

QUILD: QUantum-Regions Interconnected by Local Descriptions

MARCEL SWART,^{1,2,3} F. MATTHIAS BICKELHAUPT¹

¹Theoretische Chemie, Scheikundig Laboratorium der Vrije Universiteit, De Boelelaan 1083, NL-1081 HV Amsterdam, The Netherlands

²Institut de Química Computacional, Universitat de Girona, Campus Montilivi, E-17071 Girona, Spain

³Institució Catalana de Recerca i Estudis Avançats (ICREA), E-08010 Barcelona, Spain

Received 24 January 2007; Revised 1 August 2007; Accepted 2 August 2007

DOI 10.1002/jcc.20834

Published online 27 September 2007 in Wiley InterScience (www.interscience.wiley.com).

Abstract: A new program for multilevel (QM/QM and/or QM/MM) approaches is presented that is able to combine different computational descriptions for different regions in a transparent and flexible manner. This program, designated QUILD (for QUantum-regions Interconnected by Local Descriptions), uses adapted delocalized coordinates (Int J Quantum Chem 2006, 106, 2536) for efficient geometry optimizations of equilibrium and transition-state structures, where both weak and strong coordinates may be present. The Amsterdam Density Functional (ADF) program is used for providing density functional theory and MM energies and gradients, while an interface to the ORCA program is available for including RHF, MP2, or semiempirical descriptions. The QUILD optimization setup reduces the number of geometry steps needed for the Baker test-set of 30 organic molecules by ~30% and for a weakly-bound test-set of 18 molecules by ~75% compared with the old-style optimizer in ADF, i.e., a speedup of roughly a factor four. We report two examples of using geometry optimizations with numerical gradients, for spin-orbit relativistic ZORA and for excited-state geometries. Finally, we show examples of its multilevel capabilities for a number of systems, including the multilevel boundary region of amino acid residues, an S_N2 reaction in the gas-phase and in solvent, and a DNA duplex.

© 2007 Wiley Periodicals, Inc. J Comput Chem 29: 724–734, 2008

Key words: multilevel approach; geometry optimization; density functional theory

Introduction

Herein, we present a new multilevel QM/QM and/or QM/MM approach and program: QUILD for QUantum-regions Interconnected by Local Descriptions. QUILD allows a (complex) molecular system to be decomposed into arbitrary, interpenetrating domains such that each type of interaction can be described with a different method, i.e., the density functional that performs best for that type of interaction or process. This is a pragmatic (of course not a fundamental) approach to cope with the present-day shortcomings of density functional theory (DFT), which features density functionals that perform satisfactorily for a number but not all types of interactions and phenomena (*vide infra*).

The application of traditional multilevel (QM/QM or QM/MM) approaches within computational-chemistry studies is ever more often used, because it permits to use a highly accurate method for the most important region while treating the interactions with the surrounding regions at a lower, yet sufficiently accurate method.^{1,2} The QM/MM setup (see Fig. 1), where only the region of interest is treated with quantum chemistry methods

while the interactions with and within the surrounding regions is described with classical molecular mechanics force fields, is one of the computationally most economical multilevel approaches. However, its accuracy and applicability depend largely on the accuracy and availability of force field parameters for the system under study. Although specialized force fields are available for certain classes of chemical systems, as the AMBER,³ CHARMM,⁴ or GROMOS⁵ force fields for proteins and nucleic acids, or OPLS⁶ for organic solvents, they usually do not include the parameters for nonstandard biochemical systems or organic solvents. Moreover, the force field parameters may be specifically designed or constructed for obtaining correct macroscopic proper-

This article contains supplementary material available via the Internet at <http://www.interscience.wiley.com/jpages/0192-8651/suppmat>

Correspondence to: M. Swart; e-mail: marcel.swart@udg.es or F. M. Bickelhaupt; e-mail: fm.bickelhaupt@few.vu.nl

Contract/grant sponsor: Netherlands organization for Scientific Research (NWO-CW)

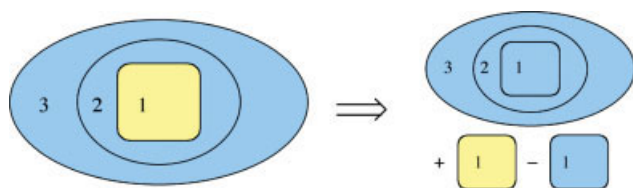


Figure 1. Schematic drawing of QM/MM setup of a protein where the active site (region 1, in yellow) is treated by QM methods, and is surrounded by the protein environment (region 2, in blue) and a solvent layer (region 3, in blue) that are both treated at the MM level. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

ties in large-scale molecular dynamics (MD) simulations. The application of the same force field parameters in small-scale geometry optimizations of relatively small systems may turn out to give (microscopic) results that are less satisfactory. For that reason, it may be more advantageous to use quantum-chemical methods whenever possible, i.e. the QM/QM multilevel approach.

Because of the computational efficiency, the availability of basis sets for the whole Periodic System, and the generally accurate results, DFT^{7–9} has become the method of choice for the majority of recent computational chemistry studies and can these days almost routinely be used for relatively large system sizes of up to a few hundred atoms.^{10–18} However, one must always remain cautious with the choice of DFT functional and/or basis set, and make sure that the particular functional is able to give a correct description for the interactions that are important for the system under study. For instance, the performance of functionals that include the recent OPTX exchange functional¹⁹ is superior to those containing Becke88 exchange²⁰ for, among others,^{21–23} the accuracy of geometries,^{24,25} spin state splittings,^{26,27} reaction barriers,^{24,28,29} or zero-point vibrational energies.²⁴ As the improvements can be linked directly to the specific formulation of the OPTX functional^{26,29} and its resulting improved performance for atomic exchange energies,¹⁹ one would naively think

that inclusion of the OPTX functional would always lead to improved performance. Unfortunately, this is not the case for weakly-bound systems, as shown recently for hydrogen-bonding³⁰ and π -stacking³¹ in DNA. Moreover, a functional that performs well^{30,32} for hydrogen-bonding interactions (BP86^{20,33}), does not necessarily give equally good results for π -stacking³¹ or reaction barriers.^{24,28,29} As a result, at present there does not seem to be a DFT functional that is equally accurate for hydrogen-bonding, π -stacking, reactions barriers, and intramolecular interactions. Therefore, for a study on for instance the structure of DNA duplexes, the aforementioned multilevel QUILD approach is needed with one DFT functional for the description of hydrogen-bonding interactions, and another for the description of π -stacking (see Fig. 2). Thus, although QUILD may be used in a similar fashion as traditional QM/QM methods, i.e., to reduce the computational cost, its main purpose is to increase the effective accuracy of a DFT (or other type of) computation.

Knowledge of molecular structure is another important aspect of scientific studies into the details of chemical and physical phenomena. Having efficient and reliable optimization techniques available is therefore a vital prerequisite, especially when studying larger systems or weakly-bound systems that are more difficult to optimize (*vide infra*). Recently,³⁴ we have proposed an adaptation of the delocalized coordinates setup as originally reported by Baker et al.³⁵; the delocalized coordinates are easy to construct and function well for strong (intramolecular) coordinates. Our adaptation enables their use also in the case of weak coordinates by employing a screening function that distinguishes weak from strong coordinates.

The purpose of this contribution is 3-fold. First, we want to introduce the program and explain the philosophy. The benefit of our multilevel approach is not only that it can be used to make the calculations cheaper and therefore feasible, but also that the best method for any type of interaction can be used as we need for DNA. For describing DNA, we use one DFT functional (see Fig. 2, in blue) for treating the complete system, and another (Fig. 2, in pink and yellow) for π -stacking between the

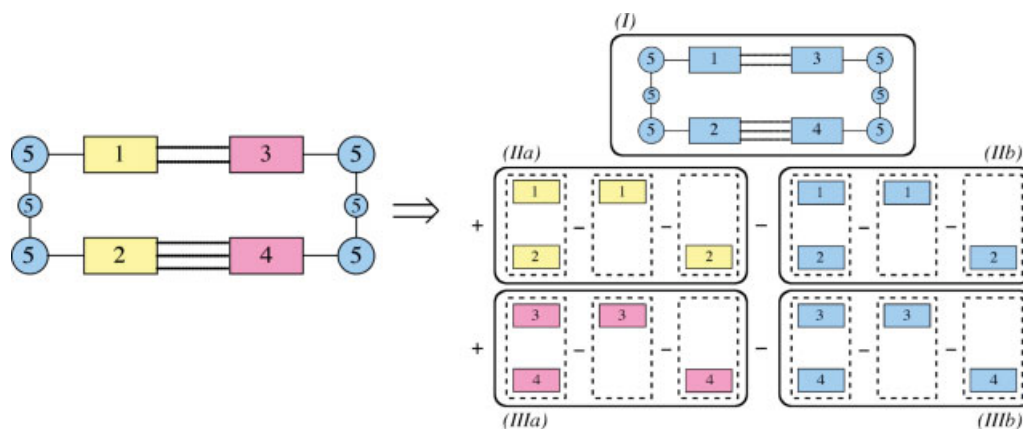


Figure 2. Schematic drawing of QM/QM treatment of DNA, where the π -stacking between the bases (regions 1–4, in yellow and pink, see also Scheme 2) is treated at a different level than the H-bonding and intramolecular interactions (numbers in italics refer to steps in QM/QM scheme, see text). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

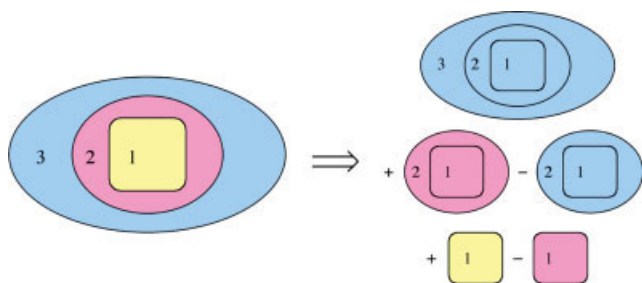


Figure 3. Schematic drawing of QM/QM/MM setup of a protein where the active site (region 1, in yellow) is treated by a high-level QM method, and is surrounded by the protein environment (region 2, in pink) treated at low-level QM, and a solvent layer (region 3, in blue) that is treated at the MM level. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

DNA bases. This is achieved by making the definition of the regions flexible, i.e., there is no need to have a layered structure as in the ONIOM approach (represented in Fig. 3). An arbitrary splitting of the total system into different regions is permitted with, therefore, possibly overlapping regions; this resembles a quilt, hence the name of the program. Second, we want to check our program by comparing the performance of the geometry optimization setup (both at RHF/STO-3G and PW91/TZ2P) with literature values for the Baker test set³⁶ of 30 organic molecules, and compare the results to optimization with the old-style optimizer in Amsterdam Density Functional (ADF). Furthermore, we compare the performance for a set of weakly bound systems, where the improvement of using the adapted delocalized coordinates is most obvious. Third, we want to show the capabilities of QUILD by showing geometry optimizations using numerical gradients, and, most importantly, the application of the multilevel QM/QM approach to a number of systems that include the treatment and accuracy of the multilevel boundary region, solvation effects on a bimolecular nucleophilic substitution (S_N2) reaction, and finally the treatment of a DNA duplex system.

Computational Details

The generation of the delocalized coordinates, screening functions, etc., and the actual geometry optimization scheme was performed within the QUILD³⁷ (QUantum-regions Interconnected by Local Descriptions) program, which has been developed for enabling QM/QM calculations, and which functions as a wrapper around the ADF,^{38,39} NEWMM, and ORCA programs. The QUILD program constructs all input files for example ADF, runs ADF, and collects all data; ADF, NEWMM (integrated into the ADF package), and ORCA are used only for the generation of the energy and gradients. The QUILD program consists of a combination of a shell script and a Fortran90 part. The Fortran90 part takes care of the more involved dealings such as input handling, setting up the coordinates, making new coordinates within geometry optimizations, and writes the input files. The shell script executes these input files consecutively, and re-executes at regular intervals the Fortran90 executable to rewrite the input files and monitor the progress being made.

A numerical evaluation of the energy gradients (2-fold finite difference with step sizes to be chosen by the user, default value 1.0×10^{-5} au; see Table S4 for dependence of optimized geometry of hydrogen-peroxide on this parameter) in QUILD enables the use of geometry optimization techniques for any methodology within either of these programs, also for those for which only the energy expression is known or implemented (for instance meta-GGA or hybrid functionals, Spin-Orbit ZORA⁴⁰ within ADF) and for which geometry optimizations are otherwise out of reach (see for example ref. 28 for geometry optimizations with meta-GGA functionals that would otherwise have been impossible). Below we report examples of using numerical gradients for Spin-Orbit ZORA, Scalar ZORA, and nonrelativistic calculations for group 16 hydrides and halogen dimers, as well as for the first singlet and triplet excited states of formaldehyde.

Similarly, the QUILD program also facilitates the use of spin-projection techniques for gradients, i.e., it enables the geometry optimization of pure spin states for systems that suffer from spin-contamination (see for example ref. 27). Moreover, as the geometry optimization and the quantum-chemistry are separated, the QUILD setup enables to use different symmetry representations for the geometry (within QUILD) and the quantum-chemistry (within ADF). Therefore, for a study on for instance metalloporphyrins, the programs allows to use C_{2v} symmetry within ADF (for convenience with the iron 3d-orbitals), and S_4 symmetry for the geometry. Although this separation could in principle be achieved also by creating a Z-matrix that enforces the S_4 symmetry, this Z-matrix may not be optimally suited for geometry optimizations.

All DFT calculations were performed with the ADF (version 2006.01)^{38,39} program developed by Baerends et al. The MOs were expanded in a large uncontracted set of Slater-type orbitals (TZ2P),⁴¹ which is of triple- ζ quality, augmented by two sets of polarization functions (e.g., 3d and 4f on C, N, O; 2p and 3d on H); the core electrons (e.g., 1s for 2nd period, 1s2s2p for 3rd period, etc.) were treated by the frozen core (FC) approximation.³⁹ An auxiliary set of s, p, d, f, and g STOs was used to fit the molecular density and to represent the Coulomb and exchange potentials accurately in each SCF cycle. The numerical integration and solving the SCF equations was in most cases performed at high accuracy (ACCINT = 8.0, SCF convergence criterium 1.0×10^{-8} ; both parameters were imposed at every geometry step) to avoid inconsistencies due to numerical noise when comparing the different optimization strategies in the first part of the results section.

Energies and gradients were mainly calculated using the local density approximation (LDA; Slater exchange and VWN⁴² correlation) with nonlocal (PW91^{43,44}) corrections added self-consistently. This xc-functional is one of the best (pure) DFT functionals for the accuracy of geometries,²⁵ with an estimated unsigned error of 0.8 pm in combination with the TZ2P basis set. Furthermore, it is able to treat weak interactions correctly.^{30,45}

The RHF and AM1 calculations were performed with the ORCA program using conventional SCF techniques, and high convergence criteria imposed (1.0×10^{-7} au for tolerance of energy, density, and DIIS equations).

The starting geometries for the Baker test set have been obtained from the supporting information in the article by

Bakken and Helgaker,⁴⁶ which, however, contains two small typographical errors (for ACANIL01 and neopentane). The corrected coordinates are given in the supporting information of the present article, which also includes the starting geometries for the molecules in the weak-coordinates set.

**Capping Atoms in QM/QM and QM/MM Calculations:
the AddRemove⁴⁷ Scheme**

In the case of multilevel approaches where the boundary between regions cuts through covalent bonds, one has to take care of the dangling bonds. For instance in the case of the application of QM/QM to a small protein, where the active site is treated at high-level QM and the remainder at low-level QM, one has to perform three calculations: one at low-level QM for the whole protein, one at high-level QM for the active site, and one at low-level QM for the active site (see Fig. 1). The energy (and gradient, Hessian) is then obtained as:

$$E_{\text{total}} = E_{\text{low}}(\text{prot.}) + E_{\text{high}}(\text{act. site}) - E_{\text{low}}(\text{act. site}) \quad (1)$$

The treatment of the active site in the latter two calculations has to deal with the dangling bonds at which the active site is cut off. This is most conveniently handled by adding capping atoms (usually hydrogens), which are positioned along the vector of the dangling covalent bond, and at a distance that corresponds to the sum of the covalent radii of the capping atom and the atom to which the capping atom is attached. Because the capping atoms are added to the active site for both the high- and low-level QM calculation (and at exactly the same position), with a presumably similar effect in both cases, the interactions of the capping atoms with the true active site atoms are in good approximation cancelled out (the total effect is removed) between the low- and high-level QM calculations. This is the AddRemove model⁴⁷ for link atoms, which was previously shown to perform well for geometries around the boundary between the QM and MM region in QM/MM calculations. The AddRemove model⁴⁷ has several advantages: it is simple and the energy and gradients are treated in similar fashion (unlike other models that project the gradients of the capping atoms onto the gradients of the real atoms):

$$g_{\text{total}} = g_{\text{low}}(\text{prot.}) + g_{\text{high}}(\text{act. site}) - g_{\text{low}}(\text{act. site}) \quad (2)$$

Furthermore, the capping atoms follow the real atoms, at a predefined distance, and therefore no artificial degrees of freedom are added by including the capping atoms.

Performance of Geometry Optimization Setup

Before presenting the results for a number of QM/QM and QM/MM applications, we would first like to show the performance of our geometry optimization setup. Baker³⁶ proposed a test set of 30 molecules that can be used for comparing different optimization strategies, with a convergence criterion for the gradient of 3.0×10^{-4} .³⁶ In our previous contribution,³⁴ we introduced an

Table 1. Number of Geometry Cycles^a for Optimizing Molecules of the Baker (30 Molecules) and Weak (18 Molecules) Test Sets, Compared for Various Geometry Optimization Schemes.

Test set	Baker, RHF/STO-3G	Baker, PW91/TZ2P	Weak, PW91/TZ2P
Baker ³⁶	240	–	–
Lindh ⁴⁹	215	–	–
Eckert ⁴⁸	196	–	–
Bakken ⁴⁶	185	–	–
This work			
ADF _{cart} ^b	–	222	748
QUILD _{cart}	–	214 (218) ^c	242 (293) ^c
QUILD _{deloc}	167	164 (171) ^c	175 (230) ^c

^aFull details can be found in Tables S1–S3 in the supporting information.

^bUsing old-style optimizer in ADF.

^cValue in parentheses is obtained when not discriminating between strong and weak coordinates, i.e., with weight $w = 1$ for all primitives, see eqs. (7)–(10) in supporting information.

additional set of weakly-bound molecules, which is extended for the current contribution to 18 molecules with their starting geometries preoptimized at MP2/6-31+G*.

Previously, we compared our own PW91/TZ2P results for the Baker test set with RHF/STO-3G data taken from the literature. This comparison is not entirely fair in the sense that it is *a priori* unclear whether the performance benefits or suffers from the more accurate PW91 treatment. Therefore, we decided to reexamine the performance of our geometry optimization setup in two steps: first, at RHF/STO-3G (through the interface with ORCA) for the Baker test set to be able to compare with values from the literature; and second, at PW91/TZ2P for the Baker and weak test sets where we compare the use of either Cartesian or delocalized coordinates within QUILD with the performance of Cartesians within the old-style optimizer in ADF.

For the Baker test set, the total number of geometry cycles needed at the RHF/STO-3G level with quasi-Newton methods ranges from 240 (Baker),³⁶ 196 (Eckert, Pulay, and Werner),⁴⁸ 215 (Lindh, Bernhardsson, Karlström, and Malmqvist)⁴⁹ to 185 (Bakken and Helgaker).⁴⁶ Note that the latter number can be reduced to 111, when the exact Hessian is used at every step (i.e. using Newton-Raphson techniques).⁴⁶ Using the QUILD scheme for geometry optimization, we need only 167 geometry steps for completely optimizing the Baker test-set (see Table 1). This is 18 geometry steps less than the results by Bakken and Helgaker, the best result from the literature so far. Our results presented here improve upon the ones reported in our previous contribution due to an improved initial Hessian, and using a dynamic trust radius (see supporting information). Note also that the Bakken–Helgaker results⁴⁶ were obtained using a rational function (RF) approach for restricting the step length, which when using extra-redundant internal coordinates was shown to be 17 steps (–9%) more efficient than the RSO approach we use. At the RHF/STO-3G level we also find an improvement when using the RF approach, but only a very modest one (a reduction with 2). More importantly, within QUILD the RF

approach performs less than RSO for the DFT calculations. Therefore, the RF approach will not be discussed any further.

Also included in Table 1 are the number of iterations needed for the Baker test set at the PW91/TZ2P level, obtained either with QUILD or the old-style optimizer in ADF. When using Cartesian coordinates, the difference between QUILD and the old-style optimizer in ADF is rather small (222 for ADF vs. 214 for QUILD). This can be improved upon significantly by using the adapted delocalized coordinates,³⁴ where we only need 164 geometry steps (a reduction of ~30%). The influence of using a screening function to distinguish between strong and weak coordinates is shown to have only a small effect on the performance for the Baker test set (see Table 1). When setting the weights [see eqs. (7)–(10) in supporting information] explicitly to one, the number of geometry steps needed increases only slightly for the Baker test set, from 164 to 171 for the adapted delocalized coordinates and from 214 to 218 for Cartesian coordinates.

The set of weakly-bound molecules contains 18 dimers, ranging from ion-dipole complexes to van der Waals (polarizability–polarizability) complexes. This weak set was studied at the PW91/TZ2P level with Cartesian and adapted delocalized coordinates in QUILD, and Cartesians in the old-style optimizer in ADF. Contrary to what we saw for the Baker set, we observe now a dramatic difference between the QUILD geometry optimization setup and the old-style optimizer in ADF. The latter needs 748 geometry steps to completely optimize all 18 molecules to a gradient maximum of 1.0×10^{-5} au, which is reduced to 242 steps when using Cartesians within QUILD (see Table 1). A further reduction to 175 geometry steps is observed when using the adapted delocalized coordinates within QUILD, i.e. a 4-fold improvement compared to the old-style optimizer in ADF.

This significant reduction in the number of steps is not only the result of distinguishing weak from strong coordinates. When using a weight of 1.0 for all coordinates [see eqs. (7)–(10) in supporting information], the number of necessary steps increases to still only 293 (Cartesian) and 230 (adapted delocalized), i.e. an increase of 20–25%. Two factors influencing the improved performance within QUILD are most likely the initial Hessian that contains less noise within QUILD, and the use of regulated GDIIS. Note also the smooth performance of the QUILD optimizer that for all of the 18 weakly-bound systems converges to the minimum within 15 cycles with the adapted delocalized coordinates (see Table S3 in the supporting information), irrespective of whether the weak interactions result from ion-dipole ($\text{H}_2\text{O} \cdots \text{H}_3\text{O}^+$), dipole–dipole ($\text{H}_2\text{O} \cdots \text{H}_2\text{O}$, $\text{HF} \cdots \text{HF}$), dipole–polarizability ($\text{Ar} \cdots \text{HF}$), or polarizability–polarizability ($\text{CH}_4 \cdots \text{CH}_4$) interactions.

Geometry Optimizations Using Numerical Gradients

The first example of using the numerical gradients setup within QUILD is given by the structures of group 16 hydrides (H_2O , H_2S , H_2Se , H_2Te , H_2Po) and group 17 dimers (F_2 , Cl_2 , Br_2 , I_2 , At_2). For these systems we performed geometry optimizations with three Hamiltonians, i.e., nonrelativistic, scalar relativistic

Table 2. Distances (Å) and Angles (deg.) for Group 16 Hydrides and Halogen Dimers Obtained with Numerical Gradient Using Nonrelativistic, Scalar Relativistic, or Spin-Orbit Relativistic Hamiltonian, at PW91/TZ2P.^a

	NR ^b	SR-ZORA ^c	SO-ZORA ^d
Gr16 dist			
H ₂ O	0.970 (0.969)	0.969 (0.969)	0.969
H ₂ S	1.353 (1.353)	1.353 (1.352)	1.353
H ₂ Se	1.482 (1.483 ^e)	1.476 (1.478 ^f)	1.476
H ₂ Te	1.684 (1.684)	1.670 (1.670)	1.671
H ₂ Po	1.781 (1.782)	1.760 (1.761)	1.770
Gr16 angles			
H ₂ O	104.37 (104.35)	104.31 (104.27)	104.31
H ₂ S	91.53 (91.68)	91.43 (91.58)	91.43
H ₂ Se	90.42 (89.68 ^e)	90.28 (89.52 ^f)	90.28
H ₂ Te	90.02 (89.89)	89.81 (89.70)	89.78
H ₂ Po	88.99 (89.07)	89.20 (89.09)	89.07
Gr17 dimer dists			
F ₂	1.419 (1.419)	1.420 (1.420)	1.420
Cl ₂	2.017 (2.018)	2.018 (2.019)	2.018
Br ₂	2.323 (2.323)	2.314 (2.317)	2.317
I ₂	2.717 ^g (2.717 ^g)	2.695 (2.695)	2.709
At ₂	2.908 ^g (2.908 ^g)	2.875 (2.875)	2.966

^aValues in parentheses refer to results with analytical gradients (that for Scalar-ZORA do not correspond exactly to the correct gradients).

^bNonrelativistic Hamiltonian, using basis sets from nonrelativistic database.

^cHamiltonian with ZORA scalar relativistic corrections included self-consistently, using basis sets from ZORA relativistic database.

^dHamiltonian with ZORA Spin-Orbit relativistic corrections included self-consistently, using basis sets from ZORA relativistic database.

^eWith corresponding all-electron basis the values are respectively 1.478 Å for the HSe distance, and 90.43° for the HSeH angle (see text).

^fWith corresponding all-electron basis the values are respectively 1.475 Å for the HSe distance, and 90.25° for the HSeH angle (see text).

^gObtained with the basis from ZORA relativistic database.

ZORA,^{39,40} and spin-orbit relativistic ZORA.^{39,40} For the first two Hamiltonians the analytical gradient is also available (although for scalar ZORA with a slight mismatch, *vide infra*), which helps to get an estimate of the accuracy of the obtained geometry. The obtained angles and distances are given in Table 2. The relativistic corrections start to have an impact on the geometry from the third period onward; for instance for H_2Se the inclusion of scalar relativistic effects reduce the SeH distance by 0.006 Å. The effect becomes larger for H_2Te (−0.014 Å) and H_2Po (−0.021 Å), and only for H_2Po do the spin-orbit corrections seem to have an effect by enlarging the HPo distance again by 0.01 Å. Moreover, for the halogen dimers does the effect of relativistic corrections start to have an impact from the third period onwards. Similar to those observed for the group 16 hydrides, the scalar relativistic corrections result in a reduction of the distance, which is counteracted by the spin-orbit corrections that lead to increasing distances. The opposing effect of scalar versus spin-orbit corrections is best shown by the At dimer, where the distance decreases by 0.03 Å due to the scalar corrections, and then increases by 0.09 Å because of the spin-orbit corrections (see Table 2).

Table 3. Geometrical Parameters (Å, deg.) for Optimized Structures of Singlet and Triplet First Excited States of Formaldehyde.

	R_{CO}	R_{CH}	$\angle HCH$	θ^a	Energy ^b
Ground. st.					
CASSCF ^c	1.210	1.118	115.1		0
CCSD(T) ^d	1.208	1.098	116.3		0
PW91/TZ2P ^e	1.207	1.115	116.0		0
Exp. ^c	1.208	1.116	116.5		0
¹ ($n \rightarrow \pi^*$) exc. st.					
CASSCF ^c	1.358	1.109	116.8	37.7	3.48
MRDCT ^c	1.335	1.116	120.2	34.5	3.50
PW91/TZ2P ^e	1.306	1.101	116.4	37.1	3.59
Exp. ^c	1.323	1.098/1.103	118.8/118.1	34.0	3.50
³ ($n \rightarrow \pi \rightarrow$) exc. st.					
CASSCF ^c	1.335	1.112	113.9	43.7	3.08
MRDCT ^c	1.313	1.100	116.3	40.0	3.22
CCSD(T) ^d	1.314	1.090	115.6	41.4	3.22
PW91/TZ2P ^e	1.305	1.106	112.9	48.9	2.88
Exp. ^c	1.307	1.084	121.8	41.1	3.12

^aOut-of-plane bending angle.^bAdiabatic relative energy w.r.t. ground state, in eV.^cFrom ref. 50.^dFrom ref. 51.^eThis work.

Odd behavior is observed for the angle of H₂Se, for which a difference of $\sim 0.7^\circ$ is observed between using either the analytical or the numerical gradient (see Table 2). As the discrepancy occurs for both the nonrelativistic and the scalar ZORA calculations, it cannot be caused by the incomplete analytical ZORA gradient. The latter is known to contain a slight mismatch between the energy expression and the potential, which leads to a geometry with zero gradient that does not exactly coincide with the energy minimum (difference of ~ 0.1 pm). The cause of the discrepancy is more likely to result from the projecting out of the FC electrons,^{39,41} as the corresponding all-electron basis set results in an H₂Se angle (both with numerical and analytical gradients) that matches the value obtained with frozen-core and numerical gradients.

The second example of using numerical gradients is the geometry optimization of the first singlet and triplet excited states of formaldehyde. In this case, we perform at each displaced geometry a TD-DFT calculation and add the excitation energy to the ground state energy at the displaced geometry to obtain the numerical derivative of the excited state energies. The optimized geometry for the ground state, first singlet excited state, and first triplet excited state is given in Table 3, where also some results from the literature^{50,51} are given for reference. The purpose of the current contribution is not to get the best possible DFT result for these excited states, but merely to show the possibility to calculate the excited state geometries. Therefore, we simply used the same DFT setup as for the other calculations (PW91/TZ2P). The obtained adiabatic excitation energies and structures of the excited states compare reasonably well with either *ab initio* or experimental data, that is, the first triplet excited state lies below the first singlet excited state, which are both found around 3.0–3.5 eV above the ground state. Moreover, the increase in

C—O distance by ~ 0.10 Å, changes in HCH angle (an increase for singlet excited state, a decrease for triplet excited state) and the out-of-plane bending observed in both the *ab initio* and experimental data are well reproduced by our approach (see Table 3).

Multilevel QM/QM and QM/MM Results

The main part of the results section will deal with examples of using the QUILD program for multilevel QM/QM or QM/MM approaches. Whenever the boundary between two regions cuts through covalent bonds, the AddRemove scheme⁴⁷ (*vide supra*) has been used to satisfy the valence of dangling bonds.

The setup of a multilevel QM/MM or QM/QM scheme is easy to construct in the QUILD program, by dividing a molecule into different regions: for instance for a DNA duplex (see Fig. 2) regions 1–4 comprise the four bases, while region 5 contains the remaining atoms. At each geometry step, a sequence of calculations is performed (see Fig. 2). First the whole system is computed at BP86 (I, in blue). Second, for the left-side π -stacked basepair (regions 1 and 2) two series of calculations are carried out: one at LDA (IIa, in yellow) and one at BP86 (IIb, in blue). These series comprise three jobs: first with the stacked base pair (regions 1 and 2), followed by a second job of one of the bases (region 1), and a third with the other base (region 2) to give only the interaction between the bases. Note that the interactions *within* region 1 and *within* region 2 are still described at BP86. Only the interaction between the two regions is described at LDA. Third and finally, the corresponding series of calculations for the right-side π -stacked basepair: one at LDA (IIIa, in pink) and one BP86 (IIIb, in blue). All individual jobs are taken care of by the QUILD program. The user only has to specify in the QUILD input block which description to use for which regions, as shown in Scheme 1 below.

The aforementioned regions (molecular fragments) and descriptions (quantum chemical methods, program packages) also have to be defined in the input (a full example input file is given in the supporting information).

The total energy expression for the QM/QM scheme is finally obtained as a combination of the different energies from the aforementioned calculations (where E_1 stands for BP86 and E_2 for LDA in this example):

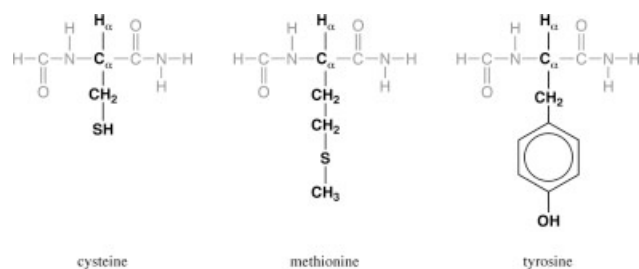
$$\begin{aligned}
 E_{\text{tot}} &= E_1(1-5) + \Delta E_2(1-2) - \Delta E_1(1-2) \\
 &\quad + \Delta E_2(3-4) - \Delta E_1(3-4) \\
 \Delta E_x(a-b) &= E_x(a-b) - E_x(a) - E_x(b)
 \end{aligned}
 \quad (3)$$

```

QUILD
INTERACTIONS
  total description 1
  intxn region 1 region 2  description 2 for description 1
  intxn region 3 region 4  description 2 for description 1
SUBEND
END

```

Scheme 1. Input example of multi-level energy expression setup.



Scheme 2. Amino acid residues studied with QM/QM, QM/MM, and QM schemes (central atoms in black comprise region A, remaining (dipeptide) atoms in grey region B).

The results of applying the aforementioned QM/QM setup on a DNA duplex is given below. The energy expression of eq. (1) is simplified whenever entire regions are replaced instead of the $\Delta E_{\alpha(a-b)}$ interactions between regions as given above. For instance, for the protein system of Figure 1, the total energy expression would simply be (where E_1 stands for MM and E_2 for DFT in this example):

$$E_{\text{tot}} = E_1(1 - 3) + E_2(1) - E_1(1)$$

Results for QM/QM, QM/MM, and QM Study on Amino Acids

The first systems we studied with multi-level approaches, consist of the dipeptide form of three amino acid residues, i.e., $\text{HCONHC}_\alpha\text{H}_\alpha\text{RCONH}_2$ molecules where **R** represents the amino acid residue sidechain (see Scheme 2). The studied residues are

cysteine (Cys), methionine (Met), and tyrosine (Tyr). In all cases were the molecules divided into two regions (see Scheme 2), the first (region A) consists of the sidechain atoms **R** with the backbone C-alpha and its attached hydrogen ($\text{C}_\alpha\text{H}_\alpha$), the second (region B) consists of the remaining (dipeptide) atoms. The molecules were studied with different multilevel choices: (i) QM at BP86/TZ2P; (ii) region A at BP86/TZ2P, region B at MM level using AMBER95 force field; (iii) region A at BP86/TZ2P, region B at BP86/DZP; (iv) region A at BP86/TZ2P, region B at BP86/DZ; (v) region A at BP86/TZ2P, region B at LDA/TZ2P; (vi) region A at BP86/TZ2P, region B at LDA/DZ; (vii) region A at BP86/TZ2P, region B at AM1; and (viii) region A at BP86/TZ2P, region B at RHF/STO-3G.

The aim of studying these three different amino acid residues is to investigate how the different multilevel scheme options influence the boundary region between different QM/MM or QM/QM regions. As the bonds at or close to this boundary are very likely to be affected most by the choice of using multilevel schemes, we compare here the bond lengths for the bonds between C_α and its bonded atoms. The $\text{C}_\alpha\text{—H}_\alpha$ distance does not change significantly (see Table 4) compared to scheme i when using either MM or DFT for region B (schemes ii–vi), but shows a somewhat larger increase (0.01 Å) when using RHF or AM1 for region B (schemes vii–viii).

Likewise, the $\text{C}_\alpha\text{—C}_\beta$ distance remains more or less constant (deviation of ~ 0.01 Å) around 1.54 Å, irrespective of the nature of the side chain and which multilevel scheme is used. This was to be anticipated as the $\text{C}_\alpha\text{—C}_\beta$ atoms are in region A that is always treated at the BP86/TZ2P level. The only exception is scheme vii, where region B is described at the semiempirical AM1 level, when the $\text{C}_\alpha\text{—C}_\beta$ distance goes to 1.59–1.60 Å. Apparently, the AM1 treatment of region B pulls that hard on

Table 4. Backbone Distances (Å) for Amino Acid Residues with Different Multilevel Schemes.

Scheme	i	ii	iii	iv	v	vi	vii	viii
Region A	BP86/TZ2P	BP86/TZ2P	BP86/TZ2P	BP86/TZ2P	BP86/TZ2P	BP86/TZ2P	BP86/TZ2P	BP86/TZ2P
Region B	BP86/TZ2P	AMBER95	BP86/DZP	BP86/DZ	LDA/TZ2P	LDA/DZ	AM1	RHF/STO-3G
Cys								
$\text{C}_\alpha\text{—N(dp)}$	1.456	1.473	1.455	1.470	1.431	1.445	1.437	1.483
$\text{C}_\alpha\text{—C(dp)}$	1.556	1.544	1.554	1.556	1.522	1.524	1.535	1.565
$\text{C}_\alpha\text{—C}_\beta$	1.536	1.541	1.537	1.532	1.535	1.532	1.602	1.546
$\text{C}_\alpha\text{—H}_\alpha$	1.097	1.099	1.098	1.096	1.100	1.100	1.110	1.105
Met								
$\text{C}_\alpha\text{—N(dp)}$	1.456	1.472	1.454	1.470	1.425	1.441	1.435	1.479
$\text{C}_\alpha\text{—C(dp)}$	1.542	1.539	1.541	1.544	1.513	1.515	1.547	1.563
$\text{C}_\alpha\text{—C}_\beta$	1.546	1.548	1.549	1.545	1.554	1.550	1.601	1.555
$\text{C}_\alpha\text{—H}_\alpha$	1.097	1.097	1.098	1.096	1.101	1.100	1.111	1.107
Tyr								
$\text{C}_\alpha\text{—N(dp)}$	1.467	1.469	1.466	1.480	1.441	1.455	1.438	1.487
$\text{C}_\alpha\text{—C(dp)}$	1.560	1.544	1.557	1.559	1.534	1.531	1.551	1.567
$\text{C}_\alpha\text{—C}_\beta$	1.535	1.553	1.533	1.529	1.540	1.530	1.589	1.552
$\text{C}_\alpha\text{—H}_\alpha$	1.101	1.101	1.102	1.100	1.105	1.104	1.114	1.113
MAD ^a		0.010	0.002	0.006	0.021	0.015	0.031	0.016
MAX ^b		0.018	0.003	0.014	0.034	0.032	0.066	0.027

^aMean absolute deviation (Å) with respect to QM values of scheme i, excluding $\text{C}_\alpha\text{—H}_\alpha$.

^bMaximum absolute deviation (Å) with respect to QM values of scheme i, excluding $\text{C}_\alpha\text{—H}_\alpha$.

C_α that the $C_\alpha-C_\beta$ distance increases by ~ 0.06 Å. At the same time, the $C_\alpha-N$ (dp) distance within scheme vii is lower than the parent QM calculation (scheme i) by 0.02–0.03 Å, while the $C_\alpha-C$ (dp) distance is more or less equal. Also with schemes v and vi, where region B is treated with LDA, do we observe a decrease of the $C_\alpha-N$ (dp) and $C_\alpha-C$ (dp) distances. Interestingly, the QM/MM calculation (scheme ii) gives geometries around the QM/MM boundary region that follow closely those of the reference QM calculation (scheme i), and the QM/QM calculations where the BP86 functional is used for region B, but with a smaller basis set (schemes iii and iv). The mean absolute deviations with respect to the QM distances of scheme i are, respectively, 0.010, 0.002, 0.006 Å for schemes ii–iv.

Apart from AM1, which leads to a significant increase of the $C_\alpha-C_\beta$ distance, all multilevel schemes studied here do not seem to disturb the boundary region significantly, when used with the AddRemove⁴⁷ link-atom scheme. However, other link-atom schemes may lead to larger distortions of the boundary region, as previously shown for the IMOMM and ONIOM2 schemes within QM/MM calculations.⁴⁷

Results for QM/MM and QM Study on S_N2 Reaction in Water

The second system we studied by a multilevel approach consists of the bimolecular nucleophilic substitution (S_N2) reaction of F^- with CH_3Cl . We studied this reaction both in the gas-phase and solvated in water, for which we used explicit solvent molecules within the QM/MM scheme. Nucleophilic substitution reactions with carbon as central particle are usually associated with a double-well energy profile, wherein the reactants first form a reactant complex (RC). The RC is stabilized by ion-dipole interactions and is separated from a similar product complex (PC) through a central barrier associated with the transition state (TS). The depth of the wells, the barrier height and the reaction energy depend on the difference in nucleophilicity between the nucleophile and the leaving group. For instance, for the S_N2 reaction between hydroxide and methylchloride, the difference in nucleophilicity is sufficiently large to give a very exothermic reaction with a single-well energy surface: the only stationary intermediate point is a PC where the hydroxide is bound to carbon and the chloride has been expelled without barrier. However, here we study the reaction of fluoride with methylchloride, where although the difference in nucleophilicity between fluoride and chloride is substantial, the typical double-well energy profile is present. Given in Figure 4 is the gas-phase energy profile at the OLYP/TZ2P level. This level of theory gives improved results for the geometries, complexation energy, and energy barriers for S_N2 reactions.^{28,52} The RC complex is stabilized with respect to reactants by 14.5 kcal/mol, and after crossing a small barrier of 0.3 kcal/mol, the PC complex is reached that is 24.2 kcal/mol more stable than the RC. Note that the value of the reaction barrier is underestimated at this level of theory by 2–4 kcal/mol, both because of the basis set and the DFT functional used.²⁸ However, the purpose of the current contribution is to focus on the effect of the solvent on the energy profile, and not as much the “best” value for the reaction barrier

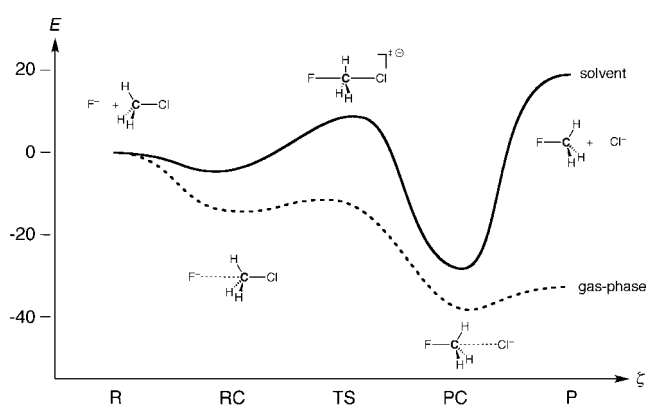


Figure 4. Energy profile (kcal/mol) for bimolecular nucleophilic substitution reaction of $F^- + CH_3Cl$ in the gas phase and in water.

that might have been obtained at the mPBE0KCIS/QZ4P level²⁸ (albeit with numerical gradients, *vide supra*). Finally, the products are found to be ~ 32 kcal/mol more stable than reactants in the gas-phase.

For studying the reaction in solvent, for each stationary point along the energy profile water molecules were placed at random within a box around the solute. The box has a cubic shape (with a side of 19 Bohrs, containing 32 water molecules) for reactants and products, and a rectangular shape (sides 19, 19 and 38 Bohrs, containing 64 water molecules) for RC, TS, and PC. Together with the solute, these solvent boxes correspond to a macroscopic density of 1.0 kg/l. The positions of the randomly placed water molecules were subsequently improved in 20 ps of MD simulations with the DRF90 program⁵³ using simulated annealing techniques. These MD simulations used an NVT ensemble⁵⁴ (at different temperatures, starting from 500 K and ending at 2 K), and a soft wall-force potential⁵³ to prevent the water molecules from evaporating from the simulation box.

The lowest energy configuration from the MD simulations was taken as starting point for QM/MM geometry optimizations within QUILD. The solute was in these QM/MM calculations again described at the OLYP/TZ2P level. The water molecules and the interaction between the solute and waters were treated with the AMBER95 force field, using the TIP3P model⁵⁵ for water. Within the QM/MM calculations in QUILD, a wall-force potential similar to that of DRF90 was applied.

For protic solvents such as water, the solvent effect is largest for those systems where the charge is localized more, and smallest where the charge is more delocalized. This can be seen clearly in the energy profile for the S_N2 reaction (see Fig. 4), where the reaction is exothermic (-32.3 kcal/mol) in the gas-phase and endothermic ($+19.6$ kcal/mol) in water. This results mostly from the solvation energy of F^- , which is ~ 43 kcal/mol larger than that of Cl^- , and thus accounts for more than 80% of the solvent effect for the products. Also the differences in solvent effect between the reactant (RC) or product (PC) complex on one side, and the TS on the other, can be understood in terms of the larger charge delocalization in the TS. Although all three systems are destabilized in water compared to the gas-phase, for the RC and PC the destabilization is a mere 10 kcal/mol with

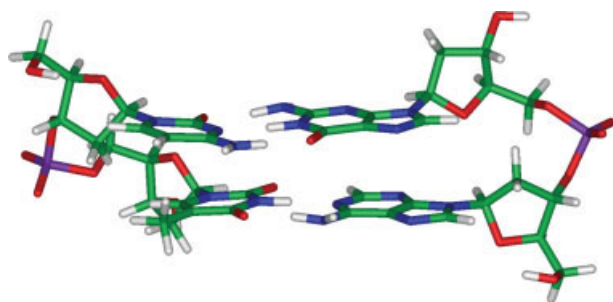


Figure 5. QM/QM optimized structure of DNA duplex (CG-TA); see Table 5 for numerical values. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

respect to the reactants, while for the TS it is ~ 25 kcal/mol (see Fig. 4).

The presence of the solvent shifts the position of the several stationary points along the reaction coordinate to some extent. The RC complex, which in the gas-phase occurs at a C—F distance of 2.51 Å, is in the solvent environment found with a C—F distance of 2.78 Å. This shift of 0.27 Å happens mainly because of the preferential interactions of F[−] with the solvent, which compete with the ion-dipole interactions of the RC complex. Because of the larger delocalization of charge in the RC complex compared to the reactants, the RC complex is destabilized in the solvent compared with the gas-phase situation; the gas-phase complexation energy of ~ 15 kcal/mol is largely reduced in the solvent to only 4.6 kcal/mol. Also the position of the TS along the reaction coordinate shifts because of the solvent environment, but unlike the RC complex that is encountered earlier, the TS occurs later, that is, in the gas-phase the C—F distance of the TS is 2.15 Å, while in solvent it shifts to 2.07 Å. Similar to the RC complex, within the PC complex the distance of the ion shifts substantially due to its solvation, that is, in the gas-phase the C—Cl distance is 3.56 Å, while in solvent it increases to 3.95 Å.

Results for QM/QM, QM/MM, and QM Study on DNA Duplexes

Finally, we would like to present an example of the QM/QM scheme for which the QUILD program has been designed. From studies on π -stacking of DNA bases and related systems,^{31,56,57} it is known that standard DFT functionals as BP86, B3LYP, or OPBE do not give a proper description of the favorable π -stacking interactions, but generally result in repulsive energy surfaces. Recent studies^{31,56} showed, however, the existence of a few DFT functionals that are able to correctly give the favorable π -stacking interactions. Interesting in this respect are the KT1 and the related KT2 functional,⁵⁸ which were constructed to reproduce the near-exact multiplicative Zhang-Morrison-Parr (ZMP) potential from coupled cluster Brueckner (BD) densities. As a result, the KT1/KT2 functionals produce NMR chemical shifts that in general show smaller deviations from experimental chemical shifts than those from standard DFT functionals. Furthermore, the ZMP potential, constructed from Brueckner (BD(T))

coupled cluster densities, gives dispersion forces in good agreement with near-exact dispersion forces.⁵⁹ Therefore, it is to be expected that the KT1, KT2 functionals behave properly for dispersion forces as well. The application of these functionals, and surprisingly also LDA, to π -stacking of DNA bases indeed showed excellent results compared to MP2 and CCSD(T) data.³¹

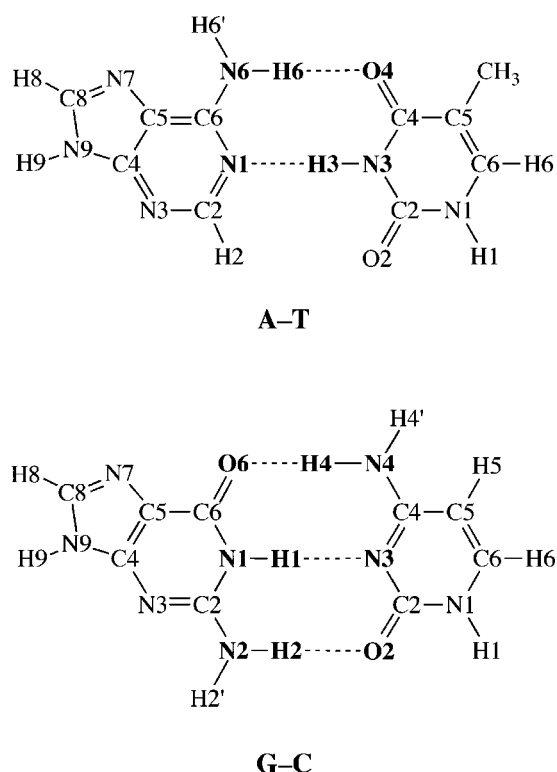
The structure of DNA is, however, not only held together by π -stacking interactions within the DNA strands, but also by hydrogen-bonding interactions between the DNA bases. The latter hydrogen-bonding interactions are strong and specific, and well understood by high-level theoretical analyses at the BP86/TZ2P level.^{30,32,60–63} Unfortunately, the functionals that give a good description for the hydrogen-bonding interactions, fail for the π -stacking interactions³¹; vice versa, the functionals that work well for π -stacking do not give good hydrogen-bonding interactions. Therefore, our QM/QM scheme where we can exploit the strong points of each functional is mandatory for a computational study on the structure of DNA. For instance, in our QM/QM scheme we can use the BP86 functional for the hydrogen-bonding (and intramolecular) interactions, and combine it with e.g. LDA for the π -stacking interactions. This is schematically indicated in Figure 2, where the total system is described by BP86 (in blue), which is combined with π -stacking interactions (described in yellow and pink) at the LDA level. As an example of this QM/QM scheme, we report here the application to a DNA (dGpdA)(dTpdC) duplex, as shown in Figure 5.

The QUILD-optimized H-bond lengths in the DNA duplex (see Fig. 5 and Table 5) agree well with those observed previously for isolated Watson–Crick pairs of nucleic acids.^{30,32,60–63} Thus, for the G–C base pair in DNA duplex we find O6—N4, N1—N3, and N2—O2 hydrogen-bond lengths of 2.74, 2.88, and 2.82 Å (see Scheme 3). And for the A–T base pair, we find N6—O4 and N1—N3 hydrogen-bond lengths of 2.85 and 2.82 Å (Scheme 3). The stacking distances of 3.0–3.1 Å between the stacked bases are somewhat shorter than experimentally observed for long DNA helices, but this is most likely due to termination effects in combination with the absence of the molecular environment (counter-ions, water, etc.) in our model system. A similar pattern emerges in the case of MD simulations of proteins in the absence of a solvent layer, which leads to compression of the protein structure.

Table 5. Geometrical Parameters (Å) for Optimized Structure of DNA Duplex (CG-TA).^a

H-bonded basepair	Distance	Stacked basepair	Distance
CG			
N4(C)—O6(G)	2.739	CT	2.949
N3(C)—N1(G)	2.884	GA	3.018
O2(C)—N2(G)	2.823		
TA			
O4(T)—N6(A)	2.853	CG-TA	3.078
N3(T)—N1(A)	2.823		

^aSee Figure 5 for a visualization of the structure, and Scheme 3 for atom numbers.



Scheme 3. Structure of isolated DNA basepairs.

Conclusions

We have presented the new program QUILD, which is designed for performing calculations through a multilevel approach in which different regions of a (supra)molecular system can be described with various *ab initio*, DFT, semiempirical, and force field methodologies. QUILD distinguishes itself from other multilevel approaches (e.g., ONIOM) by the fact that these regions can be chosen arbitrarily and may in principle overlap or interpenetrate, i.e., they do not have to conform to a shell structure. This generates a powerful flexibility that can be used not only to make computations cheaper (e.g., by treating the active center at a high level and the remainder of the system at a lower level of theory) but also to tackle each type of interaction or phenomenon with the “best” methodology for that purpose (e.g., hydrogen bonding with BP86, stacking with LDA, and transition states with OLYP). QUILD then seamlessly combines all regions and quantum chemical approaches to yield the “optimal” description for the complete system.

Another important issue tackled here is how to obtain an accurate structure, which may become troublesome for systems held together by weak interactions. A major step forward has been achieved by our adaptation of the delocalized coordinates setup by Baker et al., in which we distinguish weak from strong coordinates. Our setup for geometry optimization has also been integrated into QUILD together with a model Hessian, Quasi-Newton optimizer, regulated GDIIS, Restricted Second Order model, and a dynamic trust radius. We have tested our geometry

optimization setup for two test sets: (i) the Baker test set for *strong coordinates* (Baker test set), and (ii) a new test set of 18 *weakly-bound* molecules. For the Baker set, we need 167 geometry steps at the RHF/STO-3G level for full optimization, and 164 at the PW91/TZ2P level. The screening of the coordinates, i.e., distinguishing weak from strong coordinates, has almost no effect on the performance, but we do see a significant reduction (30%) in the number of steps when using our adapted delocalized coordinates instead of Cartesians. When using Cartesian coordinates for the Baker set, the performance with using either the old-style optimizer in ADF or QUILD is similar. However, for the set of weakly-bound molecules, this picture changes, with the old-style optimizer in ADF needing more than four times as many geometry steps as QUILD. Thus, QUILD achieves a reduction of more than 75% for the number of geometry steps needed. For the weakly-bound set, the screening of the coordinates does have a beneficial effect: it reduces the number of geometry steps by 20–25%. The QUILD program also allows for geometry optimization using numerical gradients, for approaches within ADF or ORCA for which no analytical gradient is available yet, and we have shown this setup using two examples, using nonrelativistic, scalar- and spin-orbit relativistic ZORA treatment of halogen dimers and group 16 hydrides.

Finally, we have applied the QUILD program to a number of multilevel (QM/MM or QM/QM) schemes. We studied the multilevel boundary region of amino acid residues, where we show that the boundary region does not change dramatically due to the multilevel setup when using the AddRemove link-atom scheme. Next, we investigated the solvent effects on the energy profile of a bimolecular nucleophilic substitution (S_N2) reaction, where we observe significant changes of the energy profile. Furthermore, we investigated a DNA duplex system, the ultimate goal of the QUILD project, using a QM/QM scheme in order to take full advantage of the seamless combination of different DFT functionals for different interactions.

References

1. Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J Phys Chem* 1996, 100, 19357.
2. Vreven, T.; Morokuma, K. *J Chem Phys* 2000, 113, 2969.
3. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J Am Chem Soc* 1995, 117, 5179.
4. Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S. J.; Karplus, M. *J Comput Chem* 1983, 4, 187.
5. van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P. K. H. C.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; vdf Hochschulverlag AG: Zürich, 1996.
6. Jorgensen, W. L.; Tirado-Rives, J. *J Am Chem Soc* 1988, 110, 1657.
7. Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*; Wiley-VCH: Weinheim, 2000.
8. Parr, R. G.; Yang, W. *Density Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989.
9. Dreizler, R.; Gross, E. *Density Functional Theory*; Plenum: New York, 1995.

10. Kamachi, T.; Yoshizawa, K. *J Am Chem Soc* 2003, 125, 4652.
11. Perdew, J. P.; Ruzsinszky, A.; Tao, J. M.; Staroverov, V. N.; Scuseria, G. E.; Csonka, G. I. *J Chem Phys* 2005, 123, 062201.
12. Rosa, A.; Ricciardi, G.; Baerends, E. J. *J Phys Chem A* 2006, 110, 5180.
13. Schoneboom, J. C.; Lin, H.; Reuter, N.; Thiel, W.; Cohen, S.; Ogliaro, F.; Shaik, S. *J Am Chem Soc* 2002, 124, 8142.
14. Costas, M.; Ribas, X.; Poater, A.; López Valbuena, J. M.; Xifra, R.; Company, A.; Duran, M.; Solà, M.; Llobet, A.; Corbella, M.; Usón, M. A.; Mahía, J.; Solans, X.; Shan, X. P.; Benet-Buchholz, J. *Inorg Chem* 2006, 45, 3569.
15. Sala, X.; Plantalech, E.; Romero, I.; Rodríguez, M.; Llobet, A.; Poater, A.; Duran, M.; Solà, M.; Jansat, S.; Gómez, M.; Parella, T.; Stoeckli-Evans, H.; Benet-Buchholz, J. *Chem—Eur J* 2006, 12, 2798.
16. Poater, A.; Moradell, S.; Pinilla, E.; Poater, J.; Solà, M.; Martínez, M. A.; Llobet, A. *Dalton Trans* 2006, 1188.
17. Martín, N.; Altable, M.; Filippone, S.; Martín-Domech, A.; Poater, A.; Solà, M. *Chem—Eur J* 2005, 11, 2716.
18. Duran, J.; Polo, A.; Real, J.; Benet-Buchholz, J.; Poater, A.; Solà, M. *Eur J Inorg Chem* 2003, 4147.
19. Handy, N. C.; Cohen, A. J. *Mol Phys* 2001, 99, 403.
20. Becke, A. D. *Phys Rev A* 1988, 38, 3098.
21. Hoe, W. M.; Cohen, A. J.; Handy, N. C. *Chem Phys Lett* 2001, 341, 319.
22. Baker, J.; Pulay, P. *J Chem Phys* 2002, 117, 1441.
23. Xu, X.; Goddard W. A., III. *J Phys Chem A* 2004, 108, 8495.
24. Swart, M.; Ehlers, A. W.; Lammertsma, K. *Mol Phys* 2004, 102, 2467.
25. Swart, M.; Snijders, J. G. *Theor Chem Acc* 2003, 110, 34; Erratum: *ibid* 111, 156.
26. Swart, M.; Groenhof, A. R.; Ehlers, A. W.; Lammertsma, K. *J Phys Chem A* 2004, 108, 5479.
27. Swart, M. *Inorg Chim Acta* 2007, 360, 179.
28. Swart, M.; Solà, M.; Bickelhaupt, F. M. *J Comput Chem* 2007, 28, 1551.
29. Grüning, M.; Gritsenko, O.; Baerends, E. J. *J Phys Chem A* 2004, 108, 4459.
30. van der Wijst, T.; Fonseca Guerra, C.; Swart, M.; Bickelhaupt, F. M. *Chem Phys Lett* 2006, 426, 415.
31. Swart, M.; van der Wijst, T.; Fonseca Guerra, C.; Bickelhaupt, F. M. *J Mol Model* (in press).
32. Fonseca Guerra, C.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *J Am Chem Soc* 2000, 122, 4117.
33. Perdew, J. P. *Phys Rev B* 1986, 33, 8822–8824; Erratum: *ibid* 8834, 7406.
34. Swart, M.; Bickelhaupt, F. M. *Int J Quant Chem* 2006, 106, 2536.
35. Baker, J.; Kessi, A.; Delley, B. *J Chem Phys* 1996, 105, 192.
36. Baker, J. *J Comput Chem* 1993, 14, 1085.
37. Swart, M.; Bickelhaupt, F. M. Program code QUILD, Amsterdam, 2005.
38. Baerends, E. J.; Autschbach, J.; Bérces, A.; Bickelhaupt, F. M.; Bo, C.; de Boeij, P. L.; Boerrigter, P. M.; Cavallo, L.; Chong, D. P.; Deng, L.; Dickson, R. M.; Ellis, D. E.; Fan, L.; Fischer, T. H.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Groeneveld, J. A.; Gritsenko, O. V.; Grüning, M.; Harris, F. E.; van den Hoek, P.; Jacob, C. R.; Jacobsen, H.; Jensen, L.; van Kessel, G.; Kootstra, F.; van Lenthe, E.; McCormack, D. A.; Michalak, A.; Neugebauer, J.; Osinga, V. P.; Patchkovskii, S.; Philipsen, P. H. T.; Post, D.; Pye, C. C.; Ravenek, W.; Ros, P.; Schipper, P. R. T.; Schreckenbach, G.; Snijders, J. G.; Solà, M.; Swart, M.; Swerhone, D.; te Velde, G.; Vernooijs, P.; Versluis, L.; Visscher, L.; Visser, O.; Wang, F.; Wesolowski, T. A.; van Wezenbeek, E.; Wiesenekker, G.; Wolff, S. K.; Woo, T. K.; Yakovlev, A. L.; Ziegler, T. Program code ADF; SCM: Amsterdam, 2006.
39. te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Snijders, J. G.; Ziegler, T. *J Comput Chem* 2001, 22, 931.
40. van Lenthe, E.; Baerends, E. J.; Snijders, J. G. *J Chem Phys* 1993, 99, 4597.
41. van Lenthe, E.; Baerends, E. J. *J Comput Chem* 2003, 24, 1142.
42. Vosko, S. H.; Wilk, L.; Nusair, M. *Can J Phys* 1980, 58, 1200.
43. Perdew, J. P. In *Electronic structure of Solids* 1991; Ziesche, P.; Eschrig, H., Eds.; Akademie: Berlin, 1991; p. 11.
44. Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pedereson, M. R.; Singh, D. J.; Fiolhais, C. *Phys Rev B* 1992, 46, 6671.
45. Xu, X.; Goddard, W. A., III. *Proc Natl Acad Sci USA* 2004, 101, 2673.
46. Bakken, V.; Helgaker, T. *J Chem Phys* 2002, 117, 9160.
47. Swart, M. *Int J Quant Chem* 2003, 91, 177.
48. Eckert, F.; Pulay, P.; Werner, H. J. *J Comput Chem* 1997, 18, 1473.
49. Lindh, R.; Bernhardsson, A.; Karlstrom, G.; Malmqvist, P. A. *Chem Phys Lett* 1995, 241, 423.
50. Angeli, C.; Borini, S.; Ferrighi, L.; Cimiraglia, R. *J Mol Struct (Theochem)* 2005, 718, 55.
51. Yamaguchi, Y.; Wesolowski, S. S.; Van Huis, T. J.; Schaefer, H. F., III. *J Chem Phys* 1998, 108, 5281.
52. Bento, A. P.; Solà, M.; Bickelhaupt, F. M. *J Comput Chem* 2005, 26, 1497.
53. Swart, M.; van Duijnen, P. T. *Mol Simulat* 2006, 32, 471.
54. Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon: Oxford, 1987.
55. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J Chem Phys* 1983, 79, 926.
56. Waller, M. P.; Robertazzi, A.; Platts, J. A.; Hibbs, D. E.; Williams, P. A. *J Comput Chem* 2006, 27, 491.
57. Cerny, J.; Hobza, P. *Phys Chem Chem Phys* 2005, 7, 1624.
58. Keal, T. W.; Tozer, D. J. *J Chem Phys* 2003, 119, 3015.
59. Allen, M. J.; Tozer, D. J. *J Chem Phys* 2002, 117, 11113.
60. Fonseca Guerra, C.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *Chem—Eur J* 1999, 5, 3581.
61. Fonseca Guerra, C.; Bickelhaupt, F. M. *Angew Chem Int Ed* 1999, 38, 2942.
62. Fonseca Guerra, C.; Bickelhaupt, F. M. *Angew Chem Int Ed* 2002, 41, 2092.
63. Fonseca Guerra, C.; Bickelhaupt, F. M.; Baerends, E. J. *Cryst Growth Des* 2002, 2, 239.