Verification of the structural and electrostatic properties obtained by the use of the different pseudoatom databases. (Supplementary Materials)

Joanna M. Bąk,[1] Sławomir Domagała,[2] Christian Hübschle, [3] Christian Jelsch, [2] Birger Dittrich,[3] Paulina M. Dominiak[1]*

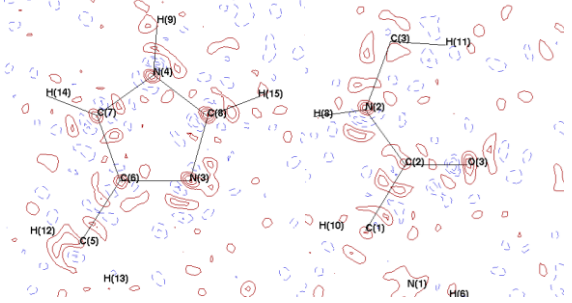
1 Department of Chemistry, University of Warsaw, ul. Pasteura 1, 02-093 Warszawa, Poland
2 Laboratoire de Cristallographie, Résonance Magnétique et Modélisations (CRM2), CNRS, UMR 7036, Institut Jean Barriol, Faculté des Sciences et Techniques, Nancy Université, BP 70239, 54506 Vandoeuvre-lès-Nancy CEDEX, France
3 Insitute fur Anorganische Chemie, Georg-August-Universitat Goettingen, Tammannstr 4, Goettingen, 37077, Germany

Experimental and computational methods:

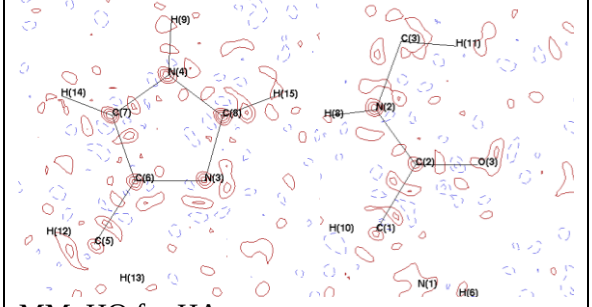
Crystals of L-His-L-Ala dihydrate were obtained by vapor diffusion methods (ethanol diffusing into a saturated aqueous solution). The high resolution single crystal X-ray measurement of L-His-L-Ala was carried out on a Bruker goniometer with a SMART 1K CCD detector with a molybdenum sealed tube source at a temperature of 100K (nitrogen gas flow cooling device). The data were collected using the omega scan method adding phi scans for completion and corrected for Lorentz and polarization effects. The scan width in omega or phi was 0.3 degrees. Different scan times were used for low- medium and high order scans with values of 5, 30 and 120 sec. Indexing, integration and scaling were performed with the supplied Bruker SAINT 7.61A software (Madison, 2008). An empirical absorption correction was applied using SADABS-2008/2 (G. M. Sheldrick, 2008), which was also used for merging. Crystallographic data are given below.

Measurement temperature [K]	100
Wavelength [Å]	0.71073
Space group	Monoclinic, $P2_1$
Unit cell dimensions: a, b, c [Å] β[°]	 4.6253(1); 14.8928(2); 8.9896(2) 90.466(1)
V [Å^3]	619.22(3)
Z	2
Crystal size [mm]	0.6 x 0.55 x 0.18
Crystal form; color	Plate; colorless
No. of measured, No of independent observed reflections	55660, 10165
Criterion for observed reflections	$F > 3\sigma(F)$
R(int) , R(sigma)	0.0195, 0.0127
Absorption coefficient μ [mm-1]	0.10
T_{min}; T_{max}	0.7230; 0.7522
$(\sin\theta/\lambda)_{max}$ [Å^{-1}]	1.24
Index ranges (after mering)	h: -11 ,11; k: -0, 37; l: 0, 22

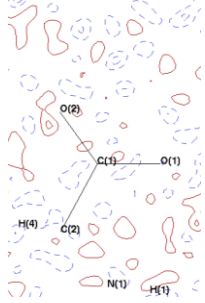
Residual densities from experimental multipolar refinements of HA and ALA are given below.



MM_TH



MM_HO for HA



MM_TH for ALA

Table 1S. Atom types used in TAAM refinements.

	EL	**EL2**	**IM**	**UB**
Water O1W, O2W	OW_HOH	O201T	O1h1h	O001
Water H1W, H2W, H3W, H4W	H1 & H2_HOH	H114T	H1o[1h]	H001

HA	**EL**	**EL2**	**IM**	**UB**
O3	O_HIS	O102	O1.5c[1.5n1c]	O104
O4	OT1 & AT2_CO2	O104	O1.5c[1.5o1c]	O101
O5	OT1 & AT2_CO2	O104	O1.5c[1.5o1c]	O101
N1	NT_NH3	N401	N1c1h1h1h	N402
N2	N_ALA	N302	N1.5c[1.5o1c]1c1h	N310
N3	ND1_HIS	N201	N1.5c[1.5n1h]1c	N201
N4	NE2_HIS	N303	N1.5c[1.5n1h]1.5c[1.5c1h]1h	N311
C1	CA_HIS	C408	C1n1c1c1h	C414
C2	C_HIS	C314	C1.5o1.5n[1c1h]1c	C304
C3	CA_HIS	C408	C1n1c1c1h	C420
C4	CT_CO2	C316	C1.5o1.5o1c	C301
C5	CB_HIS	C402	C1c1c1h1h	C407
C6	CG_HIS	C307	C1.5c[1.5n1h]1n1c	C360
C7	CD2_HIS	C306	C1.5n[1.5c1h]1.5c[1n1c]1h	C359
C8	CE1_HIS	C308T	C1.5n[1.5c1h]1.5n[1c]1h	C362
C9	CB_ALA	C401	C1c1h1h1h	C402
H5	H1_NH3	H113	H1n[1c1h1h]	H105
H6	H2_NH3	H113	H1n[1c1h1h]	H105
H7	H3_NH3	H113	H1n[1c1h1h]	H105
H8	H0_ALA	H112	H1n[1.5c1c]	H107
H9	NE2_HIS	H112	H1n[1.5c1.5c]	H114
H10	HA_HIS	H107	H1c[1n1c1c]	H103
H11	HA_ALA	H107	H1c[1n1c1c]	H103
H12	HB1_HIS	H104	H1c[1c1c1h]	H102
H13	HB2_HIS	H104	H1c[1c1c1h]	H102
H14	HD2_HIS	H102	H1c[1.5n1.5c]	H104
H15	HE1_HIS	H102	H1c[1.5n1.5n]	H104
H[CH_3]	HB1_ALA	H103	H1c[1c1h1h]	H101

GH	**EL**	**EL2**	**IM**	**UB**
O1	O_GLY	O102	O1.5c[1.5n1c]	O104
O2	OT1 & AT2_CO2	O104	O1.5c[1.5o1c]	O101
O3	OT1 & AT2_CO2	O104	O1.5c[1.5o1c]	O101
N1	NT_NH3	N401	N1c1h1h1h	N402
N2	N_HID	N302	N1.5c[1.5o1c]1c1h	N310
N3	ND1_HID	N303	N1.5c[1.5n1h]1c1h	N311
N4	NE2_HID	N201	N1.5c[1.5n1h]1c	N201
C1	CA_GLY	C407	C1n1c1c1h	C412
C2	C_GLY	C314	C1.5o1.5n[1c1h]1c	C304
C3	CA_GLY	C408	C1n1c1c1h	C420
C4	CT_CO2	C316	C1.5o1.5o1c	C301
C5	CB_HID	C402	C1c1c1h1h	C407
C6	CG_HID	C307	C1.5c[1n1h]1n1c	C361
C7	ND2_HID	C306	C1.5c[1n1c]1n1h	C358
C8	CE1_HID	C308T	C1.5n[1c1h]1.5n[1c]1h	C362
H1	H1_NH3	H113	H1n[1c1h1h]	H105
H2	H2_NH3	H113	H1n[1c1h1h]	H105
H3	H3_NH3	H113	H1n[1c1h1h]	H105

H4	HA1_GLY	H106	H1c[1n1c1h]	H102
H5	HA2_GLY	H106	H1c[1n1c1h]	H102
H6	H0_HID	H112	H1n[1.5c1c]	H107
H7	HA_HID	H017	H1c[1n1c1c]	H103
H8	HB1_HIS	H104	H1c[1c1c1h]	H102
H9	HB2_HID	H104	H1c[1c1c1h]	H102
H10	HD2_HID	H102	H1c[1.5c1n]	H104
H11	HE1_HID	H102	H1c[1.5n1.5n]	H104
H12	HD1_HID	H112	H1n[1.5c1c]	H114

HIS	**EL**	**EL2**	**IM**	**UB**
O1	OT1 & AT2_CO2	O104	O1.5c[1.5o1c]	O101
O2	OT1 & AT2_CO2	O104	O1.5c[1.5o1c]	O101
N1	NT_NH3	N401	N1c1h1h1h	N402
N2	ND1_HIS	N201	N1.5c[1.5n1h]1c	N201
N3	NE2_HIS	N303	N1.5c[1.5n1h]1.5c[1.5c1h]1h	N311
C1	CT_CO2	C316	C1.5o1.5n[1c1h]1c	C301
C2	CA_HIS	C408	C1n1c1c1h	C414
C3	CB_HIS	C401	C1c1c1h1h	C407
C4	CG_HIS	C307	C1.5c[1.5n1h]1n1c	C360
C5	CE1_HIS	C308T	C1.5n[1.5c1h]1.5n[1c]1h	C362
C6	CD2_HIS	C306	C1.5n[1.5c1h]1.5c[1n1c]1h	C59
H1	HA_HIS	H107	H1c[1n1c1c]	H105
H2	H1_NH3	H113	H1n[1c1h1h]	H105
H3	H2_NH3	H113	H1n[1c1h1h]	H105
H4	H3_NH3	H113	H1n[1c1h1h]	H114
H5	HB1_HIS	H104	H1c[1c1c1h]	H103
H6	HB2_HIS	H104	H1c[1c1c1h]	H102
H7	HE1_HIS	H102	H1c[1.5n1.5n]	H102
H8	HE2_HIS	H112	H1n[1.5c1.5c]	H104
H9	CD2_HIS	H102	H1c[1.5n1.5c]	H104

ALA	**EL**	**EL2**	**IM**	**UB**
O1	OT1 & AT2_CO2	O104	O1.5c[1.5o1c]	O101
O2	OT1 & AT2_CO2	O104	O1.5c[1.5o1c]	O101
N1	NT_NH3	N401	N1c1h1h1h	N402
C1	CT_CO2	C316	C1.5o1.5n[1c1h]1c	C301
C2	CA_ALA	C408	C1n1c1c1h	C414
C3	CB_ALA	C401	C1c1h1h1h	C402
H1	H1_NH3	H113	H1n[1c1h1h]	H105
H2	H2_NH3	H113	H1n[1c1h1h]	H105
H3	H3_NH3	H113	H1n[1c1h1h]	H105
H4	HA_ALA	H107	H1c[1n1c1c]	H103
H5	HB1_ALA	H103	H1c[1c1h1h]	H101
H6	HB2_ALA	H103	H1c[1c1h1h]	H101
H7	HB3_ALA	H103	H1c[1c1h1h]	H101

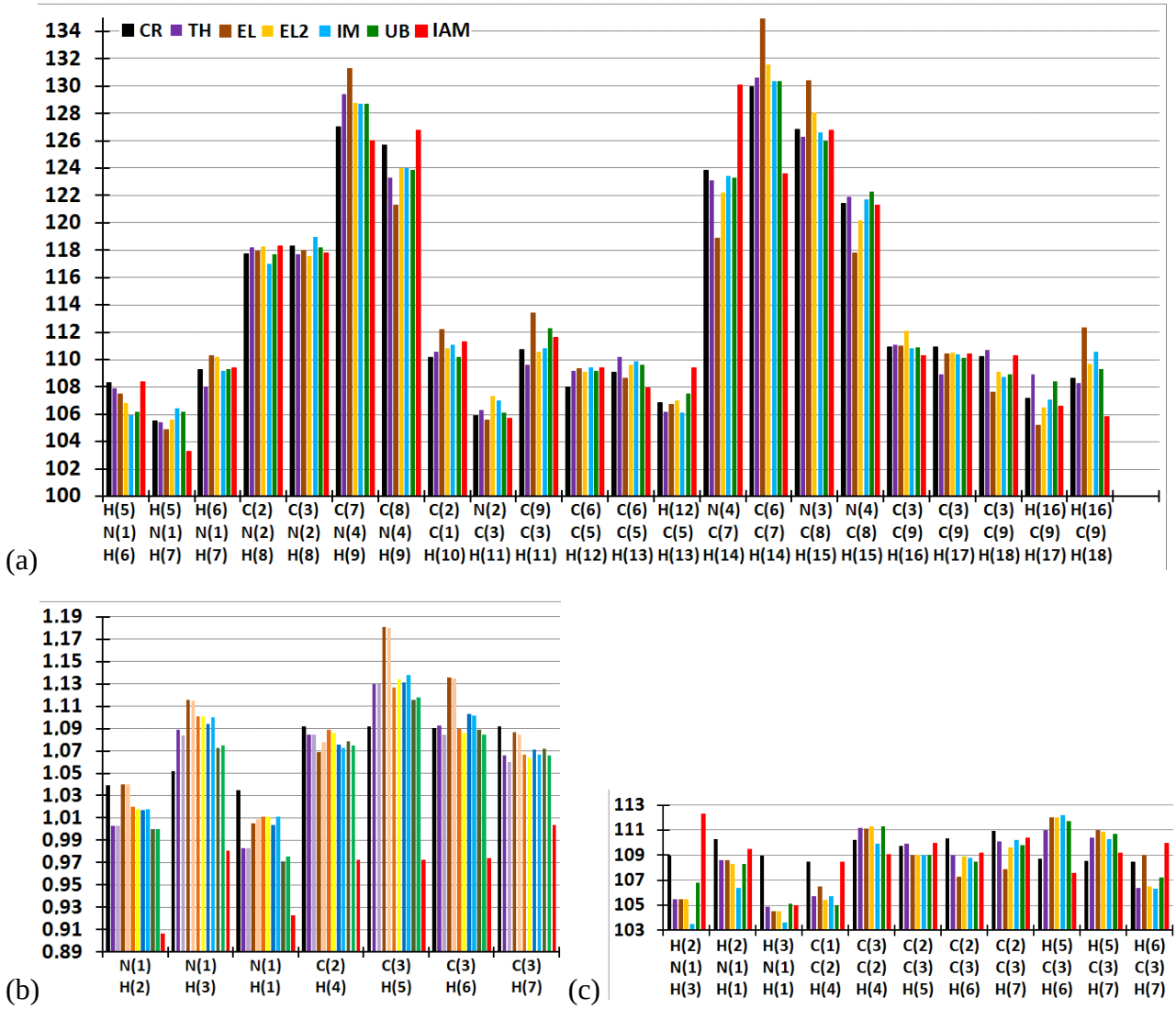


Figure 1S (a) the H-X-Y angles (°) in HA structure; (b) X-H bond lengths (Å) in ALA structure; (c) the H-X-Y angles (°) in ALA structure; obtained from IAM refinement and different TAAM refinements against high-resolution data compared to structural parameters from the reference theoretical geometries. Standard deviations from all refinements do not exceed 0.02 Å .

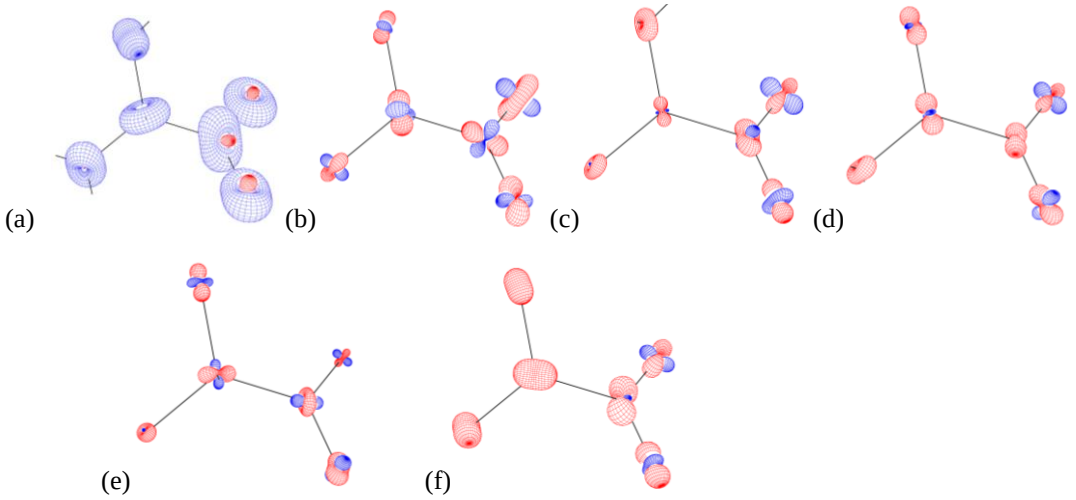


Figure 2S. The PEANUT representations of the differences in ADPs between (a) TH0.7_RB and IAM (b) TH_RB and EL_RB; (c) TH_RB and EL2_RB; (d) TH_RB and IM_RB; (e)TH_RB and UB_RB; (f) TH_RB and MM_HO. A scale of 20 was used for the representation of differences in root-mean-square deviation surfaces, except (a) where scale of 10 was used . An overestimation of the ADPs with respect to TH_RB model appears in blue.

Table 2S. Results of the Hirshfeld rigid – bond test (DMSDA) for HA.

$Z^2_{A,B}$		EL	EL2	IM	UB	MM_HO	MM_TH
O(3)	C(2)	1	2	1	1	5	-1
O(4)	C(4)	7	3	2	1	7	1
O(5)	C(4)	3	1	0	0	4	0
N(1)	C(1)	0	1	1	5	1	1
N(2)	C(2)	2	1	1	1	2	-1
N(3)	C(6)	-6	1	1	-2	0	0
N(4)	C(7)	5	-3	2	0	2	1
C(1)	C(2)	-3	-1	0	3	0	0
C(3)	C(4)	-3	1	1	2	-1	0
C(5)	C(6)	-3	-3	-3	-3	-2	-4
C(6)	C(7)	2	2	1	1	1	1

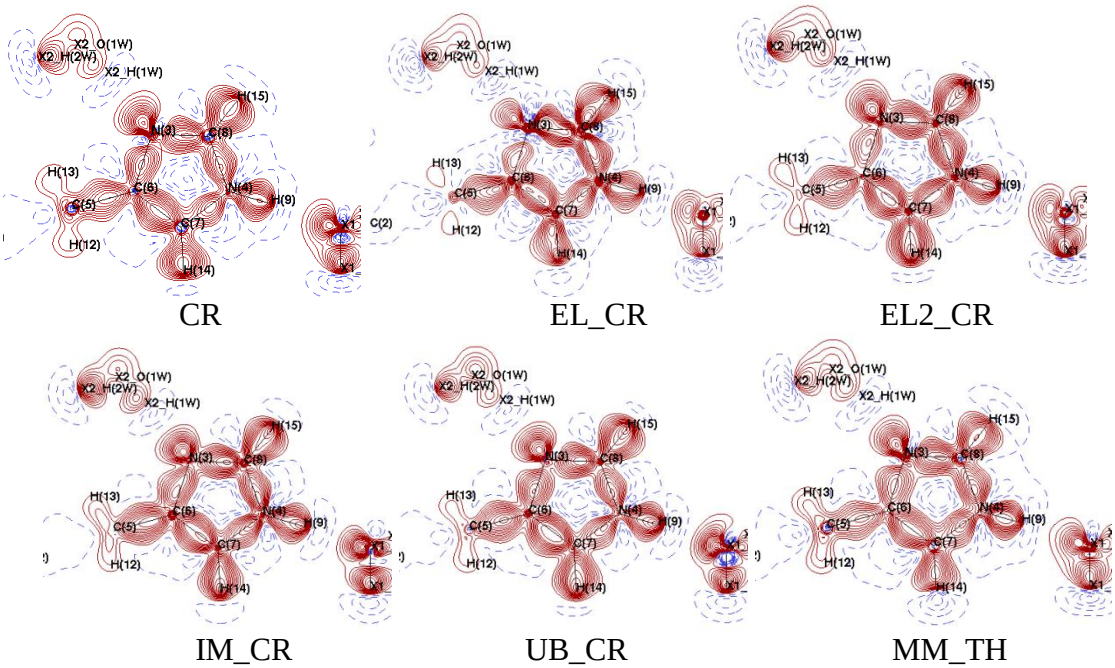


Figure 3S. Deformation density maps computed in the histidine (HA) ring plane. Contour interval is equal to 0.05 eÅ^{-3}. Positive values appear in red continuous lines; negative are in blue dashed lines.

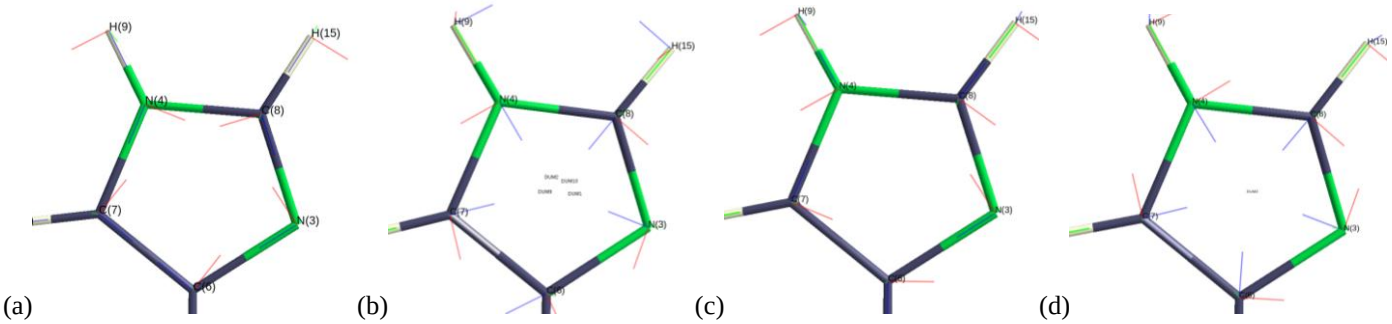
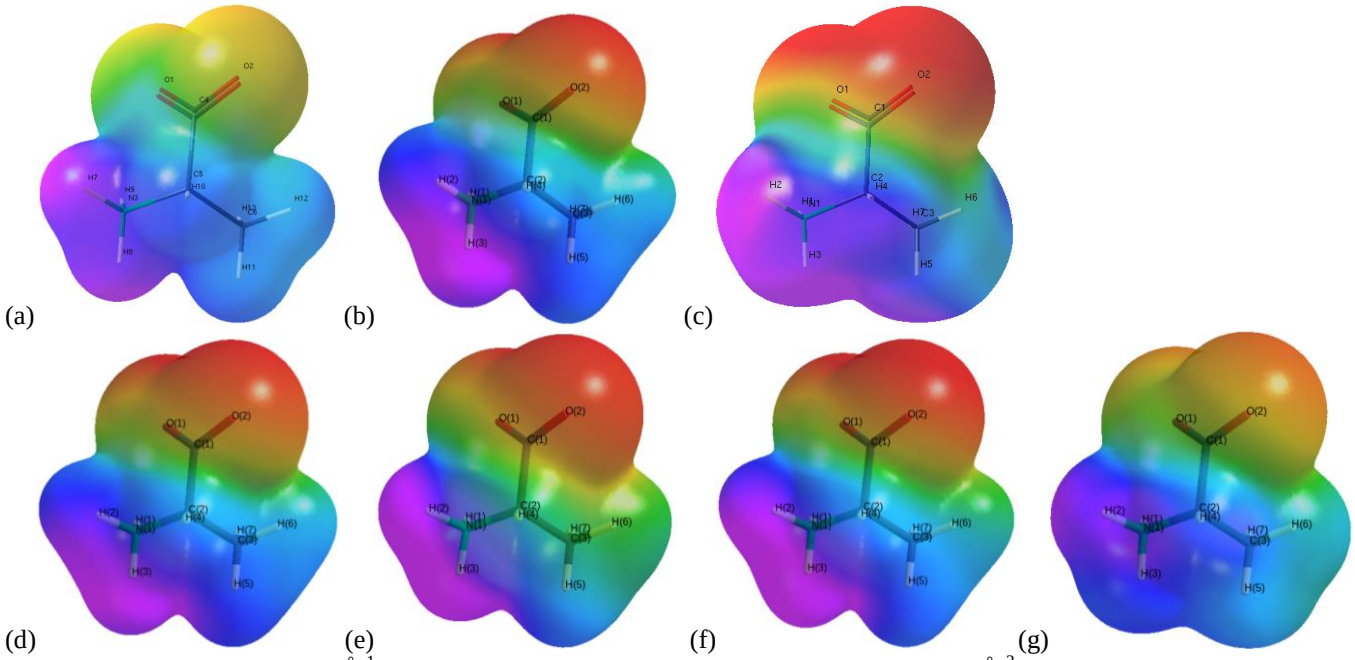


Figure 4S. Schemes of local coordinate system of the pseudoatoms in histidine ring used by the different databases: (a) ELMAM; (b) ELMAM2; (c)Invariom; (d) UBDB. X-blue; Y-red; Z-green.



(a) (b) (c)

(d) (e) (f) (g)

Figure 5S. Electrostatic potential (eÅ^{-1}) of LALA mapped on an ED isosurface at ρ=0.0135 eÅ^{-3}: (a) MP2; (b) CR; (c) EL_RB; (d) EL2_RB; (e) IM_RB; (f) UB_RB; (g) MM_TH

Definition of calculated MEPS quantities according to Murray *et al*. (2001):

The positive and negative average potential values V_{av}^{+}, V_{av}^{-} on the surface are calculated as given below:

$$V_{av}^{+} = \frac{1}{n}\sum_{i=1}^{n} V_S^{+}(r_i)$$

$$V_{av}^{-} = \frac{1}{m}\sum_{i=1}^{m} V_S^{-}(r_i)$$

The average deviation from the overall potential π, which can be interpreted as a measure of the local polarity is given as:

$$\pi = \frac{1}{m+n}\sum_{i=1}^{m+n} | V_S(r_i) - V_{av} |$$

.The positive, negative and total variances of the EPs on the surface (σ^{2+}, σ^{2-}, σ^{2}), which emphasize the extremes are given as:

$$\sigma_{tot}^2 = \sigma_{+}^2 + \sigma_{-}^2 = \frac{1}{n}\sum_{i=1}^{n} [V_S^{+}(r_i) - V_{av}^{+}]^2 + \frac{1}{m}\sum_{i=1}^{m} [V_S^{-}(r_i) - V_{av}^{-}]^2$$

.

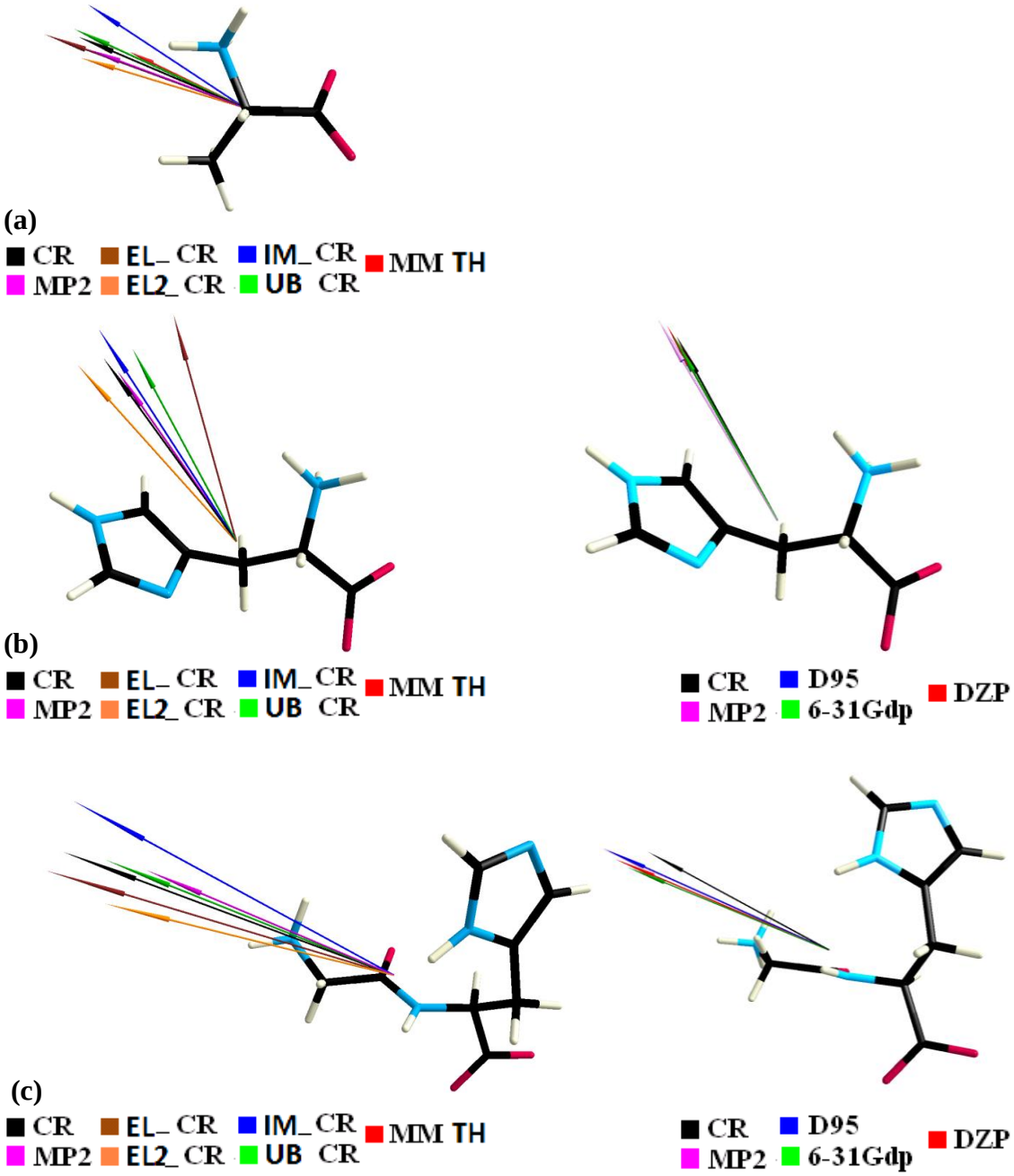


Figure 6S. The MolecoolQt representations of the molecular dipole moments vectors obtained for (a) ALA; (b) HIS and (c) GH from the different models. The vector colors indicates particular model.

Table 3S. Ees [kJ/mol] calculated for given dimers and for: (a) geometries taken from theoretical periodic calculations; (b) geometries resulting from TAAM refinements.

(a)

[kJ/mol]	symmetry	MP2	D95	6-31Gdp	DZP	CR	EL_CR	EL2_CR	IM_CR	UB_CR
HA - HA	1-x, 1/2+y, -z	-166	-160	-156	-152	-159	-139	-114	-161	-137
HA - HA	2-x,1/2 +y, -z	-121	-120	-111	-106	-105	-69	-71	-120	-98
HA -HA	1 +x, y, z	-30	-31	-4	-16	-20	-8	-63	51	-15
HA - HA	1 +x, y, 1+ z	-1	-1	1	0	1	10	-1	3	2
LALA - LALA	X, y, 1+z	-135	-132	-124	-126	-134	-141	-130	-121	-142
LALA - LALA	3/2–x, -y, 1/2+z	-30	-30	-20	-25	-30	-14	-27	-20	-28
LALA - LALA	-1/2+x, 1/2–y, 1-z	-156	-154	-136	-136	-138	-157	-136	-127	-150
HIS-HIS	1+x, y, z	-105	-101	-95	-97	-106	-34	-121	-101	-91
HIS-HIS	-x, 1-y, -z	-222	-219	-194	-192	-170	-186	-182	-189	-199
HIS-HIS	x, 1/2-y, ½+z	-122	-125	-110	-113	-113	-133	-96	-104	-111
HIS-HIS	-x, 1/2+y, 1/2-z	-60	-68	-60	-61	-45	-80	-49	-41	-63
HIS-HIS	1-x,-1/2+y, ½-z	-6	-4	-2	-2	-15	-34	-7	-15	-12
HIS-HIS	1-x, 1-y, 1-z	-40	-38	-28	-32	-38	-9	-49	-66	-40
HIS-HIS	-1+x, 1/2-y, -1/2+z	-6	-7	-5	-5	-5	21	-14	-2	5
GH-GH	-1+x, y, z	-152	-150	-134	-131	-154	-118	-166	-118	-139
GH-GH	1+x, 1+y, z	-104	-99	-98	-84	-61	-91	-59	-112	-79
GH -GH	1-x, ½+y, 1-z	-161	-157	-142	-142	-134	-105	-150	-96	-125
GH $1H_2O$	x, y, z	-59	-	-	-	-42	-34	-44	-46	-54
GH $1H_2O$	2-x, ½+y, 2-z	-82	-	-	-	-103	-62	-76	-69	-72
GH $2H_2O$	x, y, z	-60	-	-	-	-52	-42	-52	-49	-58
GH $2H_2O$	2-x, ½+y, 2-z	-79	-	-	-	-86	-72	-78	-63	-74
HA H_2O	2-x,1/2 +y, 1-z	-56	-	-	-	-78	-60	-63	-54	-47
HA H_2O	1 +x, 1+y, -1+z	-67	-	-	-	-74	-53	-63	-47	-62
HA H_2O	1 +x, 1+y, z	-61	-	-	-	-79	-4	-65	-50	-60

(b)

[kJ/mol]	1-x, 1/2+y, -z	MM_TH	MM_HO	TH_RB	EL_RB	EL2_RB	IM_RB	UB_RB
HA - HA	2-x,1/2 +y, -z	-196	-196	-145	-133	-101	-156	-130
HA - HA	1 +x, y, z	-151	-160	-102	-71	-70	-121	-99
HA -HA	1 +x, y, 1+ z	47	67	-11	-2	-55	57	-11
HA - HA	X, y, 1+z	-3	-1	3	10	2	4	5
LALA - LALA	3/2–x, -y, 1/2+z	-72	-78	-129	-140	-123	-118	-136
LALA - LALA	-1/2+x, 1/2–y, 1-z	-40	-40	-34	-16	-31	-20	-31
LALA - LALA	1-x, 1/2+y, -z	-98	-107	-139	-154	-134	-124	-149
HA H_2O	2-x,1/2 +y, 1-z	-73	-73	-73	-53	-54	-43	-45
HA H_2O	1 +x, 1+y, -1+z	-67	-74	-70	-49	-57	-42	-58
HA H_2O	1 +x, 1+y, z	-79	-80	-68	-11	-58	-49	-55

Table 4S. RMSDs in Ees [kJ/mol] between the results from periodic systems and other models.

N=10	MM_TH	MM_HO	TH_RB	EL_RB	EL2_RB	IM_RB	UB_RB
MM_TH	0	8	37	52	56	29	43
CR	37	40	7	29	27	31	17

Table 5S. The linear correlation coefficients for Ees. Highlighted coefficients indicate very strong ($r \geq 0.95$; pink) and strong ($0.95 > r \geq 0.92$; green) correlation. Number of dimers = 10).

	CR	TH_RB	EL_RB	EL2_RB	IM_RB	UB_RB	MM_TH
TH_RB	1.00	1.00					
EL_RB	0.93	0.95	1.00				
EL2_RB	0.90	0.90	0.93	1.00			
IM_RB	0.94	0.95	0.88	0.75	1.00		
UB_RB	0.96	0.98	0.97	0.94	0.93	1.00	
MM_TH	0.84	0.83	0.67	0.53	0.92	0.75	1.00
MM_HO	0.84	0.84	0.68	0.53	0.93	0.76	1.00