Supporting Information

Urea destabilizes RNA by forming stacking interaction and multiple hydrogen bonds with nucleic acid bases

U. Deva Priyakumar[†], Changbong Hyeon[‡], D. Thirumalai[1§], and Alexander D. MacKerell Jr. [*†]

[†] *Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, Maryland 21201*

[‡] *Department of Chemistry, Chung-Ang University, Seoul 156-756, Republic of Korea*

[§] *Biophysics Program, Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742*

Systems: We used two RNA constructs in our simulations. The first is a 22 nucleotide (nt) hairpin, P5GA (PDB ID: 1EOR) whose structure has been determined using NMR. The other is an oligonucleotide consisting of two complimentary strands (duplex RNA) extracted from the full RNA structure (PDB ID: 1JP0). The smaller duplex RNA with only 8nts was chosen because its conformational space can be exhaustively sampled. The results obtained using both the systems further establish that our conclusions are robust.

Urea force fields: We developed a new urea force field for use in the P5GA (hairpin) simulations (see below). For the RNA duplex the parameters for urea were taken from Ref. 1. The excellent agreement in the proposed destablization mechanism between the

different force fields lends further credence to the overall conclusions reached in this work.

Urea Force Field for CHARMM: We developed a new urea force field for simulations of the P5GA hairpin in aqueous urea solutions. Parameters for urea were optimized following the standard CHARMM protocols used previously in the simulations nucleic acids as well as other biological molecules[2,3]. Calculations were performed with the program CHARMM[4] or with the quantum mechanical program Gaussian03[5]. Briefly, internal parameters were optimized to reproduce survey data of the Cambridge Structural Database[6] or QM data (Table 1) and vibrational spectra obtained at the MP2/6-31G* level and scaled by 0.89 (Table 2)[7]. Charges and Lennard-Jones parameters were optimized to reproduce interactions of urea with water (Figure 2) obtained at the HF/6-31G* level on the MP2/6-31G* optimized gas phase geometry with the QM interaction energies scaled by 1.16 (Table 3). In addition, the free energy of aqueous solvation of urea, as calculated using the method of Deng and Roux[8], was used as target data for the optimization of the non-bonded terms. Individual simulations for the FE calculations included 5 ps of MD equilibration followed by 50 ps of sampling in a box of 125 TIP3P waters and 5 ps of equilibration and 20 ps of sampling in a box of 250 TIP3P waters.

Overall, the optimized empirical model is in good agreement with the target data. The level of agreement of the geometries is not optimal (Table 1) due to direct transfer of parameters from the CHARMM22 protein force field.[9] However, it was possible to obtain excellent agreement for the vibrational spectra including for the low frequency torsion and wag of the NH_2 moieties, which can undergo large distortions during MD

simulations. Minimum interaction energies and distances of urea with water are in good agreement with the target QM data (Table 3). The non-bond parameters were adjusted to reproduce the most favorable interactions with the C=O and NH2 moieties with the interaction distances approximately 0.2 Å less than the target values as required to obtain the proper condensed phase densities. The final free energies of solvation were -13.3 and -13.1 kcal/mol from the perturbations in the boxes of 125 and 250 waters, respectively, are in good agreement with the target experimental value of -13.8 kcal/mol.[10] In addition, the dipole moment of the final empirical model was 4.88, overestimating the MP2/6-31G* value of 4.30, an overestimation required for the non-polarizable additive force field used in this study.

The validity of the current urea force field is further established by excellent comparison between the calculated and measured heats of sublimation for base crystals. Moreover, it has been shown that the method of using experimental data and quantum mechanical calculation to obtain force field parameters is accurate even in describing the configurations of benzene dimer.[11] In addition, we also tested the robustness of the urea-induced denaturation mechanism of RNA by performing simulation for the RNA duplex using an entirely different force field.

Simulation Details: Molecular dynamics (MD) simulations for P5GA were performed using the CHARMM program[4] employing the CHARMM27 nucleic acid force field[2] and the CHARMM modified TIP3P water model.[5] The P5GA hairpin was overlaid with a pre-equilibrated truncated octahedron of water or aqueous urea of varying concentrations. The solvent box was extended at least 9 Å beyond the non-hydrogen atoms of the RNA from all the sides. Water/urea molecules were removed if one or more of the solvent

molecule's non-hydrogen atoms lie within 1.8 Å of the non-hydrogen atoms of the RNA. We added 21 sodium atoms to the resultant systems at random positions at a minimum of 3 Å from the RNA non-hydrogen atoms to maintain electrical neutrality. In all the subsequent minimizations and MD simulations, periodic boundary conditions were employed using the CRYSTAL[12] module in CHARMM. Energy minimizations were performed using the adopted basis Newton-Raphson (ABNR) method for 500 steps with mass-weighted harmonic restraints of 5.0 kcal/mol/Å on the non-hydrogen atoms of the RNA. After the initial minimization, each of these systems was subjected with a 20 ps MD simulation in the NPT ensemble followed by a 100 ps MD simulation in the NVT ensemble keeping the harmonic restraints. The short NPT simulation was carried out to allow the solvent molecules to move near the oligonucleotide and fill voids created by deleting the solvent molecules that overlapped with the RNA with the subsequent NVT simulation performed to allow full relaxation of the solvent, including the ions, around the RNA. Electrostatic interactions were treated using the particle mesh Ewald method.[13,14] Lennard-Jones interactions were truncated at 12 Å with a force switch smoothing function from 10 to 12 Å and the non-bond atom lists were updated heuristically. Production simulations were carried out at 278 K in accordance with the experimental conditions for 20 ns in the NPT ensemble with the Leapfrog integrator without any restraints. All the simulations employed an integration time step of 2 fs and the SHAKE algorithm[15] to constrain all covalent bonds involving hydrogen atoms. The NPT ensemble was achieved using Hoover chains[16] for temperature control and the Langevin piston method[17] was used to maintain a pressure of 1 ATM with a piston mass

of 600 amu and the piston collision frequency set to 0. During the production run, coordinates were saved every 2 ps for analysis.

For the duplex RNA we used the NAMD molecular dynamics package with the CHARMM27 nucleic acid force field. The ds-RNA, with each strand consisting of 4nts, was solvated in a $(42Å)^3$ cubic box. The excess charges on the phosphate groups were neutralized with 11 Na^+ and 3 Cl^- ions. To simulate the ds-RNA in aqueous urea solution we replaced water with 77, 154, and 230 urea molecules that results in 2M, 4M, and 6M urea solution. As in the simulations involving the P5GA hairpin we used periodic boundary conditions, and electrostatic interactions were treated using the particle mesh Ewald method. As in the study of P5Ga the simulations employed an integration time step of 2 *fs* and SHAKE algorithm to constrain all covalent bonds all the covalent bonds involving hydrogen atoms. The NPT ensemble was achieved using the Langevin thermostat with a friction coefficient of 5 ps^{-1} on non-hydrogen atoms. Energy minimization that removes the instability of the entire system were performed for ds-RNA, urea, ions, and water respectively. The entire system was heated from 0 to 300 K every 1.2 nsec, after which the trajectories were generated for 80 ns in aqueous solution. The length of the trajectories at 2 M urea is ~ 260ns while at 4 M and 6M they were ~ 280 ns. The coordinates of the entire system were save every ps for analyses of the data.

References:

(1) Caflisch, A.; Karplus, K. *Structure Fold. Des.* **1999, *7,*** 477-488
(2) Foloppe, N.; MacKerell, A. D., Jr. *J. Comp. Chem.* **2000**, *21*, 86-104.
(3) Mackerell, A. D. *Journal of Computational Chemistry* **2004**, *25*, 1584-1604.
(4) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187-217.

(5) Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A.; Revision C.02 ed.; Gaussian, Inc.: Wallingford CT, 2004.

(6) Allen, F. H.; Bellard, S.; Brice, M. D.; Cartwright, B. A.; Doubleday, A.; Higgs, H.; Hummelink, T.; Hummelink-Peters, B. G.; Kennard, O.; Motherwell, W. D. S.; Rodgers, J. R.; Watson, D. G. *Acta Cryst* **1979**, *B35*, 2331-2339.

(7) Scott, A. P.; Radom, L. *J. Phys. Chem.* **1996**, *100*, 16502-16513.

(8) Deng, Y. Q.; Roux, B. *Journal of Physical Chemistry B* **2004**, *108*, 16567-16576.

(9) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack Jr., R. L.; Evanseck, J.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, I., W. E.; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586-3616.

(10) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *Journal of Chemical Theory and Computation* **2005**, *1*, 1133-1152.

(11) Sherrill, CD; Sumpter, BG; Sinnokrot, MO, et al. *J. Comp. Chem.* 2009, 30, 2187-2193

(12) Field, M. J.; Karplus, M.; Harvard University: Cambridge, MA, 1992.

(13) Darden, T.; Perera, L.; Li, L. P.; Pedersen, L. *Structure with Folding & Design* **1999**, *7*, R55-R60.

(14) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T. A.; Lee, H.; Pedersen, L. G. *J Chem Phys* **1995**, *103*, 8577-8593.

(15) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comp. Physics* **1977**, *23*, 327-341.

(16) Hoover, W. G. *Phy. Rev.* **1985**, *A 31*, 1695-1697.

(17) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, R. W. *J. Chem. Phys.* **1995**, *103*, 4613-4621.

SI Table 1: Internal geometries of urea

Term	CSD	QM	Charmm
C=O	1.26±0.02	1.233(1.267)	1.225
N-C	1.33±0.03	1.381(1.3631)	1.329
N-C=O	121.0±1.9	122.6(121.2763)	124.0
N-C-N	117.9±2.0	114.8(122.0259)	112.0
N-C-N-H1		-168.3(-166.1)	-176.9
N-C-N-H2		-25.0(-13.7)	-9.4

QM data from MP2/aug-cc-pVDZ optimized structure. Values in parentheses are values from an optimization of urea in the presence of 6 water molecules with the final geometry shown in Figure 1. Bond lengths are in Å and valance and dihedral angles are in degrees.

SI Table 2: Vibrational spectra of urea from MP2/6-31G* and final empirical models. MP2/6-31G* frequencies scaled by 0.89.

	Scaled MP2/6-31G*					CHARMM				
Mode	Frequency	Assignment	%	Assignment	%	Frequency	Assignment	%	Assignment	%
1	151.3	wNH2	51	tNH2	47	151.1	wNH2	70	tNH2	29
2	382.6	wNH2	81	tNH2	20	352.0	wNH2	78	wCCO	19
3	384.7	tNH2	52	wNH2	46	432.5	tNH2	105		
4	451.5	dCCO	78			450.8	dCCO	87		
5	505.4	tNH2	83			477.2	tNH2	68	wNH2	30
6	532.2	rCCO	83			583.7	rCCO	80		
7	715.4	wCCO	108			687.6	wCCO	88	wNH2	20
8	906.1	sN-C	86			882.2	sN-C	71		
9	961.1	rNH2	73	sN-C	25	1069.1	rNH2	79	sN-C	18
10	1103.1	rNH2	74	sC=O	17	1149.8	rNH2	76	sC=O	22
11	1335.8	sN-C	52	rNH2	17	1422.1	dNH2	46	sN-C	26
		rCCO	16							
12	1529.6	dNH2	85			1580.8	dNH2	89		
13	1537.3	dNH2	91			1643.9	dNH2	49	sN-C	43
14	1674.8	sC=O	70			1779.1	sC=O	56	sN-C	20
15	3393.1	sNH	100			3401.4	sNH	99		
16	3399.5	sNH	100			3415.3	sNH	99		
17	3525.3	sNHas	100			3532.3	sNHas	99		
18	3527.8	sNHas	100			3540.9	sNHas	100		

Frequencies in cm^{-1}. Assignments represent the contribution of internal degrees of freedom to the potential energy distribution presented in percent contribution to each normal mode where s stands for bond stretching, d for valance angle deformations, r for rocking, t for torsions and w for the wagging mode.

SI Table 3: Minimum interaction energies and geometries between water and urea from QM and empirical model. Interaction orientations shown in SI Figure 2 and only the hydrogen bond distances were optimized.

Interaction energy	QM1	QM2	EMP	Diff vs. QM2
1)C=O..HOH,linear	-6.73	-6.93	-7.63	-0.70
2)C=O..HOH,120deg.	-8.61	-9.31	-9.27	0.03
3)N-H..OHH,Oside	-4.36	-4.45	-4.18	0.26
4)N-H..OHH,nonOside	-7.05	-7.09	-7.03	0.07
Average Difference (QM2)				-0.08
RMS_Difference (QM2)				0.37
Distances	QM1	QM2	EMP	
1)C=O..HOH,linear	2.02	2.01	1.74	-0.27
2)C=O..HOH,120deg.	1.96	2.01	1.72	-0.29
3)N-H..OHH,Oside	2.10	2.08	1.92	-0.16
4)N-H..OHH,nonOside	2.10	2.09	1.92	-0.17

Energies in kcal/mol and distances in Å. QM interaction energies scaled by 1.16.

QM1: non-planar structure, QM2: enforced planarity. QM interaction energies are scaled by 1.16

Orientation 2) The C=O..H angle is fixed at 120˚.

SI Table 4: Energies of interaction[a] between the base pairs in the stem region of the P5GA RNA hairpin (see Fig. 1A in the main text for the secondary structure map of P5GA).

Base Pair	0M	6M	8M
G2C21	-13.6 ± 2.5	-7.7 ± 2.4	-14.2 ± 2.2
C3G20	-17.7 ± 1.4	-16.0 ± 1.2	-10.4 ± 0.5
G4C19	-21.4 ± 0.1	-16.9 ± 1.8	-21.8 ± 0.1
A5G18	-9.8 ± 0.9	-1.0 ± 0.5	-11.7 ± 0.1
A6G17	-10.2 ± 0.5	-0.2 ± 0.0	-2.7 ± 0.3
G7U16	-5.0 ± 0.9	-0.9 ± 0.3	-1.8 ± 0.5
U8A15	-6.9 ± 1.3	-9.5 ± 0.8	-0.7 ± 0.3
C9G14	-18.1 ± 1.4	-21.0 ± 0.5	-12.5 ± 2.4
Ave[b]	-12.9	-9.2	-9.5

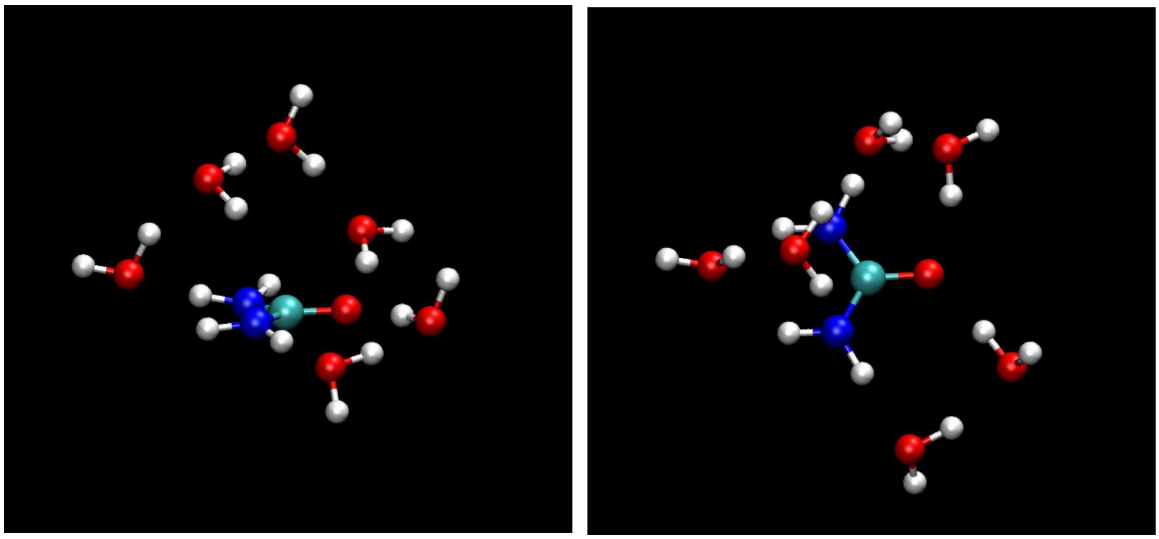
(a) All values are given in kcal/mol and errors are the standard errors. (b) The average of all the average base pair interaction energies in a given system.

SI Table 5: Decomposition of the dehydration ratio[a] (λ_{DR}) at various structural elements of the P5GA hairpin as a function of urea concentration.

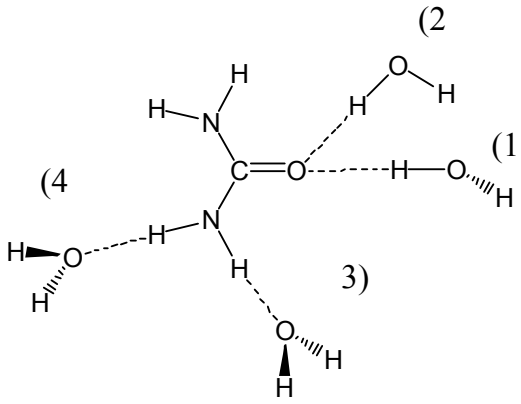
	1M[b]	**6M**	**8M**
RNA	1.94	1.11	1.26
RNA backbone	1.18	1.28	1.27
RNA bases	2.54	0.63	0.85
Stem	1.52	1.07	1.22
Stem backbone	1.16	1.29	1.28
Stem bases	1.67	0.48	0.71
Major groove	1.35	0.22	0.27
Minor groove	1.32	0.50	0.83
Pyrimidines	1.37	1.21	0.52
Purines	1.31	0.15	0.76
Loop	3.44	1.06	1.28
Loop backbone	0.98	1.12	1.16
Loop bases	6.76	0.85	1.21

(a) λ_{DR} is a quantitative measure of the decrease in the number of water molecules to the increase in the number of urea molecules in the first solvation shell of the RNA as a function of urea concentration.

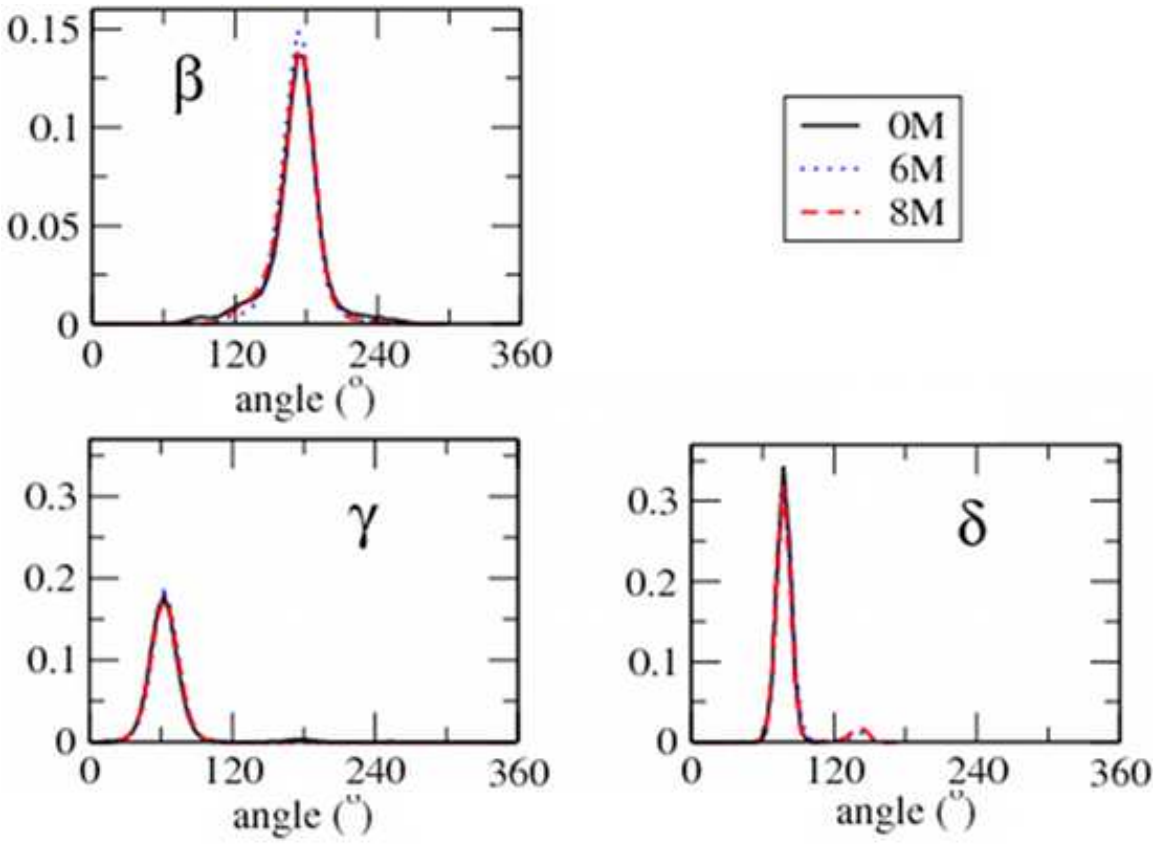
(b) Urea concentration.



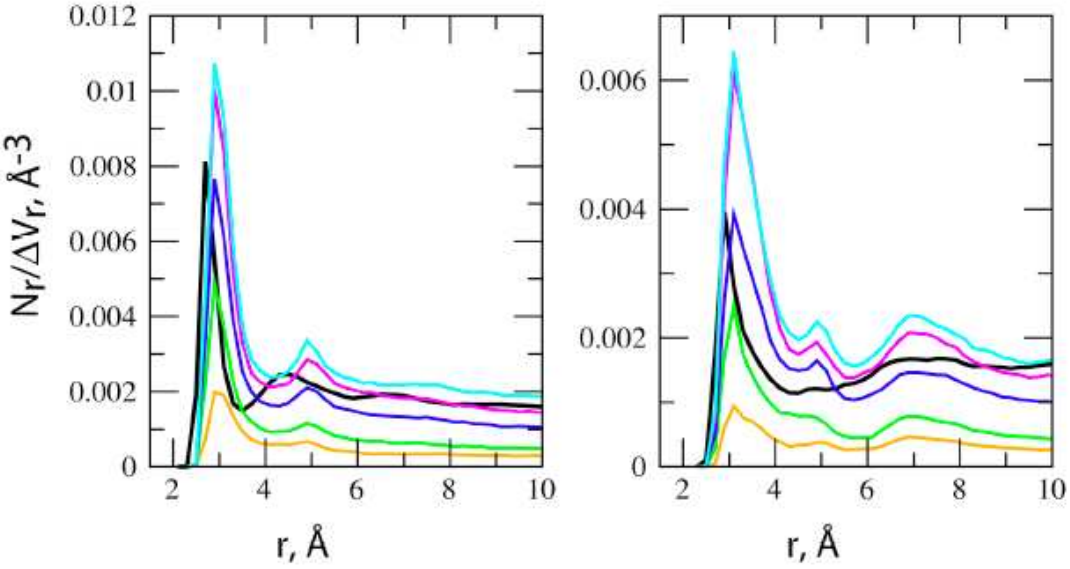
SI Figure 1: Images of a QM optimized structure of urea with 6 water molecules. The resulting configuration is used for analysis of the urea internal geometry.



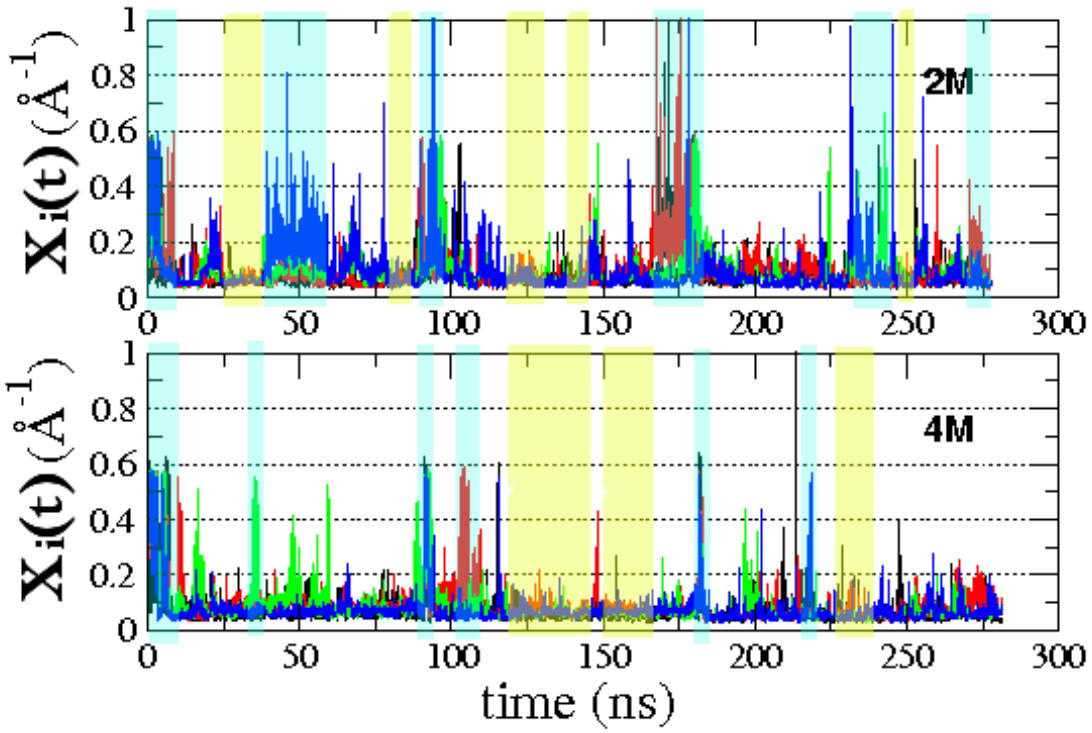
SI Figure 2: Diagram of interaction orientations between urea and water used in SI Table 3.



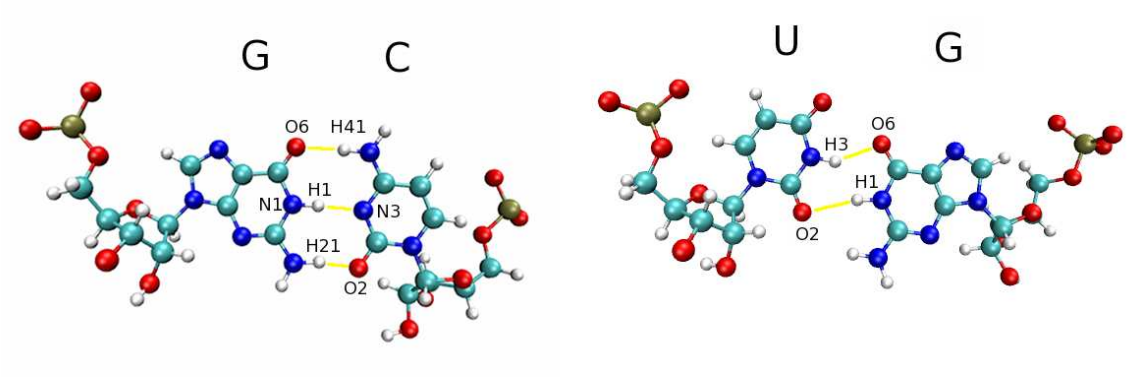
SI Figure 3: The probability distribution functions of selected phosphodiester-backbone dihedral angles for the P5GA hairpin (see Fig. 2 in the main text for definitions) at [C]=0, 6, 8M.



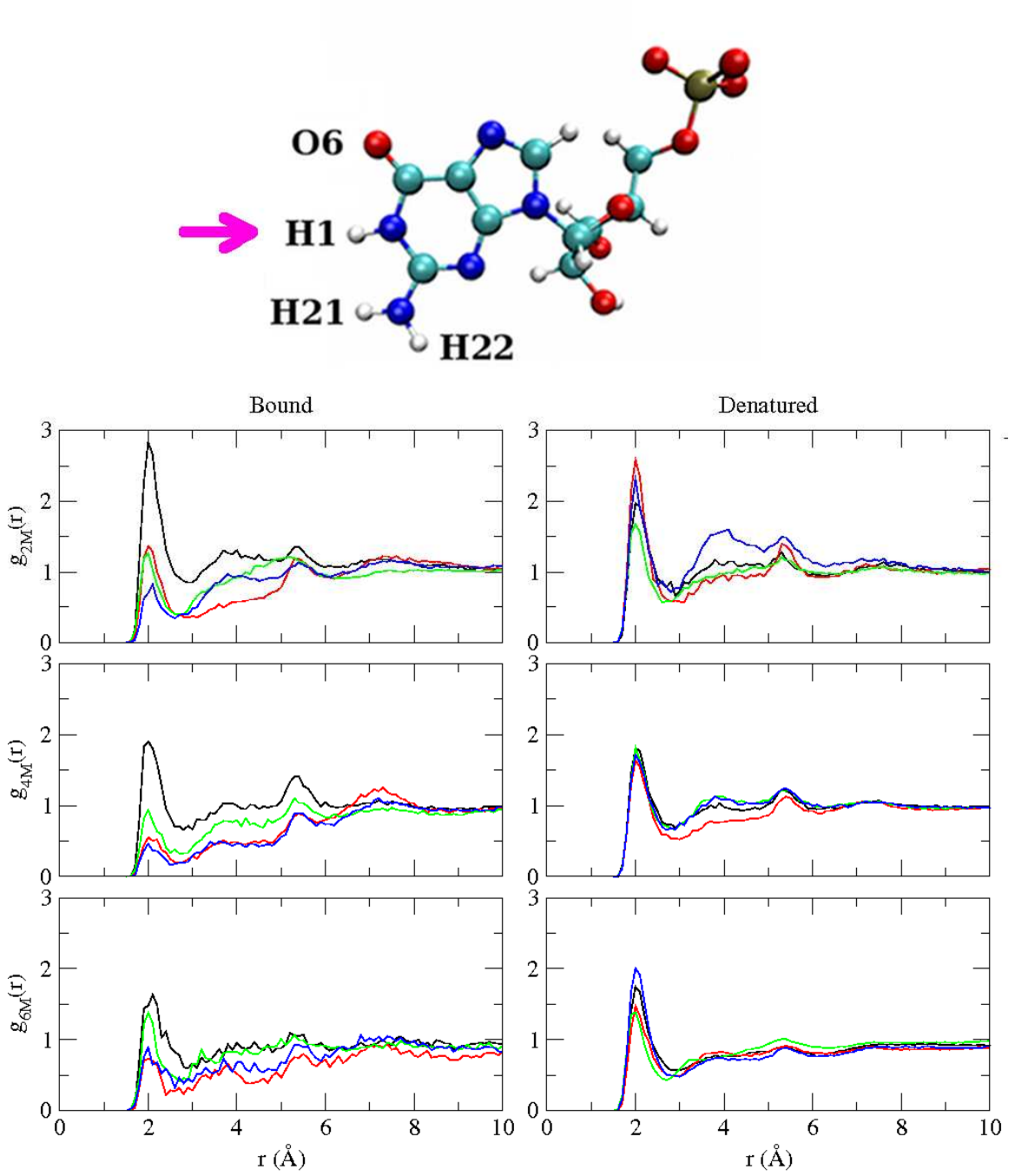
SI Figure 4: Radial distribution functions of the urea nitrogens oxygen. Results are shown for the RNA backbone atoms (left panel, includes phosphodiester and sugar oxygens) and the RNA bases (right panel). The color scheme is: 1M (orange), 2M (green), 4M (blue), 6M (magenta) and 8M (aqua) urea and for water at 0 M (black, bold). RDFs are normalized on a per atom basis. The water RDFs are scaled by 5 and 10.



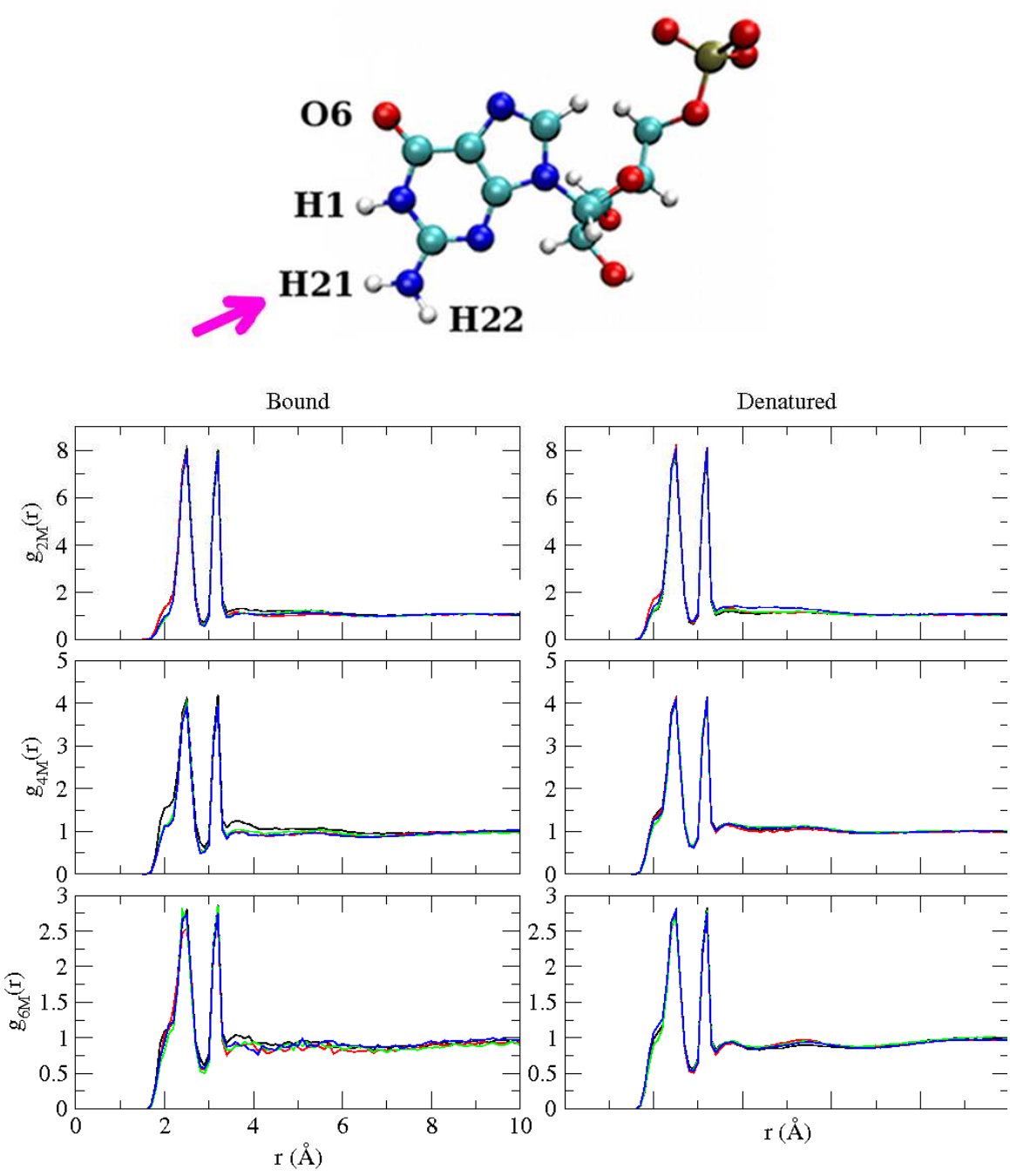
SI Figure 5: The base pair dynamics in the 8 nt RNA duplex measured using the inverse distance (1/r) are shown for [C]=2M and 4M urea.



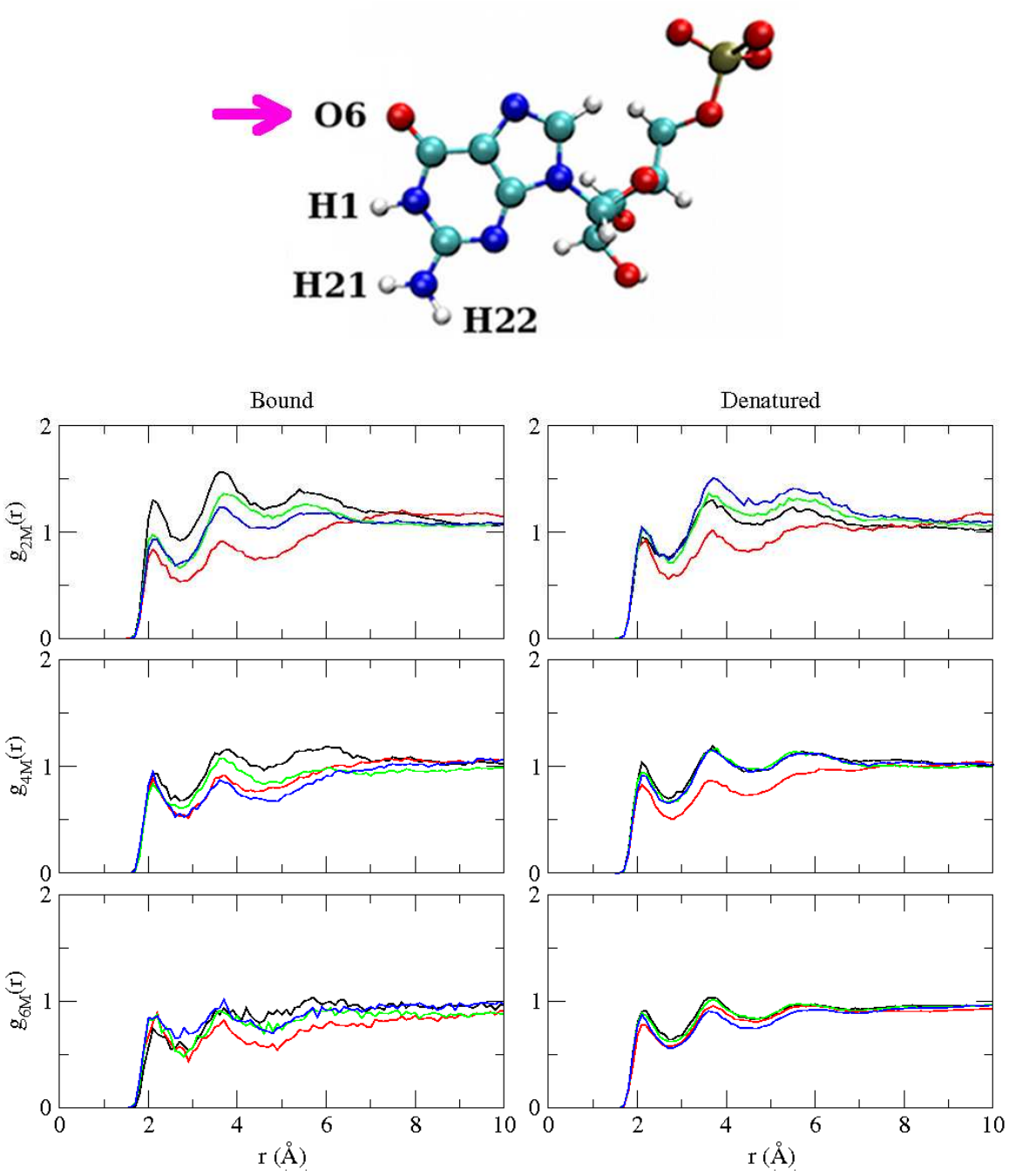
SI Figure 6: GC and UG base pairs and hydrogen bonds.



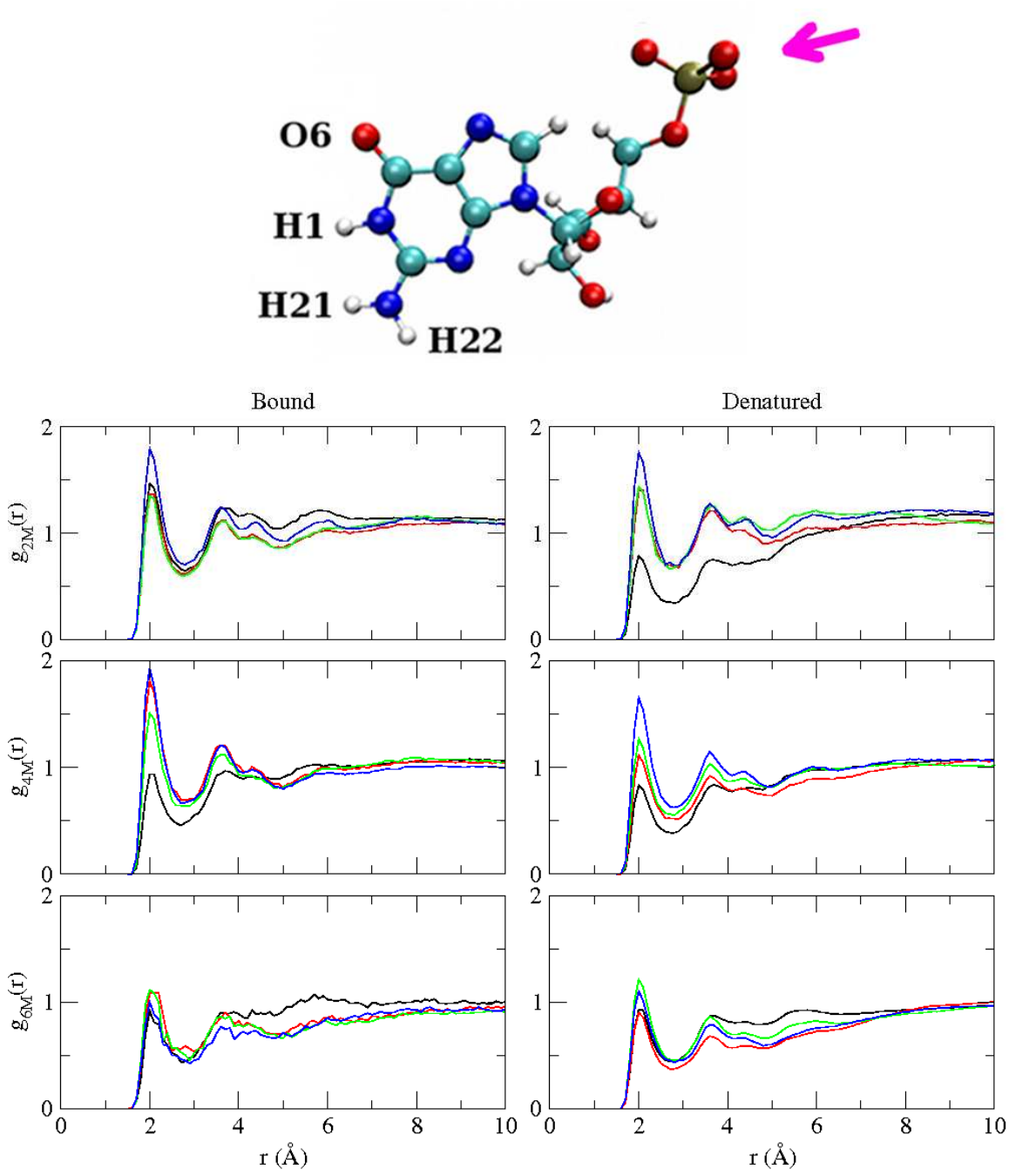
SI Figure 7: Pair functions for the urea oxygen with respect to H1 hydrogen of Guanine for the duplex RNA. Results in the panels on the left (right) are calculated using only the fraction of bound (unbound) RNA molecules.



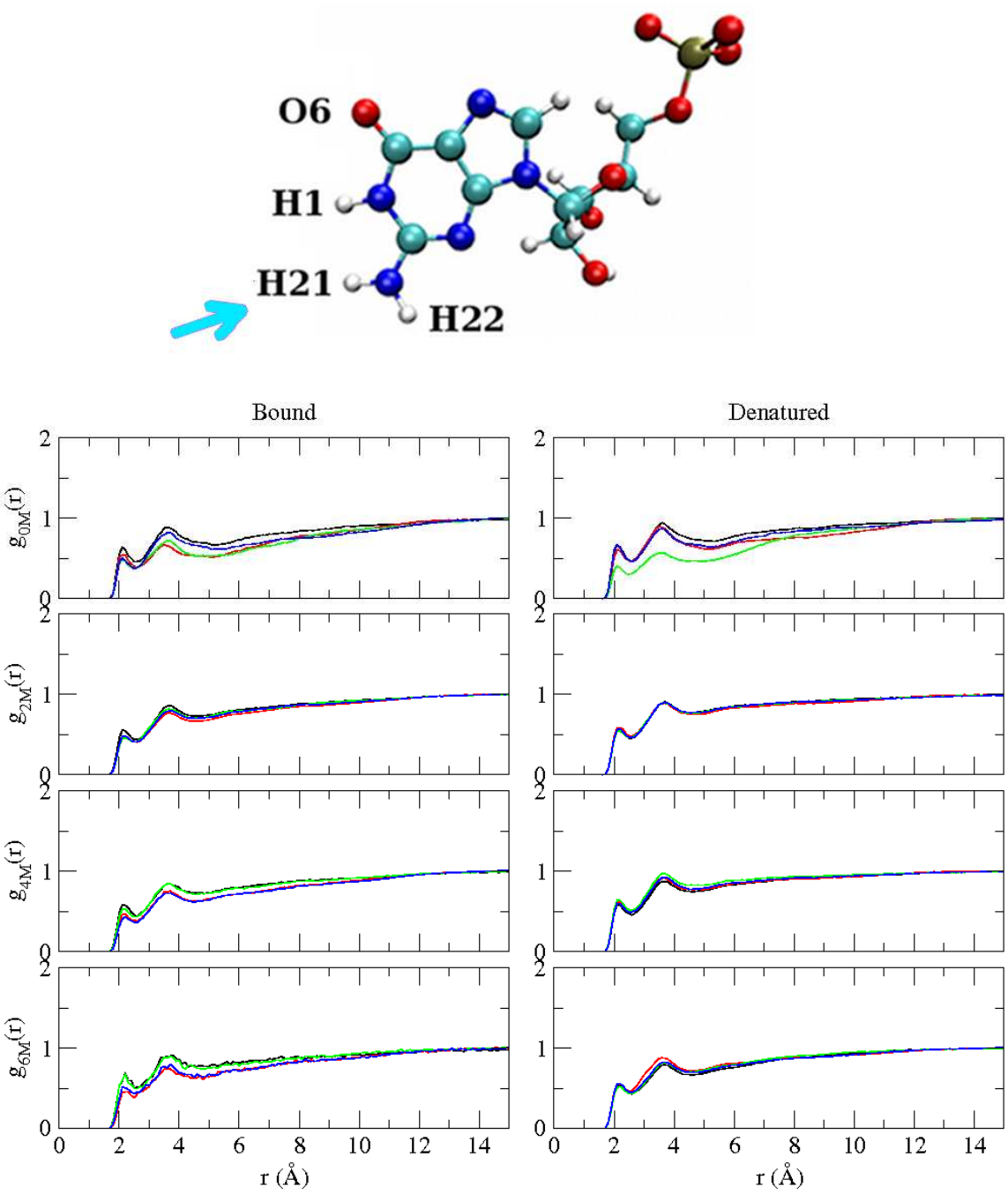
SI Figure 8: Pair functions for the urea oxygen with respect to H12 or H22 hydrogen of Guanine for the duplex RNA. Results in the panels on the left (right) are calculated using only the fraction of bound (unbound) RNA molecules.



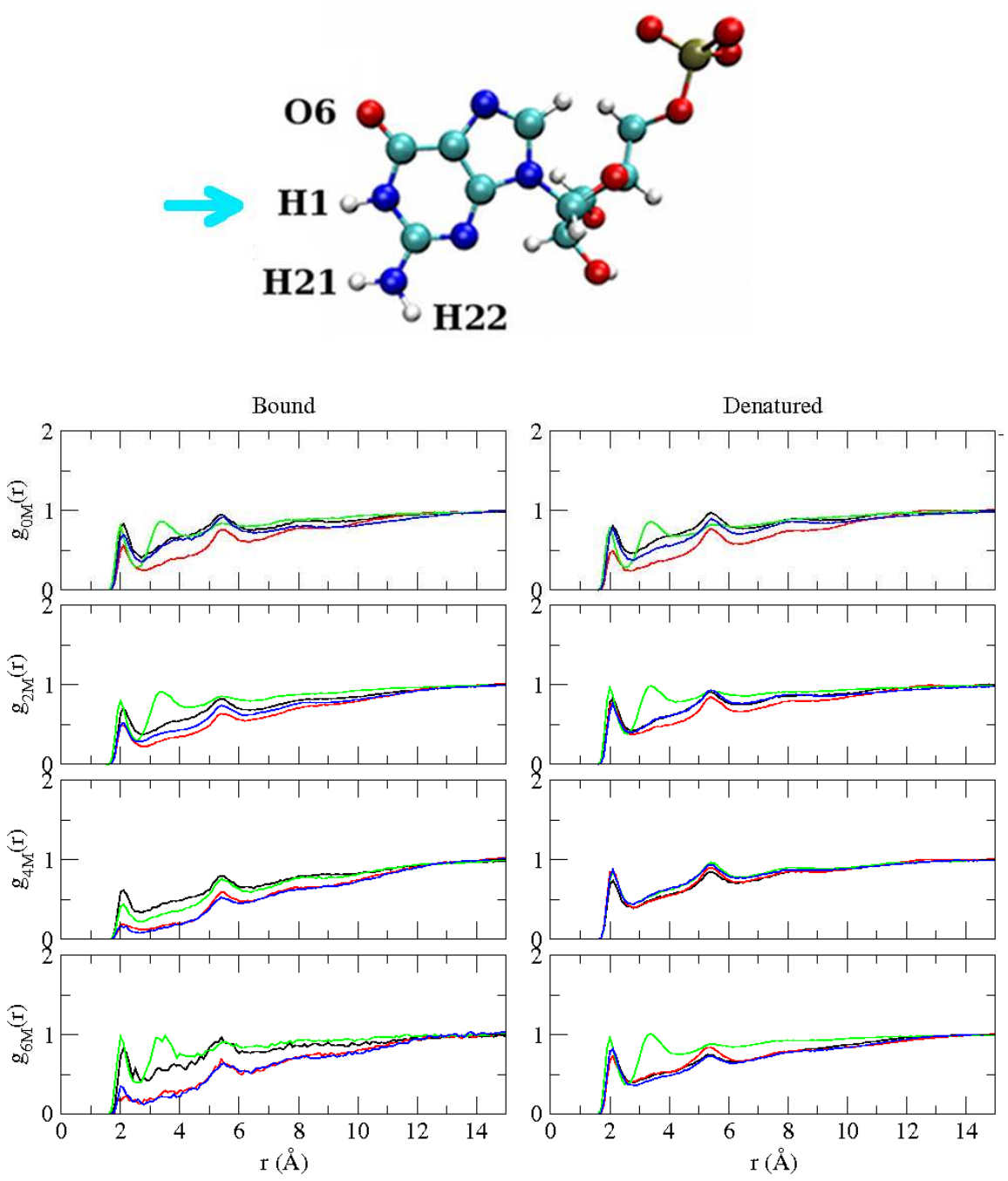
SI Figure 9: Radial distribution functions for amine hydrogen atom of urea with respect to O6 oxygen of Guanine at [C]= 2, 4, and 6 M in the duplex RNA. Results in the panels on the left (right) are calculated using only the fraction of bound (unbound) RNA molecules.



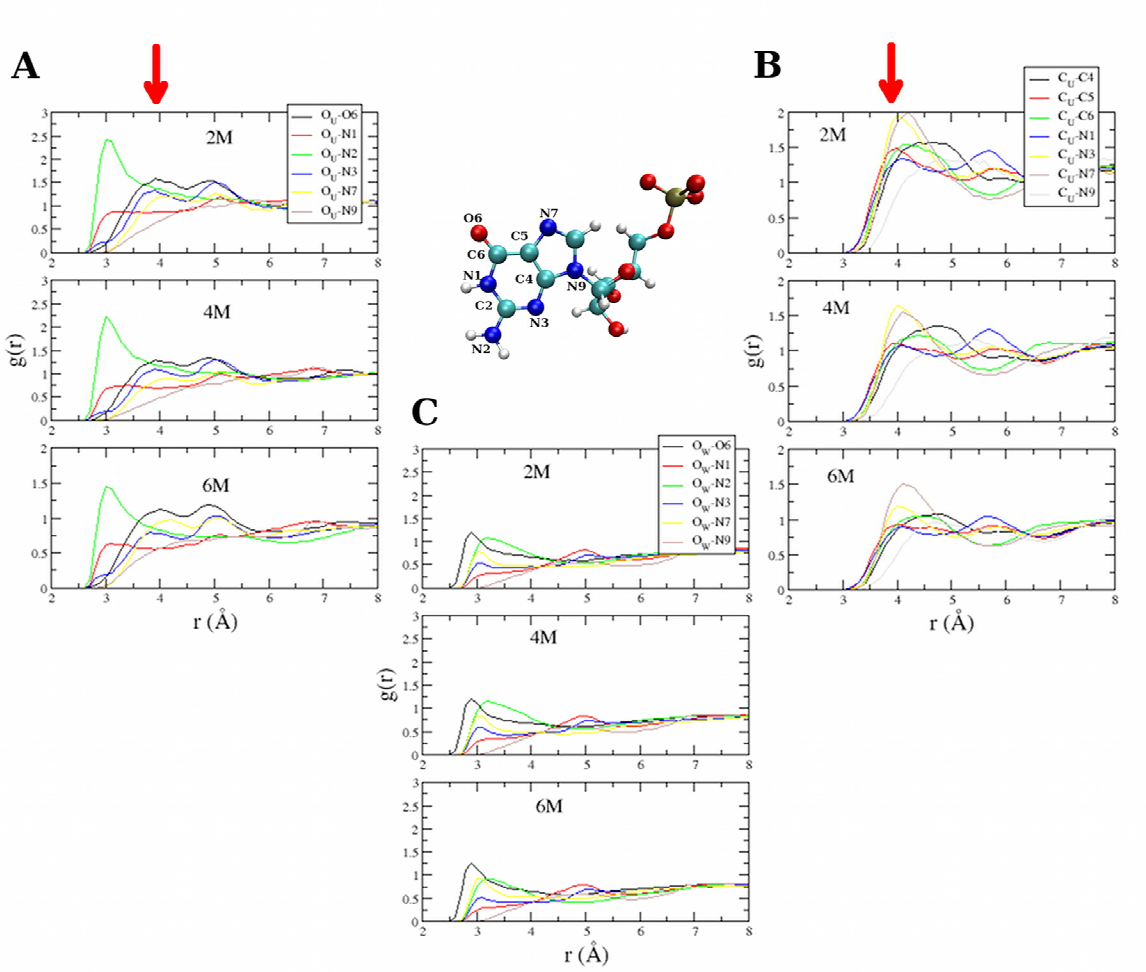
SI Figure 10: RDF for the hydrogens from amine group of urea with respect to the OP1 or OP2 oxygen of Guanine at [C] = 2, 4, and 6 M for the duplex RNA. Results in the panels on the left (right) are calculated using only the fraction of bound (unbound) RNA molecules.



SI Figure 11: RDF for the water oxygen with respect to the H21 or H22 hydrogens of Guanine. Results in the panels on the left (right) are calculated using only the fraction of bound (unbound) RNA molecules.



SI Figure 12: RDF for the water oxygen with respect to the H1 hydrogen of Guanine in duplex RNA at various concentrations of urea. Results in the panels on the left (right) are calculated using only the fraction of bound (unbound) RNA molecules.



SI Figure 13: RDF for the urea oxygen (O_U) and urea carbon (C_U) with respect to various carbon, nitrogen and oxygen atom in Guanine base when RNA duplex is in denatured state for varying urea concentrations. The broad peaks at around 4 Å, indicated with red arrows in **A** and **B**, are due to the stacks formed between urea and base group. These results are consistent with the snapshot of the urea stack shown in Fig. 3B in the main text that were obtained using a completely different urea force field. For comparison, RDF for the water oxygen (O_W) with respect to the atoms in RNA base ring are shown in **C**, which confirms the depleted distribution of water due to the hydrophobic nature of the base.