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

Involved: J. Breidung, H. Lin, F. Mariotti, S. N. Yurchenko, J. Zheng

Objective: High-level ab initio methods with large basis sets are used to compute the structures, stabilities, and vibration-rotation spectra of small molecules. The theoretical predictions need to be reliable enough to guide the spectroscopic identification of short-lived molecules and to assist in the analysis of their high-resolution vibration-rotation spectra both in the fundamental and overtone regions. They are required to be accurate also for molecules with large amplitude motions.

Results: Our ab initio studies are carried out in close cooperation with external experimental groups. Employing standard ab initio notation, we normally calculate geometries, energies, and harmonic force fields at the CCSD(T)/cc-pVQZ level, and anharmonic force fields at the MP2/cc-pVTZ or CCSD(T)/cc-pVTZ level. In our experience, theoretical results at these levels are generally adequate to provide useful guidance for our experimental partners. If highest accuracy is needed, we use even more extended basis sets and extrapolations to the complete basis set (CBS) limit. Our work in this area has been reviewed recently [428].

Bismuthine BiH_3, the heaviest of the ammonia-type hydrides XH_3, is a textbook molecule which has been synthesized in 1961, but has remained poorly characterized because of its instability and difficulties in repeating its preparation. Using a combined experimental and theoretical approach, BiH_3 has been detected in the gas phase by high resolution infrared and millimeter-wave spectroscopy [65]. The ab initio predictions are in good agreement with the observed spectra [65,322,363]. In order to achieve high accuracy for the vibrational frequencies, it is necessary to include spin-orbit effects by an appropriate configuration interaction treatment, in spite of the fact that BiH_3 is a closed-shell molecule [322]. Accurate equilibrium geometries and spectroscopic parameters are now available for BiH_3 [322,363].

Xenon compounds with a formal Xe-C bond are quite rare and unstable. Following up on our earlier work on $XeCCF_2$, the complex XeCC has been studied [18]. The molecule has a linear structure and a considerable dipole moment due to charge transfer from xenon to dicarbon. The calculations recover most of the observed red shift in the CC stretching frequency upon complexation and thus support the reported identification

of XeCC in a xenon matrix. The electronic structures of XeCC and $XeCCF_2$ are similar [18].
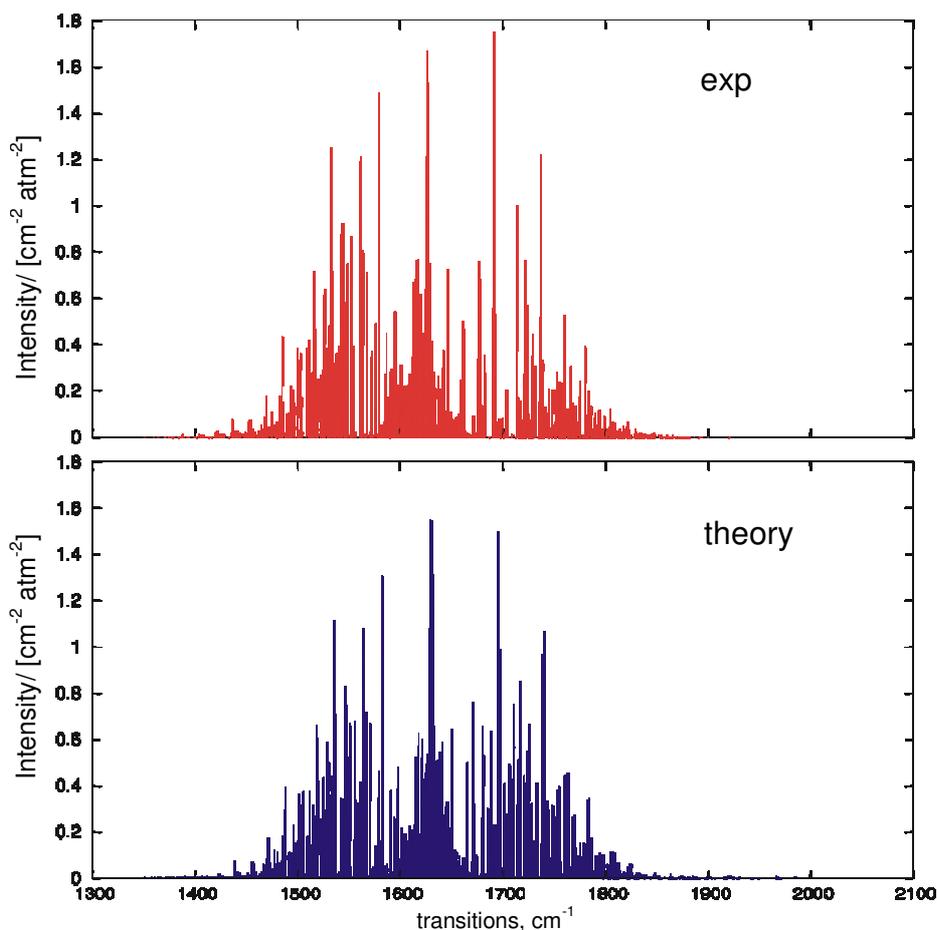
The fluorocarboxyl radical FCO_2 plays an important role in atmospheric chemistry. This short-lived species has been generated by low-pressure pyrolysis of the peroxide dimer in the gas phase and detected by high resolution spectroscopy [303]. Observation and identification of its millimeter-wave spectrum have relied on accurate ab initio predictions at the open-shell coupled cluster level [303]. The excellent agreement between experiment and theory confirms the structure of the FCO_2 radical whose ground-state parameters are now well established [303].

Vanadium hexacarbonyl $V(CO)_6$ is the only homoleptic transition-metal carbonyl that exists near room temperature as a free radical. It is expected to undergo Jahn-Teller distortion from octahedral to lower symmetry, but numerous previous investigations had not reached a consensus on the nature of the distortion. Open-shell coupled cluster calculations, and static and dynamic density functional computations, combined with new experimental vibrational spectra and a reinterpretation of published ESR data, show unambiguously that at temperatures below 20 K, $V(CO)_6$ isolated in inert matrices exhibits a static Jahn-Teller distortion to a D_{3d} structure [161], ruling out the previously suggested D_{4h} distortion. At room temperature, the molecule is highly fluxional with an average octahedral structure [161].

The preceding studies, as well as those on F_2BOH [321], CHF_3 [320], and NF_3 [170], are based on the calculation of anharmonic force fields followed by a second-order perturbation treatment of nuclear motion. This standard approach for semirigid molecules is no longer suitable for vibrationally excited polyatomic molecules with high vibrational energy and for floppy molecules with large amplitude motion. In such cases, it is necessary to resort to a variational treatment.

We have tackled this task in a case study on ammonia, the prototype of a molecule with large amplitude inversion motion, and attempted to compute its complete vibration-rotation-inversion spectrum as accurately as possible. This involves the calculation of six-dimensional (6D) potential energy surfaces at the CCSD(T) level with CBS extrapolation and corrections for core-valence correlation and relativistic effects [85], the fitting of such surfaces to a suitable analytical form, the development of an appropriate 6D variational treatment of nuclear motion, and the actual variational computations with construction and diagonalization of the corresponding Hamiltonian

matrices. The adopted variational model is based on the Hougen-Bunker-Johns approach with an Eckart-frame kinetic energy operator. The published results for the vibrational energies [85] show overall rms deviations of around 13 cm^{-1} in the range up to 15000 cm^{-1}, but less than 5 cm^{-1} for the fundamental modes, and around 1 cm^{-1} for inversion splittings. In subsequent unpublished work, we have further improved the potential energy surfaces (more points), their fitting (to higher order), and technical aspects of the variational calculations (with regard to convergence) which has reduced the overall rms deviations from around 13 to 5 cm^{-1}. Furthermore, the theoretical treatment has been extended to rotational excitations and to the evaluation of rovibrational line intensities on the basis of ab initio 6D dipole moment surfaces. To illustrate the present status of the development, the computed and observed spectra for the $\nu_4/2\nu_2$ band system of ammonia are shown in the Figure. It is obvious that the first-principles calculation reproduces the experimental spectrum remarkably well.

Intensities of the $\nu_4/2\nu_2$ bands in ammonia: Ab initio theory vs. experiment

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

Involved: K. Angermund, M. Bühl, Z. Chen, A. Fu, M. Graf, H. Hermann,
M. N. Jagadeesh, V. R. Jensen, D. Koley, S. F. Vyboishchikov

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. Further targets of our investigations are fullerenes and nanotubes.

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

Previous joint studies with the Fink group on propene polymerization have established the connection between the microstructure of zirconocene-based catalysts and the tacticity of the formed polymer, with excellent agreement between computed and observed pentad distributions at low temperatures. More recent DFT calculations indicate that the larger deviations at higher temperatures are due to the onset of epimerization (back-skip) processes not included in the original model, while the activation of other internal motions does not play a significant role according to Car-Parrinello molecular dynamics (CPMD) simulations. A systematic DFT investigation of zirconocene-catalyzed norbornene oligomerization has revealed qualitative differences compared with ethylene (e.g., with regard to the resting state and the relative energies of the relevant agostic conformers), and has given quantitative measures for stereochemical endo/exo and syn/anti preferences. In the course of this work, it has become apparent that σ-bond metathesis reactions are feasible in norbornene oligomers, with a computed free energy barrier of 12 kcal/mol in the dinorbornyl chain, which rationalizes the experimentally (X-ray) observed C7 linkage in such systems [364].

In a collaborative project with the Fürstner group, we have studied olefin metathesis reactions catalyzed by Ru-based carbene complexes. For the parent system $Ru(CH_2)(PH_3)(X)Cl_2$ with $X=PH_3$ or $X=C_3N_2H_4$ (Arduengo carbene ligand), the DFT calculations show [141] that the dissociative pathway is favored and that the incoming

olefin coordinates trans to X (without a barrier) before inserting in the rate-determining step to form the intermediate metallacycle. In an attempt to understand the stereochemistry observed in the Fürstner group during the synthesis of salicylihalamid and related model compounds (Z/E preferences) we have investigated the mechanism for the ring-closing metathesis (RCM) reaction of large sterically hindered α,ω-olefins considering four substrates, two catalysts, and two possible pathways (see Figure). We find that the stereochemistry is always determined in the second metathesis reaction, but the rate-determining step may be different for different substrates, catalysts, and pathways. Good agreement with the experimentally observed stereochemistry is only achieved when using the real catalysts and substrates in the DFT calculations. The comparison with simpler models provides insight into the role of the mesityl substituents in the catalyst and other factors that influence the stereochemistry .
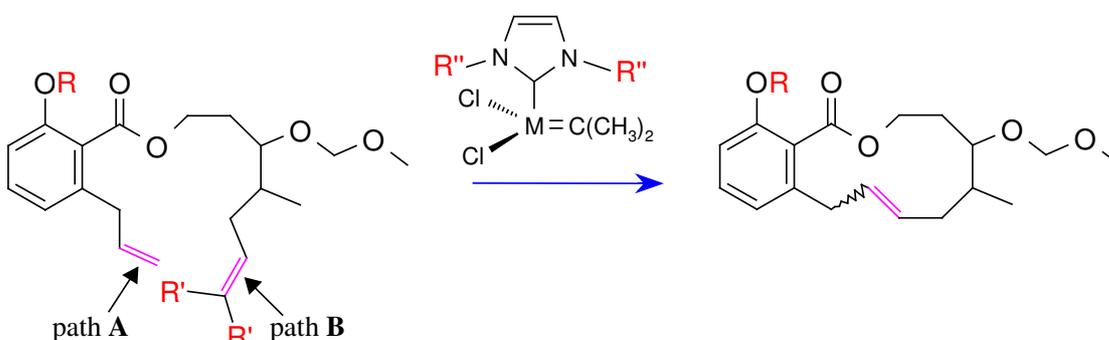
Survey of RCM reactions studied computationally: Substrates I (R=R'=H), II (R=CH$_3$, R'=H), III (R=H, R'=CH$_3$), IV (R=R'=CH$_3$); model catalyst with R''=H, real catalyst with R''=mesityl.

A cooperation with the Gooßen group addresses palladium-catalyzed cross coupling reactions. The oxidative addition of aryl halides to a Pd(0) complex is the initiating step in these reactions. According to mechanistic studies by Amatore and Jutand, three-coordinate Pd(0) complexes with a counteranion ligand are present initially, and it has therefore been proposed and commonly accepted that the oxidative addition yields five-coordinate Pd(II) intermediates. Our DFT calculations suggest a different mechanism, with the aryl halide coordinating to Pd via the halide atom to form a four-coordinate complex, that subsequently rearranges to the desired addition product without ever forming a five-coordinate intermediate [354]. This work has been triggered by analogous findings in extensive DFT studies on Pd-catalyzed cross coupling reactions between phenylboronic acid and acetic anhydride where the complete catalytic cycle (including oxidative addition, transmetalation, isomerization, and reductive elimination) has been computed in four variants in order to gain mechanistic insight.

Further collaborative DFT work within the Institute concerns the enantioselectivity of Rh-catalyzed hydrogenation of functionalized olefins (Reetz group, see also section 2.5.5), the properties of the proton sponge 4,9-dichloroquino[7,8-*h*]quinoline and its catalytically active palladium complexes (Schüth group), and the enantioselectivity of proline- and 2-methylproline-catalyzed α-alkylation of aldehydes (List group).

Several DFT studies have been carried out without involvement of experimental groups from the Institute, for example on the binding of dinitrogen to the FeFe cofactor in Fe-only nitrogenase [174] and on valence trapping in a mixed-valent dimanganese complex [329] (see also section 2.5.6). The initial activation of Pt-based precatalysts for hydrosilylation has been elucidated in an external cooperation with an industrial partner [64]. Finally, DFT calculations have been used in combination with ab initio and semiempirical calculations to investigate fullerenes and nanotubes; topics include the properties of C_{20} [335] and C_{50} [382], the NMR spectra of fullerene anions [283], heterofullerenes [66,179,223,358], the structure and stability of Si_{60} and Ge_{60} cages analogous to C_{60} [181], and the structure and reactivity of nanotubes [184,336].

Publications resulting from this research area: 64, 66, 136, 141, 174, 178-182, 223, 283, 329, 335, 336, 354, 358, 364, 382, 432

External funding: Consortium für elektrochemische Industrie GmbH (München); European COST program (working group D17/010/02)

Cooperations: R. Blom (SINTEF, Oslo, NO); M. S. Brookhart (Chapel Hill, USA); Z. Cao (Xiamen, CN); F. Diederich (Zürich, CH); G. Fink (Mülheim/Ruhr, DE); A. Fürstner (Mülheim/Ruhr, DE); L. Gooßen (Mülheim/Ruhr, DE); A. Hirsch (Erlangen, DE); S. Nagase (Okazaki, JP); M. Rabinovitz (Jerusalem, IL); M. T. Reetz (Mülheim/Ruhr, DE); P. v. R. Schleyer (Athens, USA); F. Schüth (Mülheim/Ruhr, DE)

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

Involved: Z. Chen, R. Kevorkiants, D. Khoroshun, A. Koslowski, M. Scholten, R. Steiger

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. Applications are usually motivated by requests from experimental partners or by topical chemical problems, but they also serve to explore the limits of new methods and codes.

Results: Over the past years, we have worked on semiempirical methods that go beyond the standard MNDO model by including orthogonalization corrections at the NDDO level. This has led to two new approaches labelled OM1 and OM2 (orthogonalization models 1 and 2) which offer significant improvements over established MNDO-type methods in several areas, including conformational properties, hydrogen bonds, reaction barriers, and electronically excited states. In the reporting period, the parameterization of the OM3 variant has been completed, which neglects certain of the less important OM2 terms. OM3 is somewhat faster than OM2 and has similar overall accuracy. Parameters for the elements H, C, N, O and F are now available for OM1, OM2 and OM3.

The improved description of electronically excited states in OMx methods will facilitate realistic studies of photochemical reactions in larger organic molecules. For this purpose, we have implemented the GUGACI method in a semiempirical context [237] such that general CI expansions can be treated efficiently in a relatively small active space. This in-core code is quite flexible and allows for CI calculations with up to about 100000 configurations, for all excitation classes up to full CI. It incorporates an analytic GUGACI gradient. Analytic integral derivatives have been programmed up to now only for MNDO-type methods, but not for OMx methods. For the latter, we have written a semi-analytic GUGACI gradient code which handles all the expensive steps analytically and evaluates only the integral derivatives numerically, leading to the same scaling behaviour as the fully analytic implementation.

Using the GUGACI code we have extended previous validations of OMx-CI methods for electronically excited states. In the case of vertical excitation energies, comparisons

with CASPT2 reference data for 332 singlet-singlet and 103 singlet-triplet transitions indicate typical OMx deviations of 0.4 eV overall, and of 0.3 eV for the lower valence transitions, whereas MNDO-type methods systematically underestimate the excitation energies by more than 1 eV. Further validation studies for excited-state geometries and other properties are in progress. An independent validation of OM2-CI has been done in joint work with the Elstner group (Paderborn) who find excellent agreement with high-level ab initio CI results for models of retinal proteins.

In a collaboration with the Engels group (Würzburg) we have worked on a semiempirical parameterization for excited states and chiroptic properties at the OM2-CI level using ab initio reference data. The parameter sets derived in Würzburg show significant improvements in the excited-state target properties, but at the expense of larger discrepancies for ground-state properties, and they will thus only be of limited utility. We have therefore decided to go for a second OM2-CI parameterization that aims at a balanced treatment of ground and excited states. We are currently in the process of generating the needed high-level ab initio reference data.

A number of other methodological projects have been completed in the reporting period. These include the implementation of linear scaling techniques in our semiempirical program (using the conjugate gradient density matrix search as a replacement for diagonalization), the generation of derivative code by automatic differentiation of Fortran77 programs (using the ADIFOR software), and the development of improved parameterization tools (using a combination of systematic searching, genetic algorithms, and local nonlinear least-squares minimizations).

Further methodological advances have been pursued in three external cooperations. A state specific transfer Hamiltonian has been defined and parameterized for selected reactions with regard to CCSD reference data, for use in large-scale molecular dynamics simulations (Bartlett group). An interface to the Xiamen valence-bond (VB) program has been written to explore the merits and limitations of semiempirical VB calculations (work with Wei Wu during his visit as guest professor). A smooth solvation model SCOSMO has been developed in the context of semiempirical methods with an spd-basis, which provides smooth energies and gradients by overcoming discretization errors and will therefore allow numerically more stable geometry optimizations and molecular dynamics simulations in solution (cooperation with the York group).

Concerning applications, the performance of semiempirical MNDO-type methods in fullerene chemistry has been validated by extensive comparisons with B3LYP/6-31G* results for relative energies and nucleus-independent chemical shifts [183]. This provides additional justification for employing these methods in some of our recent work on fullerenes and nanotubes [184,336,382]. In a collaboration with the Suhm group (Göttingen), we have studied large N_2O clusters containing up to 177 molecules at the AM1 level. The simulated spectra at different cluster sizes show excellent agreement with the observed spectra, and the calculations thus help to shed light on the microscopic structure of these nanosystems [359].

Publications resulting from this research area: 183, 184, 237, 286, 289, 336, 359, 382

External funding: Fonds der Chemischen Industrie

Cooperations: R. J. Bartlett (Gainesville, USA); B. Engels (Würzburg, DE); M. Elstner (Paderborn, DE); S. Patchkovskii (Ottawa, CA); M. A. Suhm (Göttingen, DE); W. Wu (Xiamen, CN); D. M. York (Minneapolis, USA)

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

Involved: A. Altun, I. Antes, S. R. Billeter, M. Bocola, J. Kästner, H. Lin, N. Otte, N. Reuter, J. C. Schöneboom, H. M. Senn, F. Terstegen, S. Thiel, T. Tuttle, D. Wang, 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: The European QUASI project (Quantum Simulations in Industry, 1998-2001) has led to the development of a flexible and general QM/MM environment with an open-ended design that allows the integration of many QM and MM codes in a modular manner. The resulting ChemShell software has been described along with the underlying theory and some illustrative applications [279]. In the reporting period, we have extended this software according to our needs. The modules for geometry optimization have been enhanced by introducing improved microiterative techniques in general, and three-layer optimizations for transition state searches in particular. In the module for molecular dynamics (MD) simulations, a number of features have been added that provide new functionality (friction dynamics, Nose-Hoover chain thermostat with corresponding integrator, additional SHAKE constraints, etc.) and better control (error estimates, convergence of thermodynamic integration, etc.). A new interface has been written for the MOLPRO code from the Werner group (Stuttgart) to access high-level ab initio QM methods within ChemShell, including local correlation approaches that are expected to be essential for large QM regions. The existing interface to the GROMOS code from the van Gunsteren group (Zürich) has been upgraded to conform to the present GROMOS version and to handle the recently proposed polarizable force fields in GROMOS. The performance of the latter in QM/MM approaches is currently evaluated in a joint project.

Our most extensive QM/MM application during the past three years has addressed the catalytic cycle of cytochrome P450cam at the B3LYP/CHARMM level (B3LYP hybrid

density functional, CHARMM force field), in close collaboration with the Shaik group (Jerusalem). Several intermediates have been characterized including the resting state [419], the pentacoordinated ferric and ferrous complexes, the last experimentally accessible intermediate Cpd 0, the yet unobserved reactive species Cpd I [120], and the product-enzyme complex [377]. Moreover, some key reaction steps have been investigated, in particular the stereospecific camphor hydroxylation by Cpd I [418] and the proton transfers leading from Cpd 0 to Cpd I. In all these cases, the comparison between B3LYP/CHARMM calculations for the complete solvated enzyme (around 25000 atoms) and B3LYP calculations for the isolated QM region (typically 40-130 atoms) allows us to assess the role of the protein environment in P450cam. In the following, we summarize a few selected results. The QM/MM calculations correctly give a low-spin resting state and high-spin pentacoordinated complexes, and they reproduce the known experimental geometries and spectroscopic parameters of these species quite well. The protein environment causes a notable differential stabilization of the high-spin states in the resting state, and it inverts the relative energies of two close-lying quintet electromers of the pentacoordinated ferrous complex. The experimentally yet unknown Compound I has almost degenerate doublet and quartet states, with the doublet slightly more stable. It adapts its electronic structure to the environment: when going from the gas phase to the enzyme, the spin density distribution changes from a predominantly sulfur-based radical to a porphyrin-based radical, due to the stabilization of electron density at sulfur by hydrogen bonding interactions with nearby residues. Camphor hydroxylation by Cpd I follows a rebound mechanism with two-state reactivity: the initial hydrogen abstraction involves a sizable barrier in both spin states (slightly lower in the doublet), whereas the subsequent radical recombination (rebound) has essentially no barrier in the doublet and a small but finite barrier in the quartet, for analogous reasons as in the gas phase (orbital correlations). The protein environment does not change the basic mechanism of hydroxylation, but it does modulate the electronic structure and the relative energies along the reaction pathway. Our QM/MM calculations do not support the recent proposal that spin density at heme side chain carboxylates is essential for catalytic activity by Cpd I. Three local minima are found in QM/MM optimizations of the product-enzyme complex which is consistent with EPR observations of three distinct species. Finally, reasonable agreement between computed and observed spectroscopic constants is obtained for heme model compounds, the resting state, and some known intermediates so that the corresponding theoretical predictions for Compound I may be helpful in future attempts to detect it by EPR or Mössbauer spectroscopy.
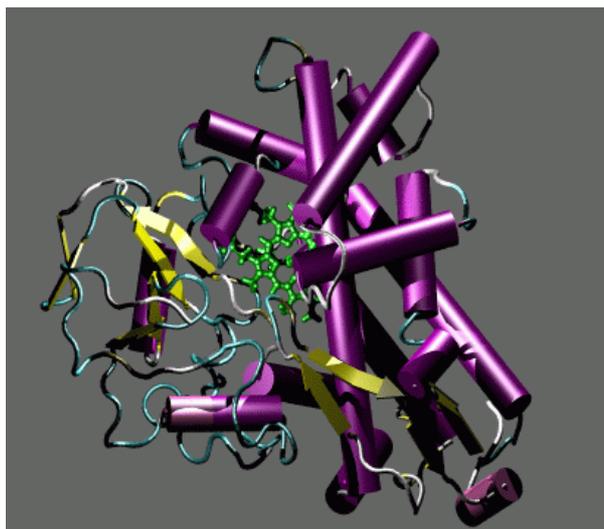
Cytochrome P450cam (heme in green)

Another enzyme project involves a collaboration with the Reetz group on directed evolution in lipases, where we employ classical MD and QM/MM calculations in an attempt to understand the enantioselectivity in the lipase-catalyzed ester hydrolysis and its optimization through successive mutagenesis. In the first stage of this project, we have performed 1 ns CHARMM-MD simulations of the relevant tetrahedral intermediates in 20 enzyme/substrate combinations, with the *R* and *S* enantiomers of α-methyldecanoic acid ester serving as substrates for the wild-type lipase from *Pseudomonas aeruginosa* and for nine selected mutants [311]. The resulting geometric and energetic changes during these MD runs allow an interpretation of some experimentally observed effects of mutations, particularly with regard to the "hot spots" at residues 155 and 162, in terms of a steric relay or the creation of a new binding pocket [311], thereby providing insight into remote and cooperative effects of mutations. In the second stage of the project, we have located the relevant stationary points for the rate-determining step (Michaelis complex, transition state, and tetrahedral intermediate) through BLYP/CHARMM geometry optimizations and evaluated the electrostatic influence of individual residues on the reaction barrier by single-point QM/MM calculations with zero MM side chain charges on the corresponding residue. In the case of *Bacillis subtilis* lipase A, this prescreening identifies His76 as a potential "hot spot" which has been confirmed experimentally. These two examples indicate that theoretical modeling can rationalize and predict "hot spots" in directed evolution.

A number of other biomolecular systems have been studied at the QM/MM level during the reporting period. An extensive QM/MM calibration study on the conversion of

dihydroxyacetone phosphate into glycerinaldehyde 3-phosphate by triosephosphate isomerase (TIM) has been completed and published [84] which supports a four-step proton transfer via an enediol, with involvement of neutral His95 acting as a proton donor. In p-hydroxybenzoate hydroxylase (PHBH), the rate-determining OH transfer from the hydroperoxide of the cofactor (flavin adenine dinucleotide) to the substrate (p-hydroxybenzoate) has been investigated at the AM1/GROMOS level; the free energy barriers from thermodynamic integration do not differ much from the average energy barriers obtained in a series of separate optimizations for different snapshots, which implies that entropic effects are not overly important in this case. In preparation of planned QM/MM studies on electronically excited states of green fluorescent protein (GFP) and photoactive yellow protein (PYP) we have derived the necessary MM parameters for the GFP chromophore and carried out exploratory classical MD simulations for GFP and PYP in the ground state [4,115]. Finally three new QM/MM projects have been started recently in the framework of external cooperations, which involve mechanistic studies on 4-oxalocrotonate tautomerase and some synthetic analogues (with E. Keinan, Haifa), on the Bergman cyclization of dynemicin A in a DNA environment (with D. Cremer, Göteborg), and on enzymatic C-F bond formation by a fluorination enzyme (with D. O'Hagan, St. Andrews).

Publications resulting from this research area: 4, 84, 115, 120, 137, 186, 210, 279, 311, 377, 418, 419

Cooperations: C. R. A. Catlow (London, UK); D. Cremer (Göteborg, SE); K.-E. Jaeger (Jülich, DE); E. Keinan (Haifa, IL); F. Neese (Mülheim/Ruhr, DE); D. O'Hagan (St. Andrews, UK); M. T. Reetz (Mülheim/Ruhr, DE); A. Schäfer (BASF, Ludwigshafen, DE); S. Shaik (Jerusalem, IL); P. Sherwood (Daresbury Laboratory, UK); W. F. van Gunsteren (Zürich, CH); H.-J. Werner (Stuttgart, DE)