

OPEN

# Computational and NMR studies of RNA duplexes with an internal pseudouridine-adenosine base pair

Indrajit Deb<sup>1,2,4,5</sup>, Łukasz Popena<sup>3,5</sup>, Joanna Sarzyńska<sup>1\*</sup>, Magdalena Małgowska<sup>1</sup>, Ansuman Lahiri<sup>2</sup>, Zofia Gdaniec<sup>1</sup> & Ryszard Kierzek<sup>1</sup>

Pseudouridine ( $\Psi$ ) is the most common chemical modification present in RNA. In general,  $\Psi$  increases the thermodynamic stability of RNA. However, the degree of stabilization depends on the sequence and structural context. To explain experimentally observed sequence dependence of the effect of  $\Psi$  on the thermodynamic stability of RNA duplexes, we investigated the structure, dynamics and hydration of RNA duplexes with an internal  $\Psi$ -A base pair in different nearest-neighbor sequence contexts. The structures of two RNA duplexes containing 5'-G $\Psi$ C/3'-CAG and 5'-C $\Psi$ G/3'-GAC motifs were determined using NMR spectroscopy. To gain insight into the effect of  $\Psi$  on duplex dynamics and hydration, we performed molecular dynamics (MD) simulations of RNA duplexes with 5'-G $\Psi$ C/3'-CAG, 5'-C $\Psi$ G/3'-GAC, 5'-A $\Psi$ U/3'-UAA and 5'-U $\Psi$ A/3'-AAU motifs and their unmodified counterparts. Our results showed a subtle impact from  $\Psi$  modification on the structure and dynamics of the RNA duplexes studied. The MD simulations confirmed the change in hydration pattern when U is replaced with  $\Psi$ . Quantum chemical calculations showed that the replacement of U with  $\Psi$  affected the intrinsic stacking energies at the base pair steps depending on the sequence context. The calculated intrinsic stacking energies help to explain the experimentally observed sequence dependent changes in the duplex stability from  $\Psi$  modification.

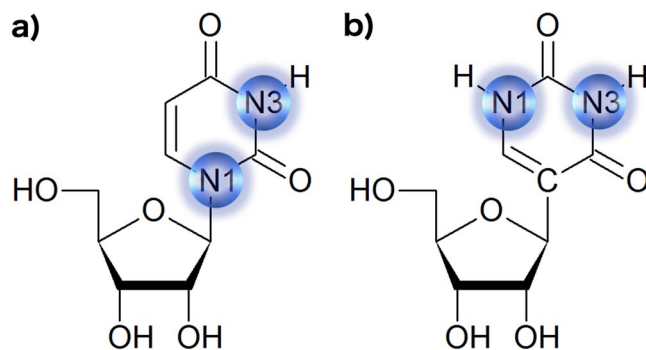
RNA features a vast collection of chemically and structurally diverse posttranscriptionally modified nucleosides. Pseudouridine ( $\Psi$ ) was the first such modification discovered<sup>1</sup> and is the most abundant.  $\Psi$  has historically been viewed as an attribute of stable noncoding RNAs including tRNA, rRNA and snRNA<sup>2</sup>. The development of next-generation sequencing methods for detection and mapping of RNA modifications revealed that  $\Psi$ s are also abundant in mRNA and long noncoding RNA. The number of  $\Psi$ s in mRNA was found to be dynamically regulated in response to cellular conditions<sup>3-6</sup>. Recent studies have discovered that  $\Psi$  participates in the control of various layers of gene expression regulation<sup>7,8</sup>. Research on  $\Psi$ , whose role ranges from fine-tuning to being functionally essential to the target RNA, is gaining renewed interest<sup>9-11</sup>.

Pseudouridine is an isomer of the nucleoside uridine in which the uracil is attached to the sugar via the C5-C1' instead of the N1-C1' glycosidic bond (Fig. 1). Compared with uridine,  $\Psi$  has an additional hydrogen bond donor (N1 imino proton) and the same Watson-Crick base pairing property. In understanding the mechanism of  $\Psi$  functionality, significant contributions have been made by structural and thermodynamic studies. Previous studies reported that the presence of  $\Psi$  mainly modulates local RNA conformation and flexibility and generally is not associated with major structural changes<sup>2</sup>. In most of the structures studied so far,  $\Psi$  contributed to the structural stability<sup>2</sup>. The impact of the  $\Psi$  modification depends on the RNA secondary structure context (stems or loops). Early NMR studies revealed that the presence of  $\Psi$  in single strand RNA favors the formation of the C3'-*endo* conformation and promotes base stacking, thus stabilizing RNA structures<sup>12</sup>. In RNA hairpins  $\Psi$  stabilizes base pairing and stacking interactions. In particular, in the anticodon-stem loop of tRNA,  $\Psi$ 39 is involved in the stabilization of the A31- $\Psi$ 39 base pair at the interface between the anticodon loop and stem region<sup>13-16</sup>. The consecutive  $\Psi$ 38 and  $\Psi$ 39 residues in tRNA<sup>His</sup> provide stacking stabilization for bases on the 3' side of the

<sup>1</sup>Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704, Poznan, Poland.

<sup>2</sup>Department of Biophysics, Molecular Biology & Bioinformatics, University of Calcutta, Kolkata, 700009, West Bengal, India. <sup>3</sup>NanoBioMedical Centre, Adam Mickiewicz University, Umultowska 85, 61-614, Poznan, Poland.

<sup>4</sup>Present address: Department of Biophysics, University of Michigan, 930 North University Avenue, Ann Arbor, Michigan, 48109, USA. <sup>5</sup>These authors contributed equally: Indrajit Deb and Łukasz Popena. \*email: [joanna.sarzyńska@ibch.poznan.pl](mailto:joanna.sarzyńska@ibch.poznan.pl)



**Figure 1.** Chemical structure of (a) uridine and (b) pseudouridine.

tRNA anticodon loop<sup>17</sup>. Also, the conserved  $\Psi$  modifications in helix 69 (H69) from large subunit ribosomal RNA have been shown to modulate the structure and conformational behavior of the stem-loop region by stabilizing the loop-closing base pair and promoting base stacking in the 3' half of the loop<sup>18,19</sup> and an additional A- $\Psi$  base pair within the loop<sup>20</sup>. The UV melting data revealed that the  $\Psi$  residues in the stem region provided thermodynamic stabilization to the RNA, whereas individual  $\Psi$ s located in the loop region contributed to a slight destabilization<sup>21,22</sup>. Moreover, structural studies of the P6.1 hairpin derived from the human telomerase RNA activation domain revealed that two  $\Psi$  modifications located in the loop increased both base-stacking and hydrogen-bonding interactions within the loop compared with the unmodified construct, resulting in higher thermodynamic stability. Here, the insertion of two  $\Psi$ s in the loop increased the RNA thermodynamic stability more than the insertion of two  $\Psi$ s in the stem<sup>23</sup>. The presence of  $\Psi$  modification within the branch-site recognition region of spliceosomal U2 snRNA induces a bulged conformation of the branch-point adenosine<sup>24,25</sup>. Replacing uridines with  $\Psi$  in RNAs containing toxic CUG or CCUG repeats that form helical conformations induced structural stabilization<sup>26</sup>. Furthermore,  $\Psi$  is known to modulate codon-anticodon interactions between mRNA and tRNA<sup>17</sup> allowing purine-purine base pairing<sup>27,28</sup>. The effect of  $\Psi$  modification on structure and thermodynamic stability is commonly attributed to the stabilization of the C3'-endo conformation, increased base stacking properties<sup>12</sup>, and the presence of an additional N1 H-bond donor that coordinates structural water molecules<sup>29,30</sup>, and all these components are considered to be interdependent<sup>17</sup>.

We previously reported the influence of pseudouridylation on the thermodynamic stability of RNA duplexes with  $\Psi$  modifications where the  $\Psi$  was located at various positions and formed base pairs with A, G, U and C<sup>31</sup>. For RNA duplexes with central  $\Psi$ -A base pairs, we determined the influences of A-U and G-C flanking base pairs on the thermodynamic stabilities. Our results indicated that the enhancement in the duplex stability from replacing U-A with  $\Psi$ -A depends on the sequence context. When U was replaced with  $\Psi$  in the duplex with the 5'-C $\Psi$ G/3'-GAC motif, we observed greater enhancement in the duplex stability (2.43 kcal/mol) than those in duplexes with other motifs (up to 0.7 kcal/mol). Moreover, in the duplex with the 5'-A $\Psi$ U/3'-UAA motif, the replacement of U with  $\Psi$  stabilized the duplex 3 kcal/mol less than what was predicted using the available nearest-neighbor parameters<sup>32</sup>.

In this work, we determined the NMR structures of two 9-mer RNA duplexes with 5'-G $\Psi$ C/3'-CAG and 5'-C $\Psi$ G/3'-GAC motifs, to examine whether structural factors could be responsible for the differences observed in the increment of thermodynamic stabilities. To further elucidate how the replacement of the middle U-A base pair with  $\Psi$ -A in different sequence contexts influences the structure, dynamics and hydration of RNA duplexes, we performed a systematic study using molecular dynamics (MD) simulations. We carried out MD simulations in explicit solvent for all duplexes with internal  $\Psi$ -A base pairs studied in previous work<sup>31</sup>, namely, duplexes with 5'-G $\Psi$ C/3'-CAG, 5'-C $\Psi$ G/3'-GAC, 5'-A $\Psi$ U/3'-UAA and 5'-U $\Psi$ A/3'-AAU motifs and their unmodified counterparts. To understand how the replacement of U with  $\Psi$  changes the stacking interactions, we applied high-level *ab initio* methods and performed calculations for eight unique base pair steps for the RNA structures derived from MD simulations.

## Results

The RNA duplexes considered in this study are listed in Table 1.

**NMR study of duplex-G $\Psi$ C and duplex-C $\Psi$ G.** Analysis of the 2D NOESY spectra (recorded in D<sub>2</sub>O) used for nonexchangeable proton assignment revealed NOE connectivity typical of A-form geometry. Expanded contour plots of the NOESY spectra corresponding to the interactions between the base H6/H8 and H1'/H5 protons with sequential pathways traced for both duplexes are shown in Fig. 2. For the  $\Psi$  residues at position 5 ( $\Psi$ 5), characteristic upfield shifts of the H1' resonances relative to the remaining H1' resonances were observed (Supplementary Tables S1, S3). This tendency is a signature of  $\Psi$  modification and is consistent with previous NMR reports regarding pseudouridine-containing RNAs<sup>15,23,33</sup>. Similar to <sup>1</sup>H, the <sup>13</sup>C chemical shifts of C1'- $\Psi$ 5 differ significantly from the ranges observed for the other anomeric carbon resonances (Supplementary Tables S2, S4). The upfield shift of the H1' and C1' resonances in the  $\Psi$  residues can be explained by the less electronegative C5 replacing the N1 atom in the N-glycosidic bond position<sup>12,30</sup>. The resonance assignments of the duplex-G $\Psi$ C and -C $\Psi$ G are listed in Supplementary Tables S1–S4.

Duplex name	Sequence	Method		Duplex name	Sequence	Method
		NMR	MD			
duplex-G $\Psi$ C	5'-UCAG <u><math>\Psi</math></u> CAGU-3' 3'-AGUCAGUCA-5'	✓	✓	duplex-GUC	5'-UCAGUCAGU-3' 3'-AGUCAGUCA-5'	✓
duplex-C $\Psi$ G	5'-UCAC <u><math>\Psi</math></u> GAGU-3' 3'-AGUGACUCA-5'	✓	✓	duplex-CUG	5'-UCACUGAGU-3' 3'-AGUGACUCA-5'	✓
duplex-A $\Psi$ U	5'-UCA <u><math>\Psi</math></u> UAGU-3' 3'-AGUAAUCA-5'		✓	duplex-AUU	5'-UCAAUUAGU-3' 3'-AGUAAUCA-5'	✓
duplex-U $\Psi$ A	5'-UCAU <u><math>\Psi</math></u> AAGU-3' 3'-AGUAAUCA-5'		✓	duplex-UUA	5'-UCAUUAAGU-3' 3'-AGUAAUCA-5'	✓

**Table 1.** List of the  $\Psi$ -modified RNA duplexes and their unmodified counterparts studied in this work. Underlined are three nucleotides in the middle of the first strand of the duplexes. Sequences of the modified or unmodified duplexes differ only in the base pairs flanking the central  $\Psi$ -A and U-A, respectively.

The NMR structures of the G $\Psi$ C- and C $\Psi$ G-duplexes were solved on the basis of NMR observables (Table 2) using the restrained molecular dynamics approach described in the Methods section. The final set of 10 structures showed that both duplexes form well-defined structures with overall RMSD for heavy atoms within 0.20 Å and 0.21 Å for duplex-G $\Psi$ C and duplex-C $\Psi$ G, respectively (Table 2, Fig. 3). To explore the conformational differences between both duplexes, we calculated the geometric parameters<sup>34</sup>. The results of the Curves+ analysis indicated that the duplexes adopted typical A-form conformations, with helical parameters and dihedral angles (backbone, sugar and glycosidic) that fall within the distributions of canonical A-RNA duplexes<sup>35</sup>. This result implied the *anti*-orientation of all the nucleobases and C3'-*endo* sugar ring conformation with calculated average pseudorotation angles and puckering amplitudes of approximately 11° and 42°, respectively. The calculated C1'-C1' distances averaged over all the base pairs in all the models were  $10.75 \pm 0.05$  Å for both duplexes. A detailed comparison of the two duplexes with the use of rigid body parameters that describe the geometry of the base pairs and sequential base pair steps (Supplementary Tables S5, S6) showed that the change of the 5' versus 3' order of the G-C and C-G Watson-Crick base-pairs adjacent to the central  $\Psi$ -A pair did not affect the local structure.

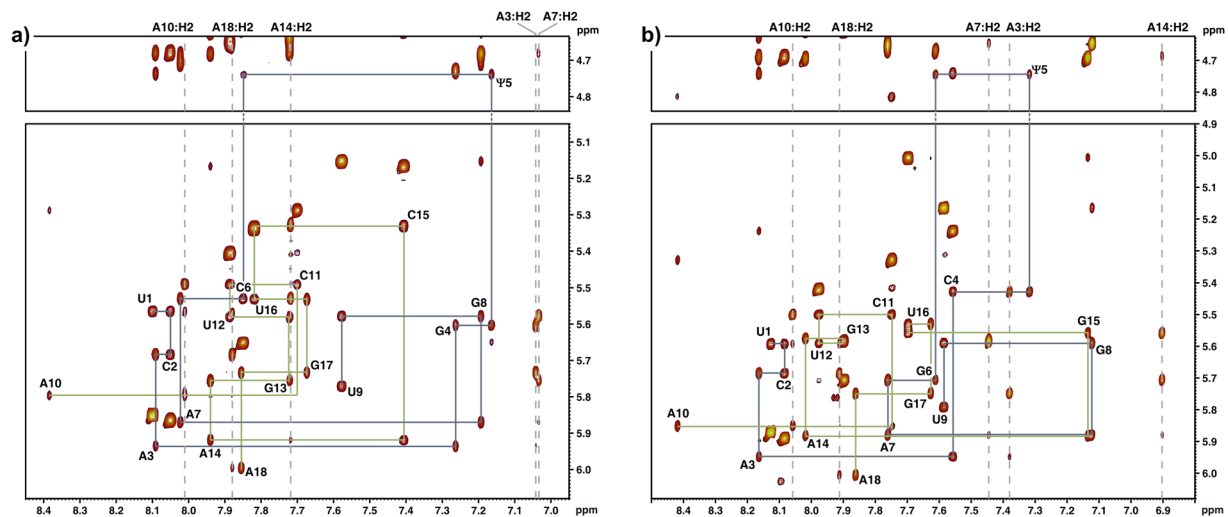
#### Molecular dynamics: Replacement of U with $\Psi$ has only subtle effects on the duplex structures.

To evaluate the effects of the  $\Psi$  modification on the structure and dynamics of the RNA duplexes with central  $\Psi$ -A base pairs in a systematic way, we carried out 500 ns MD simulations under the same conditions for all the duplexes specified in Table 1. The analysis of the MD trajectories aimed to answer the following questions: (i) for a given sequence, how does the presence of  $\Psi$  influence the conformation and flexibility of the modified duplex compared with its unmodified counterpart, (ii) whether the impact of  $\Psi$  depends on the identities of the adjacent base pairs. The analysis of the trajectories showed that both the modified and unmodified duplexes retained the initial A-form geometry with root-mean-square deviations (RMSDs) of approximately 1.4 Å (Supplementary Table S7 and Supplementary Fig. S1).

To probe the influence of the  $\Psi$  modification on flexibility of the RNA duplexes, we determined the root-mean-square fluctuations (RMSFs) per residue for the backbone atoms (P, OP1, OP2, O5', C5' and O3') and the order parameters ( $S^2$ ) for sugar C1'-H1' vectors. We found that in the duplexes with  $\Psi$  modification, the backbone flexibility was only marginally lower than that in the corresponding reference duplexes. For the duplexes with the central R $\Psi$ Y motif (R-purine and Y-pyrimidine), the decrease in RMSF was mainly localized in residues 5 and 6, whereas for the Y $\Psi$ R motif, this effect propagated along the entire strand (Supplementary Fig. S2). The  $S^2$  was higher than 0.87 when excluding terminal residues, that indicates low flexibility for the nucleotides within the duplexes, both with and without  $\Psi$  (Supplementary Fig. S3). The  $S^2$  was only marginally higher for residues in the modified strand in duplexes with Y $\Psi$ R motif.

To gain insight into the structural differences between the RNA duplexes with central  $\Psi$ -A base pairs and their unmodified counterparts at a more detailed and local level, we calculated the helical parameters<sup>34</sup> (Fig. 4, Supplementary Table S8, Supplementary Figs S4 and S5). Overall, the replacement of the base pair U-A by  $\Psi$ -A induced only slight changes in certain base pair and base pair step parameters. For example, for the  $\Psi$ 5-A14 base pairs, we observed slightly smaller values (by 3–4°) for the base pair openings compared with those obtained for the U5-A14 base pairs (Fig. 4 and Supplementary Fig. S4). The  $\Psi$ -modified duplexes displayed similar roll values at each base pair step to those of their unmodified counterparts (Supplementary Fig. S5b). Our results agree with the previously reported sequence dependence for the roll values of RNA double helices [pyrimidine-purine (YR) > purine-purine (RR) > purine-pyrimidine (RY)]<sup>36</sup>. We observed small increments in the backbone P4-P5 distances and decrements in the P3-P4 and P5-P6 distances in all the  $\Psi$ -modified strands (Supplementary Fig. S6). Further, we noticed small increments ( $\sim 0.2$  Å) in the interstrand C1'-C1' distances for the  $\Psi$ 5-A14 base pairs compared with those of the U5-A14 base pairs (Supplementary Fig. S7). The glycosidic torsion angle ( $\chi$ ) for the  $\Psi$ 5 was shifted 10° towards lower *anti* values ( $\sim 195^\circ$ ) compared with that of U5 ( $\sim 205^\circ$ ) in all the duplexes studied (Supplementary Fig. S8). In summary, we conclude that all the duplexes with central  $\Psi$ -A base-pairs maintain A-form conformations. The  $\Psi$  modification only subtly alters the local structure with marginal dependence on the sequence, which implies similar base stacking geometries in the modified and unmodified duplexes.

**Additional water bridge in the  $\Psi$ -modified duplexes.** MD simulations in explicit solvents are well suited to analyze hydration, especially regarding predictions of highly specific long-residency hydration sites<sup>37</sup>. To identify differences in the hydration patterns of  $\Psi$ -modified and unmodified duplexes, we calculated the



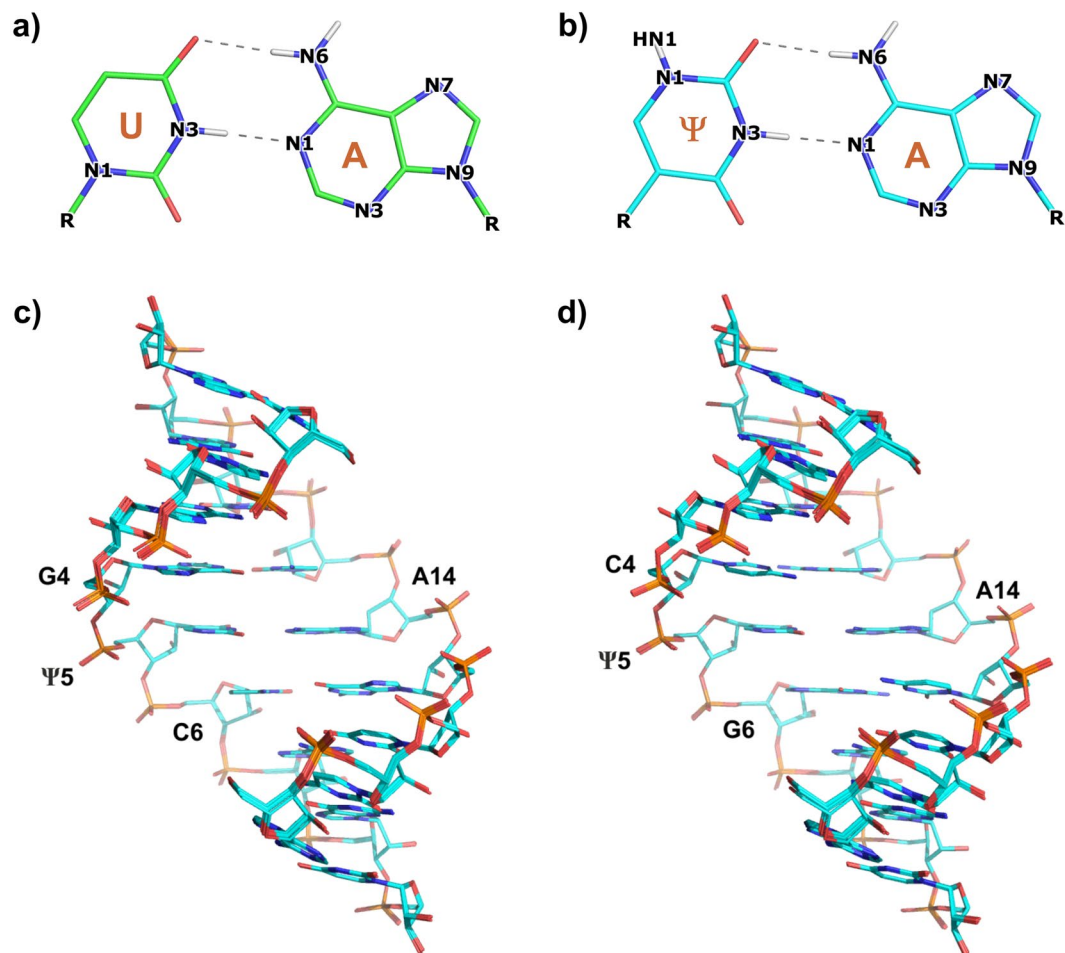
**Figure 2.** Fingerprint (H6/H8-H1') regions of the 400 ms NOESY spectra recorded in D<sub>2</sub>O at 25 °C. The sequential connectivities are traced with either a blue or a green line for both strands of the (a) duplex-GΨC and (b) duplex-CΨG.

Restraints statistics		
	Duplex-GΨC 6I1W	Duplex-CΨG 6I1V
<b>No. of restraints</b>		
Total	657	631
Distance restraints, excluding hydrogen bonds	335	310
Intra-residue	206	189
Sequential residues	119	114
Long range	10	7
Hydrogen bond	54	54
Dihedral restraints	178	177
Chirality restraints	90	90
<b>Structural refinement statistics</b>		
Violations		
Average No. of violations (<0.1 Å)	21.70 (0.48)	10.10 (0.32)
Average No. of violations (0.1–0.2 Å)	2.00 (0.00)	<0.01 (0.00)
Maximum violation (Å)	0.12 (0.00)	0.07 (0.00)
Mean torsion penalty (kcal/mol)	<0.01 (0.00)	<0.01 (0.00)
Mean distance penalty (kcal/mol)	1.39 (0.01)	0.17 (0.00)
Mean deviations from idealized geometry		
Bonds (Å)	0.01 (0.00)	0.01 (0.00)
Angles (deg)	2.46 (0.00)	2.45 (0.01)
Maximum RMSD (Å) for heavy atoms	0.20	0.21

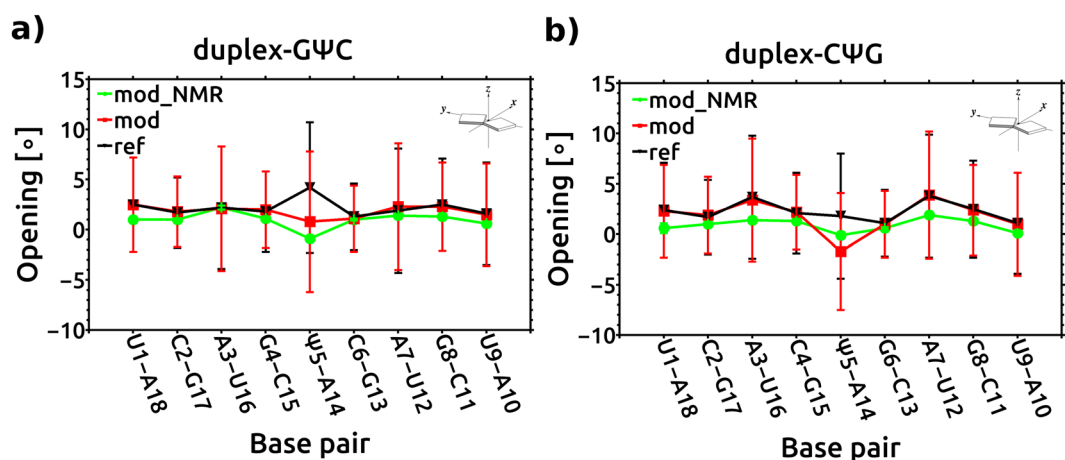
**Table 2.** Restraints and structural refinement statistics for ten structures of RNA duplexes with Ψ. Standard deviations over 10 structures are in parentheses.

radial distribution functions (RDFs) around the selected atoms in the modification region. The RDF plot of the water oxygen atoms with respect to the HN1-Ψ5 atom (Fig. 5a) for the duplex-CΨG indicates the formation of a well-defined first hydration shell between 1.5 Å and 2.5 Å with a maximum at 2.05 Å. For the corresponding hydrogen atom H5 in the unmodified U5 base, the water molecules are farther from the base. The first solvation peak of H5-U5 corresponds to the second peak around atom HN1-Ψ5, but fewer water molecules were found at this distance in the modified duplex. The distribution of water oxygens between the base and the backbone was found to be highly ordered around the geometric center between the OP2 atom and the HN1 atom of the Ψ5 residue (Fig. 5b). In general, the RDFs were not very sensitive to the sequence context of the RNA duplexes.

To compare the structure of the water around the Ψ and the corresponding U in modified and unmodified duplexes at the atomic level, we calculated the occupancy of the water-mediated hydrogen bonding bridges. In the simulations of the Ψ-modified duplexes, a frequent intramolecular water bridge was formed between the HN1 and OP2 atoms of Ψ5, with an occupancy of 28–29% (Table 3). This configuration was the most frequent water bridge seen around the duplexes, with the occupancy slightly higher than that of the water bridges between the



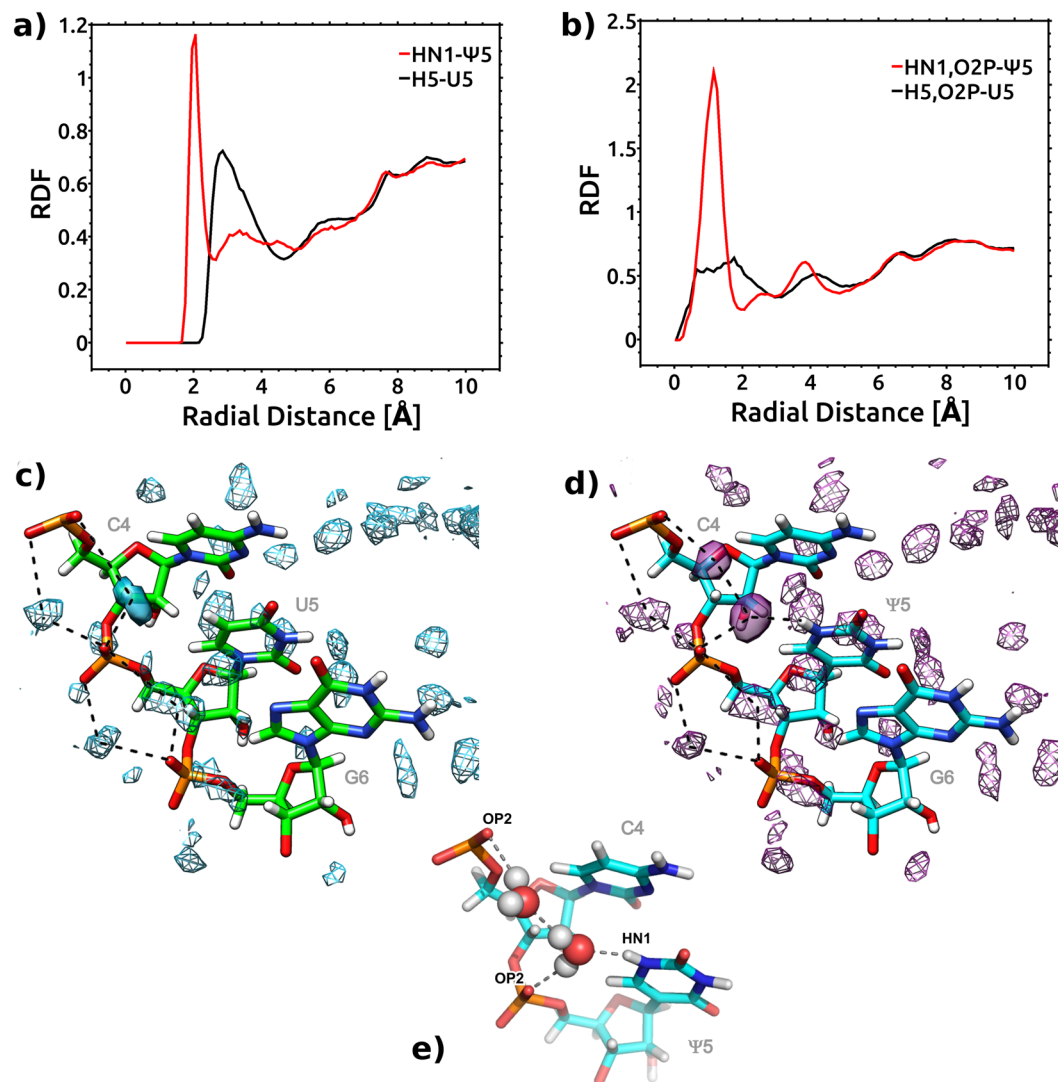
**Figure 3.** Base pairing of (a) uridine and (b) pseudouridine with adenosine. Superposition of the ten lowest-energy solution NMR structures: (c) duplex-GΨC, (d) duplex-CΨG.



**Figure 4.** Average opening values (deg) for the base pairs in (a) duplex-GΨC and (b) duplex CΨG. Green – Ψ-modified NMR models; Red – Ψ-modified duplexes from MD simulation; Black – reference, unmodified from MD simulation. Vertical lines represent standard deviations.

adjacent phosphate oxygens (up to 24%). The change in the hydration pattern due to the Ψ modification also manifested as an increase in the percent of bridging water between the adjacent phosphate oxygen atoms of residues 5 and 6 and a decrease in that between residues 4 and 5. The hydrogen bond between water oxygen and HN1-Ψ5 atom displayed a short lifetime of 0.02 ns in comparison to the long lifetime for the waters hydrogen-bonded to

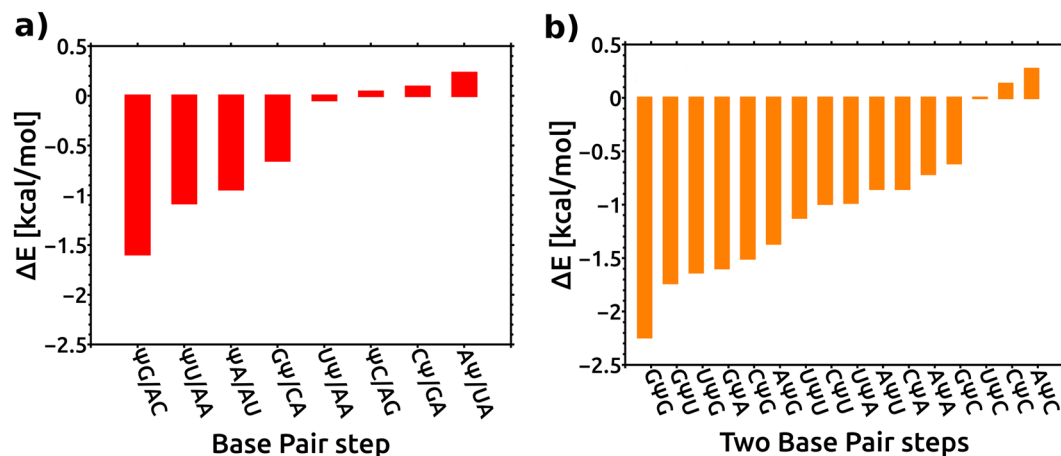




**Figure 5.** Change in the hydration pattern upon U to  $\Psi$  modification for duplex-C $\Psi$ G and duplex-CUG. RDFs of water oxygen atoms (a) around the HN1- $\Psi$ 5 and H5-U5 atoms, respectively; (b) around the geometric center of the OP2- $\Psi$ 5 and HN1- $\Psi$ 5 atoms or OP2-U5 and H5-U5 atoms. Red –  $\Psi$ -modified duplexes; Black – unmodified duplexes. Water occupancy contoured at equivalent levels and hydrogen bonding patterns (in dotted lines) viewed from the major groove. (c) Reference duplex-CUG; differences in hydration patterns are shown in solid. (d)  $\Psi$ -modified duplex-C $\Psi$ G; differences in hydration patterns are shown in solid. (e) Snapshot of two water molecules making contact between HN1, OP2 atoms of  $\Psi$  and OP2 atom of the preceding residue.

Water bridge	duplex-G $\Psi$ C		duplex-C $\Psi$ G		duplex-A $\Psi$ U		duplex-U $\Psi$ A	
	mod	ref	mod	ref	mod	ref	mod	ref
OP2(5)-W-HN1(5)	28	–	29	–	28	–	28	–
OP2(5)-W-OP1(4)	12	16	5	10	10	14	7	13
OP2(5)-W-OP2(4)	7	11	<5	9	6	9	8	13
OP2(5)-W-O5'(4)	<5	6	<5	5	<5	5	<5	<5
OP2(6)-W-OP1(5)	21	15	22	10	20	15	17	9
OP2(6)-W-OP2(5)	16	12	24	18	19	16	19	13
OP2(6)-W-O5'(5)	<5	<5	<5	<5	5	<5	<5	<5

**Table 3.** Water bridges (occupancy in %) formed by the backbone atoms near the modified residue and its unmodified counterpart. Abbreviations ‘mod’ and ‘ref’ stand for the modified and reference, unmodified duplexes, respectively.



**Figure 6.** Impact of  $\Psi$  on the intrinsic stacking energies. (a) Change in the QM stacking energy between base pairs at a given base pair step upon  $\Psi$  modification ( $\Delta E = E_{\text{modif}} - E_{\text{unmodif}}$ ).  $\Psi\text{G}/\text{AC}$  denotes the  $5'\Psi\text{G}/3'\text{AC}$  base pair step and so on. (b) Prediction of the impact of  $\Psi$  on the stacking energies at trinucleotide steps based on QM calculations.  $\text{G}\Psi\text{G}$  denotes the  $5'\text{G}\Psi\text{G}/3'\text{CAC}$  motif and so on. The base pair stacking energies at trinucleotide steps were calculated by the sum of the base pair stacking energies for two consecutive dinucleotide steps. The data are ordered in decreasing stability of eight unique dinucleotide steps/ sixteen unique trinucleotide steps in RNA duplexes.

the OP1 and OP2 atoms (0.5–0.7 ns). In modified duplexes the hydrogen-bonding lifetime to the O2P- $\Psi$ 5 and O2P of preceding residue increased to 1–1.1 ns.

To visualize the differences in the hydration patterns around the residues  $\Psi$ 5 and U5 within the duplexes, the water molecules over the MD trajectories were mapped. Figure 5d shows an additional hydration pocket around the HN1- $\Psi$ 5 hydrogen in the major groove, other than the phosphate backbone bridging water. This additional hydration site was identified in all the  $\Psi$ -modified duplexes and was not present in the unmodified reference structures (Fig. 5c). In our simulations, a water molecule bridges the HN1 and OP2 atoms within the same  $\Psi$  residue, but the water molecule is not directly bonded to the OP2 atom of the preceding residue. On the other hand, this bridging water molecule is a proton donor to another water that is coordinated with the 5'-neighboring phosphate. As a result, the HN1- $\Psi$ 5 is connected with the 5'-neighboring OP2 atom with a chain of two water molecules (Fig. 5e). These observations were similar for all the RNA duplexes studied.

**Stacking Interactions.** Our structural analysis revealed minimal effects on the base stacking geometry within the RNA duplexes from  $\Psi$  modification in different sequence contexts. However, the  $\Psi$  modification changes the electron cloud (charge distribution) compared with unmodified U and thus could influence the base stacking at base pair steps in the duplex<sup>38</sup>. To further explore the effect of  $\Psi$  on the stacking interactions, we used high level quantum mechanical (QM) method to study stacking interaction energies for eight base pair steps containing the  $\Psi$ -A and the reference U-A base pairs. The calculations were performed using the geometries of the stacked base pairs derived from explicit solvent MD simulations of the respective RNA duplexes (Supplementary Fig. S9, Supplementary Methods, Section 6). Figure 6a summarizes the changes in the stacking energies at each base pair step when U was replaced with  $\Psi$ . In addition, the base pair step stacking energies for eight unique base pair steps containing the  $\Psi$ -A base pair and their unmodified counterparts are shown in Supplementary Table S9 and Supplementary Fig. S10. These results show the clear dependence of the change in the stacking energies on the sequence context, ranging from  $-1.59$  to  $0.23$  kcal/mol. The base pair steps can be divided into two groups. In the first group,  $\Psi$  stabilizes the base pair step stacking energies, whereas in the second group, no stabilization is observed. The  $\Psi$  stabilized the base pair step stacking interactions when it was 5' to G, U and A ( $5'\Psi\text{G}/3'\text{AC}$ ,  $5'\Psi\text{U}/3'\text{AA}$  and  $5'\Psi\text{A}/3'\text{UU}$  base pair steps) and when  $\Psi$  was 3' to G ( $5'\text{G}\Psi/3'\text{CA}$  base pair step). In contrast, the  $\Psi$  at the 3' position to A, C, and U ( $5'\text{A}\Psi/3'\text{UA}$ ,  $5'\text{C}\Psi/3'\text{GU}$  and  $5'\text{U}\Psi/3'\text{AA}$  base pair steps) did not stabilize the base pair step stacking interactions. The stabilization effect was also not observed when  $\Psi$  was at 5' to C ( $5'\Psi\text{C}/3'\text{AG}$  base pair step). The largest stabilization effect,  $-1.59$  kcal/mol, was observed at the  $5'\Psi\text{G}/3'\text{AC}$  base pair step. In summary, the  $\Psi$  stabilized the stacking energies at the  $\Psi\text{R}$  base pair steps, whereas at the  $\text{Y}\Psi$  base pair steps, no stabilization was observed. At the  $\text{R}\Psi$  and  $\text{Y}\Psi$  base pair steps, the stabilization effect depended on the R and Y base identity. The analysis of the decomposition of the stacking energies into intra- and interstrand terms showed that their contributions depended on the sequence (Supplementary Table S9 and Supplementary Fig. S11). For comparison, we evaluated the stacking energies of the base pair steps using empirical molecular mechanics force field ( $\Delta E_{\text{MM}}$ ). Our MM-based calculations were performed over the trajectories, thus assessing the dynamic nature of the systems. On the average, the variability of the stacking energy for the base pair steps along the RNA trajectory was approximately 2.2 kcal/mol. For four base pair steps ( $5'\Psi\text{G}/3'\text{AC}$ ,  $5'\Psi\text{U}/3'\text{AA}$ ,  $5'\text{G}\Psi/3'\text{CA}$  and  $5'\Psi\text{A}/3'\text{AU}$ ) where the calculations at the QM level ( $\Delta E_{\text{QM}}$ ) showed stabilization in the stacking interaction energies from the  $\Psi$  modification, the MM-based calculations ( $\Delta E_{\text{MM}}$ ) displayed the same stabilization order (Supplementary Fig. S13). For the remaining four base pair steps, although the results at the QM level

and from MM-based calculations showed some discrepancies, both the  $\Delta E_{MM}$  and  $\Delta E_{QM}$  revealed that, at these base pair steps, replacement of U with  $\Psi$  change the stacking energies less than at the base pair steps from the first group. When only the contribution of the  $\Psi$  to the intrastrand stacking energy between the bases in a single strand of RNA duplex was considered, we obtained the same ordering both for the QM-based and the MM-based  $\Delta E$ s (Supplementary Fig. S14). This result agrees with previous calculations that showed that QM gas phase stacking energies are consistent with MM-based calculations obtained with an AMBER force field<sup>38–40</sup>.

Based on the changes in the stacking energies upon replacing U with  $\Psi$  calculated for the eight unique base pair steps, it was encouraging to predict the impact of  $\Psi$  on the stacking energy within the trinucleotide motifs with  $\Psi$ -A at the central position when the U-A base pairs were replaced with  $\Psi$ -A. There are 16 unique trinucleotide motifs with central  $\Psi$ -A base pairs (Fig. 6b). The predictions revealed sequence-dependent variations of the  $\Psi$  contribution to the base stacking energies. The strongest stabilization from  $\Psi$  modification is predicted for the 5'G $\Psi$ G/3'CAC motif by both the QM-based and MM-based calculations (Supplementary Fig. S15). Both calculation methods predicted no stabilization for the stacking interactions after replacing U with  $\Psi$  for the 5'U $\Psi$ C/3'AAG and 5'C $\Psi$ C/3'GAG motifs. Our results showed that the effect of pseudouridine stabilization can differ by up to approximately 2.5 kcal/mol depending on the sequence context; however, for several sequence motifs, the stabilization effect is in the range of approximately  $-1.5$  to  $-0.5$  kcal/mol. The change in the stacking energies upon replacing U with  $\Psi$  to obtain an internal  $\Psi$ -A base pair at the G $\Psi$ C and C $\Psi$ G steps correlates with the experimental result<sup>31</sup> showing a more favorable thermodynamic effect for  $\Psi$  modification of duplex-C $\Psi$ G than that of duplex-G $\Psi$ C (Supplementary Fig. S16). To visualize the sequence dependent impact of  $\Psi$ -A base pairs on the interactions with the adjacent base pairs within the RNA duplex, we generated isodensity surfaces color-coded by electrostatic potential (Supplementary Fig. S17).

In summary, analysis of our calculations suggests that the intrinsic base stacking is the main factor that contributes to the sequence dependence of the thermodynamic stabilities of the RNA duplexes due to the introduction of  $\Psi$  modifications.

## Discussion

Currently, the proposed mechanisms for the stabilization effect of  $\Psi$  include structural water coordinated by HN1- $\Psi$ , improved stacking properties and the preference for the C3'-*endo* sugar pucker<sup>12</sup>. A structural water molecule coordinated with the HN1 of the  $\Psi$  was previously seen in the electron density maps of tRNA for residues  $\Psi$ 38 and  $\Psi$ 39 at the anticodon loop<sup>29</sup> and most recently, for example, in the crystal structures of tRNA (PDB ID: 1EHZ), rRNA and in U2 snRNA-branch point duplexes (PDB ID: 3CGP, 3CGR, 3CGS). Different scenarios of interactions mediated by the pseudouridine-bound water molecule were observed: a single water molecule bridging the HN1, OP2 atoms of the  $\Psi$  and OP2 of the preceding residue<sup>29</sup> (PDB ID: 1EHZ, 3CGR, 3CGS), an intra-residue bridge only ( $\Psi$ 2621 in PDB ID: 1S72) or two water molecules between the HN1 and OP2 atoms of the  $\Psi$  and the OP2 of the preceding residue (PDB ID: 3CGP).  $\Psi$  intra-residue water bridge was present in all of the above structures whereas inter-residue water bridge to the OP2 atom of the 5' residue was formed at the loop or stem/loop interface structural context. In our duplexes, both from the NMR structures and from the MD trajectories, the distance between the HN1 atom of the  $\Psi$  and the OP2 atom of the 5' residue was too large to make a single water bridge (6.3 Å and 6.5 Å for the NMR structures of the duplex-G $\Psi$ C and duplex-C $\Psi$ G, respectively and the peak at 6.75 Å from the distance distributions in the MD simulations of all the  $\Psi$ -modified duplexes, Supplemental Fig. S18). Distances smaller than 5.4 Å (as present in the PDB ID: 3CGR, 3CGS structures with single water molecules) were observed for 3–4% of the conformations with the Y $\Psi$  motif and 8% of the conformations with the R $\Psi$ . In our MD simulations of the  $\Psi$ -modified duplexes, we observed a water bridge formed by two water molecules. No long-lived hydrogen bonds involving the HN1- $\Psi$ 5 atom were observed. Distances similar to those in our duplexes were measured between the HN1 atom of  $\Psi$  and the OP2 atom of the preceding residue in other NMR structures with  $\Psi$  embedded in a double helix (6.3 Å and 6.4 Å in the duplex PDB ID: 5OR0 and 5.1–6.9 Å in the duplex PDB ID: 2KYE, depending on the model). The presence of long-lived water bridges between the OP2 atoms of consecutive residues and the HN1 atom of  $\Psi$ 32 was previously reported from a 500 ps MD simulation of the tRNA<sup>Asp</sup> hairpin performed with SPC/E water and NH<sub>4</sub><sup>+</sup> with the AMBER FF94 force field by Auffinger and Westhof<sup>41,42</sup>. The two bridging water molecules seen in our simulations may be characteristic of the TIP3P water model. It has been reported that the choice of water model has a visible impact on the predicted structure and structural dynamics of RNA<sup>37,43</sup>.

We conclude that internal single water bridges between the HN1 and OP2 atoms of  $\Psi$  likely characterize  $\Psi$  in all structural contexts. In loops and at stem-loop junctions, this water also could form a hydrogen bond with the OP2 atom of the preceding residue and make a long-lived water bridge. In the context of a regular helix, the water molecule forms an internal bridge within  $\Psi$ , but instead of directly binding to the OP2 atom of the preceding residue, more often is hydrogen bonded by an additional water molecule, making a string of two water molecules.

The effect of  $\Psi$  associated with its base stacking properties has so far not been systematically investigated, and our work emphasizes this aspect of the  $\Psi$  contribution to the RNA duplex stability. We calculated intrinsic stacking energies at different base pair steps using the QM method and structures derived from MD simulations. The choice of the representative geometry is discussed in the Section 6 of the Supplementary Methods. Our results showed that the changes in the base stacking energies from  $\Psi$  modification depended on the sequence; in most of the sequence contexts, the  $\Psi$  stabilized the stacking interactions but not always. Although these calculations cannot be directly used for the prediction of the relative thermodynamics of duplex formation<sup>44</sup>, the calculations suggest that the stacking interactions are the key factor in the sequence context dependence of the enhanced stability of the RNA with  $\Psi$ . The duplex formation could be conceptually considered to be a two-step process: (i) the transition of single strands from an unstacked to a stacked, ordered state and (ii) the association of two preorganized single-stranded helices to form a duplex<sup>45,46</sup>. The preference for the C3'-*endo* sugar pucker conformation of  $\Psi$  appears to play a more important role in the process (i). The stabilization of the C3'-*endo* sugar



pucker conformation in single stranded RNAs has been demonstrated experimentally<sup>12</sup>. In addition, stacking interactions and structural water likely contribute to the enhanced preorganization of single stranded RNA with pseudouridine. The contribution of  $\Psi$  to single strand preorganization has recently been supported by experimental data and MD studies<sup>47</sup>. In effect,  $\Psi$  modification resulted in preorganization of single strands with nucleotides in the C3'-*endo*, *anti* and axial conformations, thereby decreasing the entropic cost of duplex formation and increasing the thermodynamic stability of the duplex. The modifications that operate by the mechanism of reducing unfavorable entropy when going from a single-stranded to a double-stranded molecule include 2'-O-methyl-modified residues, LNA nucleosides<sup>48</sup> and 2-thiouridine<sup>46</sup>. In process (ii), i.e., helix association,  $\Psi$  does not play significant role because it does not change the hydrogen bonding, and the structural water molecule is not at the binding interface of the two strands. To test if the differences in the thermodynamics of the duplex formation of the  $\Psi$ -modified and unmodified duplexes can be obtained by computational predictions, we employed the MM-PBSA method. The relative binding free energies calculated for the  $\Psi$ -modified and unmodified duplexes were between 0.61 to  $-1.47$  kcal/mol and did not follow the experimental trend (Supplemental Table S10). However, the MM-PBSA calculations correspond to process (ii) and assume that the separated single strands retained the same A-form structure as assumed by the duplex and thus do not exactly mimic the experimental measurements. The single strands are presumably more flexible than double helices. Unfortunately, the calculations of the entropic and enthalpic contributions from the single strand preorganization to the duplex formation, i.e., process (i), are beyond the scope of MM-PBSA method with the single trajectory approach (see Supplementary Methods, Section 5). The results from the MM-PBSA calculations indirectly indicated that process (i) rather than (ii) is responsible for the sequence dependent stabilization effect of  $\Psi$ . The MM-PBSA method has been previously used for predicting the thermodynamics of duplex formation for RNA duplexes with modifications that alter hydrogen bond pairing<sup>48–50</sup>. From our QM and MM calculations (Supplemental Table S9),  $\Psi$  slightly destabilized the  $\Psi$ -A base pair interactions in agreement with previous work<sup>51,52</sup>. Once the duplex is formed, all nucleotides have the C3'-*endo* conformation, and we have not seen any additional ordering of the ribose moiety of  $\Psi$  compared with that of U.

In the future, it would be interesting to evaluate the effect of  $\Psi$  on single strand conformation in different sequence contexts with both experimental and computational methods.

## Methods

**RNA synthesis.** The RNA oligonucleotides (5'-UCAG $\Psi$ CAGU-3', 5'-ACUGACUGA-3', 5'-UCAC $\Psi$ GAGU-3', 5'-ACUCAGUGA-3') were synthesized on an Applied Biosystems DNA/RNA synthesizer, deprotected and purified according to previously published procedures<sup>31,53,54</sup>. For NMR study, each of the obtained duplexes was dissolved in a volume of 200  $\mu$ l of 90% H<sub>2</sub>O and 10% D<sub>2</sub>O solution and placed into a 3 mm NMR sample tube (Supplementary Methods, Section 1).

**NMR measurements.** The NMR spectra were acquired on a Bruker Avance III 700 MHz spectrometer, equipped with a QCI CryoProbe. The two-dimensional NOESY experiments involving non-exchangeable <sup>1</sup>H resonances were recorded at 25 °C. To facilitate peak assignments and obtain additional structural information 2D NOESY spectra at 30 and 35 °C were also acquired for duplex-G $\Psi$ C and -C $\Psi$ G, respectively, with the broad spectral width of  $\sim$ 7002 Hz. High resolution DQF-COSY spectra were acquired with narrow spectral width of 2520 Hz. Natural abundance <sup>1</sup>H-<sup>13</sup>C HSQC experiments (25 °C) were performed within the spectral width of 7002 Hz in the <sup>1</sup>H dimension and 28248 Hz in the <sup>13</sup>C dimension. To increase the resolution <sup>1</sup>H-<sup>13</sup>C HSQC spectra were acquired with narrowed spectral range of 4901 Hz in <sup>1</sup>H dimension and 10570 Hz in <sup>13</sup>C dimension and 64 scans per t1 increment. Proton detected <sup>1</sup>H-<sup>31</sup>P COSY spectra were recorded at 25 °C within 2102 Hz spectral range in the <sup>1</sup>H dimension and 1417 Hz in the <sup>31</sup>P dimension. To aid <sup>1</sup>H resonance assignment a two-dimensional <sup>1</sup>H-<sup>31</sup>P hetero TOCSY-NOESY experiments were recorded at 25 °C with a DIPSI-2 sequence<sup>55</sup> using 20 ms spin-lock period and 500 ms NOESY mixing time. The spectra were acquired with a spectral width of 10504 Hz in <sup>1</sup>H and 1701 Hz in the <sup>31</sup>P dimension. Exchangeable <sup>1</sup>H resonances were analyzed using the 2D NOESY spectra obtained at 15 °C with a 100 ms mixing time. The residual water peak for homonuclear experiments in D<sub>2</sub>O was suppressed using low-power presaturation whereas solvent suppression for samples in 90% H<sub>2</sub>O/10% D<sub>2</sub>O was achieved by applying a pulse sequence using excitation sculpting with gradients<sup>56</sup>. A sodium 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) was used as an internal chemical shift reference. Collected spectra were processed with TopSpin (Bruker, Inc.) and analyzed with Felix (Felix NMR, Inc.) software.

**Structural restraints.** NOE restraints for non-exchangeable protons were obtained from analysis of 2D NOESY spectra (D<sub>2</sub>O) with mixing times of 100 and 150 ms recorded at 25 °C using the Isolated Spin Pair Approximation (ISPA) approach. An average pyridine H5-H6 peak volumes were used as a reference with a distance of 2.45 Å. Error bounds for the restraints were set to  $-15\%$  (lower) and  $+30\%$  (upper) of the calculated interproton distance. The details for the identification of Watson-Crick base pairs and hydrogen bonding pattern between  $\Psi$  and A bases, constraints for endocyclic and backbone torsion angles  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\epsilon$  and  $\zeta$ , sugar pucker conformations and orientations of the nucleobases with respect to the sugar rings are summarized in Supplementary Methods, Section 2.

**Computational details.** All molecular mechanics (MM) calculations and molecular dynamic (MD) simulations were performed with AMBER 14 simulation package and AmberTools15<sup>57</sup>. All MD simulations were carried out with CPU and GPU version of *pmemd* module using ff99bsc0 $\chi_{OL3}$  force field<sup>58–60</sup> for regular RNA and parameter set for  $\Psi$  reported by Deb *et al.*<sup>61</sup>. The 9-bp RNA duplexes were built as regular RNA duplexes in

A-RNA conformation with Nucleic Acid Builder (NAB). The  $\Psi$  modification was incorporated at the 5<sup>th</sup> position of the initial structures by rearranging the atom positions of the uridine residue.

**Structure calculations by simulated annealing method.** The structure calculations for the 9-bp  $\Psi$ -modified RNA duplexes (duplex-G $\Psi$ C and -C $\Psi$ G) were carried out with the starting coordinates in the A-RNA conformation by the simulated annealing method<sup>62</sup>. During the simulated annealing calculations, restrained MD simulations were performed using NOE structural restraints collected from the analysis of NMR spectra (see Supplementary Materials and Methods, Section 2). A generalized Born implicit solvent model was used with a salt concentration of 150 mM NaCl.

The energy minimized structure of each RNA duplex was subjected to two cycles of a simulated annealing procedure following Chen *et al.*<sup>63</sup> (see Supplementary Methods, Section 3). The 100 structures generated after the first cycle of simulated annealing runs were sorted based on the violations criteria of 0.0 kcal.mol<sup>-1</sup> penalty in torsional restraints and minimum penalty (~1.67 kcal.mol<sup>-1</sup> for duplex-G $\Psi$ C and ~0.15 kcal.mol<sup>-1</sup> for duplex-C $\Psi$ G) for distance restraints. Among these sorted structures, 20 structures, without any restraint violations higher than 0.1 Å and with minimum number of restraint violations less than 0.1 Å, were selected for the next cycle of simulated annealing run. In the second cycle, these 20 structures were further refined to extract an ensemble of 10 final structures applying the similar violation criteria (~1.39 kcal.mol<sup>-1</sup> for duplex-G $\Psi$ C and ~0.17 kcal.mol<sup>-1</sup> for duplex-C $\Psi$ G) chosen in the first cycle.

**Molecular dynamics (md) simulations.** Each of the structures was solvated with TIP3P<sup>64</sup> water molecules in truncated octahedral boxes with minimal distance of 10 Å from the solute border and periodic boundary conditions were applied. The systems were neutralized with Na<sup>+</sup> ions and KCl excess salt was added to obtain the concentration of 1.0 M by using Joung and Cheatham ion parameters<sup>65</sup> to resemble conditions of UV melting experiments. The solvated systems were minimized and equilibrated. For each system, during the production run constant pressure Langevin dynamics was performed at 300 K for a total 500 ns with a 2 fs time step and trajectory files were written at each 10 ps. The pressure was regulated using Berendsen barostat<sup>66</sup> at reference pressure of 1 atm. Electrostatics was handled with PME<sup>67</sup> method with a direct space cutoff of 10.0 Å (see Supplementary Methods, Section 4).

**Analysis of MD trajectories.** The utilities available with the *cpptraj* module of AmberTools 15<sup>68</sup> were used for the trajectory analysis. The water occupancy maps around the average MD structure were calculated using the *grid* routine and visualized using UCSF-Chimera<sup>69</sup>. Hydrogen bond formations were considered if (i) the donor-acceptor distance was  $\leq 3.0$  Å, and (ii) the donor-hydrogen-acceptor angle was  $\geq 135.0^\circ$ . The *lie* utility in the *cpptraj* program was used for the calculation of the molecular mechanics (MM) force field-based base stacking energies as a summation of pairwise electrostatic and van der Waals interaction energies between the base atoms only. Cluster analysis was performed with *cpptraj* using the average-linkage clustering algorithm (for details see Supplementary Methods, Section 5). The order parameters ( $S^2$ ) were calculated for C1'-H1' vector using the *ired* utility in the *cpptraj* program (for details see Supplementary Methods, Section 5). Structural parameters were calculated using Curves+/Canal programs<sup>34</sup>.

**Ab initio base pair step stacking energies.** The *ab initio* base pair step stacking energies were calculated in the gas phase with a density functional theory (DFT) approach at the B97D/Def2TZVPP level of theory using the GAUSSIAN09 software suite. The intra- and interstrand base stacking energies were calculated following Svozil *et al.*<sup>70</sup> (for details see Supplementary Methods, Section 6). The QM base pair step stacking energies were calculated by subtracting the hydrogen bonding interaction energies from the interaction energies of the complete four-base stacked systems (Supplementary Fig. S19) according to the equation:  $\Delta[E]_{AC,BD} = [E]_{ADBC} - ([E]_{AD} + [E]_{BC})$ , where  $[E]_{ADBC}$  represents the interaction energy of the complete four-base stacked system, and  $[E]_{AD}$  and  $[E]_{BC}$  represent hydrogen bonding interaction energies in the Watson-Crick base pairs. The interaction energies were calculated by subtracting the monomer total energies from the corresponding dimer/tetramer energies corrected for basis set superposition error (BSSE) via the counterpoise method implemented in the GAUSSIAN09 software suite. The intrastrand ( $[E]_{AB}$  and  $[E]_{CD}$ ) and interstrand ( $[E]_{AC}$  and  $[E]_{BD}$ ) interaction energies were also reported in Supplementary Table S9. Stacking energies between bases in a single strand of RNA duplex at given base step represented by the intrastrand energies are shown on Supplementary Fig. S14. The geometries of the base pair steps were derived from the average coordinates of the most populated cluster obtained from the cluster analysis (Supplementary Methods, Section 6). The four-base stacked geometries for QM calculations were prepared by superposition of previously optimized monomers of individual bases on the scaffold of the average structures.

## Data availability

Coordinates have been deposited in the Protein Data Bank (PDB IDs: 6I1W and 6I1V for the duplex-G $\Psi$ C and duplex-C $\Psi$ G, respectively) and the NMR data in the Biological Magnetic Resonance Data Bank (BMRB codes: 34324 and 34323 for the duplex-G $\Psi$ C and duplex-C $\Psi$ G, respectively).

Received: 14 February 2019; Accepted: 3 October 2019;

Published online: 07 November 2019

## References

- Davis, F. F. & Allen, F. W. Ribonucleic acids from yeast which contain a fifth nucleotide. *J. Biol. Chem.* **227**, 907–915 (1957).
- Gray, M. C. M. W. Pseudouridine in RNA: What, Where, How, and Why. I. *UBMB Life (International Union Biochem. Mol. Biol. Life)* **49**, 341–351 (2000).
- Carlile, T. M. *et al.* Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **515**, 143–146 (2014).
- Schwartz, S. *et al.* Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* **159**, 148–162 (2014).
- Li, X. *et al.* Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat. Chem. Biol.* **11**, 592–597 (2015).
- Lei, Z. & Yi, C. A radiolabeling-free, qPCR-Based method for locus-specific pseudouridine detection. *Angew. Chemie - Int. Ed.* **56**, 14878–14882 (2017).
- Karijolic, J., Yi, C. & Yu, Y. T. Transcriptome-wide dynamics of RNA pseudouridylation. *Nature Reviews Molecular Cell Biology* **16**, 581–585 (2015).
- Zhao, Y., Dunker, W., Yu, Y.-T. & Karijolic, J. The role of noncoding RNA pseudouridylation in nuclear gene expression events. *Front. Bioeng. Biotechnol.* **6**, 8 (2018).
- Ge, J. & Yu, Y. T. RNA pseudouridylation: New insights into an old modification. *Trends in Biochemical Sciences* **38**, 210–218 (2013).
- Spenkuch, F., Motorin, Y. & Helm, M. Pseudouridine: Still mysterious, but never a fake (uridine)! *RNA Biology* **11**, 1540–1554 (2014).
- Li, X., Ma, S. & Yi, C. Pseudouridine: the fifth RNA nucleotide with renewed interests. *Current Opinion in Chemical Biology* **33**, 108–116 (2016).
- Davis, D. R. Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Res.* **23**, 5020–5026 (1995).
- Durant, P. C. & Davis, D. R. Stabilization of the anticodon stem-loop of tRNA(Lys,3) by an A+-C base-pair and by pseudouridine. *J. Mol. Biol.* **285**, 115–131 (1999).
- Yarian, C. S. *et al.* Structural and functional roles of the N1- and N3-protons of  $\Psi$  at tRNA's position 39. *Nucleic Acids Res.* **27**, 3543–3549 (1999).
- Denmon, A. P., Wang, J. & Nikonowicz, E. P. Conformation effects of base modification on the anticodon stem-loop of *Bacillus subtilis* tRNA<sup>Tyr</sup>. *J. Mol. Biol.* **412**, 285–303 (2011).
- Cabello-Villegas, J. & Nikonowicz, E. P. Solution structure of psi32-modified anticodon stem-loop of *Escherichia coli* tRNA<sup>Phe</sup>. *Nucleic Acids Res.* **33**, 6961–71 (2005).
- Davis, D. R., Veltri, C. A. & Nielsen, L. An RNA model system for investigation of pseudouridine stabilization of the codon-anticodon interaction in tRNA<sup>Lys</sup>, tRNA<sup>His</sup> and tRNA<sup>Tyr</sup>. *J. Biomol. Struct. Dyn.* **15**, 1121–1132 (1998).
- Jiang, J., Kharel, D. N. & Chow, C. S. Modulation of conformational changes in helix 69 mutants by pseudouridine modifications. *Biophys. Chem.* **200–201**, 48–55 (2015).
- Jiang, J., Seo, H. & Chow, C. S. Post-transcriptional modifications modulate rRNA structure and ligand interactions. *Acc. Chem. Res.* **49**, 893–901 (2016).
- Kulik, M. *et al.* Helix 69 of *Escherichia coli* 23S ribosomal RNA as a peptide nucleic acid target. *Biochimie* **138**, 32–42 (2017).
- Sumita, M. Effects of nucleotide substitution and modification on the stability and structure of helix 69 from 28S rRNA. *RNA* **11**, 1420–1429 (2005).
- Sumita, M., Jiang, J., Santalucia, J. & Chow, C. S. Comparison of solution conformations and stabilities of modified helix 69 rRNA analogs from bacteria and human. *Biopolymers* **97**, 94–106 (2011).
- Kim, N. K., Theimer, C. A., Mitchell, J. R., Collins, K. & Feigon, J. Effect of pseudouridylation on the structure and activity of the catalytically essential P6.1 hairpin in human telomerase RNA. *Nucleic Acids Res.* **38**, 6746–6756 (2010).
- Newby, M. I. & Greenbaum, N. L. Investigation of Overhauser effects between pseudouridine and water protons in RNA helices. *Proc. Natl. Acad. Sci.* **99**, 12697–12702 (2002).
- Lin, Y. & Kielkopf, C. L. X-ray structures of U2 snRNA - Branchpoint duplexes containing conserved pseudouridines. *Biochemistry* **47**, 5503–5514 (2008).
- deLorimier, E. *et al.* Modifications to toxic CUG RNAs induce structural stability, rescue mis-splicing in a myotonic dystrophy cell model and reduce toxicity in a myotonic dystrophy zebrafish model. *Nucleic Acids Res.* **42**, 12768–12778 (2014).
- Tomita, K., Ueda, T. & Watanabe, K. The presence of pseudouridine in the anticodon alters the genetic code: A possible mechanism for assignment of the AAA lysine codon as asparagine in echinoderm mitochondria. *Nucleic Acids Res.* **27**, 1683–1689 (1999).
- Fernández, I. S. *et al.* Unusual base pairing during the decoding of a stop codon by the ribosome. *Nature* **500**, 107–110 (2013).
- Arnez, J. G. & Steitz, T. A. Crystal structure of unmodified tRNA<sup>Gln</sup> complexed with glutamyl-tRNA synthetase and ATP suggests a possible role for pseudo-uridines in stabilization of RNA structure. *Biochemistry* **33**, 7560–7567 (1994).
- Newby, M. I. & Greenbaum, N. L. A conserved pseudouridine modification in eukaryotic U2 snRNA induces a change in branch-site architecture. *RNA* **7**, 833–845 (2001).
- Kierzek, E. *et al.* The contribution of pseudouridine to stabilities and structure of RNAs. *Nucleic Acids Res.* **42**, 3492–3501 (2014).
- Hudson, G. A., Bloomingdale, R. J. & Znosko, B. M. Thermodynamic contribution and nearest-neighbor parameters of pseudouridine-adenosine base pairs in oligoribonucleotides. *RNA* **19**, 1474–82 (2013).
- Jiang, J., Aduri, R., Chow, C. S. & SantaLucia John, J. Structure modulation of helix 69 from *Escherichia coli* 23S ribosomal RNA by pseudouridylations. *Nucleic Acids Res.* **42**, 3971–3981 (2014).
- Lavery, R., Moakher, M., Maddocks, J. H., Petkeviciute, D. & Zakrzewska, K. Conformational analysis of nucleic acids revisited: Curves. *Nucleic Acids Res.* **37**, 5917–5929 (2009).
- Saenger, W. Principles of Nucleic Acid Structure. Springer-Verlag: Berlin, Heidelberg, NY (1984).
- Kailasam, S., Bhattacharyya, D. & Bansal, M. Sequence dependent variations in RNA duplex are related to non-canonical hydrogen bond interactions in dinucleotide steps. *BMC Res. Notes* **7**, 83 (2014).
- Kührová, P., Otyepka, M., Šponer, J. & Banáš, P. Are waters around RNA more than just a solvent? - An insight from molecular dynamics simulations. *J. Chem. Theory Comput.* **10**, 401–411 (2014).
- Šponer, J. *et al.* Nature and magnitude of aromatic base stacking in DNA and RNA: Quantum chemistry, molecular mechanics, and experiment. *Biopolymers* **99**, 978–988 (2013).
- Hobza, P., Kabelá, M., Šponer, J., Mejzlik, P. & Vondrášek, J. Performance of empirical potentials (AMBER, CFF95, CVFF, CHARMM, OPLS, POLTEV), semiempirical quantum chemical methods (AM1, MNDO/M, PM3), and ab initio Hartree-Fock method for interaction of DNA bases: Comparison with nonempirical beyond Hartree-Fock res. *J. Comput. Chem.* **18**, 1136–1150 (1997).
- Sponer, J., Gabb, H. A., Leszczynski, J. & Hobza, P. Base-base and deoxyribose-base stacking interactions in B-DNA and Z-DNA: a quantum-chemical study. *Biophys. J.* **73**, 76–87 (1997).
- Auffinger, P. & Westhof, E. RNA hydration: Three nanoseconds of multiple molecular dynamics simulations of the solvated tRNA(Asp) anticodon hairpin. *J. Mol. Biol.* **269**, 326–341 (1997).
- Auffinger, P. & Westhof, E. Effects of Pseudouridylation on tRNA Hydration and Dynamics: a Theoretical Approach. In *Modification and Editing of RNA* (eds Grosjean, H. & Benne, R.) 103–112 (ASM Press, 1998).

43. Bešševová, I. *et al.* Simulations of A-RNA duplexes. The effect of sequence, solute force field, water model, and salt concentration. *J. Phys. Chem. B* **116**, 9899–9916 (2012).
44. Morgado, C. A., Svozil, D., Turner, D. H. & Šponer, J. Understanding the role of base stacking in nucleic acids. MD and QM analysis of tandem GA base pairs in RNA duplexes. *Phys. Chem. Chem. Phys.* **14**, 12580–12591 (2012).
45. Holbrook, J. A., Capp, M. W., Saecker, R. M. & Record, M. T. Enthalpy and heat capacity changes for formation of an oligomeric DNA duplex: Interpretation in terms of coupled processes of formation and association of single-stranded helices. *Biochemistry* **38**, 8409–8422 (1999).
46. Larsen, A. T., Fahrenbach, A. C., Sheng, J., Pian, J. & Szostak, J. W. Thermodynamic insights into 2-thiouridine-enhanced RNA hybridization. *Nucleic Acids Res.* **43**, 7675–7687 (2015).
47. deLorimier, E. *et al.* Pseudouridine modification inhibits Muscblind-like 1 (MBNL1) binding to CCUG repeats and minimally structured RNA through reduced RNA flexibility. *J. Biol. Chem.* **292**, 4350–4357 (2017).
48. Kierzek, E. *et al.* Contributions of stacking, preorganization, and hydrogen bonding to the thermodynamic stability of duplexes between RNA and 2'-O-methyl RNA with locked nucleic acids. *Biochemistry* **48**, 4377–4387 (2009).
49. Yildirim, I., Kierzek, E., Kierzek, R. & Schatz, G. C. Interplay of LNA and 2'-O-methyl RNA in the structure and thermodynamics of RNA hybrid systems: A molecular dynamics study using the revised AMBER force field and comparison with experimental results. *J. Phys. Chem. B* **118**, 14177–14187 (2014).
50. Špačková, N. & Réblová, K. Role of inosine–uracil base pairs in the canonical RNA duplexes. *Genes* **9**, 324 (2018).
51. Chawla, M., Oliva, R., Bujnicki, J. M. & Cavallo, L. An atlas of RNA base pairs involving modified nucleobases with optimal geometries and accurate energies. *Nucleic Acids Res.* **43**, 6714–6729 (2015).
52. Seelam, P. P., Sharma, P. & Mitra, A. Structural landscape of base pairs containing post-transcriptional modifications in RNA. *RNA* **23**, 847–859 (2017).
53. Xia, T. *et al.* Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson–Crick base pairs. *Biochemistry* **37**, 14719–14735 (1998).
54. Kierzek, E. & Kierzek, R. The synthesis of oligoribonucleotides containing N6-alkyladenosines and 2-methylthio-N6-alkyladenosines via post-synthetic modification of precursor oligomers. *Nucleic Acids Res.* **31**, 4461–4471 (2003).
55. Shaka, A. J., Lee, C. J. & Pines, A. Iterative schemes for bilinear operators; application to spin decoupling. *J. Magn. Reson.* **77**, 274–293 (1988).
56. Hwang, T. L. & Shaka, A. J. Water Suppression that works. Excitation sculpting using arbitrary wave-forms and pulsed-field gradients. *J. Magn. Reson. - Ser. A* **112**, 275–279 (1995).
57. Case, D. A. *et al.* Amber 2015. University of California, San Francisco (2015).
58. Cornell, W. D. *et al.* A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179–5197 (1995).
59. Pérez, A. *et al.* Refinement of the AMBER force field for nucleic acids: Improving the description of  $\alpha/\gamma$  conformers. *Biophys. J.* **92**, 3817–3829 (2007).
60. Zgarbová, M. *et al.* 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 (2011).
61. Deb, I., Pal, R., Sarzynska, J. & Lahiri, A. Reparameterizations of the  $\chi$  torsion and Lennard–Jones  $\sigma$  parameters improve the conformational characteristics of modified uridines. *J. Comput. Chem.* **37**, 1576–1588 (2016).
62. Schmitz, M. & Steger, G. Description of RNA folding by 'simulated annealing'. *J. Mol. Biol.* **255**, 254–266 (1996).
63. Chen, J. L., Kennedy, S. D. & Turner, D. H. Structural features of a 3' splice site in influenza A. *Biochemistry* **54**, 3269–3285 (2015).
64. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
65. Joung, I. S. & Cheatham, T. E. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **112**, 9020–9041 (2008).
66. Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., Dinola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
67. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
68. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9**, 3084–3095 (2013).
69. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
70. Svozil, D., Hobza, P. & Šponer, J. Comparison of intrinsic stacking energies of ten unique dinucleotide steps in A-RNA and B-DNA duplexes. Can we determine correct order of stability by quantum-chemical calculations? *J. Phys. Chem. B* **114**, 1191–1203 (2010).

## Acknowledgements

This work was supported by the International Centre for Genetic Engineering and Biotechnology Smart Fellowship Programme [ICGEB ID#: S/IND 15-05] to I.D., National Science Centre, Poland, grants: [UMO-2013/08/A/ST5/00295], [UMO-2017/25/B/NZ1/02269] to R.K. and [UMO-2017/25/B/ST5/00971] to Z.G. This publication was also supported by the Polish Ministry of Science and Higher Education, under the KNOW program. The calculations were performed at Poznań Supercomputing and Networking Center. The authors would like to thank the European Center for Bioinformatics and Genomics (ECBiG), Poznań, Poland, for computer lab support. We are also grateful to Dr. D. Baranowski for his support during the NMR measurements.

## Author contributions

L.P., M.M., Z.G. and R.K. designed and performed the experiments. I.D. and J.S. performed calculations. I.D., L.P., J.S., A.L., and Z.G. interpreted data and wrote the manuscript. All authors approved the final version of the paper.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-019-52637-0>.

**Correspondence** and requests for materials should be addressed to J.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019