Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Interaction Preferences between Nucleobase Mimetics and Amino Acids in Aqueous Solutions

Matea Hajnic, Juan I. Osorio and Bojan Zagrovic*

Department of Structural and Computational Biology, Max F. Perutz Laboratories,

University of Vienna, Campus Vienna Biocenter 5, Vienna 1030, Austria

Supporting Information

INDEX

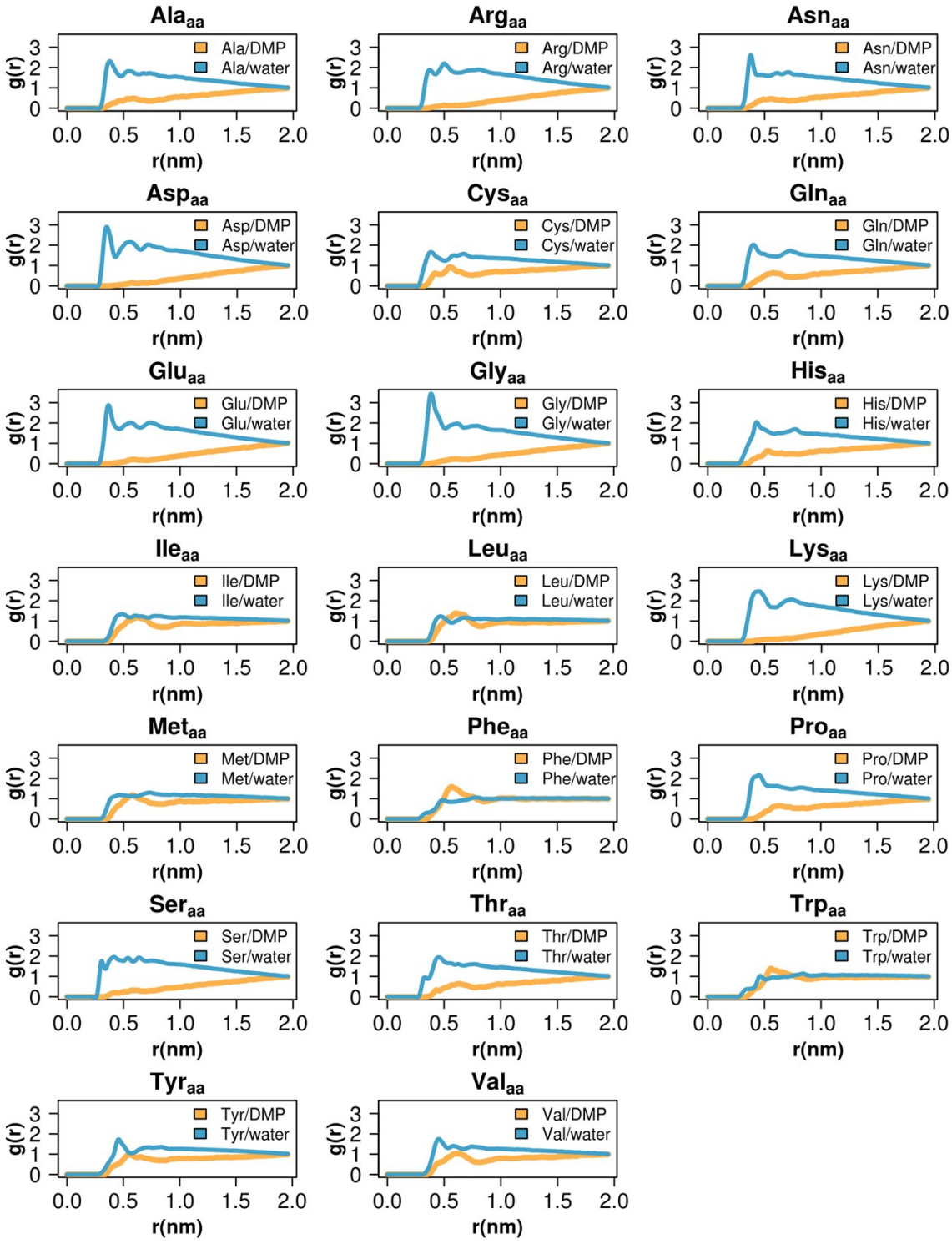

Figure S1. Radial distribution functions g(r) of amino acids in DMP-water mixtures. RDFs are calculated between center of masses of amino-acid sidechain analogs and DMP or water molecules.

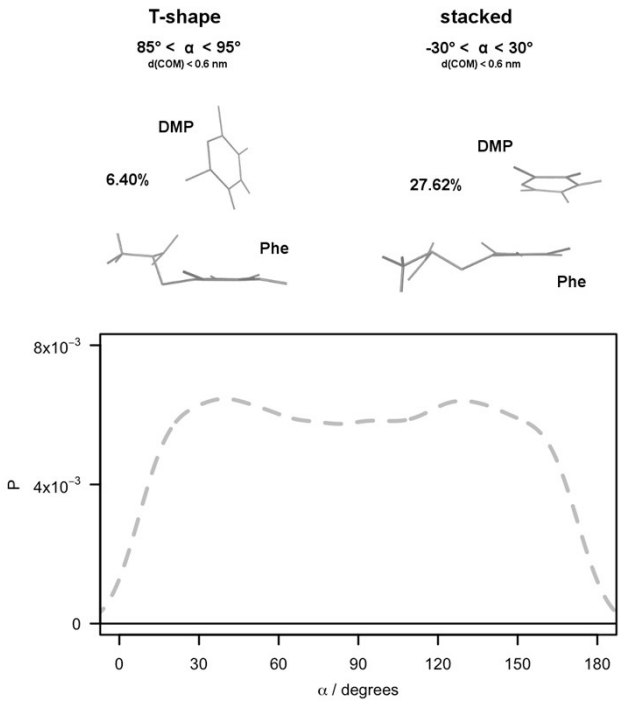

Figure S2. Distribution of angles between the phenylalanine (Phe) and DMP in aqueous solution. The angle between Phe and DMP is defined as the angle between a vector perpendicular to the phenylalanine ring plane (the normal vector) and the normal vector of the DMP ring plane. Representative examples of the "T-shape" (upper-left panel) and the stacked geometry (upper-right panel) between Phe sidechain analog and DMP together with the corresponding distribution (lower panel) of the inter-plane angles calculated for all pairs with centers of masses separated by less than 0.6 nm. All Phe-DMP angle values were derived from a 100 ns long MD simulation.

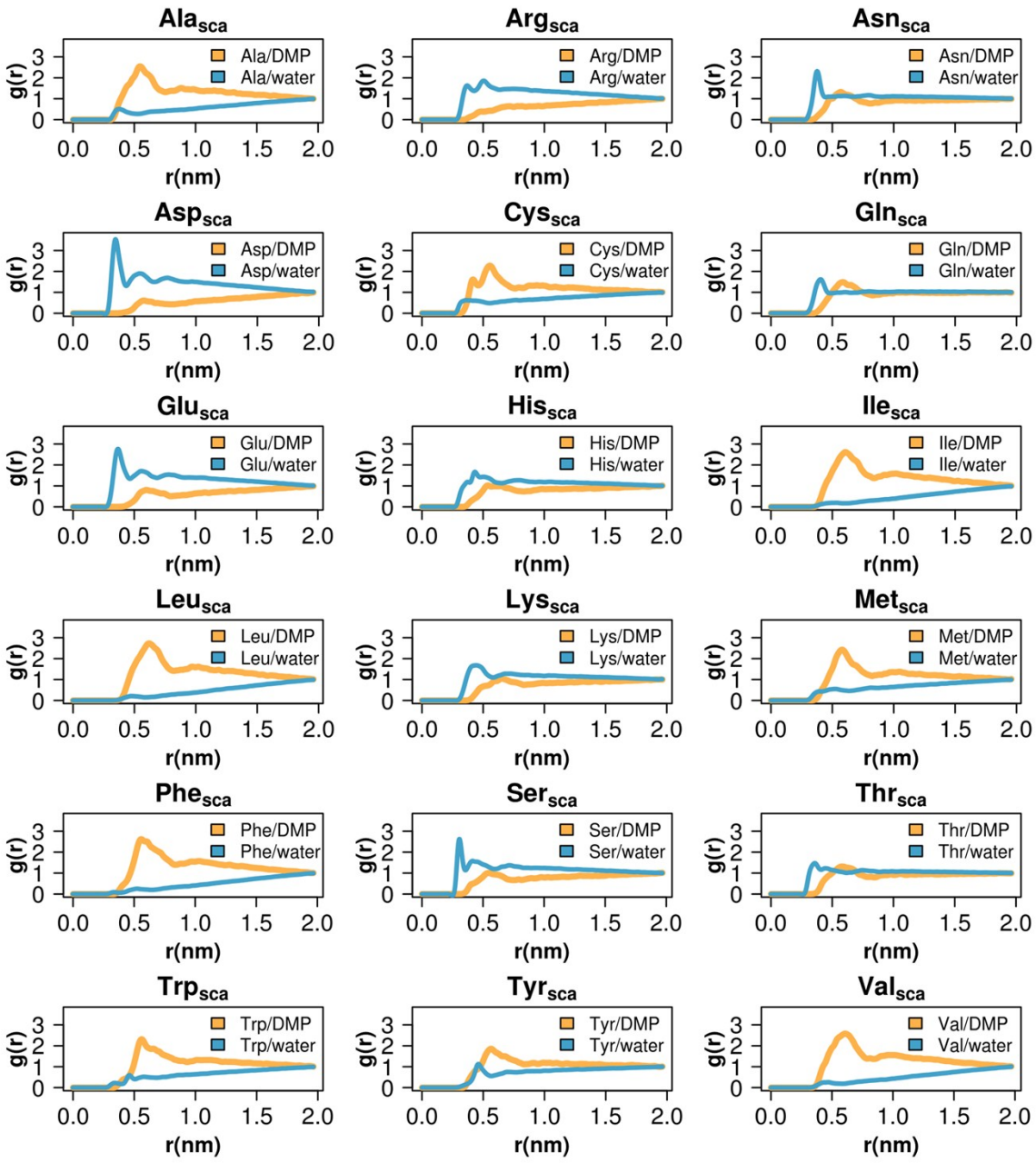

Figure S3. Radial distribution functions g(r) of amino-acid sidechain analogs in DMP-water mixtures. RDFs are calculated between center of masses of sidechain analogs and DMP or water molecules.

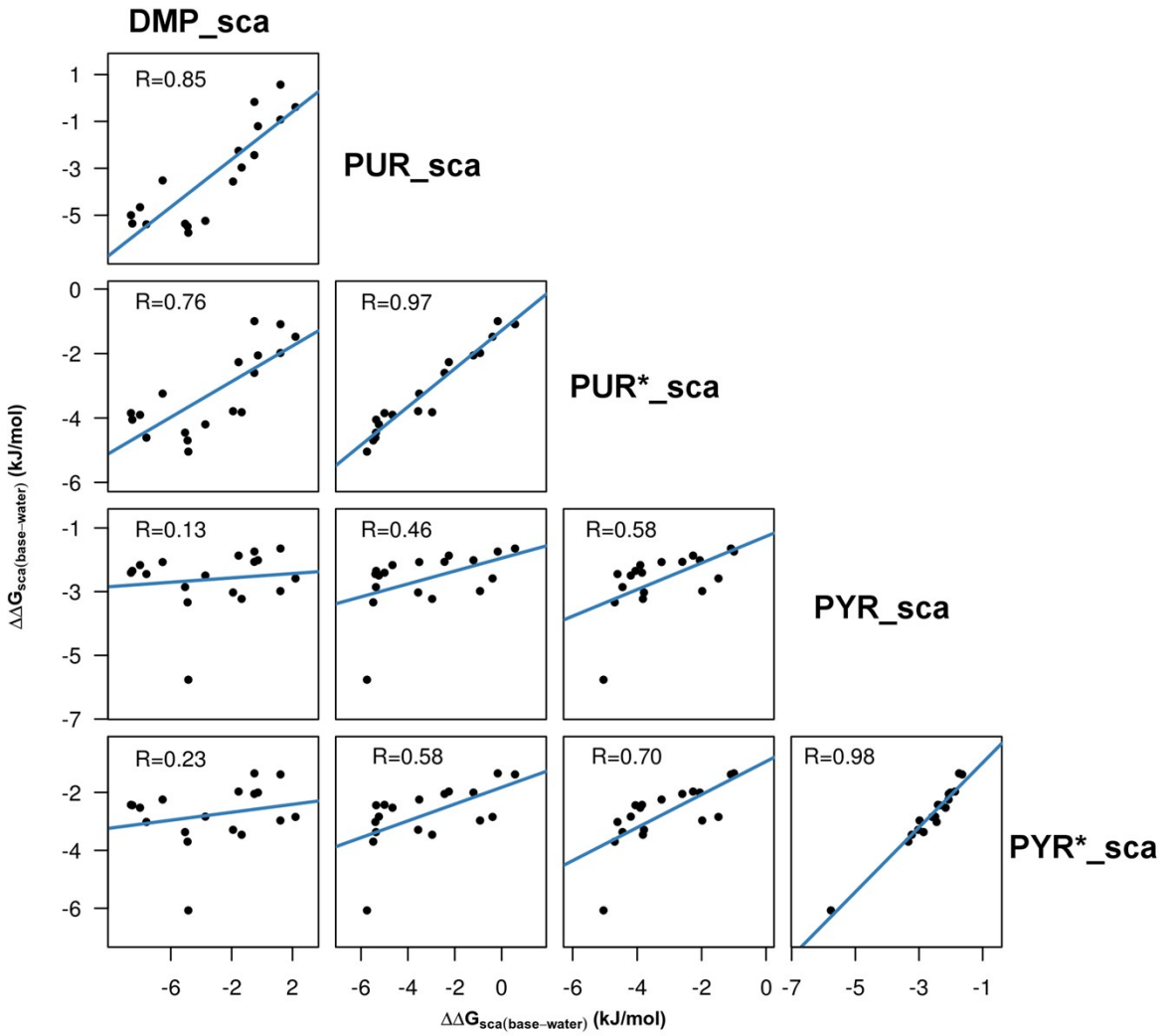

Figure S4. Pearson correlation coefficients between the relative free energy scales derived from MD simulations for systems with amino-acid sidechain analogs. *Amino-acid sidechain analog-nitrogenous base free energy scales derived from mixed systems, where both PUR^{u} and PYR^{u} molecules were present at the same time.

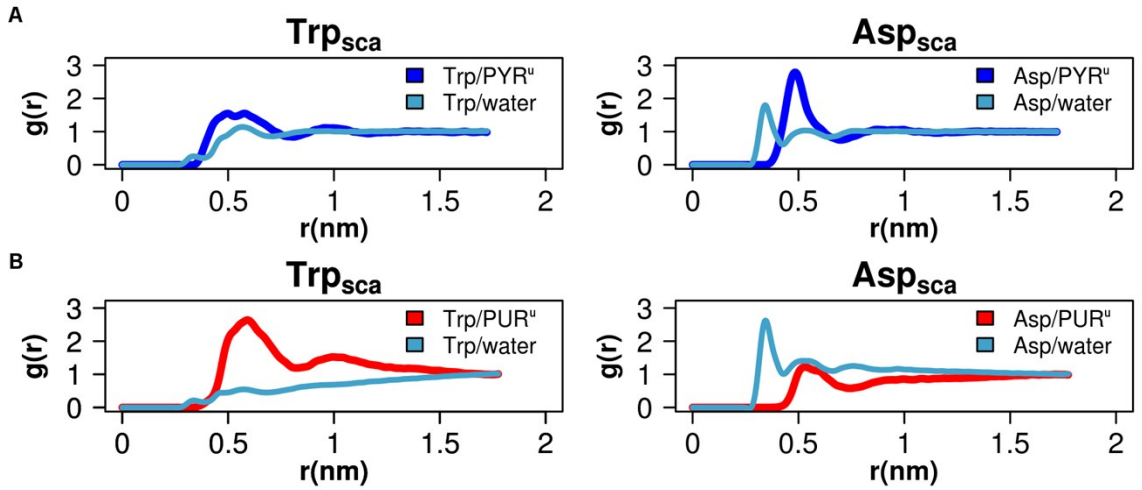

Figure S5. Radial distribution functions g(r) of amino-acid sidechain analogs. (**A**) Radial distribution function g(r) showing higher probability of finding PYR^{u} molecules close to amino-acid sidechain analog tryptophan (Trp_{sca}) than to water, and *vice versa* for aspartate (Asp_{sca}). (**B**) Radial distribution function g(r) for PUR^{u} and water for the same sidechain analogs.

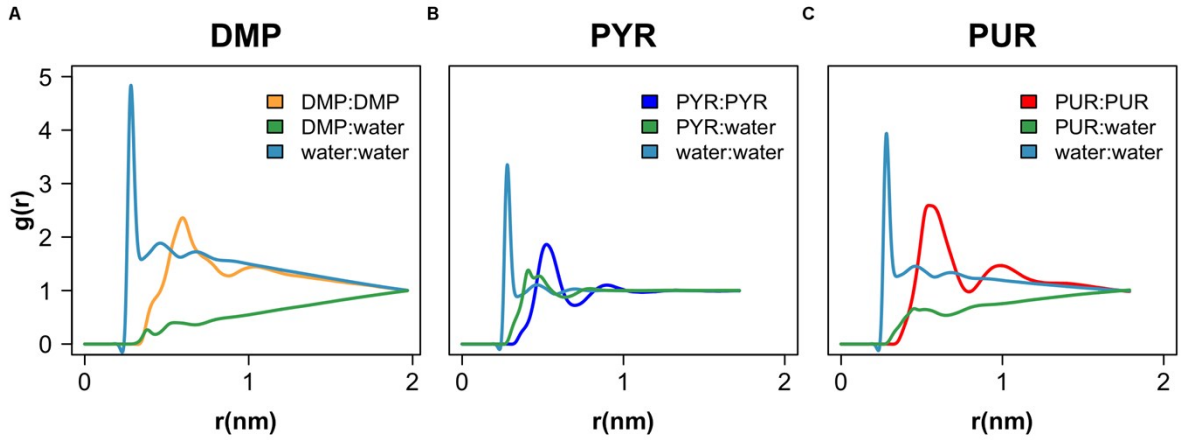

Figure S6. Radial distribution functions g(r) of nitrogenous base-nitrogenous base, nitrogenous base-water and water-water molecules in (**A**) DMP-water, (**B**) PYR^{u}-water and (**C**) PUR^{u}-water mixtures.

Table S1. Pearson correlation coefficients between relative free energy scales derived from MD simulations for systems with sidechain analogs (black) or amino acids (grey values in parenthesis).

	$\Delta\Delta G_{DMP\text{-}water}$	$\Delta\Delta G_{PYR^u\text{-}water}$	$\Delta\Delta G^*_{PYR^u\text{-}water}$	$\Delta\Delta G_{PUR^u\text{-}water}$	$\Delta\Delta G^*_{PUR^u\text{-}water}$
$\Delta\Delta G_{DMP\text{-}water}$	1.00 (1.00)				
$\Delta\Delta G_{PYR^u\text{-}water}$	0.13 (0.36)	1.00 (1.00)			
$\Delta\Delta G^*_{PYR^u\text{-}water}$	0.23 (0.47)	0.98 (0.94)	1.00 (1.00)		
$\Delta\Delta G_{PUR^u\text{-}water}$	0.85 (0.84)	0.46 (0.58)	0.58 (0.70)	1.00 (1.00)	
$\Delta\Delta G^*_{PUR^u\text{-}water}$	0.76 (0.65)	0.58 (0.61)	0.70 (0.76)	0.97 (0.87)	1.00 (1.00)

*sidechain analog and amino acid-nitrogenous base free energy scales derived from mixed systems, where both PUR^u and PYR^u molecules were present at the same time

Table S2. Spearman correlation coefficients for relative free energy scales and contact coefficient scales derived from MD simulations. In the first column, values of correlation coefficients stand for scales derived from systems with sidechain analogs, while in the second one are from systems with amino acids instead.

scale	ρ_{sca} **(ΔΔG *vs* CC)**	ρ_{aa} **(ΔΔG *vs* CC)**
DMP	-0.97	-0.90
PYR^{u}	-0.74	-0.68
PYR^{u*}	-0.90	-0.66
PUR^{u}	-0.99	-0.93
PUR^{u*}	-0.99	-0.94

*sidechain analog and amino acid-nitrogenous base free energy scales derived from mixed systems, where both PUR^{u} and PYR^{u} molecules were present at the same time

Table S3. Spearman correlation coefficients between amino-acid sidechain analog affinities for nucleobase mimetics (DMP, PYR^u and PUR^u) derived in this study and amino-acid sidechain analog affinities for the four RNA nucleobases derived in a related study using MD simulations and umbrella sampling[1].

	ADE	**CYT**	**GUA**	**URA**
DMP	0.65	0.64	0.56	0.65
PYR	0.36	0.23	0.33	0.37
PYR^{u*}	0.44	0.33	0.38	0.45
PUR	0.82	0.73	0.70	0.82
PUR^{u*}	0.83	0.76	0.70	0.83

*sidechain analog and amino acid-nitrogenous base free energy scales derived from mixed systems, where both PUR^u and PYR^u molecules were present at the same time

Table S4. Composition of MD boxes for simulated systems. In all simulation boxes, there was 1 amino acid or amino-acid sidechain analog (N_{AA}). The number of nitrogenous bases (N_{NB}) and water molecules (N_W) differed depending on the system. In the case of systems where not all amino acids had the same number of water molecules, we report a range. The box content was the same for systems with amino-acid sidechain analogs.

System	N_{AA}	N_{NB}	N_W
DMP	1	165	962-971
PYR^u	1	140	859
PUR^u	1	140	859
PYR^u:PUR^u	1	79:61	859

Table S5. List of all relative interaction free energy scales, interaction propensity scales and contact coefficients for amino acids and their sidechain analogs derived in this study. Derived scales are given as a separate Microscoft Word document **scales.doc.**

REFERENCES

1 A. de Ruiter and B. Zagrovic, *Nucleic Acids Res.*, 2014, **43**, 708–718.