International Reviews in Physical Chemistry,
Vol. 25, No. 3, July–September 2006, 353–406

Taylor & Francis
Taylor & Francis Group

Application of genetic algorithms in automated assignments of high-resolution spectra

W. LEO MEERTS*† and MICHAEL SCHMITT‡

†Molecular- and Biophysics Group, Institute for Molecules and Materials,
Radboud University Nijmegen, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands
‡Heinrich-Heine Universität Düsseldorf, Institut für Physikalische Chemie,
Gebäude 26.43.02, Universitätsstraße 1, 40225 Düsseldorf, Germany

(*Received 3 April 2006; in final form 2 May 2006*)

This paper describes an automated assignment and fitting procedure for high-resolution rotationally resolved spectra. The method is based on the application of genetic algorithms (GA) and both frequency and intensity information of these spectra is used. The basic ideas behind the GA technique is introduced and the particular fitness function critical for the success of the GA evaluation is discussed. The meta-optimization of the internal GA parameters for an optimal exploration of the error landscape of the spectrum fitting is investigated. A number of typical examples are given in which the use of automated spectrum assignments with the GA method is of crucial importance in the analysis. Examples are given for fits of very dense spectra due to overlap of a number of rovibronic spectra in conformers and isotopomers and strongly congested spectra in dimer systems. It is also shown that since the GA performs an overall fit of the complete spectrum very good information on the orientation of the transition dipole moments is gathered.

Contents

*Corresponding author. Email: leo.meerts@science.ru.nl

International Reviews in Physical Chemistry
ISSN 0144–235X print/ISSN 1366–591X online © 2006 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/01442350600785490

1. Introduction

The task of analyzing and assigning complicated spectra has been developed much over the years. The traditional manners of spectral assignment are based on finding regularities and *by eye* pattern recognition and in this way assigning quantum numbers

to the transitions. In general this is a tedious process and even a relatively simple spectrum as for example that of the rotational cool naphthalene molecule [1] can take an experienced scientist from a couple of days up to several weeks. If the analysis involves a series of spectra for example as a function of vibrational quantum number or from different isotopomers the amount of work rapidly grows out of hand. Consequently, the traditional techniques inhibit the study of larger and more complicated systems such as molecules of biological interest.

A number of tools have been developed that make use of fast computers and their graphical possibilities to facilitate the assignment *by eye*. Very widely used and with great success, in particular in the analysis of microwave and IR spectra, is the Loomis-Wood method [2]. In a Loomis–Wood diagram the occurrence of a transition is plotted versus frequency in a two-dimensional peak diagram. Subsequent frequency intervals are displayed one below the other. If the length of the intervals in the Loomis–Wood diagram is equal to $2B$ the sub-band will appear as a vertical succession of points, because the lines in a rotational sub-band of a symmetric top or linear molecule are spaced roughly by $2B$. Deviations from $2B$ show up in different slopes and curvatures of the lines connecting successive points in different lines of the Loomis–Wood diagram. Such diagrams were first used by Loomis and Wood in 1928. Because of the time required to manually create a Loomis–Wood diagram, they were not useful in the initial assignment of spectra before fast microcomputers became available. The first computer program to generate a Loomis–Wood plot was written at the Ohio State University in the 1960s [3] and the first interactive Loomis–Wood applications by Winnewisser *et al.* appeared in the 1980s [4]. Loomis–Wood programs are particularly useful for the analysis of congested spectra of symmetric tops, slightly asymmetric tops, and linear molecules. Christopher F. Neese [5] has developed an interactive Loomis–Wood assignment package,† which has been written for IGOR Pro, a program for the analysis and display of numeric data. A nice example of a recent application of the Loomis–Wood method can be found in the paper of Thompson *et al.* [6].

A number of other groups have been developing software packages that use the ideas behind the Loomis–Wood diagrams [7, 8]. The group of Kisiel has written the program AABS‡ allowing graphical analysis of extremely broadband spectra [8]. It can be used to analyse pure rotational spectra, rotation–vibration, vibronic spectra from Fourier transform interferometers or laser spectrometers.

In this frame the computer program JB95§ developed by David F. Plusquellic [9] should be mentioned. This program has been very successful in the graphical assignment (*by eye*) of high-resolution rotationally resolved spectra. The program consists of a graphical user interface based on a Windows platform. Dialog boxes provide a user interface for the calibration, linearization and manipulation of experimental data and for the generation and optimization of the simulated spectra.

† This package is available at http://fermi.uchicago.edu/freeware/LoomisWood.shtml
‡ http://www.ifpan.edu.pl/~kisiel/aabs/aabs.htm
§ The program and help are available from http://physics.nist.gov/Divisions/Div844/facilities/uvs/jb95userguide.htm

In a recent paper Morruzi [10] presented an investigation on the feasibility of automated molecular line assignment for dense rovibrational spectra. While a general-purpose, fully automated assignment procedure seems to be out of reach for the near future, he shows that a thorough investigation of the problem can lead to new, more efficient and less interactive methods, at least in reasonably favorable conditions. Interesting suggestions are provided by some modern *heuristic* problem-solving algorithms, which mimic natural processes. As a first step, he has developed a *transgenic-evolutionary* algorithm, which has successfully assigned artificial spectra of up to almost 3500 lines.

In order to try to solve the assignment problems with the help of a computer the group of Neusser [11] has developed a procedure, which directly fits the experimental data, without prior assignments. This method, which is called 'correlation automated rotational fitting', has been pioneered by Levy and coworkers [12–14], and uses the correlation between the experimental and the simulated spectrum as a measure of the quality of the fit. Unfortunately, the method still has limited applicability.

The outcome of a first study in which genetic algorithms (GA) were used to solve the automatic assignment problem was very promising and resulted in a paper by Hageman *et al.* [15]. In that paper it was shown that for a series of previously manually assigned spectra of molecules like indole, indazole, benzimidazole [16], and 4-aminobenzonitrile [17], an automatic fitting based on genetic algorithms was successful. A crucial role in the success was the development of a proper fitness function.

In a further series of papers Meerts and Schmitt showed that the GA method to automate the assignment of complicated and entangled spectra was extremely successful [18–30]. Recently the group of Neusser presented an automated analysis of the partially resolved rovibronic spectrum of p-fluorostyrene-water using a genetic algorithm based method [31].

Related methods using GA have previously been used in a variety of other spectroscopic applications such as Nuclear Magnetic Resonance [32], fluorescence/absorption spectra in polyatomic molecules [33], Raman excitation profiles in solution [34], time-resolved polarized fluorescence spectroscopy [35], Mössbauer spectroscopy [36], X-ray spectra from plasmas [37] and powder EPR spectra [38].

In this work we introduce the general GA method, the definition of the fitness function and the use of the GA technique to automatically assign high-resolution spectra. The optimization procedures and limitations are discussed. A number of typical examples are given, in which the use of automated spectrum assignments with the GA method is of crucial importance in the analysis. This ranges from fits of very dense spectra due to overlap of a number of rovibronic spectra in conformers and isotopomers, strongly congested spectra in dimer systems to a better determination of the orientation of the transition dipole moments. Important but not decisive is the reliability of the experimental intensities and the presence of a model capable to explain the observed spectra. These conditions can however be released significantly.

Although most of the examples shown concern high-resolution UV gas phase spectra, the discussed automated assignment with the help of the GA method can be applied on a much wider range of spectra such as for example from a high-resolution Fourier transform spectrometer.

2. The genetic algorithm

2.1. *Introduction*

A description of the GA used in this investigation can be found in [15, 20]. The genetic algorithm is basically a global optimizer, which uses concepts copied from natural reproduction and selection processes. For a detailed description of the GA the reader is referred to the original literature [39–41]. The GA process is schematically depicted in figure 1. We shortly introduce the elements of the GA, which will be used in the following.

- Representation of the parameters: The molecular parameters are encoded as binary or as real data type, each parameter representing a gene. A vector of all genes, which contains all molecular parameters is called a chromosome. In an initial step the values for all parameters are set to random values between lower and upper limits which have to be chosen by the user. No prior knowledge of the parameters is necessary. In our calculations in general a total of 300–500 chromosomes are randomly generated, forming a population.
- The solutions (chromosomes) are evaluated by a fitness function (or cost function), which is a measure for the quality of the individual solution. The fitness function which is used here is described in section 2.2.
- One optimization cycle, including evaluation of the cost of all chromosomes is called a generation. Generally convergence of the fit in our case is reached after 300–500 generations.
- Pairs of chromosomes are selected for reproduction and their information is combined via a cross-over process. This cross-over might take place as a one-point, two-point or uniform cross-over. As cross-over just combines information from the parent generations. It basically explores the error landscape.
- The value of a small number of bits is changed randomly. This process is called mutation. Mutation can be viewed as exploration of the cost surface. The best solutions within a generation are excluded from mutation. This elitism prevents already good solutions from being degraded.

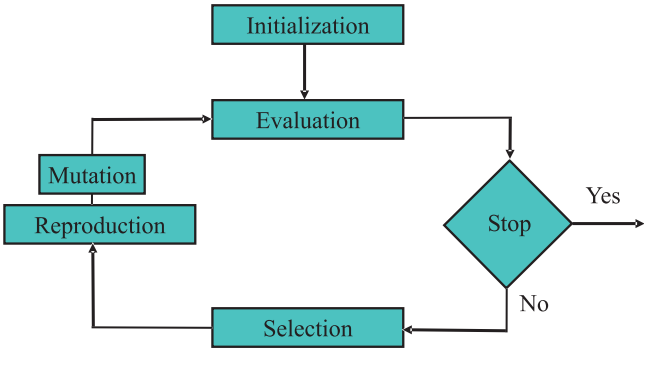

Figure 1. Schematics of the GA process.

The performance of the GA depends on internal parameters like mutation rate, elitism, cross-over probability and population size, which therefore should also be optimized for a given problem. Fortunately this meta-optimization results in similar parameters for quite different problems of optimization. The meta-optimization for some of the parameters is described in [20].

2.2. *The fitness function*

2.2.1. Definition. A detailed discussion of different fitness functions used for the assessment of similarities between spectra can be found in de Gelder *et al.* [42]. In the current discussion we will use both the terms *fitness function* (F_{fg}) and *cost function* (C_{fg}), where $C_{fg} = 100(1 - F_{fg})$ to characterize the quality of a solution. The fitness function for the fit of the spectra with N points using the genetic algorithm has been defined in equation (5) of Hageman *et al.* [15] as:

$$F_{fg} = \frac{\sum_{r=-l}^{l} w(r) \sum_{i=1}^{N} f(i)g(i+r)}{\sqrt{\sum_{r=-l}^{l} w(r) \sum_{i=1}^{N} f(i)f(i+r)} \sqrt{\sum_{r=-l}^{l} w(r) \sum_{i=1}^{N} g(i)g(i+r)}}. \tag{1}$$

In [15] C_{fg} is identical to F_{fg} in this paper. In this equation f and g represent the experimental and calculated spectra, respectively. The function $w(r)$ determines the sensitivity of the fitness function for a shift of the two spectra relative to each other. This can be rewritten by interchanging the sums and substituting $i + r = j$ to

$$F_{fg} = \frac{\sum_{i,j}^{N} f_i W_{ij} g_j}{\sqrt{\sum_{i,j}^{N} f_i W_{ij} f_j} \sqrt{\sum_{i,j}^{N} g_i W_{ij} g_j}}, \tag{2}$$

where

$$W_{ij} = w(|j - i|) \tag{3}$$

and $f_i = f(i)$ and $g_i = g(i)$.

F_{fg} in equation (2) can be interpreted as the cosine of the 'angle' between the experimental and theoretical spectrum. With the column vectors:

$$\begin{aligned} \boldsymbol{f} &= (f_1, f_2, \ldots, f_N)^T \\ \boldsymbol{g} &= (g_1, g_2, \ldots, g_N)^T \end{aligned} \tag{4}$$

and the symmetric matrix $\boldsymbol{W}$ which has the matrix elements W_{ij} we can write:

$$F_{fg} = \cos(\alpha) = \frac{(\boldsymbol{f}, \boldsymbol{g})}{\|\boldsymbol{f}\| \, \|\boldsymbol{g}\|}. \tag{5}$$

Here the inner product $(\boldsymbol{f},\boldsymbol{g})$ is defined with the metric $\boldsymbol{W}$ as:

$$(\boldsymbol{f},\boldsymbol{g}) = \boldsymbol{f}^T\boldsymbol{W}\boldsymbol{g}, \tag{6}$$

and the norm of $\boldsymbol{f}$ as $\|\boldsymbol{f}\| = \sqrt{(\boldsymbol{f},\boldsymbol{f})}$; similar for $\boldsymbol{g}$. For $w(r)$ we used a triangle function [15] with a width of the base of Δw:

$$w(r) = \begin{cases} 1 - |r|/\left(\frac{1}{2}\Delta w\right) & \text{for } |r| \le \frac{1}{2}\Delta w \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

In order for F_{fg} to serve as a good fitness function for the quality of the fit, it should have the property that it reaches its maximum value if and only if $\boldsymbol{f}$ and $\boldsymbol{g}$ are identical (apart from a normalization). This condition is fulfilled provided the matrix $\boldsymbol{W}$ is positive definite. In Appendix A it has been shown that this holds for w defined in equation (7).

2.2.2. Numerical evaluation of the fitness function. Let us now consider the numerical evaluation of the fitness function F_{fg} from equation (2) in its relation with the calculated spectrum. The calculated spectrum is obtained by a convolution of each calculated transition (k) with intensity s_k by the lineshape function L:

$$g_j = \sum_{k=1}^{N} L_{jk} s_k \tag{8}$$

with $L_{jk} = l(|j-k|)$. In matrix notation this can be rewritten as:

$$\boldsymbol{g} = \boldsymbol{L}\boldsymbol{s}. \tag{9}$$

As it turned out at least 50% of the computing time in [15] was used to perform a straightforward calculation of the fitness function F_{fg} from equation (2). For a typical GA fit F_{fg} must be calculated 150 000 times. Hence a considerable reduction in computing time can be obtained by a more effective calculation of F_{fg}. We start with a rearrangement of the order of the evaluation of equation (5) and using the properties of $\boldsymbol{W}$ and $\boldsymbol{L}$. The numerator of equation (5) is evaluated first:

$$(\boldsymbol{f}, \boldsymbol{L}\boldsymbol{s}) = \boldsymbol{f}^T\boldsymbol{W}\boldsymbol{L}\boldsymbol{s} = \tilde{\boldsymbol{f}}\boldsymbol{s} \tag{10}$$

$$\tilde{\boldsymbol{f}} = \boldsymbol{f}^T\boldsymbol{W}\boldsymbol{L}. \tag{11}$$

From equations (10) and (11) it is obvious that the effect of $w(r)$ can be interpreted as an effective line broadening of the experimental† spectrum. The use of the broadening function $w(r)$ results in a smoother error-landscape which allows an easier optimization

†A different arrangement for the equations shows that $w(r)$ also can be interpreted as an effective broadening of the calculated spectrum g. Actually the experimental and the calculated spectra are broadened relative to each other.

of the genetic algorithm process. It should be noted that the simple least squares fitness function where all spectral points have the same weight is also described by F_{fg} for the limiting case that the width of $w(r)$ is zero and $\boldsymbol{W}$ becomes the identity matrix, in this case with very sharp and deep valleys. See e.g. figure 2 from [15].

The transformed experimental spectrum from equation (11) has to be evaluated only once. Formally the sum on the right-hand side of equation (10) runs over all N points of the spectrum. However the stick spectrum array is a very sparse one: typically N is of the order of 60 000–250 000, while the number of sticks (non-zero values of s_k) is only about 1000–3000. Therefore the use of equation (10) strongly reduces the necessary computing time. The reduction of computing time is actually more dramatic since the double sum in the numerator of equation (2) over N points is reduced to a single sum over a sparse array in equation (10). Furthermore the theoretical spectrum itself $\{g_i\}$ does not have to be calculated anymore. The first term in the denominator of equation (2) also has to be evaluated only once. The second term in this denominator has to be calculated every time a value of F_{fg} is needed in the process of the genetic algorithm. Fortunately this term can also be expressed in terms of the stick spectrum $\boldsymbol{s}(\{\boldsymbol{s_k}\})$:

$$\|\boldsymbol{g}\|^2 = \boldsymbol{s}^\dagger(\boldsymbol{L}^\dagger \boldsymbol{W}\boldsymbol{L})\boldsymbol{s}. \tag{12}$$

Since $(\boldsymbol{L}^\dagger \boldsymbol{W}\boldsymbol{L})$ is a banded matrix the evaluation of $\|\boldsymbol{g}\|^2$ from equation (12) is in practice almost linear in the number of sticks rather than quadratic. Again $(\boldsymbol{L}^\dagger \boldsymbol{W}\boldsymbol{L})$ has to be evaluated only once. It turned out that the effect of the above discussed modifications of the calculation of F_{fg} was that its calculation time became negligible to the total computing time.

The use of the stick spectrum, described in this section, is limited to applications in which the lineshape parameters, like Gaussian or Lorentzian width in the Voigt profile, do not have to be fit. Inclusion of the linewidth parameters in the fit requires the reevaluation of equation (10) during the GA process, which in practice dramatically increases the computing time. Therefore a linewidth fit should be performed after a determination of all other parameters in a separate fitting procedure. Some examples are given in section 5.

3. Optimization of the GA fits

3.1. *Effect of the width (Δw) of the weight function $w(r)$ on the convergence of the GA*

In cases where the error landscape is governed by steeply descending minima, optimizers often converge to a local minimum. This problem can be avoided by an effective line broadening of the experimental or calculated spectrum as described in section 2.2.2. This relative broadening of the spectra, introduced by the weight function $w(r)$, critically determines the ability of the GA to converge to the global minimum and also the speed of convergence. The effect of smoothing the error landscape is to sense regions which are far from the minimum. Decreasing Δw improves the accuracy of the molecular parameters obtained from the fit, while narrowing the parameter space leads

to an improved sampling in the region of the minimum. This of course is the critical step in the procedure. Too strong narrowing of the parameter space leads to premature convergence of the fit – with a high probability of finding a local minimum.

To investigate the dependence of the GA procedure on the various conditions a series of test fits were performed by Meerts *et al.* [20] using a synthetic spectrum. This spectrum consists of two overlapping sub-spectra generated using the asymmetric rotor Hamiltonian with the corresponding transition frequency intensities defined in section 4.2.1. A single temperature for the calculation of the intensities has been used. The maximum J-value used in the computation of the cost function for this spectrum is 22. The number of parameters to fit was 17. It was found, that with an initial ratio of broadening function to experimental linewidth $\Delta w/\Delta_{lw}$ of 20, all runs of the GA terminated in the global minimum. Nevertheless, the deviations of the fitted parameters from the true values are large, due to the large value of Δw, which leads to broad minima at the cost surface. In a subsequent step, the fit has to be refined with a limited parameter space centred around the best fit and with a smaller value of Δw. As a rule of thumb we found that the parameter limits can be reduced to one tenth of the initial range and Δw to a half of the initial broadening. This reduction has to be checked carefully after each successive step. It was found that the fit with the reduced linewidth ratio of 10 and the reduced parameter space is already quite close to the true values. It can be further improved by decreasing $\Delta w/\Delta_{lw}$ and the parameter search space in an iterative manner until convergence for the molecular parameters is reached. In the final step, the broadening has to be set to zero. The step width for the change of the broadening function between subsequent fits depends strongly on the quality of the experimental spectrum. The better the signal to noise ratio, the larger the reduction of the parameter limits and the broadening function in a single step can be.

3.2. *Meta-optimization of the internal GA parameters*

The GA contain a number of internal parameters, which determine the performance of the fit, and have to be optimized for a given problem themselves. This has been discussed to be a major drawback of this method [43]. The speed and convergence of GA algorithms depend amongst others on the encoding of the parameters (binary or real), the cross-over type (one-point, two-point, uniform), the size of the starting population, the rate of elitism and the mutation probability.

It was shown by Meerts *et al.* that the binary encoding of the parameters leads to a smaller cost and converges more rapidly, compared to real type representation. For the real type data encoding a number of runs even did not reach the global minimum [20]. A change of the encoding depth for the binary representation from 10 bit to 20 bit virtually does not change the performance of the GA. A Gray code [44] is used throughout all calculations in order to ensure a Hamming distance of one [45]. Comparing uniform and two-point cross-over a clear advantage of the two point cross-over regarding speed of convergence and fitness of the solutions was found.

The optimum size for the starting population in terms of best cost function and speed of convergence was found to be 300. Larger populations did not improve the cost function considerably, but led to increasingly larger CPU times. The elitism, which is defined as the fraction of one generation that is passed unchanged to the next

generation has been varied between 30 and 70%. Elitism helps to prevent good solutions to be lost from one generation to the next. The optimum value has been found to be 50%. Larger values keep too many bad solutions, while for smaller values the time consumption becomes large, since many good solutions are discarded. The mutation probability should be chosen between 1 and 5%. The smaller this value gets, the better is the cost function. On the other hand the risk of being trapped in a local minimum increases. The CPU time as function of the mutation probability virtually does not change.

3.3. *Uncertainties in and correlations between the parameters from a GA fit*

After finding the parameters from a GA calculation that best fit the experimental spectrum it is in most cases also very relevant to obtain information about the uncertainties in these values and correlations between the parameters.

This information can be obtained by performing a number of GA fits with different random start values for the parameters. If proper convergence is obtained the results for each of the independent GA fits should lead to the same set of best fit values for the parameters. However, since the starting points are chosen arbitrarily, called a random seed, the final values of the parameters will be different. The spread in the results, assuming comparable cost-function values, will reflect the uncertainties in the parameters. To obtain a full statistical analysis an infinite number of independent GA runs has to be carried out.

In order to determine how many independent GA fits are needed to obtain a good estimate of the statistical uncertainties in the parameters we used the synthetic spectrum defined in section 3.1 and performed a series of GA fits with 3, 10, 20 and 28 random seeds. Since much is known about this synthetic spectrum we were also able to perform a classical fit in which quantum numbers are assigned to each individual transition, shortly called 'assigned fit' in this work. In this assigned fit those transitions are included that would be used in a typical manual assignment of the spectra, that is transitions with an intensity larger than 5% of the maximum intensity and with rotational quantum numbers up to $J = 10$. For such an assigned fit the statistical errors in the parameters are well defined see e.g. Appendix B.

The results of the series of GA fits compared to the assigned fit showed that a very good estimate of the uncertainties in the final parameters is obtained if between 10 and 20 randomly started GA fits are carried out. Of course the non-converged GA fits, which are easily recognized by their large cost function should be removed in the calculation of the statistical errors. It should also be clear that the parameter space must be sufficiently large so that the different seeds can indeed probe the minimum of the error landscape.

The relevant statistical properties that are obtained are the variance of, and covariance between the parameters. The variance of the parameter α is defined as

$$V_{\alpha\alpha} = \frac{1}{n-1}\sum_{i=1}^{n}(\alpha_i - \alpha_{av})^2, \tag{13}$$

where n is the number of performed independent GA fits (seeds), α_i the resulted value for α of the individual GA calculation i and α_{av} the average value for α of the different seeds. The statistical error σ_α in the parameter α is given by $\sqrt{V_{\alpha\alpha}}$.

The covariance between two parameters α and β is defined as

$$C_{\alpha\beta} = \frac{1}{n-1}\sum_{i=1}^{n}(\alpha_i - \alpha_{av})(\beta_i - \beta_{av}) \tag{14}$$

where the definitions β_i and β_{av} are identically defined as the corresponding α's.

The covariances are needed if parameters from a GA fit are used in consecutive calculations, for example to deduce the structural information from sets of rotational parameters (section 4.3).

4. Methods

4.1. *Experimental*

All experimental spectra are obtained with a high-resolution UV laser spectrometer. Its resolution is of the order of $0.0005\,\mathrm{cm}^{-1}$ or 15 MHz (full width at half maximum) at $35000\,\mathrm{cm}^{-1}$. This is achieved by crossing a single frequency continuous wave laser with a molecular beam and detecting the laser induced fluorescence. The schematics is depicted in figure 2.

The experimental setup for the rotationally resolved laser induced fluorescence (LIF) is described in detail elsewhere [1, 46, 47]. Briefly, it consists of a single frequency ring dye laser pumped with either with 6 W of the 514 nm line of an Ar^+ laser or with 7 W of the 515 nm line of a diode pumped cw Yb:YAG laser (ELS MonoDisk-515). The light is coupled into an external folded ring cavity for second harmonic generation (SHG). The molecular beam is formed by expanding a mostly heated sample seeded typically in 500 mbar of argon through a 100 μm hole into the vacuum. The molecular beam machine consists of three differentially pumped vacuum chambers which are linearly connected by two skimmers (1 mm and 3 mm, respectively) in order to reduce the Doppler width. The molecular beam is crossed at right angles in the third chamber with the laser beam 360 mm downstream of the nozzle. The resulting fluorescence is collected perpendicularly to the plane defined by laser and molecular beam by an imaging optics setup consisting of a concave mirror and two planoconvex lenses. The resulting Doppler width in this setup is 15 MHz (FWHM). In some experiment the Doppler width was 25 MHz, because of a slightly different arrangement of the optical system. The integrated molecular fluorescence is detected by a photomultiplier tube whose output is discriminated and digitized by a photon counter and transmitted to a PC for data recording and processing. The relative frequency is determined with a calibrated quasi-confocal Fabry–Perot interferometer with a free spectral range (FSR) of about 150 MHz. The absolute frequency was determined by recording the iodine absorption spectrum and comparing the transitions to the tabulated lines [48].

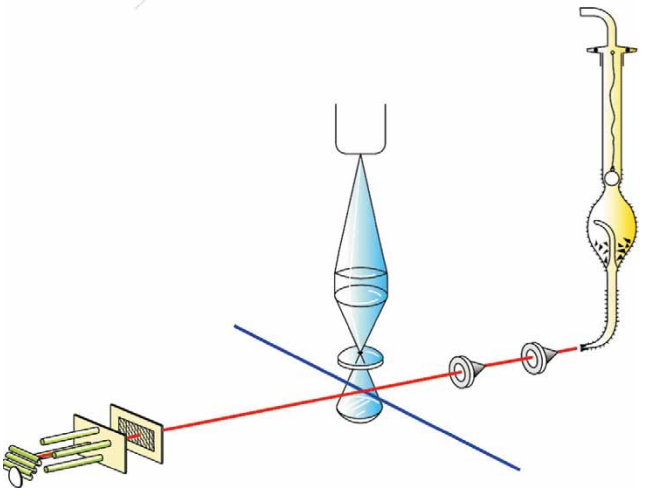

Figure 2. [Colour online] Schematics of the Nijmegen and Düsseldorf high-resolution UV spectrometers.

4.2. *The calculated spectra*

4.2.1. The asymmetric rotor Hamiltonian. For an asymmetric rotor spectrum of a transition between two electronic (or vibrational) states, the following Hamiltonian was employed [49]:

$$\widehat{H}_r = AJ_a^2 + BJ_b^2 + CJ_c^2. \tag{15}$$

Here, the $J_g (g = a, b, c)$ are the components of the body fixed angular momentum operator, and A, B and C are the three rotational constants. Let I^g be the corresponding moments of inertia than the inertial defect ΔI is defined as:

$$\Delta I = I^c - I^a - I^b. \tag{16}$$

Centrifugal distortion constants may be included if necessary through the model of a distortable rotor in Watson's A-reduced form [50, 51]:

$$\begin{aligned} H_{\text{dist}} = & -\Delta_J J^4 - \Delta_{JK} J^2 J_a^2 - \Delta_K J_a^4 \\ & - 2\delta_J J^2 \left(J_b^2 - J_c^2\right) - \delta_K \left[J_a^2 \left(J_b^2 - J_c^2\right) + \left(J_b^2 - J_c^2\right) J_a^2\right] \\ & + H_J J^6 + H_{JK} J^4 J_a^2 + H_{KJ} J^2 J_a^4 + H_K J_a^6 \\ & + 2h_J J^4 \left(J_b^2 - J_c^2\right) + h_{JK} J^2 \left[J_a^2 \left(J_b^2 - J_c^2\right) + \left(J_b^2 - J_c^2\right) J_a^2\right] \\ & + h_K \left[J_a^4 \left(J_b^2 - J_c^2\right) + \left(J_b^2 - J_c^2\right) J_a^4\right]. \end{aligned} \tag{17}$$

Hence the total rotational Hamiltonian becomes

$$H_R = H_r + H_{\text{dist}}. \tag{18}$$

4.2.2. Internal rotation Hamiltonian. If a molecule contains one or more groups that can internally rotate, additional terms in the Hamiltonian are needed to describe the effects of the internal rotation. Examples of such internal rotation (often called torsion) are the rotation of a methyl group or the coupling of an OH group with the phenyl group in for example phenol. The first is described by a $3n$-fold barrier, while the latter possesses a $2n$-fold barrier with $n \geq 1$. The torsional motion of the water moiety in, for example, phenol–H_2O is also described by a $2n$-fold barrier [52].

The coupling between the internal rotation of the $2n$ and/or $3n$ group(s) and the overall rotation of the molecule are treated in the formalism of the principal axis method (PAM) [53, 54]. The Hamiltonian for the rotation-torsion interaction can be written as:

$$\widehat{H} = \widehat{H}_R + \widehat{H}_T + \widehat{H}_{RT}, \tag{19}$$

with $\widehat{H}_R$ the rigid rotational Hamiltonian from equation (18). The torsional Hamiltonian for the mixed V_3/V_6 potential (methyl torsion) is given by:

$$\widehat{H}_T = Fp^2 + \frac{1}{2}\sum_{n=1}^{2} V_{3n}(1 - \cos 3n\alpha), \tag{20}$$

and for a two-fold V_2 potential (OH and water torsion) by:

$$\widehat{H}_T = Fp^2 + \frac{1}{2}\sum_{n=1}^{2} V_{2n}(1 - \cos 2n\alpha). \tag{21}$$

F is the torsional constant and defined by:

$$F = \frac{h}{8\pi^2 r I_\alpha} \tag{22}$$

with

$$r = 1 - \sum_{g=a,b,c} \frac{\lambda_g^2 I_\alpha}{I_g} \tag{23}$$

where λ_g $(g = a, b, c)$ are the direction cosines between the inertial axes and the axis of internal rotation. I_α is the moment of inertia of the internal rotor with respect to its symmetry axis and the I_g are the principal moments of inertia of the entire molecule. In what follows, the treatment for the $n = 2$ and $n = 3$ potentials is identical.

The coupling of internal and overall rotation is described by $\widehat{H}_{RT}$ [53, 55]:

$$\widehat{H}_R T = FW^{(1)}_{v\sigma}(\rho_a J_a + \rho_b J_b + \rho_c J_c) + FW^{(2)}_{v\sigma}(\rho_a J_a + \rho_b J_b + \rho_c J_c)^2. \tag{24}$$

For a $2n$-fold barrier the energy levels are split by $\widehat{H}_{RT}$ in two non-generate sub-torsional levels $\sigma = 0$ and $\sigma = 1$, the $A^{(2)}$ and B-states, respectively. For a $3n$-fold barrier the levels are split into a non-degenerate $\sigma = 0$ ($A^{(3)}$-state) and double degenerate $\sigma = \pm 1$ (E-state). We use the superscripts (2) and (3) to distinguish the different A sub-levels from the two and threefold internal rotation from the A rotational constant. The first order perturbation coefficients $W^{(1)}_{v\sigma}$ are zero for the non-degenerate $A^{(2)}$-, $A^{(3)}$- and B-levels. $W^{(2)}_{v\sigma}$ is non-zero for all levels.

The components of the vector $\vec{\rho}$ are given by the direction cosines $\cos\eta_g$ of the internal rotor axis in the principal axis system, the moments of inertia of the entire molecule I_g ($g = a, b, c$), and the moment of inertia of the internal rotor group I_α:

$$\rho_g = I_\alpha / I_g \cos\eta_g. \tag{25}$$

The n-th order perturbation coefficients $W^{(n)}_{v\sigma}$ in equation (24) are defined by [53]:

$$W^{(0)}_{v\sigma} = \frac{E_{v\sigma}}{F} \tag{26}$$

$$W^{(1)}_{v\sigma} = -2\langle v, \sigma | p | v, \sigma \rangle \tag{27}$$

$$W^{(2)}_{v\sigma} = 1 + 4F \sum_{v'} \frac{\left|\langle v, \sigma | p | v', \sigma \rangle\right|^2}{E_v\sigma - E'_v\sigma} \tag{28}$$

where $|v, \sigma\rangle$ are eigenfunctions of equations (20) and (21), and $E_{v\sigma}$ are the respective eigenvalues with v as the torsional state index.

4.2.3. Intensities. The intensities of the rotational lines are calculated from the eigenvectors of the effective rotational Hamiltonian and the known direction cosine matrix elements [53]. The only selection rule used is $\Delta J = 0, \pm 1$ which distinguishes P, Q, and R branches. The linestrength is proportional to:

$$A_{r'r''} \propto |\mu_a \langle r' | \Phi_{Za} | r'' \rangle|^2 + |\mu_b \langle r' | \Phi_{Zb} | r'' \rangle|^2 + |\mu_c \langle r' | \Phi_{Zc} | r'' \rangle|^2 \tag{29}$$

where the Z-axis is the space-fixed axis along the direction of the laser polarization, $|r\rangle \equiv |J, K_a, K_c\rangle$ and

$$\begin{aligned} \mu_a &= \mu \sin\phi \cos\theta \\ \mu_b &= \mu \sin\phi \sin\theta \\ \mu_c &= \mu \cos\phi. \end{aligned} \tag{30}$$

The double and single prime denote the ground and excited state, respectively, μ is the absolute value of the transition dipole moment (TDM), (ϕ, θ) the spherical coordinate angles of the transition moment vector in the molecular fixed frame (a, b, c), and Φ_{Za}, Φ_{Zb} and Φ_{Zc} are the direction cosines between the laboratory Z-axis and the molecular a, b and c axes.

In equation (30) θ is the angle between the projection vector of the transition moment on the ab-plane, and the a-axis and ϕ the angle between the TDM and the c-axis.

4.3. *The structure determination*

The program pKrFit [56] has been developed to determine the structure from the rotational constants of a set of isotopomers. pKrFit uses a gradient-based χ^2 minimizer as well as a GA based global optimizer [28]. While the gradient method is much faster, its main disadvantage is the possibility of being trapped in a local minimum, depending on the starting geometry for the fit. The genetic algorithm based fit varies the geometry parameters between given limits and leads generally to the global minimum. The GA library [57] is used in minimization mode and directly uses the least square (χ^2) value as cost function. Additionally, the change of the orientation of the transition dipole moment for different isotopomers, which contain additional structural information, can be used in the fit. The use of data from different sources can minimize the correlations between the parameters, even if the accuracy of the different data sets is very different.

If the different vibrational contributions from the different isotopomers are completely neglected the following assumption can be made:

$$I_0^g = I_{\text{rigid}}^g(r_0). \tag{31}$$

In this equation the three I_0^g's are the (experimentally determined) zero-point averaged moments of inertia with respect to the inertial axes g. The function $I_{\text{rigid}}^g(r_0)$ is calculated from the structural parameters r_0 using rigid-molecule formulas. The resulting structure is called the r_0-structure. Other structures, which contain at least partially zero-point vibrational effects can be fit, if sufficient isotopic species have been measured. These comprise the *pseudo-Kraitchman* structure (*pseudo-*r_s-structure) [58–60] and different mass-dependent structures introduced by Watson which are designated as $r_m^{(1)}$ and $r_m^{(2)}$-structures [61–64]. The *pseudo-*r_s-structure is based on the following assumption:

$$I_0^g = I_{\text{rigid}}^g(r_s) + 1/2\epsilon_{0g}. \tag{32}$$

The additional parameters ϵ_{0g} are the vibrational corrections, which are assumed to be the same for all isotopomers. Since they have to be fit together with the geometry parameters, the *pseudo-*r_s-structure can only be determined in cases where sufficient inertial parameters from different isotopomers are available.

4.4. *The software used*

The GA library PGAPack version 1.0/2005, which can run excellent on parallel processors, has been used [57]. By now the calculations have been performed on many different computer systems varying from a dual processor PC with two Pentium 2.8 GHz processors to large Linux cluster systems for example at the Dutch National Computer Facilities (http://wwww.sara.nl). In the latter case up to 64 nodes are mostly used. We have developed a package of computer programs that is used for all GA fits discussed in this paper.† Typical computing times on the dual processor PC are 7 minutes wall clock time for a single GA fit of the synthesized test spectrum (section 3.1). The model Hamiltonians that are supported in the current version of the program are of a symmetric top, of an asymmetric rotor (section 4.2.1), of molecules with internal rotation from a CH_3 group as well as with a two-fold barrier (section 4.2.2). The Hamiltonian for the three-fold two rotor case is also supported but limited to the special case of two identical rotors, symmetric to the *a*-axis, with their torsional axes in the *ab*-plane [22, 65]. All kinds of combinations, even of different type of Hamiltonians, can be made.

The program allows the direct fitting of the rotational constants, but it also has the option to fit linear combinations of the constants like $(B + C)$ and $(B - C)$ or even the inertial defect (ΔI). By setting $\Delta I = 0$ one can force planarity of the molecule.

5. Examples of the success of GA automatic assignments

5.1. *GA fits of very dense rovibronic spectra*

In the following section we will present the automated GA fits of some rovibronic spectra, which are very congested due to small rotational constants. These spectra normally do not represent a great difficulty for the GA, as will be shown below.

5.1.1. [7-D]phenol-N_2. As a first example we treat the fit of the rovibronic spectra of several isotopomers of the phenol-nitrogen cluster [19]. This is an example in which it can be shown that the results of a GA fit are identical to those of an assigned fit. The structure of phenol-nitrogen is depicted in figure 3: the nitrogen is located in the plane of the phenol, hydrogen bonded to the OH group with a bond length of 225.5 pm. In the following [7-D]phenol means replacement of hydrogen by deuterium at the hydroxy group of phenol. The rotationally resolved electronic spectrum of the electronic origin of [7-D]phenol-N_2 is shown in figure S1. The observed spectrum consists of about 400 clusters of lines, with only a few single rovibronic lines.

The molecular parameters, obtained from the GA fit with $\Delta w/\Delta_{lw} = 0$ are presented in the second column of table 1. The values given and the quoted uncertainties are the result of a statistics on ten independent GA runs, with different initial seeds. In a second

†On request, the authors make available a full version of their GA program. The program has been thoroughly tested under Linux as well several UNIX versions. The package contains the program and an extensive manual with the installation procedure. License conditions are applicable if the program is used.

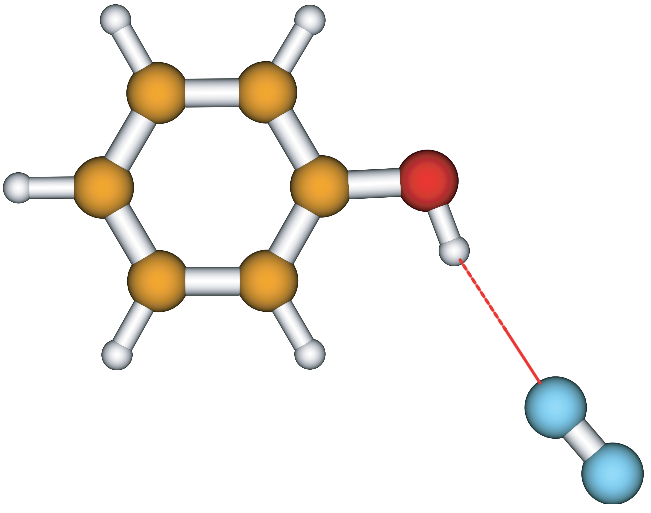

Figure 3. [Colour online] Hydrogen bonded structure of phenol-nitrogen.

Table 1. Comparison of the molecular parameters of the phenol-nitrogen cluster as obtained from an assigned assignment and from the genetic algorithm fit.

Parameter	GA fit	Assigned fit
A''(MHz)	4071.06(13)	4072.18(25)
B''(MHz)	647.89(2)	648.01(4)
C''(MHz)	559.13(2)	559.26(4)
$T(K)$	1.6(5)	
$\theta(^\circ)$	62.53(7)	
ΔA(MHz)	−140.99(6)	−141.560(91)
ΔB(MHz)	15.71(1)	15.708(13)
ΔC(MHz)	8.70(1)	8.671(9)

step, the result of the GA calculation are used to assign quantum numbers to the individual transitions and clusters of lines. Because of the high quality of the GA fit this was possible in spite of the large number of overlapping lines. With these line position assignments a second fit to the parameters of a rigid rotor Hamiltonian of equation (15) was performed. The latter fit yields better values in particular for the uncertainties of the parameters. For most parameters the values obtained from the line fit (table 1 column 1) agree within their uncertainties to the corresponding GA results. This spectrum presents an example in which the rovibronic spectrum could be fitted by the GA in a single step, without further refinement of Δw.

5.1.2. Benzonitrile-Ar. If due to experimental limitations only the outermost parts of the *P*- or the *R*-branch can be recorded and the electronic origin of a rovibronic band is missing, the task of performing an assigned fit becomes tedious or even impossible. However, in this difficult case the genetic algorithm also succeeds in finding the

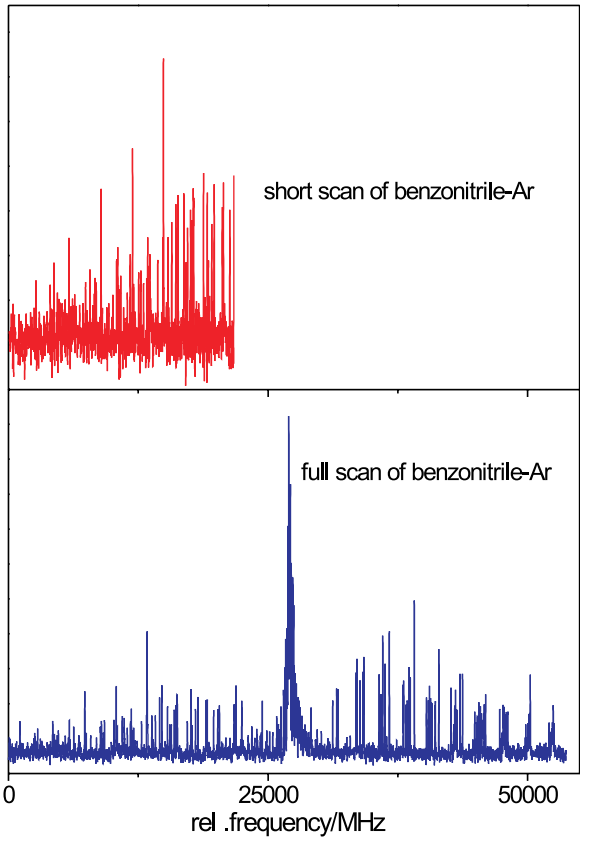

Figure 4. [Colour online] Upper trace: low-frequency part of the spectrum of benzonitrile-Ar. Lower trace: complete rovibronic spectrum. Intensities are given in arbitrary units. For details see text.

global minimum and assigning the spectrum properly. As an example we chose the spectrum of the electronic origin of benzonitrile-Ar [19], shown in the upper trace of figure 4. Obviously only the low-frequency side of the spectrum has been measured with a quite bad signal-to-noise ratio. Nevertheless, the GA was able to determine the molecular parameters. The result is given in the first column of table 2. The electronic origin is found by the GA to be 8000 MHz to the blue of the high-frequency end of the spectrum. A GA fit to the complete spectrum with good signal-to-noise (lower trace in figure 4) yields slightly different molecular parameters (second column of table 2). Nevertheless, the quality of the parameters obtained from the fit to the partial spectrum is surprisingly good. The only parameters that have relatively large deviations are the polar and azimuthal angles of the transition dipole moment. This is obviously due to the fact that the band type cannot be determined accurately from a fit of a single branch only.

Results from previous studies on Benzonitrile-Ar are also given in the last column of table 2. The current results for the excited state are more accurate than those of Helm *et al.* [11] because of the substantially lower resolution of their experiment.

Table 2. Molecular constants from a GA assignment of the partial spectrum of the origin of benzonitrile-Ar, the complete spectrum and an assigned fit. See text for details.

Parameter	*GA* fit[a]	*GA* fit[b]	Assigned fit	Other work
A''(MHz)	1343.80(150)	1347.32(19)	1347.58(18)	1347.789(12)[c]
B''(MHz)	1002.55(121)	1004.99(4)	1004.98(14)	1006.020(9)[c]
C''(MHz)	717.68(108)	718.99(4)	717.70(39)	719.817(2)[c]
θ(°)	22(3)	17.53(7)	20	
ϕ(°)	82(5)	70.05(2)	70	
$T(K)$	1.71(3)	1.68(3)	2	
ΔA(MHz)	−32.89(44)	−32.61(3)	−32.47(23)	−33.8(36)[d]
ΔB(MHz)	20.60(30)	21.08(16)	20.87(14)	21.4(27)[d]
ΔC(MHz)	6.25(15)	6.76(7)	6.80(34)	10.2(75)[d]

[a] Fit to the spectrum in the upper trace of figure 4.
[b] Fit to the spectrum in the lower trace of figure 4.
[c] MW rotational constants from reference [66].
[d] From reference [11].

Our ground state values do not completely agree with the very accurate microwave results from Dreizler *et al.* [66]. This is an indication that the uncertainties in our parameters, based on the statistical behaviour of the GA fits, are slightly underestimated.

5.1.3. 4-methylphenol and its water complex. 4-methylphenol (*p*-cresol) and its 1 : 1 complex with water are beautiful examples of the capability of the GA to entangle spectra with a complicated sub-structure. 4-methylphenol possesses two feasible large amplitude motions, the methyl rotation with a mixed V_3/V_6 potential, and the two-fold hydroxy rotation. The molecular symmetry group, which describes both motions is G_{12}. In the *p*-cresol water cluster, the OH torsion is quenched; instead an additional splitting due to the torsional motion of the water moiety with respect to the water symmetry axis is observed. This barrier is expected from the similar phenol-water system to be quite low. Thus, we have the case of two low-barrier torsional motions, which will split the origin into four sub-bands, an $A^{(3)} - E$ pair due to the threefold motion and an $A^{(2)} - B$ pair due to the two-fold rotation (e.g. section 4.2.2). Since the rotational constants of this cluster are quite small, the overlapping of four sub-spectra leads to a very congested overall spectrum, in which it is difficult, if not impossible to identify single rovibronic lines.

Because both the two-fold and threefold axes are in the plane of the phenyl group the coefficients ρ_g from equation (25) can be written as:

$$\begin{aligned} \rho_a &= I_\alpha / I_a \cos\eta \\ \rho_b &= I_\alpha / I_b \sin\eta \\ \rho_c &= 0 \end{aligned} \tag{33}$$

with η as the angle between the projection vector of the *n*-fold rotor axis on the *ab*-plane and the *a*-axis.

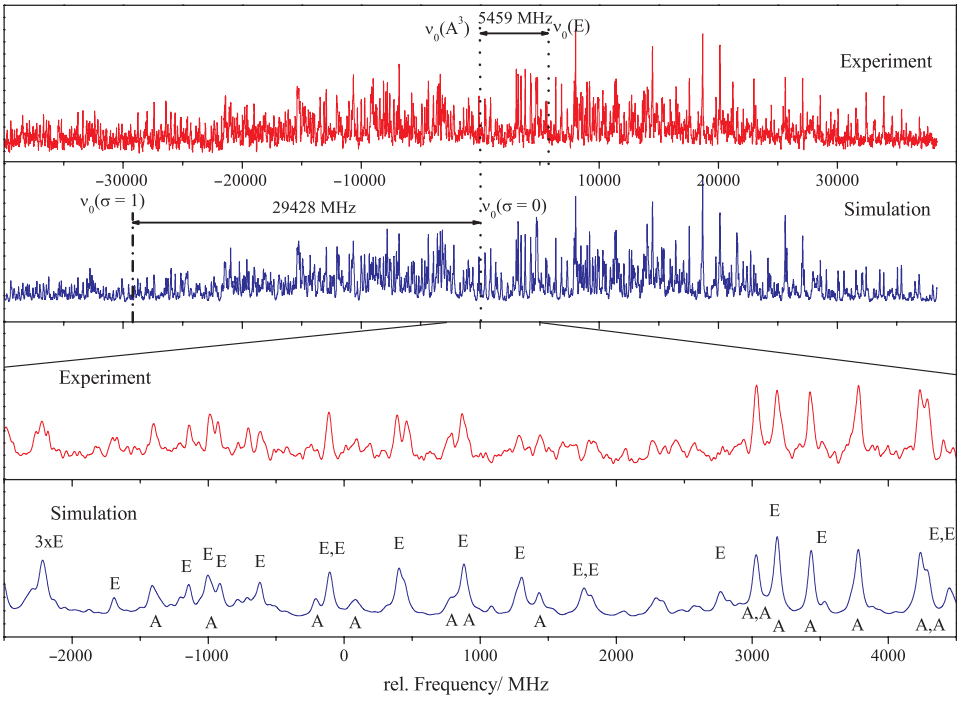

Figure 5. [Colour online] Rovibronic spectrum of the electronic origin of 4-methylphenol-water. Frequencies are given relative to the electronic origin at 34 972.873(10) cm^{-1}.

Figure S2 shows the rotational resolved spectrum of the electronic origin of *p*-cresol. It has been shown in [27] that this spectrum could fully be assigned with the GA method. As an example we will discuss here the 4-methylphenol$(H_2O)_1$ complex. Figure 5 presents the rotationally resolved electronic spectrum of the origin of the *p*-cresol-water cluster. The origin is split into an $A^{(3)}$ and an E sub-band due to the methyl torsion with a sub-torsional splitting and into an $A^{(2)}$ and B sub-band due to the water torsion. Under this assumption the spectrum has been automatically assigned with the GA method. It was found that the methyl sub-torsional splitting and the sub-torsional splitting due to the water torsion were 5459 MHz and 29428 MHz, respectively.

As in the case of phenol-water [52], the sub-torsional structure could be fit due to the two-fold water torsion using a rigid rotor model with different rotational constants for both components. Similar to the phenol-water case, the B and C rotational constants of the $(\sigma = 0)$ and $(\sigma = 1)$ components are equal within their uncertainties. The A constants of the sub-torsional components differ by about 11 MHz indicating that the torsional axis of the water torsion is more or less parallel to the main inertial *a*-axis of the cluster. This structure can be described as a trans-linear hydrogen bond configuration, with the water moiety acting as proton acceptor, like in phenol-water. A spin statistics of 1 : 3 for the ratio of $(\sigma = 0)$ to $(\sigma = 1)$ further supports a structure with the water moiety symmetric with respect to the aromatic plane.

This analysis clearly shows the convenience of the GA analysis of such a complex spectrum. Although by eye no periodicity can be observed at first glance, the GA is capable of easily finding the two splittings from the methyl torsion and the water torsion. Of course, the application of the GA based automated technique still requires the choice of the 'correct' Hamiltonian for the problem under consideration. An easy way of finding spectral splittings (and thus obtaining information about the required Hamiltonian) is to perform an autocorrelation of the spectrum. This method has been shown by Neusser's group to be helpful in the analysis of the phenol-water cluster spectrum [67] and later by Remmers *et al.* [68] in the analysis of the tunnelling spectrum of the benzoic acid dimer; see also section 5.3.1. Nevertheless, if one component of the complete spectrum amount to more than 50% of the total intensity, the GA is capable of fitting this part of the spectrum, although it might be concealed in the dense spectrum of the other component(s) and would therefore be inaccessible to a classical analysis using line position assigned fits. From the difference of the partly fitted spectrum to the experimental one the missing components can easily be recognized and subsequently fit.

The angle of the transition dipole moment with the *a*-axis is determined from the relative intensities of *a* and *b* bands to be 76.7°. The rotation of 13.3° with respect to the monomer is in good agreement with a value of 20°, that is predicted by rotation of the inertial axis upon cluster formation on the basis of the *ab initio* structure [69].

From the structures of the S_0 and S_1 of the cresol moiety determined by Myszkiewicz *et al.* [27], the structure of the p-cresol-water can be fitted to the geometric rotational constants using pKrFit. All geometry parameters in p-cresol have been kept fixed at the monomer values, the geometry of the water moiety has also been kept fixed. The symmetry constraint was furthermore imposed on the fit: the H atoms of the water moiety are symmetric with respect to the aromatic plane. Table V of Myszkiewicz *et al.* shows the results. The OO hydrogen bond length decreases by 5.0 pm upon electronic excitation, imaging the increased acidity of p-cresol upon electronic excitation, while the OOC angle and the HOOC dihedral, describing the orientation of the water moiety nearly stay constant. Two different views of the ground state structure are shown in figure 6.

5.2. *Simultaneous GA fits of a number of overlapping rovibronic spectra*

A much more demanding task than a fit of a single rovibronic spectrum is the simultaneous fit of several overlapping spectra. First of all the number of transitions within a spectral interval is multiplied, leading to very dense and congested spectra. Secondly, the number of molecular parameters is also multiplied, which generates quite a large parameter space.

Overlapping spectra occur in particular if several isotopic species are investigated. While mass resolution techniques like resonance two-photon ionization with time-of-flight mass spectroscopy, are able to separate the isotopic species with different mass, the techniques normally lack spectral resolution due to the pulse width limited resolution used in these studies. On the other hand mass selection cannot be combined with high-resolution LIF spectroscopy. Below, we show that the GA spectrum

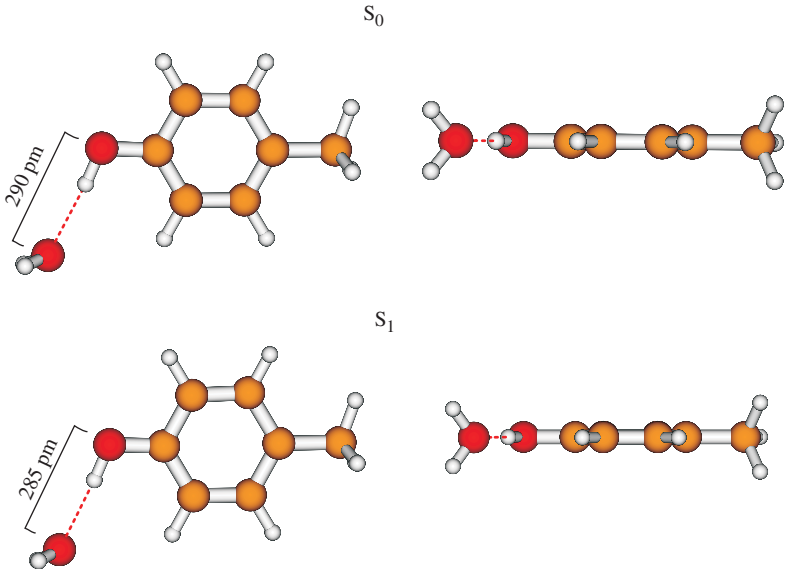

Figure 6. [Colour online] Top and side views of the ground and excited state structures of p-cresol-water. For all other geometry parameters see table V of Myszkiewicz *et al.* [27].

assignments are capable of handling overlapping spectra both from different isotopomers as well as from different conformers.

5.2.1. 7-azaindole (Pyrrolo[2,3-b]pyridine). As a first example we chose the spectrum of 7-azaindole (7AI). The spectrum of 7AI was automatically assigned using the genetic algorithm based fit [21]. The rotational constants of 7AI obtained from this GA fit are reported in table 3. For numbering and structure see figure 7.

Soft deuteration, by adding 20 mbar D_2O to the Ar seed gas prior to expansion resulted in the spectrum shown in figure 3 of [21]. Apparently a second band emerges, which can be assigned to the 7AI[ND] isotopomer. Both bands were fit together using the GA. The molecular constants of the first spectrum, like rotational constants, origin frequency, and Lorentzian width were set fixed, while the global parameters for the complete fit like temperature(s), weights, relative intensity of both spectra etc. were allowed to vary. The resulting rotational constants for 7AI[ND] for both electronic states are listed in table 3.

Higher deuteration grades were obtained by three times refluxing 7AI with an excess of DCl in D_2O (38%). This resulted in a 50 : 50 mixture of mono- and bideuterated species. Figure 8 shows the resulting high-resolution spectrum. Two new bands appeared to the blue of the two isotopomers already described. From a mass spectrum of the deuterated substance we know, that the highest deuteration grade was d_2. Thus, one of the new bands belongs to a d_2-isotopomer, the other one to a d_1-isotopomer, distinct from 7AI[ND]. We fitted the complete spectrum with the parameters of the first two bands kept fixed to the above determined parameters. The GA succeeded in

Table 3. Molecular parameters of the electronic origin band of 7-Azaindole as obtained from the genetic algorithm fit. All values are given in MHz.

	7 – AI	7 – AI[ND]	7 – AI[CD]	7 – AI[ND][CD]
A''	3928.93(2)[a]	3807.60(3)[a]	3794.95(60)	3674.46(24)
B''	1702.629(3)[a]	1684.722(2)[a]	1678.73(16)	1662.45(15)
C''	1188.128(5)[a]	1168.241(2)[a]	1164.10(8)	1144.89(5)
ν_0 [b]	0	27553.67(12)	51934.78(98)	80640.38(39)
ΔA	−183.47(11)	−173.77(6)	−172.11(10)	−162.50(8)
ΔB	1.24(5)	−0.48(3)	−0.62(4)	−0.07(5)
ΔC	−16.62(3)	−16.95(2)	−16.89(7)	−16.21(1)

[a] Values for the electronic ground states from [70].
[b] relative to the electronic origin of 7AI at 34630.74 cm^{-1}.

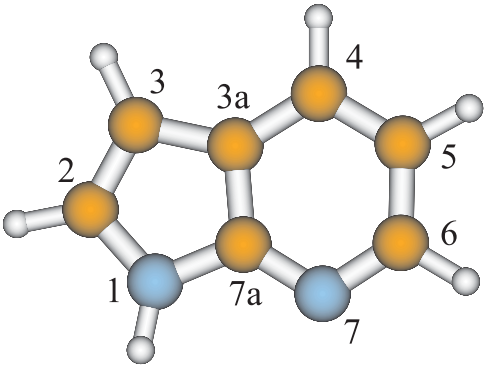

Figure 7. [Colour online] Structure and atomic numbering of 7-azaindole (Pyrrolo[2,3-b]pyridine).

finding the rotational constants for the other two isotopomers. Their values are given in table 3. A comparison to the predicted rotational constants, obtained from the *ab initio* structure [71] for all possible singly and doubly deuterated isotopomers showed immediately, that only 1*H*-Pyrrolo[2,3-b]pyridine-3-*d*1*H*-Pyrrolo[2,3-b] pyridine-1,3-d_2 (7AI[ND][CD]) match the experimentally determined rotational constants.

5.2.2. Resorcinol (1,3-dihydroxybenzene). The rovibronic spectra of the electronic origins of different resorcinol isotopomers present another example of overlapping spectra, whose parameters have to be determined simultaneously. Furthermore, this system serves as an example for a great number of spectra, that probably might also be fit using classical line position assigned fits, but with a much larger effort and time consumption. Resorcinol forms two conformers, which differ in the relative orientation of the two hydroxy groups (cf. figure 9) and are labelled as resorcinol A and B.

Due to the presence of the two hydroxy groups in meta-position, the acidity of nearly all aromatic CH groups is strongly enhanced. Therefore, deuteration leads to a large variety of different isotopomers. For each A and B conformer 10 different isotopic

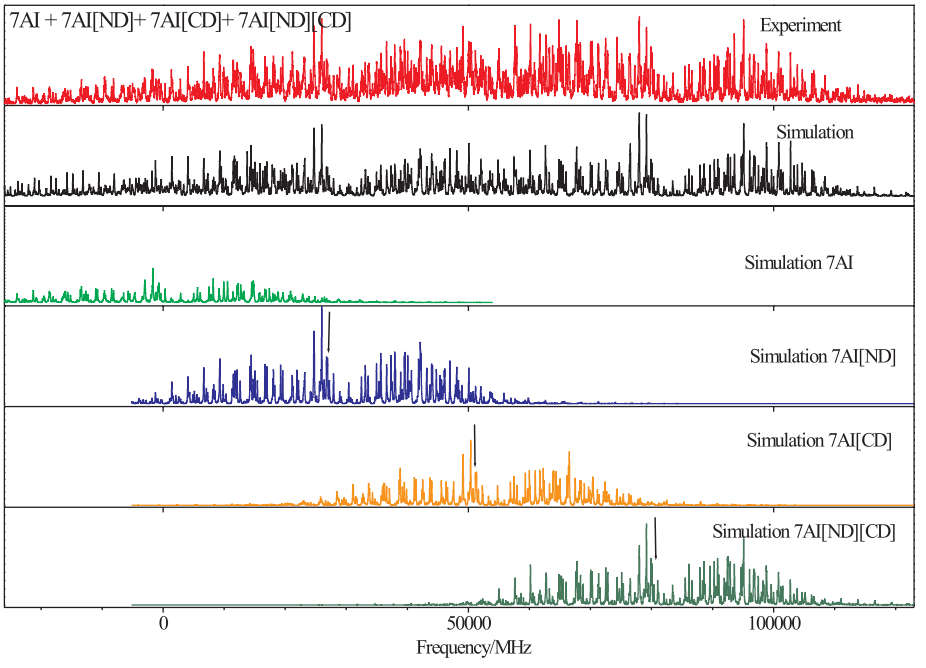

Figure 8. [Colour online] Rotationally resolved LIF spectrum of 7AI, 7AI[ND], 7AI[CD], and 7AI[ND][CD]. The upper trace gives the experimental spectrum, the second trace the simulation, using the best fit parameters. The other show simulations of the individual spectra of 7AI, 7AI[ND], 7AI[CD], and 7AI[ND][CD]. They are given only for reason of clarity, the fit has been performed using the overall spectrum. The frequency scale is relative to the origin of 7AI. The electronic origins of the other isotopomers are marked by arrows.

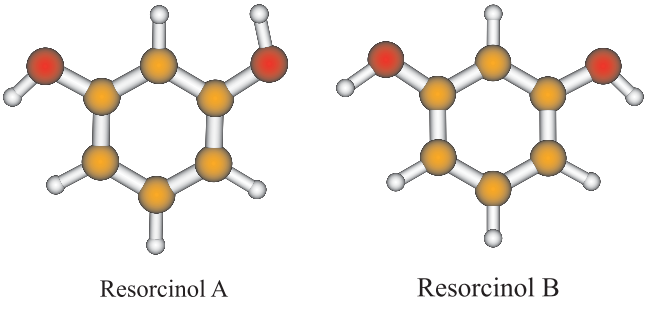

Figure 9. [Colour online] Structures of the two rotamers of resorcinol.

species have been identified, many of them partially or completely spectrally overlapping. Figure S3 shows an example of three different overlapping isotopomers of the B conformer. The manual fit of such a large number of spectra is a tedious and tiring procedure. Using the automated GA fit, we succeeded in finishing the fits and assignments of 20 spectra in only one week [26]. For the first time an unambiguous assignment of all observed isotopic species could be given, that show a nearly perfect

additivity of the zero-point energy shifts of the individual origin bands. Furthermore, on the basis of the rotational constants of 10 isotopomers is was possible to determine the *pseudo* r_s-structures [56] of both rotamers of resorcinol in their S_0 and S_1 electronic states.

5.2.3. Conformers and isotopomers of tryptamine. Conformation of molecules plays an important role in reactions of organic and biochemically relevant molecules. Still, little is known about the effect of solvents on conformations, in particular in small systems. The conformational landscape even of small aromatic molecules with an ethylamine side chain like tryptamine comprises many structurally different, but nearly equally stable conformations. Tryptamine possesses 27 low-energy conformers from three three-fold rotations in and of the alkylamino side chain (cf. figure 10): two around the C-C single bonds and one around the C-N bond. In nine of these 27 conformers the $C_\alpha H_2$–NH_2 tail of the alkylamino group is oriented perpendicular to the aromatic plane. These nine conformers have been calculated to have considerably lower energies than the remaining ones [72].

Pioneering experimental work on the different conformers of tryptamine was performed in a molecular beam using rotationally resolved LIF spectroscopy by Philips and Levy at a spectral resolution of $0.07\,cm^{-1}$ [14]. The triply deuterated conformers were investigated by Wu and Levy [73] of the same group. The conformers were analyzed using the rotational constants of undeuterated and deuterated tryptamines. Seven different conformers were identified on the basis of rotational contours and were named *A*, *B*, *C*(1), *C*(2), *D*, *E* and *F* [74]. A conformational analysis of tryptamine was reported from rotational coherence spectra taken by Connell *et al.* [75]. Caminati published the microwave structure of the *A* and *B* conformer of tryptamine and made an assignment based on *ab initio* calculations [76]. A more thorough investigation of all seven conformers was given by Carney and Zwier based on resonant ion dip infrared (RIDIR) spectra in the region of the CH alkyl stretch vibrations and by UV-UV hole-burning spectroscopy [72]. Rotationally resolved electronic spectra of all

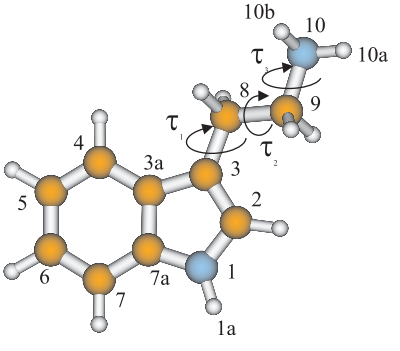

Figure 10. [Colour online] Atomic numbering and definitions of the geometry parameters used in the fit. τ_1 is the dihedral angle defined by the atoms C_2, C_3, C_8, C_9, τ_2 for C_3, C_8, C_9, N_{10} and τ_3 for the dihedral angle C_8, C_9, N_{10}, H_{10b}, where H_{10b} designates the hydrogen atom that is closest to the indole ring.

seven conformers have been taken in the group of Pratt [77]. Compared to the results presented here, one additional conformer (C(2)) has been found. The assignment to the conformers differs for two of the anti-conformers. Furthermore a splitting of 95 MHz in two of the observed spectra was identified by Nguyen *et al.*, that was traced back to a hindered internal motion, that converts two anti-conformers into each other [77].

Carney and Zwier [72] proposed a schematic nomenclature of the nine tryptamine conformers with the $C_\alpha H_2$–NH_2 tail out of plane, based on the amino group positions relative the indole ring and the orientations of the amino group lone pair. Figure 10 shows the atomic numbering, used for designation of the isotopomers and the clusters used in this section. Recently, Dian *et al.* measured directly the energy thresholds between the different conformers of tryptamine using stimulated emission pumping-hole filling and stimulated emission pumping-induced population transfer spectroscopy [78].

Since three rotational constants are not sufficient to give an unambiguous assignment to which conformation band *A* belongs, Schmitt *et al.* [25] recorded the spectra of several deuterated isotopomers. If all chemically possible deuteration grades are formed, the spectrum of each conformer consists of the undeuterated, three different singly deuterated, three different doubly deuterated, and one triply deuterated isotopomer. With the atomic numbering given in figure 10 these eight isotopomers are: tryptamine, [1b-D]tryptamine-d_1, [10a-D]tryptamine-d_1, [10b-D]tryptamine-d_1, [1b-D][10a-D]-tryptamine-d_2, [1b-D][10b-D]tryptamine-d_2, [10a-D][10b-D]tryptamine-d_2, and [1b-D][10a-D][10b-D]tryptamine-d_3. For the *A* conformer, we were only able to observe seven out of the eight possible isotopomers. Using the mixed deuteration grades, a serious problem arises for the assignment of single rovibronic lines, due to strong overlap of the different isotopomeric origins. Figure 11 shows the experimental spectrum of the seven isotopomers of the *A* conformer together with the fit of all bands and the individual results for each origin band.

The resulting inertial parameters of all isotopomers are given in table 2 of Schmitt *et al.* [25] and will be used in this section for the determination of the exact conformation of each conformer. Since seven different isotopomers contribute a total of 21 rotational constants, sufficient information for an unambiguous assignment is present.

Comparison of the rotational constants of the undeuterated tryptamine conformers with the results of *ab initio* calculations gives a first clue to the assignment to a specific structure [25]. Table 4 compares the rotational constants of the optimized MP2/6-311G(d,p) structures with the experimentally obtained rotational constants. The best agreement is found by relating band *A* to the Gpy(out) conformer, band *B* to Gpy(up), band *D* to Anti(py), band *C* to Gph(out), band *E* to Anti(up) and band *F* to Gph(up). This assignment agrees with the findings of Carney and Zwier, but is still not able to remove their ambiguity in the distinction of the Anti(ph) and Anti(py) structures. Therefore, the information has to be combined with the rotational constants of the deuterated isotopomers. Figure 12 shows the six observed conformers.

A structural fit was performed using the program pKrFit in the r_0-structure approximation [25]. The assignment of the different isotopomers of the *A* conformer is very straightforward, since the differences between calculated and experimental rotational constants are small. Starting with the known isotopomers h_3 and d_3, it was possible to assign all other isotopomers. In all cases, the agreement between the rotational

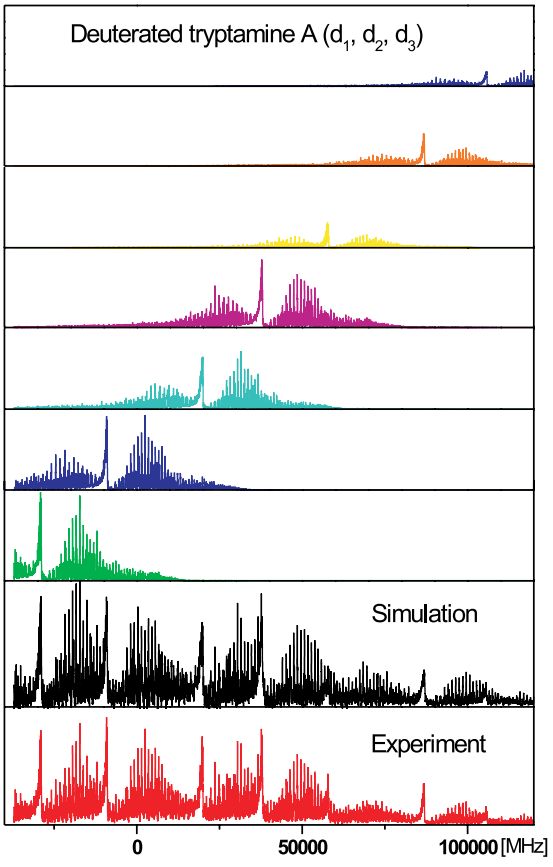

Figure 11. [Colour online] Rovibronic spectrum of the electronic origins of seven isotopomers of the *A* conformer of tryptamine. The top seven spectra are the individual results for each isotopomer origin band. The intensity of the top spectrum has been multiplied by 2 in this figure.

constants of the isotopomers predicted from the *ab initio* structures is stunning (cf. table 4). The three dihedral angles τ_1, τ_2 and τ_3 (cf. figure 10 for definition) have been varied to fit the rotational constants and are displayed in figure 12. With respect to the anti-conformers, Nguyen *et al.* [77] gave different assignments for the *D* conformer. They assigned it to the Anti(Ph) conformer, while the blue component of the *C* band was assigned to the Anti(py) conformer. Under our experimental conditions no second band in the region of *C* could be observed.

5.3. *Strongly congested spectra in dimer systems*

Symmetrically bound homodimers raise the question if the electronic excitation is localized in one of the monomer moieties or if it is completely delocalized over both monomers. Closely related is the question whether the complex conserves the inversion symmetry also in its excited state, leading to an exciton splitting, or if the symmetry is

Table 4. Comparison of rotational constants of the MP2/6-311G(d,p) optimized structures of seven conformers of tryptamine with the experimentally determined rotational constants.

	MP2/6-311G(d,p)						
	Gpy(out)	Gpy(up)	Anti(Py)	Gph(out)	Anti(Ph)	Anti(up)	Gph(up)
$E_{rel.}$(kJ/mol)	0.4083	0.5445	6.8644	0	6.8143	5.6123	1.3709
A''(MHz)	1729.9	1711.5	1767.0	1578.9	1756.2	1753.2	1588.7
B''(MHz)	687.4	686.8	614.7	772.0	619.1	616.0	753.5
C''(MHz)	556.6	555.9	476.7	564.8	478.3	476.9	566.9
Experiment	A	B	D	C		E	F
A''(MHz)	1731.02	1710.25	1767.74	1594.16		1761.40	1605.11
B''(MHz)	682.04	682.22	617.78	755.84		614.75	737.84
C''(MHz)	551.56	551.10	477.00	561.39		475.58	561.51

lost, leading to localized excitation in one monomer. Prototypes of these symmetrical homodimers are the benzoic acid dimer [68, 79], the salicylic acid dimer [80, 81], the 7-azaindole dimer [82], the 2-pyridone dimer [83–85], the o-cyanophenol dimer [86], the anthranilic acid dimer [87] and the benzonitrile dimer [29]. While for the 7-azaindole dimer and the 2-pyridone dimer delocalized excitation with an exciton splitting is found, the benzoic acid dimer, the fluorobenzoic acid dimer, the o-cyanophenol dimer and the anthranilic acid dimer show a local excitation in only one of the moieties. The strength of electronic coupling between two or more chromophores is crucial in exciton formation for energy harvesting or energy funnelling devices in nature.

5.3.1. Benzoic acid dimer revisited. Carboxylic dimers are well suited systems to study intermolecular proton tunnelling since the constituents in these dimers are connected by two intermolecular hydrogen bonds. The displacement of the two protons occurs in concert, and can be described by a single transfer coordinate in a double minimum potential, which is symmetric for an isolated dimer. Of the many carboxylic acid dimers that have been investigated, the benzoic acid dimer is one of the most thoroughly studied. The spectroscopy of this dimer has been investigated in both the crystalline phase [88–97] and the gas phase [68, 79, 98–102].

The benzoic acid dimer has been measured with by high-resolution UV spectroscopy by Remmers *et al*. [68] and is shown in figure S4. Assignments of individual transitions could not be performed, because the spectra were highly congested. In order to deduce the relevant structural information from the experimental spectrum Remmers *et al*. investigated the autocorrelation of the spectrum calculated from the correlation theorem [103]. The experimental spectrum is correlated with itself and the area of overlap between the two spectra is calculated as a function of the relative shift of the two spectra [11]. The autocorrelation spectrum, given in figure 13, shows the central

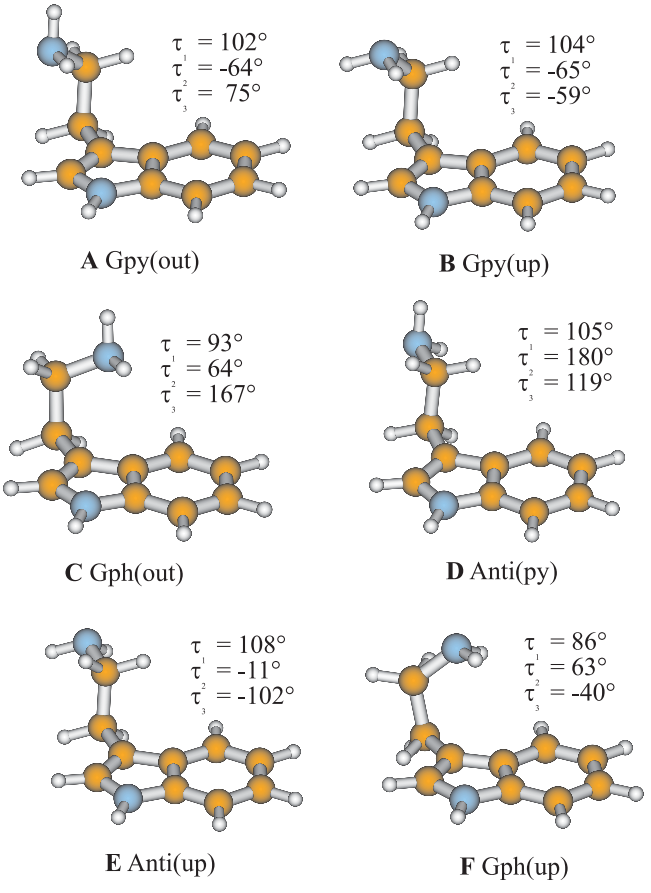

Figure 12. [Colour online] Structure of the six tryptamine conformers identified in the rovibronic analysis of the present publication. Gpy represents conformers in which the amino group position is *gauche* to the pyrrole side of the indole ring, Anti stands for a position *anti* to the indole ring, and Gph for a position *gauche* to the phenyl side. The orientations of the amino lone pair are designated by 'out' and 'py', depending on the direction of the lone pair of the amino group with respect to the rings (±60°) or 'up' (180°). This nomenclature follows the suggestion of Carney and Zwier [72].

maximum at zero frequency shift, where the spectra overlap exactly, and a maximum at ±1107 MHz. This showed that the spectrum could be identified as a composite of two components. The frequency separation between corresponding lines in the two components, denoted by Δ followed from the autocorrelation spectrum and was determined to be $\Delta = 1107 \pm 7$ MHz. Further investigation of the spectrum showed that both components are of b-type with selection rules $\Delta K_a = \pm 1$.

Aside from the above mentioned characteristics the autocorrelation spectrum showed a series of broad peaks which appear with an approximately constant spacing; see figure 13. Each of these peaks is formed from a particular *K*-stack; that is, it is formed from a series of blended Q-branch transitions originating in ground state levels with the same K_a; these transitions have the same value of ΔK_a. The spacing

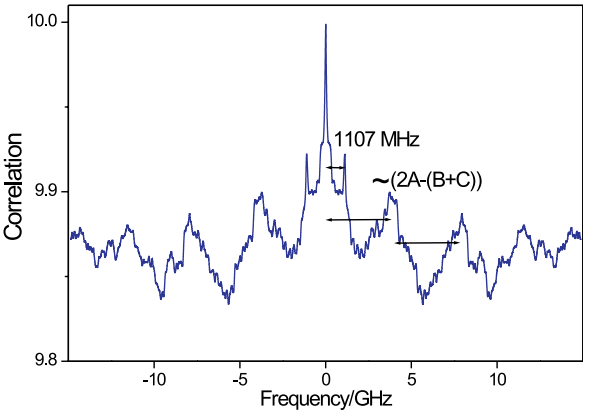

Figure 13. The autocorrelation spectrum of the $S_1 \leftarrow S_0$ transition of the benzoic acid dimer. The spacing between the sharp peaks in the middle is the frequency separation Δ between corresponding transitions in the two different components that make up the spectrum. The separation between neighbouring broad peaks is approximately equal to $[2A - (B + C)]$.

between K-stacks is approximately equal to $[2A - (B + C)]$, but it also depends weakly on ΔA. The width of each stack is determined by $\Delta(B + C)$. Throughout the spectrum the benzoic acid dimer (figure S4), a number of lines could be found with an almost constant spacing corresponding to P and R branch transitions that originate from levels with the same K_a and have the same ΔK_a. The interval between these lines depends on $(B'' + C'')$ and $\Delta(B + C)$. Consequently, in spite of the fact that it was not possible to assign individual transitions in the spectrum, a great deal of information could be extracted. By comparing the experimental spectrum with spectra simulated for various values of the molecular parameters, a range of acceptable values was determined. The results for the rotational constants are given in table 5.

Because of the recent interest in the symmetrically bound homodimers and the questions about local or delocalized bounding in the excited state of these complexes we decided to perform a GA fit of the spectrum from Remmers *et al.* [68]. The quality of the fit was excellent; similarly as obtained for the benzonitrile dimer shown in figure 15. The parameters which resulted from this fit are presented in table 5. The new, more accurate results agree within the uncertainties with those of Remmers *et al.* The GA fit also allowed a determination of the transition dipole moment. It firmly confirmed that within the experimental error the $S_1 \leftarrow S_0$ transition in the benzoic acid dimer is purely of b-type. The structure in the S_0-state, recalculated with the constants from table 5 using the program pKrFit is shown in figure 14. The geometry parameters that have been fit to the rotational constants are the OO distance, and the OHO angle. Additionally the centre of mass (CM) distance of the two monomer moieties has been determined using the relation:

$$R = \sqrt{\frac{\sum_g I_g^{\text{Dimer}} - \sum_g I_g^{\text{Monomer}_1} - \sum_g I_g^{\text{Monomer}_2}}{2\mu}} \tag{34}$$

Table 5. A comparison of the derived molecular parameters of the benzoic acid dimer in the ground and electronically excited state determined in [68] and from a recent GA fit. All values are given in MHz except (ϕ, θ), which are in degrees.

		Remmers *et al.* [68]	Current GA fit
S_0-state	A''	1923(16)	1926.87(50)
	B''	127(8)	128.19(10)
	C''	114(8)	120.27(10)
S_1-state	ΔA	−15(10)	−16.17(20)
	ΔB		−0.52(1)
	ΔC		−0.51(1)
	$\Delta(B+C)$	−1.5(7)	−1.03(2)
	Δ	1107(7)	1116(3)
	θ		86(2)
	ϕ		88(3)

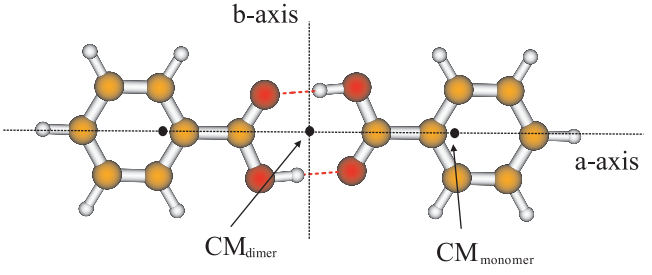

Figure 14. [Colour online] The structure of the benzoic acid dimer in the electronic ground state.

where μ is the reduced mass of the monomer moieties, and the I_g are the respective moments of inertia, described by their superscripts, which are calculated from the experimental rotational constants. For the electronic ground state the microwave inertial parameters from [104] have been taken for the benzoic acid monomers. Additionally, the CM distance of the monomer moieties obtained from the *ab initio* calculations are given for comparison. For the monomer moieties in the structure fit, we used the MP2/6-31G(d,p) geometry, that has rotational constants very close to the MW values. The calculated OO distance and OHO angle are strongly correlated, nevertheless, all runs with the GA converged to the same minimum.

For the electronically excited state no monomer experimental rotational constants are available. Therefore, some assumption on the geometry changes upon excitation of the monomer are needed. Owing to the same level of approximation of CIS calculations for the excited state and of HF calculations for the electronic ground state, quite reasonable geometry changes are obtained from the difference of these calculations. Adding these geometry differences to the ground state geometry results in an excited state geometry, that is generally close to the experimentally determined one (if available). This procedure yields the following values for the rotational constants $A = 3857\,\text{MHz}$, $B = 1239\,\text{MHz}$, and $C = 943\,\text{MHz}$. Under the assumption

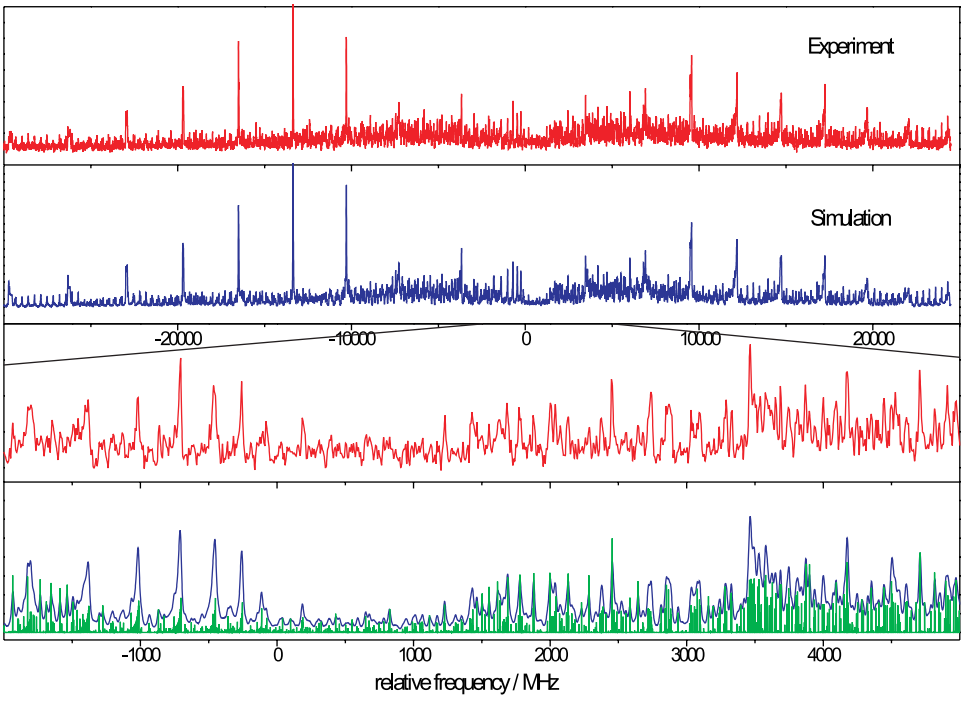

Figure 15. [Colour online] Rovibronic spectrum of the electronic origin of the benzonitrile dimer at 36 420.10 cm^{-1}. The traces show the experimental spectrum, the simulated spectrum using the rotational constants from table 7, a zoomed portion of the experimental spectrum and the zoomed simulation together with the stick spectrum.

Table 6. Comparison of calculated and experimental rotational constants, O−O distance and CM distance of the monomer moieties in the benzoic acid dimer. All calculations were performed using the 6-31G(d,p) basis set.

	S_0			S_1		
	Exp.	MP2	B3LYP	Exp.	CIS	CIS-HF
A''(MHz)	1926.87	1919	1925			
B''(MHz)	128.19	127	128			
C''(MHz)	120.27	119	120			
A'(MHz)				1910.70	1952	
B'(MHz)				127.67	125	
C'(MHz)				119.76	117	
ΔA(MHz)				−16.17		−32
ΔB(MHz)				−0.52		5
ΔC(MHz)				−0.51		4
d(OO)(pm)	261.5	267.8	262.4	263.1	269.3	
a(OHO)(pm)	178.2	178.6	178.7	177.0	179.3	
d(CM)(pm)	714.9	719.8	716.0	718.1	727.5	

of a delocalized excitation these values are used for both monomers and a CM distance of 718.1 pm is found in the excited state of the dimer.

If only one moiety is electronically excited (locally), it will be a good approximation to use the rotational constants of the ground state for one monomer and of the excited state for the other one. This approximation gives a CM distance of 717.6 pm, slightly smaller than from the assumption of delocalized excitation. pKrFit has been used to fit also the intermolecular geometry in the electronically excited state. One benzoic acid moiety has been set to the geometry parameters of the excited state as described above, for the other the ground state MP2 geometry has been used. A slightly larger OO bond length of 261.1 pm (compared to 261.5 pm in the ground state) is obtained. The deviation of the OHO bond from linearity is larger than in the electronic ground state, but due to the large correlation of the parameters, the difference is not significant. The CIS calculation predicts a local excitation, with the average aromatic CC bond lengths of one benzoic acid monomer moiety increased, compared to the ground state, while the second moiety remains in the ground state geometry. As for other systems, the difference of the rotational constants from CIS and HF calculations agree quite well with the difference of the experimental inertial parameters upon electronic excitation due to a favorable cancellation of errors, what is immediately obvious comparing the experimental CM distance and the CIS calculated CM distance.

5.3.2. Benzonitrile dimer. Like the dimer of benzoic acid, the benzonitrile dimer is supposed to have a planer arrangement of the monomer moieties. The question arises if the monomer moieties are equivalent in the ground state and in the electronically excited state [29]. The dimer has been investigated using laser induced fluorescence spectroscopy by Kobayashi *et al*. [105]. Using a rotational contour, they deduced a planar structure of the benzonitrile dimer. Takayanagi *et al*. investigated the relaxation of vibronically excited benzonitrile and the benzonitrile dimer by fluorescence dip spectroscopy [106]. They found slow decay for the monomer and a fast relaxation of the dimer, preventing the vibronic bands of the dimer to be observed.

Figure 15 presents the rovibronic spectrum of the electronic origin of the benzonitrile dimer recently studied by high-resolution UV spectroscopy [29] and the simulation using the molecular parameters from table 7, which have been determined from five independent GA runs. The lowest trace shows a zoomed part of the calculated stick spectrum and its convolution with the experimental Lorentz and Gaussian widths.

The spectrum is fit using the rigid rotor Hamiltonian with the two temperature model. It is of mixed *abc*-type and consists of about 18 000 lines in a range of 55 GHz. At a rotational temperature of about 2.8 K more than 95 J states are populated with an intensity of at least 0.5% of the strongest transition in the spectrum. The rotational constants for both electronic states, the frequency of the electronic origin, the polar angles θ and ϕ, and the Lorentzian contribution to the Voigt profile are given in table 7. The absolute orientation of the transition dipole moment is discussed in section 5.4.

As in the spectrum of the benzoic acid dimer each 'line' in the experimental spectrum is composed of a multitude of single rovibronic transitions, in many cases with very

Table 7. Molecular parameters of the electronic origin band of the benzonitrile dimer as obtained from the genetic algorithm fit. The *ab initio* calculations have been performed using the 6-31G(d,p) basis set.

	Experiment	MP2	HF	CIS
A''(MHz)	1610.99(17)	1654	1559	
B''(MHz)	187.48(7)	185	183	
C''(MHz)	168.12(7)	167	164	
$\Delta I''(u\text{Å}^2)$	−3.2727	0	0	0
$\nu_0(\text{cm}^{-1})$	36420.10(1)			
$\phi(^\circ)$	79.6(10)			90.0
$\theta(^\circ)$	58.0(10)			57.9
Δ_{Lorentz}(MHz)	10(2)			
ΔA(MHz)	−28.08(10)			−20[a]
ΔB(MHz)	−0.01(10)			0[a]
ΔC(MHz)	−0.31(10)			0[a]
A'(MHz)	1582.90(20)			1539
B'(MHz)	187.47(12)			183
C'(MHz)	168.12(12)			164
$\Delta I'(u\text{Å}^2)$	−3.4394	0	0	0

[a] Rotational constants (CIS)-rotational constants (HF).

similar intensities. The Lorentzian contribution of 10(2) MHz as obtained from the GA fit corresponds to an excited state lifetime of 16(3) ns. The relative large uncertainty for the lifetime is also a consequence of the large number of overlapping lines.

Both the quite small negative inertial defect (−3.2727 uÅ^2 for the S_0 and −3.4394 uÅ^2 for the S_1 state) and the very small *c*-type (out-of-plane) contribution to the TDM points to a planar structure in both electronic states. The deviation of the inertial defect from zero can be attributed to out-of-plane zero-point vibrational contributions from large amplitude vibrations of the floppy cluster. The inertial defect of the planar benzonitrile monomer, which has no large amplitude motions amounts to 0.085 uÅ^2.

The program pKrFit was used to determine the intermolecular structure of the benzonitrile dimer in the S_0 and S_1-state from the rotational constants. For the monomer ground state structure two different models were used. Model I is the Kraitchman structure, taken from [107]. For model II, the structure of the monomer has been fit to the microwave rotational constants of ten isotopomers of benzonitrile given in [107]. For the excited state all CC bond lengths in the aromatic ring are set to the same value, which is fit. Furthermore the C_1C_7 distance is fit, the C_7N_8 distance is kept fixed at the ground state value, and all CH bond lengths are kept fixed at a value of 107.1 pm which is a typical value for excited state aromatic CH bonds. The average CC bond length in the aromatic ring determined in this way is 142.8 pm, the C_1C_7 distance was determined to 142.7 pm.

Both monomer geometries were then kept fixed in the fit of the intermolecular geometry of the dimer. The CH-N bond lengths in the dimer using model I and II for the monomer moieties differ slightly. For monomer model I a value of 236.46(5) pm is obtained, for model II of 235.80(13) pm. This gives a rough estimate of systematic errors introduced by using different models for the monomer moieties. The resulting structure is shown in figure 16.

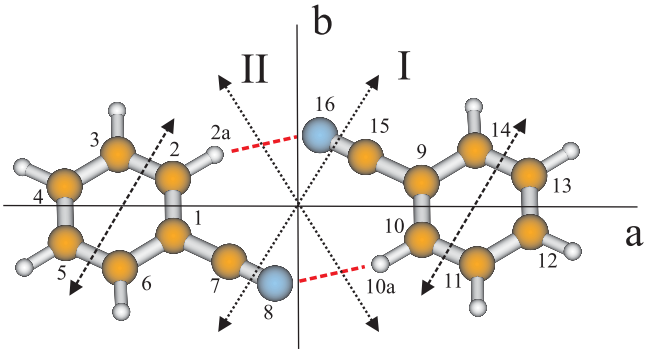

Figure 16. [Colour online] The structure of the benzonitrile dimer and orientation of the inertial axes of the cluster. The dashed double arrows indicate the transition dipole moment orientation in each monomer moiety. The dotted double arrows (I and II) show the two possible orientations of the TDM in the cluster.

5.3.3. Phenol dimer. The phenol dimer is an ideal model system to study the very sensitive equilibrium between hydrogen bonding and dispersion interaction. The main difference to the hitherto described benzoic acid dimer and the benzonitrile dimer is the inequality of the monomer moieties already in the electronic ground state, since one of the cluster constituents acts as proton donor, the other as proton acceptor. Dispersive interactions between the aromatic rings are most likely to play an important role for the relative orientation of the monomer moieties.

Figure 17 shows the rovibronic spectrum of the electronic origin of the phenol dimer together with a simulation using the molecular parameters from table 8 [30]. The lowest trace shows a zoomed part of the stick spectrum along with a convolution using the experimentally determined Lorentzian and Gaussian widths. Like in the benzoic acid dimer and the benzonitrile dimer, each line in the experimental spectrum consists of many overlapping rovibronic transitions with comparable intensities. Again here, the usefulness of the GA based fitting strategy in the automated assignment of such a congested spectrum is obvious. Given the large density and spectral overlap of the rovibronic lines, a classical assignment procedure seems to be very difficult, or even impossible.

The spectrum was fit using the rigid rotor Hamiltonian with the two temperature model. It is of mixed *abc*-type and consists of about 13 000 lines in a range of 50 GHz. At a rotational temperature of about 5 K more than 100 J states are populated with an intensity of at least 0.5% of the strongest transition in the spectrum. The fitted parameters are given in table 8.

Additionally, the rovibronic spectrum of the doubly deuterated isotopomer (d_1–d_1) has been measured. Deuteration positions are 7a and 8a (cf. figure 18). Due to incomplete deuteration, an additional weak band shows up, which can be assigned to a singly deuterated isotopomer, either in 7a (donor) or in 8a (acceptor) position. In principle, there should also be the differently monodeuterated counterpart, but despite intense scanning we could not spot an additional band. The observed singly deuterated isotopomer can be assigned to the donor deuterated species on the basis of the structural fits. Only the donor deuterated isotopomer is observed experimentally

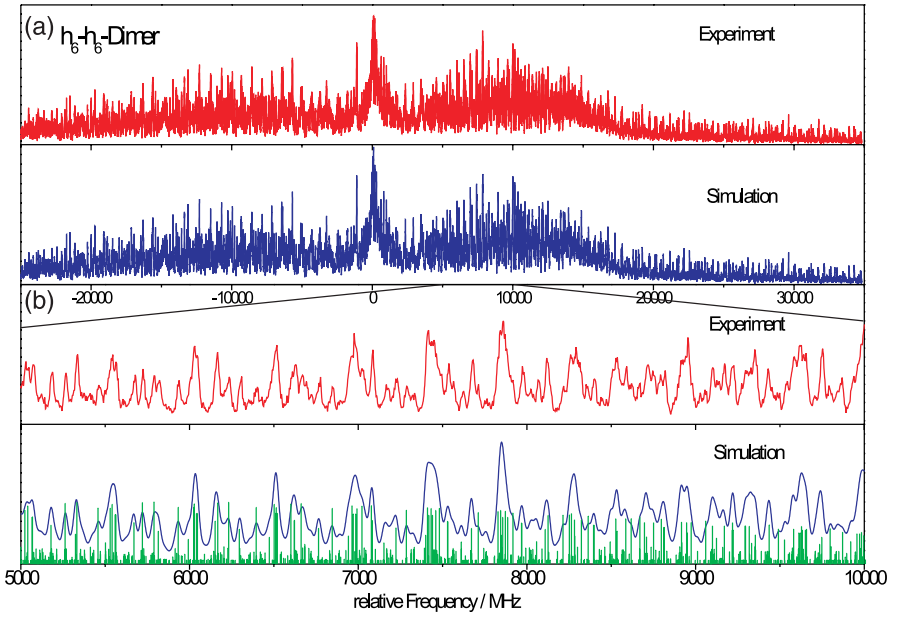

Figure 17. [Colour online] Experimental and fitted rovibronic spectrum of the electronic origin of the phenol dimer at 36 044.70 cm^{-1} using the rotational constants from table 8: (a) full spectrum; (b) 5 GHz zoomed in, together with the stick spectrum.

Table 8. Molecular parameters of the electronic origin band of the phenol dimer as obtained from the genetic algorithm fit.

	h_6-h_6	d_1-d_1	$d_1-h_6^a$	d_6-d_6	$^{13}C-^{13}C$
A''(MHz)	1416.99(39)	1376.23(10)	1399.78(124)	1239.33(12)	1413.22(15)
B''(MHz)	313.51(1)	312.84(3)	312.72(6)	287.25(1)	311.84(2)
C''(MHz)	288.11(1)	286.44(2)	287.16(6)	264.63(1)	286.70(2)
ν_0(cm^{-1})	36044.70(1)	36047.56(1)	36047.98(1)	36217.50(1)	36045.46(1)
ϕ(°)	58.2(30)	55.2(100)	66.6(23)	58.3(20)	57.4(30)
θ(°)	29.5(30)	35.7(100)	37.6(16)	40.0(35)	35.9(60)
$\Delta_{Lorentz}$(MHz)	10(3)	10(3)	10(3)	10(3)	10(3)
ΔA(MHz)	10.71(1)	4.55(1)	7.81(10)	4.28(2)	10.60(1)
ΔB(MHz)	−5.31(1)	−3.95(1)	−5.63(1)	−3.93(1)	−5.25(1)
ΔC(MHz)	−5.82(1)	−4.76(1)	−5.99(1)	−4.90(1)	−5.76(1)

[a] Donor moiety deuterated.

what is in agreement with microwave investigations of the ethylene oxide-water cluster by Caminati *et al.* [108]. They have shown that deuteration in the hydrogen bond (in case of the phenol dimer the donor position) results in a lower zero-point vibrational energy, which favors this isotopomer.

Another isotopomer which changes the rotational constants considerably more than the mono- or bideuterated ones, is the fully deuterated (d_6–d_6) isotopomer.

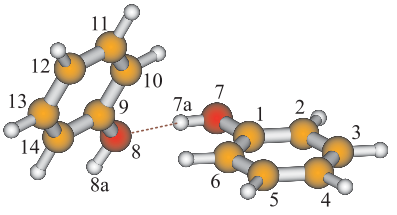

Figure 18. [Colour online] Structure and atomic numbering of the phenol dimer.

Table 9. Parameters that determine the relative intensities in the spectra of 7AI, obtained from a GA fit. For details see text.

	7-AI	7-AI[ND]
$\theta(^\circ)$	21.17(6)	17.29(1)
μ_a^2	0.8910	0.9116
μ_b^2	0.1090	0.0884
Δ_{Lorentz}(MHz)	62.52(30)	67.96(22)

The rovibronic spectrum of the electronic origin at $36\,217.50\,\text{cm}^{-1}$ was measured and fit.

The fifth isotopomer of the phenol dimer which was investigated has the C_1 and C_9 carbon atoms (cf. figure 18) replaced by ^{13}C. The resulting molecular parameters of all five isotopomers are compiled in table 8.

The rotational constants in table 8 have larger standard deviations than we would expect for a molecule with a well-defined Hamiltonian and a spectrum with good signal/noise. Part of this standard deviation results from the large number of overlapping lines, which makes it difficult even for the GA to find a unique solution.

5.4. *Orientation of the transition dipole moment*

Since the GA performs an intensity fit of the complete spectrum, much more accurate information on the orientation of the transition dipole moment is gathered than from a lineshape fit to a few individual lines. We performed an intensity fit to the rovibronic spectrum of 7-azaindole and obtained the values, presented in table 9.

The TDM for a transition to the $\pi\pi^*$ state in a planar molecule is located in the molecular plane. As a check, both angles between TDM and inertial axes, θ and ϕ were fit, resulting in values for μ_c for all isotopomers well below 1%. In all subsequent fits, the model was therefore restricted to *ab*-hybrid type. Only one angle θ is needed in this case, which is defined as angle between the TDM and the inertial *a*-axis. The value of θ of different isotopomers can be used to determine the sign, and therefore the orientation of the TDM. Isotopic substitution, that takes place off-axis results in a rotation of the respective inertial axis, but the orientation of the TDM stays the same.

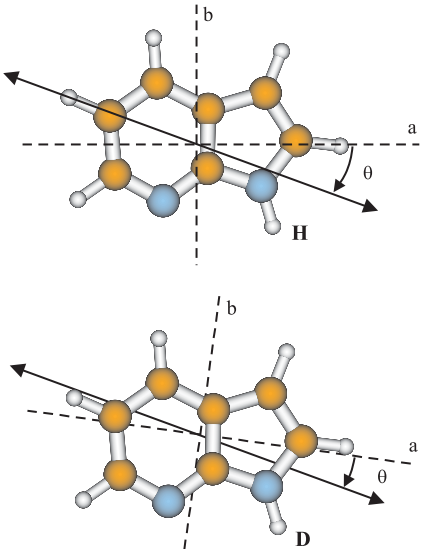

Figure 19. [Colour online] Angle θ between TDM and inertial a-axis for two different isotopomers. The direction shown defines a negative angle. The rotation of the inertial axis system upon deuteration has been exaggerated for reasons of clarity.

If the sign of the angle of the TDM with the inertial a-axis is positive, as given by most of the theoretical studies [109–111], the angle should increase for the ND tautomer and decrease for the CD tautomer, cf. figure 19. A rough estimate of the change of θ for the three isotopomers has been performed by calculating the rotation of the inertial axis system, based on the *ab initio* structure. An estimated value between 1.5° and 2.5° for the a-axis rotation shows the imperative of an exact experimental determination of θ and of a reliable fit. A rotation of 2° changes the relative intensities of a- and b-lines only by 2.5%. This fact again shows that only an intensity fit to all lines in the spectrum as performed by the GA is capable of predicting such tiny changes reliably. A fit to the overlapping spectra of 7AI and 7AI[ND] with the parameters of 7AI kept fixed at the values determined before, resulted in values of ±21.2(9)° for 7AI and of ± 17.3(9)° for 7AI[ND]. The decrease of the angle θ indicates that the absolute angle in 7AI should be –21° in contrast to most theoretical studies, that predict a positive angle.

Another example for the absolute determination of the orientation of the TDM using a combination of rotationally resolved electronic spectroscopy of several isotopomers with the GA based fitting strategy is benzimidazole [28]. In the same manner as described above for 7-azaindole, the absolute TDM orientation of benzimidazole was determined to be -32°. Table 5.4 compares the angles of the TDM vector for the two lowest electronic states, ^1L_b and ^1L_a, from various experiments with the results of electronic structure calculations at different levels of theory.

In benzonitrile dimer the angle of the transition dipole moment with the inertial a-axis upon electronic excitation of the benzonitrile dimer is 58.0° (cf. table 7). Since we only know the projection on the axis, the transition dipole might be oriented in

Table 10. Experimental and calculated TDM orientations of the two lowest excited states of benzimidazole. CASSCF calculations have been performed with the (10,9) space described in the text using the 6-31G(d,p) basis set. The TDDFT calculations have been performed using the B3-LYP functional and the TZVP basis set.

	Exp.[a]	Exp.[b]	CIS[c]	CAS[c]	TDDFT[c]
$S_1(^1L_b)$	−32	−16	−14	+36	+14
$S_2(^1L_a)$		−80	−79	−53	−34

[a] Schmitt *et al.* [28]
[b] From polarized absorption in a stretched polymer [112].
[c] In the optimized geometry of the ground state.

two ways, as shown in figure 16. Orientation I is nearly parallel to the orientation of the transition dipole in the benzonitrile monomer moieties. Under the assumption of unchanged TDM orientation upon cluster formation we favor orientation I over orientation II. The results of the *ab initio* CIS/6-31G(d,p) calculations (table 7) for the planar TDM orientation angle θ of 57.9° agree perfectly with the experimental value of 58°. Since the out-of-plane contribution to the TDM orientation is due to zero-point motions, the angle ϕ is calculated at *ab initio* level to be 90°, while experimentally we find an angle of 79.6°.

5.5. *GA fit of rotational contours*

Finally, we will show that the GA based fitting strategy can also be used in cases where the spectral resolution is so low that only rotational contours can be resolved. This case frequently occurs in electronic spectroscopy, when the excited state lifetime is considerably smaller than 1 ns, leading to large Lorentzian contributions to the line-width. Other limitations might be low experimental resolution due to limited band width of the exciting laser, as encountered frequently in pulsed laser experiments or due to limited spectral resolution of the detection scheme as found, for example in low-resolution FTIR spectroscopy. Pioneering work on the analysis of rotational band contour analysis of vibronic transitions has been done in the group of Hollas [113–117]. Due to the relatively high temperatures in their experiments and the many rotational states occupied, inertial parameters have been extracted in their work, which are of comparable quality to those from electronic spectroscopy with full rotational resolution. But even for cold jets, a band contour analysis of electronic spectra can give sufficient information to distinguish between conformers of flexible biomolecules, as has been shown in the group of Simons [118–119] or to determine connectivities in weakly bound molecular clusters like phenol-N_2 [120] (cf. section 5.1.1 for a fully resolved spectrum).

5.5.1. The FTIR spectrum of benzotriazole. As an example we choose the FTIR spectrum of benzotriazole in the region of the NH stretching vibration. Two tautomers contribute to the IR absorption in the observed region, the 1*H*- and the 2*H*-tautomer, cf. figure 20.

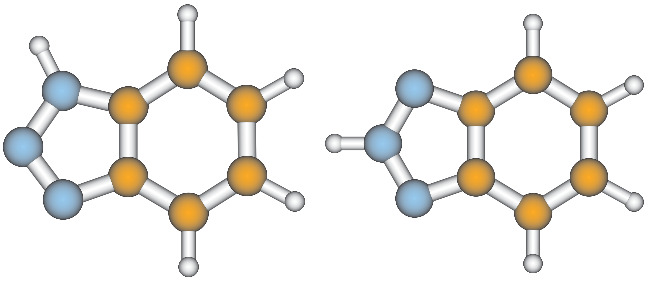

Figure 20. [Colour online] Structure of 1H-benzotriazole (left) and the 2H-tautomer (right).

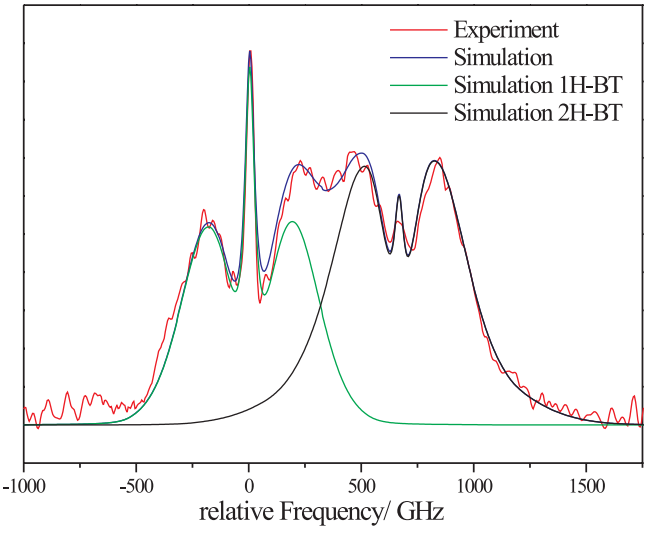

Figure 21. Experimental and simulated spectrum of the NH stretching vibration of 1H-benzotriazole and 2H-benzotriazole.

The analysis of this spectrum has been published in [121], using a classical analysis of the lineshape of this band. The parameters have been adapted manually and only the ratio of both sub-bands was fitted. The GA fit is capable of fitting both vibrations simultaneously, without any parameter restriction. Figure 21 shows the experimental contour and the best fit using the GA strategy. Of course a rotational contour with only very little resolved structure does not allow a fit of all molecular parameters, that determine the spectrum. In such cases additional information from other sources is needed. For benzotriazole, the ground state rotational constants of both tautomers are known from microwave [122] and from electronic spectroscopy [123] and can be fixed in the fit. From the structure of the tautomers in figure 20 it is clear that the 2H-tautomer must be of pure *a*-type (the dipole moment change upon NH vibration occurs purely along the *a*-axis) and the 1H-tautomer must be an *ab*-hybrid.

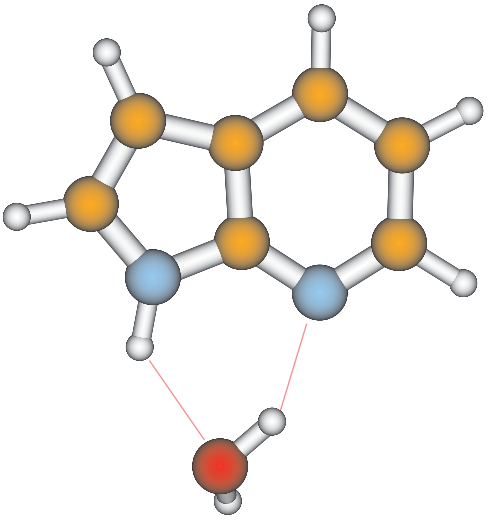

Figure 22. [Colour online] Structure of the hydrogen bonded azaindole-water cluster.

The quality of the fit to the rotational contour has greatly improved compared to the manual fit performed in [121]. Much more exact values for the relative intensities of the stretching vibrations of 1H-benzotriazole and 2H-benzotriazole can be obtained. If the intensity of both bands is recorded as a function of the temperature one can determine the energetic ordering of the tautomers and furthermore estimate their relative energy using the van't Hoff isochore.

5.5.2. The low-resolution LIF spectrum of azaindole-water. Nakajima *et al.* presented the LIF spectrum of the azaindole-water cluster with a resolution of 300 MHz ($0.01\,cm^{-1}$) [124]. This resolution is not sufficient to completely resolve single rovibronic lines. Nevertheless, the band contour could be fit using the GA based fitting technique [125]. Due to the absence of single rovibronic lines, a fit using the assignment of line positions cannot be performed here. Figure 23 shows the experimental contour, and the best simulation using the parameters from table 11.

The search space for the ground state rotational constants has been set to ± 100 MHz around inertial parameters that have been obtained from a preliminary B3LYP calculation. The changes of the rotational constants upon electronic excitation were varied ± 50 MHz symmetrically around zero. The angle θ, was allowed to vary between 0° (pure *a*-type) and 90° (pure *b*-type).

The molecular parameters determined from the GA fit to the contour in figure 23 were used in a subsequent fit of the excited state structure using a combined Franck–Condon and inertial parameter fit [125]. The fit to the ground state rotational constants gave a cyclic structure for the cluster, in which the water is doubly hydrogen bond (cf. figure 22). The geometry changes upon electronic excitation are dominated by a distinct shortening of both hydrogen bonds.

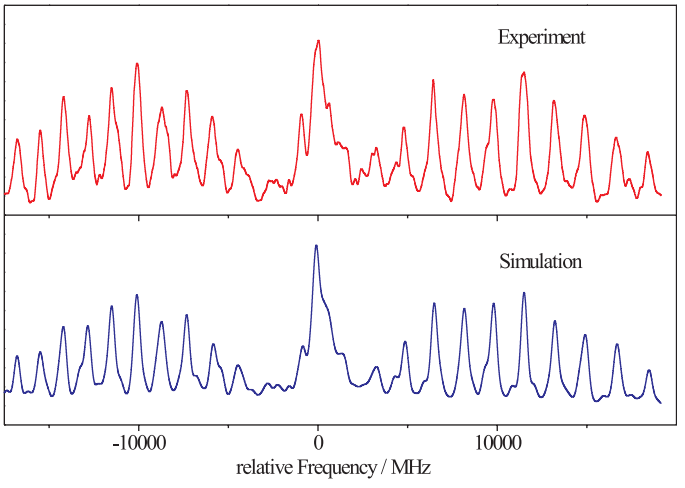

Figure 23. [Colour online] Comparison of the experimental LIF spectrum of the electronic origin of the 7AI-water cluster with the simulation using the molecular parameters from table 11.

Table 11. Molecular constants of 7AI-water from the GA fit of the LIF spectrum of [124], B3LYP prediction of the ground state parameters and search range of the parameters for the GA fit of the contour. All calculations were performed using the 6-31G (d, p) basis set.

	Exp.	B3LYP	Search range
A''(MHz)	1741.7(5)	1775	1675–1875
B''(MHz)	1336.0(4)	1349	1236–1436
C''(MHz)	761.7(2)	768	668–868
$\theta(^\circ)$	±19(3)		0–90
ΔA(MHz)	+33.1(4)		−50–50
ΔB(MHz)	+18.0(3)		−50–50
ΔC(MHz)	+11.3(2)		−50–50

6. Summary

In this paper a method for automatic assignments of complicated and congested spectra was introduced and discussed. The technique is based on the application of genetic algorithms. It has been shown to be able to assign and analyze a wide range of different spectra with complexity ranging from highly overlapping transitions to coinciding spectra of different isotopomers. The GA succeeds in determining the molecular parameters without any prior knowledge of their values.

Even spectra that are life time broadened can successful be analyzed as long as they contain sufficient structure. If the signal to noise ratio is low and/or if only a partial spectrum is available the method is often still successful.

A prerequisite of the success of the GA procedure is based on the existence of a good model for the prediction of the spectra. This seems to be the only drawback until now. However, there are many cases for which a good model prediction exists, in particular, in absorption, cavity ringdown and laser induced fluorescence spectra. Based on our

experience with the method, it is clear that in the case of small and/or local perturbations the main spectral features that conform to the model can still be extracted and hence the perturbations are isolated.

The examples discussed in this work and the references cited demonstrate the power of the GA in the application of automatically assign high-resolution spectra, spectra which can hardly be analyzed with the conventional methods. The computing power of modern PC's is more than adequate to perform the job in an acceptable time. This new technique opens the road to the analysis of the complex spectra of biomolecules and their building blocks.

Acknowledgments

We would like to thank Jos Hageman, Ron Wehrens, Lutgarde Buydens and Gerrit Groenenboom for many helpful discussions. We thank Christian Ratzer, Grzegorz Myszkiewicz, Marcel Böhm, Ivo Kalkman, Chau Vu, Robert Brause, Violetta Bednarska and Marloes van Beek for their contributions to this work and Bert Groenenboom for giving the mathematical proof presented in Appendix A. The financial support of the Deutsche Forschungsgemeinschaft (SCHM 1043/9-4 and SFB663 project A2) is gratefully acknowledged. MS wishes to thank the Nordrheinwestfälische Akademie der Wissenschaften for a grant, which made this work possible. WLM thanks Timothy S. Zwier at Purdue University for his hospitality as this paper was being written. The authors would like to thank the National Computer Facilities of the Netherlands Organization of Scientific Research (NWO) for a grant on the Dutch supercomputing facility SARA. This work was also supported by the Netherlands Organization for Scientific Research and the Deutsche Forschungsgemeinschaft in the framework of the NWO-DFG bilateral programme.

Appendix A: Conditions for the W-matrix to be positive definite

In this appendix we show that the real or complex matrix $\boldsymbol{W}$ with matrix elements $W_{ij} = w(r_i - r_j)$ is positive definite if $w(r)$ can be written as the inverse Fourier transform of a positive function $\tilde{w}(t)$. Also see [20]. Let

$$w(r) = \int_{-\infty}^{\infty} \tilde{w}(t) e^{2\pi i r t} dt \tag{A1}$$

then

$$\begin{aligned} \boldsymbol{x}^{\dagger} \boldsymbol{W} \boldsymbol{x} &= \sum_{i,j} x_i^* W_{ij} x_j \\ &= \sum_{i,j} x_i^* \int_{-\infty}^{\infty} \tilde{w}(t) \exp\left[2\pi i (r_i - r_j) t\right] dt \; x_j \\ &= \int_{-\infty}^{\infty} \left| \sum_i x_i \exp[-2\pi i r_i t] \right|^2 \tilde{w}(t) dt. \end{aligned} \tag{A2}$$

Hence, if $\tilde{w}(t) > 0$ then $\boldsymbol{x}^{\dagger}\boldsymbol{W}\boldsymbol{x} > 0$ for any vector $\boldsymbol{x}$ and so $\boldsymbol{W}$ is positive definite. If $\tilde{w}(t)$ is zero for certain values of t, the integral is still positive. Note: one should read in equation (3) $r_i = i$.

The Fourier transform of $w(r)$ defined in equation (7) is

$$\tilde{w}(t) = \int_{-\infty}^{\infty} e^{-2\pi i r t} w(r)\, dr = \frac{\Delta w}{2}\, \mathrm{sinc}^2\left(\frac{\Delta w}{2}\pi t\right), \tag{A3}$$

where $\mathrm{sinc}(x) = \sin(x)/x$ is, so $\tilde{w}(t) \geq 0$.

Appendix B: The least squares fit algorithm used in the assigned fits

A large number of linear and non-linear least squares fit methods have been developed over the years, see for example [103]. One of the more well known is the Levenberg–Marquardt method. In this appendix we describe the fit method that is used in our assigned fit calculations. We searched for a reliable fast converting method that required the minimum number of model calculations. The number of model calculations for the algorithm described in this section is $n_{\mathrm{conv}}(m+1)+1$, where m is the number of parameters to be fit and n_{conv} the number of iteration needed for a full conversion. The latter value is typically 3 to 5. For the Levenberg–Marquardt method the number of model calculations was at least two order of magnitudes larger. This is of particular importance if a model calculation is very time consuming. Strictly speaking the method we describe here it is not a non-linear algorithm since the search for the best parameters is performed by linearizing the problem around the parameter values found after the n-th iteration. However, by repeatedly recalculation the first derivatives it effectively a non-linear fit is realized.

The method we used is first discussed in [126]. Let $(f_i, i = 1, N)$ be the N experimental data. Their experimental uncertainties (standard deviations) are given by Δf_i. The corresponding calculated data are $(g_i, i = 1, N)$. The quality of the fit is expressed in the χ^2-value:

$$\chi^2 = \sum_{i=1}^{N} \frac{(f_i - g_i)^2}{\Delta f_i^2}. \tag{B1}$$

If we assume that the calculated data depend on m parameters $(\alpha_\mu, \mu = 1, m)$

$$g_i = g_i(\alpha_1, \alpha_2, \ldots, \alpha_m) = g_i(\boldsymbol{\alpha}). \tag{B2}$$

The best fit is defined for the set $\{\boldsymbol{\alpha}\}$ for which χ^2 is minimal. In order to fulfill this condition

$$\frac{\partial \chi^2}{\partial \alpha_\mu} = 2\sum_{i=1}^{N} \frac{f_i - g_i}{\Delta f_i^2} \frac{\partial g_i}{\partial \alpha_\mu} = 0. \tag{B3}$$

Let us assume that there exists an approximate set $\{\boldsymbol{\alpha}^0\}$ with

$$\alpha_\nu = \boldsymbol{\alpha}_\nu^0 + \Delta\alpha_\nu, \tag{B4}$$

for which

$$f_i = g_i(\boldsymbol{\alpha}) = g_i(\boldsymbol{\alpha}^0) - [g_i(\boldsymbol{\alpha}) - g_i(\boldsymbol{\alpha}^0)]. \tag{B5}$$

Here $[g_i(\boldsymbol{\alpha}) - g_i(\boldsymbol{\alpha}^0)]$ and $\Delta\alpha_\nu$ are assumed to be small. Equation B3 can be expanded to

$$0 = \sum_{i=1}^{N} \frac{1}{\Delta f_i^2} \left[f_i - g_i(\boldsymbol{\alpha}^0) - \sum_{\mu=1}^{m} \left(\frac{\partial g_i}{\partial \alpha_\mu} \right)_{\{\alpha^0\}} \Delta\alpha_\mu \right]$$
$$\times \left[\left(\frac{\partial g_i}{\partial \alpha_\nu} \right)_{\{\alpha^0\}} + \frac{1}{2} \sum_{\mu=1}^{m} \left(\frac{\partial^2 g_i}{\partial \alpha_\nu \partial \alpha_\mu} \right)_{\{\alpha^0\}} \Delta\alpha_\mu \right]. \tag{B6}$$

Because $f_i - g_i(\boldsymbol{\alpha}^0)$ and $\Delta\alpha_\mu$ are small, terms with $[f_i - g_i(\boldsymbol{\alpha}^0)]\Delta\alpha_\mu$ and $\Delta\alpha_\mu \Delta\alpha_\nu$ are neglected and the expression in B6 can be reduced to:

$$\sum_{i=1}^{N} \frac{1}{\Delta f_i^2} \left(\frac{\partial g_i}{\partial \alpha_\nu} \right)_{\{\alpha^0\}} \sum_{\mu=1}^{m} \left(\frac{\partial g_i}{\partial \alpha_\mu} \right)_{\{\alpha^0\}} \Delta\alpha_\mu$$
$$= \sum_{i=1}^{N} \frac{1}{\Delta f_i^2} \left[f_i - g_i(\boldsymbol{\alpha}^0) \right] \left(\frac{\partial g_i}{\partial \alpha_\nu} \right)_{\{\alpha^0\}}. \tag{B7}$$

The best set of parameters is obtained by solving the set of inhomogeneous linear equations:

$$\boldsymbol{A}\boldsymbol{\Delta}\alpha = \boldsymbol{B}. \tag{B8}$$

Here the elements of the matrix $\boldsymbol{A}$ are

$$A_{\nu\mu} = \sum_{i=1}^{N} \frac{1}{\Delta f_i^2} \left(\frac{\partial g_i}{\partial \alpha_\nu} \right)_{\{\alpha^0\}} \left(\frac{\partial g_i}{\partial \alpha_\mu} \right)_{\{\alpha^0\}}, \tag{B9}$$

of the matrix $\boldsymbol{B}$

$$b_\nu = \sum_{i=1}^{N} \frac{1}{\Delta f_i^2} \left[f_i - g_i(\boldsymbol{\alpha}^0) \right] \left(\frac{\partial g_i}{\partial \alpha_\nu} \right)_{\{\alpha^0\}} \tag{B10}$$

and

$$\boldsymbol{\Delta\alpha} = (\Delta\alpha_1, \Delta\alpha_2, \ldots \Delta\alpha_m)^T. \tag{B11}$$

Equation (B8) can be rewritten to:

$$\boldsymbol{\Delta\alpha} = \boldsymbol{CB}. \tag{B12}$$

Here $\boldsymbol{C}$ is the covariance matrix and equals $\boldsymbol{A}^{-1}$. From equations (B10)–(B12) it follows that we can write:

$$\begin{aligned}
\Delta\alpha_\nu &= \sum_{\mu=1}^{m} C_{\nu\mu} b_\mu \\
&= \sum_{\mu=1}^{m} C_{\nu\mu} \sum_{i=1}^{N} \frac{1}{\Delta f_i^2} \left[f_i - g_i(\boldsymbol{\alpha}^0)\right] \left(\frac{\partial g_i}{\partial \alpha_\mu}\right)_{\{\alpha^0\}} \\
&= \sum_{i=1}^{N} \left[f_i - g_i(\boldsymbol{\alpha}^0)\right] \frac{1}{\Delta f_i^2} \sum_{\mu=1}^{m} C_{\nu\mu} \left(\frac{\partial g_i}{\partial \alpha_\mu}\right)_{\{\alpha^0\}} \\
&= \sum_{i=1}^{N} \lambda_{i\nu} l_i
\end{aligned} \tag{B13}$$

with

$$\lambda_{i\nu} = \frac{1}{\Delta f_i^2} \sum_{\mu=1}^{m} C_{\nu\mu} \left(\frac{\partial g_i}{\partial \alpha_\mu}\right)_{\{\alpha^0\}} \tag{B14}$$

and

$$l_i = \left[f_i - g_i(\boldsymbol{\alpha}^0)\right]. \tag{B15}$$

From equation (B13) it follows that $\Delta\alpha_\nu$ can be written as a linear combination of the random variables $\{l_i\}$. It then can be shown [127, 128] that $\Delta\alpha_\nu$ is again a random variable drawn from a population whose *statistical* standard deviation is given by:

$$\sigma_\nu^2 = \sum_{i=1}^{N} \lambda_{i\nu}^2 \Delta l_i^2 + 2[\rho_{12}\lambda_1\lambda_2\Delta l_1 \Delta l_2 + \cdots]. \tag{B16}$$

Combining this with equation (B15) $\Delta l_i = \Delta f_i$, while ρ_{ij} is the correlation coefficient between l_i and l_j. We assume here that $\rho_{ij} = 0$.

$$\begin{aligned}
\sigma_\nu^2 &= \sum_{i=1}^{N} \lambda_{i\nu}^2 \Delta f_i^2 \\
&= \sum_{i=1}^{N} \frac{1}{\Delta f_i^2} \sum_{\mu=1}^{m} C_{\nu\mu} \left(\frac{\partial g_i}{\partial \alpha_\mu}\right)_{\{\alpha^0\}} \sum_{\lambda=1}^{m} C_{\nu\lambda} \left(\frac{\partial g_i}{\partial \alpha_\lambda}\right)_{\{\alpha^0\}} \\
&= \sum_{\mu=1}^{m} \sum_{\lambda=1}^{m} C_{\nu\mu} C_{\nu\lambda} \sum_{i=1}^{N} \frac{1}{\Delta f_i^2} \left(\frac{\partial g_i}{\partial \alpha_\lambda}\right)_{\{\alpha^0\}} \left(\frac{\partial g_i}{\partial \alpha_\mu}\right)_{\{\alpha^0\}} \\
&= \sum_{\mu=1}^{m} C_{\nu\mu} \sum_{\lambda=1}^{m} C_{\nu\lambda} A_{\lambda\mu} \\
&= \sum_{\mu=1}^{m} C_{\nu\mu} \delta_{\nu\mu} \\
&= C_{\nu\nu}.
\end{aligned} \tag{B17}$$

Hence the best fit parameters from equation (B4) that minimize χ^2 can be derived from equations (B4) and (B12). This calculation represents one iteration. In all practical cases the calculation should be repeated with first derivative values ($\partial g_i / \partial \alpha_\nu$) obtained at the parameter values from the previous iteration.

The calculated *statistical* standard deviations (σ_ν) in each of these parameters is given by:

$$\sigma_\nu = \sqrt{\sum_{i=1}^{N} \lambda_{i\nu}^2 \Delta f_i^2} = \sqrt{C_{\nu\nu}}. \tag{B18}$$

The correlation coefficient between the parameters ν and μ is defined as:

$$V_{\nu\mu}^{(cor)} = \frac{1}{\sigma_\nu \sigma_\mu} \sum_{i=1}^{N} \lambda_{i\nu} \lambda_{i\mu} \Delta f_i^2 = \frac{1}{\sigma_\nu \sigma_\mu} C_{\nu\mu}. \tag{B19}$$

The proof of the last equality is similar to that given in equation (B17). The standard deviation S and the variance σ of the fit are defined as:

$$S = \sqrt{\frac{\chi^2}{\sum_{i=1}^{N} \left[1/\Delta f_i^2\right]}} \tag{B20}$$

and

$$\sigma = \sqrt{\frac{\chi^2}{N-m}}\,. \tag{B21}$$

The standard deviation in parameter α_ν corrected for *non-statistical* errors is $\sigma\sigma_\nu$. Type of non-statistical uncertainties are for example errors in the model used to fit the data or an under/over estimation of the experimental errors Δf_i.

References

[1] W. A. Majewski and W. L. Meerts, J. Mol. Spec. **104,** 271 (1984).
[2] F. W. Loomis and R. W. Wood, Phys. Rev. **32,** 223 (1928).
[3] J. F. Scott and K. N. Rao, J. Mol. Spec. **20,** 461 (1966).
[4] B. P. Winnewisser, J. Reinstädtler, K. M. T. Yamada, and J. Behrend, J. Mol. Spec. **136,** 12 (1989).
[5] C. F. Neese, *International Symposium on Molecular Spectroscopy, 56th Meeting*, Columbus, Ohio, Ohio State University, June 11–15 (2001).
[6] C. D. Thompson, E. G. Robertson, and D. McNaughton, Phys. Chem. Chem. Phys. **5,** 1996 (2003).
[7] R. Brotherus, J. Comp. Chem. **20,** 610 (1999).
[8] Z. Kisiel, L. Pszczolkowski, I. R. Medvedev, M. Winnewisser, F. Lucia, and C. E.Herbst, J. Mol. Spec. **233,** 231 (2005).
[9] D. F. Plusquellic, R. D. Suenram, B. Mate, J. O. Jensen, and A. C. Samuels, J. Chem. Phys. **115,** 3057 (2001).
[10] G. Moruzzi, J. Mol. Spec. **229,** 19 (2005).
[11] R. M. Helm, H.-P. Vogel, and H. J. Neusser, Chem. Phys. Lett. **270,** 285 (1997).
[12] C. A. Haynam, D. V. Brumbaugh, and D. H. Levy, J. Chem. Phys. **81,** 2282 (1984).
[13] L. A. Philips and D. H. Levy, J. Chem. Phys. **85,** 1327 (1986).
[14] L. A. Philips and D. H. Levy, J. Chem. Phys. **89,** 85 (1988).
[15] J. A. Hageman, R. Wehrens, R. de Gelder, W. L. Meerts, and L. M. C. Buydens, J. Chem. Phys. **113,** 7955 (2000).
[16] G. Berden, W. L. Meerts, and E. Jalviste, J. Chem. Phys. **103,** 9596 (1995).
[17] G. Berden, J. van Rooy, W. L. Meerts, and K. A. Zachariasse, Chem. Phys. Lett. **278,** 373 (1997).
[18] I. Szydłowska, G. Myszkiewicz, and W. L. Meerts, Chem. Phys. **283,** 371 (2002).
[19] M. Schmitt, C. Ratzer, and W. L. Meerts, J. Chem. Phys. **120,** 2752 (2004).
[20] W. L. Meerts, M. Schmitt, and G. Groenenboom, Can. J. Chem. **82,** 804 (2004).
[21] M. Schmitt, C. Ratzer, K. Kleinermanns, and W. L. Meerts, Mol. Phys. **102,** 1605 (2004).
[22] A. E. Nikolaev, G. Myszkiewicz, G. Berden, W. L. Meerts, J. F. Pfanstiel, and D. W. Pratt, J. Chem. Phys. **122,** 084309-1-10 (2005).
[23] M. Schmitt, C. Ratzer, C. Jacoby, and W. L. Meerts, J. Mol. Struc. **742,** 123 (2005).
[24] G. Myszkiewicz, W. L. Meerts, C. Ratzer, and M. Schmitt, Phys. Chem. Chem. Phys. **7,** 2142 (2005).
[25] M. Schmitt, M. Böhm, C. Ratzer, C. Vu, I. Kalkman, and W. L. Meerts, J. Am. Chem. Soc. **127,** 10356 (2005).
[26] G. Myszkiewicz, W. L. Meerts, C. Ratzer, and M. Schmitt, Chem. Phys. Chem. **6,** 2129 (2005).
[27] G. Myszkiewicz, W. L. Meerts, C. Ratzer, and M. Schmitt, J. Chem. Phys. **123,** 044304-1 (2005).
[28] M. Schmitt, D. Krügler, M. Böhm, C. Ratzer, V. Bednarska, I. Kalkman, and W. L. Meerts, Phys. Chem. Chem. Phys., **8,** 228 (2006).
[29] M. Schmitt, M. Böhm, C. Ratzer, S. Siegert, M. van Beek, and W. L. Meerts, J. Mol. Struc. (2006) **accepted**.
[30] M. Schmitt, M. Böhm, C. Ratzer, D. Krügler, K. Kleinermanns, I. Kalkman, G. Berden, and W. L. Meerts, Chem. Phys. Chem. **7,** 1241–1249 (2006).
[31] S. Chervenkov, P. Q. W. J. E. Braun, S. Georgiev, H. J. Neusser, C. K. Nandi, and T. Chakraborty, J. Chem. Phys. **122,** 244312-1-7 (2005).
[32] G. J. Metzger, M. Patel, and X. Hu, J. Magn. Reson. B **110,** 316 (1996).
[33] J. Dods, D. Gruner, and P. Brumer, Chem. Phys. Lett. **261,** 612 (1996).
[34] M. H. Hennessy and A. M. Kelley, Phys. Chem. Chem. Phys. **6,** 1085 (2004).
[35] J. Fisza, M. Buczkowskib, M. Budzinskia, and P. Kolenderski, Chem. Phys. Lett. **407,** 8 (2005).

[36] H. Ahonen, P. A. de Souza Jr, and K. G. Vijayendra, Nucl. Instrum. Methods Phys. B **124,** 633 (1997).
[37] I. Golovkin, R. Mancini, S. Louis, R. Lee, and L. Klein, J. Quant Spectr. Rad. Transfer **75,** 625 (2002).
[38] T. Spalek, P. Pietrzyk, and Z. Sojka, J. Chem. Inf. Model. **45,** 18 (2005).
[39] J. H. Holland, *Adaption in Natural and Artificial Systems* (The University of Michigan Press, Ann Arbor MI, 1975).
[40] D. E. Goldberg, *Genetic Algorithms in Search, Optimisation and Machine Learning* (Addison-Wesley, Reading MI, 1989).
[41] I. Rechenberg, *Evolutionsstrategie – Optimierung technischer Systeme nach Prinzipien der biologischen Evolution* (Frommann-Holzboog, Stuttgart, 1973).
[42] R. de Gelder, R. Wehrens, and J. A. Hageman, J. Comput. Chem. **22**(3), 273 (2001).
[43] R. Wehrens, E. Pretsch, and L. M. C. Buydens, Anal. Chim. Acta. **388,** 265 (1999).
[44] F. Gray, Pulse Code Communication, US Patent 2 632 058, March 17, 1953.
[45] J. Schaffer (Ed.), A study of control parameters affecting online performance of genetic algorithms for function optimization, in *Proceedings of the Third International Conference on Genetic Algorithms*, San Francisco, CA (Morgan Kaufmann, 1989).
[46] G. Myszkiewicz, High resolution UV spectroscopy and laser-focused nanofabrication, PhD thesis, University of Nijmegen, Molecular and Biophysics Group, 2005.
[47] M. Schmitt, J. Küpper, D. Spangenberg, and A. Westphal, Chem. Phys. **254,** 349 (2000).
[48] S. Gerstenkorn and P. Luc, *Atlas du spectre d'absorption de la molécule d'iode 14800–20000 cm*$^{-1}$ (CNRS, Paris, 1986).
[49] H. C. Allen and P. C. Cross, *Molecular Vib-Rotors* (Wiley, New York, 1963).
[50] J. K. G. Watson, J. Chem. Phys. **46,** 1935 (1967).
[51] J. K. G. Watson, J. Chem. Phys. **48,** 4517 (1968).
[52] G. Berden, W. L. Meerts, M. Schmitt, and K. Kleinermanns, J. Chem. Phys. **104,** 972 (1996).
[53] W. Gordy and R. L. Cook, *Microwave Molecular Spectra*, 3rd edn (Wiley, New York, 1984).
[54] C. C. Lin and J. D. Swalen, Rev. Mod. Phys. **31,** 841 (1959).
[55] D. R. Herschbach, J. Chem. Phys. **31,** 91 (1959).
[56] C. Ratzer, J. Küpper, D. Spangenberg, and M. Schmitt, Chem. Phys. **283,** 153 (2002).
[57] D. Levine, PGAPack V1.0, PgaPack can be obtained via anonymous ftp from: ftp://ftp.mcs.anl.gov/pub/pgapack/pgapack.tar.z, 1996.
[58] J. Kraitchman, Am. J. Phys. **21,** 17 (1953).
[59] C. Costain, J. Chem. Phys. **29,** 864 (1958).
[60] H. D. Rudolph, Struc. Chem. **2,** 581 (1991).
[61] J. K. G. Watson, J. Mol. Spec. **48,** 479 (1973).
[62] J. G. Smith and J. K. G. Watson, J. Mol. Spec. **69,** 47 (1978).
[63] J. K. G. Watson, A. Roytburg, and W. Ulrich, J. Mol. Spec. **196,** 102 (1999).
[64] J. K. G. Watson, J. Mol. Spec. **207,** 16 (2001).
[65] X.-Q. Tan, D. J. Clouthier, R. H. Judge, D. F. Plusquellic, J. L. Tomer, and D. Pratt, J. Chem. Phys. **95,** 7862 (1995).
[66] U. Dahmen, W. Stahl, and H. Dreizler, Ber. Bunsenges. Phys. Chem. **98,** 970 (1994).
[67] R. M. Helm, H. P. Vogel, and H. J. Neusser, J. Chem. Phys. **108,** 4496 (1998).
[68] K. Remmers, W. L. Meerts, and I. Ozier, J. Chem. Phys. **112,** 10890 (2000).
[69] P. R. Richardson, M. A. Chapman, D. C. Wilson, S. P.Bates, and A. C. Jones, J. Chem. Phys. **4,** 4910 (2002).
[70] W. Caminati, S. di Bernardo, and A. Trombetti, J. Mol. Struct. **223,** 415 (1990).
[71] A. M. Graña, J. Mol. Struct. Theochem **466,** 145 (1999).
[72] J. R. Carney and T. S. Zwier, J. Phys. Chem. A **104,** 8677 (2000).
[73] Y. R. Wu and D. H. Levy, J. Chem. Phys. **91,** 5278 (1989).
[74] Y. D. Park, T. R. Rizzo, L. A. Peteanu, and D. H. Levy, J. Chem. Phys. **84,** 6539 (1986).
[75] L. L. Connell, T. C. Corcoran, P. W. Joireman, and P. M. Felker, Chem. Phys. Lett. **166,** 510 (1990).
[76] W. Caminati, Phys. Chem. Chem. Phys. **6,** 2806 (2004).
[77] T. Nguyen, T. Korter, and D. Pratt, Mol. Phys. **103,** 1603 (2005).
[78] B. C. Dian, J. Clarkson, and T. S. Zwier, Science **303,** 1169 (2004).
[79] D. E. Poeltl and J. K. McVey, J. Chem. Phys. **80,** 1801 (1984).
[80] P. B. Bisht, H. Petek, and K. Yoshihara, J. Chem. Phys. **103,** 5290 (1995).
[81] T. Yahagi, A. Fujii, T. Ebata, and N. Mikami, J. Phys. Chem. A **105,** 10673 (2001).
[82] K. Sakota and H. Sekiya, J. Phys. Chem. A **109,** 2718 (1995).
[83] A. Held and D. W. Pratt, J. Am. Chem. Soc. **112,** 8629 (1990).
[84] A. Held and D. W. Pratt, J. Chem. Phys. **96,** 4869 (1992).
[85] A. Müller, F. Talbot, and S. Leutwyler, J. Chem. Phys. **116,** 2836 (2002).

[86] F. Lahmani, M. Broquier, and A. Zehnacker-Rentien, Chem. Phys. Lett. **337–348,** 354 (2002).
[87] C. A. Southern, D. H. Levy, J. A. Stearns, G. M. Florio, A. Longarte, and T. S. Zwier, J. Phys. Chem. A **108,** 4599 (2004).
[88] J. C. Baum and D. McClure, J. Am. Chem. Soc. **101,** 2335 (1979).
[89] J. C. Baum and D. McClure, J. Am. Chem. Soc. **101,** 2340 (1979).
[90] J. C. Baum and D. McClure, J. Am. Chem. Soc. **102,** 720 (1980).
[91] J. C. Baum, J. Am. Chem. Soc. **102,** 716 (1980).
[92] S. Nagaoka, T. Terao, F. Imashiro, A. Saika, and N. Hirota, J. Chem. Phys. **79,** 4694 (1983).
[93] J. M. Clemens, R. M. Hochstrasser, and H. P. Trommsdorff, J. Chem. Phys. **80,** 1744 (1984).
[94] A. Stöckli, B. H. Meier, R. Kreis, R. Meyer, and R. R.Ernst, J. Chem. Phys. **93,** 1502 (1990).
[95] D. F. Brougham, A. J. Horsewill, A. Ikram, R. M. Ibberson, P. J. McDonald, and M. Pinter-Krainer, J. Chem. Phys. **105,** 979 (1996).
[96] M. Neumann, D. F. Brougham, C. J. McGloinm, M. R. Johnson, A. J. Horsewill, and H. P. Trommsdorff, J. Chem. Phys. **109,** 7300 (1998).
[97] C. Rambaud and H. P. Trommsdorff, Chem. Phys. Lett. **306,** 124 (1999).
[98] D. E. Poeltl and J. K. J. McVey, J. Chem. Phys. **78,** 4349 (1983).
[99] Y. Tomioka, H. Abe, N. Mikami, and M. Ito, J. Chem. Phys. **84,** 2263 (1988).
[100] G. M. Florio, E. L. Sibert, and T. S. Zwier, Faraday Discuss. **118,** 315 (2001).
[101] J. M. Bakker, L. M. Aleese, G. von Helden, and G. Meijer, J. Chem. Phys. **119,** 11180 (2003).
[102] C. K. Nandi and T. Chakraborty, J. Chem. Phys. **120,** 8521 (2004).
[103] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge University Press, Cambridge, 1992).
[104] M. Onda, M. Asai, K. Takise, K. Kuwae, K. Hayami, A. Kuroe, M. Mori, H. Miyazaki, N. Suzuki, and I. Yamaguchi, J. Mol. Struct. **482,** 301 (1999).
[105] T. Kobayashi, K. Honma, O. Kajimoto, and S. Tsuchiya, J. Chem. Phys. **322,** 1111 (1987).
[106] M. Takayagani and I. Hanazaki, J. Opt. Soc. Am. B **7,** 1898 (1990).
[107] J. Casado, L. Nygaard, and G. O. Sørensen, J. Mol. Struct. **8,** 211 (1971).
[108] W. Caminati, P. Moreschini, I. Rossi, and P. G. Favero, J. Am. Chem. Soc. **120,** 11144 (1998).
[109] J. Catalán and P. Pérez, J. Theo. Biol. **81,** 213 (1979).
[110] P. Ilich, J. Mol. Struct. **354,** 37 (1995).
[111] A. C. Borin and L. Serrano-Andrés, Chem. Phys. **262,** 253 (2000).
[112] I. Maki, K. Nishimoto, M.-A. Sugiyama, H. Hiratsuka, and Y.Tanizaki, Bull. Chem. Soc. Jpn **54,** 8 (1981).
[113] T. Cvitaš and J. Hollas, Mol. Phys. **18,** 101 (1970).
[114] T. Cvitaš and J. Hollas, Mol. Phys. **18,** 793 (1970).
[115] T. Cvitaš and J. Hollas, Mol. Phys. **18,** 801 (1970).
[116] T. Cvitaš, J. Hollas, and G. Kirby, Mol. Phys. **19,** 305 (1970).
[117] J. Christoffersen, J. M. Hollas, and G. H. Kirby, Proc. Roy. Soc. A **307,** 97 (1968).
[118] P. Butz, R. T. Kroemer, N. A. Macleod, and J. P. Simons, J. Phys. Chem. A **105,** 1050 (2001).
[119] Y. Lee, J. Jung, B. Kim, P. Butz, L. C. Snoek, R. T. Kroemer, and J. P. Simons, J. Phys. Chem. A **108,** 69 (2004).
[120] M. S. Ford, S. R. Haines, I. Pugliesi, C. E. H. Dessent, and K. Müller-Dethlefs, J. Electron Spectrosc. Relat. Phenom. **112,** 231 (2000).
[121] W. Roth, D. Spangenberg, C. Janzen, A. Westphal, and M. Schmitt, Chem. Phys. **248,** 17 (1999).
[122] F. Negri and W. Caminati, Chem. Phys. Lett. **260,** 119 (1996).
[123] G. Berden, E. Jalviste, and W. L. Meerts, Chem. Phys. Lett. **226,** 305 (1994).
[124] A. Nakajima, M. Hirano, R. Hasumi, K. Kaya, H. Watanabe, C. C. Carter, J. Williamson, and T. A. Miller, J. Phys. Chem. A **101,** 392 (1999).
[125] R. Brause, D. Krügler, M. Schmitt, K. Kleinermanns, A. Nakajima, and T. A. Miller, J. Chem. Phys. **123,** 224311 (2005).
[126] H. Bluyssen, Hyperfine structure of HDO, D_2O and HDSe by beam maser spectroscopy. PhD thesis, University of Nijmegen, Department of Molecular and Laser Physics, 1968.
[127] S. S. Wilks, *Mathematical Statistics* (John Wiley, New York, 1962), p. 82.
[128] R. H. Bacon, Am. J. Phys. **21,** 290 (1961).

Supporting online material

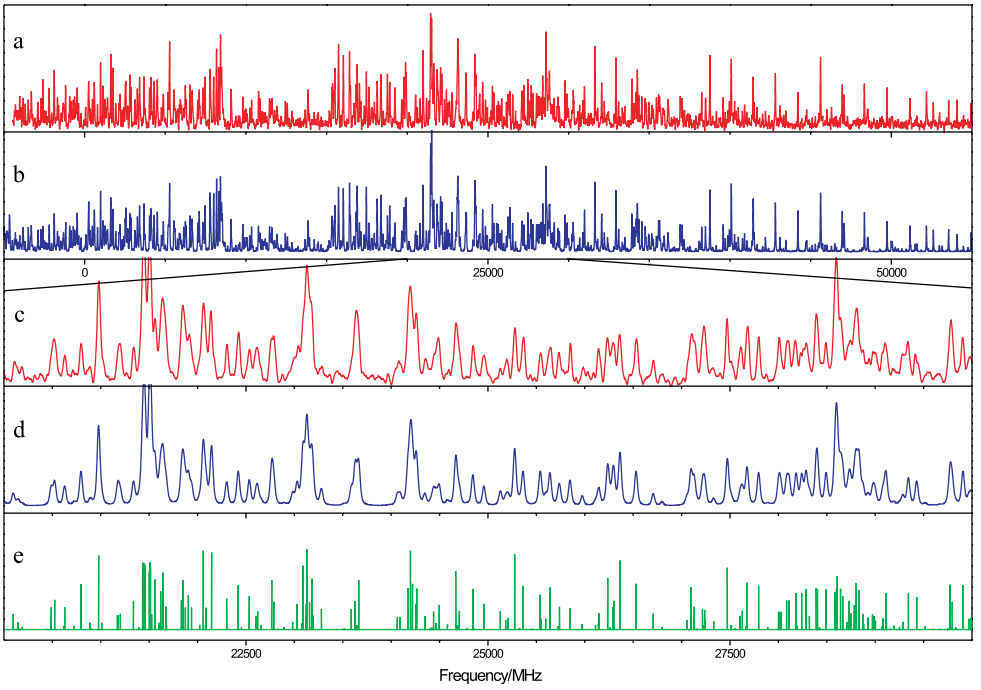

Figure S1. [Colour online] (a) Rotationally resolved electronic spectrum of the electronic origin of phenol-N_2. (b) Simulation using the inertial parameters, given in table 1. (c) Expanded view of trace (a). (d) Simulation in the same spectral range with Voigt convoluted lineshapes, using a Gaussian width of 26 MHz and a Lorentzian width of 39.4 MHz. (e) Stick spectrum in the given spectral range.

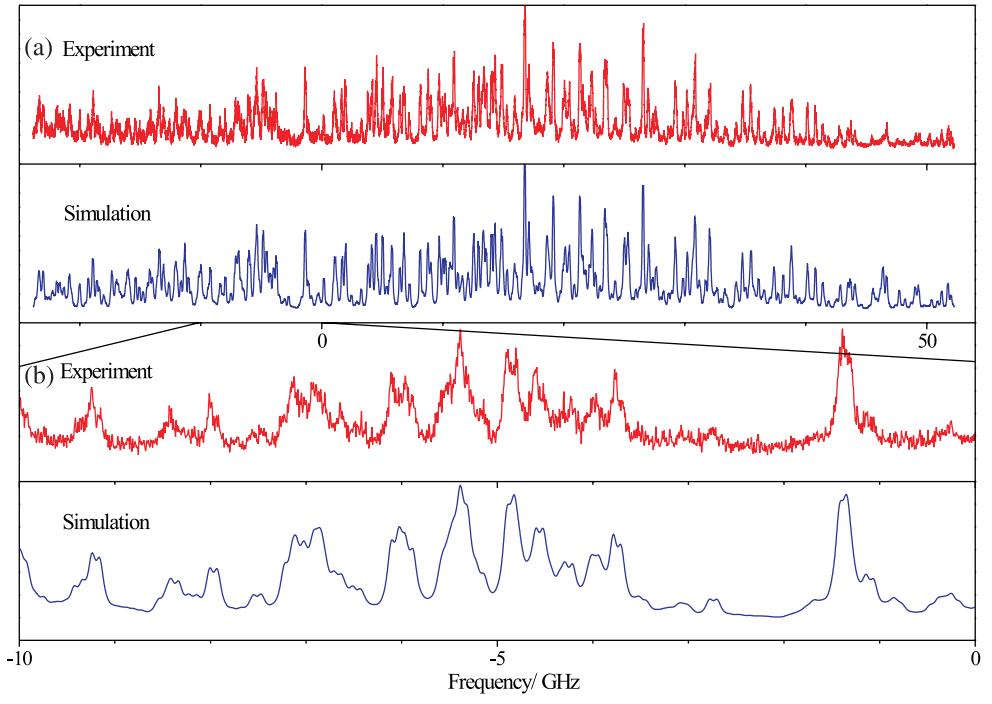

Figure S2. [Colour online] (a) Experimental and fitted rovibronic spectrum of the electronic origin of 4-methylphenol. (b) 10 GHz zoom-in of the spectrum in (a).

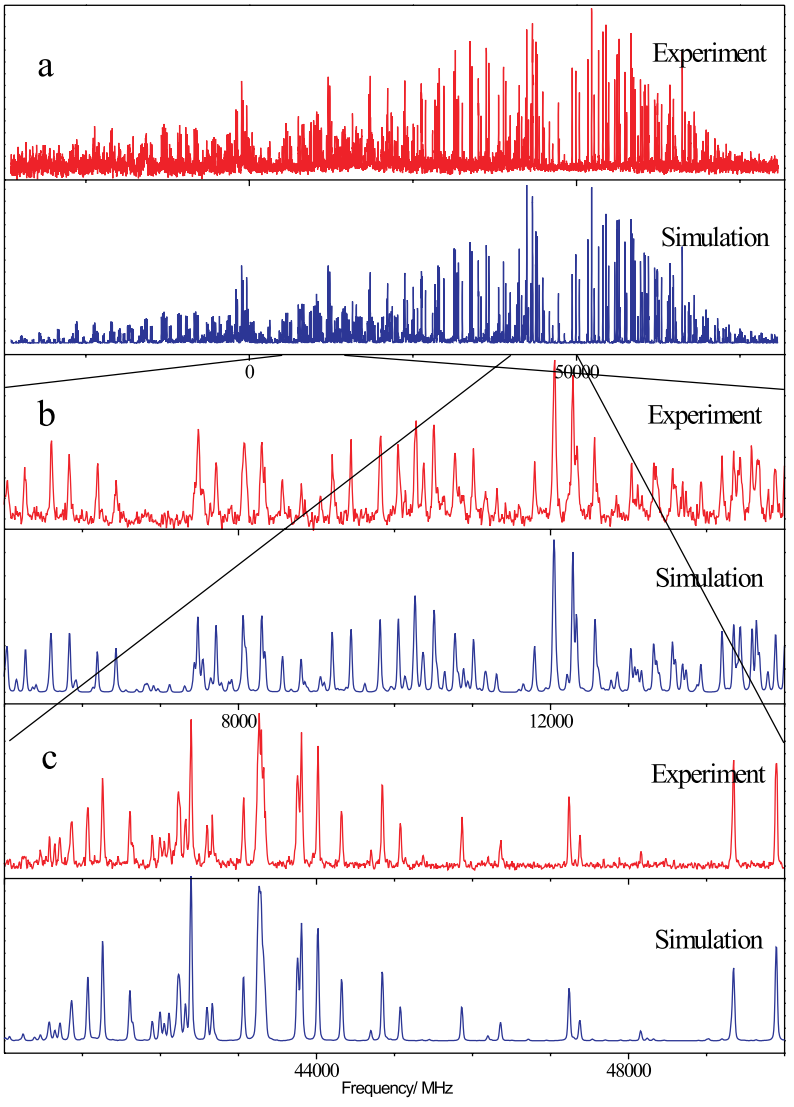

Figure S3. [Colour online] Experimental and fitted spectra of (1,2,4), (1,2,6) and (1,2,3,4) deuterated resorcinol B. (a) Full spectra. The zero in the frequency scale corresponds to the origin of the (1,2,4) isotopomer. (b) 10 GHz zoom-in of the part belonging to the (1,2,4) and (1,2,6) isotopomers. (c) 10 GHz zoom-in of the origin vicinity of the (1,2,3,4) species.

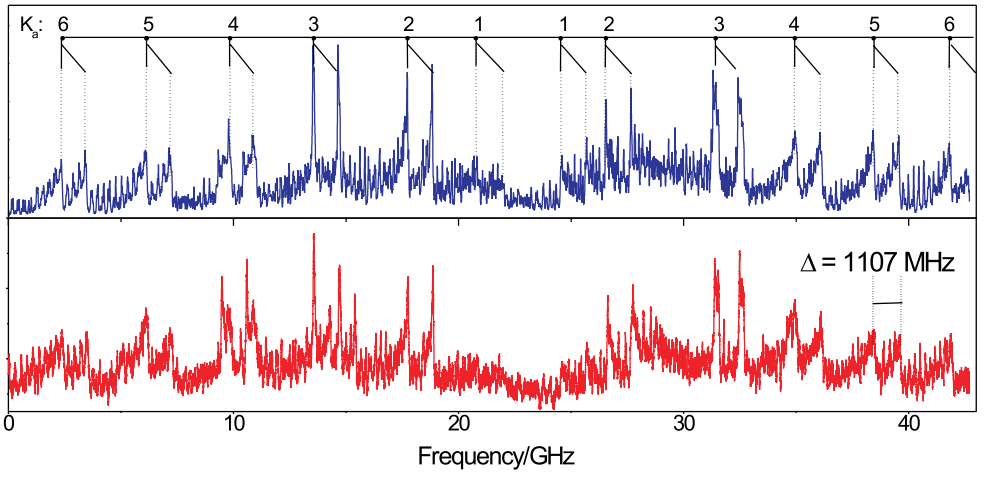

Figure S4. [Colour online] The experimental spectrum of the benzoic acid dimer (lower panel), compared with a simulation using rotational constants that are calculated from rotational constants of the benzoic acid monomer (upper panel). The upper panel shows the labelling of the *K*-stacks.