Supporting Information for:

Cation–π interactions in protein-ligand binding: theory and data-mining reveal different roles for lysine and arginine

Kiran Kumar[a‡] and Shin M. Woo[a‡], Thomas Siu[a], Wilian A. Cortopassi[a], Fernanda Duarte[b], and Robert S. Paton[a*]

[a]*Chemistry Research Laboratory, University of Oxford, 12 Mansfield Road, Oxford OX1 3TA, UK*

[b]*EaStCHEM School of Chemistry, University of Edinburgh, Joseph Black Building, David Brewster Road, Edinburgh EH9 3FJ, UK*

[‡]These authors contributed equally to this work

fernanda.duarte@ed.ac.uk; robert.paton@chem.ox.ac.uk

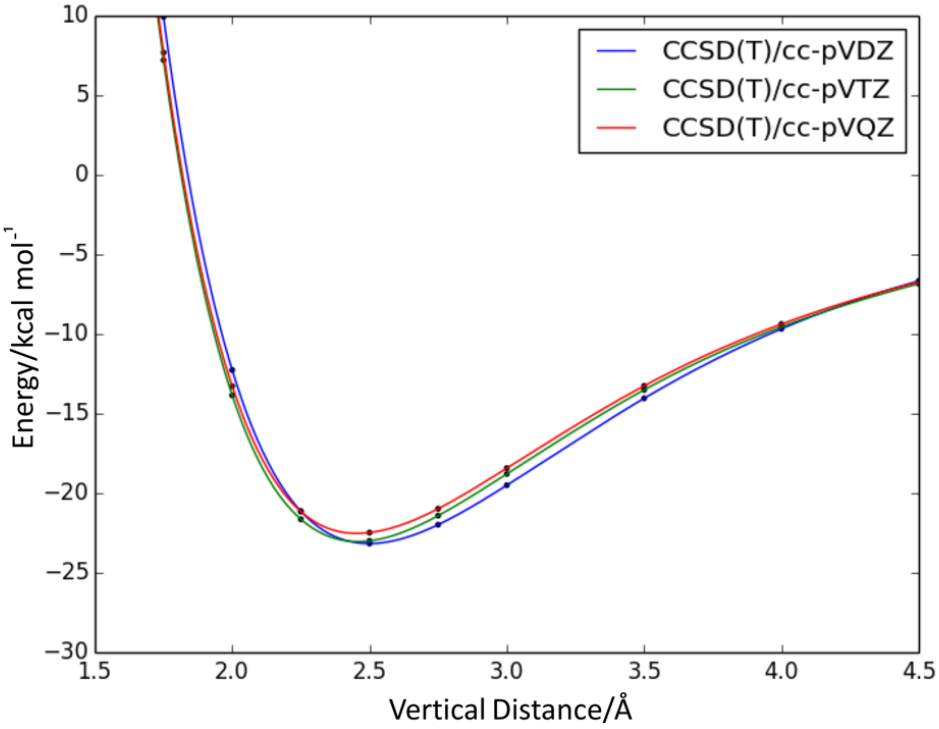
Table of contents

1 Full citation 50

Gaussian 09, Revision D.01, 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. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, 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. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. 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, Gaussian, Inc., Wallingford CT, 2016.

2 $[C_6H_6][Na]^+$ potential energy curve: basis set dependency

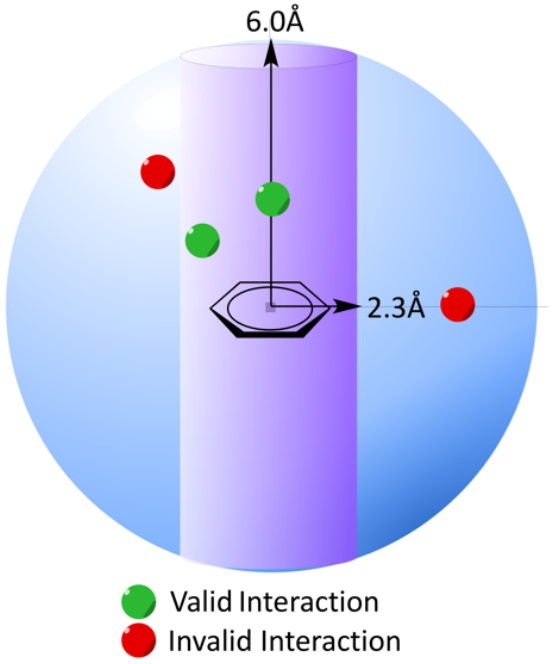
Figure S1. CCSD(T)/aug-cc-pVXZ potential energy curve of the $[C_6H_6][Na]^+$ complex



Basis Set	Equilibrium R_z (Å)	E_{int} (kcal mol^{-1})
aug-cc-pVDZ	2.50	–23.2
aug-cc-pVTZ	2.45	–23.0
aug-cc-pVQZ	2.45	–22.5

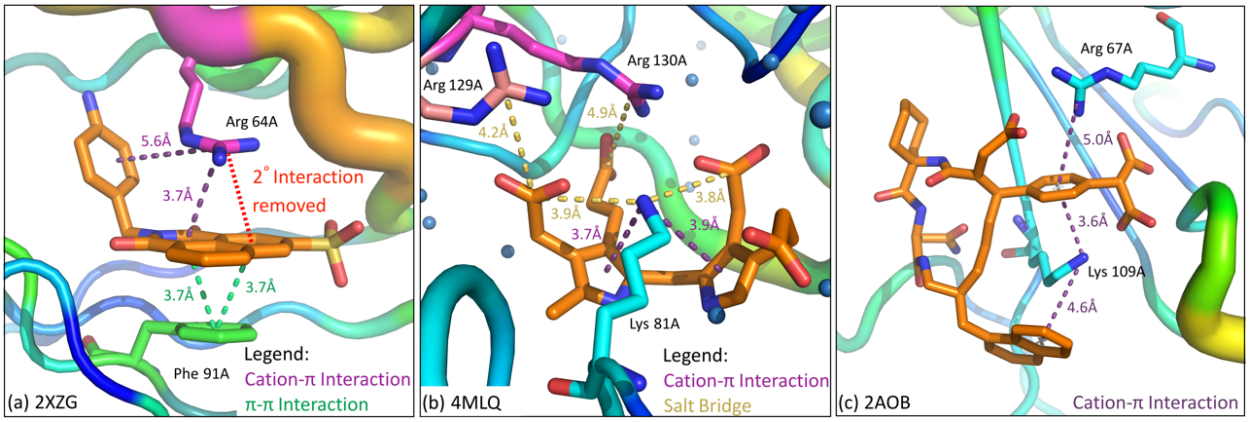
3 PLIP cation–π interaction geometric criteria

Figure S2. Geometric criteria used define a cation–π interaction. A distance cut-off of 6.0 Å (blue sphere) was defined from the centroid of the aromatic ring, along with a 2.3 Å horizontal-offset cut-off in the plane of the ring (purple cylinder).



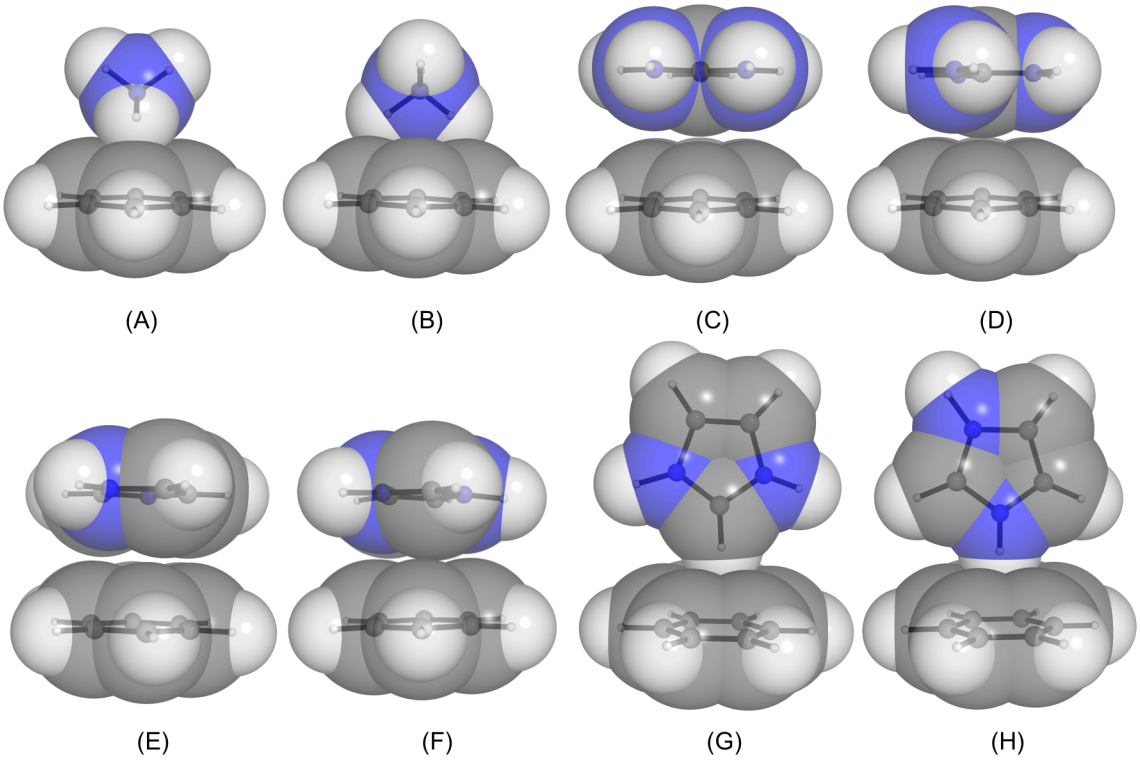
4 Interactions involving polycyclic aromatic rings

Figure S3. Examples of polycyclic aromatic rings further sub-classified as a) fused bicyclic ligand, b) sandwich complex, and c) bridged bicyclic ligand cation–π interaction.



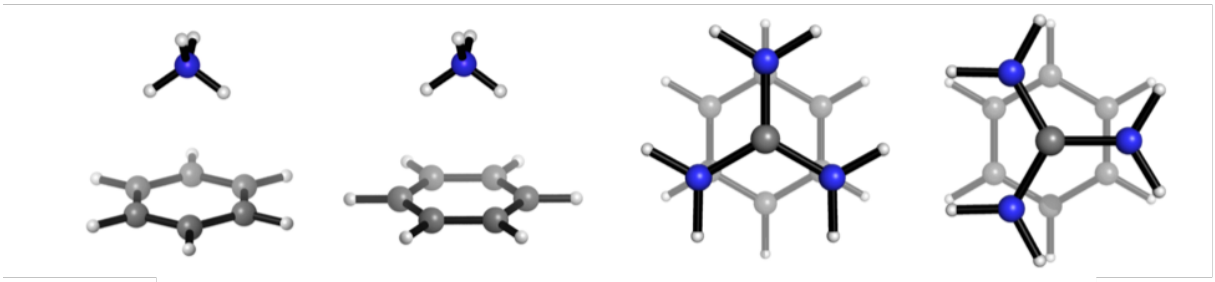
5 van der Waals Surface Representations

Figure S4. Van der Waals surfaces generated using atomic Bondi radii, for minimum energy geometries along the potential energy curve.



6 Minimum energy structures

Figure S5. SAPT2+3/aug-cc-pVDZ and DLPNO-CCSD(T)/aug-cc-pVTZ E_{int} (kcal mol^{-1}) at the minimum along the potential energy curve. Horizontal offset, R_y = 0.00 (Å) for all geometries.

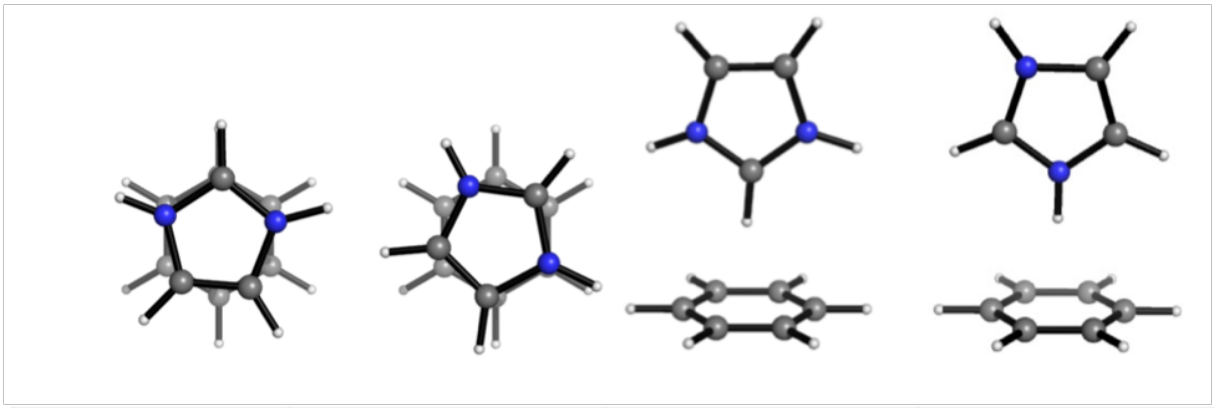


(A)
SAPT E_{int} = -19.2
CCSD(T) E_{int} = -19.2
R_z = 2.91 Å

(B)
SAPT E_{int} = -18.7
CCSD(T) E_{int} = -18.7
R_z = 2.90 Å

(C)
SAPT E_{int} = -7.1
CCSD(T) E_{int} = -7.5
R_z = 3.47 Å

(D)
SAPT E_{int} = -7.0
CCSD(T) E_{int} = -7.4
R_z = 3.47 Å



(E)
SAPT E_{int} = -7.4
CCSD(T) E_{int} = -7.7
R_z = 3.57 Å

(F)
SAPT E_{int} = -7.4
CCSD(T) E_{int} = -7.7
R_z = 3.57 Å

(G)
SAPT E_{int} = -11.8
CCSD(T) E_{int} = -11.5
R_z = 3.19 Å

(H)
SAPT E_{int} = -13.4
CCSD(T) E_{int} = -14.0
R_z = 3.01 Å

7 SAPT2+3 energy decomposition

Table S1. SAPT2+3/aug-cc-pVDZ decomposition of E_{int} values and DLPNO-CCSD(T)/aug-cc-pVTZ. Energies in kcal mol^{-1}.

Complex	Electrostatic (E_{ele})	Exchange (E_{ee})	Induction (E_{ind})	Dispersion (E_{disp})	SAPT2+3 E_{int}	CCSD(T) E_{int}
A	-13.3	14.9	-14.2	-6.6	-19.2	-19.2
B	-13.2	14.4	-13.5	-6.5	-18.8	-18.7
C	-5.2	6.2	-2.3	-5.9	-7.2	-7.5
D	-5.8	7.9	-2.4	-6.6	-6.9	-7.4
E	-5.1	6.4	-2.3	-6.5	-7.5	-7.7
F	-4.5	5.0	-2.1	-5.7	-7.3	-7.7
G	-7.9	9.5	-6.9	-6.5	-11.8	-11.5
H	-8.8	10.9	-8.2	-7.3	-13.4	-14.0

8 DLPNO-CCSD(T)/aug-cc-pVTZ calculated values

Table S2. R_z (Å) and DLPNO-CCSD(T)/aug-cc-pVTZ calculated $E_{monomer}$, $E_{complex}$, and E_{int} of cation–π complexes (**A**) and (**B**). E_π = -231.796893 a.u in the gas phase.

	(A) NH_4^+ **$E_{monomer}$ (a.u)** -56.813109		**(B)** NH_4^+ **$E_{monomer}$ (a.u)** -56.814684	
R_z (Å)	**(A)** $[C_6H_6][NH_4]^+$ **$E_{complex}$ (a.u)**	**E_{int} (kcal mol^{-1})**	**(B)** $[C_6H_6][NH_4]^+$ **$E_{complex}$ (a.u)**	**E_{int} (kcal mol^{-1})**
1.50	-288.087765	327.71	-288.135143	298.72
2.00	-288.371717	149.53	-288.389025	139.41
2.25	-288.522669	54.80	-288.529215	51.43
2.50	-288.597020	8.15	-288.599536	7.31
2.75	-288.628866	-11.84	-288.629767	-11.66
3.00	-288.639265	-18.36	-288.639891	-18.02
3.25	-288.640241	-18.98	-288.640565	-18.44
3.50	-288.637301	-17.13	-288.637826	-16.72
3.75	-288.633348	-14.65	-288.634031	-14.34
4.00	-288.629538	-12.26	-288.630299	-12.00
4.25	-288.626201	-10.17	-288.627068	-9.97
4.50	-288.623434	-8.43	-288.624362	-8.27
5.00	-288.621219	-7.04	-288.622198	-6.91
6.00	-288.617993	-5.01	-288.619041	-4.93

Table S3. R_z (Å) CPCM-MP2/cc-pVTZ calculated $E_{monomer}$, $E_{complex}$, and E_{int} of cation–π complex (**A**) in diethyl ether (dielectric constant = 4.2) and water (dielectric constant = 78.4).

R_z (Å)	$E_{int,ether}$ (kcal mol^{-1})	$E_{int,water}$ (kcal mol^{-1})
1.50	357.15	365.71
2.00	170.21	175.95
2.25	74.65	80.04
2.50	30.83	37.05
2.75	8.99	14.57
3.00	0.40	5.31
3.25	-2.56	1.62
3.50	-3.63	-0.32
3.75	-3.54	-0.91
4.00	-3.30	-1.26
4.25	-2.71	-0.97
4.50	-2.19	-0.68
5.00	-1.80	-0.50
6.00	-1.25	-0.27

Table S4. R_z (Å) and DLPNO-CCSD(T)/aug-cc-pVTZ calculated $E_{monomer}$, $E_{complex}$, and E_{int} of cation–π complexes (**C**) and (**D**). E_π = -231.800567 a.u.

	(C) **Gdm^+** **$E_{monomer}$ (a.u)** -205.441043		**(D)** **Gdm^+** **$E_{monomer}$ (a.u)** -205.441043	
R_z (Å)	**(C)** **$[C_6H_6][Gdm]^+$** **$E_{complex}$ (a.u)**	**E_{int} (kcal mol^{-1})**	**(D)** **$[C_6H_6][Gdm]^+$** **$E_{complex}$ (a.u)**	**E_{int} (kcal mol^{-1})**
1.50	-436.313400	582.46	-436.209354	647.75
2.00	-436.969408	170.81	-436.966130	172.87
2.25	-437.117579	77.83	-437.116097	78.76
2.50	-437.193159	30.40	-437.193079	30.45
2.75	-437.230276	7.11	-437.230030	7.27
3.00	-437.246628	-3.15	-437.245988	-2.75
3.25	-437.252469	-6.81	-437.252207	-6.65
3.50	-437.253562	-7.50	-437.253323	-7.35
3.75	-437.252640	-6.92	-437.252575	-6.88
4.00	-437.251313	-6.09	-437.251253	-6.05
4.50	-437.248719	-4.46	-437.248720	-4.46
5.00	-437.246811	-3.26	-437.246811	-3.26
6.00	-437.244694	-1.93	-437.245540	-2.47

Table S5. R_z (Å) and CPCM-MP2/cc-pVTZ calculated $E_{monomer}$, $E_{complex}$, and E_{int} of cation–π complex (**C**) in diethyl ether (dielectric constant = 4.2) and water (dielectric constant = 78.4).

R_z (Å)	$E_{int,ether}$ (kcal mol^{-1})	$E_{int,water}$ (kcal mol^{-1})
1.50	549.63	534.76
2.00	170.18	168.55
2.25	80.24	79.92
2.50	34.52	34.93
2.75	12.05	12.85
3.00	2.02	3.01
3.25	-1.84	-0.80
3.50	-3.02	-2.04
3.75	-3.17	-2.34
4.00	-2.90	-2.18
4.50	-2.42	-1.91
5.00	-1.60	-1.17
6.00	-0.57	-0.20

Table S6. R_z (Å) and DLPNO-CCSD(T)/aug-cc-pVTZ calculated $E_{monomer}$, $E_{complex}$, and E_{int} of cation–π complexes (**E**) and (**F**). E_π= -231.800567 a.u.

	(E) **Imi^+** **$E_{monomer}$ (a.u)**		**(F)** **Imi^+** **$E_{monomer}$ (a.u)**	
	-226.184887		-226.184887	
R_z (Å)	**(E)** **$[C_6H_6][Imi]^+$** **$E_{complex}$ (a.u)**	**E_{int} (kcal mol^{-1})**	**(F)** **$[C_6H_6][Imi]^+$** **$E_{complex}$ (a.u)**	**E_{int} (kcal mol^{-1})**
1.50	-457.575476	519.90	-457.604789	516.01
2.00	-457.866601	236.18	-457.900486	234.31
2.50	-457.977400	47.65	-457.992507	47.53
2.75	-457.995793	14.54	-458.004717	14.56
3.00	-458.002651	-0.19	-458.007788	-0.22
3.25	-458.003654	-6.04	-458.006431	-6.04
3.50	-458.002303	-7.67	-458.003901	-7.62
3.75	-458.000136	-7.44	-458.001031	-7.42
4.00	-457.997868	-6.58	-457.998407	-6.57
4.50	-457.994090	-4.77	-457.994383	-4.78
5.00	-457.991563	-3.46	-457.991759	-3.47
6.00	-457.988890	-2.00	-457.988981	-2.01

Table S7. R_z (Å) and SMD-MP2/cc-pVTZ calculated $E_{monomer}$, $E_{complex}$, and E_{int} of cation–π complex (**E**) in diethyl ether (dielectric constant = 4.2) and water (dielectric constant = 78.4).

R_z (Å)	$E_{int,ether}$ (kcal mol^{-1})	$E_{int,water}$ (kcal mol^{-1})
1.50	497.78	488.25
2.00	242.89	243.80
2.50	51.76	52.30
2.75	19.34	20.20
3.00	4.80	5.81
3.25	-1.18	-0.14
3.50	-3.17	-2.16
3.75	-3.47	-2.57
4.00	-3.17	-2.40
4.50	-2.15	-1.53
5.00	-1.38	-0.87
6.00	-0.63	-0.26

Table S8. R_z (Å) and DLPNO-CCSD(T)/aug-cc-pVTZ calculated $E_{monomer}$, $E_{complex}$, and E_{int} of cation–π complexes (**G**) and (**H**). E_{π}= -231.800567 a.u.

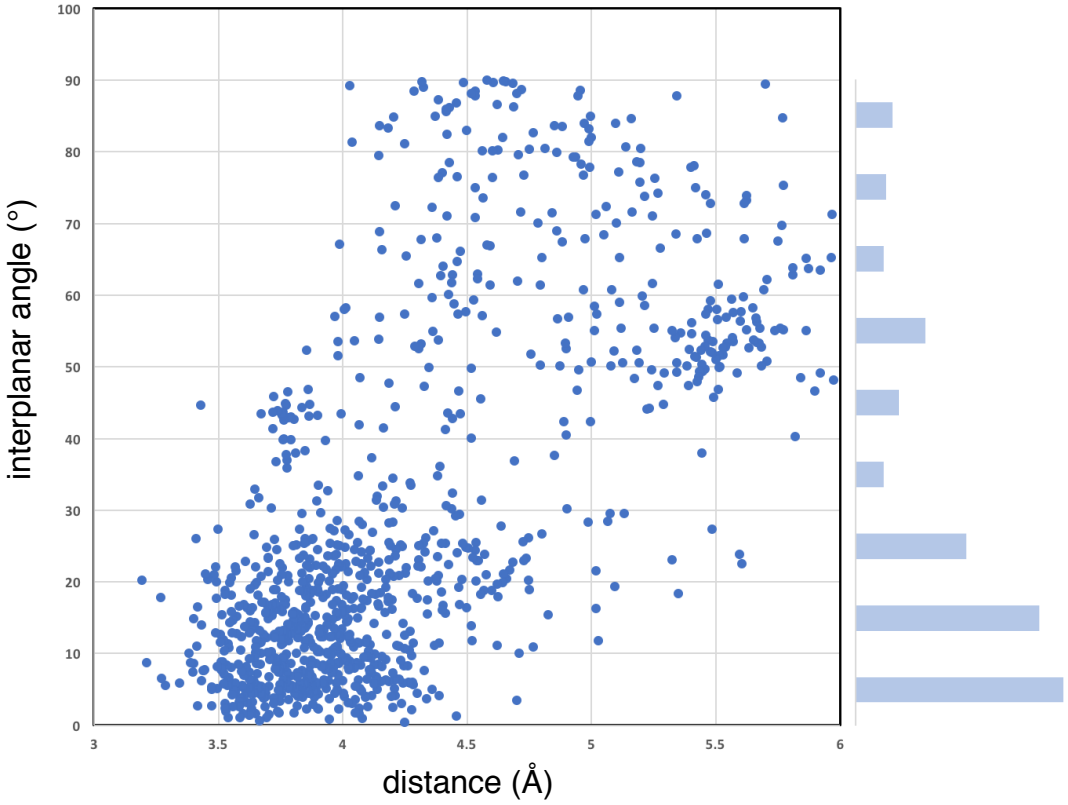
	(G) **Imi^+** **$E_{monomer}$ (a.u)** -226.184887		**(H)** **Imi^+** **$E_{monomer}$ (a.u)** -226.185434	
R_z (Å)	**(G)** **$[C_6H_6][Imi]^+$** **$E_{complex}$ (a.u)**	**E_{int} (kcal mol^{-1})**	**(H)** **$[C_6H_6][Imi]^+$** **$E_{complex}$ (a.u)**	**E_{int} (kcal mol^{-1})**
1.50	-457.156947	257.265055	-457.163683	238.870967
2.00	-457.609070	74.581396	-457.612602	53.318206
2.50	-457.909526	5.054003	-457.910256	-4.425725
2.75	-457.962289	-6.487940	-457.962794	-12.087740
3.00	-457.985750	-10.791220	-457.986358	-14.014867
3.25	-457.995086	-11.420705	-457.995625	-13.162981
3.50	-457.997670	-10.572985	-457.998138	-11.575549
3.75	-457.997317	-9.213249	-457.997830	-9.774355
4.00	-457.995947	-7.789787	-457.996477	-8.127922
4.50	-457.993056	-5.418769	-457.993618	-5.603050
5.00	-457.990972	-3.833049	-457.991525	-3.956115
6.00	-457.988646	-2.155779	-457.989197	-2.213165

Table S9. R_z (Å) SMD-MP2/cc-pVTZ calculated $E_{monomer}$, $E_{complex}$, and E_{int} of cation–π complex (**G**) in diethyl ether (dielectric constant = 4.2) and water (dielectric constant = 78.4).

R_z (Å)	$E_{int,ether}$ (kcal mol^{-1})	$E_{int,water}$ (kcal mol^{-1})
1.50	274.93	280.05
2.00	89.45	93.69
2.50	19.45	23.52
2.75	7.44	11.36
3.00	2.46	6.17
3.50	-2.27	-0.14
4.00	-2.94	-1.38
4.50	-2.48	-1.12
5.00	-1.61	-0.58
6.00	-1.04	-0.26

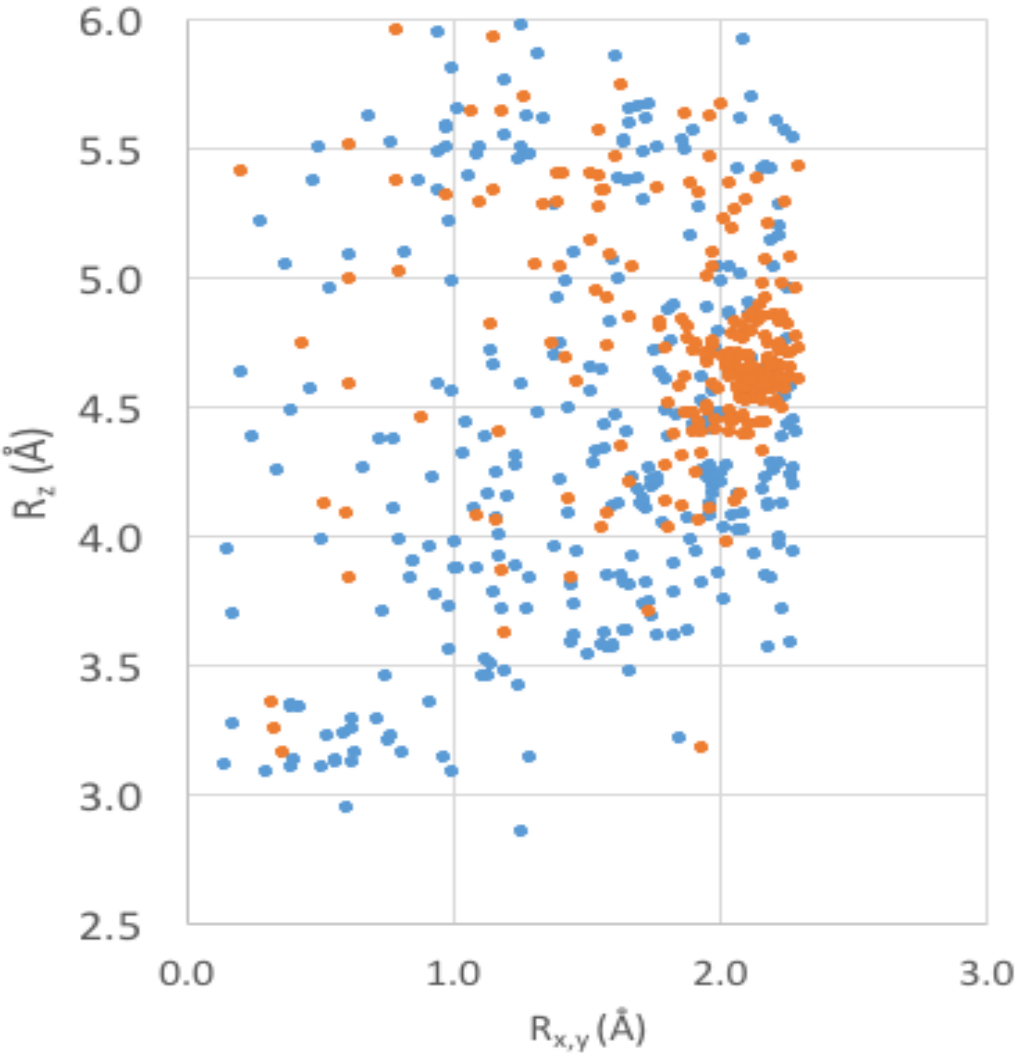
9 Orientations of Arg-aromatic interactions

Figure S6. Empirical distribution of Arg-aromatic complexes found in PDB crystal structures. The interplanar angle between arginine and the aromatic ring is plotted against the intermolecular separation of the two groups. Relative abunance in bins of 10° are shown on the right.



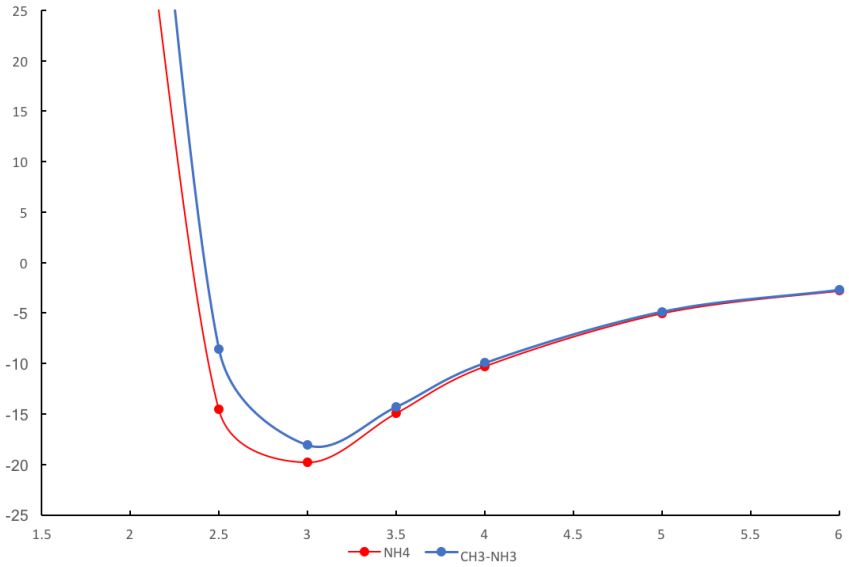
10 Lys-aromatic interactions

Figure S7. Empirical distribution of Lys-aromatic small molecule complexes found in PDB crystal structures. Orange points indicate aromatic system belongs to GTP, GDP, ATP, ADP, FAD, NAD, or derivatives.



11 Effect of methylation

Figure S8. MP2/cc-pVTZ interaction energy (kcal mol^{-1}) as a function of intermolecular separation of cation–π complexes: in red, NH_4^+-C_6H_6 (complex A); in blue $CH_3NH_3^+$-C_6H_6.



12 Cartesian coordinates

Table S10. Cartesian coordinates at DLPNO-CCSD(T)/aug-cc-pVTZ equilibrium separation.

(**A**)				(**B**)			
Atom	X	Y	Z	Atom	X	Y	Z
C	1.196745	0.716602	0.000000	C	-1.208142	-0.697418	0.000000
C	0.000000	1.394554	0.000000	C	-1.208053	0.697572	0.000017
C	-1.196745	0.716602	0.000000	C	0.000089	1.394991	0.000000
C	-1.196745	-0.716602	0.000000	C	1.208142	0.697418	0.000017
C	0.000000	-1.394554	0.000000	C	1.208053	-0.697572	0.000000
C	1.196745	-0.716602	0.000000	C	-0.000089	-1.394991	0.000017
H	2.140067	1.281657	0.000000	H	-2.160467	-1.247163	-0.000025
H	0.000000	2.494008	0.000000	H	-2.160308	1.247438	0.000006
H	-2.140067	1.281657	0.000000	H	0.000159	2.494601	-0.000025
H	-2.140067	-1.281657	0.000000	H	2.160467	1.247163	0.000006
H	0.000000	-2.494008	0.000000	H	2.160308	-1.247438	-0.000025
H	2.140067	-1.281657	0.000000	H	-0.000159	-2.494601	0.000006
--				--			
H	-0.841770	0.004911	3.501509	H	-0.000831	-0.835956	3.497879
H	0.825367	0.004911	3.501509	H	-0.000831	0.839242	3.497879
H	-0.008201	-0.790460	2.244923	H	0.819724	0.001643	2.277463
H	-0.008201	0.800282	2.244923	H	-0.821385	0.001643	2.277463

N	-0.008201	0.004911	2.897443	N	-0.000831	0.001643	2.899513

(**C**)				(**D**)			
Atom	X	Y	Z	Atom	X	Y	Z
C	1.208063	0.697555	0.000000	C	-1.208142	-0.697418	0.000000
C	-0.000068	1.394991	0.000000	C	-1.208053	0.697572	0.000017
C	-1.208132	0.697436	0.000000	C	0.000089	1.394991	0.000000
C	-1.208063	-0.697555	0.000000	C	1.208142	0.697418	0.000017
C	0.000068	-1.394991	0.000000	C	1.208053	-0.697572	0.000000
C	1.208132	-0.697436	0.000000	C	-0.000089	-1.394991	0.000017
H	2.160327	1.247406	0.000000	H	-2.160467	-1.247163	-0.000025
H	-0.000122	2.494601	0.000000	H	-2.160308	1.247438	0.000006
H	-2.160449	1.247195	0.000000	H	0.000159	2.494601	-0.000025
H	-2.160327	-1.247406	0.000000	H	2.160467	1.247163	0.000006
H	0.000122	-2.494601	0.000000	H	2.160308	-1.247438	-0.000025
H	2.160449	-1.247195	0.000000	H	-0.000159	-2.494601	0.000006
--				--			
C	0.000009	-0.004467	3.469280	C	0.005219	0.030155	3.390492
N	0.000074	1.324324	3.469280	N	-0.659177	1.180923	3.390494
H	0.859321	1.844780	3.469280	H	-1.663535	1.202043	3.390475
H	-0.858985	1.845025	3.467144	H	-0.175547	2.061386	3.392650
N	-1.150644	-0.668872	3.472194	N	1.334012	0.030155	3.390494
H	-2.031063	-0.185232	3.477093	H	1.854481	0.889394	3.390475
H	-1.171905	-1.673176	3.469852	H	1.854700	-0.828912	3.392650
N	1.150754	-0.668858	3.466578	N	-0.659177	-1.120613	3.390494
H	1.172015	-1.673150	3.469504	H	-0.175288	-2.000972	3.390475
H	2.031077	-0.185070	3.462064	H	-1.663495	-1.142009	3.392650

(**E**)				(**F**)			
Atom	X	Y	Z	Atom	X	Y	Z
C	1.208097	0.697495	0.000000	C	1.208097	0.697495	0.000000
C	0.000000	1.394991	0.000000	C	0.000000	1.394991	0.000000
C	-1.208097	0.697495	0.000000	C	-1.208097	0.697495	0.000000
C	-1.208097	-0.697495	0.000000	C	-1.208097	-0.697495	0.000000
C	0.000000	-1.394991	0.000000	C	0.000000	-1.394991	0.000000
C	1.208097	-0.697495	0.000000	C	1.208097	-0.697495	0.000000
H	2.160388	1.247300	0.000000	H	2.160388	1.247300	0.000000
H	0.000000	2.494601	0.000000	H	0.000000	2.494601	0.000000
H	-2.160388	1.247300	0.000000	H	-2.160388	1.247300	0.000000
H	-2.160388	-1.247300	0.000000	H	-2.160388	-1.247300	0.000000
H	0.000000	-2.494601	0.000000	H	0.000000	-2.494601	0.000000
H	2.160388	-1.247300	0.000000	H	2.160388	-1.247300	0.000000
--				--			
C	-0.765250	-0.934781	3.564637	C	-1.165317	-0.221730	3.645458
C	0.632187	-1.044655	3.564637	C	0.829861	0.837989	3.645458
C	0.070868	1.168401	3.564637	C	-0.547521	1.023828	3.645458
N	-1.100118	0.416132	3.564637	N	-1.008955	1.893175	3.645458
H	-2.014767	0.779644	3.564637	H	-2.248569	-0.316902	3.645458
H	-1.529992	-1.710457	3.564637	H	-0.330245	-2.283498	3.645458
H	1.206171	-1.965184	3.564637	H	2.068940	-1.009326	3.645458
H	2.212152	0.563973	3.500000	H	1.515362	1.682469	3.645458
H	0.047231	2.257376	3.564637	H	1.092418	-0.538951	3.645458
N	1.162328	0.288689	3.564637	N	-0.172510	-1.210729	3.645458

(**G**)				(**H**)			
Atom	X	Y	Z	Atom	X	Y	Z
C	1.208097	0.697495	0.000000	C	-0.000001	0.731133	5.333897
C	0.000000	1.394991	0.000000	C	-0.000002	-0.670139	5.370485
C	-1.208097	0.697495	0.000000	C	0.000008	0.006232	3.189839
C	-1.208097	-0.697495	0.000000	C	0.000005	1.136243	4.002362
C	0.000000	-1.394991	0.000000	C	0.000006	2.068663	3.687216
C	1.208097	-0.697495	0.000000	C	-0.000005	1.454230	6.148534
H	2.160388	1.247300	0.000000	H	-0.000006	-1.291514	6.259712
H	0.000000	2.494601	0.000000	H	0.000006	-2.163291	3.681594
H	-2.160388	1.247300	0.000000	H	0.000013	0.086830	2.103593
H	-2.160388	-1.247300	0.000000	H	0.000004	-1.129772	4.011223
H	0.000000	-2.494601	0.000000	H	-0.000001	0.731133	5.333897
H	2.160388	-1.247300	0.000000	H	-0.000002	-0.670139	5.370485
--				--			
C	-0.000001	0.731133	5.333897	C	-0.000001	1.119680	3.796750
C	-0.000002	-0.670139	5.370485	C	-0.000004	0.721460	5.140745
C	0.000008	0.006232	3.189839	C	0.000004	-1.143448	3.823621
N	0.000005	1.136243	4.002362	N	0.000003	-0.021500	3.000000
H	0.000006	2.068663	3.687216	H	0.000006	-0.033087	2.015831
H	-0.000005	1.454230	6.148534	H	-0.000002	2.117895	3.360780
H	-0.000006	-1.291514	6.259712	H	-0.000008	1.375150	6.006494
H	0.000006	-2.163291	3.681594	H	-0.000002	-1.346180	6.038921
H	0.000013	0.086830	2.103593	H	0.000007	-2.151623	3.411300
N	0.000004	-1.129772	4.011223	N	-0.000001	-0.713310	5.157847