Electronic Supplementary Information

To

Decoding the Enigma of RNA-Protein Recognition: Quantum Chemical Insights into Arg-fork Motifs

Raman Jangra[1], Teagan Kukhta[2], John F. Trant[2,3,*] and Purshotam Sharma[1,2,*]

[1]Computational Biochemistry Laboratory, Department of Chemistry and Centre for Advanced Studies in Chemistry, Panjab University, Chandigarh, 160014, India.

[2] Department of Chemistry and Biochemistry, University of Windsor, 401 Sunset Ave. Windsor, ON, N9B 3P4, Canada

[3] We-Spark Health Institute, 401 Sunset Ave. Windsor ON, N9B 3P4, Canada

[4] Binary Star Research Services, LaSalle, ON, N9J 3X8, Canada

**Email: psharma@pu.ac.in, j.trant@uwindsor.ca*

Table of Contents

S1. Choice of DFT functional. For evaluating the choice of DFT functionals, the models were built by methyl-capping the Arg and the interacting terminal ribonucleotides involved in Arg forks, where only interacting portions of ribonucleotide were retained. Specifically, the interacting guanidinium group of arginine was retained, which was capped at *Nε* by replacing the delta (δ) carbon with a methyl group. The interacting phosphate was capped at both *O3′* and *O5′* by replacing *C3′* and *C5′* with methyl groups. However, whenever participating in an interaction with the Arg fork, the sugar moiety was kept along with its respective nucleobase; in these cases, the associated ribonucleoside was methyl capped at both the *O3′* and *O5′* positions. If the phosphate moiety is involved in Arg fork formation along with the sugar moiety, the sugar moiety was methyl capped only at *O3′* and the phosphate was capped by replacing the *C3′* of the ribonucleotide *5′* to the phosphate, with a methyl group. Furthermore, when only the nucleobase moiety of the ribonucleotide interacts with Arg, either through hydrogen bonding (in the case of guanine) or through cation-π interaction, the associated sugar-phosphate group was removed, and the nucleobase was capped at the glycosidic nitrogen (*N9* for purines and *N1* for pyrimidines), by replacing the *C1′* with a methyl group. In the cases where the phosphate moiety of the interacting nucleobase is also involved in Arg fork formation, the entire ribonucleotide was retained in the computational model, and the ribonucleotide was methyl capped at both its *O3′* and the phosphate oxygen (*i.e.*, *O3′* of the ribonucleotide *5′* to the phosphate). Missing hydrogen atoms in the PDB were added to the heavy atoms to fulfill the requisite covalent bonding requirements. Finally, the total charge on the Arg fork was calculated according to the number of phosphate groups (–2 for each) and the guanidinium moiety of Arg (+1). These models are smaller than the original motifs and truncated at many positions and were built to examine the efficiency of different density functional theory (DFT) functionals to optimize Arg forks.

For benchmarking, three functionals – the Minnesota functional (M06-2X);[1] Becke's three-parameter hybrid Functional (B3LYP) with the D3 version of Grimme's dispersion incorporating the Becke-Johnson damping function (GD3BJ);[2-5] and the long-range corrected functional (ω-B97XD) which includes Grimme's D2 dispersion[6, 7] were employed with the 6-31G(d,p) basis set in Gaussian 16.[8] Also, to consider the effect of bulk solvent (water) on the geometries and energetics of these models, the integral equation formalism variant of the Polarizable Continuum Model (IEFPCM) was used.[9]

Supplementary Figures and Tables

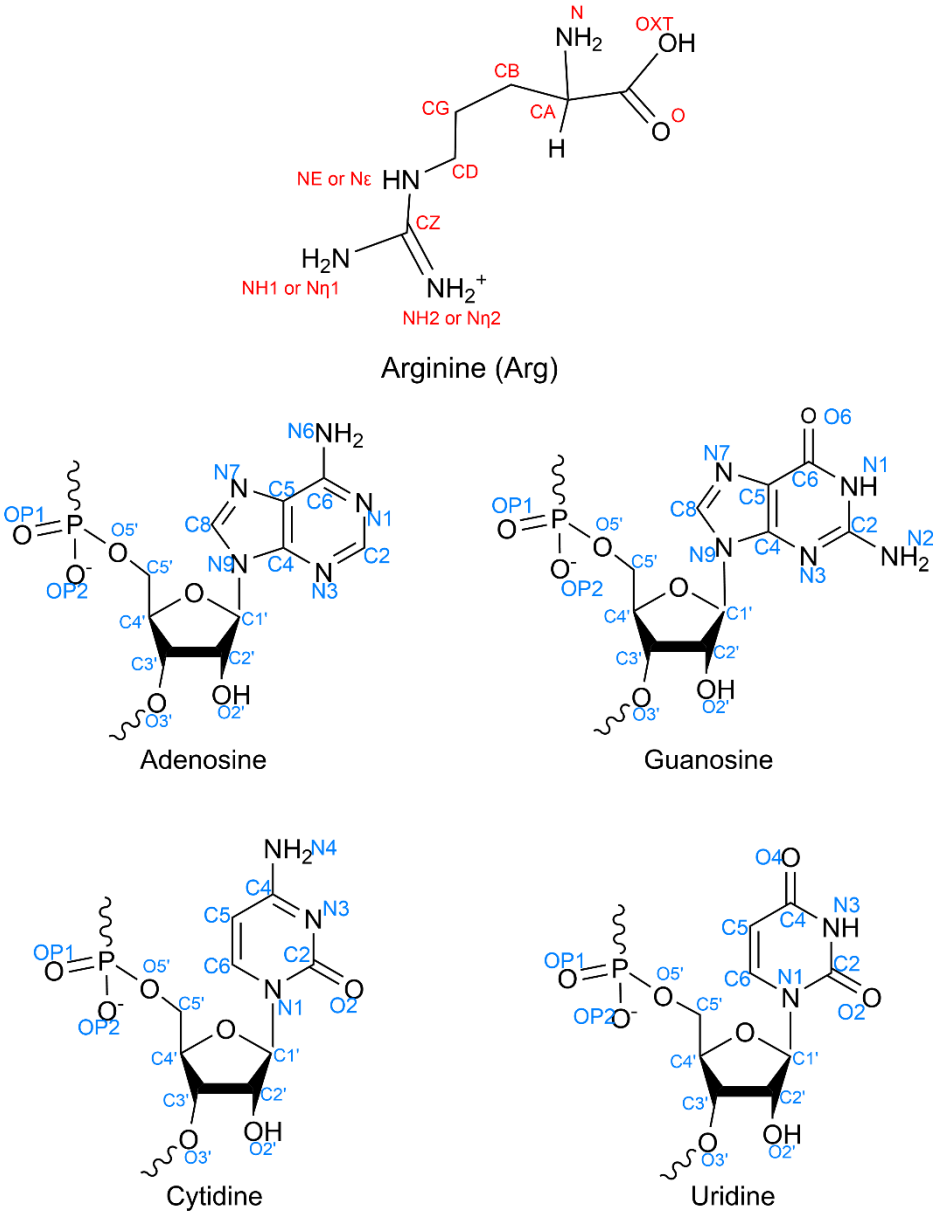

Figure S1. Structures and PDB atomic nomenclature of arginine and the ribonucleotides. Wavy lines in ribonucleotides represent the position of attachment of the next ribonucleotide in the RNA chain.

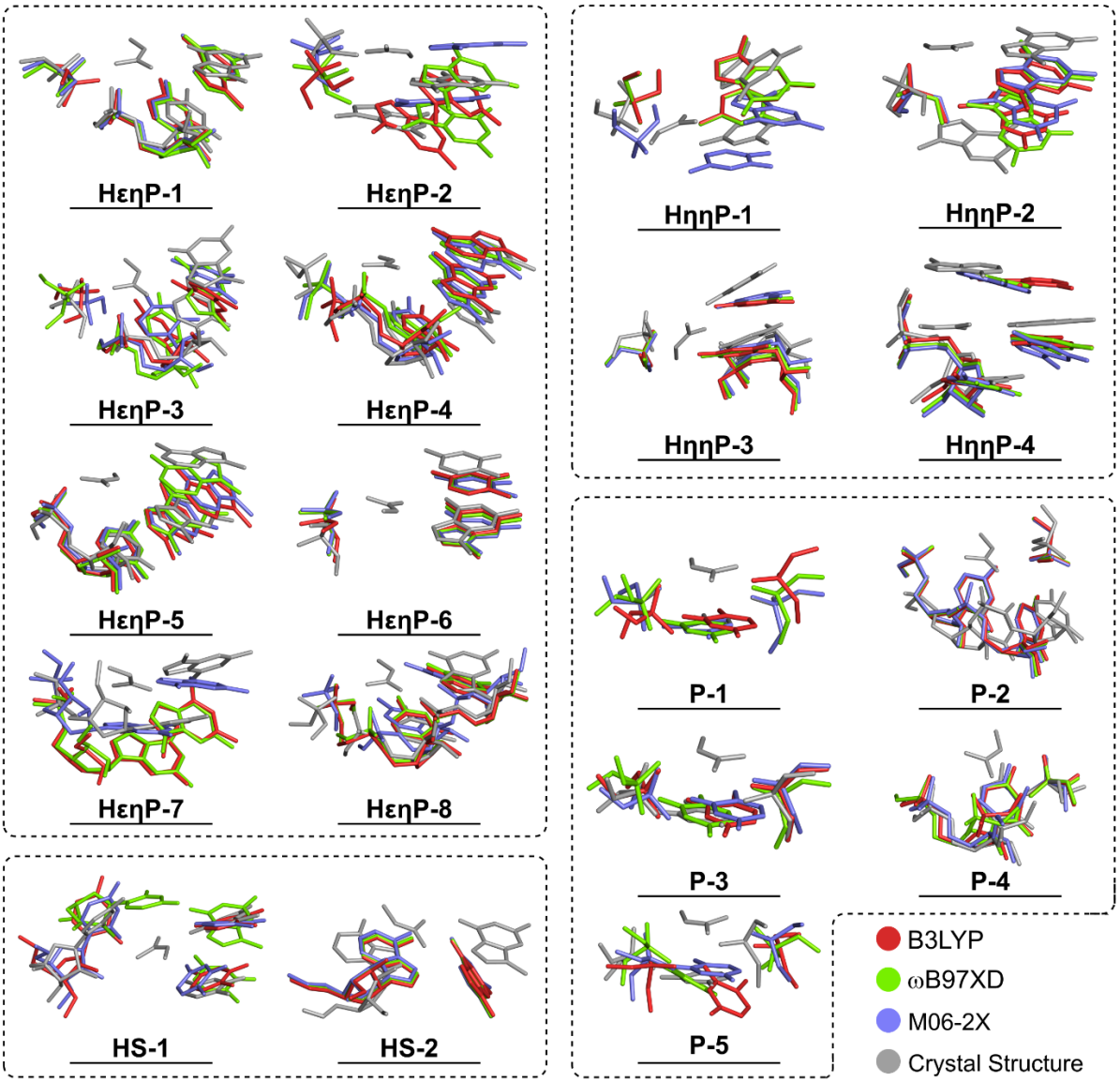

Figure S2. Superposition of the QM optimized reduced models of Arg forks on their corresponding crystallographic occurrences using PyMOL. The crystal structure occurrence of the Arg fork is shown by grey sticks, while the optimized Arg fork models using B3LYP, ωB97XD, and M06-2X methods are represented by red, green, and blue sticks respectively.

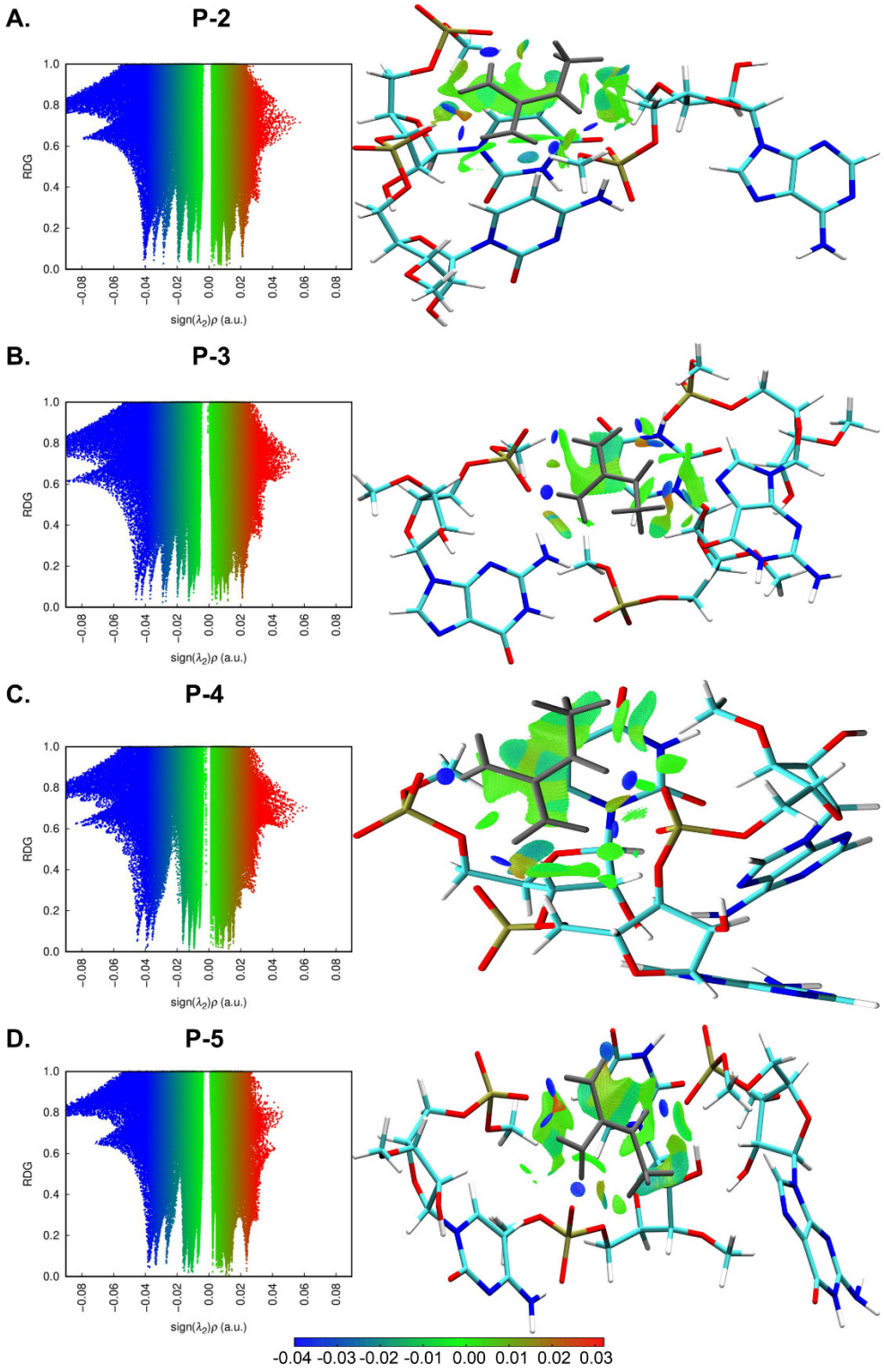

Figure S3. NCI analysis of the ω-B97XD optimized Type P Arg fork models. RDG scatter plots and isosurface plots, colored based on the values of sign(λ_2)ρ, are shown on the left and right, respectively, along with a color scale of sign(λ_2)ρ at the bottom.

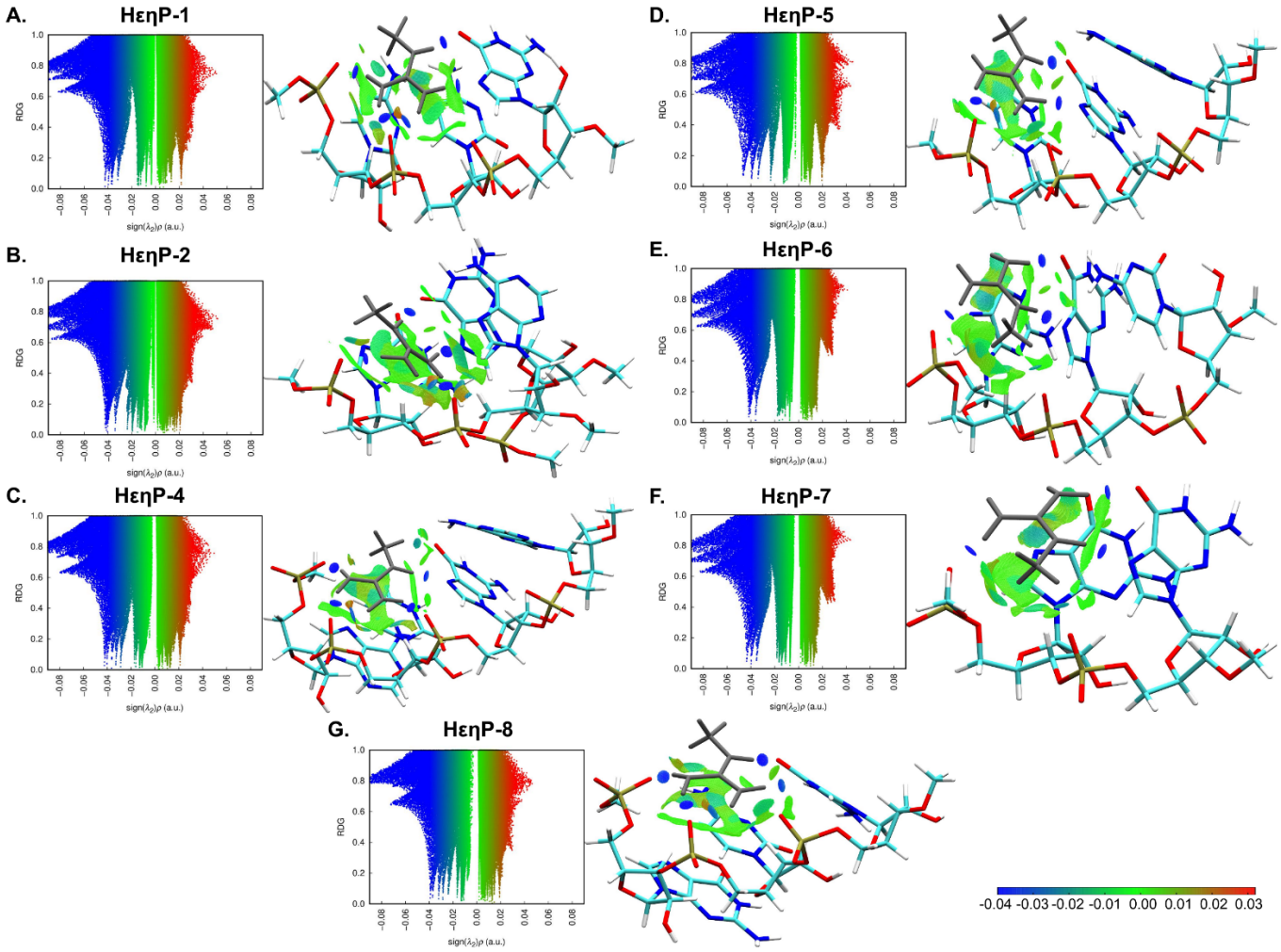

Figure S4. NCI analysis of the ω-B97XD optimized Type HεηP Arg fork models. RDG scatter plots and isosurface plots, colored based on the values of sign(λ_2)ρ, are shown on the left and right, respectively, along with a color scale of sign(λ_2)ρ at the bottom.

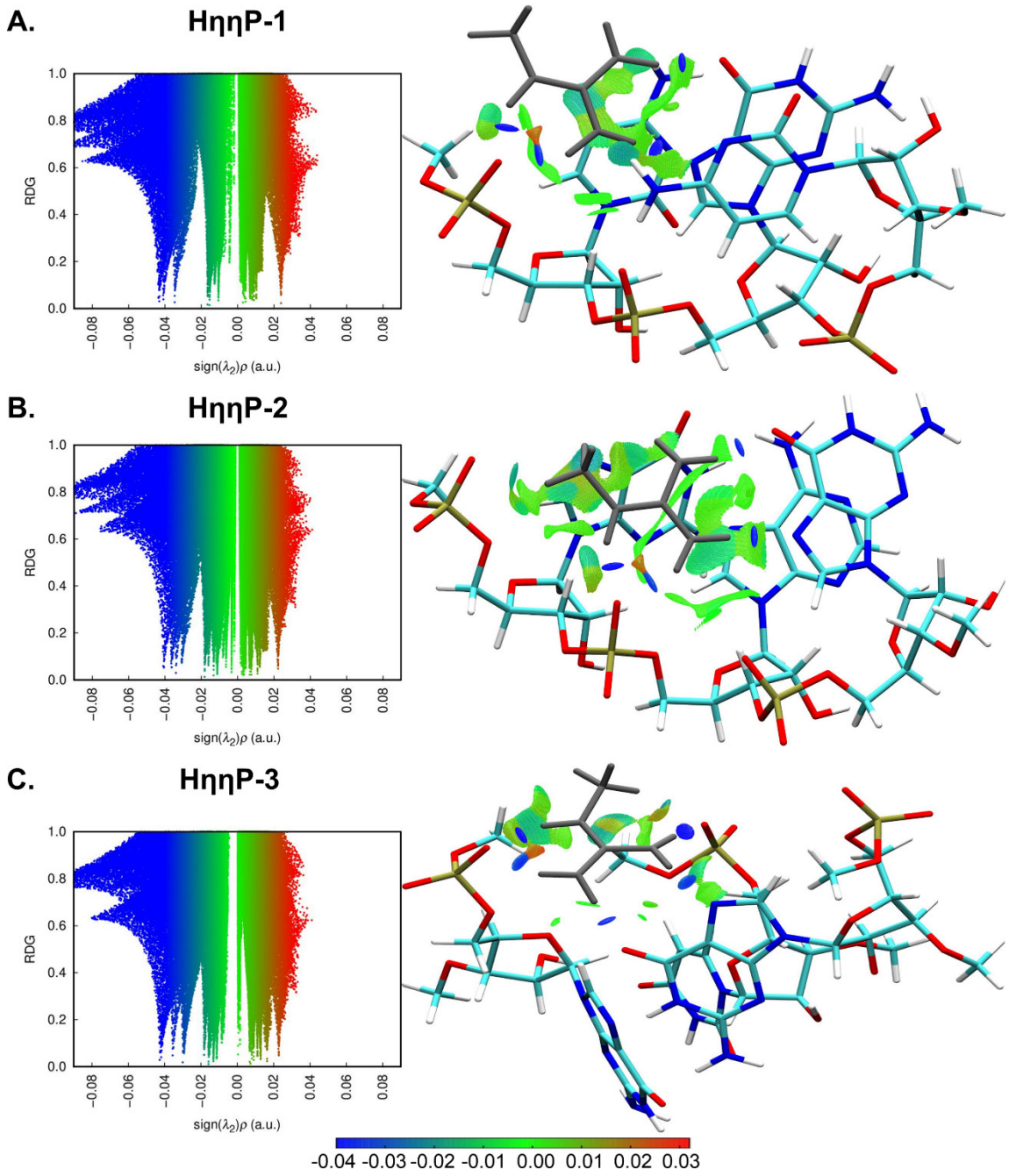

Figure S5. NCI analysis of the ω-B97XD optimized Type HηηP Arg fork models. RDG scatter plots and isosurface plots, colored based on the values of sign(λ_2)ρ, are shown on the left and right, respectively, along with a color scale of sign(λ_2)ρ at the bottom.

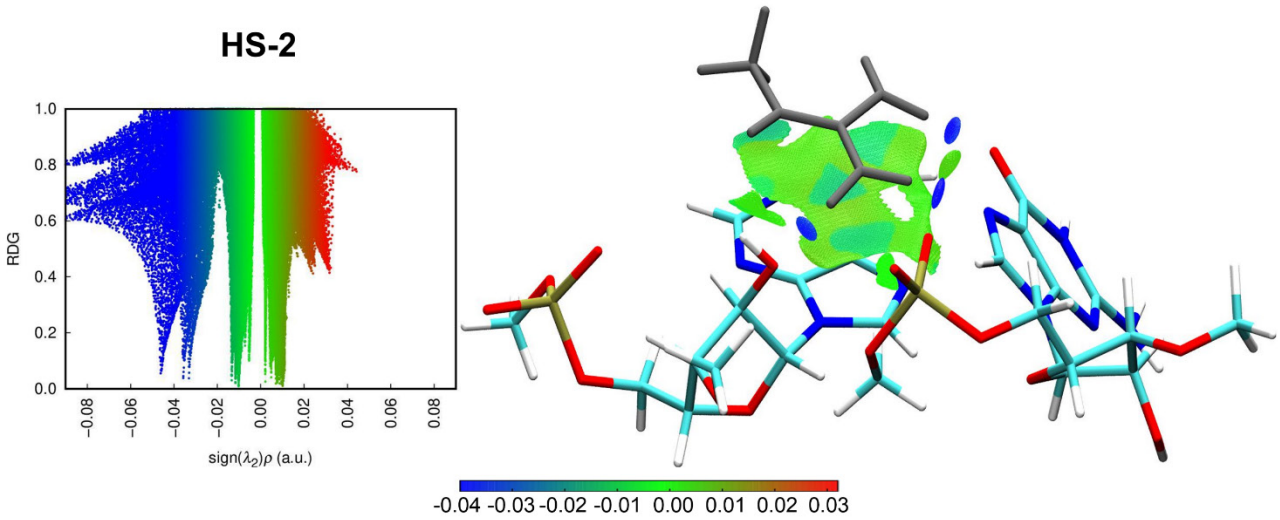

Figure S6. NCI analysis of the ω-B97XD optimized Type HS Arg fork models. RDG scatter plots and isosurface plots, colored based on the values of sign(λ_2)ρ, are shown on the left and right, respectively, along with a color scale of sign(λ_2)ρ at the bottom.

Table S1. Characteristics of distinct representatives of Type P Arg forks taken from Chavali *et al.* for this study.[10]

Nomenclature of the Arg fork	**PDB Code**	**Hydrogen-bonding interactions**	**Nucleobase(s) involved in cation-π interactions**
P-1	1JJ2	Arg(K)8 [Nε•••OP2] G(O)1354 Arg(K)8 [Nη1•••OP2] G(O)644 Arg(K)8 [Nη2•••OP1] G(O)644 Arg(K)8 [Nη2•••OP2] G(O)1354	U904
P-2	1JJ2	Arg(O)61 [Nε•••OP1] A(0)2739 Arg(O)61 [Nη1•••OP2] U(0)2736 Arg(O)61 [Nη1•••OP2] C(0)2737 Arg(O)61 [Nη2•••OP2] C(0)2737 Arg(O)61 [Nη2•••OP1] A(0)2739	U2736, C2737
P-3	1K8A	Arg(M)8 [Nε•••OP2] G(A)1354 Arg(M)8 [Nη1•••OP2] G(A)644 Arg(M)8 [Nη2•••OP1] G(A)644	U904
P-4	5AOX	Arg(B)59 [Nε•••OP1] A(C)28 Arg(B)59 [Nη1•••OP2] U(C)26 Arg(B)59 [Nη2•••OP2] A(C)27 Arg(B)59 [Nη2•••OP2] A(C)28	U26
P-5	5DM6	Arg(I)21 [Nε•••OP1] G(X)1250 Arg(I)21 [Nη1•••OP1] C(X)587 Arg(I)21 [Nη2•••OP1] C(X)587	U811

Table S2. Characteristics of distinct representatives of Type HεηP Arg forks taken from Chavali *et al.* for this study.[10]

Nomenclature of Arg fork	PDB Code	Hydrogen-bonding interactions	Nucleobase(s) involved in cation-π interactions
HεηP-1	1JJ2	Arg(C)182 [Nε•••O6] G(0)452 Arg(C)182 [Nη1•••OP1] C(0)450 Arg(C)182 [Nη1•••OP2] C(0)451 Arg(C)182 [Nη2•••OP2] C(0)451 Arg(C)182 [Nη2•••N7] G(0)452	C451
HεηP-2	1XOK	Arg(D)17 [Nε•••O6] G(B)880 Arg(D)17 [Nη1•••OP2] A(B)878 Arg(D)17 [Nη2•••N7] G(B)880 Arg(D)17 [Nη2•••OP2] A(B)878	G877
HεηP-3	2XLI	Arg(A)115 [Nε•••O6] G(A)11 Arg(A)115 [Nη1•••OP2] C(A)9 Arg(A)115 [Nη1•••OP2] C(A)10 Arg(A)115 [Nη2•••OP2] C(A)10 Arg(A)115 [Nη2•••N7] G(A)11	C10
HεηP-4	3NMU	Arg(B)334 [Nε•••O6] G(E)31 Arg(B)334 [Nη1•••OP2] A(E)29 Arg(B)334 [Nη1•••OP2] U(E)30 Arg(B)334 [Nη2•••N7] G(E)31	U30, A32
HεηP-5	3NVI	Arg(A)334 [Nε•••O6] G(E)22 Arg(A)334 [Nη1•••OP2] U(E)21 Arg(A)334 [Nη2•••OP2] U(E)21 Arg(A)334 [Nη2•••N7] G(E)22	U21, A23
HεηP-6	4L8R	Arg(C)181 [Nε•••N7] G(A)7 Arg(C)181 [Nη1•••OP1] G(A)6 Arg(C)181 [Nη2•••O6] G(A)7	C8
HεηP-7	4TUW	Arg(A)240 [Nε•••N7] G(C)9 Arg(A)240 [Nη1•••OP1] G(C)8 Arg(A)240 [Nη2•••O6] G(C)9	G8
HεηP-8	5DEA	Arg(B)10 [Nε•••O6] G(A)31 Arg(B)10 [Nη1•••OP2] G(A)29 Arg(B)10 [Nη1•••OP2] C(A)30 Arg(B)10 [Nη2•••N7] G(A)31 Arg(B)10 [Nη2•••OP2] G(A)31	C30

Table S3. Characteristics of distinct representatives of Type HηηP Arg forks taken from Chavali *et al.* for this study.[10]

Nomenclature of Arg fork	**PDB Code**	**Hydrogen-bonding interactions**	**Nucleobase(s) involved in cation-π interactions**
HηηP-1	1C0A	Arg(A)222 [Nε•••OP2] C(B)672 Arg(A)222 [Nη1•••O6] G(B)673 Arg(A)222 [Nη2•••OP2] C(B)672 Arg(A)222 [Nη2•••N7] G(B)673	C674
HηηP-2	1IL2	Arg(A)222 [Nε•••OP2] A(C)972 Arg(A)222 [Nη1•••O6] G(C)973 Arg(A)222 [Nη2•••OP2] A(C)972 Arg(A)222 [Nη2•••N7] G(C)973	G971
HηηP-3	5UD5	Arg(A)58 [Nε•••OP2] G(C)21 Arg(A)58 [Nη1•••N7] G(C)47 Arg(A)58 [Nη1•••OP1] U(C)59 Arg(A)58 [Nη2•••OP1] G(C)21 Arg(A)58 [Nη2•••O6] G(C)47	U59
HηηP-4	6CMN	Arg(A)47 [Nε•••OP2] U(D)23 Arg(A)47 [Nη1•••O6] G(D)26 Arg(A)47 [Nη2•••O5'] U(D)23 Arg(A)47 [Nη2•••N7] G(D)26	A22, U23

Table S4. Characteristics of distinct representatives of Type HS Arg forks taken from Chavali *et al.* for this study.[10]

Nomenclature of the Arg fork	**PDB Code**	**Hydrogen-bonding interactions**	**Nucleobase involved in cation-π interactions**
HS-1	1N32	Arg(G)3 [Nε•••N7] G(A)933 Arg(G)3 [Nη1•••O2'] U(A)1380 Arg(G)3 [Nη2•••O6] G(A)933	C932
HS-2	5WWF	Arg(C)99 [Nε•••O2'] A(D)2 Arg(C)99 [Nη1•••O6] G(D)4 Arg(C)99 [Nη2•••O2'] A(D)2 Arg(C)99 [Nη2•••N7] G(D)4	A2

Table S5. RMSD values obtained during benchmarking of geometries optimized reduced models using three different DFT functionals, with respect to the starting crystal structure geometry.

Arg fork	RMSD (Å)		
	B3LYP	**M062X**	**ωB97XD**
P-1	0.409	0.032	0.602
P-2	0.496	0.579	0.544
P-3	0.954	0.926	0.783
P-4	0.394	0.247	0.556
P-5	2.105	0.820	0.840
HεηP-1	0.453	0.436	0.413
HεηP-2	1.271	0.922	1.517
HεηP-3	1.499	0.960	0.468
HεηP-4	1.043	0.991	1.069
HεηP-5	0.779	0.717	0.553
HεηP-6	0.652	0.620	0.700
HεηP-7	0.810	0.788	0.817
HεηP-8	0.441	0.500	0.395
HηηP-1	1.088	1.184	0.943
HηηP-2	1.770	1.854	1.934
HηηP-3	0.946	0.892	0.898
HηηP-4	0.586	0.619	0.752
HS-1	1.011	1.000	1.392
HS-2	1.348	1.317	1.353
Average	0.950	0.811	0.870

Table S6. Interaction energies calculated at ω-B97XD/6-311+G(2df,2p) and RMSD values with respect to the starting crystal structure geometry for the full models of Type P Arg forks.

Arg fork	Interaction Energy (kcal mol^{-1})	RMSD (Å)
P-1	–41.4	1.664
P-2	–40.2	2.017
P-3	–38.3	1.539
P-4	–36.5	0.794
P-5	–36.4	2.375
Average	–38.6	1.678
Std. deviation	1.9	0.529

Table S7. Interaction energies calculated at ω-B97XD/6-311+G(2df,2p) and RMSD values with respect to the starting crystal structure geometry for the full models of Type HεηP Arg-forks.

Arg fork	Interaction Energy (kcal mol^{-1})	RMSD (Å)
HεηP-1	–40.2	0.383
HεηP-2	–38.8	0.416
HεηP-3	–40.3	0.393
HεηP-4	–41.6	0.761
HεηP-5	–33.5	1.034
HεηP-6	–31.1	0.891
HεηP-7	–29.1	0.524
HεηP-8	–38.2	0.568
Average	–36.8	0.739
Std. deviation	4.2	0.397

Table S8. Interaction energies calculated at ω-B97XD/6-311+G(2df,2p) and RMSD values with respect to the starting crystal structure geometry for the full models of Type HηηP Arg forks.

Arg fork	Interaction Energy (kcal mol^{-1})	RMSD (Å)
HηηP-1	–32.9	0.244
HηηP-2	–35.5	0.363
HηηP-3	–36.6	0.548
HηηP-4	–38.1	0.578
Average	–35.8	0.433
Std. deviation	1.9	0.137

Table S9. Interaction energies calculated at ω-B97XD/6-311+G(2df,2p) and RMSD values with respect to the starting crystal structure geometry for the full models of Type HS Arg-forks.

Arg fork	Interaction Energy (kcal mol^{-1})	RMSD (Å)
HS-1	–32.1	2.279
HS-2	–25.1	1.618
Average	–28.9	1.945
Std. deviation	3.5	0.331

Table S10. Comparison of interaction energies for two representatives of each Arg-fork type, calculated using different basis sets, on the full, optimized models.

Arg-fork Models	Interaction energy (kcal mol^{-1})			
	6-311+G(2df,2p)	**cc-pVDZ**	**aug-cc-pVDZ**	**cc-pVTZ**
HS-1	–32.1	–36.2	–32.4	–33.9
HS-2	–25.1	–27.6	–25.4	–26.1
HεηP-1	–40.2	–45.9	–40.6	–42.6
HεηP-2	–38.8	–44.3	–39.3	–41.0
P-1	–41.4	–46.7	–41.9	–43.6
P-2	–40.2	–45.7	–40.6	–42.4
HηηP-1	–32.9	–36.7	–33.2	–34.3
HηηP-2	–35.5	–40.9	–35.9	–37.9

Table S11. Average interaction energies for the full models of all four types of Arg-forks, with occurrences as identified by Chavali *et al.*[10]

Type of Arg-fork	Average interaction energy (kcal mol^{-1})	Frequency of crystal structure occurrences
P	–38.6	60
HεηP	–36.8	43
HηηP	–35.8	7
HS	–28.9	9

References:

1. Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215-241.
2. A. D. Becke, *The Journal of Chemical Physics*, 1993, **98**, 5648-5652.
3. C. Lee, W. Yang and R. G. Parr, *Physical Review B*, 1988, **37**, 785-789.
4. B. Miehlich, A. Savin, H. Stoll and H. Preuss, *Chem. Phys. Lett.*, 1989, **157**, 200-206.
5. S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456-1465.
6. J.-D. Chai and M. Head-Gordon, *Physical Chemistry Chemical Physics*, 2008, **10**, 6615-6620.
7. S. Grimme, *J. Comput. Chem.*, 2006, **27**, 1787-1799.
8. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, Williams, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, Wallingford, CT2016.
9. G. Scalmani and M. J. Frisch, *The Journal of chemical physics*, 2010, **132**, 114110.
10. S. S. Chavali, C. E. Cavender, D. H. Mathews and J. E. Wedekind, *Journal of the American Chemical Society*, 2020, **142**, 19835-19839.