2.5.1 Research Area "Ab Initio Methods" (W. Thiel)

**Involved:** J. Breidung, M. Schreiber, M. R. Silva-Junior, A. Yachmenev

Objective: Vibration-rotation and electronic spectra of small molecules are computed with high accuracy using high-level ab initio calculations with large basis sets. This includes the further development of a general variational treatment of nuclear motion that allows the prediction of rovibrational energies and intensities not only for semirigid molecules, but also for molecules with large amplitude motions. Highly correlated ab initio methods are used to provide theoretical benchmark data for the electronically excited states of representative organic chromophores.

Results: The theoretical prediction of vibration-rotation spectra requires the generation of accurate potential energy and dipole moment surfaces, followed by the variational calculation of rovibrational energies and intensities. For the former task, we employ ab initio electronic structure methods, typically coupled cluster theory with large basis sets (e.g., CCSD(T)/aug-cc-pVQZ in standard notation), possibly with complete basis set (CBS) extrapolation and corrections for core-valence correlation and relativistic effects. For the latter, we have developed and coded a variational treatment of nuclear motion that is based on the Hougen-Bunker-Johns approach with an Eckart-frame kinetic energy operator and thus also handles large amplitude motion. This has led to a general and robust variational code (TROVE) which was published in 2007.

Over the past three years, TROVE has been extended in several ways. The convergence of the calculated vibrational energies with increasing vibrational basis set was improved by introducing an extrapolation scheme analogous to the CBS schemes used in electronic structure theory [72]. A reaction path Hamiltonian (RPH) approach was implemented to enable the adiabatic separation of large amplitude motion in TROVE calculations; this procedure converges quickly and is much faster than full calculations [70]. In the course of a comprehensive study of ammonia (see below), several algorithmic improvements were introduced into TROVE to speed up the evaluation of the rovibrational matrix elements and the line strengths [279]. The code was extended to allow for thermal averaging of molecular properties using a novel matrix exponential technique that is more efficient than the standard matrix diagonalization approach; this technique can also be applied to compute the rovibrational partition function [421]. Finally, the DVR (discrete variable representation) technique was implemented as an alternative method for solving the nuclear-motion Schrödinger equation.

The TROVE program has been applied to study NH_3 [279,421], NH_3^+ [142], PH_3 [71,72], SbH_3 [422], XH_2D and XHD_2 (X=Bi, P, Sb) [280] as well as HSOH [70,281,420]. The most extensive project was the calculation of a line list for ammonia ($^{14}NH_3$) covering all rovibrational transitions that involve states up to 12000 cm^{-1} and rotational quantum number $J = 20$ (3.25 million transitions between 184400 energy levels) [279]. The TROVE computations were based on a six-dimensional potential energy surface determined at the CCSD(T)/CBS level with corrections for core-valence correlation and relativistic effects, with a subsequent very slight adjustment of selected potential parameters against experimental data, and on a corresponding CCSD(T)/aug-cc-pVQZ dipole moment surface. The rovibrational spectra calculated at T = 300 K agree very well with the available experimental results from the HITRAN database, as illustrated by the comparisons for overview spectra and for several selected low-lying band systems (see figures) [279]. It should be emphasized that the HITRAN database does not provide experimental data for ammonia in some high-energy parts of the spectrum where the theoretical results thus serve as predictions. This is of particular interest for astrophysical research which requires complete line lists at various temperatures for identification purposes. We are currently calculating a high-temperature line list for ammonia at T=1500 K.

In the case of NH_3^+ [142], PH_3 [71,72], and SbH_3 [422], improved ab initio potential energy surfaces were computed and partly refined against spectroscopic data. They were used in combination with new ab initio dipole moment surfaces [142,422] and polarizability surfaces [142] to generate accurate theoretical rovibrational spectra using variational TROVE calculations. The intensity predictions for PH_3 and SbH_3 were shown to be in accordance with local mode theory [72,422]. For the deuterated species XH_2D and XHD_2 (X = Bi, P, Sb), the formation of rotational energy clusters was observed in the vibrational ground state at high rotational excitation (up to $J = 70$); in the case of XHD_2, the centrifugal-force-induced dynamic symmetry breaking at high J leads to predominant rotation in the cluster states around one of the X-D bonds [280].
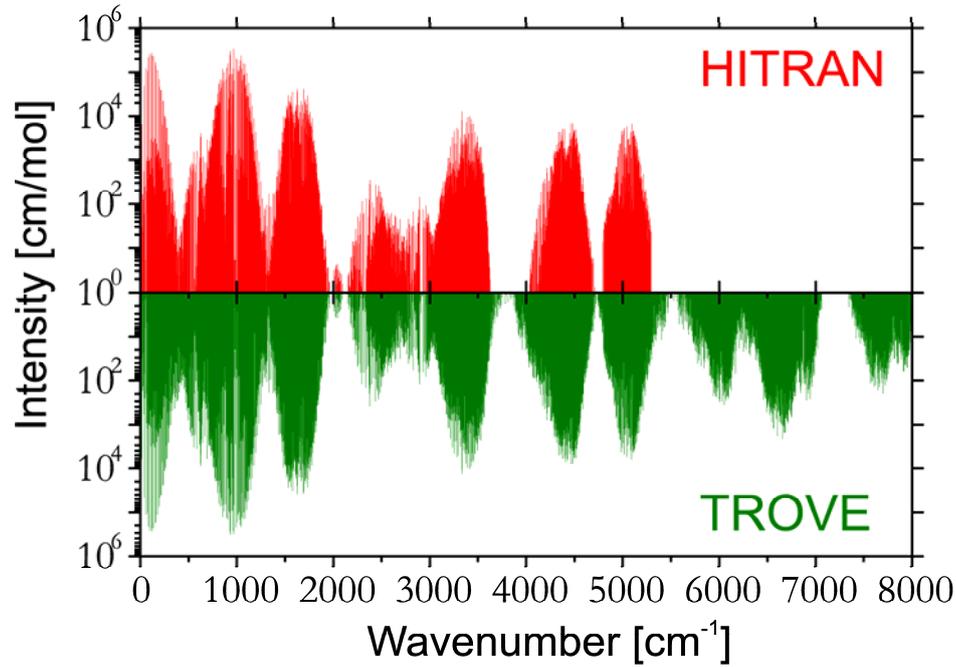
Overview of the simulated absorption (T = 300 K) spectrum (TROVE) of NH_3 compared to experiment (HITRAN), with logarithmic intensity scale.
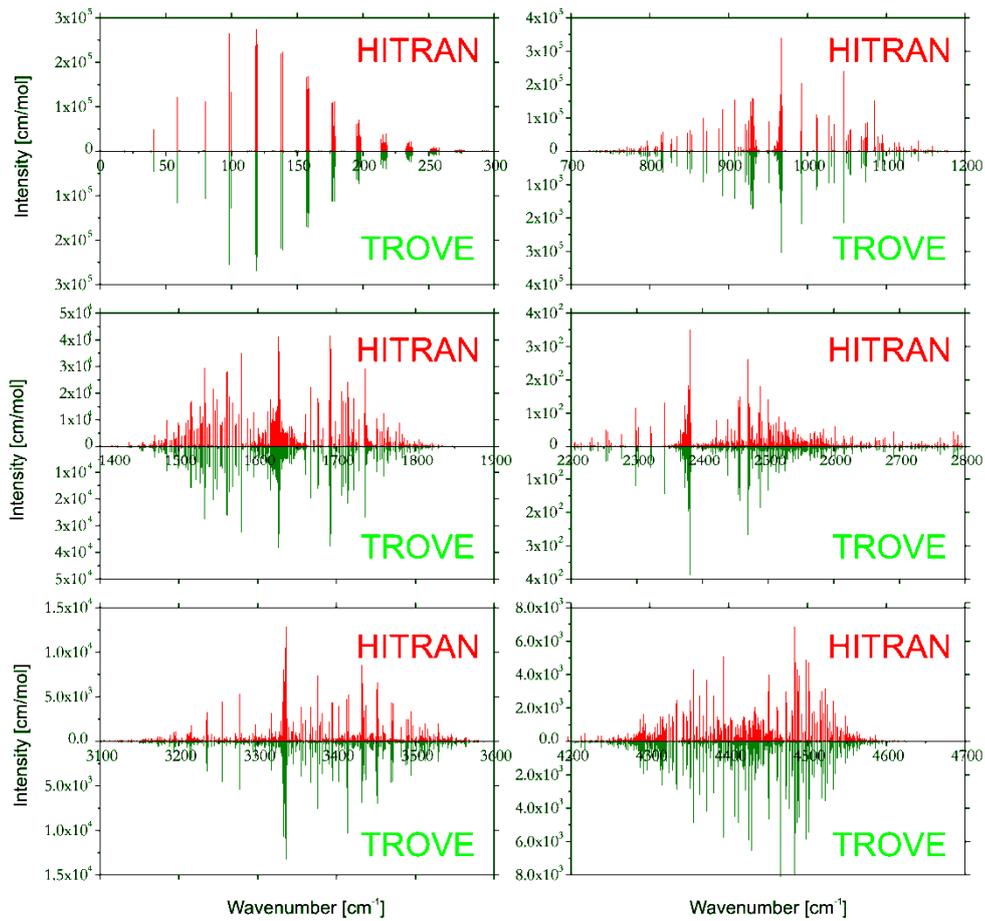
Comparison of the simulated (TROVE) and observed (HITRAN) spectra of NH_3 at T = 300 K for several low-lying band systems.

Our ab initio work on HSOH [70,281,420] was motivated by the challenge to understand the spectroscopic consequences of the large amplitude torsional motion, in particular for the rotation-torsion transitions. High-level ab initio potential energy and dipole moment surfaces were generated and served as input for variational TROVE calculations. Using the RPH approach, the observed torsional splittings could be reproduced within 0.002 cm^{-1} (up to J =40, K_a = 4), and their irregular variation with K_a could be explained in terms of a simple model of rotation-torsion motion [70]. Full-dimensional variational studies of HSOH gave unambiguous assignments of the high-resolution spectra and allowed us to rationalize the extensive perturbations in the fundamental S-H stretch region [281] and the intensity anomalies caused by the large amplitude torsional motion [420].

Further investigations in this area addressed the thermal averaging of the computed nuclear spin-spin coupling constants of NH_3 (focusing on the influence of the inversion motion) [421] and the spectroscopic refinement of the best available ab initio potential energy surface of HDO (leading to excellent agreement with experiment, root-mean-square deviation of 0.035 cm^{-1} for 4495 experimental levels) [143]. In unpublished work, we have computed a CCSD(T)/aug-cc-pVQZ potential surface for formaldehyde H_2CO and refined it against spectroscopic data (rms deviation of 0.05 cm^{-1} for 599 rovibrational levels up to $J = 5$). In thioformaldehyde H_2CS, such a refinement is not possible because of the lack of sufficient high-resolution experimental data (partly due to assignment problems caused by severe resonances); to guide the experimental analysis of the spectra, we have computed a highly accurate, purely ab initio potential energy surface that is based on CCSD(T)/CBS energies (derived from an explicitly correlated F12 treatment) and incorporates various smaller corrections (due to higher-order coupled cluster, core-valence correlation, relativistic, and non-diagonal Born-Oppenheimer effects).

In addition to the ab initio studies on vibration-rotation spectroscopy, we also use ab initio methods in other projects for reference purposes. In the area of spectroscopy, we have performed extensive benchmarking on electronically excited states [106,116,249,401,402,403]. For a set of 28 medium-size reference molecules that cover all typical organic chromophores (213 singlet and triplet excited states), systematic ab initio calculations were carried out using multi-configuration perturbation theory (CASPT2) and coupled cluster theory (CC2, CCSD, CC3), and best estimates were derived for the vertical excitation energies of 104 singlet and 63 triplet excited states [106]. The resulting reference database has subsequently been used by us [116,403] and

by several other groups for the assessment and validation of lower-level theoretical methods. On the ab initio side, we have explored various improvements to the original benchmark [106]. A comparison of noniterative and iterative triples corrections in coupled cluster treatments [CCSDR(3) vs. CC3] showed that the former are essentially as accurate as the latter, so that CCSDR(3) can be applied when CC3 is computationally not feasible [249]. Basis set extension from TZVP to aug-cc-pVTZ had in most cases relatively minor effects on the vertical excitation energies from coupled cluster [401] and CASPT2 [402] calculations, but the changes were still considered large enough to warrant an update of the recommended best estimates for the vertical excitation energies [402].

Publications resulting from this research area: 70, 71, 72, 106, 142, 143, 249, 279, 280, 281, 290, 401, 402, 420, 421, 422

External funding: European Research Training Network QUASAAR (MRTN-CT-2004-512202); Deutsche Forschungsgemeinschaft (SFB 663, project C4)

Cooperations: J.-M. Flaud (Paris, FR); T. F. Giesen (Köln, DE); P. Jensen (Wuppertal, DE); S. P. A. Sauer (Copenhagen, DK); J. Tennyson (London, UK); S. N. Yurchenko (Dresden, DE); other QUASAAR partners include A. Campargue (Grenoble, FR), L. Halonen (Helsinki, FI), M. Herman (Brussels, BE), T. Rizzo (Lausanne, CH)

2.5.2   Research Area "Density Functional Methods" (W. Thiel)

Involved: A. Anoop, J. Gonthier, B. Heggen, M. Patil, T. Tuttle, D. Wang

Objective: Density functional methods are applied in studies of transition metal and other compounds in order to understand and predict their properties. Much of the work on homogeneous transition metal catalysis and organocatalysis involves a close collaboration with experimental groups at our Institute and aims at a detailed mechanistic understanding of the reactions studied experimentally.

Results: Many of our applications of density functional theory (DFT) focus on transition metal compounds. Based on previous extensive validation our standard DFT approach normally employs the Becke-Perdew functional with an effective core potential at the metal and with medium-sized polarized basis sets. In the case of organic compounds, we normally use the B3LYP hybrid functional.

In joint work with the Fürstner and Alcarazo groups, we have studied a number of ligands and transition metal complexes that are relevant for their experimental work on homogeneous catalysis [149,176,287,298]. The computed structures were generally found to be in good agreement with the available crystal structures, and the computed metal-ligand binding energies generally reflected the experimentally observed trends. The computed molecular orbitals and natural bond order (NBO) analysis were used to characterize the electronic structure of these compounds. To be more specific, neutral and cationic 14-electron gold alkyne complexes were investigated to gain insight into the first elementary step common to many gold-catalyzed transformations; in the complexes with cyclododecyne, $\pi \rightarrow \sigma^*$ donation clearly dominates over d(Au)$\rightarrow \pi^*$ backdonation, thus explaining the highly electrophilic character of the alkyne within the coordination sphere of any chosen Au(I) fragment [176]. The π-acceptor properties of N-heterocyclic carbenes (NHCs), commonly used as ligands in gold catalysis, were studied for cyclophanic and triazo analogues of imidazopyridine-2-ylidene derivatives; it was established that the π-acceptor strength can easily be tuned in these ligands to control the course of three mechanistically distinct gold-catalyzed processes [287]. The captodative binding in carbon(0) coordination compounds was probed experimentally by auration, i.e., by the reaction between $AuCl(SMe_2)$ and $(PPh_3)CL^2$ with a variety of ligands L^2; in line with the experimental findings, the computed complexation energies predicted diauration to be possible only for two specific ligands L^2 (monoauration always), and population analysis indicated that the central carbon atom carries a

substantial negative charge not only in the free ligand, but also in the monoaurated and diaurated complexes – a counterintuitive behavior typical of carbon(0) coordination compounds [149]. A subsequent investigation of highly electron-rich imines, in particular cyclopropenylylidenimines, showed that they can be classified analogously as nitrogen(I) compounds with two lone pairs at nitrogen that are prone to interact with Lewis acids, so that they may be valuable as ligands in homogeneous catalysis [298].

Experimental work in the Reetz group has shown some time ago that rhodium catalysts with chiral monodentate phosphorous ligands can achieve asymmetric hydrogenation with high efficiency and enantioselectivity, and may thus serve as an economic alternative to the classical catalysts with bidentate ligands. Unlike the latter, the new catalysts follow the "lock-and-key" principle, i.e., the major enantiomer of the product is formed from the more stable diastereomeric prochiral catalyst-substrate complex. We have continued a detailed DFT study of the enantioselective hydrogenation of itaconic acid using a chiral $Rh(phosphonite)_2$ catalyst and calculated the previously missing parts of the catalytic cycle (i.e., reactions with solvent molecules in the initial and final phase of the cycle), to set the stage for kinetic Monte Carlo simulations aimed at understanding the origin of the observed enantioselectivity. Since such simulations depend critically on the adopted relative energies and barriers, we have checked the accuracy of the applied DFT methods in a related case [290], i.e., the asymmetric hydrogenation of two prochiral enamides using a rhodium catalyst with a bidentate phosphorus ligand, $[(R,R)\text{-MeDuPHOS}]^+$. Since this system is smaller than the Reetz system, it could be treated by accurate local pair natural orbital coupled cluster theory with single and double excitations (LPNO-CCSD). The DFT(B3LYP) results were found to be in reasonable agreement with the single-point LPNO-CCSD results, although some small deviations exist that are not entirely systematic in the minor and major reaction branches [290].

Several DFT studies have been carried out without involvement of experimental groups from the Institute [141,271,337]. In continuation of an external cooperation with an industrial partner, we have computed the mechanism of olefin hydrosilylation catalyzed by dichloro(p-cymene)ruthenium(II) and some related complexes; as in previous cases, it was found that the reaction involves an initial exchange of one of the σ-donor ligands for a hydride ligand in the induction step [271]. A combined experimental and theoretical study addressed the electronic structure of a series of iron-corrole complexes with strongly electron-deficient corrole ligands; the combination of Mössbauer data with DFT calculations provided convincing evidence for an intermediate-spin iron(III)

ion antiferromagnetically coupled to a dianionic π-radical corrole yielding an overall triplet ground state [141]. Ongoing unpublished work concerns the mechanism and the origin of selectivity in Pd-catalyzed allylic alkylation reactions.

In our benchmarks of DFT-based methods for electronically excited states, the DFT/MRCI approach (DFT combined with multi-reference configuration interaction) reproduced the ab initio reference data for vertical excitation energies best and clearly outperformed standard time-dependent DFT approaches [116]. We have applied DFT/MRCI to study the photophysics of phenalenone, a widely used singlet oxygen sensitizer, and indeed found a very high intersystem crossing rate, much larger than the fluorescence and phosphorescence rates, which rationalizes the experimentally observed behavior of phenalenone [170].

Publications resulting from this research area: 116, 141, 149, 170, 176, 271, 287, 298, 337

Cooperations: M. Alcarazo, A. Fürstner, B. List, N. Maulide, M. T. Reetz (Mülheim/Ruhr, DE); M. Braun (Düsseldorf, DE); B. Ganguly (Gujarat, IN); Z. Gross (Haifa, IL); F. Neese (Bonn, DE)

2.5.3   Research Area "Semiempirical Methods" (W. Thiel)

Involved: J. Breidung, E. Fabiano, T. Keal, A. Koslowski, Z. Lan, Y. Lu,
M. R. Silva-Junior, T. Tuttle, O. Weingart, X. Wu

Objective: This long-term project aims at the development of improved semiempirical quantum-chemical methods that can be employed to study ever larger molecules with useful accuracy. This includes the development of more efficient algorithms and computer programs. Our current focus in this area is on electronically excited states.

Results: Over the past years, we have developed semiempirical methods that go beyond the standard MNDO model by including orthogonalization corrections at the NDDO level. This has led to three new approaches labeled OM1, OM2 and OM3 (orthogonalization models 1-3) which offer significant improvements over established MNDO-type methods in several areas, including conformational properties, hydrogen bonds, reaction barriers, and electronically excited states.

During the reporting period, we have concentrated on semiempirical excited-state dynamics. In our MNDO software, we had previously implemented the GUGACI method in a semiempirical in-core version that handles general CI expansions (up to full CI) efficiently for small active spaces. The software had already included an analytic GUGACI gradient code which has now been extended to the analytic computation of the full nonadiabatic coupling matrix elements. Building on these ingredients, we have implemented [23] the surface hopping method with the fewest switches algorithm (Tully). Our code offers different integration schemes as well as different options for treating the nonadiabatic coupling and ensures proper orbital and state tracking. In an initial validation with three case studies (ethylene, methaniminium ion, and methanimine), the OM2-CI approach yielded decay times and dynamics paths similar to high-level ab initio results [23]. Subsequently more approximate switching algorithms were tested that adopt simpler models for the hopping probability; it was found that one popular variant based on a local diabatic representation shows qualitative failures and overestimates the decay times severely (compared with the standard fewest switches algorithm) [22]. In unpublished work, several decoherence corrections were implemented and tested to establish a proper default procedure.

The semiempirical surface hopping module was used to study the gas-phase excited-state dynamics of adenine [24], guanine [200], uracil, thymine, and cytosine [201] at the

OM2/MRCI level. For all these nucleobases, the OM2/MRCI energies and geometries of the relevant species (including conical intersections) showed good agreement with the available ab initio reference data. The OM2/MRCI surface hopping trajectory calculations gave sub-ps relaxation via different pathways, and the experimental decay times from time-resolved fs photoelectron spectroscopy were well reproduced in all cases. The excited-state dynamics of adenine involves excitation into the bright second excited state (L_a, $\pi\pi^*$, S_2) and then proceeds mostly (90%) by an ultrafast relaxation to the $n\pi^*$ state (S_1) followed by sub-ps decay to the ground state (see the figure for a typical trajectory); the minor decay channel (10%) leads directly to the ground state (gs) via a L_a/gs conical intersection [24]. In the other nucleobases, the dynamics also involves the lowest two excited states, but the nonadiabatic decay paths differ in their relative importance and need to be assessed separately for each molecule (e.g., because of differences in the energy level schemes) [200,201].
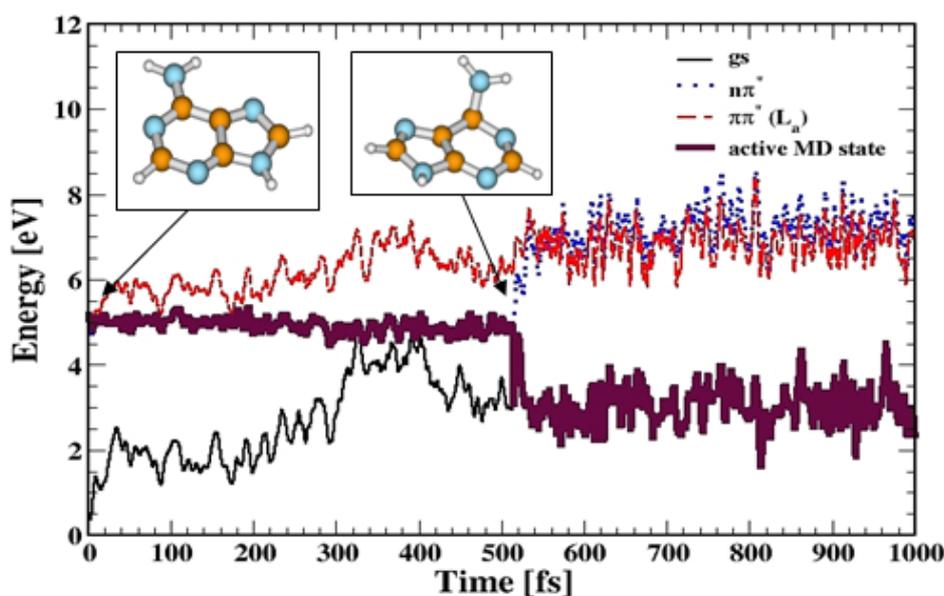
State energies for a typical trajectory following the ($n\pi^*$/gs)CI decay path in adenine.

We have also implemented our surface hopping code in the ChemShell QM/MM framework and applied it to study the nonadiabatic dynamics of nucleobases in aqueous solution (adenine, guanine) and in model DNA strands (adenine). Going from the gas phase to aqueous solution causes rather small changes in the optimum geometries (e.g., of conical intersections). The relaxation paths via the various conical intersections remain qualitatively similar, but there are changes in their relative importance because of the solvent-induced shifts in the energetics. Overall, the decay times in solution are even slightly lower than those in the gas phase, in qualitative agreement with

experiment. By contrast, the S_1 lifetime of adenine in a solvated B-type DNA oligomer model $(dA)_{10}$ is computed to be about ten times larger (ca. 5 ps) than in the gas phase, again in accordance with experiment. This is at least partly due to the higher rigidity of the DNA strand: the two relevant conical interactions are both characterized by strong out-of-plane deformations that are more difficult to reach in the DNA environment. In our DNA double strand model, solvated $(dA)_{10} \cdot (dT)_{10}$, one of these two decay paths is suppressed because the out-of-plane motion of the amino group is impeded by inter-strand hydrogen bonding (unpublished).

We are currently investigating the photoinduced dynamics of several other systems using OM2/MRCI surface hopping simulations. These include azobenzenes in the gas phase and in a condensed-phase environment, the chromophore in selected red fluorescent proteins, Feringa-type fluorene-based molecular motors, and a prototypical charge transfer molecule (4-*N,N*-dimethylamino-benzylidene malononitrile). Already published is an OM2/MRCI investigation on the photoisomerization of a protonated Schiff base that serves as a retinal model [193].

To assess the accuracy that can generally be expected in semiempirical studies of excited states, we have evaluated the performance of several popular semiempirical methods for vertical excitation energies, oscillator strengths, and excited-state dipole moments [403] using our standard ab initio benchmark set [106]. The OMx/MRCI methods were found to be much better than standard MNDO-type ground-state methods (as expected), and surprisingly also slightly better than the INDO/S excited-state method. In fact, the mean absolute deviations of the OMx/MRCI results were only somewhat higher than those obtained for standard time-dependent DFT results.

Turning to the electronic ground states, we have incorporated empirical dispersion corrections into the OMx approaches, in analogy to corresponding work by others at the DFT level; this led to significant improvements for non-covalent interactions, with mean absolute errors of 1.3-1.4 kcal/mol in the OMx-D binding energies of the complexes in the standard JSCH-2005 data base [130]. These advances support the use of dispersion-corrected OMx-D methods in studies of biomolecules, as demonstrated for the example of antibody-hapten binding [130].

Ground-state molecular dynamics (MD) simulations of large systems can be computationally very demanding even at the QM/MM level and often become practical only with semiempirical QM components. We have adopted this approach in OM3/MM

MD simulations of a QM water solute in liquid MM water to assess the merits of a polarizable water force field [39] and partly also in a systematic study of finite-temperature effects in enzymatic reactions by QM/MM free-energy calculations [259]. To further enhance the efficiency of such simulations, we have implemented a generalized solvent boundary potential (GSBP) for hybrid QM/MM MD simulations with semiempirical MNDO-type QM methods [8].

Finally, in a combined experimental and theoretical study [207], we used semiempirical methods to investigate fluorophores as optical sensors for local forces. In the case of OPV5, an oligoparaphenylene vinylene derivative with twisted backbone, correlations were observed experimentally between the external mechanical force (applied to the fluorophore embedded in a foil) and the changes in the fluorescence properties. Both the blue-shift of fluorescence and the reduced lifetime of OPV5 under tensile stress were consistent with the results of the semiempirical calculations [207].

Publications resulting from this research area: 8, 22, 23, 24, 39, 130, 193, 200, 201, 207, 259, 393, 403

Cooperations: M. Filatov (Groningen, NL); B. Hartke (Kiel, DE); G. Groenhof (Göttingen, DE); C. M. Seidel (Düsseldorf, DE)

2.5.4 Research Area "Combined Quantum Mechanical / Molecular Mechanical Methods" (W. Thiel)

Involved: M. Altarsha, A. Altun, T. Benighaus, M. Bocola, M. Doerr, E. Fabiano, Y. Hsiao, J. Kästner, T. Keal, D. Kumar, Z. Lan, Y. Lu, S. Metz, N. Otte, M. Parac, I. Polyak, E. Sanchez-Garcia, H. M. Senn, M. R. Silva-Junior, S. Thiel, T. Tuttle, D. Wang, O. Weingart, J. Zheng

Objective: This research focuses on hybrid approaches for large systems where the active center is treated by an appropriate quantum mechanical method, and the environment by a classical force field. It involves considerable method and code development. This approach allows a specific modeling of complex systems such that most of the computational effort is spent on the chemically important part. Current applications primarily address biocatalysis and aim at a better understanding of enzymatic reactions including the role of the protein environment.

Results: Combined quantum mechanical/molecular mechanical (QM/MM) methods have become a popular tool for studying reactions in complex systems such as enzymes. Typical applications make use of density functional theory (DFT) or semiempirical methods as QM component and a standard biomolecular force field (e.g., CHARMM or GROMOS) as MM component. The accuracy of QM/MM calculations can be improved by going beyond these standard methods both in the QM and MM part.

For this purpose, we have explored the use of correlated ab initio methods in two QM/MM case studies on *p*-hydroxybenzoate hydroxylase (PHBH) and aldehyde oxidoreductase (AOR), in collaboration with the Werner group [61,303]. This involved DFT(B3LYP)/MM optimizations of reaction paths and stationary points followed by single-point ab initio QM/MM energy evaluations using local correlation methods up to the LCCSD(T0) coupled cluster level. In both cases, the qualitative conclusions from the preceding DFT(B3LYP)/MM studies were confirmed, but there were non-negligible changes in the computed relative energies and barriers. Careful validation of the applied local correlation methods with regard to all computational parameters indicated convergence of the QM contribution to the barriers to within 1 kcal/mol in PHBH [61] and 2 kcal/mol in AOR [303].

In a similar vein, we have carried out a multi-reference ab initio QM(MRCI)/MM study of three intermediates in the catalytic cycle of cytochrome P450cam [2]. The calculated spin state energies are generally quite similar to those obtained at the simpler DFT(B3LYP)/MM level, despite the fact that two of these intermediates are multiconfigurational in character. This provides further justification for the use of DFT(B3LYP) in this area.

The accuracy of the MM treatment can be enhanced, in principle, by moving from standard force fields with fixed MM charges to polarizable force fields. In a cooperation with the van Gunsteren group, we had previously implemented a particular variant of polarizable force fields, the charge-on-spring model, into our ChemShell QM/MM software; this had been done through a special interface to the GROMOS MM code. Following the modular ChemShell philosophy, we have now carried out a general implementation that can be coupled with any MM code since the polarizable terms are handled in a generic ChemShell module. On the application side, we have shown that semiempirical OM3/MM MD simulations of a QM water solute in liquid MM water benefit from the use of a polarized MM model for water [39].

In QM/MM work, geometry optimization techniques are commonly employed to determine reaction paths and the relevant minima and transition states. We have extended the range of optimizers available in the DL-FIND library that is used in ChemShell, the most notable additions being the nudged elastic band method for path optimization, the dimer method for transition state searches, and three algorithms for finding conical intersections [190].

For the calculation of free energies that govern reaction rates, several methods are available in ChemShell (including thermodynamic integration, umbrella sampling, and free energy perturbation theory). These methods employ MD simulations along pre-computed reaction paths for conformational sampling. We have used these techniques to study finite-temperature effects in enzymatic reactions by QM/MM free energy simulations [259], covering reactions catalyzed by chorismate mutase (CM), fluorinase, PHBH, and cytochrome P450cam. The entropic contribution to the barrier was generally found to be rather small (mostly less than 1 kcal/mol). This suggests that barriers from QM/MM geometry optimization may often be close to free energy barriers for enzymatic reactions that involve only localized small-scale geometry changes within the active site [259].

In further methodological work, we have extended the two-layer QM/MM approach to a three-layer model by introducing boundary potentials that represent the outer part of the MM region and the bulk solvent. This offers two major advantages: conceptually, the long-range electrostatic interactions in a solvated enzyme are well described in this manner, and technically, the computational effort is reduced significantly by the strong reduction of the number of explicitly treated MM atoms (typically 2000 compared with around 30000 in standard QM/MM work). Initially, we adopted the generalized solvent boundary potential (GSBP) proposed by Roux for MM-MD simulations, and implemented it in a semiempirical QM/MM framework [8]. This required a new ChemShell module for performing Poisson-Boltzmann calculations that was thoroughly optimized for performance. Because of the associated overhead, the semiempirical QM/MM/GSBP approach is faster than the standard QM/MM treatment only for system sizes beyond 12500 atoms (e.g., by a factor of about 3 for 30000 atoms); it is thus an efficient tool for large-scale QM/MM MD simulations [8]. In a second step, we developed a new boundary potential (SMBP, solvated macromolecule boundary potential) that is conceptually similar to GSBP, but can be used with any QM method and is also efficient for geometry optimization [153]. Its generality derives from the use of virtual point charges on a sphere surrounding the explicitly modeled inner region, which represent the boundary potential arising from the outer region and the bulk solvent in the sense that they generate essentially the same electrostatic potential in the QM region. These virtual point charges and the QM wavefunction are determined in a self-consistent reaction field procedure. QM/MM/SMBP geometry optimizations on enzymes reproduce the standard QM/MM results very well, and it is therefore possible to calculate free energy profiles and barriers along QM/MM/SMBP optimized reaction paths, which reduces computation times typically by one order of magnitude without significant loss of accuracy [153]. In an initial application of these techniques, we studied the influence of long-range electrostatic interactions on the enzymatic reactions in CM and PHBH. The corresponding energetic effects were found to be non-negligible (several kcal/mol) for the hydroxylation reaction in PHBH during which there is a formal transfer of one negative charge from the substrate to the cofactor. The QM/MM/SMBP approach allows a separate analysis of the effects due to the outer MM region and the bulk solvent [294].

A final methodological advance concerns the implementation of quantum refinement into ChemShell [323]. The traditional procedure for solving protein crystal structures is to supplement the X-ray diffraction data with restraints from simple MM force fields during the refinement. These restraints are taken from QM/MM energies in quantum

refinement which should thus be more reliable especially in regions that are not well described by the available MM force fields (e.g., metal centers, inhibitors, substrates, and chromophores). Our implementation goes beyond the original proposal (Ryde) in several aspects, most notably by also allowing for an electrostatic QM/MM embedding (instead of a simple mechanical embedding). As a first application, we have used quantum refinement to derive an improved structure for the DsRed.M1 fluorescent protein, especially in the chromophore region [323]. A similar improvement was achieved for the pB_2 intermediate of the photoactive yellow protein (submitted).

Turning to QM/MM applications on enzymatic reaction mechanisms, we have continued our studies on the catalytic cycle of cytochrome P450cam, mostly at the DFT(B3LYP)/CHARMM level (partly in collaboration with the Shaik group). In the reporting period, we characterized the ferrous dioxygen and ferric peroxo complexes of P450cam [273] and compared the electronic structure of the reactive species (Cpd I) in several different heme thiolate enzymes [20]. As indicated above, ab initio QM(MRCI)/MM calculations were performed for a detailed analysis of the electronic structure of Cpd I, the ferric hydroxo complex (Cpd 0), and the hydroxo complex formed during camphor hydroxylation [2]. In addition, we have investigated several key reactions, in particular the proton transfers that generate Cpd 0 [134] and Cpd I [150,151,289] in wild-type P450cam and in several mutants. Protonation of the distal oxygen atom in Cpd 0 leads to the formation of Cpd I and water (coupling reaction), whereas protonation of the proximal oxygen atom yields the ferric resting state and hydrogen peroxide (uncoupling). In agreement with experiment, the DFT(B3LYP)/CHARMM calculations predict exclusive coupling in wild-type P450cam and exclusive uncoupling in the Thr252X (X = alanine, glycine) mutants, because in the latter case an extra water molecule can enter the hydrogen bonding network in the proton transfer path for uncoupling and thus favor this channel over coupling [150]. Detailed mechanistic insight was also gained in analogous QM/MM studies on the D251N mutant [151] and the methoxy-threonine mutant [289]. The role of single water molecules as biocatalysts in P450cam chemistry was analyzed in general terms [343]. Finally, a comprehensive review was written together with Sason Shaik to summarize the available QM/MM work on the structure, reactivity, and selectivity of P450 enzymes [400].

Concerning molybdenum enzymes, we have studied the reaction mechanism of the oxidation of acetaldehyde to acetic acid catalyzed by aldehyde oxidoreductase (AOR) and the conversion of xanthine to uric acid by xanthine oxidase (XO). In both cases, a

number of different pathways were explored to elucidate the productive binding mode of the substrate and the catalytic effect of active-site residues. The DFT(B3LYP)/CHARMM calculations on AOR favored a Lewis base catalyzed mechanism with initial deprotonation of the cofactor by Glu869 over the alternative one-step and two-step mechanisms that had been considered previously in DFT model studies; the reaction then proceeds by an attack of the activated cofactor at the substrate and a subsequent formal hydride transfer [213]. This mechanism was confirmed by high-level ab initio and free energy QM/MM calculations on AOR which did not change the mechanistic preferences and provided best estimates for the free energy barriers that differ by up to 3 kcal/mol from the DFT(B3LYP)/CHARMM barriers [303]. In the case of XO, the situation is more complicated since the xanthine substrate can adopt different orientations in the binding pocket (upside vs. upside-down) and convert between different tautomer and protonation states. Systematic DFT(B3LYP)/CHARMM calculations on XO gave a clear preference for one particular pathway: three initial proton transfers (involving Glu1261) activate both the cofactor and the substrate such that the subsequent nucleophilic and hydride transfer become more facile, through electrostatic stabilization of the rate-limiting transition state by Arg880 [214]. This is illustrated in the figure by showing the two relevant xanthine tautomers, i.e., the most stable form in the gas phase and the reactive form in the enzyme. Additional QM/MM studies on XO addressed the effects of variations in the cofactor, the substrate, and the active-site Glu802 residue; the computational results were fully consistent with the corresponding experimental findings, thus providing further support for our calculated mechanism [361].
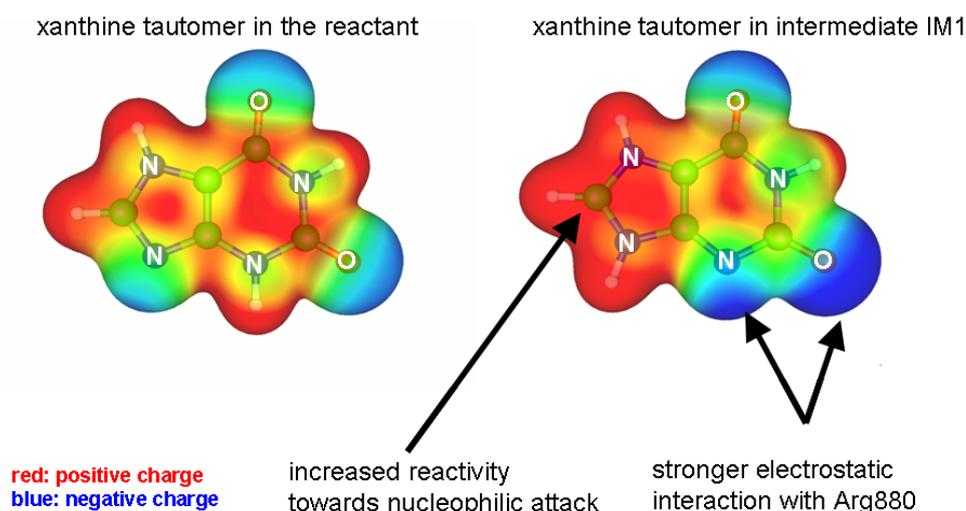
Electrostatic potentials of the most stable (left) and the reactive (right) tautomer of xanthine.

In collaboration with the Bühl group, we have assessed QM/MM models of the peroxo forms of vanadium-containing haloperoxidases (VHPOs) in terms of active-site geometries, hydrogen bonds within the active site, isotropic and anisotropic [51]V NMR chemical shifts, and TD-DFT excitation energies [185]. The models showed appreciable differences between the anisotropic chemical shifts of the different protonation states. The most likely candidates for the peroxo forms of VHPO enzymes were identified on the basis of QM/MM modeling in combination with X-ray, [51]V NMR, and UV-Vis data [185].

In collaboration with the Engels group, we have performed QM/MM studies on cysteine proteases to gain insight into the structure and stability of the resting state and the mechanisms of inhibition. The DFT/CHARMM calculations favored the zwitterionic resting state over the neutral form by about 8-9 kcal/mol due to a complex hydrogen bonding network involving several active-site residues and water molecules that stabilizes the zwitterionic form [64]. These studies also clarified the regiospecificity and inhibition potency of epoxide- and aziridine-based inhibitors [65] as well as the stereoselectivity of epoxide-based inhibitors [63]; such insights are expected to be helpful for rational drug design.

In an ongoing collaboration with the Reetz group that is motivated by their experimental work on directed evolution, we examine the mechanism of the Baeyer-Villiger oxidation reaction in cyclohexanone monooxygenase (CHMO). The currently available QM/MM results from reaction path calculations and from the optimization of the Criegee intermediate and the subsequent transition state already allow us to at least partially explain the high regioselectivity of CHMO towards certain cyclohexanone derivatives, but further work is needed to arrive at a more complete picture.

Turning to QM/MM applications on electronically excited states, we have already mentioned the semiempirical QM(OM2/MRCI)/MM surface hopping studies of nucleobases in water and in DNA strands. We have also performed QM/MM calculations of electronic absorption spectra in the condensed phase, mostly using DFT/MRCI as QM component. The corresponding protocols were established and discussed in much detail in an investigation of solvent effects on the absorption spectrum of guanine [374]. Concerning proteins, the properties and the electronic spectra of the chromophores in the red fluorescent proteins DsRed.M1 [248,393] and HcRed [405] were computed, and the influence of the LOV domain of the blue-light photosensor YtvA of *Bacillus subtilis* on the low-lying electronic states of the flavin

mononucleotide chromophore was analyzed and contrasted with the solvent shifts in water [247].

The ChemShell software that has been used in all these applications is available under a license agreement (see www.chemshell.org). The QM/MM methodology and QM/MM applications to biological systems have been reviewed [260,267].

Publications resulting from this research area: 2, 8, 20, 39, 61, 63, 64, 65, 134, 150, 151, 153, 185, 190, 213, 214, 222, 247, 248, 259, 260, 267, 273, 289, 294, 303, 323, 343, 361, 374, 393, 400, 405

Cooperations: M. Bühl (St. Andrews, UK); B. Engels (Würzburg, DE); E. Keinan (Haifa, IL); C. M. Marian (Düsseldorf, DE); F. Neese (Bonn, DE); M. T. Reetz (Mülheim/Ruhr, DE); S. Shaik (Jerusalem, IL); P. Sherwood (Daresbury, UK); S. C. Smith (Brisbane, AU); W. F. van Gunsteren (Zurich, CH); H.-J. Werner (Stuttgart, DE)

2.5.5   Research Area "Simulations of Photoexcited Molecules and Dynamical Processes" (M. Barbatti)

Involved: R. Crespo-Otero

Objective: Determination of steady and time-dependent properties of UV/vis excited molecules by means of semi-classical simulations based on ab initio quantum chemical methods.

Results:

1. Photodynamics of nucleobases and nucleobase models: We have performed excited-state dynamics simulations for all five naturally occurring DNA/RNA bases after UV-excitation. This research, carried out with our collaborators in Prague and Vienna, has allowed us to propose a very complete scenario for the mechanisms driving the internal conversion of the nucleobases. These mechanisms are schematically illustrated in the figure below.
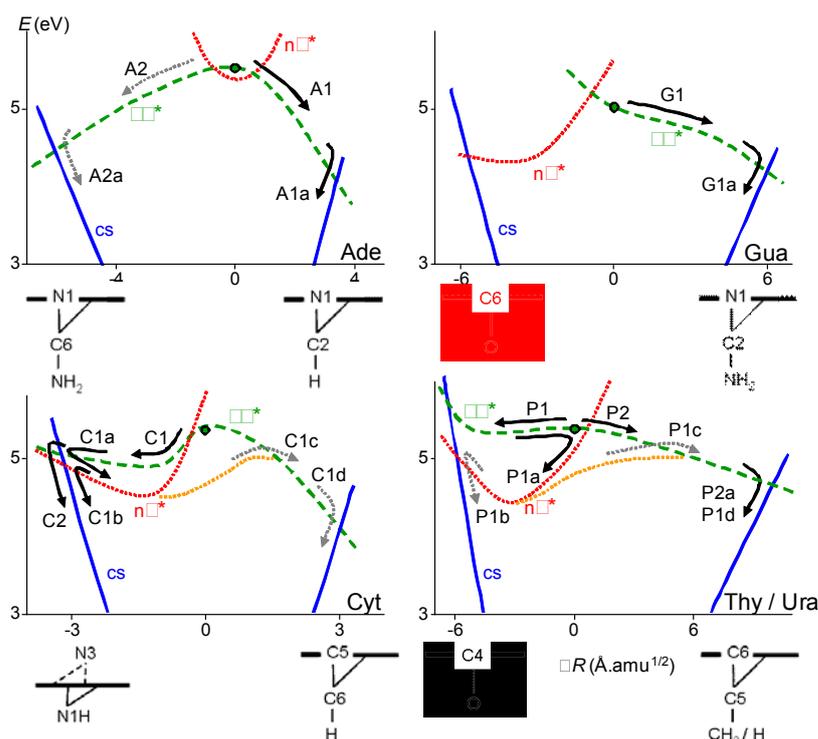
Deactivation mechanisms in UV-excited nucleobases as determined by dynamics simulations. Purine bases (ade, gua) follow homogenous pathways along the $\pi\pi^*$ state. Pyrimidine bases (cyt, thy, ura) follow inhomogeneous pathways composed by several steps.

Besides the nucleobases themselves, we have also investigated the nonadiabatic dynamics of the nucleobase models aminopyrimidine and diaminopyrimidine. The investigation of such models aims at determining the role of specific site restrictions to the internal conversion. First, we have systematically changed the substitutions in aminopyrimidine so as to understand how the imidazole group affects the out-of-plane modes of the pyrimidine ring in adenine during the dynamics. Moreover, we