Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics

Gas-phase structures and thermochemistry of neutral histidine and its conjugated acid and base.

Vanessa Riffet and Guy Bouchoux*.

Laboratoire des Mécanismes Réactionnels, Ecole Polytechnique, 91128 Palaiseau cedex, France.

Electronic Supplementary Information

ESI-1. Nomenclature describing conformers of neutral, protonated and deprotonated histidine used in Table ESI-T1, T2 and Figures ESI-F1, F2, F3.

ESI-2. Zwitterionic structures of histidine.

Figure ESI-F1. Optimized (B3LYP/6-31G(2df,p)) geometries and, into parentheses, relative G4MP2 $H°_{298}$ and $G°_{298}$ ($kJ.mol^{-1}$) of neutral histidine.

Figure ESI-F2. Optimized (B3LYP/6-31G(2df,p)) geometries and, into parentheses, relative G4MP2 $H°_{298}$ and $G°_{298}$ ($kJ.mol^{-1}$) of protonated histidine.

Figure ESI-F3. Optimized (B3LYP/6-31G(2df,p)) geometries and, into parentheses, relative G4MP2 $H°_{298}$ and $G°_{298}$ ($kJ.mol^{-1}$) of deprotonated histidine.

Table ESI-T1.Total (Hartree) and relative ($kJ.mol^{-1}$)energies, intrinsic and mixing entropies ($J.mol^{-1}.K^{-1}$) and molar fraction (%) calculated at the G4MP2level for neutral, protonated and deprotonated histidine.

Table ESI-T2.Total energies (Hartree) and thermochemical parameters($kJ.mol^{-1}$) calculated at the G4 level for Histidine, Glycine and Imidazole.

ESI-1. Nomenclature describing conformers of neutral, protonated and deprotonated histidine used in Tables ESI-T1, T2 and Figures ESI-F1, F2, F3.

Conformations of the α-amino acid group corresponding to the possible internal hydrogen bonds networks located in the OC(1)C(2)N moiety, will be denoted **I**, **I'**, **II** or **III** (Chart I). Each type of conformers, **I**, **I'**, **II** or **III**, may have C-N rotamers corresponding to different C(1)C(2)NH(H) dihedral angle values, these conformations are depicted by subscripts **A**-**E** as indicated in Chart I. We adopt the *antiperiplanar* (**a**) and *gauche* (**g+** or **g-**) denominations to describe the side chain arrangements along the C(2)C(3) bond while the two possible conformations along the C(3)C(4) sigma bonds are denoted *right* (**r**) or *left* (**l**) (Chart II).

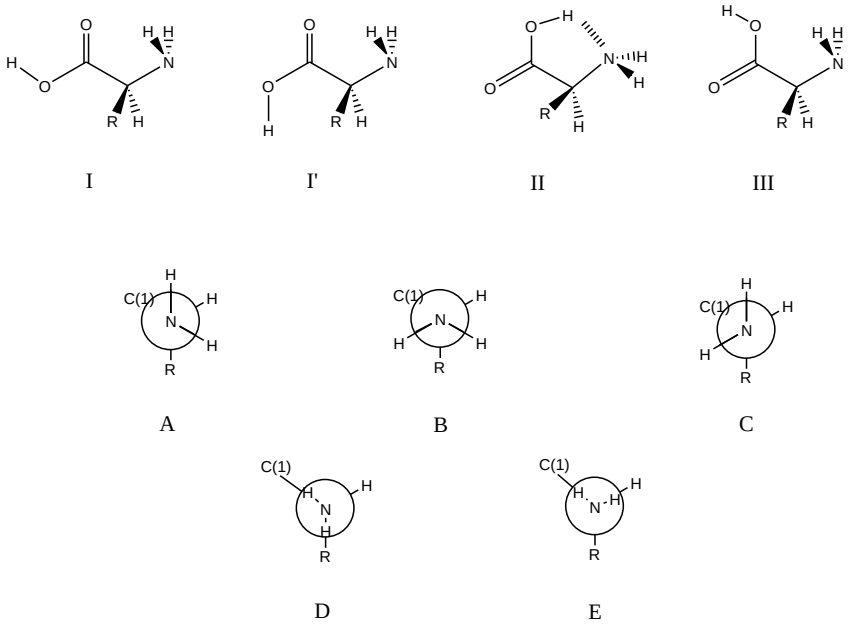


Chart I

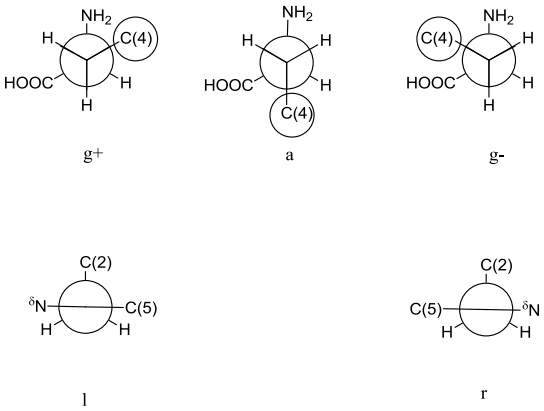
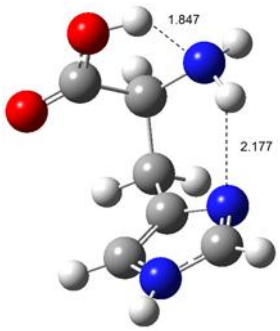


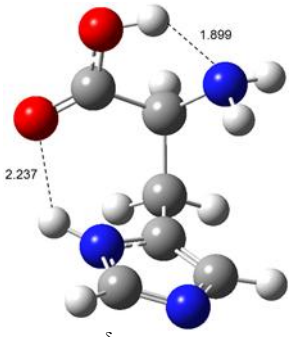
Chart II

ESI-2. Zwitterionic structures of histidine.

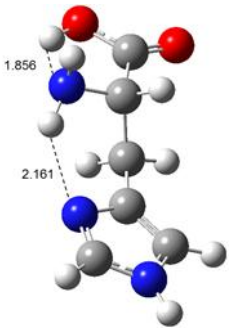
Previous studies, conducted at the B3LYP/6-311G(2df,p)//B3LYP/6-311G(d) and MP2/6-311G(2df,p)//B3LYP/6-311G(d) levels, indicated that zwitterionic forms of histidine are situated at last 50 kJ.mol^{-1} higher in energy than the most stable canonical form.[29,30] The two most stable zwitterionic forms located by Lin and coworkers[29] formally correspond to structures **His1** and **His2** where the hydroxylic proton has moved toward the nitrogen of the amino group. We observe that, starting from **His1** and **His2**, scans conducted at the B3LYP/6-31+G(d,p) level show a continuous increase in energy when the proton migrates from the OH group to the amino nitrogen atom. At this level of theory, the potential energy gap is equal to 49 (from **His1**) and 55 (from **His2**) kJ.mol^{-1} thus confirming the finding of Lin et al.[29] However, tentative of geometry optimization of the corresponding zwitterionic structures by the G4MP2 method collapsed on the canonical forms **His1** and **His2**. Similarly, zwitterionic structures resulting from protonation of the imidazole ring easily collapse to more stable canonical forms by migration of the N^{δ} proton to the carboxylate group. Thus the lone identified zwitterionic forms are characterized by large distances between these two structural moieties. This is the case of conformers **His16** and **His17** reported in Table ESI-T1 and Figure ESI-F1. These structures are however situated 168 and 174 kJ.mol^{-1} above **His1** and their occurrence in the gas phase, under usual experimental conditions, may be certainly ruled out.



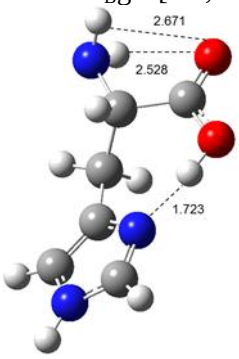
His1 His$^{\varepsilon}$II$_{B}$g-r [0.0; 0.0]



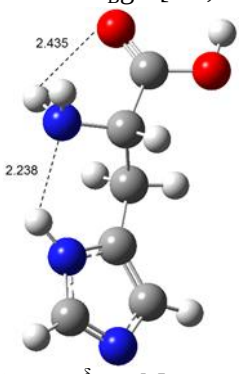
His2 His$^{\delta}$II$_{B}$g-l [2.2; 3.9]



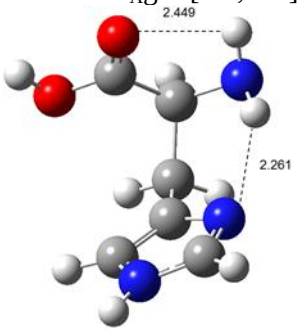
His3 His$^{\varepsilon}$II$_{A}$g+l [5.0; 4.6]



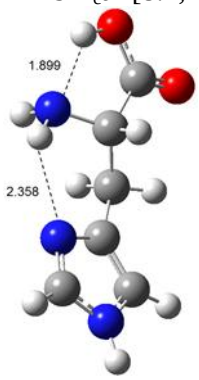
His4 His$^{\varepsilon}$I'$_{c}$ar [8.7; 6.0]



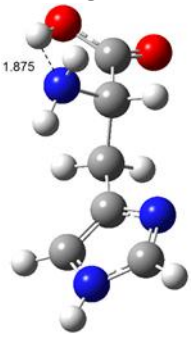
His5 His$^{\delta}$I$_{c}$g+l [9.1; 6.7]



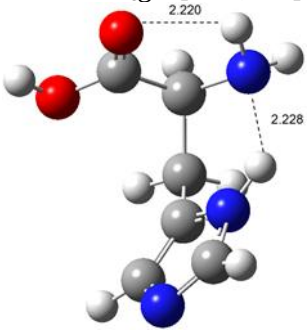
His8 His$^{\varepsilon}$I$_{c}$g-r [10.8; 7.8]



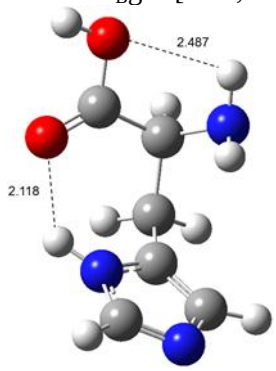
His6 His$^{\varepsilon}$II$_{B}$g+l [10.3; 9.2]



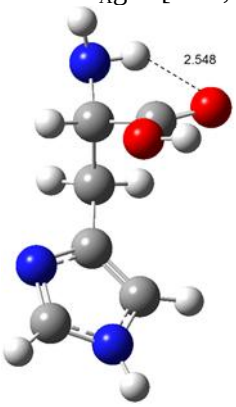
His11 His$^{\varepsilon}$II$_{A}$g+r [12.4; 9.4]



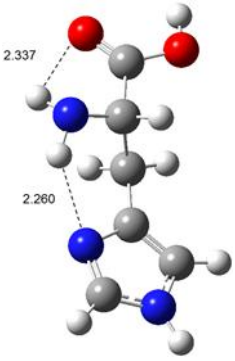
His7 His$^{\delta}$I$_{A}$g-r [10.3; 9.5]



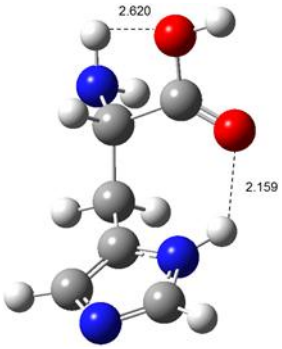
His9 His$^{\delta}$III$_{C}$g-l [11.3; 9.5]



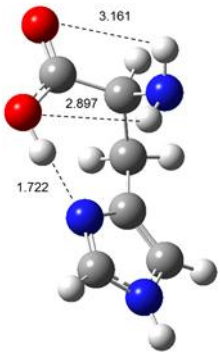
His15 His$^{\varepsilon}$I$_{c}$al [16.5; 10.5]



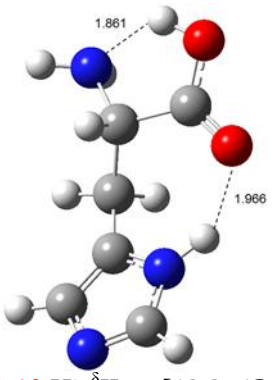
His14 His$^{\varepsilon}$I$_{B}$g+l [15.3; 12.1]



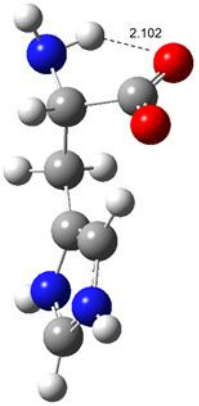
His13 His$^{\delta}$III$_{c}$ar [14.8; 13.0]



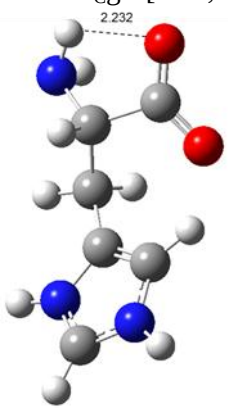
His12 His$^{\varepsilon}$I'$_{C}$g-l [13.3; 13.1]



His10 His$^{\delta}$II$_{B}$ar [12.2; 13.4]

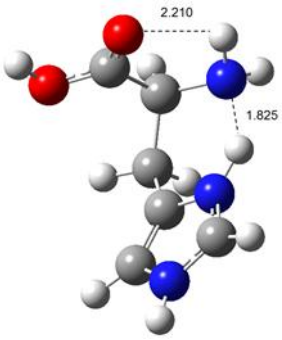


His16 ZwHis$_{E}$al [168.9;168.4]

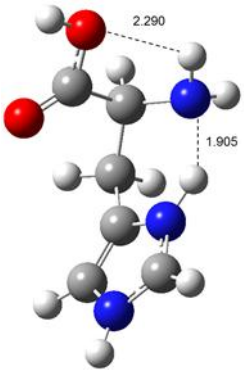


His17 ZwHis$_{C}$al [175.7;174.3]

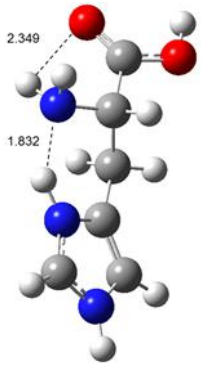
Figure ESI-F1. Optimized (B3LYP/6-31G(2df,p)) geometries and, into brackets, relative G4MP2 $H°_{298}$ and $G°_{298}$ (kJ.mol^{-1}) of neutral histidine.



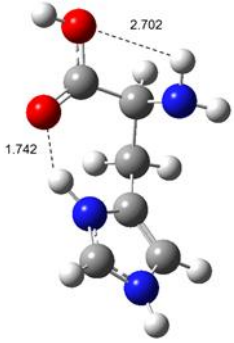
HisH1 HisHArI$_A$g-r [0.0; 0.0]



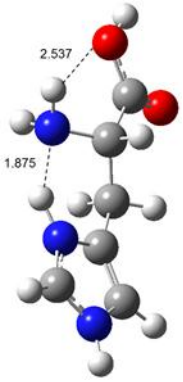
HisH2 HisHArIII$_A$g-r [2.3; 2.6]



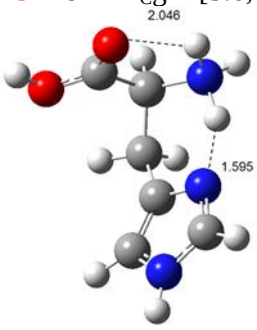
HisH3 HisHArI$_C$g+l [3.6; 3.1]



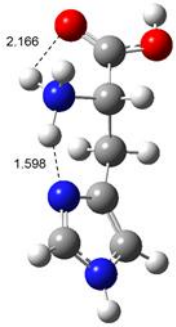
HisH4 HisHArIII$_A$g-l [5.2; 4.0]



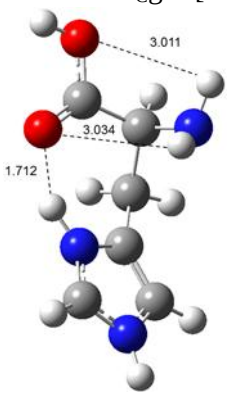
HisH8 HisHArIII$_C$g+l [8.4; 6.9]



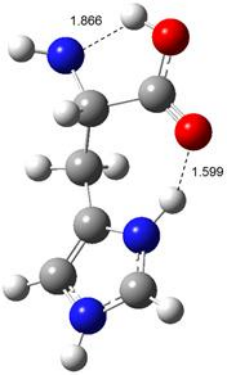
HisH5 HisH$^{\varepsilon}$Ig-r [5.8; 7.8]



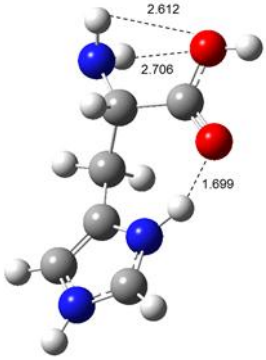
HisH6 HisH$^{\varepsilon}$Ig+l [7.1; 8.1]



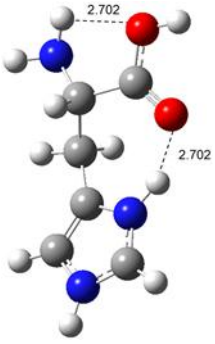
HisH9 HisHArIII$_C$g-l [8.4; 8.2]



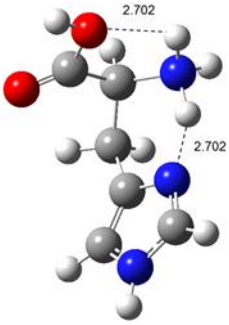
HisH7 HisHArII$_B$ar [7.3; 10.1]



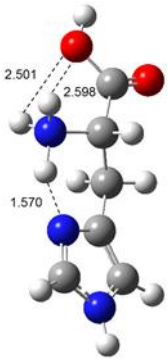
HisH10 HisHArIII$_C$ar [11.2; 10.5]



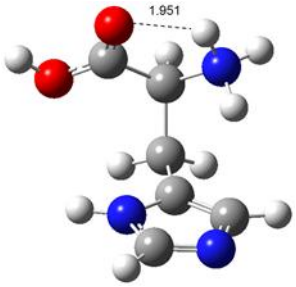
HisH11 HisHArIII$_A$ar [11.5; 11.6]



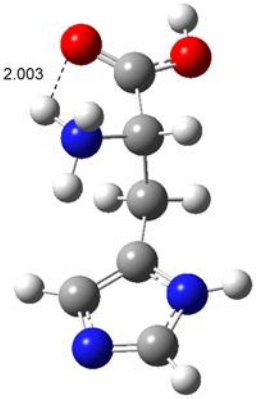
HisH12 HisH$^{\varepsilon}$IIIg-r [16.5; 17.9]



HisH13 HisH$^{\varepsilon}$IIIg+l [19.2; 18.6]

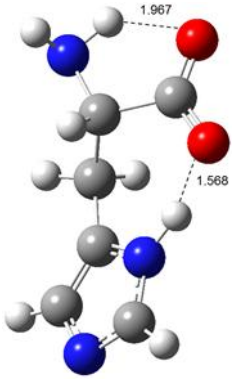


HisH14 HisH$^{\delta}$Ig-l [83.2; 82.6]

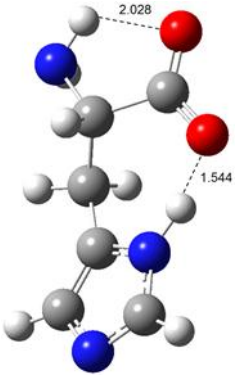


HisH15 HisH$^{\delta}$Ig+r [85.0; 83.6]

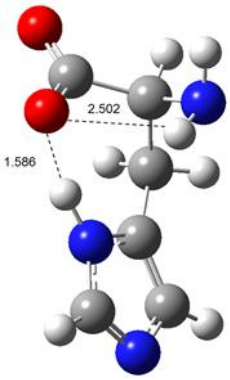
Figure ESI-F2. Optimized (B3LYP/6-31G(2df,p)) geometries and, into brackets, relative G4MP2 $H°_{298}$ and $G°_{298}$ (kJ.mol^{-1}) of protonated histidine.



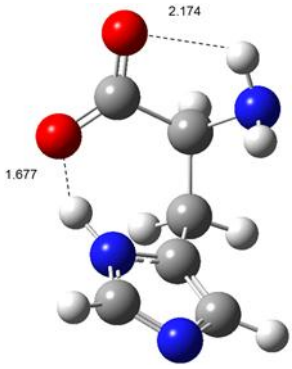
His-H2 His-H$^{\delta}_{E}$ar [0.3; 0.0]



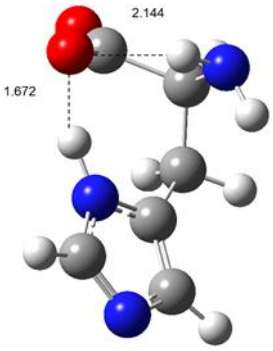
His-H1 His-H$^{\delta}_{D}$ar [0.0; 1.6]



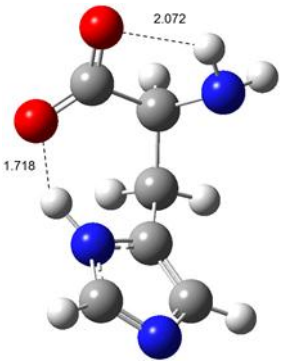
His-H4 His-H$^{\delta}_{C}$g-l [6.2; 5.5]



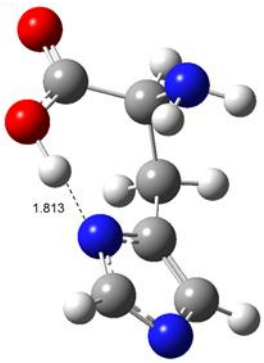
His-H3 His-H$^{\delta}_{D}$g-l [3.4; 5.6]



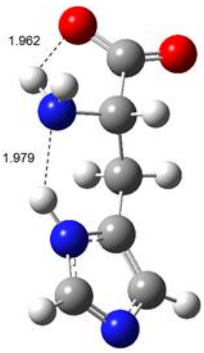
His-H5 His-H$^{\delta}_{B}$g-l [6.5; 8.7]



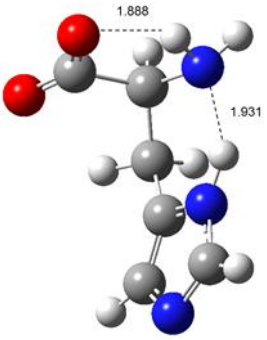
His-H6 His-H$^{\delta}_{E}$g-l [13.6; 15.0]



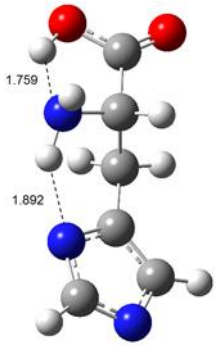
His-H7 His-HArI'$_{B}$g-l [17.9; 19.8]



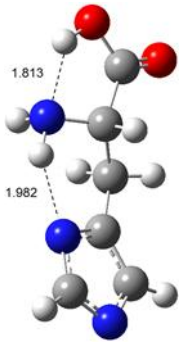
His-H8 His-H$^{\delta}_{E}$g+l [22.6; 22.8]



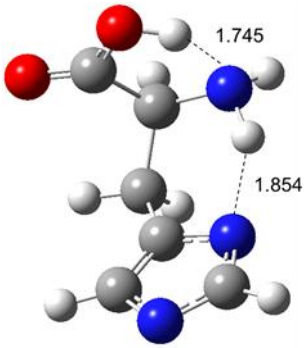
His-H9 His-H$^{\delta}_{E}$g-r [33.6; 34.5]



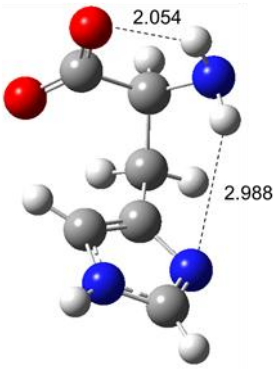
His-H10 His-HArII$_A$g+l [38.4; 39.8]



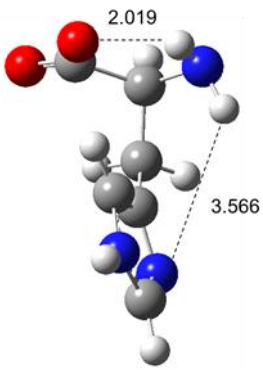
His-H11 His-HArII$_B$g+l [48.7; 49.8]



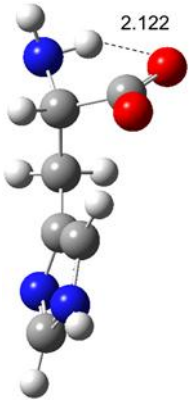
His-H12 His-HArII$_B$g-r [48.8; 50.1]



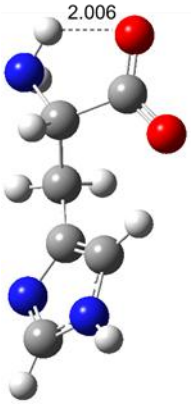
His-H13 His-H$^{\varepsilon}_C$g-r [57.6; 56.8]



His-H14 His-HεDg-r [57.8; 57.7]



His-H15 His-H$^{\varepsilon}_E$al [60.4; 59.2]



His-H16 His-H$^{\varepsilon}_D$al [66.2; 63.8]

Figure ESI-F3. Optimized (B3LYP/6-31G(2df,p)) geometries and, into brackets, relative G4MP2 $H°_{298}$ and $G°_{298}$ (kJ.mol^{-1}) of deprotonated histidine.

Table ESI-T1.Total (Hartree) and relative (kJ.mol^{-1})energies, intrinsic and mixing entropies (J.mol^{-1}.K^{-1}) and molar fraction (%) calculated at the G4MP2 level for neutral, protonated and deprotonated histidine (in bold, averaged 298 K values).

conformer name	conformer nomenclature	$H°_0$ (Hartree)	$H°_{298}$ (Hartree)	$G°_{298}$ (Hartree)	$\Delta H°_{298}$	$\Delta G°_{298}$	$S°_{298}$	x_i	$S°_{298}$mix
His1	His$^{\varepsilon}$II$_B$g-r	-548.074309	-548.063401	-548.110713	0.0	0.0	416.8	0.59	2.60
His2	His$^{\delta}$II$_B$g-l	-548.073376	-548.062571	-548.109225	2.2	3.9	411.0	0.12	2.14
His3	His$^{\varepsilon}$II$_A$g+l	-548.07247	-548.061487	-548.108968	5.0	4.6	418.3	0.09	1.85
His4	His$^{\varepsilon}$I'$_c$ar	-548.071325	-548.06008	-548.108415	8.7	6.0	425.9	0.05	1.29
His5	His$^{\delta}$I$_c$g+l	-548.071222	-548.059948	-548.108178	9.1	6.7	424.9	0.04	1.09
His8	His$^{\varepsilon}$I$_c$g-r	-548.070647	-548.059302	-548.107752	10.8	7.8	426.9	0.03	0.79
His6	His$^{\varepsilon}$II$_B$g+l	-548.070521	-548.059488	-548.107214	10.3	9.2	420.5	0.01	0.52
His11	His$^{\varepsilon}$II$_A$g+r	-548.069899	-548.058663	-548.107133	12.4	9.4	427.0	0.01	0.49
His7	His$^{\delta}$I$_A$g-r	-548.070591	-548.059493	-548.107105	10.3	9.5	419.5	0.01	0.48
His9	His$^{\delta}$III$_C$g-l	-548.070313	-548.059108	-548.107086	11.3	9.5	422.7	0.01	0.47
His15	His$^{\varepsilon}$I$_c$al	-548.068671	-548.057108	-548.106715	16.5	10.5	437.1	0.01	0.35
His14	His$^{\varepsilon}$I$_B$g+l	-548.068934	-548.057584	-548.106087	15.3	12.1	427.3	0.00	0.20
His13	His$^{\delta}$III$_c$ar	-548.069025	-548.057766	-548.105771	14.8	13.0	422.9	0.00	0.16
His12	His$^{\varepsilon}$I'$_C$g-l	-548.069393	-548.058326	-548.105718	13.3	13.1	417.5	0.00	0.15
His10	His$^{\delta}$II$_B$ar	-548.069621	-548.058768	-548.105601	12.2	13.4	412.6	0.00	0.13
His16	ZwHis$_E$al	-548.010125	-547.999063	-548.046563	168.9	168.4	418.5	0.00	0.00
His17	ZwHis$_C$al	-548.007663	-547.996495	-548.044338	175.7	174.3	421.5	0.00	0.00
			-548.0623373	**-548.109767**			**417.9**	**1.00**	**12.70**

conformer name	conformer nomenclature	$H°_0$ (Hartree)	$H°_{298}$ (Hartree)	$G°_{298}$ (Hartree)	$\Delta H°_{298}$	$\Delta G°_{298}$	$S°_{298}$	x_i	$S°_{298}$mix
HisH1	HisHArI$_A$g-r	-548.44494	-548.433983	-548.481351	0.0	0.0	417.3	0.48	2.93
HisH2	HisHArIII$_A$g-r	-548.444095	-548.433113	-548.480367	2.3	2.6	416.3	0.17	2.51
HisH3	HisHArI$_C$g+l	-548.443692	-548.432613	-548.480172	3.6	3.1	419.0	0.14	2.28
HisH4	HisHArIII$_A$g-l	-548.443201	-548.431997	-548.479838	5.2	4.0	421.5	0.10	1.90
HisH8	HisHArIII$_C$g+l	-548.441874	-548.430773	-548.478721	8.4	6.9	422.4	0.03	0.88
HisH5	HisH$^{\varepsilon}$Ig-r	-548.442484	-548.431758	-548.478391	5.8	7.8	410.9	0.02	0.69
HisH6	HisH$^{\varepsilon}$Ig+l	-548.442189	-548.431291	-548.478272	7.1	8.1	413.9	0.02	0.63
HisH9	HisHArIII$_C$g-l	-548.441959	-548.430782	-548.478225	8.4	8.2	418.0	0.02	0.60

HisH7	HisHArII$_B$ar	-548.441971	-548.43121	-548.477491	7.3	10.1	407.8	0.01	0.33
HisH10	HisHArIII$_C$ar	-548.440883	-548.4297	-548.477335	11.2	10.5	419.7	0.01	0.29
HisH11	HisHArIII$_A$ar	-548.440716	-548.429602	-548.476926	11.5	11.6	416.9	0.00	0.21
HisH12	HisH$^{\varepsilon}$IIIg-r	-548.438475	-548.427698	-548.47454	16.5	17.9	412.7	0.00	0.02
HisH13	HisH$^{\varepsilon}$IIIg+l	-548.437637	-548.426670	-548.474256	19.2	18.6	419.3	0.00	0.02
HisH14	HisH$^{\delta}$Ig-l	-548.41338	-548.402277	-548.449905	83.2	82.6	419.6	0.00	0.00
HisH15	HisH$^{\delta}$Ig+r	-548.412858	-548.401621	-548.449506	85.0	83.6	421.9	0.00	0.00
			-548.433115	**-548.480525**			**417.7**	**1.00**	**13.29**

conformer name	conformer nomenclature	H°$_0$ (Hartree)	H°$_{298}$ (Hartree)	G°$_{298}$ (Hartree)	ΔH°$_{298}$	ΔG°$_{298}$	S°$_{298}$	x$_i$	S°$_{298}$mix
His-H2	His-H$^{\delta}_E$ar	-547.552608	-547.541887	-547.588915	0.3	0.0	414.3	0.56	2.69
His-H1	His-H$^{\delta}_D$ar	-547.552617	-547.542003	-547.588295	0.0	1.6	407.9	0.29	2.99
His-H4	His-H$^{\delta}_C$g-l	-547.550471	-547.539628	-547.586809	6.2	5.5	415.7	0.06	1.43
His-H3	His-H$^{\delta}_D$g-l	-547.55124	-547.540694	-547.586796	3.4	5.6	406.2	0.06	1.42
His-H5	His-H$^{\delta}_D$g-l	-547.550129	-547.539543	-547.585601	6.5	8.7	405.8	0.02	0.59
His-H6	His-H$^{\delta}_E$g-l	-547.547444	-547.536826	-547.583204	13.6	15.0	408.6	0.00	0.08
His-H7	His-HI'$_B$g-l	-547.545822	-547.53517	-547.581388	17.9	19.8	407.2	0.00	0.01
His-H8	His-H$^{\delta}_E$g+l	-547.544129	-547.533413	-547.580235	22.6	22.8	412.5	0.00	0.00
His-H9	His-H$^{\delta}_E$g-r	-547.539736	-547.529188	-547.575759	33.6	34.5	410.3	0.00	0.00
His-H10	His-HII$_A$g+l	-547.537869	-547.527392	-547.57374	38.4	39.8	408.3	0.00	0.00
His-H11	His-HII$_B$g+l	-547.53399	-547.523461	-547.569966	48.7	49.8	409.7	0.00	0.00
His-H12	His-HII$_B$g-r	-547.533868	-547.523432	-547.569817	48.8	50.1	408.7	0.00	0.00
His-H13	His-H$^{\varepsilon}_C$g-r	-547.531009	-547.520059	-547.567276	57.6	56.8	416.0	0.00	0.00
His-H14	His-H$^{\varepsilon}_D$g-r	-547.530923	-547.519973	-547.566935	57.8	57.7	413.8	0.00	0.00
His-H15	His-H$^{\varepsilon}_E$al	-547.53001	-547.519003	-547.566355	60.4	59.2	417.2	0.00	0.00
His-H16	His-H$^{\varepsilon}_D$al	-547.527845	-547.516789	-547.564609	66.2	63.8	421.3	0.00	0.00
			-547.541659	**-547.588406**			**411.9**	**1.00**	**9.21**

Table ESI-T2.Total energies (Hartree) calculated at the G4 level for Histidine, Glycine and Imidazole.

Species	$H°_0$ (Hartree)	$H°_{298}$ (Hartree)	$G°_{298}$ (Hartree)
His1	-548.518676	-548.507768	-548.555080
His2	-548.517637	-548.506831	-548.553485
His3	-548.516708	-548.505725	-548.553205
HisH1	-548.889402	-548.878444	-548.925812
HisH2	-548.888459	-548.877477	-548.924731
HisH3	-548.888000	-548.876922	-548.924481
HisH4	-548.887419	-548.876214	-548.924056
His-H1	-547.997199	-547.986585	-548.032877
His-H2	-547.997247	-547.986526	-548.033554
Imidazole	-226.100069	-226.095333	-226.126344
ImidazoleH$^+$	-226.457854	-226.453102	-226.484193
[Imidazole-H]$^-$	-225.544932	-225.540418	-225.571049
Glycine	-284.303472	-284.296789	-284.332755
GlycineH$^+$	-284.639318	-284.632855	-284.668057
[Glycine-H]$^-$	-283.759415	-283.753213	-283.787883
H$^+$	0.000000	0.002359	