma Aldrich, USA) was used to visualize the blots.

Purification and refolding of the recombinant RT-L and RT-K

The rRT-L and rRT-K were mostly produced in *E. coli* BL21 (shuffle) strain as inclusion body form. Also, rRT-K was found in the supernatant more than rRT-L. Both the insoluble and soluble fractions were allowed to pass through the Ni-NTA agarose column after equilibration (Qiagen-CN: 31014). The column was washed with five-bed volumes of the wash buffers (100 mM NaH₂PO₄, 10 mM Tris-Cl, 8 M Urea, pH 6.3) three times. For the soluble fractions, the column was washed with another buffer containing (50 mM NaH₂PO₄, 300 mM NaCl, 10 mM Imidazole, pH 8). Finally, the insoluble and soluble

fractions were eluted with two-bed volumes of two different elution buffers (100 mM NaH₂PO₄, 10 mM Tris-Cl, 8 M Urea pH 4.5 and 3) and (50 mM NaH₂PO₄, 300 mM NaCl, 200 and 250mM Imidazole pH 8), respectively. The concentration of purified recombinant proteins was quantified by the Bradford method (Bradford 1976). The salt was removed by dialysis buffer (5 mM DTT, 1 mM EDTA, 100 mM NaCl, 50 mM Tris-HCL pH 7.5, 0.1%V/V Triton X-100, 50% Glycerol, 5% W/V Sucrose, 4 mM L-Arginine, and protease inhibitor cocktail) for 24 h.

Analysis of rnase H and reverse transcriptase activity of

RT-L and RT-K rnase H activity assay

The RNase H inactivity of rRT-K compared to rRT-L was properly assayed by the modified RT reaction in analytical techniques such as one-step RT-PCR and RT-LAMP. In this method, the SARS-CoV-2 RNA was converted to cDNA in the RT reactions as a template for rRT-L and rRT-K enzymes. The prepared cDNA was incubated for 30 min at 37 °C with DNase-I to degrade the first-strand DNA in the DNA-RNA hybrid. The treated cDNA was used in the one-step RT-PCR and RT-LAMP reactions and the product was visualized by electrophoresis and colorimetric way, respectively.

RT assay

The thermostable activity of rRT-L, rRT-K, and commercial RT (Add bio, Germany) was compared in the RT-PCR and RT-LAMP reactions. Therefore, SARS-CoV-2 RNA was used to prepare cDNA by the spike-specific reverse primer in the RT reaction. The variation parameters of the RT reaction were the temperature (50–60 °C) and incubation time (15, 20, and 30 min). The six products of the RT reaction of each rRT and commercial RT were used and compared in the PCR and LAMP reactions.

RT-PCR assay

The polymerase activities of rRT-L and rRT-K proteins were evaluated and compared with those using a commercial RT Assay Kit (Add Bio-Germany). Each 20-µL reaction contained 10 µL of 2x RT Reaction buffer, 2 µmol/L of reverse primers, 0.2–2 µL of each recombinant RT protein, and 7 µL of purified SARS-CoV-2 RNA, or 7 µL of nuclease-free H₂O as non-template control (NTC). The reverse transcription was optimized to various parameters, including a range of temperatures (50–60 °C), incubation times (15, 20, and 30 min), and each rRT protein volume (0.2–4 ul). The PCR was performed with specific primers of the SARS-CoV-2 spike gene. The PCR program was as follows: 1 cycle at 95 °C for 5 min and 40 cycles including three steps at 95 °C for 30 s, 60 °C for 40 s, and 72 °C for 30 s, followed by a final extension step of 72 °C for 5 min. All reactions were done in a PCR

system (ABI, USA), and results were analyzed from three technical repeats. The amplicons of respective fragments were analyzed via gel electrophoresis.

RT-LAMP assay

For the polymerase activity of recombinant RT proteins by the RT-LAMP analysis, the assays were carried out in triplicate and total reaction volumes of 10 μ L. Briefly, 10 μ L reactions were prepared to contain 1 μ L of 10x Bst reaction buffer; 6 mM MgSO₄; 1.4 mM each dATP, dGTP, dCTP, and dUTP; 1.6 μ M each of primers of forward inner primer (FIP) and backward inner primer (BIP); 0.2 μ M each of primers of F3 and B3; 0.4 μ M each of forward loop (FL) and backward loop (BL) primers; 0.8 M betaine; 0.2 μ L (1.6U) Bst3 (New England Biolabs, UK); 0.1 μ L (0.1U) UNG (Biotech rabbit, Germany); 0.2 μ L of each recombinant RT proteins, and 1 μ L of purified SARS-CoV-2 RNA, or 1 μ L of nuclease-free H₂O as NTC. After adding 2 μ L of SYBR green I diluted 1:100 in DMSO (Invitrogen-US) into the cap of each sample vial, the reaction mixture was set at 25 °C for 5 min and three different temperatures (55 °C, 60 °C, and 63.5 °C) for 60 min on a water bath, followed by an inactivation step of 85 °C for 5 min. Visualization of the product was detected by electrophoresis and by direct colorimetric ways too.

Statistical analyses

The SAS V.9.1 software was used for data analysis. To compare protein expression levels in the different conditions, full model two-way ANOVA was designed on the different factors including recombinant enzymes, strains, temperatures, and post-induction times. Also, the effect of interactions among treatments was computed. Multiple mean comparisons were done using the Tukey test and a p -value < 0.01 was usually chosen.

Results

In-silico analysis

The codon adaptation index (CAI) for the optimized RT-L and RT-K was 0.99. The average GC content of the RT-L and RT-K were 59.03 and 58.98. % respectively, which were the acceptable ranges and it showed the amount of energy to relax the secondary structures (Fig. 2A). Also, the negative CIS-elements of RT-L and RT-K sequences (CDS) including RNase E site, destabilizing, and false transcription termination sites were removed. The secondary structure stability of mRNA was -570.21 kcal/mol before optimizing and it was calculated -652.27 kcal/mol and -628.07 kcal/mol after optimizing for RT-L and RT-K mRNAs, respectively (Fig. 2B). Moreover, the sequence of 14 nucleotides upstream to the start codon was redesigned to the TAAGGAGGT TTTTT sequence in the pET9a vector. The 16 S rRNA

docking stability with the redesigned RBS was increased ($\Delta G = -14.87$ kcal/mol) compared to commercial RBS ($\Delta G = -8.9$ kcal/mol) (Fig. 2C). The solubility value of the rRT-K enzyme (0.306) was more than rRT-L (0.263) (Fig. 2D).

Homology modeling and docking

The 3D structure model of the mutant RT was done by the I-TASSER server. The model with a high score was selected for the other in-silico analysis. After refining by UCSF chimera1.16 and SPDB viewer software, the ERRAT server indicated the overall quality factor of the selected model was 96.82 in the primary model. In the refined model, 90.5% of residues were in the most favored region, 8.7% in the allowed region, and 0.7% in the disallowed region. Our analysis of H-bond contacts of RT-K and RT-L revealed 7 and 10 H-bond interactions with RNA templates, respectively. Furthermore, non-bonded contacts of RT-K and RT-L with RNA template were 54 and 16 contacts, respectively. The binding affinity between RT-L and RT-K with RNA template was calculated $\Delta G = -274.68$ kcal/mol and -289.57 kcal/mol, respectively (Fig. 3).

Effect of the directed mutation on the rnas H domain stability

In-silico analysis (IUpred 3.0) showed that substitution of the five hydrophobic residues (438-442aa) within the end region of the connection domain which helps the RNase H activity, to hydrophilic residues caused to form disorder region or instability in this region of rRT-K compared to rRT-L. The probability of this instability was calculated as 0.62 by IUpred 3.0 for rRT-K whereas it was 0.25 for rRT-L. The threshold of disorder probability is 0.5. Also, a point mutation of D530R in the active site of the RNase H domain increased the disorder probability to 0.5 in rRT-K compared to rRT-L which was 0.49 (Fig. 4).

SDS-PAGE and western blotting analysis

The SDS-PAGE result showed that the expression level of rRT-L and rRT-K were increased under 0.5 and 0.1 mM IPTG induction compared to 1 mM IPTG. Between two different growth temperatures (20 °C and 37 °C), the high yield of rRT-L and rRT-K proteins was observed in 0.5 mM IPTG concentration at 37 °C for both strains (Figs. 5A, B and 6). The post-induction time analysis showed that the most rRT-L and rRT-K were produced at 8 and 6 h, respectively (Figs. 5C, D and 6). Generally, ANOVA analysis of variance depicted that among treatments, just the effect of strains was not significant but the other factors including enzymes, temperature, post-induction times, and their interactions were significant at p value < 0.01 (Table S1). Consequently, the optimized yield of unpurified rRT-L and rRT-K proteins

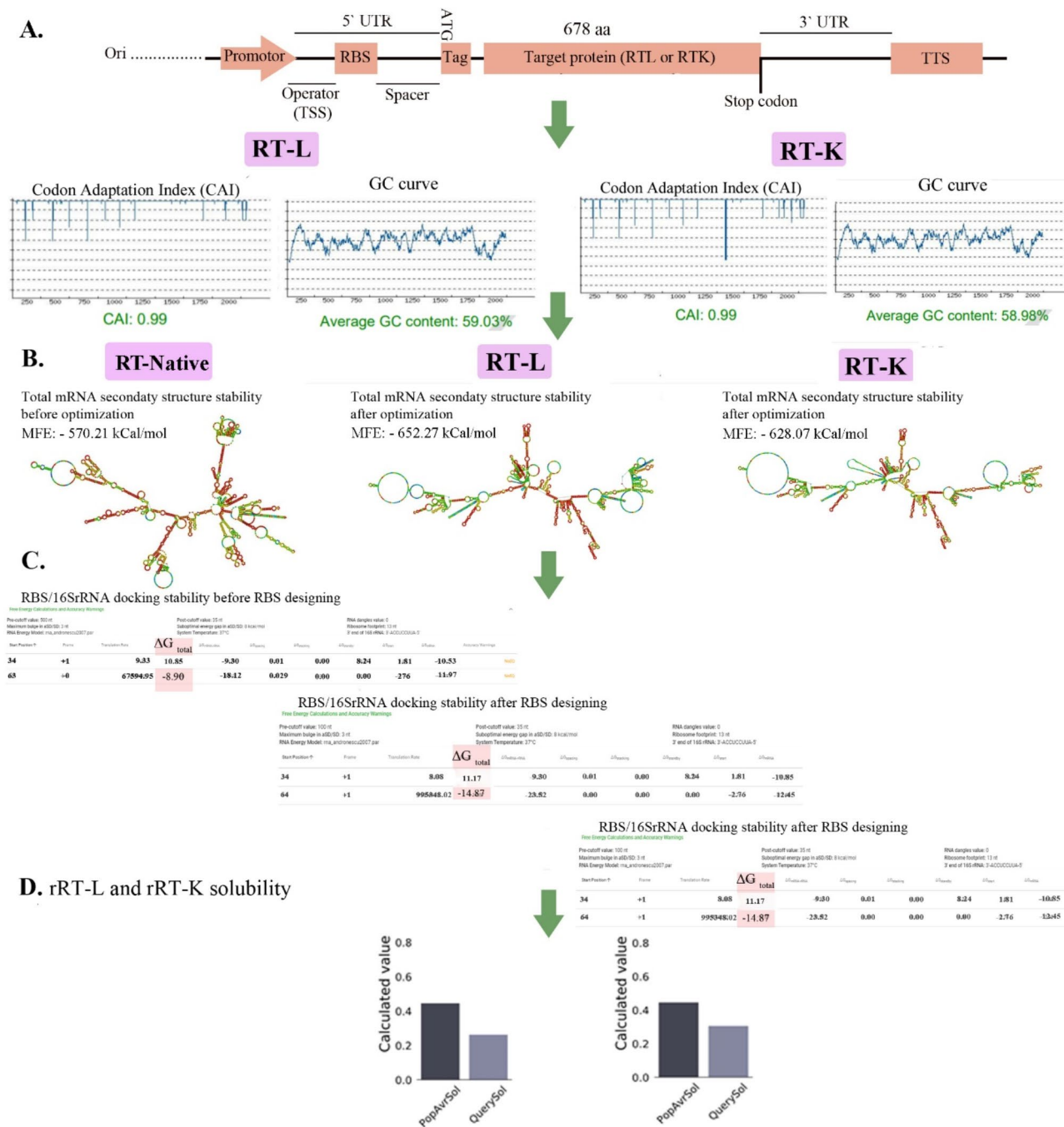


Fig. 2 Sequence optimization of native and mutant RT genes. **A** The optimized nucleic acid sequences were cloned in *Xba*I and *Bam*HI sites of pET-9a. **B** The whole minimum free energy of the secondary structures, codon adaptation index (CAI), and GC content of non-optimized and optimized RT-K and RT-L messenger RNA. **C** The minimum free energy of 16 S rRNA docking with ribosome binding site (RBS) before and after RBS designing. **D** The probability of rRT-K and rRT-L solubility

was 596.5 mg L⁻¹ and 639.65 mg L⁻¹, respectively. Also, immunoassay of unpurified and purified rRT-K and rRT-L by HRP Goat anti-mouse and anti-his monoclonal antibody depicted an intense brown color reaction with the protein size corresponding to ~ 75 kDa for rRT-L and rRT-K (Figs. 5E and 7C).

Purification of rRT-L and rRT-K proteins by Ni-NTA chromatography

Our expression results revealed that the rRT-L and rRT-K proteins were expressed more in inclusion body parts compared to the cytosol of *E. coli*. The optimized concentration of imidazole to elute the rRT-L and rRT-K proteins from the Ni-NTA column was 200mM and

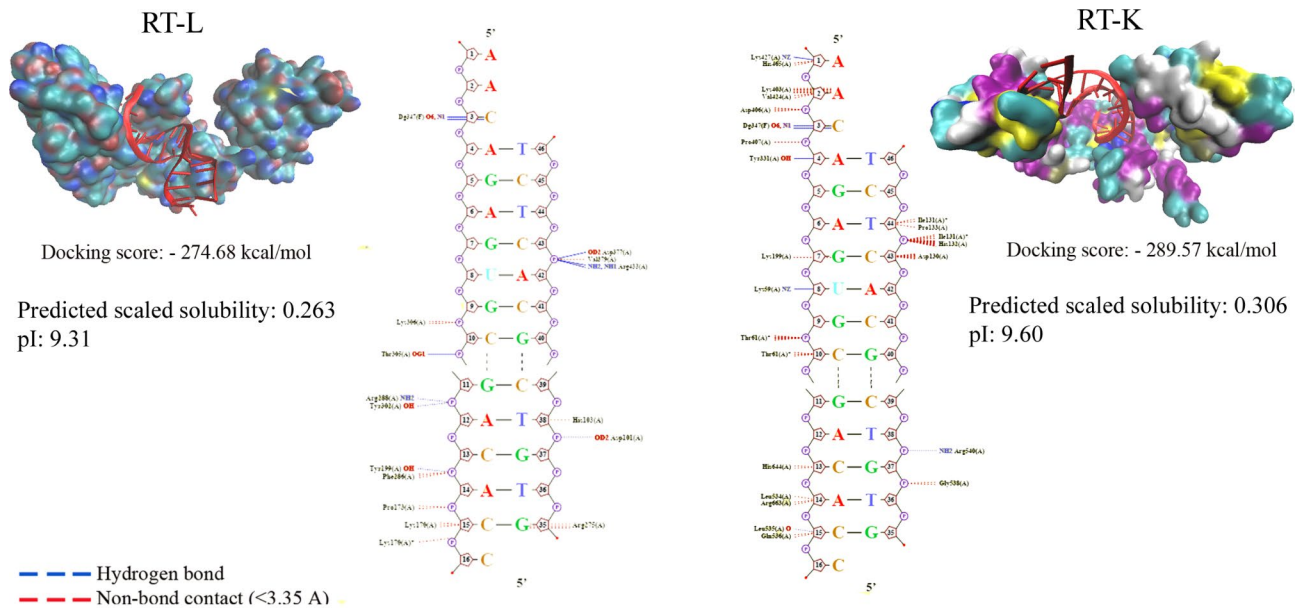


Fig. 3 Molecular docking between rRT-L and rRT-K and Sars-CoV-2 RNA as receptor and ligand, respectively, using the HDock server

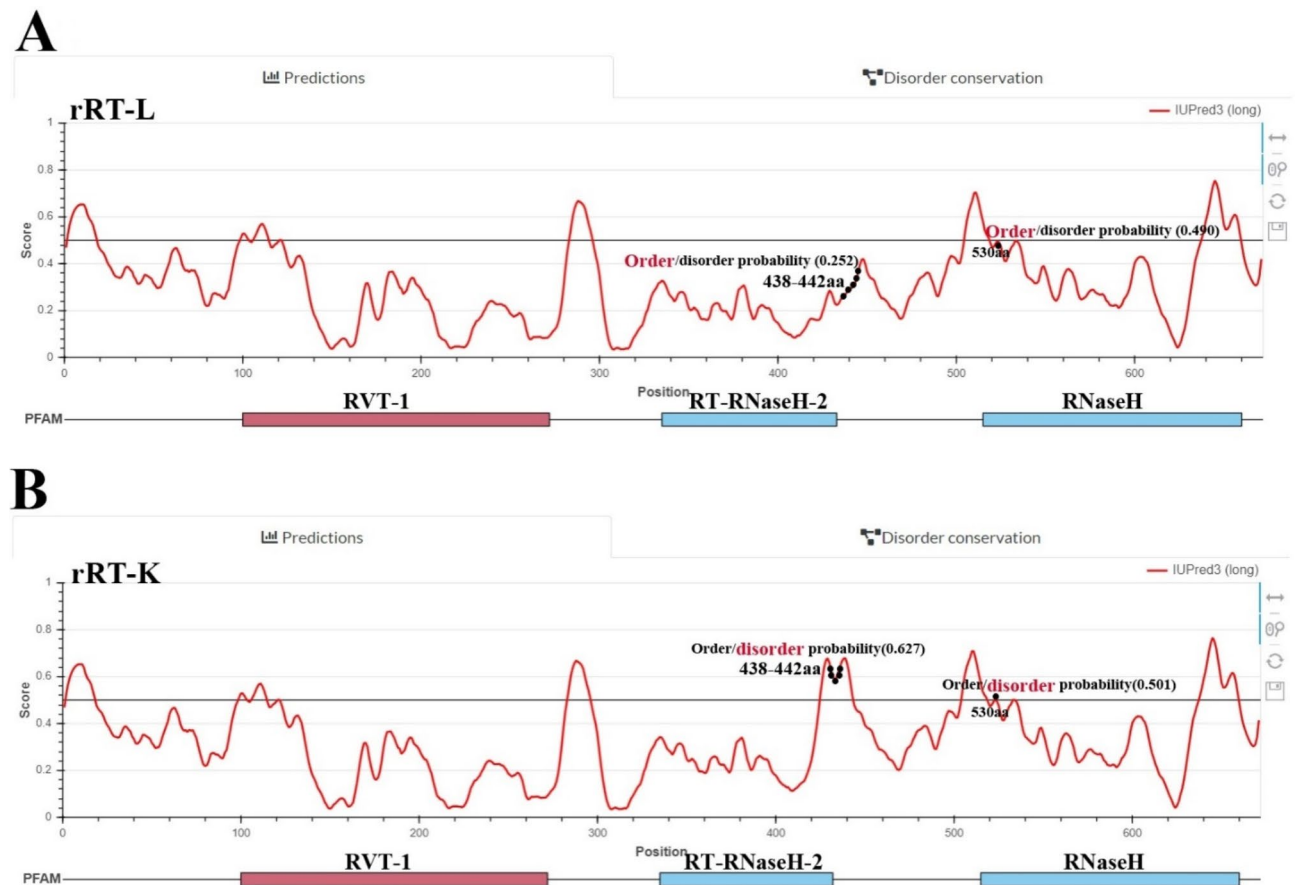


Fig. 4 The effect of site point-directed mutations on the structure stability (dis-order) of the recombinant RT-L. Prediction of structural stability before mutation (rRT-L) (A) and after mutation (rRT-K) (B)

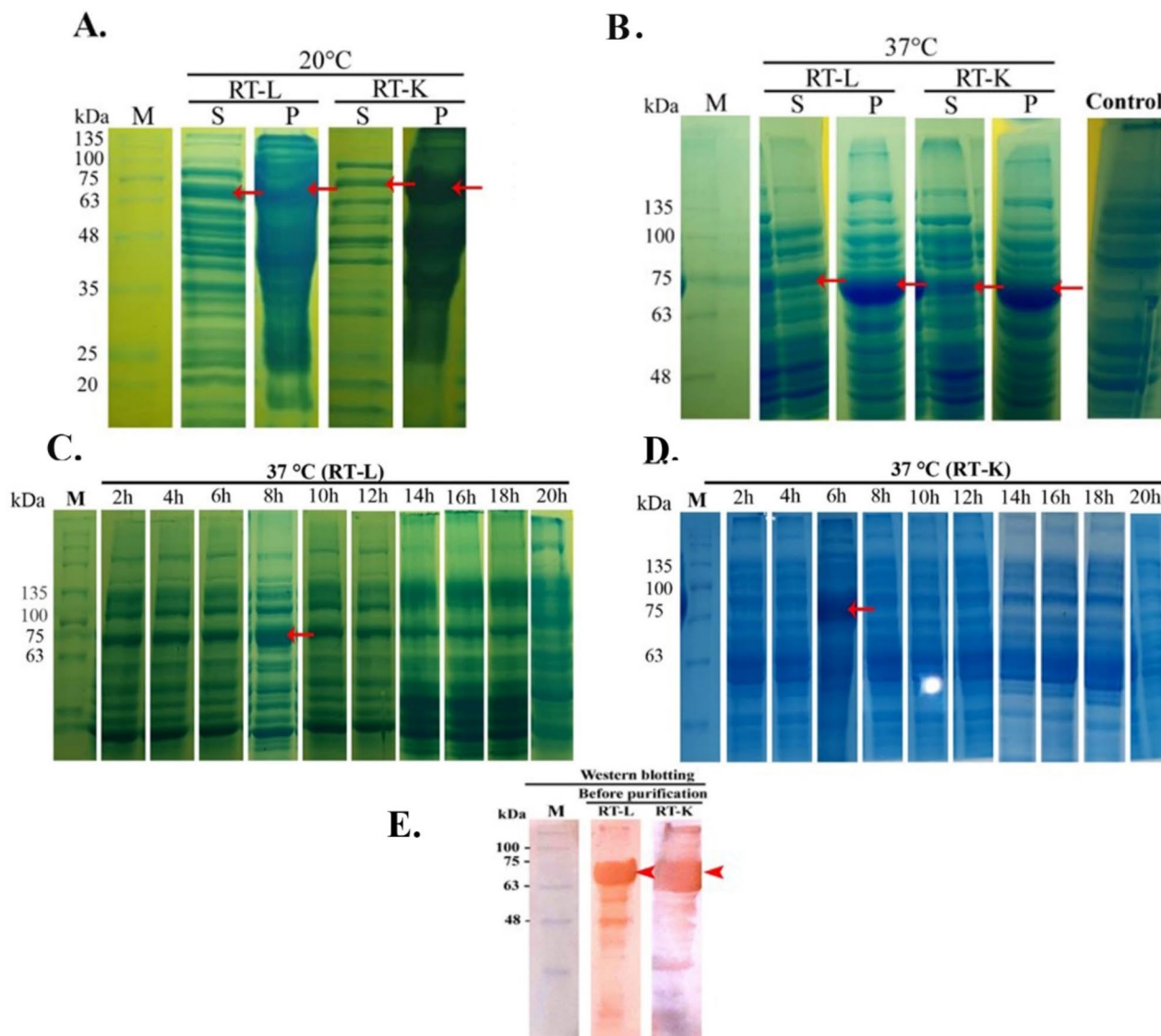


Fig. 5 **A, B** Coomassie-blue stained SDS-PAGE of rRT-L and rRT-K expression at 20 °C and 37 °C. Protein size markers (10–100 kDa) (Lane 1); supernatant fraction of cells lysate of *E. coli* BL21-rRT-L (Lane 2); pellet fraction of cells lysate of *E. coli* BL21-rRT-L (Lane 3); supernatant fraction of cells lysate of *E. coli* BL21-rRT-K (Lane 4); pellet fraction of cells lysate of *E. coli* BL21-rRT-K (Lane 5); **B** negative control *E. coli* BL21 (Lane 6). **C, D** pellet fraction of cells lysate of *E. coli* BL21 post-induction at 2, 4, 6, 8, 10, 12, 14, 16, 18, 20 h for both rRT-L and rRT-K. **E** Immunoblotting of unpurified recombinant proteins results using the anti-His monoclonal antibody. Protein molecular weight marker (10–100 kDa) (Lane 1); before purification of rRT-L (Lane 2); before purification of rRT-K (Lane 3)

250 mM in non-denaturing conditions (Fig. 7A). The optimized pH to elute the rRT-K and rRT-L proteins was 3–4.5 in the denaturing condition (Fig. 7B). The yield of purified rRT-L and rRT-K proteins was 395.55 mg L⁻¹ and 430.60 mg L⁻¹, respectively. The western blot analysis approved the purified rRT-L and rRT-K proteins (Fig. 7C).

Rnase H deficiency of rRT-K

The RNase H inactivity of rRT-K was proved by the modified RT reaction. In-silico analysis of RNase H active site sequence by the Phyre Investigator tool showed that the

substitution of D530, E568, and D659 of rRT-L to R530, G568, and R659 amino acids in rRT-K, respectively, could inactivate the RNase H domain but could not change its structure (Fig. 8A and B). In Fig. 8B (in each mutation profile), the 20 possible amino acid types are labeled along the x-axis with their one-letter code. The long colored bars indicate the probability that a mutation to the corresponding residue will harm the RNase H activity. The site active residues D530, E568, and D659 in the rRT-L sequence (indicated by the black arrow key), we can see that many mutations are likely to affect RNase H function, the highest likelihood coming from mutations

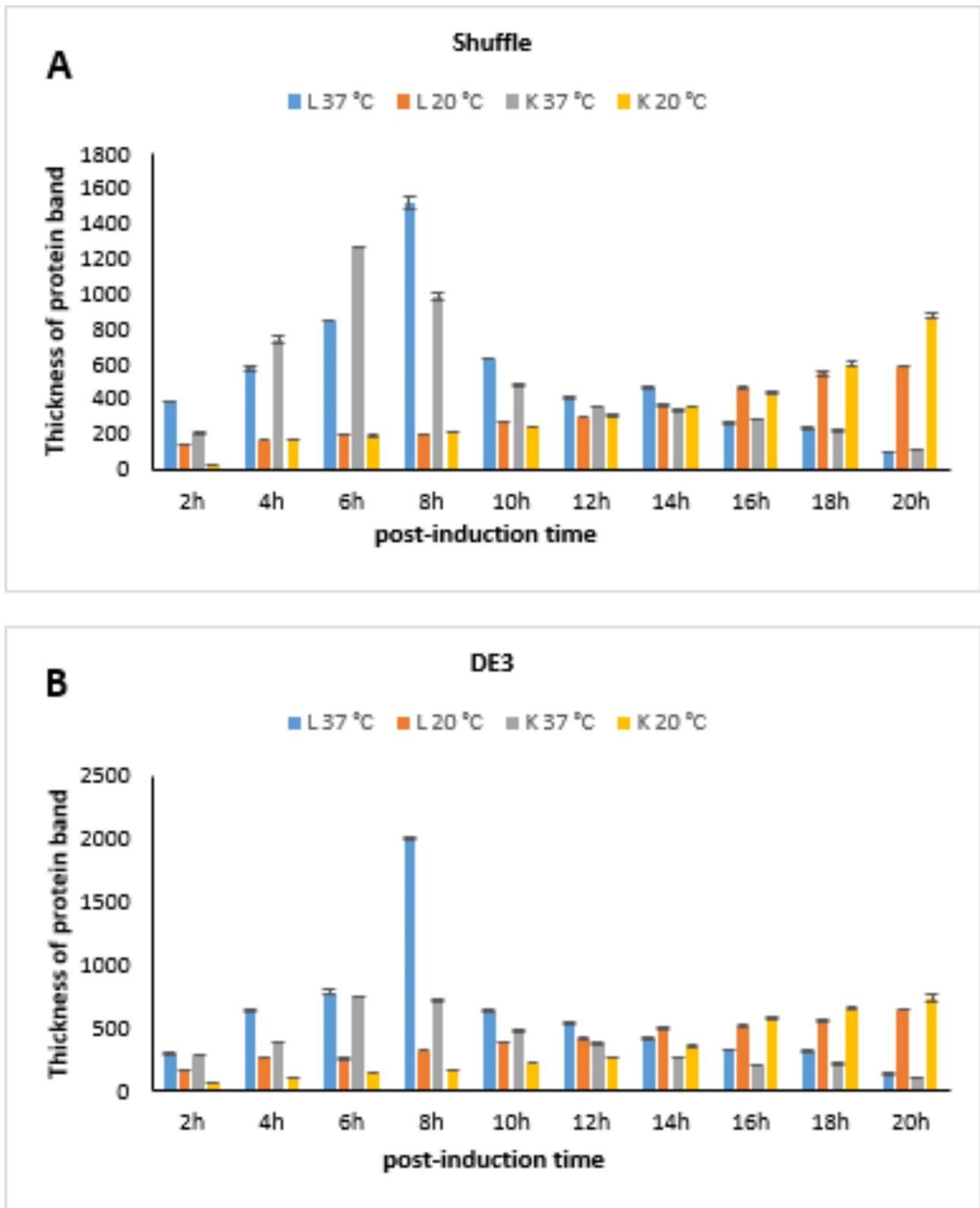


Fig. 6 Effect of temperature, and post-induction time on expression of the recombinant RT-L and RT-K enzymes in strain Shuffle (A) and DE3 (B). The error bars represent mean ±SD

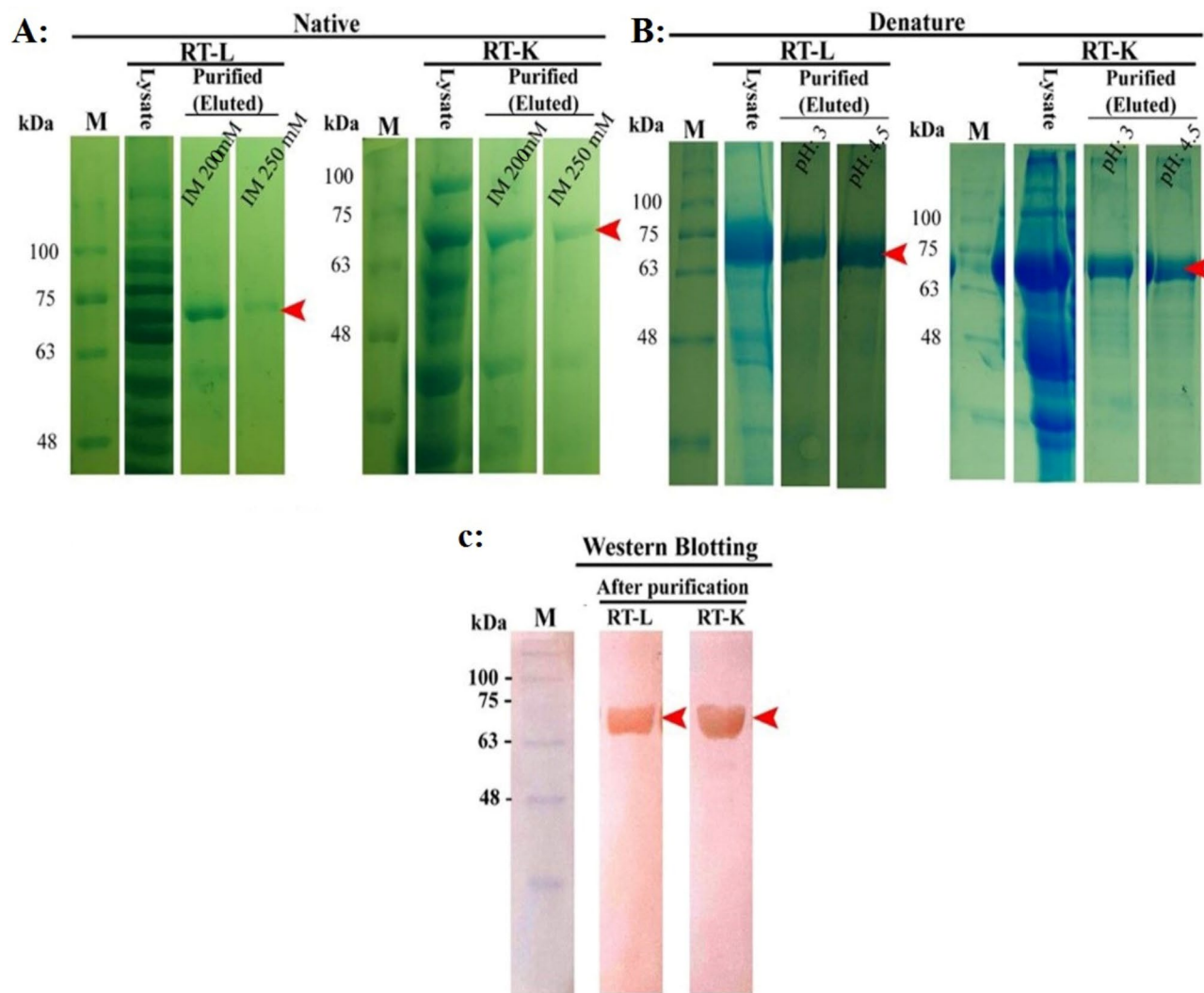


Fig. 7 Purification of rRT-L and rRT-K. **A** Purified rRT-L and rRT-K samples with the native method were separated on 12% SDS-PAGE. Protein molecular weight marker (10–100 kDa) (Lane 1); unpurified lysate (Lane 2), eluted phase contains imidazole 200mM (Lane 3), eluted phase contains imidazole 250mM (Lane 4). **B** Purified rRT-L and rRT-K samples with the denature method. Protein molecular weight marker (10–100 kDa) (Lane 1); unpurified lysate (Lane 2), eluted phase in pH3 (Lane 3), eluted phase in pH4.5 (Lane 4). **C** Immunoblotting of purified recombinant proteins results using the anti-His monoclonal antibody. Protein molecular weight marker (10–100 kDa) (Lane 1); after purification of rRT-L (Lane 2); after purification of rRT-K (Lane 3)

with long bars which are shown by orange arrow key. However, based on each sequence profile we can find the longest bar displayed by the green arrow key which is the preferred residue among the candidate mutants that does not affect the RNase H structure. These values are calculated by scanning the rRT-K sequence against a large sequence database using the iterative searching program PSI-Blast. Based on the sequence profile, the corresponding R, G, and R residues are the preferred residues in rRT-K. To validate the in-silico result, the SARS-CoV-2 RNA was converted to cDNA by both rRTs. Then, the DNA chain of the DNA-RNA hybrid was degraded by DNase-I, and the product was used as a template for the one-step RT-PCR and RT-LAMP reactions (Fig. 9A). This result showed that just the first RT reaction containing

the rRT-K enzyme could pass the second RT-PCR and RT-LAMP reactions, resulting in inactivity of the RNase H domain of rRT-K compared to rRT-L (Fig. 9B).

Thermostability assessment of rRT-L and rRT-K by RT-PCR and RT-LAMP assays

The analysis of the results of RT-PCR showed that RT-L enzymes and commercial kit face a decrease in multiplication by increasing the temperature of cDNA synthesis from 50 to 60 °C, while contrary to them, increasing the temperature led to an increase in the multiplication by the rRT-K enzyme. Also, the same results were observed for RT reaction polymerized with rRTs derived from *E. coli* cultured at 20 °C and 37 °C (Fig. 10A–D). In addition, the best temperature at which the RT-LAMP reaction

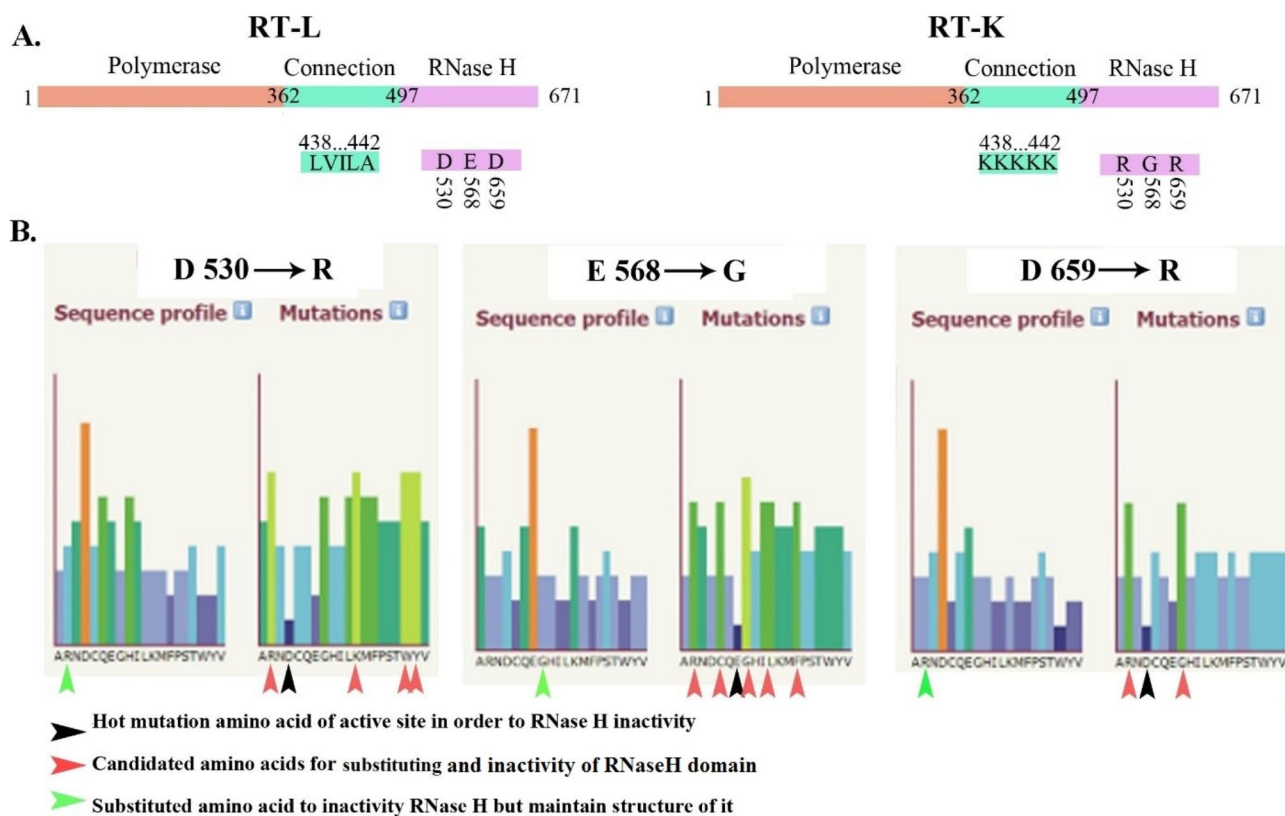


Fig. 8 Eight site-directed point mutations in the connection and RNase H domains. **A** Five (438–442aa) and three (530, 568, and 659 aa) point mutations at the RT connection and RNase H active site domains, respectively. **B** In-silico analysis of RNase H active site sequence by the Phyre Investigator tool. The substitution of D530, E568, and D659 of rRT-L to R530, G568, and R659 amino acids in rRT-K, respectively, could inactivate the RNase H domain but could not change its structure. Hot mutations amino acids (D530, E568, and D659) in the active site in the RNase H domain (Black arrow). The highest likelihood coming from mutations with long bars were candidates for substitutions (Orange arrow). Preferred residue (R530, G568, and R659) among the candidate mutants (Green arrow)

occurred was 55 °C, for the rRT-L enzyme whereas the rRT-K enzyme worked at all temperatures (Fig. 10E–G).

Discussion

MMLV reverse transcriptase is the key and well-known enzyme in the molecular biology and biotechnology field. Recombinant MMLV RT is the most widely used for cDNA synthesis, expression gene examination, mRNA, and RNA detection due to its robust polymerase activity and high fidelity. This enzyme showed good catalytic activity but its disadvantages are to have RNAase H activity, resulting in RT thermo-instability as well as insolubility (Costa et al. 2013; Fei et al. 2012; Nishimura et al. 2015; Yasukawa et al. 2008; Zajac et al. 2013). High-temperature reverse transcription helps to prevent primer dimerization, melt RNA secondary structure, and improve the specificity of binding especially if gene-specific primers are used (Bustin 2009). Also, it is necessary to use thermo-stable reverse transcriptase in one-step isothermal reactions such as RT-LAMP. To solve these problems, introducing directed mutation at specific positions is needed (Fei et al. 2012; Yasukawa et al. 2010). The

specific mutations were done in the native RT sequence to increase the solubility and thermostability of the enzyme and the thermostable polymerase activity and ribonuclease H (RNase H) inactivity of recombinant enzymes were proved by the RT-PCR and RT-LAMP assays. In the present study, five hydrophobic residues within the connection domain of rRT-K were substituted with hydrophilic residues, increasing its thermostability, and the same result was reported by Fei et al. (2012).

Another hypothesis is that the inactivity of the RNase H domain can increase the thermostability of the RT domain (Katano et al. 2016; Mizuno et al. 2010; Narukawa et al. 2021b). Thus, novel explanation was done to incorporate the specific three mutations (D530R, E568G, and D659R) in the RNase H domain to diminish its activity. These mutations were designed based on in-silico algorithms that predicted the best hot mutation in the active site of the RNase H domain without any effect on its structure which is needed for the RT domain activity (Kelley et al. 2015). On the other hand, IUPred analysis verified the aforementioned in-silico data that the directed mutations in the end region of connection and

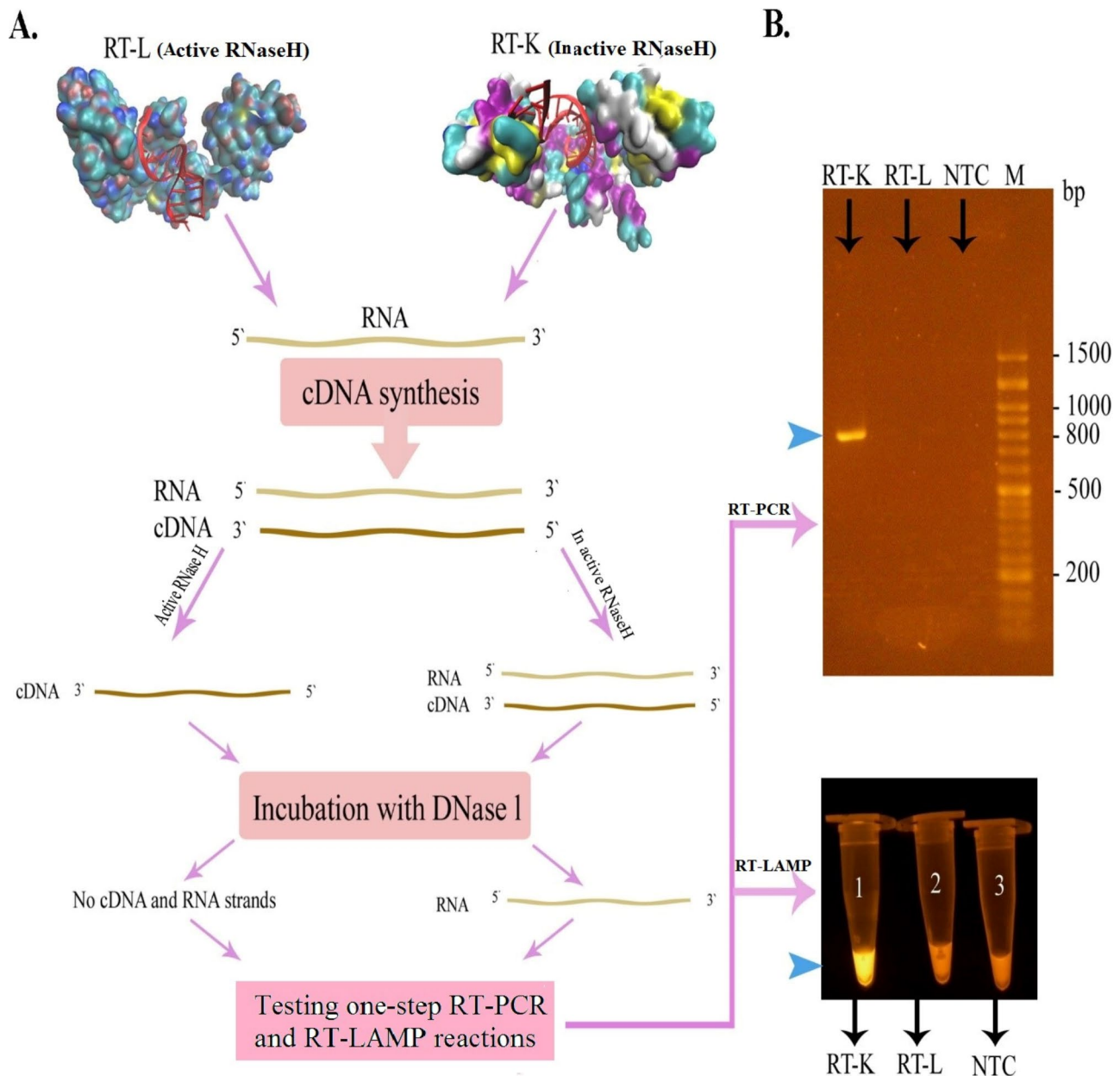


Fig. 9 RNase H activity assay. **A** The RNase H inactivity of rRT-K compared to rRT-L was assayed by the modified RT method. The SARS-CoV-2 RNA was converted to cDNA in the RT reactions as a template for rRT-L and rRT-K enzymes. The prepared cDNA was incubated for 30 min at 37 °C with DNase-I to degrade the first-strand DNA in the DNA-RNA hybrid. **B** The treated cDNA was used in the one-step RT-PCR and RT-LAMP reactions and the product was visualized by electrophoresis and colorimetric way, respectively

active site of RNase H domains changed the ordered structure of these regions to the disorder regions which caused instability in the RNase H domain of rRT-K compared to rRT-L. Recently, the effect of codon bias and its derived disorder regions on protein folding and function in prokaryotes and eukaryotes has also been reported (Liu et al. 2022; Zhou et al. 2015; Lyu and Liu 2020; Xu et al. 2013).

Confirmation of the in-silico data was proved by the modified RT reaction. Molecular docking was performed

to calculate the binding affinity of RNA with the rRT-K and rRT-L. Our in-silico experiments revealed that the three-dimensional (3D) structure of recombinant RT with specific mutations has a positive effect on the binding affinity of the RNA/ rRT-K complex. The binding energy for RNA/rRT-K was slightly higher (−289.57 kcal/mol) than that of RNA/rRT-L (−274.68 kcal/mol). Katano and coworkers (2016) showed the incorporation of four mutations (E286R/E302K/L435R/D524A) in the RT gene (variant MM4) by site-directed mutagenesis could be

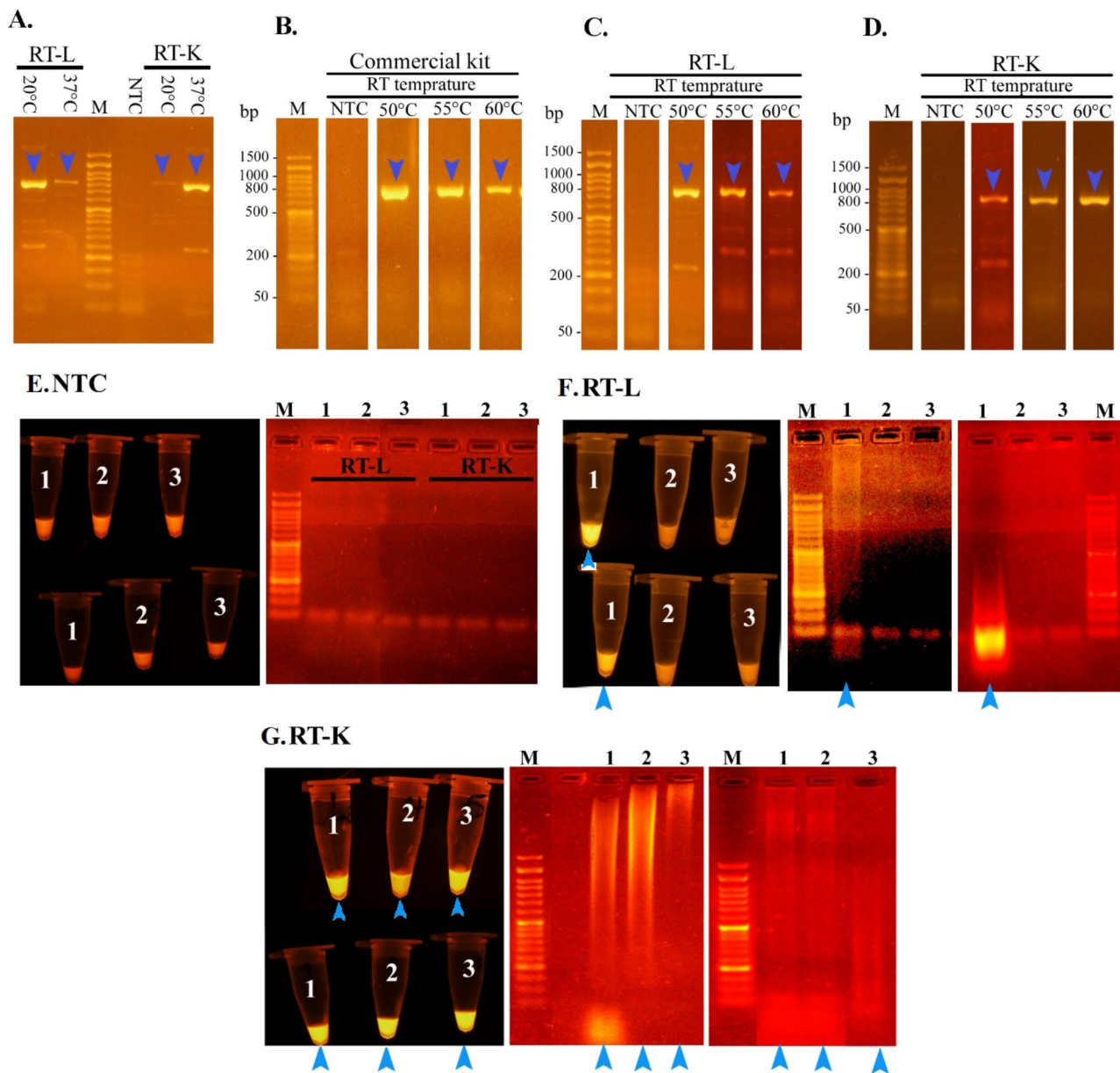


Fig. 10 Effects of thermal treatment on the DNA polymerase activity of rRTs and their related RT-PCR as well as one-step RT-LAMP results. **A** Analysis of RT-PCR results of rRT-L and rRT-K enzymes derived from *E. coli* at culture temperatures of 20 and 37 °C. Lane 1,2: RT-PCR product of rRT-L at 20 and 37 °C culture temperature; Lane 3: Marker 100 bp; Lane 4: negative control (without RNA template); Lane 4,5: RT-PCR product of rRT-K at 20 and 37°C culture temperature. **B–D** The electrophoretic pattern of RT-PCR reactions was carried out with commercial RT, purified rRT-L, and rRT-K. Lane 1: Marker 100 bp; Lane 2: negative control (without RNA template); Lane 3,4,5: RT-PCR product at 50 °C, 55 °C, and 60 °C, respectively. **E–G** The one-step RT-LAMP product of the negative control sample (without RNA template), purified SARS-CoV-2 RNA derived from the patient's swab was incubated with rRT-L, and rRT-K, respectively. The tubes and lanes numbered 1,2, and 3 point to one-step RT-LAMP product at 50 °C, 55 °C, and 60 °C, respectively. The positive results are marked with a blue arrow in the electrophoresis and colorimetric patterns

related to the increased thermal stability. They suggest that MM4 was more thermostable than wild-type and retained DNA polymerase activity when incubated at 60 °C for 10 min (Katano et al. 2016). Also, we hypothesized that the novel mutations including R530, G568, and R659 designed in this study caused to increase in the thermostable activity of rRT-K compared to

rRT-L. Because, rRT-L and commercial enzymes could not polymerize cDNA strands more than 55 °C, but rRT-K enzyme was successful at 60 °C.

Also, the cost-benefit is the versatile sequence optimization to improve the recombinant enzyme yield. Because of this, the sequence optimization including the redesigning of pET9a RBS, codon and codon pair

optimization, and secondary structure of mRNA was done. The amount of rRT expression that was reported in previous studies in *E. coli*, ranged from 2 to 23 mg L⁻¹ at shake flask condition (Nuryana et al. 2022; Rao et al. 2017). However, due to versatile optimization, the yield of rRT-L and rRT-K in this study increased to 395.55 mg L⁻¹ and 430.60 mg L⁻¹, respectively. In order to avoid secondary structure formation in the RBS site, we redesigned 14 nucleotides upstream of the start codon, resulting in the increasing interaction between 16 S rRNA and RBS site, and the loose secondary structure of the RBS region can be easily translated by ribosome machinery and increase the yield of recombinant protein. The minimum free energy of 16 S rRNA docking with RBS was calculated $\Delta G = -8.9$ kcal/mol and $\Delta G = -14.87$ kcal/mol before and after RBS designing of both RT-L and RT-K constructs, respectively. Also, the two rRTs coding sequences (CDS) were optimized to have weak secondary structures with 50 nucleotides between hairpins. The effect of these parameters on recombinant protein expression has been reported by other studies (Masoudi et al. 2021; Salis et al. 2009). Different factors such as temperature, IPTG concentration, and post-induction time maximize the production of a foreign protein in *E. coli* (Chen et al. 2009; Nuryana et al. 2022; Rao et al. 2017). We provided evidence that the optimization of IPTG concentration (0.5 mM) and temperature (37 °C) increased both rRT production. SDS-PAGE analysis revealed that the yield and activity of rRTs expressed in the Shuffle strain was more than BL21(DE3) to agree with other studies (Ahmadzadeh et al. 2020; Lobstein et al. 2012). In the present study, we focus on the utility of rRT-K, but the value of rRT-L with improved expression is also substantial. For example, in reactions that still require RNA cleavage in the RNA/DNA hybrid, rRT-K cannot be used but rRT-L will be more relevant. This study revealed that the directed mutations and the versatile sequence optimization can promise to produce thermostable commercial enzymes to decrease non-specific products of one-step RT-PCR and RT-LAMP reactions and produce cheaper RT enzymes.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13568-024-01773-6>.

Supplementary Material 1

Acknowledgements

The authors acknowledge the financial support of Golestan University of Medical Sciences which was supported by a grant from the Golestan University of Medical Science (Grant number: 111992).

Author contributions

All authors give final approval of the manuscript to be submitted. MD was involved in concept, methodology, statistical analysis, and writing the

manuscript. AY was a co-supervisor and involved in the concept, sequence optimization, in silico designing, statistical analysis, validation, and writing the manuscript. HRN was the advisor involved in the methodology and manuscript writing. AM was the advisor and was involved in the validation and editing of the manuscript. AT was a co-supervisor and was involved in concept, material support, and validation.

Availability of data and materials

The authors declare that all the data supporting the findings of this study are available within the paper, and its supplementary information is available from the corresponding author upon request.

Declarations

Ethics approval and consent to participate

IR.GOUMS.REC.1400.349.

Competing interest

The authors declare that they have no conflicting interests.

Received: 21 August 2024 / Accepted: 19 September 2024

Published online: 03 October 2024

References

- Ahmadzadeh M, Farshdari F, Nematollahi L, Behdani M, Mohit E (2020) Anti-HER2 scFv expression in *Escherichia coli* SHuffle® T7 express cells: effects on solubility and biological activity. *Mol Biotechnol* 62:18–30
- Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72(1–2):248–254
- Bustin SA (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* 55:4611–4622
- Chen Y, Xu W, Sun Q (2009) A novel and simple method for high-level production of reverse transcriptase from Moloney murine leukemia virus (MMLV-RT) in *Escherichia coli*. *Biotechnol Lett* 31:1051–1057
- Costa C, Giménez-Capitán A, Karachaliou N, Rosell R (2013) Comprehensive molecular screening: from the RT-PCR to the RNA-seq. *Transl Lung Cancer Res* 2(2):87
- Coté ML, Roth MJ (2008) Murine leukemia virus reverse transcriptase: structural comparison with HIV-1 reverse transcriptase. *Virus Res* 134(1–2):186–202
- Das D, Georgiadis MM (2001) A directed approach to improving the solubility of Moloney murine leukemia virus reverse transcriptase. *Protein Sci* 10(10):1936–1941
- Erdős G, Pajkos M, Dosztányi Z (2021) IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. *Nucleic Acids Res* 49(W1):W297–W303
- Fei X, Xuemei M, Xiansong W (2012) Soluble expression and purification of histidine-tagged Moloney murine leukemia virus reverse transcriptase by Ni-NTA affinity chromatography. *Affin Chromatogr* 17:357–368
- Hanahan D, Jessee J, Bloom FR (1991) [4] Plasmid transformation of *Escherichia coli* and other bacteria method enzymol, vol 204. Elsevier, Amsterdam, pp 63–113
- Hu W-S, Hughes SH (2012) HIV-1 reverse transcription. *CSH Perspect Med* 2(10):a006882
- Katano Y, Hisayoshi T, Kuze I, Okano H, Ito M, Nishigaki K, Takita T, Yasukawa K (2016) Expression of moloney murine leukemia virus reverse transcriptase in a cell-free protein expression system. *Biotechnol Lett* 38:1203–1211
- Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10(6):845–858
- Konishi A, Ma X, Yasukawa K (2014) Stabilization of Moloney murine leukemia virus reverse transcriptase by site-directed mutagenesis of surface residue Val433. *Biosci Biotechnol Biochem* 78(1):75–78
- Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227(5259):680
- Levin JG, Crouch R, Post K, Hu S, McKelvin D, Zweig M, Court D, Gerwin B (1988) Functional organization of the murine leukemia virus reverse transcriptase: characterization of a bacterially expressed AKR DNA polymerase deficient in RNase H activity. *J Virol* 62(11):4376–4380

- Liu K, Ouyang Y, Lin R, Ge C, Zhou M (2022) Strong negative correlation between codon usage bias and protein structural disorder impedes protein expression after codon optimization. *J Biotechnol* 343:15–24
- Lobstein J, Emrich CA, Jeans C, Faulkner M, Riggs P, Berkmen M (2012) SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microb Cell Fact* 11:1–16
- Lyu X, Liu Y (2020) Nonoptimal codon usage is critical for protein structure and function of the master general amino acid control regulator CPC-1. *mBio* 11(5):e02605–e02620. <https://doi.org/10.1128/mBio.02605-20>
- Martín-Alonso S, Frutos-Beltrán E, Menéndez-Arias L (2021) Reverse transcriptase: from transcriptomics to genome editing. *Trends Biotechnol* 39(2):194–210
- Masoudi M, Teimoori A, Tabaraei A, Shahbazi M, Divbandi M, Lorestani N, Yamchi A, Nikoo HR (2021) Advanced sequence optimization for the high efficient yield of human group A rotavirus VP6 recombinant protein in *Escherichia coli* and its use as immunogen. *J Med Virol* 93(6):3549–3556
- Mizuno M, Yasukawa K, Inouye K (2010) Insight into the mechanism of the stabilization of Moloney murine leukaemia virus reverse transcriptase by eliminating RNase H activity. *Biosci Biotechnol Biochem* 74(2):440–442
- Narukawa Y, Kandabashi M, Li T, Baba M, Hara H, Kojima K, Iida K, Hiyama T, Yokoe S, Yamazaki T (2021a) Improvement of Moloney murine leukemia virus reverse transcriptase thermostability by introducing a disulfide bridge in the ribonuclease H region. *Protein Eng Des Sel* 34
- Narukawa Y, Kandabashi M, Li T, Baba M, Hara H, Kojima K, Iida K, Hiyama T, Yokoe S, Yamazaki T (2021b) Improvement of Moloney murine leukemia virus reverse transcriptase thermostability by introducing a disulfide bridge in the ribonuclease H region. *Protein Eng Des Sel* 34:gzab006
- Nishimura K, Yokokawa K, Hisayoshi T, Fukatsu K, Kuze I, Konishi A, Mikami B, Kojima K, Yasukawa K (2015) Preparation and characterization of the RNase H domain of Moloney murine leukemia virus reverse transcriptase. *Protein Express Purif* 113:44–50
- Nuryana I, Laksmi FA, Agustriana E, Dewi KS, Andriani A, Thontowi A, Kusharyoto W, Lisdiyanti P (2022) Expression of codon-optimized gene encoding murine Moloney leukemia virus reverse transcriptase in *Escherichia coli*. *Protein J* 41(4–5):515–526
- Pinto FL, Lindblad P (2010) A guide for in-house design of template-switch-based 5' rapid amplification of cDNA ends systems. *Anal Biochem* 397(2):227–232
- Rao S, Sundararajan S, Padmanabhan S (2017) A Novel protein refolding protocol for solubilization and purification of the catalytic fragment of recombinant MMuLV reverse transcriptase overexpressed in *Escherichia Coli*. *IJISRT* 2(9):171–177
- Salis HM, Mirsky EA, Voigt CA (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* 27(10):946–950
- Tanese N, Goff SP (1988) Domain structure of the Moloney murine leukemia virus reverse transcriptase: mutational analysis and separate expression of the DNA polymerase and RNase H activities. *Proc Natl Acad Sci USA* 85(6):1777–1781
- Xu Y, Ma P, Shah P, Rokas A, Liu Y, Johnson CH (2013) Non-optimal codon usage is a mechanism to achieve circadian clock conditionality. *Nature* 495(7439):116–120. <https://doi.org/10.1038/nature11942>
- Yamchi A, Rahimi M, Javan B, Abdollahi D, Salmanian M, Shahbazi M (2024) Evaluation of the impact of polypeptide-p on diabetic rats upon its cloning, expression, and secretion in *Saccharomyces Boulardii*. *Arch Microbiol* 206(1):37
- Yasukawa K, Nemoto D, Inouye K (2008) Comparison of the thermal stabilities of reverse transcriptases from avian myeloblastosis virus and Moloney murine leukaemia virus. *J Biochem* 143(2):261–268
- Yasukawa K, Mizuno M, Konishi A, Inouye K (2010) Increase in thermal stability of Moloney murine leukaemia virus reverse transcriptase by site-directed mutagenesis. *J Biotechnol* 150(3):299–306
- Zajac P, Islam S, Hochgerner H, Lönnerberg P, Linnarsson S (2013) Base preferences in non-templated nucleotide incorporation by MMLV-derived reverse transcriptases. *PLoS ONE* 8(12):e85270
- Zhou M, Wang T, Fu J, Xiao G, Liu Y (2015) Nonoptimal codon usage influences protein structure in intrinsically disordered regions. *Mol Microbiol* 97(5):974–987. <https://doi.org/10.1111/mmi.13079>

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.