Supplementary Information

Dissecting Urea-RNA Interactions in SARS-CoV-2 Virus: Nucleobases as the Primary Denaturation Hotspots

Omkar Singh[a,b,*] and Pushyaraga P Venugopal [a,b]

[a]Computational Chemistry Laboratory, Department of Chemistry, GLA University

[b]Department of Chemistry, Indian Institute of Technology Bombay, Powai, India

***Correspondence:**

Dr Omkar Singh

[‡]Department of Chemistry, Indian Institute of Technology Bombay, 400076.

email id: omkantnirala92@gmailcom

ORCID ID: https://orcid.org/0000-0002-2943-0405

Section 1

Methods

Method 1

Urea Parameter

```
; Created by cgenff_charmm2gmx.py
[ moleculetype ]
; Name            nrexcl
Urea              3
[ atoms ]
;   nr       type  resnr residue  atom   cgnr     charge       mass  typeB    chargeB      massB
; residue    1 Urea rtp Urea q  qsum
     1         1   Urea     N1      1    -0.690    14.007   ;
     2         1   Urea      C      2     0.600    12.011   ;
     3         1   Urea     N2      3    -0.690    14.007   ;
     4         1   Urea      O      4    -0.580    15.999   ;
     5         1   Urea     H1      5     0.340     1.008   ;
     6         1   Urea     H2      6     0.340     1.008   ;
     7         1   Urea     H3      7     0.340     1.008   ;
     8         1   Urea     H4      8     0.340     1.008   ;
[ bonds ]
;   ai     aj funct            c0          c1            c2          c3
    1     2    1
    1     5    1
    1     6    1
    2     3    1
    2     4    1
    3     7    1
    3     8    1
[ pairs ]
;   ai     aj funct            c0          c1            c2          c3
    1     7    1
    1     8    1
    3     5    1
    3     6    1
    4     5    1
    4     6    1
    4     7    1
    4     8    1
[ angles ]
;   ai     aj    ak funct           c0            c1          c2          c3
    2     1     5    5
    2     1     6    5
    5     1     6    5
    1     2     3    5
    1     2     4    5
```

```
   3    2    4    5
   2    3    7    5
   2    3    8    5
   7    3    8    5
[ dihedrals ]
;   ai     aj    ak     al funct          c0          c1          c2          c3          c4          c5
   5    1    2    3    9
   5    1    2    4    9
   6    1    2    3    9
   6    1    2    4    9
   1    2    3    7    9
   1    2    3    8    9
   4    2    3    7    9
   4    2    3    8    9
[                                        dihedrals                                        ]
;     ai        aj       ak        al  funct          c0          c1          c2          c3
   2    1    3    4    2
```

Method 2

FEL calculation with eRMSD and SASA

eRMSD usually means ensemble RMSD (*effective RMSD*), and it's common in structural biology molecular simulations, mainly in RNA and DNA when comparing multiple structures (an ensemble) rather than just two single structures.

If we have an ensemble of *M* structures, the eRMSD is defined as the average RMSD over the ensemble:

$$eRMSD = \sqrt{\frac{1}{M}\sum_{k=1}^{M}(\frac{1}{N}\sum_{i=1}^{N}\left|r_i^k - r_i^{ref}\right|^2)}$$

Where M = number of structures in the ensemble, N = number of atoms per structure, r_i^k = position of atom *i* in structure k. Now we define the CVs, such as SASA and eRMSD, to perform the FEL calculation.

$$FEL = -k_B T \ln P(x, y)$$

Where x and y are the probability distribution of the CVs.

Method 3

Preferential Binding

The KB integral helps to understand the mutual affinities between the interacting molecular species in a solution. The KB integrals between components of the system can be expressed using the equation:

$$G_{\alpha\beta} = 4\pi \int_0^\infty r^2 \left[g_{\alpha\beta}(r) - 1\right] dr$$

where $g_{\alpha\beta}$ is the radial distribution function, and r is the interatomic separation. A higher value of $G_{\alpha\beta}$ indicates the inter-atomic attraction between the species α and β (either direct or mediated by other components), which is found to be very strong. The α denotes the water and protein by β, and the cosolvent by γ, so $v_{\alpha\beta}$, preferential binding[53,54] of water to protein is given by:

$$v_{\alpha\beta} = \rho_\alpha \left(G_{\alpha\beta} - G_{\beta\gamma}\right)$$

Method 4

Script used in Analysis

The codes are available at github repository:

https://github.com/Ompal12/Urea_induced_RNA_denaturation/tree/main

Section 2

Figures

Graphs of Thermodynamic Equilibrium.

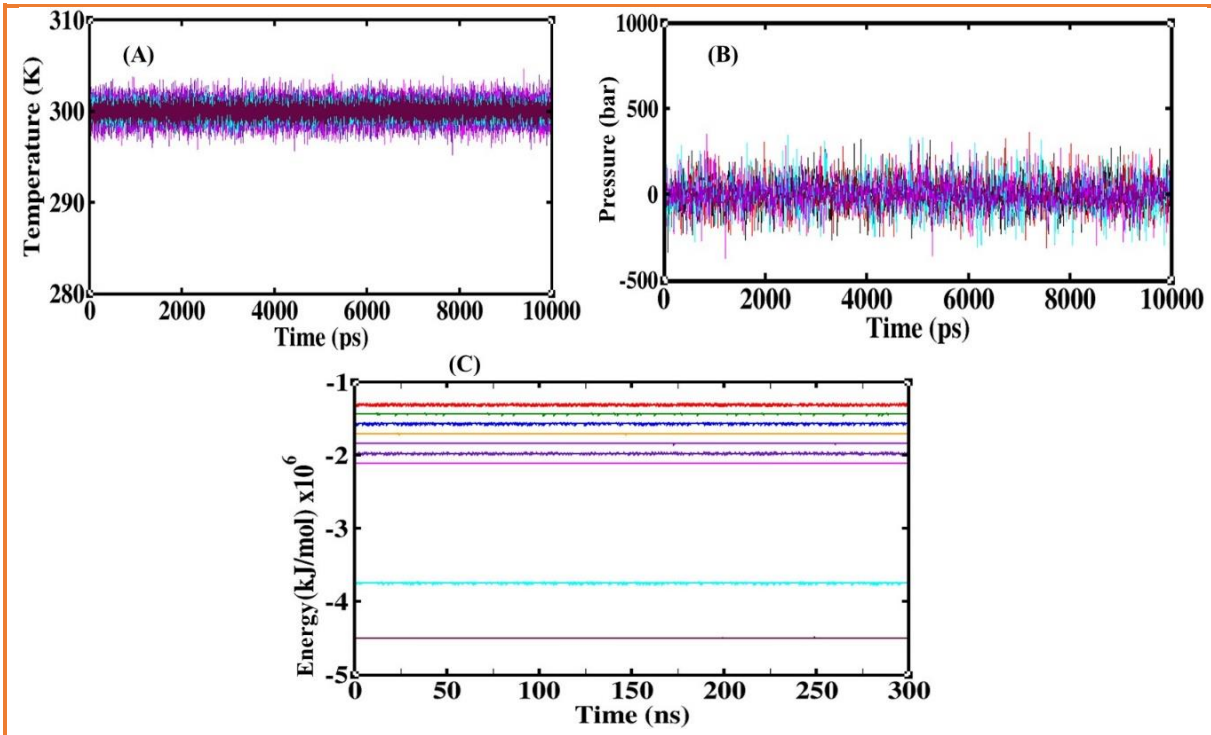


Figure S1. The thermodynamic equilibrium graph time evolution of temperature, pressure, and potential energy in all the urea concentrations.

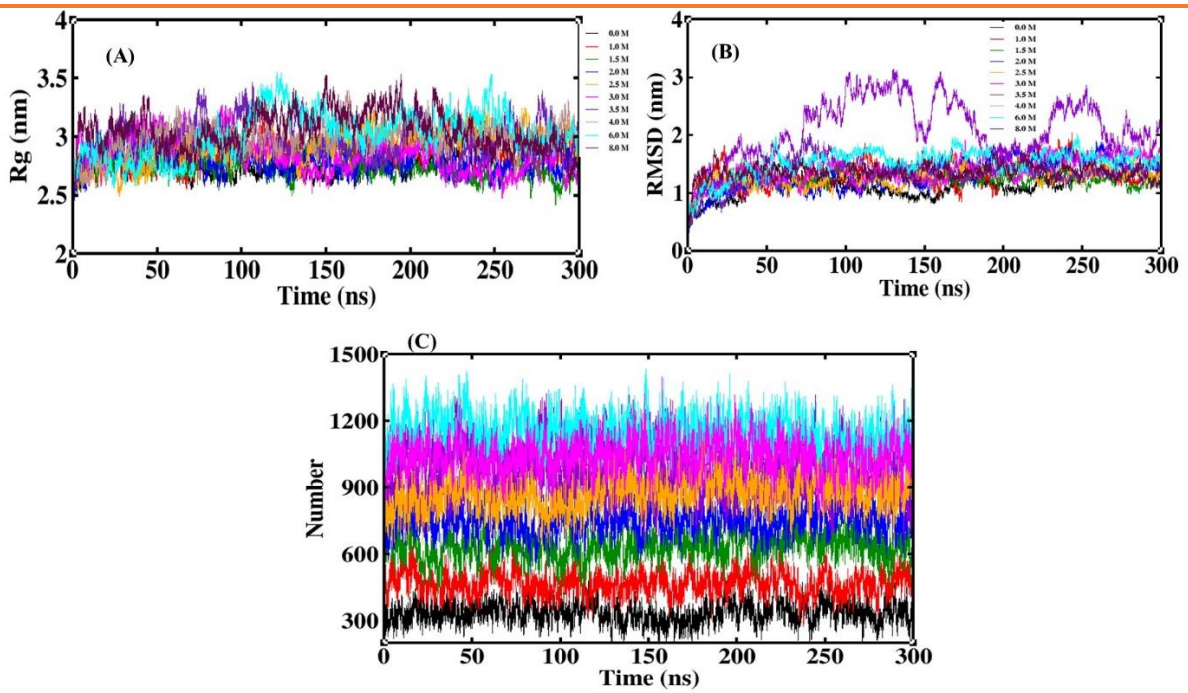


Figure S2. (A) Radius of gyration (Rg) and (B) root mean square deviation (RMSD) time evaluation graph. (C) Convergence of the simulation trajectory based on the urea molecules near the 5.4 Å RNA surface cutoff. The urea molecules were found to be constant after a 100 ns simulation.

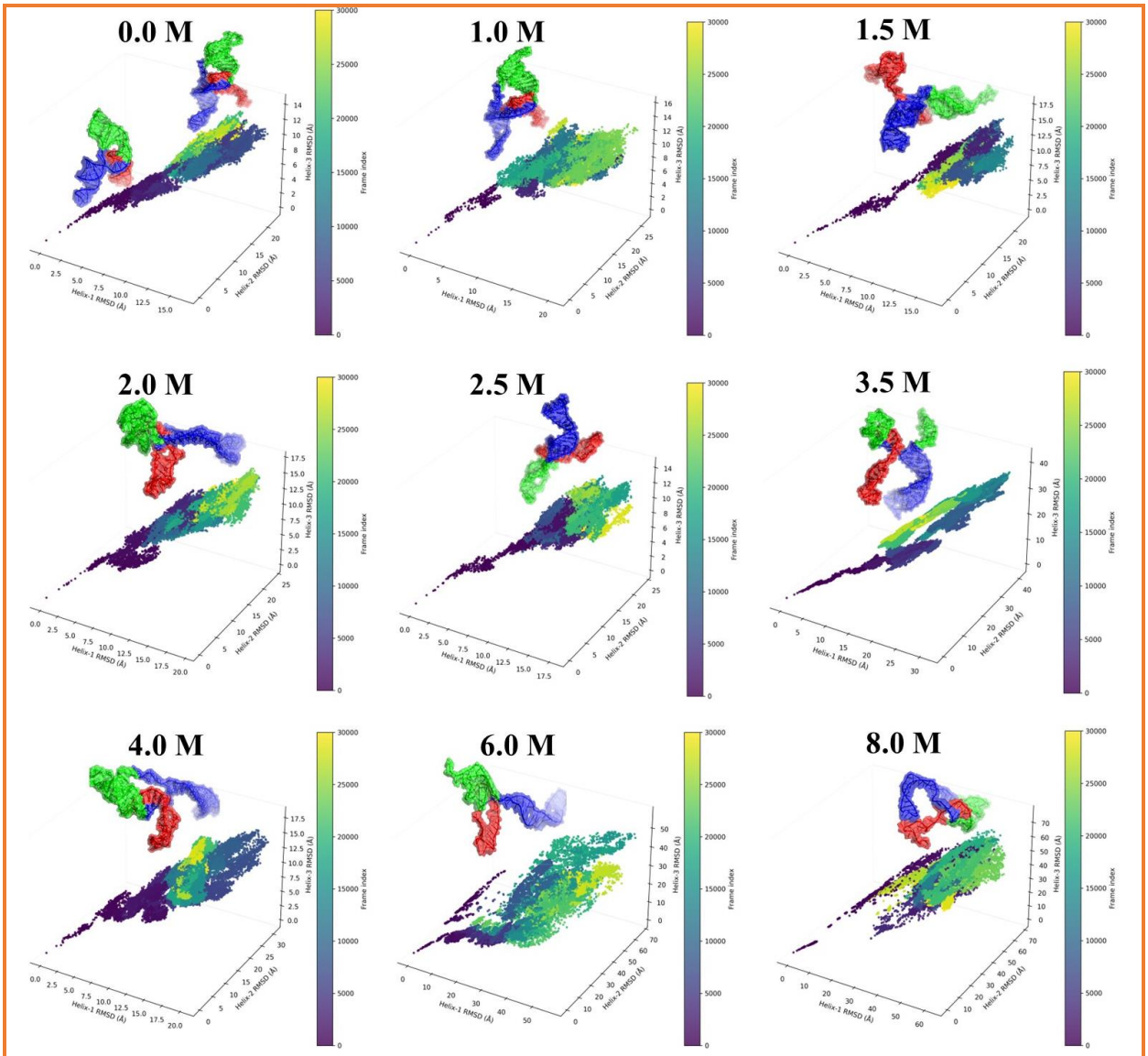
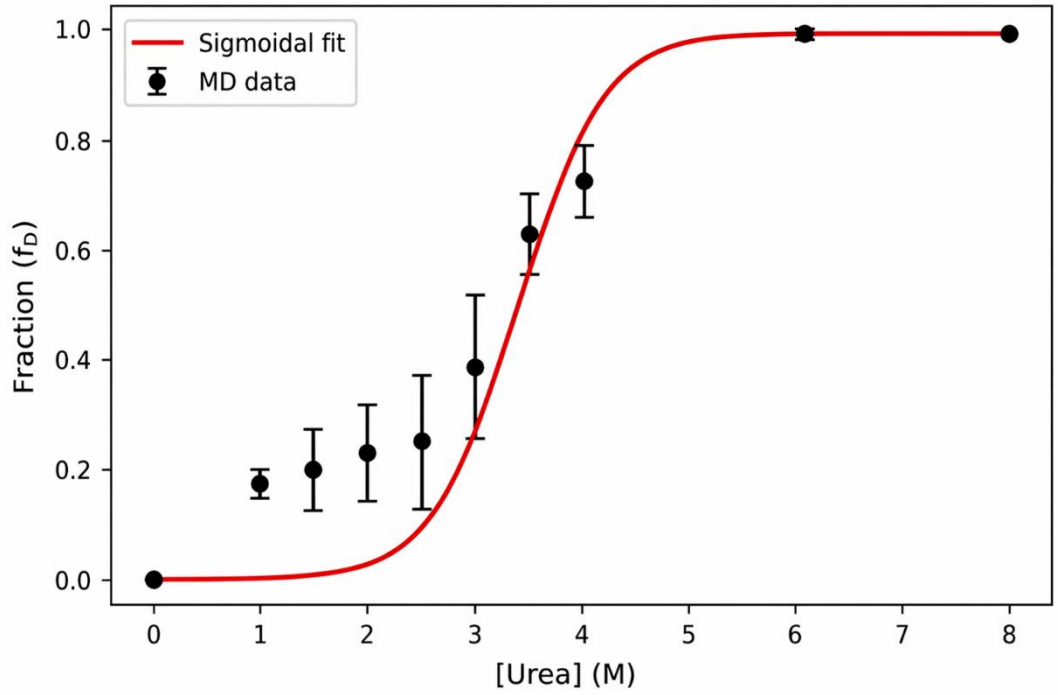


Figure S3. 3D plot illustrates the three-dimensional RMSD distributions of individual helical segments as a function of urea concentration. The color bar represents the time frame, and based on that, characterized by low and higher RMSD values across all helices, indicating the conformational topology. The cartoon surface structure represents the max helical RMSD fluctuation from the reference structure.



Figures S4. Fraction of disrupted RNA population as a function of urea concentration, calculated from RMSD relative to the reference structure, using an RMSD threshold (> 10 Å) to distinguish native and disrupted states.

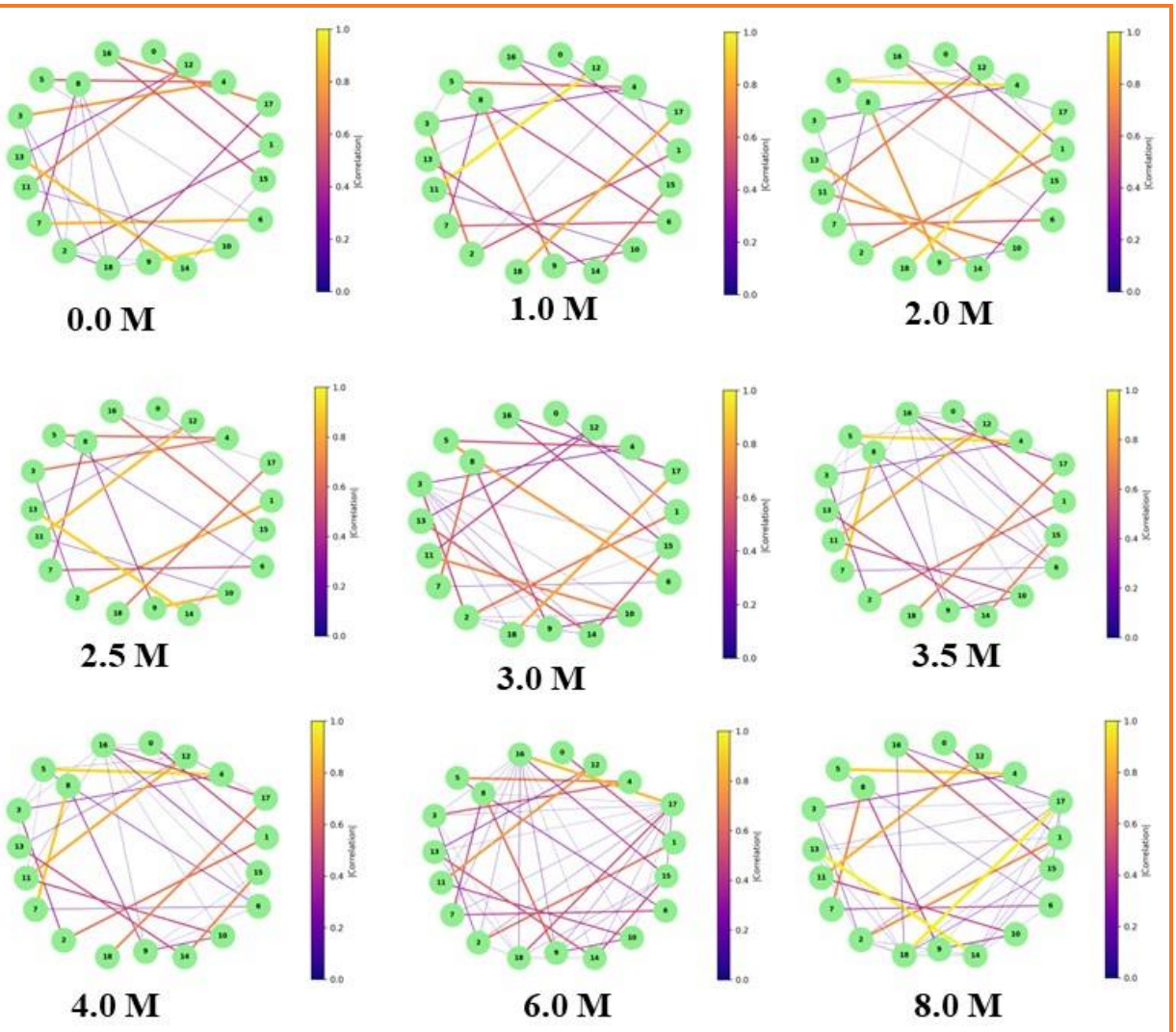


Figure S5.The network connection between the base pairs and their twist angle change with urea concentration. The graph represents the connection between the base pairs and the dynamics at higher concentrations than pure water and at lower urea (1.0-2.0 M) concentrations.

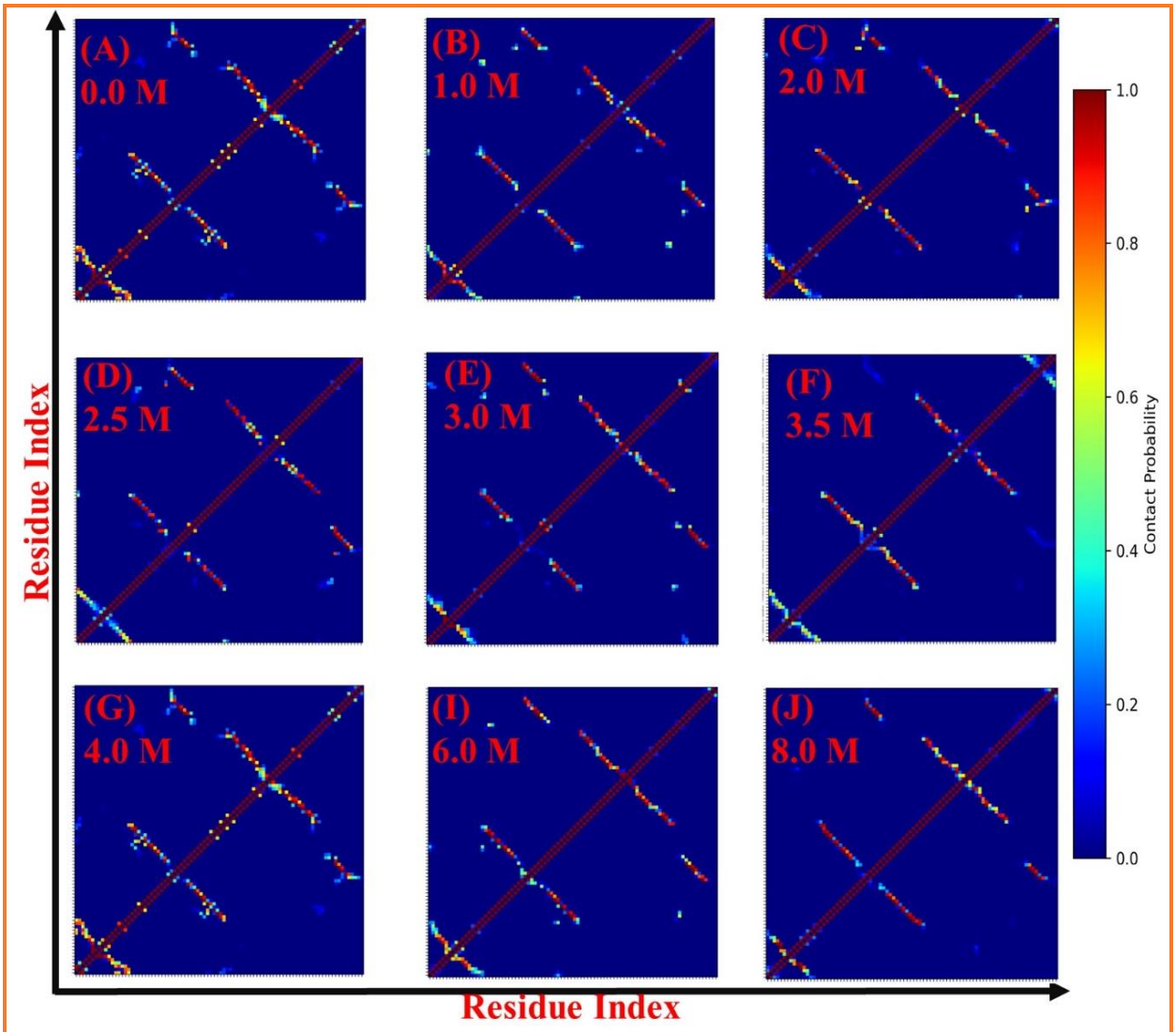
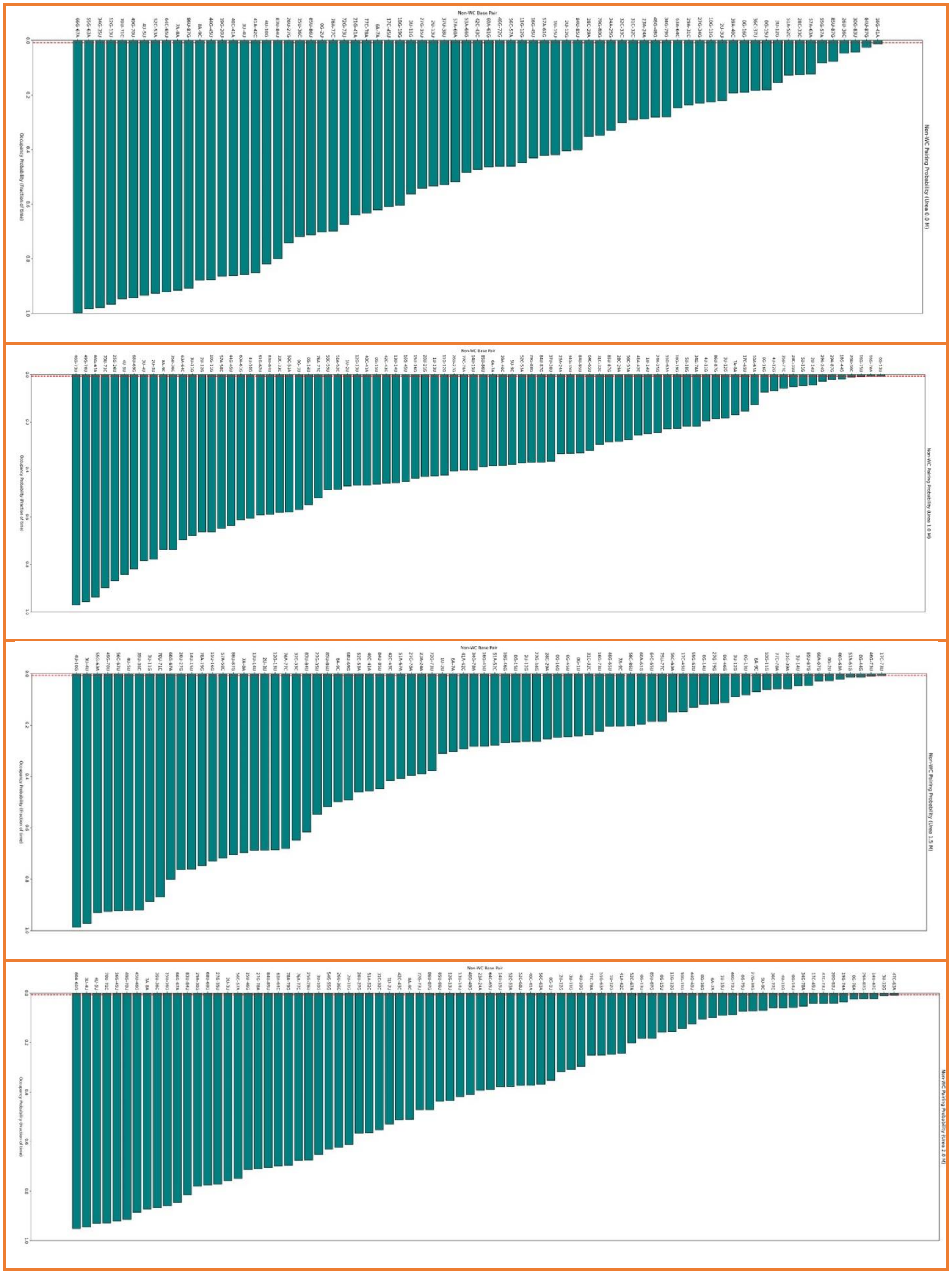
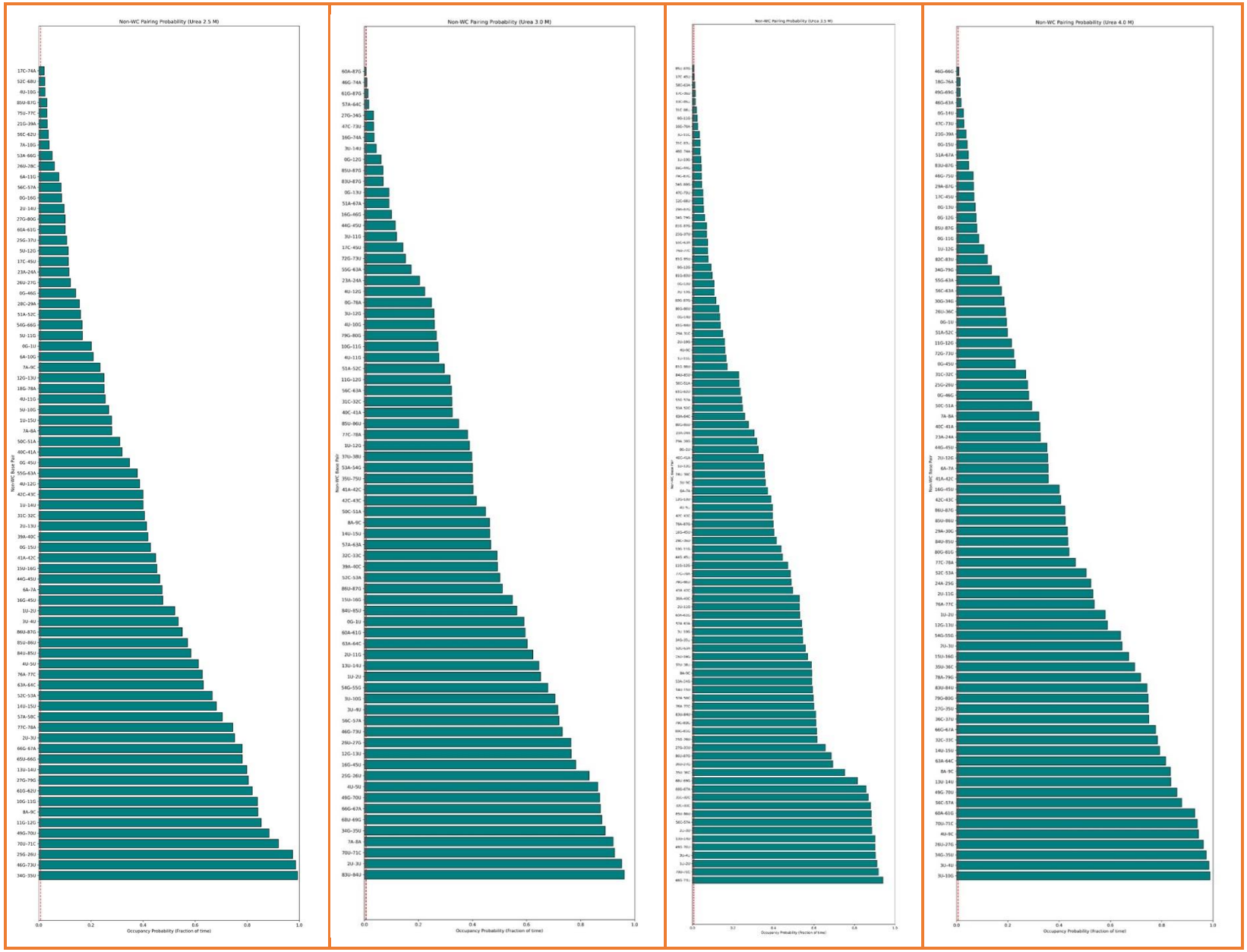


Figure S6. The residue-residue contact map in the RNA structure at different urea concentrations. In pure water and lower urea concentration, the bases interact more with a higher contact probability. At higher urea concentration, the residue-residue contact probability reduced, and the number of nucleobases with contact was also less.


Non-WC Base Pair
Occupancy Probability (Fraction of time)
Non-WC Pairing Probability (Urea 0.0 M)
Non-WC Pairing Probability (Urea 1.0 M)
Non-WC Pairing Probability (Urea 1.5 M)
Non-WC Pairing Probability (Urea 2.0 M)


Non-WC Pairing Probability (Urea 2.5 M)
Non-WC Base Pair
Occupancy Probability (Fraction of time)
Non-WC Pairing Probability (Urea 3.0 M)
Non-WC Base Pair
Occupancy Probability (Fraction of time)
Non-WC Pairing Probability (Urea 3.5 M)
Occupancy Probability (Fraction of time)
Non-WC Pairing Probability (Urea 4.0 M)
Non-WC Base Pair
Occupancy Probability (Fraction of time)

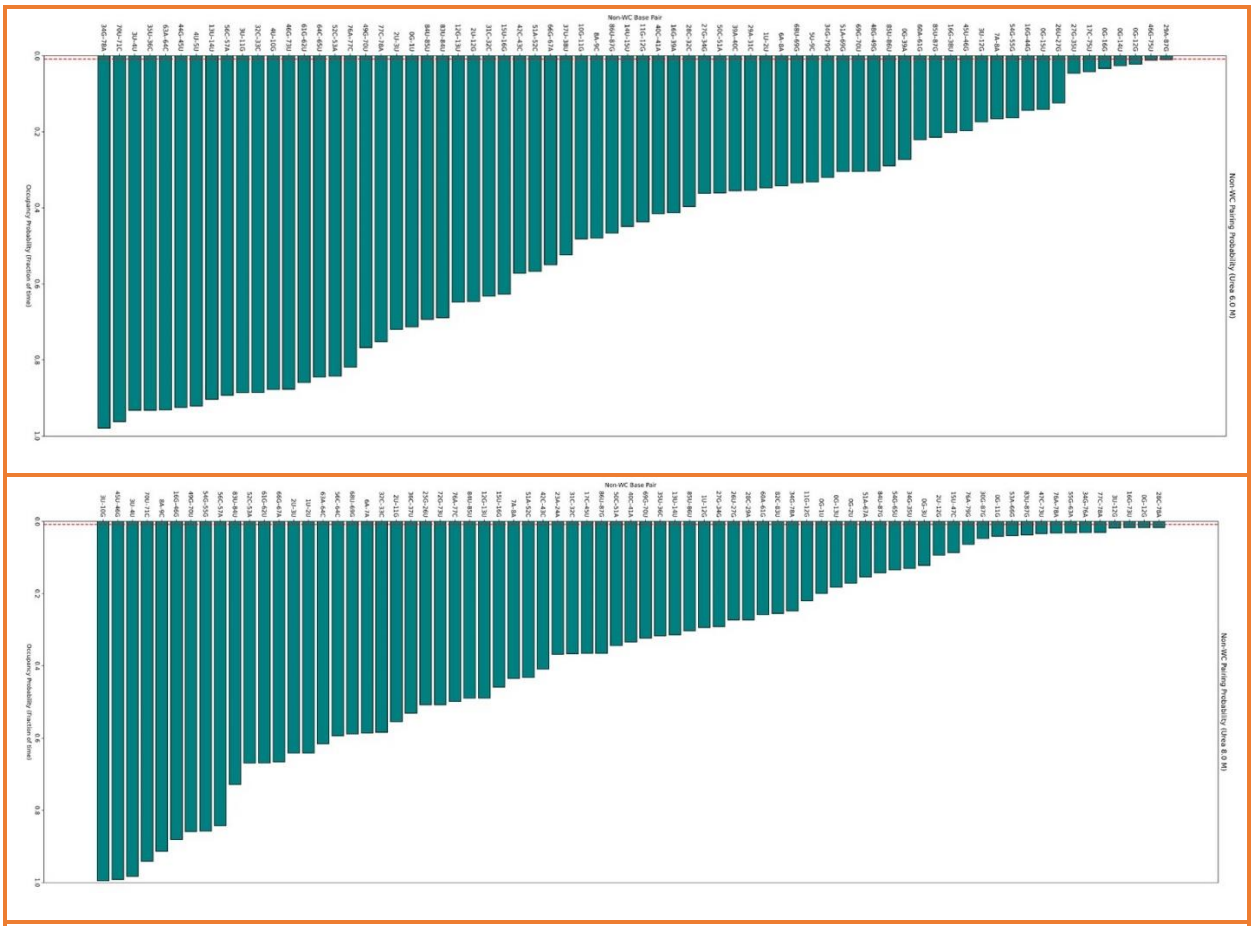


Figure S7. Non-WC Hbs and their probability distribution graph.

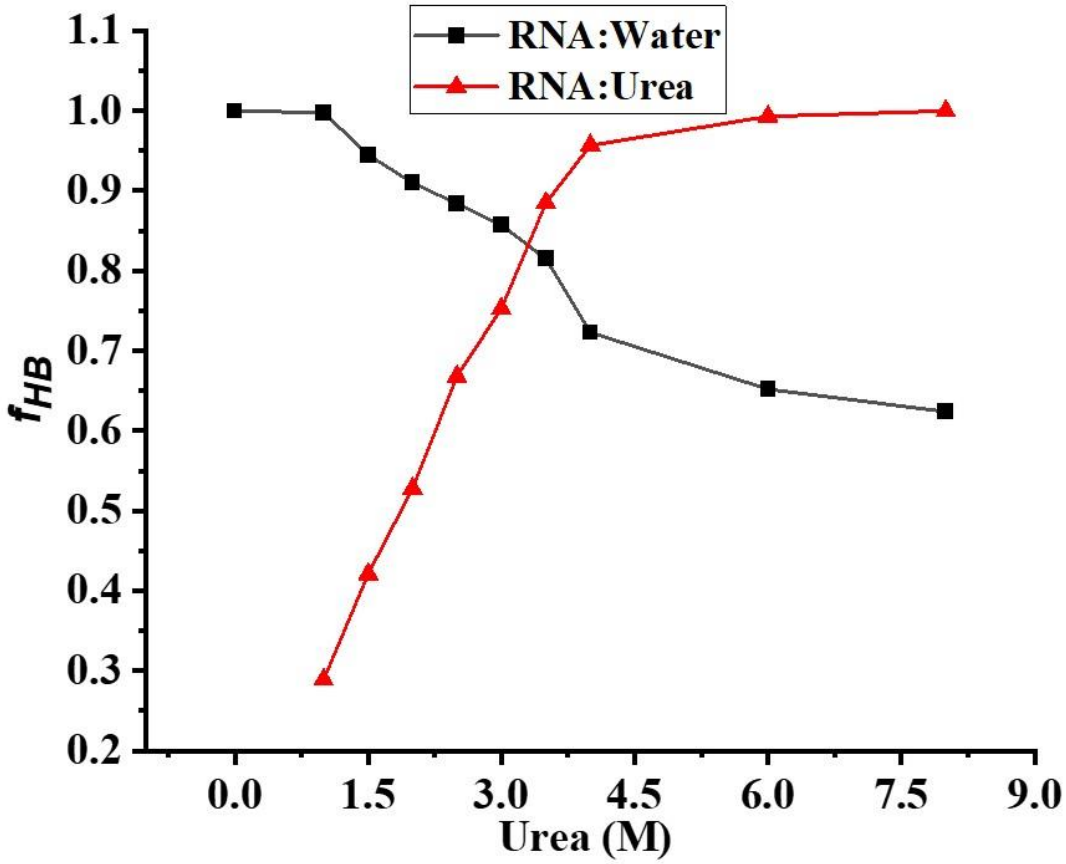


Figure S8. The fraction of the hydrogen bond of RNA: Urea and RNA: Water in urea solutions.

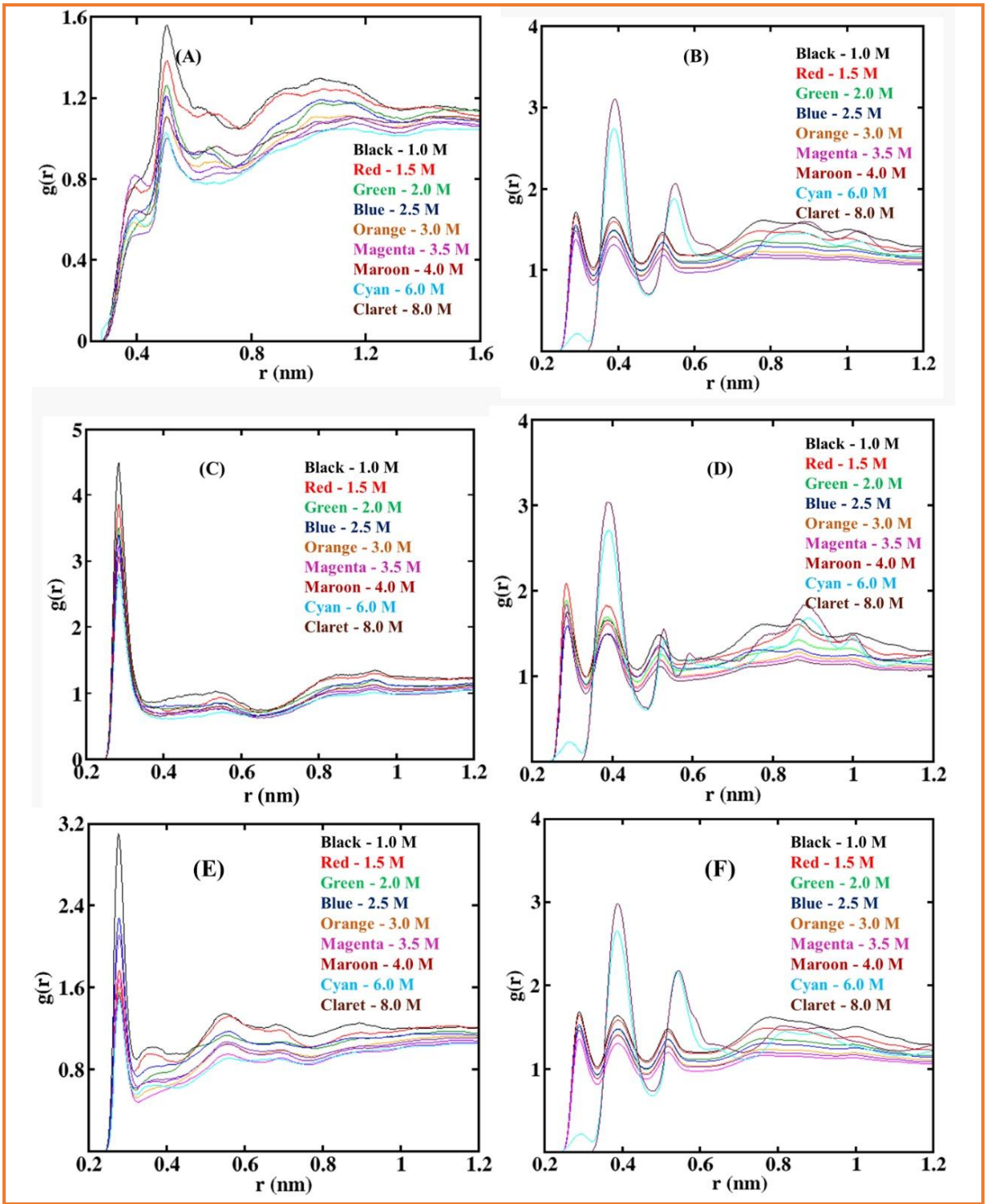


Figure S9. RDFs of (A) O_U around nitrogen of RC (B) N_U around the phosphate of RC (C) O_U around nitrogen of RG (D) N_U around the phosphate of RG (E) O_U around nitrogen of RU (F) N_U around the phosphate of RU

Section 3

Tables

Table S1. Residue-Residue contact probabilities from a trajectory generated with GROMACS, and the average contact probabilities and ΔContact values relative to 0. M reference.

Urea (M)	**Avg Contact Prob**	**$\Delta C = U_C$ – 0.0 M**
0	0.64	0
1	0.52	- 0.12
2	0.37	- 0.27
2.5	0.32	- 0.32
3	0.34	- 0.30
3.5	0.25	- 0.39
4	0.26	- 0.38
6	0.23	- 0.41
8	0.19	- 0.45

Table S2. The number of base pairs and the lifetime of these Base pairs in (ps)

0.0 M	1.0 M	1.5 M	2.0 M	2.5 M	3.0 M	3.5 M	4.0 M	6.0 M	8.0 M
52C–66G = 3299.3	52C–66G =2571.4	52C–66G =3563.6	52C–66G =769.4	52C–66G =1751.4	52C–66G =2556.8	52C–66G = 1527.2	52C–66G=1258.3	52C–66G=2532.8	52C–66G=1068.1
50C–69G =13455	50C–69G =7597.4	50C–69G =2980.7	50C–69G =1143.9	50C–69G =7376.2	50C–69G =5376.4	50C–69G= 4863.2	50C–69G=7987.8	50C–69G=1863.5	50C–69G=1364.4
25G–36C =12447.5	25G–36C =1807.7	25G–36C =7826.3	25G–36C =5655.3	25G–36C =10291.4	25G–36C =1150	25G–36C= 515.8	25G–36C=1661.7	25G–36C=12412.5	25G–36C=1618
48G–71C =10922.7	48G–71C =9946.7	48G–71C =3386.8	48G–71C =4703.2	48G–71C =6320.2	48G–71C =6454.3	48G–71C=5009.3	48G–71C=3130.3	48G–71C=2014.2	48G–71C=1871.3
54G–64C =9824	54G–64C =10679.6	54G–64C =5582.1	54G–64C =6141.3	54G–64C =5809.8	54G–64C =2806.7	54G–64C=4189	54G–64C=7143.6	54G–64C=1690.2	54G–64C=83.3
32C–80G =5898.8	32C–80G =27227.3	32C–80G =9328.1	32C–80G =29960	32C–80G =7060.7	32C–80G =14955	32C–80G=8350	32C–80G=9498.3	32C–80G=17814.3	32C–80G=24960
31C–81G =7085.7	31C–81G =3922	31C–81G =8517.1	31C–81G =1135.5	31C–81G =5028.8	31C–81G =4668	31C–81G=6750	31C–81G=5297.3	31C–81G=567.6	31C–81G=1026.8
30G–82C =3029	30G–82C =4618.2	30G–82C =7420	30G–82C =3416.7	30G–82C =23030.8	30G–82C =4225	30G–82C=1975	30G–82C=11952	30G–82C=1676.5	30G–82C=4219
19G–42C =3731.5	19G–42C =3327.8	19G–42C =3732.1	19G–42C =2890.6	19G–42C =5591.3	19G–42C =6317	19G–42C=10281	19G–42C=5514.2	19G–42C=4037.7	19G–42C=8011.3
33C–79G =6587.2	33C–79G =2499.2	33C–79G =1993.3	33C–79G =2722.7	33C–79G =2827.7	33C–79G =2300.0	33C–79G=310	33C–79G=106.2	33C–79G=1759.9	33C–79G=1313.2
21G–40C =7383.4	21G–40C =2883.5	21G–40C =3527.4	21G–40C =2946	21G–40C =1511.4	21G–40C=1101.9	21G–40C=1094.8	21G–40C=3461.6	21G–40C=1012.7	21G–40C=1082.5
18G–43C =8294.4	18G–43C =7350	18G–43C =3932	18G–43C =5726.3	18G–43C =1227.3	18G–43C =9028.8	18G–43C=1929.2	18G–43C=3791	18G–43C=1316.7	18G–43C=1389.5
24A–37U =899.6	24A–37U =449.1	24A–37U =477	24A–37U =410.8	24A–37U =500	24A–37U =494.6	24A–37U=366.2	24A–37U=572.5	24A–37U=555.2	24A–37U=455.7
23A–38U =558.7	23A–38U =434.8	23A–38U =614.3	23A–38U =707.5	23A–38U =221.3	23A–38U =445.8	23A–38U=495.3	23A–38U=728.5	23A–38U=639.7	23A–38U=653.4
22U–39A =182.2	22U–39A =246.5	22U–39A =334.8	22U–39A =279.4	22U–39A =230.7	22U–39A =167.5	22U–39A=245.8	22U–39A=317.1	22U–39A=384.1	22U–39A=268.4
17C–44G =231	17C–44G =696.4	17C–44G =945.8	17C–44G =1204	17C–44G =1636	17C–44G =1926.8	17C–44G=307.3	17C–44G=1266.8	17C–44G=181.9	17C–44G=1286.5
29A–83U =753.7	29A–83U =869.7	29A–83U =531.8	29A–83U =1261.9	29A–83U =634.3	29A–83U =69.2	29A–83U=266.7	29A–83U=714.7	29A–83U=176.1	29A–83U=190
75U–76A =78.8	75U–76A =73.6	75U–76A =69.8	75U–76A =67.7	75U–76A =83.1	75U–76A =80.4	75U–76A=71.6	75U–76A=75	75U–76A=50	75U–76A=71.1
17C–18G =71.5	17C–18G =62	17C–18G =68.5	17C–18G =60.9	17C–18G =65.1	17C–18G =64.9	17C–18G=64.3	17C–18G=63.9	17C–18G=0	17C–18G=58.6
74A–75U =875.3	74A–75U =269	74A–75U =65.6	74A–75U =71.6	74A–75U =72.4	74A–75U =75.2	74A–75U=73.2	74A–75U=70.6	74A–75U=68.9	74A–75U=72.3
51A–68U =977.7	51A–68U =119.3	51A–68U =135.8	51A–68U =155.8	51A–68U =204.4	51A–68U =141.5	51A–68U=136.4	51A–68U=183	51A–68U=97.7	51A–68U=139.8
16G–17C =1106.6	16G–17C =65.8	16G–17C =68.7	16G–17C =58.8	16G–17C =74.3	16G–17C =63.7	16G–17C=84.4	16G–17C=66.1	16G–17C=74.8	16G–17C=98.6

Section 4

CHARMM36 ff: Results
Set-II

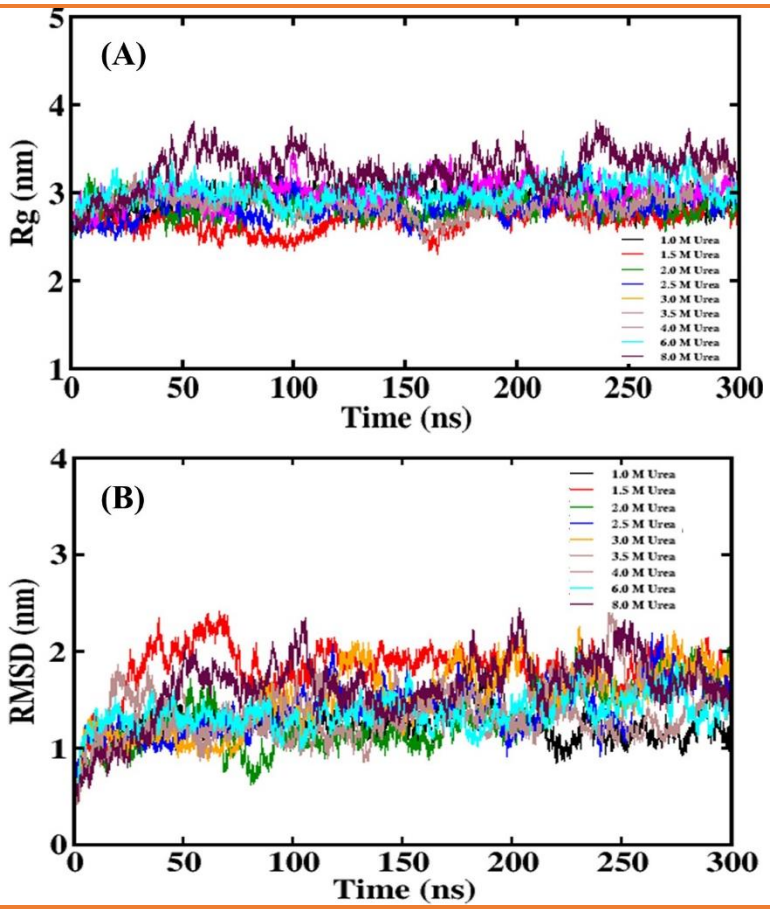


Figure S10. **(A)** Radius of gyration (Rg) and (B) root mean square deviation (RMSD) time evaluation graph.

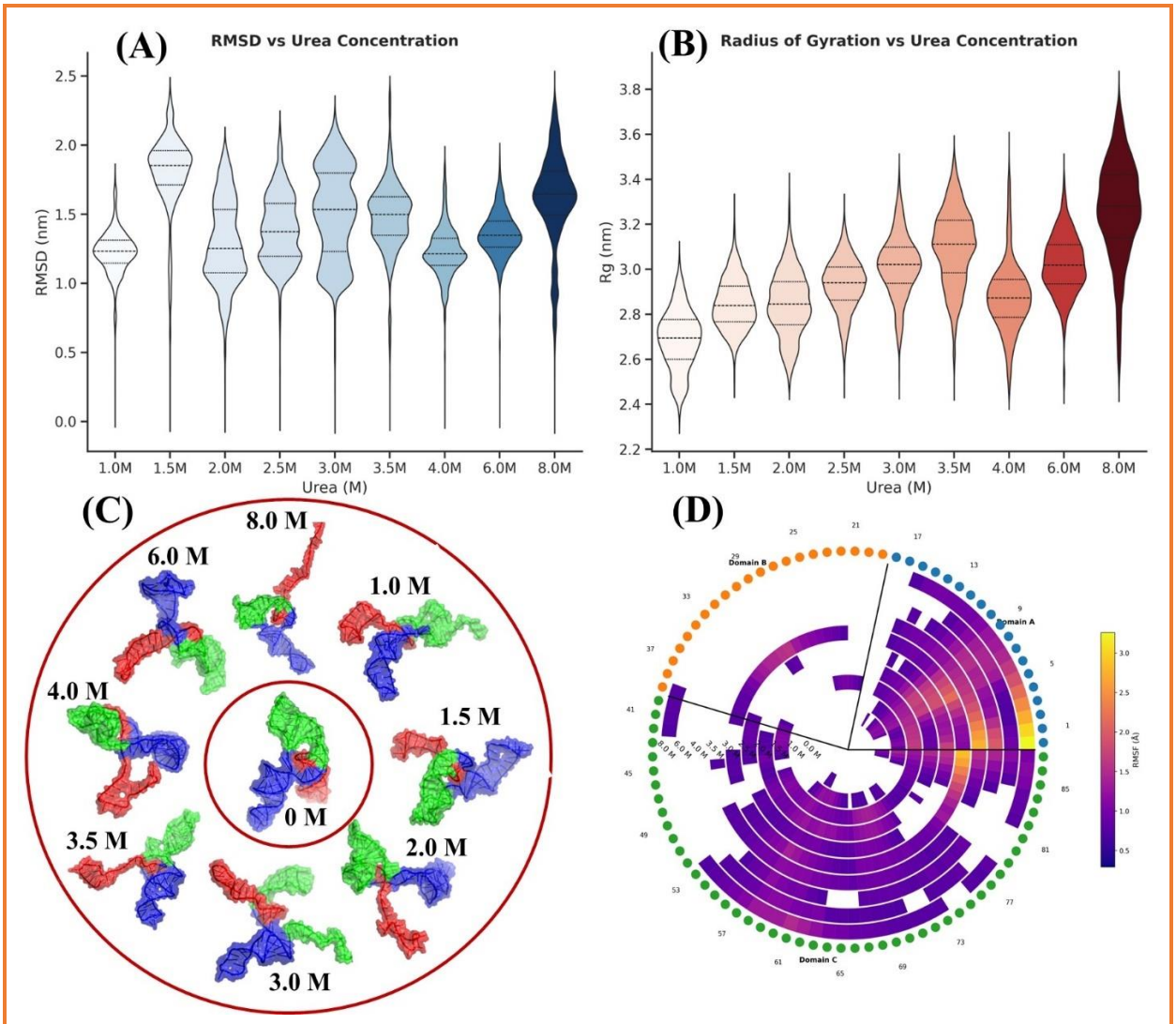


Figure S11. The upper panel presents violin distribution plots of the **(A)** root-mean-square deviation (RMSD) and **(B)** radius of gyration (Rg) of the RNA as a function of urea concentration (0–8 M). The violin distribution represents the central tendency and the full conformational distribution sampled during the simulations. **(C-D)** In the lower panel, the maximum fluctuation is evaluated across all concentrations, and the circular heat map shows the root mean square fluctuation (RMSF) of RNA structure**.** The residue that has less than 0.4 nm RMSF magnitude is not included in the heatmap (White), and the residues with higher RMSF values are encoded using a continuous color gradient.

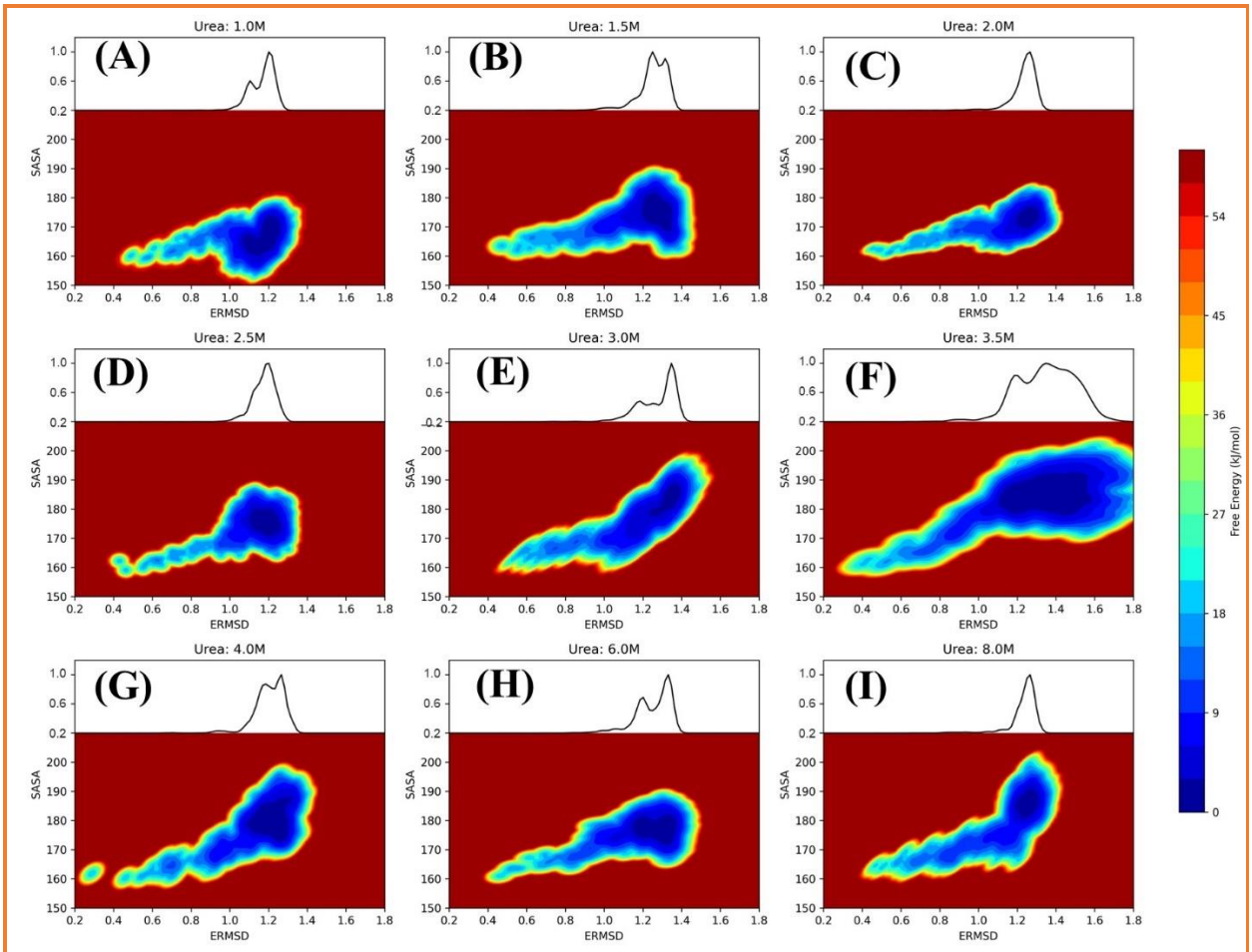


Figure S12. The two-dimensional free-energy surfaces projected onto the ensemble RMSD (eRMSD) and solvent-accessible surface area (SASA) coordinates at increasing urea concentrations.

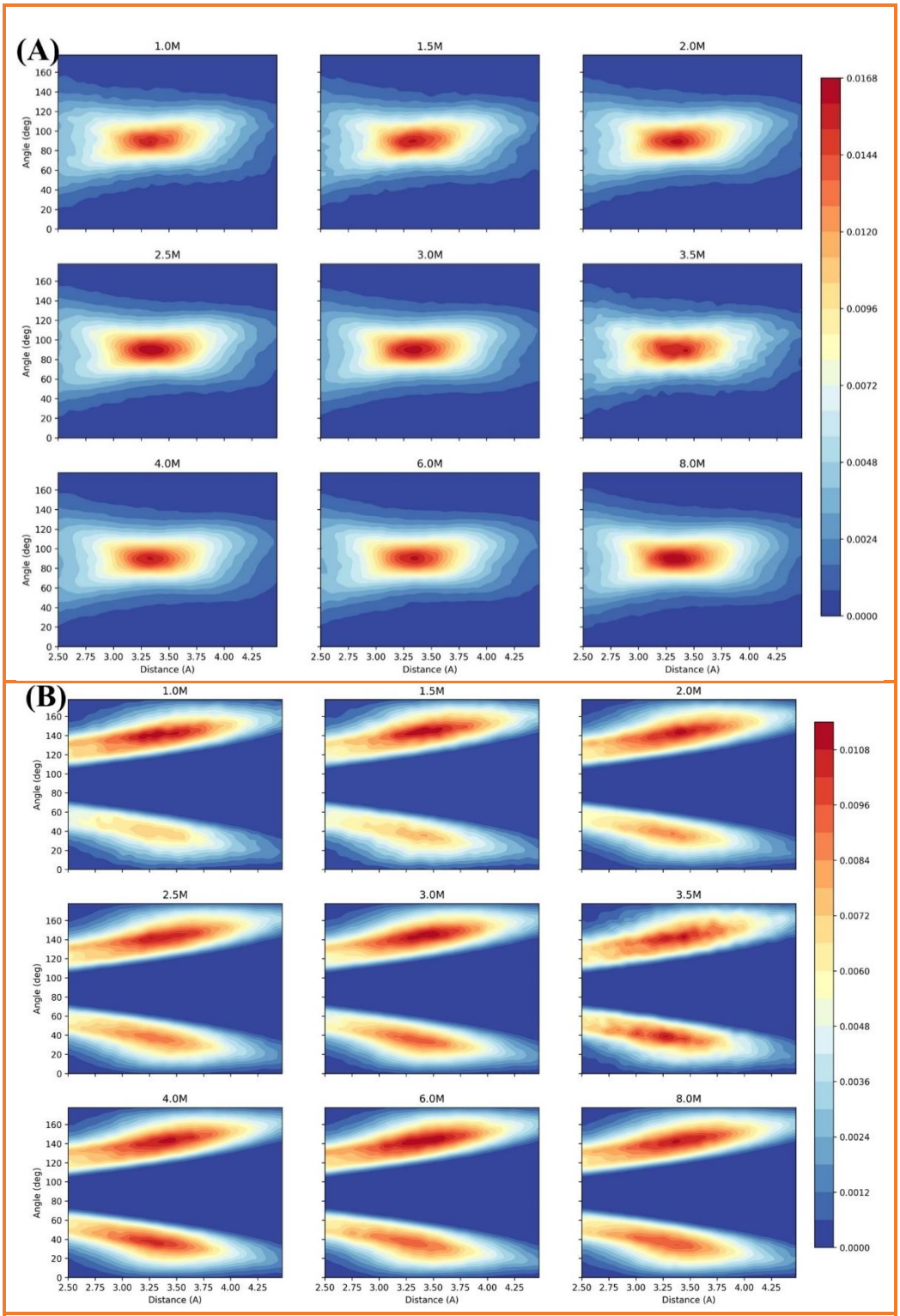


Figure S13. (A-B) Schematic representation of angle and distance definitions for quantifying π-π interactions and NH-π interactions.

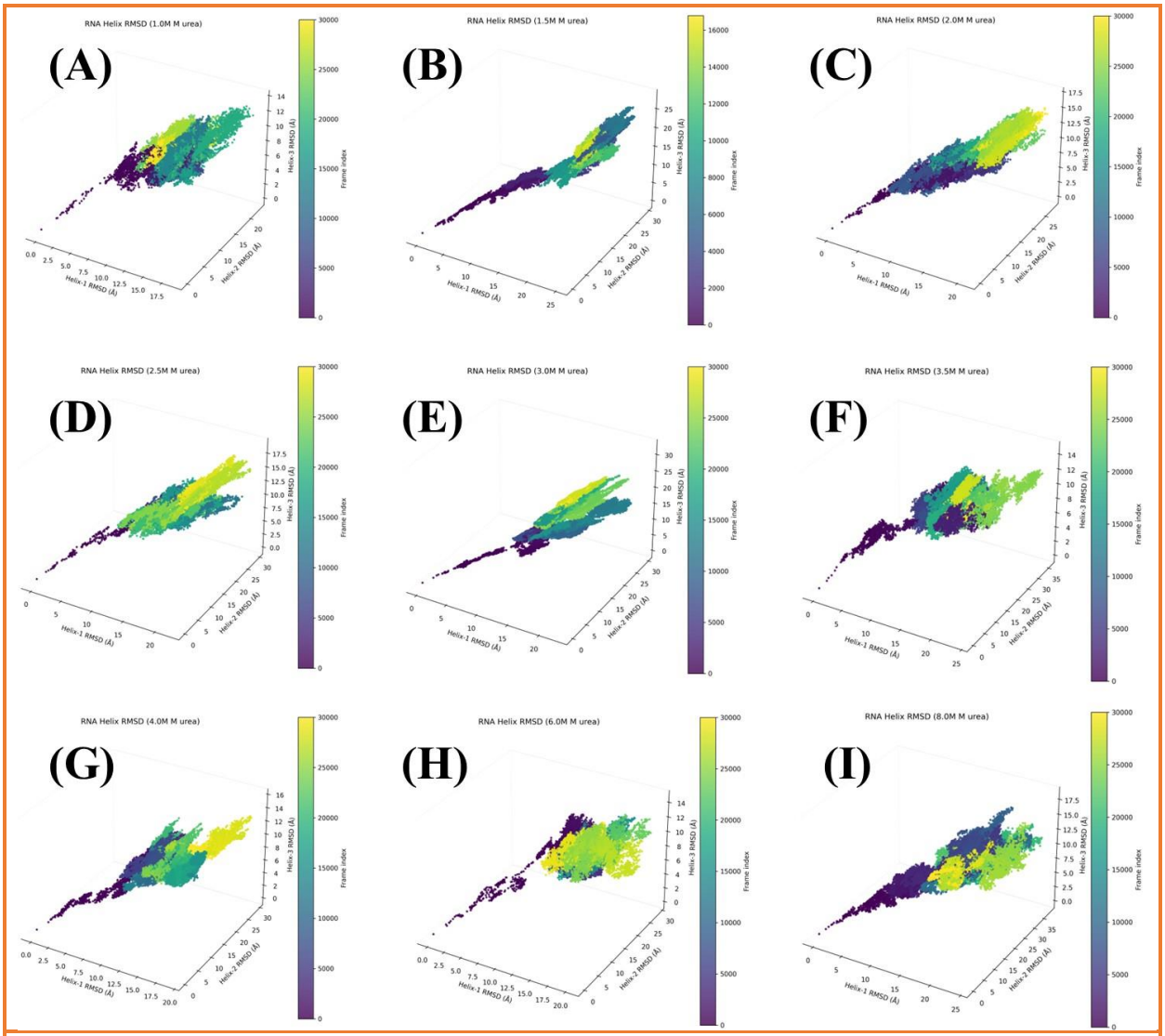


Figure S14. 3D plot illustrates the three-dimensional RMSD distributions of individual helical segments as a function of urea concentration. The color bar represents the time frame, and based on that, characterized by low and higher RMSD values across all helices, indicating the conformational topology. The cartoon surface structure represents the max helical RMSD fluctuation from the reference structure.

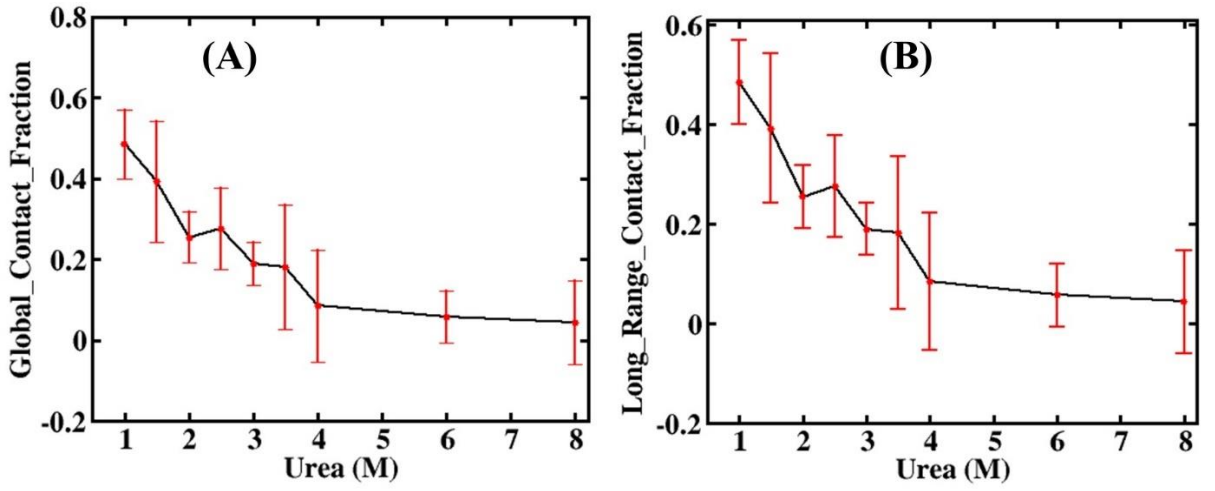


Figure S15. (A) shows the global contact order (GOC) and (B) Long-range fraction to distinguish unfolded conformers.

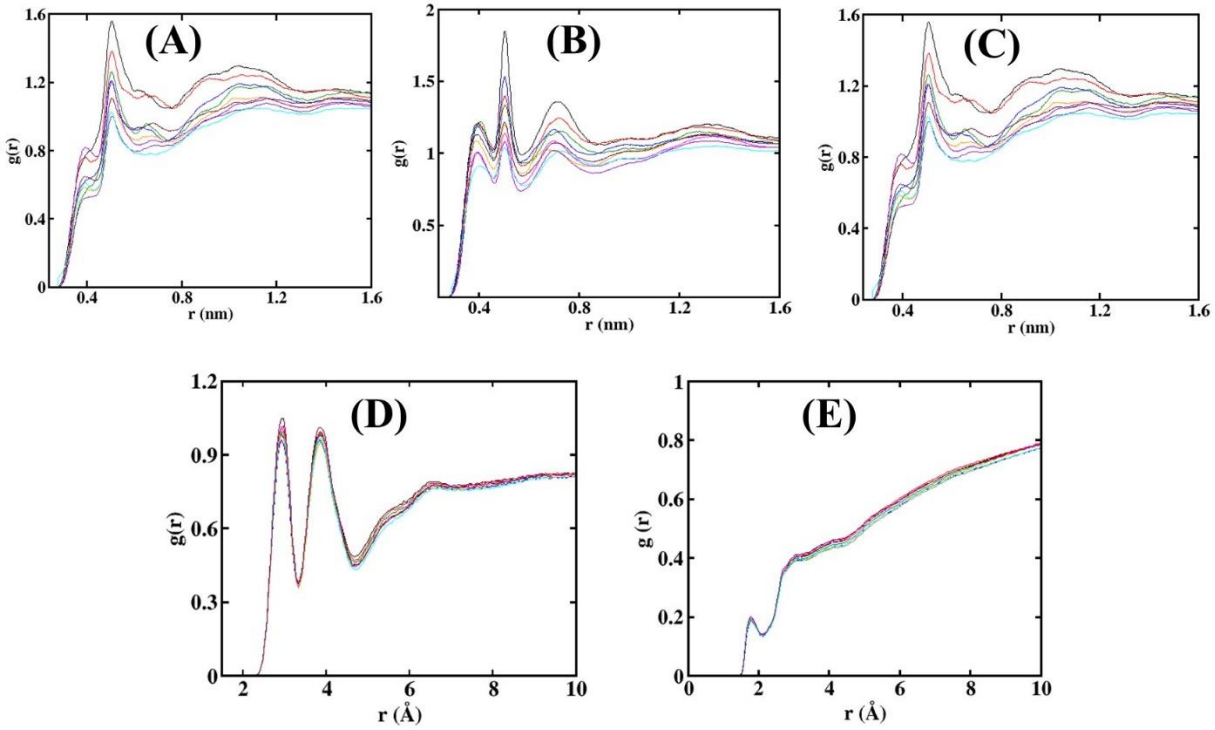


Figure S16. (A-E) The radial distribution function g(r) of urea with nucleobase in (1.0-8.0 M) concentration of urea

Section 5

AMBER99SB-ILDN ff: Results

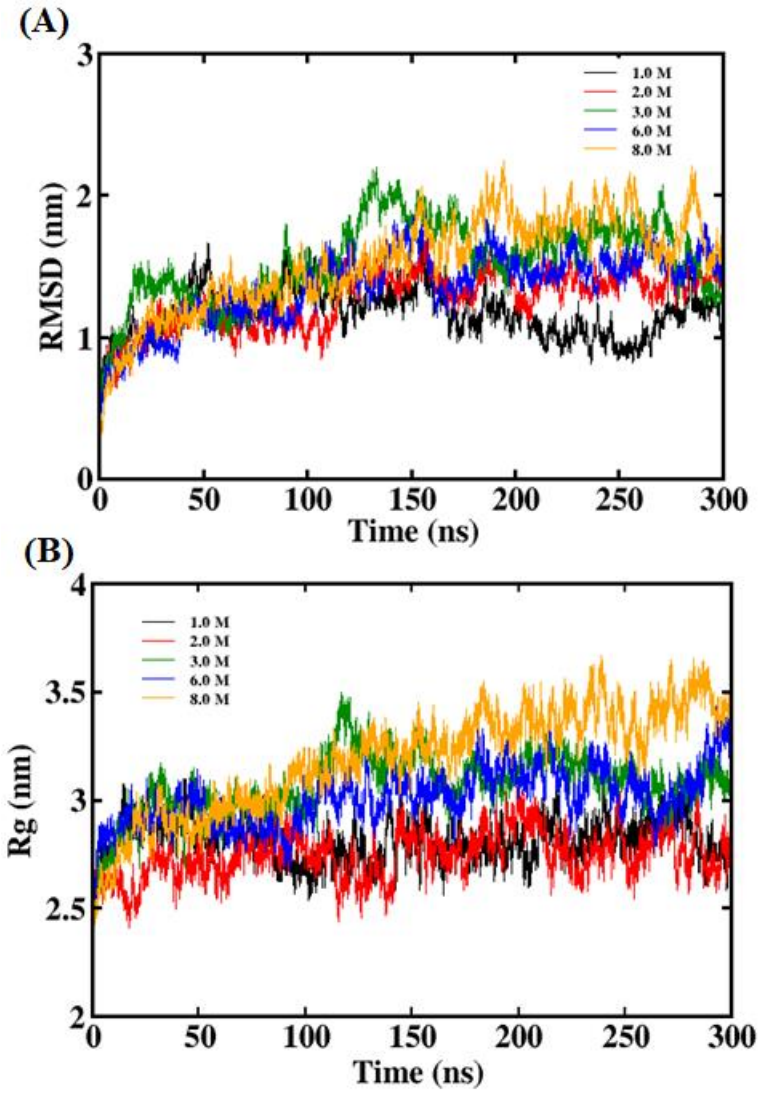


Figure S17. (A) Radius of gyration (Rg) and (B) root mean square deviation (RMSD) time evaluation graph.

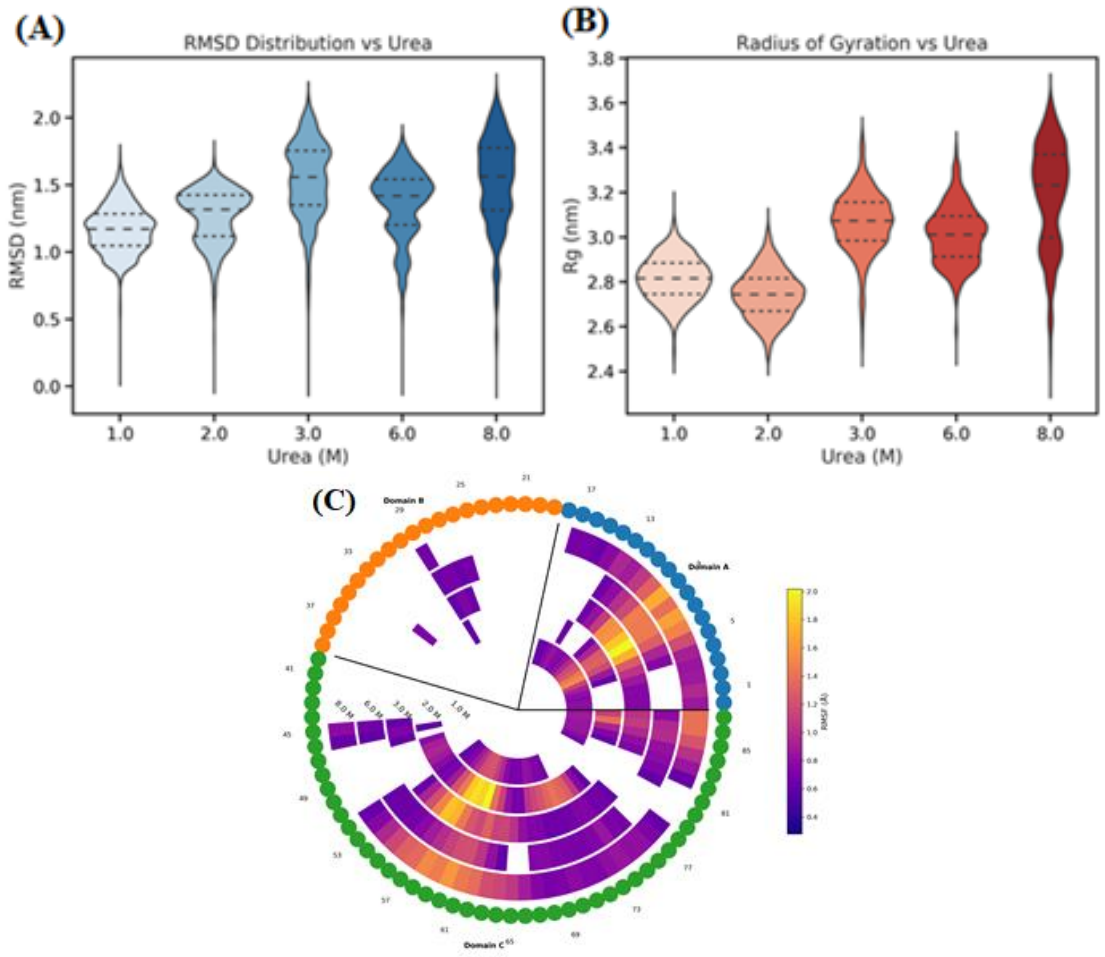


Figure S18. The upper panel presents violin distribution plots of the **(A)** root-mean-square deviation (RMSD) and **(B)** radius of gyration (Rg) of the RNA as a function of urea concentration (0–8 M). The violin distribution represents the central tendency and the full conformational distribution sampled during the simulations. **(C)** the circular heat map shows the root mean square fluctuation (RMSF) of RNA structure**.** The residue that has less than 0.4 nm RMSF magnitude is not included in the heatmap (White), and the residues with higher RMSF values are encoded using a continuous color gradient.

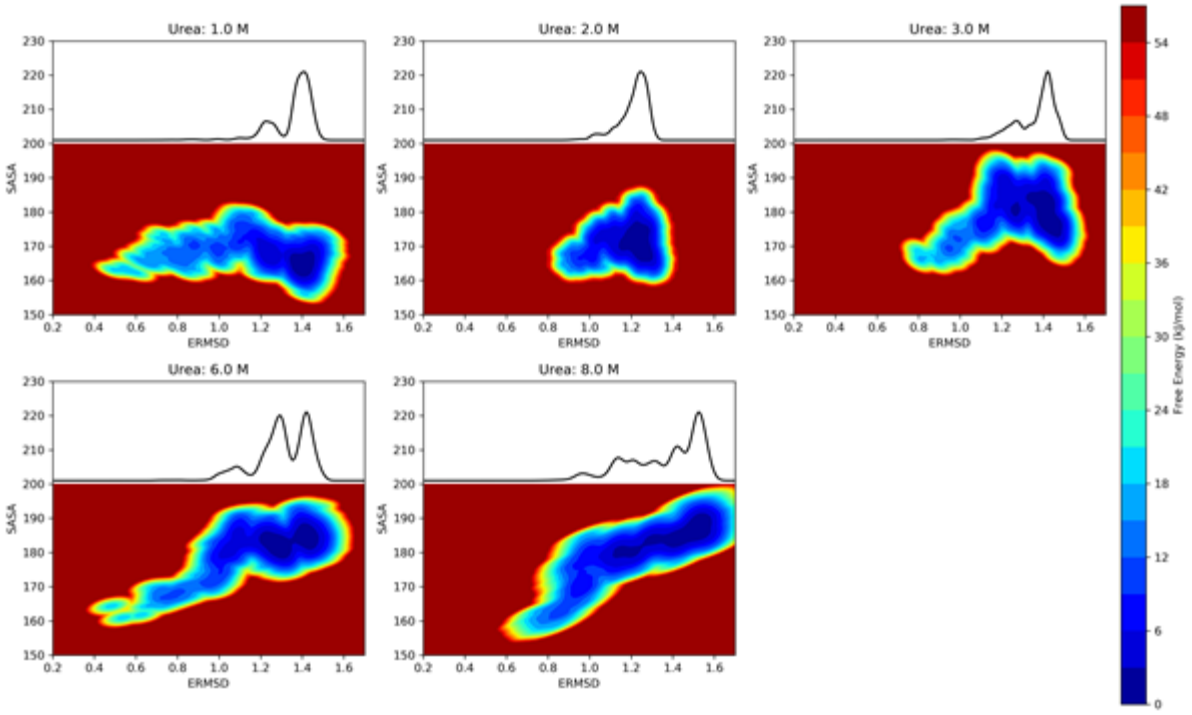


Figure S19. The two-dimensional free-energy surfaces projected onto the ensemble RMSD (eRMSD) and solvent-accessible surface area (SASA) coordinates at increasing urea concentrations.

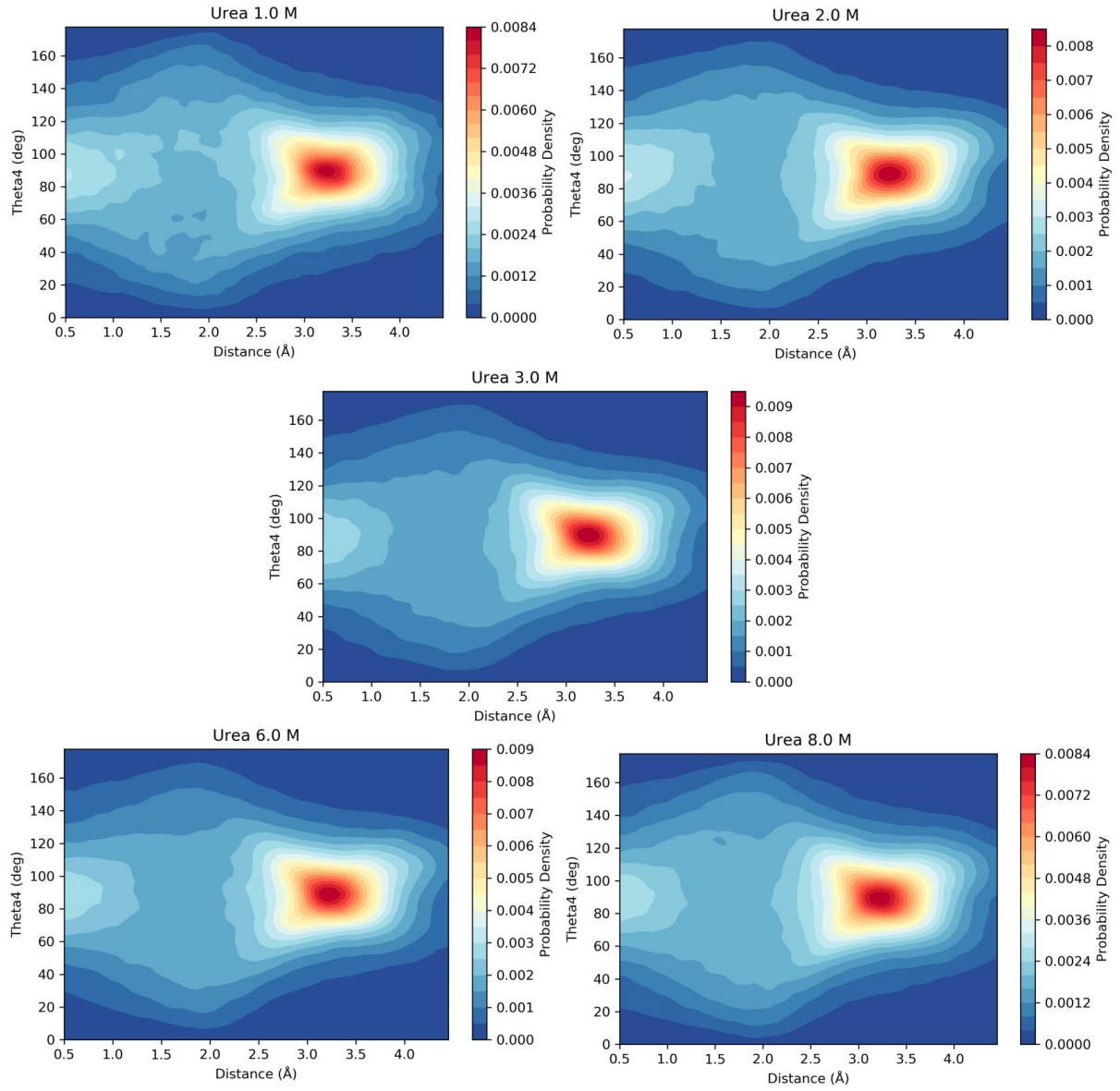


Figure S20. Schematic representation of angle and distance definitions for quantifying π-π interactions

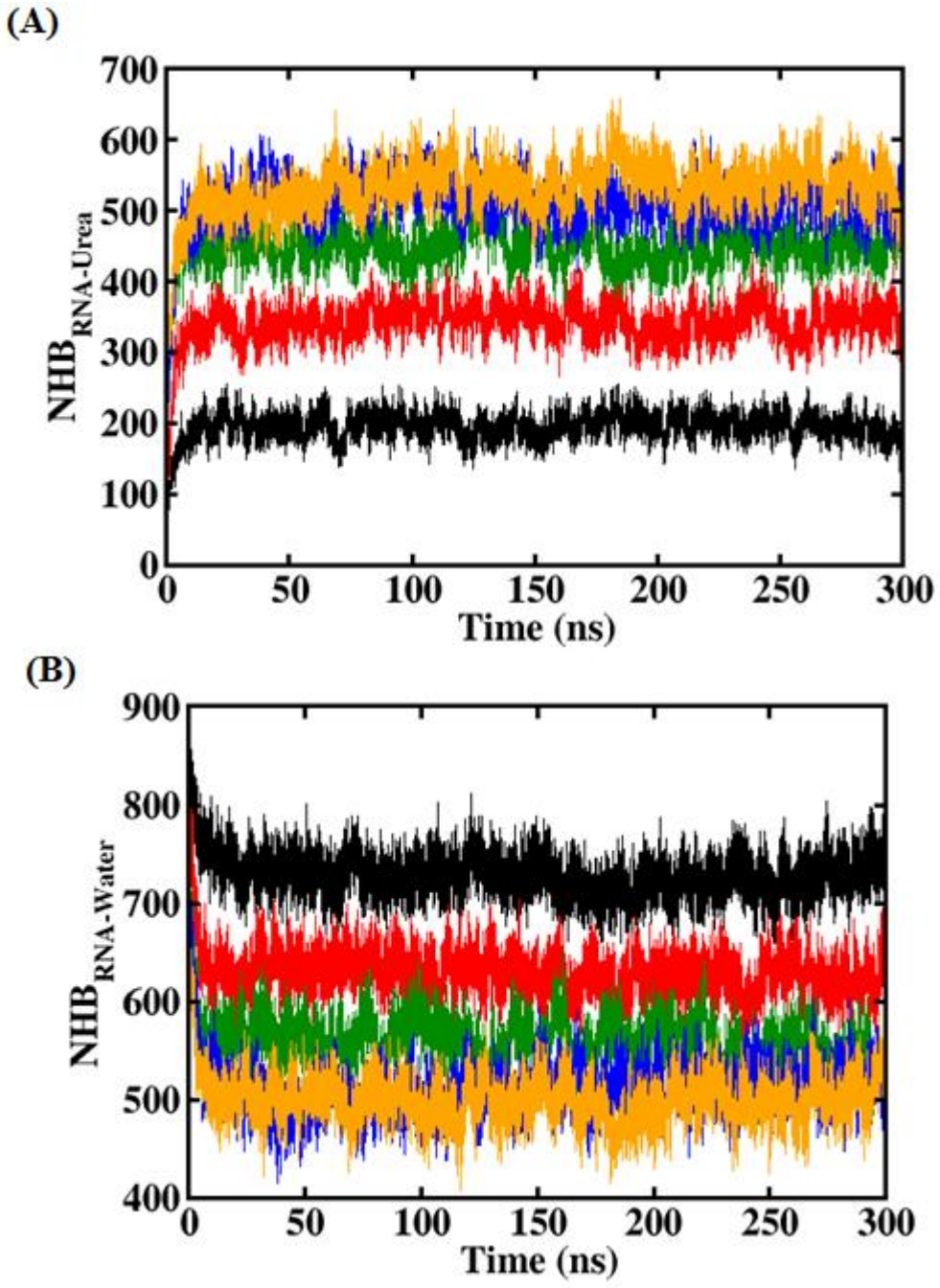


Figure S21. Hydrogen bond evolution (A) RNA- Urea (B) RNA- Water

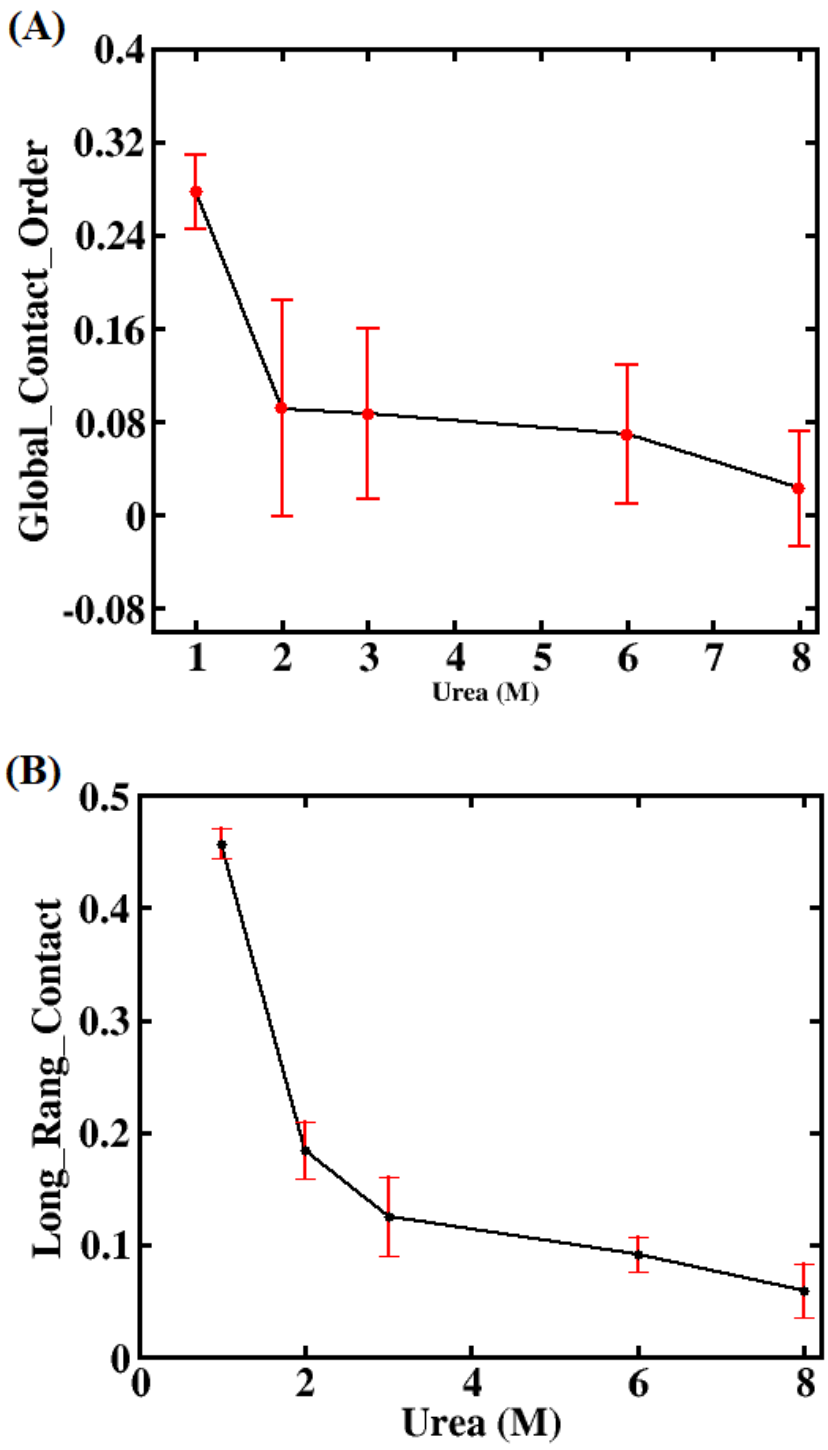
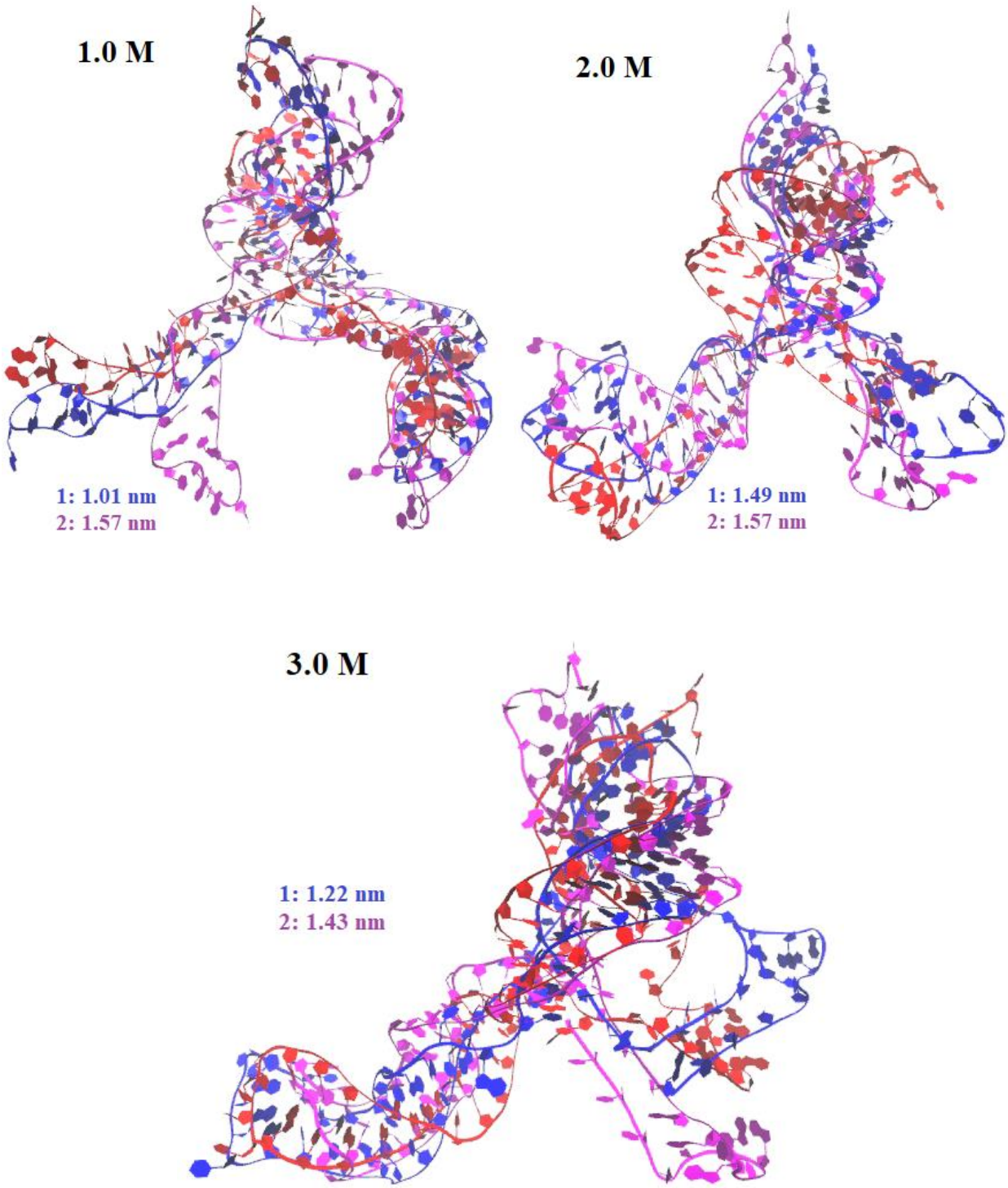


Figure S22. (A) shows the global contact order (GOC) and (B) Long-range fraction to distinguish unfolded conformers.


1.0 M
1: 1.01 nm
2: 1.57 nm
2.0 M
1: 1.49 nm
2: 1.57 nm
3.0 M
1: 1.22 nm
2: 1.43 nm

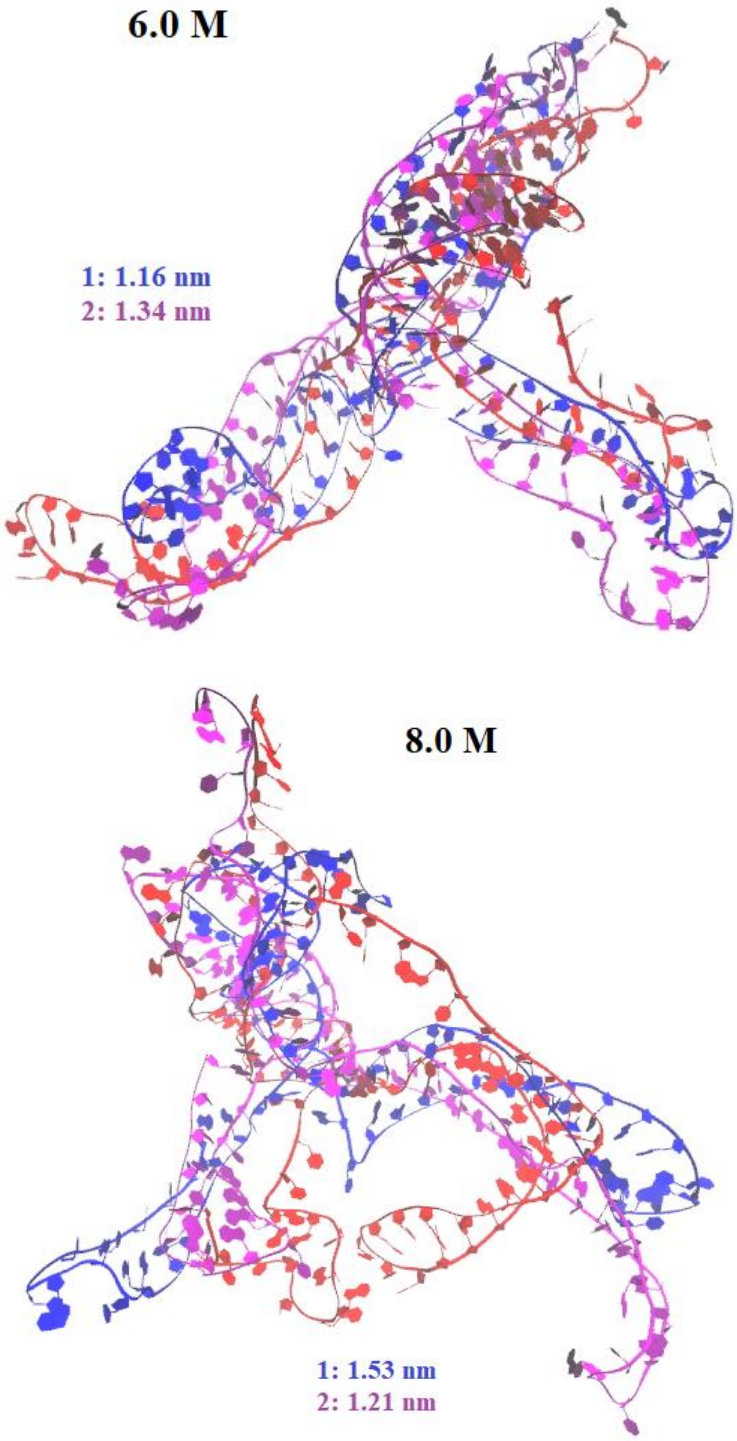


Figure S23. Overlay of structures obtained from CHARMM36 ff (245 ns: red) and AMBER99SB-ILDN ff (150 ns: blue; 300 ns: purple). RMSD for each structure obtained from AMBER99SB-ILDN ff with respect to CHARMM36 ff is provided.