

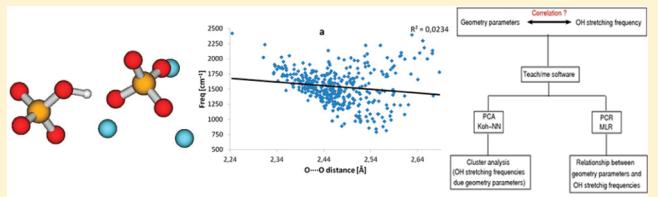
Virtually Nonexistent Correlation between the OH Stretching Frequency and the Instantaneous Geometry in the Short Hydrogen Bond of Sodium Hydrogen Bis(sulfate): Advanced Chemometrics Analysis

Gordana Pirc, Janez Mavri, Marjana Novič, and Jernej Stare*

National Institute of Chemistry, Hajdrihova 19, SI-1000 Ljubljana, Slovenia

S Supporting Information

ABSTRACT: We examined the correlation between the dynamically sampled anharmonic frequency of the OH stretching motion and the corresponding instantaneous geometric parameters associated with the structure of crystalline sodium hydrogen bis(sulfate), which is a benchmark system with an extremely short hydrogen bond. We analyzed the trajectory obtained by a conventional Car–Parrinello molecular dynamics simulation, followed by an *a posteriori* quantization of the proton motion. For statistical analysis we applied the established methodologies of multiple linear regression, principal component analysis, principal component regression, and Kohonen neural networks. No simple correlation scheme between the OH stretching frequency and any particular geometry parameter (or their combination) was found. In comparison to the established correlation schemes (e.g., Mikenda and Novak) that consider a *series of systems*, our study provides a complementary insight into the nature of hydrogen bonding of a *single system*, in the sense that it considers the important aspects of fluctuations of the environment and the resulting broadening of the OH stretching band, which cannot be adequately assessed by experiment. The absence of appreciable correlations gives strong evidence of the extreme complexity of short hydrogen bonding.



1. INTRODUCTION

Empirical correlations between various observables of hydrogen bonds (HBs) have been for long a popular tool that facilitated the characterization of this phenomenon. Providing a reliable and straightforward connection between the geometry and spectroscopic observables, the correlations represent valuable assistance for the evaluation of characteristics which are otherwise difficult to determine experimentally. An important benefit of these correlations is also their ability to outline trends related to the HB geometry which are far from being self-evident.

Almost 60 years ago, Rundle and Parasol¹ as well as Lord and Merrifield² proposed the existence of a relationship between the OH stretching frequency (ν_{OH}) and geometry of the O—H···O moiety for the strong HBs in crystalline state, and since that time a remarkable number of correlations based on experimental data^{3–13} and/or data derived from calculations^{14–16} have been reported. However, it should also be pointed out that no significantly new information emerged from the earlier correlation curves, because of the limitation of the available crystallographic and spectroscopic data.¹¹ Possibly the most well-known HB correlation schemes are those of Mikenda^{3,7,8} and Novak,¹¹ linking a number of HB geometry features of different chemical systems between themselves and vibrational properties of HB. A well-known direct application of the Mikenda relation for O—H···O hydrogen bonds is the

ultrafast vibrational dynamics study of bulk water by Bratos and co-workers, giving insight into real-time dynamics of the O···O motion.⁹ Further, in 1995 Lutz et al. published the corrected O—H(D) H bond lengths by improved experimental data via additivity of the bond valence of the intramolecular O—H(D) and intermolecular H(D)···O bonds in the materials with short HBs.⁴ In 1999 Libowitzky also found out correlation between ν_{OH} (from IR spectra) and O···O and H···O bond lengths (from structural data) of minerals, but this scheme does not satisfactorily evaluate distance–frequency correlations for very strong HBs.⁵

Despite their general usefulness, the reliability of correlations in the range of shortest HBs can be, at least to a certain part, disputed. First, a short HB is an extremely complicated interaction, which is reflected in dramatic and abrupt changes of many spectroscopic observables on shortening of the donor–acceptor separation. It is questionable if a simple two- or three-parameter correlation scheme can faithfully account for a number of large-scale and complex effects. For HBs with the O···O distance below 2.50 Å (hereafter we will restrict ourselves to the HBs of the O—H···O type, but our consideration can be generalized to other types of HB), the

Received: December 22, 2011

Revised: May 16, 2012

Published: May 17, 2012

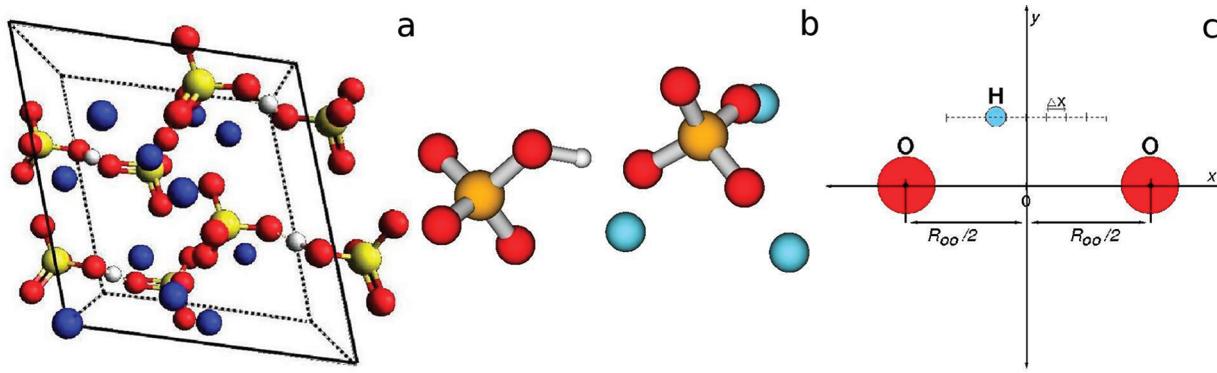


Figure 1. (a) Sodium hydrogen bis(sulfate) $[\text{Na}_3\text{H}(\text{SO}_4)_2]$ unit cell (atom colors are blue, sodium; red, oxygen; yellow, sulfur; white, hydrogen; hydrogen bonds are sketched with white dashed line) with corresponding asymmetric unit (b) and $\text{O}-\text{H}\cdots\text{O}$ moiety with selected stretching coordinate x (c).

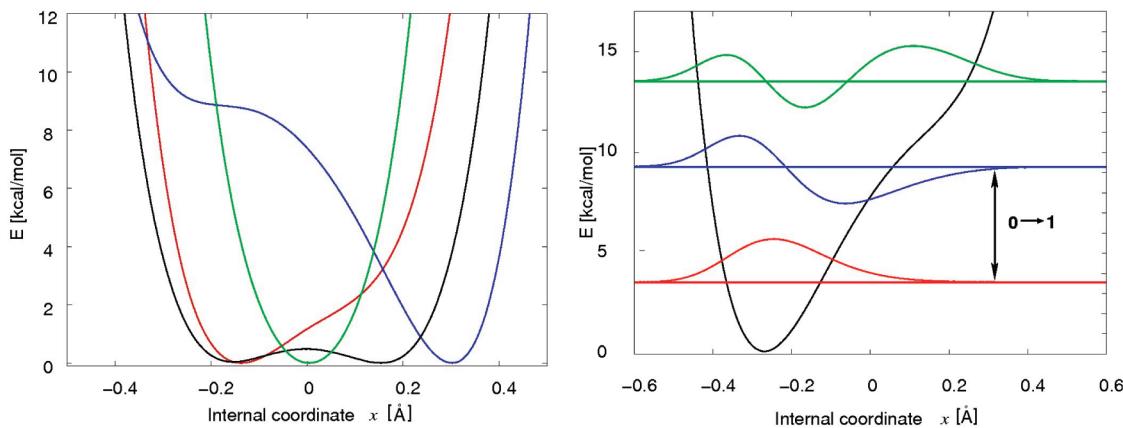


Figure 2. Selected proton potentials (left), extracted from snapshots of the CPMD simulation; for the sake of clarity only four of the 540 proton potentials are displayed; sample anharmonic proton potential together with the three lowest vibrational levels and wave functions obtained by solving variational Schrödinger equation (right). The fundamental $0 \rightarrow 1$ transition for hydrogen is indicated.

OH stretching mode can be red-shifted to lower frequencies by as much as 2000 cm^{-1} on nearly marginal decrease of the $\text{O}\cdots\text{O}$ distance. Furthermore, for shortest HBs, the effects of anharmonicity and quantum character of nuclear degrees of freedom become significant, and the coupling of the proton motion to other vibrational modes becomes extremely complicated. The stretching bands of short HBs are in many cases very broad, typically spanning a range of hundreds of wavenumbers; depicting such bands with a single parameter—usually the center of the band—is in our opinion far from being sufficient. A number of theoretical models at various levels have been proposed in order to account for the formation of strong hydrogen bonds and for band shaping,^{17–24} but many essential aspects still remain poorly understood. Second, as far as experimental vibrational spectroscopy of very short HBs is concerned, even a qualitative identification of the stretching band may be a very challenging task, let alone the precise assignment of its center, due to the fact that its contour is most often topped with other bands, Evans transmissions, etc., resulting in quite a large uncertainty in the data set on which the correlations are based. Importantly, correlation plots can exhibit quite large offsets of individual database entries from the fitted functional form in the range of shortest hydrogen bonds.²⁵ Third, while systems with a “moderate” HB with relatively large donor–acceptor separation exist in abundance, the ones with extremely short HB are relatively rare. Thus, correlation schemes feature poorer statistics in the region of the

shortest HBs due to the small data set. For all these reasons, caution is needed when attempting to devise certain characteristics of very short HBs on the basis of established correlations. Additionally, most (but not all, see, for instance, ref 26) of the correlations of the ν_{OH} and geometry parameters that have been studied so far are exclusively focused on geometry of the HB moiety and do not consider the environment (i.e., structure in the vicinity of a HB). Thus, several interesting questions about influences of the environment on the $\text{O}-\text{H}\cdots\text{O}$ moiety (e.g., related to the observed frequency shifts) are still unanswered, and finding relations between observables of HBs that cover geometry parameters beyond the most fundamental ones remains a challenge for HB research.

In the context of present work, the crucial feature of the popular correlation schemes is that they represent the relations between HB observables, *established among a number of different chemical systems*. Thus, each system represents one “point” in the correlation data set. A potential problem of such schemes originates from the fact that the real systems which are complex multidimensional objects in terms of structural parameters and spectroscopic quantities, are eventually mapped to a subspace of very low dimensionality (two in most cases). Importantly, much less is known of correlations that possibly exist *within a single chemical system, at the level of geometry fluctuations that give rise to band shaping* (yet, some studies of that kind have been published in the past^{15,20}). Namely, it is a fact that the structure is a dynamically fluctuating quantity, and the spectroscopic

observables of HBs are crucially governed by the dynamics of the fluctuating environment of the H-bonded moiety. It is of our prime interest to explore the relation between the dynamically fluctuating structure of the H-bonded system and the corresponding contribution to the broad absorption band of the OH stretching motion. In other words, while our work considers the influence of fluctuating environment on band shaping, so that both the structure and the band are represented by a number of parameters, the established correlation schemes rely on the single-value representation of the band, namely its frequency.

This paper extends our previous studies²⁵ on the assimilation of the anharmonic ν_{OH} to the corresponding geometry parameters of the snapshot structures of the sodium hydrogen bis(sulfate) ($\text{Na}_3\text{H}(\text{SO}_4)_2$, Figure 1), which is a benchmark example of materials that exhibit both superprotic behavior^{27–32} and short HB. We developed the so-called snapshot methodology based on the sampling of the instantaneous potential energy function for the hydrogen stretching motion, as influenced by the fluctuating environment.^{25,33,34} The prerequisite for our methodology is a molecular dynamics (MD) simulation (by Car–Parrinello approach in the present case, but in principle any type of MD can be used). From the MD trajectory we extracted the snapshot conformations (approximately every 250 fs in order to prevent them from being mutually correlated). Assuming that these snapshot configurations represent a faithful model of the fluctuating environment of the HB, we calculated for each snapshot the proton potential function for the hydrogen stretching motion by stepwise displacing the hydrogen atom along the O···O line while keeping all other atoms fixed (Figure 2). Next, we solved the one-dimensional vibrational Schrödinger equation for each of the snapshot potentials³³ (Figure 2), yielding a set of vibrational transitions distributed over quite a wide frequency range (Figure 3). This distribution of frequencies, also called

biomolecules, featuring both short and moderately long HBs.^{25,33,34} Despite certain simplifications undertaken in our approach, the results are encouraging, since the agreement between the calculated and observed stretching envelopes is fairly good in all cases; also, the time-averaged geometry of the system is in fair agreement with the neutron diffraction structure, in which the proton is located nearly at the midpoint of the O···O line.²⁵

Interestingly, our first attempts at correlating the anharmonic ν_{OH} to the corresponding geometry parameters (e.g., O···O or S–O distance) of the snapshot structures of sodium hydrogen bis(sulfate) were unsuccessful; no appreciable correlation was found, which is obvious even from visual investigation of the plots (Figure 4). The absence of correlation between ν_{OH} and the most representative geometric parameters of the HB suggests that the relation between the instantaneous geometry and the pertinent value of ν_{OH} is rather complicated and requires a more profound analysis.

In order to obtain better insight into the relationship between ν_{OH} and the corresponding geometry parameters of the snapshot structures of the examined compound, we applied chemometrics methods, combination of the mathematical and statistical methods, which design or select optimal measurement procedures and experiments in order to provide maximum chemical information (knowledge of chemical systems and processes) by analyzing chemical data. When considering only two variables, such as ν_{OH} and O···O distances obtained from snapshot structures (Figure 4a,b), it is easy to plot this data and to obtain the correlation between those two quantities. But when hundreds of variables are implemented into analysis, it becomes impossible to make a visual aspect to the relationship between variables in such multidimensional matrix. Therefore, one necessary part of the statistical analysis in such applications is its dimension reduction. For this purpose, several data decomposition techniques are available. We applied principal component analysis (PCA) which is recommended as the most appropriate (processing multidimensional data into two dimensions) and exploratory tool to uncover unknown trends in the data.³⁵ Additionally, we used one among different neural networks, the most suitable type of artificial neural network for clustering³⁶—Kohonen neural network self-organizing maps,³⁷ which typically produce via some nonlinear transformation of data, a two-dimensional data set, called Kohonen map. While several authors^{2,6,14,38} suggest approximately linear relationship between ν_{OH} and bond distances, such as O–H, H···O, and O···O, we were further interested in techniques for modeling and analyzing several variables, when the focus was on the relationship between a dependent variable (ν_{OH}) and one or more independent variables (geometry parameters). For this purpose, we applied calibration techniques: multiple linear regression (MLR)^{35,39} and principal component regression (PCR).³⁹ All these techniques differ in their applicability. A more comprehensive review on these various techniques can be found elsewhere (e.g., in ref 35).

The present study is to our best knowledge the first study of the correlations between the OH stretching frequency and a complete set of instantaneous geometry parameters—in addition to the O···O distance also the other bond distances, bond angles, and bond dihedral angles in the strong H-bonded crystal. In other studies,^{2,3,5,8,9,11} authors used experimental data that are by definition ensemble averages on different H-bonded crystals to establish distance–frequency correlations for

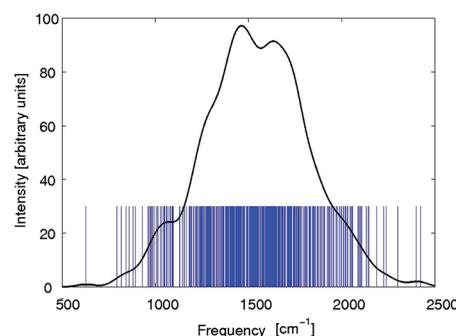


Figure 3. Distribution of anharmonic OH stretching transitions (blue vertical lines) obtained from proton potentials extracted from snapshots of the CPMD simulation. A continuous representation of this distribution (black curve) is modeled as superimposed Gaussian functions (one for each transition) with a half-width of 50 cm^{-1} .

the envelope of the OH stretching band, can be directly compared with the observed vibrational spectra; however, it should be noted that the factors which govern the intensity of transitions (e.g., the dipole moment function or polarizability, associated with infrared and Raman spectroscopy, respectively) have not been considered in our study for the sake of simplicity; all transitions that contribute to the envelope displayed in Figure 3 bear the same weight. We tested our approach on various systems, from molecular solids to aqueous

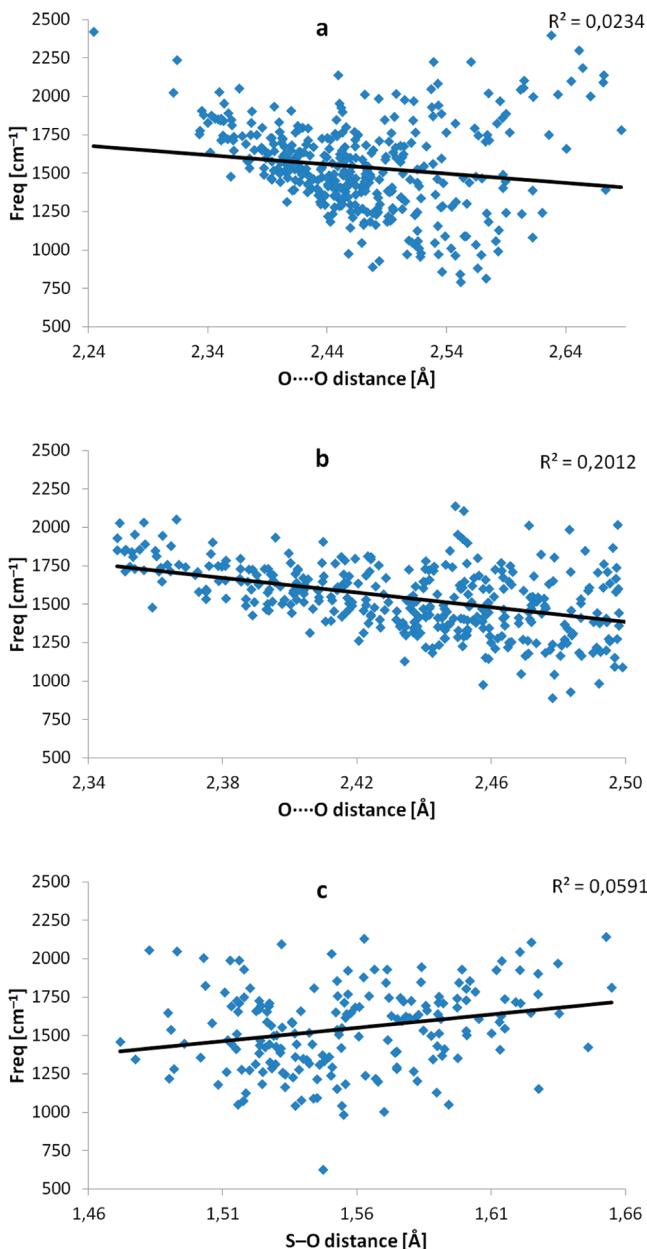


Figure 4. Simple two-dimensional plots used to obtain preliminary qualitative information about the relationship between anharmonic O—H stretching frequency and selected geometry parameters: (a) ν_{OH} vs O···O distance range, full distance range; (b) ν_{OH} vs O···O distance range, reduced distance range; (c) ν_{OH} vs S···O distance. A linear regression line with the corresponding R^2 value is displayed with each plot. From all three data plots (a–c), it is evident that we cannot devise appreciable correlation between quantities (even if the original range of O···O distance (a) was reduced (b)).

O—H···O and/or N—H···N (O) bonds. Nevertheless, it should be stressed that correlation schemes within sampled images of a single system have been used in the past. For instance, Kühn and co-workers used a closely related approach to study the nature of hydrogen bonding in adenine–uracil base pair in solution.¹⁵ Mukamel and co-workers used dynamically sampled snapshot structures for the simulation of multidimensional infrared spectra of biological systems.⁴⁰ In their study of IR and Raman band shapes of HOD in the aqueous phase, Corcelli and Skinner have utilized DFT-established correlations between solvent electric field and

frequency, transition dipole and transition polarizability to compute the IR and Raman contours of the OH and OD stretching mode from a classical MD.²⁶

It is worth noting that, as our approach is notably different even at the level of data set, the findings presented in this work are not intended to dispute the validity of the established correlations between the HB geometry and spectroscopic observables, but rather to provide a new insight into HB and demonstrate the complexity of its dynamics.

2. COMPUTATION

The present work is a continuation of our previous work²⁵ and it analyzes the dynamics of proton motion in the crystalline sodium hydrogen bis(sulfate) (Figure 1) by applying chemometrics methods on the snapshot structures along the trajectory. For the CPMD simulation details, extraction of snapshot geometries, and calculation of instantaneous proton potentials and anharmonic frequencies, the reader is referred to ref 25. It is worth stressing that we collected 135 snapshot structures and since each unit cell consists of four independent H-bonded $\text{Na}_3\text{H}(\text{SO}_4)_2$ units, we basically finished with a set of 540 anharmonic frequencies and 135 geometries of the system that can be partitioned to the 540 geometries of the $\text{Na}_3\text{H}(\text{SO}_4)_2$ unit, one per frequency.

In order to study the relationship between all calculated physical quantities, which are used to describe and characterize the O—H···O moiety and the ν_{OH} , we extended analysis of the relationship between frequency and geometry parameters up to all internal coordinates present in the unit cell. For this purpose different chemometrics methods have been applied on the matrix composed of 135×266 (1), 135×158 (2), and 135×35 (3) rows and columns, respectively. 135 rows represent snapshot structures composed of 266, 158, and 35 variables (internal coordinates in columns), respectively. The elements of the particular columns are given as a quotient of the structural parameters (interatomic distances, bond angles, dihedral angles) of the unit cell. The number of variables characterizing the structure of the system can be rationalized in the following way. The unit cell contains four $\text{Na}_3\text{H}(\text{SO}_4)_2$ formula units, totaling 56 atoms. As the number of internal degrees of freedom for a system with N atoms is $3N - 6$, the number of independent internal coordinates that depict the content of the unit cell is 162. However, as one internal coordinate of each H-bonded moiety (the stretching coordinate x , as depicted in ref 25; see also Figure 1) served as a “probe” coordinate along which the proton was displaced, we omitted it from our analysis, resulting in 158 internal coordinates of the matrix labeled (2). Note that we constructed the internal coordinate set in such a way to follow the chemical structure of the system; while this is trivial for the hydrogen bis(sulfate) anion, the three pertinent sodium ions were assigned to each hydrogen bis(sulfate) unit on the basis of minimal distances considering periodic boundary conditions when necessary. Next, since partitioning of the 12 sodium ions to the corresponding H-bonded moieties (three to each) can hardly be done unambiguously, we additionally represented the unit cell structure by considering the entire set of sodium ions to be geometrically related to each H-bonded moiety. This means that the original set of 158 internal coordinates was expanded by counting hydrogen atom and 9 excessive remote sodium ions for each H-bonded moiety, resulting in 36 redundant atoms or 108 coordinates, yielding a matrix of 266 internal coordinates (matrix (1)). Ultimately, in order to simplify the

correlation problem, we considered each individual $\text{Na}_3\text{H}(\text{SO}_4)_2$ formula unit separately, yielding 35 internal coordinates (matrix (3)).

Exploratory data analysis is often guided by additional categorical information on the data. In our case, this additional information was handled by assigning class numbers to the range of the stretching frequencies (class attributes 1, $\nu_{\text{OH}} < 1200 \text{ cm}^{-1}$; 2, $1200 \text{ cm}^{-1} \leq \nu_{\text{OH}} > 1500 \text{ cm}^{-1}$; 3, $1500 \text{ cm}^{-1} \leq \nu_{\text{OH}} > 1800 \text{ cm}^{-1}$; and 4, $\nu_{\text{OH}} \geq 1800 \text{ cm}^{-1}$). PCA and Kohonen neural network were applied for grouping the ν_{OH} due to calculated physical quantities, while two regression methods, MLR and PCR, were mainly utilized to give an additional insight into the relationship between several independent or predictor variables and a dependent criterion variable, frequencies.

All the calculations and preparation of plots in the following section were done with the Teach/me software⁴¹ using Teach/me Data Analysis application.

3. RESULTS AND DISCUSSION

Review of the Calculated Snapshot Anharmonic Frequencies and the OH Stretching Envelope. In agreement with the diverse snapshot geometries and proton potential functions, the resulting fundamental excitation frequencies for the hydrogen stretching motion, obtained via time-independent solving of the vibrational Schrödinger

Table 1. Comparison of R^2 Value (Variation That Is Predicted by the Independent Variable) in Multiple Linear Regression (MLR) Using First Five Predictors^a

model	predictors	F-part	R^2
1 ^b	dih($\text{S}^7\text{O}^{43}\text{O}^{30}\text{S}^2$)	4.26	0.0327
2 ^b	$a(\text{O}^{44}, \text{T}^{53}, \text{Na}^{10})$	3.71	0.063
3 ^b	$R(\text{S}^6-\text{O}^{42})$	3.54	0.0888
4 ^b	$a(\text{O}^{43}, \text{O}^{30}, \text{S}^2)$	3.42	0.1130
5 ^b	dih($\text{S}^8\text{O}^{44}\text{O}^{29}\text{S}^1$)	5.40	0.1359
1 ^c	$a(\text{O}^{35}, \text{S}^3, \text{O}^{27})$	5.17	0.0573
2 ^c	$R(\text{S}^8-\text{O}^{48})$	4.89	0.0928
3 ^c	dih($\text{O}^{23}, \text{O}^{35}, \text{S}^3, \text{O}^{27}$)	6.15	0.0125
4 ^c	$R(\text{S}^2-\text{O}^{22})$	6.52	0.1649
5 ^c	dih($\text{O}^{52}, \text{S}^8, \text{O}^{40}, \text{O}^{44}$)	4.00	0.2051
1 ^d	$R(\text{S}^3-\text{O}^{31})$	4.79	0.1189
2 ^d	dih($\text{Na}^{17}, \text{O}^{42}, \text{T}^3, \text{O}^{31}$)	4.60	0.1498
3 ^d	$R(\text{Na}^{17}-\text{T}^3)$	3.92	0.1789
4 ^d	dih($\text{Na}^{18}, \text{O}^{41}, \text{T}^4, \text{O}^{32}$)	3.75	0.2026
5 ^d	$R(\text{O}^{31}-\text{O}^{42})$	3.71	0.2251
1 ^e	$R(\text{S}^4-\text{O}^{32})$	6.43	0.051
2 ^e	$R(\text{O}^{32}-\text{O}^{41})$	4.28	0.095
3 ^e	$R(\text{Na}^{16}-\text{T}^3)$	5.42	0.123
4 ^e	dih($\text{S}^7, \text{O}^{43}, \text{O}^{30}, \text{S}^2$)	6.51	0.158
5 ^e	dih($\text{Na}^{10}, \text{O}^{43}, \text{T}^1, \text{O}^{29}$)	5.03	0.199

^aPredictors are marked in Figure 5. For atom numbering, see Figure 5. Legend: T , point calculated as midpoint between corresponding bridged oxygen (O) atoms. R , distance between corresponding atoms. dih, dihedral angle, the angle between two planes in a dihedron. a , bond angle, the angle that is formed between two adjacent bonds on the same atom. ^bData matrix 135×158 (targets are frequencies obtained from first asymmetric unit-I). ^cData matrix 135×158 (targets are frequencies obtained from second asymmetric unit-II). ^dData matrix 135×158 (targets are frequencies obtained from third asymmetric unit-III). ^eData matrix 135×158 (targets are frequencies obtained from fourth asymmetric unit-IV).

equation, span a wide range from 600 to 2500 cm^{-1} . This distribution of frequencies represents the band shape of the OH stretching mode due to coupling to the fluctuating environment of the H-bond (Figure 3).²⁰ The band resembles Gaussian shape and its center of gravity is at 1540 cm^{-1} . The calculated band is in reasonable qualitative agreement with the experimentally determined broad absorption in the infrared spectrum of the examined compound.^{30,42} Yet, according to Novak's correlation scheme, the calculated center of the OH stretching band corresponds to an $\text{O}\cdots\text{O}$ distance of about 2.5 \AA ,¹¹ which is notably (by more than 0.05 \AA) longer than the diffraction-determined or calculated $\text{O}\cdots\text{O}$ separation; thus, according to Novak, the present approach appears to overestimate the frequency of the OH stretching band. Indeed, in comparison to the experimental vibrational spectra of the title compound and its close analogues,^{30,42,43} our computed band appears to be slightly shifted to higher frequencies, but at the qualitative level our approach is capable of reproducing the extreme breadth of the band. Due to the complex shape of the experimental spectra—the contour of the OH stretching band is topped by other modes, cut by Evans transmissions, etc.—it is hard to devise a more quantitative comparison.

Preliminary Analysis of Data (Frequency as a Function of Geometric Parameters). In the first stage, frequencies ν_{OH} as function of geometric parameters were calculated and visualized as simple two-dimensional (2D) linear regression schemes between ν_{OH} and $\text{O}\cdots\text{O}$ or $\text{S}\cdots\text{O}(\text{H})$ distance (Figure 4). No significant correlation was observed; even in the best case when a limited range of the $\text{O}\cdots\text{O}$ distance was considered in regression, the R^2 value was just above 0.2 (see Figure 4b). Evidently, the simple correlation scheme between ν_{OH} and the $\text{O}\cdots\text{O}$ distance (Figure 4a,b), as established by Novak¹¹ and later extended by Mikenda,³ appears not to be applicable in the present case; this is in accordance with our previously published experience²⁵ and with the study of Ratajczak and Orville Thomas.⁴⁴ In the latter study, a good correlation was reported for medium-strong hydrogen bonds and the correlation breaks down for very strong hydrogen bonds. Nevertheless, it needs to be stressed that, in contrast to the latter study, our approach searches for correlation within various dynamically sampled structures of the same system and its instantaneous spectroscopic quantities (rather than doing so within a set of different systems), and is thus quite different at the very origin.

As attempts at finding simple correlation between ν_{OH} and basic geometry parameters were not successful, we extended analysis of the relationship between frequency and geometry parameters to all internal coordinates present in unit cell (or to a selected subset of coordinates). The decision to proceed with the entire cell is supported by the physical nature of the system. The main type of interaction in ionic crystals is electrostatics, which is essentially of long-range character, and therefore proton dynamics is influenced by surroundings that extend beyond the closest atoms. The aim of advanced statistical analysis was a construction of the correlation between the calculated anharmonic frequency and the internal coordinates that define the corresponding instantaneous structure of the system. We acquired 135 snapshot structures and, since each unit cell consists of four independent $\text{O}-\text{H}\cdots\text{O}$ moieties, we obtained 540 frequencies that depend on the internal coordinates. As outlined in the Computation section, we formulated the internal coordinate space by considering 35, 158, or 266 coordinates, differing in the completeness of the

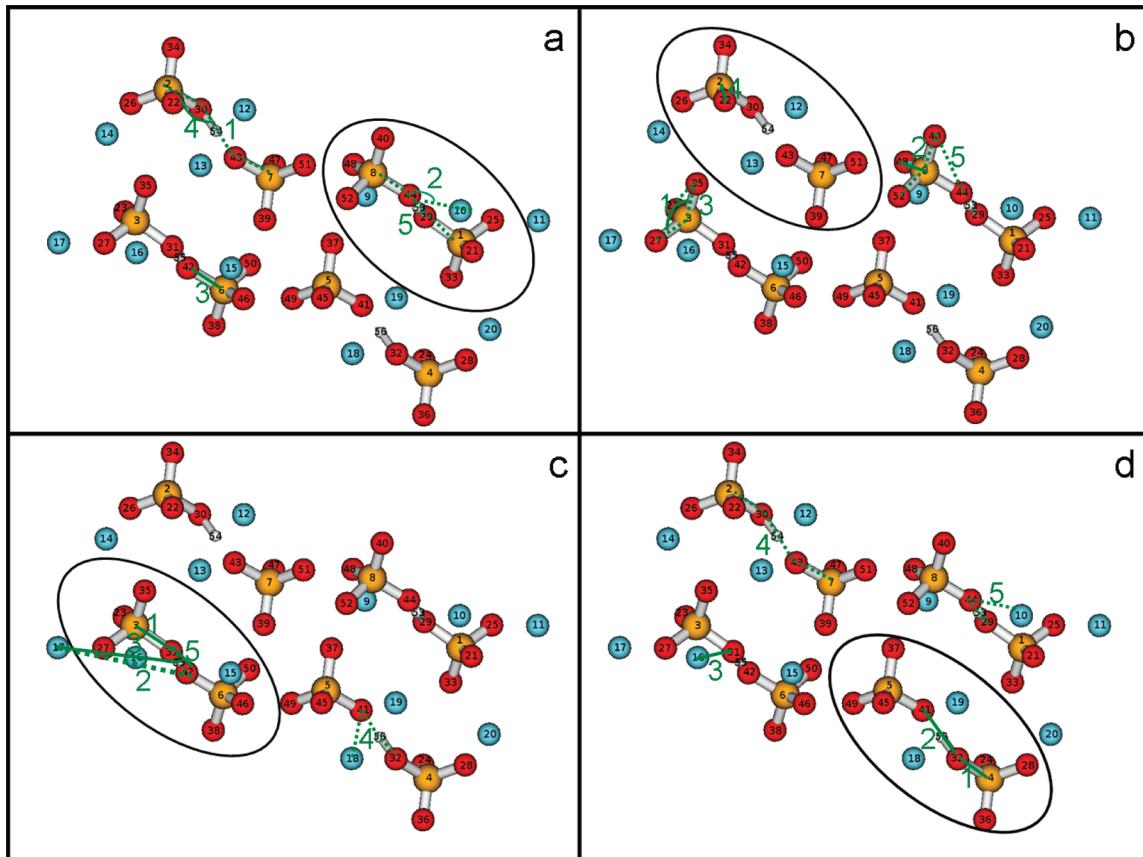


Figure 5. First five explanatory variables (numbered 1–5, in green, where 1 is the best predictor), which are the most associated with the dependent variable of interest: OH stretching frequencies for all four H-bonded moieties, denoted by the circled area (a–d). The same predictors (e.g., as first we expected O···O distance of the corresponding H-bonded moiety) of each H-bonded moiety were expected, but as we can see from this figure there is no similarity (see also Table 1). Also in all cases some of the most influential geometry parameters for the given H-bonded moiety are eventually located outside that moiety, confirming the assumption of weak and inconclusive correlation.

coordinate set and inclusion of redundant coordinates of sodium ions. We constructed the corresponding 2D data matrices and performed various analyses by the Teach/me program. We adopted the assumption that the distribution of variables follows the normal distribution.

Multiple Linear Regressions (MLR). To learn more about relationship between several independent criteria and one dependent criterion (ν_{OH}) variable, we applied the MLR technique on all matrices. We were particularly interested in whether the O···O distance is a better predictor of ν_{OH} than other geometry parameters. First, we applied the MLR technique for all four O—H···O moieties in our unit cell. We expected that the results would be similar in all four O—H···O moieties (e.g., the O···O distance would be the priority quantity). Whereas the so obtained results differed, we also applied the MLR technique on matrix (2) and the results are listed in Table 1. One way to measure the overall predictive accuracy of multiple regression models is the R^2 value, a measure of how much variation a model explains. From the results listed in Table 1 we can conclude that the model is quite reasonable but still does not provide much information about correlation between geometry parameters and ν_{OH} . Surprisingly, the results for the four hydrogen-bonded moieties are not similar. From Figure 5 and the results in Table 1 (models c and d) it is suggested that for two O—H···O moieties the S—O distances are the best predictors of the ν_{OH} , while for other two moieties (Table 1, models a and b) these coordinates appear

not to be the best predictors. Summarizing, we can conclude that in our model only very weak (if any) linear correlation between geometry parameters and ν_{OH} exists.

Principal Component Analysis (PCA). PCA was performed in order to get an overall impression about the correlation of 540 values of ν_{OH} with the corresponding geometry parameters of the snapshot structures. PCA was applied for each H-bonded moiety on the matrices composed of 266 (1), 158 (2), and 35 (3) internal coordinates, respectively. The data was additionally preprocessed in two different ways: first by using mean centered (mc) data, which means that the mean value was subtracted from individual elements, and second by using standardized (st) data which are extracted by scaling the matrix (columnwise) to zero mean and unit variance.³⁵

PCA finds eigenvectors (principal components, PCs) and eigenvalues (variances) relevant to the data by using a covariance matrix (a centering application in Teach/me software was applied) and correlation matrix (the autoscaled application in program was used for this purpose). Since our preliminary studies suggested weak correlations at best, we monitored a large set of 20 PCs rather than two or three, as common with strong correlations. We were interested in how reducing the number of variables influences the variance of the first 20 PC axes. The results are listed in Table 2, and from these data we can conclude that the number of considered internal coordinates has a pronounced influence on the PCs.

Table 2. Comparison of Variances in Principal Component Analysis (PCA) Using Different Sizes of Matrix and Two Different Scaling Procedures: Mean Centering (mc, with $m = 0.0$) and Autoscaling (st; Scale to Zero Mean and Unit Variance) of Data^a

PC	135 × 266 (1)		135 × 158 (2)		135 × 35 (3 ¹)		135 × 35 (3 ²)		135 × 35 (3 ³)		135 × 35 (3 ⁴)	
	mc	st	mc	st	mc	st	mc	st	mc	st	mc	st
1	23.40	13.50	13.51	6.12	50.18	15.15	49.05	15.94	38.79	16.90	45.51	15.37
2	16.90	11.82	13.14	5.37	28.74	9.81	30.27	11.09	32.01	9.90	34.97	10.12
3	14.15	9.82	11.02	4.80	16.62	7.85	16.98	8.48	20.78	8.72	12.68	8.43
4	11.78	7.75	10.29	4.14	3.22	7.35	2.34	6.48	6.94	6.58	5.39	7.01
5	10.30	3.70	8.89	3.61	0.50	5.87	0.52	5.90	0.66	6.19	0.54	6.62
6	4.55	3.16	8.43	3.44	0.16	4.89	0.15	4.80	0.26	5.08	0.29	5.27
7	3.04	2.60	7.48	3.14	0.14	4.67	0.13	4.40	0.12	4.19	0.13	4.51
8	2.10	2.24	6.36	2.84	0.11	4.33	0.11	3.89	0.09	3.87	0.11	4.12
9	1.67	2.16	4.96	2.67	0.07	3.65	0.08	3.59	0.07	3.22	0.08	3.92
10	1.42	2.03	4.39	2.55	0.05	3.52	0.08	3.43	0.05	3.01	0.06	3.29
11	1.35	1.84	3.31	2.39	0.04	3.05	0.06	3.22	0.05	2.87	0.05	3.01
12	1.09	1.70	2.53	2.27	0.03	2.83	0.05	2.97	0.03	2.84	0.04	2.80
13	0.93	1.58	1.74	2.07	0.03	2.80	0.03	2.76	0.03	2.69	0.03	2.68
14	0.80	1.53	1.29	2.02	0.03	2.51	0.03	2.42	0.02	2.47	0.02	2.59
15	0.61	1.46	0.84	1.85	0.02	2.30	0.03	2.23	0.02	2.34	0.02	2.20
16	0.52	1.32	0.50	1.81	0.02	2.22	0.02	2.07	0.02	2.21	0.02	2.02
17	0.44	1.28	0.19	1.75	0.01	2.01	0.01	1.95	0.01	2.14	0.02	1.84
18	0.41	1.18	0.17	1.74	0.01	1.75	0.01	1.71	0.01	1.88	0.02	1.77
19	0.34	1.11	0.11	1.64	0.01	1.70	0.01	1.53	0.01	1.79	0.01	1.64
20	0.29	1.05	0.10	1.51	0.01	1.61	0.01	1.48	0.01	1.71	0.01	1.47

^aVariance is given as percentage of the explained variance for each principal component (PC). Each PC is a linear combination of the original presentation of the system. PCs vectors are orthogonal. For the smallest matrix (3) results for each of the four hydrogen-bonded moieties are given separately.

With the most reduced matrix of internal coordinates (3) and by using autoscaled data, only 36% of variance is gathered in the first three PCs. On the other hand, in the first three PCs of transformed mean-centered data, as much as 92% of variance was explained. The reason is in a dominant role of a variable of considerably larger values (dihedral angles), because the mean centering does not scale the data to equal units. Applying PC analysis directly to irregularly spaced data makes the variables incomparable (e.g., the distances are in the approximate range between 1 and 10 Å, while the values of bond angles and dihedral angles are in the range between 0 and 360°). Transformed mean-centered data do not scale the data to comparable values; it only moves average to zero. This can reflect in that the first PC contains the majority of variance and consequently the data transformed in such a way do not cluster well, because the rest of the variables do not contribute enough to the first three PC axes. By increasing number of variables and consequently changing the content between incomparable variables, the content of variance on the first PCs decreases in the other two matrices. The problem of incomparable values of data can be overcome by interpolating data to comparable values by applying the autoscaling method. When preprocessing data by autoscaling method, the percentage of the first three variances in the resulting eigenvectors (PCs) in all matrices becomes comparable. For this reason and for the sake of brevity, we focused on presentation of autoscaled data from matrices.

The majority of the information is usually gathered in the first few PCs; in our case, only mc data in matrix (3) exhibits such behavior. In other cases, the majority of the information is contained in a much larger number of PCs. We explored a number of pairwise plots of PCs, but because of similarity of the plots only a PC1/PC2 plot of the matrix (1) is displayed (Figure 6). Unfortunately, the score versus score plot (Figure

6a) was compared and it was found that clusters were not formed. Consequently we could not identify the most influencing geometry parameter, which is in concordance with our previous study (Figure 6b). Similar results were obtained from the other two matrices.

These results led us to suspect that a more pronounced correlation between ν_{OH} and geometry might exist in individual subsets of geometric parameters (bond distances, bond angles, dihedral angles). Thus, we tried to establish the correlations separately in each subset of variables and then we also tried to examine the data with some exploratory multivariate techniques (e.g., cluster analysis). Surprisingly, there was also no correlation. In the next step we suspected that outliers (atypical infrequent observations) might influence the obtained results. As it is known from statistics, outliers have a profound influence on the slope of the regression line and consequently on the value of the correlation coefficient. To avoid such atypical observations, we first examined whether the data in columns approximately followed normal distribution and, second, we also used data without extreme values, limiting it within the $\pm 2s$ range. Again, the so-obtained results did not yield any appreciable correlation and consequently we found it not to be practical to perform such data filtering in further analysis. At the bottom line, only weak correlation between ν_{OH} and geometry parameters of the snapshot structures could be deduced by the PCA methodology.

Principal Component Regression (PCR). Since several authors^{2,4,11,20} report on the approximately linear relationship of geometrical parameters of the hydrogen bond and the ν_{OH} even below 2500 cm⁻¹, the challenge was to examine the applicability of multidimensional linear regression techniques in the present study. Although we could not prove causal relations based on correlation coefficients, we still tried to identify correlations that are due mostly to the influences of other

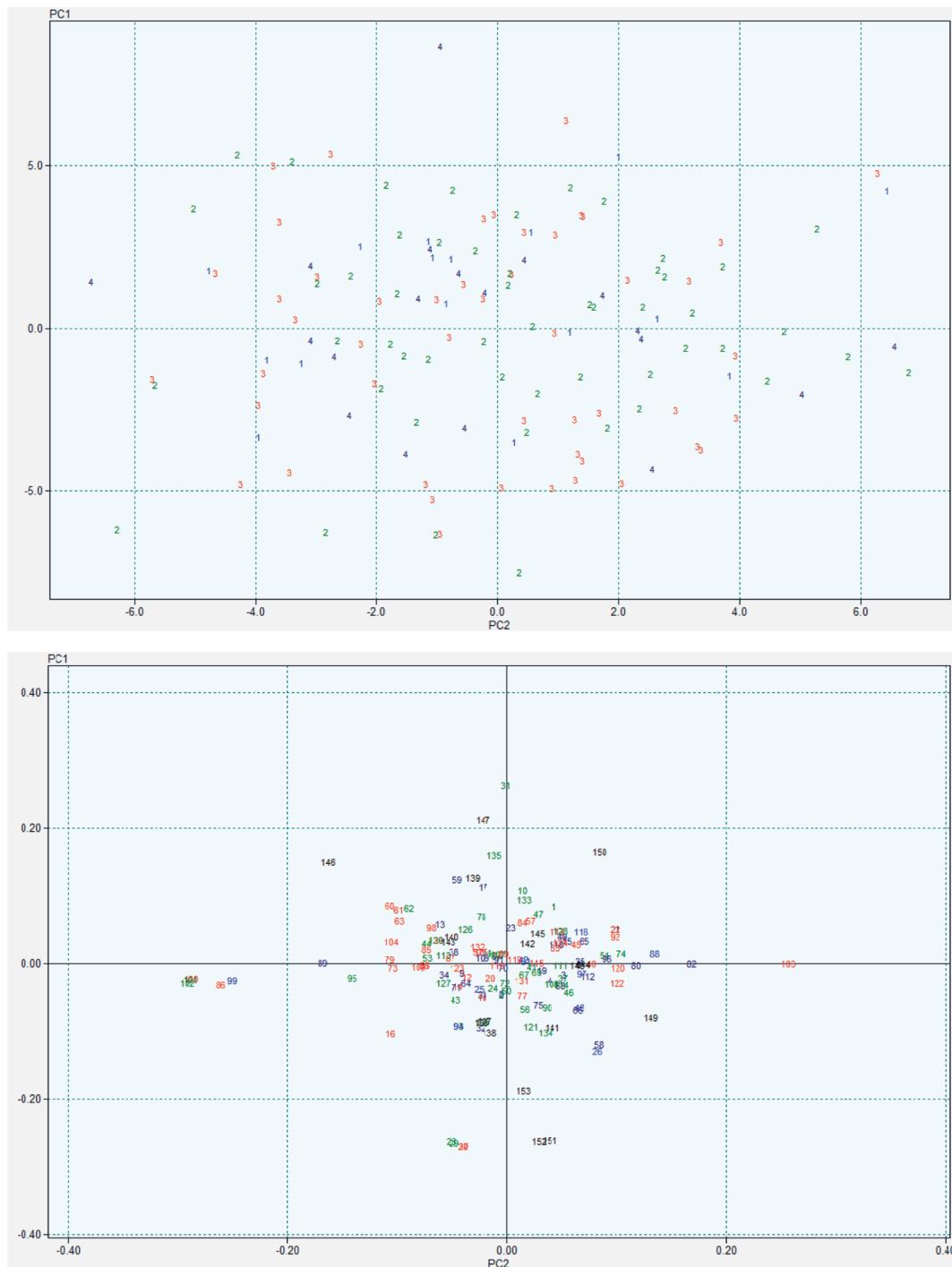


Figure 6. A large data set was projected onto two-dimensional subspace, while the variance of projected data were maximized. After vector space transformation, we obtained a more efficient description of our system. The first dimension captures the maximum variance, while the second dimension captures the maximum amount of residual variance. Plot of the first two principal components (PCs) resulting from principal component analysis (PCA) of 135 snapshot structures represented as 158 variables (geometry parameters), and it is displayed as (a) scores PC1/PC2 and as (b) loadings (vectors) against each other. Each variable is numbered (1–158) and classes are associated (see section 2) with different colors. Plot (a) shows no clusters and PC1 and PC2 do not explain geometry parameters (a).

variables—in other words, to predict a response variable from a large number of variables. For this purpose, we used PCR, which allows regression of a variable against the principal components of the rest of the variables. The same matrices as used in PCA were applied and for each H-bonded moiety an extra column with our calculated values was added; we set ν_{OH} as the target variable. Thereafter, the PCs of all other variables

were calculated. In order to obtain information about the relationship between structural parameters and ν_{OH} , and to reduce the number of variables, the regression was established by limiting the number of PCs (n PCs); this number is determined arbitrarily (usually 2 or 3), based on the trade-off between the quality of fit measured with the correlation coefficient of the regression model (increasing with the number

Table 3. Results Obtained from Principal Component Regression (PCR) Analysis Using Different Size of Matrix and the Most Appropriate Number of the Principal Components (PCs)^a

regression	no. of PCs	R ²	F-statistic
1 ^b	5	0.016	0.41
2 ^b	90	0.575	0.63
3 ^b	130	0.901	0.28
1 ^c	5	0.048	1.30
2 ^c	90	0.692	1.10
3 ^c	130	0.978	1.36
1 ^d	5	0.026	0.68
2 ^d	90	0.599	0.73
3 ^d	130	0.945	0.56
1 ^e	5	0.018	0.47
2 ^e	90	0.657	0.93
3 ^e	130	0.965	0.85
1 ^f	5	0.010	0.26
2 ^f	90	0.526	0.54
3 ^f	130	0.951	0.60
1 ^g	5	0.059	1.62
2 ^g	90	0.739	1.38
3 ^g	130	0.982	1.70
1 ^h	5	0.067	1.848
2 ^h	90	0.684	1.058
3 ^h	130	0.927	0.393
1 ⁱ	5	0.037	1.00
2 ⁱ	90	0.735	1.36
3 ⁱ	130	0.999	39.61
1 ^j	5	0.089	2.52
2 ^j	20	0.172	1.18
3 ^j	30	0.283	1.36
1 ^k	5	0.012	0.31
2 ^k	20	0.179	1.24
3 ^k	30	0.285	1.35
1 ^l	5	0.097	2.76
2 ^l	20	0.221	1.61
3 ^l	30	0.283	1.37
1 ^m	5	0.080	2.23
2 ^m	20	0.168	1.15
3 ^m	30	0.343	1.81

^aPlease note that PCR predicts a response variable from a large number of variables (takes enough eigenvectors to cover 80–90% of the variance). In normal cases the usual number of PCs is below 5. Such a large number of PCs indicates that in our case no linear correlation between the variables exists. ^bData matrix 135 × 266 (target is frequencies obtained from first asymmetric unit I). ^cData matrix 135 × 266 (target is frequencies obtained from second asymmetric unit II). ^dData matrix 135 × 266 (target is frequencies obtained from third asymmetric unit III). ^eData matrix 135 × 266 (target is frequencies obtained from fourth asymmetric unit IV). ^fData matrix 135 × 158 (target I). ^gData matrix 135 × 158 (target II). ^hData matrix 135 × 158 (target iii). ⁱData matrix 135 × 158 (target IV). ^jData matrix 135 × 35 (target I). ^kData matrix 135 × 35 (target II). ^lData matrix 135 × 35 (target III). ^mData matrix 135 × 35 (target IV).

of PCs) and the robustness assessed by the *F*-statistic (decreasing with the number of PCs). At first, we applied guidelines that are commonly used: (i) a graphical method⁴⁵ where we extracted the number of PCs by finding the place where there was a smooth decrease in the contribution of eigenvalues to the variance and (ii) by using the set of PCs whose results should satisfy the criteria that the *F*-statistic

should be greater than 1 and the quality of fit should be at least 0.75.⁴⁶ The suggested criterion breaks down in the present case. Hence, we made a stepwise selector of PCs (Table 3). The number of PCs required to explain satisfactorily the value of the variation is unreasonably large. Regardless of the large difference between the matrices, the univocal message is that no linear correlation exists between the snapshot geometry and ν_{OH} . It is otherwise known that PCs with the larger variances are not necessarily the best predictor and it could be also a predictor that there is no linear correlation at all (as we suggested in ref 25).

Kohonen Neural Network (Koh-NN). But is there in our system some nonlinear correlation? To answer this question, we applied a nonlinear mapping method Koh-NN. Two-dimensional neurons were arranged in a rectangular map (see Figure 7). Different Koh-NN architectures with 100, 121, 141, and 400 neurons (10×10 , 11×11 , 12×12 , and 20×20) were constructed. All were checked for the conflicts where two snapshot structures of different type fall into the same neuron. Several conflicts between snapshot structures of different type (different classes of frequencies, classes defined in section 2) were found, especially on the edges of the map (consequence of folding the grid). The results (Figure 7) are well separated (except the numbers marked in turquoise) and suitable for interpretation. Nevertheless, the Kohonen network architecture (Figure 7a) shows, similarly to PCA, that there is no formed cluster and that snapshot structures are not separated into four distinct groups according to predefined classes in section 2. We were also interested in formed clusters while instead of class number, id numbers were used (Figure 7b). What we can conclude from these results is that we have 135 different snapshot structures (and this was our first goal in our previous study,²⁵ to extract different structures from the trajectory) which do not form clusters. Considering all the obtained data, we can deduce that there is also no nonlinear correlation between geometry parameters and ν_{OH} .

All the above-discussed findings indicate that our model is quite complicated and does not exhibit an appreciable correlation between calculated geometry parameters and ν_{OH} . This might lead to a conclusion that, for systems with a short hydrogen bond, correlations between instantaneous fluctuating structure and the OH stretching frequency are likely to be poor or even nonexistent. The known features of short hydrogen bonding, particularly those related to the complexity of the proton motion and its coupling to the environment, support this assumption.

That said, possible pitfalls of the approach should be considered—namely, the fact that the electronic structure and vibrational frequency are a functional of the nuclear positions, thus some correlation should exist, provided that the right set of variables is used. In the context of the present study, two possible sources of error need to be emphasized. First, although we used probably the most reasonable set of internal coordinates, we limited ourselves to the content of a single unit cell. Thus, a notable portion of nuclear positions that give rise to the long-range electrostatic forces (mind that the system has a pronounced ionic character, thus electrostatics is highly relevant) were not included in the model. Second, we used a one-dimensional approximation of the quantized proton motion for the calculation of ν_{OH} . Although the approximation often yields reasonable agreement for observables (e.g., the centroid of the vibrational band), the fact is that the model is not “natural”, because it does not consider the essential three-

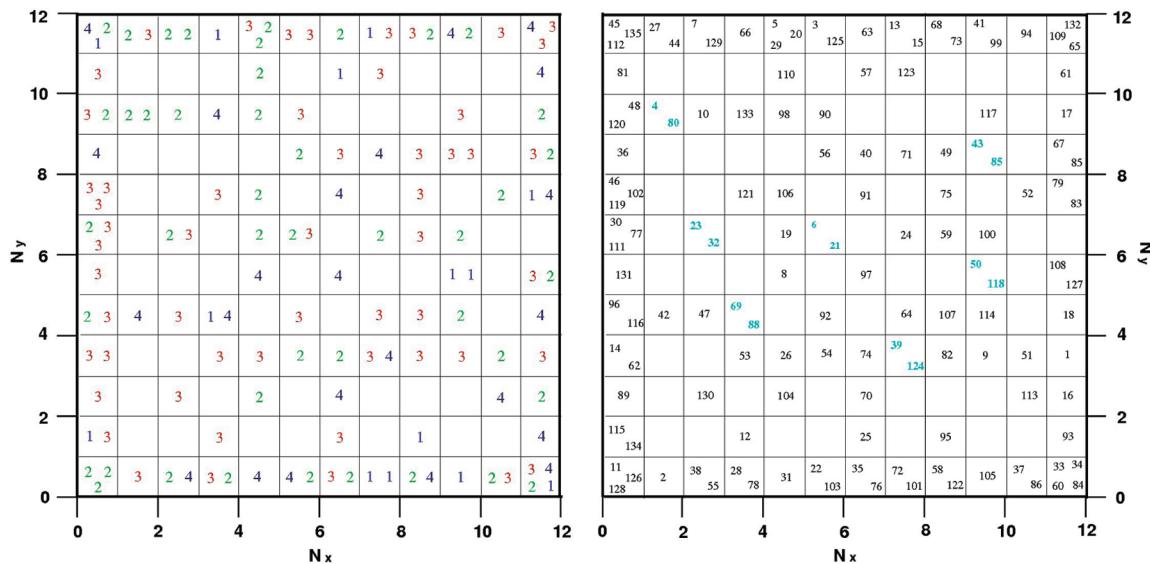


Figure 7. Grid on the left shows the structure of the Kohonen network, each rectangle symbolizing a single structure of interconnected processing units (“neurons”) which compete for the signal. Each neuron corresponds to one class of the OH stretching frequencies. Several conflicts between snapshot structures of different type are observed. The position of a single observation has no meaning; each observation is shifted just enough to make the observations attached to a particular neuron visible. The grid on the right shows the structure of the same Kohonen network representing neurons corresponding to the id number of the snapshot structure. As we can see, there are few similar structures (numbers marked in turquoise). With this representation, we try to find some clusters formed by subset of data (structures with serial numbers).

dimensional feature of the wavepacket and leaves out all the coupling with other degrees of freedom. Improvements of these drawbacks are possible by using a supercell and by treating the proton motion three-dimensionally, but at the time being they are prohibitively expensive. A possible caveat is also in the fact that the calculated envelope peaks at a slightly too high frequency value. Nevertheless, despite some inherent weak points, the absence of correlation between the instantaneous structure and the corresponding OH stretching frequency gives a pronounced message on the complexity of short hydrogen bonding, demonstrated by the present study from a rather unique viewpoint.

Confronting the absence of correlation patterns with the existing concepts that predict correlation even for very short hydrogen bonds, we need to stress again the difference at the fundamental level—while the popular correlation schemes have been established *on a data set of many different systems*, our approach attempted to find similar relations *within one system*, that is, among the several dynamically sampled instantaneous geometries and the pertinent OH stretching frequencies acquired from instantaneous proton potentials. While the present study treated the fluctuating structure as a number of different images, with each geometry parameter being an array of values, both the structure and spectroscopic quantities are treated as one number (equivalent or about equivalent to ensemble average) in the popular correlation schemes. As there exists a pronounced difference in the level of insight into the structure of matter between these concepts, our study does not attempt to undermine the established correlations (for short hydrogen bonds at least), but rather demonstrates the complexity of hydrogen bonding and stresses another poorly understood aspect of this phenomenon.

4. CONCLUDING REMARKS

Our goal was to evaluate the correlation between the constituents of the OH stretching band and the corresponding

geometry parameters of the snapshot structures of the crystalline sodium hydrogen bis(sulfate). The simple correlation scheme between ν_{OH} and the O···O distance, as found in other studies of sets of hydrogen-bonded crystals, appears not to be applicable to the present case: although we used a variety of methodologies for statistical data analysis (multiple linear regression, principal component analysis, principal multiple regression, and Kohonen neural networking), no appreciable linear or nonlinear correlation could be found. One of the possible reasons is in the fact that, for very short hydrogen bonds with extremely pronounced spectroscopic effects, the uncertainty of the established correlations is quite large. More importantly, the present work differs from many of the previous correlation studies in a crucial aspect, namely that the fluctuations of the structure (and consequently of the contribution to the OH stretching band) are considered whereas in the popular correlation schemes they are not; thus, the comparison of our results with the correlation schemes is not entirely legitimate.

In its essence, our work resembles the approach presented in refs 15 and 26, for it explores the correlation within a number of dynamically sampled images of a single system. Since the main scope of this study is to explore the correlation between geometry parameters and the pertinent frequency values, we did not consider the factors that govern the band intensity in the present work. These aspects have already been elaborated in the past by a similar methodology,²⁶ and a challenge remains to expand the present correlation studies by considering quantities such as electric field in order to get a more complete insight into the hydrogen bond dynamics. Additionally, the above-presented issue of correlation could be considered for other H-bonded systems, differing in H-bond length and strength, medium, etc., possibly leading to more generalized findings on correlation between instantaneous geometry and spectroscopic variables.

The absence of correlation demonstrates that there are various aspects of structure and dynamics of short hydrogen

bonding which are still poorly understood. As the quantum coupling of the longitudinal (stretching) proton motion to other degrees of freedom has not been explicitly undertaken in the present study, improvements of the present approach (e.g., inclusion of multiple coordinates in quantum treatment of the snapshot structures) remain a challenge for further work, albeit it needs to be stressed that many of such improvements are currently prohibitively expensive. Nevertheless, the present study clearly demonstrates that the short hydrogen bonding is an extremely complex interaction and a profound source of challenges for experimental and theoretical work.

■ ASSOCIATED CONTENT

S Supporting Information

Schematic representation of chemometrics methodologies. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: jernej.stare@ki.si.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Financial support of the Slovenian Research Agency (contract nos. P1-0012, P1-0017, and J1-2014) are gratefully acknowledged. The authors thank Mrs. Špela Župerl for her kind support in the data analysis by the Koh-NN program.

■ REFERENCES

- (1) Rundle, R. E.; Parasol, M. *J. Chem. Phys.* **1952**, *20*, 1487.
- (2) Lord, C. R.; Merrifield, R. E. *J. Chem. Phys.* **1953**, *21*, 166.
- (3) Mikenda, W. *J. Mol. Struct.* **1986**, *147*, 1.
- (4) Lutz, H. D.; Jung, C. *J. Mol. Struct.* **1997**, *404*, 63.
- (5) Libowitzky, E. *Monats. Chem.* **1999**, *130*, 1047.
- (6) Hammer, V. M. F.; Libowitzky, E.; Rossman, G. R. *Am. Mineral.* **1998**, *83*, 569.
- (7) Mikenda, W.; Steinbock, S. *J. Mol. Struct.* **1996**, *384*, 159.
- (8) Mikenda, W.; Steinbock, S. *J. Mol. Struct.* **1994**, *326*, 123.
- (9) Bratos, S.; Leicknam, J. C.; Pommeret, S. *Chem. Phys.* **2009**, *359*, 53.
- (10) Spencer, J. N.; Casey, G. J.; Buckfeld, J.; Schreibe, Hd. *J. Phys. Chem.* **1974**, *78*, 1415.
- (11) Novak, A. *Struct. Bonding (Berlin)* **1974**, *18*, 177.
- (12) Nakamoto, K.; Margoshes, M.; Rundle, R. E. *J. Am. Chem. Soc.* **1955**, *77*, 6480.
- (13) Efimov, Y. Y.; Naberukh, Yi. *Zh. Strukt. Khim.* **1971**, *12*, 591.
- (14) Pejov, L.; Spangberg, D.; Hermansson, K. *J. Phys. Chem. A* **2005**, *109*, 5144.
- (15) Kuhn, O.; Yan, Y. A. *Phys. Chem. Chem. Phys.* **2010**, *12*, 15695.
- (16) Pommeret, S.; Musat, R.; Renault, J. P.; Leicknam, J.; Bratos, S. *J. Phys.: Conf. Ser.* **2009**, *177*.
- (17) Marechal, Y.; Witkowski, A. *J. Chem. Phys.* **1967**, *48*, 3697.
- (18) Bertolasi, V.; Gilli, P.; Ferretti, V.; Gilli, G. *J. Am. Chem. Soc.* **1991**, *113*, 4917.
- (19) Zundel, G. *Adv. Chem. Phys.* **2000**, *111*, 1.
- (20) Hadži, D.; Bratos, S. The hydrogen bond. In *Recent developments in theory and experiments*; Schuster, P., Zundel, G., Sandorfy, C., Eds.; North-Holland Pub. Co.: Amsterdam, 1976; p 565.
- (21) Huggins, C. M.; Pimentel, P. C. *J. Phys. Chem.* **1956**, *60*, 1615.
- (22) Chamma, D.; Henri-Rousseau, O. *Chem. Phys.* **1999**, *248*, 53.
- (23) Chamma, D.; Henri-Rousseau, O. *Chem. Phys.* **1999**, *248*, 71.
- (24) Witkowski, A.; Wojcik, M. *Chem. Phys.* **1973**, *1*, 9.
- (25) Pirc, G.; Stare, J.; Mavri, J. *J. Chem. Phys.* **2009**, *132*, 224506.
- (26) Corcelli, S. A.; Lawrence, C. P.; Skinner, J. L. *J. Chem. Phys.* **2004**, *120*, 8107.
- (27) Sobiestianskas, R.; Banys, J.; Grigas, J.; Pawłowski, A. *Solid State Ionics* **2008**, *179*, 213.
- (28) Matsuo, Y.; Hatori, J.; Yoshida, Y.; Ikeda, Y.; Sugiyama, J.; Ikehata, S. *Phys. Lett. A* **2009**, *373*, 3470.
- (29) Matsuo, Y.; Takahashi, K.; Hatori, J.; Ikehata, S. *J. Solid State Chem.* **2004**, *177*, 4282.
- (30) Filliaux, F.; Lautie, A.; Tomkinson, J.; Kearley, G. *J. Chem. Phys.* **1991**, *154*, 135.
- (31) Homouz, D.; Reiter, G.; Eckert, J.; Mayers, J.; Blinc, R. *Phys. Rev. Lett.* **2007**, *98*.
- (32) Wood, B. C.; Marzari, N. *Phys. Rev. B* **2007**, *76*.
- (33) Stare, J.; Panek, J.; Eckert, J.; Grdadolnik, J.; Mavri, J.; Hadži, D. *J. Phys. Chem. A* **2008**, *112*, 1576.
- (34) Stare, J.; Mavri, J.; Grdadolnik, J.; Zidar, J.; Maksic, Z. B.; Vianello, R. *J. Phys. Chem. B* **2011**, *115*, 5999.
- (35) Jolliffe, I. T. *Principal component analysis*, 2nd ed.; Springer: New York, 2002.
- (36) Majcen, N.; Rajer-Kanduc, K.; Novic, M.; Zupan, J. *Anal. Chem.* **1995**, *67*, 2154.
- (37) Kohonen, T. *Self-organization and associative memory*, 3rd ed.; Springer-Verlag: Berlin ; New York, 1989.
- (38) Pimentel, P. C.; Sederholm, C. H. *J. Chem. Phys.* **1956**, *24*, 639.
- (39) Dodge, Y.; Marriott, F. H. C. International Statistical Institute. *The Oxford dictionary of statistical terms*, 6th ed.; Oxford University Press: Oxford, UK, 2003.
- (40) Schreuer, C. P., A.; Mukamel, S. *J. Am. Chem. Soc.* **2001**, *123*, 3114.
- (41) Lohninger, H. *Teach/Me Data Lab. 2.002*. In *SDL-Software Development Lohninger*; Springer: Berlin, 1999.
- (42) Videnova-Adrabinska, V. *J. Mol. Struct.* **1990**, *237*, 367.
- (43) Mikac, U.; Hadži, D.; Blinc, R. *Ferroelectrics* **2000**, *239*, 1245.
- (44) Ratajczak, H.; Orville-Thomas, W. J. *J. Mol. Struct.* **1968**, *1*, 449.
- (45) Cattell, R. B. *Multivariate Behav. Res.* **1966**, *1*, 245.
- (46) Kaiser, H. F. *Educ. Psychol. Meas.* **1960**, *20*, 141.