

# PCCP

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

*Accepted Manuscripts* are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

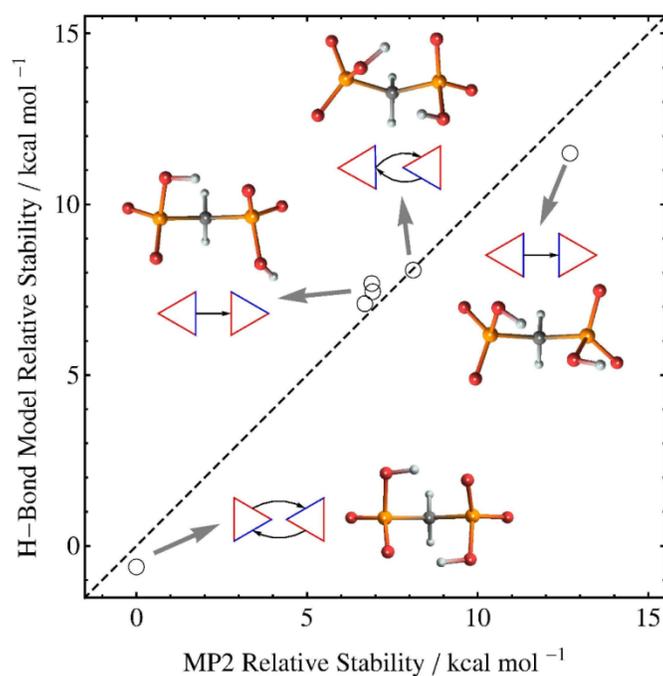
## Hydrogen Bonding Motifs in a Hydroxy-Bisphosphonate Moiety: Revisiting the Problem of Hydrogen Bond Identification

Mitra Ashouri<sup>a</sup>, Ali Maghari<sup>a</sup>, M. H. Karimi-Jafari<sup>\*b</sup>

Received (in XXX, XXX) Xth XXXXXXXXXX 20XX, Accepted Xth XXXXXXXXXX 20XX

DOI: 10.1039/b000000x

Bisphosphonates are important therapeutic agents in bone related diseases and exhibit complex H-bonding networks. To assess the role of H-bonds in their stabilities, full conformational search was performed for Methylenebisphosphonate (MBP) and 1-Hydroxyethylidene-1,1-diphosphonate (HEDP) by MP2 method in conjunction with continuum solvation model. The most stable structures and their equilibrium populations were analyzed at two protonation states via assignment of H-bonding motifs to each conformer. Geometrical and topological approaches for identification and characterization of H-bonds were compared with each other and some of the important correlations between H-bond features were described over the whole conformational space of a hydroxy-bisphosphonate moiety. Topologically derived H-bond energy obtained from local density of potential energy at bond critical points show consistent correlations with other measures such as H-bond frequency shift. An inverse power form without an intercept, predicts topological H-bond energies from hydrogen-acceptor distances with an RMS error of less than 1 kcal/mol. Consistency of this measure was further checked by building a model that reasonably reproduces relative stability of different conformers from their hydrogen-acceptor distances. In all systems, predictions of this model were improved by consideration of weak H-bonds that have no bond critical point.



Graphical Abstract

## 1. Introduction

Bisphosphonates (BP) are an important class of organophosphorous compounds with a wide range of applications from industrial to biomedical areas. Beside their use as anticorrosive or complexing agents, in recent decades they obtained an important contribution in the global pharmaceutical market for their role in treatment of a variety of bone related disease such as osteoporosis, Paget's disease and hypercalcemia due to malignancy<sup>1,2</sup>. The main biological targets of BPs are the bone mineral and the farnesyl pyrophosphate synthase (FPPS) a key enzyme of the mevalonate pathway<sup>3-6</sup>. The structure of BPs is characterized by a P-C-P backbone and two side chains on C atom. Considerable efforts have been devoted over many years to modulate the activity of BPs by changing the structure of side chains<sup>7,8</sup>. Methylenebisphosphonate (**MBP**) is the simplest form that two side chains are hydrogen atoms. Replacing them with an OH and a methyl group leads to 1-Hydroxyethylidene-1, 1-diphosphonic acid (**HEDP**) with considerable change in biomedical activity of the compound. In the most potent bisphosphonates in clinical use, one of the side chains is a hydroxyl group and the other is a bulky group with a nitrogen moiety either in an alkyl chain or within a heterocyclic structure. As a polyprotic acid, all BPs have different protonation states at different pH values and there are numerous sites in their structure that can act as donor and/or acceptor in complex intra- or intermolecular hydrogen bonding networks<sup>9</sup>. Accordingly, the hydrogen bonds play an important role in conformational flexibility of BPs in aqueous media and in interaction with biological targets<sup>10</sup>.

A few computational studies have been published on BPs<sup>11-16</sup> which some of them have been focused on their interaction with cations or solvent molecules<sup>11,12</sup> while the others try to provide some insight on the interaction of BPs with biological targets<sup>13,14</sup>. Another notable theoretical study is a DFT modelling of conformers of **HEDP** which complements experimental analysis of pH-dependent conformers of this compound<sup>15</sup>. In that work, computed Raman spectra for located conformers of **HEDP** have been used to aim in assignment of vibrational bands in experimental spectra. In a recent study<sup>16</sup>, the conformational space of pamidronate has been analyzed with DFT and *ab initio* methods and the most stable conformers of this compound have been characterized for their relative stabilities, equilibrium populations and intramolecular hydrogen bonds. It has been shown that geometrical and topological criteria for identification of H-bonds provide the same description in most cases while fail to agree with each other in some notable situations. The hydrogen bonds in a hydroxy-bisphosphonate subunit can be classified in two groups. The first group of H-bonds are those formed between phosphonate moieties and the second cases are formed between hydroxyl and one of the two phosphonate groups. In this work we aim to provide a more general analysis of H-bonding patterns and its implications for conformational diversity of a hydroxy-bisphosphonate structural module. Indeed a detailed conformational analysis of BPs is a prerequisite for study of their interactions with biological targets. Different approaches are in common use for estimation of

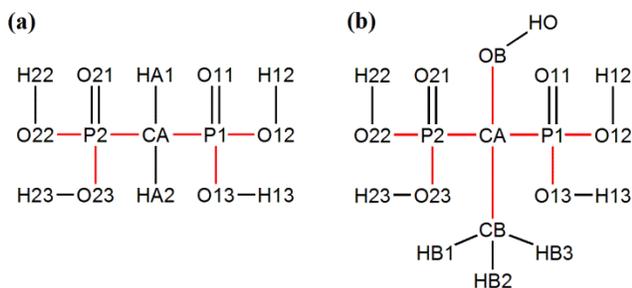
existence and strength of H-bonds<sup>17-19</sup>. In present study we used simple geometrical criteria and more sophisticated criteria based on topology of electron density obtained from quantum theory of atoms in molecules (QTAIM). By testing these criteria against each other we tried to determine the extent of agreement between QTAIM criteria of H-bond formation and simple geometrical criteria. We also paid some attention to a more general question: To what extent the qualitative and/or quantitative analysis of QTAIM results is consistent with the relative stability of different conformers? Moreover, most of the applications of QTAIM were focused on the most stable structures of considered compounds and there is less comparative information on the trend of results when applied on numerous structures of a multi conformer system. In this regard and in conjunction with the problem of H-bond characterization, it is interesting to check that whether the empirical correlations obtained over the most stable structures of different compounds are still valid for different conformations of a single molecule or not.

## 2. Methodology

A hydroxy-bisphosphonate group (Fig. 1) is a polyprotic moiety and in its fully protonated form there are 5 O-H bonds which four of them (P-O-H ones) are ionizable. Successive deprotonation of this tetraprotic acid results in different protonation states that will be denoted here as **H<sub>4</sub>L**, **H<sub>3</sub>L**, **H<sub>2</sub>L**, **HL** and **L** dropping the total charge in notations. In the case of **HEDP**, measured values of 1.21, 2.55, 6.87 and 10.66 are reported for pK<sub>a1</sub> to pK<sub>a4</sub>, respectively<sup>20</sup>. In this work we focused on **H<sub>2</sub>L** and **HL** protonation states since the analysis of species distribution via NMR controlled titrations in Ref. 20 shows that the solution of **HEDP** in pH=7 is a mixture of these states with nearly equal mol fractions.

### 2.1. Conformational search

The conformational space of **MBP** and **HEDP** was explored by systematic variation of all rotatable bonds (red bonds in Fig.1). All torsions containing hydrogen atoms were incremented in steps of 90° while for others a 60° step size was used. All symmetrically unique structures were optimized at HF/3-21G level of theory in conjunction with conductor polarized continuum model (C-PCM)<sup>21,22</sup> for inclusion of solvation effects. All unique conformers at this level were then submitted for further optimization at MP2/6-31++G(d,p) level of theory with inclusion of solvation effects via C-PCM. In the case of multi-H-bond systems where weaker intramolecular interactions play important role in conformational diversity of the system, the MP2 method is superior over DFT one though more expensive. In the same manner, the HF/3-21G is preferred in this work over semiempirical methods for prescreening of the conformational space. This was resulted in 4, 8, 22 and 37 symmetrically unique conformers for **MBP-HL**, **MBP-H<sub>2</sub>L**, **HEDP-HL** and **HEDP-H<sub>2</sub>L**, respectively. At the same level of theory, normal mode analysis was performed on final conformations to ensure their minimum energy character and also to obtain molecular partition functions and vibrational frequencies. It should be noted that the MP2 level of theory is not the ideal method for complete coverage of dispersion contributions that seems to be dominant in



**Fig.1** Structures and atom naming of fully protonated forms of **MBP** (a) and **HEDP** (b). Red bonds are explored in conformational search.

the studied systems. However, to obtain more accurate relative energies, additional single point energy calculations were performed at the MP2/6-311++G(2df,2p) level of theory.

Up to this point, all calculations were performed via the C-PCM continuum solvation model including the electrostatic contribution to the solvation free energy. Additional C-PCM calculations were performed for cavitation, dispersion and repulsion contributions of solvation free energy. The cavitation energy was calculated by the method of Pierotti and Claverie<sup>23</sup> and the empirical method of Floris and Tomasi<sup>24</sup> was used for dispersion and repulsion contributions. In all C-PCM calculations the FIXPVA tessellation method (fixed points with variable area) was used<sup>25</sup>. To obtain more accurate gradient and smoother potential energy surface, the initial number of tesserae for each atomic sphere was increased to 240. Accordingly, the calculated total free energy of each conformer can be expressed as

$$G_{tot} = E_{tot} + G_{therm}^{MP2/BS1} \quad (1)$$

$$E_{tot} = E_{gas}^{MP2/BS2} + E_{ZPE}^{PCM-MP2/BS1} + \Delta G_{solv} \quad (2)$$

$$\Delta G_{solv} = \Delta G_{solv,els}^{PCM-MP2/BS2} + \Delta G_{solv,disp}^{PCM} + \Delta G_{solv,rep}^{PCM} + \Delta G_{solv,cav}^{PCM} \quad (3)$$

where  $G_{therm}$  is the thermal correction on free energy at 298.15 K,  $E_{gas}$  and  $E_{ZPE}$  are the gas phase electronic energy and its zero point vibrational energy correction and  $\Delta G_{solv}$  is the total free energy of solvation obtained from C-PCM model including electrostatic ( $\Delta G_{solv,els}$ ), cavitation ( $\Delta G_{solv,cav}$ ), dispersion ( $\Delta G_{solv,disp}$ ) and repulsion ( $\Delta G_{solv,rep}$ ) contributions. The BS1 and BS2 are the 6-31++G(d,p) and 6-311++G(2df,2p) basis sets, respectively. Calculated relative conformational energies might be the subject of basis set superposition error (BSSE)<sup>26,27</sup>. It has been shown that the existence of BSSE can dramatically affect the relative energy of different conformers<sup>27</sup>. However, the BSSE will be diminished at the complete basis set (CBS) limit. In this regard, additional calculations were performed on some conformers of MBP to estimate the basis set related errors in calculated MP2 gas-phase relative energies. The hierarchy of correlation consistent basis sets<sup>28,29</sup> denoted as cc-pVXZ and aug-cc-pVXZ (with X=D,T,Q) were used to estimate the CBS limit. Two different CBS estimations were provided based on augmented series. The first one denoted as CBS1 was obtained via an exponential extrapolation scheme<sup>30</sup> using aug-cc-pVDZ,

aug-cc-pVTZ and aug-cc-pVQZ values of HF energy in conjunction with an  $X^{-3}$  extrapolation scheme<sup>31</sup> using aug-cc-pVTZ and aug-cc-pVQZ values of correlation energy. The second estimate denoted as CBS2 was provided by separate  $X^{-\gamma}$  extrapolation schemes<sup>32</sup> using HF (with  $\gamma=3.4$ ) and correlation (with  $\gamma=2.2$ ) energies obtained by aug-cc-pVDZ and aug-cc-pVTZ basis sets. All *ab initio* results were obtained by GAMESS suite of programs<sup>33</sup>.

For each conformer, equilibrium population was calculated according to standard techniques of statistical mechanics<sup>34,35</sup>. The lowest energy ( $E_{tot}$ ) structure in each protonation form was kept as a reference ( $C_{ref}$ ) and according to the following equation

$$C_{ref} \Leftrightarrow C_i \quad K_i = \frac{[C_i]}{[C_{ref}]} \quad (4)$$

each conformer  $C_i$  assumed to be in equilibrium with the reference conformer. The equilibrium constant  $K_i$  that by following equation

$$K_i = \frac{q_i}{q_{ref}} = K_i^{rot} K_i^{vib} K_i^{elec} = \frac{q_i^{rot} q_i^{vib} q_i^{elec}}{q_{ref}^{rot} q_{ref}^{vib} q_{ref}^{elec}} \quad (5)$$

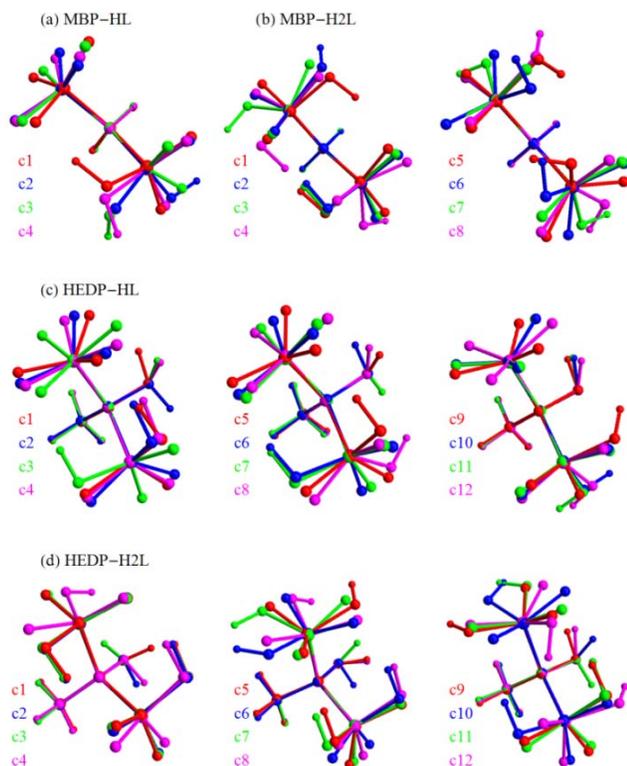
where,  $q_i$  and  $q_{ref}$  are the molecular partition function of the  $i^{\text{th}}$  and reference conformer. In this equation the equilibrium constant is factorized to the rotational ( $K^{rot}$ ), vibrational ( $K^{vib}$ ) and electronic ( $K^{elec}$ ) contributions and  $q^{rot}$ ,  $q^{vib}$  and  $q^{elec}$  are rotational, vibrational and electronic parts of molecular partition function, respectively. The fraction of each conformer in the equilibrated sample was then obtained according to the following relation

$$X_i = \frac{K_i}{\sum_i K_i} \quad (6)$$

where the sum in the denominator goes through all conformers including the reference one (Note that  $K_{ref} = 1$ ).

## 2.2. Hydrogen Bond Characterization

All hydrogen bonds in all located conformers were identified and characterized by geometrical and topological criteria. In the case of geometrical criteria, initially, an H-bond was identified when the hydrogen-acceptor distance ( $R_{HA}$ ) be less than the sum of van der Waals radii of these atoms and simultaneously the donor-hydrogen-acceptor angle ( $\theta_{DHA}$ ) be larger than 100 degrees. These loose thresholds were then tightened by assessment of the level of agreement with other criteria. Tabulated values by Bondi<sup>36</sup> were used as reference van der Waals radii. For identified H-bonds the donor-hydrogen ( $R_{DH}$ ) and donor-acceptor distances ( $R_{DA}$ ) were also considered as characteristic geometrical parameters in subsequent analysis. On the other hand, the QTAIM was utilized to define an independent set of criteria based on the topological characteristics of the electron density<sup>37</sup>. An H-bond was identified topologically if there is a bond critical point (bc) between hydrogen and acceptor atom. If so, some topological characteristics were calculated at this point including the electron density ( $\rho_{bc}$ ), its Laplacian ( $L_{bc}$ ) and the density of potential energy ( $V_{bc}$ ). Since a ring is created upon



**Fig.2** Optimized geometry of low lying conformers of **MBP** and **HEDP**. All structures were aligned from their P-C-P backbone.

formation of intramolecular H-bonds, a ring critical point ( $rc$ ) appears in topology of electron density and thus similar set of parameters ( $\rho_{rc}$ ,  $L_{rc}$  and  $V_{rc}$ ) were defined at this point as other topological characteristics of H-bonds. Finally, the empirically derived relation  $E_{HB} \approx V_{bc} / 2$  was used as a topological measure of strength of different H-bonds. This relation has been derived from correlations investigated over a set of crystal structures<sup>38</sup>. QTAIM calculations were performed by AIM2000 program<sup>39</sup>. In an assessment of different methods for estimation of H-bond energies, a reasonable correlation has been obtained between the IR frequency shift of the donor-hydrogen stretching vibration and H-bond energy<sup>19</sup>. Accordingly, the vibrational frequency shift ( $\Delta\nu$ ) was used as a third independent measure for analysis of geometrical and topological criteria of H-bonds.

This pool of H-bond data provides an opportunity to obtain some insight on following questions: *i*) To what extent the geometrical and topological criteria are in agreement for identification of H-bonds over the whole conformational space of a molecule? *ii*) How topological and geometrical characteristics of H-bonds correlate with themselves and each other? *iii*) Dose QTAIM derived topological features of H-bonds provide predictive information on the relative stability of different conformers? In this regard, linear, reciprocal and logarithmic models were examined between the best correlated pairs of H-bond properties. In each case, the goodness of the fit was measured by coefficient of determination ( $r^2$ ) and the root mean square error (RMSE) of fitted models. To clarify the patterns of hydrogen bonding in a hydroxy-bisphosphonate moiety, simple graph-like representations were designed to introduce H-bond motifs. In

each of these pictorial motifs, a triangle was used for a phosphonate group with different edge colours for protonated and bare oxygen atoms. The hydroxyl group of **HEDP** was represented by a circle and H-bonds were denoted as arrows from donors to acceptors (See Table 2). Each of these graph-like representations will be referred as an H-bond motif (denoted as **M1**, **M2**, **M3** ... for each protonation state of **MBP** and **HEDP**) and some conformers might have the same motif. Thus there is not a one to one relation between conformation numbers and motif numbers though both of them are numbered in an energy increasing order.

### 3. Results and Discussions

Calculated energetic and equilibrium quantities of all conformers of **HL** and **H<sub>2</sub>L** forms of **MBP** and some low lying conformers of **HEDP** can be found in Table 1. The full set of data was reported in ESI (Tables S1 and S2). Corresponding structures of these conformers were depicted in Fig. 2 after aligning them from their P-C-P backbone. In Table S6 values of relative MP2 gas-phase energies of some selected **MBP** conformers were reported for correlation consistent basis sets and two CBS estimations. The aug-cc-pVTZ and aug-cc-pVQZ relative energies are very similar and nearly converged to CBS estimations of relative energies. An analysis of BSSE trends in normal alkanes of various size shows that the results of aug-cc-pVQZ could be considered as free of intramolecular BSSE<sup>27</sup>. The 6-311++G(2df,2p) values used in current study are within less than 0.5 kcal/mol of aug-cc-pVQZ and both CBS estimations. Accordingly, most of the conclusions drawn here from relative energy of conformers are valid for the MP2 level of theory and are not affected considerably by basis set related errors.

In Table 2, graph-like representations of different H-bonding motifs were listed for both **MBP** and **HEDP**. Geometrical and topological features of all identified **MBP** H-bonds can be found in Table 3. The full set of H-bond data for all conformers of **HEDP** can be found in ESI (Tables S3 and S4). It would be informative to note on the distribution of H-bonds over the space of geometrical features. Among all H-bonds that were initially identified by loose geometrical criteria there are many cases that a bond critical point was not found between hydrogen and acceptor atoms.

Distribution of H-bonds over four geometrical parameters are plotted as histograms in Fig. S1. It is evident that in the space of geometrical features there is not a consistent strict decision line to discriminate topologically identified H-bonds from unidentified cases though tightening the values of geometrical thresholds can improve the level of agreement. The geometrical threshold values were tightened in motif assignment step to obtain as consistency as possible between relative stability of different conformers and their assigned motifs, while maintaining all topologically approved H-bonds in the set of geometrically identified H-bonds. Accordingly, there are three sets of H-bonds in this work. The set of H-bonds initially identified by loose geometrical criteria denoted as  $S^{loose}$ , a subset of it obtained by tightening of geometrical criteria denoted as  $S^{tight}$  and the set of topologically identified H-bonds,  $S^{top}$ , that is a subset of  $S^{tight}$ . In other words, we have  $S^{top} \subseteq S^{tight} \subseteq S^{loose}$  where the equality holds only for **MBP-H<sub>2</sub>L** subset of data. All motif definitions in

**Table 1.** H-bond motifs, populations, relative conformational energy terms and equilibrium constants of all conformers of **MBP** and low lying conformers of **HEDP**. All energetic values are in kcal/mol and are relative to the **C1** conformer at each protonation state.

conformer	motif	Pop. (%)	$E_{tot}$	$E_{gas}$	$\Delta G_{solv}$	$E_{ZPE}$	$G_{therm}$	$K$	$K^{elec}$	$K^{vib}$	$K^{rot}$
<b>MBP-HL</b>											
<b>C1</b>	<b>M1</b>	100.0	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00
<b>C2</b>	---	0.0	12.30	24.82	-12.71	0.19	-0.26	0.00	0.00	1.93	1.10
<b>C3</b>	---	0.0	13.16	23.25	-10.36	0.27	-0.36	0.00	0.00	2.65	1.09
<b>C4</b>	---	0.0	13.59	23.54	-10.09	0.14	-0.28	0.00	0.00	1.86	1.08
<b>MBP-H<sub>2</sub>L</b>											
<b>C1</b>	<b>M1</b>	100.0	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00
<b>C2</b>	<b>M2</b>	0.0	6.69	13.32	-6.11	-0.52	-1.19	0.00	0.00	2.86	1.09
<b>C3</b>	<b>M2</b>	0.0	6.89	14.97	-7.52	-0.56	-1.11	0.00	0.00	2.33	1.10
<b>C4</b>	<b>M2</b>	0.0	6.92	15.34	-7.88	-0.53	-1.08	0.00	0.00	2.29	1.09
<b>C5</b>	<b>M3</b>	0.0	8.11	9.28	-0.79	-0.38	-0.87	0.00	0.00	2.17	1.06
<b>C6</b>	<b>M4</b>	0.0	12.70	19.29	-5.86	-0.73	-1.70	0.00	0.00	4.67	1.09
<b>C7</b>	---	0.0	13.87	29.26	-14.15	-1.24	-2.91	0.00	0.00	14.27	1.19
<b>C8</b>	---	0.0	13.90	27.61	-12.48	-1.23	-2.95	0.00	0.00	15.65	1.16
<b>HEDP-HL</b>											
<b>C1</b>	<b>M1</b>	50.8	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00
<b>C2</b>	<b>M2</b>	34.4	0.62	3.94	-2.48	-0.84	-1.23	0.68	0.35	1.94	1.00
<b>C3</b>	<b>M1</b>	14.2	0.87	0.81	0.04	0.02	-0.10	0.28	0.23	1.21	1.00
<b>C4</b>	<b>M3</b>	0.2	3.70	6.50	-2.54	-0.26	-0.66	0.00	0.00	1.97	0.99
<b>C5</b>	<b>M4</b>	0.3	3.94	7.58	-3.11	-0.54	-1.40	0.00	0.00	4.25	1.00
<b>C6</b>	<b>M4</b>	0.0	5.07	9.63	-4.07	-0.49	-1.05	0.00	0.00	2.59	1.00
<b>C7</b>	<b>M4</b>	0.0	5.52	14.17	-8.58	-0.07	-0.25	0.00	0.00	1.37	1.00
<b>HEDP-H<sub>2</sub>L</b>											
<b>C1</b>	<b>M1</b>	56.8	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00
<b>C2</b>	<b>M1</b>	26.4	0.40	2.36	-1.95	-0.01	0.05	0.46	0.51	0.91	1.00
<b>C3</b>	<b>M1</b>	16.7	0.71	3.59	-2.74	-0.14	-0.12	0.29	0.30	0.97	1.00
<b>C4</b>	<b>M2</b>	0.1	4.51	8.02	-3.18	-0.33	-0.66	0.00	0.00	1.64	1.05
<b>C5</b>	<b>M2</b>	0.0	5.66	10.11	-4.26	-0.18	-0.62	0.00	0.00	2.00	1.04
<b>C6</b>	<b>M2</b>	0.0	5.81	11.25	-5.00	-0.44	-0.81	0.00	0.00	1.80	1.05
<b>C7</b>	<b>M3</b>	0.0	5.89	13.25	-7.26	-0.10	-0.41	0.00	0.00	1.62	1.05
<b>C8</b>	<b>M4</b>	0.0	5.97	10.92	-4.79	-0.16	-0.52	0.00	0.00	1.78	1.04

5

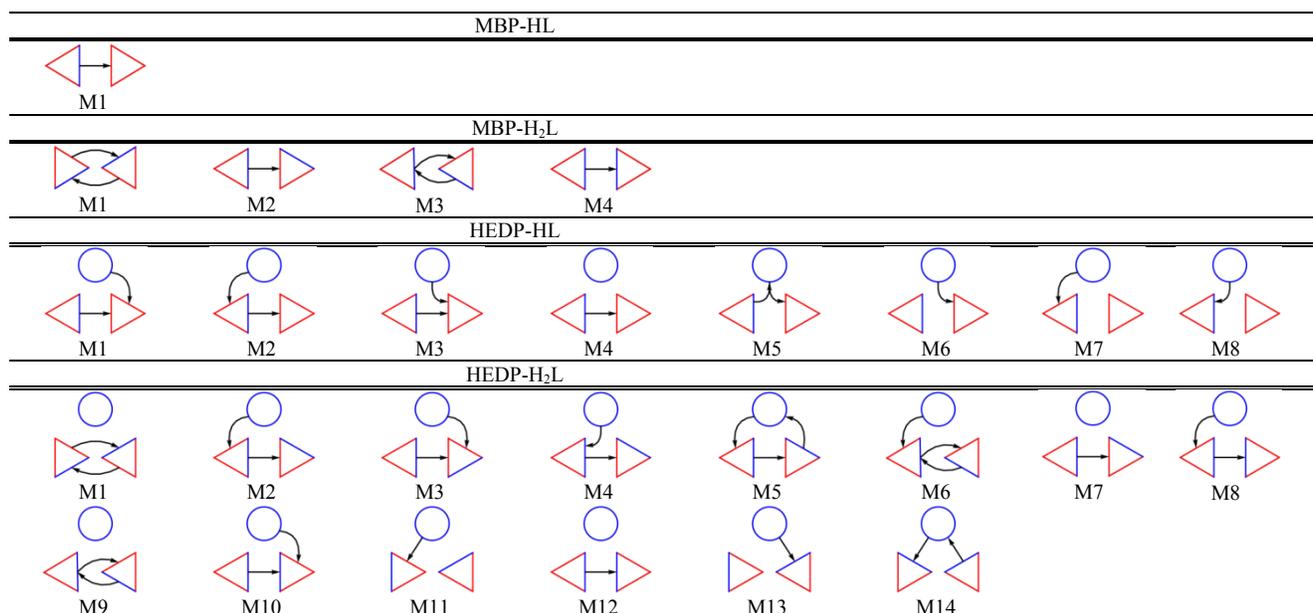
following discussions are based on  $S^{tight}$  set of H-bonds.

### 3.1. Conformers and H-Bonding Motifs

As can be seen in Table 1, the relative stability of different conformers is mainly controlled by electronic part of partition function. The rotational part is nearly the same in all conformers since the formation of H-bonds does not change the compactness of conformers, considerably. On the other hand, the vibrational part acts in opposite direction of electronic factor in most cases but is not dominant. From an entropic point of view, formation of strong H-bonds is not favoured by vibrational degrees of freedom. Accordingly, the conformers are populated by enthalpic (electronic) preference while entropic (vibrational) effects act in opposite direction. The general trend for all conformers in Tables 1, S1 and S2 is that the solvation stabilizes higher conformers (with weaker or no H-bonds) more than lower conformers (with stronger H-bonds). From the charge separation point of view, this

is an expected result since the intermolecular H-bonds act themselves as a stabilizing factor and the solvent stabilization role becomes less important. Both **HL** and **H<sub>2</sub>L** forms of **MBP** have a single populated conformer (**C1**) at 298 K while both forms of **HEDP** are a mixture of three conformers at the same conditions. In their analysis of pH-dependent protonated conformers of **HEDP** by means of FT-Raman spectroscopy and DFT calculations Cukrowski and co-workers<sup>15</sup> found two conformers for each of the **HL** and **H<sub>2</sub>L** forms of **HEDP**. The molecular mechanical prescreening of the conformational space and the shortcoming of DFT methods in complex H-bonding systems might be the reason of their loss of many conformers that are reported in present study. The singly populated **HL** and **H<sub>2</sub>L** conformers reported in Ref. 15 are the same as the most stable conformers **HL-C1** and **H<sub>2</sub>L-C1** obtained in present study for **HEDP** but current analysis shows that each of these conformers has around 50% population. In the case of **MBP**, the motif based

**Table 2** Hydrogen bonding motifs in conformers of **MBP** and **HEDP**. Each phosphonate is represented as a triangle with each edge as one of oxygen atoms (blue: protonated, red: bare). The hydroxyl group is represented by a circle and each hydrogen bond is depicted as an arrow

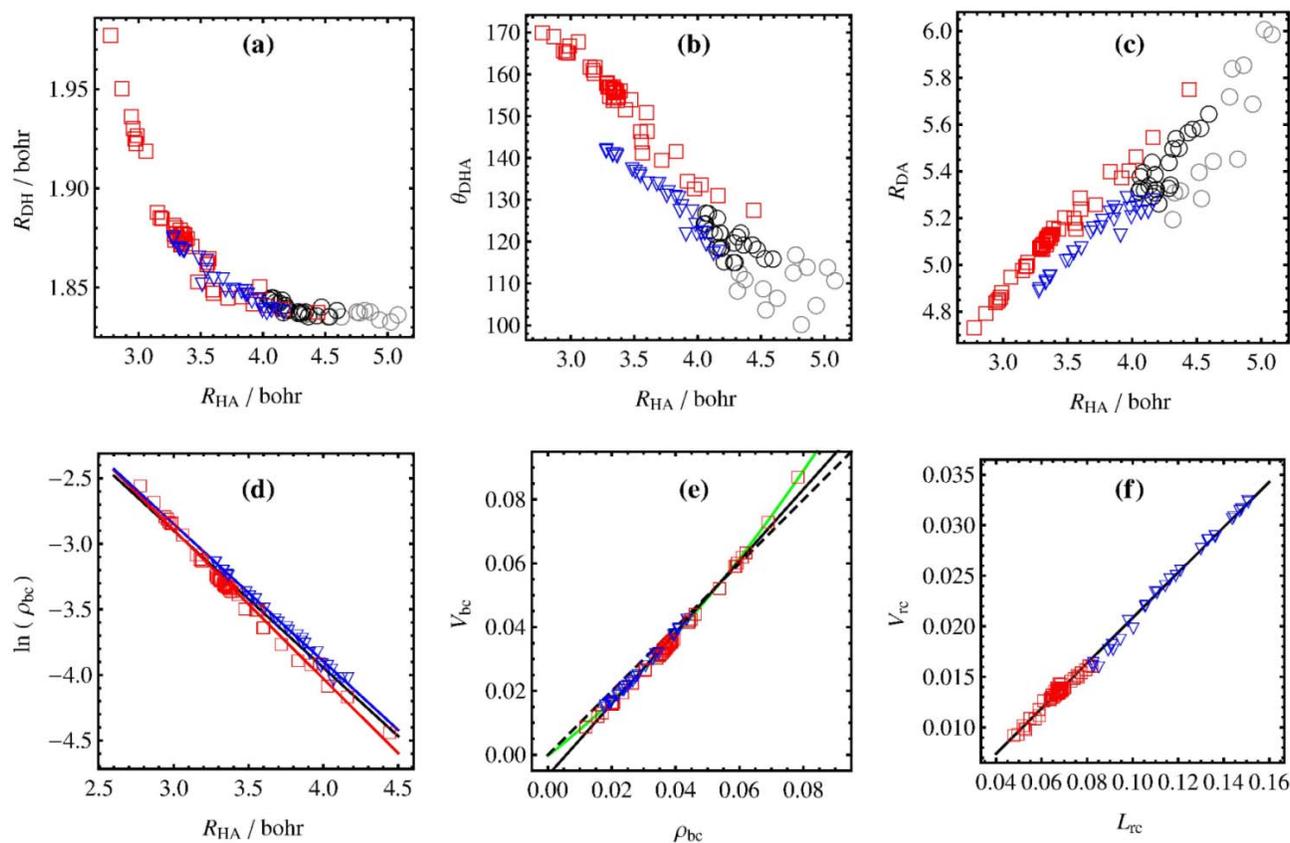


analysis of different conformers provides clear explanation of their relative energies. For **MBP-HL** only one H-bond is possible that results in **M1** motif in **C1** conformer. The other three conformers have no H-bond with minor energy differences due to relative orientations of P-OH group. In the case of **MBP-H<sub>2</sub>L** motifs, relative energies in Table 1 are clustered based on their motifs in Table 2. The **MBP-H<sub>2</sub>L-C1** has two inter-phosphonate reciprocal H-bonds that are locked via an eight-member ring. The next three conformers **C2**, **C3** and **C4** have the same motif **M2** and similar relative energies. In this motif there is one inter-phosphonate H-bond and orientation of the other P-O-H is responsible for small energy differences. The next conformer **C5** has motif **M3** that again has a locked structure as a result of two inter-phosphonate H-bonds but acceptor of the second H-bond is donor of the first one. This pattern of H-bonding locks both P-O-H groups via a smaller six-member ring in comparison with **M1**, but it seems that formation of the second H-bond is not preferred in presence of solvation effects. While the gas-phase energy of **C5-M3** is lower than the average of {**C2**, **C3**, **C4**} set (with **M2** motif) the solvation free energy provides more stabilization for these three conformers than **C5**. The last possible H-bonding motif in **MBP-H<sub>2</sub>L** is **M4** in **C6** conformer with one inter-phosphonate H-bond that the protonated oxygen is the acceptor and so is weaker than the H-bond of **M2** motif. The next two conformers **C7** and **C8** have no H-bond and thus no assigned motif. As can be seen in Table 2, the **MBP** motifs are the basis of **HEDP** motifs. The energy ordering of identified **HEDP** H-bond motifs is the same as their basic counterparts in **MBP** and the inter-phosphonate H-bonds are generally stronger than those formed between hydroxyl and phosphonate groups. The relative stability of low-lying **HEDP** conformers can be consistently explained by the type of H-bonds in their motifs. In **HL** protonation state, it is expected that the **M1** motif in which

the hydroxyl H-bond is donated to the more negative phosphonate be more favoured than the **M2** motif. This is in agreement with the gas-phase energies of **C1**, **C2** and **C3** conformers. But the solvation effect, zero point energy and thermal contributions to the free energy stabilize the **M2** motif of **C2** conformer such that it lies lower than the **C3** one in total energy scale. The next motif, **M3**, in **HL** form of **HEDP** belongs to just one conformer, **C4**, and the reason of its less stability is donation of hydroxyl H-bond to the same acceptor atom that is also donated by the inter-phosphonate H-bond. Similar conclusions can be drawn from relative stability of **HEDP-H<sub>2</sub>L** motifs. The only populated conformers **C1**, **C2** and **C3** share the same motif **M1** which keeps the basic inter-phosphonate reciprocal pattern of H-bonds. The orientation of hydroxyl group is the main factor that controls the energy difference of these three conformers. Presence of the hydroxyl group makes relative energies spreader over the energy scale of conformational space and it is not possible to draw distinct energy borders between different motifs as in the case of **MBP**.

### 3.2. H-Bond Identification and characterization

From a simplistic point of view, all H-bonds in a hydroxy-bisphosphonate moiety can be classified as five types: *i*) POH...OP, *ii*) POH...O(H)P, *iii*) COH...OP, *iv*) COH...O(H)P, *v*) POH...O(H)C. But different conformations, protonation states and H-bonding patterns results in a wide range of values for each characteristic feature especially the H-bond energy. Before dealing with the QTAIM derived H-bond energies, it is informative to see how they could be estimated from the relative energy of conformers. The energy difference between **MBP-HL-C1** conformer and the other three conformers with no H-bond can be used to estimate the strength of an inter-phosphonate POH...OP H-bond in **HL** protonation state which is around 24



**Fig.3** Important correlations between different features of hydrogen bonds. Red squares: inter-phosphonate H-bonds in  $S^{\text{top}}$ ; Blue triangles: hydroxyl-phosphonate H-bonds in  $S^{\text{top}}$ ; Black circles: H-bonds in  $S^{\text{tight}}$  but not in  $S^{\text{top}}$ ; Gray circles: H-bonds in  $S^{\text{loose}}$  but not in  $S^{\text{tight}}$ ; Black, red and blue lines represent a linear fit of whole, inter-phosphonate and hydroxyl-phosphonate data. Green curve: quadratic fit of whole data. All values are in atomic units.

kcal/mol based on gas-phase electronic energies (see Table 1). Each of the similar  $\text{POH}\cdots\text{OP}$  H-bonds in **MBP-H<sub>2</sub>L-C1** contribute around 14 kcal/mol to the stabilization of this conformer with respect to **C7** and **C8** conformers that have no H-bonds. Relative gas-phase energies of **C2**, **C3** and **C4** provide nearly the same estimation for a  $\text{POH}\cdots\text{OP}$  H-bond though it seems to be a little stronger when there is just one H-bond. For **MBP-H<sub>2</sub>L-M3** motif in **C5** (see Table 2) it is not possible to estimate H-bond energies since the H-bonds are asymmetric with respect to the type of donor and acceptor atoms. Qualitatively, it could be said that the  $\text{POH}\cdots\text{OP}$  bond is weaker than that is in **C1** to **C4** but it is stronger than the other  $\text{POH}\cdots\text{O(H)P}$  bond in **M3** motif. Indeed the later with an  $\text{H}\cdots\text{O}$  distance of 4.44 au is much weaker than the similar bond in **MBP-H<sub>2</sub>L-C6** and we can say that both H-bonds in **C5** weaken each other. Finally, the  $\text{POH}\cdots\text{O(H)P}$  in **C6** is estimated to have an energy of around 9 kcal/mol. These estimated numbers are useful for analysis of the consistency of what is provided by QTAIM on these structures. Estimation of H-bond energies from conformational energies is not possible in the case of **HEDP** since no conformer without an H-bond was identified.

Most of disagreements between geometrical and topological approaches for H-bond identification correspond to the relation between hydroxyl group and phosphonate groups, i.e. the existence or absence of a  $\text{COH}\cdots\text{OP}$  or  $\text{COH}\cdots\text{O(H)P}$  H-bond. For

example, an energy range from 4.5 to 6.2 kcal/mol is covered by **HEDP-H<sub>2</sub>L** conformers from **C4** to **C11**. This range is wider if one considers their gas-phase relative energies that are related more directly to the H-bond energies. Any approach for identification of H-bonds and quantification of their strength should ultimately be helpful in analysis of such differences in conformational space of a compound. In  $S^{\text{top}}$  set of H-bonds, just one H-bond (an inter-phosphonate one) was detected for **HEDP-H<sub>2</sub>L** conformers from **C4** to **C11** but in  $S^{\text{tight}}$  set of H-bonds there is an additional H-bond for each of these conformers that results in H-bond motifs **M2**, **M3** and **M4**. As another example, the three **HEDP-HL** conformers **C5**, **C6** and **C7** were assigned to the **M4** motif without any H-bond between hydroxyl group and any of phosphonate groups. Such an H-bond is not present in  $S^{\text{top}}$  or  $S^{\text{tight}}$  set of H-bonds. In these conformers the hydroxyl group is oriented toward the more negative doubly deprotonated phosphonate group and this interaction is responsible for energy difference of these conformers. With a slight variation of adopted geometrical criteria, one can assign an H-bond from hydroxyl to phosphonate in **C5** and **C6** conformers and explain the gas-phase energy differences between these two conformers and **C7** one. These are some of many examples show that tunable geometrical criteria are more consistent with the fuzzy nature of H-bond identification problem than the topological analysis which starts with a true-false step based on

**Table 3** Hydrogen bond properties in all conformers of **MBP**. All values are in atomic units except  $E_{HB}$  that is in kcal/mol.

	D	A	$E_{HB}$	$R_{DH}$	$\theta_{DHA}$	$R_{HA}$	$R_{DA}$	$\rho_{bc}$	$L_{bc}$	$\rho_{rc}$	$L_{rc}$	$V_{rc}$
<b>MBP-HL</b>												
<b>C1</b>	O11	O23	23.0	1.951	169.4	2.867	4.798	0.0690	0.1507	0.0168	0.0820	0.0162
<b>MBP-H<sub>2</sub>L</b>												
<b>C1</b>	O21	O13	10.4	1.878	156.2	3.354	5.128	0.0364	0.1198	0.0149	0.0677	0.0138
<b>C1</b>	O11	O23	10.2	1.877	156.1	3.361	5.135	0.0361	0.1188	0.0148	0.0683	0.0139
<b>C2</b>	O11	O23	11.3	1.878	158.3	3.288	5.081	0.0390	0.1279	0.0146	0.0687	0.0137
<b>C3</b>	O11	O22	10.7	1.875	157.0	3.331	5.109	0.0371	0.1236	0.0145	0.0676	0.0135
<b>C4</b>	O21	O13	10.9	1.876	157.4	3.313	5.097	0.0379	0.1252	0.0146	0.0692	0.0139
<b>C5</b>	O11	O23	8.4	1.863	144.2	3.557	5.185	0.0304	0.1072	0.0153	0.0685	0.0144
<b>C5</b>	O21	O11	2.9	1.838	127.9	4.441	5.756	0.0119	0.0465	0.0105	0.0478	0.0093
<b>C6</b>	O11	O21	6.2	1.845	139.8	3.718	5.264	0.0234	0.0937	0.0120	0.0572	0.0110

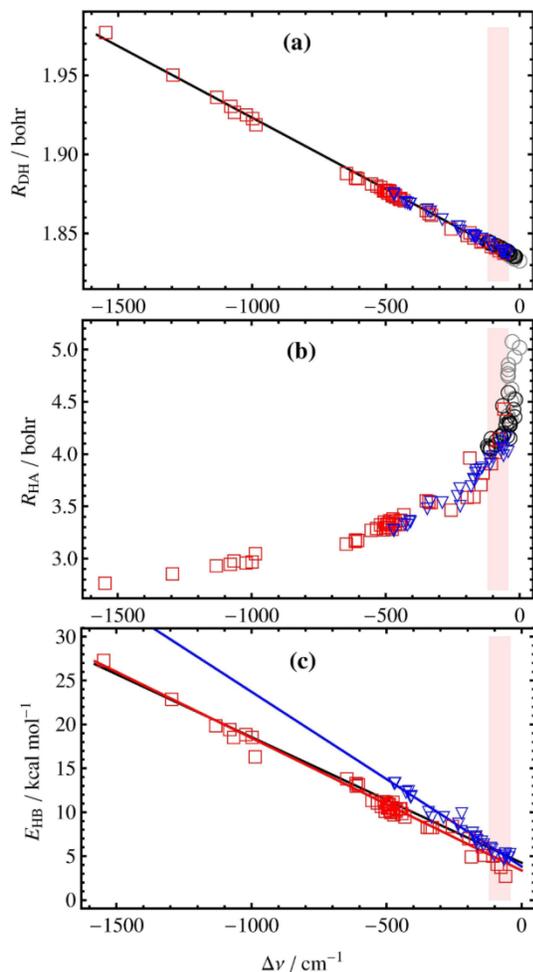
the existence or absence of a bond critical point. In other words, the true-false character of topological criteria disables this approach for explanation of some sort of diversity in conformational space while the geometrical criteria can be tuned for a satisfactory description of the origin of energy differences and fit better to the fuzzy nature of the problem. Albeit, this conclusion might be criticized as an unnecessary application of the term “H-bond” to what is just a “weak attractive H-contact”. But beyond our care about terminology, where is the rigorous border of these concepts? Here we think to H-bond as a concept that is invented to help for an explanation not to rule over it. It will be proposed in following sections that a quantitative measure of H-bond energy based on density of potential energy at “H-bond” critical point when extrapolated to “weak attractive H-contacts” results in a better prediction of relative stability of different conformers.

Though a lot of weak H-bonds (or H-contacts) might be ignored by topological approach, it has the advantage of providing a quantitative scale for strength of those that are not ignored. H-bond energies in Table 3 were calculated via an empirical formula (proposed in Ref. 38) from QTAIM derived values of local density of potential energy at H-bond critical points. The interesting result that is revealed from these data is the consistent correlation among values of  $E_{HB}$  in Table 3 and the relative stabilities of **MBP** conformers in Table 1. For example the **MBP-HL-C1** conformer with just one H-bond is around 24 kcal/mol lower in gas phase energy than **C2**, **C3** and **C4** conformers that have no H-bond, and the value of  $E_{HB}$  for the single H-bond of **C1** conformer is 23 kcal/mol. Other values of  $E_{HB}$  in **MBP-H<sub>2</sub>L** conformers are also in agreement with our previous estimations of H-bond strength in different motifs based on relative gas-phase conformational energies. In the case of **HEDP**, the relation between QTAIM derived H-bond energies and relative stability of conformers becomes somewhat noisy and less evident. There are different reasons for this difference. First of all, as can be seen in Tables 3, S3 and S4, many weak hydroxyl-phosphonate H-bonds in **HEDP** are not identified by topological criteria and thus have no value of  $E_{HB}$ . These weak interactions affect the relative stability of conformers while their role is not retrieved from topological criteria. Another reason for difference of **HEDP** with

**MBP** is the fact that the H-bonding is not the only factor for determination of relative stabilities though still dominant. Orientation of the hydroxyl group when is not involved in an H-bond, interaction of methyl group with other structural components and other steric or electronic effects act as some type of noise in any strict relation that one tries to make between H-bond energies and relative conformational stabilities.

### 3.3. Correlations between H-Bond features

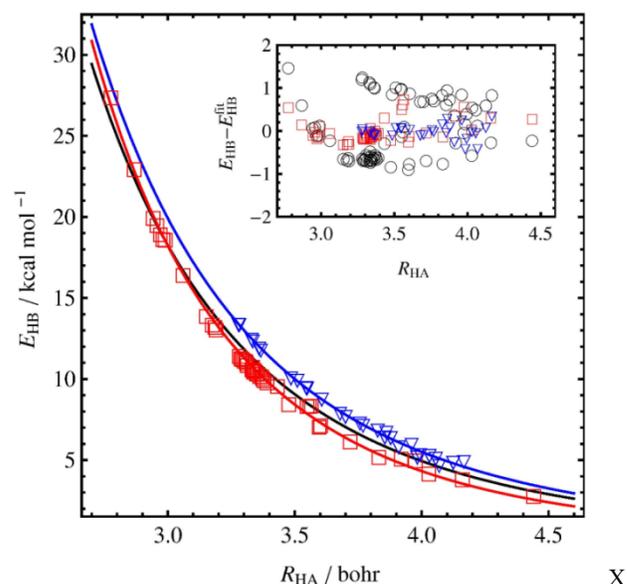
Correlations of geometrical and topological features of H-bonds are of interest for identification and classification of different types of H-bonds<sup>17,18</sup>. They also could ultimately help on derivation of better scoring formula for estimation of H-bond energies<sup>19</sup>. Important correlations between different features of H-bonds are shown in Fig. 3 and S2. Fitted functional forms and parameters, root mean square error (RMSE) and the coefficient of determination ( $r^2$ ) for all lines or curves in these figures were reported in Table S5. From the energetic point of view, the most important correlation that was found in current study is the relation between H-bond energies ( $E_{HB}$ ) derived from local density of potential energy ( $V_{bc}$ ) and the hydrogen-acceptor distance ( $R_{HA}$ ). We postpone the discussion of it to the next section and deal here with other regular trends between H-bond features. Among geometrical features, the hydrogen-acceptor distance shows better correlations with other features. The reciprocal relation between  $R_{HA}$  and  $R_{DH}$  in Figure 3a failed to fit with an inverse or exponential equation. Instead, as can be seen in Figure S2, there is a nearly linear relation between  $R_{DH}$  and  $R_{HA}^{-6}$ . On the other hand, the electron density  $\rho_{bc}$  and its Laplacian  $L_{bc}$  at bond critical point show strong correlations with  $R_{HA}$  over the whole range of H-bond strength. The reciprocal relation between  $\rho_{bc}$  and  $R_{HA}$  has an exponential form that its parameters are less dependent on the type of donor and acceptor atoms and the H-bonding motif. The relations between local density of potential energy and the electron density at bond or ring critical points are similar and most data are near the  $V = \rho$  line. However, a better regression was obtained for a quadratic relation. There are also strong linear relations between some features obtained at ring critical points. The best example is the linear fit of  $V_{rc}$  against  $L_{rc}$  shown in Figure 3f with  $r^2$



**Fig. 4** Correlation of donor-hydrogen bond length, hydrogen-acceptor distance and topological H-bond energy with vibrational frequency shift. Red squares: inter-phosphonate H-bonds in  $S^{\text{top}}$ ; Blue triangles: hydroxyl-phosphonate H-bonds in  $S^{\text{top}}$ ; Black circles: H-bonds in  $S^{\text{tight}}$  but not in  $S^{\text{top}}$ ; Gray circles: H-bonds in  $S^{\text{loose}}$  but not in  $S^{\text{tight}}$ .

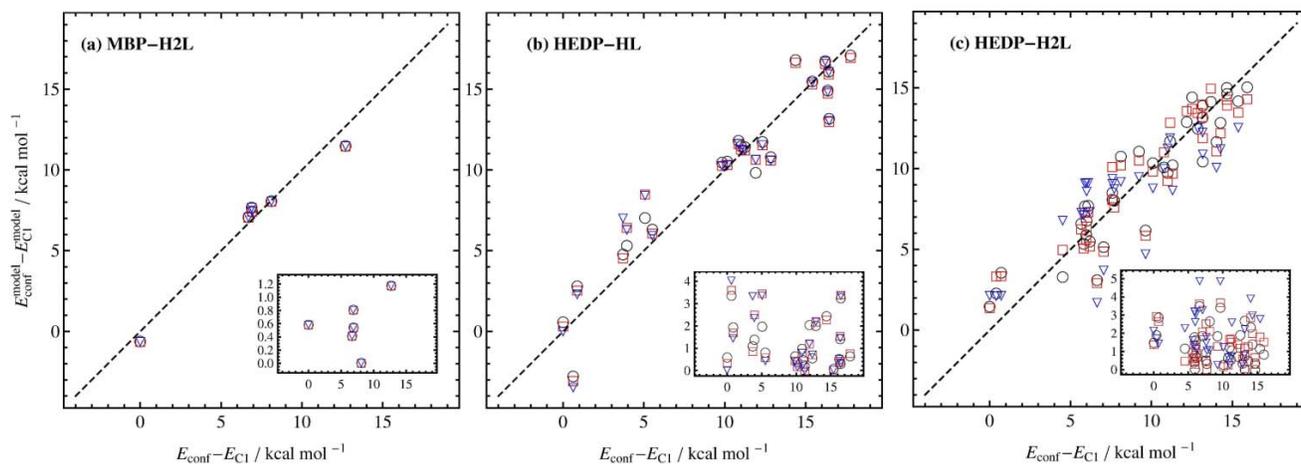
value of 0.9959 over the whole set of data. It should be noted that features calculated at ring critical points are not related to the strength of H-bonds but can discriminate the inter-phosphonate H-bonds from hydroxyl-phosphonate cases.

Another interesting quantity is the IR frequency shift of donor-hydrogen bond  $\Delta\nu$  implied by H-bonding. In Figure 4 important correlations between this feature and  $R_{\text{DH}}$ ,  $R_{\text{HA}}$  and  $E_{\text{HB}}$  are depicted in different panels. To define frequency shift, O11-H11 bond in **HEDP-HL-C21** and OB-HO bond in **HEDP-HL-C5** were chosen as reference for PO-H and CO-H bonds, respectively. These bonds participate in no H-bonds and their frequency is large enough to be a reference for red shifting H-bonds excepted in current study. However, it should be noted that, since the H-bonds under study are of similar types and among similar donor and acceptor atoms, the resulting correlations are also exist on raw values of frequencies without such an offsetting by a reference. As can be seen in Figure 4, the donor-hydrogen distance show a strong linear correlation with frequency shift but for the hydrogen-acceptor distance the



**Fig. 5** Dependence of topologically derived H-bond energies on the hydrogen-acceptor distance. Red squares: inter-phosphonate H-bonds in  $S^{\text{top}}$ ; Blue triangles: hydroxyl-phosphonate H-bonds in  $S^{\text{top}}$ ; Black, red and blue lines represent an inverse power fit of the whole, inter-phosphonate and hydroxyl-phosphonate subsets of data. Inset plot: fit residuals.

relation was not found to obey any of examined simple functional forms. In previous studies, considerable correlations were found between the strength of H-bonds and their frequency shifts<sup>19</sup>. The linear regressions obtained between the topological H-bond energy  $E_{\text{HB}}$  and  $\Delta\nu$  are in agreement with those findings and provide another support for  $E_{\text{HB}}$  values as measures of H-bond strength. However, it could be interesting to ask about the range of values that a non-topological parameter might fall in a lack of consensus with the topological approach for identification of H-bonds. For the frequency shift parameter this range is represented by the narrow pink region in Figure 4. The magnitude of this range is defined as  $|\Delta\nu_{\text{min}}^{\text{top}} - \Delta\nu_{\text{max}}^{\text{no-top}}| \approx 80 \text{ cm}^{-1}$  where  $\Delta\nu_{\text{min}}^{\text{top}} \approx -40 \text{ cm}^{-1}$  is the minimum frequency shift in the set of topologically identified H-bonds and  $\Delta\nu_{\text{max}}^{\text{no-top}} \approx -120 \text{ cm}^{-1}$  is the maximum frequency shift in the set of H-bonds that are not identified by topological criteria, *i.e.* those without a bond critical point. This range seems to be small with respect to the whole range of frequency shifts observed in the whole conformational space but as can be seen in Figure 4 numerous H-bonds lie in this range and its vicinity. This includes all members of  $S^{\text{tight}}$  and  $S^{\text{loose}}$  set of H-bonds that are not available in  $S^{\text{top}}$ . In summary, the topological approach for identification of H-bonds has the advantage of providing a useful quantitative measure of H-bond energy which seems to be consistent with most of non-topological measures such as frequency shift and hydrogen-acceptor distance but it has the critical disadvantage of ignorance of many weak H-bonds (or appearance of H-contacts) as a result of true-false nature of identification via a bond critical point. Although weak in strength, these H-bonds play an important role in fine tuning of the relative stabilities in the conformational space of a molecule and also in the stability of macromolecules such as proteins when one considers their collective effect.



**Fig.6** H-bond based prediction of relative conformational energies ( $E_{\text{conf}}^{\text{model}}$ ) in comparison with their calculated *ab initio* values ( $E_{\text{conf}} \equiv E_{\text{tot}}$ ). Blue triangles:  $S^{\text{top}}$  set of H-bonds; Red squares:  $S^{\text{tight}}$  set of H-bonds; Black circles:  $S^{\text{loose}}$  set of H-bonds. Inset plots: prediction errors in kcal/mol.

### 3.4. An H-bond-based model to score conformational stabilities

The best correlation that was obtained between topological and geometrical features is the inverse dependence of topologically derived H-bond energy ( $E_{\text{HB}}$ ) on the hydrogen-acceptor distance ( $R_{\text{HA}}$ ). In Figure 5 values of  $E_{\text{HB}}$  were plotted against  $R_{\text{HA}}$  for those cases that have a bond critical point *i.e.* the  $S^{\text{top}}$  set of H-bonds. Among different models that were examined, the best fit of data points in this figure was obtained via an inverse power model of following form

$$E_{\text{HB}} = a / R_{\text{HA}}^n \quad (7)$$

where different values of  $a$  and  $n$  parameters were obtained for the whole and different subsets of data (see Table S5). Both inter-phosphonate ( $n = 5.001$ ) and hydroxyl-phosphonate ( $n = 4.466$ ) sets of H-bonds were fitted by the same  $r^2$  value of 0.9996. The quality of the fit is also reflected by low values of RMSE that are 0.26 and 0.17 kcal/mol for inter-phosphonate and hydroxyl-phosphonate sets, respectively. A single fit of the whole data will result in a value of 4.539 for  $n$  and the fit quality is decreased to  $r^2$  and RMSE values of 0.9961 and 0.69 kcal/mol, respectively, which is still quite reasonable. Thus Equation 7 has the ability of reproducing  $E_{\text{HB}}$  within the chemical accuracy over a wide range of values. The same functional form has been used previously for a large set of intermolecular H-bonds and the value of 3.8 was proposed for parameter  $n$  from a single fit of whole data<sup>19</sup>. As a result of diversity of donor and acceptor atoms, the correlation obtained in that work is not as good as what is presented here. Evidently, a single value of  $n$  is unable to cover all types of H-bonds in different molecules. According to all of these results, another challenging test could be designed to check the reliability of topologically derived H-bond energies. Assuming that the H-bonding is the dominant factor in the energy difference of conformers, and assuming that the Equation 7 can be used for estimation of H-bond energy from hydrogen-acceptor distance, the following equation was considered as a predictive model for estimation of the relative stability of different

conformers:

$$E_{\text{conf}}^{\text{model}} = \alpha + \beta \sum_{i \in S} (a_i / R_{\text{HA},i}^n) \quad (8)$$

In this equation  $E_{\text{conf}}^{\text{model}}$  is the model energy of a specific conformer  $\text{conf} \in \{C1, C2, C3, \dots\}$  obtained by numeration of its H-bond contributions and  $\alpha$  and  $\beta$  are some constant values loosely assumed to be the same for all conformers. The sum is gone over all of H-bonds in a conformer that also belong to an  $S$  set of H-bonds where might be any of  $S^{\text{top}}$ ,  $S^{\text{tight}}$  or  $S^{\text{loose}}$ . For the case of  $S = S^{\text{top}}$  the model predictions are based on what is confirmed by topological criterion of existence of an H-bond critical point and thus the conformers will be scored by the sum of density of potential energy at their H-bond critical points. On the other hand, when the  $S$  set is considered as any of the  $S^{\text{tight}}$  or  $S^{\text{loose}}$ , fitted parameters in Equation 7 were obtained again over the  $S^{\text{top}}$  set of data but are used in Equation 8 in an extrapolatory fashion for other cases of H-bonds that have no bond critical point. Best values of parameters  $\alpha$  and  $\beta$  were obtained via a least squares regression of Equation 8 against calculated *ab initio* conformational energies ( $E_{\text{conf}} \equiv E_{\text{tot}}$ ). The model conformational energies obtained from Equation 8 are plotted against *ab initio* conformational energies in Figure 6 for **MBP-H<sub>2</sub>L**, **HEDP-HL** and **HEDP-H<sub>2</sub>L**. All values in this figure are relative to the **C1** conformer of each system and the absolute errors of model conformer energies are shown in embedded plots. In the case of **MBP-H<sub>2</sub>L**, *ab initio* conformational energies are reproduced quite well from sum of H-bond energies with an RMS error of 0.7 kcal/mol. The ability of the model is reduced for **HEDP-HL** and **HEDP-H<sub>2</sub>L** systems. In the case of **HEDP-HL** the RMS error is 1.9, 1.7 and 1.5 kcal/mol for  $S^{\text{top}}$ ,  $S^{\text{tight}}$  and  $S^{\text{loose}}$  sets, respectively. Thus the inclusion of H-bonds without a bond critical point improves the conformational energy predictions. This is even more evident for **HEDP-H<sub>2</sub>L** that the RMS error is 2.5, 1.7 and 1.5 kcal/mol for  $S^{\text{top}}$ ,  $S^{\text{tight}}$  and  $S^{\text{loose}}$  sets, respectively. As noted previously, the role of H-bonds in relative energy of conformers of **HEDP** is less dominant in comparison with **MBP** and the simple model presented by

Equation 8 is less expected to be accurate. Factors other than H-bonding that affect the relative stability of conformers are implicitly included in this model as parameters  $\alpha$  and  $\beta$  but their role is assumed to be the same for all conformers of a system. Obviously this is a crude estimate for the role of non-H-bonding factors such as the orientation of those OH groups not involved in H-bonding or other steric or electronic interactions between different structural components such as the methyl group and hydroxyl or phosphonate groups. However, the results presented in Figure 6 show that the H-bond energy estimation provided by the local density of potential energy at H-bond critical point is a reliable and consistent measure of H-bond strength over the whole conformational space of a molecule. If correlations and regularities obtained here remain valid in other systems such as peptides, there might be an interesting opportunity for development of more reliable structural scoring functions based on relations such as Equation 7, of course, after some careful atom-typing.

#### 4. Summary and Conclusion

After more than four decades, therapeutic features of bisphosphonates are still the subject of active research<sup>40</sup>. Extensive conformational search performed in this study results in 4, 8, 22 and 37 conformers for **MBP-HL**, **MBP-H<sub>2</sub>L**, **HEDP-HL** and **HEDP-H<sub>2</sub>L**, respectively. Free energy calculations including continuum solvation free energies show that at ordinary temperatures only the most stable conformer (**C1**) is populated in both protonation states of **MBP** while the **HEDP** system is a mixture of three conformers (**C1**, **C2** and **C3**) regardless of its protonation state. Possible H-bonding motifs were defined and assigned to each conformer to analyze the role of H-bonds in conformational diversity of a hydroxy-bisphosphonate moiety. The inter-phosphonate H-bonding found to be more preferred than the hydroxyl-phosphonate H-bonding patterns especially in the **HL** protonation state.

Two different approaches based on topological and geometrical criteria were compared with each other for identification and characterization of H-bonds. The most important correlations between H-bond features were analyzed within or between each of these approaches. Among geometrical parameters, the hydrogen-acceptor distance exhibits the best correlations with other parameters. Among topological features, the potential energy density at H-bond critical point and the electron density itself are possible candidates for construction of regular relations for characterization of H-bonds. As a result of true-false nature of critical point identification, numerous weak H-bonds over the conformational space are dropped from topological analysis. Though weak in strength, these H-bonds play an important role in fine regulation of the relative stability of different conformers. The tunable character of geometrical features or other non-topological measures such as the magnitude of frequency shift seems to provide a better handling of the fuzzy nature of H-bond identification problem.

On the other hand, beside the initial identification stage, the topological approach has the advantage of providing a useful quantitative measure for estimation of H-bond strength from local density of potential energy at H-bond critical point. It was shown that there is a strong correlation between this topologically

derived H-bond energy and the hydrogen-acceptor distance. Without an intercept parameter, an inverse power functional form was used to represent this correlation and over a wide range of energy values the prediction errors remain around a fraction of 1 kcal/mol. By extrapolating this relation to the set of topologically unidentified weak H-bonds, a simple model was proposed that estimates the relative stability of conformations from their hydrogen-acceptor distances. Although this model is based on a crude assumption about the role of other factors in relative conformational stabilities, its reasonable accuracy show the consistency of topologically derived measure of H-bond energy.

#### Notes

<sup>a</sup> Department of Physical Chemistry, School of Chemistry, College of Science, University of Tehran, Tehran, Iran.

<sup>b</sup> Department of Bioinformatics, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran. Fax: +98 (0)21 66404680; Tel: +98 (0)21 66969257; E-mail: [mhkarimijafari@ut.ac.ir](mailto:mhkarimijafari@ut.ac.ir)

† Electronic Supplementary Information (ESI) available: H-bond motifs, populations, relative energy terms, equilibrium constants and hydrogen bond properties of all conformers in HL and H<sub>2</sub>L forms of HEDP. Fitted functional forms, fit parameters and regression quality obtained between the best correlated features of H-bond. Benchmark calculations with correlation consistent basis sets. Distribution of all H-bonds in **MBP** and **HEDP** over geometrical parameters. See DOI: 10.1039/b000000x/

#### References

- 1 R. Graham, G. Russell, Bisphosphonates, *Bone*, 2011, **49**, 2.
- 2 M. J. Rogers, J. C. Crockett, F. P. Coxon, J. Mönkkönen, *Bone*, 2011, **49**, 34.
- 3 F. H. Ebetino, C. N. Roze, C. E. McKenna, B. L. Barnett, J. E. Dunford, R. G. G. Russell, G. E. Mieling, M. J. Rogers, *J. Organometallic Chem.*, 2005, **690**, 2679.
- 4 S. B. Gabelli, J. S. McLellan, A. Montalvetti, E. Oldfield, R. Docampo, L. M. Amzel, *PROTEINS: Structure, Function, and Bioinformatics*, 2006, **88**, 62.
- 5 F. Yin, R. Cao, A. Goddard, Y. Zhang, E. Oldfield, *J. Am. Chem. Soc.*, 2006, **128**, 3524.
- 6 S. Mukherjee, Y. Song, E. Oldfield, *J. Am. Chem. Soc.*, 2008, **130**, 1264.
- 7 J. E. Dunford, A. A. Kwaasi, M. J. Rogers, B. L. Barnett, F. H. Ebetino, R. G. G. Russell, U. Oppermann, K. L. Kavanagh, *J. Med. Chem.*, 2008, **51**, 2187.
- 8 E. Kotsikorou, E. Oldfield, *J. Med. Chem.*, 2003, **46**, 2932.
- 9 E. Matczak-Jon, V. Videnova-Adrabinska, *Coord. Chem. Rev.*, 2005, **249**, 2458.
- 10 I. Cukrowski, L. Popovic, W. Barnard, S. O. Paul, P. H. van Rooyen, D. C. Liles, *Bone*, 2007, **41**, 668.
- 11 J. P. Rasanen, E. Pohjala, T. A. Pakkanen, *J. Chem. Soc. Perkin Trans.*, 1996, **2**, 39.
- 12 J. P. Rasanen, E. Pohjala, H. Nikander, T. A. Pakkanen, *J. Phys. Chem.*, 1995, **100**, 8230.
- 13 J. Robinson, I. Cukrowski, H. M. Marques, *J. Mol. Struct.*, 2006, **825**, 134.
- 14 K. Ohno, K. Mori, M. Orita, M. Takeuchi, *Current Med. Chem.*, 2011, **18**, 220.
- 15 W. Barnard, S. O. Paul, P. H. van Rooyen, I. Cukrowski, *J. Raman Spectrosc.*, 2009, **40**, 1935.
- 16 M. Arabieh, M. H. Karimi-Jafari, M. Ghannadi-Maragheh, *J. Mol. Model.*, 2013, **19**, 427.
- 17 T. Steiner, *Angew. Chem. Int. Ed.*, 2002, **41**, 48.
- 18 S. J. Grabowski, *J. Phys. Org. Chem.*, 2004, **17**, 18.
- 19 K. Wendler, J. Thar, S. Zahn, B. Kirchner, *J. Phys. Chem. A*, 2010, **114**, 9529.

- 20 G. Hagele, Z. Szakacs, J. Ollig, S. Hermens, C. Pfaff, *Heteroatom Chem.*, 2000, **11**, 562.
- 21 V. Barone, M. Cossi, *J. Phys. Chem. A*, 1998, **102**, 1995.
- 22 M. Cossi, N. Rega, G. Scalmani, V. Barone, *J. Comput. Chem.*, 2003, **24**, 669.
- 5 23 R.A.Pierotti, *Chem. Rev.*, 1976, **76**, 717.
- 24 F. Floris, J. Tomasi, *J. Comput. Chem.*, 1989, **10**, 616.
- 25 P. Su, H. Li, *J. Chem. Phys.*, 2009, **130**, 074109.
- 26 R. M. Balabin, *J. Chem. Phys.*, 2008, **129**, 164101.
- 10 27 R. M. Balabin, *Mol. Phys.*, 2011, **109**, 943.
- 28 T. H. Dunning Jr., *J. Chem. Phys.*, 1989, **90**, 1007.
- 29 D. E. Woon, T. H. Dunning Jr., *J. Chem. Phys.*, 1993, **98**, 1358.
- 30 A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, and J. Olsen, *Chem. Phys. Lett.*, 1999, **302**, 437.
- 15 31 T. Helgaker, W. Klopper, H. Koch, and J. Noga, *J. Chem. Phys.*, 1997, **106**, 9639.
- 32 D.G. Truhlar, *Chem. Phys. Lett.*, 1998, **294**, 45.
- 33 M. W. Schmidt, K. K. Baldrige, J. A. Boatz, S. T. Elbert, M.S. Gordon, J. J. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T.L. Windus, et al, *J. Comput. Chem.*, 1993, **14**, 1347.
- 20 34 D. A. McQuarrie, *Statistical Mechanics*, HarperCollins, New York, 1976.
- 35 R. Chelli, F. L. Gervasio, C. Gellini, P. Procacci, G. Cardini, V. Schettino, *J. Phys. Chem. A*, 2000, **104**, 11220.
- 25 36 A. Bondi, *J. Phys. Chem.*, 1964, **68**, 441.
- 37 P. Popelier, Prentice-Hall: Englewood Cliffs, NJ, 2000.
- 38 E. Espinosa, E. Molins, C. Lecomte, *Chem. Phys. Lett.*, 1998, **285**, 170.
- 39 F. Biegler-König, J. Schönbohm, D. Bayles, *J. Comput. Chem.*, 2001, **22**, 545.
- 30 40. Yi-Liang Liu, Rong Cao, Yang Wang, Eric Oldfield, *ACS med. Chem. Lett.*, 2015, DOI: 10.1021/ml500528x