Supporting information for: Estimating strengths of individual hydrogen bonds in RNA base pairs: Towards a consensus between different computational approaches

Antarip Halder [a],*,† Dhruv Data,† Preethi P. Seelam [b],† Dhananjay Bhattacharyya,‡ and Abhijit Mitra*,†

†*Center for Computational Natural Sciences and Bioinformatics (CCNSB), International Institute of Information Technology (IIIT-H), Gachibowli, Hyderabad 500032, India.*

‡*Computational Science Division, Saha Institute of Nuclear Physics(SINP), 1/AF, Bidhannagar, Kolkata 700064, India*

E-mail: antarip.halder@research.iiit.ac.in; abi_chem@iiit.ac.in

[a]Current address: Solid State and Structural Chemistry Unit (SSCU), Indian Institute of Science (IISc), Bangalore 560012, India.

[b]Current address: Department of Chemistry and Biochemistry, University of Lethbridge, 4401 University Drive West, Lethbridge, Alberta, Canada T1K 3M4

Functional roles of non-coding RNAs

Over the last few decades, these newly discovered ncRNAs have trashed the established rules to make way for new ones.[1] For example, the very first discovery of the catalytic roles of RNA[2] challenged the rule that proteins are the only enzymes and not polynucleotides.[3] Subsequent discoveries of other self-splicing RNAs (Group I[4] and Group II[5] introns), tRNA processing enzyme RNase P[6] and naturally occurring RNA enzymes like the hammerhead,[7] hairpin,[8] hepatitis delta virus,[9] etc. further underscore the catalytic potential of RNA enzymes or ribozymes. Another forte of ncRNAs, that was earlier attributed exclusively to proteins, is the regulation of transcription and translation process.[10] An important example is the RNA interference (RNAi) process in which ncRNAs, specially microRNA (miRNA) and small interfering RNA (siRNA), inhibit gene expression by silencing targeted mRNAs.[11] Apart from that, different regulatory roles of miRNA in plants,[12] PIWI-associated RNAs (piRNA) in animals[13,14] and riboswitches,[15] T Boxes,[16] sRNAs,[17] etc in bacteria are known today.

Nomenclature of a base pairing interaction

As per Leontis and Westhof (LW) nomenclature scheme,[18] base pairing interaction between edge X of base M and edge Y of base N in *cis* orientation is represented as MN cXY. However, the nomenclature used in this work is slightly different. The same base pair is represented as M:N X:Y Cis. This sort of representation is also intuitive and at the same time, due to the presence of the text delimiter (:), is more convenient for parsing of metadata files, especially when modified nucleobases (represented by more than one character, *e.g.* 7MG, 2MA, etc) are involved. If the inter-base H-bonds in a particular base pair involve any of the sugar atoms (O2′, O3′, etc) a prefix ‘r’ is added with the base, e.g., G:rC W:S Cis.

Limitations of existing models that predict H-bonding energies from QTAIM based parameters

Different topology based parameters (say ρ, $\nabla^2\rho$, etc.) calculated at the Bond Critical Points of the H-bonds are known to have good correlation with the corresponding H-bonding strength. Grabowski and coworkers have proposed different derived parameters based of topology and geometry base parameters and shown that they correlate well with the strength of H-bonds. One of such example[19] is discussed below,

Grabowski's $\triangle_{com}$ measure of hydrogen bonding:

$$\triangle_{com} = \left\{\left[r_{\text{X-H}} - r^0_{\text{X-H}}/r^0_{\text{X-H}}\right]^2 + \left[(\rho^0_{\text{X-H}} - \rho_{\text{X-H}})/\rho^0_{\text{X-H}}\right]^2 + \left[(\nabla^2\rho_{\text{X-H}} - \nabla^2\rho^0_{\text{X-H}})/\nabla^2\rho^0_{\text{X-H}})\right]^2\right\}^{1/2} \tag{1}$$

where r_{X-H}, ρ_{X-H} and $\nabla^2\rho_{\text{X-H}}$ correspond to the parameters of proton donating bond involved in H-bonding, i.e., the bond length, electronic density at X-H BCP, and the Laplacian of that density, respectively, and r^0_{X-H}, ρ^0_{X-H}, and $\nabla^2\rho^0{}_{\text{X-H}}$ correspond to the same parameters of X-H bond not involved in H-bond formation.

However, the limitation of these models are that, they provide only qualitative information and do not provide us with exact H-bonding energy of the corresponding H-bonds. Getting information about the exact H-bonding energies from the QTAIM based parameters are possible from the model proposed by Espinosa et al.[20] It connects the H-bond dissociation energy D_e with the virial density V_{el} as,

$$D_e = ca_0^3 V_{el}(\vec{r}_{cp}) \tag{2}$$

where a_0 is the Bohr radius and c is the dimensionless proportionality constant, so that ca_0^3 becomes equivalent to a sort of 'effective volume'. Despite its popularity, this model has numerous limitations as discussed in Ref.[21] and is safe to use for only weak H-bonds.

For inter-molecular H-bonds, a popular approach is to assume that the stabilization energy of the molecular complexes are solely due to the contribution of the H-bonds. On the basis of this assumption, the following linear equations connecting intermolecular complex stabilization energy (SE) with electron charge density ρ^{cp} and its Laplacian $\nabla^2\rho$ have been proposed.

$$SE = \sum_{i=1}^{N} E_i \tag{3}$$

$$E_i = C_0 + C_1 \rho_i^{cp}$$

where N is the total number of H-bonds connecting the molecular complex and i is the index for the individual H-bonds. E_i is the energy of the H-bond. Different combinations of C_0 and C_1 proposed so far, have been are listed in Table 1 of Ref.[20] Here, it is to be noted that, the assumption that $SE = \sum E_i$ is not always true and the model is strongly system dependent. For example, the coefficients will be very different for the resonance assisted H-bonds found in DNA base pairs and for the water clusters. We have also illustrated in the main text the the models fail for base pairing systems.

Method of interaction energy calculation

An individual nucleobase (say, X) undergoes some structural changes (deformation) when it approaches another nucleobase (say, Y) to form a base pair (XY). Let us assume,

E_0^X = Electronic energy of the of base X in its isolated optimized geometry.

E_0^Y = Electronic energy of the base Y in its isolated optimized geometry.

E^{XY} = Electronic energy of the optimized geometry of a base pair XY.

Now, electronic energies of the 'deformed' nucleobases must be different from that of their isolated optimized geometries (E_0^X and E_0^Y). Let us further assume,

E^X = Electronic energy of the base X in its 'deformed' geometry.

E^Y = Electronic energy of the base Y in its 'deformed' geometry.

Therefore, the primary interaction energy of the base pair XY is given by,

$$\Delta E = E^{XY} - (E^X + E^Y) \tag{4}$$

This should further be corrected to account for the deformation of the individual bases from their respective isolated optimized geometries. The deformation correction is therefore calculated as,

$$E^{def} = (E^X - E_0^X) + (E^Y - E_0^Y) \tag{5}$$

Note that in equation 4, calculation of E^{XY} involves larger number of basis functions compared to those required for calculating E^X and E^Y, individually. This limitation leads to the overestimation of the interaction energy known as *Basis Set Superposition Error (BSSE).* To take care of that we have calculated the following correction term,

$$E^{BSSE} = (E^X - E_{XY}^X) + (E^Y - E_{XY}^Y) \tag{6}$$

,where E_{XY}^X and E_{XY}^Y correspond to the electronic energy of X and Y, respectively, when the calculations are performed using the same number of basis functions used for calculating the electronic energy of the base pair XY.

All the energies discussed above are calculated at 0K and do not contain the contribution of the vibrational motions that are present even at 0K (zero-point). So, we have calculated the *Zero-Point Vibrational Energy (ZPVE)* for the optimized geometry of the base pair XY (E^{ZPVE}(XY)) and the individual isolated bases X (E^{ZPVE}(X)) and Y (E^{ZPVE}(Y)). The corresponding correction term (E^{ZPVE}) is therefore calculated as,

$$E^{ZPVE} = E^{ZPVE}(XY) - (E^{ZPVE}(X) + E^{ZPVE}(Y)) \tag{7}$$

Therefore, the final corrected interaction energy is given by the following relation,

$$\begin{aligned} E^{int} &= \Delta E + E^{def} + E^{BSSE} + E^{ZPVE} && (8) \\ &= E^{XY} - (E^{X} + E^{Y}) + E^{def} + E^{BSSE} + E^{ZPVE} && (9) \end{aligned}$$

List of RNA base pairs studied in this work

List of all the natural base pairs that are available in RNABP COGEST

- **WW base pairs**: (1) A:A W:W Cis, (2) A:A W:W Trans, (3) A:C W:W Trans, (4) A:G W:W Cis, (5) A:U W:W Cis, (6) A:U W:W Trans, (7) C:C W:W Trans, (8) C:U W:W Cis, (9) C:U W:W Trans, (10) G:C W:W Cis, (11) G:C W:W Trans, (12) G:G W:W Trans, (13) G:U W:W Cis, (14) G:U W:W Trans, (15) U:U W:W Cis, (16) U:U(I) W:W Trans, (17) U:U(II) W:W Trans.

- **HH base pairs**: (1) A:A H:H Cis, (2) A:A H:H Trans, (3) A:C H:H Trans, (4) A:G H:H Cis, (5) A:U H:H Cis, (6) A:U(I) H:H Trans, (7) A:U(II) H:H Trans, (8) C:U H:H Trans, (9) G:C H:H Cis, (10) G:C(I) H:H Cis, (11) G:C(II) H:H Cis, (12) G:G H:H Trans, (13) A:G H:H Trans

- **WH base pairs**: (1) A:A W:H Cis, (2) A:A W:H Trans, (3) A:U W:H Cis, (4) C:A W:H Cis, (5) C:A W:H Trans, (6) C:C W:H Cis, (7) C:C W:H Trans,(8) G:A W:H Cis, (9) G:A W:H Trans, (10) G:G W:H Cis, (11) G:G W:H Trans, (12) G:U W:H Trans, (13) U:A W:H Cis, (14) U:G W:H Cis, (15) U:G W:H Trans, (16) U:U W:H Cis, (17) U:U W:H Trans

- **HS base pairs**: (1) A:rA H:S Cis, (2) A:rA H:S Trans, (3) A:rC H:S Cis, (4) A:rC H:S Trans, (5) A:rG H:S Cis, (6) A:rG H:S Trans, (7) A:rU H:S Cis, (8) A:rU H:S Trans, (9) C:rA H:S Cis, (10) C:rA H:S Trans, (11) C:rC H:S Cis, (12) C:rC H:S Trans, (13) C:rG H:S Cis, (14) C:rU H:S Cis, (15) C:rU H:S Trans, (16) G:A H:S Trans, (17) G:rA H:S Cis, (18) G:rC H:S Cis, (19) G:rG H:S Cis, (20) G:rU H:S Cis, (21) U:A H:S Trans, (22) U:G H:S Trans, (23) U:rA H:S Cis, (24) U:rC H:S Cis, (25) U:rG H:S Cis, (26)U:rU H:S Cis

- **WS base pairs**: (1) A:rA W:S cis, (2) A:rA W:S Trans, (3) A:rC W:S Cis, (4) A:rC W:S Trans, (5) A:rG W:S Cis, (6) A:rG W:S Trans, (7) A:rU W:S Cis, (8) A:rU W:S

Trans, (9) C:rA W:S Cis, (10) C:rC W:S Cis, (11) C:rC W:S Trans, (12) C:rG W:S Cis, (13) C:rG W:S Trans, (14) C:rU W:S Cis, (15) C:rU W:S Trans, (16) G:rA W:S Cis, (17) G:rA W:S Trans, (18) G:rC W:S Trans, (19) G:rG W:S Trans, (20) G:rU W:S Trans, (21) U:rA W:S Cis, (22) U:rA W:S Trans, (23) U:rC W:S Cis, (24) U:rG W:S Cis, (25) U:rG W:S Trans (26) U:rC W:S Trans, (27) U:rU W:S Cis, (28) U:rU W:S Trans

- **SS base pair**: (1)rA:rA S:S Cis, (2)rA:rA S:S Trans, (3)rA:rC(II) s:s Cis , (4)rA:rC(I) S:S Cis, (5)rA:rC S:S Trans, (6)rA:rG(II) S:S Cis, (7)rA:rG(I) S:S Cis, (8)rA:rG S:S Trans, (9)rA:rU(II) S:S Cis, (10)rA:rU(I) S:S Cis, (11)rA:rU S:S Trans, (12)rC:rC S:S Cis, (13)rC:rU(II) S:S Cis, (14)rC:rU(I) S:S Cis, (15)rG:rC(II) S:S Cis, (16)rG:rC(I) S:S Ci, (17)rG:rC S:S Trans, (18)rG:rG S:S Cis, (19)rG:rG S:S Trans, (20)rGrU(II)rUrG S:S Cis, (21)rG:rU(I) S:S Cis, (22)rG:rU S:S Trans, (23)rU:rU S:S Cis

List of the natural base pairs selected for this work

- **WW base pairs**: (1) A:A W:W Cis, (2) A:A W:W Trans, (3) A:C W:W Trans, (4) A:G W:W Cis, (5) A:U W:W Cis, (6) A:U W:W Trans, (7) C:C W:W Trans, (8) C:U W:W Cis, (9) C:U W:W Trans, (10) G:C W:W Cis, (11) G:C W:W Trans, (12) G:G W:W Trans, (13) G:U W:W Cis, (14) G:U W:W Trans, (15) U:U W:W Cis, (16) U:U(I) W:W Trans, (17) U:U(II) W:W Trans.

- **HH base pairs**: (1) A:A H:H Cis, (2) A:A H:H Trans, (3) A:C H:H Trans, (4) A:G H:H Cis, (5) A:U H:H Cis, (6) A:U(I) H:H Trans, (7) A:U(II) H:H Trans, (8) C:U H:H Trans, (9) G:C H:H Cis, (10) G:C(I) H:H Cis, (11) G:C(II) H:H Cis, (12) G:G H:H Trans

- **WH base pairs**: (1) A:A W:H Cis, (2) A:A W:H Trans, (3) A:U W:H Cis, (4) C:A W:H Cis, (5) C:A W:H Trans, (6) C:C W:H Cis, (7) C:C W:H Trans,(8) G:A W:H Cis, (9) G:A W:H Trans, (10) G:G W:H Cis, (11) G:G W:H Trans, (12) G:U W:H Trans,

(13) U:A W:H Cis, (14) U:G W:H Cis, (15) U:G W:H Trans, (16) U:U W:H Cis, (17) U:U W:H Trans

- **HS base pairs**:(1) A:rA H:S Cis, (2) A:rA H:S Trans, (3) A:rC H:S Cis, (4) A:rG H:S Cis, (5) A:rG H:S Trans, (6) A:rU H:S Cis, (7) A:rU H:S Trans, (8) C:rA H:S Trans, (9) C:rC H:S Cis, (10) C:rC H:S Trans, (11) C:rG H:S Cis, (12) C:rU H:S Cis, (13) C:rU H:S Trans, (14) G:A H:S Trans, (15) G:rA H:S Cis, (16) G:rG H:S Cis, (17) G:rU H:S Cis, (18) U:A H:S Trans, (19) U:G H:S Trans, (20) U:rA H:S Cis, (21) U:rC H:S Cis, (22) U:rG H:S Cis

- **WS base pairs** (1) A:rA W:S Trans, (2) A:rC W:S Cis, (3) A:rU W:S Cis, (4) A:rU W:S Trans, (5) C:rA W:S Cis, (6) C:rC W:S Cis, (7) C:rC W:S Trans, (8) C:rG W:S Cis, (9) C:rG W:S Trans, (10) C:rU W:S Cis, (11) C:rU W:S Trans, (12) G:rA W:S Cis, (13) G:rA W:S Trans, (14) G:rC W:S Trans, (15) G:rG W:S Trans, (16) G:rU W:S Trans, (17) U:rA W:S Cis, (18) U:rC W:S Cis, (19) U:rG W:S Cis, (20) U:rG W:S Trans, (21) U:rU W:S Cis, (22) U:rU W:S Trans

- **SS base pair** (1) rA:rA S:S Cis, (2) rArC(II)rCrAss Cis, (3) rA:rC(I) S:S Cis, (4) rArG(II)rGrA S:S Cis, (5) rA:rG S:S Trans, (6) rArU(II)rUrA S:S Cis, (7) rA:rU(I) S:S Cis, (8) rC:rU(I) S:S Cis, (9) rG:rC S:S Trans, (10) rG:rG S:S Cis, (11) rG:rG S:S Trans, (12) rGrU(II) S:S Cis, (13) rG:rU(I) S:S Cis

Details of E_{HB} calculation

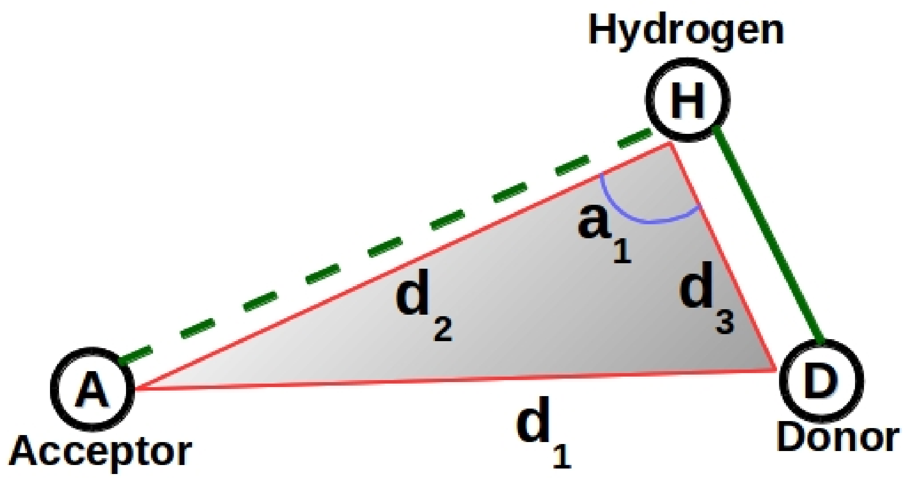


Figure S1: Representation of the geometry of a H-bond in the form of triangle.

Table S1: Red shift observed in different types of H-bonds. Calculations are performed at B3LYP-D3(BJ)/6-31+G(d,p) level. The values given within parenthesis denote the corresponding H-bonding energy in kcal mol^{-1}, as calculated from the Iogansen's relationship.

Sl.	H-bond type	Donor	Acceptor	Notation	Red shift (cm^{-1})			
					Min	Max	Average	Std. Dev.
1	N-H···N	primary N	imino N	N_I-H···N_{III}	47.6 (0.9)	434.7 (6.6)	221.3	99.4
2	N-H···N	secondary N	imino N	N_{II}-H···N_{III}	220.7 (4.4)	649.6 (8.1)	424.0	136.0
3	N-H···O	primary N	carbonyl O	N_I-H···O_c	41.9 (0.5)	475.8 (6.9)	147.0	88.8
4	N-H···O	secondary N	carbonyl O	N_{II}-H···O_c	93.2 (2.4)	496.7 (7.1)	314.1	123.2
5	N-H···O	primary N	hydroxyl O	N_I-H···O_h	41.9 (0.5)	249.4 (4.8)	120.6	65.8
6	O-H···N	hydroxyl O	imino N	O_h -H···N_{III}	135.5 (3.2)	509.6 (7.2)	314.6	113.3
7	O-H···O	hydroxyl O	carbonyl O	O_h-H···O_c	149.3 (3.5)	362.4 (5.9)	258.5	86.2
8	O-H···O	hydroxyl O	hydroxyl O	O_h-H···O_h	44.8 (0.7)	190.3 (4.0)	120.2	52.0

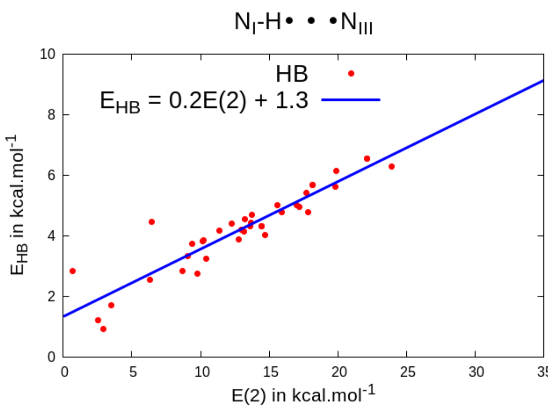
Table S2: Average values different parameters. Values calculated in B3LYP-D3(BJ) functional are given in parenthesis.

Sl.	Type of HB	Geometry	AIM parameters					NBO	Geometrical parameters					
		of BP	ρ	$\nabla^2\rho$	V	G	H_{tot}	E(2)	D-A	H-A	D-H	∠D-H-A	S_{DHA}	$\triangle_{DHA}$
1	N_I-H$\cdots N_{III}$	All	0.026	0.064	-0.017	0.016	-0.0005	13.74	3.02	2.04	1.02	165.37	6.08	0.26
			(0.025)	(0.06)	(-0.017)	(0.016)	(-0.0005)	(13.04)	(3.04)	(2.06)	(1.02)	(164.39)	(6.12)	(0.29)
		NS	0.027	0.065	-0.018	0.017	-0.0006	14.55	3.01	2.02	1.02	165.94	6.05	0.25
			(0.026)	(0.063)	(-0.017)	(0.017)	(-0.0006)	(13.84)	(3.03)	(2.03)	(1.02)	(159.5)	(6.17)	(0.28)
		S	0.025	0.069	-0.016	0.016	-0.0006	12.63	3.04	2.05	1.02	164.58	6.12	0.28
			(0.024)	(0.059)	(-0.016)	(0.015)	(-0.0004)	(11.82)	(3.056)	(2.096)	(1.02)	(159.5)	(6.17)	(-0.38)
2	N_{II}-H$\cdots N_{III}$	All	0.034	0.078	-0.022	0.020	-0.001	20.83	2.94	1.91	1.04	171.68	5.89	0.145
			(0.03)	(0.077)	(-0.022)	(0.02)	(-0.001)	(20.2)	(2.95)	(1.92)	(1.04)	(171.87)	(5.91)	(0.145)
		NS	0.036	0.084	-0.024	0.022	-0.001	23.7	2.9	1.88	1.04	172.99	5.82	0.12
			(0.0.036)	(0.0.084)	(-0.024)	(0.022)	(-0.001)	(23.55)	(2.9)	(1.88)	(1.04)	(172.95)	(5.83)	(0.124)
		S	0.027	0.063	-0.02	0.017	-0.0001	13.94	3.02	2.0	1.03	168.54	6.05	0.21
			(0.025)	(0.059)	(-0.016)	(0.016)	(-0.0007)	(11.8)	(3.05)	(2.04)	(1.03)	(169.16)	(6.12)	(0.20)
3	N_I-H$\cdots O_c$	All	0.024	0.073	-0.018	0.018	-0.00001	11.5	2.94	1.97	1.02	162.97	5.93	0.29
			(0.026)	(0.077)	(-0.020)	(0.019)	(-0.0001)	(12.48)	(2.92)	(1.94)	(1.02)	(164.35)	(5.88)	(0.26)
		NS	0.026	0.073	-0.019	0.019	-0.0004	14.08	2.95	1.97	1.02	167.27	5.94	0.22
			(0.028)	(0.079)	(-0.02)	(0.02)	(-0.0005)	(15.4)	(2.93)	(1.93)	(1.02)	(170.68)	(5.88)	(0.35)
		S	0.023	0.072	-0.018	0.018	0.0002	9.16	2.93	1.98	1.02	159.09	5.92	0.351
			(0.024)	(0.076)	(-0.019)	(0.019)	(-0.0001)	(10.14)	(2.91)	(1.96)	(1.02)	(159.28)	(5.88)	(0.35)
4	N_{II}-H$\cdots O_c$	All	0.033	0.097	-0.024	0.024	-0.0001	19.46	2.85	1.84	1.03	167.24	5.72	0.21
			(0.032)	(0.095)	(-0.024)	(0.024)	(-0.0001)	(18.88)	(2.86)	(1.86)	(1.03)	(166.48)	(5.75)	(0.23)
		NS	0.034	0.098	-0.024	0.024	-0.0002	22.28	2.85	1.82	1.03	172.765	5.7	0.12
			(0.034)	(0.0977)	(-0.025)	(0.024)	(-0.0003)	(22.28)	(2.85)	(1.82)	(1.03)	(172.75)	(5.7)	(0.12)
		S	0.03	0.097	-0.023	0.024	0.0003	14.19	2.85	1.89	1.03	156.89	5.76	0.39
			(0.03)	(0.09)	(-0.022)	(0.023)	(0.0003)	(13.21)	(2.88)	(1.925)	(1.03)	(156.01)	(5.83)	(0.41)
5	N_I-H$\cdots O_h$	All	0.019	0.055	-0.015	0.014	-0.001	7.38	3.01	2.13	1.01	146.85	6.15	0.58
			(0.02)	(0.056)	(-0.016)	(0.015)	(-0.001)	(7.86)	(3.01)	(2.1)	(1.01)	(151.35)	(6.12)	(0.5)
6	O_h -H$\cdots N_{III}$	All	0.041	0.095	-0.028	0.026	-0.002	25.785	2.78	1.81	0.99	167.925	5.58	0.19
			(0.043)	(0.098)	(-0.03)	(0.027)	(-0.003)	(26.762)	(2.76)	(1.8)	(0.99)	(166.45)	(5.55)	(0.21)
7	O_h-H$\cdots O_c$	All	0.033	0.104	-0.025	0.026	0.0004	18.115	2.88	1.95	0.98	162.23	5.81	0.32
			(0.04)	(0.12)	(-0.031)	(0.03)	(-0.0001)	(21.96)	(2.77)	(1.82)	(0.99)	(166.636)	(5.58)	(0.23)
8	O_h-H$\cdots O_h$	All	0.029	0.086	-0.022	0.022	-0.0006	14.3	2.83	1.96	0.99	150.5	5.76	0.45
			(0.032)	(0.09)	(-0.025)	(0.024)	(-0.0006)	(16.46)	(2.79)	(1.9)	(0.98)	(152.77)	(5.68)	(0.43)

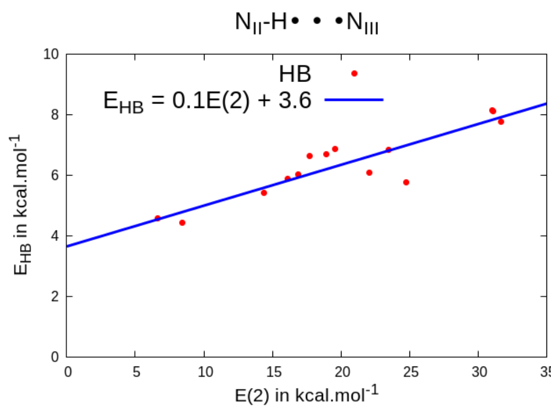
Single variable linear regression analysis

Linear relationships ($y = Ax+B$) between H-bonding strength and the individual parameters are derived from single variable linear regression analysis. Only those parameters which have high Pearson correlation coefficient (r) for all the four types of H-bonds in Table 1, are selected. In this process, 5 topological parameters (ρ, $\nabla^2\rho$, V, G and H), 1 charge transfer based parameter (E(2)) and 6 geometry based parameters (H-A distance, D-H distance, D-A distance, $\angle DHA$, S_{DHA} and Δ_{DHA}) have been considered as independent variables (x) and E_{HB} has been considered as the scalar dependent variable (y). The results are shown in the following figures (Figure S2 – S13).

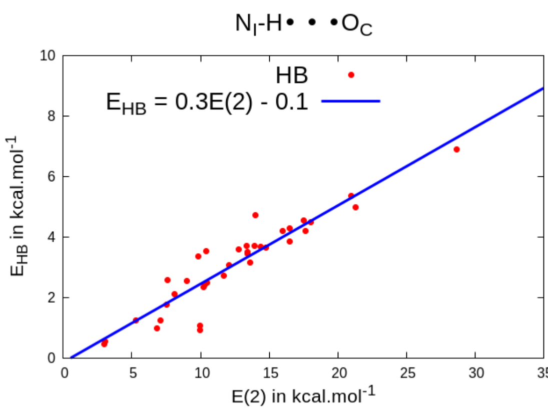
The linear regression analyses have been performed using the least square fit method as implemented in GNUPLOT package (version 5.0). The standard deviation of the residuals (σ) are reported for each case in the following figures (Figure S2 – S13). In those figures, the straight line obtained after linear regression analysis (blue continuous line) and the distribution of the raw data points (red filled circles) are shown for each parameter. Equation of the fitted straight line is also mentioned in each case. The statistical errors involved in calculation of the slope and y-intercept of all these straight lines are mentioned in Table S3 in details. Note that the analysis have been performed over all the four important H-bond types, viz. $N_I \cdots N_{III}$ (top left panel), $N_{II} \cdots N_{III}$ (top right panel), $N_I \cdots O_C$ (bottom left panel) and $N_{II} \cdots O_C$ (bottom right panel).



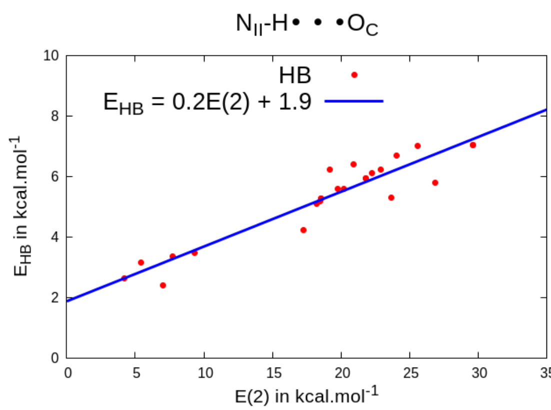
(a) σ of residuals = 0.52



(b) σ of residuals = 0.50

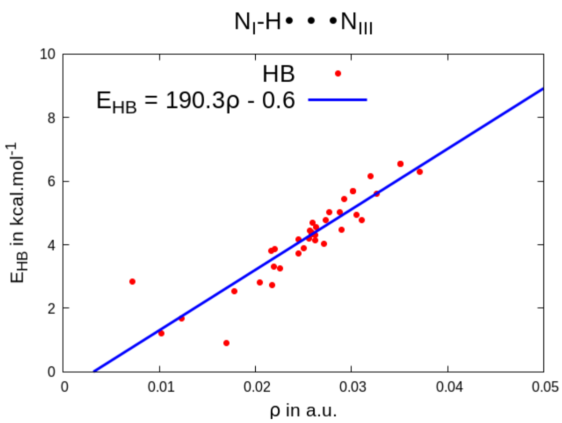


(c) σ of residuals = 0.55

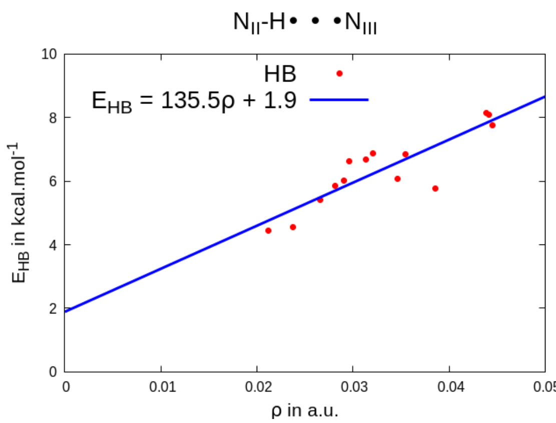


(d) σ of residuals = 0.47

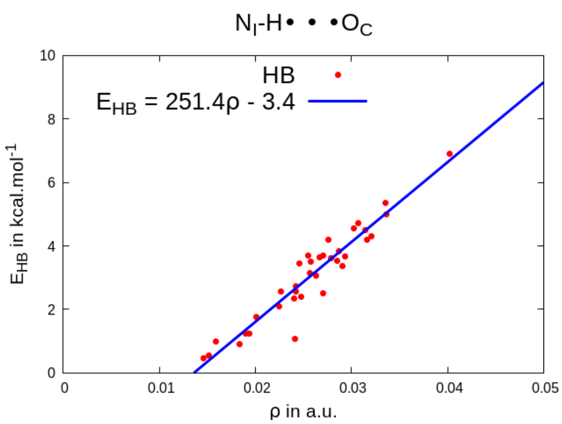
Figure S2: Single variable linear regression analysis between E$_{HB}$ vs E(2) for (a) N$_I$-H$\cdots$N$_{III}$, (b) N$_{II}$-H$\cdots$N$_{III}$, (a) N$_I$-H$\cdots$O$_C$ and (d) N$_{II}$-H$\cdots$O$_C$ type H-bonds.



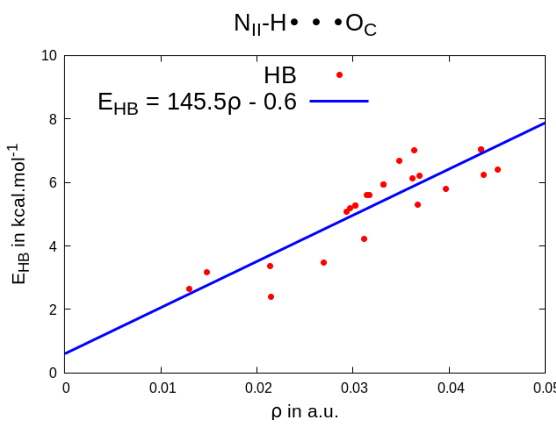
(a) σ of residuals = 0.57



(b) σ of residuals = 0.58

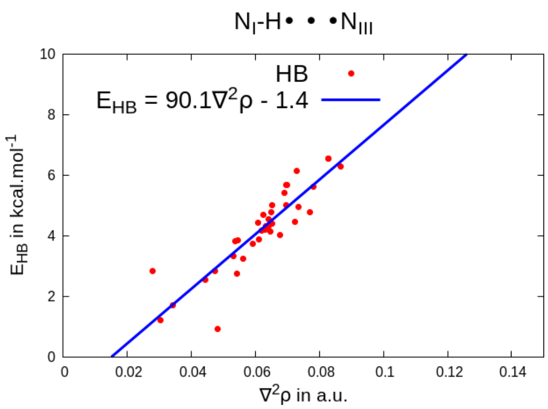


(c) σ of residuals = 0.45

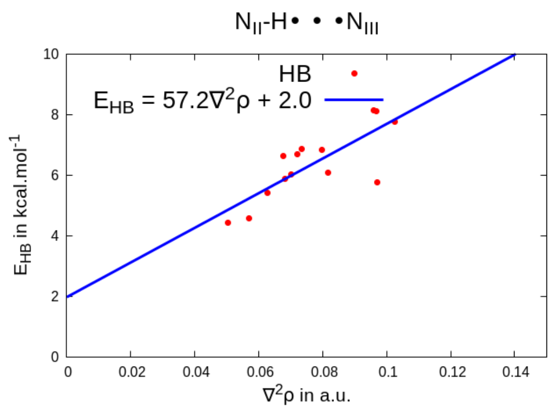


(d) σ of residuals = 0.65

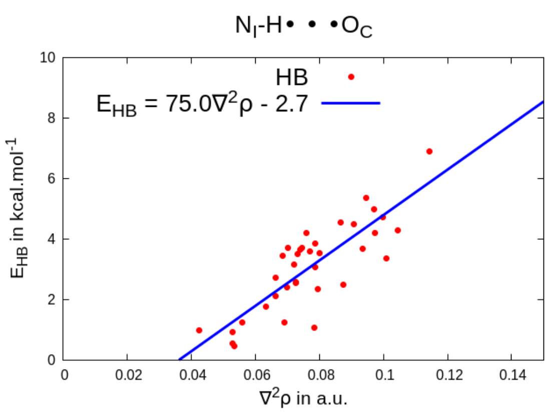
Figure S3: Single variable linear regression analysis between E$_{HB}$ vs ρ for (a) N$_I$-H$\cdots$N$_{III}$, (b) N$_{II}$-H$\cdots$N$_{III}$, (a) N$_I$-H$\cdots$O$_C$ and (d) N$_{II}$-H$\cdots$O$_C$ type H-bonds.



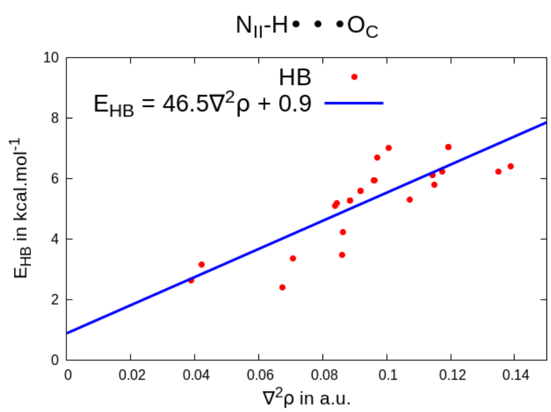
(a) σ of residuals = 0.62



(b) σ of residuals = 0.73

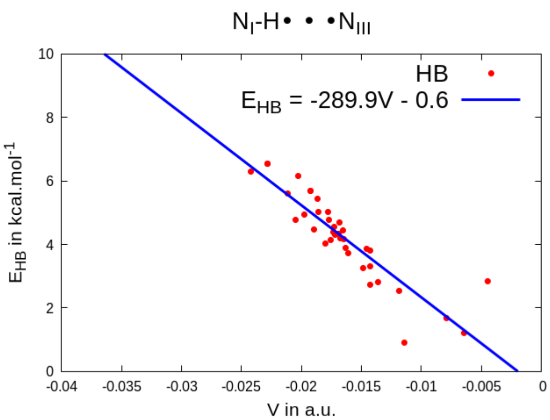


(c) σ of residuals = 0.83

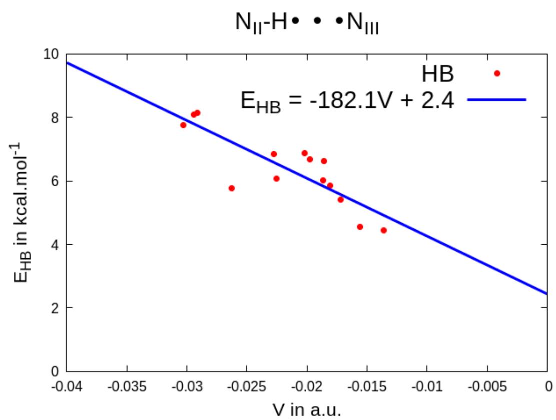


(d) σ of residuals = 0.80

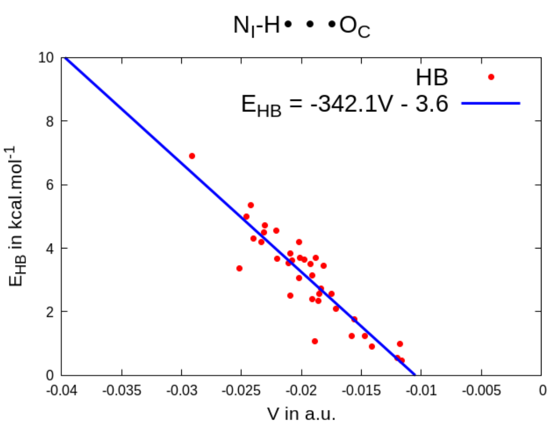
Figure S4: Single variable linear regression analysis between E_{HB} vs $\nabla^2\rho$ for (a) N_I-H$\cdots N_{III}$, (b) N_{II}-H$\cdots N_{III}$, (a) N_I-H$\cdots O_C$ and (d) N_{II}-H$\cdots O_C$ type H-bonds.



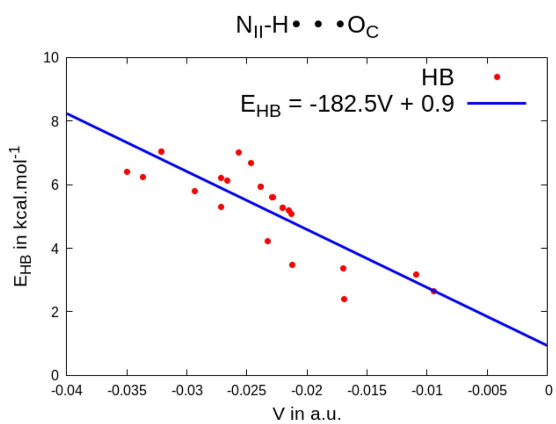
(a) σ of residuals = 0.61



(b) σ of residuals = 0.65

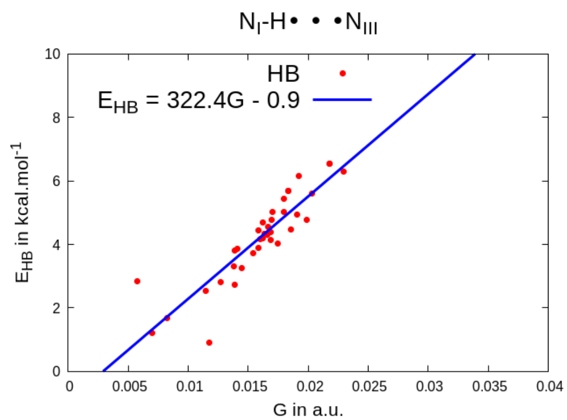


(c) σ of residuals = 0.62

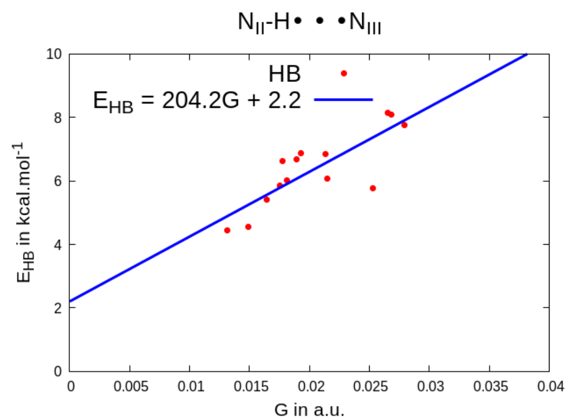


(d) σ of residuals = 0.78

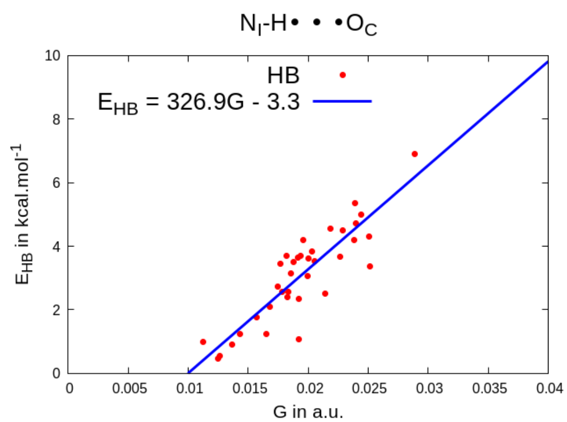
Figure S5: Single variable linear regression analysis between E_{HB} vs V for (a) N_I-H$\cdots N_{III}$, (b) N_{II}-H$\cdots N_{III}$, (a) N_I-H$\cdots O_C$ and (d) N_{II}-H$\cdots O_C$ type H-bonds.



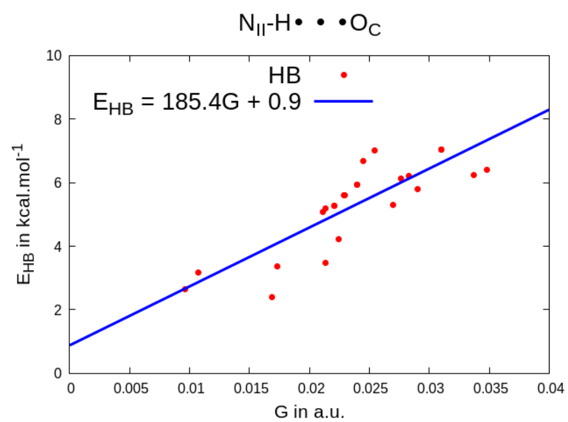
(a) σ of residuals = 0.61



(b) σ of residuals = 0.68

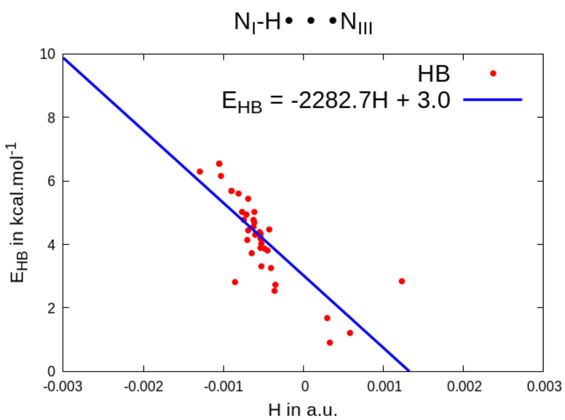


(c) σ of residuals = 0.71

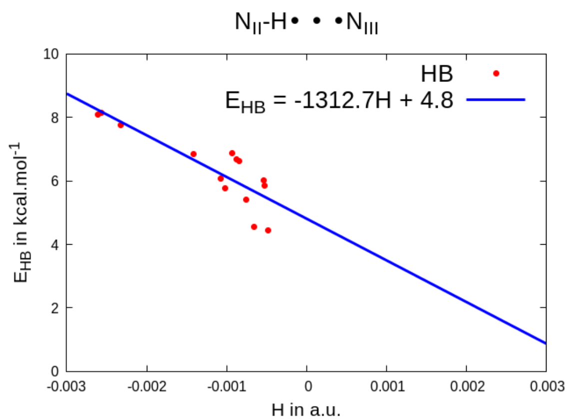


(d) σ of residuals = 0.79

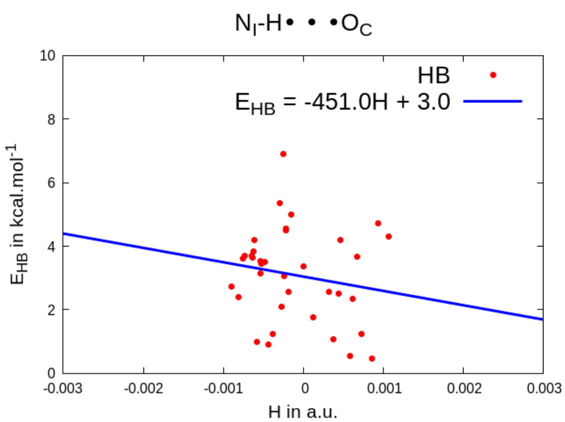
Figure S6: Single variable linear regression analysis between E_{HB} vs G for (a) N_I-H$\cdots N_{III}$, (b) N_{II}-H$\cdots N_{III}$, (a) N_I-H$\cdots O_C$ and (d) N_{II}-H$\cdots O_C$ type H-bonds.



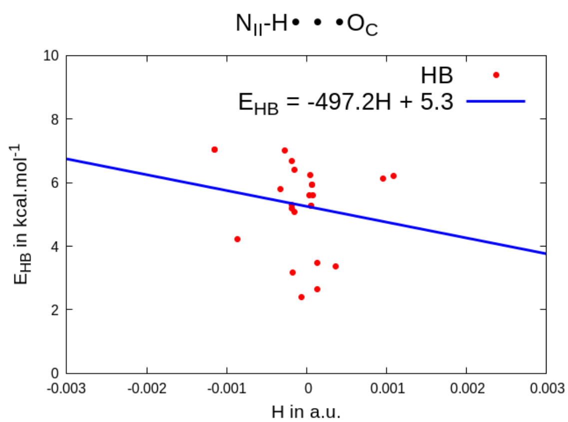
(a) σ of residuals = 0.84



(b) σ of residuals = 0.63

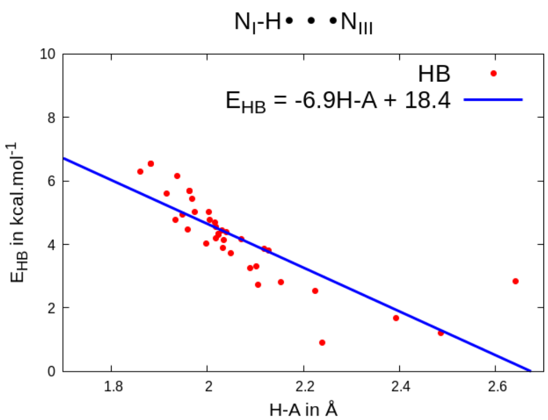


(c) σ of residuals = 1.46

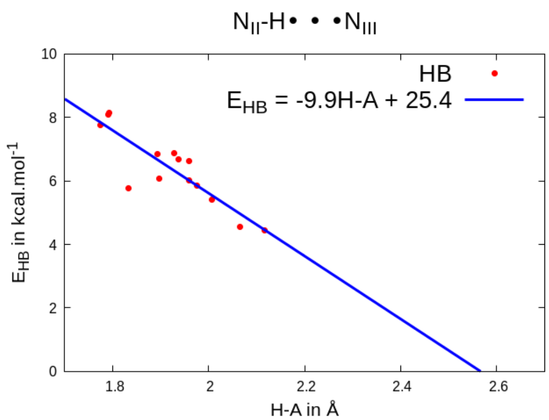


(d) σ of residuals = 1.39

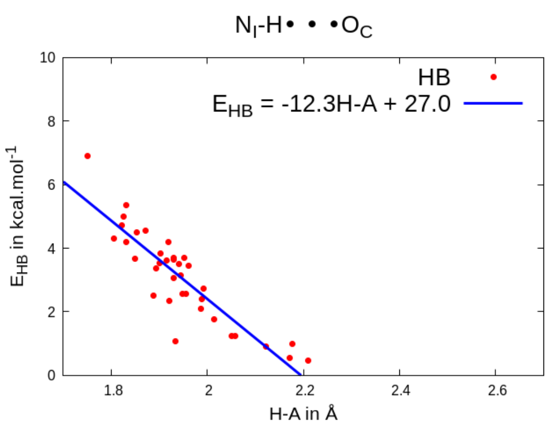
Figure S7: Single variable linear regression analysis between E_{HB} vs H_{tot} for (a) N_I-H$\cdots N_{III}$, (b) N_{II}-H$\cdots N_{III}$, (a) N_I-H$\cdots O_C$ and (d) N_{II}-H$\cdots O_C$ type H-bonds.



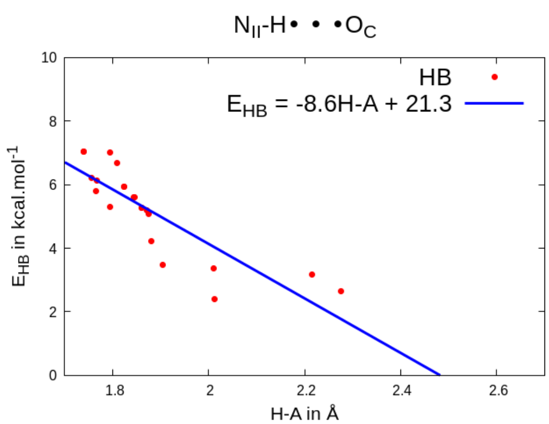
(a) σ of residuals = 0.78



(b) σ of residuals = 0.58

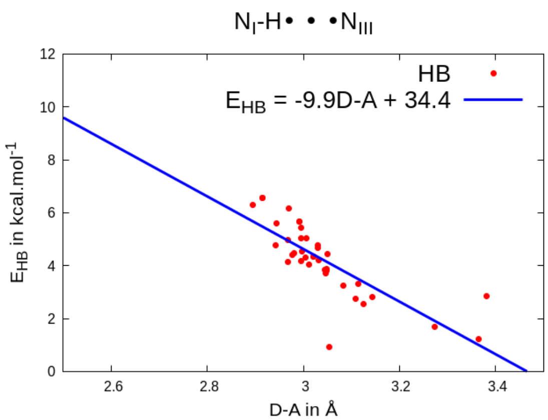


(c) σ of residuals = 0.68

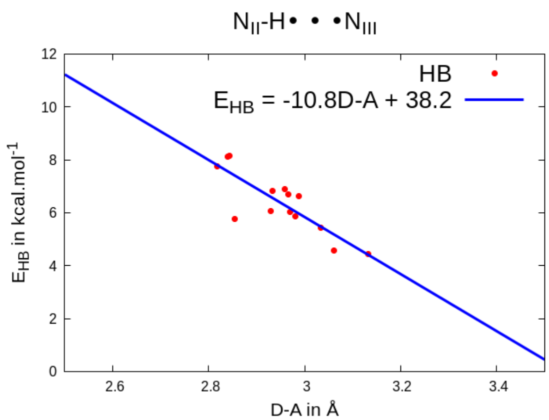


(d) σ of residuals = 0.76

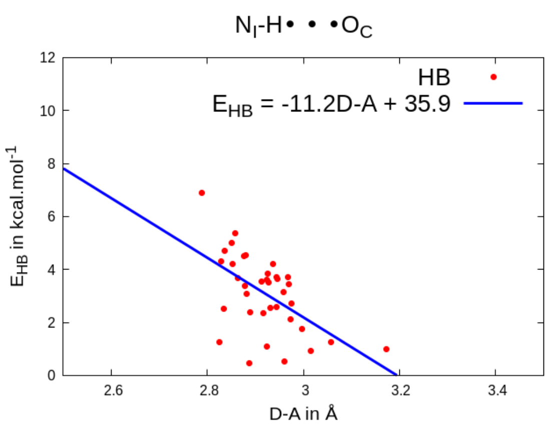
Figure S8: Single variable linear regression analysis between E_{HB} vs HA for (a) N_I-H$\cdots N_{III}$, (b) N_{II}-H$\cdots N_{III}$, (a) N_I-H$\cdots O_C$ and (d) N_{II}-H$\cdots O_C$ type H-bonds.



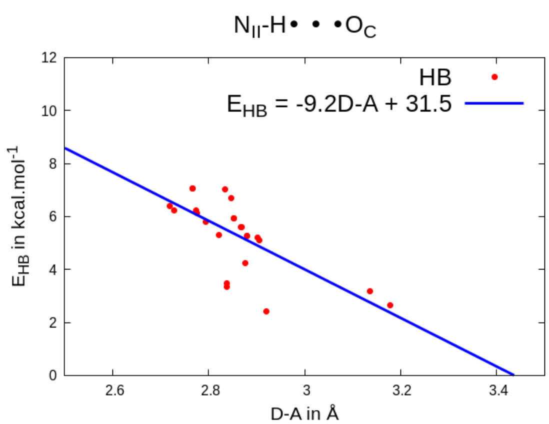
(a) σ of residuals = 0.85



(b) σ of residuals = 0.65

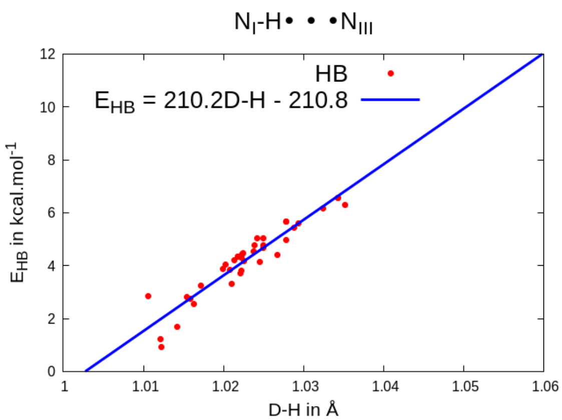


(c) σ of residuals = 1.23

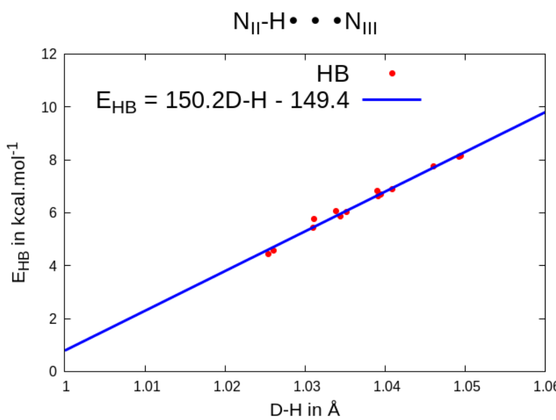


(d) σ of residuals = 1.00

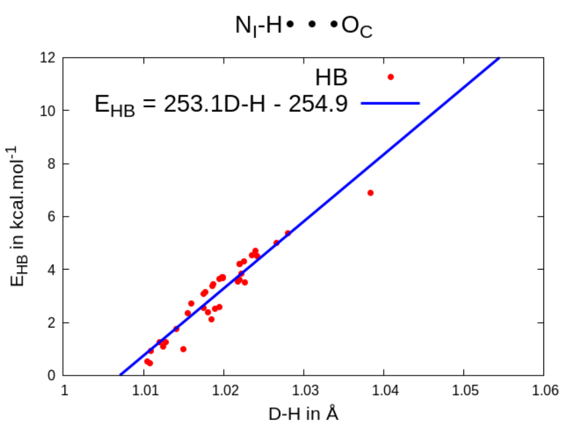
Figure S9: Single variable linear regression analysis between E$_{HB}$ vs DA for (a) N$_I$-H$\cdots$N$_{III}$, (b) N$_{II}$-H$\cdots$N$_{III}$, (a) N$_I$-H$\cdots$O$_C$ and (d) N$_{II}$-H$\cdots$O$_C$ type H-bonds.



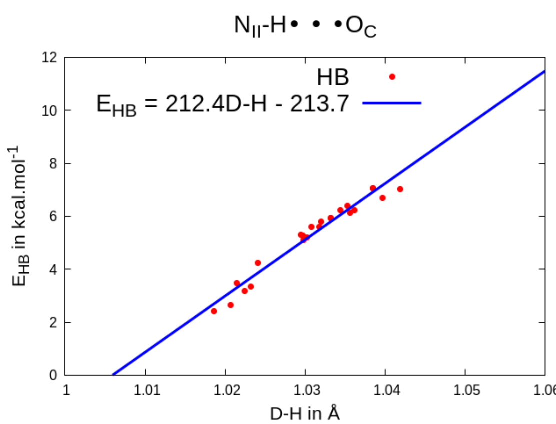
(a) σ of residuals = 0.44



(b) σ of residuals = 0.14

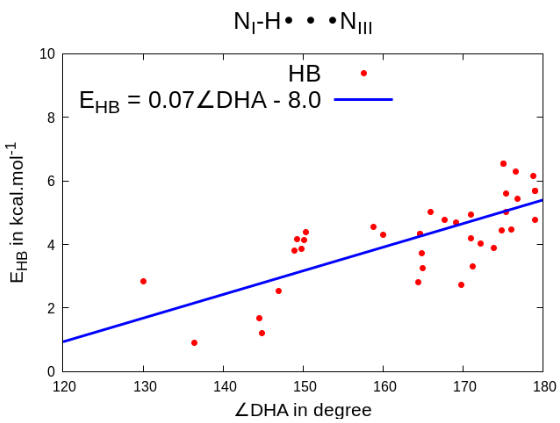


(c) σ of residuals = 0.45

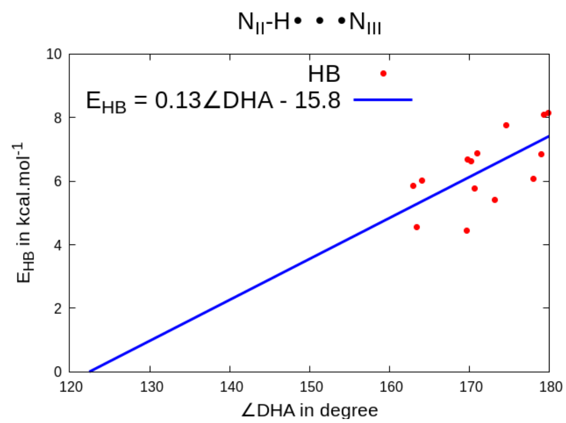


(d) σ of residuals = 0.30

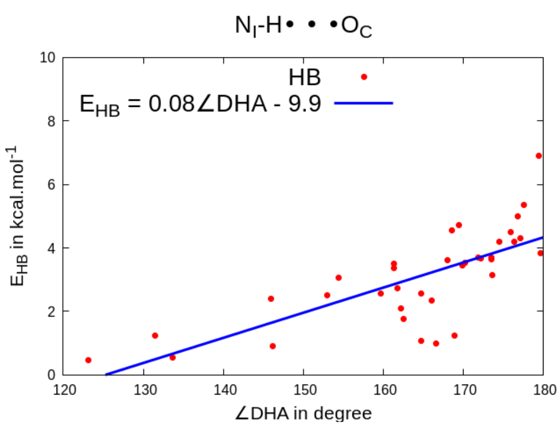
Figure S10: Single variable linear regression analysis between E_{HB} vs DH for (a) N_I-H$\cdots N_{III}$, (b) N_{II}-H$\cdots N_{III}$, (a) N_I-H$\cdots O_C$ and (d) N_{II}-H$\cdots O_C$ type H-bonds.



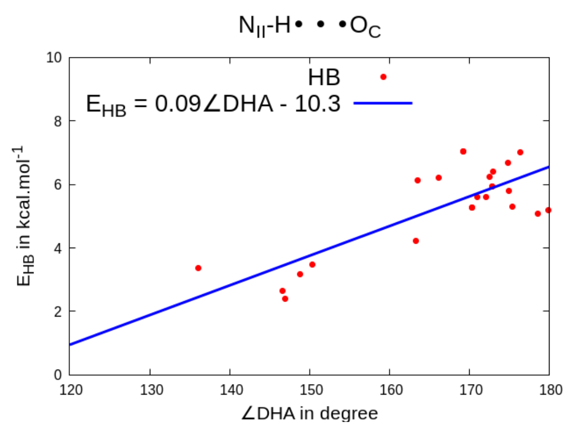
(a) σ of residuals = 0.94



(b) σ of residuals = 0.92

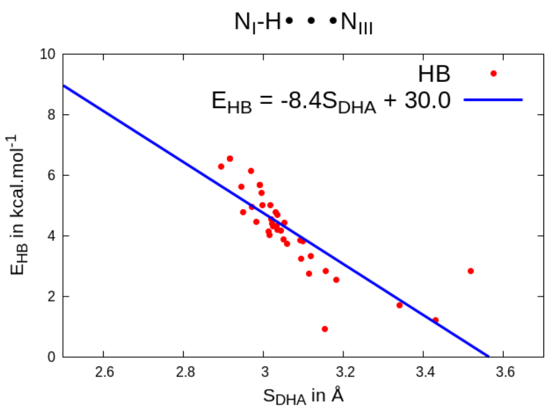


(c) σ of residuals = 1.00

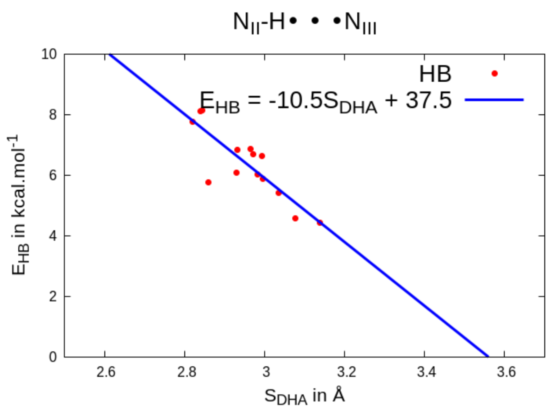


(d) σ of residuals = 0.86

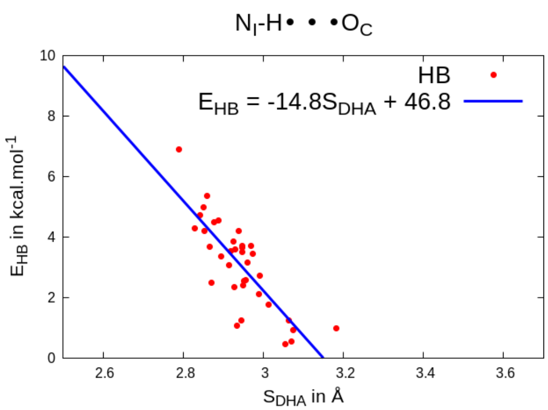
Figure S11: Single variable linear regression analysis between E$_{HB}$ vs $\angle DHA$ for (a) N$_I$-H$\cdots$N$_{III}$, (b) N$_{II}$-H$\cdots$N$_{III}$, (a) N$_I$-H$\cdots$O$_C$ and (d) N$_{II}$-H$\cdots$O$_C$ type H-bonds.



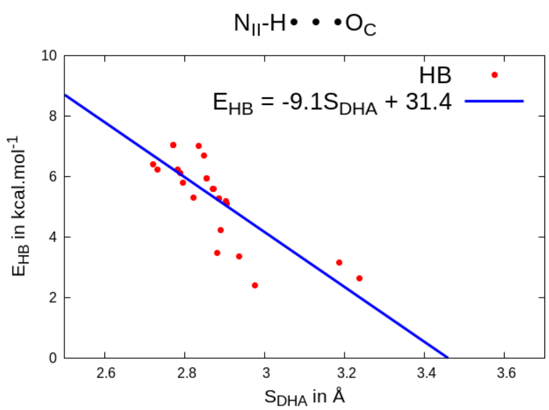
(a) σ of residuals = 0.81



(b) σ of residuals = 0.64

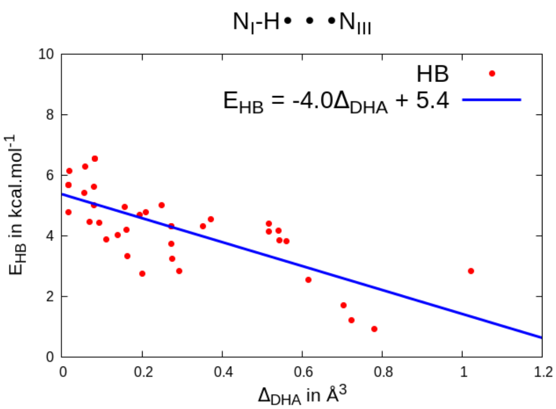


(c) σ of residuals = 0.87

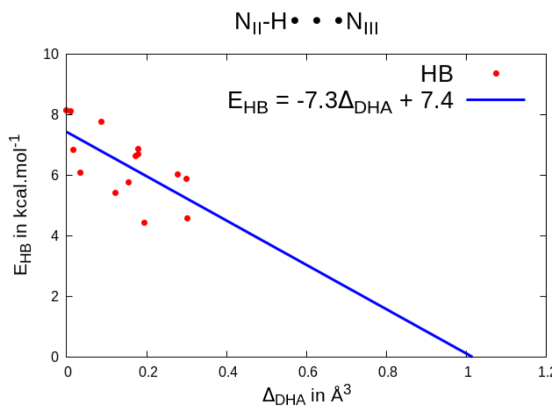


(d) σ of residuals = 0.86

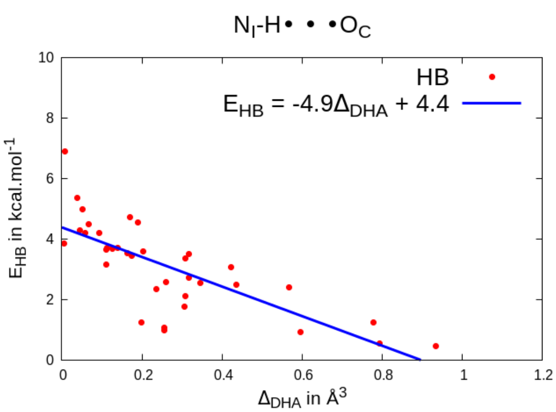
Figure S12: Single variable linear regression analysis between E_{HB} vs S_{DHA} for (a) N_{I}-$H\cdots N_{III}$, (b) N_{II}-$H\cdots N_{III}$, (a) N_{I}-$H\cdots O_{C}$ and (d) N_{II}-$H\cdots O_{C}$ type H-bonds.



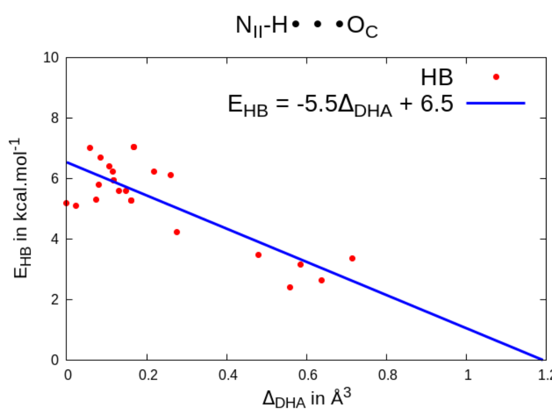
(a) σ of residuals = 0.92



(b) σ of residuals = 0.90



(c) σ of residuals = 0.97



(d) σ of residuals = 0.82

Figure S13: Single variable linear regression analysis between E_{HB} vs $\triangle_{DHA}$ for (a) N_I-H$\cdots N_{III}$, (b) N_{II}-H$\cdots N_{III}$, (a) N_I-H$\cdots O_C$ and (d) N_{II}-H$\cdots O_C$ type H-bonds.

Table S3: For each $E_{HB} = Ax + B$ type linear relationships obtained from single variable linear regression analysis, the slope (A) and y-intercept (B) are reported along with the respective standard deviations.

Parameter	H-bond	slope (A)	y-intercept (B)
E(2)	N_I-H$\cdots N_{III}$	0.22 ± 0.02 (6.871%)	1.33 ± 0.22 (16.24%)
	N_{II}-H$\cdots N_{III}$	0.13 ± 0.02 (13.3%)	3.65 ± 0.39 (10.58%)
	N_I-H$\cdots O_C$	0.26 ± 0.02 (6.78%)	-0.14 ± 0.24 (173.5%)
	N_{II}-H$\cdots O_C$	0.18 ± 0.01 (7.527%)	1.88 ± 0.27 (14.65%)
ρ	N_I-H$\cdots N_{III}$	190.34 ± 14.68 (7.713%)	-0.60 ± 0.38 (64.56%)
	N_{II}-H$\cdots N_{III}$	135.45 ± 21.24 (15.68%)	1.89 ± 0.72 (38.04%)
	N_I-H$\cdots O_C$	251.43 ± 13.83 (5.501%)	-3.42 ± 0.37 (10.71%)
	N_{II}-H$\cdots O_C$	145.50 ± 16.13 (11.08%)	0.60 ± 0.54 (88.94%)
$\nabla^2\rho$	N_I-H$\cdots N_{III}$	90.09 ± 7.60 (8.437%)	-1.35 ± 0.48 (35.68%)
	N_{II}-H$\cdots N_{III}$	57.20 ± 12.58 (22%)	1.98 ± 0.99 (49.94%)
	N_I-H$\cdots O_C$	75.03 ± 8.71 (11.61%)	-2.71 ± 0.69 (25.4%)
	N_{II}-H$\cdots O_C$	46.48 ± 6.76 (14.55%)	0.88 ± 0.66 (74.91%)
V	N_I-H$\cdots N_{III}$	-289.87 ± 24.38 (8.409%)	-0.57 ± 0.42 (73.63%)
	N_{II}-H$\cdots N_{III}$	-182.15 ± 33.96 (18.64%)	2.44 ± 0.75 (30.89%)
	N_I-H$\cdots O_C$	-342.10 ± 26.83 (7.842%)	-3.59 ± 0.53 (14.87%)
	N_{II}-H$\cdots O_C$	-182.50 ± 25.84 (14.16%)	0.94 ± 0.64 (67.9%)
G	N_I-H$\cdots N_{III}$	322.43 ± 26.91 (8.348%)	-0.93 ± 0.44 (47.44%)
	N_{II}-H$\cdots N_{III}$	204.23 ± 40.64 (19.9%)	2.20 ± 0.85 (38.49%)
	N_I-H$\cdots O_C$	326.91 ± 30.77 (9.411%)	-3.26 ± 0.61 (18.7%)
	N_{II}-H$\cdots O_C$	185.39 ± 26.33 (14.2%)	0.88 ± 0.65 (73.25%)
H_{tot}	N_I-H$\cdots N_{III}$	-2282.70 ± 299.70 (13.13%)	3.03 ± 0.21 (6.926%)
	N_{II}-H$\cdots N_{III}$	-1312.72 ± 231.50 (17.64%)	4.81 ± 0.32 (6.687%)
	N_I-H$\cdots O_C$	-451.05 ± 439.30 (97.39%)	3.05 ± 0.25 (8.15%)
	N_{II}-H$\cdots O_C$	-497.20 ± 572.50 (115.2%)	5.26 ± 0.29 (5.47%)
H-A	N_I-H$\cdots N_{III}$	-6.90 ± 0.81 (11.71%)	18.44 ± 1.67 (9.042%)
	N_{II}-H$\cdots N_{III}$	-9.90 ± 1.58 (15.95%)	25.42 ± 3.04 (11.97%)
	N_I-H$\cdots O_C$	-12.32 ± 1.10 (8.911%)	27.04 ± 2.14 (7.903%)
	N_{II}-H$\cdots O_C$	-8.57 ± 1.17 (13.7%)	21.28 ± 2.21 (10.38%)
D-A	N_I-H$\cdots N_{III}$	-9.94 ± 1.31 (13.21%)	34.44 ± 3.99 (11.59%)
	N_{II}-H$\cdots N_{III}$	-10.78 ± 2.01 (18.61%)	38.18 ± 5.92 (15.51%)
	N_I-H$\cdots O_C$	-11.25 ± 2.83 (25.13%)	35.94 ± 8.26 (22.97%)
	N_{II}-H$\cdots O_C$	-9.17 ± 1.95 (21.3%)	31.50 ± 5.59 (17.73%)
D-H	N_I-H$\cdots N_{III}$	210.20 ± 11.88 (5.653%)	-210.78 ± 12.16 (5.767%)
	N_{II}-H$\cdots N_{III}$	150.20 ± 5.17 (3.44%)	-149.41 ± 5.36 (3.587%)
	N_I-H$\cdots O_C$	253.08 ± 13.98 (5.525%)	-254.87 ± 14.25 (5.592%)

	N_{II}-H$\cdots$$O_C$	212.44 ± 9.71 (4.569%)	-213.70 ± 10.01 (4.682%)
∠DHA	N_I-H$\cdots$$N_{III}$	0.07 ± 0.01 (15.82%)	-7.99 ± 1.94 (24.29%)
	N_{II}-H$\cdots$$N_{III}$	0.13 ± 0.04 (33.77%)	-15.78 ± 7.48 (47.43%)
	N_I-H$\cdots$$O_C$	0.08 ± 0.01 (15.65%)	-9.91 ± 2.04 (20.6%)
	N_{II}-H$\cdots$$O_C$	0.09 ± 0.02 (16.32%)	-10.26 ± 2.54 (24.79%)
S_{DHA}	N_I-H$\cdots$$N_{III}$	-8.43 ± 1.03 (12.19%)	30.03 ± 3.15 (10.48%)
	N_{II}-H$\cdots$$N_{III}$	-10.55 ± 1.90 (18.06%)	37.54 ± 5.63 (15%)
	N_I-H$\cdots$$O_C$	-14.85 ± 1.84 (12.37%)	46.76 ± 5.41 (11.56%)
	N_{II}-H$\cdots$$O_C$	-9.08 ± 1.48 (16.32%)	31.39 ± 4.26 (13.58%)
Δ_{DHA}	N_I-H$\cdots$$N_{III}$	-3.95 ± 0.60 (15.14%)	5.37 ± 0.23 (4.231%)
	N_{II}-H$\cdots$$N_{III}$	-7.31 ± 2.39 (32.6%)	7.43 ± 0.42 (5.672%)
	N_I-H$\cdots$$O_C$	-4.89 ± 0.72 (14.78%)	4.39 ± 0.25 (5.705%)
	N_{II}-H$\cdots$$O_C$	-5.48 ± 0.82 (15.01%)	6.54 ± 0.25 (3.829%)

Table S4: Mean square error (MSE) and Root mean square error (RMSE) values (in kcal mol^{-1}) between the set of expected E_{HB} values and set of predicted E_{HB} values from different single parameter and multi parameter regression analysis.

Parameter	MSE			RMSE		
	weak	strong	all	weak	strong	all
V	0.40	0.47	0.45	0.63	0.68	0.67
G	0.49	0.77	0.69	0.70	0.89	0.83
H_{tot}	0.55	1.89	1.5	0.74	1.37	1.22
D-A	0.03	1.01	0.89	0.18	1.0	0.94
∠D-H-A	1.02	2.46	2.05	1.01	1.57	1.43
S_{DHA}	0.57	0.61	0.6	0.76	0.78	0.77
$\triangle_{DHA}$	1.15	1.18	1.17	1.07	1.08	1.08

Multi-variable linear regression analysis

Table S5: In the **3P model** we have performed multiple linear regression analysis taking three parameters, viz. ρ, E(2) and D-H. Performance of the model has ben estimaed on the basis of mean squared error (MSE) and varience score (R^2). Note that $R^2 = 1$ corresponds to perfect prediction. To understand the relative significance of the individual parameters 'importance analysis' has been carried out. The importance value is estimated as the difference in the corresponding R^2 values, when the corresponding variable is replaced with its average value.

H-bond		MSE	R^2	Importance
$N_I - H \cdots N_{III}$	3P model	0.15	0.92	
	ρ	0.2	0.84	0.15
	E(2)	0.19	0.85	0.14
	D-H	0.38	0.7	0.29
$N_{II} - H \cdots N_{III}$	3P model	0.0	1.0	
	ρ	0.02	0.97	0.03
	E(2)	0.03	0.97	0.03
	D-H	0.15	0.8	0.2
$N_I - H \cdots O_C$	3P model	0.1	0.95	
	ρ	0.23	0.86	0.09
	E(2)	0.23	0.85	0.1
	D-H	0.21	0.87	0.08
$N_{II} - H \cdots O_C$	3P model	0.02	0.99	
	ρ	0.06	0.93	0.06
	E(2)	0.06	0.93	0.06
	D-H	0.39	0.53	0.46

Table S6: In the **5P model** we have performed multiple linear regression analysis taking five parameters, viz. ρ, $\bigtriangledown^2\rho$, E(2), H-A and D-H. Performance of the model has ben estimaed on the basis of mean squared error (MSE) and varience score (R^2). Note that $R^2 = 1$ corresponds to perfect prediction. To understand the relative significance of the individual parameters 'importance analysis' has been carried out. The importance value is estimated as the difference in the corresponding R^2 values, when the corresponding variable is replaced with its average value.

H-bond		MSE	R^2	Importance
$N_I - H \cdots N_{III}$	5P model	0.14	0.92	
	ρ	0.16	0.87	0.05
	$\bigtriangledown^2\rho$	0.05	0.87	0.05
	E(2)	0.12	0.91	0.01
	H-A	0.19	0.85	0.07
	D-H	0.18	0.86	0.06
$N_{II} - H \cdots N_{III}$	5P model	0.0	1.0	
	ρ	0.1	0.98	0.02
	$\bigtriangledown^2\rho$	0.2	0.98	0.02
	E(2)	0.1	0.98	0.02
	H-A	0.2	0.97	0.03
	D-H	0.6	0.92	0.08
$N_I - H \cdots O_C$	5P model	0.09	0.95	
	ρ	0.16	0.9	0.05
	$\bigtriangledown^2\rho$	0.2	0.88	0.07
	E(2)	0.17	0.89	0.06
	H-A	0.2	0.88	0.07
	D-H	0.18	0.88	0.07
$N_{II} - H \cdots O_C$	5P model	0.02	0.99	
	ρ	0.05	0.94	0.05
	$\bigtriangledown^2\rho$	0.05	0.94	0.05
	E(2)	0.05	0.94	0.05
	H-A	0.05	0.94	0.05
	D-H	0.43	0.48	0.51

References

(1) Cech, T. R.; Steitz, J. A. The noncoding RNA revolution—trashing old rules to forge new ones. *Cell* **2014**, *157*, 77–94.

(2) Kruger, K.; Grabowski, P. J.; Zaug, A. J.; Sands, J.; Gottschling, D. E.; Cech, T. R. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **1982**, *31*, 147–157.

(3) Orgel, L. E. Evolution of the genetic apparatus. *J. Mol. Biol.* **1968**, *38*, 381–393.

(4) Garriga, G.; Lambowitz, A. M. RNA splicing in Neurospora mitochondria: self-splicing of a mitochondrial intron in vitro. *Cell* **1984**, *39*, 631–641.

(5) Michel, F.; Jacquier, A.; Dujon, B. Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. *Biochimie* **1982**, *64*, 867–881.

(6) Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, N.; Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **1983**, *35*, 849–857.

(7) Pley, H. W.; Flaherty, K. M.; McKay, D. B. hammerhead ribozyme. *Nature* **1994**, *372*, 3.

(8) Hampel, A.; Cowan, J. A unique mechanism for RNA catalysis: the role of metal cofactors in hairpin ribozyme cleavage. *Chem. Biol.* **1997**, *4*, 513–517.

(9) Taylor, J. M. Hepatitis delta virus. *Intervirology* **1999**, *42*, 173–178.

(10) Morris, K. V., Ed. *RNA and the Regulation of gene expression: A hidden layer of complexity*; Caister Academic Press, 2008.

(11) Bagasra, O.; Prilliman, K. R. RNA interference: the molecular immune system. *J. Mol. Histol.* **2004**, *35*, 545–553.

(12) others„ et al. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* **2000**, *408*, 86–89.

(13) Sabin, L. R.; Delás, M. J.; Hannon, G. J. Dogma derailed: the many influences of RNA on the genome. *Mol. Cell.* **2013**, *49*, 783–794.

(14) Ross, R. J.; Weiner, M. M.; Lin, H. PIWI proteins and PIWI-interacting RNAs in the soma. *Nature* **2014**, *505*, 353–359.

(15) Nudler, E.; Mironov, A. S. The riboswitch control of bacterial metabolism. *Trends. Biochem. Sci.* **2004**, *29*, 11–17.

(16) Grundy, F. J.; Henkin, T. M. The T box and S box transcription termination control systems. *Front. Biosci.* **2003**, *8*, d20–d31.

(17) Zhang, A.; Wassarman, K. M.; Ortega, J.; Steven, A. C.; Storz, G. The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. *Mol. Cell.* **2002**, *9*, 11–22.

(18) Leontis, N. B.; Westhof, E. Geometric nomenclature and classification of RNA base pairs. *RNA* **2001**, *7*, 499–512.

(19) Grabowski, S. J. A new measure of hydrogen bonding strength–ab initio and atoms in molecules studies. *Chem. Phys. Lett.* **2001**, *338*, 361–366.

(20) Espinosa, E.; Molins, E.; Lecomte, C. Hydrogen bond strengths revealed by topological analyses of experimentally observed electron densities. *Chem. Phys. Lett.* **1998**, *285*, 170–173.

(21) Nikolaienko, T. Y.; Bulavin, L. A.; Hovorun, D. M. Bridging QTAIM with vibrational spectroscopy: The energy of intramolecular hydrogen bonds in DNA-related biomolecules. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7441–7447.