RESEARCH ARTICLE

PROTEINS WILEY

# Structures of MERS-CoV macro domain in aqueous solution with dynamics: Impacts of parallel tempering simulation techniques and CHARMM36m and AMBER99SB force field parameters

Ibrahim Yagiz Akbayrak[1] | Sule Irem Caglayan[2] | Serdar Durdagi[3] |
Lukasz Kurgan[4] | Vladimir N. Uversky[5,6] | Burak Ulver[7] |
Havvanur Dervisoğlu[7] | Mehmet Haklidir[7] | Orkun Hasekioglu[7] |
Orkid Coskuner-Weber[2]

[1]Materials Sciences and Technologies, College of Sciences, Turkish-German University, Istanbul, Turkey

[2]Molecular Biotechnology, College of Sciences, Turkish-German University, Istanbul, Turkey

[3]Department of Biophysics, School of Medicine, Bahcesehir University, Istanbul, Turkey

[4]Department of Computer Science, Virginia Commonwealth University, Richmond, Virginia

[5]Department of Molecular Medicine, USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, Florida

[6]Laboratory of New Methods in Biology, Institute for Biological Instrumentation of the Russian Academy of Sciences, Federal Research Center "Pushchino Scientific Center for Biological Research of the Russian Academy of Sciences", Pushchino, Russia

[7]TUBITAK, Turkish Scientific and Technological Research Council, BİLGEM, Istanbul, Turkey

**Correspondence**
Orkid Coskuner-Weber, Department of Biophysics, School of Medicine, Bahcesehir University, Istanbul, Turkey.
Email: weber@tau.edu.tr

## Abstract

A novel virus, severe acute respiratory syndrome Coronavirus 2 (SARS-CoV-2), causing coronavirus disease 2019 (COVID-19) worldwide appeared in 2019. Detailed scientific knowledge of the members of the Coronaviridae family, including the Middle East Respiratory Syndrome Coronavirus (MERS-CoV) is currently lacking. Structural studies of the MERS-CoV proteins in the current literature are extremely limited. We present here detailed characterization of the structural properties of MERS-CoV macro domain in aqueous solution. Additionally, we studied the impacts of chosen force field parameters and parallel tempering simulation techniques on the predicted structural properties of MERS-CoV macro domain in aqueous solution. For this purpose, we conducted extensive Hamiltonian-replica exchange molecular dynamics simulations and Temperature-replica exchange molecular dynamics simulations using the CHARMM36m and AMBER99SB parameters for the macro domain. This study shows that the predicted secondary structure properties including their propensities depend on the chosen simulation technique and force field parameter. We perform structural clustering based on the radius of gyration and end-to-end distance of MERS-CoV macro domain in aqueous solution. We also report and analyze the residue-level intrinsic disorder features, flexibility and secondary structure. Furthermore, we study the propensities of this macro domain for protein-protein interactions and for the RNA and DNA binding. Overall, results are in agreement with available nuclear magnetic resonance spectroscopy findings and present more detailed insights into the structural properties of MERS CoV macro domain in aqueous solution. All in all, we present the structural properties of the aqueous MERS-CoV macro domain using different parallel tempering simulation techniques, force field parameters and bioinformatics tools.

KEYWORDS
bioinformatics, force field parameters, H-REMD simulations, MERS-CoV, simulation protocols, T-REMD simulations

# 1 | INTRODUCTION

Since the first outbreak of the severe acute respiratory syndrome (SARS) in 2003, a fatal viral disease-causing pneumonia and death was first reported in Saudi Arabia in 2012. This virus was named Middle East Respiratory Syndrome Coronavirus (MERS-CoV).[1] The current SARS-CoV-2 infection, coronaviruses and coronavirus-related infection aroused the attention of the entire world. A history of the SARS-CoV outbreak justifies these high-levels of attention. By the time the global SARS-CoV outbreak was contained, the virus spread to 26 countries, infected over 8000 people worldwide and killed almost 800. Similarly, even though MERS-CoV appeared initially in Saudi Arabia, the virus—that was new to humans—spread to several other countries in or near the Arabian Peninsula, Asia, Europe, and the United States.[2] The mortality of MERS was reported to be 4-fold higher than SARS-CoV.[3] In fact, at the end of 2019, there were a total of 2494 laboratory-confirmed cases of MERS-CoV world-wide and the MERS-CoV infection was characterized by the mortality rate of 34.4%. The current version of coronavirus, namely SARS-CoV-2, is infecting and killing more people per day than SARS and MERS combined during their existence.

Despite the history of posing threats to the human health, current knowledge of coronaviruses is rather limited. It is clear that gaining insights into the structural properties of various proteins from MERS-CoV, including the conserved macro domain within the non-structural protein 3 (NSP3), can help better understanding of the *Coronaviridae* family.[4] Since the structural properties of MERS-CoV macro domain in solution with dynamics are still poorly understood, a comparison to SARS-CoV-2 macro domain in solution with dynamics cannot be provided as well.

MERS-CoV belongs to the lineage C of β-coronaviruses (β-CoVs) that includes CoVs isolated from bats and hedgehogs. CoVs use the RNA genome to encode several structural proteins, including the spike glycoprotein (S), membrane protein (M) and nucleocapsid protein (N), and various non-structural proteins (NSPs) to facilitate its fast replication processes.[5] A single large replicase gene encodes the proteins that play a role in viral replication.[4] This gene contains two open reading frames; ORF1a and ORF1b encoding the polyproteins pp1a and pp1b, with the production of pp1b requiring a − 1 ribosome frameshift at the 3′ end of ORF1a.[6] ORF1a encodes viral proteases: main protease (M$^{pro}$) and papain like protease (PL$^{pro}$). These viral proteases play a central role in the cleavage of ORF1a and ORF1b gene products in order to produce functional NSPs.[7]

The largest NSP member of the MERS-CoV genome is the ORF1a-encoded, multifunctional and multidomain protein NSP3 that serves as a major evolutionary selection target in β-CoVs.[8] NSP3 includes N-terminal acidic domain, macro domain, SARS-unique domain, PL$^{pro}$, nucleic acid-binding domain, marker domain (G2M), transmembrane domain, and Y-domain. The macro domain received its name based on the non-histone motif of the histone variant macroH2A, which is a crucial protein module found in eukaryotes, bacteria, and archaea. The macro-domain containing proteins and enzymes play central roles in the regulation of various cellular processes. For instance, the SARS-CoV and MERS-CoV macro domains were shown to possess poly(AD)P-ribose binding affinity, which suggested that this domain regulates cellular proteins that are important for an apoptotic way via poly(ADP)-ribosylation to mediate the host response to infection.[4]

Even though X-ray structure is available for the MERS-CoV macro domain in complex with adenosine monophosphate (AMP), such structure does not capture the impact of the bulk solvent environment on protein structure and dynamics and provides a rather limited view of the underlying structural and functional residue-level characteristics. A detailed understanding of the structural properties of MERS-CoV macro domain in solution will provide the lacking structural information on CoVs and may be used for comparison with SARS-CoV-2 macro domain. In the long run, the information gleaned from such structural studies could help to design more efficient treatments including vaccines and small molecule drugs. Therefore, we present the characterization of the structural properties of MERS-CoV macro domain in aqueous solution at body temperature with dynamics at the atomic level via linking parallel tempering simulations to bioinformatics. We combine these results with several residue-level analyses that focus on the structural flexibility, presence of intrinsically disordered regions, and functional features related to the predisposition for protein-protein and protein-nucleic acid interactions. However, the chosen simulation techniques, simulation protocols and force field parameters may impact the predicted aqueous MERS-CoV macro domain structural properties. Therefore, in this study, we conduct Hamiltonian-replica exchange molecular dynamics simulations and Temperature-replica exchange molecular dynamics simulations and we also look at the impacts of CHARMM36m and AMBER99SB parameters on the calculated structural properties of aqueous MERS-CoV macro domain. For this study, we conducted three extensive different sets of parallel tempering simulations.

# 2 | MATERIALS AND METHODS

Many molecular simulation scenarios require ergodic sampling of conformations. Their energy landscapes may feature many minima and barriers between minima that can be difficult to cross at ambient temperatures over reachable simulation time scales. This means that the corresponding findings are confounded by the choice of initial conditions because such conditions determine the space region that is explored by a simulation.[9] On the other hand, replica exchange simulations seek to enhance the conformational sampling by running numerous independent replicas in different conditions, and periodically exchanging the coordinates of different ensembles (replicas).[9] Usually, temperature is used as the parameters which changes among replicas, which in turn enables conformations trapped in a local minima at a low temperature to escape by passing to a higher temperature replica. Potential energy overlap is required for efficient exchange between neighboring replicas, which results in simulations with large number of replicas especially when we investigate proteins in explicit water. Specifically, for covering a desired temperature

range, replica number grows as the square root of the number of particles, which in turn introduces limitations to the method's potential by means of computational costs. Hamiltonian replica exchange molecular dynamics (H-REMD)[10] provides a possible solution for alleviating temperature-replica exchange molecular dynamics (T-REMD) simulations-based limitations in which the different replicas are treated at a constant temperature while the Hamiltonian are used as a parameter and is reported to be more efficient in protein conformational sampling than T-REMD. As an enhanced technique, it is based on executing simultaneous replicas with different Hamiltonians of the system and enabling exchanges at a given frequency between i and j replicas at neighboring scales m and n with a probability[10] of

$$P(X_i \leftrightarrow X_j) = \min\left[1, \exp\left(\frac{-H_m(X_j) + H_m(X_i)}{k_B T}\right) + \frac{-H_n(X_i) + H_n(X_j)}{k_B T}\right]$$

where H is the Hamiltonian, T is the temperature and X are the coordinates and

$$H_m(X) = \lambda_m H_{pp} + (\lambda_m)^{1/2} H_{ps} + H_{ss}(X)$$

where $H_m$ is the Hamiltonian at scale m and $H_{pp}$, $H_{ps}$ and $H_{ss}$ represent protein, protein-solvent and solvent-solvent interaction energies. $\Lambda_m$ is the scaling factor at scale m whereby $\lambda_m \leq 1.0$. The Gromacs 5.1.4 simulation package[11] in association with PLUMED plugin (version 2.1)[12] were used to conduct the H-REMD simulations. However, the partial tempering script in PLUMED works only for AMBER and OPLS parameters, while for CHARMM parameters the scaling only applies to the epsilon term of the LJ interactions, but not to the CMAP matrix, which is an integral part of CHARMM parameters. This was corrected by using a script (see Supplementary materials section [Appendix S1]). In the H-REMD simulations, we used the CHARMM36m parameters[13] for the MERS-CoV macro domain and the TIP3P model for water.[14] We isolated the initial structure for the MERS-CoV macro domain from the publicly available crystal structure (PDB ID: 5zu7). We applied a water layer of 10 Å with 11 827 water molecules to solvate the macro domain using a cubic box. Energy minimization was performed using both the steepest descent and the conjugate gradient methods. After minimization, 500 ps of each NVT and NPT position restrained dynamics were performed with a restraining force of 1000 kJ/mol·nm² on the non-hydrogen atoms of the domain. This allowed the water molecules to equilibrate around the macro domain, thereby removing bad contacts and bringing the system closer to equilibrium. The final coordinates of the NPT equilibration were used as the initial coordinates for the unrestrained production runs. Twenty-four scaling factors ranging from $\lambda_m = 1.0$ to 0.4 were generated by a geometric distribution, which were used in the H-REMD simulation, amounting to 9.6 μs of cumulative simulation time (400 ns per replica). We use counter ions to neutralize the system. A canonical thermostat with stochastic velocity reassignment[15] with a coupling constant of 0.5 ps was used to keep each system at their requisite temperatures. For the NPT simulations, a Parrinello-Rahman barostat[16] with 1.0 bar pressure and 1.0 ps coupling constant was employed. Both van der Waals and short-range Coulombic interactions were truncated at 12 Å, and the long-range electrostatic interactions were calculated using the particle mesh Ewald method.[17] The neighbor list was updated every 10 steps with a cut-off of 12 Å. The LINCS algorithm[18] was used to constrain all bond lengths during the H-REMD simulations. An exchange between neighboring replicas was attempted every 2 ps, and the coordinates were also saved every 5 ps. The H-REMD were tested for convergence of the replica at $\lambda_m = 1.0$ from H-REMD simulations for further analysis (see Supplementary Materials section [Appendix S1]).

T-REMD simulations[19] were performed between the temperatures ranging from 280 to 320 K using 32 replicas distributed exponentially between these temperatures. We used the CHARMM36m parameters[13] and the AMBER99SB parameters[20] for the MERS-CoV macro domain and the explicit TIP3P model for water[14] for studying the impact of these chosen force field parameters on the predicted structural properties of the macro domain in water.[21] After solvating the macro domain in water by using a 10 Å water layer (11 827 water molecules), we first conduct equilibration simulations for 20 ns (per replica) using the canonical ensemble and then for additional 20 ns (per replica) using the isothermal-isobaric ensemble. We run T-REMD simulations for a total simulation time of 6.4 μs (200 ns per replica). We perform exchanges between replicas every 5 ps with a time step of 2 fs. We save trajectories every 500 steps and the LINCS algorithm was used to constrain all bond lengths in both simulations. The electrostatic and van der Waals interactions were calculated using the particle mesh Ewald (PME) method and the real-space components truncated at 12 Å. We controlled the temperature and pressure using a velocity rescaling algorithm with a relaxation time of 0.1 ps and a Parrinello-Rahman barostat[16] with a relaxation time of 2 ps and we used counterions to neutralize the charges. We calculate the structural properties of the MERS-CoV macro domain from the structures obtained after convergence from the replica closest to physiological temperature (310 K, see Supporting Materials section [Appendix S1]). We calculate the content of the secondary structure components per residue for the aqueous MERS-CoV macro domain utilizing the DSSP program both for data obtained from H-REMD simulations and different sets of T-REMD simulations using CHARMM36m and AMBER99SB parameters (see above).[22] Additionally, we determine the end-to-end distances ($R_{EE}$) and radius of gyration ($R_g$) of the MERS-CoV macro domain in water using all converged trajectories. Based on the relationship between the $R_g$ and $R_{EE}$ values, we apply the k-means clustering method to perform vector quantization and consequently to partition the structural observations into five clusters. We assign each observation to the cluster with the nearest cluster centroid that serves as a prototype of the cluster.[23] This way the structural data space is partitioned into Voronoi cells and the k-means clustering minimizes within cluster variances using squared Euclidean distances. Finally, we compute the root mean square fluctuations for each residue of the MERS-CoV macro domain in water. We compare these results to findings secured by using disorder predictors, which we describe next.

In addition, we perform residue-level analysis of the intrinsic disorder predisposition of the MERS-CoV macro domain and selected functional features related to its protein and nucleic acid binding potential. We evaluate the intrinsic disorder predisposition using a set of commonly utilized and publicly available computational tools, such as PONDR VLXT,[24] PONDR VSL2,[25] PONDR FIT,[26] and IUPred capable of predicting long and short disordered regions.[27-29] Residue-level predisposition of this domain to interact with proteins was evaluated with the state-of-the SCRIBER (SeleCtive pRoteIn-Binding rEsidue pRedictor) method.[30] SCRIBER is currently the most accurate method that predicts protein-binding residues (PBRs), and the only tool that eliminates the recently described issue of the cross-prediction of residues that interact with nucleic acids (RNA and DNA) as PBRs.[31] This allows us to accurately predict PBRs and maintain high specificity of our analysis by limiting contamination of the results by the cross-predictions. We also evaluate the nucleic acid binding potential of the MERS-CoV macro domain with the DRNApred predictor.[32] DRNApred is currently the only method that provides accurate results and successfully eliminates the cross-predictions.[32,33]

## 3 | RESULTS AND DISCUSSION

Figure 1 represents a set of the selected structures of the MERS-CoV macro domain in aqueous media that we obtained from the all-atom H-REMD simulations using the CHARMM36m parameters (A), T-REMD simulations utilizing the CHARMM36m parameters (B) and T-REMD simulations using the AMBER99SB parameters (C) at 310 K replica. Figure 2 depicts the calculated MERS-CoV macro domain secondary structure abundances per residue with dynamics using our own script. Based on these calculations, we detect six α-helix regions in the macro domain of MERS-CoV in water via H-REMD simulations. These are located between Ala25-Cys31(probability; 18%-91%), Gly50-Ser59 (probability, 90%-100%), Ala62-Lys74 (probability; 98%-100%), Val108-Asn119 (probability; 11%-99%), Pro138-Glu148 (probability; 27%-100%) and Gln160-Thr167 (probability; 11%-100%). We detect four $3_{10}$-helix conformation regions via H-REMD simulations located at Pro5-Asn8 (probability; 6%-7%), Ala25-Cys31 (probability; 0.1%-7%), Ala102-Ala104 (probability; all at 36%), Val108-Ala120 (probability; 0.6%-9%). Seven β-sheet regions exist based on H-REMD simulations; Glu10-Thr15 (probability; 29%-100%), Val18-Ile22 (probability, all at 100%), Glu34-Ala41 (probability; 61%-100%), Asp81-Gln86 (probability; 23%-100%), Asn93-Val97 (probability; all at ∼100%), Leu123-Pro127 (probability; 45%-100%) and Arg152-Val157 (probability; 70%-100%). Additionally, we detect ten turn structure regions via H-REMD simulations with higher abundancies and these are Gly1-Glu10 (probability, 9%-95%), Ile14-Cys17 (probability; 7%-100%), Asp24-Gly33 (probability; 2%-100%), Asn42-Leu45 (probability; all at ∼100%), Ala58-Gly61 (probability; 1%-100%), Ala73-Asp81 (probability; 0.2%-100%), Gly87-Asn93 (probability; 1%-100%), Asp101-Lys105 (probability; all at ∼64%), Ala117-Leu123 (probability; 17%-100%) and Leu128-Gly135 (probability; 94%-100%). While the six α-helices were

also observed in the NMR measurements,[34] they also annotate adjacent residues as helical, but this might be related to the buffer used in the experiments. NMR measurements[34] also detected seven β-sheet structure regions in MERS-CoV macro domain. However, we should mention here that the abundancies of α-helix conformations are higher with H-REMD simulations using the CHARMM36m parameters in comparison to NMR experiments.

From the T-REMD simulations using the CHARMM36m parameters (Figure 2), we find again six α-helix regions and these are located at Ala25-Tyr32 (probability; 1%-100%), Gly50-Ser59 (probability; 91%-100%), Ala62-Lys74 (probability; 95%-100%), Val108-Asn119 (3%-98%), Pro138-Glu148 (probability; 84%-100%) and Gln160-Thr167 (probability; 11%-100%). We note that results yield the same residues adopting α-helix conformation with H-REMD simulations (see above). The abundancies deviation is insignificant and the CHARMM36m parameters yield highly abundant α-helix conformations with T-REMD simulations in comparison to NMR experiments.[34] $3_{10}$-helix conformation - obtained from T-REMD simulations using the CHARMM36m parameters - differs than the one obtained from H-REMD simulations. Specifically, we find only two regions that adopt $3_{10}$-helix conformation in MERS-CoV macro domain via T-REMD simulations with probabilities higher than 1% and these are located at Ala102-Ala104 (probability; all at ∼46%) and Val108-Ala120 (probability; 1% - 25%). The probabilities obtained from T-REMD simulations differ from those obtained from H-REMD simulations using the same parameters. We detect seven β-sheet regions in the macro domain of MERS-CoV by T-REMD simulations using the CHARMM36m parameters, which is in accord with H-REMD simulations and experiments. These are located at Glu10-Thr15 (probability; 34%-100%), Val18-Leu22 (probability; all at 100%), Ser35-Ala41 (probability; 6%-100%), Asp81-Gln86 (probability; 27%-100%), Asn93-Val98 (probability; 97%-100%), Leu123-Pro127 (probability; 25%-100%) and Arg152-Val157 (probability; 77%-100%). Moreover, we find 13 turn regions with high abundancies from these T-REMD simulations and these are located at Gly1-Glu10 (probability; 1%-99%), Ile14-Cys17 (probability; 5%-100%), Lys30-Ser35 (probability; 2%-100%), Asn42-Gly48 (probability; 3%-100%), Ser59-Gly61 (probability; 7%-100%), Gln78-Gly80 (probability; all at 100%), Gly87-Asn93 (probability; 3%-100%), Asp101-Lys105 (probability; all at 54%), Asp107-Ser109 (probability; 19%), Lys116-Leu123 (probability; 7%-98%), Leu128-Gly135 (probability; 94%-99%), Arg147-Thr151 (probability; 1%-3%) and Ser126-Thr167 (probability; 6%-7%). The turn structure abundancies per residue obtained from these T-REMD simulations vary from results obtained from our H-REMD simulations (see above).

We find again six α-helix regions in the structures of MERS-CoV macro domain, which agree with NMR experiments, by T-REMD simulations using the AMBER99SB parameters (Figure 2). These are located at Ala25-Tyr32 (probability; 9%-99%), Gly50-Ser59 (probability; 71%-100%), Ala62-Lys74 (probability; 91%-100%), Val108-Ala120 (probability; 1%-94%), Pro138-Glu148 (probability; 37%-100%) and Gln160-Leu166 (67%-100%). The abundancies of α-helix are slightly smaller at some residues (Pro138-Glu148) using

the AMBER99SB parameters rather than the CHARMM36m parameters for the macro domain (see above) and show slightly better agreement with NMR experiments. The $3_{10}$-helix conformation is located at Pro5-Asn8 (probability; 31%), Ala102-Ala104 (probability; 58%), Val108-Ala120 (probability; 1%-50%), Gly132-Phe134 (probability;

7%) and represents four regions in the structures of the macro domain by T-REMD simulations using the AMBER99SB parameters. The $3_{10}$-helix conformation at Pro5-Asn8 and Gly132-phe134 could not be obtained by T-REMD simulations utilizing the CHARMM36m parameters. Residues Glu10-Thr15 (probability; 49%-78%), Val18-Leu22
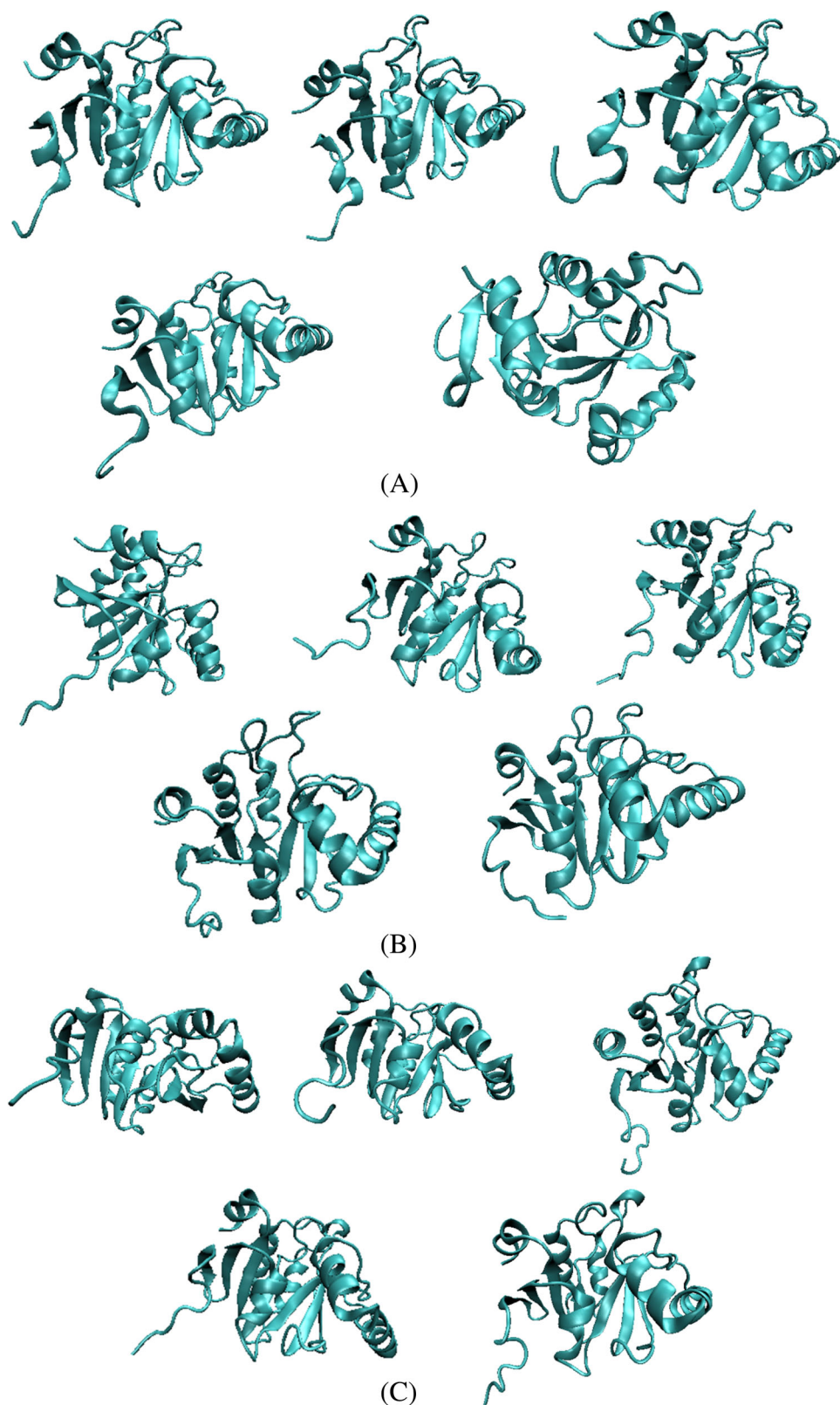


FIGURE 1    Selected structures from our H-REMD simulations using the CHARMM36m parameters, A; T-REMD simulations using the CHARMM36m parameters, B; and T-REMD simulations using the AMBER99SB parameters, C representing conformations of the MERS-CoV macro domain in aqueous media
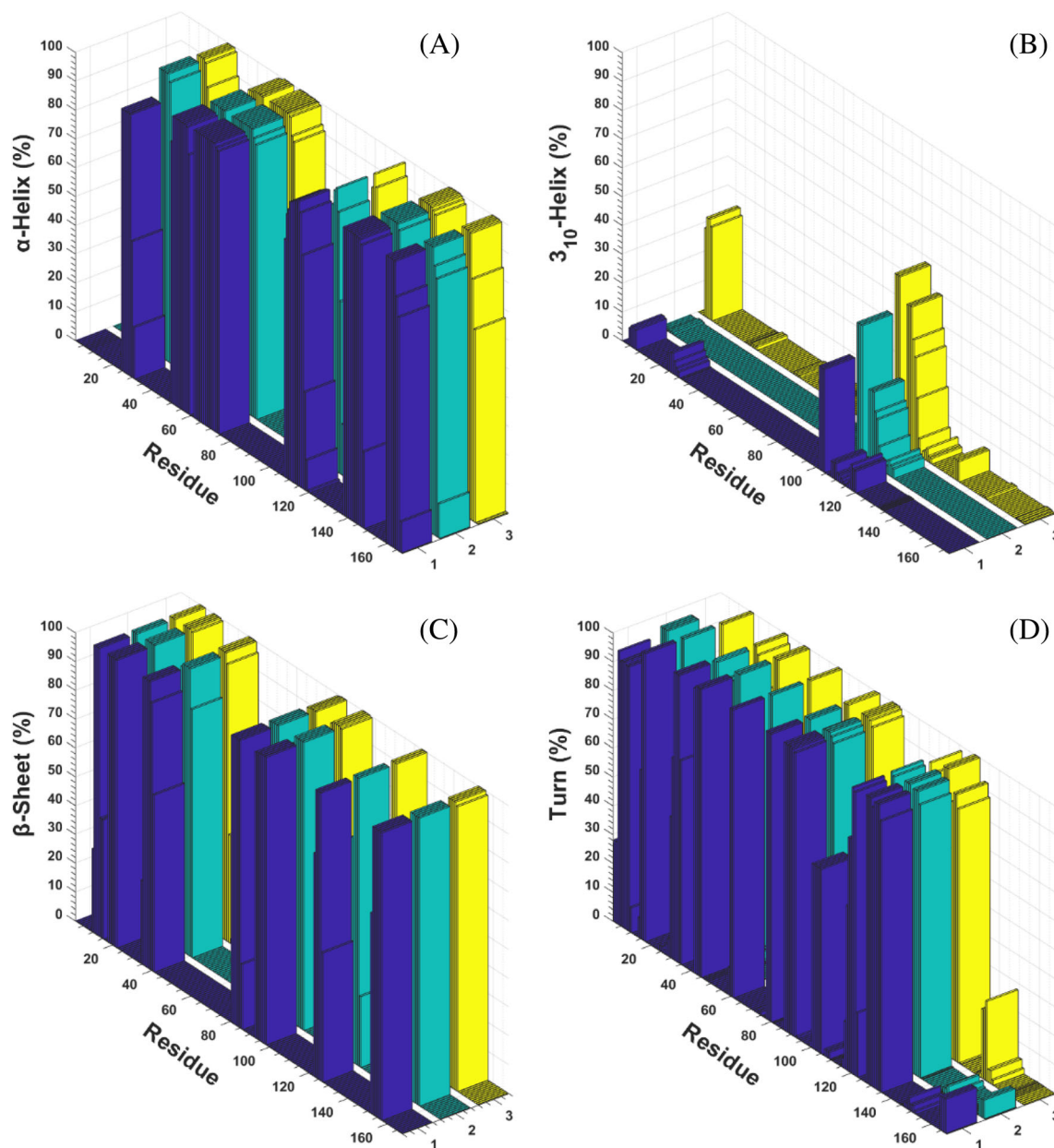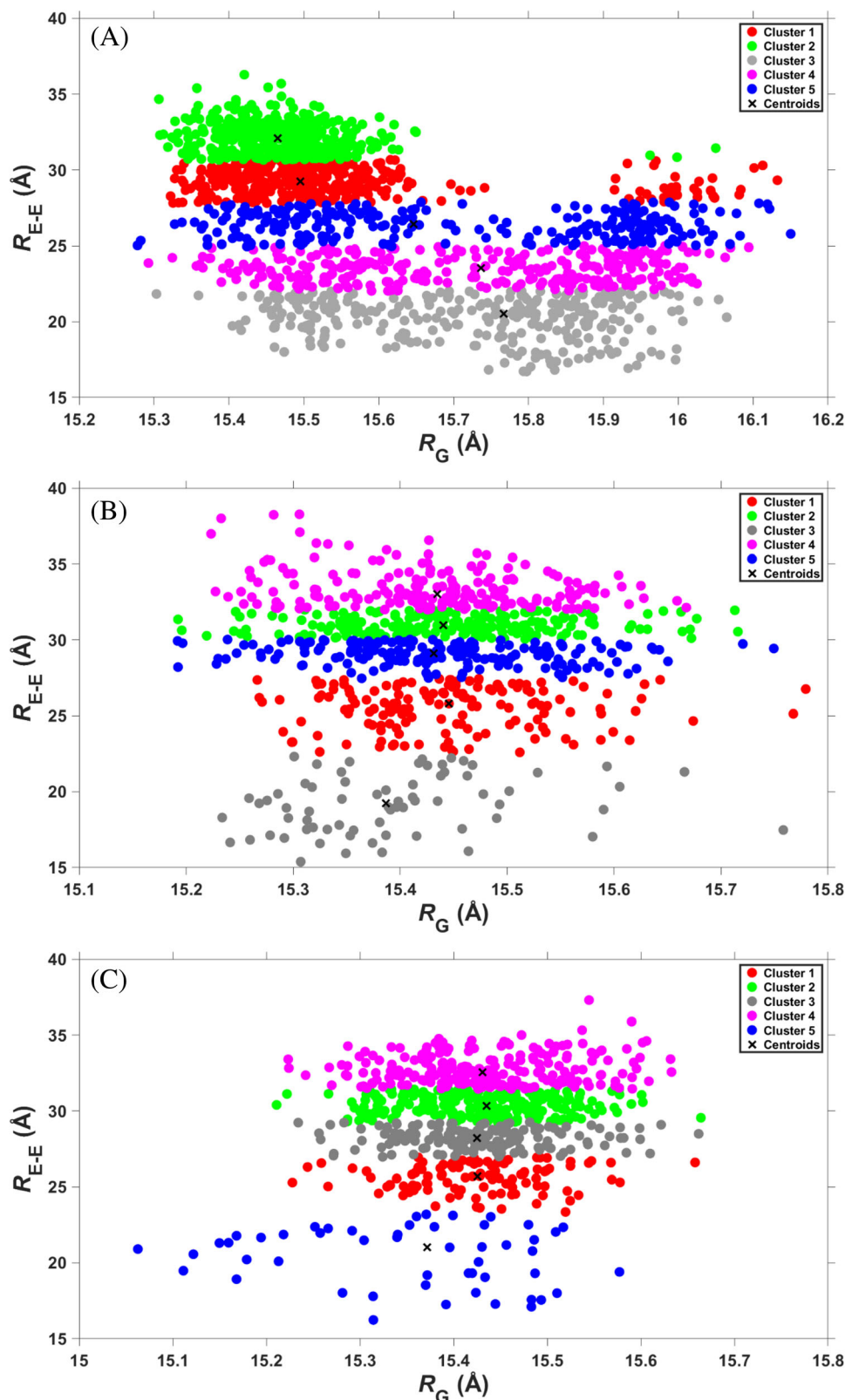
**FIGURE 2** Secondary structure elements and their residue-level probabilities recovered from the MERS-CoV macro domain structures calculated with dynamics in aqueous media: A, α-helix structure formation along with its abundances obtained from H-REMD simulations using the CHARMM36m parameters (blue), T-REMD simulations using the CHARMM36m parameters (green) and T-REMD simulations using the AMBER99SB parameters (yellow); B, 3$_{10}$-helix structure formation along with its abundances obtained from H-REMD simulations using the CHARMM36m parameters (blue), T-REMD simulations using the CHARMM36m parameters (green) and T-REMD simulations using the AMBER99SB parameters (yellow); C, β-sheet structure formation along with its abundances obtained from H-REMD simulations using the CHARMM36m parameters (blue), T-REMD simulations using the CHARMM36m parameters (green) and T-REMD simulations using the AMBER99SB parameters (yellow); D, turn structure formation along with its abundances obtained from H-REMD simulations using the CHARMM36m parameters (blue), T-REMD simulations using the CHARMM36m parameters (green) and T-REMD simulations using the AMBER99SB parameters (yellow)

(probability; 99%-100%), Ser35-Asn42 (probability; 2%-100%), Lys46 and Hsd47 (probability; 2%), Asp81-Gln86 (probability; 16%-100%), Asn93-Val98 (probability; 99%-100%), Leu123-Pro127 (probability; 13%-100%) and Arg152-Val157 (probability; 77%-100%) adopt β-sheet structure in the T-REMD simulations using the AMBER99SB parameters. In total we find eight regions of β-sheet structures. These

regions—except Lys46 and Hsd47—are in accord with H-REMD and T-REMD simulations using the CHARMM36m parameters but probabilities differ from each other. The turn structure conformation is located at Gly1-Glu10 (probability; 4%-67%), Ile14-Cys17 (probability; 2%-100%), Lys30-Ser35 (probability; 7%-100%), Ala40-Gly49 (probability; 2%-100%), Ser59-Gly61 (probability; 15%-100%), Ala73-Gly80

(probability; 2%-100%), Gly87-Lys92 (probability; 97%-100%), Gly99 (probability; 6%), Asp101-Ser109 (probability; 1%-42%), Cys114-Leu123 (probability; 1%-98%), Leu128-Gly135 (probability; 73%-100%) and Leu145-Arg152 (probability; 1%-28%) and represents 12 regions of turn structures using the AMBER99SB parameters in T-

REMD simulations. The largest discrepancies - depending on chosen simulation techniques and force fields—are detected for the $3_{10}$-helix and turn structures of the macro domain in water. A comparison with available experiments states that the simulation results obtained from T-REMD simulations using the AMBER99SB parameters for MERS-



**FIGURE 3** $R_g$ vs $R_{ee}$ values of the MERS-CoV macro domain in solution from REMD simulations that we processed with the *k* means clustering. 5 k values were used and centroids are located at (A) H-REMD simulations using the CHARMM36m parameters; $R_g = 15.50$ Å, $R_{ee} = 29.24$ Å (Centroid1), $R_g = 15.47$ Å, $R_{ee} = 32.10$ Å (Centroid 2), $R_g = 15.77$ Å, $R_{ee} = 20.52$ Å (Centroid 3), $R_g = 15.74$ Å, $R_{ee} = 23.54$ Å (Centroid 4), and $R_g = 15.65$ Å, $R_{ee} = 26.42$ Å (Centroid 5). B) T-REMD simulations using the CHARMM36m parameters; centroids are located at Rg = 15.45 Å, Ree = 25.79 Å (Centroid1), Rg = 15.44 Å, Ree = 30.95 Å (Centroid 2), Rg = 15.39 Å, Ree = 19.22 Å (Centroid 3), Rg = 15.44 Å, Ree = 33.03 Å (Centroid 4), and Rg = 15.43 Å, Ree = 29.12 Å (Centroid 5). C) T-REMD simulations using the AMBER99SB parameters; centroid are located at Rg = 15.43 Å, Ree = 25.70 Å (Centroid1), Rg = 15.44 Å, Ree = 30.33 Å (Centroid 2), Rg = 15.43 Å, Ree = 28.22 Å (Centroid 3), Rg = 15.43 Å, Ree = 33.57 Å (Centroid 4), and Rg = 15.37 Å, Ree = 21.01 Å (Centroid 5)

CoV macro domain and the TIP3P water model agree slightly better with experiments.

Figure 3 presents the results that we generate with the *k*-means clustering of the structures of MERS-CoV macro domain in water with dynamics from H-REMD simulations and from T-REMD simulations using the CHARMM36m and AMBER99SB parameters for the macro domain. We base this calculation on the radius of gyration ($R_g$) and end-to-end distance ($R_{EE}$) values. Based on H-REMD simulations using the CHARMM36m parameters for MERS-CoV macro domain, the $R_g$ values vary between 15.28 Å and 16.15 Å with a mean value of 15.61 ± 0.20 Å (Figure 3A). The $R_{EE}$ values range between 16.71 and 36.27 Å with an average value of 26.81 ± 4.45 Å (Figure 3A). Based on T-REMD simulations using the CHARMM36m parameters, Figure 3B, the $R_g$ values vary between 15.19 and 15.78 Å with a mean value of 15.43 ± 0.10 Å. The $R_{EE}$ values range between 15.38 and 36.27 Å with an average value of 29.06 ± 4.19 Å. Based on T-REMD simulations utilizing the AMBER99SB parameters, the Rg values vary between 15.10 Å and 15.66 Å with a mean value of 15.43 ± 0.10 Å (Figure 3C). Based on these T-REMD simulations using the

AMBER99SB parameters, $R_{EE}$ values range between 16.22 and 37.00 Å with an average value of 29.32 ± 3.43 Å (Figure 3C). Experimental structural studies on MERS CoV macro domain in solution are extremely limited and therefore we could not compare these results to data generated by the experiments. However, we use a set of independently computed residue-level predictions and the secondary structure analysis based on an NMR structure to contextualize and compare with our all-atom results.

The gray line in Figure 4 presents the calculated RMSF values for each residue of the MERS-CoV macro domain in aqueous media at 310 K. Based on these values, we notice more significant fluctuations (higher flexibility) in the C-terminal region of the domain even with such an extensive H-REMD and T-REMD simulation using varying parameters. The average RMSF value for the macro domain (all residues) is 1.19 ± 0.73 Å from H-REMD simulations. However, the most flexible residues are characterized by the RMSF values of up to 6.62 Å in H-REMD simulations using the CHARMM36m parameters. From T-REMD simulations—utilizing the CHARMM36m parameters—the average RMSF values for the macro domain is 1.21 ± 0.84 Å and
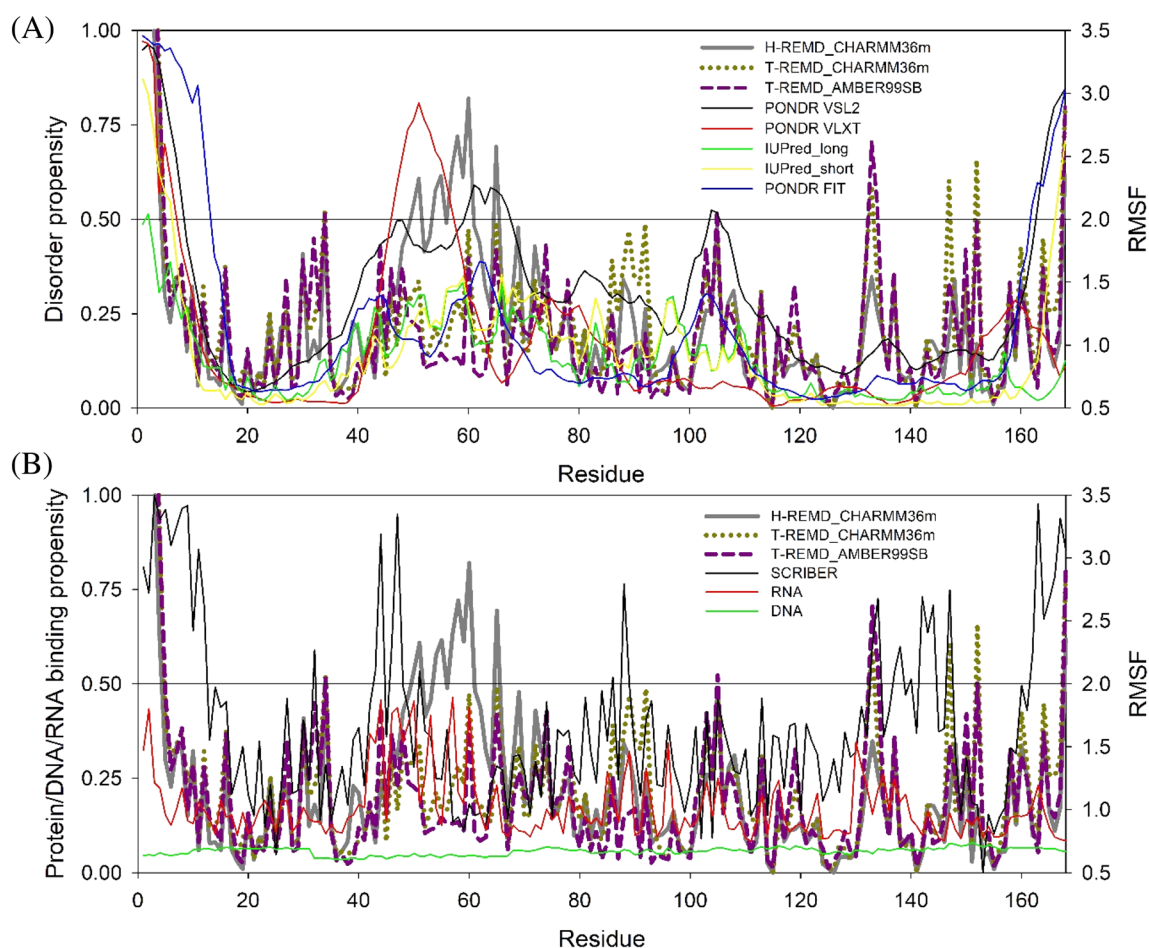


**FIGURE 4** Comparison of the structural flexibility of MERS-CoV macro domain in aqueous media with its intrinsic disorder predisposition, A and propensity for protein and nucleic acid binding, B. Structural flexibility in the aqueous media is reflected in root mean square fluctuations (RMSF) of the protein backbone as a function of the MERS-CoV macro domain residue number. Intrinsic disorder predisposition was evaluated using PONDR VLXT, PONDR VSL2, PONDR FIT, IUPred_short, and IUPres_long, A. Predisposition of this domain to interact with proteins and nucleic acids was evaluated by SCRIBER and DRNApred, respectively, B

the most flexible residues possess RMSF values up to 7.98 Å. Based on the T-REMD simulations using the AMBER99SB parameters, the average RMSF value is 1.13 ± 0.76 Å with the most flexible residues having a RMSF value up to 7.02 Å.

Figure 4A shows that some of the structural dynamic features observed in our parallel tempering simulations are correlated with the residue-level intrinsic disorder predisposition of the MERS-CoV macro domain. This is reflected in the fact that several peaks in disorder profile serve as envelopes that enclose the local RMSF peaks. However, there also some regions (eg, residues 20-38), which are predicted as mostly ordered, but which show noticeable structural fluctuations. This indicates that part of the structural fluctuations of the MERS-CoV macro domain in aqueous medium can be rooted in the intrinsic disorder predisposition of this domain, whereas other structural fluctuations are independent of the intrinsic disorder predisposition of this amino acid sequence.

We also assess propensity of this protein to interact with other proteins and nucleic acids interactions. Similar to the aforementioned disorder analysis, we annotate these interactions at the level of individual amino acids. Figure 4B illustrates that the MERS-CoV macro domain is expected to have several protein binding regions, such as residues 1-12, 32, 43-47, 51, 86-88, 133-134, 137-144, 147, and 162-168. The predicted likelihood of the protein-protein interactions for these residues exceeds the 0.5 threshold. Some of these protein-binding residues are located within the disordered or flexible regions; that is, regions characterized by the predicted disorder score exceeding 0.5 or ranging from 0.2 to 0.5, respectively. Curiously, although all highly flexible residues coincide or are located in the close proximity to the protein-binding regions/residues, not all regions with the highest protein binding potential are characterized by the highest RMSF values. Furthermore, our residue-level analysis does not find any DNA-or RNA-binding regions in the MERS-CoV macro domain. Experimental structural studies on MERS CoV macro domain in solution are extremely limited and therefore we could not compare these results to data generated by the experiments. However, we use a set of independently computed residue-level predictions and the secondary structure analysis based on an NMR structure to contextualize and compare with our results.

## 4 | CONCLUSION

We conduct H-REMD and T-REMD simulations using the CHARMM36m and AMBER99SB parameters for the MERS CoV macro domain in water and present here the results for the 310 K replica. We cover several structural properties including RMSF values with dynamics, secondary structure, and the $k$-means clustering based on radius of gyration ($R_g$) and end-to-end distance ($R_{EE}$) of the structures of MERS CoV macro domain in water with dynamics. Our findings, which rely on the RMSF values, $R_{EE}$ values and deviations, show that some of the residues are flexible. Furthermore, the global structure is compact, not very flexible, and varies only by ∼1.0 Å in water (in terms of the scale of $R_g$ fluctuations). We detected six α-helical

regions and seven β-strand regions, which are in good agreement with the available NMR measurements by H-REMD and T-REMD simulations using the CHARMM36m parameters for the macro domain. T-REMD simulations utilizing the AMBER99SB parameters for the macro domain yield six α-helical and eight β-sheet regions for the macro domain. The largest dependence on simulation techniques and force field parameters is detected for $3_{10}$-helix and turn structure formations of the MERS-CoV macro domain in water. We notice about 10 regions (H-REMD simulation using the CHARMM36m parameters), 13 regions (T-REMD simulations using the CHARMM36m parameters) and 12 regions (T-REMD simulations using the AMBER99SB parameters) with turn structure in the computed conformations of MERS CoV macro domain in water with dynamics. Additionally, we detect four (H-REMD simulations using the CHARMM36m parameters), two (T-REMD simulations using the CHARMM36m parameters) and four (T-REMD simulations using the AMBER99SB parameters) regions of $3_{10}$-helix in the structures of the macro domain in aqueous solution. We should mention here that the results depend on chosen simulation techniques and force field parameters regarding the abundancies and locations of secondary structure elements (especially $3_{10}$-helix and turn structures) and $R_{gg}$ and $R_{EE}$ values. Further experiments are required to assess the quality of simulation techniques and force field parameters because a direct comparison of all these structural properties could not be provided due to the lack of experiments in the current literature.

Based on the results of the comparison of the independently generated intrinsic disorder analysis of the MERS-CoV macro domain with the H-REMD and T-REMD simulations, we also show that only part of the structural fluctuations of this protein in aqueous medium can be attributed the local intrinsic disorder predisposition. The other structural fluctuations are independent of the local propensity of the MERS-CoV macro domain to the intrinsic disorder.

Our residue-level analysis provides some functional clues. Based on putative propensities for protein and nucleic acids interactions, we suggest that while the MERS-CoV macro domain appears not to show DNA- or RNA-binding potential, it contains several protein binding regions. Many of the corresponding PBRs are located within the disordered or flexible regions. Also, some PBRs overlap with the regions with the turn structure. Furthermore, some of the α-helices found in the MERS-CoV macro domain, especially located within the C-terminal half of the protein, were predicted to contain PBRs. Currently, we are studying the structural dynamics of various regions of different proteins from the CoV family ranging from SARS-CoV to SARS-CoV-2 with MERS-CoV in between. All in all, this study demonstrates the structural properties of the MERS-CoV macro domain in aqueous solution using different parallel tempering simulation techniques and force field parameters as well as bioinformatics.

## PEER REVIEW

The peer review history for this article is available at https://publons. com/publon/10.1002/prot.26150.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## ORCID

*Lukasz Kurgan* https://orcid.org/0000-0002-7749-0314
*Orkid Coskuner-Weber* https://orcid.org/0000-0002-0772-9350

## REFERENCES

1. Schwarz K, Groneberg DA. Current overview on MERS-CoV. *Zentralblatt Arbeitsmedizin Arbeitsschutz Ergon*. 2015;65(6):353-354. https://doi.org/10.1007/s40664-015-0062-8.

2. Subbaram K, Kannan H, Khalil Gatasheh M. Emerging developments on pathogenicity, molecular virulence, epidemiology and clinical symptoms of current Middle East respiratory syndrome coronavirus (MERS-CoV). *Hayati J Biosci*. 2017;24(2):53-56. https://doi.org/10.1016/j.hjb.2017.08.001.

3. Ford N, Vitoria M, Rangaraj A, Norris SL, Calmy A, Doherty M. Systematic review of the efficacy and safety of antiretroviral drugs against SARS, MERS or COVID-19: initial assessment. *J. Int. AIDS Soc*. 2020;23(4):e25489. https://doi.org/10.1002/jia2.25489.

4. Cho C-C, Lin M-H, Chuang C-Y, Hsu C-H. Macro domain from Middle East respiratory syndrome coronavirus (MERS-CoV) is an efficient ADP-ribose binding module: CRYSTAL STRUCTURE AND BIOCHEMICAL STUDIES. *J Biol Chem*. 2016;291(10):4894-4902. https://doi.org/10.1074/jbc.M115.700542.

5. Chafekar A, Fielding B. MERS-CoV: understanding the latest human coronavirus threat. *Viruses*. 2018;10(2):93. https://doi.org/10.3390/v10020093.

6. Snijder EJ, Bredenbeek PJ, Dobbe JC, et al. Unique and conserved features of genome and proteome of SARS-coronavirus, an early Split-off from the coronavirus group 2 lineage. *J Mol Biol*. 2003;331(5):991-1004. https://doi.org/10.1016/s0022-2836.

7. Neuman BW, Buchmeier MJ. Supramolecular architecture of the coronavirus particle. *Adv Virus Res*. 2016;96:1-27. https://doi.org/10.1016/bs.aivir.2016.08.005.

8. Forni D, Cagliani R, Mozzi A, et al. Extensive positive selection drives the evolution of nonstructural proteins in lineage C Betacoronaviruses. *J Virol*. 2016;90(7):3627-3639. https://doi.org/10.1128/JVI.02988-15.

9. Geng H, Chen F, Ye J, Jiang F. Applications of molecular dynamics simulation in structure prediction of peptides and proteins. *Comput Struct Biotechnol J*. 2019;17:1162-1170. https://doi.org/10.1016/j.csbj.2019.07.010.

10. Hritz J, Oostenbrink C. Hamiltonian replica exchange molecular dynamics using soft-Core interactions. *J Chem Phys*. 2008;128(14):144121. https://doi.org/10.1063/1.2888998.

11. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: a message-passing parallel molecular dynamics implementation. *Comput Phys Commun*. 1995;91(1–3):43-56. https://doi.org/10.1016/0010-4655.

12. Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G. PLUMED 2: new feathers for an old bird. *Comput Phys Commun*. 2014;185(2):604-613. https://doi.org/10.1016/j.cpc.2013.09.018.

13. Huang J, MacKerell AD. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem*. 2013;34(25):2135-2145. https://doi.org/10.1002/jcc.23354.

14. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 1983;79(2):926-935. https://doi.org/10.1063/1.445869.

15. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *J Chem Phys*. 2007;126(1):014101. https://doi.org/10.1063/1.2408420.

16. Parrinello M, Rahman A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J Appl Phys*. 1981;52(12):7182-7190. https://doi.org/10.1063/1.328693.

17. Darden T, York D, Pedersen L. Particle mesh Ewald: an $N \cdot \log(N)$ method for Ewald sums in large systems. *J Chem Phys*. 1993;98(12):10089-10092. https://doi.org/10.1063/1.464397.

18. Allison TC, Coskuner O, Gonzalez CA, Eds. *Metallic Systems: A Quantum Chemist's Perspective*, 1st ed. Florida: CRC Press; 2011. https://doi.org/10.1201/b10835.

19. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett*. 1999;314(1–2):141-151. https://doi.org/10.1016/S0009-2614.

20. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*. 2006;65(3):712-725. https://doi.org/10.1002/prot.21123.

21. Weber OC, Uversky VN. How accurate are your simulations? Effects of confined aqueous volume and AMBER FF99SB and CHARMM22/CMAP force field parameters on structural ensembles of intrinsically disordered proteins: amyloid-$\beta_{42}$ in water. *Intrinsically Disord Proteins*. 2017;5(1):e1377813. https://doi.org/10.1080/21690707.2017.1377813.

22. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. 1983;22(12):2577-2637. https://doi.org/10.1002/bip.360221211.

23. Likas A, Vlassis N. J. Verbeek, J. the global K-means clustering algorithm. *Pattern Recognit*. 2003;36(2):451-461. https://doi.org/10.1016/S0031-3203.

24. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. *Proteins*. 2001;42(1):38-48. https://doi.org/10.1002/1097-0134.

25. Peng K, Radivojac P, Vucetic S, Dunker AK, Obradovic Z. Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics*. 2006;7:208. https://doi.org/10.1186/1471-2105-7-208.

26. Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta*. 2010;1804(4):996-1010. https://doi.org/10.1016/j.bbapap.2010.01.011.

27. Dosztányi Z, Csizmok V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*. 2005;21(16):3433-3434. https://doi.org/10.1093/bioinformatics/bti541.

28. Dosztányi Z, Csizmók V, Tompa P, Simon I. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J Mol Biol*. 2005;347(4):827-839. https://doi.org/10.1016/j.jmb.2005.01.071.

29. Mészáros B, Erdos G, Dosztányi Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res*. 2018;46(W1):W329-W337. https://doi.org/10.1093/nar/gky384.

30. Zhang J, Kurgan L. SCRIBER: accurate and partner type-specific prediction of protein-binding residues from proteins sequences. *Bioinformatics*. 2019;35(14):i343-i353. https://doi.org/10.1093/bioinformatics/btz324.

31. Zhang J, Kurgan L. Review and comparative assessment of sequence-based predictors of protein-binding residues. *Brief Bioinform*. 2018;19(5):821-837. https://doi.org/10.1093/bib/bbx022.

32. Yan J, Kurgan L. DRNApred, fast sequence-based method that accurately predicts and discriminates DNA- and RNA-binding residues. *Nucleic Acids Res*. 2017;45(10):e84. https://doi.org/10.1093/nar/gkx059.

33. Su H, Liu M, Sun S, Peng Z, Yang J. Improving the prediction of protein-nucleic acids binding residues via multiple sequence profiles and the consensus of complementary methods. *Bioinformatics*. 2019;35(6):930-936. https://doi.org/10.1093/bioinformatics/bty756.

34. Huang Y-P, Cho C-C, Chang C-F, Hsu C-H. NMR assignments of the macro domain from Middle East respiratory syndrome coronavirus (MERS-CoV). *Biomol NMR Assign*. 2016;10(2):245-248. https://doi.org/10.1007/s12104-016-9676-9.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.