Rotationally Resolved Electronic Spectroscopy and Automatic Assignment Techniques using Evolutionary Algorithms

Michael Schmitt
Heinrich-Heine-Universität, Institut für Physikalische Chemie I,
D-40225 Düsseldorf, Germany.
and
W. Leo Meerts
Molecular and Biophysics group,
Institute for Molecules and Materials,
Radboud University Nijmegen, P.O. Box 9010,
NL-6500 GL Nijmegen, The Netherlands.

5th April 2009

Abstract

The usefulness of an evolutionary algorithm (EA) based approach to the automated evaluation of molecular parameters from various kind of spectra is shown. The applicability of the method ranges from rotationally resolved electronic spectroscopy of large molecules to nuclear magnetic resonance (NMR) spectroscopy of molecules, which are partially oriented in an anisotropic liquid-crystalline environment. The application of both the genetic algorithm (GA) and the evolutionary strategy algorithm (ES) approaches for the assignment of complex spectra and the necessity of fitting meta parameters, which are not related to the parameters of the model describing the spectra are discussed. Examples for the possible applications comprise rovibronic spectra of various aromatic water clusters, rovibronic spectra of large hetero- and homo-dimers, rovibrational spectra of the NH stretching vibrations in different tautomers of benzotriazole, and the NMR spectrum of *p*-bromo-biphenyl dissolved in a nematic crystal.

Keywords: automated analysis, evolutionary algorithms, high resolution UV spectroscopy, structure, liquid crystal NMR spectra

Contents

1 Introduction

The classical assignment of high resolution spectra is mostly based on pattern recognition by eye. Over the years a number of methods and computer programs have been developed that greatly helped the analysis of such spectra. An early attempt in the automated assignment of microwave and infrared spectra is the Loomis-Wood method (Loomis and Wood 1928), which gained wider applicability only after microcomputers became available. The group of Neusser (Helm et al. 1997) has developed the method of correlation automated rotational fitting (CARF), which directly fits the experimental data, without prior assignments. Meerts and Schmitt (2006) give a detailed overview of several of these techniques. Although the computer aided analysis meant great progress, they still need large operator interference and in general are costly in time. Furthermore, quite some experience is required from the spectroscopist and in most cases the methods are limited to well resolved spectra with not too much overlap between the transitions.

In order to be able to analyze and assign high resolution spectra automatically in a routine way we have attacked this problem with the help of evolutionary algorithms (EAs). It turned out to be a very powerful technique that made it possible to assign rotationally resolved UV laser spectra with a minimum of operator interference. The main strength of EAs is their possibility to perform global searches of the parameter space with a build in learning mechanism. The algorithms are based on concepts copied from natural reproduction and selection processes. Inherently to the fact that the EAs are global optimizers, the spectra have to be calculated a large number of times in the evaluation phase. Fortunately, with the present available fast computers and the possibility to perform vast parallel processing on large cluster machines, computing time is hardly a problem.

The process of automatic assignment of the spectra is actually based on pattern recognition guided by the EAs. Important but not decisive is the reliability of the experimental intensities and the presence of a model capable to predict the observed spectra. However, these side conditions can be released significantly in some cases as has been shown by Meerts and Schmitt (2006).

Over the years we have now analyzed and assigned numerous spectra and it was found that well resolved spectra with little overlap of spectral lines could be easily managed automatically, fast and in a routine way. This makes it possible to analyse vast amounts of experimental data and concentrate on the results rather than on the assignments. It was further shown that in cases that the spectra become more dense and the overlap of spectral lines even becomes very large, the automatic assignment procedure was still successful. Several examples are presented in the present chapter (e.g. Sect.5.2). This is of particular interest if large biological relevant molecules or biomimetics are investigated. In case of the investigation of isotopic species it often happens that several isotopologues are simultaneously present in the sample, which generally leads to a very dense spectrum due to their small spectral shifts (Schmitt et al. 2005).

In general, each type of spectrum, for which a sufficiently accurate model is available to describe the experimental data, the technique of EA fitting is applicable. Thus, the presented technique is not limited to rotationally resolved electronic spectroscopy which is the main topic of this chapter, but can instead also used for the interpretation of rovibrational spectra in the infrared (IR) (see hrs0XX) or of rotational spectra in the microwave (MW) region (see hrs0XX). The only precondition in all these cases is the availability of a Hamiltonian, which is capable of a nearly perfect description of the spectra. This precondition is not fulfilled e.g. for pure vibrational spectra, which in most cases can not be described with the necessary accuracy, since anharmonic corrections to the calculated harmonic spectra are too involved, to be applied to the large molecules considered here. Also, the application of EAs to the interpretation of MW spectra is some times hampered, due to the non reproducibility of intensities in modern Fourier transform microwave (FTMW) experiments of the Balle-Flygare type (see hrs0XX). Even though intensities are considered to be of minor importance in pattern recognition, they play an important role, if the number of lines gets very large.

In the end of this chapter we will show the application of the EA based fits to the interpretation of a completely different class of spectra, namely NMR spectra of solutes in liquid crystal solvents. Also in this case a sufficiently accurate Hamiltonian is at hand to reproduce the spectra and despite the large number of parameters to be determined, which make the manual assignment extremely tedious and time consuming, a straightforward solution on the basis of EA could be found.

2 The choice of the appropriate Hamiltonian

The essential starting point for any successful fit of a molecular spectrum is the determination of the correct Hamiltonian, needed to describe the spectrum. But even in cases where the complete Hamiltonian is not known or too time consuming to be used, an overall fit of the spectrum can be performed if the perturbations introduced by the complete Hamiltonian act only on linited ranges of the spectra.

2.1 Rotational, rovibrational, and rovibronic spectra

For most of the spectra presented here, a rigid asymmetric rotor Hamiltonian was employed (Allen and Cross 1963):

$$H_r = AJ_a^2 + BJ_b^2 + CJ_c^2. \tag{1}$$

Here $J_g (g = a, b, c)$ are the components of the body fixed angular momentum operator, A, B and C are the three rotational constants. One set of rotational constants is needed for the simulation of microwave spectra, while for rovibrational or rovibronic spectra a second set of rotational constants for the (vibrationally or electronically) excited state are necessary.

Quartic and sextic centrifugal distortion constants may be included if necessary through the model of a distortable rotor in Watson's A-reduced form (Watson 1968)

The intensities of the rotational lines are calculated from the eigenvectors of the effective rotational Hamiltonian and the known direction cosine matrix elements (Gordy and Cook 1984). The only selection rule used is $\Delta J = 0, \pm 1$ which distinguishes P, Q, and R branches. The line-strength is proportional to:

$$\begin{aligned} A_{r'r''} \quad \propto \quad & |\mu_a \langle r'|\Phi_{Za}|r''\rangle|^2 + |\mu_b \langle r'|\Phi_{Zb}|r''\rangle|^2 \\ & + |\mu_c \langle r'|\Phi_{Zc}|r''\rangle|^2 \end{aligned} \tag{2}$$

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

$$\begin{aligned} \mu_a &= \mu \ \sin\phi\cos\theta \\ \mu_b &= \mu \ \sin\phi\sin\theta \\ \mu_c &= \mu \ \cos\phi. \end{aligned} \tag{3}$$

The double and single prime denote the ground and excited state, respectively, μ is the absolute value of the transition dipole moment (TM), (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 Eq. (3) 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 TM and the c-axis. In the general case (abc-type transition) θ is called the azimuthal angle and ϕ the polar angle and the pair is written as (ϕ, θ).

The total intensity is obtained by multiplying the line strength from Eq. (2) with the temperature dependent population intensity. For simulation of gas phase spectra in a static cell the system is in thermal equilibrium and a single temperature is sufficient to calculate the distribution of rotational states using a Boltzman distribution. In molecular beams the different J states relax differently depending on the size of J. This behavior can be simulated by a two temperature model (Wu and Levy 1989):

$$n(T_1, T_2, w_T) = e^{-E/kT_1} + w_T e^{-E/kT_2}. \tag{4}$$

Here E is the energy of the lower state, w_T a weight factor, and T_1 and T_2 the two temperatures.

Thus, already for a relatively simple molecular beam spectrum with rotational resolution the number of parameters which have to be determined in the course of the fit is quite large: 6 rotational constants (3 for each state) A, B, C, eventually 10 centrifugal distortion constants (5 for each state), the center frequency ν_0, the two TDM angles θ and ϕ, the two temperatures T_1 and T_2 and a weight factor w_T from equation 4, and at least one lineshape factor determining the width of the experimental spectral line.

Any spectral perturbation which might arise from the coupling of angular momenta of internal motions of parts of the molecule with the overall angular momentum must be taken into account by using different Hamiltonians with a still larger number of parameters to be fit. Other angular momentum coupling schemes comprise Coriolis and Darling-Dennison couplings. Hamiltonians implemented in the fit program until now comprise perturbation from a single hindered three-fold rotation,(Remmers et al. 1998) from two three-fold rotors on a common torsional axis,(Tan et al. 1995) from a two-fold rotation,(Berden et al. 1996) and from axis reorientation.(Berden et al. 1995) The latter effect is due to a rotation of the inertial axes when the structure of the molecule is distorted upon electronic excitation.

2.2 The NMR Hamiltonian in a liquid crystal

The Hamiltonian (in Hz) of a molecule dissolved in a uniaxial nematic liquid crystal is given by:

$$H = \sum_{\mu} \nu_\mu I_{Z;\mu} + \sum_{\mu<\nu} J_{\mu\nu} \mathbf{I}_\mu \cdot \mathbf{I}_\nu + \sum_{\mu<\nu} D_{\mu\nu} (3 I_{Z;\mu} I_{Z;\nu} - \mathbf{I}_\mu \cdot \mathbf{I}_\nu) \tag{5}$$

where ν_μ is the resonance frequency of nucleus μ and $\mathrm{J}_{\mu\nu}$ and $\mathrm{D}_{\mu\nu}$ are the indirect spin-spin and the direct dipole-dipole coupling between nuclei μ and ν. The chemical shielding term ν_μ can have both isotropic and anisotropic parts, whereas the $\mathrm{D}_{\mu\nu}$ terms are completely anisotropic and hence do not contribute to the familiar spectra of isotropic liquids. We omit effects of the anisotropy in $\mathrm{J}_{\mu\nu}$ which is to modify slightly the observed values of $\mathrm{D}_{\mu\nu}$ and which is negligible for proton-proton couplings.

The dipolar coupling between nuclei μ and ν is:

$$D_{\mu\nu} = -\frac{h\gamma_\mu\gamma_\nu}{4\pi^2} \left\langle r_{\mu\nu}^{-3} \left(\frac{3}{2} \cos^2 \theta_{\mu\nu,Z} - \frac{1}{2} \right) \right\rangle \tag{6}$$

where $r_{\mu\nu}$ is the internuclear separation and $\theta_{\mu\nu,Z}$ is the angle between the $\mu\nu$ and magnetic field directions, and the angle brackets denote an average over solute internal and reorientational motions.

3 Theory of Evolutionary algorithms

In the past decades computational tools have been developed that are able to handle complex multi-parameter optimizations intelligently and within an acceptable time interval. The use of Evolutionary Algorithms (EAs) for solving such highly nonlinear and complex processes in science and engineering has become widespread. Although they are conceptually simple, their ability in avoiding local optima and instead finding the global optimum is remarkable and makes them suitable to handle complex optimization problems.

Evolutionary Algorithms represent a set of general purpose probabilistic search methods based on natural evolution. These algorithms mimic the concepts of natural reproduction and

selection processes. The basic idea in EAs is to create an artificial environment which encodes the search problem into biology-like terms. With the proper definition of the genetic operators the EAs let the artificial environment evolve and find the optimal solution.

Three main different EAs have been developed over the years: The genetic algorithm (GA) by Holland (1975), the Evolutionary Strategy (ES) by Rechenberg (1973) and the Evolutionary Programming by Schwevel (1993).

For the automatic assignment and analysis of the high resolution laser spectra we have mainly used the GA.(Meerts and Schmitt 2006) However recently we implemented an ES in our procedures and the comparison between GA and ES will be discussed in this chapter.

3.1 The genetic algorithm

A detailed description of the GA used in the automatic assignment and fitting of the spectra described in this chapter can be found in (Meerts and Schmitt 2006).

The molecular parameters are encoded in binary or real type, each parameter to be optimized representing a gene. The vector of all genes, which contains all molecular parameters, is called a chromosome. In an initial step the values of all parameters are set to random values between lower and upper limits which are chosen by the user. The quality of the solutions then are evaluated by a fitness function. A flow-chart of the procedure is shown in Figure 1.

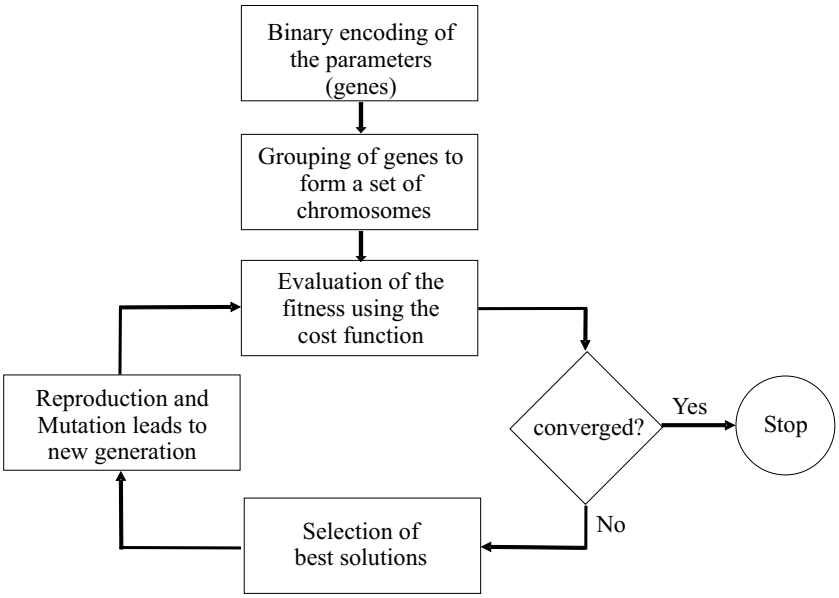

Figure 1: Schematics of the GA-process

One optimization cycle, including evaluation of the fitness of all solutions is called a generation. Pairs of chromosomes are selected for reproduction and their information is combined via a crossover process. Since crossover combines information from the parent generations, it basically explores the fitness landscape. The value of a small number of bits is changed randomly by a mutation operator. Mutation can be viewed as exploration of the fitness surface. The best solutions within a generation are excluded from mutation. This elitism prevents already good solutions from being degraded. Mutation prevents the calculation from being trapped in local minima, as is often the case with more conventional fitting routines.

3.2 The Evolutionary Strategy

The evolutionary strategy algorithm starts with one or more parent(s). A parent is a trial solution that corresponds to a set of parameters like in the genetic algorithm. From this parent an offspring of multiple children is generated. The quality or performance of these children is checked and depending on the strategy the next parent is generated. There are several different strategies for the generation of the offspring as well as the generation of the next parent. For a more detailed description see Rechenberg (1973).

The offspring is created from the parent(s) in a mutative step-size control. A drawback of the standard ES is the fact that the mutations of the decision and the strategy parameters, respectively, are subject to independent random processes. If for example an individual with a large step size undergoes only a very small change in the decision parameters and this small change turns out to yield a high fitness, the large step size will be inherited to the next generation. As a result the fitness in the next mutations may be worsen. This problem is resolved in derandomized (DR) algorithms which make the random mutations in decision and strategy parameters dependent on each other. This idea was implemented for the first as DR1 and soon improved by the concept of accumulated information (Ostermeier et al. 1994), which will be called DR2 and the history of the optimization is recorded and the evolution of the mutation ellipsoid is partially governed by past successful mutations.

A further improvement was achieved by Hansen and Ostenmeier (Hansen and Ostermeier 2001) with the Covariance Matrix Adaptation Evolution Strategy (CMA-ES). It turns out to be a particularly reliable and highly competitive evolutionary algorithm for local optimization and, surprisingly at first sight, also for global optimization (Hansen and Kern 2004). The CMA-ES does not leave the choice of strategy parameters open to the user — only the population size can be set. Finding good strategy parameters is considered to be part of the algorithm design.

Fig. 2 depicts the first four generations of an evolutionary strategy and demonstrates the effect of the chosen strategy. In general the Evolution Strategies converge faster and are more rigid than the genetic algorithm.

3.3 The fitness function for the analysis of spectra

A proper choice of this fitness function is of vital importance for the success of the EA convergence. Meerts et al. (2004) have defined a fitness function F_{fg} as:

$$F_{fg} = \frac{(\boldsymbol{f}, \boldsymbol{g})}{\|\boldsymbol{f}\| \, \|\boldsymbol{g}\|}. \tag{7}$$

Here $\boldsymbol{f}$ and $\boldsymbol{g}$ are the vector representations of the experimental and calculated spectrum, respectively. The inner product $(\boldsymbol{f}, \boldsymbol{g})$ is defined with the metric $\boldsymbol{W}$ which has the matrix elements $W_{ij} = w(|j - i|) = w(r)$ as:

$$(\boldsymbol{f}, \boldsymbol{g}) = \boldsymbol{f}^T \boldsymbol{W} \boldsymbol{g}, \tag{8}$$

and the norm of $\boldsymbol{f}$ as $\|\boldsymbol{f}\| = \sqrt{(\boldsymbol{f}, \boldsymbol{f})}$; similar for $\boldsymbol{g}$. For $w(r)$ a triangle function was used (Hageman et al. 2000) with a width of the base of Δw:

$$w(r) = \begin{cases} 1 - |r| \,/\, \left(\frac{1}{2}\Delta w\right) & \text{for } |r| < \frac{1}{2}\Delta w \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

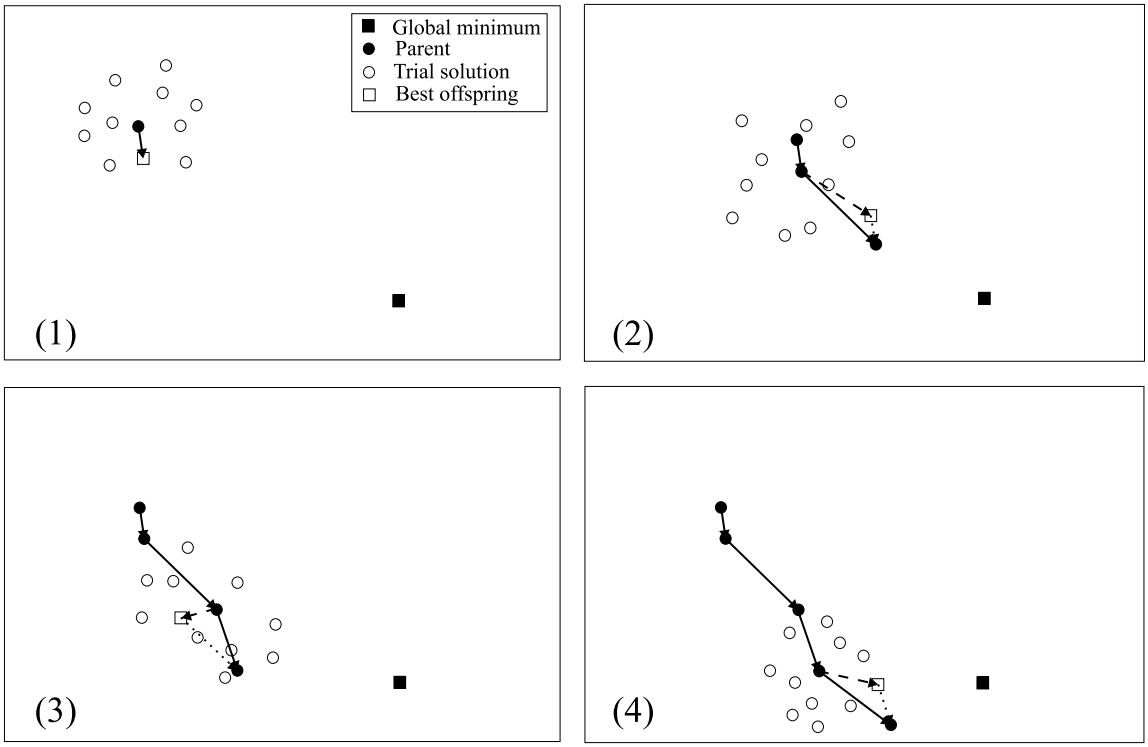

Figure 2: The first four generations of an Evolutionary Strategy: (1) An initial population is generated, and the best offspring is used as the next parent. (2) The offspring is spread over a larger area in the second generation due to the relatively large step made in the previous generation. The vector from the parent to the best offspring (dashed line) is combined with the (shortened) mutation vector of the last generation (dotted line) to generate the new parent (solid line). (3) Due to the correlation between the past two mutations the search range has been extended again in the general direction of both mutations while it has been limited in the perpendicular direction. The best offspring is now a local minimum. The memory effect of the evolutionary algorithm, which incorporates past mutation vectors into the calculation of the next parent, helps to overcome the local minimum and the next parent is still closer to the global minimum. (4) The barrier between the local and global minima has been overcome, and the optimization is progressing towards the global minimum. (Reproduced from Meerts et al. (2009) with permission from the American Institute of Physics.)

The above defined fitness function is able to smooth in a controlled way the fitness landscape and therefore allows the EA to locate the global minimum. The width of the function $w(r)$ critically determines the ability of the EA to converge to the global minimum and also the speed of convergence. The smoothing of the fitness landscape allows to sense regions far from the minimum. The EA convergence is obtained in a well defined procedure. At first the function $w(r)$ should be chosen relatively broad; $\Delta w \approx 15 - 20$ times the line widths of an individual transition in the spectrum. In this way, a first set of parameters is obtained, which still has to be refined. This is done by decreasing Δw and narrowing the limits of the parameter space to be searched in the fit. In a final calculation Δw is set to zero. Usually full EA convergence to the best set of parameters is achieved by narrowing Δw to zero in 1 or 2 steps.

3.4 Comparison between the genetic and the derandomized-ES DR2 algorithms.

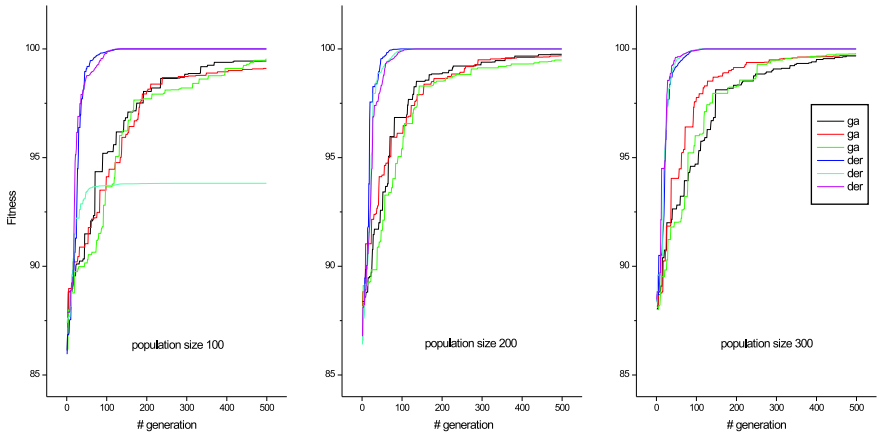

Figure 3: Comparison of three randomly started fits comparing the genetic (ga) and derandomized-ES DR2 (der) algorithms as function of the size of the population (ga) or offspring (der). In this figure the *fitness* function Eq. (7) is shown, which has a maximum values of 100.

The EAs contain a number of internal parameters, which determine the performance of the fit, and have to be optimized for a given problem themselves.

The speed and convergence of GA algorithms depend on a relative large number of internal parameters amongst others on the encoding of the parameters (binary or real), the crossover type (one-point, two-point, uniform), the size of the starting population, the rate of elitism and the mutation probability. This has been discussed to be a major drawback of this method (Wehrens et al. 1999). The meta-optimization for some of the parameters is described by Meerts et al. (2004).

The number of optimizing parameters for the derandomized-ES DR2(DR2) and CMA-ES algorithms are quite a bit smaller than of the genetic algorithm. In DR2 only a choice has to be made for offspring size and the maximum age of an individual to stay in the population. In the CMA-ES only the size of the offspring can be changed.

A comparison has been made between the GA and DR2 algorithms using the same synthetic spectrum as in Meerts et al. (2004). The result of this comparison is shown in Fig. 3. In these calculations the meta-parameter for the GA the parameters are: *Crossover Probability = 85%*, *Mutation Probability = 5%* and *Elitism = 50%*. This figure clearly shows that for this particular kind of spectrum the DR2 is faster converging and yields parameter values that are closer to the *real* ones. Although the speed of convergence is an important criterion, the total computing time is a second one. Because of the elitism factor not all individuals have to be (re)calculated in each generation in contrast to the ES. This makes the computing time for a full generation of a GA calculation typically a factor 2 faster than that of an ES calculation. However, since the ES tends to converge much faster than the GA, the convergence of a ES evolution needs typical a factor of 2 less computing time than that of a GA fit. One of the main advantages of ES over GA is the fewer number of internal strategy parameters that need to determined in a meta-optimization.

3.5 Use of the GA for the structure determination

The GA is not only utilized for an automated assignment of the rovibronic spectra, but also for the (non-linear) fit of the molecular structure to the experimentally determined rotational constants. Ratzer et al. (2002) have presented a computer program to determine molecular structures from the rotational constants of a set of different isotopologues using a gradient-based χ^2 minimizer. This program was complemented with a GA based global optimizer(Schmitt, Krügler, Böhm, Ratzer, Bednarska, Kalkman and Meerts 2006). The global minimizer has been extended to allow for intermediate local minimization steps rather than simple cost function evaluations. This technique was originally described by Li and Scheraga as part of a simulated annealing (SA) exploration of the potential hypersurface of proteins. (Li and Scheraga 1987) The algorithm, which is also known as *basin-hopping* (Wales and Doye 1997), was successfully used in combination with GA and/or SA to determine minimum-energy structures of fullerenes and atomic clusters. (Deaven and Ho 1995)

To increase the number of (independent) data points in the geometry determinations both the rotational constants as well as the angles of the transition dipole moment (TDM) of the respective isotopologue with the main inertial axes are taken into account. Both experimental data sets are completely independent from each other and can be extracted from the GA (or ES) fits of the spectra.

The cost function to be minimized in a fit of the structure to the rotational constants of k different isotopologues is the weighted sum of squared residuals: (Brandt 1998)

$$\chi^2 = \chi^2_{rot} = \Delta \boldsymbol{y}^T \boldsymbol{W}_{rot} \Delta \boldsymbol{y}. \tag{10}$$

In this equation $\boldsymbol{y}^T = (B^0_{a,1}, B^0_{b,1}, B^0_{c,1}, ..., B^0_{a,k}, B^0_{b,k}, B^0_{c,k})$ represents the vector of $n = 3k$ measured and calculated rotational constants $\boldsymbol{y}_{exp}$ and $\boldsymbol{y}_{calc}$ of the k isotopologues, respectively, $\Delta \boldsymbol{y} = \boldsymbol{y}_{exp} - \boldsymbol{y}_{calc}$ is their corresponding vector of residuals and $\boldsymbol{W}_{rot}$ denotes a $n \times n$ positive-definite weight matrix which is obtained from the covariance matrix of the n experimental rotational constants. The rotational constants directly depend on the Cartesian coordinates of the molecule, thus the mathematical relation between those n inertial values and the internal coordinates b_j, $j = 1..m$, of a hypothetical equilibrium structure is straightforward.

There exists no direct relation between the angles of the TDM with the main inertial axes, and the internal coordinates b_j of the molecule similar to that of the inertial parameters. But we can deduce an indirect relation by recognizing that the rotation of the inertial frame of a given reference isotopologue into any other isotopologue also implies a corresponding change of the TDM vector. The TDM vector $\boldsymbol{\mu}_{ref} = (\mu_{ref,a}, \mu_{ref,b}, \mu_{ref,c})^T$ of the reference isotopologue is transformed into the TDM vector $\boldsymbol{\mu}_{other}$ of another isotopologue via

$$\boldsymbol{\mu}_{other} = \boldsymbol{R}^T_{other} \boldsymbol{R}_{ref} \boldsymbol{\mu}_{ref}. \tag{11}$$

In this equation $\boldsymbol{R}_i$ denotes the 3×3 orthogonal rotation matrix which transforms the initial geometry of the i^{th} isotopologue into its own inertial system. Using both rotational constants and TDM orientations, we define the cost function as

$$\chi^2 = \chi^2_{rot} + \chi^2_{TM} = \Delta \boldsymbol{y}^T \boldsymbol{W}_{rot} \Delta \boldsymbol{y} + \Delta \boldsymbol{z}^T \boldsymbol{W}_{TDM} \Delta \boldsymbol{z} \tag{12}$$

where either $\boldsymbol{z}^T = (\theta_1, \phi_1, ..., \theta_r, \phi_r)$ or $\boldsymbol{z}^T = (\lambda^2_{a,1}, \lambda^2_{b,1}, ..., \lambda^2_{a,r}, \lambda^2_{b,r})$ in the general abc case depending on the available r transition dipole moment data. In the special case of ab hybrid spectra, the vector $\boldsymbol{z}$ simplifies to $\boldsymbol{z}^T = (\theta_1, ..., \theta_r)$ or $\boldsymbol{z}^T = (\lambda^2_{a,1}, ..., \lambda^2_{a,r})$, respectively.

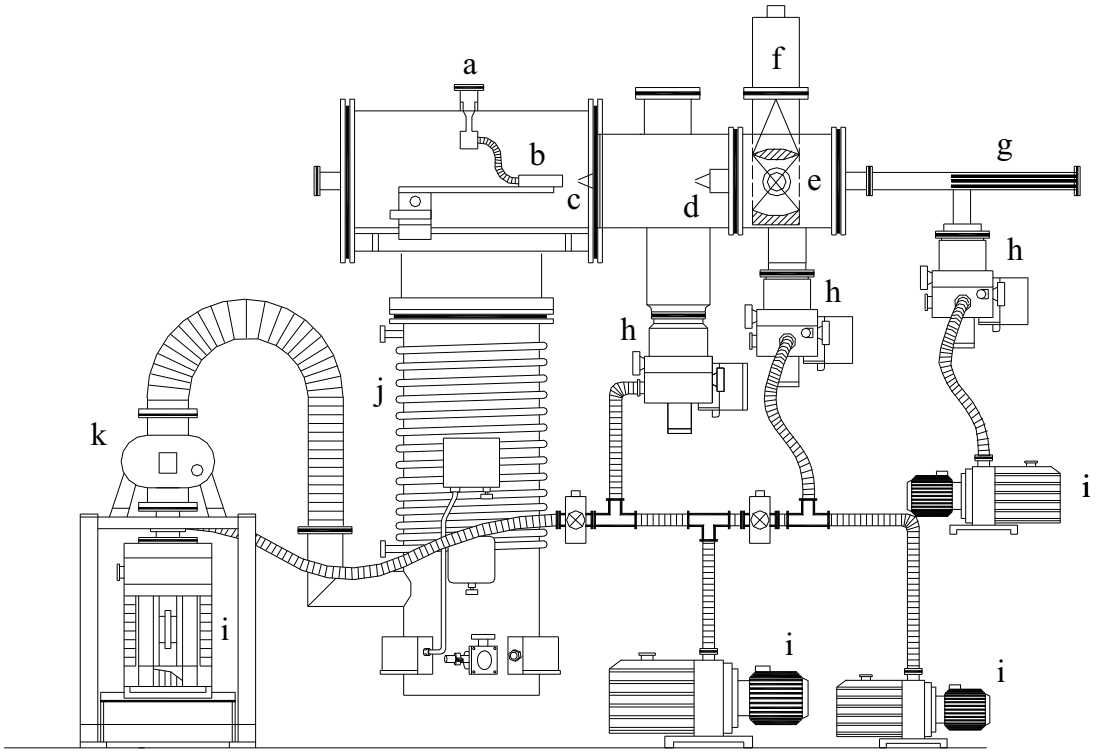

Figure 4: Molecular beam machine used for the experiments described in this chapter. (a) probe container; (b) nozzle; (c,d) beam skimmers; (e) imaging optics; (f) photomultiplier; (g) quadrupole mass spectrometer; (h) turbo-molecular pumps; (i) rotatory pumps, (j) oil diffusion pump; (k) roots blower pump.

Similarly to the weight matrix $\boldsymbol{W}_{rot}$ of the rotational constants, $\boldsymbol{W}_{TDM}$ is the weight matrix obtained by the covariance matrix of the TDM data.

4 Rotationally resolved electronic spectroscopy

Since most spectra, that are dealt with in this chapter are rovibrationally resolved electronic spectra, we will explain the experimental setup and some of the experimental implications in more detail.

The term "high resolution electronic spectroscopy" is used in the following chapter only for rotationally resolving techniques of electronic spectroscopy. Two prerequisites must be fulfilled in order to achieve this goal. Firstly, a laser system is needed that emits in the visible (400 - 750 nm) and/or in the UV (400 - 250 nm) range, with a resolution of about 1 MHz or better. At a typical wavelength of 300 nm a bandwidth of 1 MHz is equivalent to a resolution of 1 in 10^9. Secondly, an experimental set up, in which the Doppler broadening is minimized and at the same time the temperature of the sample is reduced in order to decrease the number of populated rovibronic states. Especially for large molecules, this is an indispensable requirement, since the spectra get more and more complicated with increasing number of transitions.

The problem of Doppler broadening in electronic spectra is outstandingly severe, since the Gaussian broadening due to the Doppler effect is proportional to the excitation frequency. Thus, while of minor importance in microwave spectroscopy (frequencies between 0.3 and 300

GHz), it turns a major problem in electronic spectroscopy (frequencies exceeding 1 PHz). Thus, the Doppler broadening in the UV is more than 10^6 times larger than in the microwave region. The reason for the Doppler broadening in gas phase experiments is the large velocity distribution of the molecules in the direction of the laser beam. Any technique aiming at a reduction of the Doppler width must deal with this velocity distribution. The application of seeded molecular beams (Ross 1966, Scoles 1988), which are crossed at right angles with the laser considerably reduces the Doppler broadening. By seeding we refer to the technique of mixing a small amount of the molecules under investigation with a surplus of a rare gas, prior to expansion. A further reduction of the Doppler width can be achieved by confining the molecular beam to a small angular range about the beam axis using one or more skimmers. By means of this geometric limitation of the molecular beam in the direction of the laser beam a reduction of the velocity distribution in a thermal ensemble by a factor of 200 can be achieved. Thus, the Doppler width which in a thermal ensemble of Argon at 373 K at a central wavelength of 300 nm amounts to about 2 GHz can be reduced to 10 MHz by pure geometric skimming.

Further reduction of the Doppler width can be achieved by geometric filtering of the central part of the emitted fluorescence. Figure 4 shows a typical molecular beam setup used for rotationally resolved electronic spectroscopy. The molecular beam, which is formed by expansion of the probe (a), seeded in Argon through the nozzle (b) is skimmed twice (c,d) for a maximum reduction of the Doppler width. The beam is crossed at right angles with the laser light and the fluorescence is collected at right angles though imaging optics (e) onto a photomultiplier (f). The molecular beam position can be controlled by means of a quadrupole mass spectrometer (g).

Apart from cooling the seeded molecules in the molecular beam of nobles gases to rotational temperatures below 2 K, condensation of the seed molecules takes place, making weakly bound complexes experimentally accessible. At typical vibrational temperatures of a few 10 K and rotational temperatures below 2 K hydrogen bound complexes and even van der Waals bound complexes are stable species and can be investigated spectroscopically in molecular beams. High stagnation pressures and large nozzle diameters are necessary prerequisites for cluster formation, since the rate of the cluster formation determining three body collisions scales with $P_r^2 D/T_r^2$ where P_r is the stagnation pressure in the reservoir, T_r is the reservoir temperature and D is the nozzle diameter. From this follows, that high pumping speeds are necessary for observing larger clusters in molecular beams.(Scoles 1988)

Several detection schemes are possible. The most common one is the use of laser induced fluorescence (LIF), in which the emitted fluorescence acts as probe for the absorption of the laser light by the molecules under investigation (Majewski and Meerts 1984, Majewski et al. 1989, Schmitt et al. 2000).

Another possible detection scheme is the multiphoton ionization (MPI) (Sussmann et al. 1994), which has the advantage compared to LIF of yielding mass resolved spectra, which is especially advantageous if molecular complexes are investigated with similar absorption frequencies. On the other hand the resolution in MPI experiments is limited by the necessary use of pulsed lasers with a temporal band width of a few nanoseconds to the Fourier limit of the pulse duration. Typical resolutions which can be achieved in these experiments are on the order of 70 - 100 MHz.

Other detection schemes comprise sensitized phosphorescence as probe Uijt de Haag and Meerts (1989), in cases were a strong intersystem crossing results in a weak or even completely absent fluorescence signal, or the direct measurement of absorption of radiation in the molecular beam via optothermal techniques Becucci et al. (1997) in cases were radiationless

processes dominate, which lead to highly vibrationally excited ground state molecules.

Another complementary method for determination of rotational constants in different electronic states must be mentioned, when talking about rotationally resolved electronic spectroscopy. Felker and co-workers established the so-called rotational coherence spectroscopy (RCS) (Felker 1992), which was in the last few years considerably improved by Riehn (2002). While the rotational resolution in the experiments described here, is achieved by limiting the spectral band width as described in the preceding chapter, RCS utilizes a completely different approach. Like the Fourier transform (FT) describes the mathematical relationship between time and frequency domain, this approach is utilized in RCS by measuring the time which a molecule needs for rotation per radian. For large molecules, these times are at the order of a few picoseconds. Thus, to measure the rotational recurrences of large molecules, exciting laser pulses with picosecond or even sub-picosecond duration are necessary. Between spectrally and temporarily resolving spectroscopic techniques exists a complementary relationship, since the frequency distances of the rotational lines decrease with increasing moment of inertia, while the distance between two recurrences increases.

In the following it is shown, how the evolutionary methods can be utilized in the practical analysis of very complex spectra. We chose rotationally resolved electronic spectra in a molecular beam, FTIR spectra in the gas phase and liquid crystal NMR spectra to show the broad applicability of the method.

5 Application of EA and GA to the fit of rotationally resolved electronic spectra

5.1 Structure and stabilization of aromatic water complexes

Water complexes of polar aromatic molecules have found considerable interest as model systems for solvation processes, ground state and electronically excited state acidities, and dynamical processes, connecting the primarily excited state with other electronically excited states or with the electronic ground state. In all examples we present here, a splitting of the rovibronic spectra into two components by a tunneling motion of the water moiety is found. While for the smaller of these systems, the rotational constants are so large, that despite the doubling of lines by the tunneling splitting a rotational resolution of the lines is achieved, larger systems like the p-cyanophenol$(H_2O)_1$ or the tryptamine$(H_2O)_1$ cluster constitute a problem due to congestion of the spectra like in the case of the aromatic dimers. Another examples of a water complex with strongly overlapping rovibronic lines is the p-cresol$(H_2O)_1$ cluster, in which a splitting due to the tunneling motion of the water moiety and an additional splitting from the internal rotation of the methyl group leads to four overlapping torsional subbands.

5.1.1 Phenol-water

Among the most studied hydrogen bound clusters of aromatic molecules the phenol$(H_2O)_1$ cluster certainly ranks at the top. From early days of laser spectroscopy in molecular beams on this cluster was subject to a multitude of experimental and theoretical investigations.(Abe et al. 1982, Lipert and Colson 1988, Stanley and Castleman 1991, Schütz et al. 1993, Berden et al. 1996)

The spectra and their assignment The rovibronic spectrum of the electronic origin of phenol$(H_2O)_1$ at 35996.47 cm^{-1} shown in Figure 5 is an example for a spectrum of a hydrogen bonded cluster, that can easily be analyzed and fit using classical line position assignment techniques. Indeed, the original spectrum was completely analyzed manually.(Berden et al. 1996) Due to a hindered internal rotation of the water moiety in the cluster, the spectrum is split into two subtorsional components, $\sigma = 0$ and $\sigma = 1$ with a splitting of 25.4 GHz between the electronic origins of both subtorsional components. The separate simulations of each component is shown in Figure 5. The lowest trace, which shows a zoomed portion of the simulated spectrum along with the stick spectrum reveals, that most of the lines in the spectrum are composed of single rovibronic transitions, making a line position assigned fit *easy* in this case.

Determination of the structural parameters The rotational constants of the phenol-water cluster, which are compiled in Table 1 show, that all heavy atoms of phenol$(H_2O)_1$ must be in one plane. From the rotational constants an inertial defect of -2.4 u$Å^2$ for the ground state and of -2.8 u$Å^2$ for the electronically excited states is calculated. This is in agreement with a structure, in which the two hydrogen atoms of the water moiety are symmetrically out of the aromatic plane by about 1 Å. The fact that the rotational B and C constants of the $\sigma = 0$ and the $\sigma = 1$ subtorsional components are equal and the only difference is in the A rotational constant immediately leads to the conclusion that the axis of the internal rotation of the water moiety is parallel to the a-axis of the cluster. Using the rotational constants of the monomer and the phenol-water cluster, the position of the center of mass (CM) of the water moiety with respect to the position of the CM of phenol can be calculated. Since the position of the phenolic O-atom and the water O-atom with respect to each CM are known, the hydrogen bond length OO can be calculated. The complete procedure is described in Ref. (Berden et al. 1996) and yields bond lengths of 293 pm for the electronic ground state and of 289 pm for the electronically excited state. The above considerations lead to the structure of the phenol-water cluster as shown in Figure 6. The barrier to torsion of the water moiety is determined to be 180 and 130 cm^{-1} for the S_0 and the S_1 state, respectively.

5.1.2 p-Cyanophenol-water

The rotationally resolved spectrum of the electronic origin of the p-cyanophenol$(H_2O)_1$ cluster has a considerably higher line density than the spectrum of phenol$(H_2O)_1$. Two main reasons exist for that. Firstly, the A, B, and C rotational constants are smaller than in the phenol-water cluster, with the latter two being nearly half of the size as in phenol-water. Secondly, the changes of the rotational constants B and C upon electronic excitation (ΔB and ΔC) are much smaller in the p-cyanophenol-water cluster compared to phenol-water(cf. Table 1).

The spectra and their assignment The rovibronic spectrum of the electronic origin of p-cyanophenol$(H_2O)_1$ given in Figure 7 shows the same kind of torsional splitting due to the hindered water torsion as the phenol$(H_2O)_1$ spectrum, although the splitting between the subtorsional origins is only 8.1 GHz compared to 25.4 GHz.(Jacoby et al. 2006) The smaller subtorsional splitting is caused by a much higher barrier to the water torsion compared to phenol$(H_2O)_1$, which can be traced back to an additional interaction of one of the water oxygen lone pairs, with the ortho CH bond, which is much more acidic than in any other hydroxybenzene derivative (cf. the benzonitrile dimer in section 5.2.2).

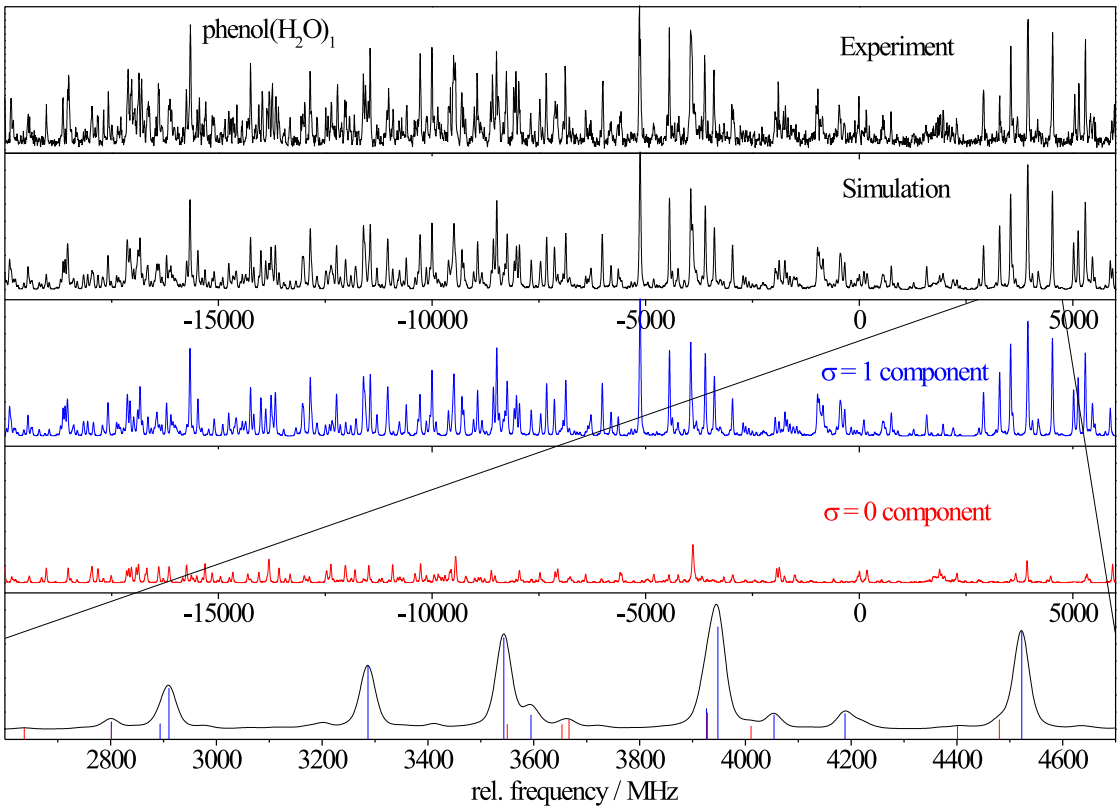

Figure 5: Rovibronic spectrum of the electronic origin of phenol$(H_2O)_1$ at 35996.47 cm^{-1}. The traces show the experimental spectrum, the simulated spectrum, the simulations of the subtorsional components $\sigma = 0$ and $\sigma = 1$ and zoomed portion of the simulation together with the stick spectra of both components.

The lowest trace in Figure 7 which shows the simulation along with the stick spectra of both subtorsional components reveal impressively why the use of automated assignment techniques is mandatory in this case. Up to 50 single rovibronic transitions contribute to each feature in the spectrum, which appears to be a single "line".

Determination of the structural parameters The same considerations, as made for the phenol$(H_2O)_1$ cluster also hold for the p-cyanophenol$(H_2O)_1$ cluster. The orientation of the water moiety in the cluster can be described as *trans*-linear. The fit of the hydrogen bond lengths in the ground and the electronically excited state reveal, that both hydrogen bonds are shorter ($r_{OO}(S_0)$=283 pm; $r_{OO}(S_1)$=269 pm) than in the case of phenol-water. Also the decrease of the hydrogen bond length upon electronic excitation is larger than in phenol-water. Both facts can be traced back to the acidity of p-cyanophenol. While phenol is a moderate acid in its electronic ground state (pK_a=9.86), its acidity increases upon electronic excitation by nearly four orders of magnitude to a pK_a of 6.(Granucci et al. 2000) p-cyanophenol is a much stronger acid in the ground state (pK_a=7.74) and simultaneously, the increase of acidity upon electronic excitation to a pK_a of 3.33 in larger than in phenol.

Also, the acidity of the *ortho* CH position is considerably increased compared to phenol. With the strong reduction of the OO hydrogen bond length, both oxygen lone pairs can overlap to form two hydrogen bonds. The stronger bond is formed to the hydroxy group, the weaker one to the CH group (cf. Figure 8).

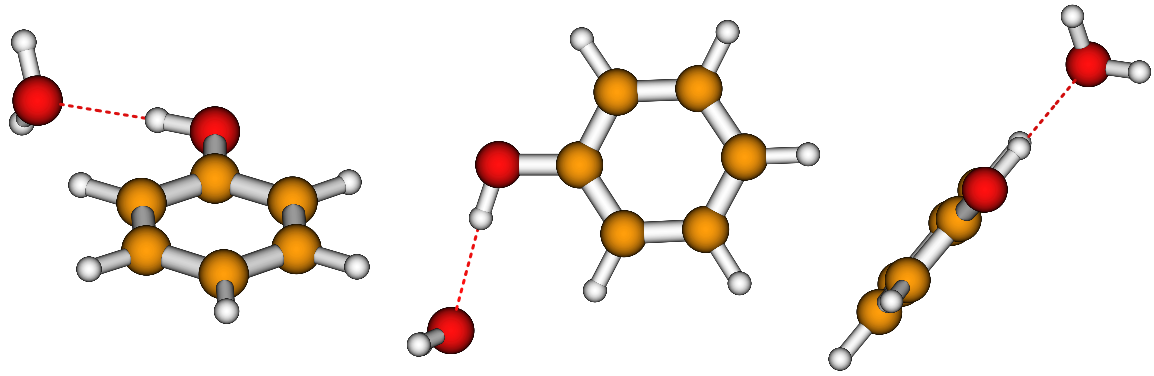

Figure 6: The structure of the phenol-$(H_2O)_1$ cluster.

5.1.3 p-Methylphenol-water

The interpretation of the methylphenol$(H_2O)_1$ water cluster spectrum bears the highest difficulties among the hitherto described hydroxybenzene-water clusters. Not only, that the rotational constants are smaller than in phenol-water, but there is an additional complication which arises from the internal torsional motion of the methyl group. While in phenol-water and p-cyanophenol-water the spectrum was split into two subtorsional components ($\sigma = 0$ and $\sigma = 1$), in the p-cresol-water complex four subtorsional components have to be considered. Each of the $\sigma = 0$ and $\sigma = 1$ components is further split by the threefold methyl rotation in a set of A and doubly degenerate E levels (Myszkiewicz et al. 2005). The two subtorsional bands of the water torsion and the respective subtorsional bands of the methyl torsion could be fitted independently, showing that the two motions are practically decoupled.

The spectra and their assignment Figure 9 shows the rovibronic spectrum of the electronic origin of p-methylphenol$(H_2O)_1$ at 34971.89 cm^{-1}. At first sight the higher line density compared to phenol$(H_2O)_1$ and to p-cyanophenol$(H_2O)_1$ cluster is obvious. The origins of the $\sigma = 0$ and $\sigma = 1$ are marked in the simulation trace of Figure 9, together with the $A - E$ splitting of each component. The origins of the $\sigma = 0$ and $\sigma = 1$ components are 29.4 GHz apart, compared to 25.4 GHz for phenol-water and 8.1 GHz for p-cyanophenol-water, alluding similar torsional barriers for the water torsion in p-methylphenol-water and phenol-water. The molecular parameters from the GA fit of p-methylphenol$(H_2O)_1$ are compiled in Table 1 and are compared to those of phenol$(H_2O)_1$ and of p-cyanophenol$(H_2O)_1$.

Determination of the structural parameters Again, as for phenol$(H_2O)_1$, and p-cyanophenol$(H_2O)_1$ a *trans* linear hydrogen bond arrangement is found with a OO hydrogen bond length of 290 pm in the electronic ground state and of 285 pm in the electronically excited state, close to the value of unsubstituted phenol and considerably longer than in p-cyanophenol$(H_2O)_1$. The barrier in the electronic ground state is determined to be 183 cm^{-1}, close to the value of phenol-water.

5.1.4 Tryptamine-water

The case of the tryptamine$(H_2O)_1$ cluster is insofar exceptional, that the tryptamine monomer moiety exists in seven different conformational forms in the gas phase (Nguyen et al. 2005). First Park et al. (1986) and later Connell et al. (1990) and Schmitt et al. (2005) showed how

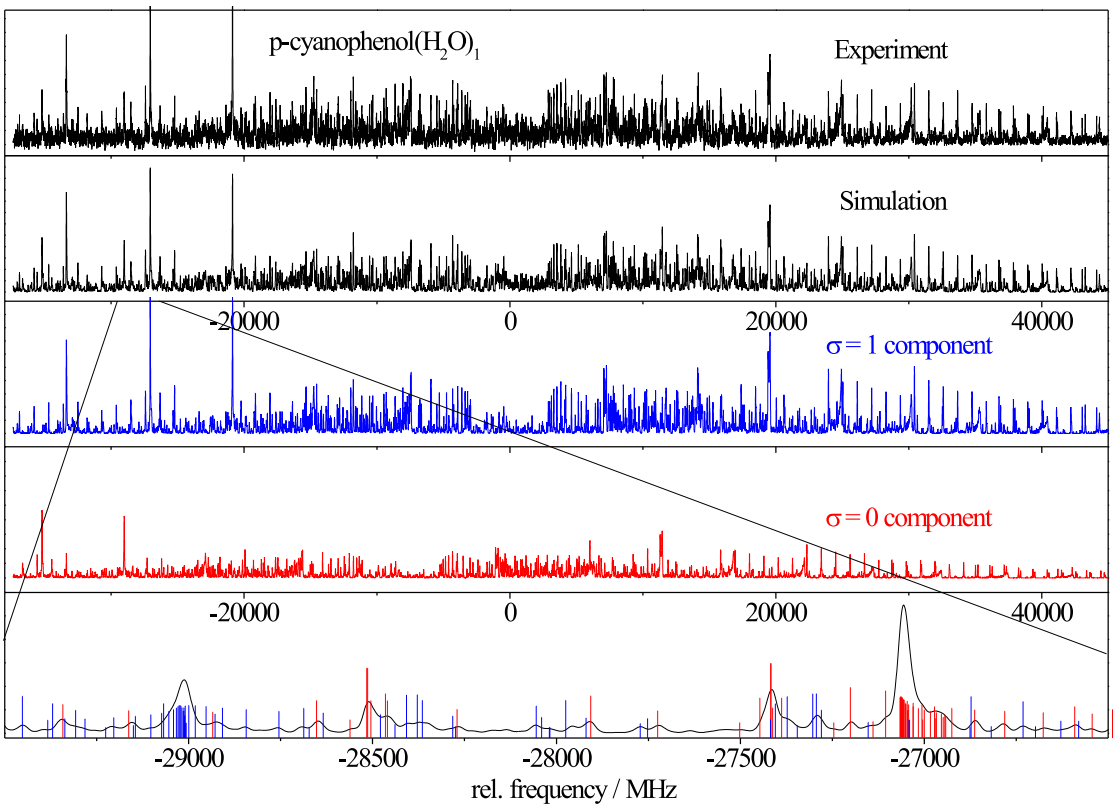

Figure 7: Rovibronic spectrum of the electronic origin of p-cyanophenol$(H_2O)_1$ at 35304.46 cm^{-1}. The traces show the experimental spectrum, the simulated spectrum, the simulations of the subtorsional components $\sigma = 0$ and $\sigma = 1$ and zoomed portion of the simulation together with the stick spectra of both components.

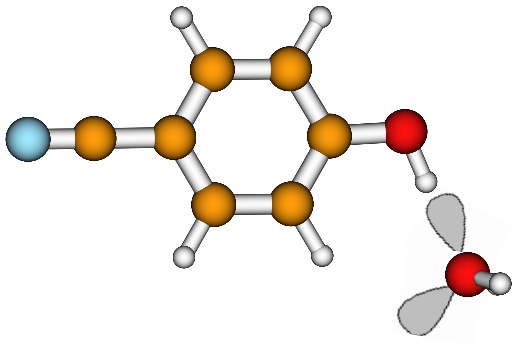

Figure 8: The structure of the p-cyanophenol-$(H_2O)_1$ cluster.

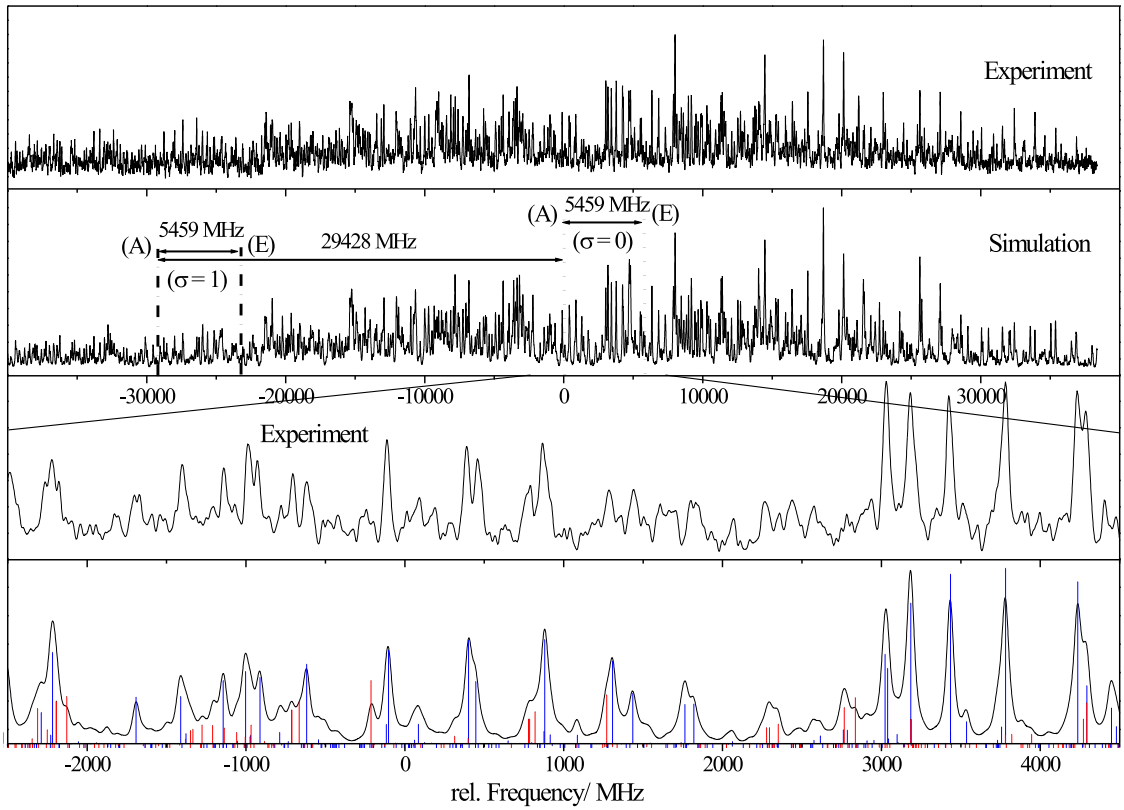

Figure 9: Rovibronic spectrum of the electronic origin of p-methylphenol$(H_2O)_1$ at 34971.89 cm^{-1}. The traces show the experimental spectrum, the simulated spectrum, the simulations of the subtorsional components $\sigma = 0$ and $\sigma = 1$ and zoomed portion of the simulation together with the stick spectra of both components.

Table 1: Comparison of the molecular parameters of the electronic origin bands of the phenol$(H_2O)_1$, the p-cyanophenol$(H_2O)_1$, and the p-methylphenol$(H_2O)_1$ cluster.

	phenol$(H_2O)_1$	p-cyanophenol$(H_2O)_1$	p-methylphenol$(H_2O)_1$
$A''(\sigma = 0)$/MHz	4291.49(4)	3462.16(50)	3663.48(244)
$B''(\sigma = 0)$/MHz	1092.1445(2)	587.87(8)	765.60(59)
$C''(\sigma = 0)$/MHz	873.7271(2)	503.62(6)	637.00(48)
$A''(\sigma = 1)$/MHz	4281.76(1)	3459.95(6)	3652.19(53)
$B''(\sigma = 1)$/MHz	1092.3254(1)	588.00(4)	765.98(31)
$C''(\sigma = 1)$/MHz	873.9082(1)	503.75(2)	637.71(31)
V_2''/cm^{-1}	180	275	183
ν_0/cm^{-1}	35995.62(1)	35304.46(2)	34971.891(10)
$\Delta\nu$/GHz	25.455(10)	8.1494(10)	29.428(7)
$\Delta A(\sigma = 0)$/MHz	-102.7(6)	-48.69(24)	-72.61(345)
$\Delta B(\sigma = 0)$/MHz	8.4(2)	-0.11(1)	9.18(83)
$\Delta C(\sigma = 0)$/MHz	0.89(6)	-1.08(1)	3.76(68)
$\Delta A(\sigma = 1)$/MHz	-114.36(2)	-50.39(5)	-75.82(75)
$\Delta B(\sigma = 1)$/MHz	8.30(2)	-0.13(1)	8.37(44)
$\Delta C(\sigma = 1)$/MHz	0.94	-1.06(1)	3.90(44)
$\Delta V_2/cm^{-1}$	-30	-92	-58

this structural diversity collapses into a single conformer upon cluster formation with one water molecule.

The spectra and their assignment Contrary to the phenol-water clusters described in the preceding sections, the vibronic spectrum of the origin of the tryptamine $(H_2O)_1$ cluster does not show a torsional splitting. Figure 10 presents the experimental spectrum with the simulation using the best parameters from a GA fit. As in the water clusters of the phenol derivatives, a multitude of lines contributes to each feature observed in the experimental spectrum, what made the use of GA for the automated spectral assignment and determination of the molecular parameters given in Table 2 indispensable.

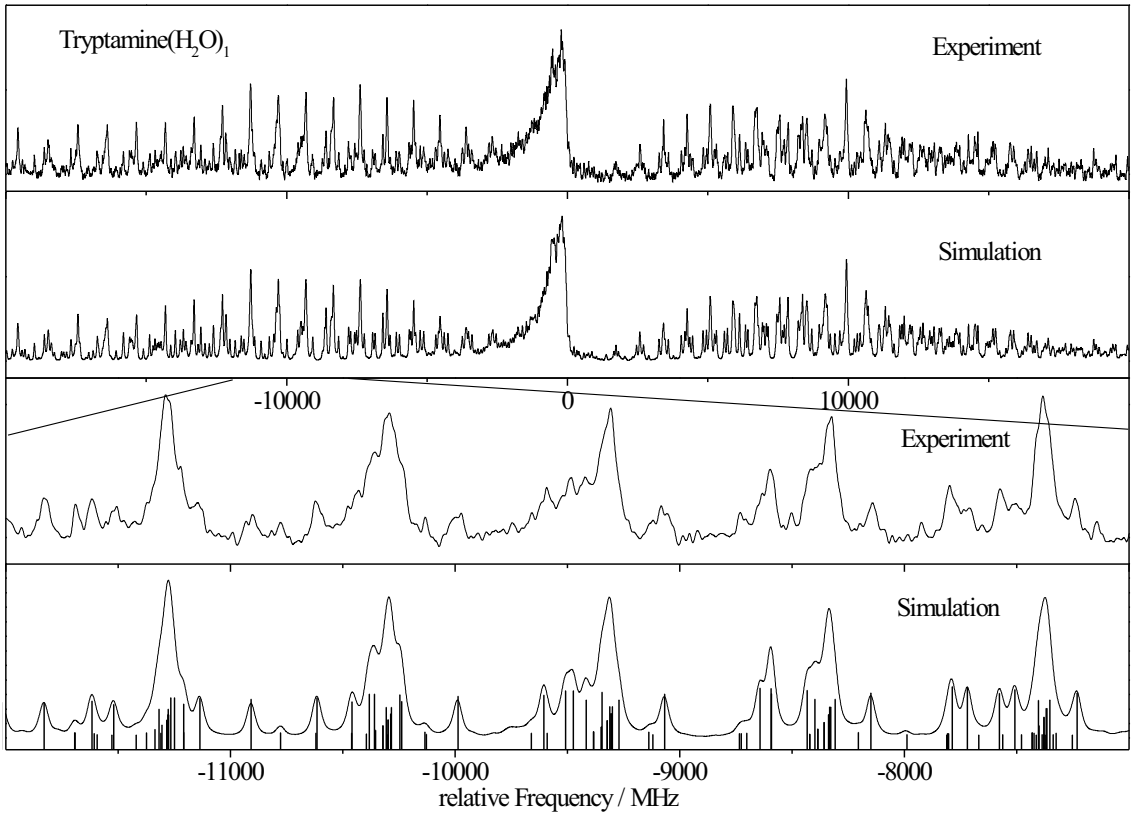

Figure 10: Rovibronic spectrum of the electronic origin of the water cluster of the Gpy(out) conformer of tryptamine together with the best simulation, a zoomed portion of the experimental spectrum and the zoomed simulation together with the stick spectrum.

Determination of the structural parameters Extensive ab initio calculations are necessary to determine the binding position of the water molecule to the tryptamine moiety in the cluster. Several positions are possible: the water might bind to the pyrrolic NH group via a hydrogen bond, with the indole acting as proton donor, similar to the indole-water cluster. Secondly the water can bind to the ethylamino group via a hydrogen bond. Three cases must be distinguished here: (i) The amino group might act as proton acceptor, the water as proton donor. (ii) The amino group is the proton donor via the H_a atom. (iii) The amino group is the proton donor via the H_b atom. From comparison of the rotational constants and the stabilization energies, only the structure, shown in Figure 11 is considered, which immediately

Table 2: Molecular Parameters for the tryptamine-water cluster determined from GA fits, as described in the text. The MP2 calculations were performed with the G03 program package (Frisch et al. 2003) using the 6-311G(d,p) basis set. RICC2 calculations were performed with Turbomole V5.8 using the VTZP basis set from the turbomole library (Ahlrichs et al. 1989) for both the ground and first electronically excited $\pi\pi^*$ states.

	Experiment	RIMP2	RICC2 (S_0 and L_b)	RICC2 (S_0 and L_a)
$A''(MHz)$	1465.78	1469	1464	1464
$B''(MHz)$	483.43	471	498	498
$C''(MHz)$	397.99	387	410	410
ν_0/cm^{-1}	34957.11	-		32549
$\Delta A(MHz)$	-6.82	-		-76
$\Delta B(MHz)$	-4.71	-		+39
$\Delta C(MHz)$	-3.72	-		+26

gives the answer to the question, why one out of seven tryptamine conformers is frozen out, upon water complexation. In the structure shown, the water acts as proton donor to the amino group, but at the same time is proton acceptor to the CH group of the pyrrole ring. Since the Gpy(out) conformer is the only one, in which the distance of the CH group and the amino group is small enough to be bridged by a single water molecules, this explains the exclusive appearance of only one conformer.

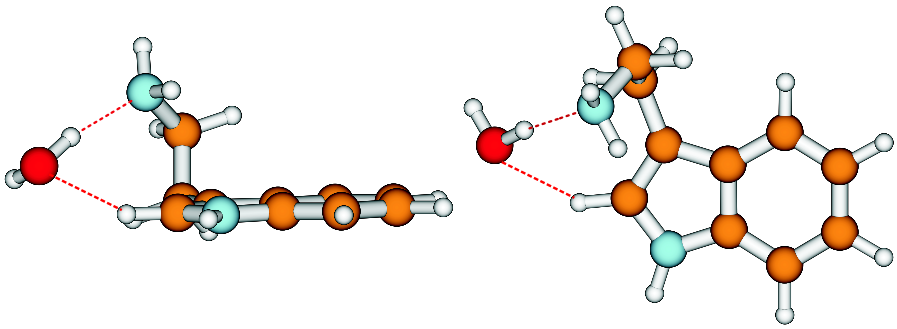

Figure 11: The structure of the tryptamine-$(H_2O)_1$ cluster.

Here, as in other complexes of indole derivatives with polar solvent molecules, the question arises about the electronic nature of the electronically excited state. Two close lying electronically excited states, which are historically labeled by L_a and L_b, might change their energetic ordering upon cluster formation with polar molecules, since the L_a state is highly polar, whereas the polarity of the L_b state is much smaller. In the present case comparison of the structure from ab initio calculations, shown in Table 2 with the experimental rotational constants clearly shows, that the experimentally observed electronic origin of the water cluster remains of L_b type, as the monomer.

5.2 Global vs. local electronic excitation in symmetric and antisymmetric aromatic dimers

Many publications have raised the question if the electronic excitation in symmetrically bound homodimers of aromatic compounds is localized in one of the monomer moieties or is completely delocalized over both monomer moieties. Prototypes of these symmetrical dimers are the benzoic acid dimer investigated by Poeltl and McVey (1984) and by Remmers et al. (2000), the salicylic acid dimer (Yahagi et al. 2001), the 2-pyridone dimer (Held and Pratt 1992, Müller et al. 2002), the o-cyanophenol dimer (Lahmani et al. 2002) and the benzonitrile dimer (Schmitt, Böhm, Siegert, van Beek and Meerts 2006). While for the 7-azaindole dimer and the 2-pyridone dimer delocalized excitation is found, the benzoic acid dimer, the fluorobenzoic acid dimer and the o-cyanophenol 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 funneling devices in nature.

5.2.1 The phenol dimer

Among others, the phenol dimer is an ideal model system to study the very sensitive equilibrium between hydrogen bonding and van der Waals interactions. The spectroscopy of complexes of phenol with different solvent molecules have been subject of a plethora of publications in the last decades. In most of these complexes phenol acts as a proton donor with respect to the solvent molecule.

The phenol dimer takes a special position, since one of the cluster constituents acts as proton donor, the other as proton acceptor, both of which are accessible spectroscopically as was shown by Fuke and Kaya (1983). Van der Waals interactions between the π-systems of the aromatic rings are most likely to play an important role for mutual arrangement of the monomer moieties, while for most of the other phenol clusters the structure of the cluster is mainly governed by dipole-dipole interactions.

The spectra and their assignment Figure 12 shows the rovibronic spectrum of the electronic origin of the phenol dimer together with the best GA fit using the molecular parameters from Table 3.

The lowest trace of 12 shows a zoomed part of the convoluted simulation along with the unconvoluted stick spectrum. Most lines in the spectrum are superpositions of up to 15 single rovibronic transitions, that contribute with similar intensities. Such heavy congestion of lines can in general only be handled using a EA based fitting strategy. The large density of rovibronic lines avoids a classical assignment procedure or at least makes it a incredibly tedious work.

The spectrum is fit using a rigid rotor Hamiltonian. It consists of about 13000 lines in a range of less than 2 cm^{-1}. An additional challenge lies in the fact, that the spectrum is composed of nearly equal *a*-, *b*-, and *c*-type fractions. 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 molecular parameters obtained from the fit are given in Table 3. Other isotopologues, which are indispensable for the structure determination have been measured and are gathered in Table 3.

Determination of the structural parameters The intermolecular parameters were determined using the GA based fit of the structural parameters to the rotational constants and

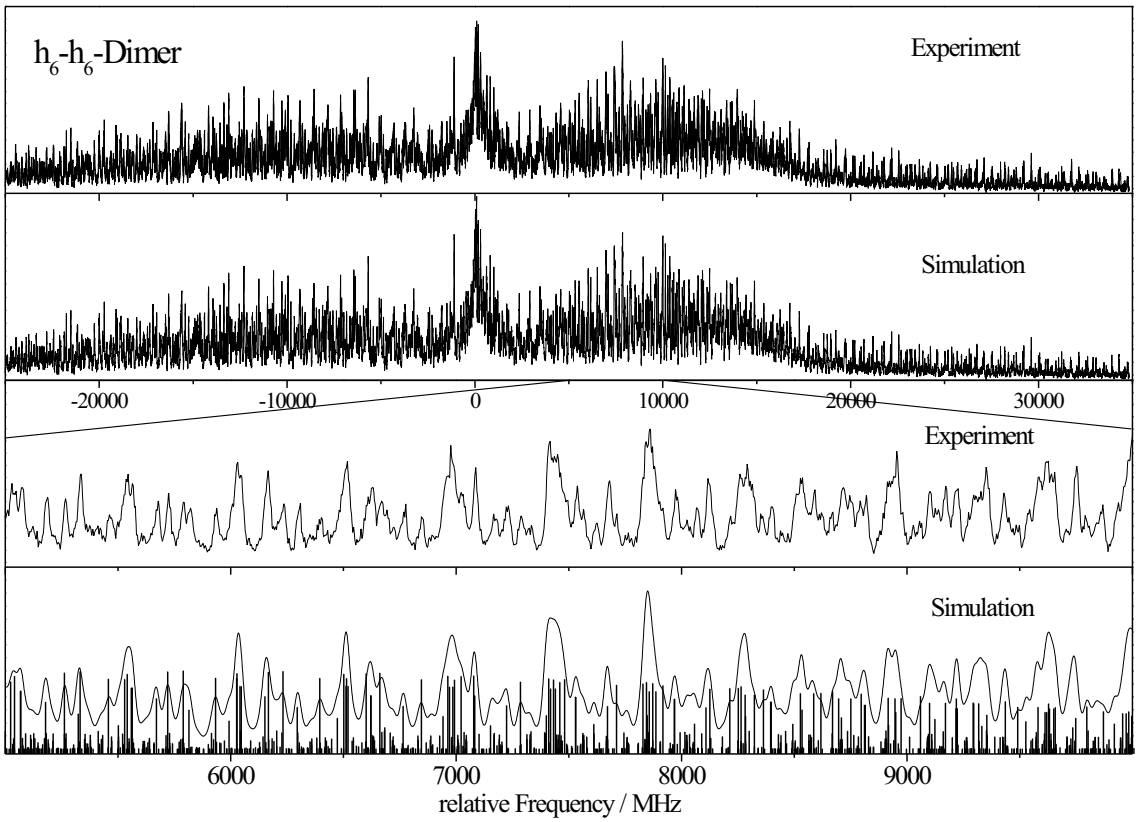

Figure 12: Rovibronic spectrum of the electronic origin of the $h_6 - h_6$ isotopologue of the phenol dimer at 36044.70 cm^{-1}. The following traces show the simulated spectrum using the rotational constants from Table 3, a zoomed portion of the experimental spectrum and the zoomed simulation together with the stick spectrum.

Table 3: Molecular parameters of the electronic origin band of five isotopologues of the phenol dimer obtained from genetic algorithm fits. Rotational constants and their differences in the electronically excited state are given in MHz, the angles of the transition dipole moment ϕ and θ with the inertial axes are given in degrees.

	$h_6 - h_6$	$d_1 - d_1$	$d_1 - h_6$	$d_6 - d_6$	^{13}C-^{13}C
A''	1416.99(39)	1376.23(10)	1399.78(124)	1239.33(12)	1413.22(15)
B''	313.51(1)	312.84(3)	312.72(6)	287.25(1)	311.84(2)
C''	288.11(1)	286.44(2)	287.16(6)	264.63(1)	286.70(2)
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)
ΔA	10.71(1)	4.55(1)	7.81(10)	4.28(2)	10.60(1)
ΔB	−5.31(1)	−3.95(1)	−5.63(1)	−3.93(1)	−5.25(1)
ΔC	−5.82(1)	−4.76(1)	−5.99(1)	−4.90(1)	−5.76(1)

to the TDM orientations of the five isotopologues given in Table 3 as described in section 3.5. Table 4 and Figure 13 show the bond lengths, angles and dihedral angles which determine the relative orientation of the to monomer moieties with respect to each other for the ground and the first excited state.

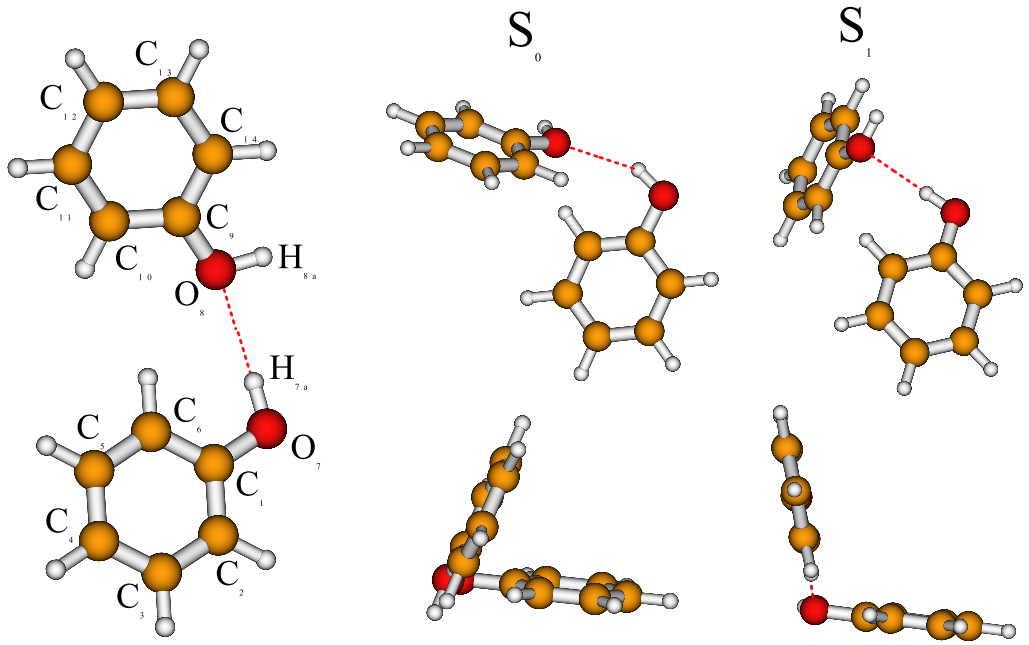

Figure 13: Left: Atomic numbering of the phenol dimer, used in the definitions of Table 4. Right: Two different views of the S_0 (r_s) and the S_1 (r_s) geometry of the phenol dimer from Table 4.

Table 4: Experimentally determined pseudo-Kraitchman (r_s) geometry parameters of the phenol dimer in the S_0 and in the S_1-state. All distances are given in pm, angles and dihedral angles in degrees. For the fit of the intermolecular structure of the dimer in the S_1 state we used a locally excited donor moiety.

	$S_0(r_s)$	$S_1(r_s)$
$r(H_{7a}O_8)$	236.9(54)	174.4(16)
$a(O_7H_{7a}O_8)$	150.5(20)	179.9(1)
$a(C_9O_8H_{7a})$	139.8(14)	135.7(9)
$d(O_8H_{7a}O_7C_1)$	9.4(47)	63.4(9)
$d(C_9O_8H_{7a}O_7)$	62.8(46)	−114.8(9)
$d(C_{10}C_9O_8H_{7a})$	−178.2(22)	−101.5(14)

The geometries shown in Table 4 and Figure 13 give clear evidence, that the hydrogen bond strength of the phenol dimer increases upon electronic excitation, while the dispersion interaction between the monomer moieties decreases.

5.2.2 The benzonitrile dimer

The dimer of benzonitrile displays an exception from the other cases in this chapter, since no O-H$\cdots$O hydrogen bond is formed. Nevertheless, the monomer moieties are bound by a

hydrogen bond, which is a (weaker) C-H$\cdots$N bond in this case. Since the C-H groups in ortho position to the nitrile group are quite acidic, there can be formed a quite strong hydrogen bond between the nitrile group acting as proton acceptor and the CH group acting as proton donor. Of main concern is the symmetry and the structure of the dimer in the ground and excited states.

The spectra and their assignment Figure 14 shows the rovibronic spectrum of the electronic origin of the benzonitrile dimer together with the best GA fit using the molecular parameters from Table 5. The lowest trace shows a zoomed part of the convoluted simulation along with the calculated stick spectrum.

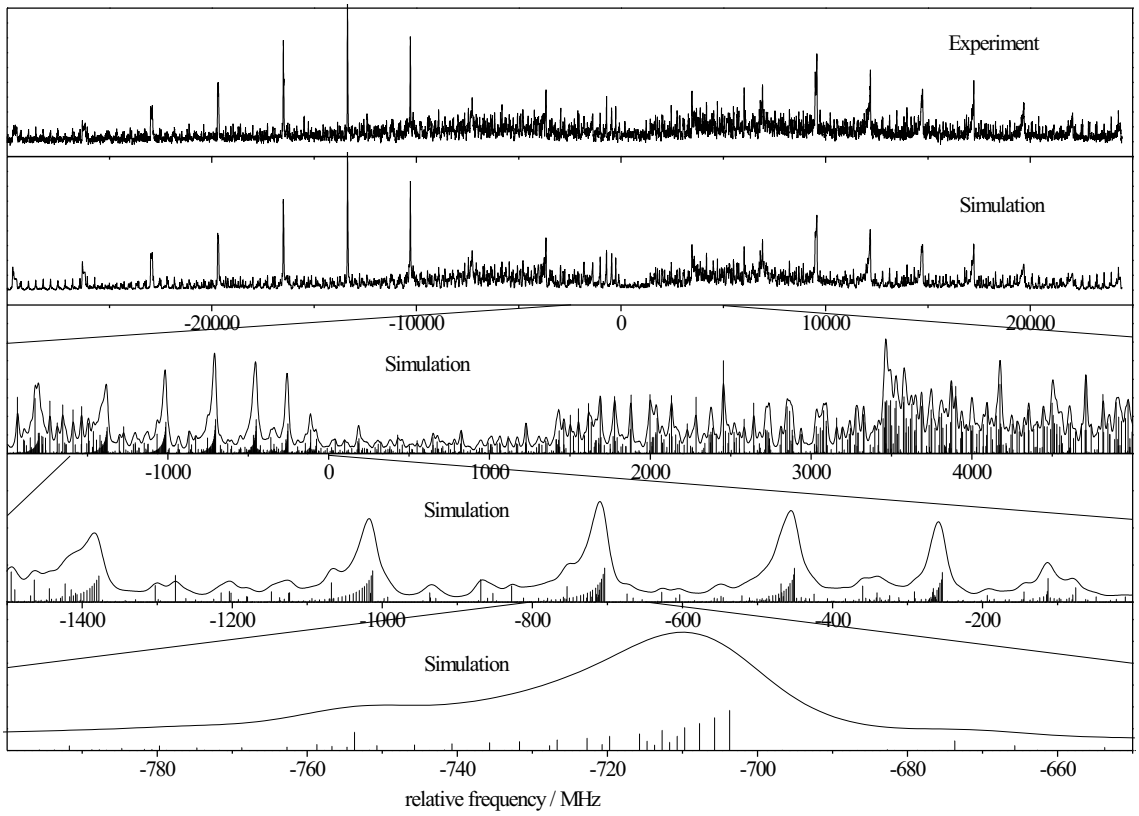

Figure 14: Rovibronic spectrum of the electronic origin of the benzonitrile dimer at 36420.10 cm^{-1}. The traces show the experimental spectrum, the simulated spectrum using the rotational constants from Table 5, a zoomed portion of the experimental spectrum and the zoomed simulation together with the stick spectrum.

The spectrum is fit using a rigid rotor Hamiltonian. It is of mixed *abc*-type and consists of about 18000 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 origin frequencies, the polar angles θ and ϕ, defining the orientation of the transition dipole moment in the inertial frame, and the Lorentzian contribution to a Voigt profile with 15 MHz Gaussian contribution are given in Table 5.

Comparison of the stick spectrum with its convolution using the experimental line widths immediately reveals, why a "classical" fit with line position assignments is impossible for a molecular system of that size. Each "line" in the experimental spectrum is composed of up

Table 5: Molecular parameters of the electronic origin band of the benzonitrile dimer as obtained from the genetic algorithm fit. Standard deviations are given in parentheses. The MP2 calculations were performed with the G03 program package (Frisch et al. 2003) using the 6-311G(d,p) basis set. RICC2 calculations were performed with Turbomole V5.8 using the VTZP basis set from the turbomole library (Ahlrichs et al. 1989) for both the ground and first electronically excited $\pi\pi^*$ states.

	Experiment	MP2	RICC2
A''/MHz	1610.99(17)	1654	1645
B''/MHz	187.48(7)	185	186
C''/MHz	168.12(7)	167	167
$\Delta I''$/uÅ^2	-3.2727	0	0
ν_0/cm^{-1}	36420.10(1)	-	37851
ΔA/MHz	-28.08(10)	-	-28
ΔB/MHz	-0.01(10)	-	0
ΔC/MHz	-0.31(10)	-	0
$\Delta\Delta I$/uÅ^2	-0.1667	0	0

to 40 single rovibronic transitions, in many cases of very similar intensities. Figure 14 shows an expanded view of the simulated stick spectrum along with the convoluted spectrum using a Gaussian width of 15 MHz and a Lorentzian contribution of 10 MHz.

Determination of the structural parameters The structure of the benzonitrile dimer has been determined from the rotational constants of a single isotopologues. Therefore, quite a few model assumptions have to be made. For the electronic ground state, the two monomer moieties have been fit to the microwave transitions from Ref. (Casado et al. 1971). Subsequently, the intermolecular parameters of the dimer are fit, assuming unchanged monomer geometries upon cluster formation. The result is shown in Figure 15.

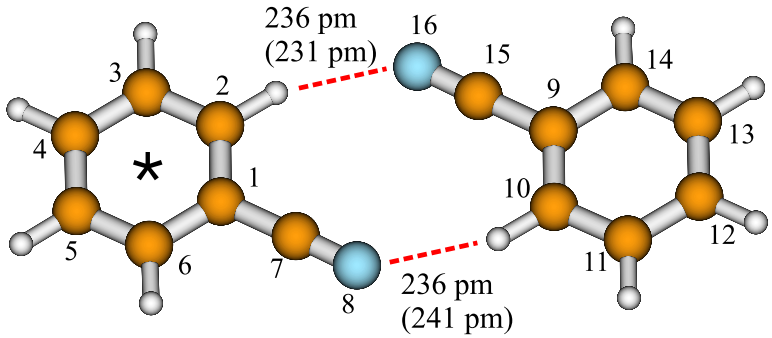

Figure 15: The structure of the benzonitrile dimer. The hydrogen bond lengths for the electronic ground state and the excited state (in parentheses) are given as obtained from the structure fit to the rotational constants. The asterisk denominates the benzonitrile moiety, which is electronically excited.

The S_1 structure for the dimer is fitted assuming local excitation in one of the monomer moieties. The geometry of one monomer was fitted to the rotational constants of the excited state of benzontrile of Borst *et al.*(Borst et al. 2001), while the other was kept fixed at the

Kraitchman geometry of the ground state. For the fit of the excited state of the monomer all CC bond lengths in the aromatic ring are treated equally and are 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.

In this way we could fit the excited state structure of the dimer, resulting in one considerably shorter N_8H_{10a} bond length in the electronically excited state compared to the ground state and one longer one.

The benzonitrile dimer exhibits a local excitation of one of the cluster moieties like the benzoic acid dimer. Obviously, the electronic coupling between the two monomer moieties is not strong enough to allow a delocalized electronic excitation.

5.2.3 The benzoic acid dimer

One of the most well known symmetric aromatic homo-dimers is the benzoic acid dimer. Since the C_1C_7 axes of the monomer moieties are nearly on a straight line, the A rotational constant of the cluster will be larger than in case of the benzonitrile dimer. Nevertheless, the cluster is extended considerably more in the direction of the A axis compared to the benzonitrile dimer, leading to much smaller B and C rotational constants. As will be seen in the following, the need for an automated assignment tool like ES is even larger in this case, than in any other, described hitherto.

The spectra and their assignment As in the case of the phenol dimer several isotopologues have been investigated. In the following we show the results for only one of them, exemplarily. The investigated isotopologues are the undeuterated (d_0d_0), two singly deuterated (d_1d_0 and d_0d_1), a doubly deuterated (d_1d_1) and the fully deuterated (d_5d_5) isotopologue. Single deuteration (d_1) in the above examples of isotopologues takes place at the hydrogen of the carboxy group. Figure 16 shows the rotationally resolved electronic origin of the d_0d_0 isotopologue of the benzoic acid dimer. The lowest trace shows a zoomed part of the convoluted simulation along with the calculated stick spectrum. The red and the blue sticks represent transitions within the two sublevels, caused by the tunneling of the two carboxy protons in the dimer. In such a dense spectrum one cannot talk about rovibronic lines, this is clearly a limit where the term rovibronic contour is more appropriate, despite the high experimental resolution. Each feature in this contour consists of a few hundred single rovibronic transitions, clearly showing why the utilization of an evolutionary strategy is indispensable in this case. The benzoic acid dimer spectra were the first to be fit using the derandomized evolutionary strategy (ES DR2), cf. section 3.2.

Determination of the structural parameters The main concern about the structure of the benzoic acid dimer is the symmetry of the cluster in the ground and electronically excited state. While all experimental investigations for the ground state agree in a C_{2h} symmetry, i.e. identical monomer moieties, the symmetries of the electronically excited state and the transition states for both ground and excited states are in question. The investigations described here show, that the symmetry of benzoic acid dimer changes to C_s upon electronic excitation. Structurally this can be explained by a tilting of the two monomer moieties, caused by the fact that they are no longer equivalent in the excited state. Since electronic excitation is as-

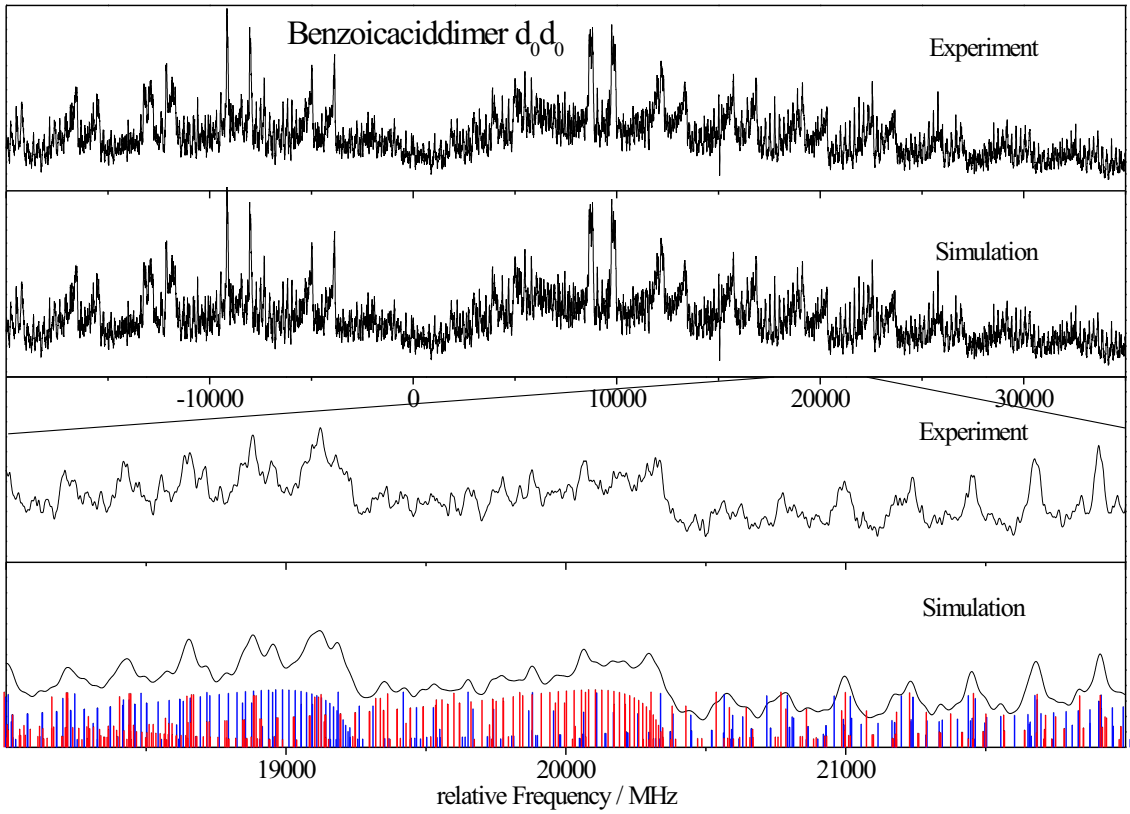

Figure 16: Rovibronic spectrum of the electronic origin of the d_0d_0 isotopologue of the benzoic acid dimer at 35723.79 cm^{-1}. The traces show the experimental spectrum, the simulated spectrum using the rotational constants from Table 6, a zoomed portion of the experimental spectrum and the zoomed simulation together with the stick spectrum.

Table 6: Molecular parameters of the electronic origin band of the benzoic acid dimer as obtained from the EA fit. Standard deviations are given in parentheses. The MP2 calculations were performed with the G03 program package (Frisch et al. 2003) using the 6-311G(d,p) basis set. RICC2 calculations were performed with Turbomole V5.8 using the VTZP basis set from the turbomole library (Ahlrichs et al. 1989) for both the ground and first electronically excited $\pi\pi^*$ states.

	Experiment	MP2	RICC2
A''/MHz	1924.99(20)	1919	1925
B''/MHz	128.06(14)	127	128
C''/MHz	120.15(14)	119	120
$\Delta I''$/uÅ^2	-2.5139	0	0
ν_0/cm^{-1}	35723.786(1)	-	36789
ϕ/°	73.18	-	89.9
θ/°	79.00	-	90.0
ΔA/MHz	-16.07(9)	-	-16
ΔB/MHz	-0.52(12)	-	-1
ΔC/MHz	-0.51(12)	-	0
$\Delta I'$/uÅ^2	-2.8210	-	0

sociated with a change of acidity as in the case of water clusters of several phenol derivatives (cf. section 5.1.2).

5.3 Determination of a large number of molecular parameters

Evolutionary strategies are especially successful in cases were the parameter space is very large and/or unstructured. In the following two examples we will present results of spectral fits, in which a large number of parameters have to be fit simultaneously.

5.3.1 The $2_0^1 4_0^3$ rovibrational band of the $\widetilde{A}^1A_2 \leftarrow \widetilde{X}^1A_1$ electronic transition of formaldehyde

Motsch et al. (2008) have presented the rotationally resolved room temperature absorption spectrum of the $2_0^1 4_0^3$ and the $2_0^2 4_0^1$ the rovibronic bands of formaldehyde in the $\widetilde{A}^1A_2 \leftarrow \widetilde{X}^1A_1$ electronic transition. Due to the high temperature, high J-states are populated and sextic centrifugal distortion terms in the fit have shown to be necessary. The appropriate nuclear spin statistics (Ka even levels = 1; Ka odd levels = 3) has also been taken into account. In this case the number of parameters to be fit amounts to 6 rotational constants (3 for each state), 10 quartic centrifugal distortion constants (5 for each state), 14 sextic centrifugal distortion constants (7 for each state), and one central frequency = 31 parameters. This is a typical case for the application of evolutionary strategies, since the advantage of these techniques over classical (local) optimizers is increasing with increasing parameter space.

5.3.2 The case of many overlapping isotopologues: tryptamine

Another case of high number of parameters to be fit occurs, when several isotopologues of one molecule absorb within the rovibronic band width. A case that has been studied extensively by us is the tryptamine system. Addition of D_2O to tryptamine results in a mixture of 8 different readily forming isotopologues due to three acidic positions (the pyrrolic NH and the two amino H atoms) with the possibility of H↔D exchange: 1 undeuterated, 3 singly deuterated, 3 doubly deuterated, and one triply deuterated isotopologues.(Schmitt et al. 2005) All eight rovibronic spectra completely overlap, and a total of 61 parameters have to be fit for the whole spectrum(six rotational constants and a central frequency for each of the isotopologues and additionally two temperatures, a temperature weight factor (cf. equation 4) and two line shape parameters. Regarding the high line density due to the small rotational constants and the large number of parameters to be fit no other method as evolutionary strategies seem to be able to tackle this problem.

6 GA fit of rovibronic or rovibrational contours

The examples, shown above of very large molecular complexes pose the question if the fitting technique is still applicable when the information from individual transitions or from groups of transitions is completely lost due to heavy spectral overlap. In the following we will show, that the GA based fitting strategy can also be used in cases, were the spectral resolution is so low, that only rotational contours can be resolved. This might occur in electronic spectroscopy, when the life time of the excited state is considerably smaller than 500 ps, leading to large Lorentzian broadening. Other limitations might be a 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.

6.1 The contours of adenine rovibronic spectra

The medium resolved (the spectral resolution in these pulsed laser experiment is 0.04 cm^{-1} corresponding to 1200 MHz) LIF spectra of selected vibronic bands of adenine, shown in Figures 17 and 19 are an example of insufficient rotational resolution in electronic spectra due to a very short excited state lifetime.(Lee et al. 2006) The upper trace in Figure 17 shows the experimental LIF spectrum of the C-band of adenine, which has been subject to a long lasting controversy about its nature. Although, no single rovibronic features are observed a general structure with P, Q, and R branch is visible. The second trace shows the GA fit of the band. The stick spectrum, which is shown in a zoomed central part of the simulated spectrum clearly shows, that the fact that no resolved spectrum as in the so far discussed cases is observed but only a rovibronic contour is not a consequence of a very high line density, but indeed caused by a broadening of each individual line by the short life time of adenine. The lowest trace shows a Voigt convolution of the stick spectrum with a Gaussian and a Lorentzian width of 25 MHz, respectively. This would be equivalent to a (hypothetical) excited state life time of 6 ns, which is a typical value for indole-like chromophores. Clearly, a beautifully resolved spectrum could be expected with such a long life time.

The fit of a rovibronic contour like shown in Figure 17 is not as straightforward as the GA fits described in the previous sections. Many molecular parameters have to be fixed since the fit problem is severely under determined. Fortunately the microwave spectrum of adenine was measured, so that the ground state rotational constants are known. Thus there were only a limited number of parameters to fit: Three changes of rotational constants upon electronic excitation, the direction of the transition dipole moment (TDM), the rotational temperature, and the Lorentzian line width. The most important parameters are those which determine the direction of the TDM, because they give important indications to the nature of the electronically excited state.

Other bands, that appear in the low resolution spectrum of adenine are the D and the E band. Both bands have rovibronic contours that are completely different from that of the adenine C band. Figure 19 shows the experimental spectrum and the simulation of the E band, which is very similar to the contour of the D band.

The main contribution to the intensity of the C band arises from *c*-type transitions, indicating an out-of-plane transition moment, while for the E and D bands mainly in-plane transition moment components (*a* and *b*) dominate. Using this information the C band can be identified as the electronic origin of the (out-of-plane) polarized $n\pi^*$ state, the D band as the origin of the (in-plane) polarized $\pi\pi^*$ state, and the E band as an out of plane vibrational band of the $n\pi^*$ state. This out-of-plane vibration is forbidden, but gains intensity via a vibronic coupling (Herzberg 1966) of the $n\pi^*$ and the $\pi\pi^*$ state. The large intensity of the electronically forbidden E band points to a large vibronic coupling between these two states.

6.2 The Fourier transform infrared spectrum of benzotriazole

Another example of a spectrum for which only the rotational contour of a vibrational band in spite of a fully resolved rovibrational spectrum was obtained is the Fourier transform infrared (FTIR) spectrum of benzotriazole in the gas phase which was examined by Roth et al. (1999). Two tautomers contribute to the IR absorption in the observed region, namely the 1*H*- and

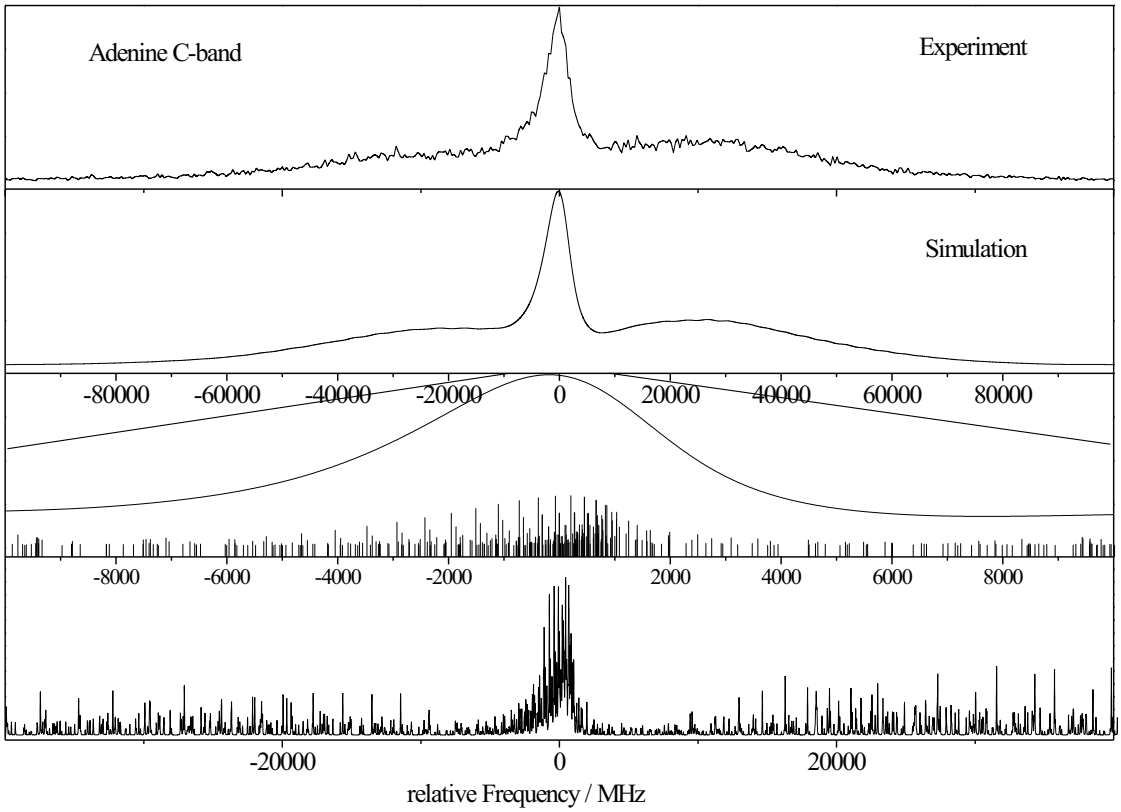

Figure 17: Rovibronic spectrum of the C-band of adenine at 36061.8 cm^{-1} along with the simulation using the best parameters from the GA fit. The subsequent trace shows a zoomed part of the spectrum with the stick spectrum overlayed. The lowest trace shows a convolution of the complete stick with a Voigt profile using Gaussian and Lorentzian widths of 25 MHz, respectively.

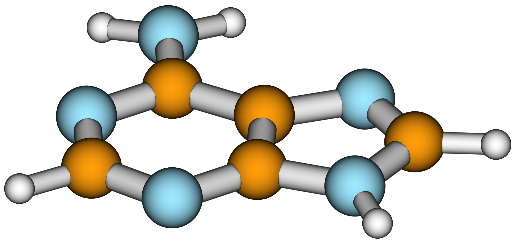

Figure 18: The structure of adenine.

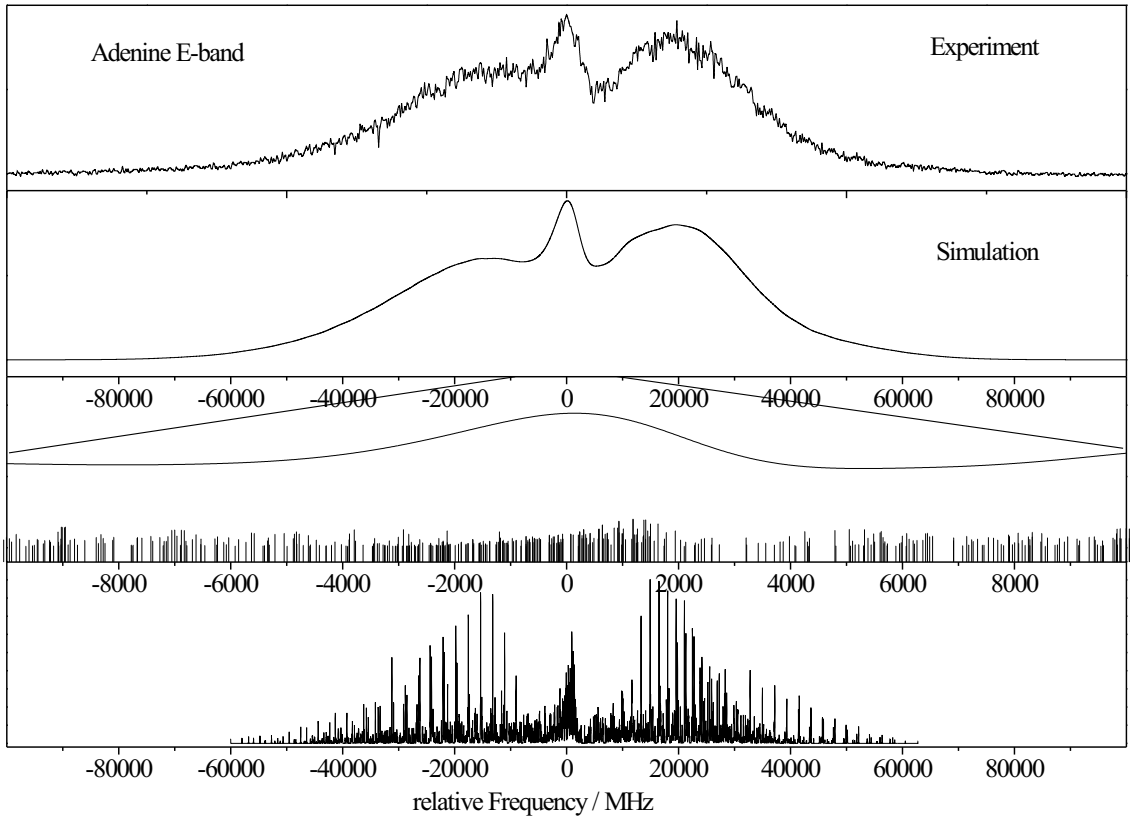

Figure 19: Rovibronic spectrum of the E-band of adenine at 36247.7 cm^{-1} along with the simulation using the best parameters from the GA fit. The subsequent trace shows a zoomed part of the spectrum with the stick spectrum overlayed. The lowest trace shows a convolution of the complete stick with a Voigt profile using Gaussian and Lorentzian widths of 25 MHz, respectively.

the $2H$-tautomer, which are shown in Fig. 20. The reason for the lack of rotational resolution in this example are the high temperature (400 K) and connected therewith the very high line density and Doppler width along with the limited resolution (0.5 cm^{-1}) of the FTIR spectrometer. Figure 21 shows the FTIR spectrum, which has been taken between 3450 and 3550 cm^{-1} in the region of the NH stretching vibration. The NH stretching vibrations of the two tautomers have different absorption frequencies, and also different band types. The directions of the changes of the dipole moments are shown in the inertial axis system for each tautomer in Figure 20. Obviously, in the $2H$ tautomer a pure a-type band will be observed, while for $1H$-benzotriazole an ab-hybrid type can be expected.

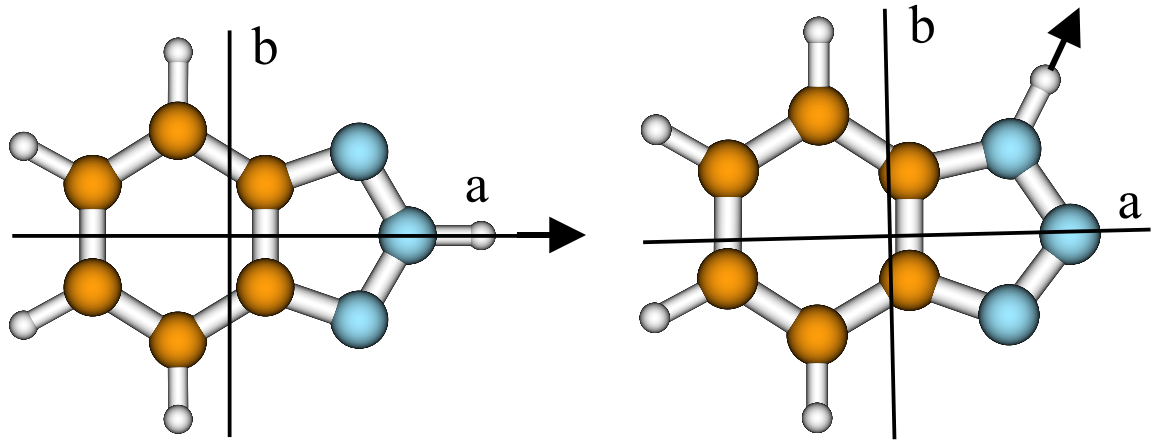

Figure 20: Structure of $1H$-benzotriazole (right) and $2H$-benzotriazole (left). The arrow shows the direction of the dipole moment change upon NH stretching vibration.

Also in this case, the GA fit is simplified by the fact that the ground state rotational constants are known from microwave spectroscopy (for the $1H$-tautomer) and from rotationally resolved electronic spectroscopy (for the $2H$-tautomer). The NH-stretching band of the $2H$-tautomer must be a pure a-type band from symmetry considerations (cf. Figure 20). The hybrid type of $1H$-benzotriazole, the changes of the rotational constants upon vibrational excitation, the center frequencies of both bands, and the relative intensity of both bands have been fit. The lowest trace in Figure 21 shows a small portion of the simulated stick spectrum. The blue stick represent transitions of the $2H$-tautomer, the red sticks of the $1H$-tautomer. Clearly, t is the high temperature of the sample, which makes it impossible in this case to obtain rovibrational resolution.

6.3 Concluding remarks to the fit of rotational contours

The above examples clearly show, that the technique of automated fits using the GA also is successful for spectra which lack full rotational resolution. Of course, the more spectral features are resolved, the more accurate is the determination of the molecular parameters and the less parameters have to be fixed in the fit. High temperatures are sometimes advantageous when the resolution is limited by the experimental set up, since the large number of rotational levels which are occupied at high temperatures, makes spectral features, which are causes by band heads more easily visible. In this context, the reader is referred to the pioneering work of Hollas on the analysis of rotational band contour analysis of vibronic transitions at elevated temperatures.(Hollas 1982)

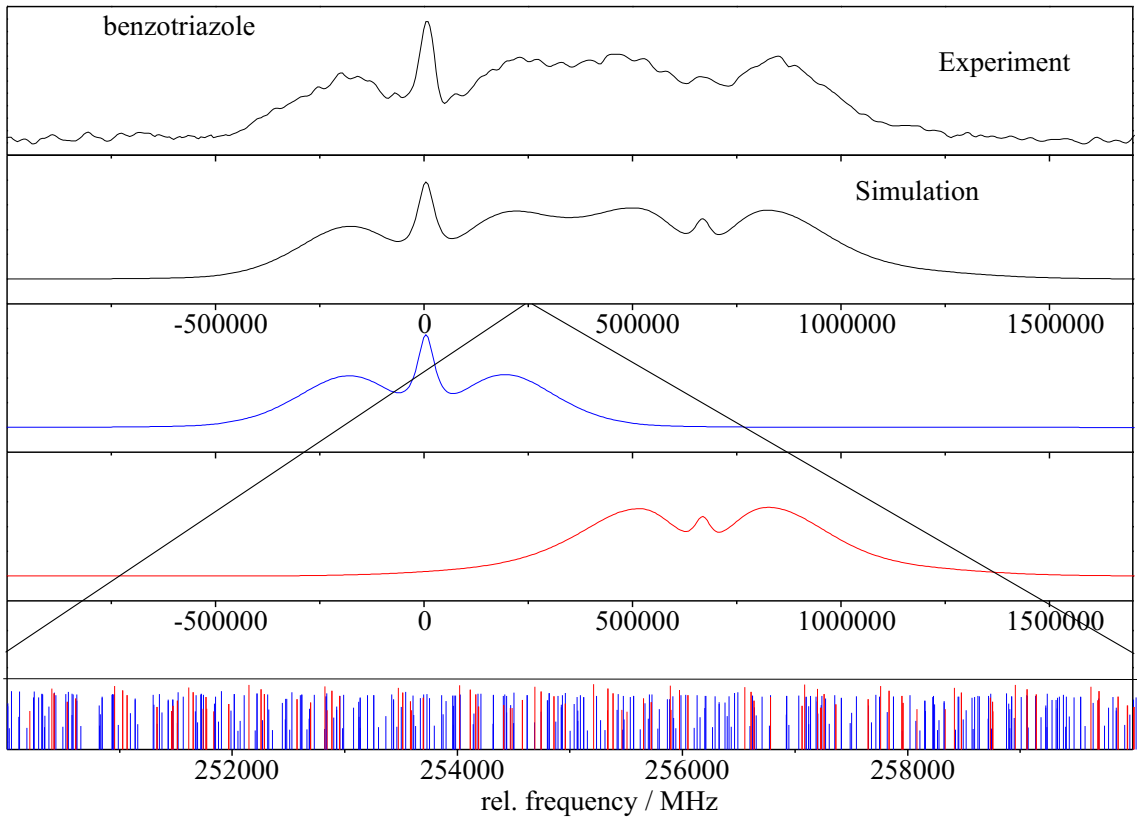

Figure 21: FTIR Spectrum of benzotriazole, taken at 400 K, along with the individual simulations of both tautomers. The lowest trace shows the stick spectra of both tautomers in a zoomed 10 GHz window.

7 Analysis of liquid crystal NMR spectra

As further example of the use of EA we will now discuss a completely different type of high resolution spectra: Nuclear Magnetic Resonance (NMR) spectra of molecules which are partially oriented in an anisotropic liquid-crystalline environment. The NMR spectroscopy of orientationally ordered solutes in anisotropic liquids yields a wealth of information about both the solute and the intermolecular interactions between the solute and the solvent (Burnell and de Lange 2003). The extension to solutes that can exist in several symmetry-unrelated conformers is both interesting and challenging. The conformer distribution in such flexible molecules is governed by the total potential which in the condensed phase includes both intra- and inter-molecular terms. This distribution can also be affected by orientational ordering. The extent of both isotropic and anisotropic effects in the solvent is difficult to measure. Because dipolar couplings between nuclei on different rigid subunits of a flexible molecule are very sensitive to conformer geometry and orientational order they offer an excellent route to gaining insight into this difficult problem. Up to now the measurement of such dipolar couplings has been hampered by the difficulty of analysis of the complicated spectra that are obtained.

The number of independent intramolecular pair interactions determines the complexity of assignment of these NMR spectra. This number is small for molecules with a limited number of spins (≤ 6) and high symmetry. For solutes with more spins and less symmetry the number of different dipole-dipole interactions increases rapidly, thus making the analysis of complex NMR spectra notoriously difficult. This in spite of the fact that in most cases the resolution is sufficient high that little overlap in the transitions occurs. The origin of the difficulties in the analysis lies in the nature of the NMR Hamiltonian: The spectra usually lack easy recognizable regularities. Many of such spectra have therefore remained intractable for decades.

The dominating terms in the Hamiltonian of a molecule dissolved in a uniaxial nematic liquid crystal are given in Eq. 5. The number of dipolar interactions $D_{\mu\nu}$ grows strongly with the number of spins in the molecule and their values can easily differ by an order of magnitude. In an attempt to reduce the number of parameters and to estimate the strength of the dipole-dipole interactions Meerts et al. (2007) have used the relations between the Saupe order matrix S_{kj} and the direct dipole-dipole interactions:

$$D_{\mu\nu} \;=\; -\frac{h\gamma_\mu\gamma_\nu}{4\pi^2}\left\langle r_{\mu\nu}^{-3}\right\rangle \sum_{k,l} S_{kl}\cos\theta_k\cos\theta_l \tag{13}$$

where θ_k is the angle between the k molecule-fixed Cartesian axis and the vector connecting nuclei μ and ν at internuclear distance $\mathrm{r}_{\mu\nu}$. The angular brackets denote averaging over all internal motions.

The order matrix S_{kl} is defined as averages over reorientational motion of the second-order Legendre polynomials:

$$S_{kl} \;=\; \frac{1}{2} < 3\cos\theta_{Z,k}\cos\theta_{Z,l} - \delta_{kl} > \tag{14}$$

where $\cos\theta_{Z,k}$ is the angle between the molecule-fixed k axis and the space-fixed Z axis which lies along the liquid-crystal director. The S tensor is traceless and symmetric.

In this way it is possible to make a crude prediction of the dipole-dipole interaction from an estimated structure of the molecule.

The way Meerts et al. (2007) tackled the NMR spectrum of the nine-spin system p-bromobiphenyl dissolved in a nematic crystal was by a two step process. In the first step a relatively

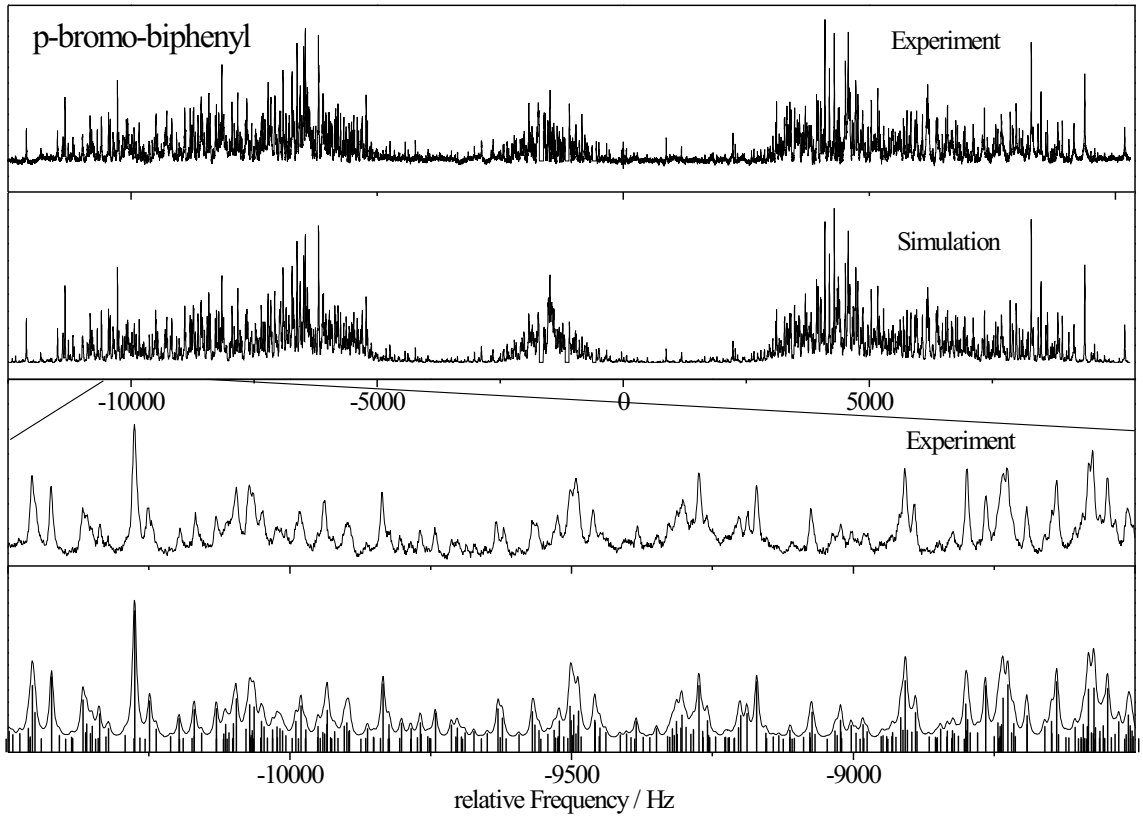

Figure 22: ^{1}H NMR spectrum of *p*-bromo-biphenyl dissolved in 'magic mixture'. The experimental spectrum is compared with the final fit, a zoomed portion of the experimental spectrum and the zoomed simulation together with the stick spectrum. The horizontal scales are in Hz. (Reproduced from Meerts et al. (2007) with permission from Elsevier Publishers.)

crude geometrical structure of the solute is assumed and two independent order parameters, chemical shielding parameters and the dihedral angle between the two rings are varied in a GA-fitting procedure until the best correspondence with the experimental spectrum is obtained. Owing to the approximations made in this step, complete agreement cannot be obtained. However, in practice the choice of geometry does not have to be perfect in order to obtain sufficient common features between calculated and experimental spectra. Approximate dipolar couplings are calculated in this step. In a second GA-calculation procedure, these quantities are used as starting values for a fit in which all dipole-dipole couplings and chemical shielding parameters are varied until convergence is reached. Exceptionally good agreement is obtained in step two because no constraints are placed on the dipolar coupling and chemical shielding parameters. A typical result is presented in Figure 22.

Acknowledgment

We 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 and Marloes van Beek for their contributions to this work. The financial support of the Deutsche Forschungsgemeinschaft (SFB663 project A2) is gratefully acknowledged. M. S. likes to thank the Nordrheinwestfälische Akademie der

Wissenschaften for a grant, which made this work possible. The authors 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 (SCHM1043/10-1).

References

Abe H, Mikami N and Ito M 1982 *J. Phys. Chem.* **86**, 1768–1771.

Ahlrichs R, Bär M, Häser M, Horn H and Kölmel C 1989 *Chem. Phys. Letters* **162**, 165–169.

Allen H C and Cross P C 1963 *Molecular Vib-Rotors* Wiley New York.

Becucci M, Lakin N M, Pietraperzia G, , Salvi P R, Castellucci E and Kerstel E R T 1997 *J. Chem. Phys.* **107**, 10399–10405.

Berden G, Meerts W L and Jalviste E 1995 *J. Chem. Phys.* **103**, 9596–9606.

Berden G, Meerts W L, Schmitt M and Kleinermanns K 1996 *J. Chem. Phys.* **104**, 972–982.

Borst D R, Korter T M and Pratt D W 2001 *Chem. Phys. Letters* **350**, 485–490.

Brandt S 1998 *Data Analysis* 3 edn Springer Verlag Berlin.

Burnell E E and de Lange C A, eds 2003 Kluwer Academic Dordrecht, The Netherlands.

Casado J, Nygaard L and Sørensen G O 1971 *J. Mol. Struct.* **8**, 211–224.

Connell L L, Corcoran T C, Joireman P W and Felker P M 1990 *Chem. Phys. Letters* **166**, 510–516.

Deaven D M and Ho K M 1995 *Phys. Rev. Letters* **75**, 288–291.

Felker P M 1992 *J. Phys. Chem.* **96**, 7844–7857.

Frisch M J, Trucks G W, Schlegel H B, Scuseria G E, Robb M A, Cheeseman J R, Montgomery J A, Jr., Vreven T, Kudin K N, Burant J C, Millam J M, Iyengar S S, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson G A, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox J E, Hratchian H P, Cross J B, Adamo C, Jaramillo J, Gomperts R, Stratmann R E, Yazyev O, Austin A J, Cammi R, Pomelli C, Ochterski J W, Ayala P Y, Morokuma K, Voth G A, Salvador P, Dannenberg J J, Zakrzewski V G, Dapprich S, Daniels A D, Strain M C, Farkas O, Malick D K, Rabuck A D, Raghavachari K, Foresman J B, Ortiz J V, Cui Q, Baboul A G, Clifford S, Cioslowski J, Stefanov B B, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin R L, Fox D J, Keith T, Al-Laham M A, Peng C Y, Nanayakkara A, Challacombe M, Gill P M W, Johnson B, Chen W, Wong M W, Gonzalez C and Pople J A 2003 'Gaussian 03, revision a.1' Gaussian, Inc., Pittsburgh, PA.

Fuke K and Kaya K 1983 *Chem. Phys. Letters* **94**, 97–101.

Gordy W and Cook R L 1984 *Microwave Molecular Spectra* 3 edn Wiley New York.

Granucci G, Hynes J, Millié P and Tran-Thi T H 2000 *J. Am. Chem. Soc.* **122**, 12243–12253.

Hageman J A, Wehrens R, de Gelder R, Meerts W L and Buydens L M C 2000 *J. Chem. Phys.* **113**, 7955–7962.

Hansen N and Kern S 2004 *in* X Yao et al., eds, 'Parallel Problem Solving from Nature PPSN VIII' Vol. 3242 of *LNCS* Springer pp. 282–291.

Hansen N and Ostermeier A 2001 *Evolutionary Computation* **9**(2), 159–195.

Held A and Pratt D W 1992 *J. Chem. Phys.* **96**, 4869–4876.

Helm R M, Vogel H P and Neusser H J 1997 *Chem. Phys. Letters* **270**, 285–291.

Herzberg G 1966 *Molecular Spectra and Molecular Structure, III. Electronic Spectra and Electronic Structure of Polyatomic Molecules* van Nostrand Reinhold Company New York.

Holland J H 1975 *Adaption in Natural and Artificial Systems* MI: The University of Michigan Press Ann-Arbor.

Hollas J M 1982 *High Resolution Spectroscopy* Butterworth & Co Publishers London.

Jacoby C, Böhm M, Vu C, Ratzer C and Schmitt M 2006 *ChemPhysChem* **7**, 448–455.

Lahmani F, Broquier M and Zehnacker-Rentien A 2002 *Chem. Phys. Letters* **337–348**, 354.

Lee Y, Kim B, Schmitt M and Kleinermanns K 2006 *J. Phys. Chem. A* **110**, 11819–11823.

Li Z Q and Scheraga H A 1987 *Proc. National Academy of Science* **84**, 6611–6615.

Lipert R J and Colson S D 1988 *J. Chem. Phys.* **89**, 4579–4585.

Loomis F W and Wood R W 1928 *Phys. Rev* **32**, 223–236.

Majewski W A and Meerts W L 1984 *J. Mol. Spec.* **104**, 271–281.

Majewski W A, Plusquellic D F and Pratt D W 1989 *J. Chem. Phys.* **90**, 1362–1367.

Meerts W L, de Lange C A, Weber C J C and Burnell E E 2007 *Chem. Phys. Letters* **441**, 342–364.

Meerts W L, de Lange C A, Weber C J C and Burnell E E 2009 *J. Chem. Phys.* **139**, 044504(1–8).

Meerts W L and Schmitt M 2006 *Int. Rev. Phys. Chem.* **25**, 353–406.

Meerts W L, Schmitt M and Groenenboom G 2004 *Can. J. Chem.* **82**, 804–819.

Motsch M, Schenk M, Zeppenfeld M, Schmitt M, Meerts W L, Pinkse P W and Rempe G 2008 *J. Mol. Spec.* **252**, 25–30.

Müller A, Talbot F and Leutwyler S 2002 *J. Chem. Phys.* **116**, 2836–2847.

Myszkiewicz G, Meerts W L, Ratzer C and Schmitt M 2005 *J. Chem. Phys.* **123**, 044304–1–044304–7.

Nguyen T, Korter T and Pratt D 2005 *Mol. Phys.* **103**, 1603–1613.

Ostermeier A, Gawelczyk A and Hansen N 1994 *Step-Size Adaptation Based on Non-Local Use of Selection Information. In Parallel Problem Solving from Nature (PPSN3)* Springer. this reference seems incomplete?

Park Y D, Rizzo T R, Peteanu L A and Levy D H 1986 *J. Chem. Phys.* **84**, 6539–6549.

Poeltl D E and McVey J K 1984 *J. Chem. Phys.* **80**, 1801–1811.

Ratzer C, Küpper J, Spangenberg D and Schmitt M 2002 *Chem. Phys.* **283**, 153–169.

Rechenberg I 1973 *Evolutionsstrategie - Optimierung technischer Systeme nach Prinzipien der biologischen Evolution* Frommann-Holzboog Stuttgart.

Remmers K, Jalviste E, Mistrk I, Berden G and Meerts W L 1998 *J. Chem. Phys.* **108**, 8436–8445.

Remmers K, Meerts W L and Ozier I 2000 *J. Chem. Phys.* **112**, 10890–10894.

Riehn C 2002 *Chem. Phys.* **297**, 283.

Ross J, ed. 1966 *Molecular Beams* Advances in Chemical Physics Interscience Publishers New York, London, Sydney.

Roth W, Spangenberg D, Janzen C, Westphal A and Schmitt M 1999 *Chem. Phys.* **248**, 17.

Schmitt M, Böhm M, Ratzer C, Vu C, Kalkman I and Meerts W L 2005 *J. Am. Chem. Soc.* **127**, 10356–10364.

Schmitt M, Böhm M, Siegert C R S, van Beek M and Meerts W L 2006 *J. Mol. Struct.* **795**, 234–241.

Schmitt M, Krügler D, Böhm M, Ratzer C, Bednarska V, Kalkman I and Meerts W L 2006 *Phys. Chem. Chem. Phys.* **8**, 228–235.

Schmitt M, Küpper J, Spangenberg D and Westphal A 2000 *Chem. Phys.* **254**, 349–361.

Schütz M, Bürgi T, Leutwyler S and Fischer T 1993 *J. Chem. Phys.* **98**, 3763–3776.

Schwevel H P 1993 *Evolution and Optimium Seeking* John Wiley & Son New York.

Scoles G, ed. 1988 *Atomic and Molecular Beam Methods* Vol. 1 Oxford University Press New York, Oxford.

Stanley R J and Castleman, Jr. A W 1991 *J. Chem. Phys.* **94**, 7744–7756.

Sussmann R, Neuhauser R and Neusser H J 1994 *Can. J. Phys.* **72**, 1179.

Tan X Q, Clouthier D J, Judge R H, Plusquellic D F, Tomer J L and Pratt D 1995 *J. Chem. Phys.* **95**, 7862–7871.

Uijt de Haag P A M and Meerts W L 1989 *Chem. Phys.* **135**, 139–147.

Wales D J and Doye J P K 1997 *J. Phys. Chem.A* **101**, 5111–5116.

Watson J K G 1968 *J. Chem. Phys.* **48**, 4517.

Wehrens R, Pretsch E and Buydens L M C 1999 *Anal. Chim. Acta* **388**, 265.

Wu Y R and Levy D H 1989 *J. Chem. Phys.* **91**, 5278–5284.

Yahagi T, Fujii A, Ebata T and Mikami N 2001 *J. Phys. Chem. A* **105**, 10673–10680.