Data-informed reparameterization of modified RNA and the effect of explicit water models: Application to pseudouridine and derivatives

Nivedita Dutta,[†] Indrajit Deb,[†] Joanna Sarzynska,[‡] and Ansuman Lahiri[*], [†]

[†]Department of Biophysics, Molecular Biology and Bioinformatics, University of Calcutta, 92, Acharya Prafulla Chandra Road, Kolkata 700009, West Bengal, India

[‡]Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland

Correspondence to: Ansuman Lahiri ([*]E-mail: albmbg@caluniv.ac.in)

ORCID of the authors: Nivedita Dutta (0000-0002-8371-9007), Indrajit Deb (0000-0002-2722-582X) Joanna Sarzynska (0000-0002-0500-2238), Ansuman Lahiri (0000-0002-7398-4114)

ABSTRACT

Pseudouridine is the most abundant post-transcriptional modification in RNA. We have previously shown that the FF99-derived parameters for pseudouridine and some of its naturally occurring derivatives in the AMBER distribution either alone or in combination with the revised **γ** torsion parameters (parmbsc0) failed to reproduce their conformational characteristics observed experimentally (Deb I, et al. J. Chem. Inf. Model. 2014, 54 (4):1129–1142; Deb I, et al. J. Comput. Chem., 2016, 37:1576−1588; Dutta N, et al. J. Chem. Inf. Model. 2020, 60 (10):4995–5002). However, the application of the recommended bsc0 correction did lead to an improvement in the description not only of the distribution in the $\boldsymbol{\gamma}$ torsional space but also of the sugar pucker distributions. In an earlier study, we examined the transferability of the revised glycosidic torsion parameters ($\boldsymbol{\chi}_{\mathrm{IDRP}}$) for Ψ to its derivatives. We noticed that although these parameters in combination with the AMBER FF99-derived parameters and the revised $\boldsymbol{\gamma}$ torsional parameters resulted in conformational properties of these residues that were in better agreement with experimental observations, the sugar pucker distributions were still not reproduced accurately. Here we report a new set of partial atomic charges for pseudouridine, 1-methylpseudouridine, 3-methylpseudouridine and 2′-O-methylpseudouridine and a new set of glycosidic torsional parameters ($\boldsymbol{\chi}_{\mathrm{ND}}$) based on chosen glycosidic torsional profiles that most closely corresponded to the NMR data for conformational propensities and studied their effect on the conformational distributions using REMD simulations at the individual nucleoside level. We have also studied the effect of the choice of water model on the conformational characteristics of these modified nucleosides. Our observations suggest that the current revised set of parameters and partial atomic charges describe the sugar pucker distributions for these residues more accurately and that the choice of a suitable water model is important for the accurate description of their conformational properties.

KEYWORDS

RNA modification, pseudouridine, glycosidic torsion, conformational characteristics, hydration pattern

INTRODUCTION

Post-transcriptional modifications have been known to be crucial in the regulation of the structure, stability and function of RNA molecules. The MODOMICS database currently lists 172 such modifications [1]. Pseudouridine (Ψ) was the first post-transcriptional modification discovered [2-4] and is one of the most abundant modifications. Pseudouridine, an isomer of uridine (U), was identified as 5-ribosyluracil and was called the fifth nucleoside [5-8]. This modified residue contains a C-C base-sugar bond, i.e., in the case of pseudouridine, the uracil base is attached to the sugar by a C1′-C5 bond unlike the C1′-N1 glycosidic linkage found in uridine (Figure 1 (a)). Hence, in contrast to uridine, pseudouridine contains an additional ring nitrogen atom (N1 imino atom) which acts as an additional hydrogen bond donor and is found to be protonated at physiological pH [3,9].

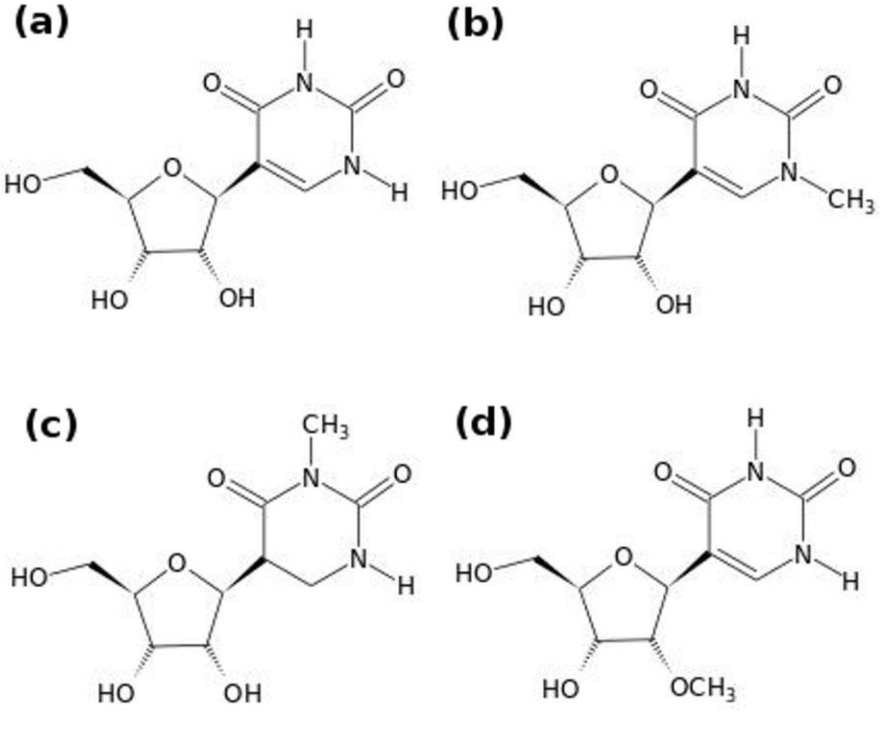

Figure 1. Structures of (a) pseudouridine, Ψ (PSU); (b) 1-methylpseudouridine, m^{1}Ψ (1MP); (c) 3-methylpseudouridine, m^{3}Ψ (3MP); and (d) 2′-O-methylpseudouridine, Ψm (MRP).

Pseudouridine was reported to be the most commonly observed modification in the stable RNAs, i.e., tRNA, rRNA and snRNA[3]. Further studies involving high-throughput sequencing methods and transcriptome mapping revealed the abundance of pseudouridine as an epigenetic modification, i.e. in mRNA as well as in long noncoding RNA (lncRNA) [10-14]. Several experimental and theoretical studies suggest the important contribution of pseudouridine to the structure, dynamics and thermal stability of RNA [15-21]. This modification has been found to reduce the motion of the neighbouring bases, stabilize the C3′-endo conformation and enhance the stability and the stacking propensity in a context-dependent manner [15,20-23]. Newby and Greenbaum studied the interaction between Ψ and water in the pre-mRNA branch-site helix and reported that a water-ΨHN1 hydrogen bond contributes to the stabilization of the unique observed architectural features of this helix [18].

In 2016, we reported that the reoptimized set of glycosidic torsion parameters ($\boldsymbol{\chi}_{\mathrm{IDRP}}$) for pseudouridine developed by us, were sufficient to improve the description of the conformational distribution of the glycosidic torsion space but the description of the sugar pucker distribution for Ψ was still not accurate [24]. In another study in 2020, we checked the transferability of these parameters ($\boldsymbol{\chi}_{\mathrm{IDRP}}$) to the derivatives of Ψ and observed that the $\boldsymbol{\chi}_{\mathrm{IDRP}}$ parameters combined with the AMBER FF99-derived parameters [25] and the revised set of $\boldsymbol{\gamma}$ torsional parameters predicted the conformational properties of these residues which were in general agreement with the experimental (NMR) data but failed to describe the sugar pucker distributions accurately [26].

In the present study we report a new set of glycosidic torsional parameters ($\boldsymbol{\chi}_{\mathrm{ND}}$) and a new set of partial atomic charges for pseudouridine (Ψ), 1-methylpseudouridine ($m^1\Psi$), 3-methylpseudouridine ($m^3\Psi$) and 2′-O-methylpseudouridine (Ψm) (Figure 1). We have compared the results obtained with these parameters with those previously obtained with the FF99 parameters and the FF99 parameters in combination with the $\boldsymbol{\chi}_{\mathrm{IDRP}}$ parameters and bsc0 $\boldsymbol{\gamma}$ torsional parameters.

In the earlier studies, multiple schemes [27] and/or general schemes [28] were chosen for the quantum mechanical scan and the molecular mechanical energy profiles were fitted with those with the objective that the re-optimized parameters will be able to explore, preferentially, any of the four quadrants (NORTH/*syn*, NORTH/*anti*, SOUTH/*syn*, SOUTH/*anti*) of the conformational preferences. In the present work, we calculated the quantum mechanical glycosidic torsional energy profiles for five different initial conformations. Then a particular scheme was chosen which outperformed other schemes in reproducing QM profile that was in agreement with the experimentally observed conformational preference. Next, the MM profile was fitted to the chosen QM profile. Additionally, the partial charges were newly generated at the individual modification level before generating the MM profile to incorporate the effect of electrostatic interactions. As a proof of concept, we have chosen pseudouridine and three of its derivatives as a (small) closely related test set that includes molecules with different chemical moieties. For additional validation of our parameter sets, we examined their performance in predicting the conformational and hydration characteristics of the ssRNA trimers and tetramers containing pseudouridine.

It has been reported in recent studies that the choice of water model has a significant impact on the predicted RNA structure and dynamics [29,30]. Kührova et al.; based on their study involving the simulation of canonical A-RNA duplexes using explicit water models; i.e. TIP3P[31], TIP4P/2005[32], TIP5P[33] and SPC/E[34], reported that the TIP5P water model was not found to be optimal for simulating RNA systems [29]. Here, we have investigated the impact of the choice of explicit water models on the conformational characteristics and hydration pattern of Ψ, $m^1Ψ$, $m^3Ψ$, and Ψm.

METHODS

Ab initio potential energy surface (PES) scan

Preparation of the initial geometries

For the initial geometries of the modified nucleosides Ψ (PSU), m^{1}Ψ (1MP), m^{3}Ψ (3MP), and Ψm (MRP), we have used the mean values for bonds, angles and dihedral angles corresponding to the ribose sugar following Gelbin et al. (1996) [35] and considered planar geometries for the bases. The three-letter codes of the modified residues are according to Aduri et al. (2007) [25]. These structures were prepared using the molecular structure editor MOLDEN [36]. The geometries of the modified nucleosides were kept either in the C3'-endo/g$^+$ conformation or in the C2′-endo/g$^+$ conformation and for that the corresponding torsional angles were fixed at definite values. The value of the **γ** dihedral angle (O5′-C5′- C4′-C3′) was fixed at 54° (which corresponds to the g$^+$ conformation) as observed in the A-form RNA [37]. To compel the nucleoside geometries to stay in the C3'-endo conformation, the values of the ***δ*** (C5′-C4′-C3′-O3′) and O4′-C1′- C2′-C3′ dihedral angles were fixed at 81° and -24°, respectively. To constrain the geometries to the C2'-endo sugar pucker conformation the value of the ***δ*** (C5′-C4′-C3′-O3′) and O4′-C1′-C2′-C3′ dihedral angles were set to 140° and 32° respectively. Five initial geometries, i.e., SC1, SC2, SC3, SC4 and SC5 (Table S1) with constrained values of the H5T-O5′-C5′-C4′ and C1′-C2′-O2′-HO2′ torsional angles were prepared for each of the modified nucleosides, to either promote or restrict the base-sugar hydrogen bonding interactions by maintaining the nucleosides either in C3′-endo or in C2′-endo sugar pucker conformation. The schemes SC1-SC4 were chosen following the values of the torsional angles corresponding to the four schemes chosen in Yildirim et al. [27] and SC5 was chosen based on the *syn* scheme as mentioned in Deb et al. [24]. SC4 also corresponds to the *anti* scheme as mentioned in Deb et al. [24]. For the SC4 conformational scheme, the H5T-O5′-C5′-C4′ and C1′-C2′-O2′-HO′2 dihedrals were respectively constrained to 174° and 93° and due to that the O5′-H···O4 base-sugar hydrogen bonding interaction is restricted and O2′-H···O4 base-sugar hydrogen bonding interaction is facilitated and hence the geometries corresponding to PSU and its derivatives are compelled towards *anti* conformation which is not the predominant conformation for these nucleosides. For the SC5 scheme, the values of the H5T-O5′-C5′-C4′ and C1′-C2′-O2′-HO′2 dihedrals were respectively constrained to 60° and -153° to promote the O5′-H···O4 and restrict the O2′-H···O4 base-sugar hydrogen bonding interactions and hence to force a *syn* conformation which is predominant for PSU and its derivatives [38]. The SC1 and SC2 conformational schemes were kept in the C2′-endo conformation while SC3-SC5 were kept in the C3′-endo conformation. To prevent any hydrogen bonding interaction between H3T or O2′ and base, so that these interactions cannot affect the glycosidic torsion energy profile, the C4′-C3′-O3′-H3T torsion was fixed at -148° for all the initial geometries. The initial structures corresponding to each of the five conformational schemes are shown in Figure S1. The geometry which corresponds to the SC5 conformational scheme for each of the modified nucleosides (along with the atom names) is shown in Figure S2.

Quantum mechanical scan

All the quantum mechanical calculations were performed using the GAUSSIAN09 software suite [39]. For all the five initial geometries for each of the modified nucleosides, a gas phase PES scan was executed around the glycosidic torsion angle (O4′-C1′-C5-C6) with an increase in its value by 5° resulting in 72 conformations for each nucleoside geometry. Optimization of the structures, during the PES scan, was carried out using the HF/6-31G* level of theory. During the geometry optimization step, the dihedral angles mentioned in Table S1, were kept frozen with the objective of obtaining a smooth QM energy profile. The QM energies (E_{QM}) corresponding to each of the 72 geometry optimized conformations (for each scheme) were calculated using the MP2/6-31G* level of theory. Out of the five quantum mechanical energy (E_{QM}) profiles around χ, we have chosen one particular conformational scheme, i.e. SC5, because the lowest energy minimum for this scheme corresponded to the *syn* region of the glycosidic torsional space (Figure S3) and experimental (NMR) studies for pseudouridine and its derivatives, under study, reported a preference for the *syn* conformation [54-56]. Additionally, the value of energy corresponding to the global minimum of that profile was found to be the least compared to other schemes (Figure S3).

RESP fitting

The new set of partial atomic charges for each of the modified nucleosides was developed corresponding to the lowest energy conformation of the quantum mechanical energy profile of the chosen scheme, i.e. SC5, by RESP [40,41] fitting (Restrained Electrostatic Potential fitting) method using the R.E.D. version III.52

perl program [42]. The partial atomic charges for the atoms of each of the nucleosides are listed in the supporting information (Table S2).

Molecular mechanical (MM) energy minimization

For the calculation of the molecular mechanical (MM) energies (E_{MM}) corresponding to the 72 quantum mechanically (QM) optimized geometries, we have used the AMBER16 software package [43] (Figure 2). During the MM energy minimizations, the dihedral angles (as mentioned in Table S1) were restrained to the values corresponding to the QM optimized geometries by applying a force constant of 1500 Kcal/mol $Å^2$. The starting structures for the MM energy minimization step were the structures equivalent to the QM optimized geometries obtained from the PES scan. The 5′-phosphate group was replaced with a hydrogen (5′-OH) and a hydrogen atom (3′-OH) was added to the 3′ end of the original topology provided by Aduri et al. [25] to create the topologies for all the modified nucleosides used in this study with the parameters corresponding to the 5′-OH and 3′-OH groups taken from the FF99 force field parameter set [44]. During the MM energy minimization, all the glycosidic torsion parameters corresponding to the Aduri et al. [25] parameter set were set to zero for all the modified nucleosides. Minimizations were carried out using the steepest descent method followed by the conjugate gradient method in order to obtain a smooth glycosidic torsional energy profile for each residue. To incorporate the non-bonded interactions during the energy minimization in vacuum, a long range cut-off of 12 Å was used.

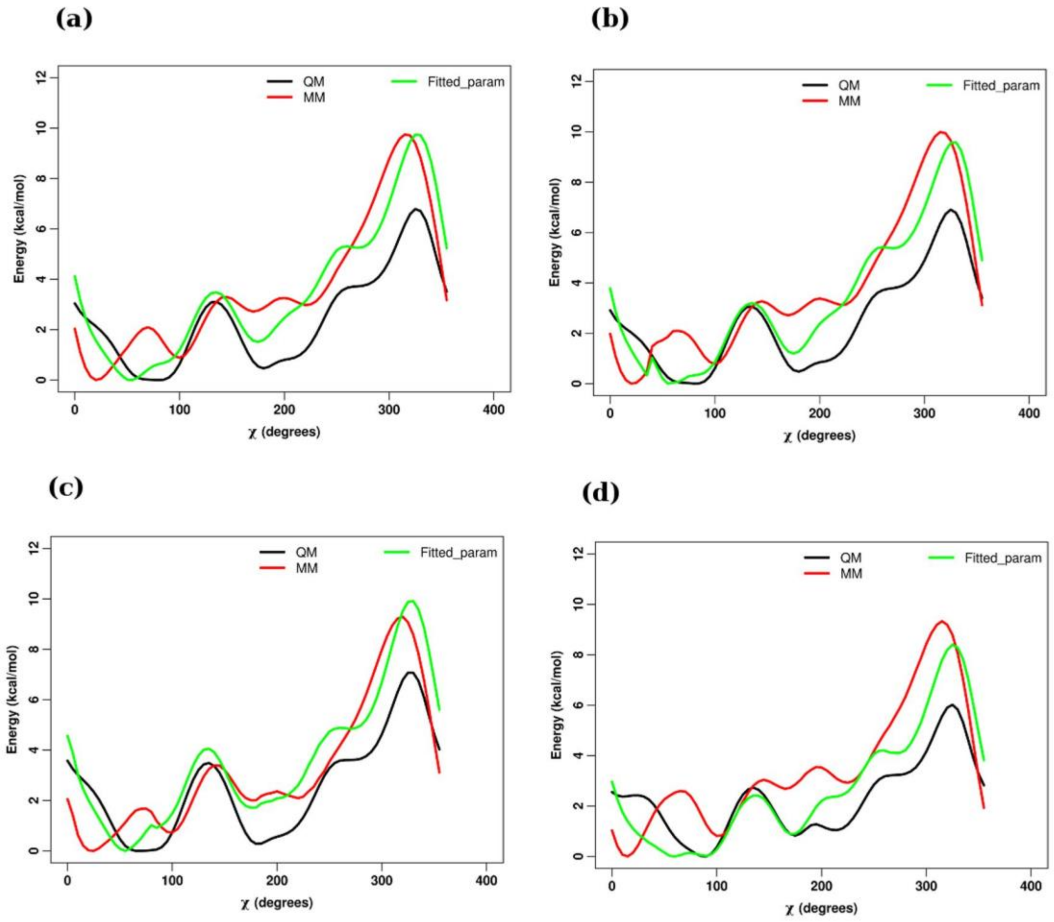

Figure 2 Energy profiles of the χ torsional angles (O4′-C1′-C5-C4) for (a) PSU, (b) 1MP, (c) 3MP and (d) MRP residues corresponding to QM calculations (black), MM calculations with the FF99 parameter sets keeping the glycosidic torsion parameters zero (red) and MM calculations with the FF99 parameter sets combined with the newly derived χ torsional parameters and the newly developed partial atomic charges (FF99_χ_{ND}) (green) by fitting the difference between the QM and MM energies. The minimum energies were set to zero for convenience. The ranges 30°-90° and 170°-300° for the χ torsional angles along the X-axis, correspond to the *syn* and *anti* base orientations respectively.

Fitting χ torsion potentials

The potential energy due to the glycosidic torsion angle is represented by the difference (E_{CHI}) between the QM energy (E_{QM}) and MM energy (E_{MM}) and is given by the following equation:

$$E_{CHI} = E_{QM} - E_{MM} \quad (1)$$

The 72 values for E_{CHI} obtained from eq. (1) were fitted to the Fourier series as shown in eq. (2): $E_{CHI} = \sum_{n=1}^{4} \quad [Vn\,\{1 + cos(n\chi - \boldsymbol{\phi} n)\}] \qquad (2)$

Where $\boldsymbol{\chi}$ represents the glycosidic torsion angle; i.e. the dihedral around (O4′-C1′-C5-C6) and V_n represents the potential energy barrier around the glycosidic torsion angles ($\boldsymbol{\chi}$) and $\boldsymbol{\phi}_n$ is the phase angle.

System preparation

The starting structures in this study were taken from the original PDB format files for each of the four modified ribonucleoside residues corresponding to their quantum mechanically optimized geometries provided by Aduri et al. [25], and available in the AMBER 2018 package. These initial structures of these modified ribonucleosides were in a NORTH/anti/g+ conformation. The FF99_$\boldsymbol{\chi}_{IDRP}$_bsc0 [24] parameter set for Ψ was obtained from Deb et al. [24], and FF99_$\boldsymbol{\chi}_{IDRP}$_bsc0 [24] parameter sets for $m^1\Psi$, $m^3\Psi$, and Ψm residues were obtained from Dutta et al. [26]. The FF99_$\boldsymbol{\chi}_{ND}$_bsc0 parameter sets for Ψ, $m^1\Psi$, $m^3\Psi$, and Ψm residues were prepared by combining our newly derived $\boldsymbol{\chi}$ torsional parameters ($\boldsymbol{\chi}_{ND}$) and the revised $\boldsymbol{\gamma}$ parameters developed by Pérez et al. [45] (parmbsc0) with the required bond, angle and torsional parameters for each modification from the AMBER provided parameters derived from Aduri et al. parameters [25]. The revised $\boldsymbol{\gamma}$ torsional parameters were incorporated by replacing the atom type that described the terms corresponding to the $\boldsymbol{\gamma}$ torsion in the default topology files with the torsional terms provided in the revised parmbsc0 force field. The newly developed partial atomic charges for the atoms (except for some atoms as mentioned in the supporting information) of each of the four modified ribonucleosides were introduced replacing the partial atomic charges of these atoms in the preparatory file (prepin) provided by Aduri et al. [25]. We used these revised parameter sets for energy minimization and MD simulation steps. The revised force field parameter sets for Ψ, $m^1\Psi$, $m^3\Psi$, Ψm (FF99_$\boldsymbol{\chi}_{ND}$_bsc0) are given in the supporting information.

The modified ribonucleosides Ψ, m^{1}Ψ, m^{3}Ψ, and Ψm were separately simulated using the FF99_χ_{IDRP}_bsc0 and FF99_χ_{ND}_bsc0 parameters respectively. Detailed description of the force field parameters used in this study are provided in Table 1. The newly derived glycosidic (χ) torsion parameters are listed in Table 2.

Table 1. Relevant details of the force fields used in this study

Force fields	Applied revised parameters for torsions	Definition of the applied revised torsional terms
FF99	None	AMBER provided parameters for Ψ, m^{1}Ψ, m^{3}Ψ, Ψm nucleosides developed by Aduri et al. [25].
FF99_bsc0	γ	AMBER provided parameters for Ψ, m^{1}Ψ, m^{3}Ψ, Ψm nucleosides developed by Aduri et al. [25] in combination with revised γ torsion parameters developed by Pérez et al. [45] (parmbsc0).
FF99_χ_{IDRP}_bsc0	χ and γ	For Ψ, FF99_χ_{IDRP}_bsc0 parameters obtained from by Deb et al. [24] and for its three derivatives (m^{1}Ψ, m^{3}Ψ, and Ψm), FF99_χ_{IDRP}_bsc0 parameters [24,25,45] modified by the introduction of required bond, angle and torsional parameters for each modification from the AMBER provided parameters derived from Aduri et al. parameters [25] (obtained from Dutta et al. [26]).
FF99_χ_{ND}_bsc0	χ and γ	Revised glycosidic torsion parameters (χ_{ND}) for Ψ, m^{1}Ψ, m^{3}Ψ, and Ψm nucleosides and revised γ torsion parameters developed by Pérez et al. [45] (parmbsc0) in combination with the required bond, angle and torsional parameters for each modification from the AMBER provided parameters derived from Aduri et al. parameters [25] along with the newly developed set of partial atomic charges for each of these modified nucleosides.

Table 2. Revised **χ** torsion parameters for Ψ, $m^1\Psi$, $m^3\Psi$, and Ψm.

Modified nucleosides	**Torsional angle**	**n**	**V_n**	**ϕ**
Ψ (PSU)	O4′-C1′-C5-C6	1 2 3 4	1.00081 0.757522 1.41029 0.340171	180.738622 0.758978 180.459801 1.19649
$m^1\Psi$ (1MP)		1 2 3 4	0.761663 0.660492 1.4897 0.347886	180.683487 0.713806 180.419626 1.10914
$m^3\Psi$ (3MP)		1 2 3 4	0.751305 0.871664 1.33006 0.402064	180.505381 0.801897 180.466825 1.14129
Ψm (MRP)		1 2 3 4	1.06577 0.872603 1.52327 0.322335	180.37802 0.711823 180.474343 1.07613

Replica exchange molecular dynamics simulations

All replica exchange molecular dynamics (REMD) simulations [46] were performed using the multi-sander approach in AMBER 16 [43] in explicit water. To study the effect of the water model on the conformations of these nucleosides, REMD simulations were carried out using the combination of the FF99_χ_{IDRP}_bsc0 and FF99_χ_{ND}_bsc0 force fields with each of the TIP3P[31], TIP4P-Ew[47] and SPC/E[34] water models and the hydration patterns for pseudouridine and its three derivatives corresponding to the different force field-water model combinations were analyzed. The modified nucleoside residues Ψ, $m^1\Psi$, $m^3\Psi$, Ψm were solvated with TIP3P or TIP4P-Ew or SPC/E water molecules in truncated octahedral boxes with a closest distance of 9 Å between any solute atom and the edge of the box.

Energy minimization of the solvated system was carried out in two steps. For the first set of energy minimization which consisted of 500 steps of steepest descent followed by 500 steps of conjugate gradient optimization, the nucleosides were held fixed with the help of a positional restraining force of 500 kcal/mol $Å^2$. The next set of energy minimization was performed without any positional restraining force and consisted of 1000 steps of steepest descent followed by 1500 steps of conjugate gradient optimization. Equilibration of the energy minimized systems was carried out in two steps. In the first step, the systems were heated from 0K to 300K temperature in 20 ps with a 2 fs time step using a constant volume dynamics by the application of a 10 kcal/mol $Å^2$ positional restraining force. In the second step of equilibration, whole systems were equilibrated in the absence of any restrain, at 300K temperature for 200 ps with a 2 fs time step using constant pressure dynamics (reference pressure of 1 atm and pressure relaxation time of 2 ps). After the completion of the equilibration steps, the final coordinates obtained were used as the starting coordinates for the REMD simulations. In the REMD equilibration step before the REMD production run, each of the systems was equilibrated at 16 target temperatures that spanned over a range from 300K to 400K (i.e. at T = 300.0 K, 305.8 K, 311.7 K, 317.8 K, 323.9 K, 330.2 K, 336.6 K, 343.1 K, 349.7 K, 356.5 K, 363.4 K, 370.5 K, 377.6 K, 384.9 K, 392.4 K and 400.0 K) and this step was carried out for 1 ns with a 2 fs time step with constant volume dynamics. These equilibrated systems were used for the REMD production runs consisting of 2000 cycles in constant volume. 4000 steps of MDs were performed with a 2 fs time step before the attempted exchange between the neighbouring replicas at the temperatures mentioned above. The REMD production runs generated simulation of 16 ns for each of the replicas, yielding a total simulation of 256 ns in aggregate. For each system-force field and water model combinations, three independent sets of REMD simulations were performed.

For propagation of the trajectories, Langevin dynamics (with random velocity scaling with 1 ps^{-1} collision frequency) was used. The SHAKE algorithm [48] was used to constrain the bonds which involved hydrogen atoms. Particle mesh Ewald (PME) was used for handling the electrostatic interactions. To include nonbonded interactions, a long range cutoff of 8 Å was used.

Analysis of conformational ensembles

For the analysis of the simulated ensembles we calculated the distribution of sugar pucker conformations, distribution of the *syn* or *anti* conformations of the glycosidic torsion angle (**χ**) and the distribution of the **γ** torsional angle over different conformational states.

The convention followed for the atom names and the dihedral angle nomenclatures was as given in Saenger [38]. The magnitude of the pseudorotation angle was calculated following Altona and Sundaralingam [49]. The pseudorotation angular space was divided into C3′-endo/NORTH (270°≤ P< 90°) and C2′-endo/SOUTH (90° ≤ P< 270°) regions of sugar puckering[50], which allowed us to directly compare simulated conformational distributions and the equilibrium distributions of the pseudorotation angle (P) as reported in the NMR data.

In our analysis, the **χ** torsional angle is defined by the atoms O4′-C1′-C5-C4 (for all the modified nucleosides) and was considered to be in the *anti* conformation if its magnitude was within the angular range of 170°-300° and in the *syn* conformation if it was within the angular range of 30°-90° [35,51,52]. The values that were beyond these ranges were referred to as others [35,51,52].

For the calculation of the γ torsional angle, the conformational space with respect to the torsional angle consisting of the atoms O5′-C5′-C4′-C3′ was divided into the conformations referred as g+ (for 60°±30°), g- (for 300°±30°), trans (180°±30°) and others (outside the ranges mentioned for the other conformations).

We analysed the hydrogen bonding characteristics, radial distribution function (RDF) for each of the four residues and the distribution of the θ torsion angle (H2′-C2′-O2′-HO2′) for the Ψ, $m^1\Psi$, $m^3\Psi$ residues. For the calculation of the pseudorotation angle P, the χ, γ, and θ torsion angles, hydrogen bonds and RDFs, cpptraj tool from Ambertools18 [53] was used. RDFs of water oxygen atoms around the HN1 atom was calculated for each of the Ψ, $m^3\Psi$ and Ψm residues and RDFs of water oxygen atoms around the HN3 atom was calculated for each of the Ψ, $m^1\Psi$ and Ψm residues. Hydrogen bond formations were taken into account if the distance between the donor and the acceptor atoms was ≤ 3 Å and the donor-hydrogen-acceptor angle was ≥ 135°. The water occupancy maps around the average MD structure (the average MD structures were obtained from 600 frames corresponding to each of the four conformations i.e NORTH/*syn*, SOUTH/*syn*, NORTH/*anti* and SOUTH/*anti* conformations from a set of 16 ns REMD simulations) of Ψ corresponding to the FF99_χ_{ND}_bsc0 and TIP3P force field and water model combination were calculated using the grid routine in cpptraj tool and visualization was done using UCSF-Chimera [54].

RESULTS AND DISCUSSION

In an earlier study [26] we validated the revised parameter sets for pseudouridine (Ψ) (FF99_χ_{IDRP}_bsc0) [24] and checked the transferability of these parameters to the four pseudouridine derivatives i.e. $m^1\Psi$, $m^3\Psi$, Ψm and $m^1acp^3\Psi$ and our observations indicated that the revised parameters for Ψ were transferable to the

Ψ derivatives. In the present study we reoptimized the parameters for the glycosidic torsion angle individually for Ψ and its three derivatives $m^1\Psi$, $m^3\Psi$ and Ψm and developed new sets of partial atomic charges for each of these residues and compared the conformational ensembles. The REMD simulations were carried out using the combination of the force fields i.e. FF99_χ_{IDRP}_bsc0 and FF99_χ_{ND}_bsc0 with the TIP3P, TIP4P-Ew and SPC/E water models. The results are written and discussed below.

Pseudorotation angle (P)

With the AMBER FF99 parameter sets, the distribution of the pseudorotation angle was observed to have a smaller population of the NORTH sugar pucker conformation compared to the experimentally observed population for each of the modified residues except for Ψ [26] (Table S3, Figure 3). Inclusion of the revised **γ** torsion parameters (parmbsc0) with the AMBER FF99 parameter sets resulted in an improvement in the propensity of the NORTH sugar pucker conformation for all the Ψ-derivatives. But with the FF99_bsc0 parameters, the propensity of the NORTH sugar pucker conformation for Ψ was significantly lower than the experimentally observed value [26].

For Ψ, the FF99_χ_{ND}_bsc0 force field in combination with the TIP3P and the SPC/E models generated a population of the NORTH sugar pucker conformation which were in general much closer to the experimentally observed population than those generated by the FF99_χ_{IDRP}_bsc0 force field in combination with each of the water models in this study. But FF99_χ_{ND}_bsc0 in combination with the TIP4P-Ew water model generated a much greater population of the NORTH conformers of Ψ than the other force field and water model combinations and also the experimentally observed population. However, FF99_χ_{IDRP}_bsc0 + TIP4P-Ew reproduced the experimental value of the NORTH population for $m^1\Psi$ better than all the other force field-water model combinations. In the case of $m^3\Psi$, it was observed that, the FF99_χ_{ND}_bsc0 force field in combination with each of the three water models generated a population of the NORTH sugar pucker conformation which agreed better with the NMR results than what was observed with the FF99_χ_{IDRP}_bsc0 force field in combination with each of the water models in this study. For Ψm, the FF99_χ_{ND}_bsc0 + TIP3P and FF99_χ_{ND}_bsc0 + SPC/E combinations generated population of the NORTH conformers in better agreement with the NMR results than those generated by the other force field-water model combinations.

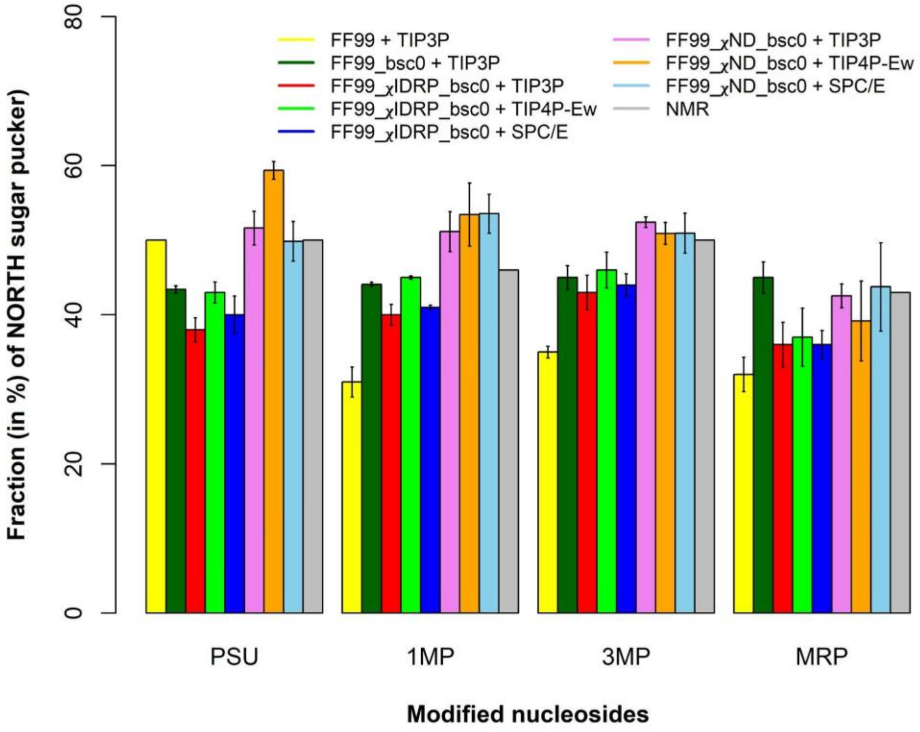

Figure 3. Fraction (in %) of NORTH sugar pucker in equilibrium ensembles of Ψ (PSU), $m^1\Psi$ (1MP), $m^3\Psi$ (3MP), and Ψm (MRP) at 300 K. The values reported here are the averages (along with the standard deviations) calculated from three independent sets of 16 ns REMD simulations.

Glycosidic torsion angle (χ)

For each of the modified nucleosides under this study, experimental (NMR) studies reported preference for the *syn* conformation [55-57]. The FF99 and FF99_bsc0 parameters, for each of the modified residues predicted an excess population of *anti* conformers (>90%) [26,58]. Earlier, we reported that FF99_χ_{IDRP}_bsc0 + TIP3P shifted the equilibrium towards the *syn* conformation. The FF99_χ_{IDRP}_bsc0 parameter sets in combination with each of the TIP4P-Ew and SPC/E water models also generated a much greater population of *syn* conformation in good agreement with the NMR data than that obtained with the FF99 parameter sets (Table S4, Figure 4). With the revised parameter sets FF99_χ_{ND}_bsc0 in combination with each of the three water models, the modified residues adopted a much greater population of the *syn* conformation than what was predicted by the default AMBER parameters. But for each modified nucleoside, the population of syn conformers predicted by the FF99_χ_{ND}_bsc0 parameters were lower than what was predicted by the FF99_χ_{IDRP}_bsc0 parameters for each of the water models under this study.

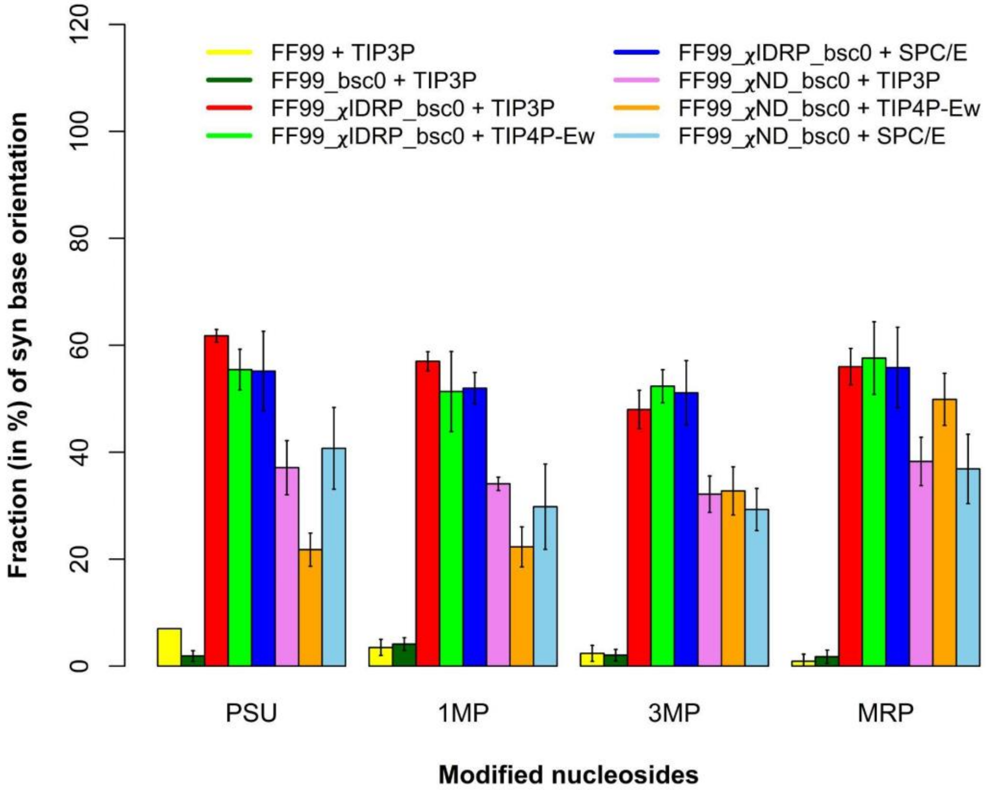

Figure 4. The fraction (in %) of *syn* base orientation in the equilibrium ensembles of Ψ (PSU), m^{1}Ψ (1MP), m^{3}Ψ (3MP), and Ψm (MRP) at 300 K. The values reported here are the averages (along with the standard deviations) calculated from three independent sets of 16 ns REMD simulations. The modified nucleosides Ψ, m^{1}Ψ (1MP), m^{3}Ψ (3MP), and Ψm (MRP) have been reported to prefer the *syn* conformation [55-57].

Gamma torsion angle (γ)

In our earlier studies, we reported that, with the FF99 parameter sets, the g+ population was much lower than the experimentally observed population for pseudouridine and its derivatives [26,58]. In the present study, it was observed that all the force field and water model combinations predicted the g+ population greater than what was predicted with the FF99 parameter sets, but also than the experimentally observed population (Table S5, Figure 5). As was reported earlier [26], in the present study also we observed that the inclusion of the revised **γ** torsion parameters developed by Pérez et al.[45] (parmbsc0) shifted the equilibrium almost exclusively towards the g+ conformation (~90%).

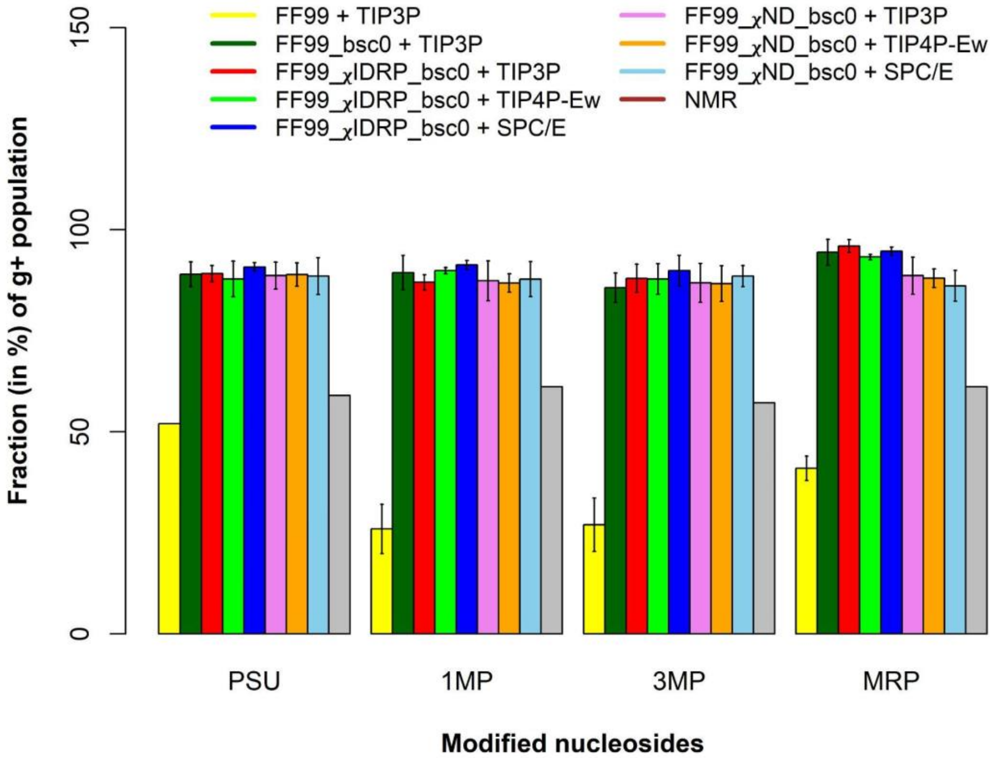

Figure 5. The fraction (in %) of g^+ population in the equilibrium ensembles of Ψ (PSU), $m^1\Psi$ (1MP), $m^3\Psi$ (3MP), and Ψm (MRP) at 300 K. The values reported here are the averages (along with the standard deviations) calculated from three independent sets of 16 ns REMD simulations.

Correlation of the pseudorotation equilibrium with the glycosidic torsion angle (χ)

The two-dimensional scatter correlation plots of pseudorotation angle (P) vs glycosidic torsion angle (χ) revealed that for all the ribonucleosides in this study, with FF99_χ_{IDRP}_bsc0 + TIP3P there was a significantly large population of the SOUTH/*syn* conformations (Figures S4-6). With FF99_χ_{IDRP}_bsc0 + TIP4P-Ew,, there were almost equal populations of SOUTH/*syn* and NORTH/*syn* conformations for all the four modified nucleosides, but the population of the SOUTH/*syn* conformers was a little higher in each case. The FF99_χ_{IDRP}_bsc0 + SPC/E force field-water model combination also predicted a higher population of SOUTH/*syn* conformers than the others. In general, with the FF99_χ_{ND}_bsc0 force field in combination with the TIP3P and the SPC/E water models, almost equal populations of the SOUTH/*syn* and NORTH/*anti* conformers were observed for each of the modified residues. The combination FF99_χ_{ND}_bsc0 + TIP4P-Ew predicted a large population of NORTH/*anti* conformers for Ψ and $m^1\Psi$ nucleosides. But for $m^3\Psi$, this force field-water model combination predicted almost equal populations of the SOUTH/*syn* and NORTH/*anti* conformers. With FF99_χ_{ND}_bsc0 + TIP4P-Ew, Ψm preferentially adopted the SOUTH/*syn* conformation.

Correlation of the pseudorotation equilibrium with the gamma torsion angle (γ)

From the two-dimensional correlation maps (two-dimensional scatter plots), it was observed that the FF99_χ_{IDRP}_bsc0 force field in combination with each of the three water models in the present study, predicted a greater population of the SOUTH/g+ conformers followed by that of the NORTH/g+ conformers for each of the modified nucleosides (Figures S7-9). In general, with the FF99_χ_{ND}_bsc0 parameter sets in combination with each of the three water models, we observed that there were almost equal populations of the NORTH/g+ and SOUTH/g+ conformers for all the residues. With FF99_χ_{ND}_bsc0+TIP4P-Ew, Ψ preferentially adopted the NORTH/g+ conformation while Ψm preferentially adopted the SOUTH/g+ conformation. The populations of the g- and trans conformers were extremely low due to the inclusion of the $\boldsymbol{\gamma}$ torsion parameters developed by Pérez et al. [45] (parmbsc0) as was observed in our earlier study [26].

Hydrogen bonding

The hydrogen bonds except O5′-H5T---O4 (Figure 6) and O2′-HO2′---O4 hydrogen bonds were observed to be negligible (Tables 3-4). With each of the force field-water model combinations, for all the modified residues (not applicable to Ψm), it was observed that the number of conformers with O2′-HO2′---O4 hydrogen bonding interaction were very small and much lesser than that of the O5′-H5T---O4 hydrogen bonding interaction.

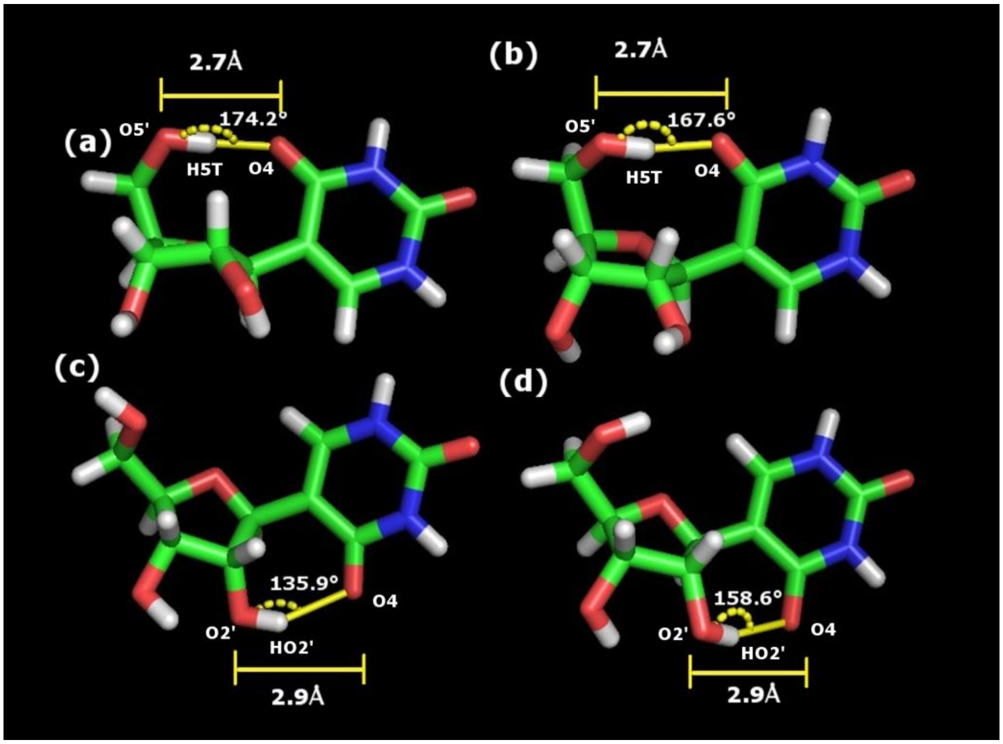

Figure 6. (a) Snapshots of (a) the O5′-H5T---O4 hydrogen bond in Ψ (PSU) residue in the *syn* conformation of the glycosidic torsion and NORTH conformation of sugar pucker; (b) the O5′-H5T---O4 hydrogen bond in Ψ (PSU) residue in the *syn* conformation of the glycosidic torsion and SOUTH conformation of sugar pucker; (c) the O5′-H5T---O4 hydrogen bond in Ψ (PSU) residue in the *anti* conformation of the glycosidic torsion and NORTH conformation of sugar pucker; (d) the O5′-H5T---O4 hydrogen bond in Ψ (PSU) residue in the *anti* conformation of the glycosidic torsion and SOUTH conformation of sugar pucker All

the snapshots are taken from a set of 16 ns REMD simulations corresponding to the FF99_χ_{ND}_bsc0 force field and TIP3P water model combination.

For Ψ, FF99_χ_{IDRP}_bsc0+TIP3P, FF99_χ_{ND}_bsc0+TIP3P and FF99_χ_{ND}_bsc0+SPC/E force field-water model combinations predicted greater populations of conformers with O5′-H5T---O4 hydrogen bonding interaction than what were predicted by the other force field-water model combinations. With FF99_χ_{ND}_bsc0+TIP4P-Ew, the number of O5′-H5T---O4 hydrogen bonding interactions for Ψ were significantly lower than what were predicted by the other force field-water model combinations. For $\mathrm{m}^1\Psi$, FF99_χ_{ND}_bsc0+TIP3P predicted a greater number of O5′-H5T---O4 hydrogen bonding interactions than those predicted by the other force field-water model combinations. All the force field-water model combinations generated almost equal populations of conformers with O5′-H5T---O4 hydrogen bonding interaction for the $\mathrm{m}^3\Psi$ residue. For Ψm, FF99_χ_{ND}_bsc0 force field parameters predicted greater number of O5′-H5T---O4 hydrogen bonding interactions than what was predicted by FF99_χ_{IDRP}_bsc0 in combination with each of the water models in this study. For Ψ and $\mathrm{m}^1\Psi$, the FF99_χ_{ND}_bsc0 force field predicted slightly greater population of conformers with O2′-HO2′---O4 hydrogen bonding interaction than the FF99_χ_{IDRP}_bsc0 parameter sets with each of the water models. But for $\mathrm{m}^3\Psi$, all the force field-water model combinations generated similar populations of conformers with O2′-HO2′---O4 hydrogen bonding interaction.

Table 3. Percent (%) occurrence of O5′-H5T---O4 hydrogen bond (at 300K).

Force fields	**Water models**	**Hydrogen bonding atoms**	**Ψ** **(PSU)**	**($\mathrm{m}^1\Psi$)** **1MP**	**($\mathrm{m}^3\Psi$)** **3MP**	**(Ψm)** **MRP**
FF99_χ_{IDRP}_bsc0	TIP3P	**O5′-H5T---O4**	20±0.82	18±1.3	16±3.0	19±2.2

FF99_χ_{IDRP}_bsc0	TIP4P-Ew		15±2.3	14±2.2	15±2.6	16±1.4
FF99_χ_{IDRP}_bsc0	SPC/E		17±2.3	15±0.41	15±2.0	17±2.3
FF99_χ_{ND}_bsc0	TIP3P		20±2.1	22±3.0	17±2.3	23.4±3.1
FF99_χ_{ND}_bsc0	TIP4P-Ew		10±0.9	12±2.2	18±3.0	28±2.8
FF99_χ_{ND}_bsc0	SPC/E		21±4.3	15±5.3	17±1.4	22±4.0

Table. 4. Percent (%) occurrence of O2′-HO2′---O4 Hydrogen bond (at 300K).

Force fields	**Water models**	**Hydrogen bonding atoms**	**Ψ (PSU)**	**(m^{1}Ψ) 1MP**	**(m^{3}Ψ) 3MP**
FF99_χ_{IDRP}_bsc0	TIP3P	**O2′-HO2′---O4**	2.4±0.34	2.7±0.58	3.1±0.09
FF99_χ_{IDRP}_bsc0	TIP4P-Ew		2.5±0.21	2.4±0.25	3.0±0.18
FF99_χ_{IDRP}_bsc0	SPC/E		2.2±1.1	2.6±0.27	2.9±0.72
FF99_χ_{ND}_bsc0	TIP3P		3.4±0.27	4.0±0.12	3.5±0.25
FF99_χ_{ND}_bsc0	TIP4P-Ew		4.8±0.53	6.0±0.83	3.2±0.64
FF99_χ_{ND}_bsc0	SPC/E		3.1±0.26	4.2±0.58	3.1±0.15

Radial distribution function

From the RDF plot of water oxygen atoms with respect to the HN1 atom of Ψ, it was observed that the FF99_χ_{IDRP}_bsc0 force field in combination with each of the water models predicted the formation of a well-defined first hydration shell between 1.5 Å to 2.5 Å having a maximum at ~2 Å (Figures S10-12). This observation was consistent with that of the recent report by Deb et al. [21]. The FF99_χ_{ND}_bsc0 force field in combination with each of the water models also predicted the formation of a well-defined first hydration shell between 1.5 Å to 2.5 Å having a maximum at ~2 Å around the Ψ-HN1 atom. For the HN1 atoms of m^{3}Ψ and Ψm also, all the force field and water model combinations predicted the formation of a well-defined first hydration shell between 1.5 Å to 2.5 Å having a maximum at ~2 Å. For the Ψ and m^{3}Ψ residues, the concentration of the water molecules around the HN1 atom was observed to be slightly higher with the FF99_χ_{ND}_bsc0 force field than what was observed with the FF99_χ_{ND}_bsc0 force field for each of the water models while for Ψm the concentration was observed to be slightly lower with the FF99_χ_{ND}_bsc0 force field than what was observed with the FF99_χ_{ND}_bsc0 force field for each of the water models. From the RDF plots of water oxygen atoms with respect to the HN3 atoms of Ψ, m^{1}Ψ and Ψm nucleosides, with each of the force field and water model combinations, the formation of a well-defined first hydration shell was observed between 1.5 Å to 2.5 Å having a maximum at ~2 Å. For Ψ and m^{3}Ψ, the concentration of the water molecules around the HN3 atom was observed to be similar with each of the force field and water model combinations. Interestingly, for Ψm, the FF99_χ_{ND}_bsc0 force field parameters predicted a higher concentration of water molecules around the HN3 atom than what was predicted by FF99_χ_{IDRP}_bsc0 in combination with each of the water models. The hydration pattern around pseudouridine (Ψ) corresponding to the FF99_χ_{ND}_bsc0 and TIP3P force field and water model combination is shown in Figure 7.

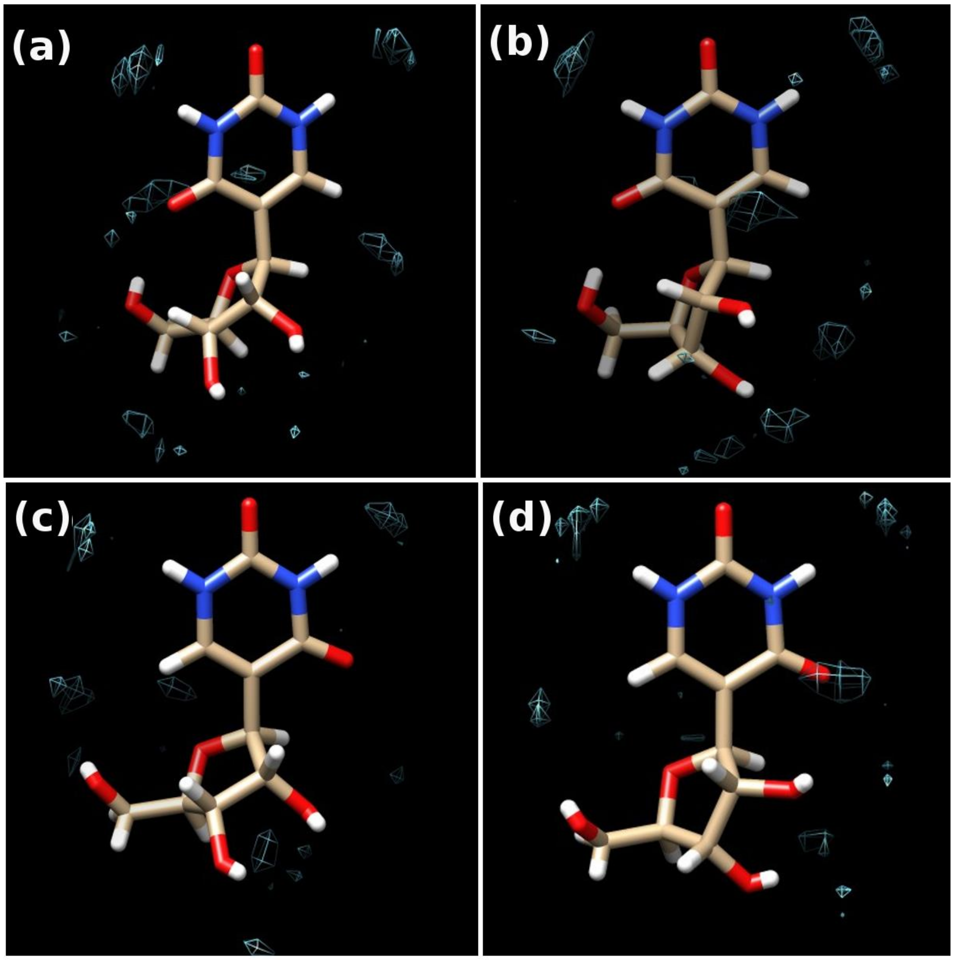

Figure 7. Hydration pattern of Ψ (PSU) in (a) NORTH/*syn* (b) SOUTH/*syn*, (c) NORTH/*anti* and (d) SOUTH/*anti* conformations corresponding to the FF99_χ_{ND}_bsc0 force field and TIP3P water model combination. Water occupancy contoured at equivalent levels visualized using UCSF-chimera [54].

Orientation of the 2'-hydroxyl group of Ψ, m^{1}Ψ and m^{3}Ψ nucleosides

The orientation of the 2′-hydroxyl groups of RNA has been reported to have a significant contribution to the stability of the A-form RNA helices [59] and also in RNA-protein interactions [60]. The A-RNA duplex has been suggested to be stabilized by a network consisting of water-mediated hydrogen bonds mediated by the 2′ hydroxyl groups and also the extensive individual hydration of the 2′ hydroxyl groups [61,62]. Kührova et al. (2014) reported that the choice of water model has significant effect on the orientation of the 2′-OH atom of nucleotides and hence also on the entire RNA structure [29]. The **θ** torsion angle (H2′-C2′-O2′-HO2′) populates three regions, the O3′ domain (value of **θ** between 50-140°), the O4′ domain (value of **θ** between 175-230°) and the base domain (value of **θ** between 270-345°), for C3′-endo sugar pucker conformation [29,63]. It has been reported that the 2’-OH group when oriented towards the base domain can act as a hydrogen bond donor to a water molecule and when it is oriented towards the O3′ domain it can accept a hydrogen bond from the same water molecule [59,61]. NMR studies at low temperatures suggested that the 2′-OH group can be oriented either towards the O3′ domain or towards the base domain and the predominant orientation of the 2′-OH group is reported to be towards the O3′ domain [64]. In the present study, we checked the effect of the combinations of the FF99_$\boldsymbol{\chi}_{\mathrm{IDRP}}$_bsc0 and FF99_$\boldsymbol{\chi}_{\mathrm{ND}}$_bsc0 force fields with the three different water models TIP3P, TIP4P-Ew and SPC/E, on the orientation of the 2′-OH atom corresponding to the Ψ, $m^1\Psi$, and $m^3\Psi$ residues (Figure 8). The distribution of the **θ** torsion angle (H2′-C2′-O2′-HO2′) angle was similar for each of the three water models for each modified residue. But the distribution differed between the two force fields. The O3′-domain was predominantly sampled (followed by the base-domain) by all the force field-water model combinations in agreement with the experimental and theoretical studies [29,64]. The population of the conformers with the 2′-OH atom oriented towards the O4′-domain were significantly lower than the population of the conformers with the 2′-OH atom oriented towards the other two domains. While a prominent peak was observed at the O4′-domain with the FF99_$\boldsymbol{\chi}_{\mathrm{IDRP}}$_bsc0 force field in combination with each of the three water models, the FF99_$\boldsymbol{\chi}_{\mathrm{ND}}$_bsc0 force field did not predict the same.

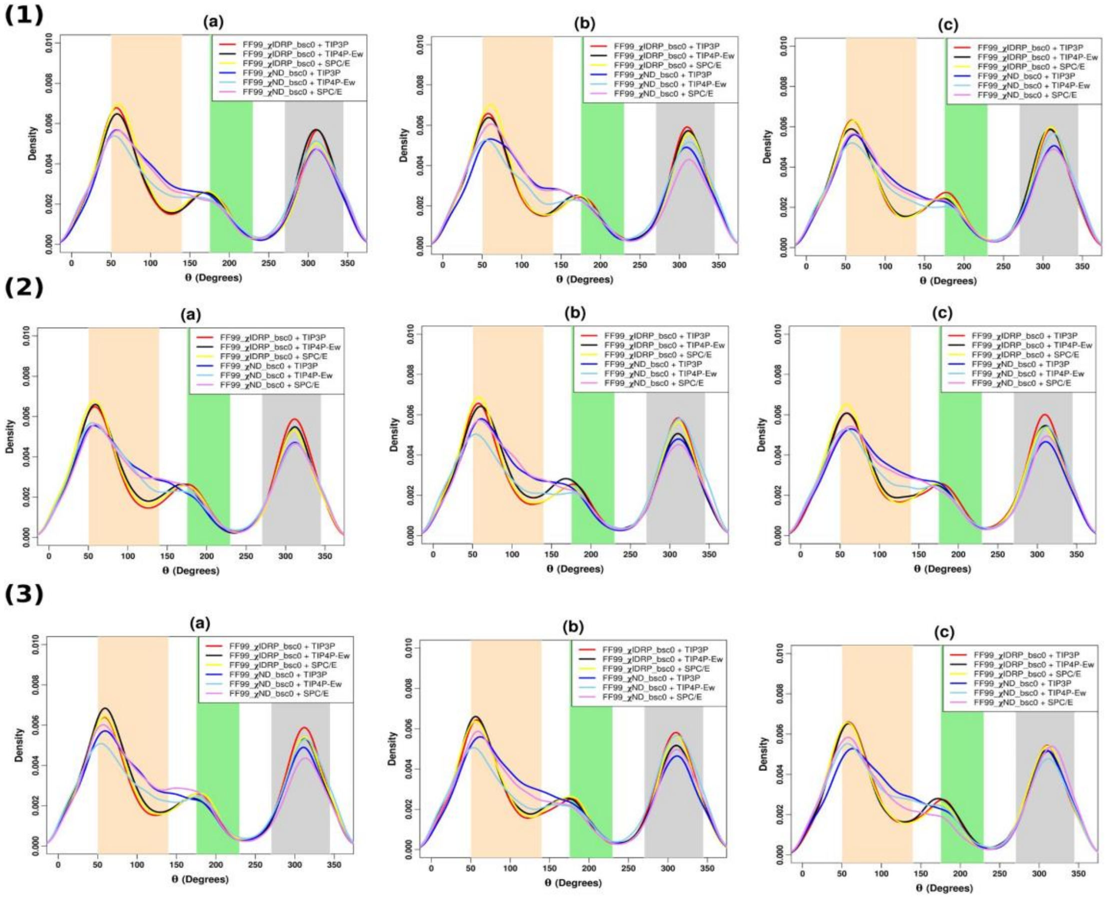

Fig. 8. Distribution of the **θ** torsion angle (H2′-C2′-O2′-HO2′) for (a) PSU, (b)1MP, and (c) 3MP, corresponding to FF99_χ_{IDRP}_bsc0 and TIP3P, FF99_χ_{IDRP}_bsc0 and TIP4P-Ew, FF99_χ_{IDRP}_bsc0 and SPC/E, FF99_χ_{ND}_bsc0 and TIP3P, FF99_χ_{ND}_bsc0 and TIP4P-Ew, and FF99_χ_{ND}_bsc0 and SPC/E force field and water model combinations for the (1) first set, (2) second set, (3) third set of 16 ns REMD simulations (at 300 K). The **θ** torsion angle (H2′-C2′-O2′-HO2′) populates the following three regions: O3′ domain (value of **θ** between 50-140°) (indicated in bisque rectangle); O4′ domain (value of **θ** between 175-230°) (indicated in light green rectangle) and the base domain (value of **θ** between 270-345°) (indicated in grey rectangle), for C3′-endo sugar pucker conformation [29,63].

CONCLUSIONS

In the present study we derived a revised set of of partial atomic charges and glycosidic torsional parameters (χ_{ND}) for the nucleosides Ψ, m^{1}Ψ, m^{3}Ψ, and Ψm following a data-informed approach. At the individual nucleoside level, the partial atomic charges and glycosidic torsional parameters (χ_{ND}) were calculated by applying RESP fitting method to the lowest energy conformation of the quantum mechanical energy profile of a chosen conformational scheme and fitting the molecular mechanics energy profile to that scheme-specific quantum mechanical energy profile, respectively. The choice of a particular conformational scheme was dictated by the NMR results that reported a preference for the syn conformation for pseudouridine and its derivatives under study[55-57] and thereafter looking for the scheme that had the lowest energy value for the syn conformation. The consequences of the application of the revised set of glycosidic torsional parameters (χ_{ND}) in combination with the revised γ torsion parameters (parmbsc0) developed by Pérez et al. [45] and the AMBER FF99-derived parameters[25] for these modified nucleosides were analysed using replica exchange molecular dynamics simulations. The newly derived parameters were validated by comparing the simulated conformational preferences with the available experimental (NMR) data as well as with the observations in Dutta et al. [26]. REMD simulations were carried out using the FF99_χ_{IDRP}_bsc0 [24] and FF99_χ_{ND}_bsc0 force fields in combination with each of the TIP3P, TIP4P-Ew and SPC/E water models. Three independent REMD simulations (each of 16 ns) were carried out in 16 temperature windows ranging from 300 to 400 K, resulting in 768 ns of simulation time in total.

It was observed that there were significant differences in the description of the conformational properties of each of the modified nucleosides by different combinations of force fields and water models. The revised force field parameter sets (FF99_χ_{ND}_bsc0) with the TIP3P water model was able to closely reproduce the experimentally observed sugar pucker preferences for each of the modified nucleosides in this study. The accuracy of the prediction of the population of the C3'-endo/NORTH conformers might be important for accurate reproduction of the C3'-endo/NORTH pucker conformation associated with the A-form RNA structures.

In general, the newly developed force field parameters (FF99_χ_{ND}_bsc0) in combination with each of the water models under this study shifted the distribution of the base orientation for each of the modified nucleosides towards the *syn* conformation in contrast to the excess of *anti* conformations predicted by the AMBER FF99 and AMBER FF99_bsc0 parameters [25,45]. But the population of the *syn* conformers predicted by the FF99_χ_{ND}_bsc0 force field was observed to be less than that predicted by the FF99_χ_{IDRP}_bsc0 force field parameters. The choice of water model was not found to influence the description of the base orientation to a significant extent for the FF99_χ_{IDRP}_bsc0 force field parameters. However, the FF99_χ_{ND}_bsc0 force field in combination with the TIP4P-Ew water model resulted in a somewhat smaller population of the *syn* conformers in the case of Ψ and m^{1}Ψ nucleosides and a significantly greater population of the *syn* conformers for Ψm than what were observed with the other two water models.

In earlier studies from our group [24,26], we reported that, at the single nucleoside level, the inclusion of the revised **γ** torsion parameters (parmbsc0) developed by Pérez et al. [45] along with the FF99_χ_{IDRP} parameter sets did not reproduce the experimentally observed population of the g+ conformers, but predicted a much larger g+ population for pseudouridine and its derivatives. We also noted that the large population of g+ conformers observed with the FF99_χ_{IDRP}_bsc0 parameters might be necessary to maintain the g+ conformation of a nucleotide as is observed in the standard A-form of RNA [37]. In the present study, we observed that the newly derived FF99_χ_{ND}_bsc0 parameter sets also predicted a large population of the g+ conformation for each of the modified residues. The populations of g+ conformers for all the nucleosides under this study, predicted by each of the force field and water model combinations were similar and were much larger than that predicted with the FF99 parameters.

The observations from the calculations of the number of O5′-H5T---O4 and O2′-HO2′---O4 hydrogen bonding interactions for each of the modified nucleosides in this study, suggested that O5′-H5T---O4 hydrogen bonding interaction contributes to the stabilization of the syn base orientation[38] while the O2′-HO2′---O4 hydrogen bonding interaction may facilitate the *anti* base orientation.

The differences in the hydration pattern of the modified nucleosides were better revealed by the radial distribution function calculations. All the force field-water model combinations predicted similar distances of the first hydration shell corresponding to the water molecules around the HN1 atoms of Ψ, m^{3}Ψ, and Ψm residues and the HN3 atoms of Ψ, m^{1}Ψ and Ψm residues. In general, FF99_χ_{ND}_bsc0 parameter sets predicted greater numbers of water molecules around the HN1 atoms of Ψ and m^{3}Ψ nucleosides but lesser number of water molecules around the HN1 atoms of Ψm than what were predicted by the FF99_χ_{IDRP}_bsc0 parameters in combination with each of the water models. Although the number of water molecules around the HN3 atoms were observed to be similar for each of the residues with each of the force field-water model combinations, interestingly with the FF99_$\boldsymbol{\chi}_{\text{IDRP}}$_bsc0 parameter sets the number of water molecules around the HN3 atom of Ψm was greater than that predicted by FF99_$\boldsymbol{\chi}_{\text{IDRP}}$_bsc0 in combination with each of the water models.

The orientation of the 2′-OH atom was observed to be similar with for each of the modified residues under this study, all the force field water model combinations predicted the predominant orientation of the 2′-OH atom towards O3′ which is consistent with previous NMR results [64]. Interestingly, with the FF99_$\boldsymbol{\chi}_{\text{IDRP}}$_bsc0 force field the presence of a prominent peak at the O4′-domain was observed in the distribution of the **θ** torsion angle (H2′-C2′-O2′-HO2′).

We are further validating the revised sets of parameters for the modified nucleosides under this study, based on their performance in reproducing the conformational and hydration characteristics of the single-stranded and double-stranded RNA contexts.

SUPPORTING INFORMATION.

AMBER preparatory files including the newly developed partial atomic charges for pseudouridine (Ψ) [PSU_ND.prepin], 1-methylpseudouridine ($m^1\Psi$) [1MP_ND.prepin], 3-methylpseudouridine ($m^3\Psi$) [3MP_ND.prepin], 2′-O-methylpseudouridine (Ψm) [MRP_ND.prepin]. Revised parameter sets i.e. AMBER frcmod files for pseudouridine (Ψ) [PSU_FF99_Chi_ND_bsc0.frcmod], 1-methylpseudouridine ($m^1\Psi$) [1MP_FF99_Chi_ND_bsc0.frcmod]; 3-methylpseudouridine ($m^3\Psi$) [3MP_FF99_Chi_ND_bsc0.frcmod] and 2′-O-methylpseudouridine (Ψm) [MRP_FF99_Chi_ND_bsc0.frcmod]. Table S1. Frozen and restrained dihedrals during QM optimization in PES scan and MM energy minimizations. Table S2. Partial atomic charges for Ψ, $m^1\Psi$, $m^3\Psi$, and Ψm ribonucleosides. Table S3. Propensity (in %) for NORTH sugar puckering of Ψ, $m^1Ψ$, $m^3Ψ$, and Ψm ribonucleosides. Table S4. Fraction (in %) of base orientation states for Ψ, $m^1Ψ$, $m^3Ψ$, and Ψm ribonucleosides. Table S5. Fraction (in %) of **γ** conformational states for Ψ, $m^1Ψ$, $m^3Ψ$, and Ψm ribonucleosides. Figure S1. Structures of Ψ, $m^1Ψ$, $m^3Ψ$, and Ψm ribonucleosides corresponding to the five conformational schemes. Figure S2. Conformational scheme (SC5) used in this work for the Ψ, $m^1Ψ$, $m^3Ψ$, and Ψm ribonucleoside residues (along with the atom names). Figure S3. Energy profiles around **χ** torsional angles from QM calculations corresponding to the five schemes for the Ψ, $m^1Ψ$, $m^3Ψ$, and Ψm ribonucleoside residues. Figures S4-6. Population distribution of the Ψ, $m^1Ψ$, $m^3Ψ$, and Ψm residues corresponding to the different force field and water model combinations with the pseudorotation angle (P) along the x-axis and the glycosidic torsion angle (**χ**) along the y-axis for the three independent sets of 16 ns REMD simulations respectively. Figures S7-9. Population distribution of the Ψ, $m^1Ψ$, $m^3Ψ$, and Ψm residues corresponding to the different force field and water model combinations with the pseudorotation angle (P) along the x-axis and the gamma torsion angle (**γ**) along the y-axis for the three independent sets of 16 ns REMD simulations respectively. Figures S10-12. RDFs of water oxygen atoms around the HN1 atom of the Ψ, $m^3Ψ$, and Ψm residues corresponding to the different force field and water model combinations for the three independent sets of 16 ns REMD simulations respectively. Figures S13-15. RDFs of water oxygen atoms around the HN3 atom of the Ψ, $m^1Ψ$, and Ψm residues corresponding to the different force field and water model combinations for the three independent sets of 16 ns REMD simulations respectively.

ACKNOWLEDGMENTS

The authors acknowledge the support provided by the SERB funding (EMR/2016/007753). N.D. acknowledges support from the DST INSPIRE Junior Research Fellowship (DST/INSPIRE Fellowship/2018/IF180895). The simulations and calculations were performed at the Poznan Supercomputing and Networking Center. The authors would like to thank Aditya Kumar Sarkar and Rupak Pal for technical help.

Conflicts of interest

There are no conflicts to declare.

Author Contributions

N.D., and A.L. jointly conceived and designed the study. N.D. generated the data and analyzed the results. All authors reviewed the manuscript and approved the final version.

REFERENCES

1. Boccaletto P, Machnicka MA, Purta E, et al (2018) MODOMICS: a database of RNA modification pathways. 2017 update. Nucleic Acids Research 46:D303–D307

2. Cohn WE, Volkin E (1951) Nucleoside-5′-Phosphates from Ribonucleic Acid. Nature 167:483–484

3. Davis FF, Allen FW (1957) Ribonucleic acids from yeast which contain a fifth nucleotide. J Biol Chem 227:907–915

4. Charette M, Gray MW (2000) Pseudouridine in RNA: what, where, how, and why. IUBMB Life 49:341–351

5. Scannell JP, Crestfield AM, Allen FW (1959) Methylation studies on various uracil derivatives and on an isomer of uridine isolated from ribonucleic acids. Biochimica et Biophysica Acta 32:406–412

6. Yu C-T, Allen FW (1959) Studies of an isomer of uridine isolated from ribonucleic acids. Biochimica et Biophysica Acta 32:393–406

7. Cohn WE (1959) 5-Ribosyl uracil, a carbon-carbon ribofuranosyl nucleoside in ribonucleic acids. Biochim Biophys Acta 32:569–571

8. Cohn WE (1960) Pseudouridine, a Carbon-Carbon Linked Ribonucleoside in Ribonucleic Acids:

Isolation, Structure, and Chemical Characteristics. Journal of Biological Chemistry 235:1488–1498

9. Hall KB, McLaughlin LW (1992) Properties of pseudouridine N1 imino protons located in the major groove of an A-form RNA duplex. Nucleic Acids Research 20:1883–1889

10. Carlile TM, Rojas-Duran MF, Zinshteyn B, et al (2014) Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. Nature 515:143–146

11. Schwartz S, Bernstein DA, Mumbach MR, et al (2014) Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. Cell 159:148–162

12. Li X, Zhu P, Ma S, et al (2015) Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. Nat Chem Biol 11:592–597

13. Zaringhalam M, Papavasiliou FN (2016) Pseudouridylation meets next-generation sequencing. Methods 107:63–72

14. Lei Z, Yi C (2017) A Radiolabeling-Free, qPCR-Based Method for Locus-Specific Pseudouridine Detection. Angew Chem Int Ed Engl 56:14878–14882

15. Hall KB, McLaughlin LW (1991) Properties of a U1/mRNA 5’ splice site duplex containing pseudouridine as measured by thermodynamic and NMR methods. Biochemistry 30:1795–1801

16. Davis DR, Veltri CA, Nielsen L (1998) An RNA model system for investigation of pseudouridine stabilization of the codon-anticodon interaction in tRNALys, tRNAHis and tRNATyr. J Biomol Struct Dyn 15:1121–113

17. Yarian CS, Basti MM, Cain RJ, et al (1999) Structural and functional roles of the N1- and N3-protons of psi at tRNA’s position 39. Nucleic Acids Res 27:3543–3549

18. Newby MI, Greenbaum NL (2002) Investigation of Overhauser effects between pseudouridine and water protons in RNA helices. Proc Natl Acad Sci U S A 99:12697–12702

19. Hudson GA, Bloomingdale RJ, Znosko BM (2013) Thermodynamic contribution and nearest-

neighbor parameters of pseudouridine-adenosine base pairs in oligoribonucleotides. RNA 19:1474–1482

20. Kierzek E, Malgowska M, Lisowiec J, et al (2014) The contribution of pseudouridine to stabilities and structure of RNAs. Nucleic Acids Research 42:3492–3501

21. Deb I, Popenda Ł, Sarzyńska J, et al (2019) Computational and NMR studies of RNA duplexes with an internal pseudouridine-adenosine base pair. Sci Rep 9:16278

22. Davis DR (1995) Stabilization of RNA stacking by pseudouridine. Nucleic Acids Res 23:5020–5026

23. Auffinger P, Westhof, E. (1998**)** in Modification and Editing of RNA, eds. Grosjean H, Benne R (Am. Soc. Microbiol., Washington, DC): 103–112

24. Deb I, Pal R, Sarzynska J, Lahiri A (2016) Reparameterizations of theχTorsion and Lennard-JonesσParameters Improve the Conformational Characteristics of Modified Uridines. Journal of Computational Chemistry 37:1576–1588

25. Aduri R, Psciuk BT, Saro P, et al (2007) AMBER Force Field Parameters for the Naturally Occurring Modified Nucleosides in RNA. Journal of Chemical Theory and Computation 3:1464–1475

26. Dutta N, Sarzynska J, Lahiri A (2020) Molecular Dynamics Simulation of the Conformational Preferences of Pseudouridine Derivatives: Improving the Distribution in the Glycosidic Torsion Space. J Chem Inf Model 60:4995–5002

27. Yildirim I, Stern HA, Kennedy SD, et al (2010) Reparameterization of RNA χ Torsion Parameters for the AMBER Force Field and Comparison to NMR Spectra for Cytidine and Uridine. Journal of Chemical Theory and Computation 6:1520–1531

28. Zgarbová M, Otyepka M, Sponer J, et al (2011) Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. J Chem Theory Comput 7:2886–2902

29. Kührová P, Otyepka M, Šponer J, Banáš P (2014) Are Waters around RNA More than Just a Solvent? - An Insight from Molecular Dynamics Simulations. J Chem Theory Comput 10:401–411

30. Auffinger P, Westhof E (1997) RNA hydration: three nanoseconds of multiple molecular dynamics simulations of the solvated tRNA Asp anticodon hairpin 1 1Edited by J. Karn. Journal of Molecular Biology 269:326–341

31. Jorgensen WL, Chandrasekhar J, Madura JD, et al (1983) Comparison of simple potential functions for simulating liquid water. The Journal of Chemical Physics 79:926–935

32. Abascal, JLF, Vega C (2005) A general purpose model for the condensed phases of water: TIP4P/2005. The Journal of Chemical Physics. 123, 234505

33. Mahoney MW, Jorgensen WL (2000) A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. Journal of Chemical Physics. 112, 8910−8922.

34. Berendsen HJC, Grigera JR, Straatsma TP (1987) The missing term in effective pair potentials. The Journal of Physical Chemistry 91:6269–6271

35. Gelbin A, Schneider B, Clowney L, et al (1996) Geometric Parameters in Nucleic Acids: Sugar and Phosphate Constituents. Journal of the American Chemical Society 118:519–529

36. Schaftenaar G, Noordik JH (2000) Journal of Computer-Aided Molecular Design. 14:123–134

37. Richardson JS, Schneider B, Murray LW, et al (2008) RNA backbone: consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). RNA 14:465–481

38. Saenger W (1984) Principles of Nucleic Acid Structure. Springer Advanced Texts in Chemistry

39. Frisch MJ, Trucks GW, Schlegel HB, et al (2009) Gaussian 09; Gaussian, Inc.: Wallingford, CT.

40. Bayly CI, Cieplak P, Cornell W, Kollman PA (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. The Journal of Physical Chemistry 97:10269–10280

41. Cieplak P, Cornell WD, Bayly C, Kollman PA (1995) Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. Journal of Computational Chemistry 16:1357–1377

42. Vanquelef E, Simon S, Marquant G, et al (2011) R.E.D. Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. Nucleic Acids Research 39:W511–W517

43. Case DA, Cheatham TE, Darden T, et al (2005) The Amber biomolecular simulation programs. Journal of Computational Chemistry 26:1668–1688

44. Wang J, Cieplak P, Kollman PA (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? Journal of Computational Chemistry 21:1049–1074

45. Pérez A, Marchán I, Svozil D, et al (2007) Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. Biophysical Journal 92:3817–3829

46. Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. Chemical Physics Letters 314:141–151

47. Horn HW, Swope WC, Pitera JW, et al (2004) Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. J Chem Phys 120:9665–9678

48. Ryckaert J-P, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. Journal of Computational

Physics 23:327–341

49. Altona C, Sundaralingam M (1972) Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation. J Am Chem Soc 94:8205–8212

50. IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN) (**1983)** Abbreviations and Symbols for the Description of Conformations of Polynucleotide Chains. Recommendations 1982. Eur. J. Biochem., 131 (1): 9–15

51. Foloppe N, Hartmann B, Nilsson L, MacKerell AD Jr (2002) Intrinsic Conformational Energetics Associated with the Glycosyl Torsion in DNA: A Quantum Mechanical Study. Biophysical Journal. 82 (3): 1554–1569

52. Foloppe N, Nilsson L (2005) Toward a Full Characterization of Nucleic Acid Components in Aqueous Solution: Simulations of Nucleosides. Journal of Physical Chemistry B 109 (18): 9119–9131

53. Case DA, Ben-Shalom IY, Brozell, SR, et al (2018) AMBER 2018; University of California: San Francisco.

54. 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

55. Chang Y-C, Herath J, Wang TH-H, Chow CS (2008) Synthesis and solution conformation studies of 3-substituted uridine and pseudouridine derivatives. Bioorg Med Chem 16:2676–2686

56. Neumann JM, Bernassau JM, Guéron M, Tran-Dinh S (1980) Comparative conformations of uridine and pseudouridine and their derivatives. Eur J Biochem 108:457–463

57. Desaulniers J-P, Chui HM-P, Chow CS (2005) Solution conformations of two naturally occurring RNA nucleosides: 3-methyluridine and 3-methylpseudouridine. Bioorg Med Chem 13:6777–6781

58. Deb I, Sarzynska J, Nilsson L, Lahiri A (2014) Conformational preferences of modified uridines: comparison of AMBER derived force fields. J Chem Inf Model 54:1129–1142

59. Fohrer J, Hennig M, Carlomagno T (2006) Influence of the 2′-Hydroxyl Group Conformation on the Stability of A-form Helices in RNA. Journal of Molecular Biology 356:280–287

60. Hou Y-M (2001) An important 2'-OH group for an RNA-protein interaction. Nucleic Acids Research 29:976–985

61. Egli M, Portmann S, Usman N (1996) RNA hydration: a detailed look. Biochemistry 35:8489–8494

62. Olsen GL, Bardaro MF Jr, Echodu DC, et al (2009) Hydration dependent dynamics in RNA. J Biomol NMR 45:133–142

63. Auffinger P, Westhof E (1997) Rules governing the orientation of the 2′-hydroxyl group in RNA. Journal of Molecular Biology 274:54–63

64. Nozinovic S, Fürtig B, Jonker HRA, et al (2010) High-resolution NMR structure of an RNA model system: the 14-mer cUUCGg tetraloop hairpin RNA. Nucleic Acids Res 38:683–694