

Mechanism of Preferential Packaging of Negative Sense Genomic RNA by Viral Nucleoproteins in Crimean-Congo Hemorrhagic Fever Virus

Mohammad Reza Dayer · Mohammad Saaid Dayer ·
Seyedeh Elham Rezatofighi

Published online: 30 January 2015
© Springer Science+Business Media New York 2015

Abstract The Crimean-Congo Hemorrhagic Fever (CCHF) is an infectious disease of high virulence and mortality caused by a negative sense RNA nairovirus. The genomic RNA of CCHFV is enwrapped by its nucleoprotein. Positively charged residues on CCHFV nucleoprotein provide multiple binding sites to facilitate genomic RNA encapsidation. In the present work, we investigated the mechanism underlying preferential packaging of the negative sense genomic RNA by CCHFV nucleoprotein in the presence of host cell RNAs during viral assembly. The work included genome sequence analyses for different families of negative and positive sense RNA viruses, using serial docking experiments and molecular dynamic simulations. Our results indicated that the main determinant parameter of the nucleoprotein binding affinity for negative sense RNA is the ratio of purine/pyrimidine in the RNA molecule. A negative sense RNA with a purine/pyrimidine ratio (>1) higher than that of a positive sense RNA (<1) exhibits higher affinity for the nucleoprotein. Our calculations revealed that a negative sense RNA expresses about 0.5 kJ/mol higher binding energy per nucleotide compared

to a positive sense RNA. This energy difference produces a binding energy high enough to make the negative sense RNA, the preferred substrate for packaging by CCHFV nucleoprotein in the presence of cellular or complementary positive sense RNAs. The outcome of this study may contribute to ongoing researches on other viral diseases caused by negative sense RNA viruses such as Ebola virus which poses a security threat to all humanity.

Keywords Crimean-Congo Hemorrhagic Fever Virus · Nucleoprotein · Negative sense RNA · Positive sense RNA · Molecular dynamic simulation

Abbreviations

CCHFV	Crimean-Congo Hemorrhagic Fever Virus
MD	Molecular dynamics
RMSD	Root mean square displacement
RMSF	Root mean square fluctuation
SAS	Solvent accessible surface

M. R. Dayer (✉)
Department of Biology, Faculty of Science, Shahid Chamran
University of Ahvaz, Ahvaz, Iran
e-mail: mrdayer@scu.ac.ir

M. S. Dayer
Department of Parasitology and Medical Entomology, School
of Medical Sciences, Tarbiat Modares University, Tehran, Iran
e-mail: dayer@modares.ac.ir

S. E. Rezatofighi
Division of Virology and Microbiology, Department of Biology,
Faculty of Science, Shahid Chamran University of Ahvaz,
Ahvaz, Iran
e-mail: e.tofighi@scu.ac.ir

1 Introduction

Crimean Congo Hemorrhagic Fever (CCHF) is a tick borne disease caused by a virus belongs to *Nairovirus* genus of *Bunyaviridae* family called Crimean Congo Hemorrhagic Fever Virus or briefly CCHFV [1–3]. In human, this virus leads to a severe infectious disease of high virulence and high case-fatality rate ($\sim 30\%$) classified by World Health Organization (WHO) as a biosafety level 4 pathogen [4–7]. The incidence of such a life threatening disease is well distributed all around the world including Asian, Middle East, Southeast Europe and African countries [7–11].

CCHFV is a tri-segmented negative sense RNA virus with S (small), M (medium) and L (large) segments [12, 13]. These segments of RNA in CCHFV encode nucleoprotein, glycoprotein and polymerase, respectively [14–17]. As in other negative sense RNA viruses, RNA of CCHFV is enwrapped by nucleoproteins forming a complex called ribonucleoprotein particle (RNP) which protects genomic RNA against degradation by host cell nucleases and helps packaging the newly synthesized genomic RNA to form virions [15, 18–21].

The CCHFV nucleoprotein comprises 482 amino acid residues in its primary structure and 18 alpha helices in the secondary structure as its predominant regular structure. The alpha helices include about 54 % of amino acid residues whereas beta strand structures include only 2.5 % of amino acids [22, 23]. In the three dimensional structure, nucleoprotein has two domains; a stalk and a head domains [24].

Sequence analyses indicated the existence of highly positively charged regions along the nucleoprotein that suggest its ability to bind negative back-bones of RNA molecules [19, 25, 26]. A region comprising residues Lys³³⁹, Lys³⁴³, Lys³⁴⁶, Arg³⁸⁴, Lys⁴¹¹, His⁴⁵³ and Gln⁴⁵⁷ and a region with Arg¹³⁴, Arg¹⁴⁰ and Gln⁴⁶⁸ in the head domain, on the one hand, and a region with Arg¹⁹⁵, His¹⁹⁷, Lys²²², Arg²⁸² and Arg²⁸⁶ in the stalk domain, on the other hand, provide suitable binding motifs for a genomic RNA [24, 26]. Considering the low sequence similarity of nucleoproteins in *Bunyaviridae* family, it is thought that the binding properties of nucleoproteins are not sequence specific [5–7, 15, 16]. The genomic RNA of negative sense RNA viruses, as in CCHFV, is in opposite sense to mRNA. Therefore, it should be transcribed to complementary RNA by viral polymerase prior to becoming suitable for translation by host cell machinery [26–28].

It is well known that the viral polymerase specifically recognizes and binds the genomic RNA encapsidated in RNPs rather than the naked one and transcribe it to a positive RNA [29–31]. However, the presence of the nucleoprotein is essential for safe elongation of the transcription process as it likely plays a pivotal helicase activity to prevent pairing of the genomic RNA and the newly synthesized complementary RNA as a double stranded structure [30, 32–34]. There are reports suggesting that the helicase activity results from the nucleoprotein different affinities for negative and positive sense RNA strands i.e. with higher affinity for a negative (viral) RNA and lower affinity for a positive sense (complementary) RNA [35–38].

In the present work, by analyzing genomic sequences of RNA viruses either with negative or positive sense, performing different docking experiments and carrying out molecular dynamic (MD) simulations, we undertook to

study the mechanism conferring different affinities to CCHFV nucleoprotein for negative and positive sense RNAs'. The outcomes of this study may give a better understanding of packaging mechanism for CCHFV negative sense genomic RNA.

2 Methods and Materials

2.1 Coordinate Structures Preparation

The available two coordinate structures for monomeric and trimeric forms of CCHFV nucleoprotein were obtained from Protein Data Bank archive under PDB ID numbers: 3U3I, 4AQG and 4AQF respectively [39]. These structures were constructed based on crystallographic data at 2–3 Å of resolutions prior to being used as starting structures to perform experiments.

For the purpose of this study, the negative and positive sense RNAs were prepared by dissociating a short double strand RNA of 21 nucleotides obtained from PDB (ID: 2KE6) and a long double stranded RNA of 42 base pairs which had been constructed using ArgusLab, a free docking software [40]. The sequence of the negative strand of the short RNA was: 5'AAUUUAAAAAUACAAUCAAGC3' with a purine/pyrimidine ratio of 1.65, whereas that of positive strand was: 5'GCUUGAUUGUAUUUUUAAAU U3' with a purine/pyrimidine ratio of 0.615. Also, the sequence of the negative strand of the long RNA was: 5'GUGACGUGACGUGACGUGACGUGAGUGACGUG-ACGUGACGUG3' with a purine to pyrimidine ratio of 1.47, whereas that of the positive strand was: 5'CACGUCACGUCACGUCACUCACGUCACGUCCA-CGUCACGUCAC3' with a purine to pyrimidine ratio of 0.68.

The coordinate structures of both short and long RNAs (negative and positive senses) were hydrated and energy minimized by gromacs prior to being used in docking or MD experiments. It should be noted that these positive and negative RNAs were selected to have distant purine to pyrimidine ratio enough to magnify their difference in terms of their *physicochemical properties in a bid to elucidate their differential behaviors*.

2.2 Docking Experiments

Docking experiments were carried out on hydrated and optimized structures of CCHFV nucleoproteins and RNA molecules using Hex software version 6.3 (<http://www.loria.fr/~ritchied/hex/>) [41]. Docking results were scored based on their energy and the first 100 docked structures were averaged and used for energy calculations.

2.3 Molecular Dynamic Experiments

For MD experiments, the best docked structures of CCHFV nucleoprotein in complex with short and long negative or positive sense RNAs were placed separately in the center of rectangular boxes having dimensions of $6.97 \times 7.48 \times 9.60$, $6.97 \times 7.50 \times 9.65$, $9.64 \times 9.65 \times 13.20$ and $8.94 \times 9.85 \times 13.49$ nm respectively. The simulated boxes were filled and covered with a water layer modeled by the SPC/E model within a radius of 1.0 nm.

2.4 Simulation Settings

Setting up the systems: GROMACS 4.5.5 with double precision implemented on UBUNTU-12.04 were used for MD simulations using amber99sb-ildn force field for parameterizations [42]. Systems charge neutralities were checked by GROMACS machine and neutralized by adding sodium ions. Systems were optimized for system constituents including solvent, ions and hydrogen atoms by more than 1,400 steps of energy minimization using steepest descent algorithm with the total energy below 350 kJ/mol. LINCS and SETTLE algorithms were used to constrain bond length and water geometry. Short 500 ps—MD simulations with bond restraints were carried out prior to full length simulations [43]. Final 20 ns—MD simulations were performed at 37 °C using Berendsen thermostat and at 1 atmosphere pressure using Berendsen barostat for coupling temperature and pressure respectively. Electrostatic interactions were treated with particle mesh Ewald (PME) and the neutral pH was set using Asp, Glu, Arg, and Lys amino acids in ionized forms [44, 45].

2.5 Statistical Analysis

The results were analyzed statistically using the Statistical Package for the Social Science (SPSS-PC, version 15. SPSS, Inc., Chicago, IL). The parameters were considered significantly different at $p < .05$.

3 Results

The transcription of viral negative sense RNA by a polymerase to a complementary positive sense RNA results in replacement of purine nucleotides (G and A) with their pyrimidine counterparts (C and U) and vice versa. Depending on proportional nucleotide content of a parental RNA, the complementary chain may differ in its overall purine to pyrimidine ratio when contrasted to the parental chain. In order to calculate purine/pyrimidine ratios of negative and positive RNAs, we analyzed a large set of

genomic information of well characterized negative and positive sense RNA viruses in terms of their corresponding sequences. Tables 1 and 2 list the negative and positive sense RNA viruses obtained from www.ncbi.nlm.nih.gov/nucleotide for the same purpose.

Sequence analyses showed that the average purine/pyrimidine ratios are 1.10 ± 0.06 and 0.96 ± 0.05 (Average \pm SD) for negative and positive sense RNA respectively. Although, the difference seemed trivial, but statistical analyses indicated significantly higher ratio for negative sense RNA compared to positive sense one at p value $< .05$. Given their higher abundance, the larger purine bases bind stronger (via stacking and/or hydrogen bonding) to counter amino acids residues on CCHFV nucleoprotein than pyrimidine bases. In order to test this hypothesis, we first constructed four mono nucleotides of G, A, C and U by ArgusLab software and optimized their chemical structures. The optimized nucleotides then were docked to CCHFV nucleoproteins (3U3I and 4AQG for monomeric and 4AQF for trimeric forms) using Hex 6.3 and blind mode of docking. Binding energy of the best 100 docked structures then averaged as binding affinity. As expected, our data showed that the binding energy of mononucleotides to 3U3I nucleoprotein were -454.58 , -128.84 , -122.78 and -122.56 kJ/mol for G, A, U and C respectively. The same patterns were seen when using monomeric 4AQG and trimeric 4AQF nucleoproteins as targets (data not shown). As evidenced, purine nucleotides show higher binding energies than pyrimidine nucleotides. Furthermore, the content of each nucleotides in the genomes of negative sense RNA viruses calculated as percentage (mean \pm SD) were 31.20 ± 2.44 , 21.09 ± 2.07 , 26.46 ± 1.97 and 21.22 ± 2.09 for A, G, T and C nucleotide respectively, showing prominently higher content for A nucleotide. The same calculation for genome composition of positive sense RNA viruses indicated 28.49 ± 2.10 , 20.45 ± 2.57 , 31.74 ± 3.14 and 19.31 ± 3.87 % for A, G, T and C nucleotide contents respectively. Interestingly, when the actual count of genomic nucleotides for negative and positive sense RNA viruses (as listed in Tables 1, 2) are multiplied by their corresponding energies, the resulting total binding energy of negative sense RNA become significantly higher than that of positive sense RNA by 0.5 kJ/mol per nucleotide. This energy builds up a significant barrier for positive sense RNA to be packaged, as genetic material, by CCHFV nucleoprotein instead of/or concomitantly with negative sense RNA during virus assembly. This confers the negative sense RNA with a greater affinity for CCHFV nucleoproteins and hence preferential binding. As evidenced, this binding preference is driven by higher proportion of adenine rather than other nucleotides. What remains to be answered, at this stage, is whether the preferential binding of CCHFV nucleoproteins

Table 1 Negative sense RNA viruses for four different families obtained from www.ncbi.nlm.nih.gov/nucleotide

Family	Genus	Negative stranded viruses
<i>Bornaviridae</i>	<i>Bornavirus</i>	Borna disease virus
<i>Bunyaviridae</i>	<i>Nairovirus</i>	Crimean-Congo hemorrhagic fever
<i>Bunyaviridae</i>	<i>Nairovirus</i>	Rift Valley fever virus
<i>Filoviridae</i>	<i>Ebolavirus</i>	Bundibugyo ebolavirus
<i>Filoviridae</i>	<i>Ebolavirus</i>	Cote d'Ivoire ebolavirus
<i>Filoviridae</i>	<i>Marburgvirus</i>	Marburg marburgvirus
<i>Filoviridae</i>	<i>Ebolavirus</i>	Zaire ebolavirus
<i>Paramyxoviridae</i>	<i>Avulavirus</i>	Avian paramyxovirus 3
<i>Paramyxoviridae</i>	<i>Avulavirus</i>	Avian paramyxovirus 6
<i>Paramyxoviridae</i>	<i>Avulavirus</i>	Avian paramyxovirus 7
<i>Paramyxoviridae</i>	<i>Avulavirus</i>	Avian paramyxovirus 9
<i>Paramyxoviridae</i>	<i>Morbillivirus</i>	Canine distemper virus
<i>Paramyxoviridae</i>	<i>Morbillivirus</i>	Dolphin morbillivirus
<i>Paramyxoviridae</i>	<i>Ferlavirus</i>	Fer-de-lance virus
<i>Paramyxoviridae</i>	<i>Henipavirus</i>	Hendra virus
<i>Paramyxoviridae</i>	<i>Rubulavirus</i>	Human parainfluenza virus 2
<i>Paramyxoviridae</i>	<i>Rubulavirus</i>	Human parainfluenza virus 4b
<i>Paramyxoviridae</i>	<i>Rubulavirus</i>	Mapuera virus
<i>Paramyxoviridae</i>	<i>Morbillivirus</i>	Measles virus
<i>Paramyxoviridae</i>	<i>Rubulavirus</i>	Mumps virus strain
<i>Paramyxoviridae</i>	<i>Avulavirus</i>	Newcastle disease virus
<i>Paramyxoviridae</i>	<i>Henipavirus</i>	Nipah virus
<i>Paramyxoviridae</i>	<i>Respirovirus</i>	Bovine parainfluenza virus 3
<i>Paramyxoviridae</i>	<i>Morbillivirus</i>	Peste des petits ruminants virus
<i>Paramyxoviridae</i>	<i>Rubulavirus</i>	Simian virus 41
<i>Paramyxoviridae</i>	<i>Pneumovirus</i>	Human respiratory syncytial virus
<i>Paramyxoviridae</i>	<i>Metapneumovirus</i>	Avian metapneumovirus
<i>Rhabdoviridae</i>	<i>Lyssavirus</i>	European bat lyssavirus 1
<i>Rhabdoviridae</i>	<i>Cytorhabdovirus</i>	Lettuce necrotic yellows virus
<i>Rhabdoviridae</i>	<i>Lyssavirus</i>	Rabies virus
<i>Rhabdoviridae</i>	<i>Ephemerovirus</i>	Bovine ephemeral fever virus
<i>Rhabdoviridae</i>	<i>Novirhabdovirus</i>	Infectious hematopoietic necrosis virus
<i>Rhabdoviridae</i>	<i>Nucleorhabdovirus</i>	Potato yellow dwarf virus
<i>Rhabdoviridae</i>	<i>Perhabdovirus</i>	Perch rhabdovirus
<i>Rhabdoviridae</i>	<i>Sigmavirus</i>	<i>Drosophila melanogaster</i> sigma virus
<i>Rhabdoviridae</i>	<i>Vesiculovirus</i>	Vesicular stomatitis Indiana virus
<i>Rhabdoviridae</i>	<i>Tibrovirus</i>	Tibrogargan virus

to the negative sense RNA is a sequence specific property or not. Using single-stranded oligonucleotides with 2-8 nucleotides long, our serial docking experiments revealed that the binding interface of CCHFV nucleoprotein and RNA involves only 3–4 nucleotides and their counter residues. Therefore, to study the effect of nucleotide sequence on binding energy, we constructed 64 possible sequences of trimeric and 256 possible sequences of tetrameric nucleotides using G, A, C and U nucleotides and optimized their chemical structures. These nucleotide structures were then docked to optimized structures of CCHFV

nucleoproteins (3U3I and 4AQG for monomeric and 4AQF for trimeric forms). ANOVA analysis of the best 100 resulted structures from each docking experiments showed that the binding energies of trimeric and tetrameric sequences are significantly sequence independent (p value $< .05$). In other words, the binding energy of oligonucleotides is only dependent to the total content of purine bases but not to the sequence of nucleotides in RNA string.

In the next step, a short RNA (21 nucleotides) as well as a long RNA (42 nucleotide) both in negative and positive sense states were used and docked to monomeric (3U3I and

Table 2 Positive sense RNA viruses for seven different families obtained from www.ncbi.nlm.nih.gov/nucore

<i>Family</i>	<i>Genus</i>	Positive Stranded Viruses
<i>Arteriviridae</i>	<i>Arterivirus</i>	Equine arteritis virus
<i>Coronaviridae</i>	<i>Alphacoronavirus</i>	Avian infectious bronchitis virus
<i>Coronaviridae</i>	<i>Gammacoronavirus</i>	Beluga Whale coronavirus SW1
<i>Coronaviridae</i>	<i>Deltacoronavirus</i>	Bulbul coronavirus
<i>Coronaviridae</i>	<i>Torovirus</i>	Equine torovirus ne
<i>Coronaviridae</i>	<i>Alphacoronavirus</i>	Feline coronavirus
<i>Coronaviridae</i>	<i>Deltacoronavirus</i>	Munia coronavirus HKU13
<i>Coronaviridae</i>	<i>Betacoronavirus</i>	Murine coronavirus
<i>Coronaviridae</i>	<i>Alphacoronavirus</i>	Porcine epidemic diarrhea virus
<i>Coronaviridae</i>	<i>Betacoronavirus</i>	SARS coronavirus
<i>Coronaviridae</i>	<i>Alphacoronavirus</i>	Scotophilus bat coronavirus 512
<i>Coronaviridae</i>	<i>Alphacoronavirus</i>	TGEV Purdue P115
<i>Coronaviridae</i>	<i>Deltacoronavirus</i>	Thrush coronavirus HKU12
<i>Coronaviridae</i>	<i>Bafinivirus</i>	White bream virus
<i>Dicistroviridae</i>	<i>Aparavirus</i>	Acute bee paralysis virus
<i>Dicistroviridae</i>	<i>Cripavirus</i>	Cricket paralysis virus
<i>Iflaviridae</i>	<i>Iflavirus</i>	Infectious flacherie virus
<i>Marnaviridae</i>	<i>Marnavirus</i>	Heterosigma akashiwo RNA virus
<i>Mesoniviridae</i>	<i>Alphamesonivirus</i>	Nam Dinh virus
<i>Picornaviridae</i>	<i>Aphthovirus</i>	Foot-and-mouth disease virus
<i>Picornaviridae</i>	<i>Aquamavirus</i>	Aquamavirus A
<i>Picornaviridae</i>	<i>Avihepatovirus</i>	Duck hepatitis A virus
<i>Picornaviridae</i>	<i>Cardiovirus</i>	Encephalomyocarditis virus
<i>Picornaviridae</i>	<i>Cosavirus</i>	Cosavirus A
<i>Picornaviridae</i>	<i>Dicipivirus</i>	Cadicivirus A
<i>Picornaviridae</i>	<i>Enterovirus</i>	Enterovirus C
<i>Picornaviridae</i>	<i>Erbovirus</i>	Equine rhinitis B virus
<i>Picornaviridae</i>	<i>Hepatovirus</i>	Hepatitis A virus
<i>Picornaviridae</i>	<i>Kobuvirus</i>	Aichivirus A
<i>Picornaviridae</i>	<i>Parechovirus</i>	Human parechovirus
<i>Picornaviridae</i>	<i>Salivirus</i>	Salivirus A
<i>Picornaviridae</i>	<i>Sapelovirus</i>	Porcine sapelovirus
<i>Picornaviridae</i>	<i>Senecavirus</i>	Seneca Valley virus
<i>Picornaviridae</i>	<i>Teschovirus</i>	Porcine teschovirus
<i>Picornaviridae</i>	<i>Tremovirus</i>	Avian encephalomyelitis virus
<i>Roniviridae</i>	<i>Okavirus</i>	Gill-associated virus

4AQG) and trimeric (4AQF) forms of CCHFV nucleoproteins. Figure 1 plots average binding energies obtained for the best 100 structures of each negative and positive sense RNAs with different nucleoprotein structures. As indicated, negative sense RNA exhibited significantly higher binding energy than positive sense RNA (p value $< .05$) with either monomeric (3U3I and 4AQG) or trimeric (4AQF) nucleoproteins as docking targets. This finding suggests that CCHFV nucleoproteins bind more tightly to negative sense RNA than positive one. Figure 1, also, shows that irrespective of their senses, long RNAs have comparatively higher affinities to nucleoprotein than short RNAs.

Based on the results of aforementioned docking experiments, we then selected CCHFV nucleoproteins-RNA complexes of maximum binding energies for positive and negative sense RNAs (both short and long) to carry out MD simulations. The complexes were then placed in cubic boxes filled with SPCE water at 37 °C and 1 atmosphere pressure and energy minimized to lower than 300 kJ/mol prior to MD simulations. Table 3 summarizes simulation history for these systems.

The simulation trajectories, all of 20 ns duration, were then processed using gromacs commands. In all cases, the same outputs were obtained for short and long RNA-nucleoprotein complexes. Figure 2a shows root mean

Fig. 1 Binding energy for CCHFV nucleoproteins (monomers of 3U3I, 4AQG and trimer of 4AQF) to negative and positive sense RNAs' with short and long length obtained from blind docking experiments using Hex 6.1 software

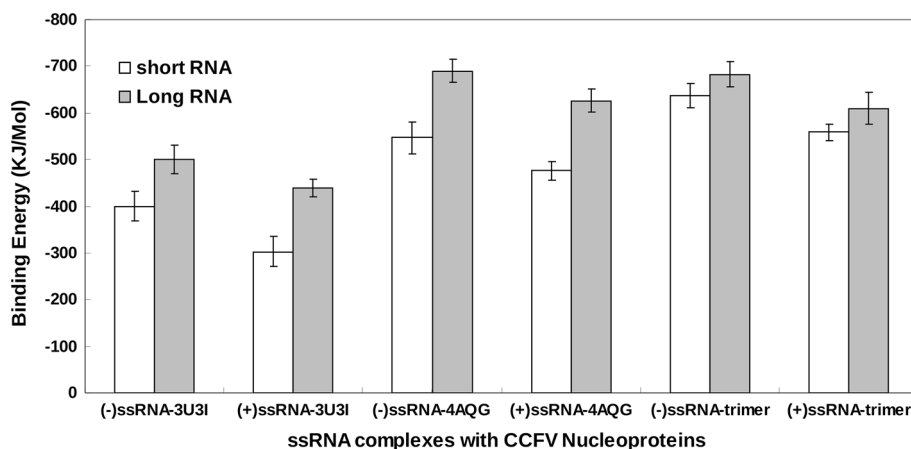


Table 3 simulation history for the simulated systems of CCHFV with short and long RNA either with positive or negative sense nature

	Short RNA		Long RNA	
	Negative sense	Positive sense	Negative sense	Positive sense
Box size	$6.97 \times 7.50 \times 9.65$	$6.97 \times 7.48 \times 9.65$	$9.64 \times 9.65 \times 13.20$	$8.94 \times 9.85 \times 13.49$
Total atoms	50,253	50,224	121,884	118,164
No. of solvent	14,069	14,072	37,745	36,507
Sodium ions	18	18	36	36
Temperature	37	37	37	37
Pressure	1 bar	1 bar	1 bar	1 bar
Total time (ns)	20	20	20	20
Time step	0.001 ps	0.001 ps	0.001 ps	0.001 ps

square displacement (RMSD) curve of CCHFV nucleoprotein during simulation of its complexes with both negative and positive sense RNAs. The curve indicates that CCHFV nucleoprotein undergoes similar patterns of structural alterations in the presence of either senses of RNA which ultimately converge towards equilibrated states. As the structural alterations exerted by MD force are similar during these simulations, therefore, both systems may be used for further comparative studies.

Figure 2b illustrates RMSD change of RNA chains against protein back-bone for both negative and positive sense RNA complexes. This curve provides a good tool to explore the movement of RNA relative to protein back-bone during simulation. The initial phase of RMSD curve (0–5,000 ps) marks a sharper increase in RMSD for the negative sense RNA in contrast to positive sense RNA. Towards formation of RNA-CCHFV nucleoprotein complexes, the faster movement of negative sense RNA means that this docked RNA chain has a closer position to that in the final conformation and so proceeds faster to reach the final state. After this initial phase has elapsed, the negative sense RNA reaches a more stable state which restricts further movement, hence, its RMSD curve progresses with

steeper slope than that of its counter RNA (Fig. 2b). However, the positive sense RNA does not reach a stable state until about 12,000 ps of simulation after which both systems of RNA chains attain certain stable conformations with no further reasonable increase in RMSD values.

The root mean square fluctuation (RMSF) of nucleoprotein alpha carbons during 20 ns of simulation is shown in Fig. 3. The RMSF curve is a very useful index for protein flexibility during simulation with hot (or more flexible) points being at the peaks of the curve. As depicted, the total RMSF of the negative sense RNA is significantly lower (~ 8 tenth) than that of the positive sense RNA. The lower RMSF of CCHFV nucleoprotein complex with negative sense RNA compared to its complex with positive sense RNA indicates lower flexibility and therefore more stable conformation of the former complex. This confirms the postulated preferential binding of the negative sense RNA to CCHFV nucleoprotein based on binding energy calculations (Fig. 1).

The trend of mean square displacement (MSD) curve (Fig. 4) for the negative and positive sense RNAs during simulation is a reliable index for the extent of movement or diffusion of RNA chains inside CCHFV nucleoprotein. This

Fig. 2 **a** Propagation in RMSD curve of CCHFV nucleoprotein against its initial state with time obtained for 20 ns simulation at 37 °C, pH 7, 1 atmosphere of pressure and in explicit water box. **b** Change in RMSD curve of RNA in complex with CCHFV nucleoprotein in contrast to nucleoprotein backbone extracted from 20 ns simulation at 37 °C, pH 7, 1 atmosphere of pressure and in explicit water box

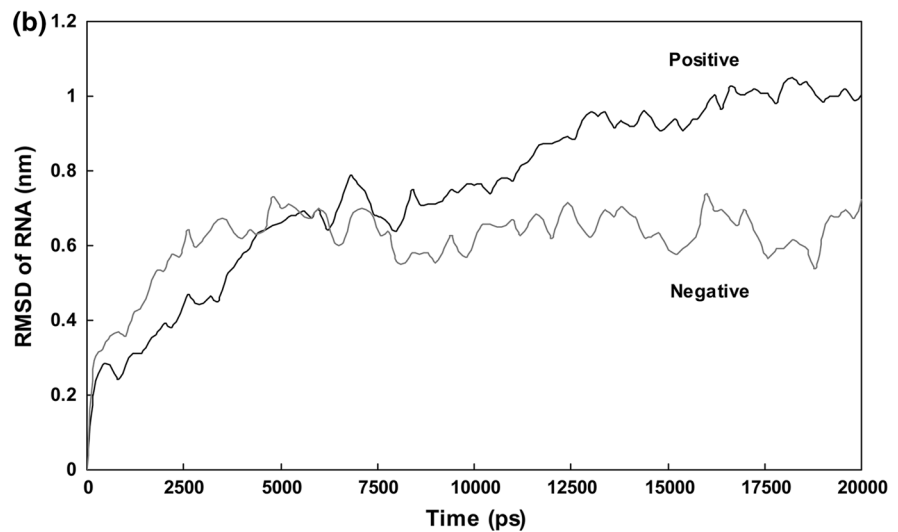
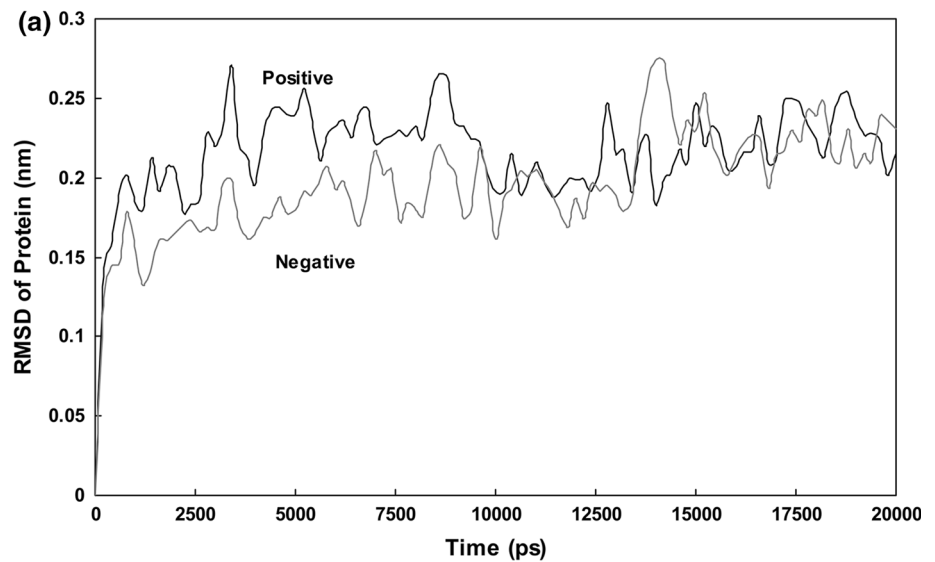
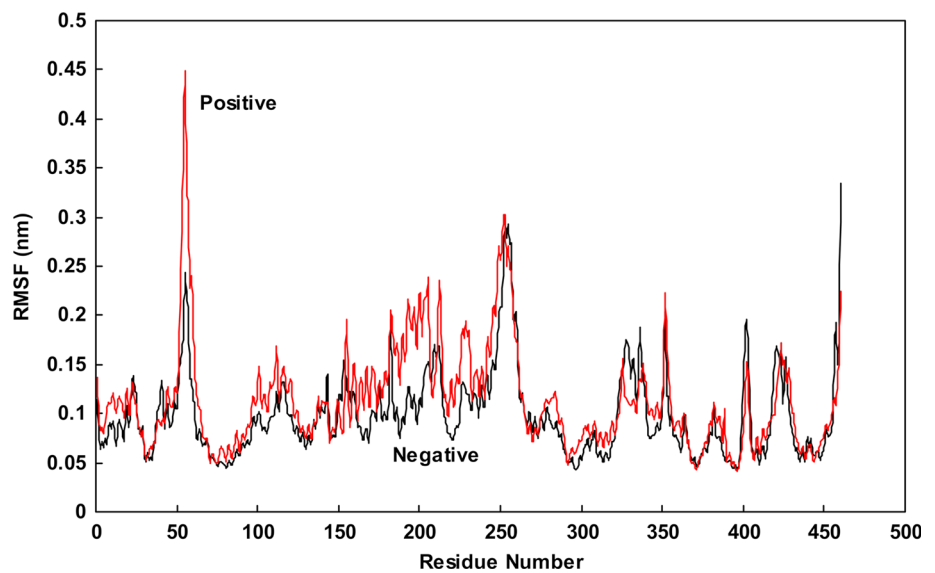


Fig. 3 RMSF plot of CCHFV nucleoprotein in complex with positive and negative sense RNA obtained from 20 ns simulation in 37 °C and 1 atmosphere of pressure in SPCE water box



parameter indirectly reflects the stability of RNA-nucleoprotein complex against applied MD forces. The sharper increase of MSD curve for the positive sense RNA indicates reduced stability of its complex with CCHFV nucleoprotein and its free movement inside protein during simulation. This finding is yet another indication on the stability of the negative sense RNA-nucleoprotein complex.

Figure 5 shows the hydrophobic solvent accessible surface (SAS) of the nucleoprotein for complex formation with negative and positive RNA during simulation. Increase in SAS during simulation could be interpreted as to be the result of structural alterations of nucleoprotein which leads to orientation of hydrophobic residues towards outside for RNA binding. Showing an elevated SAS curve, the negative sense RNA seems to exert more structural alterations in CCHFV nucleoprotein during simulation than its counter RNA does. This is again another indication of a stronger binding of the negative sense RNA with the nucleoprotein. The stronger binding of the negative sense RNA is also confirmed by a significant decrease in the content of the protein secondary structure of alpha helix (p value $< .05$) from 35.14 ± 2.5 to 30.2 ± 3.0 % (mean \pm SD) as shown in Fig. 6. This finding also is in agreement with increased changes in nucleoprotein structure caused by the presence of negative sense RNA.

Figure 7a represents the final conformation of positive and negative sense RNA complexes with CCHFV nucleoprotein. In order to clarify the binding pattern, we only showed the secondary structure elements of the protein instead of showing all atoms. Helices of α_5 and α_6 which are placed in the vicinity of CCHFV binding site are shown in blue, while RNA molecules are shown as balls and sticks. As seen, negative sense RNA (right scheme) fits

better to CCHFV binding site in the region between head and stalk domains. In contrast, positive sense RNA (left scheme) is placed somewhat far from binding site. Figure 7b shows more details about structures of the binding sites. The positively charged groups lining the binding sites which make electrostatic interactions with negative backbone of RNA are labeled. These binding sites are extracted from the first and last frame of simulated systems trajectories for positive (top couple) and negative (bottom couple) sense RNA complexes using WebLab Viewer Lite 4.0 software. As depicted, in the top-left image, the positive sense RNA binds first to Lys¹¹⁹, Arg¹⁷⁷, Arg¹⁹⁵, Arg²²⁵ and Lys⁴⁷³. After 20 ns simulation, RNA experiences a slight alteration (top right image) and gets attached to an additional positive residue of Arg²⁰². While, in negative sense RNA, the binding site (bottom-left image) comprises Lys¹¹⁹, Arg¹⁷⁷, Arg¹⁹⁵, Arg²²⁵ and Lys²⁹² and Arg²⁹⁸ with two more positive residues. This finding verifies that the negative sense RNA binds more tightly to CCHFV nucleoprotein. This is because, upon performing 20 ns simulation, the negative sense RNA comes in contact with three extra positive residues including Lys²⁵¹, Lys²⁸² and Lys²⁸⁶ (bottom right). Figure 7a–b reconfirm the fact that the negative sense RNA binds better and fits more effectively than the positive sense RNA to its binding crevice throughout simulation.

4 Discussion

The agent of the Crimean-Congo Hemorrhagic Fever is a negative sense RNA virus which causes one of most life threatening and lethal infectious diseases in human [10, 11]. This virus is, therefore, worthy to be extensively

Fig. 4 Mean Square Displacement curve of positive and negative sense RNA during simulations of their complexes with CCHFV nucleoprotein for 20 ns at 37 °C, pH 7, 1 atmosphere of pressure and in explicit water box

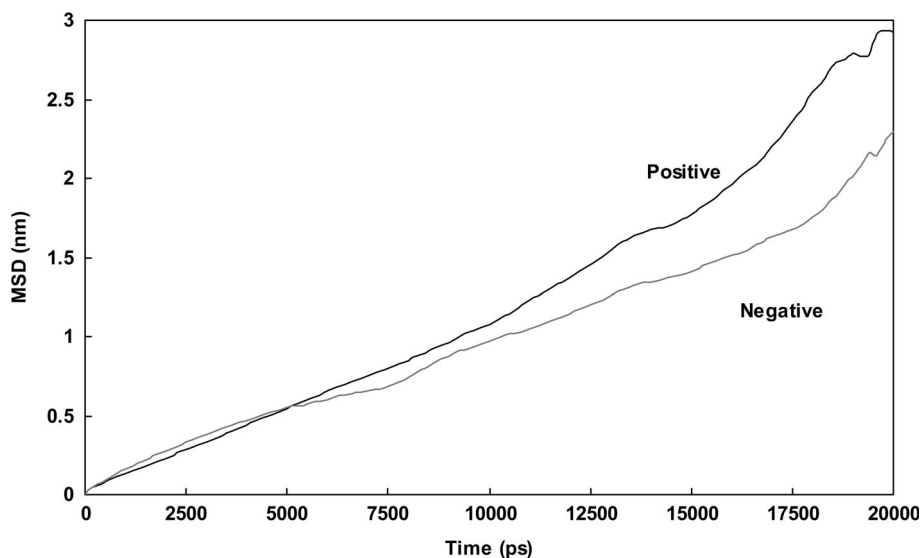


Fig. 5 Hydrophobic solvent accessible surface (SAS) plots for CCHFV nucleoprotein complex with positive and negative sense RNA with time from our simulations at 37 °C, pH 7, 1 atmosphere of pressure and in the presence SPCE water box

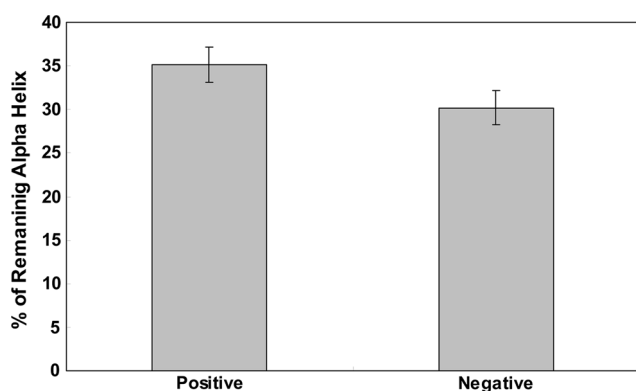
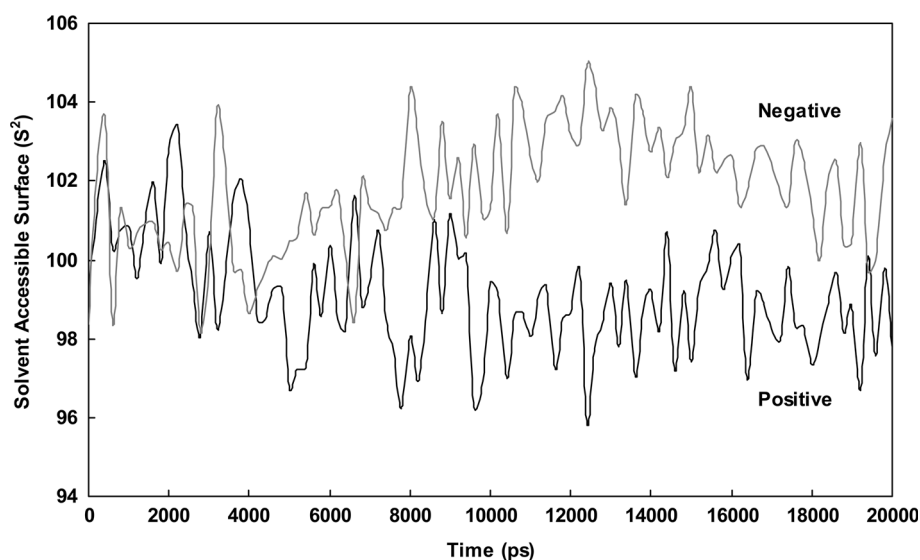


Fig. 6 Percent of remaining alpha helix structures of CCHFV nucleoproteins after simulation of their complexes with negative and positive sense RNA for 20 ns in contrast to their initial structures

studied using different methodologies [24–26]. The main objective of the present study was to pick up some useful hints for better understanding of infection process of such a terrible disease. It seemed, therefore, sensible to start with the negative sense properties of CCHFV and its ability to be enwrapped by viral nucleoprotein [15, 16].

The preferential packaging of the negative sense RNA by CCHFV nucleoprotein in viral assembly seems to be assisted by specific structure recognition characteristics which differentiates between negative sense RNA and positive sense one. This differentiation may be performed based on nucleotides composition and/or their assortment (sequence) within RNA chains [42, 44–47]. In the first run, we collected the complete sequences of genomic RNA for positive and negative senses viruses listed in Tables 1 and 2 [39]. Using Microsoft Excel, we calculated the total numbers of four nucleotides including G, A, C and U for

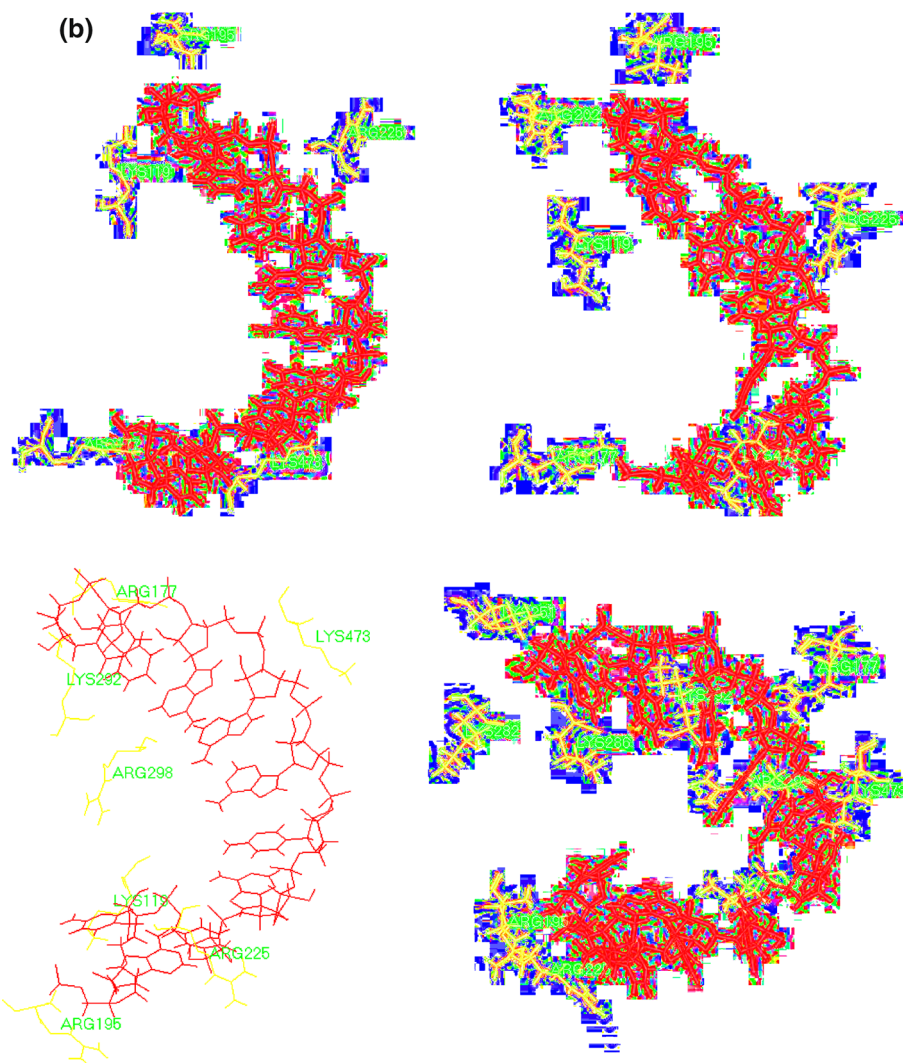
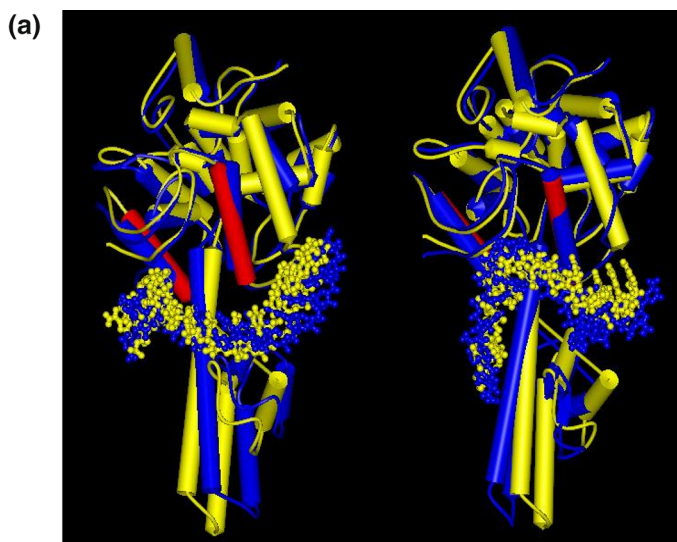
each virus listed in the tables. Then, we calculated the ratio of purines to pyrimidines for the negative and positive sense RNA viruses.

Our results show that the negative sense RNA has higher ratio of purines to pyrimidines (ratio > 1) than its counter positive sense RNA which has a ratio of less than unity (ratio < 1). Having bigger bases as well as polar groups, purine nucleotides such as G with highest binding energy and A with higher frequency amongst nucleotides interact stronger with their amino acid counterparts on CCHFV nucleoprotein. Our docking experiments confirm higher binding energies of 0.5 kJ/mol per nucleotide for negative sense RNA. This energy seems to be the main cause for preferential binding of the negative RNA to CCHFV nucleoprotein.

Our MD simulations indicate that the negative sense RNA makes a more stable complex with CCHFV nucleoprotein (Fig. 2a–b). The retained diffusion of the negative RNA inside nucleoprotein with lowered slope of MSD curve (Fig. 4) also confirms this claimed stability. Attenuated curve of RMSF for CCHFV nucleoprotein in complex with the negative sense RNA compared to the positive sense RNA (Fig. 3) indicates the formation of a more stable complex with lower flexibility in CCHFV nucleoprotein. Moreover, curves of the hydrophobic solvent accessible surface (SAS) (Fig. 5) and the significant decrease in secondary structure (Fig. 6) confirmed stronger interactions between the negative sense RNA and CCHFV nucleoprotein during simulation hence inducing structural alterations of CCHFV nucleoprotein. Finally the extracted structures of binding site residues indicated the presence of more positive residues in CCHFV for negative sense RNA binding (Fig. 7a–b).

Fig. 7 a Binding pattern of positive (*left*) and negative (*right*) sense RNA to binding crevice of CCHFV nucleoprotein, obtained from last frame of simulation trajectories for 20 ns at pH 7, 37 °C, 1 atmosphere and in SPCE water boxes.

b representative of the binding site residues in contact to positive sense RNA before (*top-left*) and after (*top-right*) simulation and to negative sense RNA *bottom-left* and *bottom-right*, before and after simulation, respectively



5 Conclusion

It was shown that Lassa virus nucleoprotein binding site for RNA is closed by α_5 and α_6 helices in closed conformation of trimeric structures. Outward movement of these helices in gating mechanism open the binding site and permit RNA entrance and binding to virus nucleoprotein [46]. Nevertheless the negative and the positive sense RNAs here bind almost to the same binding site reported for Lassa Virus (Fig. 7). However, our simulation trajectories indicate that there is neither a significant movement in α_5 and α_6 helices nor a significant disturbance in residues string of 112–122 to support the gating mechanism. Instead, we conclude that CCHFV nucleoprotein may act via different mechanism. Co-crystallization of CCHFV nucleoproteins with negative and positive sense RNAs, X-ray crystallography and performing MD simulations in conditions similar to those in virus seems to be the only way to elucidate the precise mechanism underlying RNA binding to CCHFV nucleoprotein.

Acknowledgments The financial support of Shahid Chamran University of Ahvaz and Tarbiat Modares University, Tehran, are acknowledged.

Conflicts of interest There are no conflicts of interest to disclose.

References

- Vincent MJ, Sanchez AJ, Erickson BR, Basak A, Chretien M, Seidah NG, Nichol ST (2003) Crimean-Congo hemorrhagic fever virus glycoprotein proteolytic processing by subtilase SKI-1. *J Virol* 77(16):8640–8649
- Nichol ST (2001) Bunyaviruses. In: Knipe DM, Howley PM (eds) *Fields virology*, vol 2. Lippincott Williams and Wilkins, Philadelphia, pp 1603–1633
- Osborne JC, Elliott RM (2000) RNA binding properties of bunyamwera virus nucleocapsid protein and selective binding to an element in the 5' terminus of the negative-sense S segment. *J Virol* 74(21):9946–9952
- Hewson R, Gmyl A, Gmyl L, Smirnova SE, Karganova G, Jamil B, Hasan R, Chamberlain J, Clegg C (2004) Evidence of segment reassortment in Crimean-Congo haemorrhagic fever virus. *J Gen Virol* 85(Pt 10):3059–3070
- Wang Y, Dutta S, Karlberg H, Devignot S, Weber F, Hao Q, Tan YJ, Mirazimi A, Kotaka M (2012) Structure of Crimean-Congo hemorrhagic fever virus nucleoprotein: superhelical homo-oligomers and the role of caspase-3 cleavage. *J Virol* 86(22):12294–12303
- Whitehouse CA (2004) Crimean-Congo hemorrhagic fever. *Antiviral Res* 64(3):145–160
- Ergönül O (2006) Crimean-Congo haemorrhagic fever. *Lancet Infect Dis* 6(4):203–214
- Duh D, Nichol ST, Khristova ML, Saksida A, Hafner-Bratkovic I, Petrovec M, Dedushaj I, Ahmeti S, Avsic-Zupanc T (2008) The complete genome sequence of a Crimean-Congo hemorrhagic fever virus isolated from an endemic region in Kosovo. *Virol J* 5:7
- Elliott RM (2009) Bunyaviruses and climate change. *Clin Microbiol Infect* 15(6):510–517
- Maltezou HC, Papa A (2010) Crimean-Congo hemorrhagic fever: risk for emergence of new endemic foci in Europe? *Travel Med Infect Dis* 8(3):139–143
- Yu XJ, Liang MF, Zhang SY, Liu Y, Li JD, Sun YL, Zhang L, Zhang QF, Popov VL, Li C, Qu J, Li Q, Zhang YP, Hai R, Wu W, Wang Q, Zhan FX, Wang XJ, Kan B, Wang SW, Wan KL, Jing HQ, Lu JX, Yin WW, Zhou H, Guan XH, Liu JF, Bi ZQ, Liu GH, Ren J, Wang H, Zhao Z, Song JD, He JR, Wan T, Zhang JS, Fu XP, Sun LN, Dong XP, Feng ZJ, Yang WZ, Hong T, Zhang Y, Walker DH, Wang Y, Li DX (2011) Fever with thrombocytopenia associated with a novel bunyavirus in China. *N Engl J Med* 364(16):1523–1532
- Schmaljohn CS, Hooper JW (2001) *Bunyaviridae: the viruses and their replication*. In: Knipe DM, Howley PM (eds) *Fields virology*, vol 2. Lippincott Williams and Wilkins, Philadelphia, pp 1581–1602
- Lukashev AN (2005) Evidence for recombination in Crimean-Congo hemorrhagic fever virus. *J Gen Virol* 86(Pt 8):2333–2338
- Rudolph MG, Kraus I, Dickmanns A, Eickmann M, Garten W, Ficner R (2003) Crystal structure of the borna disease virus nucleoprotein. *Structure* 11(10):1219–1226
- Morikawa S, Saijo M, Kurane I (2007) Recent progress in molecular biology of Crimean-Congo hemorrhagic fever. *Comp Immunol Microbiol Infect Dis* 30(5–6):375–389
- Vorou R, Pierrotsakos IN, Maltezou HC (2007) Crimean-Congo hemorrhagic fever. *Curr Opin Infect Dis* 20(5):495–500
- Clerx JP, Casals J, Bishop DH (1981) Structural characteristics of nairoviruses (genus *Nairovirus*, Bunyaviridae). *J Gen Virol* 55 (Pt 1):165–178
- Ferron F, Li Z, Danek EI, Luo D, Wong Y, Coutard B, Lantey V, Charrel R, Canard B, Walz T, Lescar J (2011) The hexamer structure of Rift Valley fever virus nucleoprotein suggests a mechanism for its assembly into ribonucleoprotein complexes. *PLoS Pathog* 7(5):e1002030
- Raymond DD, Piper ME, Gerrard SR, Smith JL (2010) Structure of the Rift Valley fever virus nucleocapsid protein reveals another architecture for RNA encapsidation. *Proc Natl Acad Sci USA* 107(26):11769–11774
- Overby AK, Pettersson RF, Neve EP (2007) The glycoprotein cytoplasmic tail of Uukuniemi virus (Bunyaviridae) interacts with ribonucleoproteins and is critical for genome packaging. *J Virol* 81(7):3198–3205
- Huiskonen JT, Overby AK, Weber F, Grünwald K (2009) Electron cryo-microscopy and single-particle averaging of Rift Valley fever virus: evidence for GN-GC glycoprotein heterodimers. *J Virol* 83(8):3762–3769
- Deyde VM, Khristova ML, Rollin PE, Ksiazek TG, Nichol ST (2006) Crimean-Congo hemorrhagic fever virus genomics and global diversity. *J Virol* 80(17):8834–8842
- Severson W, Xu X, Kuhn M, Senutovitch N, Thokala M, Ferron F, Longhi S, Canard B, Jonsson CB (2005) Essential amino acids of the hantaan virus N protein in its interaction with RNA. *J Virol* 79(15):10032–10039
- Guo Y, Wang W, Ji W, Deng M, Sun Y, Zhou H, Yang C, Deng F, Wang H, Hu Z, Lou Z, Rao Z (2012) Crimean-Congo hemorrhagic fever virus nucleoprotein reveals endonuclease activity in bunyaviruses. *Proc Natl Acad Sci USA* 109(13):5046–5051
- Rudolph MG, Kraus I, Dickmanns A, Eickmann M, Garten W, Ficner R (2003) Crystal structure of the borna disease virus nucleoprotein. *Structure* 11(10):1219–1226
- Albertini AA, Wernimont AK, Muziol T, Ravelli RB, Clapier CR, Schoehn G, Weissenhorn W, Ruigrok RW (2006) Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science* 313(5785):360–363
- Luo M, Green TJ, Zhang X, Tsao J, Qiu S (2007) Structural comparisons of the nucleoprotein from three negative strand RNA virus families. *Virol J* 4:72

28. Ng AK, Zhang H, Tan K, Li Z, Liu JH, Chan PK, Li SM, Chan WY, Au SW, Joachimiak A, Walz T, Wang JH, Shaw PC (2008) Structure of the influenza virus A H5N1 nucleoprotein: implications for RNA binding, oligomerization, and vaccine design. *FASEB J* 22(10):3638–3647
29. Chenavas S, Estrozi LF, Slama-Schwok A, Delmas B, Di Primo C, Baudin F, Li X, Crépin T, Ruigrok RW (2013) Monomeric nucleoprotein of influenza A virus. *PLoS Pathog* 9(3):e1003275
30. Baudin F, Bach C, Cusack S, Ruigrok RW (1994) Structure of influenza virus RNP. I. Influenza virus nucleoprotein melts secondary structure in panhandle RNA and exposes the bases to the solvent. *EMBO J* 13(13):3158–3165
31. Iseni F, Baudin F, Blondel D, Ruigrok RW (2000) Structure of the RNA inside the vesicular stomatitis virus nucleocapsid. *RNA* 6(2):270–281
32. Arnheiter H, Davis NL, Wertz G, Schubert M, Lazzarini RA (1985) Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis. *Cell* 41(1):259–267
33. Honda A, Uéda K, Nagata K, Ishihama A (1988) RNA polymerase of influenza virus: role of NP in RNA chain elongation. *J Biochem* 104(6):1021–1026
34. Ruigrok RW, Crépin T, Kolakofsky D (2011) Nucleoproteins and nucleocapsids of negative-strand RNA viruses. *Curr Opin Microbiol* 14(4):504–510
35. Masters PS, Banerjee AK (1988) Complex formation with vesicular stomatitis virus phosphoprotein NS prevents binding of nucleocapsid protein N to nonspecific RNA. *J Virol* 62(8):2658–2664
36. Curran J, Marq JB, Kolakofsky D (1995) An N-terminal domain of the Sendai paramyxovirus P protein acts as a chaperone for the NP protein during the nascent chain assembly step of genome replication. *J Virol* 69(2):849–855
37. Mavrakis M, Méhouas S, Réal E, Iseni F, Blondel D, Tordo N, Ruigrok RW (2006) Rabies virus chaperone: identification of the phosphoprotein peptide that keeps nucleoprotein soluble and free from non-specific RNA. *Virology* 349(2):422–429
38. Carter SD, Surtees R, Walter CT, Ariza A, Bergeron É, Nichol ST, Hiscox JA, Edwards TA, Barr JN (2012) Structure, function, and evolution of the Crimean-Congo hemorrhagic fever virus nucleocapsid protein. *J Virol* 86(20):10914–10923
39. Stiffin RM, Sullivan SM, Carlson GM, Holyoak T (2008) Differential inhibition of cytosolic PEPCK by substrate analogues. Kinetic and structural characterization of inhibitor recognition. *Biochemistry* 47(7):2099–2109
40. Thompson MA (2004) Molecular docking using ArgusLab, an efficient shape-based search algorithm and the AScore scoring function. ACS meeting, Philadelphia, vol 172, CINF 42, PA
41. Macindoe G, Mavridis L, Venkatraman V, Devignes MD, Ritchie DW (2010) HexServer: an FFT-based protein docking server powered by graphics processors. *Nucleic Acids Res* 38:445–449
42. Dayer MR, Dayer MS (2013) Whiskers-less HIV-protease: a possible Way for HIV-1 deactivation. *J Biomed Sci* 20(1):67
43. Lee B, Richards FM (1971) The interpretation of protein structures: estimation of static accessibility. *J Mol Biol* 55(3):379–400
44. Baker WR, Kintanar A (1996) Characterization of the pH titration shifts of ribonuclease A by one- and two-dimensional nuclear magnetic resonance spectroscopy. *Arch Biochem Biophys* 327(1):189–199
45. Peters GH, van Aalten DM, Svendsen A, Bywater R (1997) Essential dynamics of lipase binding sites: the effect of inhibitors of different chain length. *Protein Eng* 10(2):149–158
46. Hastie KM, Liu T, Li S, King LB, Ngo N, Zandonatti MA, Woods VL Jr, de la Torre JC, Saphire EO (2011) Crystal structure of the Lassa virus nucleoprotein-RNA complex reveals a gating mechanism for RNA binding. *Proc Natl Acad Sci USA* 108(48):19365–19370