

Hydrogen Bond Dynamics of Histamine Monocation in Aqueous Solution: How Geometric Parameters Influence the Hydrogen Bond Strength[†]

Gordana Pirc,^a Jernej Stare,^a Janez Mavri,^a and Robert Vianello^{b,*}

^aLaboratory for Biocomputing and Bioinformatics, National Institute of Chemistry, Hajdrihova 19, SI-1000 Ljubljana, Slovenia

^bQuantum Organic Chemistry Group, Ruđer Bošković Institute, Bijenička cesta 54, HR-10000 Zagreb, Croatia

RECEIVED JUNE 26, 2014; REVISED AUGUST 11, 2014; ACCEPTED AUGUST 26, 2014

Abstract. Chemometric statistical approaches involving multiple linear regression (MLR) and principal component analysis (PCA) were employed on a set of 42 distinct snapshot structures of the physiological histamine monocation in aqueous solution along the Car-Parrinello molecular dynamics trajectory, in order to obtain a better insight into the relationship between the geometry parameters of the system and the resulting ν_{NH} stretching frequencies. A simple 2D linear regression of ν_{NH} with $\text{N}_{\text{amino}} \cdots \text{O}_{\text{water}}$ distances gave a very poor correlation ($R^2 = 0.42$), but both MLR and PCA with the inclusion of four directly bonded water molecules offered a notably predictive model that is even able to distinguish two classes of structures based on the Cl^- counterion position. Taking into account waters from the first, second and third solvation shells, sequentially diminished the overall predictive ability of the model, yet increased the number of useful predictors that, in the largest model with 51 solvent molecules, all correspond to bulk water, implying that both chemometric methods are consistent in suggesting that fundamental histamine N–H stretching vibrations are very complex in nature and strongly coupled to the fluctuating environment.

Keywords: hydrogen bond, histamine monocation, multiple linear regression, principal component analysis

INTRODUCTION

Histamine is a biogenic amine and is mostly known for its role as an inflammation mediator. Its functions as a neuromodulator and neurotransmitter in the brain are less well established, but its known scope of actions is expanding. Recently it has gained importance as a signaling molecule in the processes of sleep-wake cycle, appetite control, learning, memory and emotion. Moreover, its signaling paths seem to be involved in conditions such as depression, schizophrenia, Alzheimer's disease and epilepsy.¹ Consequently, processes involving histamine synthesis, transport, metabolism and binding to macromolecules have a significant physiological relevance. In the brain, histamine is synthesized from histidine by a specific enzyme L-histidine decarboxylase, and is taken up into synaptic vesicles by the vesicular monoamine transporter-2.² It is released into synaptic cleft upon depolarisation stimuli and like every other neurotransmitter it has to be removed from synaptic cleft in a relatively short time. The primary histamine metabolising enzyme in the brain seems to be

an intracellular histamine-*N*-methyl-transferase, which produces *N*-methyl-histamine that is later converted to the corresponding aldehyde by monoamine oxidase (MAO). In other pathways, histamine is degraded by diamine oxidase (DAO) and to some extent directly by MAO.

Concerning the nature of molecular interactions between histamine and its larger biological systems, these are dominated by the histamine hydrogen bonding ability,³ just as it is the case when it is surrounded with solvent water molecules under physiological conditions. Therefore, the hydration of histamine is of considerable interest, particularly with regard to structures and vibrational spectra. Histamine is composed of an imidazole ring and an aminoethyl side-chain, and according to its $\text{p}K_{\text{a}}$ values (5.8 for the imidazole ring nitrogen atom and 9.4 for the aliphatic amino group),⁴ histamine is predominantly a monocation (96 %) protonated at the free-amino position.⁵ In cells, histamine recognition induces (at least) partial desolvation, implying that a portion of water molecules from hydration shells must be removed to make a way for the interac-

[†] Dedicated to Dr. Mirjana Eckert-Maksić on the occasion of her 70th birthday.

* Author to whom correspondence should be addressed. (E-mail: robert.vianello@irb.hr)

tion with proteins and enzymes. A very illustrative example is provided by the histamine receptors H1–H4, which are all activated through the hydrogen bonding to histamine,⁶ suggesting that it is highly desirable to properly understand its interactions with physiological water and biological macromolecules. The latter is in line with recent molecular dynamics simulations by Keserü and co-workers⁷ that emphasized the role of water-histamine interactions at the binding site of the human histamine H4 receptor, being in agreement with several previous reports.⁸

Recently, we modeled the nature of the hydration of physiological histamine monocation by using Car-Parrinello molecular dynamics simulations (CPMD)⁹ and a static cluster model.¹⁰ In the former study, we simulated the N–H stretching band envelope, which are degrees of freedom that predominantly influence the strength of the hydrogen bonding as probed by the vibrational spectroscopy. The approach involved the construction of one dimensional proton potentials on snapshot structures extracted from a CPMD trajectory, followed by solving the one-dimensional vibrational Schrodinger equation. By separating the amino N–H vibrations of the protonated ethylamino chain from those within the imidazole ring in the experimental spectra, we clearly demonstrated that the former moiety forms stronger hydrogen bonds with the surrounding water molecules,^{9,10} which should be significant for the biological function of histamine. It remains a challenge to identify how the fluctuating intra- and intermolecular geometric parameters influence the N–H stretching frequencies, which is the aim of the present work. The most important factor seems to be the proton donor-proton acceptor distance among histamine amino nitrogens and water oxygen atoms. Careful comparisons between the experimental donor-acceptor distances (taken from the frozen crystal structures) and the corresponding stretching frequencies for a series of related systems are known as Novak¹¹ and Mikenda¹² relations. Analysis of the CPMD trajectory offers an advantage that it is performed for several configurations of the same system and that all other geometric features could be considered. This approach was previously used in the case of a very strong hydrogen bonding in the crystals of sodium hydrogen bis(sulfate),^{13,14} where, remarkably, no appreciable correlation between the instantaneous geometric parameters and the pertinent OH stretching frequency could be found, demonstrating the extreme complexity of the short and strong hydrogen bond. In contrast, aqueous histamine forms relatively long and weak hydrogen bonds with water;^{9,10} hence part of the motivation for the present study is a comparison of the geometry-frequency correlation patterns between strong and weak hydrogen bonds.

COMPUTATIONAL DETAILS

As part of our interest in examining the hydrogen bond dynamics of the histamine monocation in aqueous solution,^{9,10} the present work analyzes the snapshot structures along the CPMD trajectory with chemometric methods – a combination of the mathematical and statistical approaches designed to provide maximum chemical information of the chemical system by analyzing chemical data. For the CPMD simulation details, extraction of snapshot geometries, and the calculation of instantaneous proton potentials and thus derived anharmonic frequencies, the reader is referred to reference 9. Briefly, we collected 42 distinct geometries during 10 ps simulation, involving the histamine monocation, 93 water molecules and Cl[−] counterion, and since each snapshot structure consists of four N_{amino}···O_{water} moieties, we tailed a set of 168 anharmonic frequencies.

In order to study the relationship between geometric parameters and the corresponding ν_{NH} frequencies, we started the analysis by considering all internal coordinates of the studied system (termed predictors throughout the text), based on the fact that the main type of interactions in the simulation box is electrostatic in nature, which is essentially long-range in character, therefore the proton dynamics is controlled by surroundings that extend beyond the nearest atoms. However, this made a visual aspect of such multi-dimensional correlation a very difficult and challenging task. To overcome this problem, we reduced the data dimension by employing the Principal Component Analysis (PCA), being recommended as the most appropriate method for preprocessing multi-dimensional data into two dimensions.¹⁵ We were also interested in finding a relationship between a single dependent variable (ν_{NH}) and several independent variables (geometry parameters), for which we utilized the Multiple Linear Regression (MLR) as a calibration technique.^{15,16}

As first, MLR and PCA analysis have been applied on the matrix of $42 \times n$ dimensions, corresponding to 42 snapshot structures with n internal coordinates in columns (see Table 1). The elements of particular columns are given as a quotient of the structural parameters (bond distances, valence angles, dihedral angles). The total number of variables is obtained in the following way. For the system with N atoms, the number of internal degrees of freedom is $3N-6$, yielding 888 unique internal coordinates that describe each snapshot structure with 298 atoms.⁹ Since one internal coordinate of each H-bonded moiety served as a “probe” coordinate along which the proton was displaced in a stepwise manner when calculating the corresponding proton potentials, we omitted it from our analysis. For this purpose, the periodic boundary conditions and replacement of water molecules were also considered. Our

Table 1. Matrix dimensions used for the chemometric analysis. The internal coordinate space was formulated by considering different number of water molecules (according to the distance between the nitrogen N_{amino} and the oxygen O_{water} of bulk water) for the each N–H \cdots O hydrogen-bonded moiety

Interactions $N_{\text{amino}}\cdots O_{\text{water}}$	$d(N\cdots O)^{(a)}$ / Å	$N(\text{atoms})$	$N(\text{H}_2\text{O})$	$N(\text{int. coord.})$	$N(\text{distances})$	$N(\text{angles})$	$N(\text{torsions})$	matrix ^(b)
closest H ₂ O	< 3.0	31	4	87	30	29	28	42×30 (M1)
1 st shell waters	< 4.5	61	15	177	60	59	58	42×60 (M2)
2 nd shell waters	< 6.4	130	37	384	129	128	127	42×129 (M3)
3 rd shell waters	< 7.2	172	51	510	171	170	169	42×171 (M4)
all H ₂ O		298	93	888	297	296	295	42×297

^(a) Euclidean distance $N_{\text{amino}}\cdots O_{\text{water}}$ between the amino nitrogen and oxygen from bulk water.

^(b) 42 rows represent snapshot structures composed of 30, 60, 129, 171 and 297 variables (internal coordinates in columns).

initial inspection suggested poor correlation between all geometry parameters and ν_{NH} at best, giving evidence that the investigated HB interactions are extremely complex as was the case with the strong hydrogen bond in sodium hydrogen bis(sulfate).¹⁴ However, in contrast to the virtually nonexistent correlation reported in the latter study, here we analyzed each of the individual subsets of geometric parameters separately (bond distances, bond angles, dihedral angles), and found that when examining only bond distances there appears to be a certain correlation with ν_{NH} values. Therefore, MLR and PCA have been applied on the reduced matrices of $42 \times m$ dimensions, where m denotes the number of bond distances, yielding matrices **M1**–**M4** with $m = 30, 60, 129$ and 171 , respectively (Table 1). We further refined the process by sequentially increasing the number of considered water molecules, starting with four directly bonded molecules, and proceeding by employing the distance criteria of $d(N_{\text{amino}}\cdots O_{\text{water}}) < 4.5$ Å, $d(N_{\text{amino}}\cdots O_{\text{water}}) < 6.4$ Å, and $d(N_{\text{amino}}\cdots O_{\text{water}}) < 7.2$ Å for the first, second and third solvation shells, corresponding to 15, 37 and 51 waters, respectively (Table 1, Figure 1). In addition, we categorized all datasets by assigning class numbers, firstly to the range of Cl \cdots O distances (class attributes **CIO1**: $d(\text{Cl}\cdots\text{O}) < 3.5$ Å; **CIO2**: $d(\text{Cl}\cdots\text{O}) \geq 3.5$ Å), and secondly, to the N–H stretching frequencies (class attributes **NH1**: $\nu_{\text{NH}} < 2400$

cm^{-1} ; **NH2**: $2700 \leq \nu_{\text{NH}} < 2400$ cm^{-1} ; **NH3**: $3000 \leq \nu_{\text{NH}} < 2700$ cm^{-1} ; **NH4**: $\nu_{\text{NH}} > 3000$ cm^{-1}). All calculations and preparation of plots were done with the Teach/Me software¹⁷ using Teach/Me Data Analysis application.

RESULTS AND DISCUSSION

Review of the Calculated Anharmonic Frequencies and the N–H Stretching Envelope

The calculated fundamental excitation frequencies span a wide range from 3229 cm^{-1} to 2136 cm^{-1} , being a consequence of very diverse snapshot geometries used to calculate proton potentials of different shapes.⁹ This Gaussian-type frequency distribution represents the N–H stretching envelope (Figure 2b) having the maximum absorption between 2820 cm^{-1} and 2730 cm^{-1} , and the center of gravity at 2799 cm^{-1} . According to the correlation scheme of Novak,¹⁸ the latter frequency corresponds to the $N_{\text{amino}}\cdots O_{\text{water}}$ distance of about 2.75 Å, which is around 0.1 Å shorter than the calculated CPMD-averaged value of 2.85 Å.⁹ However, it should be noted that in contrast to the dynamic aqueous environment of the present system, Novak and co-workers established their scheme on a dataset of solid crystalline structures, hence this offset is likely to be of little significance. Another possible reason for a disagreement is

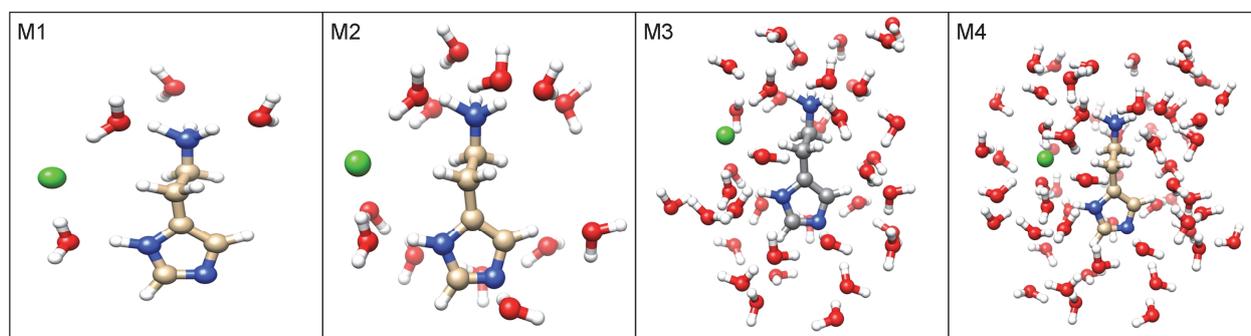


Figure 1. Molecular models of the histamine monocation in aqueous solution constructed by using different number of water molecules in hydration shells: 4 (**M1**), 15 (**M2**), 37 (**M3**), and 51 (**M4**).

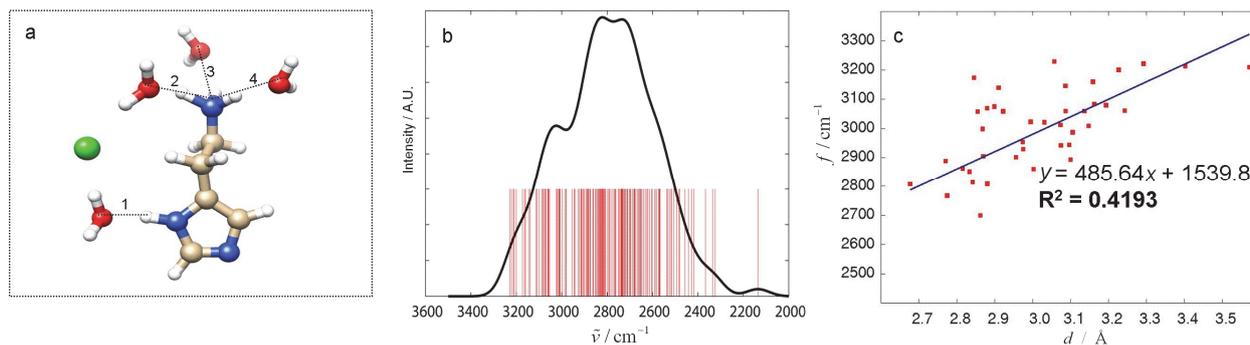


Figure 2. (a) Structure of the investigated histamine monocation with the labeled H-bonded moieties (HB1–HB4); (b) Distribution of the anharmonic NH stretching transitions (red vertical lines) obtained from proton potentials extracted from the snapshot structures along the CPMD trajectory. A continuous representation of this distribution (black curve) is obtained as a superposition of the Gaussian functions (one for each transition) with a half width of 50 cm^{-1} ; (c) Simple two-dimensional plots used to obtain preliminary qualitative information about the relationship between anharmonic N–H stretching frequencies and the corresponding $N_{\text{amino}}\cdots O_{\text{water}}$ distances. The calculated standard errors of the slope, interception, and regression assume 92.71, 280.1, and 106.2, respectively.

that the applied DFT method introduced systematic error in the calculation of proton potentials that is reflected in shifted frequencies. Nevertheless, our computed results are in excellent agreement with the experimental spectra in terms of both the shape and position of the N–H stretching band.⁹

Multiple Linear Regression (MLR)

In the first stage, ν_{NH} frequencies were visualized in a simple 2D linear regression with only $N_{\text{amino}}\cdots O_{\text{water}}$ distances (Figure 2c), which yielded a very poor correlation ($R^2 = 0.42$), but still considerably better than found for the strong hydrogen bonding in crystalline sodium hydrogen bis(sulfate) ($R^2 \approx 0.2$ at best). Since we were particularly interested in whether the hydrogen bond donor-acceptor distance is a better predictor of ν_{NH} than other geometry parameters, at first, we applied the MLR technique to all four $N_{\text{amino}}\cdots O_{\text{water}}$ moieties solvated by four directly bonded water molecules (Figure 2a). Since the mentioned fragments are located in different environments, for example relative to the position of the Cl^- counterion, the obtained results differed, which led us to apply the MLR analysis to larger matrices of different sizes (see later). One way to express the overall predictive accuracy of multiple regression models is the R^2 value,¹⁹ a measure of how much variation a model explains. With short $N_{\text{amino}}\cdots O_{\text{water}}$ distances, the model is quite reasonable yielding R^2 values above 0.75 (Table 2), and for the hydrogen bonds labeled **HB1–HB3** (Figure 2a) it provides significant correlation between geometry parameters and ν_{NH} , while for the last one (labeled **HB4**) we found a weak correlation between data. Interestingly, for each of the four $N_{\text{amino}}\cdots O_{\text{water}}$ moieties one should consider different predictors to describe the corresponding ν_{NH} values in the best way. In other words,

we did not find a single set of predictors that would work well with all four N–H groups. In addition, it turned out that the data in the full matrix were distinguished according to the $\text{Cl}\cdots O_{\text{water}}$ distances (see Computational Details), which led us to apply the MLR technique on smaller matrices with $17 \times m$ (**CIO1**) and $25 \times m$ (**CIO2**) dimensions (Figure 3), which increased the R^2 values in both cases. From the data we can conclude that the length of the $C_{\alpha}\text{--}N_{\text{amino}}$ distance is the most influential predictor in both the protonated $-\text{NH}_3^+$ and ring-amino group, labeled as d7 and d1 in Figure 3, respectively.

Secondly, we upgraded our model by considering the coordinates of 15 water molecules from the first hydrated shell all with $d(N_{\text{amino}}\cdots O_{\text{water}}) < 4.5\text{ \AA}$ (Table 2 and Figure 4). It follows that the water molecule labeled with Y is a useful predictor of only one $N_{\text{amino}}\cdots O_{\text{water}}$ moiety, while water labeled X is not listed as predictor in studied cases (Figure 4b). For all four H-bonded moieties the common predictors within the histamine monocation are d2 and d7, together with d19 from the bulk. Further improvement was made by including more distant second hydration shell solvent molecules in the analysis ($d(N_{\text{amino}}\cdots O_{\text{water}}) < 6.4\text{ \AA}$) yielding a cluster of 37 waters. Again, $\text{Cl}\cdots O_{\text{water}}$ distances turned out to be good predictors, together with the majority of other predictors that are all corresponding to bulk water molecules. From the histamine monocation, one should consider d7, listed in **HB1**, **HB2** and **HB3**, and d1 listed in **HB2** and **HB3** moieties. It is interesting that no predictors are listed in **HB4**. Finally, third hydration shell ($d(N_{\text{amino}}\cdots O_{\text{water}}) < 7.2\text{ \AA}$) with 51 water molecules was also considered. In this case, $\text{Cl}\cdots O_{\text{water}}$ distances are not good predictors, whereas the majority of predictors are geometry parameters of bulk water molecules. It is interesting that, the predictor d8 ($N_{\text{amino}}\cdots O_{\text{water}}$ distance) can be

Table 2. Comparison of R^2 values in the multiple linear regression (MLR) using predictors P with $R^2 > 0.75$

H-bonded moiety ^(a)	HB1			HB2			HB3			HB4												
	1	1a	1b	2	2a	2b	3	3a	3b	4a	4b											
1	P d27	R^2 0.7819	P d10	R^2 0.7665	P d23	R^2 0.7519	P d7	R^2 0.7926	P d21	R^2 0.7710	P d6	R^2 0.7782	P d1	R^2 0.8200	P d29	R^2 0.7769	P d12	R^2 0.8494	P d11	R^2 0.7710		
2	d16	0.7626	d15	0.8354	d1	0.8146	d2	0.7873	d3	0.8510	d5	0.8098	d9	0.7953	d16	0.8750	d1	0.8050	d1	0.9081	d20	0.8252
3	d3	0.7666	d30	0.8917	d6	0.8629	d29	0.8027	d9	0.8785	d19	0.8633	d7	0.8241	d21	0.8981	d2	0.8515	d27	0.9493	d14	0.8643
4	d17	0.7709	d31	0.9493	d7	0.9419	d19	0.8149	d23	0.9123	d28	0.8879	d20	0.8335	d23	0.9204	d30	0.8780	d21	0.9684	d15	0.8921
5	d2	0.7749	d26	0.9832	d4	0.9525	d28	0.8197	d11	0.9365	d8	0.9100	d18	0.8446	d4	0.9588	d15	0.9004	d15	0.9802	d12	0.9175
6	d11	0.7777	d18	0.9951	d8	0.9623	d8	0.8264	d5	0.9788	d18	0.9313	d17	0.8555	d13	0.9787	d24	0.9164	d18	0.9850	d27	0.9325
7	d13	0.7991	d7	0.9991	d18	0.9691	d7	0.8387	d18	0.9914	d2	0.9538	d28	0.8616	d26	0.9893	d10	0.9391	d5	0.9942	d7	0.9533
8	d31	0.8006	d23	0.9997	d11	0.9761	d16	0.8450	d10	0.9980	d11	0.9638	d15	0.8650	d18	0.9930	d25	0.9462	d17	0.9989	d19	0.9640
9	d9	0.8024	d5	0.9999	d28	0.9829	d9	0.8494	d14	0.9998	d12	0.9800	d8	0.8681	d11	0.9978	d8	0.9545	d6	0.9999	d18	0.9799
10	d8	0.8044	d16	1.0000	d17	0.9886	d11	0.8518	d28	1.0000	d14	0.9860	d29	0.8722	d31	0.9997	d9	0.9654	d22	1.0000	d21	0.9882
11	d10	0.8055	d13	1.0000	d9	0.9896	d25	0.8538	d19	1.0000	d22	0.9894	d11	0.8760	d10	1.0000	d22	0.9790			d9	0.9922
12	d23	0.8063	d6	1.0000	d24	0.9915	d5	0.8562			d26	0.9917	d13	0.8839			d16	0.9882			d22	0.9966
13	d26	0.8072			d16	0.9921	d14	0.8575			d1	0.9955	d27	0.8915			d12	0.9953			d17	0.9989
14	d25	0.8083			d21	0.9940	d21	0.8591			d30	0.9975	d23	0.8981			d4	0.9979				
15	d29	0.8095			d5	0.9954	d24	0.8613			d27	0.9991	d21	0.8997			d3	0.9986				
16	d6	0.8101			d31	0.9962	d12	0.8641			d25	0.9999	d25	0.9007			d21	0.9993				
17	d5	0.8102			d29	0.9967	d17	0.8642			d31	1.0000	d5	0.9012			d18	0.9997				
18	d24	0.8102			d30	0.9979	d10	0.8643					d30	0.9020			d6	1.0000				
19																						
20																						
21																						

^(a) The numbering of the hydrogen bonded moieties **HB1–HB4** is presented in Figure 1a and used consistently throughout the text.

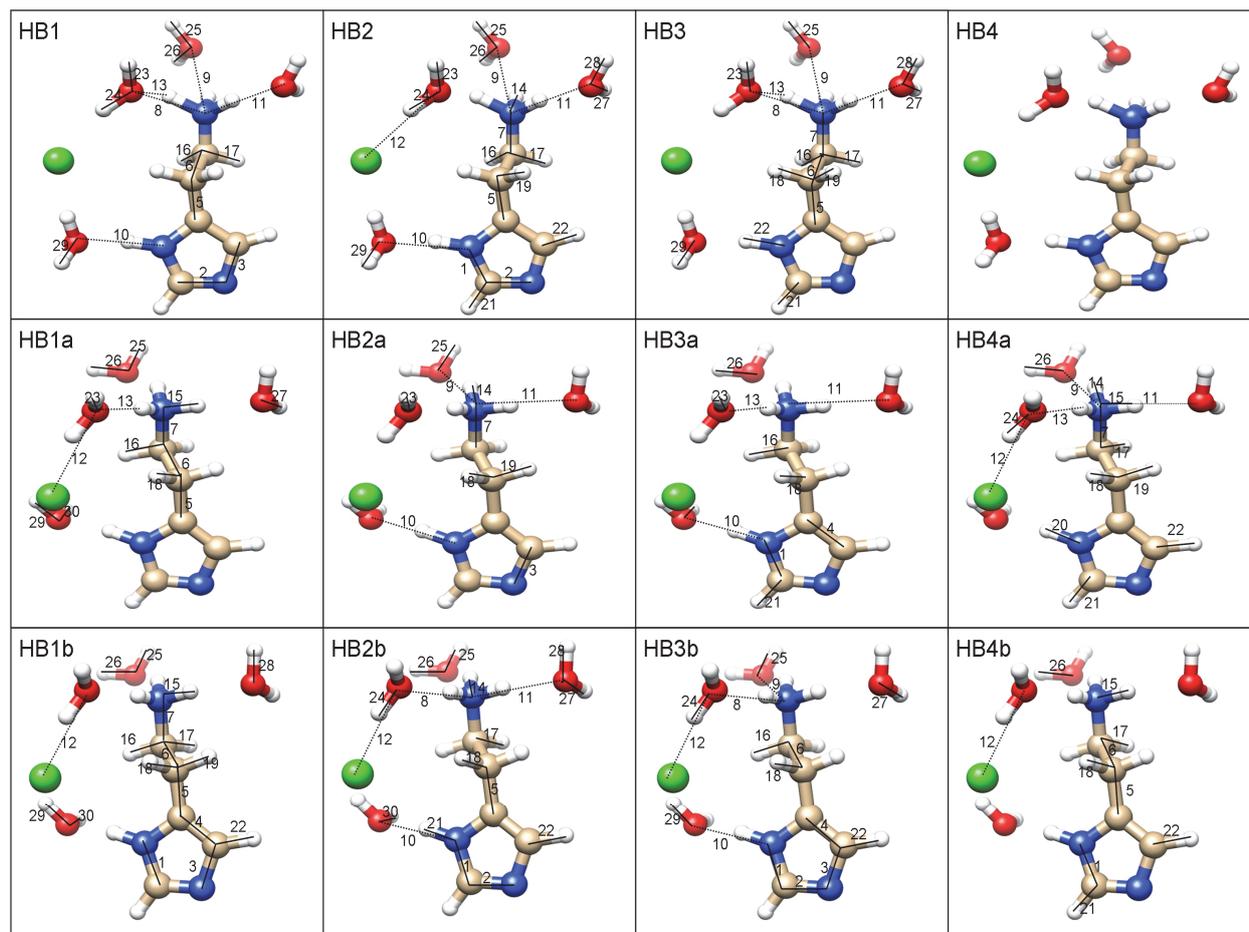


Figure 3. First N exploratory variables (numbered according to the predictor names listed in Table 2), which are the most significant predictors for the hydrogen bonds **HB1**–**HB4** (denoted with numbers "1"–"4") in the model with four directly bonded water molecules. Labels "a" and "b" correspond to **CIO1** and **CIO2** classes of structures, respectively.

found in **HB1**, **HB2** and **HB4**, and not in **HB3** moiety. Moreover, the important predictors from histamine monocation are again d7 in **HB1**, and d6 and d3 in **HB4** group.

Summarizing this section, we can conclude that the linear correlation between geometry parameters and individual frequencies exists. It is also worth stressing out, that by increasing the number of solvent

water molecules, the features related to the complexity of the proton motion and its coupling to the environment become more apparent.

Principal Component Analysis (PCA)

PCA was performed in order to get an insight into the overall correlativity of all 168 ν_{NH} values with the

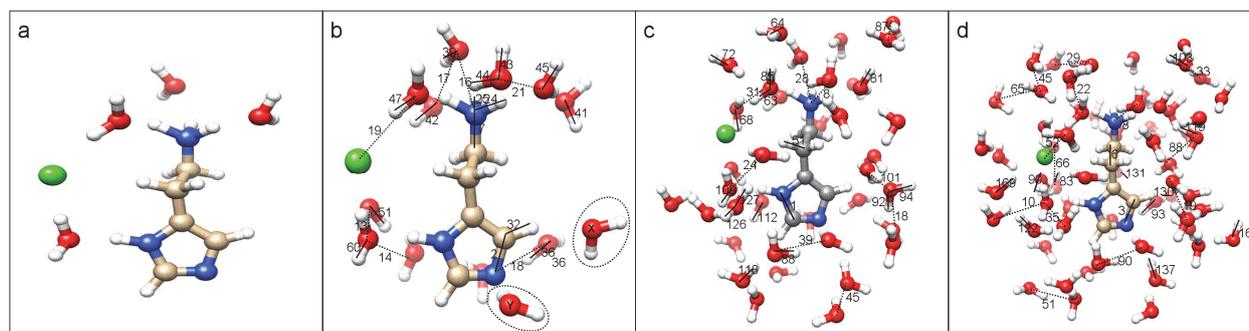


Figure 4. First N predictors of the **HB4** moiety in models with different number of solvent water molecules.

Table 3. Comparison of the variances^(a) in the Principal Component Analysis (PCA) using different matrix sizes

N(PC)	HB-group and type of H ₂ O interaction											
	HB1	HB2	HB3	HB4	HB1a	HB2a	HB3a	HB4a	HB1b	HB2b	HB3b	HB4b
	closest H ₂ O				closest H ₂ O (geometry class CIO1)				closest H ₂ O (geometry class CIO2)			
1	97.71	97.76	97.95	97.71	58.89	59.64	62.27	59.00	83.75	83.75	84.20	83.72
2	0.76	0.75	0.72	0.76	16.42	16.62	16.52	16.46	6.38	6.42	6.38	6.43
3	0.63	0.64	0.60	0.63	8.73	8.85	8.21	8.72	4.43	4.43	4.38	4.41
4	0.32	0.32	0.24	0.32	6.26	6.06	4.03	6.27	2.07	2.06	1.92	2.07
5	0.15	0.15	0.15	0.15	2.79	2.75	2.98	2.79	1.33	1.31	1.20	1.32
6	0.12	0.11	0.07	0.12	2.08	2.02	1.63	2.04	0.47	0.48	0.49	0.48
7	0.06	0.05	0.05	0.06	1.39	0.75	0.94	1.40	0.40	0.42	0.33	0.42
8	0.05	0.03	0.03	0.05	0.72	0.67	0.71	0.68	0.27	0.28	0.23	0.27
9	0.03	0.03	0.03	0.03	0.55	0.59	0.56	0.53	0.21	0.20	0.21	0.21
10	0.02	0.02	0.02	0.02	0.41	0.39	0.41	0.39	0.20	0.18	0.19	0.19
	1 st shell H ₂ O				2 nd shell H ₂ O				3 rd shell H ₂ O			
1	18.36	18.36	18.36	18.22	11.91	11.91	11.91	11.73	11.22	11.22	11.22	11.22
2	11.67	11.67	11.67	11.59	8.25	8.25	8.25	8.11	8.54	8.54	8.54	8.54
3	10.86	10.86	10.86	10.82	7.55	7.55	7.55	7.22	8.02	8.02	8.02	8.02
4	8.34	8.34	8.30	8.30	6.63	6.63	6.63	6.65	7.16	7.16	7.16	7.16
5	6.90	6.90	6.97	6.97	6.34	6.34	6.34	6.37	6.67	6.67	6.67	6.67
6	6.29	6.29	6.33	6.33	5.40	5.40	5.40	5.20	5.75	5.75	5.75	5.75
7	5.46	5.46	5.43	5.43	4.84	4.84	4.84	4.80	5.34	5.34	5.34	5.34
8	5.22	5.22	5.17	5.17	4.62	4.62	4.62	4.59	4.91	4.91	4.91	4.91
9	3.92	3.92	3.92	3.92	4.34	4.34	4.34	4.12	4.75	4.75	4.75	4.75
10	3.36	3.36	3.36	3.33	3.77	3.77	3.77	3.57	4.15	4.15	4.15	4.15

^(a) Euclidean distance $N_{\text{amino}} \cdots O_{\text{water}}$ between the amino nitrogen and oxygen from bulk water.

^(b) 42 rows represent snapshot structures composed of 30, 60, 129, 171, and 297 variables (internal coordinates in columns).

matching geometry parameters. Similarly to the MLR, PCA was applied for each H-bonded moiety **HB1–HB4** on matrices composed of 42 rows and sets of 30, 60, 129, and 171 variables in columns (**M1–M4**), the latter corresponding to the number of considered internal bond distances in each case. The data was additionally preprocessed by using mean centered data, in other words, by subtracting the overall mean value from individual elements.

PCA analyzes data by yielding eigenvectors (principal components, PCs) and the corresponding eigenvalues (variances) by using the covariance matrix of the whole dataset. The majority of the information is usually gathered in the first few PCs; in our case, the smallest matrix **M1** already exhibits such behavior as already around 98 % variance is explained in the first two principal components in all four $N_{\text{amino}} \cdots O_{\text{water}}$ moieties (Table 3), with the most influencing variables being distances d10, d11, d12, and d13 (Figure 5). We explored a number of pair-wise plots of principal components and compared the corresponding score vs. score displays, and additionally categorized all datasets by

assigning class numbers according to a range of both $Cl^- \cdots O_{\text{water}}$ distances (**CIO1** and **CIO2**) and N–H stretching frequencies (**NH1–NH4**) (see Supplementary Materials). Interestingly, in **M1**, apart from the obvious $N_{\text{amino}} \cdots O_{\text{water}}$ distances, one of the most influencing parameters is $Cl^- \cdots O_{\text{water}}$ distance (d12), in agreement with the MLR analysis presented earlier, which is basically responsible that the results clusters in **CIO1** (having 25 structures) and **CIO2** (having 17 structures) classes, with the prevalence of each determined by d9 and d14 variables, respectively (Table 3).

By increasing the number of water molecules in the model one also increases the number of the most influencing geometry parameters while diminishing the significance of the two classes based on the counterion position. In the case of the largest model **M4** (51 water molecules), almost all significant geometry parameters correspond to bulk water molecules. The only exception is provided by d8, which is related to the $N_{\text{amino}} \cdots O_{\text{water}}$ moiety.

Further, we were interested in how changing the number of variables affects variances in the first ten

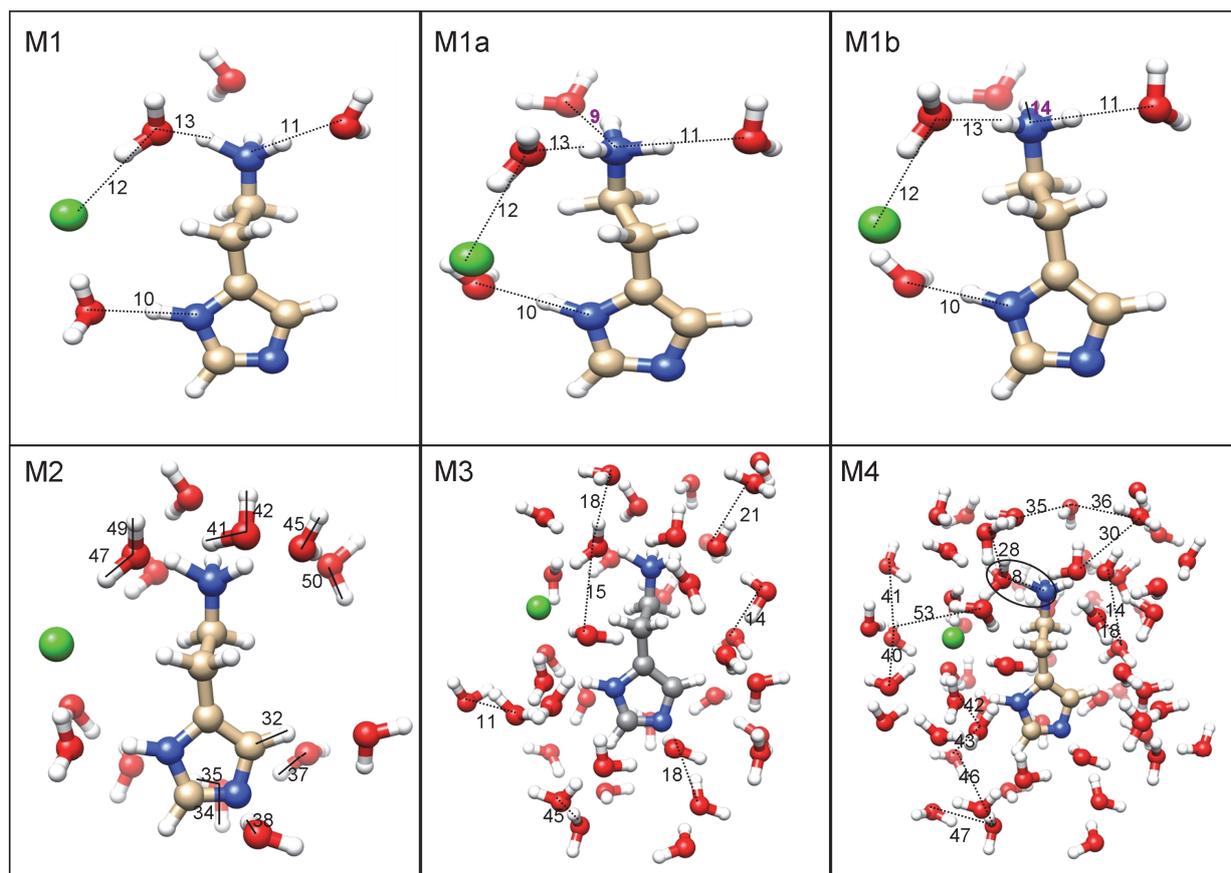


Figure 5. The most influencing variables obtained by the PCA analysis in the constructed models. Insets correspond to different number of considered solvent water molecules, being 4 (**M1**), 15 (**M2**), 37 (**M3**) and 51 (**M4**), whereas **M1a** and **M1b** denote **ClO1** and **ClO2** classes of structures, respectively.

PC axes, and found that it exerts pronounced influence (Table 3). By increasing the number of variables, the content of variances decreases, while the percentage of the first three variances in the resulting PCs changes only distinctly. This could be ascribed to the Cl⁻ counterion, which has a more profound effect in the smaller models, and to the complexity of the proton motion and its coupling to its environment. On the other hand, when comparing percentages of the first three variances within similar models (e.g. **M1** model with the first shell waters) they are almost identical (Table 3). It follows that with PCA method we obtained similar conclusion as from the MLR study – instantaneous fluctuating structure has a pronounced effect on the proton motion and its dynamic, which is very complex and differently coupled to the environment.

Along with other established computational methods, the presented approach will contribute towards understanding of the complexity of the hydrogen bonding dynamics in polar environments reflected in the vibrational spectra^{20–24} and its relevance for biocatalysis.²⁵

CONCLUDING REMARKS

The aim of our study was to evaluate the correlation between the frequencies of the N–H stretching band and the corresponding geometry parameters of the snapshot structures of the histamine monocation in aqueous solution. A simple two-dimensional correlation scheme between ν_{NH} and the matching $N_{\text{amino}} \cdots O_{\text{water}}$ distances appeared not to have significant predictive ability. On the other hand, more advanced techniques, such as multiple linear regression (MLR) and principal component analysis (PCA), applied on preprocessed data yielded reasonably accurate predictors. For example, around 98 % variance in the data is well explained by the first two principal components in all four $N_{\text{amino}} \cdots O_{\text{water}}$ moieties. Moreover, with PCA method applied on the cluster involving only four directly attached water molecules we were able to distinguish two classes of structures based on the counterion position. When the analysis was additionally tuned by assigning class numbers according to the value of the $\text{Cl} \cdots O_{\text{water}}$ distances, the percentage of the variance in the first two principal components in all $N_{\text{amino}} \cdots O_{\text{water}}$ moieties decreases to around 90 %.

By increasing the number of water molecules in the model one also increases the number of the most influencing geometry parameters and diminishes the relevance of the mentioned two structural classes that depend on the position of the Cl^- counterion. With the largest model investigated here (51 water molecules), all influential geometry parameters arise from bulk water, and only around 20 % of the variance is explained in the first two principal components of four $\text{N}_{\text{amino}} \cdots \text{O}_{\text{water}}$ groups. This might lead to a conclusion that a correlation between instantaneous fluctuating histamine structure and the N–H stretching frequencies is likely to be poor, and that the dynamic of the proton motion is very complex and strongly coupled to the fluctuating environment. The results obtained from the MLR analysis support this assumption.

Despite the fact that the structure–frequency correlation found for a set of dynamically sampled instantaneous snapshot structures *within one system* is only moderate, it represents a clear advancement in comparison with the previously investigated system of crystalline sodium hydrogen bis(sulfate),¹⁴ and demonstrates once again that hydrogen bonding is an extremely complex interaction. For the mentioned inorganic material, no appreciable linear or even non-linear correlation could be found, indicating the limitations of the applied one-dimensional quantum treatment. Remarkably, the present system features much weaker and longer hydrogen bonds, which is reflected in the observed correlations between the instantaneous geometric parameters and the corresponding N–H stretching frequencies.

Supplementary Materials. – Supporting informations to the paper, containing selected pair-wise plots of principal components PC1 vs. PC2 together with the corresponding score vs. score displays, are enclosed to the electronic version of the article. These data can be found on the website of *Croatica Chemica Acta* (<http://public.carnet.hr/ccaaa>).

Acknowledgements. G. P., J. S. and J. M. would like to thank the Slovenian Research Agency for financial support in the framework of the program group P1–0012 and within the corresponding research project contract No. J1–2014. R. V. gratefully acknowledges the European Commission for an individual FP7 Marie Curie Career Integration Grant (contract number PCIG12–GA–2012–334493). Part of this work was supported by COST Action CM1103.

REFERENCES

1. K. Yanai and M. Tashiro, *Pharmacol. Ther.* **113** (2007) 1–15.
2. A. Merickel and H. Edwards, *Neuropharmacology* **34** (1995) 1543–1547.
3. P. Gilli, L. Pretto, V. Bertolasi, and G. Gilli, *Acc. Chem. Res.* **42** (2009) 33–44.
4. D. G. Cooper, R. C. Young, G. J. Durant, and C. R. Ganelin, in: *Comprehensive medicinal chemistry: the rational design, mechanistic study and therapeutic application of chemical compounds*, C. Hansch (ed.), Pergamon Press, Oxford, U.K., 1990, 323–421.
5. W. L. Nelson, in: *Foye's Principles of Medicinal Chemistry*. T. L. Lemke, D. A. Williams, V. F. Roche, and S. W. Zito (eds.), 7th edition, Baltimore, Philadelphia: Lippincott Williams & Wilkins, 2013.
6. S. J. Hill, C. R. Ganelin, H. Timmerman, J. C. Schwartz, N. P. Shankley, J. M. Young, W. Schunack, R. Levi, and H. L. Haas, *Pharmacol. Rev.* **49** (1997) 253–278.
7. B. Jójárt, R. Kiss, B. Viskolcz, G. M. Keserű, *J. Chem. Inf. Model.* **48** (2008) 1199–1210, and references cited therein.
8. (a) J. Henin, B. Maigret, M. Tarek, C. Escricout, D. Fourmy, C. Chipot, *Biophys. J.* **90** (2006) 1232; (b) L. Rivail, C. Chipot, B. Maigret, I. Bestel, S. Sicsic, M. Tarek, *J. Mol. Struct. (Theochem)* **817** (2007) 19–26.
9. J. Stare, J. Mavri, J. Grdadolnik, J. Zidar, Z. B. Maksić, and R. Vianello, *J. Phys. Chem. B* **115** (2011) 5999–6010.
10. R. Vianello and J. Mavri, *New J. Chem.* **36** (2012) 954–962.
11. R. Blinc, D. Hadži, and A. Novak, *Bericht. Buns. Phys. Chem.* **64** (1960) 567.
12. W. Mikenda, *J. Mol. Struct.* **147** (1986) 1–15.
13. G. Pirc, J. Stare, and J. Mavri, *J. Chem. Phys.* **132** (2010) 224506.
14. G. Pirc, J. Mavri, M. Novič, and J. Stare, *J. Phys. Chem. B* **116** (2012) 7221–7231.
15. I. T. Jolliffe, *Principal component analysis*, 2nd edition, Springer: New York, 2002.
16. Y. Dodge and F. H. C. Marriott, *International Statistical Institute. The Oxford dictionary of statistical terms*, 6th edition, Oxford University Press: Oxford, UK, 2003.
17. H. Lohninger, Teach/Me Data Lab. 2.002. in: *SDL-Software Development Lohninger*, Springer: Berlin, 1999.
18. A. Lautié, F. Froment, and A. Novak, *Spectrosc. Lett.* **9** (1976), 289–299.
19. H. F. Kaiser, *Educ. Psychol. Meas.* **20** (1960) 141–151.
20. M. Z. Breła, M. J. Wójcik, M. Boczar, and R. Hashim, *Chem. Phys. Lett.* **558** (2013) 88–92.
21. N. Rekik and M. J. Wójcik, *Chem. Phys.* **369** (2010) 71–81.
22. M. Boczar, R. Kurczab, and M. J. Wójcik, *Vibrat. Spectr.* **52** (2010) 39–47.
23. N. Issaoui, N. Rekik, B. Oujia, and M. J. Wójcik, *Int. J. Quant. Chem.* **110** (2010) 2583–2602.
24. N. Došlić, M. Abdel-Latif, and O. Kuehn, *Acta Chim. Slov.* **58** (2011) 411–424.
25. S. C. L. Kamerlin, J. Mavri, and A. Warshel, *FEBS Lett.* **584** (2010) 2759–2766.

SUPPLEMENTARY MATERIALS

Plots of the first two principal components (PCs) resulting from the principal component analysis of 42 snapshot structures represented as N variables (N – number of distances in the constructed models; see Table 1 and Figure 1), displayed as PC1/PC2 scores (a–d; results for the **HB4** moiety in the model with 4 directly bonded water molecules **M1**) and as loadings (vectors) against each other (e–h).

