THE OVERLAP POPULATION - A QUANTUM CHEMICAL CRITERION FOR THE HYDROGEN BOND ANALYSIS IN NUCLEIC ACIDS STRUCTURES

C. Bendic*

abstract: Up to now, due to the dimension of the biopolymers, and to the limited computer resources, only geometrical criteria were used to identify possible candidates for the attractive interactions like hydrogen bonds (HBexplore, HBplus programs). We have proposed Mulliken overlap population as a quantitative quantum chemical criterion, not only for the identification of possible interactions, but also for the evaluation of their relative strength. Overlap population (O.P) was calculated in the frame of the Extended Hückel method using SHB_interactions program. A comparison between the results obtained with this program and those obtained using geometrical criteria (HBexplore), was made. To this purpose, both RNA and a DNA structure (1rnk and 1dsc PDB codes) were used. The base-base H-bonds found in 1rnk and 1dsc using geometrical criteria have overlap population in the range 0.005-0.15. Our results show that there is a clear delimitation between H-bond overlap populations when the acceptor is an oxygen atom and those when the acceptor is a nitrogen atom, the D – H…N bonds having overlap populations greater than D – H…O bonds, for the same range of distances between hydrogen and acceptor atom. Overlap population criterion allows detecting the possible weaker H-bonds

Introduction

The hydrogen bonds are one of the most important intra- and intermolecular interactions in biological macromolecules [1], and are responsible for the structural and functional differences in RNA and DNA.

Up to now, due to the dimension of the biopolymers, and to the limited computer resources, only geometrical criteria were used to identify possible candidates for the attractive interactions. The tools available for H-bond analysis, like the corresponding modules in structure determination or modelling packages, do not have the flexibility required. Even the complex programs HBexplore [2] and HBPLUS [3] are limited to a selection of potential H-bonds, followed by a classification and a statistical analysis, based only on geometrical criteria.

Now we propose Mulliken overlap population (O.P) [4] as a quantitative quantum chemical criterion, not only for the identification of possible interactions, but also for the evaluation of their relative strength.

---

* Department of Physical Chemistry, University of Bucharest, 4-12 Bd. Elisabeta, Bucharest 70034, Romania

Analele Universităţii din Bucureşti – Chimie, Anul XIV (serie nouă), vol. I-II, pg. 233-240

The objective of the present paper was to make a comparative study between the results obtained using this quantum criterion and the results based only on using geometrical criteria (HBexplore).

Computer Software

Mulliken overlap populations were calculated in the frame of the Extended Hückel (EH) method [5] using SHB_interactions program [6]. This program was especially conceived in order to identify and analyse intermolecular interactions in biopolymers using overlap populations as selection criterion. The program can be obtained on request from the author. The source code for the SHB_interactions program, written in C, is available at http://gw-chimie.math.unibuc.ro/staff/cbendic/shb/SHB_interactions.html. More details about how to use the program and some samples, are delivered together with the source code and the make file.

The source code can be modified and any other desired method that permits the overlap population calculation, can be added and used in the place of the EH method.

The choice of the overlap population as a quantitative quantum chemical criterion, able to measure the strength of atom-atom intermolecular interactions, is justified qualitatively: the more positive is the electronic population of atomic overlap distribution $\chi^*_A\chi_B$ (A and B are two neighbouring nuclei), the greater the overlap distribution contributes to atom-atom interaction [7]. The chemical bond is a classical example.

We have chosen EH as quantum chemistry calculation method because this method get an efficient run time and provides, due to our algorithm, the possibility to obtain the electronic structure properties of huge molecules and to perform such calculations for a large set of biopolymer structures in a relative short time.

The usage of EH method is justified also because it is the only semiempirical method that does not use the ZDO approximation and allows the direct calculation of Mulliken overlap population. The simplicity of this method and the approximations that are made are compensated by the use of overlap population as a relative criterion in the interpretation of the results.

SHB_interactions requires structural information in the PDB format [8]. As for the structures determined by diffraction methods, the positions of H atoms are usually not given, the program in the present form, can use only NMR structures that contain also the coordinates of hydrogen atoms. An alternative way, which allows X-ray structures to be used, is to add hydrogen atoms with a protein-nucleic acid manipulation program.

SHB_interactions program uses as input two files: the first one is the PDB file pdbxxxx (xxxx is the PDB code), and the second one is a file named xxxx, which represents a possible interaction table. This table lists all the pairs of the residue numbers that correspond to the different residues (nucleotide-nucleotide, ligand-nucleotide, molecule-molecule etc.) which possess atoms placed at a distance less than 3.9 Å (value adopted by HBexplore for the maximum donor-acceptor distance) and that can interact with each other.

The output file of HBexplore [2] named xxxx_hbx_anal may be also used as an xxxx file, as it contains the tables of H-bonds sorted according to the residue types involved.

Using this table of residue numbers, SHB_interactions program cuts off from the PDB structure the corresponding residue pairs, adds hydrogen atoms to satisfy the oxygen and phosphorus valences, and performs EH calculation of the overlap population. Finally, the program lists all the overlap population values greater than 0.0005.

Results and Discussion

The analysis of the different types of atom-atom interactions performed with SHB_interactions program [6] on a large number of chemical compounds (hydrocarbons, alcohols, amino acids, pyrimidinic and purinic bases, proteins and nucleic acids) has shown that covalent bonds (bond length up to 1.6 Å) have overlap populations in the range 0.3 – 2.8 [9].

In order to compare the results obtained using overlap population as a quantum chemical criterion (SHB_interactions) for the identification of possible intermolecular interactions, and the results obtained using geometrical criteria (HBexplore), two NMR structures were used: a 34-nucleotide RNA pseudoknot that causes efficient frame shifting in mouse mammary tumour virus (PDB code: 1rnk) [10] and a double-stranded oligonucleotide, (5'-d(*Gp*Ap*Ap*Gp*Cp*Tp*Tp*C)-3'), complexed with actinomycin D (PDB code: 1dsc) [11]. Actinomycin D is an antibiotic used clinically for the treatment of highly malignant tumours that contains a 2-aminophenoxazin-3-one chromophore and two identical cyclic pentapeptide lactones.

The H-bonds and the corresponding overlap populations, obtained with SHB_interactions program, are presented in the Table 1 and 2. In these tables the overlap population values between hydrogen atoms and different acceptors are listed in the increasing order. The last five columns of the tables contain the geometrical parameters used by HBexplore for the selection of these hydrogen bonds. It can be observed that all base-base H-bonds selected using HBexplore geometrical criteria were found by SHB_interactions, with overlap population values in the range 0.005-0.15.

The results obtained outline the capability of the overlap population to substitute all the five geometrical parameters used by HBexplore. All H-bonds selected by HBexplore were found by SHB_interactions except 1H6:A33 – O4:U8 H-bond; in this case A33 and U8 are practically located in two parallel planes, and although geometrical criteria were fulfilled, this bond was rejected by SHB_interactions because a negative value of the overlap population was obtained for it (Table 1) and therefore, the interaction between 1H6 and O4 is not possible.

Besides, the use of the overlap population, as a quantum selection criterion, presents the advantage that offers the possibility of the evaluation of the relative strength of different H-bonds, and also of the detection not only of the weaker H-bonds, as those presented at the end of Table 1 and 2, but also of any other atom-atom intermolecular interactions.

Table 1. Base-base hydrogen bonds of 1rnk found by SHB_interactions program

H atom			Acceptor			O.P	Type	Geometrical parameters (Hbexplore)*				
								rHA	rDA	aDHA	aHAA1 (aHAAm)	aDAA1 (aDAAm)
2H2	G	4	N1	A	27	0.1476	b-b	1.59	2.57	163.7	172.2	167.0
1H6	A	27	N3	G	17	0.0921	b-b	1.74	2.74	171.2	173.9	177.1
H1	G	10	N3	C	31	0.0631	b-b	1.76	2.7	153.9	161.9	170.6
H1	G	2	N3	C	18	0.0619	b-b	1.78	2.77	165.0	169.5	174.7
H1	G	9	N3	C	32	0.0609	b-b	1.78	2.75	160.0	163.4	170.5
H1	G	17	N3	C	3	0.0599	b-b	1.78	2.75	160.9	164.5	171.3
H1	G	11	N3	C	30	0.0589	b-b	1.78	2.73	156.3	165.1	172.6
H1	G	15	N3	C	5	0.0578	b-b	1.81	2.8	164.1	167.8	173.3
H1	G	1	N3	C	19	0.0566	b-b	1.82	2.73	148.6	167.9	176.8
H1	G	4	N3	C	16	0.0508	b-b	1.88	2.85	159.9	171.0	173.5
H1	G	29	N3	C	12	0.0482	b-b	1.87	2.86	163.1	169.0	173.8
H3	U	8	N1	A	33	0.0355	b-b	2.05	2.85	134.0	179.7	178.1
1H4	C	30	O6	G	11	0.0378	b-b	1.67	2.64	159.3	134.9	127.4
1H4	C	3	O6	G	17	0.0357	b-b	1.62	2.59	157.8	136.3	134.4
1H4	C	18	O6	G	2	0.0350	b-b	1.64	2.64	170.9	130.7	128.5
H3	U	8	O2	U	34	0.0329	b-b	1.92	2.66	127.5	129.6	144.9
1H2	G	11	O2	C	30	0.0288	b-b	1.87	2.86	166.9	120.0	117.9
1H2	G	4	O2	C	16	0.0288	b-b	1.81	2.76	156.3	123.0	125.1
1H2	G	10	O2	C	31	0.0277	b-b	1.9	2.87	160.4	120.7	119.8
1H2	G	9	O2	C	32	0.0271	b-b	1.87	2.84	161.6	127.1	125.7
1H2	G	2	O2	C	18	0.0259	b-b	1.88	2.88	167.9	122.8	122.0
1H2	G	15	O2	C	5	0.0257	b-b	1.87	2.87	170.3	128.3	125.2
1H2	G	17	O2	C	3	0.0251	b-b	1.88	2.86	162.5	129.7	127.3
1H2	G	29	O2	C	12	0.0241	b-b	1.87	2.88	173.3	121.2	119.7
1H4	C	5	O6	G	15	0.0229	b-b	1.77	2.61	138.2	132.1	133.6
1H4	C	32	O6	G	9	0.0217	b-b	1.8	2.73	152.3	127.5	128.3
1H4	C	16	O6	G	4	0.0176	b-b	1.84	2.81	159.0	130.3	128.5
1H4	C	19	O6	G	1	0.0173	b-b	1.78	2.61	136.6	108.9	118.9
1H4	C	12	O6	G	29	0.0170	b-b	1.83	2.77	152.9	134.5	128.5
1H4	C	31	O6	G	10	0.0167	b-b	1.83	2.78	155.5	122.1	122.2
H3	U	13	O6	G	28	0.0162	b-b	1.97	2.91	153.7	123.4	118.9
1H2	G	1	O2	C	19	0.0114	b-b	2.1	3.09	167.4	122.0	118.6
1H6	A	33	O4	U	8	-0.0099	b-b	1.75	2.62	141.5	113.9	120.8
2H2	G	7	N1	A	6	0.0025	b-b	2.62	3.11	109.9	161.3	170.9
1H2	G	28	O2	U	13	0.0076	b-b	2.59	3.34	131.3	158.0	153.9

*D, donor; A, acceptor; H, hydrogen; A1, Am bonded neighbours of A; rHA, distance H…A; rDA, distance D…A; aHAA1(Am), angle HAA1(Am); aDAA1(Am), angle DAA1(Am) [2].

Table 2. Base-base hydrogen bonds of 1dsc found by SHB_interactions program

H atom	Acceptor	O.P	Type	Geometrical parameters (Hbexplore)*				
				rHA	rDA	aDHA	aHAA1 (aHAAm)	aDAA1 (aDAAm)
H3 T B 14	N1 A A 3	0.0508	b-b	1.82	2.80	158.7	178.9	173.6
H3 T A 6	N1 A B 11	0.0430	b-b	1.81	2.81	170.6	161.7	162.7
H3 T B 15	N1 A A 2	0.0365	b-b	1.86	2.85	164.5	160.8	164.7
H3 T A 7	N1 A B 10	0.0355	b-b	1.88	2.87	165.4	179.8	165.1
H1 G A 1	N3 C B 16	0.0290	b-b	2.03	3.02	165.5	164.3	168.7
H1 G B 9	N3 C A 8	0.0267	b-b	2.05	3.05	164.6	163.5	168.1
H1 G A 4	N3 C B 13	0.0258	b-b	2.01	3.00	163.2	157.8	163.1
H1 G B 12	N3 C A 5	0.0175	b-b	2.16	3.17	171.9	163.4	165.8
2H2 G B 12	O2 C A 5	0.0483	b-b	1.75	2.76	164.7	138.0	143.2
2H2 G A 4	O2 C B 13	0.0439	b-b	1.76	2.75	159.0	137.0	142.0
2H2 G A 1	O2 C B 16	0.0332	b-b	1.82	2.83	167.3	157.8	159.3
2H2 G B 9	O2 C A 8	0.0310	b-b	1.84	2.85	167.2	158.7	159.9
2H6 A A 2	O4 T B 15	0.0153	b-b	2.11	3.08	158.5	165.6	159.5
2H6 A B 10	O4 T A 7	0.0141	b-b	2.13	3.11	159.3	164.9	159.1
2H4 C B 16	O6 G A 1	0.0135	b-b	2.02	3.02	157.0	163.8	160.3
2H4 C A 8	O6 G B 9	0.0129	b-b	2.05	3.04	156.7	164.6	161.0
2H4 C B 13	O6 G A 4	0.0077	b-b	2.21	3.18	150.5	167.0	158.3
2H6 A B 11	O4 T A 6	0.0059	b-b	2.34	3.31	157.5	160.6	154.4
2H6 A A 3	N3 T B 14	0.0024	b-b	3.11				
2H6 A B 11	N3 T A 6	0.0012	b-b	3.48				
2H4 C A 5	O6 G B 12	0.0017	b-b	2.56				

*D, donor; A, acceptor; H, hydrogen; A1, Am bonded neighbours of A; rHA, distance H…A; rDA, distance D…A; aHAA1(Am), angle HAA1(Am); aDAA1(Am), angle DAA1(Am) [2].

Another interesting result is outlined in Figure 1, that presents the overlap populations corresponding to the base-base H-bonds as a function of the hydrogen-acceptor distance (r_{HA}), both for 1rnk and 1dsc structures (Figure 1a,b). It may be observed, that there is a clear delimitation between H-bond overlap populations when the acceptor is oxygen atom and those when the acceptor is nitrogen atom. Besides, D – H…N bonds have overlap populations greater than D – H…O bonds (D is donor atom), for the same distance between hydrogen and acceptor atom.

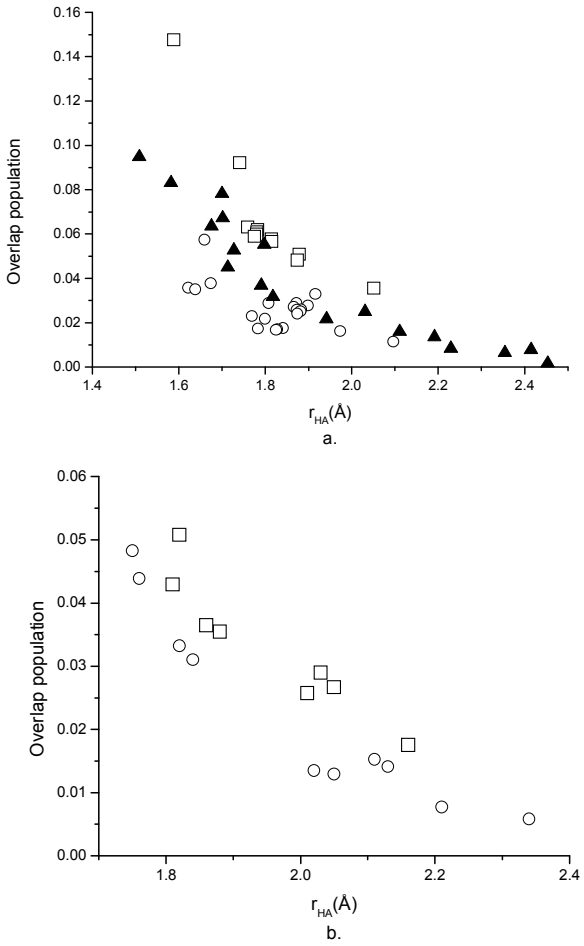

Fig.1: *Overlap population for H-bonds* vs. *H…A distance: a. 1rnk structure; b. 1dsc structure;*
□ - *D – H…N base-base H- bonds;* o - *D – H…O base-base H- bonds;* ▲ - *backbone- backbone H-bonds;*

The other types of H-bonds (backbone-base and backbone-backbone) are much less frequent in the DNA and RNA structures. However, these H-bonds are very important for RNA, especially for the unusual structures like the pseudoknot.

Although present in the case of 1rnk structure (Figure 1a), the backbone-backbone H-bonds are practically inexistent in 1dsc structure. The examination of the data in Figure 1a also outlines that the overlap populations for the backbone-backbone H-bonds are intermediate between D – H…N and D – H…O base-base H-bonds. This is an evidence for the

capability of the overlap population to make distinction between different H-bond types, and allows comparative analysis of the results.

Actinomycin D, that is present in the 1dsc structure, contains two identical cyclic pentapeptides lactones, and therefore 1dsc is a good example to illustrate that SHB_interactions program can be applied not only to nucleic acids but also to proteins.

It is interesting to note that the interactions implying C-H group as a potential hydrogen donor with different acceptor atoms correspond to overlap populations up to 0.03 (Table 3), i.e. in a range where classical hydrogen bonds are also observed.

Table 3. C – H …O hydrogen bonds found by SHB_interactions program in actinomycin D

H atom		Acceptor		Type	rHA	O.P
HA	THR C 23	O*	PXZ C 17	C-H...O	2.2956	0.0321
HA	THR C 18	O	PXZ C 17	C-H...O	2.3061	0.0288
2HG	PRO C 20	O*	PXZ C 17	C-H...O	1.9739	0.0234
HA	DVA C 19	O	THR C 18	C-H...O	2.3542	0.0181
HB	THR C 18	O	MVA C 22	C-H...O	2.7882	0.0174
1HN	SAR C 26	O	PRO C 25	C-H...O	2.3598	0.0157
HB	mVA C 27	O	SAR C 26	C-H...O	2.0789	0.0152
1HN	SAR C 21	O	PRO C 20	C-H...O	2.8015	0.0113
2HD	PRO C 20	O	DVA C 19	C-H...O	2.4265	0.0103

The results obtained with SHB_interactions program point out that much controversial C-H···O hydrogen bonds [12,13] are present and have an important contribution to the atom-atom interaction between the two cyclic pentapeptides.

Conclusions

In conclusion we may infer that in spite of the limitation of the EH method, the results obtained with SHB_interactions program allow a rationalization of the H-bonds in the DNA and RNA, as well as protein structures. Our results suggest that the overlap population criterion can be a useful tool not only for the H-bonds identification, but also for their classification according to the strength and allow an estimate of their contribution to the stability of different structures.

The results obtained outline the capability of the overlap population criterion to substitute all the five geometrical parameters used by HBexplore. Besides, the use of the overlap population as a quantum selection criterion presents the advantage to detect not only the weaker H-bonds, but also any other atom-atom intermolecular interactions.

REFERENCES

1. Jeffrey, G.A. and Saenger, W. (1991) **Hydrogen Bonding in Biological Structure**, Springer-Verlag, New York, 569.
2. Lindauer, K., Bendic, C. and Sühnel, J. (1996) *CABIOS*, **12**, 281-289;
3. McDonald, I.K. and Thornton, J.M. (1994) *J. Mol. Biol*., **238**, 777-793.
4. Mulliken, R.S. (1955) *J.Chem. Phys*. , **23**,1833-1840.
5. Hoffmann, R. (1963) *J.Chem.Phys*, **39**,1397-1412; (1964) *J.Chem.Phys*, **40**,2745.
6. Bendic, C. (2004) http://gw-chimie.math.unibuc.ro/staff/cbendic /shb/shb _interactions.html.
7. Dence, J.B. and Diestler, D.J. (1987) **Intermediate Physical Chemistry**, Jhon Wiley & Sons, New York, 95-98.
8. Berstein, F.C., Koetzle, T.F., Williams, G., Mayer, E.F., Bryce, M.D., Rodgers, J.R, Kennard, O., Simanouchi,T. and Tasumi, M. (1977) *J. Mol. Biol.,* 112,535-542.
9. Bendic C., (2005) *IEJMD,* to be published.
10. Shen, L.X. and Tinoco, I.J.,Jr. (1995) *J. Mol. Biol*.,**247**,963-978
11. Lian, C., Robinson, H. and Wang, A. H. J. (1996) *J.Am.Chem.Soc*., **118**, 8791.
12. Starikov, E.B and Steiner, T. (1997) *Acta Cryst*., **1997**, D53, 345-347
13. Brandl, M., Lindauer, K., Meyer, M. and Sühnel, J. (1999) *Theor. Chem. Acc.* ,**101**, 103-113