

Ab Initio Treatment of Complex Systems
Kohn-Sham Orbitals for Multi Reference Methods
and the Base Pairing Properties of Xanthine

Dissertation zur Erlangung des
naturwissenschaftlichen Doktorgrades
der Bayerischen Julius-Maximilians-Universität Würzburg

vorgelegt von
Thomas Hupp
aus Hafenlohr am Main

Würzburg 2003

Eingereicht am: _____

bei der Fakultät für Chemie und Pharmazie

1. Gutachter: _____

2. Gutachter: _____

der Dissertation

1. Prüfer: _____

2. Prüfer: _____

3. Prüfer: _____

des Öffentlichen Promotionskolloquiums

Tag des Öffentlichen Promotionskolloquiums: _____

Doktorurkunde ausgehändigt am: _____

meinen Eltern

Danksagung

An dieser Stelle möchte ich allen danken, die es mir ermöglicht haben, diese Arbeit anzufertigen, indem sie mir sowohl bei Problemen und Fragen geholfen, als auch dazu beigetragen haben, die arbeitsreiche Zeit zu verschönen.

An erster Stelle möchte ich die Mitarbeiter vom AK Engels erwähnen. Mein Dank gilt vor allem Julian Henn, mit dem ich zeitweise sogar die Wohnung geteilt habe und der im Spülen locker eine AEG Favorit ersetzen kann. Patrick Musch hat mich bereits im Studium begleitet und in den AK Engels gebracht. Christian Sturm, Björn Dietrich und die F-Praktikanten Marco Schwieder und Franziska Jenke waren an der Untersuchung der Alanyl-PNA-Systeme beteiligt. Durch die gute Zusammenarbeit aller ist es gelungen, ein völlig neues Gebiet in unserem Arbeitskreis zu etablieren. Zuletzt möchte ich Mario Arnone danken, der sich nie darüber beschwert hat, dass ich seine Kaffeetasse fast täglich verwende, sowie den "Neuen" im AK, Milena Mladenovich und Sebastian Schlund, die durch fleißiges Korrekturlesen noch so einige Fehler aus der Arbeit ausmerzen konnten.

Ein weiterer Dank für den unermüdlichen Brot- und Toilettenpapiernachschub in der Endphase dieser Arbeit geht an meine WG-Mitbewohnerin Anke Meisner.

Die vorliegende Arbeit entstand in drei Kooperationen:

Hervorzuheben ist die Zusammenarbeit mit Herrn Prof. Dr. Andreas Göring. Ihm möchte ich für die Bereitstellung des lokalen Hartree-Fock Codes und die vielen angenehmen und fruchtbaren Diskussionen danken.

Die Untersuchung der Basenpaarungen durch Herrn Prof. Dr. Klaus Weisz und Frau Dr. Eline Basilio Janke mit Tieftemperatur NMR-Verfahren stellt eine wertvolle Ergänzung der vorliegenden Arbeit dar. Die Kooperation mit Herrn Prof. Dr. Ulf Diederichsen und Herrn Dr. Markus Hoffmann begann bereits im Rahmen meiner Diplomarbeit und setzte sich über

die Jahre, auch nach dem Wechsel des AK Diederichsen nach Göttingen, fort. Die Zusammenarbeit war sehr schön und hat diese Arbeit um viele interessante Themen bereichert.

Weiterhin danke ich der Stiftung Stipendienfonds im Verband der Chemischen Industrie für ein Doktorandenstipendium und dem Graduiertenkolleg 690 Elektronendichte für die finanzielle Unterstützung.

Mein besonderer Dank gilt Herrn Prof. Dr. Bernd Engels, dem es gelungen ist, mich für die Fragestellungen vor allem auch in Zeiten, in welchen es nur langsam vorwärts ging, zu begeistern. Er hat es verstanden, die Arbeit zu leiten und mir gleichzeitig die Freiheit gegeben, eigene Ideen und Interessen zu verfolgen, welches der Grundstock für ein erfolgreiches wissenschaftliches Arbeiten ist.

Contents

Preface	i
1 Introduction to Part1	1
1.1 The role of orbitals in multi reference approaches	1
1.2 Orbitals in HF- and KS-approaches	7
2 Kohn Sham Orbitals for Ab Initio Methods	23
2.1 MRCI-calculations for ethene	24
2.2 Relaxation of the orbitals describing the core ion of the excited states	30
2.3 Discussion of the water molecule	32
2.4 Discussion of ethene	48
2.5 MRCI-computations for carbonmonoxide	52
2.6 MRCI-computations for formaldehyde	57
2.7 MRCI-computations for acetone	58
2.8 MRPT2 for the ground state	61
2.9 Excitation energies of carbonmonoxide from MRPT2	63
2.10 Excitation energies of formaldehyde from MRPT2	67

2.11	Excitation energies of acetone from MRPT2	68
2.12	Rydberg-valence mixing of the 1^1B_{1u} -state of ethene	69
3	Introduction to Part 2	77
4	Base Pairing of Xanthine and Xanthine Derivatives	85
4.1	The stability of alanyl PNA	85
4.1.1	Regioisomerism of xanthine	86
4.1.2	Homodimers of N7-xanthine	90
4.1.3	Stacked tetramers	93
4.1.4	Hydrogen bonding and stacking of other alanyl PNA systems	95
4.1.5	Stability model for alanyl-PNA	97
4.2	Base pairing of N9-xanthine	102
4.2.1	Homodimers of N9-xanthine	102
4.2.2	Weak pairings of N9-xanthine	104
4.2.3	Base pairing of hypoxanthine	106
4.2.4	Strong hydrogen bonds	109
4.2.5	Comparison with experiment	111
4.2.6	Mutagenicity	113
4.2.7	Pairings of hypoxanthine to canonical pyrimidine bases	115
4.2.8	Pairings of neutral xanthine to canonical pyrimidine bases	117
4.2.9	Anionic pairings of xanthine with canonical bases	121
5	Summary	127
6	Zusammenfassung	137
A	Free Energies for the Deprotonation in a buffered solution	157

Preface

The topic of the present work is the characterization of the structure and stability of bioorganic systems. In particular it deals with base paired systems that are governed by a complex interplay of different weak interactions such as hydrogen bonding, base stacking and solvent interactions. One major part of this study is the implementation of new approaches for a better overall description of the mentioned systems with a special emphasis on the spectroscopic characterization (*e.g.* UV-Vis, CD, NMR) of organic and bioorganic structures and aggregations.

The computational effort of the methods used to directly describe electronically excited states, in particular the MRCI- and the MRPT2-method, strongly depends on the quality of the underlying orbitals. The basic idea followed in the present work was to reduce the computational effort of these methods by applying the Local Hartree-Fock (LHF) orbitals. LHF-orbitals are optimized at a computational cost roughly comparable to that of the HF method, while their virtual orbitals possess the quality of natural orbitals, that normally are obtained from highly expensive CI-methods.

After the implementation of the LHF-code into the DIESEL-MRCI program package, the orbitals were tested for the description of excited states for a set of test molecules. These computations showed that LHF-orbitals are an excellent starting point for the MRCI- and the MRPT2-method. Since the MR-wavefunction exhibits a significantly faster convergence, the LHF-orbitals lead to a strong reduction of the computational effort of both methods. These properties in the future should allow the application of MR-methods also for bioorganic systems, which, however, demands further improvements in the MRPT2-code.

However, the present program code does not allow the treatment of very large systems for

reasons like integral handling and storage of the coefficient vector. Therefore, the characterization of the structure and stability of the bioorganic base paired systems has been performed using standard theoretical methods like density functional theory, Møller-Plesset perturbation theory, and also force field methods.

The treated systems include different types of base pairings of which all involve the non-canonical nucleobase xanthine. One topic in this context is to find an explanation for the unexpectedly high stability of the xanthine-alanyl-PNA double strand. Another topic deals with the question, why xanthine, that is formed from guanine in DNA under the influence of free radicals, *e.g.*, from cigarette smoke, is able to form mutagenic base pairs with the pyrimidine base thymine. The analysis of bioorganic systems also includes the assignment of NMR spectra using theoretical methods, like for example the determination of base pairing modes by the chemical shifts of the H-bonding protons in low temperature ^1H -NMR spectroscopy.

Consequently, the present work consists of two apparently unrelated parts:

- In the first part, LHF Kohn-Sham orbitals are tested in multireference configuration interaction and multireference perturbational approaches. These methods shall be used in the future for the characterization of bioorganic systems.
- The second part analyzes the unusually high stability of DNA analogous double strands that are built from xanthine dimers. Furthermore, the mutagenicity of xanthine nucleotides that are formed in DNA under chemical stress is studied.

Chapter 1

Introduction to Part1

1.1 The role of orbitals in multi reference approaches

Multi Reference (MR) approaches^[1-5] are important methods for the computation of excited states, because they are able to cover both static and dynamical correlation effects which is important for a balanced description of the electronic excitation spectrum. One example is the description of Rydberg and excited valence states. An incomplete inclusion of the correlation energy leads to a redshift of the Rydberg series in the computed spectrum, since the contribution of dynamical correlation to Rydberg states is much lower than that to valence states or to electronic ground states.^[6,7] While static correlation needs to be treated in a Configuration Interaction (CI)-procedure, dynamical correlation to a sufficient amount can also be accounted for by perturbation theory.^[2-5] The CASPT2 approach by Roos and co-workers^[8] for instance treats static correlation effects by a CASSCF (Complete Active Space Self Consistent Field) approach while the dynamical correlation is included *via* second order perturbation theory (PT2).

The computational effort of MR-approaches strongly depends on the quality of the underlying orbitals. Orbitals that well capture the nature of the considered electronic states lead to compact wave functions, and make small zeroth-order wave functions possible, while accurate computations with inappropriate orbitals demand very large zeroth order and as a consequence very large total wave functions. Orbitals from a previous Hartree-Fock (HF)

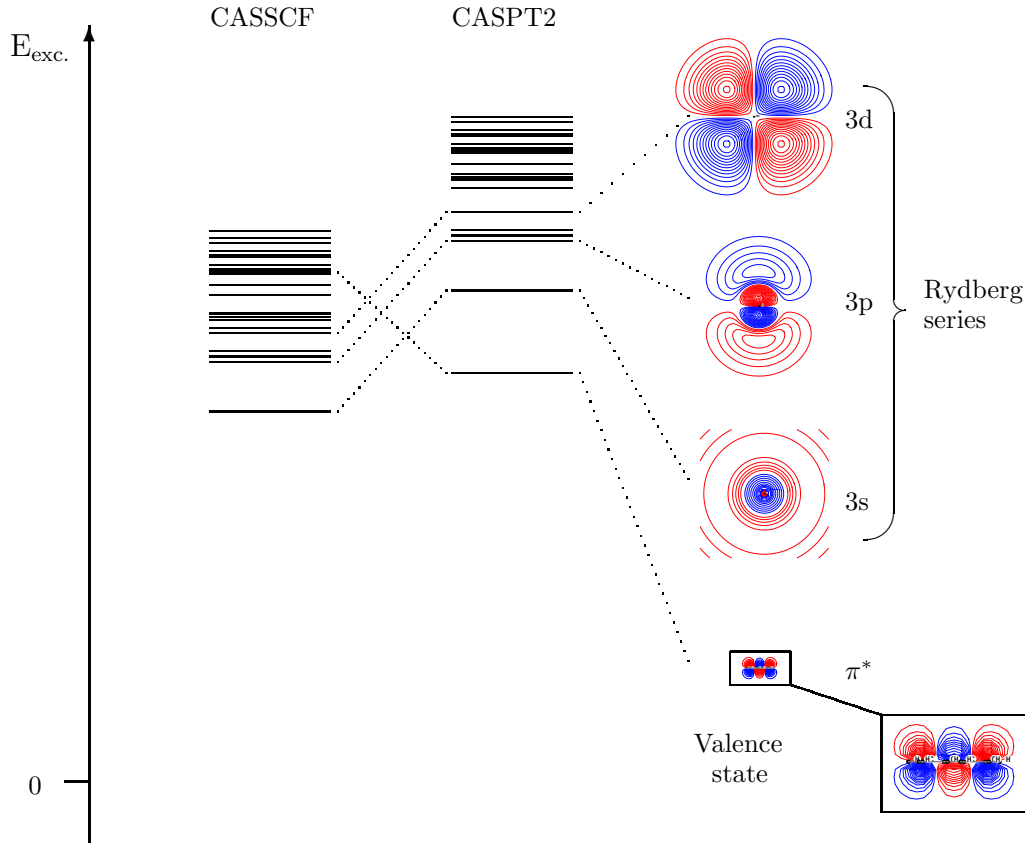


Fig. 1.1: Energetic ordering of Rydberg and valence states in CASSCF and CASPT2. In addition, idealized contour plots of Rydberg and valence orbitals are given. For CASSCF the Rydberg series is shifted to lower excitation energies.

calculation are an example for less suited orbitals. The unoccupied HF-orbitals are too diffuse since they are optimized in the field of all N electrons of the system and not in the physically correct field of $N-1$ electrons which is employed for the occupied orbitals. As a result, they are unable to describe static correlation and also do not reflect the electronic nature of excited states. To correct this, MR-computations based on HF-orbitals need large reference spaces to take static and dynamical correlation into account sufficiently.

Multiconfigurational-SCF- (MCSCF) or CASSCF-orbitals^[9–11] allow a much better CI-description since the orbitals within the active space are explicitly optimized to describe the character of a given state. Orbitals outside the active space are also better suited than virtual HF-orbitals since they “feel” a more attractive averaged potential, however, their quality is much lower than the quality of the orbitals within the active space since the potential still is

that of an N electron instead of an $N-1$ electron system. MCSCF- or CASSCF-approaches have some other disadvantages. First, they are quite expensive, especially if the size of the molecule increases or the symmetry of the considered system decreases. The second disadvantage is the often poor convergence behavior of the MCSCF- or CASSCF-calculations.

To improve the quality of the one-electron basis several procedures were advised. One example is natural orbitals (NO) which are obtained by the diagonalization of the one electron density matrix of a previous CI-Computation. NOs obtained from a full CI represent the best one-particle basis^[12] but also NOs obtained from a small CI-expansion give a better description of dynamical correlation than for example HF-orbitals.

Improved Virtual Orbitals (IVO) represent another possible one-particle basis. They are obtained from a modified HF approach in which the virtual orbitals are optimized in an assumed $N-1$ electron potential instead of the normal HF-potential.^[13-16] Consequently, this approach contains a certain arbitrariness and the orbital transformation has to be done at least twice. IVOs were proposed by Potts *et al.*^[17] for the construction of CAS-CI wavefunctions. A MRPT2-procedure based on IVO-orbitals has been implemented by Grimme *et al.*^[18]

One possible alternative to these orbitals are Kohn-Sham-(KS)-orbitals.^[19-21] In contrast to HF-orbitals, KS-orbitals are eigenfunctions to a Hamiltonian operator with the same local multiplicative potential for all orbitals. In principle, both virtual and occupied KS-orbitals are optimized in a potential of $N-1$ electrons. Therefore, virtual KS-orbitals have a more compact form than HF-orbitals. Likewise, the orbital energies, which for a long time were considered to be merely auxiliary quantities, can be seen as well defined approximations for excitation energies.^[22,23] Furthermore, KS-orbitals have been shown to be very useful in the interpretation of electronic structures and interactions of molecules.^[24]

In the last years, a large number of methods that combine density functional theory (DFT) and CI-methods have been developed.^[25-27] The majority of these approaches tries to describe dynamical correlation via a density functional, since CI methods describe dynamical correlation in a very inefficient way. Static correlation on the other hand, which in conventional DFT is only poorly described, is treated very accurately by a CI-procedure. The main problem, these combination methods have to cope with, is the double counting of the correlation energy.

Other approaches only try to take advantage of the orbital shape, but not of energy eigenvalues or other approximated energy expressions. However, the use of conventional KS-orbitals in ab initio CI-Methods to compute electronically excited states does not seem to offer large advantages in comparison to HF-orbitals.^[28,29]

Perturbational approaches based on the KS-determinant like the Görling-Levy-perturbation theory^[30,31] may suffer from a divergent perturbation series,^[29,32] which is shown in fig. 1.2 which gives the total energies of Raghleigh-Schrödinger perturbation series using a KS single particle operator and compares this to energies from Møller-Plesset and full-CI. Fig. 1.2 shows that the convergence problem seems to be more severe in approaches based on KS-orbitals than in Møller-Plesset perturbation theory, because the energy differences that enter the denominators of the perturbational energy corrections are smaller. This difficulty should be much less severe in multireference perturbation theory, since excited configurations that contribute most, and which in single reference perturbation theory would lead to an overestimation of the energy correction, belong to the CI-part of the calculation. Hirao and coworkers^[33] for example used GGA-orbitals in a CAS-CI-MRPT2-approach. Depending on the chosen CAS-space, this approach leads to very accurate results, which, however, do not seem to depend very much on the type of orbitals chosen as one-electron basis. In conventional density functional methods like the local density approximation (LDA) or generalized gradient approximations (GGA),^[21] the use of approximate exchange functionals prevents the cancellation of erroneous Coulomb self-interactions. These self-interactions lead to a wrong asymptotic decay of the KS potential. As a consequence, the energy of the Highest Occupied Molecular Orbital (HOMO)^[34] is shifted upwards, and is only a very poor approximation of the ionization potential. In anions, the HOMO often is not bound at all.^[35] Furthermore, due to the wrong asymptotic behavior, Rydberg states are also not bound in conventional DFT.

A fundamental solution to this problem is to employ the exact local multiplicative KS exchange potential which is determined by an integral equation, the EXX equation.^[36-38] A numerical grid method to solve this equation is known for quite some time^[36] but restricted to spherical systems, *i.e.*, atoms. Therefore recently basis set methods to solve the EXX equation were proposed^[37] and first implemented for solids^[39] and later also for molecules.^[40-43] While EXX implementations for solids are numerically stable, problems

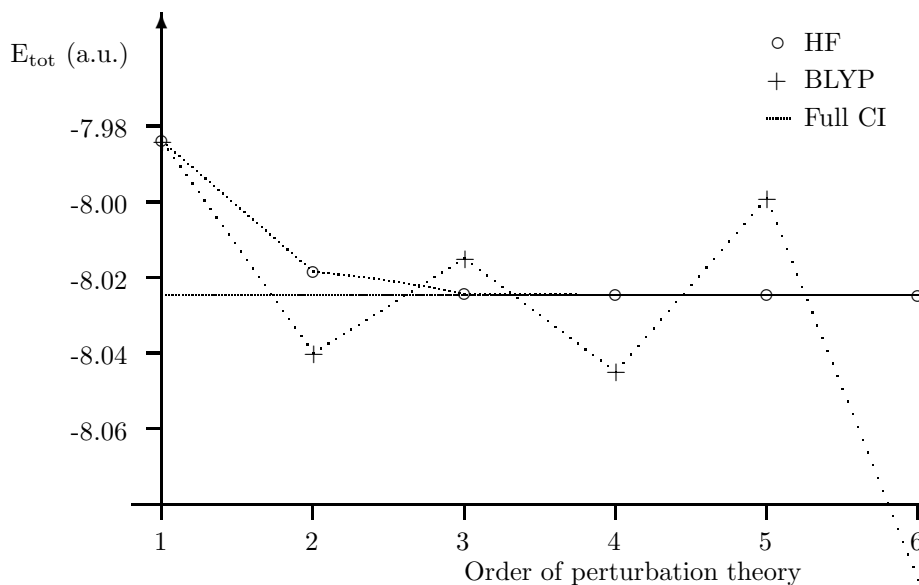


Fig. 1.2: Results of Rayleigh-Schrödinger perturbation series for LiH based on HF and BLYP zeroth order Hamiltonians compared with Full CI results. Data from Warken.^[29]

occur for molecules.^[43,44] To solve this problem very recently an effective EXX approach, the localized Hartree-Fock (LHF) method, was introduced^[44] (compare also^[45]). The LHF method yields numerically stable exchange potentials and orbital spectra.

Because the KS exchange potential is a local multiplicative potential it acts in the same way on all orbitals, occupied as well as unoccupied ones. Moreover, it exactly cancels the self-interacting terms of the Coulomb potential. Thus neither occupied nor unoccupied LHF-orbitals suffer from Coulomb self-interactions. As a result physically correct orbital and eigenvalue spectra including Rydberg series are obtained and the orbitals are well characterized in chemical terms as, *e.g.*, bonding, anti-bonding and so on. Veseth^[42,46] used EXX-Orbitals expanded into Slater functions in a CI of about 30-40 hand selected configurations to describe excitation energies of diatomic molecules. The results obtained this way are in general very promising, however, in some cases the computed excitation energies severely deviate from the experimental values.

This work tests whether Kohn-Sham orbitals and in particular LHF-orbitals represent useful one electron basis sets for multireference methods such as MRCI (sections 2.1 to 2.7) and

MRPT2 (sections 2.8 to 2.11). In section 2.1, LHF-orbitals are compared with HF-, BLYP- and partly CASSCF-orbitals with respect to their ability to serve as a single electron basis for the calculation of excited states of ethene within a subsequent MRSD-CI calculation. As a criterion to test the quality of an orbital set, the number of reference configurations that is needed to push the error in the MRSD-CI+Q calculation below 0.1 eV is used. Taking the water and the again the ethene molecules, the role of orbital relaxation in the occupied valence orbitals is extensively studied in the sections 2.2 and 2.4. In the calculation of excited states orbital relaxation effects can lead to problems, since the KS and HF-orbitals tested represent ground state orbitals.

In the the sections 2.5 to 2.7, it is tested how well LHF-orbitals describe the excited states of carbonmonoxide, formaldehyde and acetone in MRCI-calculations. Carbonmonoxide and acetone seem to represent quite difficult test cases for EXX-orbitals. For carbonmonoxide, for example, Veseth^[42] found difficulties in the CI description of Rydberg states if he employed EXX-orbitals. It is investigated, to what extend these difficulties are related to the quality of the EXX-Rydberg orbitals and how this behavior is influenced by orbital relaxation effects. Formaldehyde and acetone are chosen, because even though both are electronically very similar molecules, the ability of EXX-orbitals to describe their excited states seems to be quite different for both molecules. Additionally, all three represent a series in which the molecular size is systematically increased. The LHF-orbitals are compared with HF-, BLYP- and B3LYP-Orbitals. B3LYP-orbitals were included, because time-dependend DFT hybrid functionals give more accurate excitation energies than pure GGA functionals for carbonmonoxide^[69] and formaldehyde.^[48,69] Even though this might mainly be due to error compensation, it could be taken as a hint that the underlying orbitals might be a better starting point to describe the corresponding excited states.

In the following sections, is is tested whether LHF-orbitals are also suited for an MRPT2 approach. In this part, it is first (section 2.8) studied to what extend the overestimation of the second order corrections in single state perturbation theory can be eliminated in an MR approach. In the next step (sections 2.11 to 2.9), excited states are computed with an MRPT2 approach based on LHF-orbitals.

Major parts of chapter 2 are published in refs. 49-51.

Before the results of the analysis are presented in chapter 2, in the next section gives a more

detailed description of the differences between the HF method and Kohn-Sham approaches.

1.2 Orbitals in HF- and KS-approaches

The simplest description of a many particle system that obeys the Pauli principle is the Slater determinant of the single particle wavefunctions. Both the HF- and the KS-approach try to describe a given many particle system by a single Slater determinant. The principal difference between both approaches is that KS-orbitals in contrast to HF-orbitals are minimized under the additional condition that the orbitals come from a local potential. The KS-counterpart to the HF method is the exact exchange-only (EXX) KS-method, because both the EXX- and the HF-methods are based on the same total energy functional. In contrast to EXX-methods, the starting point for virtually all common density functionals is the description of exchange and correlation by the local density approximation. The different treatment of the orbitals by the three approaches is described in the following sections. The introductory descriptions given in for the HF and KS ansatz roughly follows the description of standard textbooks of theoretical chemistry and density functional theory.^[12,19,21,38]

The HF method

The systems that are treated in this work are many electron systems described by the Hamiltonian operator

$$\hat{H} = \hat{T} + \hat{W} + \hat{V} \quad (1.1)$$

with the kinetic-energy operator

$$\hat{T} \equiv \sum_i \hat{t}_i = -\frac{1}{2} \sum_i \hat{\Delta}_i, \quad (1.2)$$

the Coulomb interaction of the electrons

$$\hat{W} \equiv \sum_j \sum_{i>j} \frac{1}{r_{ij}} \quad (1.3)$$

that move in an external potential

$$\hat{V} \equiv \sum_i \hat{v}_{ext}(\mathbf{r}_i) \quad (1.4)$$

that is typically determined by the Coulomb potential of the nuclei. In eqs. 1.2 to 1.4 uppercase symbols always denote total energy operators, while lowercase symbols always stand for the one or two electron operators. The goal is now to find a good approximation to the many electron wavefunction Ψ that solves the Schrödinger equation

$$\hat{H}\Psi = E\Psi. \quad (1.5)$$

The simplest physically sound approximation to the exact N-electron wavefunction Ψ is the antisymmetrized product of N single electron wavefunctions $\chi(\mathbf{x})$, the *Slater determinant*:^[12]

$$\begin{aligned} \Psi = \tilde{\Psi}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) &= (N!)^{-1/2} \begin{vmatrix} \chi_i(\mathbf{x}_1) & \chi_j(\mathbf{x}_1) & \dots & \chi_N(\mathbf{x}_1) \\ \chi_i(\mathbf{x}_2) & \chi_j(\mathbf{x}_2) & \dots & \chi_N(\mathbf{x}_2) \\ \dots & \dots & \dots & \dots \\ \chi_i(\mathbf{x}_N) & \chi_j(\mathbf{x}_N) & \dots & \chi_N(\mathbf{x}_N) \end{vmatrix} \\ &= |\chi_i(\mathbf{x}_1)\chi_j(\mathbf{x}_2)\dots\chi_N(\mathbf{x}_N)\rangle \end{aligned} \quad (1.6)$$

In equation 1.6 $(N!)^{-1/2}$ is the normalization factor, which by definition is contained in the short-hand notation that is given in the bottom line. The spin orbitals $\chi(\mathbf{x})$ are the product of spatial orbital and spin function $\varphi_j(\mathbf{r}) * \alpha(s)$.

The total energy of the Slater determinant is

$$\begin{aligned} E_{tot} = \langle \Psi | \hat{H} | \Psi \rangle &= \sum_{i=1}^N \int d^3r \chi_i^*(\mathbf{x}) [\hat{t} + \hat{v}_{ext}] \chi_i(\mathbf{x}) + \\ &\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \left(\int \int d\mathbf{x} d\mathbf{x}' \frac{\chi_i(\mathbf{x}) \chi_i(\mathbf{x}) \chi_j(\mathbf{x}') \chi_j(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} - \right. \\ &\left. \int \int d\mathbf{x} d\mathbf{x}' \frac{\chi_i(\mathbf{x}) \chi_j(\mathbf{x}) \chi_j(\mathbf{x}') \chi_i(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \right). \end{aligned} \quad (1.7)$$

In equation 1.7 the total energy is described as the sum over kinetic and external potential energy, the classical coulomb interaction of the spin orbitals and the exchange interaction which is the result of the electron correlation induced by the Pauli principle.^[12] For systems having an even number of electrons, eqn. 1.7 can be reduced to a functional depending only on the spatial orbitals $\varphi_j(\mathbf{r})$:

$$E_{tot} = 2 * \sum_{i=1}^{N/2} \int d^3r \varphi_i^*(\mathbf{r}) [\hat{t} + \hat{v}_{ext}] \varphi_i(\mathbf{r}) +$$

$$\sum_{i,j=1}^{N/2} (2 * \int \int d\mathbf{r}d\mathbf{r}' \frac{\varphi_i(\mathbf{r})\varphi_i(\mathbf{r})\varphi_j(\mathbf{r}')\varphi_j(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} - \int \int d\mathbf{r}d\mathbf{r}' \frac{\varphi_i(\mathbf{r})\varphi_j(\mathbf{r})\varphi_j(\mathbf{r}')\varphi_i(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}). \quad (1.8)$$

In chemist's notation, 1.7 is abbreviated as

$$E_{tot} = 2 \sum_{i=1}^{N/2} (i|\hat{h}|i) + \sum_{i,j=1}^{N/2} [2(ii|jj) - (ij|ji)]. \quad (1.9)$$

with the one electron operator $\hat{h} = \hat{t} + \hat{v}_{ext}$. According to the variational principle, the best description of a system is the one that minimizes the total energy functional. Consequently, in the HF scheme, the total energy functional is minimized with respect to the orbitals:

$$\frac{\delta}{\delta\varphi_i(\mathbf{r})} (E_{tot} - \epsilon (\sum_j \langle \varphi_j | \varphi_j \rangle - N)) = 0. \quad (1.10)$$

In 1.10 the Lagrangian multiplier ϵ ensures a constant electron number. The minimization leads to the HF equations:

$$h(1)\varphi_i(1) + \sum_{j=1}^N \left[\int d\mathbf{r}' |\varphi_j(\mathbf{r}')|^2 r_{12}^{-1} \right] \varphi_i(\mathbf{r}) - \sum_{j=1}^N \left[\int d\mathbf{r}' \varphi_j(\mathbf{r}') \varphi_i(\mathbf{r}') r_{12}^{-1} \right] \varphi_j(\mathbf{r}) = \epsilon_i \varphi_i(\mathbf{r}), \quad (1.11)$$

which are often written as effective one electron equations:

$$\hat{F}\varphi_i(\mathbf{r}) = \epsilon_i \varphi_i(\mathbf{r}) \quad (1.12)$$

with the Fock-operator

$$\hat{F} = \hat{h}_i + \hat{u}_i + \hat{v}_{x,i}^{NL} \quad (1.13)$$

that is composed of the one electron part \hat{h} , the Coulomb potential \hat{u}_i

$$\hat{u}_i = \sum_{j=1}^N \left[\int d\mathbf{r}' |\varphi_j(\mathbf{r}')|^2 r_{12}^{-1} \right] \quad (1.14)$$

and the nonlocal exchange potential \hat{v}_x^{NL} . Non-locality means that the HF exchange potential is not one single potential that is the same for all orbitals, but it can only be defined for each orbital individually, since the orbital itself enters the expression for \hat{v}_x^{NL} .^[38]

$$\hat{v}_{x,i}^{NL} F\varphi_i(\mathbf{r}) = - \sum_{j=1}^N \left[\int d\mathbf{r}' \varphi_j(\mathbf{r}') \varphi_i(\mathbf{r}') r_{12}^{-1} \right] \frac{\varphi_j(\mathbf{r})}{\varphi_i(\mathbf{r})}. \quad (1.15)$$

Density functional theory in the Kohn-Sham formalism

The basis of modern DFT is the theorem of Hohenberg and Kohn,^[52] that states that the ground state wavefunction Ψ as well as the external potential v is uniquely determined by the ground state density:

$$\Psi \leftrightarrow \rho(\mathbf{r}) \leftrightarrow v(\mathbf{r}), \quad (1.16)$$

and as a consequence all observables, *i.e.*, the expectation values of hermitian operators of a static many particle system are also functionals of its ground state density ρ . In particular, the total energy can be written as a functional of the density:

$$E[\rho(\mathbf{r})] = \langle \Psi_n[\rho(\mathbf{r})] | \hat{H} | \Psi_n[\rho(\mathbf{r})] \rangle = \langle \Psi_n[\rho(\mathbf{r})] | \hat{T} + \hat{W} + \hat{V} | \Psi_n[\rho(\mathbf{r})] \rangle \quad (1.17)$$

$$= T[\rho(\mathbf{r})] + W[\rho(\mathbf{r})] + \int \rho(\mathbf{r})v d\mathbf{r}. \quad (1.18)$$

If $\rho(\mathbf{r})$ is the exact ground state density, then $E[\rho(\mathbf{r})]$ equals the exact ground state energy. For any other density $\rho(\mathbf{r}) \neq \rho(\mathbf{r})_0$, the Hohenberg-Kohn variation principle

$$E[\rho(\mathbf{r})_0] > E[\rho(\mathbf{r})] \quad (1.19)$$

holds. To obtain the best possible ground state density and the best possible total energy one has to minimize the energy with the side condition of a constant of electrons, which leads to the Euler-Lagrange equation:^[38]

$$\frac{\delta E[\tilde{\rho}(\mathbf{r})]}{\delta \rho(\mathbf{r})} - \mu = 0. \quad (1.20)$$

In equation 1.19 the kinetic energy and Coulomb interaction do not depend on the external potential v_0 of the particular system, it can be seen to be universal. Kinetic energy and Coulomb interaction will be abbreviated as

$$F[\rho(\mathbf{r})] \equiv \langle \Psi_n[\rho(\mathbf{r})] | \hat{T} + \hat{W} | \Psi_n[\rho(\mathbf{r})] \rangle = T[\rho(\mathbf{r})] + W[\rho(\mathbf{r})]. \quad (1.21)$$

The functional $F[\rho(\mathbf{r})]$ is independent of the external potential v_0 , *i.e.*, it has the same functional form for all systems with a fixed particle-particle interaction.

Nearly all applications of DFT in quantum chemistry use the Kohn-Sham formulation of DFT.^[53] The Kohn-Sham formalism introduces a system of N *noninteracting* electrons that move in an external local potential v_s which has the property to yield exactly the same density as the potential v of the interacting system. Since the Hohenberg-Kohn theorem does not depend on the particular form of the electron-electron interaction, it is also valid if $W = 0$. As a consequence, v_s must be unique:

$$v(\mathbf{r}) \leftrightarrow \rho(\mathbf{r}) \leftrightarrow v_s(\mathbf{r}). \quad (1.22)$$

The many particle Hamiltonian operator (1.1) can now be written as the sum over single electron Hamiltonians (1.23):

$$\hat{H} = \sum_{i=1}^N \hat{h}_s(i) = \sum_{i=1}^N \left(-\frac{1}{2} \hat{\Delta}_i + v_s(\vec{\mathbf{r}}_i) \right) \quad (1.23)$$

The single electron wavefunctions for the noninteracting system can be written as

$$\left(-\frac{1}{2} \Delta_i + v_s(\mathbf{r}_i) \right) \phi_i(i) = \epsilon_i \phi_i(i). \quad (1.24)$$

The corresponding density is

$$\rho(\mathbf{r}) = \sum_{i=1}^N |\phi_i(\mathbf{r})|^2. \quad (1.25)$$

The KS-wavefunction, that is the Slaterdeterminant of the KS-orbitals, cannot be considered to be an approximation to the exact wavefunction, but it is a property precisely defined by the exact density of the system.^[54] The total energy of this noninteracting system is given by:

$$E_s[\rho(\mathbf{r})] = T_s[\rho(\mathbf{r})] + \int d^3r [\rho(\mathbf{r})] V_s(\mathbf{r}). \quad (1.26)$$

Since approaches that evaluate all energy terms by explicit density functionals like the Thomas Fermi model do not allow an acceptable accuracy for kinetic energy, Kohn and Sham introduced the exact orbital dependend expression for the kinetic energy functional $T_s[\rho(\mathbf{r})]$.

$$T_s[\rho(\mathbf{r})] = \sum_{i=1}^N \int d^3r \phi_i^*[\rho](\mathbf{r}) \left[-\frac{\nabla^2}{2} \right] \phi_i[\rho](\mathbf{r}). \quad (1.27)$$

Like in the Thomas-Fermi model, Kohn and Sham used the classical Coulomb interaction for the potential energy, while the exchange and Coulomb correlation is summed up in the exchange-correlation term $E_{xc}[\rho(\mathbf{r})]$. This leads to the following expression for $F[\rho(\mathbf{r})]$:

$$F[\rho(\mathbf{r})] = T[\rho(\mathbf{r})] + W[\rho(\mathbf{r})] = T_s[\rho(\mathbf{r})] + \overbrace{\frac{1}{2} \int \int d\mathbf{r} d\mathbf{r}' \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}}^{\text{Coulomb interaction}} + E_{xc}[\rho(\mathbf{r})] \quad (1.28)$$

The definition of exchange correlation energy is obtained, if $F[\rho(\mathbf{r})]$ of the *noninteracting* system is compared to $F[\rho(\mathbf{r})]$ of the interacting system in equation 1.21:

$$E_{xc}[\rho(\mathbf{r})] = T[\rho(\mathbf{r})] - T_s[\rho(\mathbf{r})] + W[\rho(\mathbf{r})] - \frac{1}{2} \int \int d\mathbf{r} d\mathbf{r}' \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \quad (1.29)$$

Adding the interaction with the external potential to equation 1.29, the total energy functional can now be expressed as:

$$E[\rho(\mathbf{r})] = \int d\mathbf{r} \rho(\mathbf{r}) v + T_s[\rho(\mathbf{r})] + \frac{1}{2} \int \int d\mathbf{r} d\mathbf{r}' \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + E_{xc}[\rho(\mathbf{r})] \quad (1.30)$$

Exchange and correlation in KS-theory

So far, the exchange-correlation energy has simply been defined as the difference between the exact expression for $F[\rho(\mathbf{r})]$ and the approximations made in equation 1.28. E_{xc} can be further partitioned into exchange E_x and correlation energy E_c :

$$E_{xc} = E_x + E_c \quad (1.31)$$

Since the ground state wavefunction for in the KS formalism is given the a Slater determinant of the KS-orbitals, the exact exchange energy in the KS formalism has the same expression as in HF-theory

$$E_x^{KS} = \frac{1}{2} \sum_{i,j}^{occ} \int \int d\mathbf{r} d\mathbf{r}' \frac{\phi_i^*(\mathbf{r})\phi_j^*(\mathbf{r}')\phi_i(\mathbf{r}')\phi_j(\mathbf{r})}{|\mathbf{r} - \mathbf{r}'|}. \quad (1.32)$$

In equation 1.32 the exchange energy, like the kinetic energy before, is an explicit functional of the orbitals. Why can it be called an implicit functional of the density? In KS theory

the energy is minimized under the subsidiary condition that the orbitals come from a local potential. If the local exchange potential $v_x = \frac{\delta E_x}{\delta(\rho)}$ is known, the exchange energy can be expressed as an explicit functional of the density:

$$E_x \rho(\mathbf{r}) = \int d\mathbf{r} v_x(\mathbf{r}) \rho(\mathbf{r}) \quad (1.33)$$

As a consequence, as long as the restriction that the orbitals that enter an orbital dependent functional come from a local potential is fulfilled, the explicitly orbital dependent functional can be considered to be an implicit density functional.

Before KS approaches that use the exact exchange functional are described, common DFT methods that approximate the exchange (and correlation) by an explicit density functional are described. The basic model of merely all these approximations is the *uniform electron gas*, which represents a system in which the electrons move in a structureless (uniform) positive potential. The expression for the exchange potential of such a system was already formulated by Bloch and Dirac in the late 1920's:

$$v_x = -\sqrt[3]{3\rho(\mathbf{r})}\pi \quad (1.34)$$

In 1988, Becke introduced the gradient correction $\Delta E_x^{B88}[\rho(\mathbf{r})]$ to the LDA expression of the exchange energy:^[55]

$$E_x^{B88}[\rho(\mathbf{r})] = E_x^{LDA}[\rho(\mathbf{r})] + \Delta E_x^{B88}[\rho(\mathbf{r})] \quad (1.35)$$

with

$$\Delta E_x^{B88}[\rho(\mathbf{r})] = \beta \rho(\mathbf{r})^{1/3} \frac{x^2}{1 + 6\beta x \sinh^{-1} x} \quad (1.36)$$

and

$$x = \frac{|\hat{\nabla}\rho(\mathbf{r})|}{\rho^{3/4}}. \quad (1.37)$$

The correction term $\Delta E_x^{B88}[\rho(\mathbf{r})]$ contains the free parameter β , by which $E_x^{B88}[\rho(\mathbf{r})]$ was fitted to the exchange energy of noble gas atoms from helium to radon.

The correlation energy for the uniform electron gas is more difficult to evaluate. In 1980 Vosko, Wilk und Nusair derived an analytical expression for the correlation energy, which is commonly known as the VWN functional. A gradient corrected functional for the correlation energy has been proposed by Lee, Yang and Parr (LYP).^[56] In this work, it is used in combination with the B88-exchange (BLYP) as the standard example for a common GGA-functional. Since the explicit form of the LYP-functional is rather lengthy, it is not given here,

but it can be taken from standard textbooks on quantum chemistry and density functional theory.^[21,78]

The energy functional can be further improved, if the exchange energy is partly described by the HF-exchange. One of these so called hybrid functionals is the Becke-3-parameter (B3) functional:^[57]

$$E_{xc}^{B3LYP} = a E_x^{LDA} + (1 - a) E_x^{KS} + b \Delta E_x^{B88} + c E_c^{LYP} + (1 - c) E_c^{VWN} \quad (1.38)$$

The three parameters used in E_{xc}^{B3LYP} were determined as $a = 0.80$, $b = 0.72$ and $c = 0.81$ by fitting to experimental atomization energies and ionization potentials.

The LHF-method

If explicitly orbital-dependent functionals like the exact exchange, but also orbital dependent correlation functionals are used, the evaluation of $\delta E_{xc}/\delta\rho$ leads to the Optimized Potential Method.^[38] The exact exchange (EXX) potential of a closed-shell system obeys the integral equation^[36,37,58]

$$\begin{aligned} & \int d\mathbf{r}' X_s(\mathbf{r}, \mathbf{r}') v_x(\mathbf{r}') \\ &= 4 \sum_a^{\text{occ.}} \sum_s^{\text{unocc.}} \phi_a(\mathbf{r}) \phi_s(\mathbf{r}) \frac{\langle \phi_s | \hat{v}_x^{\text{NL}} | \phi_a \rangle}{\varepsilon_a - \varepsilon_s}. \end{aligned} \quad (1.39)$$

In equation (1.39) $X_s(\mathbf{r}, \mathbf{r}')$ stands for the KS response function, $v_x(\mathbf{r}')$ for the KS exchange potential, ϕ_a and ϕ_s denote real valued occupied and unoccupied spatial KS-orbitals with the eigenvalues ε_a and ε_s , while \hat{v}_x^{NL} designates a nonlocal exchange operator which has the form of the HF exchange operator. In the EXX method, this integral equation has to be solved in each iteration of the SCF process to construct the exchange potential. A numerical grid-procedure for the solution of the EXX-equation has been known for quite some time,^[36] but it is limited to spherical systems. In the last years, basis set methods employing auxiliary basis sets were proposed to solve equation (1.39).^[37,59] These methods were first implemented for solids,^[39,59,60] later for molecules.^[40–43,61] While the implementations for solids were numerically stable, for molecules problems occurred.^[43,44,61] An alternative EXX-approach is the local HF by Della Sala and Görling^[44] (LHF). The LHF approach is derived

in ref.^[44] by comparing the HF to the exact exchange-only KS formalism. As explained in section the HF determinant is defined as the single Slater determinant that minimizes the total energy. The exchange-only KS determinant is the Slater determinant that minimizes the total energy under the additional condition that its orbitals are eigenstates of an one-particle Hamiltonian operator with a local multiplicative potential. Due to this additional condition the two determinants are different and the exchange-only KS determinant always yields a slightly higher total energy.

$$E^{HF} \leq E^{EXX} \quad (1.40)$$

The LHF method is derived by making the assumption that the two determinants are identical. This is an approximation, however, a very good one and the resulting LHF method yields *de facto* identical results as a full EXX method and therefore can be considered to be an effective exact exchange method. Note that even if the two determinants were identical this would not mean that the corresponding canonical orbitals and their eigenvalues had to be identical. It only means that the occupied orbitals are related by a unitary transformation but the individual orbital and their eigenvalues could be completely different. In practice, it turns out that the difference between exchange-only KS and HF determinants (however, not the differences between their orbitals and eigenvalues) is negligible.^[36,44] For small and medium sized molecules and for a given atomic basis set the overlap of the HF and the exchange-only KS determinants deviate from one (identity of the determinants) by only $10^{-5} - 10^{-6}$ a.u., while total energies differ by 0.01% or less. If the HF and exchange-only KS determinant were identical then one Slater determinant would be an eigenstate of both the many-particle HF and the many-particle exchange-only KS equation:

$$[\hat{T} + \hat{V}_{ext} + \hat{U} + \hat{V}_x^{NL}]\Phi = E_s^{HF}\Phi \quad (1.41)$$

$$[\hat{T} + \hat{V}_{ext} + \hat{U} + \hat{V}_x]\Phi = E_s^{EXX}\Phi \quad (1.42)$$

The operators \hat{T} , \hat{V}_{ext} , \hat{U} , \hat{V}_x^{NL} and \hat{V}_x in Eqs. 1.41 and 1.42 are the many-electron operators corresponding to the kinetic energy, the external potential, the Coulomb potential, the non-local HF operator and the local KS exchange potential. E_s^{HF} and E_s^{LHF} are the sums over the corresponding HF/EXX orbital energies. From manipulating the latter two equations

the expression

$$v_x^{LHF}(\mathbf{r}) = v_x^S(\mathbf{r}) + \frac{2}{\rho(\mathbf{r})} \sum_{(a,b) \neq (N,N)}^{occ.} \phi_a(\mathbf{r})\phi_b(\mathbf{r}) [\langle \phi_b | \hat{v}_x^{LHF} | \phi_a \rangle - \langle \phi_b | \hat{v}_x^{NL} | \phi_a \rangle], \quad (1.43)$$

with the Slater potential^[62]

$$v_x^S(\mathbf{r}) = -\frac{2}{\rho(\mathbf{r})} \sum_{a,b}^{occ.} \phi_a(\mathbf{r})\phi_b(\mathbf{r}) \int d\mathbf{r}' \frac{\phi_b(\mathbf{r}')\phi_a(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}, \quad (1.44)$$

can be derived^[44] for the LHF exchange potential of a closed shell system. In eqs. 1.43 and 1.44, ϕ_a and ϕ_b are occupied orbitals, $\rho(\mathbf{r})$ denotes the electron density, N is the label of the highest occupied molecular orbital, and \hat{v}_x^{NL} is a nonlocal exchange operator of the form of the HF exchange operator but constructed from KS-orbitals. In the sum on the right-hand side of equation 1.43, the term which $\phi_a = \phi_b = \phi_N$ has been excluded (see ref.^[44]). Eq. (1.43), in contrast to the expression yielding the exchange potential in a full EXX treatment, contains only occupied orbitals. It can be solved iteratively along with the usual self-consistency process in an efficient and numerically stable manner, see Della Sala and Görling.^[44,63] Constructing the exchange potential according to eq. (1.43) can be considered as a localization of a nonlocal exchange operator \hat{v}_x^{NL} of the HF type. Therefore the approach has the name localized HF method. The LHF method has been implemented in the TURBOMOLE program package.^[44,63,64] Recently also a restricted open shell LHF approach has been implemented.^[65]

HF-orbitals

Since the present work is mainly interested in the properties of the orbitals, it is worth taking closer look at the different treatment of the orbitals in HF- and KS-theory. In the HF equations, occupied and unoccupied virtual orbitals are treated in a completely different way. The orbital energy for an occupied orbital is

$$\epsilon_a = \langle \phi_a | \hat{h} | \phi_a \rangle + \sum_{x=1}^N (\langle \phi_a(r_1)\phi_x(r_2) | \phi_a(r_1)\phi_x(r_2) \rangle - \langle \phi_a(r_1)\phi_x(r_2) | \phi_x(r_1)\phi_a(r_2) \rangle) \quad (1.45)$$

For the occupied orbitals the terms in which $\phi_x = \phi_a$, the self-interacting terms, are canceled:

$$\langle \phi_a(\mathbf{r}_1)\phi_a(\mathbf{r}_2)|\phi_a(\mathbf{r}_1)\phi_a(\mathbf{r}_2) \rangle - \langle \phi_a(\mathbf{r}_1)\phi_a(\mathbf{r}_2)|\phi_a(\mathbf{r}_1)\phi_a(\mathbf{r}_2) \rangle = 0 \quad (1.46)$$

Because of the cancellation of the self-interaction, an occupied HF-orbital of an N-electron system only interacts with N-1 other electrons.

An important physical meaning is of the energies for the occupied orbitals obtained, if the ionization potential is computed as the energy difference between the ground state Slater determinant and the Slater determinant of a system from which an electron has been removed from spin orbital φ_c , while the other orbitals are left unchanged (“frozen orbital” approximation):

$$IP = {}^{N-1}E - {}^N E_0 = -(c|h|c) - \sum_a ((aa|cc) - (ac|ca)) = \epsilon_c. \quad (1.47)$$

The ionization potential equals the negative of the orbital energy. The approximation of the ionization potentials by the orbital energies is known as Koopmans’ theorem. It is based on the approximation that the correlation energy that is not treated in Hartree-Fock theory is compensated by the neglect of orbital relaxation in the “frozen orbital” approximation.

For an unoccupied orbital no terms in equation

$$\epsilon_i = \langle \phi_i|\hat{h}|\phi_i \rangle + \sum_{x=1}^N (\langle \phi_i(r_1)\phi_x(r_2)|\phi_i(r_1)\phi_x(r_2) \rangle - \langle \phi_i(r_1)\phi_x(r_2)|\phi_x(r_1)\phi_i(r_2) \rangle) \quad (1.48)$$

are canceled. This means, that an unoccupied HF-orbital of an N-electron system interacts with all N electrons of the system. This means that an unoccupied orbital of a neutral system “feels” a neutral potential. It is as if an electron had been added to the system to produce an N+1 electron state, and ϵ_i represented the energy of the extra electron. Correspondingly, the electron affinities EA can be approximated by the energies of virtual HF-orbitals.

$$EA = {}^N E_0 - {}^{N+1}E_i = -(i|h|i) - \sum_a ((aa|ii) - (ai|ia)) = \epsilon_i. \quad (1.49)$$

In contrast to the Koopmans’ theorem for ionization energies, for electron affinities, the neglect of correlation and orbital relaxation have the same sign and do not cancel each other but they are summed up. As a consequence, Koopmans’ electron affinities are usually very

poor.

This work mainly uses virtual orbitals to describe either ground state correlation or excited states. For both purposes HF-orbitals should not be very useful: In order to efficiently describe static correlation, the virtual orbitals need to have a rather compact shape. However, since HF-orbitals usually describe unbound states, they are highly diffuse. Furthermore, HF-orbitals also do not describe excited states, because they do not describe a particle conserving system, but a system with one additional electron. The anions described by the virtual orbitals are more diffuse than most of the excited states (except for some highly diffuse Rydberg states). Due to their diffuseness, especially the description of valence states is very difficult, while they are somewhat more useful for the description of Rydberg states, which are much more diffuse than valence states.

In this work, HF virtual orbitals of cationic systems ($N-1$ electrons) are sometimes used to describe the Rydberg states of a neutral molecule. In this case a Rydberg state is not seen as an excited state of a neutral ground state, but as a cation (in the core of a Rydberg state) to which a weakly bound electron is added. For the Rydberg states, orbital relaxation and correlation only play minor roles because the interaction of the Rydberg electron with the cationic core is only weak. Also virtual orbital energy

$$\epsilon_i = (i|\hat{h}|i) + \sum_{x=1}^{N-1} ((i|x|x) - (ix|xi)) \quad (1.50)$$

$$(1.51)$$

in this case is a very good approximation to the negative of the energy necessary to remove the Rydberg electron from the molecule.

KS orbitals

In KS-theory, the nonlocal exchange potential is replaced by the local exchange potential.

$$\epsilon_a = (a|\hat{h}|a) + \sum_{x=1}^N (aa|xx) - (a|v_x|a) \quad (1.52)$$

This means that all KS-orbitals are eigenfunctions to the same single particle Hamiltonian. In the exact KS potential, both occupied and virtual orbitals are optimized in a field of

N-1 electrons. As a consequence, KS virtual orbital energies are not approximation for the electron affinity, but the difference between the energies of occupied and unoccupied orbitals can be seen as a well defined approximation of excitation energies.^[22,23] This is in good agreement with common chemical intuition, that understands virtual orbitals in a particle conserving way. Virtual orbitals are usually described as entities into which electrons can be excited rather than spaces to absorb additional electrons. While this is not realized in the HF model, the KS model represent represents such a particle conserving system. For example how KS-orbitals are used to describe chemical interactions see Bickelhaupt and Baerends.^[24] Besides the properties of the virtual orbitals, KS-orbitals in principle offer another advantage over orbitals obtained from other *ab initio* methods. The KS theory is an exact MO theory, since the occupied MOs yield the exact ground state density. Occupied MOs of exact KS theory take into account all interactions including electron correlation. However, this argument certainly does not hold for LHF-orbitals, which do not include correlation and only partly for the common LDA- and GGA-orbitals. Even though LDA- and GGA-orbitals include a large fraction of the correlation energy, they often suffer from erroneous self interactions. This is because if exchange is treated approximately, the Coulomb self interaction does not necessarily cancel the self-exchange:

$$(aa|aa) \neq (a|v_x|a). \quad (1.53)$$

LDA and GGA orbitals

For methods based on the local density approximation are applied to the hydrogen atom, the self interaction error is in the order of 10^{-3} of the total energy. The self interaction error is found in all systems and becomes severe in systems in which no error cancellation is found such as the description of bond dissociations and biradical reactions.^[21] Relevant for the present work is that the self-interaction error affects the asymptotic behavior of the KS potential. For a point far away from the molecule, the nuclear potential equals the coulomb potential of the orbitals. As a consequence, the fall-off is only determined by the decay of the

exchange potential, which for LDA is $\rho^{1/3}$. Since ρ decays exponentially¹, also the exchange, and thus the total potential decays exponentially instead of $-1/r$. The wrong asymptotic decay of the KS-potential affects both the spatial form of the orbitals and their energies. Since the spatial form of KS-orbitals will be discussed in great detail later in this work, here only the orbital energies are discussed, which are important, since they serve as input entities or “first order approximation” in TDDFT.^[48,66] The wrong exponential decay shifts all occupied orbitals to higher energies^[34] as it is shown in fig. 1.3. The negative of the HOMO energy for GGA- and LDA-orbitals is often strongly underestimates the ionization potential. In anions, the HOMO often is not bound at all.^[35] Also the virtual orbitals are also shifted to higher energies. In many cases only very few states of the Rydberg series are bound. However, since the HOMO is also shifted to higher energies while the HOMO-LUMO gap is often correct, especially the excitation energies for Rydberg state are often under rather than overestimated by the orbital energy difference and the TDDFT approach.

Asymptotic behavior of the LHF potential

The asymptotic behavior of the KS-potentials v_s can be improved by using asymptotic corrections,^[67–69] however, this does not fundamentally solve the problem of erroneous self-interactions. The fundamental solution of this problem is to employ the exact local multiplicative KS exchange potential in the KS self-consistent field (SCF) process, as it has been described for the EXX-approaches like the LHF. In EXX-approaches, Coulomb self-interactions are exactly canceled. This means, that each orbital is optimized in the field of $N-1$ occupied orbitals, which in contrast to the HF-method is valid for all orbitals including

¹For the exact KS potential it can be easily shown that the density decay exponentially for $r \rightarrow \infty$:

For a point far away from the system, an electron should only feel the potential $-1/r$, since the interaction of the nuclei and the other $N-1$ electrons cancel. The KS equation then can be reduced to

$$\left(-\frac{1}{2r} \frac{d^2}{dr^2} r - \frac{1}{r}\right) \phi_i(r) = \epsilon_i(r) \phi_i(r), \quad (1.54)$$

which is identical to the hydrogen system with the corresponding solutions. The asymptotic form of $\epsilon_i(r)$ is

$$\phi_i(r) \sim e^{-\sqrt{-2\epsilon_i}}. \quad (1.55)$$

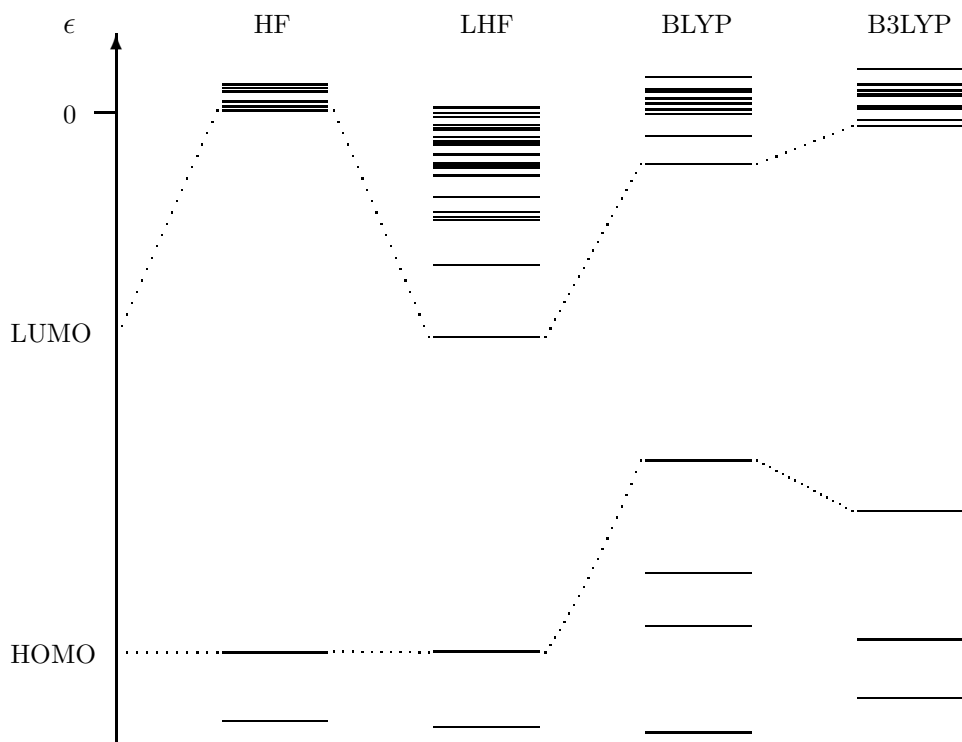


Fig. 1.3: Schematic HF-, exact exchange KS-, and GGA-eigenvalue spectrum

the unoccupied ones. It can be easily shown that the LHF-potential falls off like $-1/r$ for $r \rightarrow \infty$:

It has already been said that for a point far away from the system, nuclear and Coulomb potential cancel each other and the asymptotic behavior is mainly determined by the exchange potential. The LHF exchange potential in equation 1.43 can be written as a sum of the Slater potential and a correction term:

$$v_x(\mathbf{r}) = v_x^S(\mathbf{r}) + v_x^{\text{corr}}(\mathbf{r}). \quad (1.56)$$

For $r \rightarrow \infty$ the correction term

$$v_x^{\text{corr}}(\mathbf{r}) = \frac{2}{\rho(\mathbf{r})} \sum_{(a,b) \neq (N,N)}^{\text{occ.}} \phi_a(\mathbf{r})\phi_b(\mathbf{r}) [\langle \phi_b | \hat{v}_x^{\text{LHF}} | \phi_a \rangle - \langle \phi_b | \hat{v}_x^{\text{NL}} | \phi_a \rangle] \quad (1.57)$$

vanishes exponentially. This is because the factors $\phi_a^{KS}\phi_b^{KS}/\rho$ that enter the correction term vanish if not both ϕ_a^{KS} and ϕ_b^{KS} equal the HOMO, because the electron density in the asymptotic region decays as the square of the HOMO and the ratios ϕ_a^{KS}/ϕ_N^{KS} asymptotically

vanish exponentially for $a \neq N$ due to the asymptotic behavior of the KS-orbitals. The term in which $\phi_a = \phi_b = \phi_N$ had been eliminated from 1.57. The fall-off behavior of the potential in the asymptotic region is therefore only determined by the Slater potential

$$\begin{aligned}
 v_x^S(\mathbf{r}) &= -\frac{2}{\rho(\mathbf{r})} \sum_{a,b}^{occ.} \phi_a(\mathbf{r})\phi_b(\mathbf{r}) \int d\mathbf{r}' \frac{\phi_b(\mathbf{r}')\phi_a(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \\
 &= -\frac{2\phi_N(\mathbf{r})\phi_N(\mathbf{r})}{\rho(\mathbf{r})} \int d\mathbf{r}' \frac{\phi_b(\mathbf{r}')\phi_a(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \\
 &= -\frac{1}{r} \int d\mathbf{r}' \phi_b(\mathbf{r}')\phi_a(\mathbf{r}') = -\frac{1}{r}, \tag{1.58}
 \end{aligned}$$

By the same argument as given for the correction term, the terms $\phi_a^{KS}\phi_b^{KS}/\rho$ vanish and only the term with $\phi_a^{KS} = \phi_b^{KS} = \phi_N^{KS}$ is left of the sum. $2\phi_N^2$ in the asymptotic region is equal to the density and can be canceled. In the next step $|\mathbf{r}-\mathbf{r}'|$ is set equal to r since \mathbf{r} is large compared to \mathbf{r}' , so that the integration can be reduced to the norm of the HOMO, which is one, and the total potential decays as $-1/r$. Due to its correct asymptotic behavior, the LHF-potential yields a physically correct eigenvalue spectrum as it is shown in fig. 1.3. The eigenvalue spectrum contains the antibonding orbitals and the complete Rydberg series. The effects of the correct $-1/r$ asymptotic decay on the structure of the virtual orbitals will be discussed in great detail in the next chapter.

Chapter 2

Kohn Sham Orbitals as One Electron Basis Set for Ab Initio Methods

To test the quality of a given set of orbitals in a CI-method, two principal approaches are possible: One is to compute the desired states with the same computational effort and to compare the accuracy that is reached (for example in the computed excitation energies). The other possibility is to compute all states in the same accuracy and to compare the computational effort that is needed to reach this accuracy. In this work, the second approach is chosen. The main reason to choose the second approach this is that the main property that determines whether a one electron basis set is useful for a CI-method or not is the convergence of the CI method. Therefore, it is straightforward to chose the convergence itself as a quality criterion.

2.1 MRCI-calculations for ethene

The first example of this analysis are the excited states of ethene that are computed with the MRCI-approach. Table 1 gives the computed excitation energies². As the quality criterion it also contains the number of the reference configurations which have to be taken into account to push the error of the computation to 0.1 eV or less. More information about the composition of the reference spaces of a few selected states can be taken from Table 2 which contains the weights of the reference configurations in the MR-CI wave function. Differences between the structure of orbitals obtained from various ansätze can be taken from Figure 1. Exemplary it gives the contour plots of the π^* -orbital (b_{2g} symmetry) and of the four lowest

²Computational details for section 2.1:

To obtain a better comparability with the literature, the geometry employed by Serrano-Andrés *et al.*^[70] with $R_{CC} = 1.339$ Å, $R_{CH} = 1.086$ Å and $\angle_{HCC} = 117.6^\circ$ was used to calculate the vertical excitation energies of ethene. The ethene molecule lies in the yz plane with the z-axis going through the carbon atoms. A (12s7p3d) basis set^[71] contracted to [8s7p3d] was employed for carbon and a 6s3p1d un-contracted basis set for hydrogen. Primitive gaussian functions with the exponents 0.012138 (s-type), 0.0042482 (s-type), 0.0080150 (p-type), 0.0028052 (p-type), 0.028512 (d-type) were added on the carbon atoms.

The CASSCF-orbitals were optimized for each given state employing a (2,11)CASSCF-space. For the 1^3A_g -state a (2,12)CASSCF-calculation was performed. For all CASSCF-calculations the MOLCAS program package^[72] was used.

In the LHF calculations, the Slater potential was constructed exclusively with the algorithm described in eq. (25), (26) and (27) of Della Sala and Görling.^[63] In the asymptotic region far from the molecule the LHF exchange potential was constructed using asymptotic continuations of the KS-orbitals with the parameter M_t determining the asymptotic region set to 10.

The reference spaces for the MRCI-computations (multireference singles-doubles-CI) were constructed by the following selection procedure: If n states are to be computed, a selective MRCI with a reference wavefunction of the n+3 configuration state functions lowest in energy is performed. The selection is done by Epstein Nesbith perturbation theory with a selection threshold of 0.1 mH. Configurations of the resulting wavefunction contributing more than a given c^2 -value are again selected as the new reference wavefunction for another selective MRCI which is repeated until a constant reference space is reached. A more detailed description of this selection procedure is given in section 2.9. To reach an accuracy of 0.1 eV and less, selection thresholds between $c^2 > 0.008 - 0.005$ had to be chosen. The influence of higher excitations is estimated by the normalized form of the Davidson correction.^[73] The complete valence shell is correlated. All MRSD-CI and MRPT2^[4,5] calculations were performed with the DIESEL-MR-CI program package^[74] in combination with the MOLCAS program package.^[72]

lying orbitals of b_{3u} symmetry (π -symmetry).

Except for HF-orbitals the reference space of the X 1A_g ground state is composed of two configurations. Besides the groundstate configuration $\Psi_0 = |1a_g^2 1b_{1u}^2 2a_g^2 2b_{1u}^2 1b_{2u}^2 3a_g^2 1b_{3g}^2 1b_{3u}^2\rangle$ also the configuration obtained from the $\pi^2 \rightarrow \pi^{*2}$ double excitation is found to possess a weight larger than 0.02 (Table 2). The actual numbers vary to some extent (LHF: 0.023; CAS: 0.026; BLYP: 0.022) but the overall picture is similar for these single-electron bases. Despite this similarity the shapes of the π^* -orbitals differ (Figure 1). Both the LHF and the BLYP π^* -orbital have the form of the intuitive π^* -orbital. The π^* -orbital obtained from a groundstate (2,11) CASSCF-computation,^[70] however, does not have this intuitive shape but possesses an additional nodal plane which results since the orbital is optimized for an efficient description of correlation effects. The energetically lowest HF-orbital of π^* -symmetry (b_{2g}) is very diffuse and cannot be identified as the intuitive anti-bonding π^* -orbital. Instead the character of the intuitive π^* -orbital is dissolved in various higher lying orbitals. As a consequence the corresponding reference space consists of three configurations instead two.

This interpretation is clearly confirmed by the results for the lowest lying triplet state, 1^3B_{1u} , where 6 reference configurations are needed for an accurate calculation if HF-orbitals are employed. Employing LHF- or BLYP-orbitals only one reference configuration possessing a weight of more than 0.9 is needed for an accurate computation showing that both approaches provide very appropriate orbitals to describe this excited state. On the other hand employing the orbitals of a (2,11) CASSCF computation for the 1^3B_{1u} state one additional configuration is found which has to be included in the description of this lowest lying excited triplet state. It is worth to note that the orbitals obtained from a (2,11) CASSCF computation for the X 1A_g ground state are not suitable for a description of the 1^3B_{1u} state. This underlines the above interpretation that the CASSCF-orbitals are optimized to describe correlation effects but do not reflect the intuitive orbitals describing the excited states of the molecule.

As shown in previous works, the computational determination of the 1^1B_{1u} state is very complicated. The maximum of the absorption band is found at 7.66 eV, but for the vertical excitation Serrano-Andres *et al.* suppose a value of 8.0 while the value of 7.66 eV is attributed to a somewhat twisted geometry.^[70] Recently, based on new computations, Lischka and coworkers located the vertical transition around 7.7 eV.^[75]

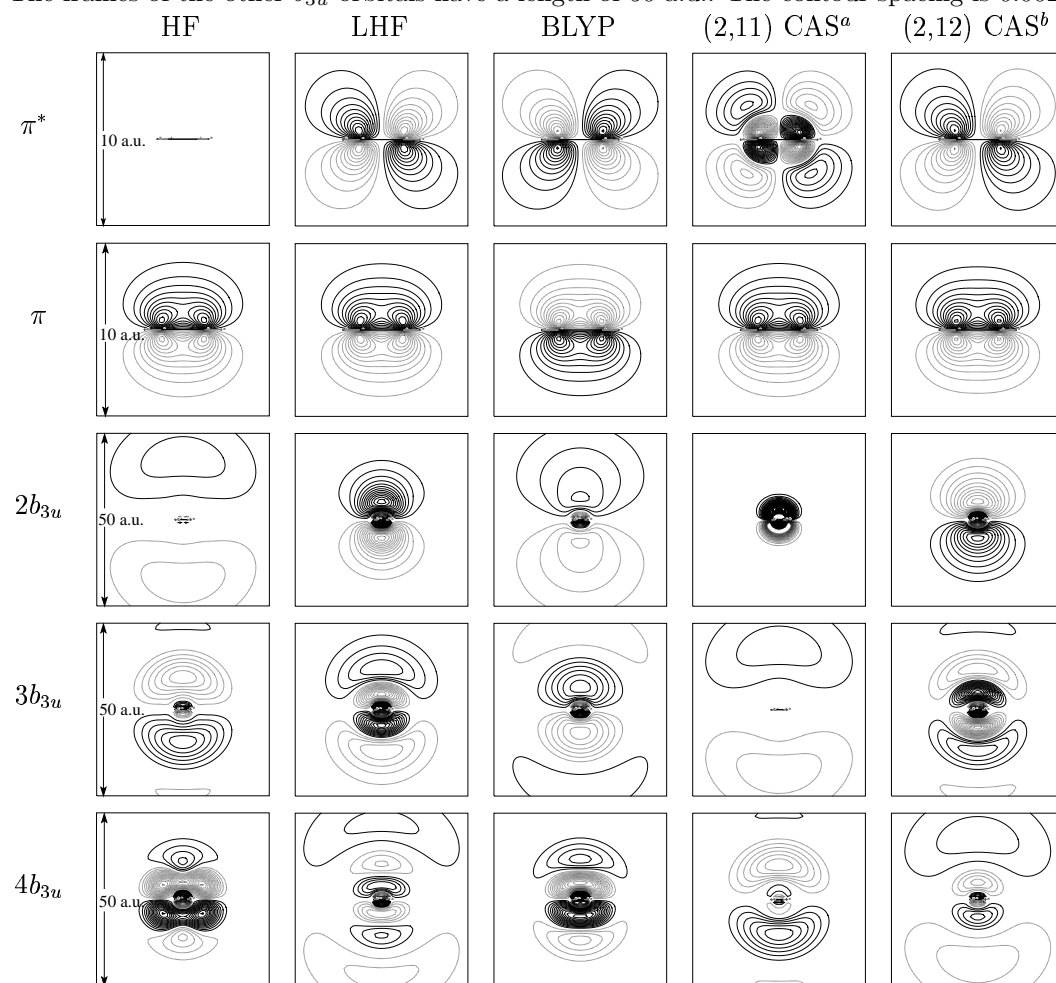
Tab. 2.1: Computed excitation energies and number of reference configurations being necessary to obtain an accuracy of about 0.1 eV. The number contains those reference configurations contributing with a value of c^2 higher than 0.004 to the MR-CI wave function. The reference space contains additional configurations which had to be included, if several electronic states were computed at the same time. The total number of configurations is given in parenthesis.

state	class.	HF		LHF		BLYP		CAS		exp. ^[a]
		n_{Ref}	ΔE	n_{Ref}	ΔE	n_{Ref}	ΔE	n_{Ref}	ΔE	
X $^1A_{1g}$	Ψ	3	-	2	-	2	-	2	-	-
1 $^1B_{1u}$	$\pi \rightarrow \pi^*$	3	7.96	3	7.94	3	7.95	1	8.17	7.66 ^[b]
1 $^1B_{3u}$	$\pi \rightarrow 3s$	9	7.17	1	7.22	1	7.24	1	7.26	7.11
1 $^1B_{1g}$	$\pi \rightarrow 3p\sigma$	4	7.86	1	7.88	3	7.82	-	-	7.80
1 $^1B_{2g}$	$\pi \rightarrow 3p\sigma$	7	7.96	1	8.00	3	7.96	-	-	7.90
2 1A_g	$\pi \rightarrow 3p\pi$	4 (7)	8.28	1 (2)	8.28	2 (4)	8.31	-	-	8.28
2 $^1B_{3u}$	$\pi \rightarrow 3d\sigma$	10 (10)	8.61	2 (4)	8.63	5 (5)	8.66	-	-	8.62
2 $^1B_{1u}$	$\pi \rightarrow 3d\pi$	-	-	4 (4)	9.21	3 (3)	9.19	-	-	9.33
1 $^1B_{2u}$	$\pi \rightarrow 3d\delta$	-	-	1	9.12	1	9.13	-	-	9.05
1 $^3B_{1u}$	$\pi \rightarrow \pi^*$	6	4.45	1	4.48	1	4.49	2	4.43	4.36
1 $^3B_{3u}$	$\pi \rightarrow 3s$	9	7.06	1	7.10	1	7.14	-	-	6.98
1 $^3B_{1g}$	$\pi \rightarrow 3p\sigma$	5	7.79	1	7.83	2	7.87	-	-	7.79
1 $^3B_{2g}$	$\pi \rightarrow 3p\sigma$	7	7.91	1	7.97	2	8.00	-	-	-
1 3A_g	$\pi \rightarrow 3p\pi$	4	8.13	1	8.13	3	8.11	1	8.09	8.15
2 $^3B_{3u}$	$\pi \rightarrow 3d\delta$	-	-	2 (3)	8.57	6 (6)	8.55	-	-	8.56

^[a]Experimental data are taken from Serrano *et al.*^[70]

^[b]The value of the vertical excitation energy of the 1^1B_{3u} -state is widely discussed in the literature. Serrano *et al.* give 8.0 eV, while the 7.66 eV is attributed to a somewhat twisted geometry.^[70] Recent computations by Lischka and coworkers predict a vertical excitation energy around 7.7 eV.^[75] For this case, the error may exceed 0.1 eV. The number of reference configurations, in particular in the case of CAS, therefore does not reflect the quality of the orbitals (details see text).

Fig. 2.1: Spatial structure of the π^* -orbital and of the four lowest lying orbitals of b_{3u} -symmetry. The frame length for the π^* and the π -orbitals is 10 a.u. and a contour spacing of 0.025 a.u. is used. The frames of the other b_{3u} -orbitals have a length of 50 a.u.. The contour spacing is 0.0025 a.u..



^a obtained from (2,11) CASSCF for the ground state, the $3b_{3u}$ and $4b_{3u}$ are not included the CAS-space.

^b obtained from (2,12) CASSCF for the 1^3A_g state, the $4b_{3u}$ is not included the CAS-space.

Tab. 2.2: Composition of the reference spaces of selected electronic states.

Zust	class.	HF	LHF	BLYP	CAS
X 1A_g	Ψ_0	0.8892	0.8828	0.8814	0.8824
	$\pi^2 \rightarrow 1b_{2g}^2$	-	0.0230	0.02677	0.0222
	$\pi^2 \rightarrow 5b_{2g}^2$	0.0067	-	-	-
	$\pi^2 \rightarrow 5b_{2g}^1 6b_{2g}^1$	0.0061	-	-	-
1 $^1B_{1u}$	$\pi \rightarrow 1b_{2g}$	0.0905	0.7210	0.7691	0.8983
	$\pi \rightarrow 2b_{2g}$	0.4590	0.1308	0.0481	-
	$\pi \rightarrow 3b_{2g}$	0.3339	0.0312	0.0613	-
2 1A_g	$\pi \rightarrow 2b_{3u}$	0.1078	0.8799	0.4868	-
	$\pi \rightarrow 3b_{3u}$	0.5749	-	0.3967	-
	$\pi \rightarrow 4b_{3u}$	0.1233	-	-	-
	$\pi \rightarrow 5b_{3u}$	0.0795	-	-	-
	Ψ_0	0.0002 ^a	0.0001	0.0001	-
	$\pi^2 \rightarrow 1b_{2g}^2$	-	0.0026	0.0008	-
	$\pi^2 \rightarrow 5b_{2g}^2$	0.0005	-	-	-
	$\pi^2 \rightarrow 5b_{2g}^1 6b_{2g}^1$	0.0011	-	-	-
1 3A_g	$\pi \rightarrow 2b_{3u}$	0.0795	0.8928	0.3484	0.9123
	$\pi \rightarrow 3b_{3u}$	0.5208	-	0.4151	-
	$\pi \rightarrow 4b_{3u}$	0.1704	-	0.1277	-
	$\pi \rightarrow 5b_{3u}$	0.1190	-	-	-
2 $^3B_{3u}$	$\pi \rightarrow 4a_g$	-	0.0017	0.0250	-
	$\pi \rightarrow 5a_g$	-	0.7488	0.0625	-
	$\pi \rightarrow 6a_g$	-	0.1399	0.0679	-
	$\pi \rightarrow 7a_g$	-	-	0.2023	-
	$\pi \rightarrow 8a_g$	-	-	0.4906	-
	$\pi \rightarrow 9a_g$	-	-	0.0392	-

While LHF- and BLYP-orbitals need only 1 reference configuration to characterize the 1^3B_{1u} state they need 3 reference configurations to compute the corresponding singlet state 1^1B_{1u} (Table 1). The reference space consists of the main configuration describing the single excitation $\pi \rightarrow 1b_{2g}(\pi^*)$ from Ψ_0 which possesses a weight of 0.7210 (LHF) and 0.7691 (BLYP). For both types of orbitals the single excitations $\pi \rightarrow 2b_{2g}$ and $\pi \rightarrow 3b_{2g}$ represent two addi-

tional reference configurations. Both excitations are the dominant reference configurations for the next state of this symmetry (2^1B_{1u} from $\pi \rightarrow d_\pi$ Rydberg excitation), showing the interaction between the two states which was already discussed in the literature.^[70,76,77] Although the weights of this additional reference configurations differ to some extent going from LHF- to BLYP-orbitals the overall picture is very similar.

For a good representation of the 1^1B_{1u} state HF also needs only 3 reference configurations. However, as can be seen from Table 2 the emerging picture is completely different. Employing HF-orbitals the two single excitations $\pi \rightarrow 2b_{2g}$ and $\pi \rightarrow 3b_{2g}$ represent the dominant contributions to this state. This shows again, that the intuitive character of the π^* -orbital is dissolved in various higher lying orbitals. Employing CASSCF-orbitals optimized for the given state, only one reference configuration is found, however, the computed excitation energy is about 0.2 eV higher than those obtained with HF-, LHF-, or BLYP-orbitals. As already pointed out by Serrano et al. CASSCF-orbitals seem to be less efficient in the description of such mixing.^[70] This can also be seen from the work of Lischka and coworkers who needed very large reference spaces to push the excitation energy below 7.8 eV.^[75]

For the electronic Rydberg states arising due to the $\pi \rightarrow 3s$ Rydberg transitions, namely the 1^1B_{3u} and the 1^3B_{3u} state both LHF- and BLYP-orbitals already provide a good representation if only one reference configuration is involved. The electronic state arising due to the $\pi \rightarrow 3p_\sigma$ transitions seems to be more complicated, however. While LHF-orbitals seem to be still perfect, three reference configurations are needed if BLYP-orbitals are employed as single-electron basis. The additional reference configurations possess weights of about 0.11 and 0.06. An enlargement of the reference space used for the LHF ansatz changes the excitation energies by less than 0.05 eV. HF-orbitals again do not provide the intuitive orbitals reflecting the excited states.

A similar situation is found for the 2^1A_g and the 1^3A_g states which represent the $\pi \rightarrow 3p_\pi$ -Rydberg transitions. Again only LHF provides orbitals reflecting the intuitive character of these Rydberg orbitals while for BLYP-orbitals the character is dissolved into various orbitals. This can be seen in the shape of the b_{3u} orbitals which represent the Rydberg np_π sketched in Figure 1. The LHF-orbitals nicely reflect the np_π Rydberg series in the intuitive ordering starting with the least diffuse one. The BLYP np_π Rydberg series starts with the very diffuse $2b_{3u}$ and $3b_{3u}$ orbitals. The inner region of the $3b_{3u}$ BLYP-orbital resembles

the $2b_{3u}$ LHF-orbital. As a result employing the BLYP-orbitals the reference spaces for the 2^1A_g and the 2^3A_g states are dominated by configurations arising from both the $\pi \rightarrow 2b_{3u}$ and the $\pi \rightarrow 3b_{3u}$ excitation while the configuration arising due to the $\pi \rightarrow 4b_{3u}$ excitation possesses a much smaller weight.

The electronic states arising due to $\pi \rightarrow 3d$ -Rydberg excitations seem to be even more complicated. Also for LHF-orbitals a correct description of the 2^1B_{3u} and the 2^1B_{1u} needs more than one reference configuration. In this case the analysis of the wave function points more to a real interaction between two states than to problems in the orbital representation. Summarizing, the first example shows that LHF-orbitals should be well suited as single electron basis for MR-CI computations. This underlines that they reflect intuitive chemical concepts and can be identified as, *e.g.*, anti-bonding or Rydberg orbitals. Also BLYP-orbitals are usable for the description of some states but are less suitable for the description of higher lying states. As expected HF-orbitals represent the worst set of orbitals. As known before, CAS orbitals are well suited for MR-CI computations.

2.2 Relaxation of the orbitals describing the core ion of the excited states

When an electron is excited from an occupied into a previously unoccupied orbital, the orbitals of the cationic core contract. Furthermore, density sometimes is redistributed along molecular axes, *i.e.*, the polarization of bonds can change. Since this work uses ground state orbitals, it needs to be understood how the orbital relaxation is described in the MR-CI-computations and whether this might lead to errors in the computation of the excited states. In this section, a number of energy variables is introduced that will be used in the following discussion of orbital relaxation.

For a precise calculation of electronically excited states, orbital relaxation effects, static, and dynamical correlation have to be considered. All of these effects are taken into account in the complete MRSD-CI+Q energy ($E_{\text{MR-CI+Q}}$). On the other hand, none of these effects are included in the energy expectation value (E_{det}) of the configuration state functions, which in

sections ?? and 2.4 are called determinants³ which in a simple one-particle picture represent the excited states and which are built from HF or KS orbitals optimized for the groundstate by carrying out excitations in the corresponding HF or KS determinant. Static correlation energy is normally defined as the energy lowering with respect to the restricted (open) shell Hartree-Fock (R(O)HF) energy introduced by adding enough flexibility in the wave function to be able to qualitatively describe the system.^[78] In this study, as static correlation is defined by the difference between the simple determinantal energy (E_{det}) and the energy obtained for E_{ref} if the state is treated by a small CI expansion within the space consisting only of the reference configurations. This definition is slightly different to the general one since the reference space was chosen in a way that the deviations of excitation energies obtained from the MRSD-CI+Q calculations from experimental values are lower than 0.1 eV. However, both are close enough to justify this definition. Finally the energy difference between E_{ref} and $E_{\text{MRCI+Q}}$ gives the sum of the remaining effects namely the orbital relaxation and dynamical correlation. Within the text this sum will be abbreviated as $E_{\text{corr}}^{\text{rel,dyn}}$. In the case of Rydberg states which represent systems with a weakly interacting electron cycling around a positively charged core, orbital relaxation effects of the core orbitals can be estimated by computing the corresponding cationic system employing an ROHF ansatz. The occupied orbitals from this ROHF computation include orbital relaxation effects although probably to a slightly larger extent than actually found for Rydberg states of the neutral molecule. Furthermore, the virtual orbitals for the cations should reflect the correct shapes of Rydberg orbitals since within the ROHF computation they “feel” the potential of a singly charged core which is the correct potential for Rydberg electrons. Orbital relaxation effects (E_{rel}) arising due to the loss of the electron can be estimated from the energy difference between E_{det} for the cation computed with ground state orbitals and the ROHF energy for the cation. Finally, subtracting E_{rel} from $E_{\text{corr}}^{\text{rel,dyn}}$ which includes both dynamical correlation and orbital relaxation gives information on the magnitude of dynamical correlation. Within the tables,

³All excited states considered here possess two singly occupied orbitals. For triplet states the situation is simple since if both orbitals are occupied with α spin electrons. This leads to one determinant which already possesses the correct S^2 and M_s quantum number. For the singlet states two determinants arise. Only proper linear combinations of both are eigenfunctions to the S^2 operator. These linear combinations are called configuration state functions (CSF). However, to avoid confusion within the text the term “determinant” is used for both triplet and singlet states and E_{det} for the corresponding energies.

excitation energies obtained on the various levels are given. Excitation energies estimated with orbital energy differences ΔE_{orb} are also given in the tables.

2.3 Discussion of the water molecule

Table 2.3 gives computed excitation energies⁴ for the water molecule that are obtained if LHF-orbitals are used as one electron basis. Besides the excitation energies also the three energetically lowest ionization energies are included. Except for the 1^1A_1 state all reference spaces are built up from only two configurations, *e.g.*, the configurations $1b_1 \rightarrow 3p_x$ and $1b_1 \rightarrow 3d_{xz}$ in the calculation of the excited states 1^1A_2 and 2^1A_2 . This represents the smallest possible reference space, because always two states, *i.e.* two roots, are computed at the same time. As seen from table 1 $\Delta E_{\text{MRCI+Q}}$ values obtained on the basis of LHF-orbitals are in excellent agreement with their experimental counterparts. Computations with larger reference spaces change $\Delta E_{\text{MRCI+Q}}$ by less than 0.05 eV. Both findings show that the minimal reference spaces (two configurations for two states) are already sufficient for an accurate description if LHF-orbitals are employed. Additionally, the small differences between ΔE_{det} and ΔE_{ref} underline that the two states computed together do not mix, *i.e.*, the nature of a given electronically excited state is dominated by only one determinant (configuration state function) characterized by the one-particle excitation given in the second column of table 1. This behavior is excellent for the assignment of electronically excited states. It would be very advantageous if the properties of the LHF-orbitals described above lead to the possibility to accurately estimate the electronic spectrum without or partly without the consideration of the remaining effects namely dynamical correlation and electron relaxation. However, as shown in table 2.3, this is impossible. A computation employing the orbital energies of the

⁴Computational details for section 2.3:

All calculations were performed at the experimental geometry being $R_{\text{OH}} = 0.975 \text{ \AA}$, and $\angle_{\text{HOH}} = 116.5^\circ$.^[79] The water molecule is chosen to lie in the yz plane with the z-axis being collinear to the C_2 -rotation axis. (12s7p3d) basis sets^[71] contracted to [8s7p3d] were employed for oxygen and a 6s3p1d un-contracted basis set for hydrogen. Primitive gaussian diffuse functions with the exponents 0.0320 (s-type), 0.0100 (s-type), 0.0066 (s-type), 0.0030 (s-type), 0.0280 (p-type), 0.0100, (p-type) 0.0540 (p-type), 0.0150 (d-type), 0.0090 (d-type), 0.0032 (d-type) were added on the oxygen atom.

Tab. 2.3: Computed excitation energies of the Rydberg states of water employing LHF-ground state orbitals. Only minimal reference spaces were used. For more information and an explanation of the various quantities see text. For the groundstate: $E_{\text{det}} = E_{\text{ref}} = E_{\text{LHF}} = -76.06045$ a.u.; $E_{\text{MRCI+Q}} = -76.32745$ a.u.. All energies in eV.

state	assignment	ΔE_{orb}	ΔE_{det}	ΔE_{ref}	$\Delta E_{\text{MRCI+Q}}$	$\Delta E_{\text{exp}}^{[a]}$	$E_{\text{corr}}^{\text{rel,dyn}}$
X 1A_1	Ψ_0	-	-	-	-	0.00	-7.26
2 1A_1	$3a_1 \rightarrow 3s/1b_1 \rightarrow 3p_y$	10.18 ^[b]	11.41	10.95	9.66	9.7	-8.55
1 1A_2	$1b_1 \rightarrow 3p_x$	10.08	10.31	10.32	9.15	9.1	-8.41
2 1A_2	$1b_1 \rightarrow 3d_{xz}$	11.85	12.05	12.05	10.65		-8.67
1 1B_2	$1b_1 \rightarrow 3d_{xy}$	12.21	12.32	12.31	10.83		-8.74
2 1B_2	$1b_1 \rightarrow 4d_{xy}$	12.92	12.95	12.97	11.49	11.40	-8.74
1 1B_1	$1b_1 \rightarrow 3s$	8.13	8.66	8.66	7.39	~ 7.5	-8.53
2 1B_1	$1b_1 \rightarrow 3p_z$	10.99	11.09	11.09	9.77	~ 10.0	-8.59
1 3A_1	$3a_1 \rightarrow 3s$	10.18	10.44	10.43	9.34	9.30	-8.41
2 3A_1	$1b_1 \rightarrow 3p_y$	11.00	10.77	10.79	9.57		-8.48
1 3A_2	$1b_1 \rightarrow 3p_x$	10.08	10.06	10.05	9.00	~ 9.0	-8.32
2 3A_2	$1b_1 \rightarrow 3d_{xz}$	11.85	11.98	11.98	10.82		-8.67
1 3B_2	$1b_1 \rightarrow 3d_{xy}$	12.21	12.27	12.29	10.82		-8.74
2 3B_2	$3a_1 \rightarrow 3p_x$	12.92	12.24	12.23	11.24		-8.61
1 3B_1	$1b_1 \rightarrow 3s$	8.13	8.09	8.09	7.07	~ 7.1	-8.26
2 3B_1	$1b_1 \rightarrow 3p_z$	10.99	10.99	10.99	9.76		-8.70
2B_1	$1b_1 \rightarrow \infty$	13.79	13.77		12.34	12.62	-8.59
2A_1	$3a_1 \rightarrow \infty$	15.83	15.94		14.60	14.74	-8.31
2B_2	$1b_2 \rightarrow \infty$	19.62	19.56		18.86	18.51	-7.97

^[a]Experimental excitation energies from Cai *et al.*,^[80] ionization potentials from Brundle.^[81]

^[b]Computed for $3a_1 \rightarrow 3s$.

LHF-orbitals predicts excitation energies which are about 1 - 2 eV too high. Also the ionization energies are overestimated by the same number. Moreover, multiplet splittings, *e.g.*, the difference between the 1^1A_2 and the 1^3A_2 excitation energies, are principally inaccessible by orbital energy differences. The differences between ΔE_{orb} and ΔE_{det} are smaller than 1 eV but for singlet states the agreement between theory and experiment becomes even worse if ΔE_{det} instead of ΔE_{orb} is used, even though in ΔE_{det} multiplet splitting shows up. As expected, for triplet states only small differences between ΔE_{orb} and ΔE_{det} are found. As already discussed above the situation does not change if the interaction between the reference configurations is taken into account (ΔE_{ref}) showing that static correlation effects are of minor importance if the MRSD-CI+Q calculation is based on LHF-orbitals. The situation is only improved if the remaining effects dynamical correlation and orbital relaxation, are also included in the calculation. The fact that computed excitation energies are then shifted by about 1 eV or more to lower values shows that these effects ($E_{\text{corr}}^{\text{rel,dyn}}$) are more important for the excited than for the ground state. Indeed, Table 1 shows that $E_{\text{corr}}^{\text{rel,dyn}}$ is somewhat more than 1 eV more negative for excited states than for the ground state.

Table 2.3 shows that an accurate description of the 1^1A_1 state needs three instead of two reference configurations. However, this does not result from shortcomings of the employed LHF-orbitals but results from an avoided crossing between two states of a_1 symmetry taking place at a bond angle of about 100° . This avoided crossing was discussed to explain the vapor-phase dissociation of H_2O induced by photons or electrons having an energy of 10 eV or more.^[82,83]

Summarizing, LHF-orbitals seem to be well suited as one electron basis for subsequent MRSD-CI computations of the excited states, since only a minimal number of reference configurations is needed. As a consequence an assignment of the excited states to a single excitation (*e.g.*, $1b_1 \rightarrow 3p_x$ for the 1^1A_2 state) is straightforward in most cases. Deviations from this behavior point to real interaction between two electronically excited states. This is supported by the shape of the LHF-orbitals of a_1 symmetry depicted in figure 2.2. However, the strong influence of the effects summarized in $E_{\text{corr}}^{\text{rel,dyn}}$ is quite intriguing and might in some cases limit the usefulness of LHF-orbitals for the description of excited states.

The results obtained with HF ground state orbitals are summarized in table 2.4. In difference to LHF-orbitals the sizes of reference spaces necessary to obtain an accuracy of about 0.1 to

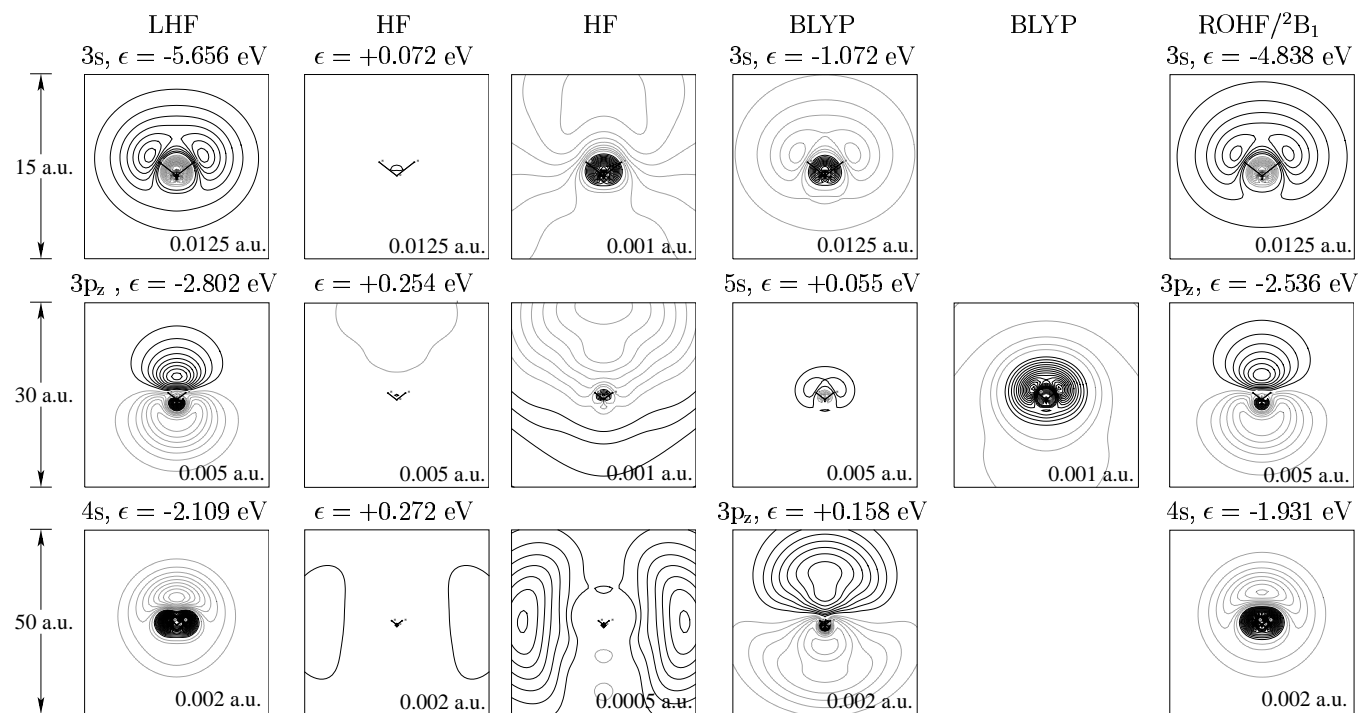
0.2 eV for the vertical excitation energies range from 6 to 9. Only for states of b_2 symmetry already 3 reference configurations are sufficient. The importance of large reference spaces is also reflected in the differences between ΔE_{det} and ΔE_{ref} . If the whole reference space is taken into account the error in most cases is reduced by more than one 1 eV showing that static correlation effects are very large if HF-orbitals are used. ΔE_{ref} still possesses errors of 1 - 2 eV underlining that also dynamical correlation and orbital relaxation effects cannot be neglected. The size of these effects in the total energies, summarized in $E_{\text{corr}}^{\text{rel,dyn}}$, ranges around 8.8 - 9.0 eV, *i.e.*, it possesses the same magnitude as for LHF-orbitals. This indicates that the larger reference spaces are necessary for the description of the electronic characters of the excited states. Their characters can only be accessed by an overlay of several configurations, since the shapes of the HF-orbitals do not mirror the form of Rydberg orbitals. This can be seen from the shape of the HF-orbitals of a_1 symmetry depicted in figure 2.2. In comparison to LHF-orbitals the virtual HF-orbitals are not only too diffuse, they do not even reflect the general shape. This behavior is expected since virtual HF-orbitals “feel” the potential of the neutral system instead of the potential of a positively charged core. However, once the character of an excited state is correctly described by a linear combination of several configurations, the remaining effects, *e.g.*, dynamical correlation and orbital relaxation, possess the same magnitude as those found for the LHF-orbitals. The discussion shows that for HF-orbitals a clear assignment of the excited states is problematic. Consequently, an assignment and the corresponding orbital energies are omitted in table 2.4. The physical interpretation can be taken from table 2.3.

The results obtained when BLYP-orbitals are employed for the MRSD-CI-computation of the excited states of the water molecule are summarized in tables 2.5 and 2.6. While table 2.5 gives the excitation energies computed with minimal reference spaces (2 reference configurations for 2 states), table 2.6 summarizes the values obtained for extended reference spaces. A comparison between table 2.3 (LHF results) and table 2.5 (BLYP results) shows small but remarkable differences between both sets of orbitals.

Tab. 2.4: Computed excitation energies of the Rydberg states of water using HF-orbitals. Energies are given in eV. n_{ref} gives the number of reference configurations with $c^2 \geq 0.1$. If the actual number of reference configurations used in the calculation differs from the number being important for the description of the state, the total number is given in brackets. For the groundstate: $E_{\text{det}} = E_{\text{ref}} = E_{\text{HF}} = -76.06398$ a.u.; $E_{\text{MRCI+Q}} = -76.32886$ a.u. All energies in eV.

state	n_{ref}	ΔE_{det}	ΔE_{ref}	$\Delta E_{\text{MRCI+Q}}$	ΔE_{exp}	$E_{\text{corr}}^{\text{rel,dyn}}$
X 1A_1	1	-	-	-	-	-7.21
2 1A_1	6 (7)	12.59	11.55	9.82	9.7	-8.93
1 1A_2	3 (9)	12.51	10.77	9.07	9.1	-8.92
2 1A_2	4 (9)	12.73	12.21	10.61		-8.81
1 1B_2	3	12.94	12.44	10.83		-8.81
2 1B_2	3	12.98	13.07	11.49	11.40	-8.78
1 1B_1	3 (7)	12.20	9.50	7.28	~ 7.5	-9.42
2 1B_1	3 (7)	12.67	11.86	10.01	~ 10.0	-9.06
1 3A_1	9 (15)	14.65	10.86	9.27	9.30	-8.80
2 3A_1	6 (15)	12.51	11.14	9.54		-8.80
1 3A_2	2 (6)	12.90	11.07	9.06	~ 9.0	-9.21
2 3A_2	4 (6)	12.73	12.54	10.91		-8.83
1 3B_2	3	12.94	12.43	10.83		-9.21
2 3B_2	3	12.99	13.06	11.49		-8.83
1 3B_1	3 (7)	12.60	9.24	6.85	~ 7.1	-9.59
2 3B_1	4 (7)	12.60	11.82	9.96		-9.07
2B_1	1	13.87		12.33	12.62	-8.74
2A_1	1	15.92		14.61	14.74	-8.51
2B_2	1	19.54		18.87	18.51	-7.87

Fig. 2.2: Shapes of the three lowest lying orbitals of a_1 -symmetry of the water molecule. Box sizes and the spacing of the contour lines are given in the plot. For HF- and BLYP-orbitals two columns with different contour spacing are given because the structures of the highly diffuse orbitals are not visible in the picture obtained with the higher contour spacing.



The influence of differences in orbital structure from of both sets on the excitation energies are best reflected in ΔE_{det} . On this level, the computations employing either LHF- or BLYP-orbitals strongly overestimate the excitation energies. LHF-orbitals are somewhat superior since they predict lower excitation energies. For states describing excitations into the 3d and the 3p Rydberg orbitals the deviations between both sets of orbitals vary from 0.8 to 1.5 eV, if the 3s or the 4d orbitals are involved smaller deviations of about 0.3 eV and 0.1 eV, respectively, are found. For the triplet states the deviations are even larger (0.5 - 1.8 eV). The differences between both sets of orbitals are also reflected within the singlet triplet gaps (ΔE_{ST}) which on this computational level is equal to the difference in the exchange integral of both singly occupied orbitals. In general the ΔE_{ST} values of the computations employing LHF-orbitals are slightly larger (0 - 0.3 eV) which is the reason that LHF and BLYP results for triplet states differ more than the data for singlet states.

Since the ionization energies only depend on the shapes of the occupied orbitals, the ionization energies obtained on the determinantal level employing either LHF- or BLYP-orbitals differ less (0.2 - 0.3 eV) than the excitation energies. This is in line with the general finding that the occupied orbitals are more similar than the virtual orbitals.

The large differences found between the ΔE_{orb} and ΔE_{det} values obtained with BLYP- and LHF-orbitals, respectively, arise since in contrast to E_{det} the orbital energies are determined by the shapes of the respective orbitals and the functional form of the corresponding exchange potentials. Comparing LHF- and BLYP-orbitals both types of quantities are different. Generally, the large difference between ΔE_{orb} and ΔE_{det} in the LHF as well as the BLYP case shows that if DFT-orbitals are used for subsequent CI calculations, it is problematic to use the DFT orbital energies for the computation of diagonal elements of the CI matrix employing simply exchange integrals of DFT-orbitals for the correction. In this case diagonal and off-diagonal matrix elements are computed on a different footing.

Employing BLYP-orbitals in combination with minimal reference spaces leads to $\Delta E_{\text{MRCI+Q}}$ values which agree quite nicely with the experimental excitation energies. The errors are lower than 0.2 eV in most cases showing an only slightly worse agreement with the experimental values than the predictions obtained with LHF-orbitals. With a difference of 0.55 eV to the LHF result the 2^1A_2 state is an exception, but no experimental results exist for this state.

Tab. 2.5: Computed excitation energies of the Rydberg states of water using BLYP-orbitals. For more information see text or tables 1 and 2. For the groundstate: $E_{\text{BLYP}} = -76.45412$ a.u., $E_{\text{det}} = E_{\text{ref}} = -76.04977$ a.u., $E_{\text{MRCI+Q}} = -76.32529$ a.u.. All energies in eV.

state	assignment	n_{ref}	ΔE_{orb}	ΔE_{det}	ΔE_{ref}	$\Delta E_{\text{MRCI+Q}}$	ΔE_{exp}	$E_{\text{corr}}^{\text{rel,dyn}}$
X 1A_1	Ψ_0	1		-	-	-	-	-7.50
2 1A_1	$3a_1 \rightarrow 3s$	1 (2)	8.20	11.62	11.63	9.59	9.7	-9.54
1 1A_2	$1b_1 \rightarrow 3p_x$	1 (2)	7.27	11.82	11.80	9.31	9.1	-9.99
2 1A_2	$1b_1 \rightarrow 3d_{xz}$	1 (2)	7.42	13.15	13.16	11.20		-9.46
1 1B_2	$1b_1 \rightarrow 3d_{xy}$	2 (2)	7.43	13.14	12.79	10.81		-9.48
2 1B_2	$1b_1 \rightarrow 4d_{xy}$	2 (2)	8.01	13.04	13.40	11.50	11.40	-9.44
1 1B_1	$1b_1 \rightarrow 3s(4a_1)$	1 (2)	6.13	9.04	9.04	7.25	~ 7.5	-9.27
2 1B_1	$1b_1 \rightarrow 3p_z(6a_1)$	1 (2)	7.36	12.50	12.59	10.21	~ 10.0	-9.79
	$1b_1 \rightarrow 5s(5a_1)$	1 (2)	7.26	12.75	12.78	10.69		-9.58
1 3A_1	$3a_1 \rightarrow 3s$	1 (2)	8.20	10.93	10.93	9.20	9.30	-9.23
2 3A_1	$1b_1 \rightarrow 3p_y$	1 (2)	7.36	12.37	12.46	9.97		-9.89
1 3A_2	$1b_1 \rightarrow 3p_x$	1 (2)	7.27	11.78	11.77	9.18	~ 9.0	-10.08
2 3A_2	$1b_1 \rightarrow 3d_{xz}$	1 (2)	7.42	13.15	13.16	11.20		-9.46
1 3B_2	$1b_1 \rightarrow 3d_{xy}$	2 (2)	7.43	13.14	12.79	10.80		-9.48
2 3B_2	$3a_1 \rightarrow 2p_x$	2 (2)	9.35	13.04	13.40	11.50		-9.44
1 3B_1	$1b_1 \rightarrow 3s$	1 (2)	6.13	8.66	8.64	6.89	~ 7.1	-9.25
2 3B_1	$1b_1 \rightarrow 3p_z$	1 (2)	7.36	12.74	12.77	10.66		-9.60
2B_1	$1b_1 \rightarrow \infty$	1	7.205	14.06		12.16	12.62	-9.39
2A_1	$3a_1 \rightarrow \infty$	1	9.278	16.19		14.44	14.74	-9.25
2B_2	$1b_2 \rightarrow \infty$	1	13.156	19.80		18.71	18.51	-8.58

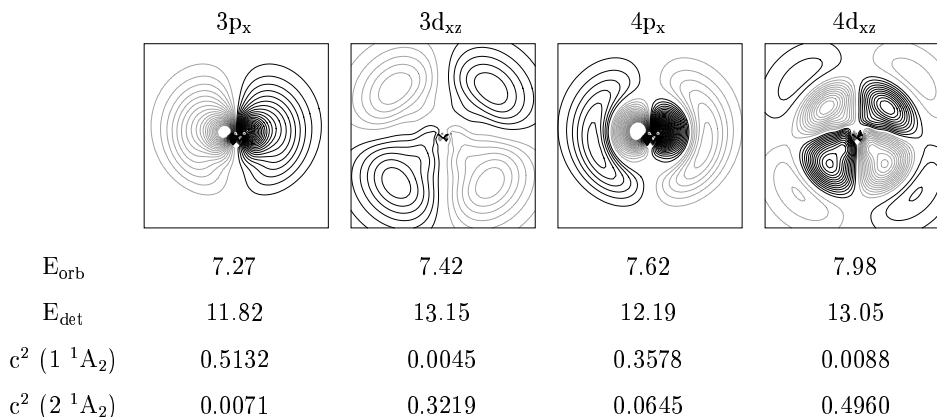
Tab. 2.6: Computed excitation energies of the Rydberg states of water using BLYP-orbitals and enlarged reference spaces. For more information see text or tables 1 and 2. For the groundstate: $E_{\text{BLYP}} = -76.45412$ a.u., $E_{\text{det}} = E_{\text{ref}} = -76.04977$ a.u., $E_{\text{MRCI+Q}} = -76.32529$ a.u.. All energies in eV.

state	n_{ref}	ΔE_{ref}	$\Delta E_{\text{MRCI+Q}}$	ΔE_{exp}	$E_{\text{corr}}^{\text{rel,dyn}}$
$2\ ^1A_1$	1 (3)	11.60	9.58	9.7	-9.52
$1\ ^1A_2$	2 (4)	11.05	9.05	9.1	-9.52
$2\ ^1A_2$	2 (4)	12.81	10.79		-9.52
$1\ ^1B_2$	3 (3)	12.58	10.66		9.42
$2\ ^1B_2$	2 (3)	13.23	11.31	11.40	-9.41
$1\ ^1B_1$	1 (4)	8.99	7.24	~ 7.5	-9.25
$2\ ^1B_1$	1 (4)	11.77	9.67	~ 10.0	-9.59
$1\ ^3A_1$	1 (5)	10.92	9.18	9.30	-9.24
$2\ ^3A_1$	3 (5)	11.19	9.42		-9.27
$1\ ^3A_2$	3 (6)	10.44	8.83	~ 9.0	-9.11
$2\ ^3A_2$	4 (6)	12.32	10.46		-9.35
$1\ ^3B_2$	3 (3)	12.57	10.65		-9.41
$2\ ^3B_2$	2 (3)	13.22	11.31		-9.40
$1\ ^3B_1$	1 (4)	8.59	6.88	~ 7.1	-9.25
$2\ ^3B_1$	3 (4)	11.74	9.64		-9.60

However, while the extension of the reference spaces does not change the energies based on LHF-orbitals the computed $\Delta E_{\text{MRCI+Q}}$ values based on BLYP-orbitals values change considerably if the reference spaces are enlarged (table 2.6). A description of the composition of the extended reference spaces for the 1A_2 and 1B_2 states can be taken from figs. 2.3 and 2.4.

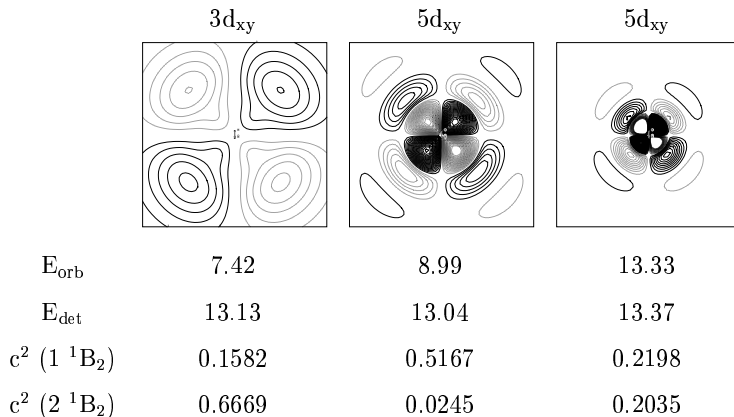
The minimal reference space for the computation of the 1A_2 states contains the configurations $|1a_1^2 2a_1^2 1b_2^2 3a_1^2 1b_1^1 2b_2^1\rangle$ and $|1a_1^2 2a_1^2 1b_2^2 3a_1^2 1b_1^1 3b_2^1\rangle$ which are obtained from the ground state configuration by single excitations out of the $1b_1$ into the $2b_2$ and the $3b_2$ orbitals, respectively. In the extended reference space the configurations arising due to the single excitations out of the $1b_1$ into the $4b_2$ and the $5b_2$ orbitals, respectively, are added. Going from 2 to 4 reference configurations the excitation energies of the 1A_2 states are shifted from 9.31 eV ($1\ ^1A_2$) to 9.05 eV and from 11.2 eV to 10.79 eV ($2\ ^1A_2$) bringing both into better

Fig. 2.3: Shapes of virtual BLYP-orbitals of b_2 symmetry of the water molecule. The box size is 60 a.u., the contour spacing is 0.001 a.u.. In addition, orbital energy differences ($\epsilon_{\text{HOMO}} - \epsilon_{\text{orb}}$) and the expectation values for the 1 and 2 1A_2 states using a single determinant are given. The c^2 -values refer to the contribution to the 1 and 2 1A_2 states. Energies are given in eV.



agreement with the LHF results (9.15 eV, 10.65 eV). Going from the smaller to the extended reference space also the characters of the states change. From figure 2.3 it is seen that for the 1 1A_2 state the configuration $|1a_1^2 2a_1^2 3a_1^2 1b_2^2 1b_1^1 4b_2^1\rangle$ possesses a weight of 0.3578 within the CI expansion. For the extended reference space, the weight of the $|1a_1^2 2a_1^2 3a_1^2 1b_2^2 1b_1^1 4b_2^1\rangle$ configuration is only 0.5132, while its weight in the description with the minimal reference space was 0.8518. The admixture of the higher configurations arises since the $2b_2$ orbital which has a $3p$ nodal structure (the additional node cannot be seen in the figure) is so diffuse that contributions from the higher lying configuration are needed to correct the density in the vicinity of the molecule. The $4b_2$ orbital which is singly occupied in the second important configuration represents the $4p$ on the basis of its nodal structure but at the same time is more compact within the vicinity of the molecule than the $3p$ orbital. A similar pattern is seen for the 2 1A_2 state, however, this time the character of the state changes completely if the extended reference space is used since the $3b_2$ representing the $3d$ due to its nodal structure is far too diffuse. This explains the large error found for the minimal reference space. It is interesting to note that for the BLYP approach orbital and determinantal energies give different sequences of the states. While the series given by the orbital energies follows the nodal structure the sequence given by E_{det} is more influenced by the compactness of the orbitals.

Fig. 2.4: Shapes of virtual BLYP-orbitals of a_2 symmetry of the water molecule. The box size is 60 a.u., the contour spacing 0.001 a.u.. In addition, orbital energy differences ($\epsilon_{\text{HOMO}} - \epsilon_{\text{orb}}$) and the expectation values for the 1B_2 states using a single determinant are given. The c^2 -values refer to the contribution to the 1 and 2 1B_2 states. Energies are given in eV.



Both triplet states of a_2 symmetry are also shifted (9.18 eV to 8.83 eV for the $1 {}^3A_2$ and 11.20 eV to 10.46 eV for $2 {}^3A_2$, respectively) upon using an extended reference space. However, while for the singlet states the agreement to LHF or experimental results improves for the triplet states the results with the larger reference spaces are considerable lower than the LHF (9.0 eV and 10.82 eV) or experimental values (9.0 eV for the $1 {}^3A_2$). A similar behavior is found for all other states. This underestimation is in line with the computed ionization thresholds which, employing BLYP-orbitals, are underestimated by about 0.5 eV. The reasons for the underestimation will become clearer after the discussion of the computations involving orbitals optimized for the cation.

As already mentioned, this study shows that the differences between LHF- and BLYP-orbitals are not too pronounced at least for the energetically lowest ones. Nevertheless, an assignment of the states is already more complicated since more reference configurations are needed if BLYP-orbitals are employed. Another difference is found for the $2 {}^1A_1$ state which employing LHF-orbitals possesses contributions from the $|1a_1^2 2a_1^2 1b_2^2 3a_1^1 1b_1^2 4a_1^1\rangle$ and the $|1a_1^2 2a_1^2 1b_2^2 3a_1^1 1b_1^1 2b_1^1\rangle$ configuration while for BLYP-orbitals the $|1a_1^2 2a_1^2 1b_2^2 3a_1^1 1b_1^2 4a_1^1\rangle$ configuration is sufficient as can be seen from the differences between tables 1 and 2.5, 2.6. At a first glance this seems to show the superiority of BLYP-orbitals but actually proves the opposite. First of all as discussed above, the LHF-orbitals correctly reflect an avoided

crossing between the $2\ ^1A_1$ and the $3\ ^1A_1$ states (named $\tilde{B}\ ^2A_1$ and $\tilde{D}\ ^2A_1$ states in the discussion of Theodorakopoulos *et al.*^[82,83]). Additionally from table 2.3 it is obvious that the excitation energies of the corresponding triplet states $1\ ^3A_1$ and $2\ ^3A_1$ computed with LHF-orbitals already agree quite nicely with the experimental results if only two reference configurations are involved. Employing BLYP-orbitals in combination with a minimal reference space (table 2.5) the energy gap between both states is predicted too large. Only if a larger reference space is used the spacing is predicted correctly whereby the $2\ ^3A_1$ state is shifted to a lower energy and the $1\ ^3A_1$ state nearly remains unchanged. This behavior results, because as seen from figure 2.2 BLYP gives a 3s orbital which is very similar to the corresponding LHF or cationic orbital while its $3p_y$ orbital is too diffuse. As a consequence the $1\ ^3A_1$ (generated from the ground state by a $3a_1 \rightarrow 3s$ excitation) is already well described with the minimal reference space while the $2\ ^3A_1$ (generated from the ground state by a $3a_1 \rightarrow 3p$ excitation) demands a larger space (table 2.5 and 2.6). The unbalanced description of both orbitals is also the key to the missing configuration mixing of the $2\ ^1A_1$ which for BLYP-orbitals is suppressed since states involving the $3p_y$ orbitals are too high in energy relative to those involving the 3s orbital.

The reasons for the underestimation of excitation energies found for MRSD-CI+Q-calculations employing BLYP-orbitals are connected with the ability of BLYP-orbitals to describe the cation. This is indicated by the computations of the cation employing the different sets of orbitals which are summarized in table 2.7. Employing HF- or LHF-orbitals optimized for the ground state of the neutral water molecule ($X\ ^1A_1$) for the description of three cationic states (2B_1 , 2A_1 and 2B_2) the reference configurations corresponding to the three states possess weights of about 0.91 within the complete CI wavefunctions while the reference configuration possesses a weight of 0.94 if the orbitals are employed for the computation of the $X\ ^1A_1$ state itself. A similar weight (0.94) is found if BLYP-orbitals are used to describe the $X\ ^1A_1$ state, but if they are used to compute the cation the weights are only around 0.89, *i.e.*, considerably smaller than the weights obtained with HF- or LHF-orbitals. This shows that BLYP-orbitals optimized for the neutral system are considerably less suited for the description of the cation than the corresponding HF- or LHF-orbitals. The reason for this finding may lie in the well known wrong asymptotic behavior of the GGA potentials which possess an exponential instead of a $-1/r$ asymptotic decay.^[34] As a consequence beside

Tab. 2.7: c^2 -values of the corresponding reference configuration for the ground and the 2B_1 , 2A_1 and 2B_2 ionic states of water, using orbitals from a LHF, HF and BLYP ground state calculations. The ROHF-orbitals were obtained from the calculation of the 2B_1 state. For more information see text.

state	LHF	HF	BLYP	ROHF/ 2B_1
X 1A_1	0.94398	0.94370	0.94354	0.87984
2B_1	0.91111	0.91029	0.89565	0.95457
2A_1	0.91293	0.91262	0.89754	
2B_2	0.92188	0.92229	0.90719	

the virtual also the occupied BLYP-orbitals are slightly too diffuse. While this does not seem to hamper the description of the neutral system problems arise if the same orbitals are used for the description of the cation which possesses a more compact electronic structure.

As expected, since Rydberg states represent systems composed of a weakly interacting electron circling around a cationic core, similar small weights for the reference spaces are found for the excited states. Within the excited states the weights of the reference spaces are larger than 0.91 for the LHF-orbitals but only around 0.89 if BLYP-orbitals are involved. As seen from table 2.7 for the neutral groundstate both sets of orbitals lead to weights larger than 0.94. As a consequence of the strong difference in the weights found for the BLYP computations the Davidson correction overestimates the correction due to higher excitations and predicts energy values which are too low in comparison to the ground state. This also explains the nice agreement found if minimal reference spaces are used in BLYP computations. The overestimation of the influence of higher excitations by the Davidson correction compensates the errors arising due to an incomplete reference space. Due to the higher weights found for LHF-orbitals the Davidson correction works much better.

Differences between LHF- and BLYP-orbitals can also be seen from the values of $E_{\text{corr}}^{\text{rel,dyn}}$, which for BLYP-orbitals are about an 1 eV larger than for the LHF-orbitals. It is interesting to remember that the computations using HF-orbitals and those obtained with LHF-orbitals possess very similar $E_{\text{corr}}^{\text{rel,dyn}}$ values, although much larger reference spaces are needed. However, it could be that some effects which for the BLYP-orbitals are summarized in $E_{\text{corr}}^{\text{rel,dyn}}$ are already included in the static correlation of the HF-orbitals. The larger negative val-

Tab. 2.8: Computed excitation energies of the Rydberg states of water in eV. For the description of the excited states orbitals from a ROHF-calculation of the 2B_1 cationic state were used. Only minimal reference spaces were used. For more information see text or tables 1 and 2. For the groundstate: $E_{\text{det}} = E_{\text{Min}} = E_{\text{HF}} = -76.06398$ a.u.; $E_{\text{MRCI+Q}} = -76.32886$ a.u.. All energies in eV.

state	assignment	ΔE_{det}	ΔE_{ref}	$\Delta E_{\text{MRCI+Q}}$	ΔE_{exp}	$E_{\text{corr}}^{\text{rel,dyn}}$
$2\ {}^1A_1$	$1a_1 \rightarrow 3s/b_1 \rightarrow 3p_y$	(8.69)	8.76	9.78	9.7	-6.19
$1\ {}^1A_2$	$1b_1 \rightarrow 3p_x$	8.10	8.11	9.26	9.1	-6.06
$2\ {}^1A_2$	$1b_1 \rightarrow 3d_{xz}$	9.52	9.52	10.79		-5.93
$1\ {}^1B_2$	$1b_1 \rightarrow 3d_{xy}$	9.60	9.64	11.01		-5.84
$2\ {}^1B_2$	$1b_1 \rightarrow 4d_{xy}$	10.29	10.28	11.66	11.4	-5.83
$1\ {}^1B_1$	$1b_1 \rightarrow 3s$	6.55	6.55	7.48	~ 7.5	-6.27
$2\ {}^1B_1$	$1b_1 \rightarrow 3p_z$	8.61	8.61	9.91	~ 10.0	-5.91
$1\ {}^3A_1$	$3a_1 \rightarrow 3s$	8.88	8.88	9.42	9.3	-6.66
$2\ {}^3A_1$	$1b_1 \rightarrow 3p_y$	8.43	8.42	9.67		-5.95
$1\ {}^3A_2$	$1b_1 \rightarrow 3p_x$	8.00	8.00	9.10	~ 9.0	-6.10
$2\ {}^3A_2$	$1b_1 \rightarrow 3d_{xz}$	9.49	9.49	10.77		-5.92
$1\ {}^3B_2$	$1b_1 \rightarrow 3d_{xy}$	9.63	9.63	11.00		-5.84
$2\ {}^3B_2$	$3a_1 \rightarrow 3p_x$	10.28	10.28	11.66		-5.83
$1\ {}^3B_1$	$1b_1 \rightarrow 3s$	6.20	6.19	7.12	~ 7.1	-6.29
$2\ {}^3B_1$	$1b_1 \rightarrow 3p_z$	8.57	8.58	9.86		-5.93
2B_1	$1b_1 \rightarrow \infty$	11.12		12.51	12.62	-5.81

ues of $E_{\text{corr}}^{\text{rel,dyn}}$ for BLYP-orbitals could also result from the incorrect behavior found for the Davidson correction.

The quality of the three sets of orbitals is also nicely reflected in the shape of the virtual orbitals depicted in figure 2.2. While HF-orbitals are more or less structureless BLYP-orbitals at least mirror the shape of the Rydberg orbitals, but are too diffuse. Additionally the $4s$ orbital possesses a lower orbital energy than the $3p_z$ orbital, *i.e.*, the sequence of the Rydberg orbitals is not correctly described. As an exception the description of the $3s$ orbital, as already seen in the excitation energies, is much better. One reason could be its higher valence character. Our discussion underlined that LHF-orbitals are well suited for the description of the excited states but also showed that dynamical correlation and orbital relaxation effects ($E_{\text{corr}}^{\text{rel,dyn}}$ around -8.6 eV) are larger for the excited states than for the ground state ($E_{\text{corr}}^{\text{rel,dyn}}$

≈ -7.3 eV). The excited states possess two unpaired electrons, *i.e.* their dynamical correlation should be smaller than the dynamical correlation found for the ground state which possesses only doubly occupied orbitals. Consequently orbital relaxation effects are expected to be the reason for the larger $E_{\text{corr}}^{\text{rel,dyn}}$ values of the excited states.

As already discussed ROHF computations of the cation offer a way to estimate the orbital relaxation effects arising due to electronic excitations into Rydberg states, since the occupied orbitals already include the relaxation due to the positively charged core, and the virtual orbitals are optimized within an N-1 electron potential and thus do not “see” an electron too much as in an HF calculation of the neutral molecule. The ROHF calculations were performed for the cation obtained by taking one electron out of the $1b_1$ orbital (2B_1). The shapes of the lowest lying virtual orbitals of a_1 -symmetry can be taken from figure 2.2. The orbitals are nearly identical to the LHF-orbitals, but they seem to be slightly more diffuse. The computations performed with the cationic orbitals are summarized in table 2.8. The excitation energies are computed with respect to a ground state energy obtained from a SD-CI computation employing ground state HF-orbitals of the neutral molecule. This reference is used since as expected from the small weights discussed in combination with the data of table 5 the orbitals obtained for the cation are not suited to describe the ground state of the neutral molecule. Furthermore, a minimal (two configurations for two roots) reference space gives a sufficient description for all states, except the $2\ {}^1A_1$ state. Strong differences between table 2.3 and table 2.8 are found in all energy quantities except $\Delta E_{\text{MRCI+Q}}$. The assumption that the orbital relaxation effects are already accounted for if cationic orbitals are employed is supported by the much smaller $E_{\text{corr}}^{\text{rel,dyn}}$ values given in table 2.8. The consideration that the remaining dynamical correlation for the excited states is smaller than in the ground state is also supported since the $E_{\text{corr}}^{\text{rel,dyn}}$ values given in table 2.8 are smaller than the $E_{\text{corr}}^{\text{rel,dyn}}$ value obtained for the ground state which is around 7.2 eV if LHF- or HF-orbitals are employed. As already discussed, the energy difference between E_{det} for the cation computed with ground state orbitals and the ROHF energy for the cation gives the size of the orbital relaxation effects ($E_{\text{corr}}^{\text{rel}}$) arising due to the missing electron. For the 2B_1 state a value of -2.75 eV is computed. Subtracting this value from $E_{\text{corr}}^{\text{rel,dyn}}$ of the 2B_1 state given in table 2.4 gives an estimate of the size of the dynamical correlation ($E_{\text{corr}}^{\text{dyn}} = -5.99$ eV). The small

difference between this value and $E_{\text{corr}}^{\text{rel,dyn}}$ given in table 6 (-5.81 eV)⁵ is expected since both sets of orbitals are slightly different in the description of correlation. The comparison of $E_{\text{corr}}^{\text{rel}}$ and $E_{\text{corr}}^{\text{dyn}}$ shows that for the water cation orbital relaxation effects are about half as large as the effects arising due to dynamical correlation.

To compute the relaxation effects for the LHF- or BLYP-orbitals the corresponding restricted open shell computations had to be performed, which is not possible at the moment. For LHF, however, it can be exploited that the occupied LHF- and HF-orbitals are very similar.^[44] From E_{det} of the ${}^2\text{B}_1$ cationic state computed with LHF-orbitals (-75.55411 a.u.) and the ROHF energy (-75.65510 a.u.) an estimate of -2.75 eV for the relaxation energy for LHF-orbitals is obtained. The value is identical to the one obtained for HF-orbitals, which is in line with the above mentioned assumption. Subtracting the expectation values of the excited determinants constructed with ROHF- and LHF-orbitals, respectively, gives a rough estimate of the orbital relaxation effects for the Rydberg states. The computed values vary between -1.8 and -2.8 eV . The values follow the expected trend: The lower lying states possess smaller values while the higher lying states converge towards the value of -2.75 eV found for the ionized ${}^2\text{B}_1$ state. For BLYP-orbitals a similar estimation of the orbital relaxation effects is troublesome since the occupied BLYP- and HF-orbitals are different.

Especially the strong influence of the dynamical correlation explains why the studies performed by Veseth^[42] or Bourř^[28] found good agreement in some cases but bad one in others. The differences did not arise due to the description of the Rydberg orbitals but due to errors arising from the truncation of the CI expansion employed in both studies. Both approaches could not include dynamical correlation and orbital relaxation in a reliable manner. The deviations found for the carbonmonoxide molecule by Veseth^[42] are explained in more detail in section 2.5.

The question remains whether the various effects depend more on the nature of the orbital from which the excitation takes place or whether it more correlates with the character of the orbital into which the electron is excited. The study on the water molecule mainly gives information about the influence of the orbital into which the excitation takes place since mainly excitations out of the $1b_1$ orbital are computed. For the excitations out of the $1b_1$

⁵This value only gives the dynamical correlation since the orbitals were optimized for the cation, *i.e.*, orbital relaxation is already accounted for.

orbital table 2.3 shows that $E_{\text{corr}}^{\text{rel,dyn}}$ differ by about 0.4 eV depending on the character of Rydberg orbital. Table 2.3 also shows that $E_{\text{corr}}^{\text{rel,dyn}}$ values depends stronger on the nature of the orbital than on the question whether the singlet or the triplet state is computed. The influence of the orbital from which the excitation takes place ($1b_1$ and $3a_1$) seems to be less important as can be seen from table 2.3 but the nature of both orbitals (the $1b_1$ orbital represents a non-bonding orbital but also the $3a_1$ orbital possess some lone pair character) seems to be too similar for a definite answer. This is shown from the $E_{\text{corr}}^{\text{rel,dyn}}$ values computed for both ionization limits. They differ by only 0.28 eV if the 2B_1 and the 2A_1 states are compared. Stronger differences could be expected from excitations out of the $1b_2$ orbital but these excitations are too high in energy.

2.4 Discussion of ethene

To study the influence of the orbital out of which the excitation takes place the excitation energies for some selected triplet states of ethene⁶ have been computed which arise from excitations out of the π orbital and the highest lying σ orbital, respectively, of the previous test case, the ethene molecule. These results are summarized in table 2.9. This section will only focus on $E_{\text{corr}}^{\text{rel,dyn}}$ since the usefulness of the various sets of orbitals for the description of the Rydberg series of the ethene molecule was already discussed in section 2.1. The energetical data besides $E_{\text{corr}}^{\text{rel,dyn}}$ are only given for completeness. In table 2.9 the differences (≈ 1 eV) between $E_{\text{corr}}^{\text{rel,dyn}}$ computed for excitations out of the π or σ orbital, respectively, are obvious. With respect to the orbitals into which the excitation takes place a difference (≈ 0.5 eV) between excitations into the π^* (valence-valence excitations) and into the Rydberg orbitals

⁶Computational details for sections 2.4 and 2.12:

The computations for ethene in section 2.4 were performed at the experimental geometry by Serrano-Andrés *et al.* described above. We employed a (12s7p3d) basis set^[71] contracted to [8s7p3d] for carbon and a 6s3p1d un-contracted basis set for hydrogen. In section 2.1, diffuse gaussian functions were added to each carbon atom. However, for diffuse gaussian functions it is not necessary to add gaussian functions of the same exponent on each atom. Therefore, in this computation primitive gaussian functions were only added on the symmetry center of the molecule. The corresponding exponents are: 0.0192746 (s-type), 0.00770987 (s-type), 0.00308395 (s-type), 0.00123358 (s-type), 0.0342015 (p-type), 0.013680 (p-type), 0.00547224 (p-type), 0.218889 (p-type), 0.041354 (d-type), 0.016541 (d-type), 0.00661671 (d-type).

Tab. 2.9: Computed vertical excitation energies of selected triplet states of ethene using LHF ground state orbitals. States representing excitations from the HOMO (π) and the HOMO-1 (σ) into the same virtual orbital are compared. For the groundstate: $E_{\text{det}} = -78.05440$ a.u., $E_{\text{ref}} = -78.07390517$ a.u., $E_{\text{MRCI+Q}} = -78.42283$ a.u. All energies in eV.

state	assignment	ΔE_{orb}	ΔE_{det}	$\Delta E_{\text{ref}}^{\text{[a]}}$	$\Delta E_{\text{MRCI+Q}}$	$\Delta E_{\text{exp.}}$	$E_{\text{corr}}^{\text{rel,dyn}}$
1 $^3B_{1u}$	$\pi \rightarrow \pi^*$	5.94	3.78	4.29	4.48	4.36 ^[b]	-9.30
2 $^3B_{1g}$	$\sigma \rightarrow \pi^*$	7.97	8.87	9.40	8.48		-10.42
1 $^3B_{3u}$	$\pi \rightarrow 3s$	7.30	6.88	7.41	7.09	6.98 ^[c]	-9.82
1 $^3B_{3g}$	$\sigma \rightarrow 3s$	9.33	10.70	11.18	9.79		-10.88
2 $^3B_{3u}$	$\pi \rightarrow d\sigma$	8.59	8.39	8.92	8.61	8.57 ^[c]	-9.81
2 $^3B_{3g}$	$\sigma \rightarrow d\sigma$	10.62	12.11	12.69	11.41		-10.78
$^2B_{3u}$	$\pi \rightarrow \infty$	10.18	10.18	10.71	10.46	10.51 ^[d]	-9.76
$^2B_{3g}$	$\sigma \rightarrow \infty$	12.20	13.91	14.44	13.31	12.79 ^[e]	-10.62

[a] The difference of ΔE_{ref} to ΔE_{det} is mainly due to the $\pi^2 \rightarrow \pi^{*2}$ contribution to the ground state which leads to an energy lowering of 0.53 eV, [b] van Veen,^[85] [c] Wilden and Corner,^[86] [d] Lide^[79] and [e] Bieri and Asbrink.^[87]

is found, while the difference of $E_{\text{corr}}^{\text{rel,dyn}}$ for excitations into different Rydberg orbitals is very small. The $E_{\text{corr}}^{\text{rel,dyn}}$ values obtained for the Rydberg states approach the values obtained for the corresponding ionization limits, while $E_{\text{corr}}^{\text{rel,dyn}}$ computed for the valence states are absolutely larger. A strong difference between the computed and the measured $^2B_{3g}$ ionization limit is found, but a CCSD(T) computation employing a TZ2P basis set predicts the same theoretical result.^[84] Therefore it has to be expected that the difference between experiment and theory results since the $^2B_{3g}$ state possess a different geometrical structure than the ground state of the neutral molecule. Taking this into account shifts the computed ionization limit to 12.8 eV. However, since always vertical excitations are discussed, the vertical ionization limit will be used for the remaining discussions. To estimate the orbital relaxation effects the calculation has been repeated employing orbitals obtained from an ROHF calculation of the corresponding ionized states $^2B_{3u}$ and $^2B_{3g}$, respectively. These results are summarized in table 2.10. While orbitals from a ROHF calculation of the $^2B_{3u}$ state are taken for excitations out of the π orbitals, orbitals from a ROHF calculation of the $^2B_{3g}$ state are taken for the σ excitations. Table 2.11 lists the computations of the neutral ground

Tab. 2.10: Computed vertical excitation energies of selected triplet states of ethene using ROHF-orbitals. For states involving excitations out of the π orbital, orbitals optimized for the cationic ${}^2B_{3u}$ state have been employed, while orbitals optimized for the cationic ${}^2B_{3g}$ state were used for states arising due to excitations out of the σ orbital. HF-orbitals were used to calculate the groundstate. For the groundstate: $E_{\text{det}} = E_{\text{ref}} = E_{\text{HF}} = -78.06206317$ a.u., $E_{\text{MRCI+Q}} = -78.42207$ a.u.. All energies in eV.

state	assignment	ΔE_{det}	ΔE_{ref}	$\Delta E_{\text{MRCI+Q}}$	$\Delta E_{\text{exp.}}$	$E_{\text{corr}}^{\text{rel,dyn}}$
1 ${}^3B_{1u}$	$\pi \rightarrow \pi^*$	4.52	4.49	4.43	4.36	-9.86
2 ${}^3B_{1g}$	$\sigma \rightarrow \pi^*$	9.66	9.55	8.43		-10.82
1 ${}^3B_{3u}$	$\pi \rightarrow 3s$	5.88	5.87	7.06	6.98	-8.60
1 ${}^3B_{3g}$	$\sigma \rightarrow 3s$	10.21	10.21	9.77		-10.23
2 ${}^3B_{3u}$	$\pi \rightarrow d\sigma$	7.24	7.24	8.57	8.56	-8.47
2 ${}^3B_{3g}$	$\sigma \rightarrow d\sigma$	11.54	11.54	11.30		-10.04
${}^2B_{3u}$	$\pi \rightarrow \infty$	8.94		10.47	10.51	-8.26
${}^2B_{3g}$	$\sigma \rightarrow \infty$	13.33		13.31	12.79	-9.81

state and of the two ionization limits employing various sets of orbitals. HF labels the results obtained with HF-orbitals optimized for the ground state of the neutral system while ROHF gives the results obtained with ${}^2B_{3u}$ or ${}^2B_{3g}$ ROHF-orbitals, respectively. The columns entitled LHF and BLYP summarize the values obtained with LHF- or BLYP-orbitals optimized for the neutral ground state while the rows entitled with c^2 give the weights of the most dominant configuration of the CI wavefunctions. ΔE_{ion} gives the computed ionization limits (MRSD-CI+Q). The results are obtained from SD-CI computations. Only for the ground state of the neutral system computations involving LHF- or BLYP-orbitals also MRSD-CI computations have been performed. Their reference spaces include the ground state configuration and the $\pi^2 \rightarrow \pi^{*2}$ double excitation. The results are given in two additional rows for the groundstate. The $E_{\text{corr}}^{\text{rel,dyn}}$ values in parenthesis also include the static correlation of this computation which is normally not contained in $E_{\text{corr}}^{\text{rel,dyn}}$. The picture painted by table 2.11 is very similar to that found for the water molecule (table 2.7). On the basis of the weights of the most dominant configuration the quality of HF-, LHF- and BLYP-orbitals is very similar for the description of the groundstate of ethene but differences appear if these orbitals are used for the description of the ${}^2B_{3u}$ state. As already found for the water molecule BLYP-

Tab. 2.11: Comparison of c^2 -values, correlation energies and the first and second ionization potential of ethene using groundstate LHF-, HF- and BLYP-orbitals. In the last column results obtained with ROHF-orbitals optimized for the two ionic states are listed individually. The ground state data in this column were calculated using the orbitals of the ${}^2B_{3u}$ state. For more information see text. All energies in eV.

state		LHF	HF	BLYP	ROHF
$X {}^1A_1$	c^2	0.89908	0.89888	0.89837	0.87776
	$E_{\text{corr}}^{\text{rel,dyn}}$	-9.94	-9.80	-10.28	-11.31
	c^2	0.90574 ¹		0.90458 ^[a]	
	$E_{\text{corr}}^{\text{rel,dyn}}$	-9.49 (-10.03 ^[b])		-9.81 (-10.38 ²)	
${}^2B_{3u}$	c^2	0.89146	0.89139	0.88897	0.91669
	$E_{\text{corr}}^{\text{rel,dyn}}$	-9.74	-9.60	-9.95	-8.27
	E_{ion}	10.46	10.40	10.44	10.47
${}^2B_{3g}$	c^2	0.87622	0.87347	0.87624	0.89082
	$E_{\text{corr}}^{\text{rel,dyn}}$	-10.63	-10.62	-10.80	-9.81
	E_{ion}	13.31	13.21	13.39	13.31

[a] The reference space of the ground state also includes the $\pi^2 \rightarrow \pi^{*2}$ -doubly excited reference configuration (see text).

[b] Total electron correlation contribution, including dynamical and static correlation. For more information see text.

orbitals seem to be less appropriate although the differences are smaller for ethene than for water. Employing ground state orbitals of the neutral system for the description of the ionic ${}^2B_{3g}$ state the weights of the leading configuration generally decrease further but this time BLYP-orbitals possess the same weight as LHF-orbitals. The orbital relaxation effect for the ${}^2B_{3u}$ state is only about -1.3 eV while an even smaller value of -0.7 eV is estimated for the ${}^2B_{3u}$ state being both considerably smaller than the orbital relaxation effects estimated for the water molecule which were around -3 eV. Subtracting the orbital relaxation effect from $E_{\text{corr}}^{\text{rel,dyn}}$ estimates for the dynamical correlation contribution can be obtained. For excitations out of the π orbital values of about 8.5 eV for the Rydberg states are obtained, while a value of about 10 eV is found for excitations out of the σ orbital. This shows that the composition of $E_{\text{corr}}^{\text{rel,dyn}}$ for water and ethene is quite different since orbital relaxation effects are smaller and dynamical correlation is larger. While for water orbital relaxation effects

give a third of the $E_{\text{corr}}^{\text{rel,dyn}}$ they contribute only about a tenth for ethene. Finally, table 7 supports that deficiencies in the description of orbital relaxation and electron correlation are the underlying reasons for the findings of Veseth and Bouř, who found that that Kohn Sham orbitals as a basis of subsequent CI calculations sometimes give good but sometimes also bad agreement with experimental results. For states arising due to excitations out of the π orbitals ($1\ ^3B_{1u}$, $2\ ^3B_{3u}$ and $2\ ^3B_{3u}$) already ΔE_{det} or ΔE_{ref} give good approximations for the excitation energies with deviations from experiment of only about 0.2 - 0.4 eV. This behaviour results since for the excited states orbital relaxation effects are nearly canceled by the decrease of the dynamical correlation effects with the consequence that the $E_{\text{corr}}^{\text{rel,dyn}}$ values of ground and excited states differ by only 0.4 eV. However this only holds for these states. For the remaining states arising due to excitations from the σ orbitals the error cancellation breaks down leading to stronger errors in ΔE_{det} and ΔE_{ref} .

2.5 MRCI-computations for carbonmonoxide

In section 2.3, the computed excitation energies of the water molecule that were obtained using BLYP-orbitals did not converge towards the experimentally determined values when the reference space were extended, but to lower values. This poses a severe problem for the usage of the computational effort that is needed to reach a certain accuracy as a quality criterion, because this case shows that in some cases due to error cancellation a smaller computational effort leads to a more accurate excitation energy. As a consequence, the quality criterion needs to be replaced by a criterion that is less susceptible to error cancellation. In this section, the number of reference configurations that is needed to reach convergence in the excitation energies from MRCI+Q computations is used as a quality criterion.

The results are summarized in table 2.5⁷, which gives the excitation energies for the lowest

⁷Computational details for section 2.5, 2.8 and 2.9:

The calculations for carbonmonoxide were performed at the experimental geometry with $R_{\text{CO}} = 1.1283\ \text{\AA}$. The molecule is chosen with the z-axis being collinear to the CO-bond. Carbonmonoxide was calculated in the C_{2v} point group. A (12s7p3d) basis set^[71] contracted to [8s7p3d] for oxygen and carbon was employed. Primitive diffuse Gaussian functions with the exponents 0.039421656 (s-type), 0.0121380 (s-type), 0.0075 (s-type), 0.0042482 (s-type), 0.035011523 (p-type), 0.008015, (p-type) 0.0028052 (p-type), 0.0285120 (d-type),

two ${}^1\Pi$ -states of carbonmonoxide as a function of the number of reference configurations. The reference spaces for the excited states are constructed from single excitations out of the HOMO- into virtual π -orbitals. The configurations are successively added to the reference space according to their contribution (c^2 -value) to a Singles-CI wavefunction. For all orbital types, the reference space of the ground state consists of the HF or KS determinant Ψ and the $\pi_x^2 \rightarrow \pi_x^{*2}$, $\pi_y^2 \rightarrow \pi_y^{*2}$, $\pi_x^2 \rightarrow \pi_y^{*2}$ and the $\pi_y^2 \rightarrow \pi_x^{*2}$ doubly excited configurations (see section 2.8).

When the A ${}^1\Pi$ -state is computed with HF-orbitals, a computation using only one reference

Tab. 2.12: Vertical excitation energies from MRCI+Q-computations for the A ${}^1\Pi$ - and the E ${}^1\Pi$ -state of carbonmonoxide using HF-, LHF-, BLYP- and B3LYP-orbitals and increasing numbers ($n_{Refconf}$) of reference configurations. In addition, the weight (c^2) of the reference configurations in the MRCI-wavefunction with $n_{Refconf} = 8$ is given. Ground state energies: HF: -113.13789 H; LHF -113.13767 H; BLYP: 113.13428 H; B3LYP: -113.13534 H; $E_{exp}^{[102]}$ A ${}^1\Pi$: 8.51 eV; E ${}^1\Pi$: 11.53 eV. Excitation energies in eV.

		$n_{Refconf}$								c^2 for $n = 8$
		1	2	3	4	5	6	7	8	
HF	A ${}^1\Pi$	8.71	8.43	8.43	8.41	8.41	8.40	8.39	8.39	0.876
	E ${}^1\Pi$		11.38	11.32	11.18	11.14	11.14	11.14	11.14	0.865
LHF	A ${}^1\Pi$	8.55	8.52	8.52	8.51	8.50	8.51	8.50	8.50	0.881
	E ${}^1\Pi$		11.26	11.29	11.26	11.26	11.26	11.26	11.26	0.871
BLYP	A ${}^1\Pi$	8.64	8.63	8.63	8.61	8.60	8.60	8.60	8.59	0.890
	E ${}^1\Pi$		11.70	11.52	11.42	11.41	11.40	11.39	11.36	0.880
B3LYP	A ${}^1\Pi$	8.61	8.60	8.58	8.58	8.58	8.58	8.58	8.58	0.880
	E ${}^1\Pi$		11.79	11.65	11.57	11.35	11.35	11.35	11.35	0.879

configuration for the excited state gives an energy of 8.61 eV. When two reference configurations are used, this value is significantly lowered to 8.43 eV. Both values are quite close to the experimental value of 8.51 eV. A further increase of the reference space does not lead to a significant change in the computed excitation energies, *e.g.* a computation involving eight reference configurations leads to an energy of 8.39 eV. For the E ${}^1\Pi$ -state convergence to a value of 11.14 eV is reached after 5 configurations (one of these configurations is used to

0.0042 (d-type), 0.002 (d-type) were added on the carbon atom.

describe the A $^1\Pi$ -state which represents the first root in the CI-computation). The experimental value for the E $^1\Pi$ -state is 11.53 eV.

LHF-orbitals lead to a much faster convergence in the computed excitation energies. For the A $^1\Pi$ -state a computation with only one reference configuration predicts an excitation energy of 8.55 eV, while a computation with eight reference configurations gives 8.50 eV. For the computation of the E $^1\Pi$ -state two configurations lead to an excitation energy of 11.26 eV. A further enlargement of the reference space does not change this value.

Using BLYP-orbitals, the excitation energies computed for the A $^1\Pi$ -state is lowered very little when the reference space is increased. The excitation energy for the E $^1\Pi$ state required 4 configurations to converge around an excitation energy of 11.4 eV. B3LYP-, like BLYP-orbitals, allow a very fast convergence for the A $^1\Pi$ -state, but the convergence for the E $^1\Pi$ -state is very slow, the converged excitation energy of 11.35 eV is only reached with more than 6 configurations.

LHF-, BLYP- and B3LYP-orbitals give a very compact description of the A $^1\Pi$ -state, also the description based on HF-orbitals only required two configurations. For the E $^1\Pi$ -state, only LHF-orbitals allow a very compact description, for all other orbitals tested more than 4 configurations are needed to reach constant energies. However, the converged energies for the E $^1\Pi$ -state are not always in good agreement with the experimentally determined excitation energy 11.53 eV. The energies computed for the E $^1\Pi$ -state with LHF-orbitals are 0.3 eV below the experimentally determined value. For HF-orbitals the converged energy underestimates the experimental value even by 0.4 eV, the results obtained with BLYP- (11.39 eV) and B3LYP-orbitals (11.35 eV) are in a somewhat better agreement with the experimental data.

In the case of HF-orbitals the underestimation is partly due to the poor description of static correlation, even though configurations representing $\pi^2 \rightarrow \pi^{*2}$ excitations were included in the reference wavefunction. However, the diffuseness of HF-orbitals does not allow a good description of static correlation effects. The major part of the error, however, is due to a poor description of the cationic core of the excited states by ground state orbitals obtained from uncorrelated methods like HF and LHF. This is proven by the weights of the reference wavefunction (c^2 -values) for the computations with $n_{ref}=8$ (table 1). The error in the corresponding excitation energies directly correlates with the smaller c^2 -values. In addition to the

c^2 -values in table 2.12, table 2 gives the weights for the ground state of CO ($c^2(\Psi)$), and the $^2\Sigma^+$ state $c^2(^2\Sigma^+)$ and the $^2\Pi$ state of CO^+ . The c^2 -values obtained for the E $^1\Pi$ -states in table 2.12 are identical to the c^2 -values that are obtained if the ground state orbitals are employed for the computation of the $^2\Sigma^+$ ground state state of CO^+ in table 2. This correlation shows that the low c^2 -values obtained for the 2 $^1\Pi$ -state result from the description of the cationic core and are completely independent of the quality of the Rydberg orbitals. Since the Davidson procedure only gives a correct estimate of the influence of higher excitations, when c^2 -values of ground and excited states do not differ too much, the small c^2 -values for the excited as well as for the ionized state lead to a breakdown of the Davidson correction and subsequently to an underestimation of the corresponding excitation energies. A similar finding was observed in the computation of the excited states of water using BLYP-orbitals in section 2.3.

Table 2 also contains the c^2 -values for the $^2\Pi$ cationic state of carbonmonoxide and for the ground and ionized states of formaldehyde and acetone. For HF- and LHF-orbitals the differences between the c^2 -values of ground and first ionized states in these cases are always smaller than those found in the case of the ground and the $^2\Sigma^+$ -state of carbonmonoxide. Consequently, since no overestimation problems in the Davidson correction occur for these cases, the computed excitation energies always converge towards the experimentally determined values. Furthermore, for the $^2\Pi$ -state of carbonmonoxide and the 2B_2 -states of formaldehyde and acetone, HF- and LHF-orbitals tend to give better c^2 -values than for example BLYP-orbitals, like it was described in section 2.3. This behavior should be due to the

Tab. 2.13: Weights (c^2) of the reference configurations in the MRCI-wavefunction of ground ($c^2(\Psi)$) and monocationic states of carbonmonoxide ($c^2(^2\Sigma^+)$, $c^2(^2\Pi)$), formaldehyde and acetone ($c^2(^2B_2)$). The orbitals used were obtained from HF- or Kohn-Sham ground state computations.

	CO			CH ₂ O		C(CH ₃) ₂ O	
	$c^2(\Psi)$	$c^2(^2\Sigma^+)$	$c^2(^2\Pi)$	$c^2(\Psi)$	$c^2(^2B_2)$	$c^2(\Psi)$	$c^2(^2B_2)$
HF	0.915	0.865	0.887	0.908	0.878	0.862	0.840
LHF	0.923	0.871	0.885	0.914	0.883	0.863	0.841
BLYP	0.922	0.879	0.877	0.914	0.873	0.862	0.833
B3LYP	0.922	0.879	0.881	0.915	0.877	0.864	0.838

well known wrong asymptotic behavior of GGA potentials that fall off exponentially instead of $-1/r$.^[34] As expected, the c^2 -values for B3LYP-orbitals are always located in between the values obtained with HF- and BLYP-orbitals. The full vertical excitation spectrum of car-

Tab. 2.14: Excitation energies of carbonmonoxide and sizes of reference spaces (in brackets) in the corresponding MRCI calculation. The number of states determined in one calculation is denoted as n_{states} . Ground state energies: see table 2.12.

state	class.	n_{states}	HF	LHF	BLYP	B3LYP	exp. ^[a]
A $^1\Pi$	$n \rightarrow \pi^*$	1	8.43 (2)	8.55 (1)	8.64 (1)	8.61 (1)	8.51
I $^1\Sigma^-$	$\pi \rightarrow \pi^*$	1	9.85 (6)	10.00 (2)	9.89 (2)	10.00 (2)	9.88
D $^1\Delta^-$	$\pi \rightarrow \pi^*$	2	9.93 (7)	10.08 (3)	9.93 (3)	9.97 (3)	10.23
		2		10.16 (5)	10.15 (5)	10.12 (5)	
B $^1\Sigma^+$	$n \rightarrow 3s$	3	10.46 (5)	10.57 (3)	10.67 (5)	10.59 (6)	10.78
C $^1\Sigma^+$	$n \rightarrow 3p_z$	3	10.97 (11)	11.11 (3)	11.25 (6)	11.21 (7)	11.4
E $^1\Pi$	$n \rightarrow 3p$	2	11.14 (5)	11.26 (2)	11.42 (4)	11.36 (5)	11.53
$^2\Sigma^+$	$n \rightarrow \infty$	1	13.54 (1)	13.67 (1)	13.80 (1)	13.80 (1)	14.01 ^[b]
$^2\Pi$	$\pi \rightarrow \infty$	1	16.77 (1)	16.81 (1)	16.64 (1)	16.76 (1)	16.91 ^[c]

[a] All excitation energies from Nielsen *et al.*,^[88] [b] Lide^[79] and [c] Chong *et al.*^[89]

bonmonoxide is shown in table 2.14. Like in the case of the E $^1\Pi$ -state, a large number of up to 7 reference configurations is needed to compute each state when the MRCI-calculations are based on HF-orbitals. All excitation energies are found 0 to 0.5 eV below the experimentally determined values. Also the first and second ionization potentials are given. The first ionization potential computed to 13.54 eV with HF-orbitals is strongly underestimated by 0.47 eV, while for the second ionization potential a value of 16.77 eV is found which is in a somewhat better agreement with the experimental value of 16.91 eV.

Like for the $^1\Pi$ -states discussed above, LHF-orbitals give a very compact description of all excited states of carbonmonoxide. For the A $^1\Pi$ -, B $^1\Sigma^+$ - and the C $^1\Sigma^+$ -state, only one or two reference configurations per root are needed to obtain converged energies. Compared to the experimental values of 10.78 and 11.4 eV, the computed excitation energies of 10.57 and 11.11 eV for the B $^1\Sigma^+$ - and C $^1\Sigma^+$ -states are shifted to lower energies, like it was already explained for the E $^1\Pi$ -state. The I $^1\Sigma^-$ - and the D $^1\Delta^-$ -states represent linear combinations of $\pi \rightarrow \pi^*$ -excited configurations. For the I $^1\Sigma^-$ -state an energy of 10.00 eV

was computed with two reference configurations, for the $D^1\Delta^-$ -state and energy of 10.08 eV is computed using 3 reference configurations. This value is brought closer to the experimental value of 10.23 eV if the $\pi_x^2 \rightarrow \pi_x^{*2}$ and $\pi_y^2 \rightarrow \pi_y^{*2}$ doubly excited configurations are also considered.

For valence states, the CI-description using BLYP- or B3LYP-orbitals is of a comparably good quality as for LHF-orbitals. For Rydberg states, the description based on BLYP- or B3LYP-orbitals requires many more reference configurations. As it was shown for the $E^1\Pi$ -state, Rydberg states need to be constructed from 4-6 reference configurations in the case of BLYP-orbitals, and 6-7 reference configurations when B3LYP-orbitals are used while for LHF-orbitals 2-3 reference configurations suffice.

2.6 MRCI-computations for formaldehyde

Table 2.15⁸ shows that as in the case of carbonmonoxide, a very large number of reference configurations is needed to describe the excited states of formaldehyde when HF-orbitals are used. 4-8 reference configurations are needed to describe each excited state.

LHF-orbitals allow a distinctly better description of the excited states of formaldehyde than HF-orbitals. The number of reference configurations in table 2.15 is always identical to the number of states determined in a computation.

BLYP-orbitals give a very good description of the valence (1^1A_2) and the 3s Rydberg (1^1B_2) states. For the 3p-Rydberg states, the description is poorer. In this case 3-4 reference configurations are needed to describe each state. If the states are computed with the minimal

⁸Computational details for sections 2.15, 2.8 and 2.10:

The calculations for formaldehyde were performed at the experimental geometry with $R_{CO} = 1.1283 \text{ \AA}$, $R_{CH} = 1.1283 \text{ \AA}$ and $\angle_{OCH} = 116.5^\circ$.^[79] The molecule is chosen to lie in the yz plane with the z -axis being collinear to the CO-bond.

(12s7p3d) basis sets^[71] contracted to [8s7p3d] for oxygen and carbon and a 6s3p1d un-contracted basis set for hydrogen was used for formaldehyde. On the carbon atom, diffuse Gaussian functions with the exponents 0.039421656 (s-type), 0.01213800 (s-type), 0.0075000 (s-type), 0.004248200 (s-type), 0.0342015 (p-type), 0.008015000 (p-type), 0.002805200 (p-type), 0.028512 (d-type), 0.0042 (d-type), 0.002 (d-type) were added. The reference space was selected as described for carbonmonoxide using for excited states only single excitations out of the HOMO into the virtual orbitals.

set of reference configurations (1 configuration per CI root), the computed excitation energies show large deviations from experiment of about 0.6 eV for the 3p Rydberg-states. The ionization potentials computed with BLYP-orbitals (10.63 eV) is lower than that computed with LHF-orbitals (10.73 eV; exp. 10.88 eV). This corresponds to the smaller c^2 -values found in the computation of the cationic states given in table 2.13 and demonstrates the lower ability of BLYP-orbitals to describe the cationic core of the Rydberg states of formaldehyde.

The tendencies found for BLYP-orbitals are even worse for B3LYP-orbitals. While the valence 1^1A_2 -state is very well described, especially the 3p-Rydberg states are very problematic. An excitation energy of 9.30 eV compared to an experimental value of 7.97 eV is obtained for the 2^1B_2 -state if only the two most contributing main configurations are used as reference space. Only if the number of reference configurations is increased to 7, an energy of 7.91 eV is obtained. The ionization potential is computed as 10.67 eV.

Tab. 2.15: Excitation energies of formaldehyde from MRCI+Q-computations. Groundstate energies: HF: -114.31411 H; LHF: 114.31309 H; BLYP: -114.30941 H; B3LYP: -114.31092 H. For details see table 2.14.

state	class	HF	LHF	BLYP	BLYP	B3LYP	B3LYP	exp. ^[1]
1^1A_2	$n \rightarrow \pi^*$	3.80 (4)	3.91 (1)	3.82 (1)	3.82 (1)	3.86 (1)	3.86 (1)	3.94
1^1B_2	$n \rightarrow 3s$	7.14 (6)	7.15 (1)	7.05 (1)	7.05 (1)	7.22 (1)	7.18 (4)	7.09
2^1B_2	$n \rightarrow 3p_z$	7.94 (8)	8.03 (2)	8.61 (2)	8.02 (4)	9.30 (2)	7.91 (7)	7.97
2^1A_1	$n \rightarrow 3p_y$	8.06 (6)	8.10 (2)	8.63 (2)	8.09 (3)	9.00 (2)	8.19 (5)	8.12
2^1A_2	$n \rightarrow 3p_x$	8.24 (6)	8.30 (2)	9.03 (2)	8.39 (3)	9.34 (2)	8.28 (6)	8.38
2B_2	$n \rightarrow \infty$	10.66 (1)	10.73 (1)	10.63 (1)		10.67 (1)		10.88 ^[2]

[a] All data from Allen and Tozer,^[90] for more recent experimental data see also Liu *et al.*^[91] [b] Lide.^[79]

2.7 MRCI-computations for acetone

In all cases studied so far, LHF potential yielded better virtual orbitals than the BLYP method. For the acetone molecule, this is different. Table 2.16⁹ shows that the only state

⁹Computational details for sections 2.7 and 2.11:

The geometry for acetone was optimized with the B-P86 functional in the TZVP basis set using analytical

that is well described by LHF-orbitals is the 1^1A_2 -valence state. For all Rydberg states, the excitation energies are overestimated by about 0.2 eV when limited reference spaces were used. In addition, the energies converge rather slowly towards the experimental values when the reference spaces are enlarged. This leads to quite large reference spaces with up to 8 reference configuration for the 1^1B_2 -states.

HF-orbitals again require even larger reference spaces, but astonishingly, BLYP-orbitals

Tab. 2.16: Excitation energies of acetone from MRCI+Q-computations. Ground state energies: HF -192.69670 H; LHF: -192.69545 H; BLYP: -192.41356 H; B3LYP: -192.69338 H. For details see table 2.14.

state	class.	HF	LHF	LHF	BLYP	BLYP	B3LYP	B3LYP	exp. ^[a]
1^1A_2	$n_y \rightarrow \pi^*$	4.39 (8)	4.45 (1)	4.45 (1)	4.40 (1)	4.27 (4)	4.49 (1)	4.30 (5)	4.38
1^1B_2	$n_y \rightarrow 3s$	6.56 (7)	6.64 (1)	6.55 (4)	6.49 (1)	6.47 (3)	6.60 (1)	6.51 (5)	6.35
2^1A_2	$n_y \rightarrow 3p_x$	7.29 (8)	7.56 (2)	7.36 (4)	7.48 (2)	7.31 (4)	7.70 (2)	7.26 (5)	7.36
2^1A_1	$n_y \rightarrow 3p_y$	7.57 (4)	7.68 (2)	7.53 (3)	7.44 (2)	7.44 (2)	7.60 (2)	7.45 (4)	7.41
2^1B_2	$n_y \rightarrow 3p_z$	7.51 (7)	7.76 (2)	7.41 (8)	7.75 (2)	7.39 (5)	7.84 (2)	7.48 (5)	7.45
2^1B_2	$n_y \rightarrow \infty$	9.57 (1)	9.59 (1)		9.53 (1)		9.58 (1)		9.8 ^[b]

[a] All data from ter Steege *et al.*^[93] and Merchan *et al.*^[94] and [b] Chong.^[23]

give an acceptable description off all computed states except the 2^1B_2 -state. For the other states, an accurate excitation energy is already obtained with the limited reference space. The error for the 2^1B_2 -state at this level is about 0.3 eV when the size of the reference wavefunction is limited to only two reference configurations. The Rydberg series of acetone begins at a lower energy than the Rydberg series of formaldehyde, and the lower BLYP-Rydberg-orbitals of acetone do not seem to be too much affected by the wrong fall off behavior of the Kohn-Sham potential. As is was described above, B3LYP-orbitals are of lower quality than BLYP-orbitals, which can be seen from the excitation energies determined with the restricted reference spaces in row 9 of table 2.16 that deviate by 0.1 to 0.4 eV from experiment and the gradients as implemented in the TURBOMOLE program package.^[64] The molecule is chosen to have the carbon and oxygen atoms in the yz plane with the z-axis being collinear to the CO-bond. For acetone, the TZVP basis set^[92] was used. It represents an (11s6p1d) basis set contracted to [5s3p1d]. The TZVP basis set was extended by 0.019274694 (s-type), 0.0121380 (s-type), 0.00770987770 (s-type), 0.00308395111 (s-type), 0.035011523 (p-type), 0.01368060 (p-type), 0.00547224 (p-type), 0.028512 (d-type), 0.04135446090 (d-type), 0.04135446090 (d-type) primitive Gaussian functions which were placed on the carbonylic carbon atom.

larger reference spaces of 4-5 configurations that are needed to reach converged excitation energies.

For the ionization potentials, all types of orbitals leads to very similar values between 9.59 eV (LHF) to 9.53 eV (BLYP). The experimental ionization potential is 9.8 eV. For the lower

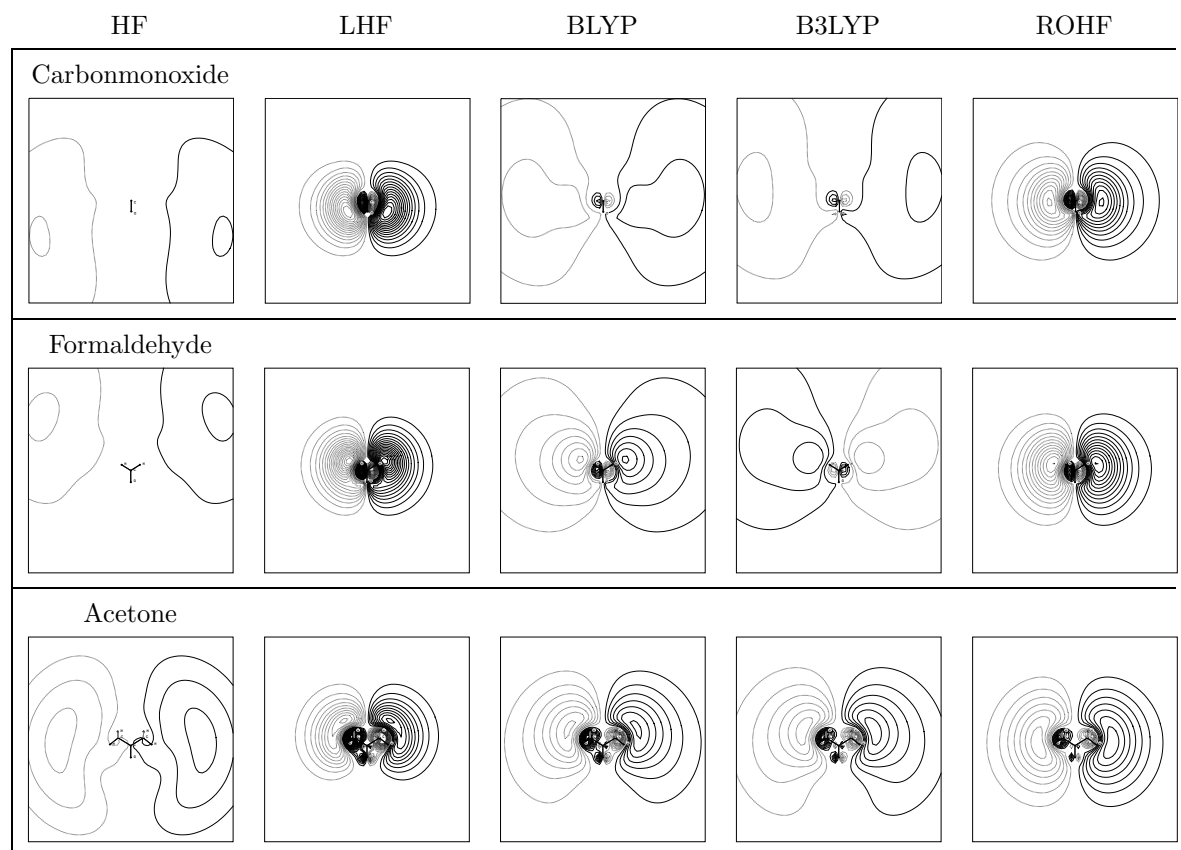


Fig. 2.5: The lowest virtual orbital of b_2 symmetry ($3p_x$). Frame length: 40 a.u., contour spacing: 0.03 a.u.

Rydberg states of the acetone molecule, LHF-orbitals give a rather poor description, while BLYP-orbitals tend to describe at least the energetically lowest Rydberg states much better. This result can be better understood from the spatial structure of the orbitals involved. In figure 2.5, the $3p_x$ orbitals of carbonmonoxide, formaldehyde and acetone are compared. The $3p_x$ orbitals from an ROHF calculation of the corresponding 2B_2 -cation shall serve as reference since MRCI computations show that they provide a very good description of the Rydberg $2{}^1A_1$ state. As reference also orbitals explicitly optimized for the $2{}^1A_1$, *e.g.* from a CASSCF calculation could have been taken. However, the ROHF orbitals have been pre-

ferred since they are optimized on a similar level of sophistication as HF- or LHF-orbitals. For carbonmonoxide and acetone, the HF-orbitals shown are highly diffuse and lack any structure while the LHF-orbitals closely resemble the orbitals of the 2B_2 cation. The only difference is that they appear to be slightly more compact than the latter. This corresponds to the results from the MRCI calculation of the corresponding states, which both are well described by LHF-orbitals. The $3p_x$ BLYP orbital is very diffuse for carbonmonoxide, and somewhat less diffuse in the case of formaldehyde. B3LYP-orbitals represent a mixture of BLYP- and HF-orbitals, which makes them even more diffuse than BLYP-orbitals. As a consequence, a very large number of reference configurations was needed in the MRCI calculation to construct the corresponding $2\ ^1A_1$ state. Going from carbonmonoxide to formaldehyde to acetone, the lowest virtual orbital of b_2 symmetry from the ROHF computation becomes somewhat more diffuse. In contrast, the corresponding HF- and all KS-orbitals studied tend to be more compact for acetone than for carbonmonoxide or formaldehyde. While the HF-orbital still is far too diffuse to give an accurate description of the $2\ ^1A_1$ state, the LHF-orbital in this case is too compact, while the BLYP-orbitals in the case of acetone give an acceptable description of the corresponding Rydberg state.

2.8 MRPT2 for the ground state

While KS-orbitals seem to be well suited for MRCI computations, their usage in perturbation theories based on only one reference configuration usually leads to very large 2nd order corrections and even to a divergence in the perturbational series.^[?,29,32] To what extent this overestimation can be reduced in a multireference approach is tested in table 2.17¹⁰ which summarizes the results for the ground state of carbonmonoxide. It gives reference energies (E_{ref}) and the correlation contribution resulting from an enlargement of the reference space ($E_{corr/ref}$) and the CI ($E_{corr}(SDCI)$) or the PT2 ($E_{corr}(PT2)$) treatment on top of the reference spaces. $E_{corr/ref}$ is obtained by subtracting E_{ref} of the multi- from E_{ref} of the single reference case. The total correlation energies of the treatment can be obtained by adding $E_{corr/ref}$ to $E_{corr}(SDCI)$ or to $E_{corr}(PT2)$, respectively.

¹⁰For computational details see section 2.5.

Tab. 2.17: Comparison of reference energies (E_{ref}), static ($E_{corr/ref}$) and dynamical (E_{corr}) correlation from single and multi reference SDCI and PT2 for carbonmonoxide. $E_{corr/ref}$ is obtained by subtracting E_{ref} for the single reference case ($n_{ref} = 1$) from the E_{ref} for the multi reference case ($n_{ref} = 3$ or 5). Energies in H.

Orbitals	$n_{ref}^{[a]}$	E_{ref}	$E_{corr/ref}$	E_{corr} (SDCI)	E_{corr} (PT2)
HF	1	-112.78342	-	-0.31945	-0.35420
LHF	1	-112.77465	-	-0.32522	-0.37067
BLYP	1	-112.75820 ^[b]	-	-0.33588	-0.39917
B3LYP	1	-112.76753 ^[b]	-	-0.32950	-0.38627
HF	3	-112.79001	-0.00659	-0.31688	-0.33228
LHF	3	-112.80175	-0.02710	-0.30768	-0.32176
BLYP	3	-112.78796	-0.02976	-0.31712	-0.34138
B3LYP	3	-112.79179	-0.02426	-0.31459	-0.33766
HF	5	-112.79341	-0.00999	-0.31511 ^[c]	-0.330894
LHF	5	-112.81085	-0.03620	-0.30169 ^[c]	-0.31609
BLYP	5	-112.79594	-0.03774	-0.31207 ^[c]	-0.33687
B3LYP	5	-112.79934	-0.03181	-0.30993 ^[c]	-0.33350

[a] Reference configurations: 1: Ψ ; 3: $\Psi, \pi_x^2 \rightarrow \pi_x^{*2}, \pi_y^2 \rightarrow \pi_y^{*2}$; 5: $\Psi, \pi_x^2 \rightarrow \pi_x^{*2}, \pi_y^2 \rightarrow \pi_y^{*2}, \pi_x^2 \rightarrow \pi_x^{*2}, \pi_y^2 \rightarrow \pi_y^{*2}$.

[b] $E_{BLYP} = -113.35096$ H; $E_{B3LYP} = -113.31089$ H.

[c] For the Davidson corrected total energies (MRCI+Q) see table 2.12.

We first concentrate on the CI approach since it is strictly variational. While the HF approach gives the lowest energy for a single determinant wavefunction, KS-orbitals in all cases lead to higher correlation energies within the SDCI ($E_{corr}(\text{SDCI})$). Nevertheless, their total energies within the SDCI treatment are still higher than the corresponding HF value (HF: -113.10287 H, LHF: -113.09987 H, BLYP: -113.09408 H, B3LYP: -113.09703 H). Also the correlation contributions arising from the enlargement of the reference space are distinctly larger (factor 3-4) if KS-orbitals instead of HF-orbitals are employed. However, the correlation effects arising from the additional SDCI excitations on top of the reference spaces consisting of 5 reference configurations are larger if HF-orbitals are used, while the LHF-

orbitals give the lowest contribution. Summing up all effects for the computation with 5 reference configurations (Ψ , $\pi_x^2 \rightarrow \pi_x^{*2}$, $\pi_y^2 \rightarrow \pi_y^{*2}$, $\pi_x^2 \rightarrow \pi_y^{*2}$ and $\pi_y^2 \rightarrow \pi_x^{*2}$) the LHF-orbitals give the lowest energy (E_{tot} (MRCI) = -113.11254 H) being only about 0.004 H lower than the energies obtained with the other orbitals.

Employing the PT2 approach¹¹ all orbitals give higher correlation contributions than found in the corresponding CI calculations, but going from 1 over 3 to 5 reference configurations E_{corr} (PT2) absolutely decreases much faster than E_{corr} (SDCI). As a consequence the total energies of the PT2 approach increase if one goes from 1 reference configuration to 3 reference configurations, *e.g.* for LHF-orbitals the total energy decreases from -113.14532 H to -113.12351 H. The total energy again increases slightly if 5 instead of 3 reference configurations are used (LHF = -113.12694 H). A similar behaviour is found for all other orbitals. Obviously, this shows, that all orbitals lead to an overestimation of correlation effects if only 1 reference configuration is employed. Furthermore, this effect is considerably larger for KS-orbitals than for HF-orbitals. In the MRPT2 approach, however, this deficiency of KS-orbitals seems to be remedied. Please note that the correlation contribution arising from the PT2 treatment on top of the reference space of 3 configurations is 0.32176 H if LHF-orbitals are used, while for HF-orbitals 0.33228 H are obtained. For 1 reference configuration these values are 0.37067 H and 0.35420 H, respectively. The values obtained with BLYP-orbitals are somewhat larger, however.

The same trends is obtained for the formaldehyde molecule in table 2.18¹² and acetone (not given). This indicates that LHF-orbitals are well suited for an MRPT2 treatment.

2.9 Excitation energies of carbonmonoxide from MRPT2

In the next step, it is tested how well HF-, LHF-, BLYP- and B3LYP-orbitals describe excited states in an MRPT2 approach. To minimize the arbitrariness that is always involved in the construction of the reference wavefunction and to have a meaningful comparison of the orbital properties, all reference wavefunctions were determined by the same selection

¹¹For all PT2 calculations the approach of Pulay *et al.*^[2] was used. In this approach no orbital energies, but the real energy expectation values for the given configurations are employed.

¹²For computational details see section 2.15.

Tab. 2.18: Comparison of reference energies (E_{ref}), static ($E_{corr/ref}$) and dynamical (E_{corr}) correlation from single and multi reference SDCI and PT2 for formaldehyde. $E_{corr/ref}$ is obtained by subtracting E_{ref} for the single reference case ($n_{ref} = 1$) from the E_{ref} for the multi reference case ($n_{ref} = 2$). Energies in H.

Orbitals	$n_{ref}^{[a]}$	E_{ref}	$E_{corr/ref}$	E_{corr} (SDCI)	E_{corr} (PT2)
HF	1	-113.91483	-	-0.26050	-0.39631
LHF	1	-113.90445	-	-0.36734	-0.41490
BLYP	1	-113.88557 ^[b]	-	-0.38035	-0.44582
B3LYP	1	-113.89662 ^[b]	-	-0.37267	-0.43089
HF	2	-113.91867	-0.00384	-0.35905 ^[c]	-0.37808
LHF	2	-113.92769	-0.02324	-0.35228 ^[c]	-0.37389
BLYP	2	-113.91235	-0.02678	-0.36286 ^[c]	-0.39601
B3LYP	2	-113.91936	-0.02274	-0.35810 ^[c]	-0.38802

[a] Reference configurations: 1: Ψ ; 2: $\Psi, \pi^2 \rightarrow \pi^{*2}$.

[b] $E_{BLYP} = -114.56161$ H; $E_{B3LYP} = -114.49704$ H.

[c] For the Davidson corrected total energies (MRCI+Q) see table 2.15.

procedure described in the following.

Out of an orbital space consisting of the orbitals of the ground state configuration plus four virtual orbitals of each irreducible representation, the $n + 3$ configuration state functions having lowest energy are searched (n is the number of states computed in each irreducible representation). These configurations constitute the reference wavefunction of a selective MRCI. The selection in this MRCI is done by Epstein Nesbeth perturbation theory using a selection threshold of 0.1 mH. Configurations of the resulting wavefunction contributing more than a given c^2 -value are again selected as the new reference wavefunction for another selective MRCI. This procedure is repeated as described until the reference wavefunction remains unchanged, which is reached after 2-5 cycles. In table 2.19¹³ excitation energies for carbonmonoxide are displayed that were computed using LHF-orbitals and different selection thresholds for the c^2 -values which determine the size of the reference space. For valence states, already the reference space obtained with the largest selection threshold ($c^2 > 0.002$) gives quite reliable excitation energies, that are within a range of 0.2 eV of the experimen-

¹³For more computational details see section 2.5

Tab. 2.19: MRPT2 excitation energies in eV of carbonmonoxide as a function of the size of the reference space. The size is indicated by the selection threshold used for the weights of the configurations which were included in the reference space. The actual size of the reference spaces can be taken from table 2.20. For more details see text.

state	class.	$c^2 >$					exp.
		0.002	0.0015	0.001	0.0007	0.0005	
A $^1\Pi$	$n \rightarrow \pi^*$	8.32	8.35	8.35	8.34	8.40	8.51
I $^1\Sigma^-$	$\pi \rightarrow \pi^*$	9.94	9.96	9.96	9.98	9.99	9.88
D $^1\Delta^-$	$\pi \rightarrow \pi^*$	10.05	10.05	10.06	10.06	10.07	10.23
B $^1\Sigma^+$	$n \rightarrow 3s$	9.45	10.28	10.62	10.62	10.50	10.78
C $^1\Sigma^+$	$n \rightarrow 3p_z$	11.00	10.97	11.20	11.20	11.12	11.4
E $^1\Pi$	$n \rightarrow 3p$	11.09	11.30	11.30	11.30	11.31	11.53

tally determined values. For the computed Rydberg states this reference space gives larger errors of up to 1.3. If the selection threshold is decreased, the computed energies become more reliable with an accuracy better than 0.3 eV for all states, however, selection thresholds smaller than $c^2 > 0.001$ do not lead to a higher precision in the computed excitation energies. For decreasing selection thresholds, the computed excitation energies converge toward the values computed by the MRCI+Q approach in table 2.14. The number of reference

Tab. 2.20: Number of reference configurations selected for carbonmonoxide in each irreducible representation (of C_{2v}) for different selection thresholds ($c^2 >$). n_{states} denotes the number of states computed in each irreducible representation. For the corresponding excitation energies see table 2.19.

$C_{\infty v}$	C_{2v}	n_{states}	$c^2 >$				
			0.002	0.0015	0.001	0.0007	0.0005
$\Psi, D \ ^1\Delta^-, B \ ^1\Sigma^+, C \ ^1\Sigma^+$	A_1	4	26	31	41	49	60
I $^1\Sigma^-$	A_2	1	9	10	10	13	13
A $^1\Pi, E \ ^1\Pi$	B_1	2	19	24	29	39	46

configurations selected for a given c^2 -value is given in table 2.20. For the 4 states computed in A_1 -symmetry 26 reference configurations were selected if a c^2 -value of 0.002 was used, while for the A_2 -state 9 and for the two $^1\Pi$ states that were computed in B_1 -symmetry, 19 reference configurations were selected. Decreasing the threshold to 0.001, 41, 10 and 29

reference configurations were selected, respectively.

Since the selection threshold $c^2 > 0.001$ seems to offer the best balance between accuracy and computational effort, it will be used to compare the different sets of orbitals (table 2.21). The reference spaces computed with a threshold of $c^2 > 0.001$ for all molecules contain about 10 reference configurations per state computed with LHF-, BLYP- and B3LYP-orbitals, *i.e.* the number of reference configurations selected is very similar for the different types of KS-orbitals. When HF-orbitals were used, about 20 reference configurations per state are selected.

Employing this set of reference configurations, for carbonmonoxide, excitation energies computed with the MRPT2 approach using HF-orbitals give very strong deviations from experiment. In the computation shown in table 2.21 the I $^1\Sigma^-$ and D $^1\Delta$ valence states cannot be located at all. If a larger number of states is computed, valence states can be found, however, the computed excitation energies still show very strong deviations from experiment. Rydberg states are better described than valence states. For Rydberg states the computed excitation energies show deviations from experiment up to 0.65 eV.

The energies computed using LHF-orbitals tend to underestimate the experimentally de-

Tab. 2.21: Excitation energies in eV of carbonmonoxide computed by MRPT2 using a selection threshold of $c^2 > 0.001$ for the reference wavefunction. Ground state energies: HF -113.12343 H; LHF: -113.12723 H; BLYP: -113.13122 H; B3LYP: -113.13074 H.

state	class.	HF	LHF	BLYP	B3LYP	exp. ^a
A $^1\Pi$	$n \rightarrow \pi^*$	8.07	8.35	8.32	8.29	8.51
I $^1\Sigma^-$	$\pi \rightarrow \pi^*$	12.33	9.96	9.98	10.23	9.88
D $^1\Delta$	$\pi \rightarrow \pi^*$	-	10.06	9.97	10.05	10.23
B $^1\Sigma^+$	$n \rightarrow 3s$	10.83	10.62	10.64	10.73	10.78
C $^1\Sigma^+$	$n \rightarrow 3p_z$	12.15	11.20	11.76	12.19	11.4
E $^1\Pi$	$n \rightarrow 3p$	11.19	11.31	11.55	11.86	11.53
$^2\Sigma^+$	$n \rightarrow \infty$	14.20	14.24	14.23	14.32	14.01
$^2\Pi$	$\pi \rightarrow \infty$	17.82	17.46	17.53	17.66	16.91

[a] see table 2.14

termined values by 0.2 eV. An exception is the I $^1\Sigma^-$ state which is computed to be 9.96 eV compared to the experimental value of 9.88 eV. In accordance with the results from the

MRCI computations one might expect that the systematic underestimation of the experimental excitation energies might be due to a somewhat deficient description of the $^2\Sigma^+$ cation which represents the core of the computed Rydberg states. However, the ionization potential which is predicted to be at 14.24 eV overestimates the experimental by 0.2 eV. Very similar values for the ionization potential are obtained with HF- (14.20 eV) and BLYP-orbitals (14.23 eV), while the ionization potential computed using B3LYP-orbitals of 14.32 eV is somewhat higher.

The excitation energies computed with BLYP-orbitals have a comparable accuracy as the data obtained using LHF-orbitals. While most of the computed energies deviate no more than 0.2 eV from the the experimental values, the computed excitation energy for the C $^1\Sigma^+$ state of 11.76 eV lies 0.36 eV above the experimental value.

B3LYP-orbitals lead to less accurate excitation energies than BLYP-orbitals. While the excitation energies for the A $^1\Pi$ -, the D $^1\Delta$ and the B $^1\Sigma^+$ state are comparable to those computed with LHF- or BLYP-orbitals, the excitation energies for the I $^1\Sigma^+$ -, the C $^1\Sigma^+$ - and the E $^1\Pi$ state are predicted to lie 0.35, 0.46 and 0.66 eV above the experimental values.

2.10 Excitation energies of formaldehyde from MRPT2

Employing the same MRPT2 approach as described for carbonmonoxide, the energies computed for formaldehyde¹⁴ using HF-orbitals also tend to show large deviations from experiment. The vertical excitation energy of 3.63 eV for the 1 1A_2 -valence state is about 0.3 eV below the experimental value, while the 2 1B_2 with 9.02 eV is predicted by about 1 eV too high in energy. The 1 1B_2 , 1 1A_1 and 2 1A_2 having values of 7.09, 8.51 and 8.50 eV are predicted somewhat more precisely.

LHF-orbitals allow a very reliable prediction of all computed vertical excitation energies. For the computed Rydberg states, the deviation from the experimentally determined values is less than 0.1 eV. BLYP-orbitals give a good description of the $n \rightarrow \pi^*$ -valence- and the 3s Rydberg state, but the computed energies for the 3p Rydberg states deviate by 0.2 - 0.5 eV from experiment. Energies computed with B3LYP-orbitals have an error of 1 - 2 eV for all

¹⁴For computational details see section 2.15.

excited states computed for formaldehyde.

Tab. 2.22: Excitation energies in eV of formaldehyde computed by MRPT2 using a selection threshold of $c^2 > 0.001$ for the reference wavefunction. Ground state energies HF: -114.29663 H; LHF: -114.3081 H; BLYP: -114.30755 H; B3LYP: -114.30561 H

state	class	HF	LHF	BLYP	B3LYP	exp. ^a
1 ¹ A ₂	$n \rightarrow \pi^*$	3.63	3.76	3.81	3.75	3.94
1 ¹ B ₂	$n \rightarrow 3s$	7.09	7.13	7.08	7.21	7.09
2 ¹ B ₂	$n \rightarrow 3p_z$	9.02	7.92	8.46	8.13	7.97
1 ¹ A ₁	$n \rightarrow 3p_y$	8.51	8.08	8.39	8.35	8.12
2 ¹ A ₂	$n \rightarrow 3p_x$	8.50	8.28	8.59	8.57	8.38
² B ₂	$n \rightarrow \infty$	11.28	11.13	11.38	11.35	10.88

[a] see table 2.15

2.11 Excitation energies of acetone from MRPT2

While for carbonmonoxide and formaldehyde a poor performance in the MRCI computations roughly correlates to larger energy deviations in the excitation energies computed with the MRPT2 approach, for acetone, HF-orbitals perform much better in the MRPT2- than in the MRCI computations (Tab. 2.23¹⁵). Apart from the 1 ¹A₂ valence state for which an excitation energy of 3.90 eV compared to an experimental value of 4.38 eV is computed, HF-orbitals allow a precise computation of the excited states of acetone with an error of less than 0.1 eV in comparison to experimental data.

The excitation energies computed using LHF-orbitals show deviations from the experiment of 0.2 eV for the 1 ¹A₂ valence state and 0.3 to 0.4 eV for the Rydberg states. Like it was already seen in the MRCI computations, the description of the Rydberg states of acetone based on LHF-orbitals seems to be problematic.

In contrast, for acetone excitation energies computed with BLYP- and B3LYP-orbitals seem to be far more reliable than those computed with LHF-orbitals. Excitation energies for the 3s- and the 3p Rydberg states never deviate more than 0.2 eV from experiment. The error

¹⁵For Computational details see section 2.7

that is found seems to be a systematic red shift for BLYP-orbitals. Energies computed using B3LYP-orbitals agree even better with experimental data than the energies obtained with BLYP-orbitals.

Tab. 2.23: Excitation energies in eV of acetone computed by MRPT2 using a selection threshold of $c^2 > 0.001$ for the reference wavefunction. Ground state energies: HF: -192.66650 H; LHF: -192.67543 H; BLYP: -192.68902 H; B3LYP: -102.68712 H.

state	class.	HF	LHF	BLYP	B3LYP	exp. ^a
1^1A_2	$n_y \rightarrow \pi^*$	3.90	4.17	4.13	4.14	4.38
1^1B_2	$n_y \rightarrow 3s$	6.29	6.02	6.25	6.53	6.35
2^1A_2	$n_y \rightarrow 3p_x$	7.31	6.94	7.28	7.36	7.36
2^1A_1	$n_y \rightarrow 3p_y$	7.48	7.00	7.24	7.32	7.41
2^1B_2	$n_y \rightarrow 3p_z$	7.40	7.19	7.43	7.47	7.45
2B_2	$n_y \rightarrow \infty$	9.36	9.50	9.49	9.47	9.8

[a] see table 2.16

2.12 Rydberg-valence mixing of the 1^1B_{1u} -state of ethene

The very first example of this work was the electronic computation of the excitation spectrum of ethene. The excited states of ethene here are computed again, this time using the MRPT2 approach described in section 2.9. The singlet excitation energies obtained with HF-, LHF- and BLYP-orbitals are given in table 2.24¹⁶. These values are compared to excitation energies obtained from MRPT2 computations using ROHF-orbitals optimized for the $^2B_{3u}$ cation and CASPT2 results from Serrano-Andres *et al.*^[70]

Computations using HF-orbitals give an excitation spectrum which is blue shifted by about 2 eV. LHF-orbitals lead to a better agreement with experiment, the average error is 1 eV. The energies computed using BLYP-orbitals tend to overestimate the experimental values somewhat more than the spectrum computed with HF-orbitals. Especially the 1^1B_{2u} - and 2^1B_{3u} states are overestimated by 0.44 and 0.35 eV, respectively. These results roughly correspond to those obtained in section 2.1 for the MRCI ansatz. However, for the MRPT2

¹⁶For computational details see section 2.4

Tab. 2.24: Excitation energies in eV of ethene computed by MRPT2 using a selection threshold of $c^2 > 0.001$ for the reference wavefunction. Ground state energies HF: H; LHF: H; BLYP: H; B3LYP: H.

state	class.	HF	LHF	BLYP	ROHF	CASPT2 ^[a]	exp. ^[a]
1 ¹ B _{3u}	$\pi \rightarrow 3s$	7.31	7.24	7.35	7.17	7.17	7.11
1 ¹ B _{1u}	$\pi \rightarrow \pi^*$	7.85	7.70	7.75	8.34	8.40	7.66
1 ¹ B _{1g}	$\pi \rightarrow 3p\sigma$	7.99	7.95	8.03	7.84	7.85	7.80
1 ¹ B _{2g}	$\pi \rightarrow 3p\sigma$	8.17	8.08	8.18	7.95	7.95	7.90
2 ¹ A _g	$\pi \rightarrow 3p\pi$	8.42	8.30	8.45	8.24	8.40	8.28
2 ¹ B _{3u}	$\pi \rightarrow 3d\sigma$	8.79	8.75	9.04	8.49	8.66	8.62
1 ¹ B _{2u}	$\pi \rightarrow 3d\delta$	9.11	9.08	9.16	8.9	9.18	9.05
2 ¹ B _{1u}	$\pi \rightarrow 4d\pi$	9.21	9.23	9.40	9.25	9.31	9.33
1 ³ B _{1u}	$\pi \rightarrow \pi^*$	4.37	4.39	4.42	4.45	4.39	4.36

[a]: CASPT2 excitation energies are taken from Serrano-Andrés *et al.*^[70]

approach, HF-orbitals seem to be slightly better suited than BLYP-orbitals.

The ROHF-orbitals from the computation of the ²B_{3u}-cation give a very good description for all Rydberg states. The error is usually below 0.1 eV, which is about the same accuracy as that found in the CASPT2 computations by Serrano-Andres *et al.*. This shows that for the computation of excited states, an ROHF calculation of the core ion often leads to orbitals that have a quality comparable to the more expensive CASSCF-approach. The 1 ³B_{1u}-state (T-state) is correctly predicted to be around 4 eV. However, even though the Rydberg- and the T-states are computed in a very high accuracy, both approaches lead to a very large error of about 0.7 eV for the singlet antibonding valence (V) state of ethene. The erroneous prediction of the 1 ¹B_{1u}-state is quite problematic since the transition into the V-state dominates the excitation spectrum of ethene, because the oscillator strengths of valence states are much larger than those of Rydberg states. Furthermore, in liquid phase or solvent, Rydberg states are quenched by the solvent cavity, so that the excitation spectrum is completely dominated by transitions into the valence states.

The basic problem which leads to the large deviation in the excitation energies computed by the latter two MRPT2-approaches is that the reference-CI- or the CASSCF-level only includes a very small fraction of the total correlation energy, while most of the dynamical

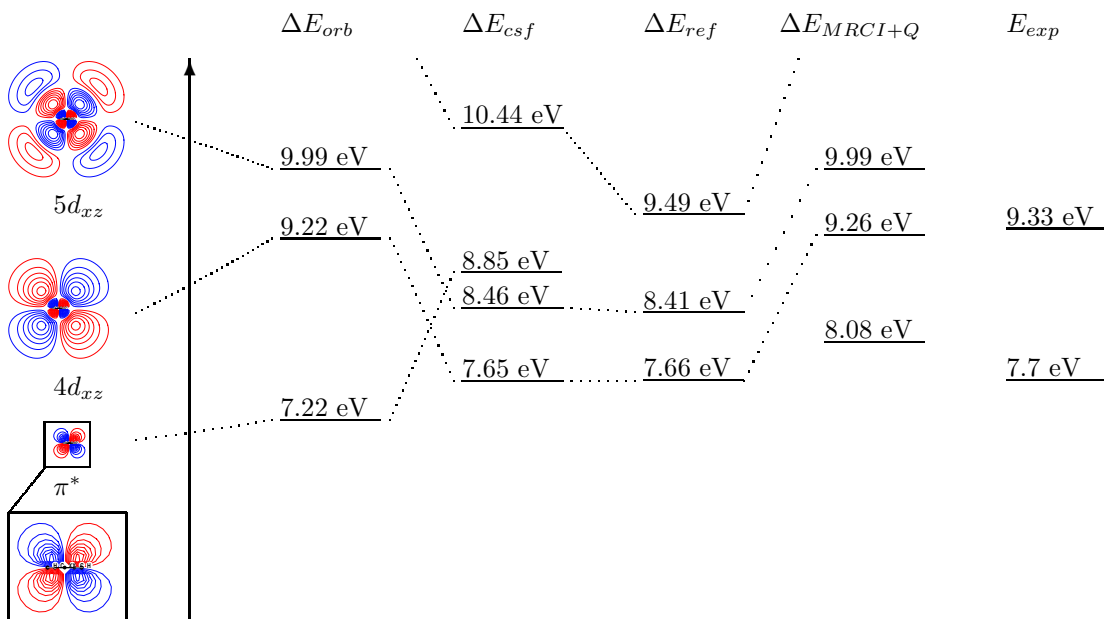


Fig. 2.6: Excitation energies of ethene computed from the sum of virtual orbital energies of virtual orbitals and the ionization potential ($\Delta E_{orb} = IP + \epsilon_{virt}$; $IP = 10.51 \text{ eV}^{[79]}$), from the energies of the configuration state functions (ΔE_{csf}), the reference CI (ΔE_{Ref}) and the full MRCI+Q-computation (ΔE_{MRCI+Q}). The orbitals for these computations were generated from an ROHF-computation of the $^2B_{3u}$ -cation. Ground state energies: $E_{LHF} = 78.05440 \text{ a.u.}$, $E_{Ref} = -78.07391 \text{ a.u.}$, $E_{MRCI+Q} = -78.42283 \text{ a.u.}$ LHF-orbitals were used to compute the ground state.

correlation energy is neglected. Since Rydberg and valence states have different contributions of dynamical correlation, valence and Rydberg states on the reference-CI level may get close to each other, which leads to an erratic Rydberg-valence mixing. However, in single-state MRPT2, the reference wavefunction is “frozen” at the reference level, and in the cases in which such mixing takes place, represents a very poor approximation to the fully correlated system. To remove this deficiency, Finley *et al.*^[95] developed the multi state-CASPT2-approach, which was applied to the ethene molecule. For the 1^1B_{1u} -state it predicts an excitation energy of 8.0 eV.

However, when HF-, LHF- and BLYP-orbitals are used, already a single state MRPT2 is able to give a correct excitation energy for the 1^1B_{1u} -state of ethene. To analyze why HF- or KS-orbitals give a correct prediction of the excitation energy, this section compares how the excitation energies and the corresponding wavefunctions change if increasing fractions of

the correlation energy are included in the computation.

Fig. 2.6 gives excitation energies obtained with ROHF-orbitals optimized for the ${}^2B_{3u}$ cation. Since HF virtual orbital energies represent electron affinities, the virtual orbital energies were added to the ionization potential to obtain ΔE_{orb} . ΔE_{orb} is 7.22 eV for the $1\ {}^1B_{1u}$, 9.22 for the $2\ {}^1B_{1u}$ - and 9.99 eV for the $1\ {}^1B_{1u}$ state¹⁷. This simple calculation already allows a good prediction of the experimental energies of 7.66 and 9.33 eV for the 1 and $2\ {}^1B_{1u}$ state. Fig. 2.7 also depicts the spatial form of the three energetically lowest orbitals of b_{2g} symmetry. It shows a very compact antibonding orbital and the two 3d and 4d Rydberg orbitals. It is interesting to note that the orbital following the π^* -orbital in the literature is often characterized as a $3d_{xz}$ Rydberg orbital.^[77] However, since the π^* and the first ($3d_{xz}$) Rydberg orbital have the same angular nodal structures, orthogonality must be ensured by the radial part of the Rydberg orbital. Since the lowest d_{xz} Rydberg orbital then has one additional radial node, following Wiberg *et al.*,^[96] it here is characterized as a $4d_{xz}$ orbital. The energies of the single configuration state functions (CSF) ΔE_{csf} have a different order than those obtained from the orbital energies. The lowest CSF with 7.65 eV is the one that describes the excitation into the $4d_{xz}$, followed by 8.46 eV for $\pi \rightarrow 5d_{xz}$ -excitation while the $\pi \rightarrow \pi^*$ -excited state is located at 8.85 eV. This ordering is found, because the ROHF procedure explicitly optimized the core-orbitals of the Rydberg state. Furthermore, due to the low interaction of the Rydberg orbital with the core, Rydberg states are generally downshifted if correlation energy is neglected or only partially taken into account.

In the next step, a small CI is performed which consists of the configurations $\pi \rightarrow 1b_{2g}$ to $\pi \rightarrow 4b_{2g}$. Since this CI represents the reference-CI for the following MRCI-computation, the corresponding excitation energy is called ΔE_{Ref} . In this case, the 4d and the 5d Rydberg states are located at 7.66 and 8.41 eV, the next state which also has mainly Rydberg character is found at 9.49 eV. However, in the reference CI no valence state can be assigned. Instead, all states of ${}^1B_{1u}$ symmetry exhibit a very strong valence contribution, which is

¹⁷In the following MRCI-computations 4 reference configurations were included. However, since the basis set does not contain gaussian function that are diffuse enough to describe a $6d_{xz}$ -orbital, the the $4b_{2g}$ -orbital has a positive orbital energy. The corresponding configuration is only included to increase the flexibility of the CI to describe the other states, however, in the figures 2.6 and 2.7 energies corresponding to the 6 d Rydberg state that are above the ionization potential are not given.

Tab. 2.25: Structure of the wavefunction for the lowest state of figure 2.6

assignment	$\Psi(CSF)$	$\Psi(RefCI)$	$\Psi(MRCI)$
$\pi \rightarrow 1b_{2g}(\Pi^*)$	0.0000	0.3397	0.7601
$\pi \rightarrow 2b_{2g}(4d_{xz})$	1.0000	0.5927	0.0927
$\pi \rightarrow 3b_{2g}(5d_{xz})$	0.0000	0.0531	0.0295
$\pi \rightarrow 4b_{2g}(6d_{xz})$	0.0000	0.0144	0.0138

shown in table 2.25 for the lowest state of figure 2.6. The character of the state is only weakly ($c^2 = 0.5927$) dominated by the $\pi \rightarrow 4d_{xz}$ excitation, while there is a very strong valence contribution of $c_{\pi \rightarrow \pi^*}^2 = 0.3397$.

The lowest state of 1^1B_{1u} symmetry has an excitation energy of 8.08 eV at the MRCI+Q level. The full MRCI wavefunction of the 1^1B_{1u} -state has a completely different structure than the first root of the reference CI. The $\pi \rightarrow \pi^*$ excitation dominates the wavefunction with $c^2 = 0.7601$, while there is some Rydberg contribution of $c^2 = 0.0927$ ($\pi \rightarrow 4d_{xz}$). The inclusion of dynamical correlation energy at the MRCI+Q-level leads to a stabilization of the ground state. As a consequence, the complete spectrum is shifted to shorter wavelengths. Only at the MRCI level the lowest state becomes a clear valence state, while the second state of 1^1B_{1u} symmetry in this case is the 4d Rydberg state with some valence contribution. The character of the computed states is completely reversed if an increasing fraction of correlation energy is included in the computations.

Figure 2.7 gives excitation energies computed with LHF-orbitals on different levels of sophistication. ΔE_{orb} gives the excitation energy computed from the orbital energy difference. For the V state, this value is 5.92 eV, which, as expected, is exactly in between the experimental singlet excitation energy of 7.66 eV and the triplet excitation energy of 4.36 eV. The excitation energy for the 2^1B_{1u} state of 9.21 eV is already very close to the experimental value of 9.33 eV. The virtual LHF-orbitals have nearly the same spatial form as the ROHF-orbitals of the 2^1B_{3u} cation.

If the excitation energy is computed from the energy difference of the single configuration state functions, the valence state is shifted upwards to 8.31 eV, while the energy of the Rydberg state is lowered to 7.75 eV. Even though both states are placed inverse, they are still

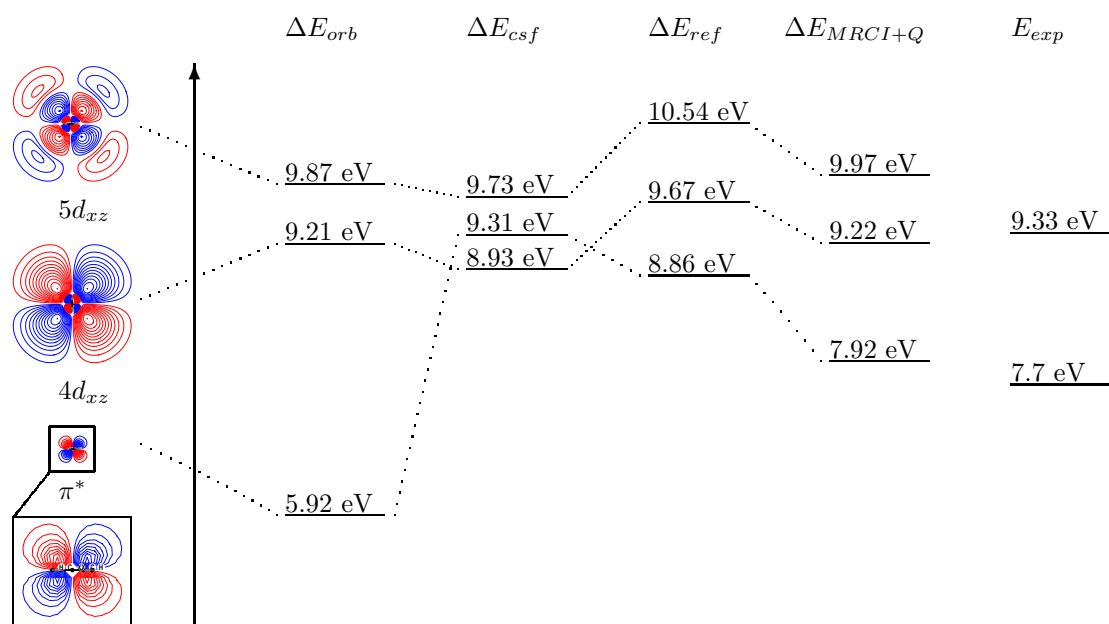


Fig. 2.7: Excitation energies of ethene computed with LHF-orbitals from the orbital energy difference (ΔE_{orb}), from the energies of the configuration state functions (ΔE_{csf}), the reference CI (ΔE_{Ref}) and the full MRCI+Q-computation (Δ_{MRCI+Q}). $\epsilon_{HOMO} = 10.179$ eV. Ground state energies: $E_{LHF} = 78.05440$ a.u., $E_{Ref} = -78.07391$ a.u., $E_{MRCI+Q} = -78.42283$ a.u.

Tab. 2.26: Structure of the wavefunction (c^2 -values) of the lowest state of figure 2.7

assignment	$\Psi(CSF)$	$\Psi(RefCI)$	$\Psi(MRCI)$
$\pi \rightarrow 1 b_{2g}(\Pi^*)$	0.0000	0.5213	0.7357
$\pi \rightarrow 2 b_{2g}(4d_{xz})$	1.0000	0.3341	0.0771
$\pi \rightarrow 3 b_{2g}(5d_{xz})$	0.0000	0.1180	0.0507
$\pi \rightarrow 4 b_{2g}(6d_{xz})$	0.0000	0.0264	0.0207

rather close in energy, and in contrast to the computation based on the ROHF-orbitals of the $^2B_{3u}$ -cation, the 5d-Rydberg configuration having a relative energy of 9.73 eV remains above the valence configuration. When LHF-orbitals are used, the CSFs of Rydberg states keep their correct energetic position. This behavior is due to error cancellation since both orbital relaxation and correlation are neglected as it has been explained in section 2.4.

The reference CI predicts an excitation energy of 8.28 eV for the lower state of 1^1B_{1u} -symmetry. Table 2.26 shows that the 1^1B_{1u} state can be characterized by an excitation into the antibonding orbital ($c^2 = 0.5213$), with a strong contribution of $c^2 = 0.3341$ of the configuration that represents the excitation into the 4d-Rydberg-orbital. However, in contrast to the computations shown before, the lowest root of the reference-CI in this case is the valence state. For the 2^1B_{1u} Rydberg state, an excitation energy of 9.30 eV is predicted, which is very close to the experimental value of 9.33 eV. The state can correctly be characterized as a 4d Rydberg state. The error cancellation from the neglect of both orbital relaxation and correlation allows a correct relative ordering of valence and Rydberg states. The high accuracy of the computation of the Rydberg state is due to the error compensation of the neglect of orbital relaxation and correlation at the level of a small CI as it was explained in section 2.4.

The full MRCI+Q computation predicts energies of 7.92 eV for the V-state and 9.22 eV for the 2^1B_{1u} -state. The c^2 -values in table 2.26 for the full MRCI computation show a less mixed character with $c^2 = 0.7357$ for the configuration $\pi \rightarrow \pi^*$ and only 0.0927 for $\pi \rightarrow 4d_{xy}$.

In contrast to ROHF-orbitals, the error cancellation for LHF-orbitals allows a correct ordering of the states already at the reference-CI level. A correct relative ordering is an important

Tab. 2.27: Structure of the reference wavefunction (c^2 -values) of the 1^1B_{1u} -state selected with a selection threshold $c^2 > 0.001$.

assignment	1^1B_{1u}		1^3B_{1u}	
	Ψ (ROHF)	Ψ (LHF)	Ψ (ROHF)	Ψ (LHF)
$\pi \rightarrow 1b_{2g}(\Pi^*)$	0.4496	0.7392	0.9698	0.9827
$\pi \rightarrow 2b_{2g}(4d_{xz})$	0.4616	0.1457	0.0096	0.0025
$\pi \rightarrow 3b_{2g}(5d_{xz})$	0.0620	0.0486	0.0069	0.0036
$\pi \rightarrow 4b_{2g}(6d_{xz})$	0.0188	0.0205	0.0064	0.0046
\sum (others)	0.0079	0.0458	0.0076	0.0066

precondition for a good construction of the reference wavefunction for single state MRPT2, because the reference wavefunction cannot readapt in single state MRPT2 due to effects of static correlation. The reference wavefunction for the MRCI computations described here was chosen somewhat smaller than the reference wavefunction for the MRPT2-computations. The latter, that are given in table 2.27, consist of 11 CSFs for LHF- and 9 CSFs for the ROHF-orbitals. The larger reference space allows a somewhat better description of the corresponding states than the 4 CSFs that were used in the MRCI-computation. Table 2.27 shows that for ROHF-orbitals, the reference wavefunction represents an equal mixture of the $\pi \rightarrow \pi^*$ -valence and the $\pi \rightarrow 4d_{xz}$ -Rydberg state, while the reference wavefunction based on LHF-orbitals already reflects the MRCI-wavefunction of table 2.26 with the dominant configuration $\pi \rightarrow \pi^*$ ($c^2 = 0.7392$) and some Rydberg contribution ($c_{\pi \rightarrow \pi^*}^2 = 0.1457$). For comparison, table 2.27 also gives the reference wavefunction for the 1^3B_{1u} -state. Since the 1^3B_{1u} -state is much lower in energy, it shows no Rydberg-valence mixing. The reference wavefunction for both sets of orbitals is strongly dominated by the excitation into the antibonding orbital, and all orbital sets lead to correct excitation energies. This last example shows that the better positioning of Rydberg relative to valence states may offer large advantages for MRPT2 approaches. However, since it is only analyzed here for an individual test case, a generalization of this finding requires further studies.

The summary for this part is given in chapter 5, while next chapter turns to the base paired and base stacked systems of xanthine.

Chapter 3

Introduction to Part 2

Base pairing in DNA and PNA

An unequivocal base pair recognition is the precondition for the successful replication and transcription of Ribo Nucleic Acid (RNA) and Desoxyribo Nucleic Acid (DNA).^[97] The correct recognition of the complementary base is mainly determined by two factors: The restrictions due to the double helix topology, which limits the available space and geometry of each base pair, and the H-bond donor-acceptor pattern of the nucleobases.

The restrictions of the double helix topology result from the constraints imposed by the helical sugar-phosphate backbone in combination with the geometry of the other base pairs, and their effect is enhanced by the tightness of the binding pocket of DNA-polymerase.^[98,99] Consequently, all base pairs must have the same size, which means that a purine base can only form a pair with a pyrimidine base and *vice versa*.^[101] This excludes for instance purine-

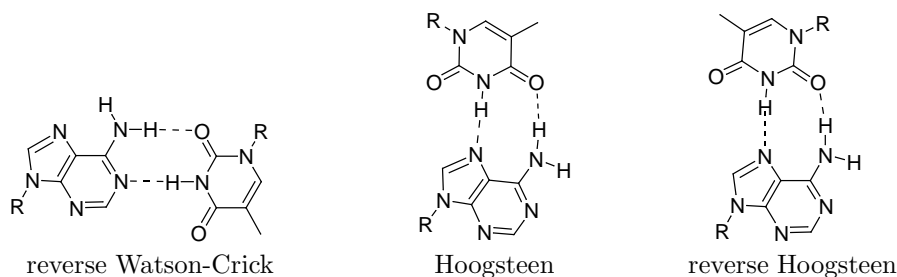


Fig. 3.1: Non-canonical base pairs of adenine and thymine



Fig. 3.2: Canonical Watson-Crick pairings of guanine, cytosine, adenine and thymine

purine pairs. Additionally, the helix topology limits the orientation of the nucleobases to the Watson-Crick pairing mode, while other combinations like the reverse Watson-Crick, Hoogsteen and reverse Hoogsteen pairing modes, illustrated in figure 3.1, cannot be realized within a DNA double strand. While the geometrical restrictions limit the possible pairings to purine-pyrimidine combinations in the Watson-Crick pairing mode, the hydrogen bond donor/acceptor pattern of nucleobases determines that adenine only pairs to thymine and guanine to cytosine like it is shown in fig. 3.2.

However, next to the hydrogen bonding pattern and the steric restrictions induced by double helix topology and DNA polymerase, also base stacking interactions, and H-bonding of the nucleobases to DNA-polymerase strongly contribute to the unambiguity of the template base recognition.^[100] The stability of DNA double strands results from the H-bond strength (not to be set equal to the hydrogen bonding pattern, which denotes the hydrogen bonding geometry), stacking and solvent interactions. Only this complex interplay of various interactions makes it possible that the base pair recognition is at the same time not only highly selective, but also reversible, so that next to an unequivocal base pair recognition also the separation of the double strand into single strands is possible.

To better understand the complex interplay between the different interactions, model systems are needed, in which the strong limitations induced by the backbone are lifted, so that a wider spectrum of possible pairings can be studied. One such model system is Peptide Nucleic Acid (PNA), in which the ribosyl phosphodiester backbone is replaced by a peptide chain as it is shown in fig. 3.3. PNA received considerable interest in 1991 when Nielsen *et al.*^[102] proposed aminoethylglycin-PNA as an antisense agent. The basis of the pharmaceutical activity of aminoethylglycin-PNA is the formation of double strands with messenger-RNA that inhibits the expression of the corresponding proteins. Another PNA system, which, however, does not bind to DNA or RNA, is PNA with a regular alanyl back-

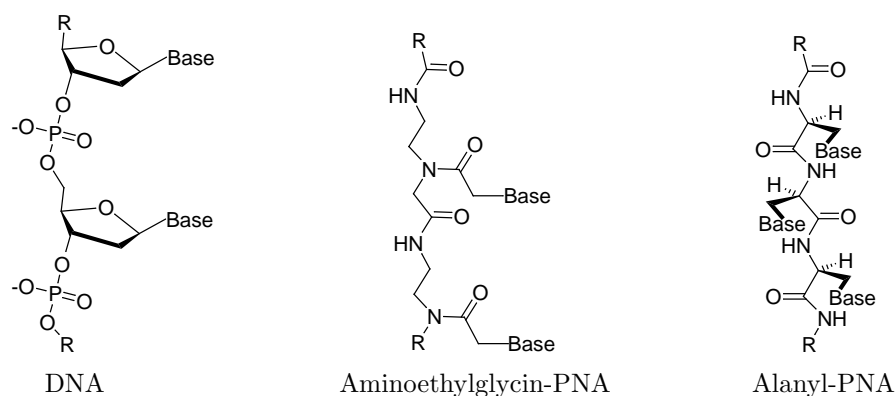


Fig. 3.3: Structures of DNA, Aminoethylglycin-PNA and Alanyl-PNA

bone^[103] (alanyl-PNA). Even though at the present time alanyl-PNA has no prospects of serving as a pharmaceutical agent, it presents an even better model system to study base pairing processes. This is because of the particular structural properties of alanyl-PNA:

In alanyl-PNA, the distance of each xanthine unit in the backbone is about 3.6 Å, which is close to the ideal stacking distance of 3.4 Å.^[104] If the nucleoamino acids in the backbone of alanyl-PNA have alternating configurations, alanyl-PNA forms linear instead of helical double strands. The backbone is then forced into a half β -sheet conformation.

The linear backbone topology allows a wide variety of different pairing modes. Examples are given in table 3.1 for the canonical nucleobases.^[105] Next to pyrimidine-purine pairings, the possible combinations also include purine-purine base pairs, which are of considerable interest, since they are often observed in RNA chemistry, while pyrimidine-pyrimidine pairings have not been found to form stable double strands in alanyl-PNA.

Tab. 3.1: UV-melting temperatures (T_m) of alanyl-PNA hexamers of the sequence $XXYXYY$ or $XXXXXX$ (6 μ M, 0.1 M NaCl, 0.01 M $\text{Na}_2\text{HPO}_4/\text{H}_3\text{PO}_4$, WC = Watson-Crick, rev WC = reverse Watson-Crick, rev H = reverse Hoogsteen)^[105]

		adenine	thymine	guanine	cytosine
adenine	pairing mode	rev WC / rev H	rev H	rev H	rev WC
	T_m	21 °C	25 °C	32 °C	20° C
guanine	pairing mode		WC	rev WC / rev H	WC
	T_m		28° C	41° C	58° C

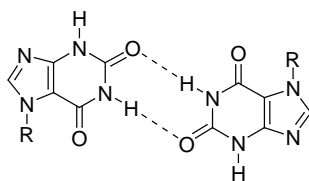


Fig. 3.4: Dimer of xanthine

One major analytic tool for analyzing base paired systems is temperature dependent UV-vis spectroscopy. In this method the melting temperature T_m is determined, at which 50% of double strands are dissociated. The T_m -values of several alanyl-PNA hexamers of the canonical bases^[105] are given in table 3.1. The melting temperatures of hexameric PNA double strands based on two-dentate pairings are located between 20 and 32 °C, while those that are able to form a third H-bond have melting temperatures between 41 and 58 °C. Also the T_m value of the non-canonical xanthine-2,6-diaminopurine alanyl-PNA hexamers (XXDXDD) of 54 °C is well within the range of those for double strands built up from basepairs with three H-bonds. An exceptionally stable double strand is the self-paired hexameric xanthine-alanyl-PNA, which has a melting temperature of 48 °C. This melting temperature is in the range of the T_m -values hexameric alanyl-PNA double strands that involve three hydrogen bonds, even though in the xanthine self pairing only two hydrogen bonds can be formed.^[106]

Xanthine and xanthine derivatives

Well known examples of xanthine compounds are the pharmacologically active methylated derivatives caffeine, theophylline, or theobromine (fig. 3.5), which act as stimulants and diuretics. They are found for instance in *Coffea arabica*, *Camellia sinensis* and *Theobroma cacao* which are consumed in the form of coffee, tea and chocolate.^[107] As a consequence, purine bases of the xanthine type have always been of great interest especially for theoretical

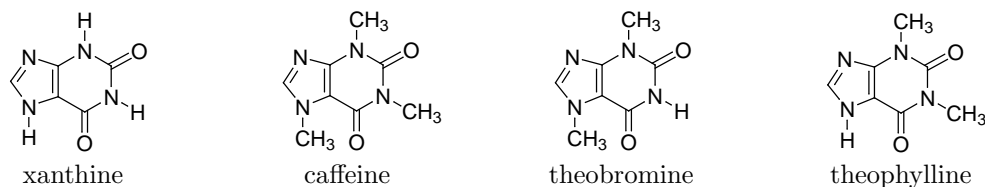


Fig. 3.5: Xanthine and methylated xanthine derivatives caffeine, theobromine and theophylline

chemists. Furthermore, xanthine is an intermediate on the pathway of purine degradation and both xanthine and its desoxy derivative hypoxanthine are converted to uric acid by the action of xanthine oxidase^[108] (fig. 3.6). An inhibition of this oxidation reaction by xanthine derivatives is of considerable pharmaceutical interest, because high plasma levels of uric acid (hyperuricaemia) are linked to the development of gout.

In the context of base pairing properties it is even more important that N9-xanthine and hypoxanthine are major purine deamination products derived from adenine and guanine, respectively, and that both are known to be highly mutagenic lesions.^[109,110] In physiological medium, the deamination reaction can either occur spontaneously by hydrolysis, or due to reactions with free radicals like OH and NO,^[111,112] or nitrous acid^[113,114] (fig. 3.6). Xanthine and hypoxanthine are for example found to be the major lesions in epidermal skin or calf thymus after exposure to cigarette smoke.^[115] It is obvious, that the transformation of

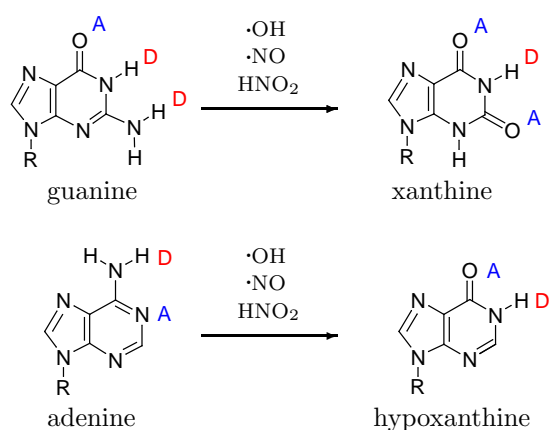


Fig. 3.6: Deamination of guanine and adenine due to the reaction with nitric acid or free radicals. D denotes an H-bond donating group, while A stands for an H-bond accepting functionality.

adenine to hypoxanthine simply reverses the H-bond donor acceptor pattern of the purine base (fig. 3.7), so that, in difference to adenine which pairs to thymine, hypoxanthine preferentially forms base pairs to cytosine. Since this base pairing leads to the substitution of thymine with cytosine in the complementary strand, which represents a point mutation, it is called mutagenic.

Fig. 3.8 compares the donor-acceptor pattern of guanine and xanthine. Like for guanine, also



Fig. 3.7: Inversion of the H-bond donor acceptor pattern and exchange of the complementary base going from hypoxanthine to thymine. D denotes an H-bond donating group, while A stands for an H-bond accepting functionality.

for xanthine a Watson-Crick pairing with cytosine should be possible, even though the two hydrogen bonds of the base pair need to compensate the repulsive electrostatic interaction of the carbonylic oxygen atoms. Since this base pairing does not change the base sequence of the complementary strand, it is called non-mutagenic. A mutagenic Watson-Crick pairing with thymine (which changes the base sequence of the complementary strand since cytosine is replaced by thymine) does not seem to be feasible, since it only involves repulsive interactions. This, however, is in contrast to the experimental finding that xanthine-thymine base pairs are inserted into DNA at a rate comparable to that of the xanthine-cytosine base pairing.^[116] To date no structures of DNA duplexes containing xanthine have been determined by X-ray crystallography or NMR. As a consequence the pairing modes of xanthine to canonical bases and the stabilizing interactions are not known. The identification of the possible pairing modes is the second major topic of the next chapter.

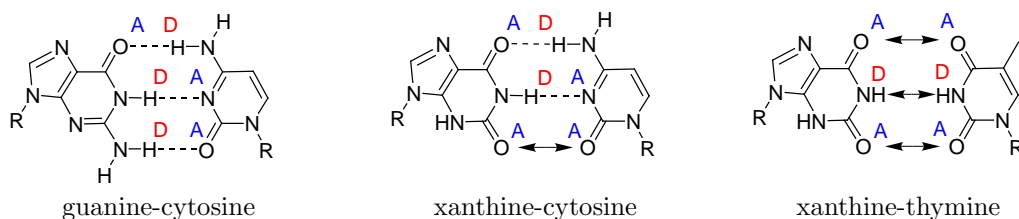
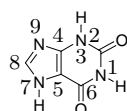


Fig. 3.8: Change in the donor acceptor pattern going from guanine to xanthine, and Watson-Crick pairings of xanthine to cytosine and thymine. D denotes an H-bond donating group, while A stands for an H-bond accepting functionality.

Summarizing, the second part of this work has two fundamental motivations:

- One is to explain the unusual stability of selfpairing xanthine-alanyl PNA double strands. To explain the stability of the xanthine-xanthine pairing, a model based on tautomeric forms of xanthine has already been proposed earlier.^[106,118] However, in refs. 106 and 118, the xanthinyl nucleoside amino acids and the respective oligomers have been erroneously described as N9 regioisomers¹⁸. A reliable determination of the correct regioisomer of xanthine by theoretical means is shown in section 4.1.1. In the sections following, all possible pairing modes of the N7-regioisomer of xanthine and those of other alanyl PNA pairings are compared with respect to hydrogen bonding and stacking. Section 4.1.5 finally sets up a model that explains the high stability of the xanthine-alanyl PNA double strand based on hydrogen bonding and the stacking interaction only.
- The basic question of section 4.2 is how xanthine is able to form base pairs with thymine that can be incorporated into DNA at a rate comparable to that of the xanthine-cytosine base pairing. Section 4.2 first studies the general base pairing properties of N9-xanthine. In this context, symmetric dimers of xanthine are studied and compared to other hydrogen bonded systems, among these the dimer of hypoxanthine. The predictions made are tested by comparison with experimental data. The reason to study symmetric homodimers is that the interaction energies in hydrogen bonded systems are small compared to the accuracy of the methods. In symmetric pairings, the number of possible interactions is decreased, while their size is doubled, which makes an evaluation of the different effects more reliable. In the next step, the Watson-Crick base pairs of neutral and deprotonated xanthine and for comparison also of hypoxanthine with canonical pyrimidine bases are computed to identify the possible canonical/non-canonical base pairing modes and to understand the mutagenicity of xanthine formed in DNA.

¹⁸Numbering of atoms in xanthine:



Chapter 4

Base Pairing of Xanthine and Xanthine Derivatives

4.1 The stability of alanyl PNA

Before the factors responsible for the stability of the xanthine alanyl PNA double strand was examined, the regioisomer of xanthinyl nucleic acid that builds up the hexamer needed to be identified. The alanyl backbone was attached to xanthine **1** by a nucleophilic ring opening of Boc-serinelactone **2** which is shown in fig. 4.1.^[106] The main products of this reaction are the Boc protected N7- and N9-xanthinyl nucleo amino acids **3** and **4**. The reaction products were characterized by ¹H-NMR and ¹³C-NMR spectroscopy, however, both NMR-methods by themselves do not allow to distinguish the regioisomers **3** and **4**, since no difference in the ¹H-NMR coupling pattern can be seen in the ¹H-NMR spectrum for the **3** and **4**. A correct assignment was possible by an evaluation of the experimental spectrum using a theoretical approach.

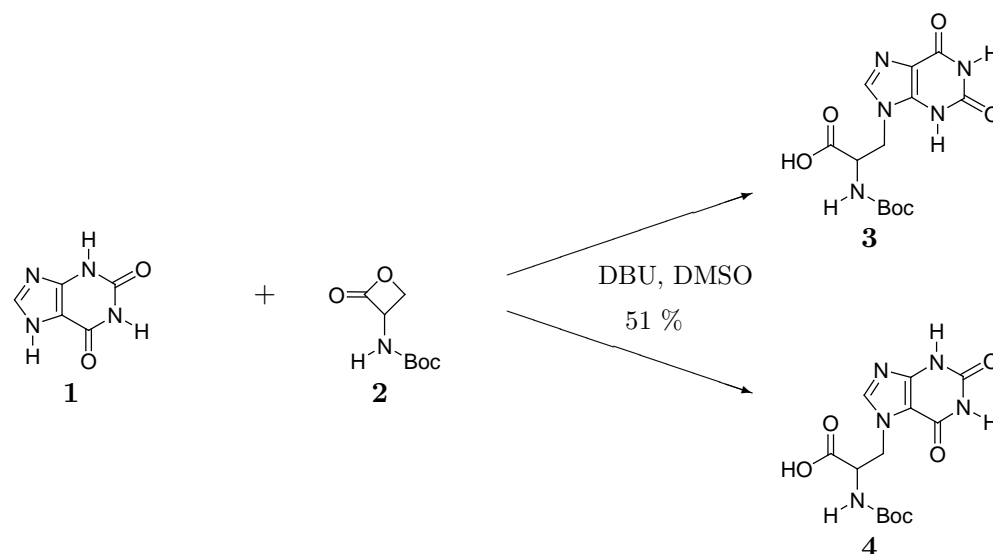


Fig. 4.1: Alternative reaction pathways during the ring opening of Boc-serinelactone to form Boc-protected xanthinyl nucleosides

4.1.1 Regioisomerism of xanthine

Since ^1H -NMR-spectra are strongly influenced by the solvent environment and usually only show a poor agreement with NMR-spectra measured at room temperature, here only the evaluation based on the ^{13}C -NMR spectrum is given. The experimental ^{13}C -NMR spectrum of the Boc-protected xanthinyl nucleoside and the computed ^{13}C -NMR spectra of N7- and N9-xanthine **5** and **6** (fig. 4.2) are given in table 4.1¹⁹.

¹⁹Computational details for section 4.1.1:

The geometries of the two tautomers of xanthine were optimized using the BLYP^[55,56] functional in combination with the TZVP basis set,^[92] which represents an (11s6p1d) AO basis in a [5s3p1d] contraction. For the calculation of three and four center integrals, the Resolution of Identity (RI)^[119,120] approximation was used. Minima were checked by frequency calculations from which also zero point vibrational energies, thermal corrections and entropies necessary for the computation of ΔG were taken (BLYP/TZVPP). NMR-spectra were computed using the HF and the MP2 approach in combination with the TZVP basis set. The energies of the two regioisomers of xanthine were characterized by the MP2 ansatz^[78] in combination with a TZVPP basis set^[92] using the RI approximation and the B3LYP-^[57] and PBE0-^[121,122] functionals also in combination with the TZVPP basis set. The TZVPP basis set^[92] represents an (11s6p2d1f) AO basis in a [5s3p2d1f] contraction. Solvent effects were estimated using the COnductor like Screening MOdel (COSMO)^[123] with a dielectric constant of 46.7 to simulate the DMSO solvent. Natural atomic orbital population analyses were performed using the B3LYP functional in combination with the 6-31G(d) basis set^[124] using the Gaussian98



Fig. 4.2: Regioisomers of xanthine

First, the accuracy of the computational methods needs to be tested without knowing the regioisomer described by the experimental spectrum. Of the five carbon atoms of xanthine, C4, C5 and C8 belong to the pyrimidine ring and consequently should be influenced by the substitution pattern, while the chemical shifts of the C2 and C6 carbon atoms should be very similar for both regioisomers, so that the evaluation of the computational method is done by a comparison of the computed and experimental chemical shifts for C2 and C6. For C2, the HF approach predicts a chemical shift of 165.4 ppm (**5**) or 164.8 (**6**) ppm, while the MP2 chemical shifts are 153 ppm and 152.6 ppm, respectively, which both are much closer to the experimental value of 151.6 ppm. Also the HF chemical shifts for C6, that are around 171 ppm, strongly deviate from the experimental chemical shift of 156.0 ppm, again, this value is in very good agreement with the chemical shifts of 157.7 and 157.8 ppm that were obtained by the MP2-approach. The comparison shows, that the MP2 approach allows a very precise prediction of the ^{13}C -chemical shifts, which is accurate enough to allow an unambiguous determination of the regioisomer, while the chemical shifts cannot be predicted precisely enough if correlation is not taken into account like it is the case in the HF-method.

Tab. 4.1: Computed and experimental C^{13} -NMR chemical shifts of N7- and N9-xanthine.

			C2	C6	C4	C5	C8
5	N7-H-xanthine	HF/TZVP	165.4	170.8	165.1	110.7	159.0
5	N7-H-xanthine	MP2/TZVP	153.0	157.8	150.6	115.3	134.2
6	N9-H-xanthine	HF/TZVP	164.8	170.6	154.1	122.1	143.9
6	N9-H-xanthine	MP2/TZVP	152.6	157.7	140.1	123.4	130.0
exp.			151.6	156.0	149.9	106.8	144.0

program package.^[125] All other computations were performed using the TURBOMOLE program package.^[64]

After this assessment of the accuracy for the HF- and MP2-method in the computation of ^{13}C chemical shifts of xanthine, theoretically and experimentally determined NMR shifts for the carbon atoms C4, C5 and C8 are compared to determine the regioisomer. For C4, the shift determined by the MP2 approach for the N7 regioisomer **5** of 150.6 ppm is in a very good agreement with the experimental value of 149.9 ppm, while for the N9 regioisomer **6** a value of 140.1 ppm is predicted.

For the carbon atoms C5 and C8 the agreement of the values computed for **5** (deviations of 8 and 10 ppm) is not as good as for the other carbon atoms, however, the difference to the experimentally determined chemical shifts is still much smaller for **5** than for **6**, for which deviations of 14 and 17 ppm are found. Summarizing the computations, the comparison of experimental and computed ^1H -NMR-spectra shows that the Boc protected xanthinyl nucleic amino acid can be identified to be the N7 regioisomer of xanthine.



Fig. 4.3: Xanthine derivatives methylated at N7 (**7**) and N9 (**8**)

For C5 and C8 the deviation from the experimental value of about 10 ppm is larger than for all other atoms, where the agreement with experiment is within 2 ppm. Since C5 and C8 are the carbon atoms neighboring the substituted N9-atom, it might be suspected, that the deviation stems from a too reduced model system since the alkyl rest is replaced by a hydrogen atom.

Tab. 4.2: Computed and experimental ^{13}C -NMR chemical shifts of N-7-CH₃- and N-9-CH₃-xanthine.

			CH ₃	c-2	c-6	c-4	c-5	c-8
7	N-7-CH ₃ -xanthine	MP2/TZVP	38.7	153.0	159.1	152.1	116.9	137.9
8	N-9-CH ₃ -xanthine	MP2/TZVP	33.9	152.8	157.9	141.7	124.8	134.5
	exp.		(47.7)	151.6	156.0	149.9	106.8	144.0

To test the model system, the hydrogen atoms in **5** and **6** were replaced by methyl groups to give **7** and **8** (fig. 4.3), however, no further improvement in the computed chemical shifts can be seen for the values given in table 4.2. However, also these computations show that the

main product of the alkylation reaction shown in figure 4.1 is the N7-substituted regioisomer. The formation of N7-xanthinyl nucleic acid is in good agreement with the thermodynamic stability of both isomers. Table 4.3 gives the energy difference between N7-H-xanthine **5** and N9-H-xanthine **6** both in gas phase and a dielectric solvent with $\epsilon = 46.7$ (DMSO). The computed energies given in table 4.3 show a preference of about 9 kcal/mol of the N7-regioisomer. This difference is slightly decreased for ΔG .

Tab. 4.3: Energies in kcal needed to tautomerize N7-H xanthine to N9-H-xanthine.

	ΔE	ΔG	ΔE
	$\epsilon = 1$	$\epsilon = 1$	$\epsilon = 46.7$
MP2/TZVPP	8.9	6.2	3.2
B3LYP/TZVPP	9.2	6.4	3.2
PBE0/TZVPP	9.4	6.6	3.8

The energy difference of the two regioisomers can easily be explained if electrostatic interactions are taken into account. Fig. 4.4 shows that for **5** the distance between the negatively charged N9 atom and the N3-hydrogen atom is only 2.74 Å, while the distance between N7-hydrogen atom and the oxygen atom of the C6=O group is 3.09 Å. In contrast to that, the two positively charged hydrogen atoms and the negatively charged N9 and (C8-) oxygen-atoms repel each other for N9-xanthine **6** (distances of 2.84 and 3.12 Å, respectively). The

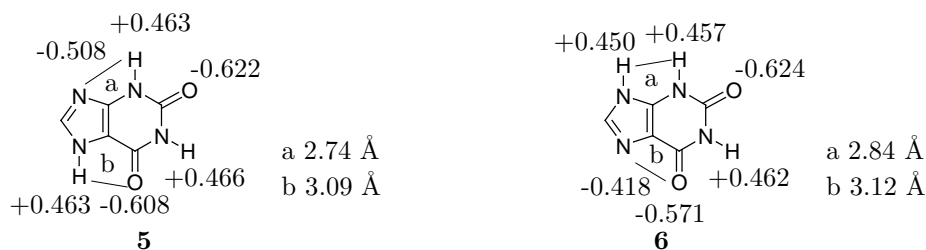


Fig. 4.4: Atomic charges of **5** and **6**

fact, that the energetic difference of the two regioisomers is mainly caused by electrostatic interactions is in line with the decrease of the tautomerization energy in dipolar solvent, since solvent effects decrease electrostatic attractions and repulsions. Furthermore, no significant difference is found between the energies obtained using the MP2 approach and the B3LYP or PBE functionals. Since the energy difference between the two regioisomers is mainly determined by electrostatic interactions also the HF-method or the MP2 method with smaller

basis sets lead to very similar results (Compare Ha *et al.*^[141]).

4.1.2 Homodimers of N7-xanthine

Section 4.1.1 has shown that the xanthine-alanyl PNA described in ref. 106 is built from N7-xanthinyl nucleoside amino acids. For N7-xanthine, the large number of ten possible hydrogen bonded bidentate base pairs can be formed, because 3 H-bond acceptor and 2 H-bond donor functionalities alternate. Fig. 4.5 shows the six dimers that are bound only *via* the Watson-Crick site of the xanthine, and fig. 4.6 gives the four additional dimers that involve the Hoogsteen site of xanthine.

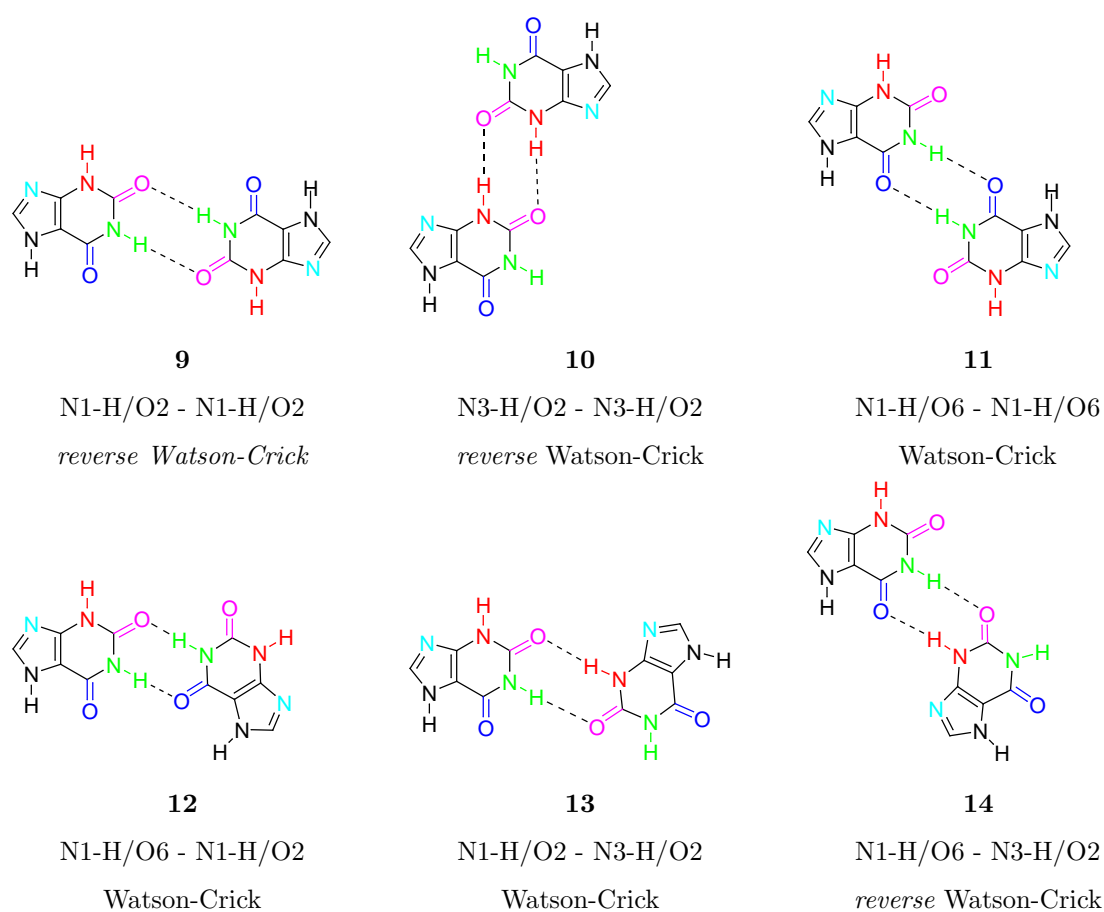


Fig. 4.5: Possible hydrogen bonded dimers of N7-xanthine

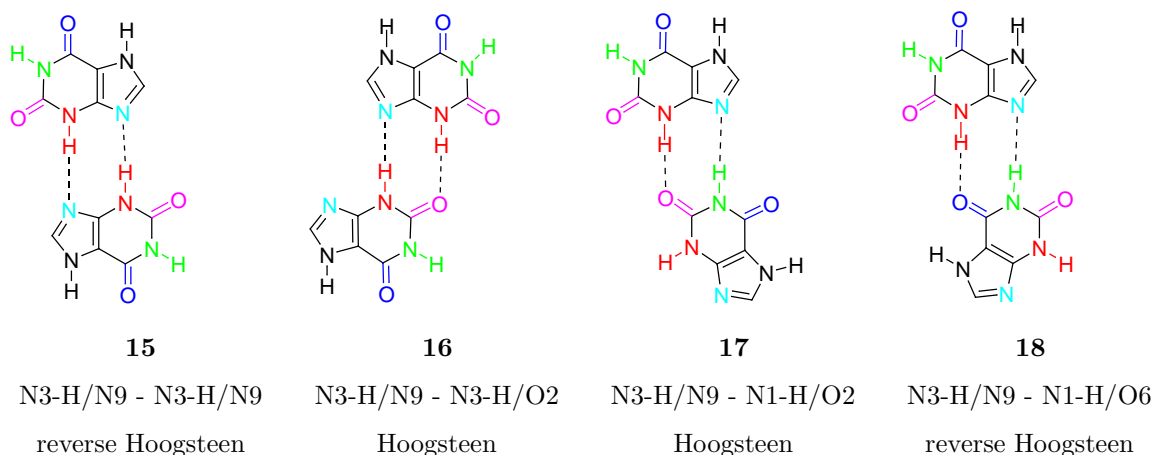


Fig. 4.6: Possible hydrogen bonded dimers pairing via the Hoogsteen site of N-7-xanthine

Table 4.4²⁰ gives dimerization energies computed for the xanthine base pairs **9** to **18**. The dimerization energies computed with the MP2-method range from -10.9 kcal/mol for pairing **11** to -14.1 kcal/mol for the symmetric pairing **10**²¹. The most stable pair is the reverse Hoogsteen pair **15**, which is expected, because a pyrimidine nitrogen atom is a better H-bond acceptor than a carbonylic oxygen atom. Other pairings involving only one Hoogsteen site are slightly weaker, the MP2-method predicts dimerization energies of -13.9

²⁰Computational details for section 4.1.2:

The dimerization energies were computed using geometries that were optimized with the BLYP^[55,56] functional in the TZVP^[92] basis set. The RI^[119,120] approximation was used for the calculation of the three and four center integrals. Minima were checked by frequency calculations. The dimerization energies were computed with the BLYP- and MP2-approach^[78] in combination with the TZVPP^[92] basis set and using the RI approximation. The basis set superposition error (BSSE) was corrected according to the Boys-Bernardi counterpoise procedure.^[126] All computations were performed using the TURBOMOLE program package.^[64]

²¹Thermodynamic data:

For all computations of dimerization energies, only the differences on the absolute energies are given. Since the dimerization changes the structure and force constants of the monomers only little, ΔH for the dimerization differs from E_{dim} by only +1 to +2 kcal/mol, proportional to the size of E_{dim} . Furthermore, the computed entropy contribution at room temperature disfavors the dimerization by more than 20 kcal/mol, but does not differ significantly for the individual pairings. However, the computed dimers only represent a fragment of the complete system, so that the relative contribution of the entropy is much smaller. Furthermore, in the real system, the dimer formation liberates solvent molecules, which leads to an increase of entropy for the dimer.

Tab. 4.4: Dimerization energies of the ten possible pairing modes of N7-H-Xanthine. All energies in kcal/mol.

dimer	E_{dim}			
	MP2/TZVPP	BLYP/TZVP	MP2/TZVPP	BLYP/TZVP
	$\epsilon = 1$	$\epsilon = 1$	$\epsilon = 78$	$\epsilon = 78$
9	-11.6	-8.6	-5.4	-2.6
10	-14.1	-11.8	-7.3	-3.8
11	-10.9	-9.2	-5.0	-2.7
12	-11.3	-9.0	-5.2	-2.6
13	-12.8	-10.3	-6.1	-3.2
14	-11.5	-10.2	-5.3	-3.2
15	-14.8	-12.0	-6.6	-4.4
16	-13.9	-10.9	-6.7	-3.6
17	-13.1	-10.1	-6.1	-3.7
18	-13.0	-10.1	-6.1	-3.6

kcal/mol for pairing mode **16** and around -13 kcal/mol for the pairings **17** and **18**. The dimerization energies computed with the BLYP-functional are about 2 kcal/mol smaller than the MP2-values. Since the stability measurements were performed in aqueous solution, the dimerization energies were also computed with $\epsilon = 78$ using the conductor like screening model. For the polar environment, the MP2 energies are reduced by a factor of two, while the BLYP energies are reduced by a factor of 3. While the MP2-values are more reliable for the gas phase,^[127] the values for polar environment are overestimated, because the solvent effect is only considered in the HF part, while the perturbational treatment considers the solvent effect only indirectly *via* the modification of the orbitals, so that the dimerization energies of around -3 to -4 kcal/mol determined by the BLYP approach can be considered to be more reliable. The hydrogen bond strengths found here are in line with values found for other two-dentate systems like the adenine-thymine bases pair or the thymine-thymine pairing,^[127] which all have dimerization energies around -10 kcal/mol. Hydrogen bonding by itself therefore is not the factor that leads to the unusually high stability of the xanthine alanyl-PNA double strand.

4.1.3 Stacked tetramers

In contrast to hydrogen bonded base pairs, the geometry of stacked base pairs is mainly determined by the constraints of the backbone, *i.e.* the topology of alanyl-PNA. A free optimization of stacked dimers or tetramers would lead to rather artificial structures, and it needs to be considered whether a pairing mode can be realized at all within a given backbone geometry. To take into account the restrictions imposed by the backbone, the full hexamers were optimized using the MMFF94 force field. Out of these structures, the xanthine-xanthine tetramers were cut out and the remaining dangling bonds at the N7 positions were substituted by methyl groups, so that a structure like it is shown in fig. 4.7 results. The interaction energies obtained for the stacked xanthine dimers are shown in table 4.5²².

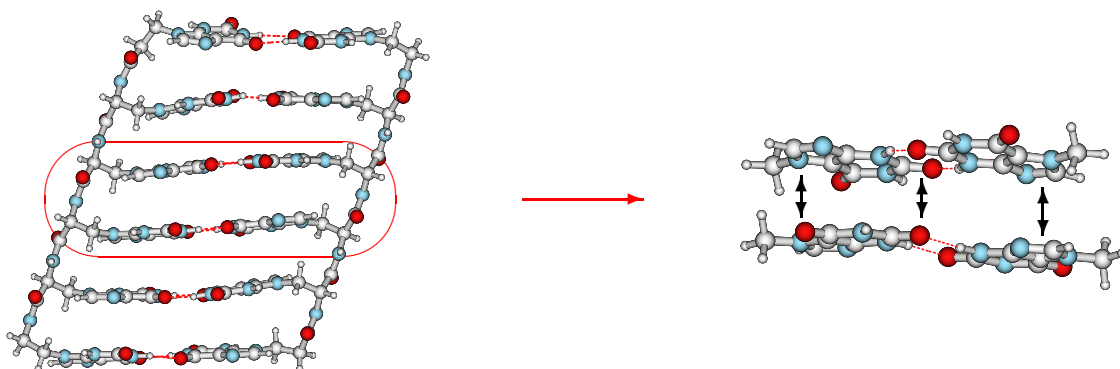


Fig. 4.7: Generation of stacked tetrameric structures

The computed stacking energies show a strong variation for the different pairing modes and

²²Computational details for section 4.1.3:

To compute stacking energies, the full alanyl-PNA hexamers, for which only the terminal groups were replaced by hydrogen atoms, were optimized using the MMFF94 force field^[128] as implemented in the SYBYL program package.^[129] The alanyl backbone was deleted and the resulting methylene group at the N7-atom was saturated by a hydrogen atom. For further details see text. The stacking interaction was then computed for two central xanthine pairs using the MP2 approach. These MP2 calculations were performed in the TZVP^[92] basis set and the RI approximation.^[119,120] To exclude the BSSE error, ghost orbitals were used for the monomers. Solvent effects were estimated using the COSMO^[123] with a dielectric constant of 78 to simulate the aqueous environment, geometries were always taken from gas-phase computations. The MP2-computations were performed using the TURBOMOLE program package.^[64]

vary from -8.8 kcal/mol for pairing **17** to -18.4 kcal/mol for pairing **13**. The pairings **11**, **12**, **14** and **15** that are not listed in table 4.5 do not lead to stable hexameric double strands. The stacking energies obtained here are considerably lower than the the interaction energies obtained in the literature for stacked dimers of -9 to -10 kcal/mol,^[127] however, since these values refers to freely optimized stacked pairs, their stacking geometry usually cannot be realized within any stacked structure involving a backbone.

The strong variation found in the gas phase disappears if solvent effect are included *via* the conductor like screening model. This behaviour is expected, because the extreme values for E_{stack} are due to very unfavorable or due to very favorable electrostatic interactions, which both are decreased in polar environment. Since most of the charge concentrations are located on the heteroatoms at the outside of the tetramer, the computations for solvent environment correctly reflect the physical system. In polar solvent, the tetramers can be grouped into pairings **9**, **10** and **13** that pair *via* the Watson-Crick site with a stacking energy around -14 kcal/mol, and pairings **16**, **17** and **18** that pair *via* the Hoogsteen side with about -11 kcal/mol. Since virtually all literature data is computed for freely stacked dimers, an evaluation of the stacking interaction and its correlation to the melting temperatures is not possible within the xanthine-alanyl PNA system only. However, an evaluation is possible by comparison to the corresponding values of other alanyl PNA systems, which will be done in the next section.

Tab. 4.5: Stacking energies E_{stack} of the pairing modes **9**, **10**, **13**, **16**, **17** and **18** of N7-H-Xanthine. All energies in kcal/mol.

tetramer	E_{stack}	
	MP2/TZVPP, $\epsilon = 1$	MP2/TZVP, $\epsilon = 78$
9	-12.1	-13.9
10	-13.5	-14.2
13	-18.4	-13.9
16	-9.2	-10.8
17	-8.8	-11.4
18	-14.1	-11.0

4.1.4 Hydrogen bonding and stacking of other alanyl PNA systems

In the previous two sections, the size of hydrogen bonding and stacking has been determined for dimers of N7-xanthine. However, it is not clear how these values correlate with the stability of a hexameric double strand of PNA, which is determined by the melting temperature T_m . This section compares hydrogen bonding and stacking for different alanyl PNA systems,

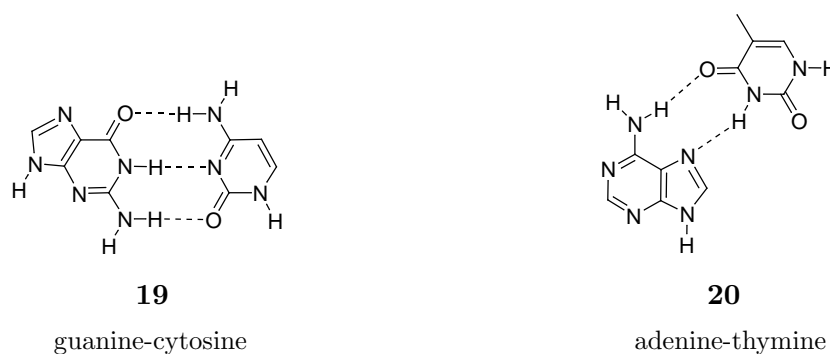


Fig. 4.8: Pairing of guanine and cytosine (canonical Watson-Crick) of adenine-thymine (Hoogsteen)

for which the pairing mode is either known or consists only of a small number of possible combinations. The systems chosen are the guanine-cytosine alanyl-PNA, which forms pairings in the Watson-Crick mode **19** and the adenine-thymine pairing, which is realized in the Hoogsteen pairing mode **20**,^[105] since the Hoogsteen pairing mode is slightly more stable than the regular Watson-Crick pairing mode.^[127] Both pairings are shown in fig. 4.8. The third example is the 2,6-diaminopurine-xanthine pairing, for which a melting temperature of 54 °C has been determined.^[106] For this pairing, the tridentate modes **21** - **24** illustrated in fig. 4.9 are possible.

Table 4.6²³ gives dimerization energies for the base pairs **19** to **24**. For the guanine-cytosine base pair **19**, the MP2-method predicts a dimerization energy of -25.7 kcal/mol, the BLYP value of -23.4 kcal/mol is somewhat smaller. For adenine-thymine (**20**), -14.4 kcal/mol and -11.7 kcal/mol are computed, respectively. The MP2-values are about 2 kcal/mol larger than the corresponding energies from the literature.^[127] The larger values for H-bonded dimers are obtained, because the very flexible TZVPP basis set allows a better description of the hydrogen bonds, while computations described Sponer *et al.*^[127] in all cases used less flexible

²³For computational details see sections 4.1.2 (computation of H-bonded dimers) and 4.7 (computation of stacked tetramers).

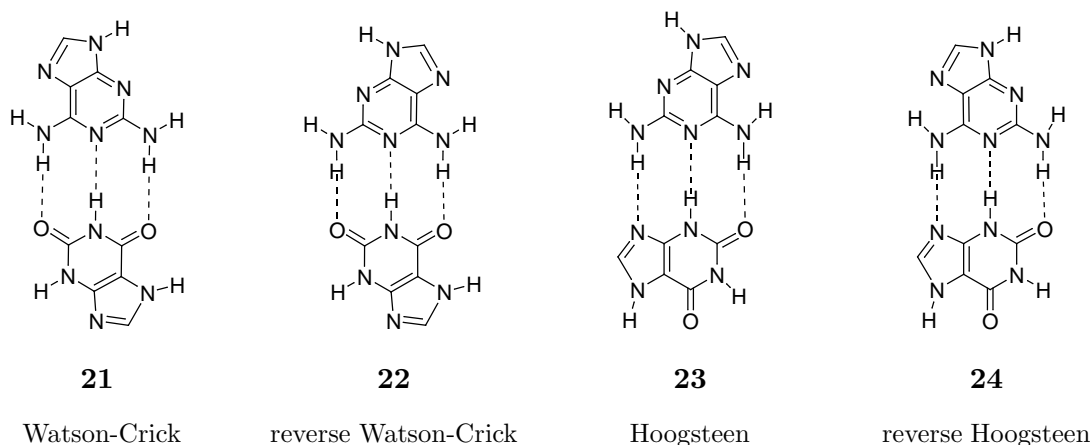


Fig. 4.9: Possible hydrogen bonded pairs of N7-xanthine with diaminopurine

basis sets.

Tab. 4.6: Dimerization energies E_{dim} of the pairings guanine-cytosine, adenine-thymine and xanthine-2,6-diaminopurine (XD). All energies in kcal/mol.

		E_{dim}			
		MP2/TZVPP	BLYP/TZVP	MP2/TZVPP	BLYP/TZVP
		$\epsilon = 1$	$\epsilon = 1$	$\epsilon = 78$	$\epsilon = 78$
19	guanine-cytosine	-25.7	-23.4	-14.1	-7.8
20	adenine-thymine	-14.4	-11.7	-6.4	-3.2
21	XD Watson-Crick	-15.7	-13.2	-6.8	-4.0
22	XD reverse Watson-Crick	-15.7	-13.1	-7.0	-4.0
23	XD Hoogsteen	-16.3	-13.7	-7.1	-4.1
24	XD reverse Hoogsteen	-16.5	-13.8	-7.4	-4.3

The 2,6-diaminopurine-xanthine base pairs have total binding energies of about -15.7 kcal/mol for dimers bound *via* the Watson-Crick site (**21** - **22**) and -16.3 kcal/mol and -16.5 kcal/mol for the Hoogsteen (**23**) and reverse Hoogsteen (**24**) basepairs. The corresponding BLYP-values computed in the TZVP basis set again are about 2 kcal/mol lower than the MP2-energies. Compared to the binding energy of the guanine-cytosine base pair **19**, the H-bonds for the 2,6-diaminopurine-xanthine pairings are rather weak and close to those of the two-dentate adenine-thymine pairing or the dimerization energies computed for the different self pairings of N7-xanthine. The weakness of the hydrogen bonds is due to repulsive secondary

interactions between (a) the protons, and (b) the nitrogen and oxygen atoms in the hydrogen bonding region. The explanation is similar to that found for the 2,6-diaminopurine-uracil pairing.^[130] In aqueous environment, all H-bonds are significantly weakened. Similar to the results obtained for N7-xanthine pairings, the MP2-method predicts a decrease of E_{dim} by a factor of 2 and BLYP a factor of 3.

Like the H-bonds, also the stacking interaction in table 4.7 for the guanine-cytosine tetramer of -12.1 kcal/mol pairing is higher than the stacking interaction for the adenine-thymine pairing of -10.2 kcal/mol. The 2,6-diaminopurine-xanthine pairing has the strongest stacking interaction with values of up to -15.9 kcal/mol. A dipolar environment leads to a strengthening of the 2,6-diaminopurine-xanthine stacking interaction of about 1 kcal/mol, the guanine-cytosine stacking interaction is insignificantly strengthened by 0.6 kcal/mol, while the adenine-thymine pairing remains unaffected.

Tab. 4.7: Stacking energies obtained from MP2/TZVP-computations. All energies in kcal/mol.

		MP2/TZVP	MP2/TZVP
		$\epsilon = 1$	$\epsilon = 78$
19	guanine-cytosine	-12.1	-12.7
20	adenine-thymine	-10.2	-10.2
21	xanthine-2,6-diaminopurine	-16.5	-14.3
22	xanthine-2,6-diaminopurine	-14.7	-15.9
23	xanthine-2,6-diaminopurine	-10.0	-10.5

4.1.5 Stability model for alanyl-PNA

To explain and predict the stabilities of alanyl PNA double strands, a model is set up, which might appear to be rather crude, however, it will be seen that it is able to predict the melting temperatures T_m of the alanyl PNA double strands quite accurately. It is assumed, that the total stabilization of the hexameric double strand is only composed of the energy contributions from hydrogen bonding E_{dim} and stacking E_{stack} . The sum of the latter gives

$E_{\text{H-b./stack}}$.

$$E_{\text{H-b./stack}} = 6 * E_{\text{dim}} + 5 * E_{\text{stack}} \quad (4.1)$$

The factors 5 and 6 simply count the number of the respective interactions (5 stacking interactions between 6 H-bonded dimers). The assumption that the double strand stabilization energy can be computed as the sum of hydrogen bonding and stacking only deliberately neglects, *e.g.* a major part of the backbone strain and the molecular effect of the aqueous environment. In the next step $E_{\text{H-b./stack}}$ is set equal to the ΔH of the dimer formation, and it is also assumed that the entropy change upon dimer formation is very similar for the different hexameric alanyl PNA compounds. This could be the case because of the similarity of the systems compared. The melting temperature is reached when

$$\Delta G = \Delta H - T_m \Delta S = 0. \quad (4.2)$$

If the assumptions made here are correct, the melting temperature correlates linearly with $E_{\text{H-b./stack}}$.

First the energy contributions from H-bonding and stacking are analyzed separately.

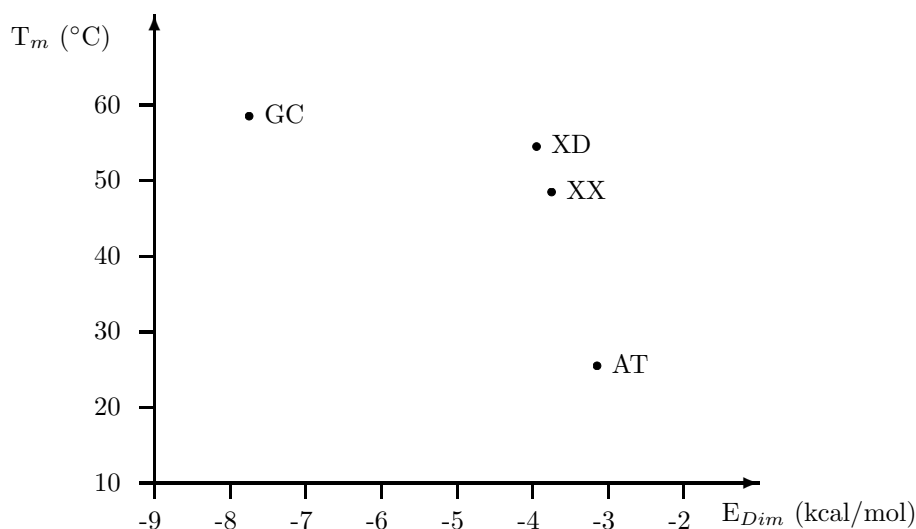


Fig. 4.10: Correlation of the H-bonds strengths ($E_{\text{dim}} * 6$ with the UV-vis melting temperature T_m of hexameric of alanyl-PNA of the sequence XXYXYYY or XXXXXX . AT denotes adenine-thymine, XX xanthine-xanthine, XD xanthine-diaminopurine and GC guanine-cytosine.)

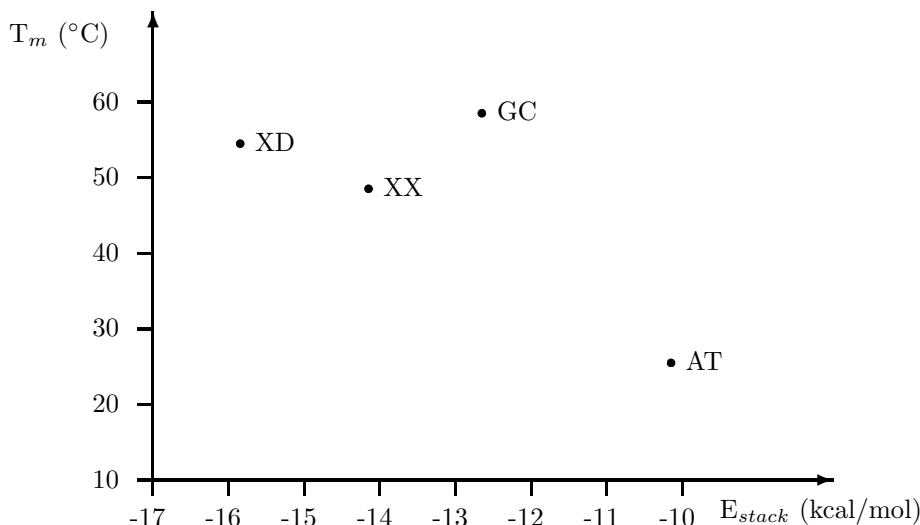


Fig. 4.11: Correlation of stacking energies with the melting temperature T_m of hexameric alanyl-PNA.

For more information see figure 4.10

Fig. 4.10 correlates the dimerization energies of H-bonded dimers of the four kalanyl-PNA systems studied so far with the melting temperatures of the respective hexameric alanyl PNA duplexes. For fig. 4.10, the energies computed with the BLYP functional in combination with the COSMO to mimic solvent effects were chosen. The reason to prefer BLYP over MP2 is that MP2, which is more reliable in the gas phase, in combination with the COSMO severely seems to overestimate the H-bond strengths. Experimentally determined H-bond strengths for nucleobases are usually between -5 and -7 kcal/mol^[131] for the gas phase, which is in good agreement with all computed values, while the contribution in aqueous environment is -0.5 to -2 kcal/mol.^[132-134] These values are in good agreement with the values computed by the BLYP method.

For the xanthine-xanthine (XX) and 2,6-diaminopurine-xanthine (DX) base pairs simply the pairing mode with the highest value for E_{dim} that can be realized within a hexameric double strand was chosen, *i.e.* the pairings **10** (XX) and **22** (DX). From the comparison of the sum over all hydrogen bonds with the melting temperatures of the adenine-thymine and the guanine-cytosine pairs in fig. 4.10, melting temperatures of around 30 $^{\circ}$ C for xanthine-xanthine and 35 $^{\circ}$ C for the 2,6-diaminopurine-xanthine pairs are predicted. This shows that

no good correlation based on hydrogen bonding only is possible for these pairings. If the stability of the alanyl PNA double strands were determined by hydrogen bonding only, the stability of both double strands would be much lower than found experimentally, which shows that this reduced form of the proposed model is too simplified to explain the T_m -values.

Correlating only the stacking energies (fig. 4.11), the melting temperature of the xanthine-xanthine alanyl-PNA would be around 70 °C and for the xanthine-2,6-diaminopurine pairing even higher. Also the stacking taken by itself is insufficient. However, if the sum over both H-bonding and stacking ($E_{Dim/H-bonds} * 6 + E_{stack} * 5$, the factors 5 and 6 count the numbers of interactions) is correlated with the T_m -values, a linear correlation is found (fig. 4.12). The correlation still needs to be tested for further systems, however, already at this point, it is quite surprising to find a linear correlation without using a single fitting parameter, because this means that effects of the entropy and the molecular water environment either cancel or are similar for all systems compared.

In this model, the stacking energies dominates by a factor of about 1 to 4 over the stabilization from H-bonding. This result is similar to what is assumed for DNA, where the stacking interaction is supposed to play a more important role in the stabilization than H-bonding.^[135,136]

This simple model shows that the high stability of the xanthine alanyl-PNA double strand, but also that of the tri-dentate xanthine-2,6-diaminopurine pairing is due to an increased stacking interaction that compensates the weakness of the hydrogen bonds. However, it also shows that the hydrogen bonding and base stacking interactions can indeed be called the determining factor for the stability of xanthine-Alanyl PNA²⁴.

²⁴Further computations:

Starting from the computations presented here, the studies concerning the alanyl PNA systems have been continued by B. Dietrich in the scope of his diploma thesis,^[137] and at the present time are part of the doctoral project of C. Sturm. Within these two projects, the full alanyl-PNA hexamers are described by force field methods, and the minima are determined by Monte Carlo and molecular dynamics approaches. Solvent effects are considered by explicit water molecules within a box of periodic boundary conditions. So far, these computations show that the stability of xanthine alanyl PNA might either be due to an increased stacking contribution of the hexamer built up from pairing mode **10** as it is described here, or because of the very efficient solvation of a hexamer that realizes Hoogsteen pairing mode **24**. Since this effect is due to the molecular water environment, it is not considered in the data presented here.

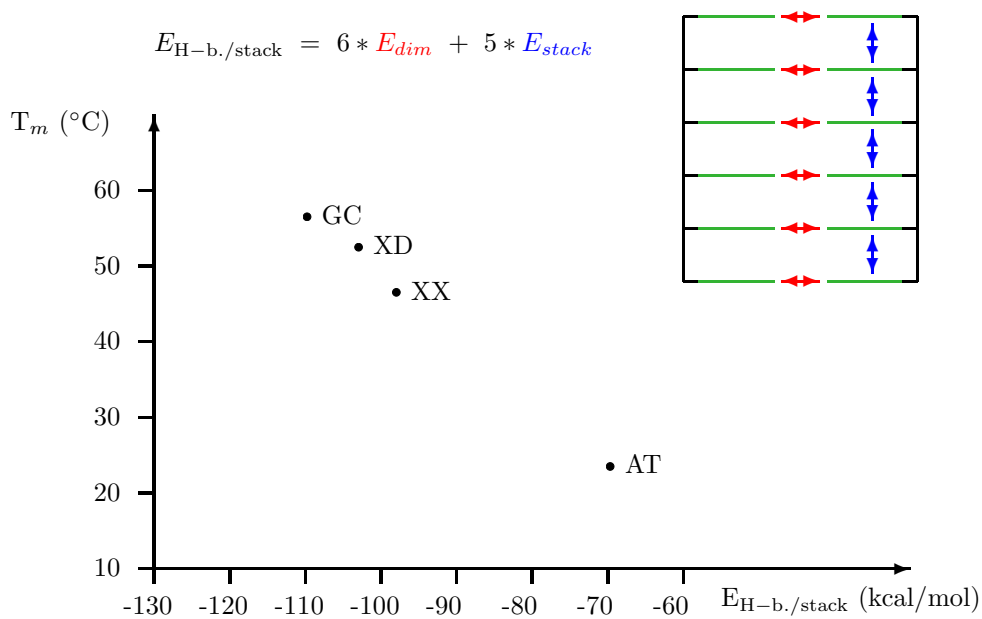


Fig. 4.12: Correlation of the stabilization energy $E_{\text{H-b./stack}}$ resulting from the sum over all H-bond and stacking interactions with the experimental melting temperatures (T_m) of hexameric xanthine alanyl PNA duplexes

4.2 Base pairing of N9-xanthine

The thermodynamically disfavored N9-regioisomer of xanthine plays a very important (unwanted) role in nature, since N9-xanthine is formed in DNA from guanine due to nitrogen loss and subsequent “mismatched” pairings of xanthine with canonical bases are one of the major causes of point mutations. The following sections study the base pairing properties of N9-xanthine.

4.2.1 Homodimers of N9-xanthine

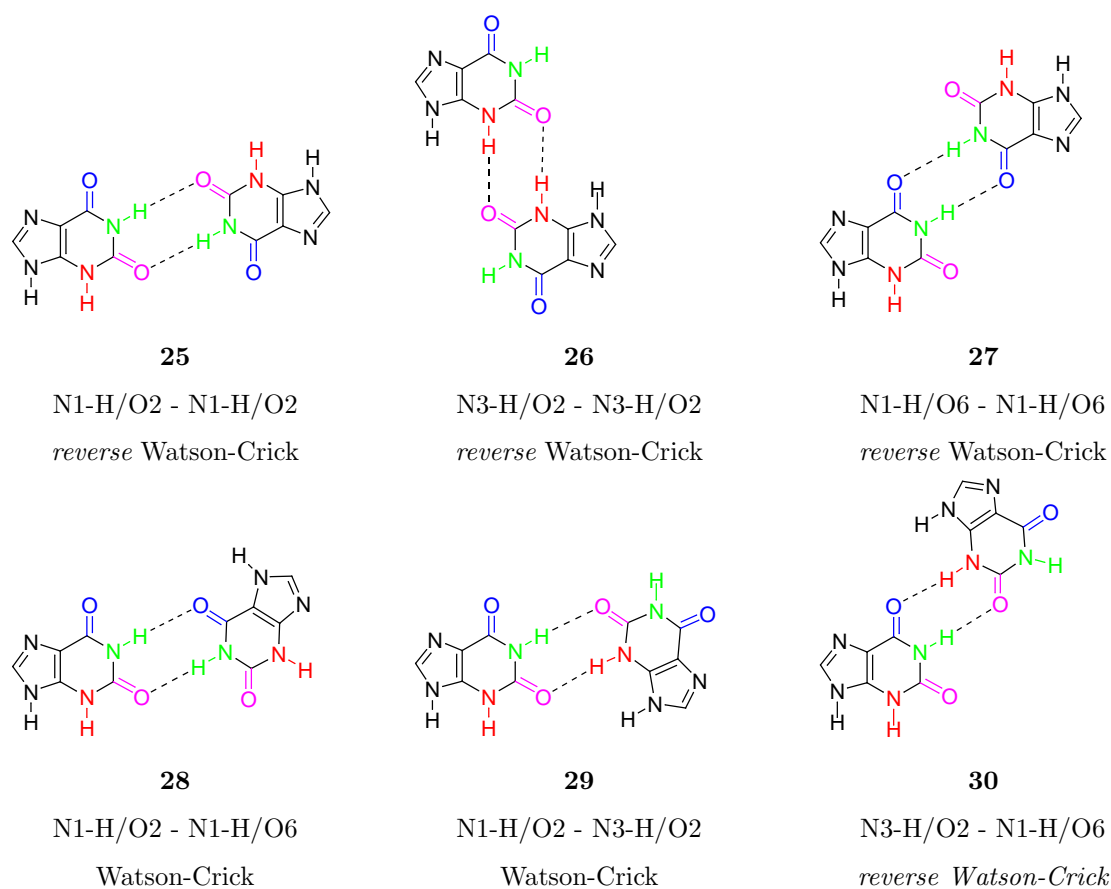
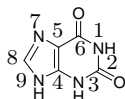


Fig. 4.13: Possible hydrogen bonded dimers of N9-xanthine

²⁵Numbering of atoms in xanthine:



In contrast to its N7-regioisomer, N9-xanthine can only form two-dentate base pairs *via* the Watson Crick site of the xanthine molecule. Scheme 4.13 shows the six possible two-dentate dimerization modes of diketo xanthine. The respective dimerization energies given in table 4.8²⁶ show remarkable differences regarding the strength of the individual hydrogen bonds. Next to the MP2 values, table 4.8 gives dimerization energies obtained with the B3LYP functional. Here, the B3LYP functional is used instead of the BLYP functional since the energies obtained with the B3LYP functional in the gas phase agree somewhat better with the MP2-values than energies obtained with the BLYP functional. While the dimerization energies computed by the MP2-approach for the reverse Watson-Crick pairs **25** and **27** and Watson-Crick pair **28** ($E_{dim} = -10.6, -10.9$ and -10.4 kcal/mol, respectively) lie in the lower range of the stabilization energies known for two-dentate base pairs, the dimerization energy of the strongest bonded base pair N3-H/O2 - N3-H/O2 (**26**) is twice as large ($E_{dim} = -22.0$ kcal/mol). Also given are the dimerization energies for polar environment with $\epsilon = 40$ and $\epsilon = 78$, computed with the B3LYP functional and the conductor like screening model. In polar solvent, the dimer formation is weakened by a factor of 3 to 4, but the relative order roughly remains the same. In section 4.2.5, the dimerization energies for the dielectric constant of $\epsilon = 40$ will be compared to experimental data.

Why does the hydrogen bond strength of the different pairings show this wide variation?

²⁶Computational details for sections 4.2 to 4.2.8:

The dimerization energies of N9-xanthine were computed using geometries that were optimized with the B3LYP functional in the 6-31++G(d,p) basis set.^[124] Minima were checked by frequency computations. The geometry optimization were performed with the Gaussian98 program package.^[125] Since bond lengths differed by less than 0.01 Å for regular and 0.03 Å for hydrogen bonds, the geometries for hypoxanthine and 4,5-dihydroxanthine were optimized with the BLYP functional in the TZVP^[92] basis set using the RI^[119] approximation which allows a much more efficient optimization. The dimerization energies were determined using the MP2 approach in the RI-approximation, and the B3LYP-functional^[57] in combination with the TZVPP basis set.^[92] The basis set superposition error (BSSE) was corrected according to the Boys-Bernardi counterpoise procedure.^[126] Solvent effects were estimated using the COSMO^[123] with a dielectric constant of $\epsilon = 40$ to simulate the freon solvent and $\epsilon = 78$ for aqueous environment. ¹H-NMR-spectra were computed using the HF and the MP2 approach in combination with the TZVP basis set. Natural bond orbital population analyses were performed using the B3LYP functional in combination with the 6-31G(d) basis set using the Gaussian98 program package.^[125] All computations except those for which it is stated otherwise were performed using the TURBOMOLE program package.^[64]

Tab. 4.8: Dimerization energies of xanthine pairs 1-6

pair	MP2/TZVPP	B3LYP/TZVPP	B3LYP/TZVPP	B3LYP/TZVPP
	$\epsilon = 1$	$\epsilon = 1$	$\epsilon = 40$	$\epsilon = 78$
25	-10.6	-8.8	-3.1	-2.8
26	-22.0	-20.5	-5.1	-4.5
27	-10.9	-9.6	-3.7	-3.3
28	-10.4	-8.8	-3.3	-2.9
29	-14.1	-13.0	-3.7	-3.5
30	-17.5	-15.7	-4.4	-4.3

The strength of the individual H-bonds seems to be connected with the hydrogen bond donor unit (N1-H or N3-H) involved, while both hydrogen bond accepting carbonyl groups seem to lead to similar bond strengths when they form a hydrogen bond with the same hydrogen donor unit. Dimers in which both hydrogen bonds involve the the N3-H donor unit (mode **26**) have the highest dimerization energy of -22.0 kcal/mol, followed **29** and **30** that involve one N3-H functionality (14.1 and 17.5 kcal/mol, respectively), while pairings that only involve the N1-H functionality as hydrogen bond donor contain hydrogen bonds that have a strength of -5 to -6 kcal/mol ($E_{dim} = -10.4$ to -10.9 kcal/mol), which is the standard H-bond strength of nucleobases in the gas phase.^[127]

Summarizing the computed results, the xanthine dimers are bound by two types of H-bonds with quite different bond strengths of around -5 and -10 kcal/mol, respectively (gas pase values). This means that in order to explain the variations in the dimerization energies of xanthine, it is sufficient to study only two representative examples in which only one of the two H-bond types is realized. Consequently, in the following discussion, only the pairing modes **25** and **26** will be discussed as representative examples.

4.2.2 Weak pairings of N9-xanthine

To understand the differing H-bond strengths of the xanthine dimers, simplified model systems were computed which allow a definite characterization of the relevant interactions. Table 4.9 gives dimerization energies for the amides **31-33** illustrated in fig. 4.14. The sim-

ple acetamide **31** has a dimerization energy of -14.4 kcal/mol, which is at the upper limit of the values of -5 to -7 kcal/mol normally found for hydrogen bonds in the gas phase.^[127] The bond strength correlates with the increased acidity of N-H bonds formed by sp^2 -nitrogen atoms since the higher s character allows a better stabilization of negative charges.^[138] Compared to the dimerization energy of the amide **31**, the dimerization energy of the dixanthine **25** is about 3.6 kcal/mol smaller, while that of **26** are increased by 7.6 kcal/mol.

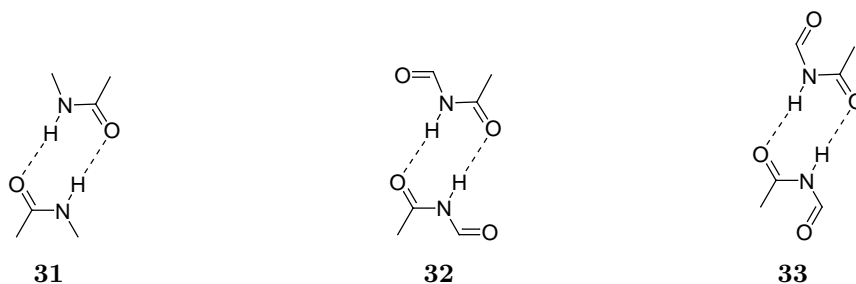


Fig. 4.14: Hydrogen bonded dimers of N-methylacetamide **31**, *cis* N-formylacetamide **32** and *trans* N-formylacetamide **33**

Here, it is discussed why the hydrogen bonds of **25** are weakened compared to **31**. The discussion for the high strength of the H-bonds in **26** is given in section 4.2.4. In **32** and **33**, the N-methyl group is replaced by a formyl group which either has *s-cis* or *s-trans* orientation relative to the N-H group. When the carbonyl group has *s-cis*-orientation (**32**), the dimerization energy is decreased by about 5 kcal/mol compared to that of **31**. The weakening of the hydrogen bonds is the result of secondary electrostatic repulsions between the H-bonded and the not H-bonded carbonyl-group, which is illustrated in fig. 4.15, where it can be seen that the H-bond angle deviates from the ideal value of 180° by about 8° and the distance of the H-bonded and the other oxygen atom is 3.95 Å. In addition to the *s-cis* stereoisomer, also the *s-trans* stereoisomer has been computed to separate the electronic effect due to the conjugation of the carbonyl group from the electrostatic effects. The *s-trans* conjugation in **33** leads to a slight decrease of E_{dim} of about 1 kcal/mol which is small compared to the decrease of E_{dim} caused by *s-cis* oriented formyl group in **32**.

For pairing mode **25** of xanthine, the repulsive secondary electrostatic effects are similar to those found in **32** since both compounds have similar local geometries. In line with the findings for **32**, also for the xanthine dimer **25** the NH-O-angle deviates by about 9° from the idealized H-bond angle of 180° which is shown in fig. 4.15. The H-bond lengths and the

Tab. 4.9: Dimerization energies the amides dimers **31-33**

pair	MP2/TZVPP	B3LYP/TZVPP	B3LYP/TZVPP
	$\epsilon = 1$	$\epsilon = 1$	$\epsilon = 78$
31	-14.4	-13.1	-3.6
32	-9.2	-7.9	-1.9
33	-13.1	-12.2	-2.9

distances of the H-bond participating and the non-participating oxygen atom agree within 0.1 Å to those of diamide **32**.

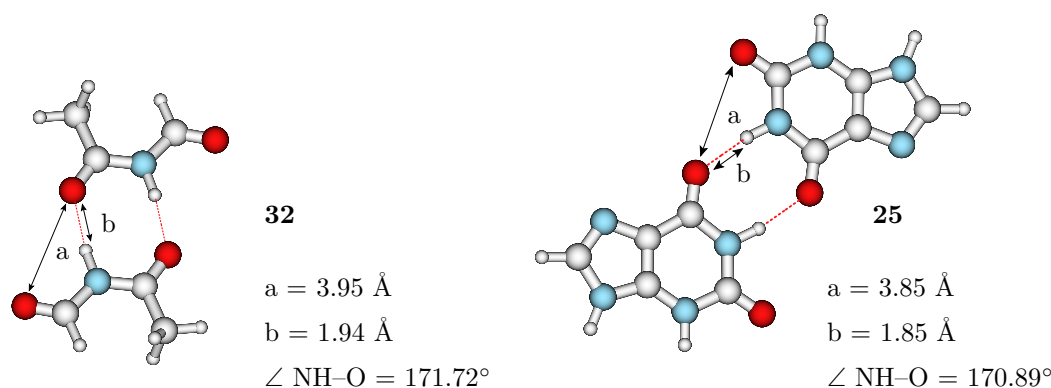
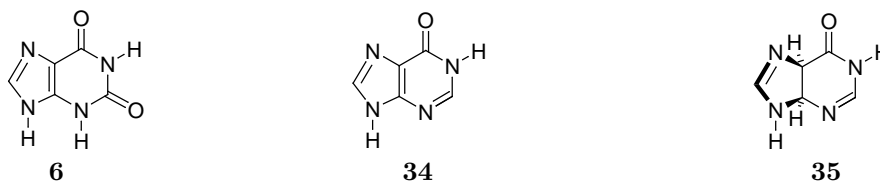


Fig. 4.15: Hydrogen bond lengths, -angles and the distances of the carbonyl groups of **32** and **25**

Summarizing, the lower hydrogen bond strength of dimers involving the N1 hydrogen atom is mainly due to unfavorable secondary electrostatic repulsions of the carbonyl groups. Similar electrostatic effects have been described for tri- and four-dentate base pairs and hydrogen bonded macromolecular systems.^[130,154]

4.2.3 Base pairing of hypoxanthine

In line with the arguments given in the last section, a replacement of the C2=O carbonyl group of xanthine by a functionality that only involves C-X bonds of little polarity, the dimerization energy of the base pair derived from **25** should be increased to values around -14 or -15 kcal/mol. One example is the non-canonical nucleobase hypoxanthine **34** (fig. 4.16), which has a C-H group next to the N1-H functionality. The only possible two-dentate dimerization mode **36** of hypoxanthine is shown in fig. 4.17. Table 4.10 gives the dimeriza-

Fig. 4.16: N9-xanthine **6**, hypoxanthine **34** and 4,5-dihydrohypoxanthine **35**

tion energies of dimer **36**. The dimer, having a dimerization energy of -18.6 kcal/mol (MP2), is about 8 kcal/mol more stable than the the corresponding base pair **25** of xanthine. The dimerization energy of **36** is much higher than the -14 to -15 kcal/mol that were expected from the comparison with N-methyl acetamide **31**.

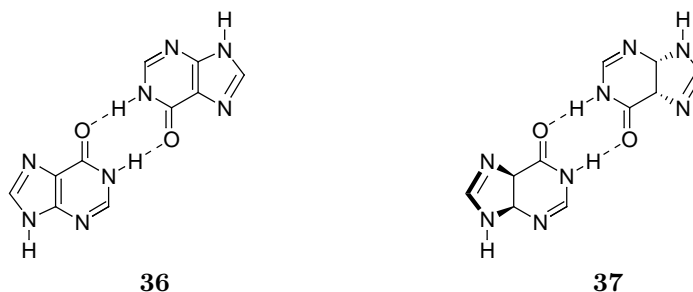


Fig. 4.17: Dimer of hypoxanthine and dihydrohypoxanthine

The energy computed using the B3LYP approach shows the same decrease. One possible explanation for the high strength of the hydrogen bond of hypoxanthine is that the dimer formation leads to a better aromatic stabilization of the pyrimidine ring (see resonance structures **34b** and **34c** of fig. 4.18). The size of the aromatic contribution to the hydrogen

Tab. 4.10: Dimerization energies for hypo- and 4,5-dihydrohypoxanthine

pair	MP2/TZVPP	B3LYP/TZVPP	B3LYP/TZVPP	B3LYP/TZVPP
	$\epsilon = 1$	$\epsilon = 1$	$\epsilon = 40$	$\epsilon = 78$
36 hypoxanthine	-18.6	-17.2	-5.0	-4.7
37 dihydrohypoxanthine	-15.8	-14.0	-3.9	-3.9

bonding can easily be determined:

If the central C4=C5-double bond of hypoxanthine is hydrated, the H-bonding functionalities (C6=O and N1-H) should not be distorted too much, but the aromatic resonance structures

34b and **34c** in fig. 4.18 are no longer possible, hence the aromaticity of the pyrimidine ring is lifted.

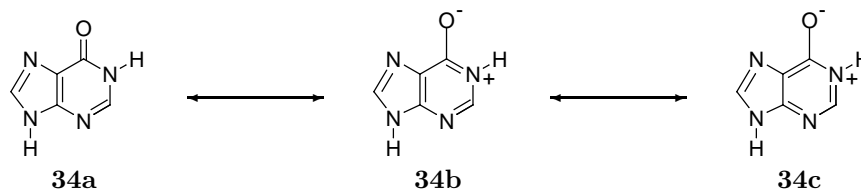


Fig. 4.18: Resonance structures of hypoxanthine

The dimerization energy for 4,5-dihydroxanthine **37** is -15.8 kcal/mol, which is comparable to that of N-methyl acetamide **31**. Subtracting E_{dim} of **37** from E_{dim} of **36** gives the aromatic stabilization of the dimerization formation that is -2.8 or -3.2 kcal/mol computed by the MP2- or B3LYP-method, respectively. In polar environment, the aromatic stabilization is still about -1 kcal/mol.

A short note:

The high dimerization energy of the H-bonded guanine-cytosine base pair, which leads to the high stability of the DNA double strand parts predominantly built from these base pairs, is usually explained by very favorable secondary electrostatic interactions.^[130]

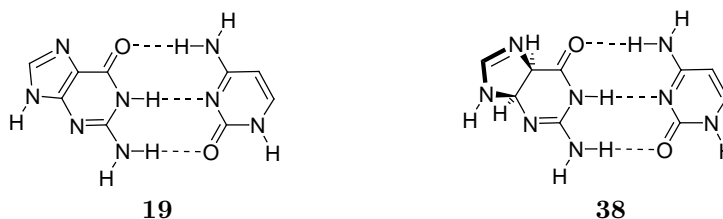


Fig. 4.19: Canonical guanine-cytosine, and 4,5-dihydroguanine-cytosine base pairs

However, next to the favorable secondary electrostatic interactions also in the case of the guanine-cytosine base pair, the H-bonds are strengthened due to an increased aromaticity for guanine. Table 4.11 that gives the dimerization energies for the guanine-cytosine and the non-aromatic 4,5-dihydroguanine-cytosine basepairs from fig. 4.19. For guanine-cytosine,

the aromatic stabilization is little more than -2 kcal/mol in the gas phase, and -1.2 kcal/mol in aqueous environment.

Tab. 4.11: Dimerization energies for the guanine-cytosine and 4,5-dihydroguanine-cytosine base pairs

		MP2/TZVPP	B3LYP/TZVPP	B3LYP/TZVPP
pair		$\epsilon = 1$	$\epsilon = 1$	$\epsilon = 78$
19	guanine-cytosine	-25.7	-24.4	-8.7
38	4,5-dihydroguanine-cytosine	-23.6	-21.9	-7.5

4.2.4 Strong hydrogen bonds

To understand which factors lead to the unusually high stability of pairing mode **26**, the base pair is compared to the corresponding dimer of uracil **39** and to the base pair **10** of the regioisomer N9-xanthine (see fig. 4.20). Table 4.12 compares the dimerization energy of diuracil **39** to that of pair **26** of N9-xanthine, and also to the dimerization energy of the corresponding base pair **10** of N7-xanthine. The dimerization energy for **39** of -17.2 kcal/mol is about 3 kcal/mol higher than that of the diamide **31**. The higher stability of diuracil **39**

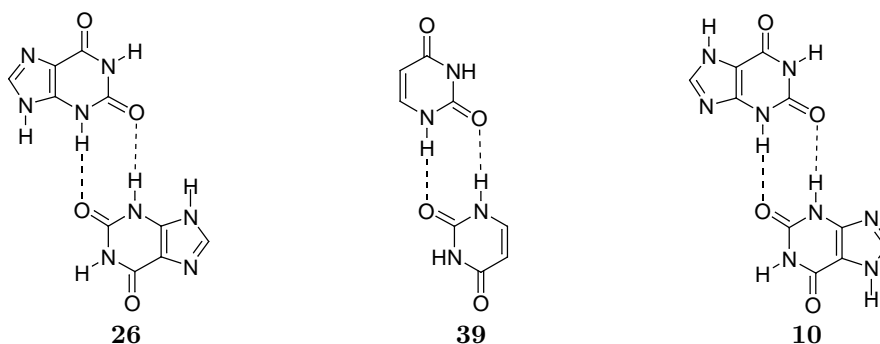


Fig. 4.20: Dimers of N9-xanthine **26**, uracil **39** and N7-xanthine **10**, that only hydrogen bonds with the N3-H donor functionality.

has already been studied in the literature, where similar dimerization energies have been obtained,^[139] **39** has been recomputed here to have an identical computational level for the comparison with the xanthine pairings.

Interestingly, depending on how the histidine ring is annelated, the N3H-O hydrogen bond is either reinforced or weakened. In line with the arguments given in section 4.2.2, the

Tab. 4.12: Comparison of dimerization energies of dixanthine **26** with diuracil **39** and of dixanthine **39**, the N7-regioisomer of **26**.

		MP2/TZVPP	B3LYP/TZVPP	B3LYP/TZVPP
pair		$\epsilon = 1$	$\epsilon = 1$	$\epsilon = 78$
26	N9-xanthine	-22.0	-20.5	-4.5
39	uracil	-17.2	-15.9	-4.3
10	N7-xanthine	-14.1	-12.2	-3.7

annulation of the histidine ring either places a positively charged hydrogen atom next to the N3-H group, which stabilizes the the hydrogen bond since the positive partial charge attracts the oxygen atom, or it places the negative partial charge of the nitrogen lone pair next to the N3-H group which repels the carbonylic oxygen atom of the complementary base.

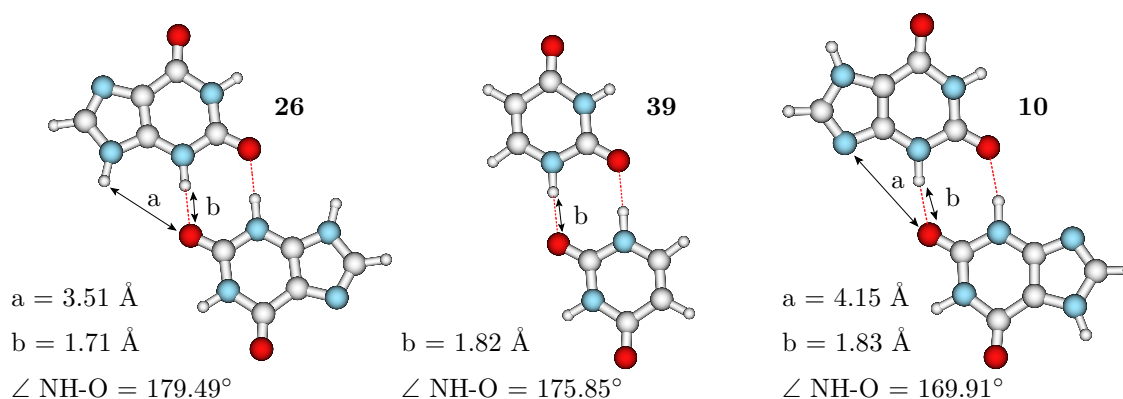


Fig. 4.21: Hydrogen bond lengths, -angles and the distances of the carbonyl groups of **26**, **39** and **10**

This again is confirmed by the geometric characteristics of the H-bonds. The H-bond angle N3H-O is 169.91° for the N7-regioisomer (pairing **10**). In the case of pairing **26** of the N9-regioisomer, the angle is 179.49° , while for the uracil pairing a value of 175.85° has been computed. For **39**, the deviation of the N3H-O-angle from the idealized value of 180° cannot be explained based on the computations given here.

Since the stabilization of the H-bonds of **26** and the destabilization of those of **10** is mainly due to secondary electrostatic interactions, the difference in the dimerization energies is strongly decreased in aqueous environment. The B3LYP computation using the COSMO predicts that for $\epsilon = 78$ the dimerization energies differ by only 0.8 kcal/mol, compared

to about 8 kcal/mol in the gas phase. In contrast, the stabilization from the increase of aromaticity for hypoxanthine is decreased by only a factor of 3.5 from -2.8 kcal/mol to -0.8 kcal/mol. The comparison shows that electronic effects are less affected by the solvent environment than electrostatic effects and for systems in very polar solvents like water always need to be taken into consideration.

This analysis shows that the stability of the xanthine-dimer **26** is mainly due to favorable secondary electrostatic interactions in addition to the electronic effect of the C4=C5 double bond. In contrast, the H-bonds of pairing **25** are weakened due to repulsive secondary electrostatic interactions. For the pairings **27** and **28** that were not explicitly analyzed here, the explanation for the weakness of the hydrogen bonds is in line with that of **25**, for the pairings **29** and **30**, stabilizing and destabilizing effects for the individual hydrogen bonds have to be summed up.

4.2.5 Comparison with experiment

Xanthine

An evaluation of the predictions was possible by analyzing the dimer xanthosine, which is the ribosyl derivative of xanthine, and inosine, the ribosyl derivative of hypoxanthine, using low temperature $^1\text{H-NMR}$ -techniques. The experimental work was performed by Weitz and Basilio.^[140] The advantage of low temperature NMR over room temperature NMR is that the isolated base pairs can be identified, because no fluctuation over several structures takes place. To have a better correspondence with the experimental conditions, the dimerization energies have also been computed for solvent environment using the B3LYP functional in combination with the conductor like solvation model using $\epsilon = 40$. The results have already been given in table 4.8 for xanthine and in table 4.10 for hypoxanthine. In polar solvent all hydrogen bonds are weakened by a factor of 3 to 4. Pairing mode **25** remains most stable with a dimerization energy of -5.1 kcal/mol, mode **30** has an energy of -4.4 kcal/mol, and the dimerization energies of the pairings **25**, **27**, **28** and **29** are between -3.1 and -3.7 kcal/mol. This comparison shows that the pairing modes **26** and **30** remain unusually strong also in polar environment. For hypoxanthine, a high stabilization energy of -5.0 kcal/mol is com-

puted also for the solvent environment.

The experimental $^1\text{H-NMR}$ spectrum measured at -126°C in freon solution²⁷ shows a very good agreement with the $^1\text{H-NMR}$ spectrum that was computed with the HF-approach in a TZVP basis set²⁸ for pairing mode **28**. The chemical shifts for the two N1-hydrogen atoms of the computed $^1\text{H-NMR}$ -spectrum and those for the C8-hydrogen atom differ no more than 0.16 ppm from the experimental values. One exception is the N3-hydrogen atoms for which chemical shifts of 6.69 and 6.59 ppm are computed. These values strongly deviate from the measured chemical shift of 10.69 ppm for the two hydrogen atoms. The explanation for this deviation is the formation of an internal hydrogen bond within the xanthosine molecule as it is illustrated in the fig. 4.22 between C6-oxygen atom of the ribosyl sugar and the N3-hydrogen atom. The computed chemical shift for the N3-hydrogen atom of inosine **40** of 10.30 ppm²⁹ is in much better agreement with the experimental value of 10.69 ppm.

²⁷Measured $^1\text{H-NMR}$ spectrum (freon):

N1H 11.88 ppm (1 H); N1H 11.68 ppm (1 H); N3H 10.69 ppm (2 H); C8H 7.47 ppm (2 H), H1' (5,67 ppm), H4'/H2'/H3' (4,21 ppm, 4,17 ppm, 4,05 ppm).

²⁸Computed $^1\text{H-NMR}$ spectrum (HF/TZVPP):

25 N1H 11.51 ppm (2 H); (N9H 7.76 ppm (2 H)); C8H 7.48 ppm (2 H); N3H 6.74 ppm (2 H).

26 N3H 13.38 ppm (2 H); (N9H 8.25 ppm (2 H)); C8H 7.59 ppm (2 H); N1H 7.39 ppm (2 H).

27 N1H 12.41 ppm (2 H); (N9H 7.83 ppm (2 H)); C8H 7.42 ppm (2 H); N3H 6.69 ppm (2 H).

28 N1H 11.86 ppm (1 H); N1H 11.84 ppm (1 H); (N9H 7.74 ppm (1 H); N9H 7.71 ppm (1 H)); C8H 7.50 ppm (1 H); C8H 7.48 ppm (1 H); N3H 6.69 ppm (1 H); N3H 6.59 ppm (1 H).

29 N1H 12.16 ppm (1 H); N3H 12.12 ppm (1 H); (N9H 8.02 ppm (1 H); N9H 7.80 ppm (1 H)); C8H 7.52 ppm (1 H); C8H 7.47 ppm (1 H); N1H 7.35 ppm (1 H); N3H 6.63 ppm (1 H).

30 N3H 13.27 ppm (1 H); N1H 12.46 ppm (1 H); (N9H 8.27 ppm (1 H); N9H 7.80 ppm (1 H)); C8H 7.53 ppm (1 H); C8H 7.52 ppm (1 H); N1H 7.40 ppm (1 H); N3H 6.73 ppm (1 H).

²⁹Computed $^1\text{H-NMR}$ spectrum (HF/TZVP):

40 N3H 10.30 ppm (1 H); N1H 7.31 ppm (1 H); C8H 7.59 ppm (1 H); 7.40 ppm (1 H); N3H 6.73 ppm (1 H); H1' 5,27 ppm (1 H), H2' 4.25 ppm (1 H), H3' 4.15 ppm (1 H), H4' 4.08 ppm (1 H), H5'a 4.04 ppm (1 H), H5'b 3.69 ppm (1 H).

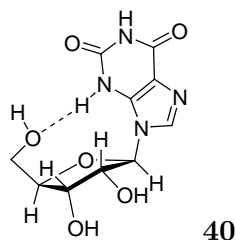


Fig. 4.22: Internal hydrogen bond of xanthosine

Hypoxanthine

As expected, a rather high association constant of 200 M^{-1} has been experimentally found for hypoxanthine.^[140] For the dimer of hypoxanthine, the computation predicts a chemical shift of 14.61^{30} ppm that is even stronger than the chemical shift predicted for pairing mode **26** of xanthine, even though in the computations the hydrogen bonds of **26** were even slightly stronger. In this case, the measured value for the N3-Hydrogen bond is somewhat lower than the computed chemical shift (13.87 ppm).

4.2.6 Mutagenicity

It has been explained in chapter 3 that N9-xanthine and hypoxanthine are formed in DNA due to nitrogen loss from guanine and adenine. This process can for example be induced by the reaction of the canonical nucleobases with nitric acid or free radicals such as OH or NO. Both lesions are considered to be highly mutagenic.^[109] Point mutations are found if xanthine forms base pairs with thymine, and hypoxanthine with cytosine (mut). Pairings of xanthine to cytosine also take place, but they are not mutagenic (non-mut). How stable are such pairings?

In principle, a very large number of base paired combinations of xanthine and hypoxanthine with canonical bases is thinkable, however, already the DNA-topology limits the choice of possible base pair combinations to the Watson-Crick pairing mode of purine with pyrimi-

³⁰Computed ¹H-NMR spectrum (HF/TZVP):

40 N1-H 14.61 ppm (2H), (N9-H 8.80 ppm (2H),) C2-H 8.65 ppm (2H), C8-H 7.67 ppm (2H)

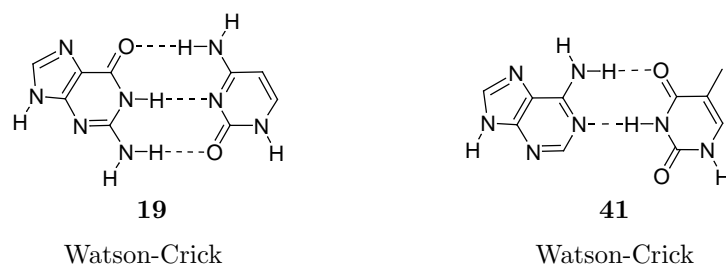


Fig. 4.23: Canonical base pairs guanine-cytosine (**19**) and adenine-thymine (**41**)

dine bases. Furthermore, the tight binding pocket of DNA-polymerase enforces a very strict geometric selection of the substrate base pair,^[98–100] so that the base pairs to a large extent need to resemble the canonical guanine-cytosine (**19**) and adenine-thymine (**41**) base pairs illustrated in fig. 4.23. Starting from this information, in the following sections only Watson-Crick base pairs of xanthine and hypoxanthine with pyrimidine bases are taken into consideration. The pairings are classified into those that have the exact backbone geometry of a canonical Watson-Crick base pair, which will be named Watson-Crick basepairs and those, for which only a distorted Watson-Crick pairing mode is possible, are denoted distorted Watson-Crick. For these pairings, it is assumed that, as long as the deviation from the Watson-Crick geometry is small, the incorporation of the base pair into DNA is still possible, even though the rate is significantly lowered, which also means that the incorporation can be enhanced by very favorable interactions such as strong hydrogen bonding. This assumption is in good agreement with the finding that the incorporation efficiency of the guanine-thymine mismatch base pair is significantly lower than that of the canonical base pairs, but still much higher than that of all other mismatched pairs of canonical bases.^[100] The classification is illustrated in fig. 4.24 for pairings of hypoxanthine. The pairing of

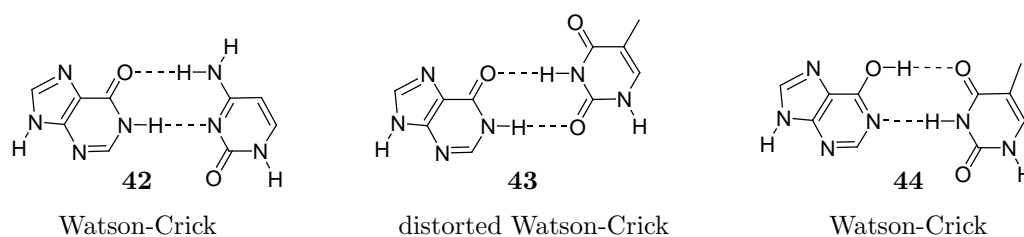


Fig. 4.24: Mutagenic base pair of hypoxanthine and cytosine (**42**), the non-mutagenic base pair of hypoxanthine and thymine (**43**), and a hypoxanthine-thymine base pair involving a tautomeric form of hypoxanthine (**44**)

hypoxanthine and cytosine **42** represents a Watson-Crick base pair, because both the purine and the pyrimidine base of **42** can exactly be superimposed to the purine and pyrimidine bases of **19** or **41** given in fig. 4.23, respectively. The hypoxanthine-thymine pairing **43** represents a distorted Watson-Crick pairing mode, since the thymine ring is shifted compared to that of the canonical adenine-thymine base pair **41** (red), as it can be seen from figure 4.25. The pairing **44** that involves the enol tautomer of hypoxanthine again has Watson-Crick pairing mode.

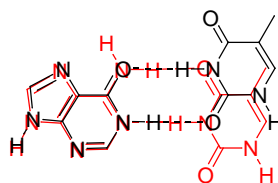
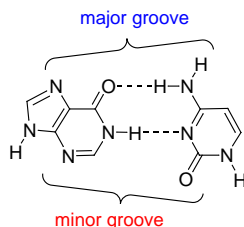


Fig. 4.25: Overlay of the non-mutagenic base pair of hypoxanthine and thymine (**43**) (black) and the canonical base pair **41** of adenine and thymine (red) to illustrate the geometric deviation of pairing **43** from the exact Watson-Crick geometry

4.2.7 Pairings of hypoxanthine to canonical pyrimidine bases

Table 4.13 gives the dimerization energies for the pairings **42** to **44**. Base pair **42** corresponds to an adenine-cytosine mismatch, and is therefore mutagenic. It has a very large dimerization energy of -20.2 kcal/mol computed by the MP2 approach. The B3LYP value is -18.1 kcal/mol, which is reduced to -5.0 kcal/mol if the aqueous environment is taken into account. Next to this relatively high stability, pairing **42** also perfectly mimics a canonical Watson-Crick base pair since its major groove³¹ is identical to that of the guanine-cytosine

³¹Major and minor groove:



Watson-Crick pairing, while its minor groove imitates an adenine-thymine pairing, which leads to a kinetically favored template effect.^[142] The high stability of base pair **42** is (a)

Tab. 4.13: Binding energies for the canonical base pairs and for the pairing of hypoxanthine to canonical pyrimidine bases in kcal/mol. For pairings that involve tautomeric forms of xanthine, the dimerization energies refer to the formation starting from keto-hypoxanthine. WC denotes that the pairing has the exact geometry of a canonical Watson-Crick base pair, dist refers to a distorted Watson-Crick pairing mode. In column mut (mutagenicity), non-mut denotes a pairing that does not lead to a point mutation in the complementary strand, while a pairing labeled mut leads to a point mutation.

pair	mode	mut	MP2/TZVPP	B3LYP/TZVPP	B3LYP/TZVPP
			$\epsilon = 1$	$\epsilon = 1$	$\epsilon = 78$
19 guanine-cytosine	non-mut	WC	-25.7	-24.9	-8.7
41 adenine-thymine	non-mut	WC	-13.8	-11.5	-3.6
42 hypoxanthine-cytosine	mut	WC	-20.2	-18.2	-5.0
43 hypoxanthine-thymine	non-mut	dist	-13.5	-11.9	-3.7
44 hypoxanthine-thymine	non-mut	WC	-12.3	-9.0	+2.9

due to favorable secondary electrostatic interactions between the N3-hydrogen atom and the carbonyl group, and (b) due to the aromatic stabilization of the bonding for hypoxanthine (see fig. 4.18). Both effects have been intensively discussed in the previous sections. These results, showing a very high stability of the hypoxanthine-cytosine pairing, are in good agreement of the experimental data concerning the inosine-cytosine base pair.^[143-145]

The non-mutagenic pairing of hypoxanthine with thymine **43**, which corresponds to the canonical adenine-thymine pairing is only possible either in a distorted Watson-Crick geometry (**43**) or in a pairing involving tautomeric forms (**44**). The dimerization energies in table 4.13 show that **43**, having a dimerization energy of -13.5 or -11.9 kcal/mol in the gas phase (MP2 and B3LYP, respectively) and -3.7 kcal/mol in solvent environment, represents a stable pairing, even though it is weaker than the hypoxanthine-cytosine pairing. In DNA, **43** is disfavored by its distorted Watson-Crick geometry.

If the geometric restrictions of DNA are lifted, like it is the case for *t*RNA, the uracil analogue of pairing **43** is indeed formed. Since the lifting (“wobbling”) of the topological restrictions of DNA is the precondition for the formation of this base pair, the corresponding theory

has been named the “wobble theory”.^[146] The pairing **44** which would possess the right Watson-Crick geometry, but involves a tautomeric form of hypoxanthine, is disfavored in polar environment.

4.2.8 Pairings of neutral xanthine to canonical pyrimidine bases

The diketo form of xanthine can only realize one single Watson-Crick base pairing mode with cytosine, which is depicted in fig. 4.26, because the alternative distorted Watson-Crick base pair **46** involving the N3-hydrogen atom of xanthine can be excluded because of the steric hindrance shown in fig. 4.26 and since its geometry deviates too strongly from the exact Watson-Crick geometry.

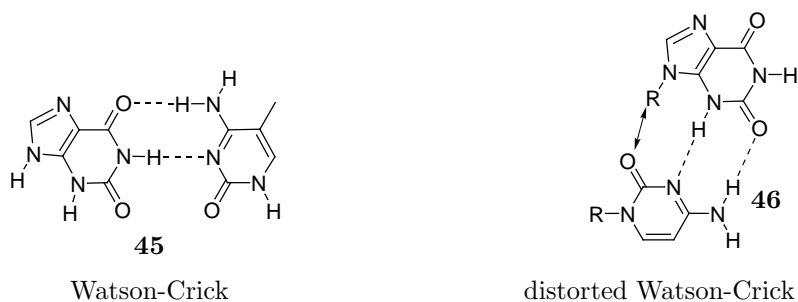


Fig. 4.26: Pairings of xanthine and cytosine

Since the two carbonyl groups repel each other in **45**, xanthine is rather weakly bound to cytosine. MP2 predicts a binding energy of -12.3 kcal/mol for **45**, B3LYP -10.3 kcal/mol for the gas phase and -2.3 kcal/mol for aqueous environment.

Furthermore, it needs to be considered, that **45** is twisted by 30°, which is shown in fig. 4.27.

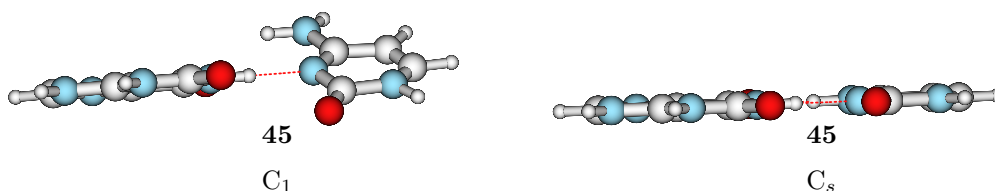
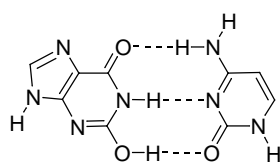


Fig. 4.27: Twisted (C_1) and planar (C_s) structure of xanthine-cytosine pairing **45**. The twist-angle of the molecular planes in the C_1 -structure is 30.01°

The planarization, which allows a better stacking interaction, costs about 0.5 kcal/mol. The xanthine-cytosine base pair **45** has no H-bond accepting functionality in N3-position. It is known that H-bond donor groups in the active site of xanthine polymerase bind to the N3-acceptor site of purine (or the C2=O carbonyl group of pyrimidine bases) in the minor groove of the newly formed base pair,^[147–150] however, it could be, that the neighboring C2=O carbonyl functionality instead of the N3 atom serves as acceptor site for DNA polymerase.

**47**

Watson-Crick

Fig. 4.28: Watson-Crick pairing of tautomeric xanthine to cytosine

Another stable Watson-Crick base pair of xanthine and cytosine is obtained if xanthine is replaced by its enol 1-H-2-enol-6-keto-xanthine (see table 4.14). Even though the tautomerization energy for the rotamer **48b** of 10.5 kcal/mol ($\Delta G = 11.6$ kcal/mol) is rather high, the base pair has a very high dimerization energy (with respect to the formation out of diketo-xanthine) in the gas phase of -22.8 kcal/mol (MP2). **47** profits from very favorable secondary interactions between the N1-hydrogen atom and the cytosine carbonyl group, and from an aromaticity contribution like it has been discussed for the hypoxanthine dimer in section 4.2.3.

Tab. 4.14: Energies in kcal needed to tautomerize N7-H xanthine to 1-H-2-enol-6-keto-xanthine.

	48a			48b		
	ΔE	ΔG	ΔE	ΔE	ΔG	ΔE
	$\epsilon = 1$	$\epsilon = 1$	$\epsilon = 78$	$\epsilon = 1$	$\epsilon = 1$	$\epsilon = 78$
MP2/TZVPP	1.9	0.8	-	8.4	9.5	-
B3LYP/TZVPP	3.7	2.6	11.1	10.5	11.6	14.3

The B3LYP value of -18.5 kcal/mol underestimates the dimerization energy by 4.3 kcal/mol, which is mainly due to the overestimation of tautomerization energy by the B3LYP functional by about 2 kcal/mol. In polar solvent, the tautomerization is strongly disfavored, so that **47** is predicted to be only slightly more stable than **45**.

Tab. 4.15: Binding energies for xanthine to canonical bases in kcal/mol. For pairings that involve tautomeric forms of xanthine, the dimerization energies refer to the formation out of diketo-xanthine. The columns mode and mut are defined as described in table 4.13. For nonplanar base pairs (C_1) also the energies of the planar structures (C_s) are given.

				MP2/TZVPP	B3LYP/TZVPP	B3LYP/TZVPP	
pair		mut	mode	sym	$\epsilon = 1$	$\epsilon = 1$	$\epsilon = 78$
45	xanthine-cytosine	non-mut	WC	C_1	-12.3	-10.3	-2.3
				C_s	-11.5	-9.9	-1.8
47	xanthine-cytosine	non-mut	WC	C_s	-22.8	-18.5	-2.5
50	xanthine-thymine	mut	dist	C_s	-11.0	-9.2	-3.7
51	xanthine-thymine	mut	dist	C_s	-10.9	-9.4	-3.7
52	xanthine-thymine	mut	WC	C_s	-13.5	-8.5	+6.4
53	xanthine-thymine	mut	WC	C_1	+3.8	+7.5	+11.4
				C_s	+4.1	+7.7	+11.4
54	xanthine-thymine	mut	WC	C_1	-1.3	+0.5	+7.5
				C_s	-1.2	+0.6	+7.6

So far, it has not been considered that xanthine has a pK_a -value of 5.7^[153] and is deprotonated at physiological pH (fig. 4.29, for computed pK_a -values see also Rogstad *et al.*^[117]). Having a pK_a -value of 5.7, the deprotonation at physiological pH of 7.4^[146] is exothermic with $\Delta G = -2.3$ kcal/mol³². This value needs to be subtracted from the dimerization energies in table 4.15 since the involved deprotonated species are higher in energy than the

³²The free energy for the deprotonation:



is computed as:

$$\Delta G = 2.303RT(pK_s - pH). \quad (4.4)$$

For the derivation of eqn. 4.4 see appendix A.

deprotonated global minimum. As a consequence, **45** and **47** are only weakly or not bound at all, and should *e.g.* not be stable as free base pairs in aqueous solution, if also the entropy is considered, that is negative for the dimer formation³³. The low affinity of xanthine to cytosine should decrease, but not inhibit the incorporation into DNA, since the pairing is still favored geometrically, which is the main precondition for an incorporation.^[99]

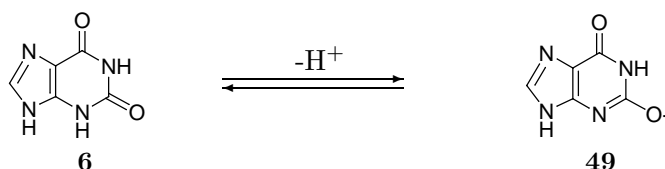


Fig. 4.29: Deprotonation of xanthine

To cause a point mutation, xanthine has to pair with thymine. Such pairings are possible in the distorted Watson-Crick geometries **50** and **51** (Fig. 4.30). Both seem to represent stable pairings in gas phase and in aqueous environment even if the -2.3 kcal/mol are subtracted to take into account the acidity of N9-xanthine (table 4.15). However, this stability is only about 1 kcal/mol higher than that of **50** and **51**, which should be too small to compensate the geometrical deviation of **50** and **51** from the exact Watson-Crick pairing geometry.

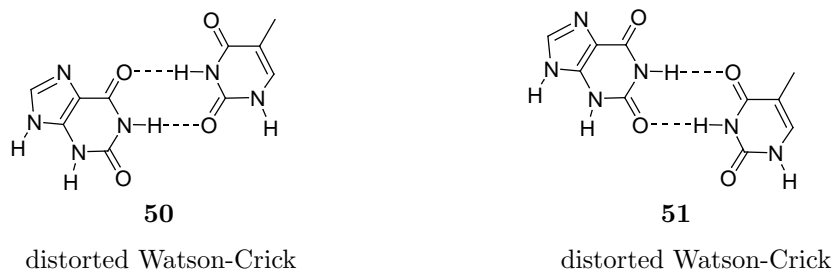


Fig. 4.30: Pairings of xanthine and thymine

Within an “exact” Watson-Crick geometry only pairings **52**, **53** and **54** involving tautomers of **49** can be realized for the xanthine-thymine combination. However, such pairings are unstable either already in the gas phase (**53**) or clearly unstable in aqueous environment (**52**)

³³As explained in footnote 21, in the computations shown here the entropy is not given, because a correct computation of the entropy requires the consideration of explicit water molecules which is not possible in the scope of the ab initio computations.

and **54**).

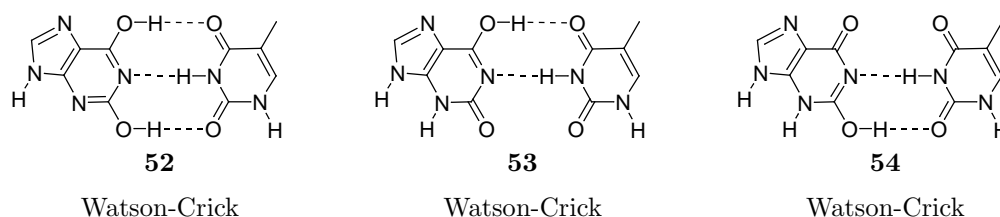


Fig. 4.31: Pairings of xanthine and thymine involving tautomeric forms of xanthine

Summarizing, neutral N9-xanthine is able to form stable Watson-Crick pairs with cytosine, but not with thymine. Stable xanthine-thymine pairings are only possible in distorted Watson-Crick geometries, however, these pairings do not have an H-bond accepting functionality in N3-position. As a consequence, neutral xanthine cannot lead to the incorporation of mutagenic xanthine-thymine base pairs into the DNA double strand.

4.2.9 Anionic pairings of xanthine with canonical bases

In the next step, also the anionic base pairings of xanthine are taken into consideration. Since anion properties are largely determined by solvent stabilization, only computations³⁴ considering a polar environment are discussed here, while gas phase data for anions can hardly be compared to properties of anions in polar solvent. Figure 4.34 summarizes the computed dimerization energies. The pairing of deprotonated xanthine to cytosine **55** (fig. 4.32) has a dimerization energy of around zero. The energy stabilization from the H-bonds

³⁴Computational details for section 4.2.9:

GGA-functionals have difficulties in the description of anions due to the self-interaction error.^[35] This drawback is partially remedied in the B3LYP functional due to the admixture of Hartree-Fock exchange. Consequently, for the optimization of anionic dimers, the B3LYP-functional was used in combination with the TZVP basis set^[92] that was extended by diffuse primitive gaussian functions with the exponents 0.040 (s- and p-type) for carbon, 0.053 (s- and p-type) for nitrogen and 0.068 (s- and p-type) for oxygen,^[151] the basis set will be abbreviated as TZVP(+). Dimerization energies were computed using the B3LYP functional in combination with the TZVPP basis set^[92] extended by the diffuse primitive gaussian functions given above (denoted as TZVPP(+)). For all computations the solvent effect was taken into account by the Conductor like solvation model.^[123] All computations were performed using the TURBOMOLE program package.^[64]

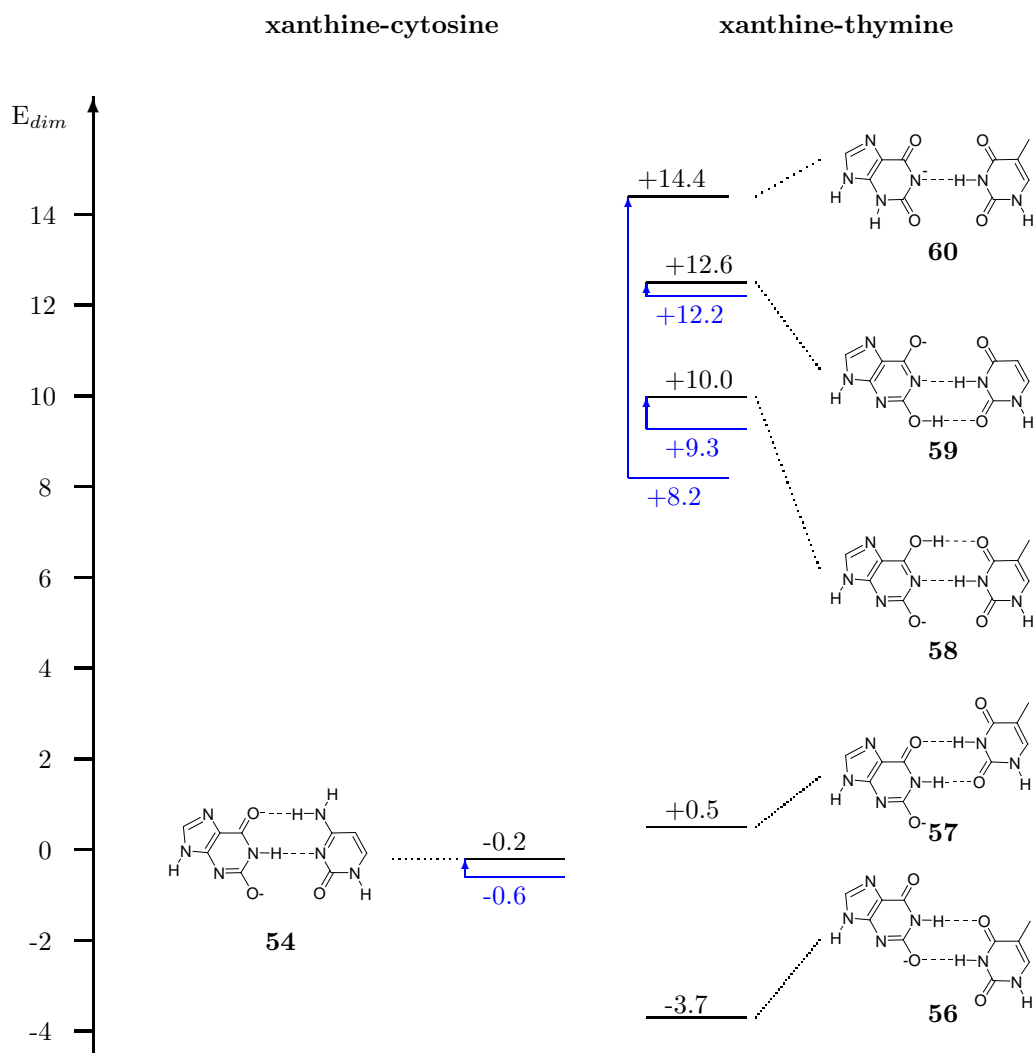


Fig. 4.34: Dimerization energies of the anionic Watson-Crick pairs of xanthine and cytosine or thymine (B3LYP/TZVPP(+); $\epsilon = 78$). The blue lines represent the dimerization energies of the nonplanar structures of **54**, **58**, **59** and **60**. Energies in kcal/mol.

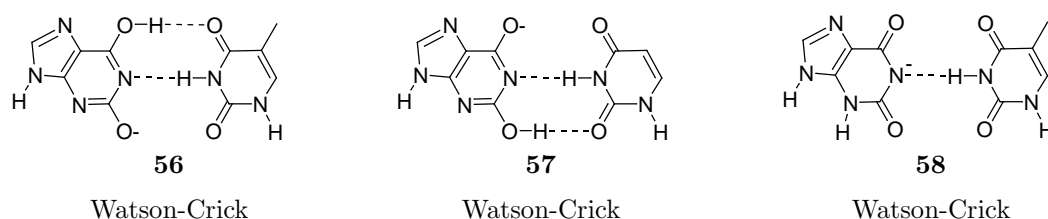


Fig. 4.35: Pairings of xanthine and thymine involving anionic tautomeric forms

The pairings **58**, **59** and **60** (fig. 4.35) involving tautomeric forms of the N3C2-O enolate of xanthine **49** have not been found to represent stable formations. Since pairing **60** has no second H-bond that fixes the dimer in a planarized geometry, its minimum structure is twisted so that the carbonyl groups have a maximum distance as it is shown in fig. 4.36.

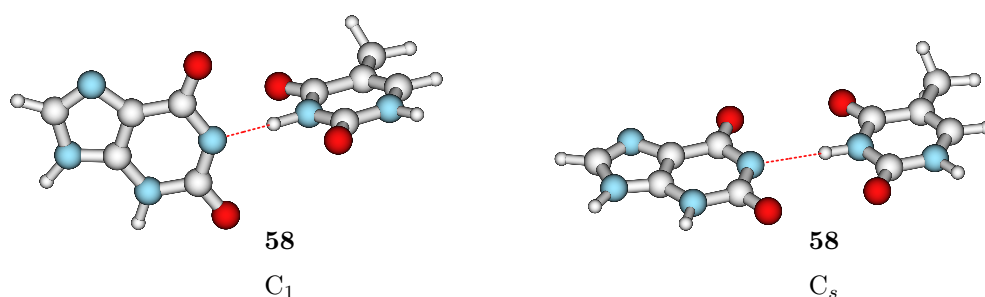


Fig. 4.36: Twisted (C_1) and planar (C_s) structure of anionic pairing of xanthine to thymine **58**. The twist angle of the molecular planes in the C_1 -structure is 88.59°

For the mutagenic potential of xanthosine, the calculations lead to the following picture:

- Pairings that realize the exact Watson Crick geometry are only possible for the pairing of xanthine to cytosine, however, the base pairs are only weakly bound. The dimerization energies of both the neutral, for which a pairing involving a tautomer of xanthine seems to be slightly favored, and the anionic pairing is around 0 kcal/mol, so that the incorporation of the xanthine-cytosine base pair into DNA is solely possible, because the base pair fulfills the geometric demands of DNA polymerase, but it is not enhanced by the H-bond interactions. The finding that in the Watson-Crick pairing mode xanthine has almost no affinity to cytosine is in good agreement with the experimental

result that the cytosine-xanthine base pair is incorporated into DNA at a much lower rate than the cytosine-guanine base pair, which has a very strong hydrogen bonding ($E_{Dim} = -8.7$ kcal/mol).

- To explain why the xanthine-thymine base pair can be incorporated into DNA, the anionic pairings of xanthine to thymine have to be considered. In its deprotonated form, xanthine has much a higher affinity to thymine than to cytosine, which is due to the distorted Watson-Crick pairing **55** in which the negative charge of the xanthine anion located at the C2-oxygen atom can be stabilized by an H-bond formed by the sp^2 N1-hydrogen atom. However, the geometry of pairing **55** deviates from the exact geometry of the canonical Watson-Crick base pairs, so that the fit into the binding pocket of DNA polymerase is worsened. Even though the turn over rate for the distorted pairings is significantly lower, the decrease should be compensated by the higher stability of the base pair, so that the turnover rate is comparable to that of xanthine-guanine.

Chapter 5

Summary

The present work consists of two parts. The first one deals with theoretical questions and tests the performance of orbitals obtained from a self-interaction free KS method, the LHF-approach, in multireference *ab initio* methods. The purpose of this part is to enable a more efficient computation of excitation energies, which is important for the spectroscopic characterization of many organic and bioorganic molecules.

The second part focuses on bioorganic questions and studies the base pairing properties of the purine base xanthine in order to explain, *e.g.*, the unusually high stability of selfpairing xanthine alanyl-PNA double strands and the mutagenicity of xanthine formed in DNA.

LHF-orbitals in MR *ab initio* methods

In contrast to HF- and standard DFT-methods, the LHF-approach leads to a fully bound virtual orbital spectrum, because Coulomb self interactions are exactly canceled in the LHF-ansatz. Furthermore, the energies of the occupied orbitals are not upshifted, like it is the case for standard DFT-methods, so that Koopmans' theorem remains valid. In line with this, also the occupied LHF-orbitals are somewhat more compact than standard DFT-orbitals.

The present work shows that both properties are of great benefit for MR methods. The virtual LHF-orbitals are well optimized and allow an efficient description of excited states and static correlation in both MRCI- and MRPT2-approaches. Furthermore, the higher

compactness of the occupied LHF- compared to standard DFT-orbitals leads to a better description of the center ion of Rydberg states.

However, for each of the two advantages mentioned at least one example molecule has been found, for which LHF-orbitals actually perform worse than HF- and/or standard DFT-orbitals. This shows, that even though LHF virtual orbitals allow an excellent MRCI- and MRPT2-description for the electronically excited states of a large number of molecules, this cannot be generalized and their performance needs to be tested for each individual case.

MRCI

In detail, the excited states of the test molecules ethene, water, carbonmonoxide, formaldehyde and acetone were computed using the MRCI-approach. For the first two test molecules water and ethene, the quality of the orbitals was determined by the number of reference configurations that is needed to reach a given accuracy in the MRCI+Q-computation. LHF-orbitals were compared to HF-, BLYP- and, for ethene, also to CASSCF-orbitals.

For both molecules, the number of configurations in the reference space needed to obtain the required accuracy usually equals the number of states computed when LHF-orbitals are used, which means that the MRCI-wavefunctions are dominated by one configuration only. The corresponding virtual LHF-orbitals can be easily classified into antibonding or the 3s-, 3p- etc. orbitals of the Rydberg series. The only exception is the $^1B_{1u}$ -valence state of ethene, that shows Rydberg-valence mixing. In this case, the Rydberg-valence mixing found is physical and not an artefact produced by the underlying LHF-orbitals.

In contrast to the compact CI-description based on LHF-orbitals, the use of HF-orbitals requires very large reference spaces of up to ten configurations per state to obtain the demanded accuracy, since virtual HF-orbitals are very diffuse and the CI needs to reconstruct the density of the Rydberg state from a large number of configurations. The performance of BLYP-orbitals is strongly determined by the erroneous exponential decay of the BLYP-potential. BLYP-orbitals become increasingly diffuse with increasing energy, while the inner part of the BLYP Rydberg orbitals remains relatively dense. The lower excited states are well described in the MRCI-computations by BLYP-orbitals, but for Rydberg states with higher quantum numbers up to 6 reference configurations are needed to describe a given

state. CASSCF-orbitals, like LHF-orbitals, allow a very good description of the excited states of ethene. However, this is only the case if the CASSCF-orbitals are explicitly optimized for the given states, while ground state CASSCF-orbitals sometimes show a wrong nodal structure.

In the next step, the role of relaxation of the orbitals describing the core ion of the excited states was studied. For the water molecule, both LHF- and HF-orbitals lead to similar excitation energies in the MRCI+Q computation. However, when BLYP-orbitals are used the MRCI+Q energies do not converge towards the values determined experimentally with increasing reference spaces, but tend to underestimate the experimental data. The reason for the underestimation of the excitation energies is that the wrong asymptotic behavior of the BLYP potential leads to a too slow decay of the density. This erroneous decay does not lead to difficulties in the description of the ground state. However, if the same orbitals are used to describe the more contracted core density of an excited state, the same error leads to smaller c^2 -values. The core density of an excited state is more contracted because when an electron is excited from an occupied into a previously unoccupied orbital, the remaining occupied orbitals, which for an excited state describe the core ion, contract. As a consequence of the smaller c^2 -values, the Davidson correction overestimates the full-CI energy for the respective states. For other cases like the ethene molecule, the orbital relaxation only plays a minor role in the computation of the excitation energies, since an error cancellation as assumed in Koopmans' theorem takes place and the decrease of correlation energy about equals the energy contribution of the orbital relaxation.

Since the MRCI+Q energies not in all cases converge towards the experimentally determined excitation energies, the first quality criterion had to be redefined and was replaced by the number of singly excited configurations in the reference space, that is needed to reach convergence in the MRCI computation. Since the single excitations take place out of one defined orbital, the refined quality criterion tests the properties of the virtual orbitals only, while the deviation of the computed excitation energy from the experimental values tests the ability of the occupied orbitals to describe the core orbitals of the excited states.

Using the refined quality criterion, the three test molecules carbonmonoxide, formaldehyde and acetone were computed. Even though carbonmonoxide, formaldehyde and acetone are electronically rather similar molecules, the performance of LHF-orbitals in the description of

their excited states with MRCI methods differs considerably. LHF-orbitals excellently capture the electronic character of the excited states of formaldehyde, but for carbonmonoxide and acetone, problems of completely different nature occur. For carbonmonoxide, the core orbitals of the excited states seem to be imperfectly described. For acetone, the difficulties are related to the spatial form of the virtual LHF-orbitals, which seem to be somewhat too compact to give a good description of the corresponding states.

For BLYP-orbitals, the properties of the virtual orbitals are again strongly related with the erroneous asymptotic decay of the KS-potential. As a result, BLYP-orbitals seem to be a good choice to describe excited states up to a certain energy level, but above this threshold, the virtual orbitals become highly diffuse and no longer allow a good description of the corresponding excited states. Fortunately, BLYP-orbitals well describe the situation for acetone, which is due to the larger size of acetone compared to formaldehyde or carbonmonoxide, so that also the space that is not affected by the wrong decay of the BLYP-potential is larger. For carbonmonoxide, formaldehyde and acetone also B3LYP-orbitals were tested. As expected the results obtained with B3LYP-orbitals simply represent an averaging of those obtained using HF- and BLYP-orbitals.

MRPT2

It was shown by computations for the ground states of carbonmonoxide and formaldehyde, that the overestimation of the 2nd order correction in the single reference case is eliminated if configurations representing double excitations into antibonding orbitals are included in the reference spaces. This inclusion also allows a more efficient description of static correlation effects.

The MRPT2-computations for excited states were performed for carbonmonoxide, formaldehyde, acetone and ethene. Since MRPT2-methods demand larger reference spaces than the MRCI-method, an automatic selection procedure was used to determine the reference configurations. LHF-orbitals give a good description for the excited states of formaldehyde, ethene, and with some limitations also for those of carbonmonoxide. Difficulties of the MRCI computation to describe the core orbitals of carbonmonoxide with LHF-orbitals do not seem to play a major role in the considered MRPT2-approach. For the excitation energies

of formaldehyde, the MRPT2 method possesses the same precision as the MRCI-method. However, the excitation energies of acetone computed on the basis of LHF-orbitals show a systematic underestimation of the experimental excitation energies especially for Rydberg states. Despite these problems, LHF-orbitals seem to give a better positioning of Rydberg relative to valence states which allows a better description of Rydberg-valence mixings. This was shown for the Rydberg-valence mixing of the $^1B_{1u}$ -states of ethene, which represents a difficult case for the CASPT2- or for MRPT2-methods using orbitals optimized for the center ion of the Rydberg states.

In comparison to LHF-orbitals, HF-orbitals in several cases lead to large deviations of the computed energies from experimental values. An exception is the Rydberg states of acetone that can be predicted very precisely with the MRPT2-approach based on HF-orbitals. Like for the MRCI-method also for the MRPT2-approach BLYP-orbitals only give a good description of valence and lower lying Rydberg states, while the description of higher lying Rydberg states is difficult. Except for the carbonmonoxide molecule, where B3LYP-orbitals lead to somewhat less accurate excitation energies, BLYP- and B3LYP-orbitals perform very similar in the MRPT2-computations.

Base pairing of xanthine

In the second part of the present work, the base pairing properties of xanthine and xanthine derivatives were studied. The purpose of this part was to find an explanation for the unexpectedly high stability of the xanthine alanyl PNA double strand. Furthermore, it was analyzed, why xanthine, that is formed from guanine in DNA under chemical stress, is able to form mismatched base pairs with the pyrimidine base thymine.

Stability of xanthine alanyl PNA

In the first step, the regioisomer present in the considered alanyl PNA was identified to be the N7-regioisomer of xanthine by a theoretical analysis of the ^{13}C -NMR spectrum. To analyze the stability of the xanthine self-pairing, a simplified model was set up, in which the stability of the PNA double strand was explained solely by the energy contributions from H-bonding and base stacking.

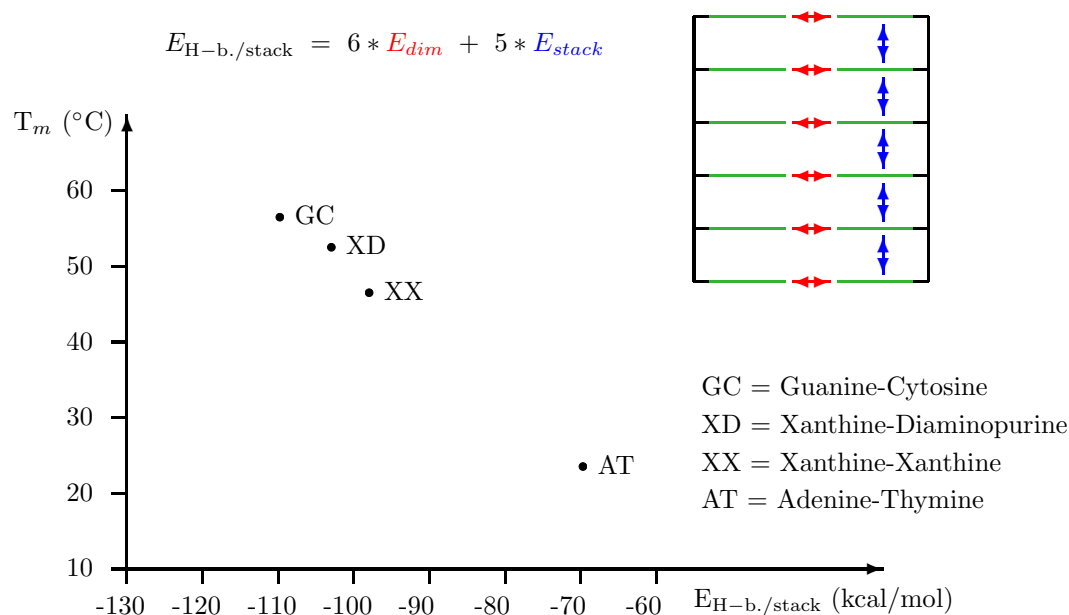


Fig. 5.1: Correlation of the stabilization energy $E_{\text{H-b./stack}}$ resulting from the sum over all H-bond and stacking interactions with the melting temperature (T_m) of hexameric xanthine alanyl PNA. The factors 6 and 5 simply sum up the number of interactions as it is shown by the blue and red arrows in the illustration.

For that purpose, the dimerization and stacking energies for the xanthine-xanthine, guanine-cytosine, adenine-thymine and xanthine-2,6-diaminopurine base pairs were computed using DFT and MP2 methods. Solvent effects were taken into account by the conductor like screening model. The influence of the peptide backbone on the stacking geometry was considered by force field optimizations.

While the individual contributions from hydrogen bonding and stacking do not correlate with the melting temperature T_m , the sum of both correlates linearly with T_m (fig 5.1). This correlation is somewhat surprising, because this means that the effects of the entropy and the molecular water environment either cancel or are similar for all systems compared. In this model, the stability of the xanthine selfpairing mainly stems from an enlarged stacking interaction, while the H-bonds give only minor contributions to the stability of the xanthine selfpaired double strand of alanyl-PNA.

Base pairing properties of N9-Xanthine

The computation of the base pairing properties of N9-xanthine revealed a strong variation in the individual H-bond strengths for the selfpairing of xanthine, that range from -4 to -11 kcal/mol in the gas phase and -2.5 to -5 kcal/mol in polar solvent. By comparison with model systems it was shown that the strong variance of the H-bond strength is mainly due to attractive or repulsive secondary electrostatic interactions. For the homodimer of hypoxanthine it was shown that the increase of aromaticity in the pyrimidine ring upon dimer formation leads to a strengthening of the hydrogen bonds. Electronic effects are important, because they are less affected by solvent interaction than purely electrostatic interactions.

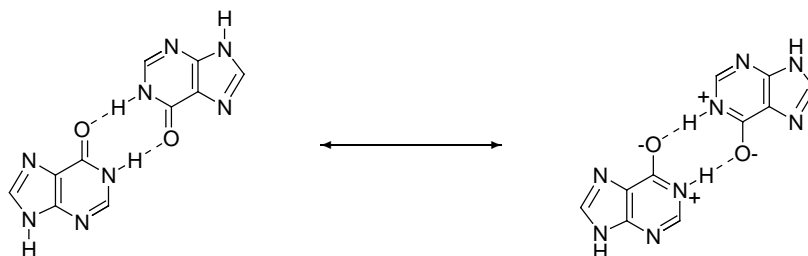


Fig. 5.2: Stabilization of the dimer of hypoxanthine due to an increase of aromaticity

The predictions made have been tested experimentally by the characterization of xanthosine (the ribosyl derivative of xanthine) and inosine (the ribosyl derivative of hypoxanthine) dimers using low temperature $^1\text{H-NMR}$ spectroscopy. The assignment of the spectra was performed by a comparison with computed spectra. In the experimental study xanthosine did not form one of the predicted unusually stable dimers, but preferred a rather weak pairing mode, since the steric of the ribosyl group inhibits the formation of energetically more favorable pairings. Instead, the N3-H functionality that would lead to the unusually stable pairing modes is stabilized by an internal hydrogen bond to an oxygen atom of the ribosyl sugar. For inosine, the predicted strong association has indeed been found.

Mutagenicity of hypoxanthine and xanthine

Several neutral and anionic Watson-Crick base pairs of xanthine were computed with MP2- and DFT-methods in order to explain the mutagenicity of hypoxanthine and xanthine. Also basepairs involving tautomeric forms of xanthine and hypoxanthine were considered. To evaluate the dimerization energies found, the dimers were classified into pairings that have the exact geometry of the canonical base pairs and those that realize a distorted Watson-Crick pairing mode.



Fig. 5.3: Watson-Crick pairings of anionic and tautomeric xanthine with cytosine

The computations show that a stable pairing which realizes the exact geometry of a canonical Watson Crick base pairing is only possible for the pairing of xanthine to cytosine (fig. 5.3), however, the base pairs are only weakly bound. The dimerization energies of both the neutral and the anionic pairing is around 0 kcal/mol, so that the xanthine-cytosine base pairs are incorporated into DNA solely because the base pairs fulfill the geometric demands of DNA polymerase, but it does not profit from any additional stabilization due to hydrogen

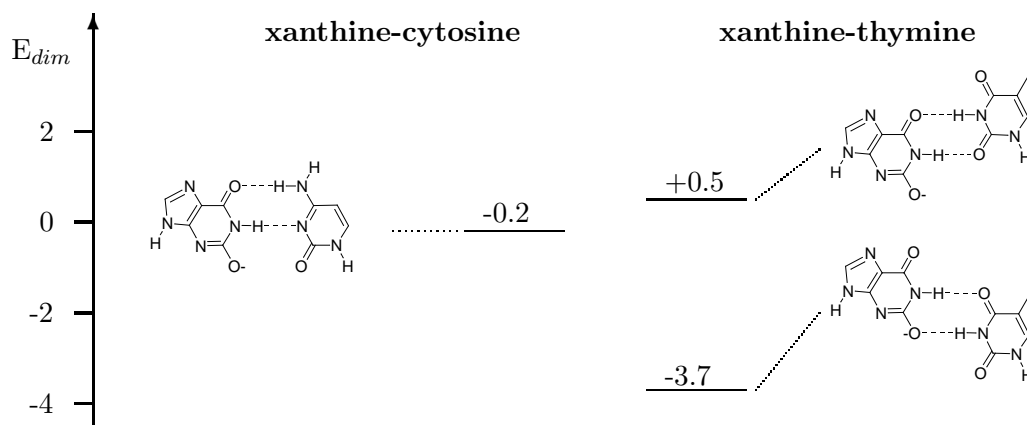


Fig. 5.4: Dimerization energies of the most stable anionic Watson-Crick pairings of xanthine and cytosine or thymine (B3LYP/TZVPP(+); $\epsilon = 78$). Energies in kcal/mol.

bonding. The finding that in the Watson-Crick pairing mode xanthine has almost no affinity to cytosine is in correspondence with the experimental result that the cytosine-xanthine base pair is incorporated into DNA at a much lower rate than the cytosine-guanine base pair, which has a very strong hydrogen bonding.

While the affinity of xanthine to cytosine is very low, the computations predict that xanthine is able to form a stable Watson-Crick pairing with thymine (fig. 5.4). However, the pairing has a somewhat distorted Watson-Crick geometry, so that its high stability is outbalanced by the worsened fit to the binding pocket of DNA-polymerase. As a consequence, the xanthine-thymine pairing is incorporated into DNA not at a faster, but only at a rate comparable to that of the xanthine-cytosine pairing

Chapter 6

Zusammenfassung

Die vorliegende Arbeit besteht aus zwei Teilen. Der erste untersucht die Eignung von LHF-Orbitalen für Multireferenzverfahren. Das Ziel dieses Teils ist eine effizientere Berechnung angeregter Zustände, was zur spektroskopischen Charakterisierung vieler organischer und bioorganischer Moleküle wichtig ist.

Der zweite Teil befasst sich mit bioorganischen Fragestellungen und untersucht die Paarungseigenschaften der Purinbase Xanthin. Es wird unter anderem die unerwartet hohe Stabilität der Xanthin Selbstpaarung in Alanyl-PNA erklärt und es wird untersucht, auf welche Weise Xanthin in der DNA mutagene Fehlpaarungen mit Thymin eingehen kann.

LHF-Orbitale als Einteilchenbasis für MR ab initio Methoden

Im Unterschied zu HF- und Standard-DFT-Methoden führt der LHF-Ansatz zu einem vollständig gebundenen Orbitalspektrum, da Coulomb-Selbstwechselwirkungen im LHF-Ansatz exakt korrigiert werden. Durch die Korrektur der Coulomb-Selbstwechselwirkungen sind im LHF-Ansatz auch die Energien der besetzten Orbitale nicht wie in Standard-DFT-Methoden zu höheren Werten verschoben, so dass das Koopmans' Theorem gültig bleibt und die besetzten LHF-Orbitale etwas kompakter als Standard-DFT-Orbitale sind.

Die vorliegende Arbeit zeigt, dass beide Eigenschaften deutliche Vorteile für MR-Verfahren darstellen. Die virtuellen LHF-Orbitale sind gut optimiert und erlauben eine effiziente

Beschreibung sowohl angeregter Zustände als auch statischer Korrelationseffekte in MRCI- und MRPT2-Ansätzen. Weiterhin führt die kompaktere Struktur der besetzten LHF-Orbitale zu einer besseren Beschreibung des kationischen Rumpfes von Rydbergzuständen.

Andererseits wurden zu beiden genannten Vorteilen auch jeweils ein Beispielmolekül gefunden, in dem die Vorteile nicht zum Tragen kommen, und zu deren Beschreibung Orbitale aus HF- oder Standard-DFT-Methoden besser geeignet sind. Diese Beispiele zeigen, dass jeder Einzelfall für sich getestet werden muss, auch wenn die angeregten Zustände der meisten Moleküle sehr gut mit LHF-Orbitalen beschrieben werden können.

MRCI

Im Einzelnen wurden die angeregten Zustände der Testmoleküle Ethen, Wasser, Kohlenmonoxid, Formaldehyd und Aceton mit dem MRCI-Ansatz berechnet. Die Qualität der Orbitale wurde zunächst über die Anzahl der Referenzkonfigurationen bestimmt, die erforderlich ist, um eine festgelegte Mindestgenauigkeit in der MRCI-Rechnung zu erreichen. Dieser Ansatz wurde zur Berechnung des Ethen- und des Wassermoleküls verwendet. Die LHF-Orbitale wurden mit HF-, BLYP- und für Ethen auch mit CASSCF-Orbitalen verglichen.

In beiden Fällen kann unter Verwendung von LHF-Orbitalen die geforderte Genauigkeit bereits erreicht werden, wenn jeder Zustand nur über eine einzige Referenzkonfiguration beschrieben wird, womit auch die MRCI-Wellenfunktion von nur einer einzigen Konfiguration dominiert ist. Die in den Referenzkonfigurationen einfach besetzten virtuellen LHF-Orbitale können im Unterschied zu HF- oder BLYP-Orbitalen eindeutig als antibindende oder 3s-, 3p- usw. Orbitale der Rydberg-Serie klassifiziert werden. Eine Ausnahme stellt der $^1B_{1u}$ -Zustand von Ethen dar, in dem es zu einer starken Rydberg-Valenzmischung kommt. Im Falle des $^1B_{1u}$ -Zustands von Ethen ist die Rydberg-Valenzmischung allerdings physikalisch und kein Artefakt, das durch die Verwendung von LHF-Orbitalen erzeugt wird.

In Unterschied zur effizienten CI-Beschreibung mit LHF-Orbitalen erfordert die Verwendung von HF-Orbitalen sehr große Referenzräume aus bis zu 10 Konfigurationen pro Zustand, um die geforderte Genauigkeit zu erreichen. Virtuelle HF-Orbitale sind sehr diffus, weshalb im CI viele Konfigurationen benötigt werden um die Dichteverteilung eines

Rydbergzustands zu rekonstruieren. Die Eignung von BLYP-Orbitalen ist vom fehlerhaften asymptotischen Verhalten des BLYP-Potentials bestimmt: Mit zunehmender Orbitalenergie werden BLYP-Orbitale mehr und mehr diffus, gleichzeitig besitzen die jeweiligen Orbitale im Inneren (d.h. im molekülnahen Bereich) oft eine sehr kompakte Struktur. Dadurch erlauben BLYP-Orbitale eine gute Beschreibung energetisch tief liegender angeregter Zustände, während zur Beschreibung energetisch höher liegender Rydbergzustände bis zu 6 Referenzkonfigurationen benötigt werden, um die elektronische Struktur eines einzelnen Zustands zu erfassen. CASSCF-Orbitale erlauben eine sehr gute Beschreibung der angeregten Zustände von Ethen, allerdings müssen sie für jeden einzelnen Zustand optimiert werden.

Im nächsten Schritt wurde die Relaxation der Orbitale untersucht, die das Zentralion der Rydbergzustände beschreiben. Für das Wassermolekül wurden mit HF- und LHF-Orbitalen bei unterschiedlichem Aufwand ähnlich genaue Anregungsenergien erhalten. Werden jedoch BLYP-Orbitale verwendet, so konvergieren die berechneten Anregungsenergien bei Vergrößerung des Referenzraumes nicht gegen die experimentell ermittelten Werte, sondern neigen dazu diese zu unterschätzen. Die Ursache dafür ist der Selbstwechselwirkungsfehler, welcher im BLYP-Verfahren zu einem zu langsamen Abfall der Dichte führt. Im Grundzustand wirkt sich dieser Fehler nur wenig aus. Werden jedoch die gleichen Orbitale verwendet um einen angeregten Zustand zu beschreiben, so führt der zu flache Abfall der Kernionendichte zu kleineren c^2 -Werten. In diesem Fall sind die besetzten BLYP-Orbitale zu diffus um den kationischen Kern des angeregten Zustandes korrekt zu beschreiben. Wenn ein Elektron von einem besetzten in ein unbesetztes Orbital angeregt wird, vermindert sich nämlich für die anderen besetzten Orbitale die Kernabschirmung, wodurch es zu einer Orbitalkontraktion kommt. Die zu kleinen c^2 -Werte wiederum führen zu einem Versagen der *Full-CI* Korrektur. In anderen Fällen wie den angeregten Zuständen des Ethenmoleküls spielt die Orbitalrelaxation nur eine untergeordnete Rolle, da es zu einer Fehlerkompensation von Orbitalrelaxation und Korrelation kommt, wie sie auch in Koopmans' Theorem gefordert wird.

Da die MRCI+Q-Energien nicht in jedem Fall auf die experimentellen Anregungsenergien hin konvergieren, musste das Qualitätskriterium neu formuliert werden. Es wurde durch die Anzahl der Referenzkonfigurationen ersetzt, die notwendig ist, um Konvergenz in den berechneten MRCI+Q Energien zu erreichen. Da die Einfachanregungen aus festgelegten Orbitalen erfolgen, bewertet das verbesserte Qualitätskriterium nur die Eigenschaften der

virtuellen Orbitale, während die Eignung der besetzten Orbitale über die Abweichung der berechneten Anregungsenergien von den experimentellen Werten, bzw. über die c^2 -Werte der MRCI-Wellenfunktion beurteilt wird.

Mit Hilfe des neu formulierten Qualitätskriteriums wurden die drei Testmoleküle Kohlenmonoxid, Formaldehyd und Aceton berechnet. Obwohl alle drei Moleküle elektronisch sehr ähnlich zu sein scheinen, beschreiben LHF Orbitale ihre angeregten Zustände auf sehr unterschiedlich Weise. Während sie den elektronischen Charakter der angeregten Zustände von Formaldehyd ausgezeichnet erfassen, treten für Kohlenmonoxid und Aceton Schwierigkeiten unterschiedlicher Natur auf. Im Falle von Kohlenmonoxid scheinen die Orbitale zur Beschreibung des Kernions schlecht geeignet zu sein. Für das Aceton wiederum resultieren die Schwierigkeiten aus der räumlichen Struktur der virtuellen LHF-Orbitale, die zu kompakt sind, um den elektronischen Charakter der dazugehörigen angeregten Zustände angemessen zu erfassen.

Die Ergebnisse, die für die drei Testmoleküle Kohlenmonoxid, Formaldehyd und Aceton mit HF- und BLYP-Orbitalen erhalten wurden, zeigen ähnliche Trends wie sie bereits für das Ethenmolekül erhalten wurden. Interessanterweise führen BLYP-Orbitale zu einer relativ guten Beschreibung des Acetonmoleküls. Für Kohlenmonoxid, Formaldehyd und Aceton wurden auch B3LYP-Orbitale getestet. Wie zu erwarten liegen die erhaltenen Anregungsenergien stets zwischen denen von HF- und BLYP-Orbitalen.

MRPT2

Zur Verwendung von KS-Orbitalen in MRPT2-Ansätzen wurde zunächst gezeigt, dass die häufig auftretende Überschätzung der Korrelationsenergie in der zweiten Ordnung der Störungstheorie in einem Multireferenz-Ansatz verschwindet, wenn die aus Doppelanregungen in die π^* -Orbitale resultierenden Konfigurationen in den Referenzräumen berücksichtigt werden. Zusätzlich ermöglicht die Berücksichtigung dieser Konfigurationen auch eine effizientere Beschreibung der statischen Korrelation.

Im nächsten Schritt wurden die angeregten Zustände von Kohlenmonoxid, Formaldehyd, Aceton und Ethen berechnet. Da für das MRPT2-Verfahren generell größere Referenzräume als für das MRCI-Verfahren benötigt werden, wurden die Referenzräume über ein automa-

tisches Selektionsverfahren bestimmt und die Genauigkeit der berechneten Anregungsenergien verglichen. In dem gewählten MRPT2-Ansatz erlauben LHF-Orbitale eine gute Beschreibung der angeregten Zustände von Kohlenmonoxid, Formaldehyd und Ethen. Die Schwierigkeiten im MRCI, die mit der Beschreibung der Orbitale des Kernions der angeregten Zustände von Kohlenmonoxid verbunden waren, scheinen im MRPT2-Ansatz keine größere Rolle zu spielen. Wie bereits in den MRCI-Rechnungen gezeigt wurde, beschreiben die LHF-Orbitale die Rydbergzustände von Aceton fehlerhaft, was im MRPT2 zu einer systematischen Unterschätzung der Anregungsenergien führt.

Anhand des $^1B_{1u}$ -Zustandes von Ethen, der einen schwierigen Fall für die CASPT2-Methode darstellt, wurde zusätzlich gezeigt, dass LHF-Orbitale zu einer günstigeren relativen Lage von Valenz und Rydbergzuständen im Referenz-CI führen, wodurch eine korrektere Beschreibung der Rydberg-Valenzmischung möglich ist. Im Vergleich zu LHF-Orbitalen führen HF-Orbitale im MRPT2-Ansatz in vielen Fällen zu deutlichen Abweichungen der berechneten Anregungsenergien von den experimentell bestimmten Werten. Eine Ausnahme sind die Rydbergzustände von Aceton, die ausgehend von HF-Orbitalen sehr genau mit dem MRPT2-Ansatz berechnet werden können. Wie bereits für den MRCI-Ansatz beschrieben, ergeben BLYP-Orbitale eine sehr genaue Berechnung der Valenzzustände und der energetisch tief liegenden Rydbergzustände, während mit zunehmender Anregungsenergie die Rydbergzustände schlechter erfasst werden. Mit Ausnahme des Kohlenmonoxids, für das B3LYP-Orbitale etwas schlechter als BLYP-Orbitale geeignet sind, zeigten sich in den MRPT2-Rechnungen keine größeren Unterschiede zwischen BLYP- und B3LYP-Orbitalen.

Basenpaarung von Xanthin

Im zweiten Teil der vorliegenden Arbeit wurden die Paarungseigenschaften von Xanthin und Xanthinderivaten untersucht. Ziel dieses Teils war es, eine Erklärung für die unerwartet hohe Stabilität des Xanthin Alanyl-PNA Selbstpaarung zu finden. Weiterhin wurde untersucht, weshalb Xanthin, das in der DNA u.a. unter chemischem Stress gebildet wird, mutagene Fehlpaarungen mit der Pyrimidinbase Thymin eingehen kann.

Stabilität der Xanthin Alanyl PNA

Zunächst wurde durch den Vergleich experimenteller und berechneter ^{13}C -NMR-Spektren das Regiosomer von Xanthin bestimmt, welches zu der ungewöhnlich hohen Stabilität der Xanthin-Xanthin-Selbstpaarung in Alanyl-PNA verantwortlich ist. Zur Untersuchung der Stabilität der Xanthin-Selbstpaarung wurde ein stark vereinfachendes Modell aufgestellt,

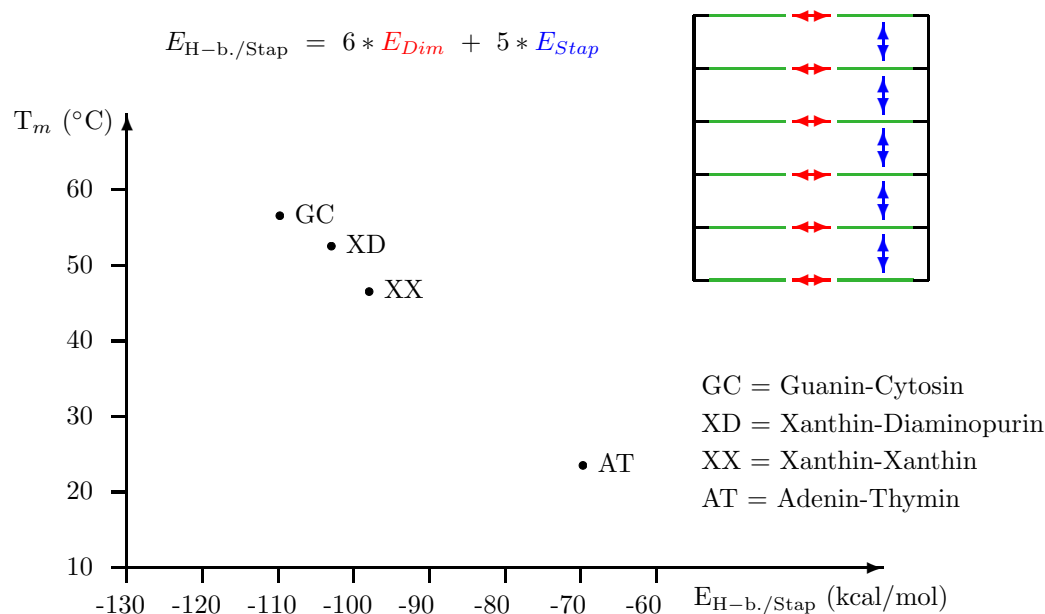


Abb. 6.1: Korrelation der Gesamtstabilisierungsenergie $E_{\text{H-b./Stap}}$, die sich aus der Summe aller H-Brücken und Stapelungswechselwirkungen ergibt, mit den experimentell ermittelten Schmelztemperaturen hexamerer Alanyl-PNA. Die Faktoren 5 und 6 ergeben sich aus der Anzahl der im Hexamer auftretenden Wechselwirkungen.

in dem die Stabilität der PNA-Duplexe nur über die Energiebeiträge aus den Wasserstoffbrücken (E_{Dim}) und der Basenstapelung (E_{Stap}) bestimmt wird. Die Dimerisierungs- und Stapelungsenergien unterschiedlicher Paarungen wurden mit DFT- und MP2-Methoden bestimmt. Solvenseffekte wurden über ein Kontinuummodell erfasst und der Einfluß des peptidischen Rückgrats auf die Stapelungsgeometrie wurde durch Kraftfeldmethoden berücksichtigt.

Während die einzelnen Energiekomponenten aus den H-Brücken und der Basenstapelung keinen eindeutigen Zusammenhang zu den Schmelztemperaturen erkennen lassen, korreliert die Summe aus beiden linear mit den experimentell ermittelten T_m -Werten (Abb. 6.1). Dies bedeutet, dass die Beiträge aus der Entropie, der molekularen Wasserumgebung und der Rückgratspannung sich entweder aufheben oder für alle behandelten Systeme sehr ähnlich sind. Die Stabilität der Xanthin-Xanthin- und die der 2,6-Diaminopurin-Xanthin-Paarung, ergibt sich durch einen erhöhten Stapelungsbeitrag der Purinpaarungen, während die Wasserstoffbrücken der Xanthin Selbstpaarung nur wenig zur Stabilisierung des Xanthin-Xanthin- und des Xanthin-Diaminopurin-Alanyl-PNA-Doppelstrangs beitragen.

Paarungseigenschaften von N9-Xanthin

Zur Untersuchung der Paarungseigenschaften von N9-Xanthin wurden zunächst H-verbrückte Homodimere von Xanthin untersucht. Hierbei wurden extreme Variationen in den Bindungsstärken der einzelnen H-Brücken gefunden, die sich zwischen -4 bis -11 kcal/mol in der Gasphase und -2.5 bis -5 kcal/mol im Solvens betragen. Durch Vergleich mit Modellsystemen konnte die starke Varianz der H-Brückenstärke auf anziehende bzw. abstoßende sekundäre elektrostatische Wechselwirkungen zurückgeführt werden. Weiterhin wurde das Homodimer von Hypoxanthin untersucht, bei dem die H-Brücken durch eine Erhöhung der Aromatizität im Pyrimidinring zusätzlich verstärkt werden, was zu einer deutlichen Stabilisierung des Dimers führt (Abb. 6.2). Elektronische Effekte müssen vor allem deshalb berücksichtigt werden, da sie im Unterschied zu rein elektrostatischen Effekten deutlich weniger von der Solvensumgebung beeinflusst werden.

Die Vorhersagen zu den Xanthin- und Hypoxanthin-Homodimeren wurden experimentell mittels Tieftemperatur ^1H -NMR-Untersuchungen der Dimere von Xanthosin (dem Ribo-

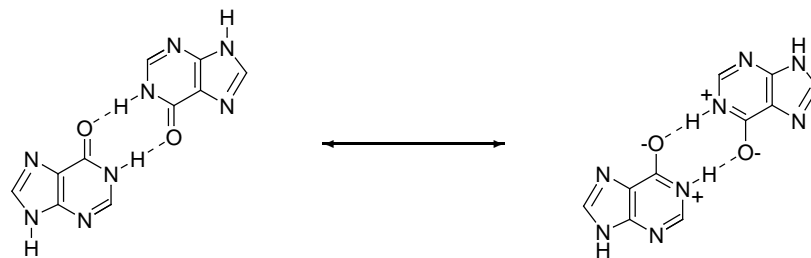


Abb. 6.2: Stabilisierung des Inosindimers über die Erhöhung der Aromatizität

sylderivat von Xanthin) und Inosin (das Ribosylderivat von Hypoxanthin) in Freonlösung überprüft. Die Auswertung der Spektren erfolgte wiederum mit Hilfe berechneter Spektren. Es zeigte sich, dass Xanthosin aufgrund der sterischen Abstoßung des Ribosylrestes nicht in einer der zu erwartenden sehr stabilen Modi paart, sondern einen eher schwach gebundenen Modus vorzieht. Für Inosin konnte auch experimentell eine sehr hohe Assoziationskonstante bestimmt werden.

Mutagenität von Hypoxanthin und Xanthin

Zur Erklärung der Mutagenität von Hypoxanthin und Xanthin wurden verschiedene neutrale und anionische Watson-Crick Basenpaarungen von Hypoxanthin und Xanthin mit Pyrimidinbasen berechnet. Hierbei wurden u. a. auch tautomere und anionische Formen von Xanthin berücksichtigt. Zur Bewertung der erhaltenen Dimerisierungsenergien wurden die Paarungen danach klassifiziert, ob ihre Geometrien mit denen der kanonischen Basenpaarungen deckungsgleich sind, oder ob sie in einer verzerrten Watson-Crick Geometrie vorliegen, was die Einbaurate in die DNA aufgrund des räumlichen Anspruchs der DNA-Polymerase vermin-



Abb. 6.3: Watson-Crick Paarungen von anionischem und tautomerem Xanthin mit Cytosin

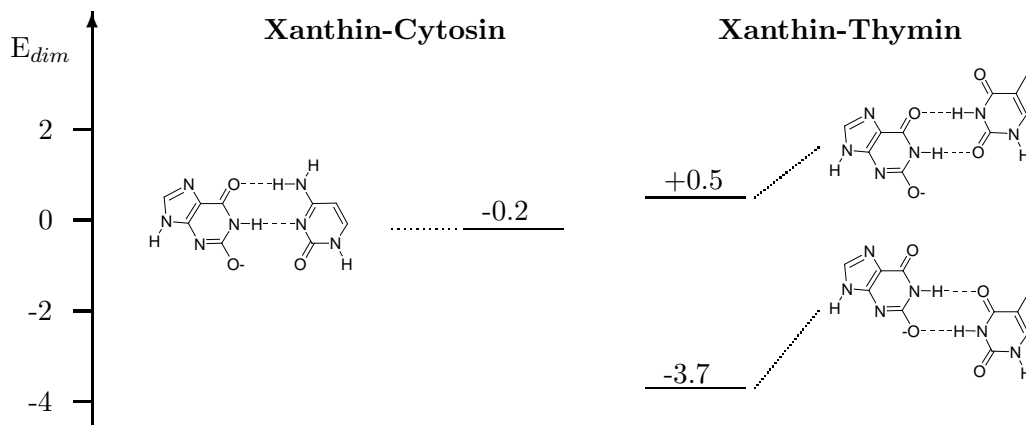


Abb. 6.4: Dimerisierungsenergien anionischer Watson-Crick Paarungen von Xanthin und Cytosin bzw. Thymin. Energien in kcal/mol.

dert.

Die Rechnungen zeigen, dass Xanthin nur mit Cytosin Watson-Crick-Paarungen eingehen kann (Abb. 6.3), welche jedoch nur sehr schwach gebunden sind. In der neutralen Form scheint eine dreizählige Basenpaarung unter Beteiligung einer tautomeren Form des Xanthins etwas stabiler zu sein als die zweizählige Paarung von Diketoxanthin mit Cytosin. Da die Dimerisierungsenergie sowohl der neutralen als auch der anionischen Basenpaarung nur wenig unter 0 kcal/mol liegt, ist der Einbau der Xanthin-Cytosin-Paarung in die DNA zwar aufgrund der günstigen Geometrie möglich, wird aber nicht durch einen Energiebeitrag aus den H-Brücken verstärkt. Die im Vergleich zur Guanin-Cytosin Paarung deutlich geringere Affinität von Xanthin zu Cytosin ist im Einklang mit dem experimentellen Befund, dass die Cytosin-Xanthin Paarung deutlich langsamer als die Guanin-Cytosin Paarungen in die DNA eingebaut werden.

Während die Rechnungen nur eine geringe Affinität von Xanthin zu Cytosin vorhersagen, scheint das Anion von Xanthin in der Lage zu sein, eine sehr stabile Basenpaarung mit Thymin einzugehen (Abb. 6.4). Allerdings muss die Dimerisierungsenergie die schlechtere Anpassung in die Bindungstasche der DNA-Polymerase ausgleichen, da die Paarung in einer etwas verzerrten Watson-Crick Geometrie vorliegt. Insgesamt wird die Paarung daher nicht schneller in die DNA eingebaut, wie erwartet aufgrund der H-Brückenstärken, stattdessen besitzt sie eine ähnliche Einbaurrate wie die geometrisch günstigere aber weniger stabile Xanthin-Cytosin Paarung.

Bibliography

- [1] I. Shavitt, Mol. Phys. **94**, 3 (1998).
- [2] K. Wolinsky, H. L. Sellers and P. Pulay, Chem. Phys. Lett. **140**, 225 (1987).
- [3] K. Wolinsky and P. Pulay, Chem. Phys. Lett. **90**, 3467 (1989).
- [4] R. B. Murphy and R. P. Messmer, Chem. Phys. Lett. **183**, 443 (1991).
- [5] R. B. Murphy and R. P. Messmer, J. Chem. Phys. **97**, 4170 (1992).
- [6] B. O. Roos, M. P. Fülcher, P.-Å. Malmqvist, M. Merchán and L. Serrano-Andrés, Theoretical studies of electronic spectra of organic molecules, in: S. R. Langhoff, ed. *Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy* (Kluwer Academic Publishers, Dordrecht, The Netherlands, 1995), 357.
- [7] B.O. Roos, K. Andersson, M.P. Fülcher, P.-Å Malmqvist, K. Pierloot and M. Merchán, Multiconfigurational perturbation theory: Applications in electronic spectroscopy, in: I. Prigogine and S. A. Rice, eds. *Advances in Chemical Physics: New Methods in Computational Quantum Mechanics Vol. XCIII:219* (John Wiley & Sons, Dordrecht, New York, 1996.).
- [8] K. Andersson, P.-Å.Malmquist, B. O. Roos, A. J. Sadley and K. Wolinski, J. Phys.Chem. **94**, 1990 (1990).
- [9] P. E. Siegbahn, A. Heiberg, B. O. Roos and B. Levy, Phys. scripta **21**, 323 (1980).
- [10] B. O. Roos, P. R. Taylor and P. E. Siegbahn, Chem. Phys. **48**, 157 (1980).

- [11] B. O. Roos, *Int. J. Quantum Chem.* **S14**, 175 (1980).
- [12] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry* (McGraw-Hill, New York, 1989).
- [13] S. Huzinaga and C. Arnau, *Phys. Rev. A* **1**, 1285 (1970).
- [14] S. Huzinaga and C. Arnau, *J. Chem. Phys.* **54**, 1948 (1971).
- [15] K. Morokuma and S. Iwata, *Chem. Phys. Lett.* **16**, 192 (1972).
- [16] E. R. Davidson, *J. Chem. Phys.* **57**, 1999 (1972), S. T. Elbert, S. E. Langhoff and E. R. Davidson, *ibid* **57**, 2005 (1972).
- [17] D. M. Potts, C. M. Taylor, R. K. Chauduri and K. F. Freed, *J. Chem. Phys.* **114**, 2529 (2001).
- [18] S. Grimme and M. Waletzke, *Phys. Chem. Chem. Phys.* **2**, 2075 (2000).
- [19] R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, Oxford, 1989).
- [20] R. M. Dreizler and E. K. U. Gross, *Density Functional Theory* (Springer, Heidelberg, 1990).
- [21] W. Koch and M. C. Holthausen, *A Chemist's Guide to Density Functional Theory* (Wiley-VCH, New York, 2000).
- [22] A. Görling, *Phys. Rev. A* **54**, 3912 (1996).
- [23] D. P. Chong, O. V. Gritsenko and E. J. Baerends, *J. Chem. Phys.* **116**, 1760 (2002).
- [24] F. M. Bickelhaupt and E. J. Baerends, *Rev. Comp. Chem.* **15**, 1 (2000).
- [25] J. Gräfenstein and D. Cremer, *Chem. Phys. Lett.* **316**, 569 (2000).
- [26] S. Grimme and M. Waletzke, *J. Chem. Phys.* **111**, 5645 (1999).
- [27] R. Pollet, A. Savin, T. Leininger and H. Stoll, *J. Chem. Phys.* **116**, 1250 (2002).
- [28] P. Bouř, *Chem. Phys. Lett.* **345**, 331 (2001).

- [29] M. Warken, Chem. Phys. Lett. **237**, 256 (1995).
- [30] A. Görling and M. Levy, Phys. Rev. B **97**, 13105 (1993).
- [31] M. Ernzerhof, Chem. Phys. Lett., **263**, 499 (1996).
- [32] M. Seidl, J. P. Perdew and S. Kurth, Phys. Rev. Lett. **84**, 5070 (2000).
- [33] Y. Nakao, Y-K. Choe, K. Nakayama and K. Hirao, Mol. Phys. **100**, 729 (2002).
- [34] E. J. Baerends and O. V. Gritsenko, J. Phys. Chem. A **101**, 5383 (1997).
- [35] M. Weimer, F. Della Sala and A. Görling, Chem. Phys. Lett. **372**, 538 (2003).
- [36] J. D. Talman and W. F. Shadwick, Phys. Rev. A **14**, 36 (1976).
- [37] A. Görling and M. Levy, Phys. Rev. A **50**, 196 (1994).
- [38] T. Grabo, T. Kreibich, S. Kurth and E.K.U. Gross, in: V. I. Anisimov, ed. *Strong Coulomb Correlations in Electronic Structure Calculations: Beyond the Local Density Approximation* (Gordon and Breach, Amsterdam 1999), 203.
- [39] M. Städele and J. A. Majewski, P. Vogl, and A. Görling, Phys. Rev. Lett. **79**, 2089 (1997).
- [40] A. Görling, Phys. Rev. Lett. **83**, 5459 (1999).
- [41] S. Ivanov, S. Hirata and R. J. Bartlett, Phys. Rev. Lett. **83**, 5455 (1999).
- [42] L. Veseth, J. Chem. Phys. **114**, 8789 (2001).
- [43] S. Hirata, S. Ivanov, I. Grabowski, R. Bartlett, K. Burke, and J. D. Talman, J. Chem. Phys. **115**, 1635 (2001).
- [44] F. Della Sala and A. Görling, J. Chem. Phys. **115**, 5718 (2001).
- [45] O. V. Gritsenko and E. J. Baerends, Phys. Rev. A **64**, 042506 (2001).
- [46] J. Álvarez Ruiz, M. Coreno, P. Erman, A. Kivimäki, E. Melero García, K.C. Prince, M. de Simone, E. Rachelev, R. Richter, J. Rius i Riu and L. Veseth, Chem. Phys. Lett. **372**, 139 (2003).

- [47] D. J. Tozer and N. C. Handy, *J. Chem. Phys.* **109**, 10180 (1998).
- [48] R. Bauernschmitt and R. Ahlrichs, *Chem. Phys. Lett.* **256**, 454 (1996).
- [49] T. Hupp, B. Engels, F. Della Sala and A. Görling, *Chem. Phys. Lett.* **360**, 175 (2002).
- [50] T. Hupp, B. Engels, F. Della Sala and A. Görling, *Z. Phys. Chem.* **217**, 133 (2003).
- [51] T. Hupp, B. Engels and A. Görling, *J. Chem. Phys.* **119**, 11591 (2003).
- [52] P. Hohenberg and W. Kohn, *Phys Rev.* **136**, B864 (1964).
- [53] W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).
- [54] K. Burke, *The ABC of DFT* (available at: www.dft.rutgers.edu/kiero/beta).
- [55] A. D. Becke, *Phys. Rev. B.* **38**, 3098 (1988).
- [56] C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B* **37**, 785 (1988).
- [57] A. D. Becke, *J. Chem. Phys.* **98**, 5648 (1993).
- [58] R. T. Sharp and G. K. Horton, *Phys. Rev.* **90**, 317 (1953).
- [59] A. Görling, *Phys. Rev. B* **53**, 7024 (1996).
- [60] M. Städele and M. Moukara, J. A. Majewski, P. Vogl, and A. Görling, *Phys. Rev. B* **59**, 10031 (1999).
- [61] S. Hamel, M. E. Casida and D. R. Salahub, submitted.
- [62] J. C. Slater, *Phys. Rev.* **81**, 385 (1951).
- [63] F. Della Sala and A. Görling, *J. Chem. Phys.* **116**, 5374 (2002).
- [64] TURBOMOLE, Versions **5. - 5.** R. Ahlrichs, M. Bär, H.-P. Baron, R. Bauernschmitt, S. Böcker, M. Ehrig, K. Eichkorn, S. Elliott, F. Furche, F. Haase, M. Häser, H. Horn, C. Huber, U. Huniar, M. Kattaneck, C. Kölmel, M. Kollwitz, K. May, C. Ochsenfeld, H. Öhm, A. Schäfer, U. Schneider, O. Treutler, M. v. Arnim, F. Weigend, P. Weis, H. Weiss, University of Karlsruhe, Germany, since 1988.

- [65] F. Della Sala and A. Görling,
- [66] E. J. Baerends, in *Density Functional Theory, time independent and time-dependent* (available at: www.kfunigraz.ac.at/tchwww/sax/maripfarr/prog02_old.html).
- [67] R. van Leeuwen and E. J. Baerends, *Phys. Rev. A* **49**, 2421 (1994).
- [68] M. Grüning, O. V. Gritsenko, S. J. A. van Gisbergen, and E. J. Baerends, *J. Chem. Phys.* **114**, 652 (2001).
- [69] D. J. Tozer and N. C. Handy, *J. Chem. Phys.* **109**, 10180 (1998).
- [70] L. Serrano-Andrés, M. Merchán, I. Nebot-Gil, R. Lindh, B. Roos, *J. Chem. Phys.* **98**, 3151 (1993).
- [71] F. Della Sala and A. Görling, *Int. J. Quantum. Chem.*, **91**, 131 (2003).
- [72] MOLCAS Version5. K. Andersson, M. Barysz, A. Bernhardsson, M.R.A. Blomberg, D.L. Cooper, T. Fleig, M.P. Fülscher, C. de Graaf, B.A. Hess, G. Karlström, R. Lindh, P.-Å. Malmqvist, P. Neogrady, J. Olsen, B.O. Roos, A.J. Sadlej, M. Schütz, B. Schimmelpfennig, L. Seijo, L. Serrano-Andres, P.E.M. Siegbahn, J. Stalring, T. Thorsteinsson, V. Veryazov and P.-O. Widmark, Lund University, Sweden, 2000.
- [73] J. E. Del Bene, E.A. Stahlberg, Shavitt, *Int. J. Quantum. Chem. Symp.*8, 455, (1990).
- [74] M. Hanrath, B. Engels, *J. Chem. Phys.* **225**, 197 (1997).
- [75] T. Müller, M. Dallos, H. Lischa, *J. Chem. Phys.* **100**, 7176 (1999).
- [76] L. E. McMurchie, E. R. Davidson, *J. Chem. Phys.* **67**, 5613 (1977).
- [77] R. J. Buenker, S. D. Peyerimhoff, *Chem. Phys.* **9**, 75 (1975).
- [78] F. Jensen, *Introduction to Computational Chemistry* (Wiley & Sons, Chichester, 1999).
- [79] Handbook of Chemistry and Physics, D.R. Lide, Ed. in Chief; CRC Press: Boca Raton, 1998-1999.
- [80] Z.-L. Cai, D.J. Tozer and J.R. Reimers, *J. Chem. Phys.* **113**, 7084 (2000).

- [81] C.R. Brundle, M.B. Robin, N.A. Kuebler and H. Basch, *J. Am. Chem. Soc.* **94**, 1451 (1972).
- [82] G. Theodorakopoulos, C.A. Nicolaides, R.J. Buenker and S.D. Peyerimhoff, *Chem. Phys. Lett.* **89**, 164 (1982).
- [83] G. Theodorakopoulos, I.D. Petsalakis, R.J. Buenker and S.D. Peyerimhoff, *Chem. Phys. Lett.* **105**, 253 (1984).
- [84] M. Nooijen and V. Lotrich, *J. Chem. Phys.* **113**, 4549 (2000).
- [85] (a) E.H. van Veen, *Chem. Phys. Lett.* **41**, 540 (1976); (b) W. H. Flicker, O.A. Mosher and A. Kuppermann, *Chem. Phys. Lett.* **36**, 56 (1975).
- [86] D.G. Wilden and J. Corner, *J. Phys. B* **13**, 1009 (1980).
- [87] G. Bieri and L. Asbrink, *J. Electron Spectrosc. Relat. Phenom.* **20**, 149 (1980).
- [88] E. S. Nielsen, P. Jørgensen and J. Oddershede, *J. Chem. Phys.* **73**, 6238 (1980).
- [89] D.P. Chong, O.V. Gritsenko and E.J. Baerends, *J. Chem. Phys.* **116**, 1760 (2002).
- [90] M. J. Allen and D. J. Tozer, *J. Chem. Phys.* **113**, 5183 (2000).
- [91] J. Liu, H.-T. Kim and S. L. Anderson, *J. Chem. Phys.* **114**, 9797 (2001).
- [92] A. Schäfer, C. Huber and R. Ahlrichs, *J. Chem. Phys.* **100**, 5829 (1994).
- [93] D. H. A. ter Steege, A. C. Wirtz and W. J. Buma, *J. Chem. Phys.* **116**, 547 (2002).
- [94] M. Merchan, B. O. Roos, R. McDiarmid and X. Xing, *J. Chem. Phys.* **104**, 1791 (1996).
- [95] J. Finley, P.-Å. Malmqvist, B.O. Roos and L. Serrano-Andrés, *Chem. Phys. Lett.* **288**, 299 (1998).
- [96] K. B. Wiberg, C. M. Hadad, J. B. Foresman and W. A. Chupka, *J. Phys. Chem.* **96**, 10756 (1992).
- [97] W. Saenger, *Principles of Nucleic Acid structure* (Springer, New York, 1994).

- [98] E. T. Kool, *Annu. Rev. Biophys. Biomol. Struct.* **30**, 1, (2001).
- [99] E. T. Kool, J. C. Morales and K. M. Guckian, *Angew. Chem.* **112**, 1046, (2000).
- [100] U. Diederichsen, *Angew. Chem. Int. Ed.* **37**, 1655, (1998).
- [101] S. Neidle *Nucleic Amino Acid structure* (Oxford University Press, Oxford, 1999).
- [102] P. E. Nielsen, M. Engholm, R. H. Berg and O. Buchard, *Science* **254**, 1497 (1991).
- [103] U. Diederichsen, in: U. Diederichsen, T. K. Lindhorst, B. Westmann, L. A. Wessjohann ed. *Bioorganic Chemistry - Highlights and New Aspects* (VCH Weinheim 1999), 255.
- [104] U. Diederichsen, *Angew. Int. Ed. Engl.* **35**, 445 (1996).
- [105] U. Diederichsen *Lineare Nucleinsäure-Analoga mit peptidischem Rückgrat* Habilitationsschrift, TU-München (1999).
- [106] M. F. H. Hoffmann, A. M Brückner, T. Hupp, B. Engels and U. Diederichsen, *Hel. Chim. Acta.* **83**, 2580 (2000).
- [107] G. Eisenbrand, P. Schreier *Römpp Lexikon Lebensmittelchemie* (Georg Thieme, Weinheim 1995).
- [108] G. Rastelli, L. Constantino and A. Albasini, *J. Am. Chem. Soc.* **119**, 3007, (1997).
- [109] Y. W. Kow, *Free Rad. Bio. & Med.*, **33**, 886 (2002).
- [110] V. Vongchampa, M. Dong, L. Ginipalli and P. Dedon, *Nucleic Acids Res.* **31**, 1045 (2003).
- [111] D. A. Wink, K. S. Kasprzak, C. M. Maragos, R. K. Elespuru, M. Misra, T. M. Dunams, T. A. Cebula, W. H. Koch, A. W. Andrews and J. S. Allen, *Science* **254**, 1001 (1991).
- [112] J. L. Caulfeld, J. S. Wishnok and S. R. Tannenbaum *J. Biol. Chem.* **273**, 12689 (1998).
- [113] R. Shapiro, S. Pohl, *Biochemistry* **7**, 448 (1968).
- [114] R. Shapiro, S. J. Shiuey, *Biochim. Biophys. Acta* **174**, 403 (1969).

- [115] J. P. Spence, J. Jenner, K. Chimel, O. I. Aruoma, C. E. Cross, R. Wu, B. Halliwell
FEBS lett. **375**, 179 (1995).
- [116] G. E. Wuenschell, T. R. O'Connor and J. Termini, *Biochem.* **42**, 3608 (2003).
- [117] K. N. Rogstad, Y. H. Jang, L. C. Sowers and W. A. Goddard III, *Chem. Res. Toxicol.*
16, 1455 (2003).
- [118] T. Hupp Diplomarbeit, (Würzburg, 2000).
- [119] K. Eichkorn, O. Treutler, H. Öhm, M. Häser, R. Ahlrichs, *J. Chem. Phys.* **242**, 652
(1995).
- [120] F. Weigend, M. Häser, H. Patzelt, R. Ahlrichs, *J. Chem. Phys.* **294**, 143 (1998).
- [121] Perdew, J.P., Burke, K., Ernzerhof, M., *Phys. Rev. Lett.* **77**, 3865 (1996).
- [122] Perdew, J.P., Ernzerhof, M., Burke, K., *J. Chem. Phys.* **105**, 9982 (1996).
- [123] A. Klamt, G. J. Schürmann, *J. Chem. Soc. Perkin Trans.* **2**, 799 (1993).
- [124] M. J. Frisch, J. A. Pople, J. S. Binkley, *J. Chem. Phys.* **80**, 3265 (1984).
- [125] Gaussian 98 (Revision A.7), M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria,
M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, R. E. Stratmann,
J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M.C. Strain,
O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, A.
Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma,
D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V.
Ortiz, B. B. Stefanov, G. Liu , A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts,
R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C.
Gonzalez, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, J.
L. Andres, M. Head-Gordon, E. S. Replogle and J. A. Pople, Gaussian, Inc. Pittsburgh,
PA,(1998).
- [126] S. F. Boys, F. Bernardi, *Mol. Phys.* **19**, 553 (1970).
- [127] P. Hobza, J. Sponer, *J. Chem. Rev.* **99**, 3247 (1999).

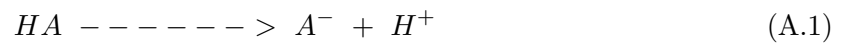
- [128] T. A. Halgren, *Comp. Chem* **29**, 730 (1996).
- [129] SYBYL, Tripos Associates, 1699 Hanley Road, Suite 303, St. Louis, MO, 63144.
- [130] W. L. Jorgensen and J. Pranata, *J. Am. Chem. Soc.* **112**, 2008 (1990).
- [131] A. R. Fersht, *Trends Biochem. Res.* **12**, 301 (1987).
- [132] F. H. Martin, M. M. Castro, F. Aboul-ela and I. Tinoco Jr, *Nucleic Acid Res.* **13**, 8927 (1985).
- [133] Y. Kawase, S. Iwai, H. Inoue, K. Minura and E. Ohtsuka, *Nucleic Acid Res.* **14**, 7727 (1986).
- [134] D. H. Turner, N. Sugimoto, R. Kierzeck, S. D. Dreiker, *J. Am. Chem. Soc.* **109**, 3783 (1987).
- [135] M. Petersheim and D. H. Turner, *Biochemistry* **22**, 256 (1983).
- [136] S. Bommarito, N. Peyret and J. SantaLucia, *Nucleic Acids Res.* **28**, 1929 (2000).
- [137] B. Dietrich, Diplomarbeit, (Würzburg, 2002).
- [138] J. Huheey, E. Keiter, R. Keiter *Anorganische Chemie, Prinzipien von Struktur und Reaktivität* (Walter de Gryter, Berlin 1995), p. 229.
- [139] Martin Kratchovíl, Ola Engkvist, Jiří Šponer, Pavel Jungwirth and P. Hobza, *J. Phys. Chem. A* **102**, 6921 (1998).
- [140] T. Hupp, E. Basilio, K. Wietz and B. Engels, manuscript in preparation.
- [141] T.-K. Ha, H.-J. Keller, R. Gunde and H.-H. Gunthard, *J. Phys. Chem. A* **103**, 6612 (1999).
- [142] M. R. Valentine and J. Termine, *Nucleic Acids Res.* **29**, 1192, (2001).
- [143] F. H. Martin, M. M. Castro, F. Aboul-ela and I. Tinoco Jr., *Nucleic Acids Res.* **13**, 8927, (1985).

- [144] E. Ohtsuka, S. Matsuki, M. Ikehara, Y. Takahashi and K. Matsubara, *J. Biol. Chem.* **260**, 1605, (1985).
- [145] M. Hill-Perkins, M. D. Jones and P. Karran *Mut. Res.* **162**, 153 (1986).
- [146] J. M. Berg, J. L. Tymozko and L. Stryer, *Biochemistry*, (W. H. Freeman and Company), New York, 2001).
- [147] S. H. Eom, J. Wang and T. A. Steitz, *Nature* **382**, 278 (1996)
- [148] K. Bebenek, W. A. Beard, T. A. Darden, L. Li, R. Prasad, , B.A. Luxon, D.G. Gorenstein, S.H. Wilson and T. A. Kunkel, *Nat. Struct. Biol.* **4**, 194, 1997.
- [149] S. Doublié, S. Tabor, A. M. Long, C. C. Richardson and T. Ellenberger, *Nature* **391**, 278 (1998).
- [150] H. Huang, R. Chopra, G. L. Verdine and S. C. Harrison, *Science* *282*, 1669, (1998).
- [151] R. Poirier, R. Kari and I. G. Csizmada, *handbook of gaussian basis sets* (Elsevier, Amsterdam, 1985).
- [152] M. F. Goodman, *Proc. Natl. Acad. Sci. USA* **94**, 10493 (1997).
- [153] G. Stoychef, B. Kierdaszuk and D. Shugar, *Eur. J. Biochem.* **269**, 4048 (2002).
- [154] D. C. Sherrington and K. A. Taskinen, *Chem. Soc. Rev.* **30**, 2008 (112) (2001).

Appendix A

Free Energies for the Deprotonation in a buffered solution

The free energy for the deprotonation^[a]:



is defined as

$$\Delta G = \Delta G^0 + 2.303RT * \log \frac{[A^-][H^+]}{[HA]}. \quad (\text{A.2})$$

For equilibrium ($\Delta G = 0$) A.2 leads to

$$\Delta G^0 = 2.303RT pK_s. \quad (\text{A.3})$$

A.3 and the definition of the pH-value are now substituted into A.2 to give

$$\Delta G = 2.303RT(pK_s + \log \frac{[A^-]}{[HA]} - pH). \quad (\text{A.4})$$

For standard concentrations $[HA] = [A^-] = 1$ the second term of A.4 can be set to zero, because

$$\log \frac{[A^-]}{[HA]} = 0, \quad (\text{A.5})$$

so that the standard free energy for the deprotonation at a given pH-value can be computed as:

$$\Delta G = 2.303RT(pK_s - pH). \quad (\text{A.6})$$

[a] C. E. Mortimer and U. Müller, *Chemie*, 8. Auflage, (Thieme-Verlag, Stuttgart, 2003).

LEBENS LAUF

PERSÖNLICHE DATEN

Name	Thomas Andreas Hupp
Geburtsdatum	15.9.1973
Geburtsort	Würzburg
Adresse	Peterstr. 6 97070 Würzburg

SCHULBILDUNG

1980-1984	Grundschule Hafenlohr
1984-1990 und 1991-1994	Balthasar-Neumann-Gymnasium Marktheidenfeld, Abschluss: Abitur
1990-1991	Cheyenne Mountain High School, Colorado Springs, USA

ZIVILDIENTST

07/1994 -09/1995	Sozialstation St. Elisabeth Marktheidenfeld
------------------	---

STUDIUM

10/1995 -09/2000	Studium der Chemie an der Universität Würzburg
07/1998 -10/1998	Praktikum bei Procter & Gamble, Rom, Italien Thema: Entwicklung von Reinigungsmitteln
03/1999 -07/1999	École Supérieure Chimie Physique Électronique de Lyon, Frankreich, Thema: Synthese von Zuckern
02/2000 -09/2000	Diplomarbeit im Arbeitskreis von Prof. Dr. Bernd Engels Titel: <i>Theoretische Untersuchungen zur Tautomerie und Basenpaarung von Xanthin und Uracil</i>
09/2000	Diplom Preis der Fakultät für Chemie und Pharmazie zum Diplom
seit 11/2000	Arbeiten an der Dissertation unter Prof. Dr. Bernd Engels: Titel: <i>Ab Initio Treatment of Complex Systems: Kohn-Sham Orbitals for Multi Reference Methods and the Base Pairing Properties of Xanthine</i> gefördert durch ein Stipendium des Fonds der Chemischen Industrie

Publikationen

M. F. H. Hoffmann, A. M Brückner, T. Hupp, B. Engels and U. Diederichsen: Specific Purine-Purine Base Pairing in Linear Alanyl-Peptide Nucleic Acids, *Hel. Chim. Acta.* **83**, 2580 (2000).

B. Dietrich, T. A. Hupp, B. Engels: Binding Properties in Peptide Nucleic Acids in: Horst Rollnik, Hrsg., *NIC Symposium 2001* (Dietrich Wolf, 2002).

T. Hupp, B. Engels, F. Della Sala and A. Görling: Orbitals from a self-interaction free Kohn-Sham Potential as a single electron basis for ab initio methods, *Chem. Phys. Lett.* **360**, 175 (2002).

T. Hupp, B. Engels, F. Della Sala and A. Görling: Analysis of Kohn-Sham Orbitals for Subsequent MRSD-CI Calculations of Excitation Energies, *Z. Phys. Chem.* **217**, 133 (2003).

T. Hupp, B. Engels and A. Görling: Advantages and Limitations of Kohn-Sham Orbitals as Single Electron Basis for Multireference Configuration Interaction and Multireference Perturbation Theory, *J. Chem. Phys.* **119**, 11591 (2003).

Hiermit erkläre ich an Eides statt, daß ich die Dissertation

“Ab Initio Treatment of Complex Systems
Kohn-Sham Orbitals for Multi Reference Methods
and the Base Pairing Properties of Xanthine”

selbständig angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt habe.

Ich erkläre weiterhin, daß diese Dissertation weder in gleicher oder anderer Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Ich habe früher außer den mit dem Zulassungsgesuch urkundlich vorgelegten Graden keine weiteren akademischen Grade erworben oder zu erwerben versucht.

Würzburg, 15.12.2003

(Thomas Hupp)