



# Mutational analysis of catalytic site domain of CCHFV L RNA segment

Neha Kaushal<sup>1</sup> · Manoj Baranwal<sup>1</sup>

Received: 23 October 2022 / Accepted: 24 February 2023 / Published online: 6 March 2023  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

## Abstract

**Introduction** Crimean-Congo haemorrhagic fever virus (CCHFV) has tripartite RNA genome and is endemic in various countries of Asia, Africa and Europe.

**Method** The present study is focused on mutation profiling of CCHFV L segment and phylogenetic clustering of protein dataset into six CCHFV genotypes.

**Results** Phylogenetic tree rooted with NCBI reference sequence (YP\_325663.1) indicated less divergence from genotype III and the sequences belonging to same genotypes have shown less divergence among each other. Mutation frequency at 729 mutated positions was calculated and 563, 49, 33, 46 and 38 amino acid positions were found to be mutated at mutation frequency intervals of 0–0.2, 0.21–0.4, 0.41–0.6, 0.61–0.8 and 0.81–1.0 respectively. Thirty-eight highly frequent mutations (0.81–1.0 interval) were found in all genotypes and mapping in L segment (encoded for RdRp) revealed four mutations (V2074I, I2134T/A, V2148A and Q2695H/R) in catalytic site domain and no mutation in OTU domain. Molecular dynamic simulation and in silico analysis showed that catalytic site domain displayed large deviation and fluctuation upon introduction of these point mutations.

**Conclusion** Overall study provides strong evidence that OTU domain is highly conserved and less prone to mutation whereas point mutations recorded in catalytic domain have affected the stability of protein and were found to be persistent in the large population.

**Keywords** Phylogenetic tree · OTU · Catalytic domain · Molecular dynamic simulation · Mutation analysis

## Introduction

Tick-borne infections are emerging at a high rate due to globalization, climate change and migration. Crimean-Congo haemorrhagic fever (CCHF), a tick-borne disease, is predominantly caused by CCHF virus (CCHFV) and hyalomma ticks serve as vector for CCHFV. CCHF virus (CCHFV) is endemic in Africa, Asia and Europe [1]. One hundred forty outbreaks of CCHFV have been reported since 1967 [2] and 10,000–15,000 cases with 500 deaths are recorded every year [3]. In 1944, CCHFV outbreak was first reported in Crimean region of Eastern Europe. In India, CCHFV was first reported in Ahmedabad in 2011, and recently, it has been reported in Rajasthan, 2019 [4]. CCHFV exhibits 10–40% of fatality rate in humans [5].

Molecular clock and phylogenetic analysis of CCHFV infected patients in India revealed that 2010–2011 outbreak strain shares homology with Tajikistan [6] and 2015 outbreak strain exhibits homology with Afghanistan CCHFV strain [7]. These reports suggest that CCHFV has emerged in India due to migration of ticks from endemic regions (Afghanistan and Tajikistan) [6]. CCHFV has six genotypes (I–VI) classified based on their preferential geographical region [8]: genotypes I–III (belongs to Africa), genotype IV (Asia), genotype V (Europe) and genotype VI (Greece) [8, 9]. A survey was conducted for evaluating seroprevalence of CCHFV in domestic animals around the globe. They have found seropositive cattles (in 34 countries), sheep (25 countries), goats (15 countries) and camels (11 countries) whilst no seropositive animals were found in New Zealand, Australia, Germany and Netherland; thereby, no CCHFV cases have been reported from these countries [10] and concluded that CCHFV is most prevalent in areas with seropositive domestic animals.

CCHFV tripartite negative sense RNA genome consists of RNA polymerase (large segment, L), envelope glycoprotein (GP) (medium segment, M) and nucleocapsid protein (NP)

✉ Manoj Baranwal  
manoj.baranwal@thapar.edu

<sup>1</sup> Department of Biotechnology, Thapar Institute of Engineering and Technology, Patiala, Punjab 147004, India

(small segment, S) [11]. Viral RNA genome encapsulated by nucleoprotein forms ribonucleoprotein, which bind with RNA-dependent RNA polymerase (RdRp) and engage in viral replication [12] whereas GP interacts with ribonucleoprotein to mediate virus assembly [12]. Structural GPs encoded by M segment are responsible for viral attachment and formation of neutralizing antibodies [12]. L segment has two domains: ovarian tumour domain and catalytic site domain [13, 14]. Ovarian tumour domain has deubiquitination (DUB) activity and participates in evading host immune response which suggests DUB inhibitors can be designed for CCHF treatment [15]. Hawman et al. have reported three mutations (S2007N, V2686 and P3281L) in L segment of CCHFV by passaging of CCHFV Hoti strain in type I interferon deficient mice and reported that the mutated residues are highly conserved in all CCHFV genotypes [16].

Functions of proteins are primarily executed by their functional and structural elements termed as domains [17]. Zheng et al. reported that virus host interaction is primarily mediated by domains, as domains are most stable and conserved regions of protein [18]. Mutations can affect stability, virus host interaction and catalytic activity of proteins as it has been addressed in the literature that Q266L mutation in H7N9 strain of influenza virus enhances binding affinity of H7N9 strain with its receptors and also led to influenza pandemic [19]. Single amino acid mutant models of fibroblast growth factor receptor 1 involved in cancer have shown effect of mutation on interaction ability of proteins. Mutant models have shown a smaller number of hydrogen bonds in comparison to native models and can also destabilize the protein [20, 21]. Salinas et al. have reported that mutations can alter binding site pocket shape and its ligand identification properties [22]. Another study described enhancement of stability due to mutations (A49L and Q106T) in interleukin-4 [23]. Mutations in spike protein of SARS-CoV-2 reduce infectivity upon glycosylation at N331 and N343 and enhance infectivity upon D614G mutations whereas few mutations cause antibody resistance [24]. Molecular dynamic simulation study of effect mutation on binding affinity of SARS-CoV-2 with human ACE2 receptor described that all SARS-CoV-2 variants (delta plus, kappa, mu, iota, lambda and C.1.2) exhibit higher binding affinity than wild type (PDBid: 6MOJ) [25]. Considering these facts, the current work is aimed to assess the phylogenetic relationship and frequency of mutations in L segment of CCHFV around the globe, followed by identification of conserved spots in L segment and evaluation of effect of mutations on stability and function of domains by using *in silico* tools and molecular dynamic simulation study. Interestingly, phylogenetic analysis suggests that sequences belonging to same genotype have less divergence and genotype III sequences are more closely related to reference sequence as it also belongs to genotype III. Thereafter, mutation profiling revealed that OTU domain of CCHFV L

segment has no highly frequent mutations representing its conserved in nature whilst four highly frequent mutations (V2074I, I2134T/A, V2148A and Q2695H/R) were found in catalytic site domain. In molecular dynamic simulation study of wild and six mutant catalytic site domains, it was found that mutant catalytic site domain has shown large deviation and fluctuations over 50 ns simulation run.

## Methodology

### Data acquisition and phylogenetic analysis

Protein sequences of CCHFV for L segment with human host were downloaded from Virus Pathogen Resource Database (ViPRbrc) in FASTA format. Multiple sequence alignment (MSA) was carried out by Clustal W online server. Clustal W aligns sequence-by-sequence weighting, gap penalty and weight matrix [26]. Multiple sequence alignment executed by Clustal W was further taken as input for mutation analysis and building phylogenetic tree.

Mega X provides good platform for evolutionary studies [27]. Evolutionary analysis of protein sequence dataset was carried out by constructing phylogenetic tree by maximum likelihood with 1000 bootstrap replications using molecular evolutionary genetic analysis (Mega X) software. Tree was clustered according into six genotypes of CCHFV reported in the literature [9, 28].

### Mutation profiling

Mutations were recorded in all sequences across the length of L segment w.r.t. reference sequence (Accession No.: YP\_325663.1) using BioEdit 7.2 sequence alignment editor software [29]. BioEdit 7.2 software is a user-friendly software and entails various applications, such as amino acid or nucleotide variation identification, pairwise alignment, multiple sequence alignment and phylogenetic tree [29]. Mutation frequency was calculated as the number of sequences with mutated amino acid divided by the total number of sequences.

### Structural and solubility analysis

Solubility of proteins is crucial for assessment of protein homeostasis which is necessary for stability and function of proteins [30]. Solubility of mutant models was evaluated using SODA web server. SODA calculates difference in physicochemical properties of wild and mutant molecules. SODA prediction is based on five components: secondary structure propensity ( $\alpha$ -helix and  $\beta$ -sheets), aggregation, intrinsic disorder and hydrophobicity [31]. HOPE server was employed for assessment of effect of point mutation on structure and function of protein. HOPE (Have yOur Protein Explained)



and Nigeria, which belongs to genotype III. Therefore, USA strain of CCHFV was also classified as genotype III (Fig. 1).

## Mutation profiling

Phylogenetic analysis revealed divergence among sequences belonging to different genotypes; hence, mutation analysis was carried out. Amino acid mutations were recorded in 106 sequences with respect to (w.r.t.) reference sequence (YP\_325663.1) across 3945 amino acid length of CCHFV L RNA segment and 729 positions were found to be mutated. Mutation frequencies were classified into five intervals (0–0.2, 0.21–0.4, 0.41–0.6, 0.61–0.8 and 0.81–1.0) (Table 1). The majority of mutations were found to be less frequent, like 563 mutations were below 0.2. The number of mutations which lies in the range of 0.81–1.0 was considered highly frequent and 38 mutations were identified in this interval (Table 1). These mutations were observed to be present in all genotypes. Highly frequent mutations were mapped in L segment schematic diagram to locate in different domains of L segment (Fig. 2). In 0.81–1.0 frequency interval, 24 amino acids have been mutated into single variant and 14 amino acids into more than one amino acid variant (Fig. 2). L segment has two domains: ovarian tumour (OTU) domain and catalytic site domain [14]. Four highly frequent mutations were found in catalytic site domain at V2074I (0.99), I2134T/A (0.95), V2148 (0.99)

and Q2695H (0.87) whereas no mutation was found in OTU domain (Fig. 2).

Furthermore, we have identified the peptide having 90% conservancy and then looked for the location of highly frequent mutations. Four and thirteen conserved fragments were identified in OTU and catalytic site domain respectively (Table 2). Out of the four mutated amino acid positions, two mutated amino acids V2074I and V2148A were found to be present in 90% conserved fragments (Table 2). I2134T/A has mutation frequency of 0.95, but it was not recorded in the 90% conserved fragments, as two amino acid variations were observed at this position, whereas Q2695H has mutation frequency of 0.87; therefore, glutamine (Q) amino acid was recorded in the conserved fragment.

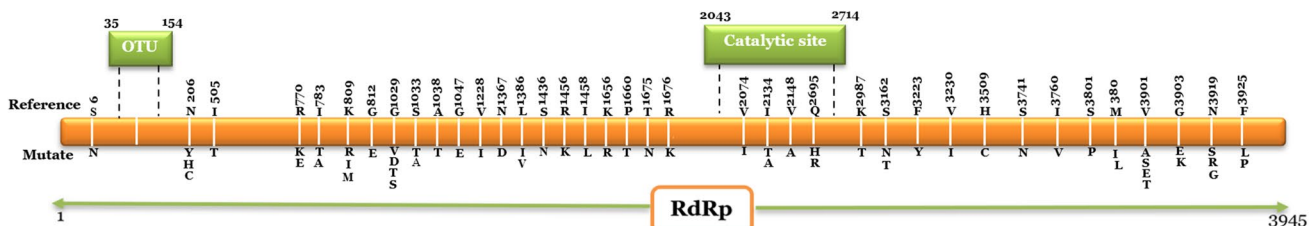
In addition, 42 new sequences released till 21 June 2022 (after 9 August 2021) were downloaded from NCBI database and analysed for whether the selected four highly frequent mutations of catalytic site are persistent or have been mutated over time. It was observed that three mutated amino acids in catalytic domain (V2074I, I2134T/A and V2148A) (Fig. 2) are persistent in all 42 recent sequences. Although Q2695H mutation was also located in 41 sequences, one sequence (Accession no.: QYF06534.1) collected from Gujarat, India, in 2019 displayed Q2695R mutation. Thereafter, highly frequent mutations identified in catalytic site domain were further evaluated to study the effect of point mutations on the stability of catalytic site domain with the aid of various bioinformatics tools and molecular dynamic simulation study.

**Table 1** No. of mutated amino acid positions at different frequency intervals

| Frequency interval | Total mutated amino acid positions |
|--------------------|------------------------------------|
| 0.0–0.2            | 563                                |
| 0.21–0.4           | 49                                 |
| 0.41–0.6           | 33                                 |
| 0.61–0.8           | 46                                 |
| 0.81–1.0           | 38                                 |
| Total              | 729                                |

## Structural and solubility analysis

The selected highly frequent mutations (V2074I, I2134A, I2134T, V2148A, Q2695R and Q2695H) identified in catalytic site domain (Fig. 2) were evaluated for structural and solubility analysis using Project HOPE and SODA online web server respectively. SODA web server reported that solubility of catalytic domain is increased upon introduction of point mutations at V2074I, I2134T, I2134A and V2148A, whereas Q2695H and Q2695R mutations have decreased the solubility of protein (Table 3). Project



**Fig. 2** Schematic representation of mutation allocated in 0.81–1.0 frequency interval. OTU, ovarian tumour domain. All mutations have been identified in all genotypes of CCHFV. The mutations represented in catalytic domain are persistent in recent strains as well

**Table 2** Ninety percent conserved fragments found in L segment domains

| Domain                              | 90% conserved fragments (Location)                                                                                                             |
|-------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------|
| Ovarian tumor domain (OTU) (35-154) | GDGNCFYHSIAELTMPNKT (35-54)                                                                                                                    |
|                                     | AARKYYQEEPEA (68-59)                                                                                                                           |
|                                     | RMLSDNEWGSTLEASMLAKEMG (92-113)                                                                                                                |
|                                     | TIHWTVAASDEVEAGIKFGDGDVFTAVNLLHSGQTHFDA (115-154)                                                                                              |
| Catalytic site (2043-2714)          | PSVSFTKKEEVLIKRLKRNFLKQHNLEIMETVNL*FFAALSAPWCLHYKALESYLVRHPEILDCGSKEDC (2043-2109)                                                             |
|                                     | LTLLDLSVSKLLVCLY (2111-2126)                                                                                                                   |
|                                     | SLKLGFLVKYA*VTLFTSNGEPFSLSLNDGGLDLHLKTTDEKLLHQTKIVFAKIGLSGN (2138-2196)                                                                        |
|                                     | YDFIWTQMIANSNFVCKRLTGRSTGERLPRSVRSKVIYEMVKLVGETGMAILQQLAFAQALNYEHR FYAVLAPKAQLGGARDLLVQETGKVMHATTEMFSRNLLKTTSDDGLTNPHLKETILNVGLDCL (2198-2329) |
|                                     | NMRNLGDKPISEGSNLVNFYKVICISGDNTKWGPIHCCSFSGMMQQVLKNVPDWCSFYKLTFIKILNC RQVEIPAGSIKKILNLVRY (2331-2417)                                           |
|                                     | LCSKGGVEQ (2419-2427)                                                                                                                          |
|                                     | DNLDSWDGNDTVKFLVTTYISKGLMALNSYNHMGQGIHHATSSVLTSLAAVLFEELAIFYLK (2439-2500)                                                                     |
|                                     | SLPQTTVHVEHAGSSDDYAKCIVVTG (2502-2527)                                                                                                         |
|                                     | ELYSQYDETFWKHACRLKNFTA AVQRCCQMKDSAKTLV (2532-2569)                                                                                            |
|                                     | DCFLEFYSEFMMGYRVTPAVIKFMFTGLINSSVTPQSLMQACQVSSQQAMYSVPLVTNT (2571-2631)                                                                        |
|                                     | FTLLRQQIFFNHVEDFIRRYG (2633-2653)                                                                                                              |
|                                     | LTLGTLSPFGRLFVPTYSGLVSSA VALEDAEVIARAAQTL (2655-2694)                                                                                          |
|                                     | MNSVSIQSSSLTTLDSL (2696-2712)                                                                                                                  |

\*Mutation found in 0.81–1.0 frequency (highlighted in green text)

**Table 3** Effect of point mutations on stability and solubility of catalytic site domain

| Mutation | SODA     | HOPE                                                                                                                                                                                                                         |
|----------|----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| V2074I   | 2.075*   | The mutations might disturb the function of catalytic domain as the mutant residues are present in a domain important for the activity of RdRp protein.                                                                      |
| V2148A   | 114.171* |                                                                                                                                                                                                                              |
| I2134T   | 16.861*  | The mutant residue occurs more often at this position than the wild residue in the sequence and mutant residues with similar properties have been observed, so it is possible that your mutant residue is accepted here too. |
| I2134A   | 21.163*  |                                                                                                                                                                                                                              |
| Q2695H   | -8.76    |                                                                                                                                                                                                                              |
| Q2695R   | -0.972   |                                                                                                                                                                                                                              |

\*Mutant models displaying solubility upon introduction of point mutation are highlighted in green text

HOPE server stated that V2074I and V2148A may disrupt the function, as both residues are located in the domain which is responsible for main activity of protein. I2134T/A and Q2695H/R are not damaging to protein as mutant

residues are more acceptable and found often than the wild residues in at both positions (Table 3). Furthermore, the stability of catalytic domain was also evaluated by molecular dynamic simulation study.

## Homology modelling and refinement of wild and mutant domains

To evaluate the effect of single amino acid variation on stability and compactness of CCHFV catalytic site domain, homology modelling of wild and mutant catalytic domain was carried out. Six mutant models were developed by introducing single point mutation in each model. Three-dimensional structure of catalytic site domain is not available in protein data bank (PDB) database. The PDB structure of Ebola virus RdRp protein was used as template (PDBid: 7YES), as it displayed maximum identity with catalytic site domain; therefore, it was used for homology modelling. Wild catalytic site domain was constructed using template (PDBid: 7YES), followed by refinement and energy minimization of wild model. Thereafter, refined and minimized wild model was used as template for the construction of the six mutant models. All mutant models were also refined and minimized before molecular dynamic simulation study. Wild and six mutant models displayed 90.3% and 91.4–91.8% (Table S2) amino acids in favourable regions in Ramachandran plots respectively (Figure S1 and Figure S2). Superimposed images of wild and mutant models suggest that modelled protein of all mutant models displayed deviation in the range of 1.139–1.689 Å RMSD from wild type model in PyMOL software (Fig. 3 and Table S2). All the mutant

residues are superimposed over the wild residue whilst I2134A and V2148A displayed minimum deviation of 1.189 Å and 1.139 Å respectively upon superimposition (Fig. 3).

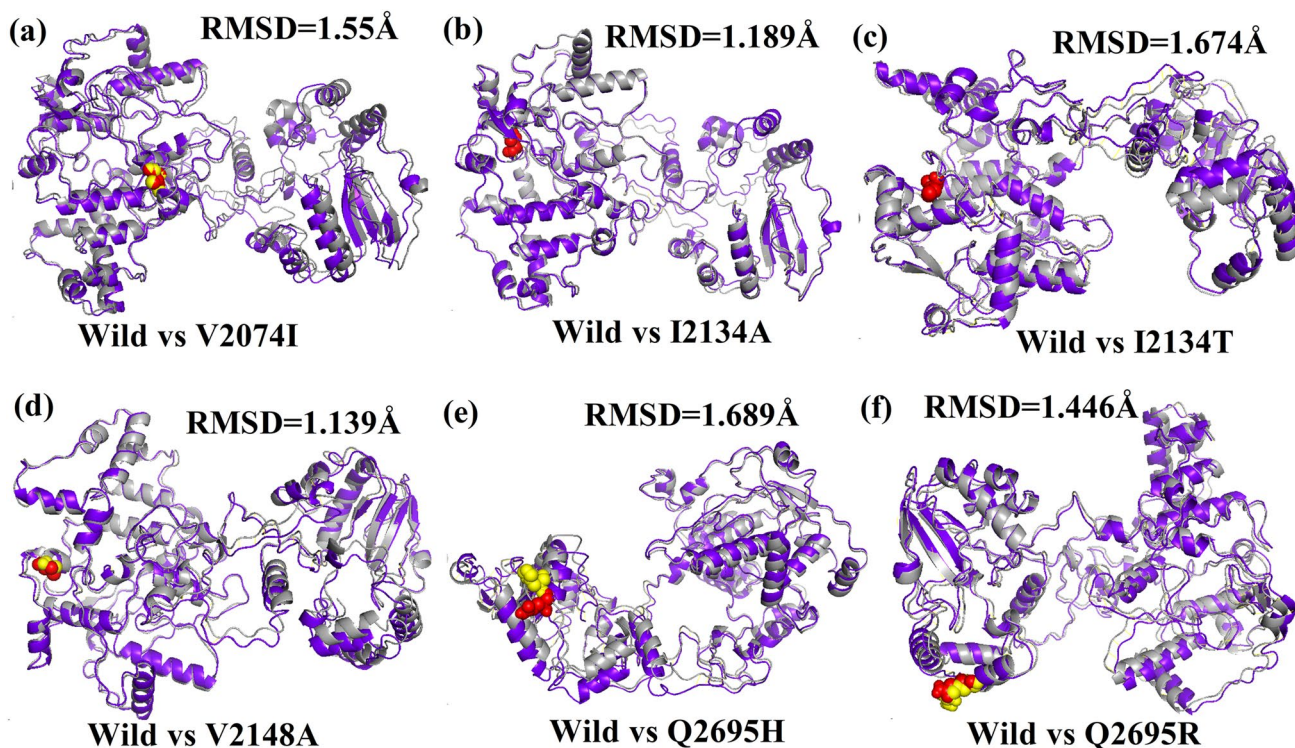
## Molecular dynamic simulation study

The stability of wild and six mutant models of catalytic site domain was assessed by MD simulation study of 50 ns in charmm27 force field in TIP3P water box. All the models were solvated with SOL molecules and neutralized by 3  $\text{Cl}^-$  ions each.

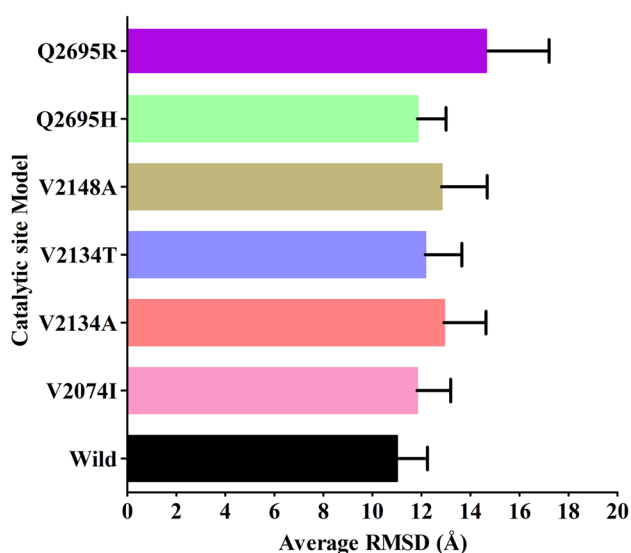
In root mean square deviation (RMSD) plot, wild type model has shown less average deviation ( $11.01 \pm 1.24$  Å) than V2074I ( $11.82 \pm 1.37$  Å), I2134A ( $12.92 \pm 1.72$  Å), I2134T ( $12.16 \pm 1.49$  Å), V2148A ( $12.82 \pm 1.87$  Å), Q2695R ( $11.84 \pm 1.16$  Å) and Q2695H ( $14.7 \pm 2.51$  Å) (Fig. 4).

RMSD trajectories revealed that wild type model gained convergence after 5 ns and V2074I and Q2695H mutant models gained convergence before 5 ns, whereas I2134A/T and V2148A gained convergence after 10 ns (Fig. 5). Among all models, Q2695R displayed large deviation during 50 ns simulation and gained convergence after 25 ns.

Root mean square fluctuation (RMSF) graph displayed fluctuations at each amino acids in the catalytic site domain of RdRp protein; therefore, V2074I model has been mutated



**Fig. 3** Pictorial representation of template model superimposed on (a) V2074I, b I2134A, c I2134T, d V2148A, e Q2695H and f Q2695R. V2148A and I2134A displayed minimum deviation

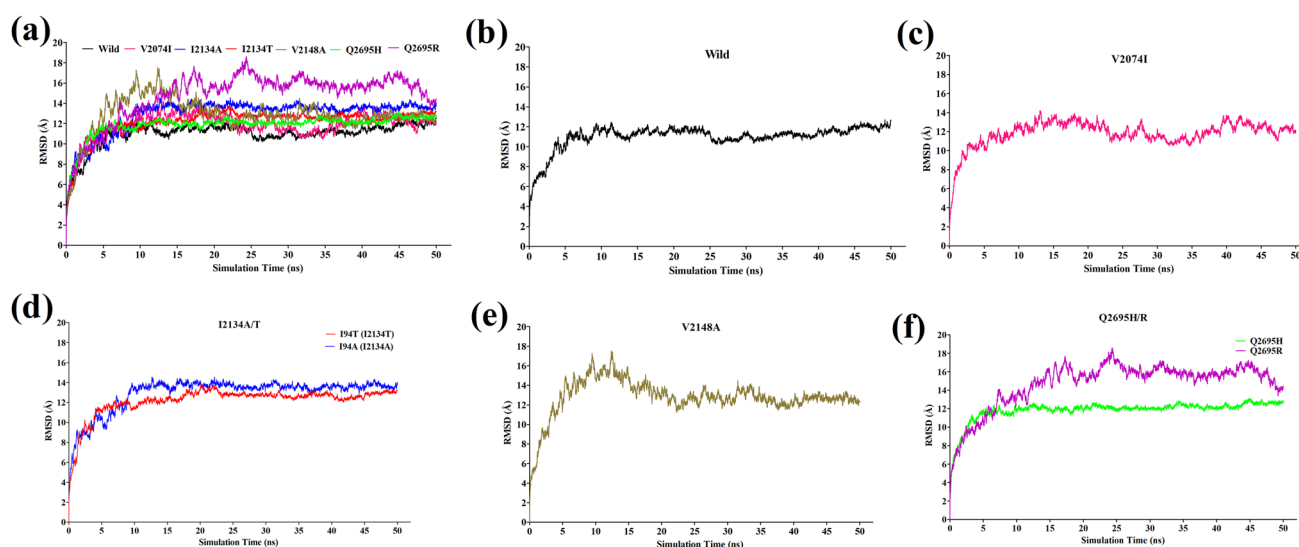


**Fig. 4** Average RMSD plot of wild and mutant models. All mutant models have shown more deviation than wild model

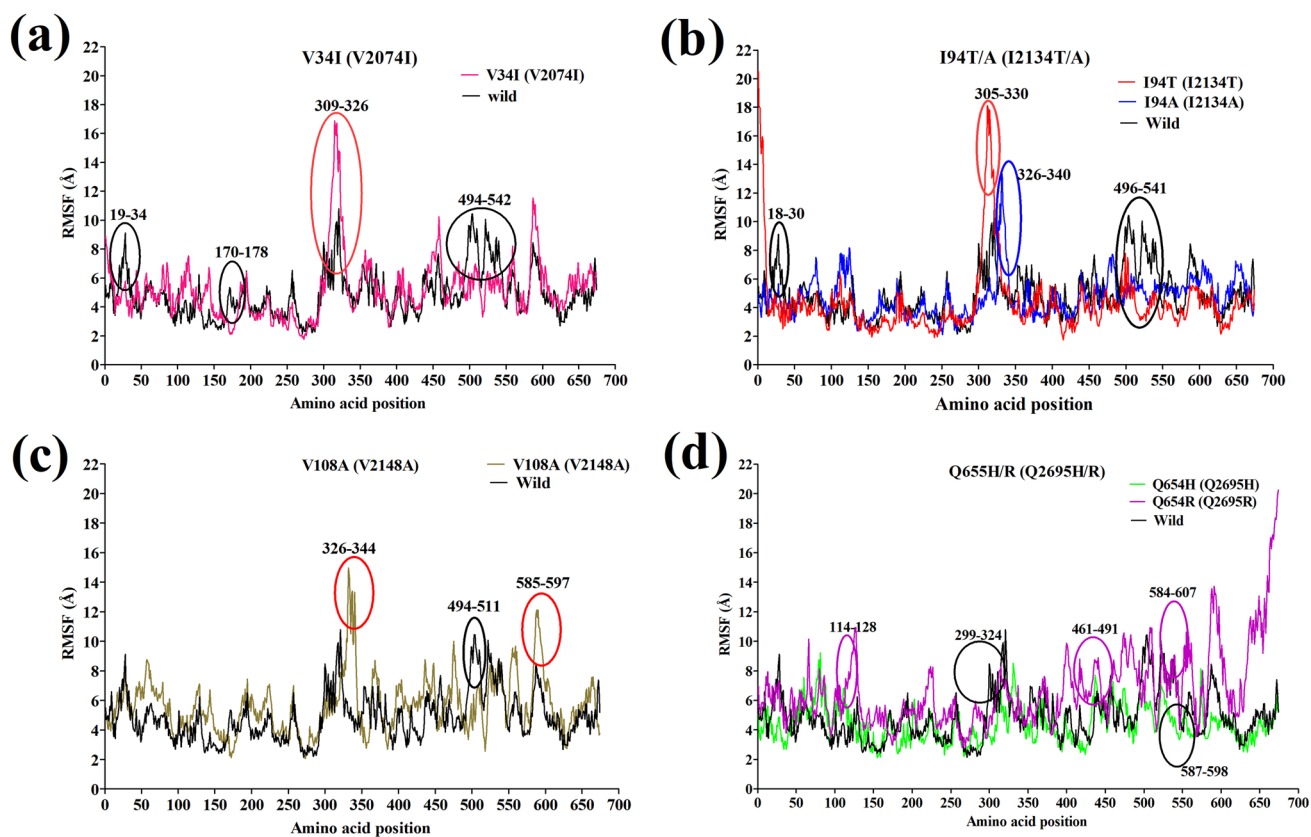
at 34 position in catalytic site domain, I2134T/A at 94 position, V2148A at 108 position and Q2695H/R at 652 position. All mutant models displayed more RMSF values w.r.t. wild model, as core residues of all mutant models have shown large fluctuations than wild model. All mutant models except Q2695H/R displayed large fluctuations in the central residues (between 305 and 344) w.r.t. wild model. As V2074I and I2134T have large RMSF difference between 305 and 326 amino acid (aa) residues, I2134A and V2148A displayed large RMSF difference between 326 and 344 aa residues (Fig. 6). On the other hand, mutant models have

also shown stability w.r.t. wild model at some positions, as V2074I gained stability at 19–34 aa residues, 170–178 aa residues and 494–542 aa residues (Fig. 6). I2134T/A have shown less RMSF values than wild model at 18–30 aa residues and 496–541 aa residues. V2148A model has shown stability at 494–511 aa residues. Among Q2695H and Q2695R, Q2695R is highly unstable and have shown large fluctuations w.r.t. wild model. Among all mutant models, V2074I, V2148A and Q2695H exhibited minimum fluctuation, whereas other mutant models displayed large fluctuation upon single amino acid mutation.

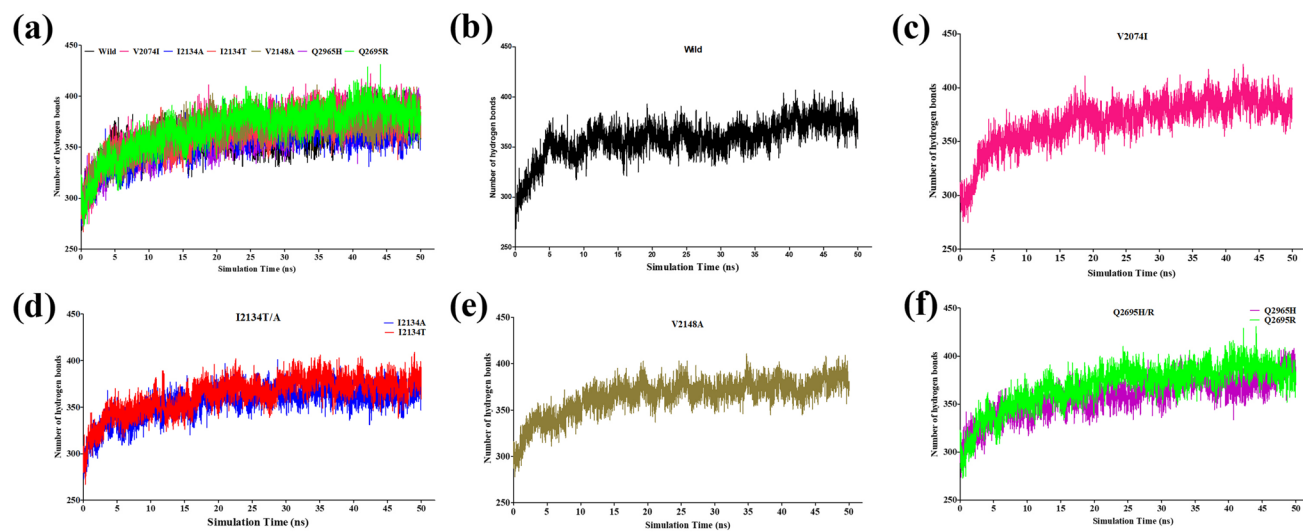
Overall, RMSF plots decipher that catalytic site domain lost stability upon introduction of mutant residues and large difference in RMSF values observed in the mutant models in comparison with wild model which suggests that point mutations have apparently affected the stability of catalytic site domain. The rationale behind the large fluctuation and deviation in mutant models was further assessed by hydrogen bond analysis. It was observed that the number of mean intra-hydrogen bonds in mutant models has been increased w.r.t. wild model (358.2). Among all mutant models, V2074I has maximum mean hydrogen bonds (369.1), followed by Q2695R (368.6), V2148A (364.9) and I2134T (363.1). Q2695H exhibits similar number of mean hydrogen bonds (358.3) w.r.t. wild model, whereas the number of mean hydrogen bonds was reduced in I2134A (355.4). Wild, V2074I, I2134A, I2134T, V2148A, Q2695H and Q2695R displayed 349, 359, 347, 352, 356, 347 and 357 hydrogen bonds respectively at 25th percentile. At 75th percentile, 370, 384, 368, 377, 378, 373 and 385 mean hydrogen bonds have been recorded in wild, V2074I, I2134A, I2134T, V2148A, Q2695H and Q2695R models respectively (Fig. 7). Overall



**Fig. 5** RMSD plot of catalytic site domain representing deviation for 50 ns simulation run of (a) all models, b wild model, c V2074I, d I2134T/A, e V2148A and f Q2695H/R. All mutant models have shown more deviation with respect to wild model



**Fig. 6** RMSF plot of catalytic site domain representing fluctuation at each amino acid residue, **a** V34I (V2074I) vs wild model, **b** I94T/A (I2134T/A) vs wild model, **c** V2148A vs wild model, **d** Q2695H/R vs wild model



**Fig. 7** Average number of intramolecular hydrogen bonds in **(a)** both wild and mutant models, **b** wild, **c** V2074I, **d** I2134T/A, **e** V2148A, **f** Q2695H/R

MD trajectories deciphered that local interaction has been affected as a result of a single mutant residue in the catalytic site domain. Out of six mutant models, four mutations

(V2074I, I2134T, V2148A and Q2695R) have induced local interactions, whereas I2134A displayed reduction in local interactions. Q2695H displayed minimum deviation and



fluctuation during 50 ns simulation as well as local interactions remained stable due to mutation.

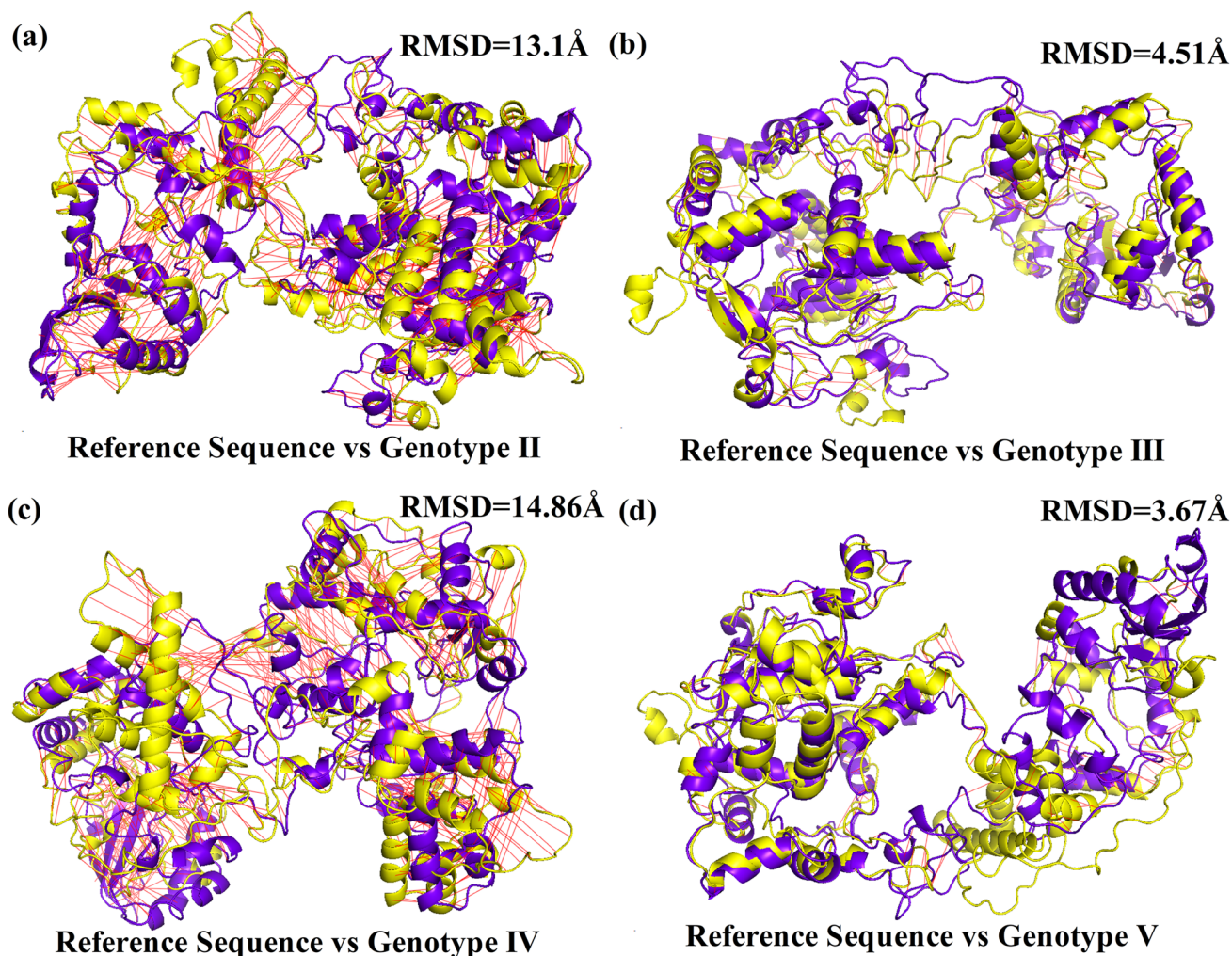
### Structural variation of catalytic site domain among CCHFV genotypes

CCHFV has been classified into six genotypes based on their topological preferences. One representative protein sequence of each CCHFV L segment for genotypes II, III, IV and V (Table S3) was taken from the protein dataset to analyse the structural variation. As protein sequences of genotype I and genotype VI of CCHFV L segment have not been reported till date, therefore, their structural variation was not analysed. Thereafter, three-dimensional structure of each CCHFV L segment genotype catalytic site domain was executed by taking modelled catalytic site domain of NCBI reference sequence (Accession No.: YP\_325663.1) of CCHFV L segment by homology modelling. Three-dimensional

structure of each genotype was visualized and superimposed with wild model to calculate the RMSD value using PyMOL software. Upon superimposition of each genotype on reference sequence, large structural variation was observed in genotype II (RMSD=13.1 Å) and genotype IV (RMSD=14.86 Å) (Fig. 8). Genotype III and genotype V displayed less structural variation as they have shown RMSD value of 4.51 Å and 3.67 Å respectively (Fig. 8). Moreover, NCBI reference sequence also belongs to genotype III; therefore, less variation was observed between genotype III and reference sequence.

### Discussion

Mutations aid in the survival of virus, resistance against drugs and evolution of new strains. Mutation profiling provides insight about the effect of point mutation on virus



**Fig. 8** Pictorial representation of catalytic site domain of NCBI reference sequence with (a) genotype II, b genotype III, c genotype IV and d genotype V

virulence, structure, replication and host invasion [36, 37]. A study described the effect of multi-variable point mutation (D53Q/E/H/W) in HIV on the stability of envelope glycoprotein. Out of four mutations, D53W displayed high binding affinity than wild protein [38]. Another study evaluated the effect of point mutation on Zika virus polyprotein in mice and Vero cells and found that few mutations reduced replication fitness whilst few mutations aided virus persistence and transmission in future outbreaks [39]. In the current study, four highly frequent mutations were found to be located in the catalytic domain of RNA-dependent RNA polymerase (RdRp) protein and molecular dynamic simulation analysis revealed that these point mutations have affected the stability of protein.

CCHFV genome has three RNA segments (L, M and S segment) with fatality rate of 10–40% and has been classified into six genotypes based on their geographical predominance [35, 40]. Phylogenetic tree developed with 107 sequences represented less divergence among the sequences belonging to same CCHFV genotypes and results were in concordance with studies reported in the literature [8, 9, 35, 41]. NCBI reference sequence (Accession No.: YP\_325663.1) belongs to genotype III; therefore, genotype III sequences displayed less evolutionary distance from reference sequence in the phylogenetic tree, albeit genotypes II, IV and V have shown large distance in the phylogenetic tree. As phylogenetic tree displayed divergence among CCHFV genotypes, thereby, mutation profiling of 106 sequences was executed.

Researchers have reported *in vitro* and *in vivo* studies describing effect of point mutation in viral RdRp on the viral replication and ribavirin resistance. G64S point mutation in poliovirus [42], A372V point mutation in Coxsackie virus [43], L123F point mutation in human enterovirus 71 [44], R84H point mutation in foot and mouth disease virus [45], V43I point mutation in influenza-A virus [46] and V793I and G806R point mutation in West Nile virus [47] have caused ribavirin resistance, better survival of virus and enhanced polymerase fidelity (rate of error generation during polymerization) [42, 44–47]. Mutations in spike protein of SARS-CoV2 reduces infectivity upon glycosylation at N331 and N343 and enhances infectivity upon D614G mutations whereas few mutations cause antibody resistance [48]. Considering these facts, mutation profiling in 106 sequences w.r.t. NCBI reference sequence (Accession No.: YP\_325663.1) of L segment (encoding for RdRp protein) was carried out and has reported a total of 729 mutated amino acid positions among 3945 amino acids. The majority of the mutations were less frequent and 38 amino acid positions were in the frequency interval of 0.81–1.0 which was considered highly frequent mutations.

Function of proteins is primarily executed by their functional and structural elements termed as domains [17]. L segment has two domains: ovarian tumour domain (OTU)

and catalytic site domain. OTU has deubiquitination (DUB) activity and participates in evading host immune response which suggests DUB inhibitors can be designed for CCHF treatment [15]. Catalytic site domain of RdRp plays key role in RNA-dependent polymerization (<https://prosite.expasy.org/rule/PRU00539>), which is crucial for virus survival [49]. Hence, the highly frequent mutations were mapped in L segment and four mutations were found in catalytic site (V2074I, I2134T/A, V2148 and Q2695H) but OTU domain has no mutated residue. Out of four mutated amino acid positions, two mutations were located in 90% conserved fragments, whereas I2134T/A was not located in these conserved fragments as two amino acid variations were identified and Q2695H has 0.87 mutation frequency; therefore, it was not found in the 90% conserved fragment. In addition, persistence of four mutated positions identified in the catalytic site domain was assessed in 42 recent sequences. It was observed that three mutated amino acid positions (V2074I, I2134T/A and V2148A) are persistent in all 42 sequences. Although Q2695H mutation was also located in 41 sequences, one sequence (Accession no.: QYF06534.1) displayed Q2695R mutation.

Researchers have reported that catalytic activity of protein decreases when the stability of protein increases due to mutation [50]. Saini et al. described enhancement of stability due to mutations (A49L and Q106T) in interleukin-4 [23]. Salinas et al. have reported that mutations can alter binding site pocket shape and its ligand identification properties [22]. To study the effect of point mutations on the solubility and stability of catalytic domain, three-dimensional structure of wild and six mutant models was generated using MODELLER. Ramachandran plot of mutant models displayed more amino acids in favourable regions in comparison with wild model, as wild and six mutant models displayed 90.3% and 91.4–91.8% amino acids in favourable regions respectively. Solubility analysis deciphered that solubility of protein might increase due to V2074I, I2134T/A and V2148A point mutations, whereas Q2695H/R point mutation might decrease the solubility. HOPE server suggests that V2074I and V2148A might disrupt the protein function as they are located near highly conserved region, whereas I2134T/A and Q2696H/R may not interfere protein function, as the mutant residues are more acceptable at these positions than the wild residue. In addition, researchers have studied the effect of point mutations on the stability of protein aided molecular dynamic simulation study [51, 52]. It has been reported that I591D point mutation in alpha-dystroglycan caused instability of protein [51], whilst another study has reported less RMSD deviation in mutant models than wild model of serum and glucocorticoid-regulated kinase 1 (SGK1) protein [52]. Zhang et al. observed less deviation in wild type of human cytochrome P450 A2 protein than mutant F186L mutant model [53]. Therefore, stability of

wild and mutant catalytic domain was executed for 50 ns using GROMACS v2021.3 and the topology of each molecule was created in charmm27 force field in TIP3P water box. Overall, all mutant models displayed more deviation and fluctuation w.r.t wild model. Among six mutant models, V2074I, I2134T/A and Q2695H displayed minimum deviation over 50 ns simulation run but RMSF plot of V2074I model displayed high RMSF values of core aa residues than wild model, whereas Q2695H model displayed minimum deviation as well as less fluctuation in the core amino acid residues. Along with Q2695H model, RMSF plot of V2074I and I2134T/A models also described minimum fluctuation at majority core aa residues. At multi-variable positions, it was observed that I2134T and Q2695H are more stable than I2134A and Q2695R respectively. The results predicted by project HOPE server are comparable with molecular simulation study, as HOPE server predicted that in I2134T/A and Q2695H models, these mutant residues often occur at these positions and exhibit similar properties and therefore might not disrupt the stability of protein, and interestingly, these models displayed comparative deviation and fluctuation w.r.t. wild model during 50 ns simulation run as well. Our findings are in agreement with a study reported by Khan et al., describing increase in flexibility and deviation of residues in mutant models of ribosomal protein S1 w.r.t. wild type [54]. Another study has evaluated the effect of mutations on intramolecular contacts and reported the number of contacts in Rab5a mutants: Ala30Pro and Ala30Arg contacts have been increased to 100 and 202 respectively [55]. Additionally, structural variation of catalytic site domain of NCBI reference sequence was executed among all genotypes and large structural variation was observed in genotype II (RMSD: 13.1 Å) and genotype IV (RMSD: 14.86 Å). These structure variations are in accordance with phylogenetic analysis where genotypes II and IV are distant from reference sequence. However, genotype V is also distant from reference sequence but there is less structure variation (RMSD). One of the reasons could be that catalytic domain may be not undergoing much structure change in this genotype V. Moreover, all the mutant residues are part of catalytic site domain responsible for RNA polymerization, and it may affect the viral fidelity and efficacy for RdRp-targeted drugs, as other viruses have gained resistance due to point mutation in their RdRp protein [42, 44–47].

## Conclusion

Due to high mutation rate of viruses, it is difficult to develop diagnostic kits and treatment which can be used over years. The current study aimed at identification of highly frequent mutations and study of effect of mutations on the stability of CCHFV domain. It was found that OTU

domain is a highly conserved region of CCHFV L segment as no highly frequent mutation was found in it, whereas four mutations were found in catalytic site domain. Interestingly, all mutant models displayed less stability and more deviation than wild model in molecular dynamic simulation run of 50 ns and conclude that catalytic site domain has affected stability upon introduction of those mutations and displayed large amino acid fluctuation. All catalytic site domain mutations were found to be persistent in all genotypes and recent strains as well; therefore, they might lead to resistance against RdRp-targeted drugs due to high fidelity as observed in case of other viruses.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00894-023-05487-7>.

**Acknowledgements** We would like to thank the scientific community.

**Author contribution** Neha Kaushal—original draft; data curation; formal analysis. Manoj Baranwal—conceptualization; methodology; supervision; review and editing

## Declarations

**Ethical approval** Not required

**Consent to participate** Not applicable

**Conflict of interest** The authors declare no competing interests.

## References

- Shrivastava N, Shrivastava A, Ninawe SM et al (2019) Development of multispecies recombinant nucleoprotein-based indirect elisa for high-throughput screening of crimean-congo hemorrhagic fever virus-specific antibodies. *Front Microbiol* 10:1–14. <https://doi.org/10.3389/fmicb.2019.01822>
- Appannanavar SB, Mishra B (2011) An update on crimean congo hemorrhagic fever. *J Glob Infect Dis* 3:285–292. <https://doi.org/10.4103/0974-777X.83537>
- ECDC (2022) Factsheet about Crimean-Congo haemorrhagic fever. In: *Eur Cent Dis Prev Control*. <https://www.ecdc.europa.eu/en/crimean-congo-haemorrhagic-fever/facts/factsheet>. Accessed 13 Sep 2022
- Lahariya C, Goel MK et al (2012) Emergence of viral hemorrhagic fevers: is recent outbreak of crimean congo hemorrhagic fever in India an indication. *J Postgrad Med* 58:39
- WHO (2023) Crimean-Congo haemorrhagic fever. In: *World Heal Organ*. [https://www.who.int/health-topics/crimean-congo-haemorrhagic-fever#tab=tab\\_1](https://www.who.int/health-topics/crimean-congo-haemorrhagic-fever#tab=tab_1). Accessed 31 Jan 2023
- Yadav PD, Cherian SS, Zawar D et al (2013) Genetic characterization and molecular clock analyses of the Crimean-Congo hemorrhagic fever virus from human and ticks in India, 2010–2011. *Infect Genet Evol* 14:223–231. <https://doi.org/10.1016/j.meegid.2012.10.005>
- Yadav PD, Patil DY, Shete AM et al (2016) Nosocomial infection of CCHF among health care workers in Rajasthan, India. *BMC Infect Dis* 16:1–6. <https://doi.org/10.1186/s12879-016-1971-7>

8. Lukashev AN, Klimentov AS, Smirnova SE et al (2016) Phylogeography of Crimean-Congo hemorrhagic fever virus. *PLoS One* 11:1–14. <https://doi.org/10.1371/journal.pone.0166744>
9. Monsalve Arteaga L, Muñoz Bellido JL, Negredo AI et al (2021) New circulation of genotype V of Crimean-Congo hemorrhagic fever virus in humans from Spain. *PLoS Negl Trop Dis* 15:e0009197. <https://doi.org/10.1371/journal.pntd.0009197>
10. Spengler JR, Bergeron É, Rollin PE (2016) Seroepidemiological studies of Crimean-Congo hemorrhagic fever virus in domestic and wild animals. *PLoS Negl Trop Dis* 10:1–28. <https://doi.org/10.1371/journal.pntd.0004210>
11. Farzani TA, Földes K, Ergünay K et al (2019) Immunological analysis of a CCHFV mRNA vaccine candidate in mouse models. *Vaccines* 7:1–17. <https://doi.org/10.3390/vaccines7030115>
12. Carter SD, Surtees R, Walter CT et al (2012) Structure, function, and evolution of the Crimean-Congo hemorrhagic fever virus nucleocapsid protein. *J Virol* 86:10914–10923. <https://doi.org/10.1128/jvi.01555-12>
13. Akutsu M, Ye Y, Virdee S et al (2011) Molecular basis for ubiquitin and ISG15 cross-reactivity in viral ovarian tumor domains. *Proc Natl Acad Sci U S A* 108:2228–2233. <https://doi.org/10.1073/pnas.1015287108>
14. Uniprot (2022) Q6TQR6 L\_CCHFI. <https://prosite.expasy.org/rule/PRU00539>. Accessed 13 Sep 2022
15. Srinivasan P, Kumar SP, Karthikeyan M et al (2011) Epitope-based immunoinformatics and molecular docking studies of nucleocapsid protein and ovarian tumor domain of Crimean-Congo hemorrhagic fever virus. *Front Genet* 2:1–9. <https://doi.org/10.3389/fgene.2011.00072>
16. Hawman DW, Meade-White K, Leventhal S et al (2021) Immunocompetent mouse model for Crimean-Congo hemorrhagic fever virus. *Elife* 10:1–22. <https://doi.org/10.7554/eLife.63906>
17. Basu MK, Poliakov E, Rogozin IB (2009) Domain mobility in proteins: functional and evolutionary implications. *Brief Bioinform* 10:205–216. <https://doi.org/10.1093/bib/bbn057>
18. Zheng LL, Li C, Ping J et al (2014) The domain landscape of virus-host interactomes. *Biomed Res Int* 2014. <https://doi.org/10.1155/2014/867235>
19. Maginnis MS (2018) Virus-receptor interactions: the key to cellular invasion. *J Mol Biol* 430:2590–2611
20. Doss CGP, Rajith B, Garwas N et al (2012) Screening of mutations affecting protein stability and dynamics of FGFR1-A simulation analysis. *Appl Transl Genomics* 1:37–43. <https://doi.org/10.1016/j.atg.2012.06.002>
21. Piao L, Chen Z, Li Q et al (2019) Molecular dynamics simulations of wild type and mutants of SAPAP in complexed with shank3. *Int J Mol Sci* 20. <https://doi.org/10.3390/ijms20010224>
22. Ramírez-Salinas GL, García-Machorro J, Quiliano M et al (2015) Molecular modeling studies demonstrate key mutations that could affect the ligand recognition by influenza AH1N1 neuraminidase. *J Mol Model* 21. <https://doi.org/10.1007/s00894-015-2835-6>
23. Saini S, Jyoti-Thakur C, Kumar V et al (2018) In silico mutational analysis and identification of stability centers in human interleukin-4. *Mol Biol Res Commun* 7:67–76. <https://doi.org/10.22099/mbrc.2018.28855.1310>
24. Li YC, Bai WZ, Hashikawa T (2020) The neuroinvasive potential of SARS-CoV2 may play a role in the respiratory failure of COVID-19 patients. *J Med Virol* 92:552–555. <https://doi.org/10.1002/jmv.25728>
25. Celik I, Khan A, Dwivany FM et al (2022) Computational prediction of the effect of mutations in the receptor-binding domain on the interaction between SARS-CoV-2 and human ACE2. *Mol Divers* 26:3309. <https://doi.org/10.1007/s11030-022-10392-x>
26. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680. <https://doi.org/10.1093/nar/22.22.4673>
27. Kumar S, Stecher G, Li M et al (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 35:1547–1549. <https://doi.org/10.1093/molbev/msy096>
28. Nasirian H (2020) New aspects about Crimean-Congo hemorrhagic fever (CCHF) cases and associated fatality trends: a global systematic review and meta-analysis. *Comp Immunol Microbiol Infect Dis* 69:101429
29. Hall T, Biosciences I, Carlsbad C (2011) BioEdit: an important software for molecular biology. *GERF Bull Biosci* 2:60–61
30. Calloni G, Vabulas RM (2020) Proteome-scale studies of protein stability. Elsevier Inc.
31. Paladin L, Piovesan D, Tosatto SCE (2017) SODA: prediction of protein solubility from disorder and aggregation propensity. *Nucleic Acids Res* 45:W236–W240. <https://doi.org/10.1093/nar/gkx412>
32. Venselaar H, te Beek TAH, Kuipers RKP et al (2010) Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics* 11:1–10. <https://doi.org/10.1186/1471-2105-11-548>
33. Eswar N, Webb B, Marti-Renom MA et al (2006) Comparative protein structure modeling using modeller. *Curr Protoc Bioinforma* 15:5.6.1–5.6.30. <https://doi.org/10.1002/0471250953.BI0506S15>
34. Pettersen EF, Goddard TD, Huang CC et al (2004) UCSF Chimera - a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612. <https://doi.org/10.1002/jcc.20084>
35. Wampande EM, Waiswa P, Allen DJ et al (2021) Phylogenetic characterization of Crimean-Congo hemorrhagic fever virus detected in African blue ticks feeding on cattle in a Ugandan abattoir. *Microorganisms* 9:1–12. <https://doi.org/10.3390/microorganisms9020438>
36. Kosuge M, Furusawa-Nishii E, Ito K et al (2020) Point mutation bias in SARS-CoV-2 variants results in increased ability to stimulate inflammatory responses. *Sci Rep* 10:1–9. <https://doi.org/10.1038/s41598-020-74843-x>
37. Shishir TA, Jannat T, Bin NI (2022) An in-silico study of the mutation-associated effects on the spike protein of SARS-CoV-2, Omicron variant. *PLoS One* 17:1–21. <https://doi.org/10.1371/journal.pone.0266844>
38. Conceição de Souza R, de Medeiros MG, Siqueira AS et al (2016) Investigating the effects of point mutations on the affinity between the cyanobacterial lectin microvirin and high mannose-type glycans present on the HIV envelope glycoprotein. *J Mol Model* 22:1–9. <https://doi.org/10.1007/s00894-016-3137-3>
39. Collette NM, Lao VHI, Weilhammer DR et al (2020) Single amino acid mutations affect Zika virus replication in vitro and virulence in vivo. *Viruses* 12:1–20. <https://doi.org/10.3390/v12111295>
40. Abuova G, Karan L, Pshenichnaya N et al (2022) Results of genotyping of the Congo-Crimean hemorrhagic fever virus circulating in Southern Kazakhstan. *Int J Infect Dis* 116:S121. <https://doi.org/10.1016/j.ijid.2021.12.286>
41. Fazlalipour M, Baniyasi V, Pouriayevali M et al (2019) Crimean-Congo hemorrhagic fever virus Asia 2 genotype in Qeshm Island, southern Iran: a case report. *J Vector Borne Dis* 56:276–279. <https://doi.org/10.4103/0972-9062.289389>
42. Bordería AV, Rozen-Gagnon K, Vignuzzi M (2016) Fidelity variants and RNA quasispecies. *Exp Syst* 392:303. [https://doi.org/10.1007/82\\_2015\\_483](https://doi.org/10.1007/82_2015_483)
43. Levi LI, Gnädig NF, Beaucourt S et al (2010) Fidelity Variants of RNA Dependent RNA polymerases uncover an indirect, mutagenic activity of amiloride compounds. *PLoS Pathog* 6:1–14. <https://doi.org/10.1371/journal.ppat.1001163>

44. Meng T, Kwang J (2014) Attenuation of human enterovirus 71 high-replication-fidelity variants in AG129 mice. *J Virol* 88:5803–5815. <https://doi.org/10.1128/jvi.00289-14>
45. Zeng J, Wang H, Xie X et al (2013) An increased replication fidelity mutant of foot-and-mouth disease virus retains fitness in vitro and virulence in vivo. *Antiviral Res* 100:1–7. <https://doi.org/10.1016/j.antiviral.2013.07.008>
46. Cheung PPH, Watson SJ, Choy KT et al (2014) (2014) Generation and characterization of influenza A viruses with altered polymerase fidelity. *Nat Commun* 51(5):1–13. <https://doi.org/10.1038/ncomms5794>
47. Van Slyke GA, Arnold JJ, Lugo AJ et al (2015) Sequence-specific fidelity alterations associated with West Nile virus attenuation in mosquitoes. *PLoS Pathog* 11:1–21. <https://doi.org/10.1371/journal.ppat.1005009>
48. Li Q, Wu J, Nie J et al (2020) The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. *Cell* 182:1284–1294.e9. <https://doi.org/10.1016/j.cell.2020.07.012>
49. NCBI (2022) Catalytic domain. In: NIH. The region of an enzyme, to cause the enzymatic reaction. <https://www.ncbi.nlm.nih.gov/mesh?Db=mesh&Cmd=DetailsSearch&Term=%22Catalytic+Domain%22%5BMeSH+Terms%5D#:~:text>. Accessed 14 Sep 2022
50. Studer RA, Dessailly BH, Orengo CA (2013) Residue mutations and their impact on protein structure and function: detecting beneficial and pathogenic changes. *Biochem J* 449:581–594. <https://doi.org/10.1042/BJ20121221>
51. Pirolli D, Sciandra F, Bozzi M et al (2014) Insights from molecular dynamics simulations: structural basis for the V567D mutation-induced instability of zebrafish alpha-dystroglycan and comparison with the murine model. *PLoS One* 9. <https://doi.org/10.1371/journal.pone.0103866>
52. AlAjmi MF, Khan S, Choudhury A et al (2021) Impact of deleterious mutations on structure, function and stability of serum/glucocorticoid regulated kinase 1: a gene to diseases correlation. *Front Mol Biosci* 8:1–14. <https://doi.org/10.3389/fmolb.2021.780284>
53. Zhang T, Liu LA, Lewis DFV, Wei DQ (2011) Long-range effects of a peripheral mutation on the enzymatic activity of cytochrome P450 1A2. *J Chem Inf Model* 51:1336–1346. <https://doi.org/10.1021/ci200112b>
54. Khan MT, Khan A, Rehman AU et al (2019) Structural and free energy landscape of novel mutations in ribosomal protein S1 (rpsA) associated with pyrazinamide resistance. *Sci Rep* 9:1–12. <https://doi.org/10.1038/s41598-019-44013-9>
55. Khan FI, Aamir M, Wei DQ et al (2017) Molecular mechanism of Ras-related protein Rab-5A and effect of mutations in the catalytically active phosphate-binding loop. *J Biomol Struct Dyn* 35:105–118. <https://doi.org/10.1080/07391102.2015.1134346>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.