Scientific Computing & Modelling

ADF QM/MM Manual

ADF Program System
Release 2013

Table of Contents

1 Introduction

1.1 General remarks

About this Document

The NEWQMMM subkeyword is available since adf2008.01, which allows large QM/MM calculations. The pdb2adf documentation is available since adf2005.01.

This document describes the QM/MM option within ADF, how to use it, how to set up inputs, what features are available, what its limitations are, and so on. This manual assumes that the reader already has experience with ADF and has some basic knowledge of molecular mechanics (MM) and combined QM/MM theories. A brief overview of the combined QM/MM methods is included.

This document is organized in the following manner. First, concepts and naming conventions that are used throughout the document are introduced. In Chapter 2, setting up a QM/MM simulation with ADF is discussed with a detailed description of all the input options (sections 2.1 and 2.2) and the Force Field files (section 2.3). Section 2.4 contains a 'walk through' of how to set up QM/MM jobs.

The combined QM/MM code and the documentation are in continuous development. We appreciate any comments, bug reports and suggestions for its improvement.

Summary of Functionality

Currently, the QM/MM implementation within ADF is based on a modified version [1] of the 'IMOMM' scheme of Maseras and Morokuma [2] (called the IMOMM/ADF scheme); alternatively, the recently developed AddRemove scheme [3] is available. The molecular mechanics code has been designed to be as flexible as possible, allowing for many levels of customization. As a result of this flexibility, operation of the program requires the user to have some experience with molecular mechanics methods. At the same time, ADF remains the main driver to control the simulation of the whole QM/MM system, since one of the objectives of the implementation has been to treat the MM subsystem as a perturbation to the QM system.

We summarize the current functionality and limitations of the implementation:

- Morokuma's IMOMM [2] QM/MM coupling scheme (IMOMM/ADF).
- AddRemove [3] coupling scheme.
- AMBER95 [4] and SYBYL (Tripos) [5] force fields provided.
- Free form and modifiable force field parameter file.
- Ability to mix and match force field functions.
- Geometry optimization, linear transit and transition state optimizations (In these optimizations, the MM region is fully optimized between each SCF cycle or in other words between each QM geometry step).
- All ADF input 'styles' available (Cartesian, Z-matrix, mixed...).
- Any number of covalent bonds can cross the QM-MM boundary.
- Global optimization of the MM subsystem available (simulated annealing and grid search algorithms).

Some notable limitations are:

- Symmetry constraints cannot be applied, unless the MM subsystem is frozen.
- Geometry constraints involving MM atoms are limited.
- Not more than one QM atom can be bonded to a single MM atom.
- Not compatible with some ADF extensions such as NMR, ESR, COSMO, IRC and excited states.

- The AddRemove model is available only when using Cartesian coordinates.
- The conjugate gradient optimizer is available only for Cartesian coordinates.

Applicability

This QM/MM implementation evolved from research on organometallic complexes and catalytic systems. Since these systems are generally under 1000 atoms in size, the program had not been optimized to handle large macromolecular systems such as enzymes. In particular, the non-bonded code was not efficient for very large systems. It has been adjusted in the ADF2002.01 release, in order for the code to work efficiently on both small and large systems. Moreover, the calculation of the MM forces has been parallellized to further increase the efficiency and applicability. With the current release, systems with up to several thousand atoms have been tested without problems. Although the input is not designed to handle the topology of macromolecules such as amino acids and peptides, it can be constructed quite easily. For the large systems this amounts however to some 30000 lines; therefore, tools (pdb2adf) have been developed for creating input files automatically for a given PDB-file. This tool is available through a link on the 'Contributed Software' part of the SCM web site http://www.scm.com, but is now also part of the official release starting with adf2005.01.

Currently, only the AMBER95 and SYBYL force fields are included. This might also limit the applicability. However, the force field parameters and potentials are fairly customizable and other force fields are easily added.

The utility pdb2adf

Starting from the adf2005.01 version the utility pdb2adf is available in the official release. This utility creates an ADF input file from a PDB file, for a subsequent QM/MM calculation using ADF. This tool has been developed by Marcel Swart. Previously this utility could be found in the contributed software page. Starting from adf2008.01 there is support for the NEWQMMM subkey.

1.2 Concepts and Terminology

LINK bonds

When performing a QM/MM simulation, one often wants to partition the system such that some covalent bonds cross the QM/MM boundary. These so-called 'link' bonds demand special attention in any QM/MM implementation. The link bonds are a critical aspect of the QM/MM method used here and a good understanding of the concepts is essential. In this section we describe how they are treated in ADF and we introduce the nomenclature that is used throughout the manual.

a. Full QM/MM or "real" system

b. Model QM system

c. Atom labeling used in ADF QM/MM

d. Relationship of coordinates in a LINK bond

$$X_{MM} = \alpha X_{cap} + (1-\alpha)X_{QM}$$

$$\alpha = \frac{R}{r}$$
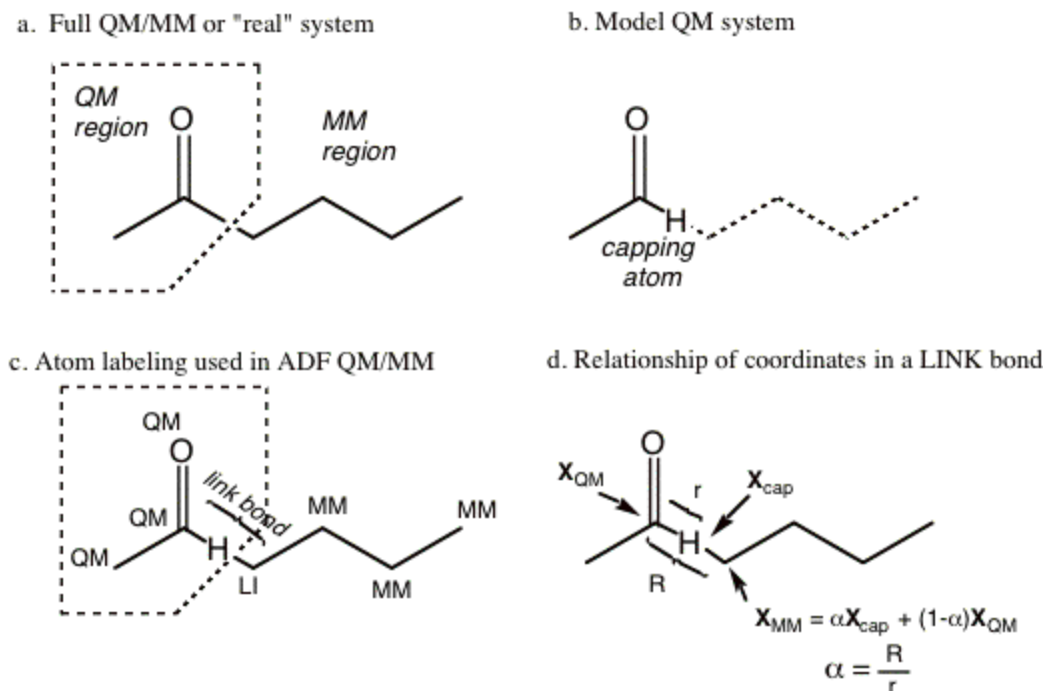
Figure 1-1 Example of QM/MM partitioning and details of the naming conventions adopted in ADF.

Figure 1-1a depicts a simple molecular system that has been divided into QM and MM regions by the dotted polygon. In this example there is only one link bond, or one covalent bond that traverses the QM-MM boundary (that cross the dotted polygon). When a covalent bond traverses the QM-MM boundary, the electronic system of the QM region must in some way be truncated across this bond. Several methods of dealing with this problem have been proposed in the literature. By far the most commonly adopted method, which was originally introduced by Singh and Kollman [6], involves capping the QM system with a 'dummy atom' or what we call a 'capping atom' (We use capping atom to avoid confusion with the dummy atoms used in ADF). Since the pioneering work of Singh and Kollman [6] many variations of the basic capping atom approach have evolved. In ADF, one of the adopted approaches is the one developed by Maseras and Morokuma which has been given the name the 'Integrated Molecular Orbital and Molecular Mechanics' or the IMOMM method by the authors. The key feature of the capping atom approach is that the electronic structure calculation is performed on what is referred to as the 'QM model' system where the MM region is removed and replaced with capping dummy atoms (often hydrogen but not necessarily so). Figure 1-1b depicts the QM model system corresponding to the example system presented in Figure 1-1a. The capping atom satisfies the valence requirements of the QM region and allows for a standard electronic structure calculation to be performed on the QM fragment. It is important to realize that the capping atom is not part of the real system, but is simply an atom that is introduced to truncate the electronic system of the QM region. This is why it is often referred to as the dummy atom. For every 'link' bond there are three atoms of importance, the capping atom and the two atoms that are part of the 'real' link bond - one from the QM region and one from the MM region. Figure 1-1c illustrates the three atoms involved in the link bond. From this point on, we will refer to the MM atom that is part of the 'real' link bond as the link atom; it is labeled 'LI' in Figure 1-1c. Although both the QM and MM atoms that are part of the link bond could be considered 'link' atoms, we designate only the MM atom as the link atom, because it has a special place in the ADF QM/MM input. It is this atom that is replaced by the capping atom in the electronic structure calculation of the QM model system.

Although the capping atom approach is convenient from the standpoint of the electronic structure calculation, the 'extra' capping atoms complicates the situation, as they do not exist in the real system, see Figure 1-1c. For each link bond, there are potentially three extra nuclear degrees of freedom (corresponding to the Cartesian coordinates of the capping atom) that are not present in the real system. In the IMOMM/ADF implementation [1] we alleviate the problem by removing the MM atom that is part of the 'real' link bond

as a free variable. Instead we define its position in terms of the QM atom it is bonded to and the capping atom that replaces it in the QM model system. More specifically, the MM link atom is constrained to lie along the bond vector of the capping atom bond, via the simple relationship expressed in equation 1.1 and depicted in Figure 1-1d. Here, X_{MM}, X_{cap} and X_{QM} refer to the Cartesian coordinates of the subscripted atoms and a is a constant defined as the ratio of the real link bond length to that of the length of the capping bond.

$$X_{LI} = a \, X_{cap} + (1 - a) \, X_{QM} \quad (1.1)$$

For each link bond, there is a unique a parameter that is held constant throughout the simulation. Since the capping atom is often at a shorter distance than the real MM atom, alpha is usually greater than unity. For example, when a Hydrogen capping atom is used to cap a C-C single bond, a is around 1.38. Note that the energy depends on the value chosen for this parameter a, and comparisons between two systems are meaningful only when the values for them are chosen equal.

Although the position of the MM atom is not an independent variable (or free degree of freedom), the bond length of the link bond can change during a geometry optimization. If the capping bond in the model QM system stretches or contracts, so does the link bond in the full system. Note that any forces exerted on the LI atom are projected onto the connected QM atom and onto the capping atom. For more details see references [1, 2].

In the AddRemove model [3] the position dependence is reversed. Instead of the *real* classical LI atom following the *artificial* capping atom, it is the position of the capping atom that is constrained:

$$X_{cap} = X_{QM} + R_{eq} \, U_{LI\text{-}QM} \quad (1.2)$$

Here R_{eq} is the equilibrium distance (from the force field !) for the QM-cap bond, while $U_{LI\text{-}QM}$ is the unit vector pointing from QM to LI.

This means also that the artificial degrees of freedom of the *capping* atom are being removed ! Moreover, after every QM geometry optimization cycle, the interactions of the *artificial* capping atoms with the real QM atoms are corrected for by subtracting the corresponding MM interactions [3]. In this way, the capping atoms are Added and Removed. It has been shown [3] that this choice gives a much better description for the bonds involved in the QM/MM boundary region than the IMOMM/ADF scheme.

The ADF QM/MM Hybrid Potential

This section summarizes how the QM/MM hybrid potential is constructed in the IMOMM and AddRemove methods. A more detailed and formal discussion can be found in references [3] [2, 6]. The two basic components of the QM/MM potential are the potential arising from the electronic structure calculation of the QM model system and the potential arising from the molecular mechanics force field calculation.

In the IMOMM method the potential of the QM model system acts as a base where additional molecular mechanics potentials are added. When there are no covalent bonds that cross the QM/MM boundary the situation is straightforward. For example, consider a QM/MM simulation in which there are two molecules, one in the QM region and the other in the MM region such that no bonds cross the boundary. All MM potentials needed to define the MM molecule are included. Additionally, all non-bonded MM potentials between QM and MM atoms are included. All bonded MM potentials within the QM molecule are discarded because they are accounted for by the QM calculation.

When there are covalent bonds that cross the QM/MM boundary, the question of which MM potentials to accept and which to discard is not so easy to answer. Consider the system shown in Figure 1-2a, with one covalent bond that traverses the QM/MM boundary. Shown in Figure 1-2b is the equivalent QM model system with a capping hydrogen atom. In the IMOMM approach, MM potentials are only included if they depend on atoms that have no equivalent in the QM model system. Hence, any MM potential in which all atoms involved are QM atoms are NOT included in the total QM/MM potential, for instance the C_2-C_1 bond

stretching or the C_2-N_3-H_4 angle bending potentials. Furthermore, the C_5-N_3 bond stretching potential is also not included, because an equivalent in the QM model system exists, namely the N_3-H_{cap} bond. The QM potential is assumed to adequately model the link bond. The same is true for the C_2-N_3-C_5 bending potential. Again there is an equivalent in the QM model system that involves the capping hydrogen atom. The rule therefore also implies that any MM potentials in which only QM or LI atoms are involved, are NOT included in the hybrid QM/MM potential. On the other hand, all MM potentials that involve at least one or more MM atoms are included. For example, C_2-N_3-C_5-O_6 torsion potential is included because there is no equivalent in the QM model system and the O_6 atom is a pure MM atom.

There is only one exception. It involves the non-bonded interactions between QM atoms and LI atoms. From the rules above this MM potential should be discarded. However, in the IMOMM method this potential is included. The reasoning is that this interaction in the real system is not adequately modeled in the QM model system.
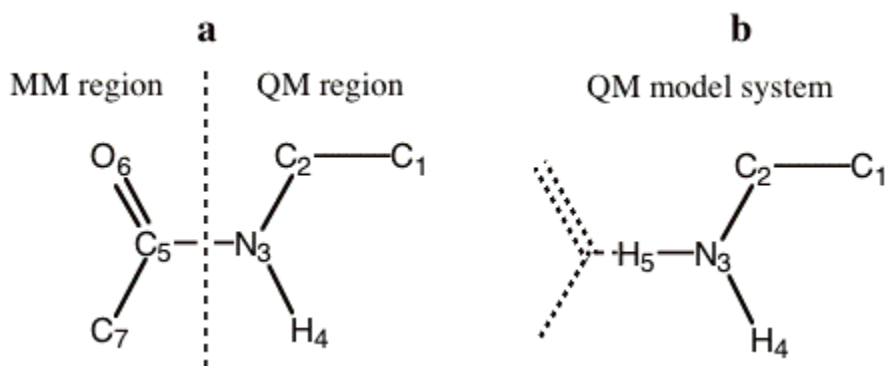
Figure 1-2 a) QM/MM partitioning. b) The equivalent QM model system. The numeric subscripts simply refer to the atom numbering.

In the AddRemove model [3], things are less complicated. The classical LI atom is treated as a normal MM atom with corresponding MM potentials to both the MM and QM atoms. The same goes for the MM correction potentials of the capping atoms (only with the real QM atoms and other capping atoms). For the AddRemove model, it's best to view the setup as the setup of two systems, one with all *real* QM and MM atoms (case a) and one with the real QM atoms and the capping atoms (case b). For both you need to build up the force field; in the former case (a), all interactions involving only QM atoms are ignored (these are already present through the QM calculation), while in the latter case (b) all interactions without contributions from the capping atoms are ignored. The interactions of case (a) are the normal MM/MM and QM/MM interactions, while the interactions of case (b) are used for correcting the QM interactions of the capping atoms. This also means that an atomic force field type should be assigned to the capping atoms, which is being handled in the LINK_BONDS block.

Nomenclature and Terminology

This section summarizes the naming conventions that are used throughout this document. Some of the terminology has already been described in the previous section. Since the nomenclature describing the link bonds can be somewhat confusing we recommend that special attention be given to this section.

Full QM/MM system vs. QM model system

When performing a combined QM/MM simulation, the molecular system is divided into QM and MM regions as shown in Figure 1-1a. We will refer to the total hybrid system as the 'full system' or sometimes we will refer to it as the 'real system'. The 'QM model system' is the capped system for

which the electronic structure calculation is performed. Figure 1-1a shows the full system and Figure 1-1b depicts the corresponding QM model system.

Link bonds and capping atoms

The 'link bonds' are those covalent bonds that cross the QM-MM boundary in the full QM/MM system. A link bond involves one atom that belongs to the MM region and one from the QM region, see Figure 1-1a. The 'capping atoms' (sometimes termed dummy atoms) refer to the atoms that are used to cap the valence in the model QM system. Capping atoms are not part of the full system. The 'capping bond' is the covalent bond in the QM model system that corresponds to the link bond in the real system. The terms capping atom or capping bond only refer to the model QM system, whereas the term LINK only refers to bonds or atoms in the real system.

QM, MM and LI atom types

In the ADF QM/MM input, each atom in the full system must be designated as a QM, MM or LI atom, where LI refers to link. Figure 1-1c shows these designations for the example system. Although two atoms are involved in a LINK bond, we only designate the atom in the MM region as the LI atom. We do so because this atom corresponds to a capping atom in the QM model system. In systems where there are no covalent bonds that cross the QM/MM boundary, there will be no LI atoms.

Atom types

There are several different meanings of the term atom and atom type that have arisen because of the hybrid nature of the QM/MM method, see the following items.

QM/MM type or QM/MM atom type (MM, QM or LI)

Refers to the partitioning of the full system into QM and MM regions. As described above, there are only three QM/MM types allowed: QM, MM, and LI. Each atom in the full system is assigned a QM/MM atom type in the MM_CONNECTION_TABLE subkey block.

Force Field atom type

The atom type used in the force field calculation. Each atom in the full system is assigned a force field atom type in the MM_CONNECTION_TABLE subkey block of the input. Force field atom types assigned to each atom must correspond to atom types defined in the Force field file. Force field atom types must be assigned to all atoms, even the QM atoms because the non-bonded interactions between QM and MM atoms are treated by a molecular mechanics potential.

ADF atoms or ADF fragments

The atom types used for the electronic structure calculation of the model system. These are the atoms or fragments defined in the FRAGMENTS key block in a standard ADF calculation. The ADF fragment types used for the capping atoms are defined in the LINK_BONDS subkey block. Note that capping atoms can only be single atom fragments, not compound fragments as allowed by ADF.

Partitioning into QM and MM regions

In a QM/MM simulation the basic question is how to partition the system into QM and MM regions. When studying an active site of a catalyst, for example, one must decide where to put the QM/MM boundary. Putting the boundary too close to the reaction center will question the chemical realism of the model. On the other hand, if one makes the boundary too far away, the computational expense of the QM calculation may become problematic. Each system is different in this respect and the user must make the proper tests to validate the appropriateness of the QM/MM partitioning used. We strongly suggest that the reader examines the literature on QM/MM methods and understands the basic limitations of the approach.

Below we give examples of QM/MM partitioning that should not or can not be used. For comparison, we also give some representative examples of QM/MM partitioning that the program does allow. In the examples, the region enclosed in the dotted polygon represents the QM region and the atoms labeled with 'LI' are the so-called Link atoms.

First, the QM/MM boundary should not cut across double, triple or aromatic bonds as shown in Figure 1-3. In these examples, a simple capping atom does not satisfy the valence of the QM fragment and the electronic structure of the QM model system would be drastically different from that of the 'real' system.
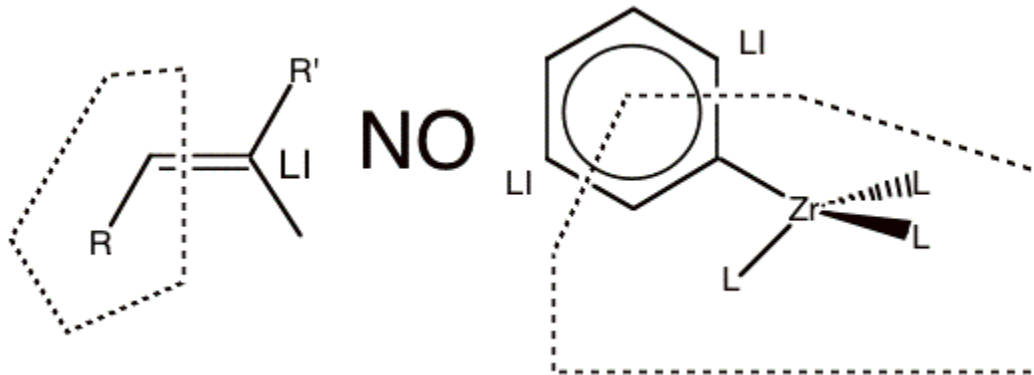
Figure 1-3 Examples of partitioning that should not be used because the link bonds are double or aromatic bonds.

Next, figure 1-4a depicts examples of partitioning that are not allowed because the LI atom has a covalent bond to more than one QM atom. A LI atom can only be bonded to one QM atom. Figure 1-4b shows the opposite, which is allowed. In other words, one QM atom can be bonded to more than one LI atom. This is due to the partitioning scheme that was used and the geometric relationship expressed in Equation 1.1, which restricts the position of the link atom, based on the QM and dummy atom. Note that there is no limit to the number of LI atoms or link bonds, just that each LI atom can only be bonded to one QM atom.

a NOT ALLOWED
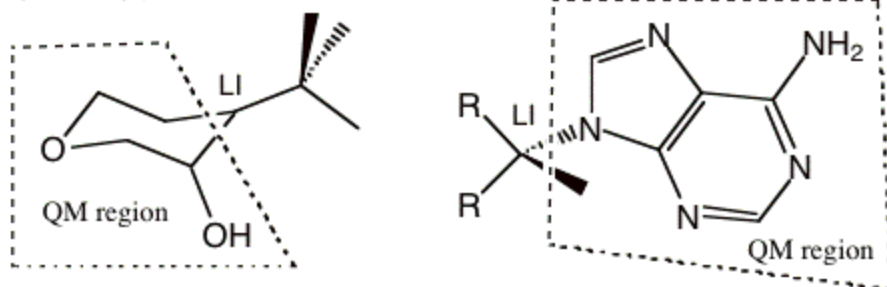
b ALLOWED
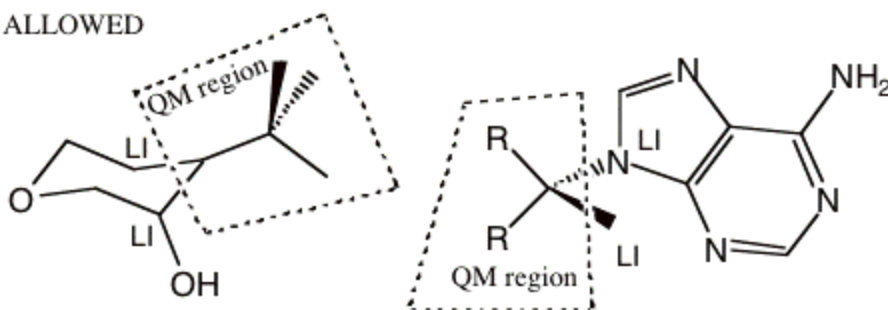
Figure 1-4 a) Examples of partitioning that are not allowed because the LI atom has a covalent bond to more than one QM atom. b) The allowed reverse of the examples showed in (a). A LI atom **can** only bond to one QM atom.

Then, figure 1-5 provides some representative examples of partitioning that the program does allow. Example **a** shows a typical solute-solvent QM/MM partitioning where there are no link bonds at all. Example **b** depicts two separate molecules each possessing a QM and a MM region. We emphasize that any number of molecules and link bonds can be used. Recall that in the IMOMM/ADF method all link bonds have a different a parameter associated with them, each specified by the user, upon which the energy depends (and which are difficult to choose or generalize). In example **b** there would be four independent a parameters. Example **c** seems very similar to the earlier example in Figure 1-3. The difference is that the ring in Figure 1-cd is not aromatic and consequently the link bonds in example **d** cross *single* bonds. Example **d** shows a single molecule, with two QM regions separated a MM region. For this example, two equivalent pedagogic representations of the sample partitioning are displayed. Example **e** is a complex organometallic system that we have tested the QM/MM approach on.
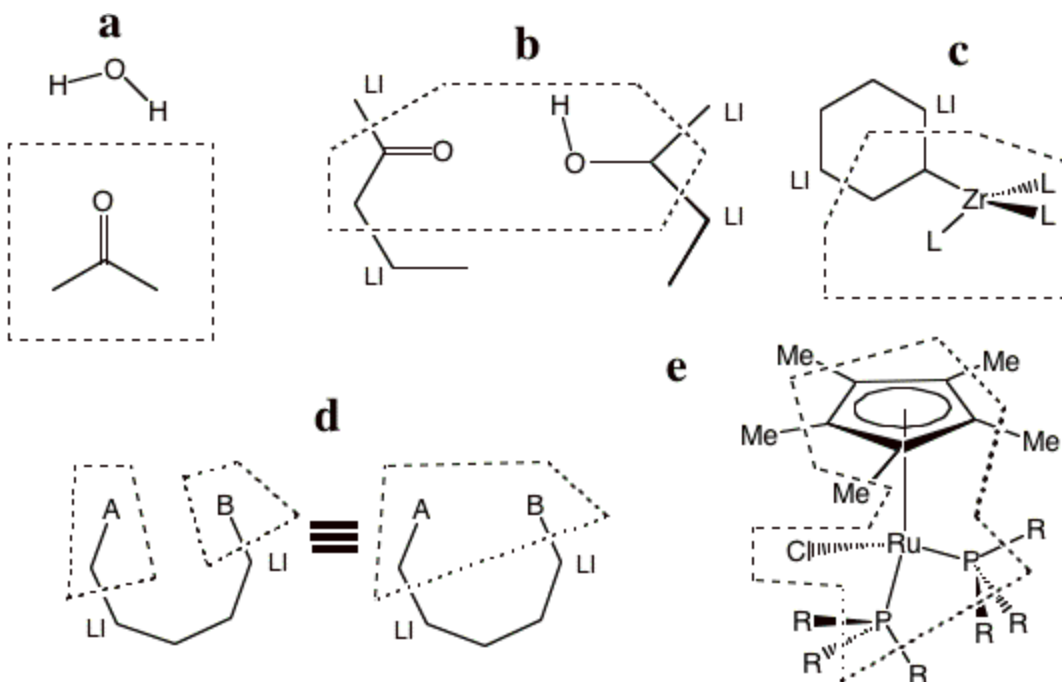
Figure 1-5 Representative examples of QM/MM partitioning that can be used in ADF QM/MM ADF.

1.3 Implementation

ADF QM/MM

The basic philosophy of the ADF QM/MM implementation has been to treat the MM region as a perturbation to the QM model system. The atoms of the QM model system are controlled by ADF as they would in a regular pure QM run, whereas the MM atoms are treated 'separately'. For example, when geometry optimization is performed, the Hessian matrix is only generated / updated for the QM model system. This is possible because at each geometry step the MM subsystem is fully optimized (with the QM and LI atoms frozen). One should be aware however that there are not necessarily six zero eigenvalues of the Hessian, as the QM system is coupled to the MM system (and therefore not free to rotate/translate !). When internal coordinates are used to define the structure of the complex, the atoms of the QM model system are optimized using the internal coordinate system, whereas the MM subsystem is fully optimized at each step within the Cartesian coordinate system. Figure 1-6 shows the flow control of the ADF QM/MM implementation.

The geometry optimization of the MM subsystem is controlled separately from the QM system. For example, the default convergence criteria for the MM subsystem are far stricter than that of the QM system. Furthermore, there are options to use global optimization techniques on the MM subsystem, see the detailed description of input options later in this manual.

standard ADF initialization

initialize MM system

⬇

new coordinates of QM model system

⬇

- build new MM cartesian coordinates starting from QM coordinates
- fully optimize the MM subsystem with the QM atoms frozen
- determine the MM energy contribution to the total energy and the MM forces acting on the QM atoms

⬇

optimize wave function of QM model system
(in presence of MM charge density-if applicable)
determine the DFT (QM) energy and gradients

⬇

add MM gradients to the QM gradients of the model system

⬇

use ADF optimization routines to optimize the geometry of the QM model system with the combined QM/MM gradients

geometry convergence checks performed by standard ADF routines

⬇

finish

Figure 1-6. Schematic representation of the ADF QM/MM implementation

Limitations, Bugs and Deficiencies

Listed here is a list of known bugs, deficiencies and limitations of the ADF QM/MM implementation. Please contact us if you find new entries to this list.

- The MOPAC input option introduced with ADF99 can not be used with the QM/MM option.
  The QM/MM feature of ADF requires the traditional ADF input styles.
  However, all options within the tradition ADF input format can be used, such as CART, ZMAT, and ZCART.
- Geometry constraints cannot be applied to MM atoms when using internal coordinates for optimization.
  Currently constraints can only applied to MM atoms when the optimization is performed in Cartesian coordinates.
- Linear transit coordinates cannot involve any MM atoms.
- See also the Limitations part at the beginning of this document.

1.4 GUI-support

Note that the graphical user interface ADF-GUI enables all users to set up complicated calculations with a few mouse clicks, and provides graphical representations of calculated data fields, see the GUI documentation.

2 ADF QM/MM

2.1 Input: The Essentials

Overview

Compared to a standard ADF run, there are two additional input components required to run an ADF QM/MM simulation:

- The QMMM key block has to occur in the ADF input file
- A separate force field parameter file.

In this section we describe how to set up an ADF QM/MM simulation, assuming an appropriate force field file is already available. A full description of all of the options and of the FORCE FIELD file is provided in later sections.

Example Input

Let Figure 2-1 show the molecular system that we want to simulate and the intended partitioning. Note that this is the same system as shown in Figure 1-1 except that a non-bonded water molecule has been added to the MM region.
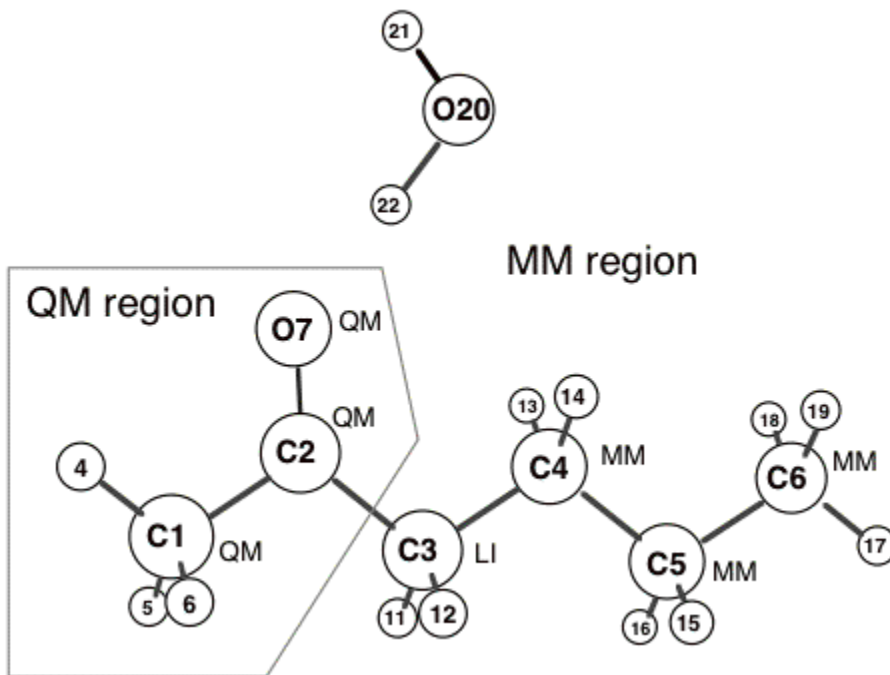
Figure 2-1. Sample structure. Atom numberings and the QM/MM atom types (QM, MM and LI) are shown.

Example 2.1 below shows the ADF input file for the simulation with the geometry defined in Cartesian coordinates. The initial atomic coordinates of the full QM/MM system are defined as they would be in a normal ADF run with the ATOMS and GEOVAR key blocks. Notice that the subkey blocks within the QMMM key block (e.g. MM_CONNECTION_TABLE) end with SUBEND. If they were closed by 'END' ADF would consider the QMMM key block complete at that point.

Example 2.1:

```
Title example QM/MM input
Fragments
 C C.t21
 H H.t21
 O O.t21
End

Symmetry NOSYM
Charge 0 0

ATOMS
 1 C 0.00000 0.00000 0.00000
 2 O 1.48700 0.00000 0.00000
 3 C -0.76430 1.32381 0.00000
 4 C -0.76428 -1.32378 -0.00002
 5 H -0.50028 -1.89680 -0.89230
 6 H -0.50028 -1.89683 0.89224
 7 H -1.83868 -1.12409 -0.00002
 8 H -1.40364 1.39261 -0.88358
 9 H -1.40364 1.39261 0.88358
 10 C 0.22461 2.48208 0.00000
 11 H 0.85235 2.42365 -0.89260
 12 H 0.85235 2.42365 0.89260
 13 C -0.53689 3.80103 0.00000
 14 H -1.16478 3.85787 -0.89260
 15 H -1.16478 3.85787 0.89260
 16 C 0.45222 4.95913 -0.00000
 17 H 1.08538 4.90988 -0.88932
 18 H 1.08538 4.90988 0.88932
 19 H -0.08590 5.91026 0.00000
 20 H 2.43700 1.64545 0.00000
 21 O 3.03926 2.50556 -0.00000
 22 H 3.96191 2.28678 -0.45094
END
```

QMMM
 FORCE_FIELD_FILE sybyl.ff
 RESTART_FILE mm.restart
 OUTPUT_LEVEL=1
 WARNING_LEVEL=1
 ELSTAT_COUPLING_MODEL=0
 LINK_BONDS
 1 - 3 1.38000 H
 SUBEND

 MM_CONNECTION_TABLE
 1 C_2 QM 2 3 4
 2 O_2 QM 1
 3 C_3 LI 1 8 9 10
 4 C_3 QM 1 5 6 7
 5 H QM 4
 6 H QM 4
 7 H QM 4
 8 H MM 3
 9 H MM 3

```
 10 C_3 MM 3 11 12 13
 11 H MM 10
 12 H MM 10
 13 C_3 MM 10 14 15 16
 14 H MM 13
 15 H MM 13
 16 C_3 MM 13 17 18 19
 17 H MM 16
 18 H MM 16
 19 H MM 16
 20 H MM 21
 21 O_3 MM 20 22
 22 H MM 21
 SUBEND
END

GEOMETRY
 ITERATIONS 20
 CONVERGE E=1.0E-3 GRAD=0.0005
 STEP RAD=0.3 ANGLE=5.0
 DIIS N=5 OK=0.1 CYC=3
END

XC
 LDA VWN
 GGA POSTSCF Becke Perdew
End

Integration 3.0

SCF
 Iterations 60
 Converge 1.0E-06 1.0E-6
 Mixing 0.20
End
End Input
```

Defining the Coordinates

The initial coordinates of the full QM/MM system are defined with the ATOMS and GEOVAR key blocks as in a normal ADF run. All input methods allowed by ADF can be used with the exception of the new MOPAC option introduced in ADF1999. For example the coordinates can be defined in Cartesian coordinates as in the example above, by a Z-matrix or by the mixed Z-CART method. It is important to realize that only the initial coordinates of the full or real system are required as input. There is no need to define the coordinates of the QM model system or of the dummy capping atoms. The partitioning of the system into QM and MM regions, and the parameters required to define the capping atoms are given in another part of the input. The program will automatically generate the QM model system and the position of the capping atoms.

IMPORTANT: There is a strict rule concerning the order of the atoms in the ADF input during a QM/MM simulation. Namely, **the QM atoms and the LI atoms must precede any MM atoms in the input**. The QM and LI atoms can come in any order as long as they come before any MM atoms. The program will check this and abort if this is violated.

ADF atoms and fragments

The typical ATOMS key block in ADF has the following format:

```
ATOMS
  {n} atom coordinates {f=fragment}
  ...
END
```

For the QM atoms, the atom labels and fragments should be defined as in a normal ADF run. The atom labels for the MM atoms are not read by the program. Instead, the MM force field atom types for all atoms are defined in the MM_CONNECTION_TABLE subkey block. Similarly if fragments are defined for the MM atoms, the program also ignores these. This is also true of the LI atoms, since in the calculation of the QM model system, capping atoms replace link atoms. The replacement atom type used for the electronic structure calculation is defined in the LINK_BONDS subkey block.

QMMM key block

The QMMM key block, which is in bold face in example 2.1, is a mandatory key block. This key block is necessary to invoke a combined QM/MM simulation. It contains the connection table and the force field atom types needed to define the molecular mechanics potential. If link bonds are present then it defines the necessary parameters for each link bond.

This section also explains how to specify the Force Field file to use and various other options.

`FORCE_FIELD_FILE`

> Keyword (required, default = amber95.ff)
> This keyword simply defines the full path of the force field file to be used for the molecular mechanics potential. The location of the force field file is given after the keyword. The full path can be given, or just the file name. In the latter case, the program looks in the current directory that ADF is executing in. Examples:

> ```
> FORCE_FIELD_FILE /home/username/sybyl.ff
> ```

> ```
> FORCE_FIELD_FILE sybyl.ff
> ```

`MM_CONNECTION_TABLE`

> Subkey block (required)
> This key block defines the connection table, the force field atom types and the partitioning of the full system into QM and MM regions. It is critical that the atoms specified in this key block are in the same order as in the ATOMS key block. This is important, because the program may not detect this type of input error and you would get ridiculous results.

> ```
> MM_CONNECTION_TABLE
>  n FF_LABEL MM_TYPE connection numbers
>  ...
> END
> ```

The labels are defined in the following table.

input column		
1	N	atom number
2	FF_LABEL	Force field atom type. These labels correspond to the atom types defined in the force field file. They can be up to four characters long. Xx defines dummy atoms.
3	MM_TYPE	QM, MM or LI
4-	connection numbers	These define to which atoms the current atom has a covalent bond. These connections are used to generate

		the molecular mechanics potential. Currently, a maximum of 6 connections is allowed per atom.

The connection table should be a fully redundant one. In other words, if atom #1 is bonded to atom #5, they each should have the other atom listed in their connections.
Example:

```
1 C_2 QM 2 3 4 5
2 O_2 QM 1
3 H QM 1
4 C_3 QM 1 5 6 7
5 Cu QM 4 1
6 H QM 4
```

A fully non-redundant connection table is also supported. In such a connection table, once a bond is mentioned, it is not mentioned again. In other words, the connection list for any atom cannot contain an atom, which precedes it in the atom numbering.
Example:

```
1 C_2 QM 2 3 4 5
2 O_2 QM
3 H QM
4 C_3 QM 5 6 7
5 Cu QM
6 H QM
```

These two connection tables are equivalent. Connection tables that are semi-redundant might cause problems. We recommend the fully redundant connection table.

LINK_BONDS

Required for systems with LINK bonds
This key block required for systems with covalent bonds that cross the QM/MM boundary. These bonds are referred to in this document as the link-bonds. Each link bond has a constant parameter '*alpha*' associated with it, which is defined as the ratio of the bond length in the real system and of the capping bond in the model QM system. See Section 1 or reference [1] more details. To determine the *alpha* parameters for each link bond, one can take the capping atom bond distance in a 'pure QM' calculation of the QM model system and ratio it to the corresponding bond distance in the real system. These ratios are typically around 1.30 to 1.50 when hydrogen as a capping atom. The LINK_BONDS subkey block has the following format:

```
LINK_BONDS
 atom_a - atom_b alpha replacement_fragment
[addremove_force_field_type]
 ....
SUBEND
```

Example:

```
LINK_BONDS
 15 - 3 1.42 H H1
 8 - 1 1.40 Cl.dzp Cl
SUBEND
```

The integers *atom_a* and *atom_b* refer to the numbering of the two atoms involved in the link bond. One of the atoms will be a LI type atom whereas the other will be a QM type atom. *Atom_a* and *atom_b* must be separated by " - " with at least one space between the integer and the hyphen. In other words '3 - 4' is OK, but not '3- 4' '3 -4'. Atoms need not be in any particular order, and the order of the link bonds is

19

also not important. Following this is the *alpha* parameter for that specific bond. The *replacement_fragment* is the ADF atom used for the capping atom in the electronic structure calculation of the QM model system. Often the capping atom is a hydrogen atom, however, it need not be. The *replacement_fragment* must be present in the FRAGMENT key block in the ADF input file. The *addremove_force_field_type* need only be present for the AddRemove model [3], and indicates the force field type of the capping atom (similar to FF_LABEL in the MM_CONNECTION_TABLE block). Important note: It is very important to realize that the Hamiltonian depends on the a parameters used. Thus, when comparing relative energies for example, one has to take care that the a's corresponding to the same bonds are identical.

Restarts

For the most part, restarts with ADF QM/MM are the same as in a standard ADF run. In other words to specify a restart, one needs to use the RESTART keyword. This signals the QM/MM extension to read the data from the QM/MM restart file as opposed to the input. It is important to emphasize that the **QM/MM extension has its own restart file**. Thus, the user will have to keep track to two restart files, the standard ADF restart file (TAPE21) and the QM/MM restart file. By default the QM/MM restart file is 'mm.restart', but you can change this by using the RESTART_FILE keyword **within the QMMM key block**. A truncated example of an ADF QM/MM restart input is given below:.

```
RESTART ADF_restart.file &   {same as in a standard ADF run}
 NOHESS
END

QMMM
 RESTART_FILE mm_restart.file {optional, with default filename of 'mm.restart'}
 FORCE_FIELD_FILE sybyl.ff
 OUTPUT_LEVEL=1
 WARNING_LEVEL=1
 ...
 ...
 ...
END
```

At the moment, when the RESTART keyword is used, the QM/MM extension will always look for the QM/MM restart file. There is no way to bypass this. However, the QM/MM restart file is simply a text file whose contents resemble those of the QMMM key block in the ADF QM/MM input and you can easily modify it.

It is also important to note that the RESTART_FILE both specifies the file to be read at the beginning of a restarted run, and the file that will be written to. In an ADF QM/MM run that was restarted, the initial QM/MM restart file is overwritten.

Geometry Constraints and Fixing Coordinates of MM atoms

Constraints can be applied to coordinates in the same way they are done with a standard ADF run (i.e. most commonly through the GEOVAR key block). There are no limitations to applying geometry constraints to QM or LI atoms, however, there exist some limitations to applying geometry constraints to MM atoms, particularly when using a Z-matrix.

When a geometry optimization is performed in Cartesian coordinates there is almost no limitation to applying constraints to MM atoms. One important exception is that linear transit constraints can not involve MM atoms. At the moment constraints cannot be applied to the coordinates that define the position of the MM atoms when internal (ZMAT) coordinates are used. Note that constraints still can be applied to the coordinates of the QM atoms, but just not to the MM atoms.

Symmetry constraints can not be applied with a QM/MM run, unless all of the MM atoms are frozen. Furthermore, the program will not check that the atoms of the MM region satisfy the imposed symmetry.

Miscellaneous Notes

Using ADF Dummy Atoms

It is often necessary to use dummy atoms when defining a Z-matrix for a geometry optimization. (Here we are **not** referring to the capping dummy atoms) Dummy atoms are allowed in a QM/MM simulation and they can be used just as they would in a normal ADF run. There are a few things to remember when using dummy atoms in a QM/MM run.

If a dummy atom is to be used to define the coordinates of QM atoms in a Z-matrix for geometry optimization, the dummy atom must precede any MM atoms in the atom list. In this case one should consider the dummy atom as part of the model QM system.

Atom should be given the Xx force field atom type in the MM_CONNECTION_TABLE key block. (In this way the atom is excluded from the non-bonded pair list)

Do not make any bonds to the dummy atom in the connection table; otherwise the program may require the user to define molecular mechanics potentials involving the dummy atom.

Dummy atoms are allowed in the MM region, however, they will only be used to define the initial coordinates. This is because the MM subsystem is always optimized in Cartesian space where dummy atoms are not necessary.

One can consider dummy atoms as part of the model QM system. For example, the following is not allowed because the dummy atom, which is assigned to the QM region, comes after atoms 5 and 6, which are MM, atoms.

```
MM_CONNECTION_TABLE
 1 C_2 QM 2 3 4
 2 O_2 QM 1
 3 C_3 LI 1 8 9 10
 4 C_3 QM 1 5 6 12
 5 C_3 MM 4
 6 C_3 MM 4>
 7 Xx QM
 8 H MM 3
 ...
SUBEND
```

The example below is almost identical to the above example except that the dummy atom is assigned as a MM atom. Although this is allowed, this dummy atom will be optimized as a MM atom.

```
MM_CONNECTION_TABLE
 1 C_2 QM 2 3 4
 2 O_2 QM 1
 3 C_3 LI 1 8 9 10
 4 C_3 QM 1 5 6 12
 5 C_3 MM 4
 6 C_3 MM 4
 7 Xx MM
 8 H MM 3
 ...
SUBEND
```

The final example below, is probably what is wanted. Here the dummy atom is a QM atom and comes before any MM atoms. In this way, the dummy atom can be used to define the Z-matrix with the QM model system.

```
MM_CONNECTION_TABLE
 1 C_2 QM 2 3 4
 2 O_2 QM 1
 3 C_3 LI 1 8 9 10
 4 C_3 QM 1 5 6 12
 5 Xx QM
 6 C_3 MM 4
 7 C_3 MM 4
 8 H MM 3
 ...
SUBEND
```

Linear Transit Runs.

The QM/MM option can be used with the LINEAR TRANSIT option in ADF. However, the linear transit parameter or coordinate cannot involve any MM atoms. In other words, the linear transit parameter can only be defined in terms of QM atoms only.

Frequency Runs with QM/MM

At the moment, a FREQUENCY run can not be performed when using a QM/MM potential in ADF. This option has been implemented but has been disabled in this release for further testing.

2.2 QMMM keyblock options

Introduction

Chapter 2 explains setting up an ADF QM/MM simulation. This section describes the available options that you can define in the QMMM key block. This section is essentially a reference source. The main components of the QMMM key block, the MM_CONNECTION_TABLE, the FORCE_FIELD_FILE, and the LINK_BONDS key blocks that are described in detail in the previous chapter are repeated here (with some additional notes). Please note that the global optimization options are not well tested and are prone to crashing the run. If one is interested in using these options, please be aware of this fact. We appreciate reports of any failures.

Example Input

In this section we simply provide a few examples of the QM/MM key block. In some examples, the MM_CONNECTION_TABLE and LINK_BONDS subkey blocks are not filled.

Example 3.1 This example depicts a global optimization of the MM region with the simulated annealing-like optimizer available in the QM/MM program. The global search involves 100 ps of MD at 1000 K with 100 structures sampled in regular intervals during the simulation. Each of the 100 structures is then partially optimized, and then the 10 best are fully optimized. At the end of this, the lowest energy structure is used for the QM/MM run. Note that in this MD search, the QM atoms, including the link atoms are frozen.

```
QMMM
  FORCE_FIELD_FILE sybyl.ff
  OUTPUT_LEVEL 1
```

```
      WARNING_LEVEL 1
      ELSTAT_COUPLING_MODEL 1
      OPTIMIZE
        GLOBAL
          METHOD MD_SEARCH
          FREQUENCY ONCE
        SUBEND
      SUBEND
      MD_SEARCH
        TIME{PS} 100.0
        N_STRUCTURES 100
        TEMPERATURE 1000.0
      SUBEND
      MM_CONNECTION_TABLE
        ...
      SUBEND

      LINK_BONDS
        ...
      SUBEND
    END
```

Example 3.2 In this example, custom charges are assigned to some of the atoms. Charges for atoms that were not given specific charges in the QMMM key block are assigned on a per atom-type basis from the force field file. Also note that this example has no LINK_BONDS subkey block. This is only allowed if there are not link bonds, as in the example in Figure 1-2a.

```
QMMM
 FORCE_FIELD_FILE sybyl.ff
 OUTPUT_LEVEL 1
 WARNING_LEVEL 1
 ELSTAT_COUPLING_MODEL 1
 MM_CONNECTION_TABLE
   ...
 SUBEND

 CHARGES
   0.4
   0.3
   3 -0.1
 SUBEND
END
```

Description of Options

FORCE_FIELD_FILE

> Keyword (required, default = amber95.ff)
> This keyword simply defines the full path of the force field file to be used for the molecular mechanics potential. The location of the force field file is given after the keyword. The full path can be give, or just the file name. In the latter case, the program checks the current directory that ADF is executing in.
> Examples:

```
FORCE_FIELD_FILE /home/username/sybyl.ff
```

```
FORCE_FIELD_FILE sybyl.ff
```

```
NEWQMMM
```

Keyword (Default is not to use the subkey NEWQMMM)

Key to be used for more efficient QM/MM calculations, work in progress. It also allows more QM/MM atoms than in a default calculation. **This key should be used ONLY with the amber force field.** This key also offers the possibility to use the new QM/MM input format, which can, for example, be made with the utility *pdb2adf*. The old input format remains working if one includes this NEWQMMM subkey.

```
OUTPUT_LEVEL
```

Keyword (Default = 1)

The integer following this keyword specifies the amount of output to be printed to the ADF output file.
0: minimal output
1: normal output
2: trouble shooting output
OUTPUT_LEVEL 2 is recommending for initially setting up a job. However, once the job is set up properly this output level is probably too verbose.

```
WARNING_LEVEL
```

Keyword (Default = 1)

The performs some checking of the input, ranging from examining all interatomic distances to examining the input order of the QM, LI and MM atoms. The integer following this keyword specifies how many warnings to report and when to stop the run due to the warning.
-1: Report only the most severe warnings, and never stop the run. Useful when user is knowingly violating the 'rules'.
0: Report severe warnings and only stop at 'fatal' errors.
1: Report all warnings, stop at severe and fatal errors. This is the default.
2: Report all warnings and stop at any of them. Useful when initially setting up a job.

```
MDC_LEVEL
```

Keyword (Default = 2)

The integer following this keyword specifies the level of the Multipole Derived Charge analysis [7] used in conjunction with ELSTAT_COUPLING_MODEL=4. Note that the default value has been changed compared to ADF2007.01
1: MDC-m charges are used to update charges of QM system.
2: MDC-d charges used.
3: MDC-q charges used.

```
MM_CONNECTION_TABLE
```

Subkey block (required)

This key block defines the connection table, the force field atom types and the partitioning of the full system into QM and MM regions. It is critical that the atoms specified in this key block are in the same order as in the ATOMS key block. This is important because it is difficult for the program to detect this type of input error.

```
MM_CONNECTION_TABLE
 n FF_LABEL MM_TYPE connection numbers
 ...
SUBEND
```

The labels are defined in the following table.

input column		
1	n	atom number

2	FF_LABEL	Force field atom type. These labels correspond to the atom types defined in the force field file. They can be up to four characters long. Xx defines dummy atoms.
3	MM_TYPE	QM, MM or LI
4-	connection numbers	These define the atoms to which the current atom has a covalent bond. These connections are used to generate the molecular mechanics potential. Currently, a maximum of 6 connections is allowed per atom.

The connection table should be a fully redundant one. In other words, if atom #1 is bonded to atom #5, they each should have the other atom listed in their connections.
Example:

```
1 C_2 QM 2 3 4 5
2 O_2 QM 1
3 H QM 1
4 C_3 QM 1 5 6 7
5 Cu QM 4 1
6 H QM 4
```

A fully non-redundant connection table is also supported. In such a connection table, once a bond is mentioned, it is not mentioned again. In other words, the connection list for any atom cannot contain an atom that precedes it in the atom numbering.
Example:

```
1 C_2 QM 2 3 4 5
2 O_2 QM
3 H QM
4 C_3 QM 5 6 7
5 Cu QM
6 H QM
```

These two connection tables are equivalent. Connection tables that are semi-redundant might cause problems. We suggest using the fully redundant connection table.

LINK_BONDS

Required for systems with LINK bonds
This key block required for systems with covalent bonds that cross the QM/MM boundary. These bonds are referred to in this document as the link-bonds. Each link bond has a constant parameter '*alpha*' associated with it, which is defined as the ratio of the bond length in the real system and of the capping bond in the model QM system. See Section 1 and the Appendix for more details. To determine the *alpha* parameters for each link bond, one can take the capping atom bond distance in a 'pure QM' calculation of the QM model system and ratio it to the corresponding bond distance in the real system. These ratios are typically around 1.30 to 1.50 when hydrogen as a capping atom. The LINK_BONDS subkey block has the following format:

```
LINK_BONDS
 atom_a - atom_b alpha replacement_fragment addremove_force_field_type
 ....
SUBEND
```

Example:

```
LINK_BONDS
  15 - 3 1.42 H H1
  8 - 1 1.40 Cl.dzp Cl
```

```
      SUBEND
```

The integers *atom_a* and *atom_b* refer to the numbering of the two atoms involved in the link bond. One of the atoms will be a LI type atom whereas the other will be a QM type atom. *Atom_a* and *atom_b* must be separated by " - " with at least one space between the integer and the hyphen. In other words '3 - 4' is correct, but not '3- 4' '3 -4'. Atoms need not be in any particular order, and the order of the link bonds is also not important. Following this is the *alpha* parameter for that specific bond. The *replacement_fragment* is the ADF atom used for the capping atom. Often the capping atom is a hydrogen atom, however, it need not be. The *replacement_fragment* must be present in the FRAGMENT key block in the ADF input file. The *addremove_force_field_type* need only be present for the AddRemove model [3], and indicates the force field type of the capping atom (similar to FF_LABEL in the MM_CONNECTION_TABLE block).

Important note: It is very important to realize that the Hamiltonian depends on the a parameters used. When comparing relative energies for example, one needs to take care that the a's corresponding to the same bonds are identical.

CHARGES

Key block (optional)

This key block defines the initial charges on each atom based on their atom **number**. Atom numbers must be carefully specified because the program does not assume any order. Charges can also be assigned based on their atom **type** from the force field file in the CHARGE PARAMETERS key block. If this key block does not specify charges, the program looks for charge assignments from the force field file. If charges are not assigned in either this key block or the force field file, then a charge of 0.0 is assigned. When polarizable electrostatic coupling is invoked, the charges for QM and LI atoms are not read, because the MM point charges interact with the QM charge distribution. With ELSTAT_COUPLING_MODEL=4, these charges (for the QM system) are used only in the first cycle of the geometry optimization: after each cycle, the QM charges are replaced with the Multipole Derived Charges.

Example:

```
      CHARGES
       3 -0.10000
       1 0.05000
       2 0.05000
      SUBEND
```

ELSTAT_COUPLING_MODEL

Keyword (optional, default=1)

This keyword controls the type of electrostatic model that is used, including whether true electrostatic coupling between the QM and MM regions is evoked.

Coupling Model	Description
0	Electrostatics OFF
1	Simple electrostatic coupling, where there is no polarization of the QM wave function, i.e. pure MM coupling.
2	Electrodynamic coupling, where the point charges in the MM region can polarize the wave function. **This option is not currently functional.**
4	Simple electrostatic coupling, like option 1. But the point charges of the QM region are updated throughout the geometry optimization using the Multipole Derived Charge analysis [7] (MDC-x level depending on MDC_LEVEL).

```
OPTIMIZE
```

Key block (optional)
This key block allows the user to modify the geometry optimization settings. An example of a key block with many available options is shown below:

```
OPTIMIZE
 MAX_STEPS 1000
 MAX_GRAD 0.001
 PRINT_CYCLES 20
 METHOD BFGS

 GLOBAL
 METHOD GRID
 FREQUENCY ONCE
 SUBEND

 GRID
 INCREMENT 20.0
 BOND 2 - 4
 BOND 2 - 3
  SUBEND
 SUBEND
```

Sub-options to this key are described next.

```
OPTIMIZE: MAX_STEPS
```

Keyword (optional, default = 1000)
This keyword defines the maximum number of optimization steps allowed before the optimization is discontinued.

```
OPTIMIZE: MAX_GRADIENT
```

Keyword (optional, default = (0.01 kcal/mol)/Angstrom)
This keyword allows the user to change the convergence criteria. For now, the optimization is considered converged when the maximum gradient on any MM atom is less than MAX_GRADIENT. The default value will provide gradients that are very small, especially when compared to the convergence criteria specified most electronic structure codes. NOTE: The gradients on the QM atoms due to the MM potentials are not accounted for in the convergence criteria. Lrge MM forces can exist on the QM atoms after the optimization.

```
OPTIMIZE: ENERCVG
```

Keyword (optional, default = 0.001 kcal/mol)
This keyword allows the user to change the convergence criteria for the energy between successive cycles. Can only be used in case of NEWQMMM.

```
OPTIMIZE: METHOD
```

Keyword (optional, default = BFGS, available: BFGS, STEEPEST_DESCENT, CONJGRAD)
For the most part, the default quasi-Newton optimizer with BFGS Hessian update scheme is very stable, and converges well. Other optimizers available are the steepest descent method (STEEPEST_DESCENT) and conjugate gradient (CONJGRAD). STEEPEST_DESCENT and CONJGRAD are almost always less efficient than the BFGS optimizer (particularly close to the minimum). It is notable that the Hessian based BFGS method requires more memory than the STEEPEST_DESCENT method and so for very large systems may be problematic to use. In that case, it is best to use the CONJGRAD method.

27

```
OPTIMIZE: MM_NOTCONVERGED
```

Keyword (optional, default = 1)
This keyword defines what should happen if the MM geometry is not fully optimized after
MAX_STEPS steps; set this to zero for large (biochemical) systems where it may be problematic to
get the optimization to converge fully in a limited number of steps (1000): the QMMM run will
continue as if the MM-optimization had converged.

```
OPTIMIZE: FIX_MM_GEOMETRY
```

Keyword (optional, default = .false.)
If this keyword is specified in the OPTIMIZE subblock, the MM system will be frozen, i.e. no
geometry optimization will be done on any of its atoms. If NEWQMMM is included use
'IRUNTYPE_QMMM 0' as separate keyword (outside the OPTMIZE subblock).

```
OPTIMIZE: PRINT_CYCLES
```

Keyword (optional, default = 100)
PRINT_CYCLES represents the number of optimization cycles between which the optimization
status is printed and the MM restart file is written.

```
OPTIMIZE: GLOBAL
```

Sub key block (optional) - CURRENTLY IN A BETA STATE
This subkey block controls the global optimization options in the program. Currently the global
optimization option has not been thoroughly tested and should be considered to be in beta form.
The normal optimizers are designed only to locate the "nearest" local minimum and therefore you
are not guaranteed to find the best overall structure, which is termed the global minimum structure.
Currently, only two global optimization algorithms have been implemented:
- Molecular dynamics based optimizer related to a simulated annealing algorithm
- A grid search, which generates conformations by rotations about bonds specified by the user.
Both optimizers generate a number of structures (100s to 1000s), which are all partially optimized.
The partially optimized structures are then sorted based on their energies. The best 10 of these
structures are then fully optimized. The best of these fully optimized structures is kept and assumed
to be the global minimum. The global optimization is only applied to the MM region with the QM
atoms frozen. Therefore, the structure can only be considered the global minimum structure on the
constrained surface where the QM atoms and QM charge density are frozen.

```
   OPTIMIZE
     GLOBAL
       METHOD MD_SEARCH
       FREQUENCY ONCE
     SUBEND
     MD_SEARCH
       TIME{PS} 100.0
       N_STRUCTURES 100
       TEMPERATURE 1000.0
     SUBEND
   SUBEND
```

Global optimization is not the default. Therefore, to invoke a global optimization, the GLOBAL
subkey block must exist. It is important to note that the subkey blocks that control the global
optimization schemes are subkey blocks of the OPTIMIZE key block and not sub-sub key blocks
within the GLOBAL subkey block. The above example demonstrates this.

```
OPTIMIZE: GLOBAL: METHOD
```

Keyword (optional, default = MD_SEARCH)
This key block specifies the global optimization method to be used. To date there are only two methods, MD_SEARCH which is the default and GRID. More detail one how these methods work is given in the description of the MD_SEARCH and GRID subkey blocks.

OPTIMIZE: GLOBAL: FREQUENCY

Keyword (optional, default = ONCE)
This key block specifies how often the global optimization algorithm, if it is specified, is called. Since the global optimization is very time consuming it is not recommended that it be used every QM iteration. The default is that it is done only on the first iteration. The options available are tabulated below.

ONCE	Only at the first iteration
EVERY_TIME	At every iteration
N_CYCLES X	At each X-th iteration including the first. Here X is the integer following the "N_CYLES" keyword. e.g. N_CYCLES 4

OPTIMIZE: MD_SEARCH

Subkey block (optional, default settings specified below)
The MD_SEARCH method involves performing molecular dynamics on the MM subsystem at a high temperature. The high temperature dynamics allows the MM subsystem to "get out of" the local minimum of the initial structure and explore other regions phase space, potentially leading to lower energy structures. During the molecular dynamics, structures are sampled at specified intervals and stored. When the dynamics is complete, the stored structures are optimized and sorted in terms of their energy. This procedure is similar to simulated annealing, except that the temperature of the dynamics is not ramped up and down in a cyclic fashion. At the beginning, the dynamics is immediately pulsed up to the specified temperature with a random excitation on each of the free MM degrees of freedom. An example of the key block, with good settings is given below.

```
MD_SEARCH
  TIME{PS} 100.0
  N_STRUCTURES 100
  TEMPERATURE 700.0
SUBEND
```

In the above example, the MM subsystems are heated up to a temperature of 700 Kelvin. Dynamics is run for a total of 100 picoseconds, with a total of 101 structures sampled (100 plus the initial structure). Each structure is sampled every 1.0 picoseconds. The default timestep is 0.5 femtoseconds, and therefore in the above example 200,000 timesteps will be performed.
This global search technique is the most general and robust of the two methods implemented. It is therefore the default global optimization method. This subkey block is optional, since the default settings should work reasonably with most systems.

OPTIMIZE: GRID

Subkey block (optional, required if method selected)
The GRID method provides a systematic search for global minimum by rotating about specified covalent bonds in the MM subsystem. This method is only efficient for small systems or systems where the conformational variability is confined to torsions involving a few bonds. The user must specify the bonds that are to be rotated in the search, up to a maximum of 10, and the increment (in degrees) by which the bonds are to be rotated between subsequent structures. The program does not allow bonds that are
completely within the QM subsystem (link bonds are allowed, however, or
part of a ring system.
Finally, since QM atoms cannot be rotated, at least one of the two fragments resulting from splitting

the specified bond must contain no QM atoms. An example of the key block is shown below where three bonds are rotated, in 60° increments. 216 structures (6x6x6) will be generated corresponding to a full 360° rotation about the three bonds in 60° increments and all combinations thereof.

```
GRID
 INCREMENT 60.0
 BOND 7 - 6
 BOND 8 - 7
 BOND 9 - 8
SUBEND
```

NOTE: It is important to realize that the program uses the connection table specified in the input to determine which atoms to rotate.

MASSES

Key block (optional)
This is used to assign custom masses to individual atoms. If no custom masses are specified, then the default masses defined in the force field file are used. Below is example input.

```
MASSES
 15 32.066
 8 2.0
SUBEND
```

The first column is an integer specifying the atom number and the second column is a real specifying the custom mass of that atom in atomic mass units. The atoms need not be in any particular order and it is not necessary to specify custom masses for all atoms. It should be noted that only masses of the MM atoms and the link atoms could be customized. Masses of the QM atoms and the capping atoms are taken from the QM code.

2.3 The Force Field File

This required file must contain the force field parameters and the MM potential for each kind of MM interaction. Although predefined force field files (AMBER and SYBYL) are provided with ADF, these force field files can be customized. For example, one may want to change a particular force constant, or one may need to introduce a new atom type, for instance a transition metal. This section provides a detailed description of the force field file.

General Notes

Format

The force field file is keyword driven with each key block defining parameters for each molecular mechanics interaction type such as bond types, angle types, torsion types, ...etc. The key block begins with the keyword, such as "BONDS".

The lines that actually contain the parameters are sandwiched between two lines that contain "========". The lines between the keyword and the first line containing "========" are not read by the program. These lines are intended for the user to define the columns as shown below. There can be as many lines between the keyword and the first '=====' as needed.

Example:

30

```
BONDS
Atoms   pot   K                 ro      Notes
i - j         type  (kcal/molA^2)   (Ang)
==============================
...
CA  CA  1     938.0             1.400   amber95
CT  CT  1     620.0             1.526   amber95
...
==============================
```

Force Field Atom types

The force field atom types are the labels given to each atom in the real system that is used to generate the molecular mechanics portion of the QM/MM potential. These are separate from the atomic fragment types used by ADF for the electronic structure calculation. It is important to realize that QM atoms will have both an atomic fragment type and a force field atom type.

There are some limitations to the force field label types that the user can specify:

- Labels can be a maximum of four characters long, with no spaces.
- The atom types are case sensitive.
- They can contain letters, numbers and other characters except ',' '.' or '=' and tabs.

Example of atom types that are not compatible with the program: C.3, C 3, C=3, C_sp3, *

Examples of atom types that are correct: C_3, C3, Csp3, and C*

Wild Cards

Wild cards can be specified with the asterisk, '*'. Wild cards can be specified for angles, torsions and out-of-plane bends. Please refer to the specific sections for the limitations.

CAUTION: When using wild cards, place the wild cards at the beginning of the data section, beginning with the parameters with the most wild cards and ending with those that posses the least wild cards.

Example:

```
*     C_3   *     100.310   111.000   two wild cards
*     C_3   C_3   100.310   111.000   one wild card
C_2   C_3   C_3   100.310   111.000   no wild cards
```

If this ordering is not followed, then the wild card parameters will over-ride the specific parameters.

Dummy Atoms

If the QM program uses dummy atoms, they are specified 'Xx' or 'XX'. The program will automatically remove dummy atoms from the non-bonded pair list. Please note that if the user specifics bonds to the dummy atom in the connection table, the program will look for the appropriate parameters in the force field file. The program does not filter out this possibility because sometimes it is useful to specify MM bonds to dummy atoms).

Miscellaneous Notes

- Do not remove the '======' separator lines.
- Units are in kcal/mol, Angstroms, degrees, amu unless otherwise specified
- Sections can be in any order; i.e. BENDS can come before BONDS.
- All keywords are case sensitive and most are in ALL-CAPS
- Input is all free format
- Blank lines will be ignored
- Comment lines can be added to parameter data sections by beginning the line with the '#' symbol.

Example:

```
H H 1.0080
HC H 1.0080
  # example of comment line denoted with # mark.
H1 H 1.0080
H2 H 1.0080
```

4.3 A (partial) Example File

In this section we provide an example force field file to illustrate the format of the file. Only a limited number of parameters are included. A detailed description of each section of the force field file is provided in the next section.

```
FORCE_FIELD_SETTINGS
==============================
ELSTAT_1-4_SCALE          1.0000
VDW_1-4_SCALE             1.0000
VDW_DEFAULT_POTENTIAL     1       (1:6-12 2:exp-6 3:exp purely repulsive)
DIELECTRIC_CONSTANT       1.000
==============================


MASSES & ATOM LABELS
-----------------------------------
force_field  atomic
atom_type    symbol  mass      NOTES
========================
C_3          C       12.0110   sp3 hybridized carbon
C_2          C       12.0110   sp2 hybridized carbon
C_1          C       12.0110   sp1 hybridized carbon
C_ar         C       12.0110   aromatic
N_3          N       14.0070
N_2          N       14.0070
O_3          O       15.9990
========================
```

BONDS Ebond = 0.5*K(r-ro)**2

```
---------------------------------------
Atoms        pot
i - j        type   K         R          NOTES
==============================
C_2   C_2    1      1340.00   1.335      WHITE_77
C_2   C_3    1      639.00    1.501      WHITE_75
C_3   C_3    1      633.60    1.540      *
C_3   N_2    1      760.20    1.440      *
==============================
```
BENDS Ebend = 0.5*k(a-ao)^2

```
-------------------------------------
Atoms                    pot
i - j - k                type    K          theta      NOTES
==========================================
*       C_2     *        1       78.79      120.00     WHITE_77
*       C_3     *        1       65.66      109.50     WHITE_77
*       C_ar    *        1       78.79      120.00     *
C_ar    C_2     N_2      1       131.31     120.00     *
C_3     C_3     C_ar     1       78.79      109.50     *
==========================================
```

TORSIONS

```
------------------------------------------
Atoms                    pot
i - j - k - l            type    k          per        NOTES
=================================================
*     C_2    C_2    *     2      12.5000    -2.0
*     C_1    C_3    *     2      0.0000      1.0
C_2   C_2    C_3    *     2      0.1260     -3.0
C_3   C_2    C_3    *     2      0.1260      3.0
H     C_2    C_3    *     2      0.2740      3.0
*     C_ar   C_ar   C_ar  2      2.3500     -2.0
*     C_2    C_3    C_2   2      0.1260      3.0
*     C_2    C_3    C_3   2      0.1260      3.0
C_3   C_3    C_3    C_3   0      0.5000      3.0              no torsion potential
C_2   C_2    C_3    C_2   2      0.1260     -3.0
C_3   C_3    N_2    C_2   1      0.5000      4      180.0  This and the next 3 lines
&                                0.1500      3      180.0  are part of a multi-component
&                                0.5300      1      0.0    Fourier potential
C_3   C_3    C_2    N_2   1      0.1000      4      0.0
&                                0.0700      2      0.0    '&' is a continuation marker
=================================================
```

OUT-OF-PLANE

```
--------------------------
Atoms                    pot
i - j - k - l            type    K          NOTES
===================================
*     *      C_2    *     2      480        TRIPOS_85
*     *      N_2    *     2      120        TRIPOS_85
H     H      N_2    C_3   2      120        TRIPOS_85
C_3   H      N_2    *     2      120        TRIPOS_85
===================================
```

VAN DER WAALS
```
atom(s)Emin      Rmin      gamma    NOTES
============================
C_3     0.1070   3.4000    12.00
C_2     0.1070   3.4000    12.00
C_ca    0.1070   3.4000    12.00
C_ar    0.1070   3.4000    12.00
C_1     0.1070   3.4000    12.00
N_3     0.0950   3.1000    12.00
N_2     0.0950   3.1000    12.00
N_2 -
N_2 2   0.0950   3.1000    12.00    purely repulsive potential for this pair
============================
```

```
type        charge(e)  NOTES
================
OW          -0.82      TIP3P water model
HW          0.41       TIP3P water model
================
```

Section by Section Description

FORCE_FIELD_SETTINGS

Key block (required)
This key block specifies various global options for the force field file, mostly concerned with the treatment of the non-bonded potentials.

```
FORCE_FIELD_SETTINGS
===============================
ELSTAT_1-4_SCALE            0.5
VDW_1-4_SCALE               0.5
VDW_DEFAULT_POTENTIAL       1        (1:6-12 2:exp-6 3:exp purely repulsive)
DIELECTRIC_CONSTANT         1.000
===============================
```

ELSTAT_1-4_SCALE & VDW_1-4_SCALE

Most force fields scale the non-bonded interactions by a factor of 0.5 if the atoms are the terminal atoms of a defined torsion. This scaling factor, which is termed the 1-4 scaling factor, can also be different for the electrostatic potential and for the Van der Waals potentials and thus they are separately defined in the input.

VDW_DEFAULT_POTENTIAL

This keyword defines what kind of potential is used for the non-bonded van der Waals interactions. The potential types have been assigned integer values as defined in the following table.

VDW potential type		constants required (in order)
0	no potential	none
1	Lennard-Jones 12-6 $$E_{vdw}(R) = D_o\left[\left(\frac{R_o}{R}\right)^{12} - 2\left(\frac{R_o}{R}\right)^6\right]$$	D_o, R_o
2	Exponential-6 or Buckingham $$E_{vdw}(R) = D_o\left\{\left[\left(\frac{6}{\zeta-6}\right)e^{\zeta\left(1-\frac{R}{R_o}\right)}\right] - \left[\left(\frac{\zeta}{\zeta-6}\right)\left(\frac{R_o}{R}\right)^6\right]\right\}$$	D_o, R_o, ζ ζ=12.0 is standard
3	$$E_{vdw}(R) = D_o\left(\frac{6}{\zeta-6}\right)e^{\zeta\left(1-\frac{R}{R_o}\right)}$$ Purely Repulsive	D_o, R_o, ζ
4	Purely Attractive (dispersion term) $$E_{vdw}(R) = -2D_o\left(\frac{R_o}{R}\right)^6$$	D_o, R_o

DIELECTRIC_CONSTANT

Default = 1.00
This defines the dielectric constant used for the calculation of the electrostatic interactions. For example, 1.00 = vacuum and 80 is that of bulk liquid water. Currently, only a constant dielectric has been implemented.

BONDS

Key block (required)
This key block specifies the potential type and parameters for each kind of MM bond stretching interaction. An example is given below.

```
BONDS
Atoms          pot    K                ro        NOTES
i - j          type   (kcal/molA^2)    (Ang)
========================================
CA    CA   1      938.0            1.400     amber95
CT    CT   1      620.0            1.526     amber95
HC    Zr   0      0.0                        no potential found
========================================
```

The first two columns are the atom types (up to four characters long) and the third column is an integer specifying the potential type.

BOND potential type		constants required (in order)
0	no potential	none
1	simple harmonic: $E_b{}^{ij} = 1/2\ K\ (R_{ij} - R_o)^2$ AMBER95, Sybyl	K, R_o

BENDS

Key block (required)
This key block specifies the potential type and parameters for each kind of MM bond angle interaction. An example is given below.

```
BENDS
Atoms               pot    k            ao       NOTES
i - j - k           type   (kcal/mol)   deg
==========================================
*      CA    *    1      70.00        120.00   example of wild card
*      CA    CA   1      126.00       120.00
CA     CA    N2   1      140.00       120.10   amber95 N2-CA-CM
CA     CA    CT   1      140.00       120.00   amber95
==========================================
```

The first three columns specify the atom types and the fourth column is an integer specifying the potential type. The angle bend potential types are described in the table below with the additional constants required.

BEND potential type		constants required (in order)
0	no potential	none
1	theta harmonic: $E_\theta{}^{ijk} = 1/2\ K_\theta\ (\theta_{ijk} - \theta_o)^2$ AMBER95, SYBYL	K_θ, θ_o (θ in degrees)

Notice that wild cards can be specified for both terminal positions of the bend or just one as in the example above. It is important that the parameters be ordered from the least specific (those containing the most wild cards) to the most specific parameters.

TORSIONS

Key block (required)
This key block specifies the potential type and parameters for each kind of MM bond torsion interaction. For the bond stretching and bending potentials, only one potential has to date been implemented since both AMBER and SYBYL both use simple harmonic potentials. However, AMBER and SYBYL use different functional forms to represent the torsion potentials, each with their own set of parameters. The AMBER and SYBYL torsional potentials used in this program are defined in the table below.

TORSION potential type		constants required (in order)
0	no potential	none
1	AMBER: $$E_{tors} = \frac{K_{tor}}{N_{tors}}[1 + \cos(n\phi - \phi_o)]$$	K_i, n_i (periodicity-integer), $\phi_{o,i}$ (phase shift)
2	SYBYL: $$E_{tors} = \frac{1}{2}K_{tor}[1 + \cos(n\phi_o)\cos(n\phi)]$$	K, s

Notice that the two potentials have a different number of parameters. For example, when the program reads 'potential type' number 1, it will expect three parameters K_i, n_i, $\phi_{o,i}$. Further notice that the AMBER torsional potential is a sum of Fourier components (this is what the index i refers to).

Below is an example of the TORSIONS key block, made up of AMBER force field types.

```
TORSIONS
Atoms                    pot           per.    shift
i - j - k - l            type   k      n       to      NOTES
==================================================
*      CV   NB   *       1      2.4000 2       180.0   JCC,7,(1986),230
*      CW   NA   *       1      1.5000 2       180.0   JCC,7,(1986),230
&                               0.1000 3       0.0
C      N    CT   C       1      0.2000 2       180.0
N      CT   C    N       1      0.4000 4       180.0
&                               1.3500 2       180.0
&                               0.7500 1       180.0
CT     CT   N    C       1      0.5000 4       180.0
==================================================
```

Most AMBER torsion potentials are not specific to all four atoms i-j-k-l, but only on the central two, j-k. Wild cards are specified with the '*' symbol as illustrated above. Again, the ordering is important. The parameters should be ordered from least specific (those containing the most wild cards) to most specific. The AMBER torsion potential can be composed of more than one Fourier component for a single torsion potential. Additional Fourier components are specified with the '&' continuation symbol as in the example above. At the moment, up to 6 Fourier components are allowed. Notice that the individual components need not be specified in any particular order. In the above example key block, there are only 5 torsional potentials defined, not 8. Two of the potentials are composed of more than one Fourier component as indicated by the '&' continuation line.

Below is an example of the TORSIONS key block for the SYBYL force field. Notice that the potential

36

types are all '2'. There are fewer parameters and no multi component potentials. Also, some potentials are defined with two or only one wild card.

```
TORSIONS
-------------------------------------------
Atoms                   pot
i - j - k - l           type      k        per      NOTES
=================================================
*     C_ar  S_3   *     2       1.0000    3.0       *
*     S_3   S_3   *     2       0.0000    2.0       EXP
C_2   C_2   C_3   *     2       0.1260   -3.0       WHITE_77
C_3   C_2   C_3   *     2       0.1260    3.0       WHITE_77
H     C_2   C_3   *     2       0.2740    3.0       *
*     C_ar  C_ar  C_ar  2       2.3500   -2.0       *
*     C_2   C_3   C_2   2       0.1260    3.0       WHITE_77
*     C_2   C_3   C_3   2       0.1260    3.0       WHITE_77
*     C_2   C_3   H     2       0.2740    3.0       WHITE_77
*     C_3   C_3   H     2       0.3200    3.0       MC_88
O_2   C_2   C_3   C_3   2       0.7000   -3.0       JL_ES_
O_co  C_2   C_3   C_3   2       0.7000   -3.0       MAC_1
C_2   C_3   C_3   C_2   2       0.0400    3.0       WHITE_77
C_2   C_3   C_3   C_3   2       0.1260    3.0       WHITE_77
=================================================
```

One can also mix different potential types within the same force field file, as illustrated below. In this example, three are three potentials. The first two are SYBYL type potentials whereas the last one is a multi component AMBER potential.

```
H     C_2   C_3   *     2       0.2740    3.0
*     C_ar  C_ar  C_ar  2       2.3500   -2.0
N     CT    C     N     1       0.4000    4        180.0
&                               1.3500    2        180.0
&                               0.7500    1        180.0
```

OUT-OF_PLANE

Key block (required)
This key block specifies the potential type and parameters for each kind of MM out of plane bend. This potential is sometimes referred to as the inversion potential or improper torsions (depending on the force field). The potential types currently supported are provided in the table below.

out-of-plane potential type	description	constants required (in order)
0	no potential	none
1	AMBER: $E_{inv} = K [1 + \cos(n\varphi - \varphi_0)]$	K, n, φ_0 (n=2, φ_0 = 180° for planar, n=3, φ_0 = 120° for tetrahedral)
2	SYBYL: $E_{oopl} = K d^2$ d is the distance of the plane in Å	K

An example of the key block for the AMBER type potentials is given below. It is important to realize that the atom k is the **atom k is the central atom**. (We have adopted the somewhat odd standard of AMBER in this respect).

```
OUT-OF-PLANE
-------------------------
Atoms                   pot
i - j - k - l           type      K        to       NOTES
```

```
==============================================
*       *       CA      H4      1           1.10        180.0       bsd.on C6H6 nmodes
*       *       CA      H5      1           1.10        180.0       bsd.on C6H6 nmodes
*       O2      C       O2      1          10.50        180.0       JCC,7,(1986),230
*       N2      CA      N2      1          10.50        180.0       JCC,7,(1986),230
*       CT      N       CT      1           1.00        180.0       JCC,7,(1986),230
CK      CB      N*      CT      1           1.00        180.0
==============================================
```

VAN DER WAALS

Key block (required)
This key block specifies the potential type and parameters for each kind of MM van der Waals interaction between two atoms. A sample key block is shown below:

```
atom(s)    type    emin      rmin        alpha      NOTES
==========================================
CA                 -.0860    3.81600     12.00      amber95
HA                 -.0150    2.91800     12.00      amber95
Ni - HA    2       -.0480    2.7         12.00      NOTE potential type
Ni - CA    D       -.0480    2.7         12.00      default potential
==========================================
```

The van der Waals key block is somewhat different than the previous key blocks, because generally not every atom pair is defined with its own parameters. Rather, the parameters are assigned on a per atom basis and then special combination rules are used to construct the parameters for each atom pair combination. For this reason, a default potential type is defined in the FORCE_FIELD_SETTINGS key block.

VDW potential type		constants required (in order)
0	no potential	none
1	Lennard-Jones 12-6 $$E_{vdw}(R) = D_o \left[\left(\frac{R_o}{R} \right)^2 - 2 \left(\frac{R_o}{R} \right)^6 \right]$$	D_o, R_o
2	Exponential-6 or Buckingham $$E_{vdw}(R) = D_o \left\{ \left(\frac{6}{\zeta-6} \right) e^{\zeta\left[1 - \frac{R}{Ro}\right]} - \left[\left(\frac{\zeta}{\zeta-6} \right) \left(\frac{R_o}{R} \right)^6 \right] \right\}$$	D_o, R_o, x x=12.0 is standard
3	Purely Repulsive $$E_{vdw}(R) = D_o \left(\frac{6}{\zeta-6} \right) e^{\zeta\left[1 - \frac{R}{Ro}\right]}$$	D_o, R_o, x

For each type of van der Waals interaction, the program first scans the key block for pair specific parameters. For pair specific potentials, the default potential type can be replaced by any of the available potentials. The three sample lines below specify pair-specific potentials. The two atom types must be separated by a hyphen with spaces between the hyphen and the atom type. Following the specification of the atom pair, the potential type is defined. If D or d is specified here, then this means to use the default potential type. Following the potential type are the parameters needed for that potential type (see above table).

```
CA - CA    1       0.0860    3.81600     12.00      amber95
Ni - HA    0
Ni - CA    D       0.0480    2.7         12.00      default potential type
```

38

If a pair specific parameter can't be found, then the program looks for individual atom parameters corresponding to each of the atom types in the pair. The pair specific parameters are then constructed from combination of the two individual atom parameters using the following combination rules:

VDW potential type		
1	Lennard-Jones 12-6	$D_{ij} = (D_i * D_j)^{1/2}$, $R_{ij} = (R_i + R_j)/2$
2	Exponential-6 or Buckingham	$D_{ij} = (D_i * D_j)^{1/2}$, $R_{ij} = (R_i + R_j)/2$ $\zeta_{ij} = (\zeta_i * \zeta_j)^{1/2}$
3	Purely Repulsive	$D_{ij} = (D_i * D_j)^{1/2}$, $R_{ij} = (R_i + R_j)/2$ $\zeta_{ij} = (\zeta_i * \zeta_j)^{1/2}$

When individual atom parameters are not used, no potential type is specified since the default potential type is always used. An example is given below.

```
CA    0.0860    3.81600    12.00    amber95
HA    0.0150    2.91800    12.00    amber95
```

The ability to define pair specific parameters is especially useful for those force fields that have different combination rules than used in the program. For example, Jorgensen's TIP3P water force field uses geometric averages for both D_{ij} and R_{ij}.

MASSES & ATOM LABELS

Key block (required)
This key block specifies the default masses for each MM atom type and the element label for each MM atom type. In an ADF QM/MM run, the element label defined for each atom type is the label used for printing out to the LOGFILE. This allows one to easily cut and paste the generated coordinates to a molecule viewing program without having to go in and changing all of the "CT"s to "C"s.

A sample key block is shown below:

```
MASSES & ATOM LABELS
=================================
Ni               Ni            58.70
CM               C             12.011
CA               C             12.011
CT               C             12.011
HC               H             1.0079
HA               H             1.0079
=================================
```

The first column is the MM atom type, the second is the label used for printing and the third column is the mass of the atom type. The atoms do not have to be specified in any particular order.

CHARGES

Key block (optional)
This key block specifies the parameters for the charges on the atoms by atom type. To date only the initial charge is available, however if some sort of charge equilibration scheme was introduced the parameters would go here. NOTE: initial charges can also be specified on a per atom basis in the MM INPUT file.

```
CHARGES
atoms           initial
label           charge
```

```
========================
OW              -0.8
HW              0.4
========================
```

2.4 Setting up a QM/MM Simulation: a 'Walk Thru'

In this section we provide a detailed 'walk thru' of the process of setting up an ADF QM/MM simulation. There will be two examples, the first being a fairly straightforward example and the second one being fairly complex.

Example A: Cytocine

This is a straightforward example, where the input necessary to perform a QM/MM simulation of cytosine (Figure 5-1) will be constructed.

Step 1. Partitioning the System and the Model QM system

First one must decide where to partition the system into QM and MM regions. This is actually a very important step since the partitioning can be considered the 'original sin'. Much thought and testing should be put into deciding where to place the QM/MM boundary. In this example, we have chosen the partitioning depicted in Figure 5-1a in order to keep the example simple. In this figure the QM region enclosed in the dotted polygon, with two covalent bonds crossing the QM/MM boundary. One must also choose an appropriate QM model system for which the electronic structure calculation will be performed. To preserve the sp^3 hybridization of the carbon center in the QM region, we must keep the carbon tetravalent. Thus, we will cap the two dangling bonds with dummy or capping hydrogen atoms. One can use any monovalent atom such as H or F, but H is probably best. The reason that monovalent atoms should be used for capping atoms is that one does not want capping atom to have any 'dangling' bonds. Capping or dummy groups can not be used. Figure 5-1b, depicts the QM model system with two capping hydrogen atoms. Thus, the electronic structure calculation will be performed on methane such that the capping hydrogen atoms lie along the bond vector of the link bond in the real system as shown in Figure 5-1b.

a. QM/MM cytosine   b. QM model system
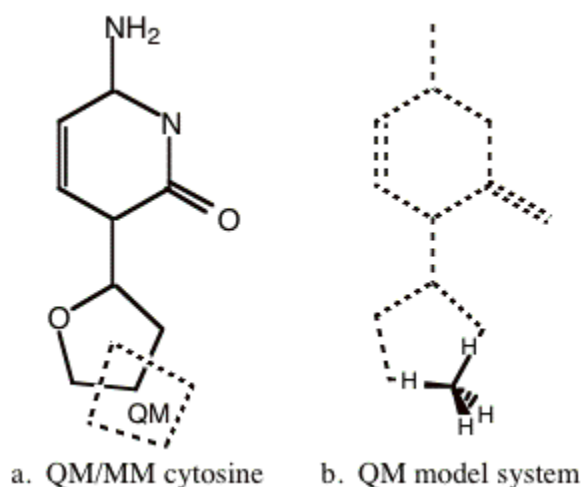
Figure 5-1 Cytocine QM/MM example model. a) Shows the whole system with the atoms enclosed in the dotted polygon making up the QM system. b) Shows the equivalent QM model system. The remainder of the

cytocine molecule is shown ghosted to demonstrate the relationship between the model system and the full system. The QM model system consists of a closed shell methane molecule.

Step 2. Labeling of Atoms (QM, MM or LI)

Once a partitioning of the system has been established, one needs to designate each of the atoms in the full system as QM, LI or MM atom type. For the example system, these designations are shown in Figure 5-2a, where the atoms that are not labeled are MM atoms.

a. QM, MM, LI labels   **b. atom numbering**   **c. AMBER95 atom-types**
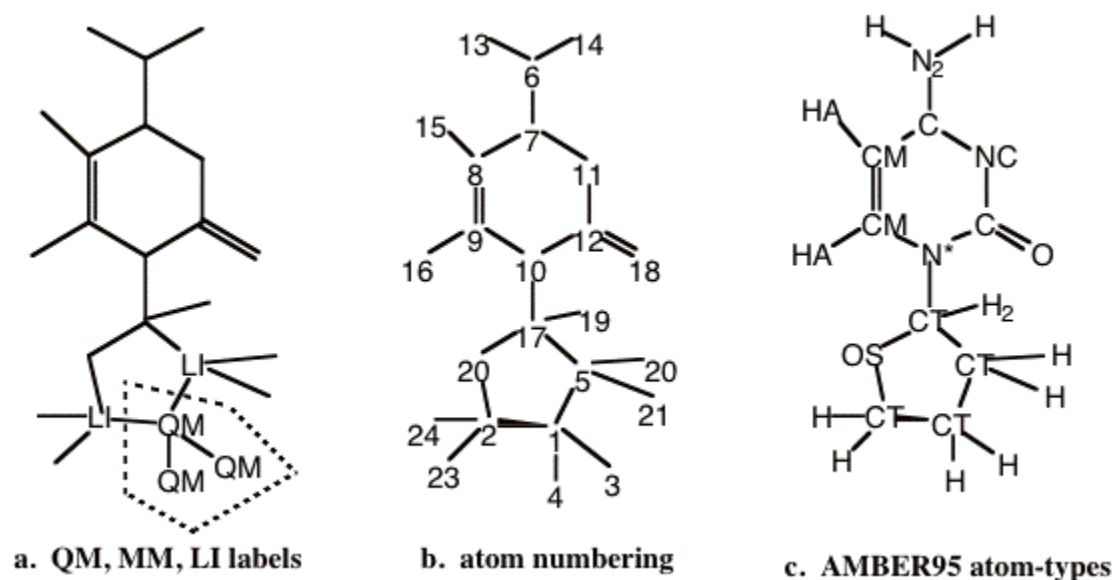
Figure 5-2 Labeling of the example model. a) QM, MM and LI designations. Atoms not labeled are 'MM' atoms. The dotted polygon encloses the QM region of the model. b) Atom numbering of the entire system. Note that the QM and LI atoms precede any MM atoms. c) The AMBER95 force field atom type designations.

All atoms within the dotted polygon are 'QM' type atoms. The atoms outside of the QM region will either be MM or LI atoms depending on whether they are part of a link bond or not. The covalent bonds that cross the QM/MM boundary are termed the link bonds. In the example system, there are two such covalent bonds. The atoms that lie on the MM side of the link bonds are labeled the LI atoms. Thus, these are all of the atoms that lie outside the dotted polygon that have a covalent bond to QM atoms. If there are no covalent bonds that traverse the QM/MM boundary, then there will be no LI atoms.

Step 3. Renumbering of Atoms

There is a strict rule concerning the ordering of the atoms, based on the QM, MM, or LI atom type designation. All QM and LI atoms must come before any MM atoms. A valid atom numbering for the example system is shown in figure 5-2b. Here the atoms labeled 'QM' or 'LI' in Figure 5-2a are the first five atoms of the molecular system.

Step 4. ADF QM/MM input: Atomic coordinates

Now we can begin to construct the input. We will begin with the atomic coordinates. For this example, we will optimize the geometry of the complex in Cartesian coordinates. Coordinates of the whole QM/MM complex or the 'real' complex should be defined here. DO NOT define the coordinates of the capping atoms.

The program will calculate their positions, and add them automatically. The definition of the coordinates is done exactly as they are in a standard ADF run. Below is the ATOMS key block for our example system.

```
ATOMS Cartesian
  1 C  1.94807  3.58290 -0.58162
  2 C  1.94191  3.61595  1.09448
  3 H  1.69949  4.49893 -1.05273
  4 H  2.99455  3.17964 -0.86304
  5 C  0.94659  2.40054 -0.92364
  6 N -1.74397 -3.46417  0.31178
  7 C -1.00720 -2.20758  0.33536
  8 C -1.66928 -1.00652  0.31001
  9 C -0.92847  0.25653  0.34895
 10 N  0.43971  0.26735  0.38232
 11 N  0.36409 -2.20477  0.28992
 12 C  1.09714 -0.95413  0.22469
 13 H -2.89781 -3.50815  0.31746
 14 H -1.21484 -4.49217  0.31721
 15 H -2.80940 -0.93497  0.30550
 16 H -1.55324  1.21497  0.33885
 17 C  1.23309  1.44017  0.30994
 18 O  2.58277 -1.01636  0.23914
 19 H  2.37276  1.25557  0.29984
 20 O  1.02358  2.43085  1.50880
 21 H  1.17136  1.95097 -1.87367
 22 H -0.10600  2.77333 -0.80348
 23 H  1.62170  4.54039  1.51392
 24 H  2.99608  3.28749  1.41345
END
```

Step 5. Connection Table and MM force field types

In order to construct a molecular mechanics potential, the program needs to know the connectivity of the molecular system and the molecular mechanics force field atom-type designations. In this example we are using the AMBER95 force field of Kollman and coworkers [4]. The appropriate AMBER95 atom-types for this molecule are shown in Figure 5-2c. No new atom types need to be introduced to the standard AMBER95 force field to treat this system. However, if this were needed, then the force field file would have to be modified.

Next a connection table needs to be constructed. For this program this needs to be done on an atom by atom basis. Either a fully redundant connection table or a fully non-redundant connection table is acceptable. A redundant connection table refers to one in which the covalent bonds are defined for all atoms. For example, if X is bonded to Y, in the connections for atom X, a bond is defined to atom Y. For the connections to atom Y, a bond is also defined to atom X even though the bond has already been defined. In a non-redundant connection table, when a bond is defined in the connections for atom X, it is not again defined in the connections for atom Y.

We now can begin to construct part of the input, namely the MM_CONNECTION_TABLE subkey block of the QMMM key block. For this example, the MM_CONNECTION_TABLE key block is given below.

```
MM_CONNECTION_TABLE
 1 CT QM 2 3 4 5
 2 CT LI 1 20 23 24
 3 HC QM 1
 4 HC QM 1
 5 CT LI 1 17 21 22
```

```
 6 N2 MM 7 13 14
 7 CA MM 6 8 11
 8 CM MM 7 9 15
 9 CM MM 8 10 16
10 N* MM 9 12 17
11 NC MM 7 12
12 C MM 10 11 18
13 H MM 6
14 H MM 6
15 HA MM 8
16 H4 MM 9
17 CT MM 5 10 19 20
18 O MM 12
19 H2 MM 17
20 OS MM 2 17
21 HC MM 5
22 HC MM 5
23 H1 MM 2
24 H1 MM 2
SUBEND
```

The first column is simply the atom number. The atoms defined here MUST be in the same order as defined in the ATOMS key block provided in the previous section. Again, we do not include the capping atoms. The second column shows the AMBER95 atom-types for our system, displayed in Figure 5-2c. The third column is the MM, QM or LI designation. Notice that the QM and LI atoms appear before any MM atoms. The remaining columns are reserved for the connection table. In the above example, a fully redundant connection table is provided.

Step 6. LINK_BONDS

When there are covalent bonds that cross the QM/MM boundary, the LINK_BONDS subkey block is required. Since one only defines the 'real system in both the ATOMS key block and the MM_CONNECTION_TABLE subkey block, this key block defines both the initial position of the capping atom and what kind of ADF fragment atom will be used as a capping atom. In this example we have two link bonds, both of which will be 'capped' with capping hydrogen atoms as shown in Figure 5-1b. Below is the LINK_BONDS subkey block for our example.

```
LINK_BONDS
  1 - 5 1.380 H
  1 - 2 1.375 H
SUBEND
```

The first part of the input specifies the atoms involved in link bonds. Here QM atom 1 forms link bonds with atoms 5 and 2. The column in the input is the link bond a parameter, which is defined as the ratio between the capping bond length in the QM model system and the bond length of the corresponding link bond in the real system. This ratio can be determined by taking the necessary bond lengths from a pure QM calculation of the model QM system, and the bond length from the whole complex. If those are not available, they can be taken from tabulated bond lengths or bond lengths of similar bonds in other complexes. There is an independent a parameter for each link bond. It is VERY IMPORTANT to emphasize that the total energy of the QM/M system is dependent upon the a parameters. Thus, if one is comparing the energetics of two conformational isomers calculated with the QM/MM method, this comparison is only valid if the a parameters used are the same. In our example, ratios of 1.38 and 1.375 were used. This is somewhat typical ratio of C-H to C-C bond lengths, in aliphatic hydrocarbons. The last column in the LINK_BONDS input refers to the ADF fragment for which will be used for the capping atom in the electronic structure calculation. Please, note this fragment must be present in the FRAGMENTS key block of the ADF input.

Step 7. Assignment of Atomic Charges

Perhaps the most dubious aspect of the QM/MM approach involves the non-bonded electrostatic interaction between the QM and MM regions. The ADF QM/MM extension currently only supports placement of static point charges on MM atoms. At the moment, you have two options. First, you can chose to have the MM point charges to interact with the electron density of the QM model system, thereby allowing the wave function of the QM system to be polarized. Alternatively, you can assign static point charges to the QM atoms which interact with MM point charges as would happen if the whole system were treated with a molecular mechanics force field. In this example, we will choose the latter, using the standard AMBER95 charges cytosine. To specific how the electrostatic interactions between the two regions are treated, one uses the ELSTAT_COUPLING_MODEL keyword in the QMMM key block and sets it equal to 1.

In ADF QM/MM the atomic point charges can be assigned on an atom-type basis, where the point charges are taken from the force field file. It can also be defined on a per atom basis, where a unique charge is assigned to each atom in the molecular system in the CHARGES subkey block. Since the charges in AMBER95 are assigned according to the nucleic or amino acid, we must assign the charges on a per-atom basis. Given below is the CHARGES subkey block with the appropriate AMBER95 point charges assigned to the system. The first column in this subkey block is the atom numbering. It is important to use the right atom number instead because the program actually determines the charges on each atom individually by searching for the atom number within this key block. Charges don't have to be in order.

```
ELSTAT_COUPLING_MODEL=1
CHARGES
 1  0.0000
 2  0.0000
 3  0.0000
 4  0.0000
 5  0.0000
 6 -0.9530
 7  0.8185
 8 -0.5215
 9  0.0053
10 -0.0484
11 -0.7584
12  0.7538
13  0.4234
14  0.4234
15  0.1928
16  0.1958
17  0.0066
18 -0.6252
19  0.2902
20 -0.2033
21  0.0000
22  0.0000
23  0.0000
24  0.0000
SUBEND
```

Step 8. Remainder of the QMMM key block

The ADF QM/MM input is almost complete. Now only a few settings need to be defined in the QMMM key block. The remainder of the QMMM key block is given below.

```
FORCEFIELD_FILE /usr/bob/QMMM_data/amber95.ff
RESTART_FILE mm.restart
```

44

```
    OUTPUT_LEVEL=1
    WARNING_LEVEL=2
```

The FORCEFIELD_FILE defines the filename of the force field file to be used. If the force field file is not in the running directory of the ADF job, then the full path needs to be specified. The RESTART_FILE specifies the name of the QM/MM restart to be written. If the job is a restart itself, this keyword also specifies the QM/MM restart file to read.

The OUTPUT_LEVEL specifies how much output to print during the course of the ADF QM/MM run. OUTPUT_LEVEL=1 is good for most purposes. Using an OUTPUT_LEVEL=2 is good when trouble shooting, but probably provides too much output when the job is running normally. The WARNING_LEVEL keyword specifies when to stop the job. When it is set to 2, the run stops at any spot where a potential QM/MM problem is detected. This is good when first setting up a job because the program attempts to point out potential problems.

Step 9. Putting it all together: The whole ADF QM/MM input

The whole ADF QM/MM input for the sample system is given below. The following will be a QM/MM geometry optimization performed in Cartesian coordinates with no constraints. Some comments are provided in bold.

```
Title CYT amber95 test - CARTESIAN GEOMETRY OPTIMIZATION NO CONSTRAINTS
Fragments
 C T21.C.III.1s  Notice that only fragments for the calculation of
 H T21.H.III  model system are needed.
End

Symmetry NOSYM

Charge 0 0  This refers to the charge of the QM model system, not the 'real' system

ATOMS Cartesian
 1 C 1.94807 3.58290 -0.58162
 2 C 1.94191 3.61595 1.09448
 3 H 1.69949 4.49893 -1.05273
 4 H 2.99455 3.17964 -0.86304
 5 C 0.94659 2.40054 -0.92364
 6 N -1.74397 -3.46417 0.31178
 7 C -1.00720 -2.20758 0.33536
 8 C -1.66928 -1.00652 0.31001
 9 C -0.92847 0.25653 0.34895
10 N 0.43971 0.26735 0.38232
11 N 0.36409 -2.20477 0.28992
12 C 1.09714 -0.95413 0.22469
13 H -2.89781 -3.50815 0.31746
14 H -1.21484 -4.49217 0.31721
15 H -2.80940 -0.93497 0.30550
16 H -1.55324 1.21497 0.33885
17 C 1.23309 1.44017 0.30994
18 O 2.58277 -1.01636 0.23914
19 H 2.37276 1.25557 0.29984
20 O 1.02358 2.43085 1.50880
21 H 1.17136 1.95097 -1.87367
22 H -0.10600 2.77333 -0.80348
23 H 1.62170 4.54039 1.51392
24 H 2.99608 3.28749 1.41345
```

```
END

QMMM
 FORCEFIELD_FILE amber95.ff
 RESTART_FILE mm.restart
 OUTPUT_LEVEL=1
 WARNING_LEVEL=2
 ELSTAT_COUPLING_MODEL=1
 LINK_BONDS
   1 - 5 1.38000 H
   1 - 2 1.38030 H
 SUBEND
 MM_CONNECTION_TABLE
   1 CT QM 2 3 4 5
   2 CT LI 1 20 23 24
   3 HC QM 1
   4 HC QM 1
   5 CT LI 1 17 21 22
   6 N2 MM 7 13 14
   7 CA MM 6 8 11
   8 CM MM 7 9 15
   9 CM MM 8 10 16
   10 N* MM 9 12 17
   11 NC MM 7 12
   12 C MM 10 11 18
   13 H MM 6
   14 H MM 6
   15 HA MM 8
   16 H4 MM 9
   17 CT MM 5 10 19 20
   18 O MM 12
   19 H2 MM 17
   20 OS MM 2 17
   21 HC MM 5
   22 HC MM 5
   23 H1 MM 2
   24 H1 MM 2
 SUBEND
 CHARGES
   1 0.0 CT
   2 0.0 CT
   3 0.0 HC
   4 0.0 HC
   5 0.0 CT
   6 -0.9530 N2
   7 0.8185 CA
   8 -0.5215 CM
   9 0.0053 CM
   10 -0.0484 N*
   11 -0.7584 NC
   12 0.7538 C
   13 0.4234 H
   14 0.4234 H
   15 0.1928 HA
   16 0.1958 H4
   17 0.0066 CT
   18 -0.6252 O
```

```
   19 0.2902 H2
   20 -0.2033 OS
   21 0.0000 HC
   22 0.0000 HC
   23 0.0000 H1
   24 0.0000 H1
 SUBEND
END

GEOMETRY
 ITERATIONS 20
 CONVERGE E=1.0E-3 GRAD=0.0005
 STEP RAD=0.3 ANGLE=5.0
 DIIS N=5 OK=0.1 CYC=3
END

XC
 LDA VWN
 GGA POSTSCF Becke Perdew
End

Integration 3.0

SCF
 Iterations 60
 Converge 1.0E-06 1.0E-6
 Mixing 0.20
 DIIS N=10 OK=0.500 CX=5.00 CXX=25.00 BFAC=0.00
End

End Input
```

In the above example, the geometry was defined with Cartesian coordinates and the geometry optimization was also done in Cartesians. The same input could also have easily been defined with a Z-matrix in the ATOMS key block.

Example B: Pd$^+$-Ethene pi-complexation Linear Transit

This example is more complex to demonstrate some problems that one might encounter in a more advanced problem. For instance, the simulation will involve the customization of the standard Tripos force field, the use of dummy atoms in the MM region and QM region and the use of constraints. Figure 5-3a depicts the system that we intend to simulate. More specifically we wish to determine the reaction profile of removing the

olefinic substrate from the metal center of Pd$^+$-phosphino-ferrocenyl-pyrazole complex. For this purpose we wish to perform a linear transit run with ADF, whereby the distance between the metal center and the midpoint of the olefinic carbons is used as a reaction coordinate. The linear transit geometry optimization will be done in internal coordinates. This reaction coordinate is shown as the arrow line in Figure 5-3. This example originates from our research on similar neutral bis-trichlorosilyl compounds. The system has been changed slightly to introduce additional technical considerations when using dummy atoms and linear transit calculations. The QM/MM calculations of these related compounds reveal that the approximations introduced in this system are quite reasonable. Also, in terms of predicting the geometry of this class of complexes, the QM/MM method performs exceptionally well.
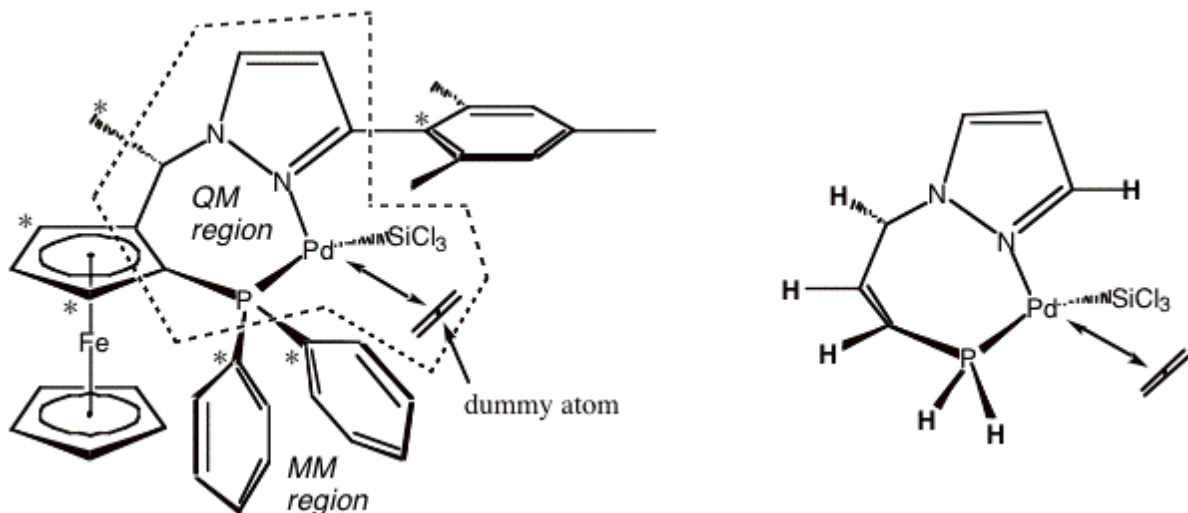
a b

Figure 5-3 Example system Pd$^+$-ethene pi-complex. a) Full system, with linear transit coordinate indicated by the arrow line. The QM/MM boundary is shown as the dotted polygon with the QM region residing inside. The 'LI' atoms are denoted with the asterisks. b) The QM model system with the capping hydrogen atoms depicted in bold.

Step 1. Partitioning the System and the Model QM system.

First one must decide where to partition the system into QM and MM regions. For this example, we have decide to partition the system as illustrated in Figure 53b, whereby the QM region is contained in the dotted polygon. The corresponding QM model system, for which the electronic structure calculation will be performed, is depicted in Figure 5-3b. In the model QM system the link atoms have been replaced by capping hydrogen atoms. Notice that the QM/MM boundary cuts through the cyclopentadienyl ring of the ferrocenyl ligand. Based on experimental studies of this complex, it is assumed that the ferrocenyl ligand acts only as a spectator ligand and can be modeled effectively on a steric basis only. Using an olefinic group will approximate the sp2 hybridization of the Cp rings. Here, special care must be taken to preserver the structural features of the Cp ring. For example the C-C bond distance in the ferrocenyl ligand is approximately 1.45 Ang whereas it is about 1.34 Ang in an olefin. This will be elaborated on later. The replacement of the phenyl phosphine in the real system by hydrogen phosphine will have some consequences due to the different electronic properties of the substituents. It is known that the phenyl substitution on the phosphine is more electron withdrawing than the hydrogen substituent. The replacement of the phenyl phosphine with hydrogen phosphine will result in a contraction of the Pd-P bond in the QM model system and hence the Pd-P bond will be too short in our QM/MM model.

Step 2. Labeling of Atoms

Once a partitioning of the system has been established, one needs to label each of the atoms in the full system as QM, LI or MM atom type. For the example system, all atoms contained within the dotted polygon in Figure 5-3a are 'QM' atoms. All atoms that are marked with asterisks in Figure 5-3a are 'LI' atoms and finally the remaining atoms outside the dotted polygon are 'MM' atoms. The dummy atom that we need in order to define our reaction coordinate is designated a QM atom. This dummy atom will be made to lie midway between the two olefinic carbon atoms of the ethene moiety. It is important to realize that the linear transit constraint cannot involve any MM atoms. Two dummy atoms representing the center of the Cp rings of the ferrocenyl ligand will also be introduced. They we be part of the MM subsystem.

Step 3. Renumbering of Atoms

It will be re-emphasize that there is a strict rule concerning the ordering of the atoms, based on their QM, MM, or LI atom type designation. All QM and LI atoms must come before any MM atoms. This rule also applies to the dummy atoms. All atoms in the QM model system shown in Figure 5-3b and their equivalent LI atoms in the real system must come first. Given below are the Cartesian coordinates of the initial geometry with the atoms renumbered. Although the optimization will be performed in internal coordinates, this is a complex example, and it might help the reader to examine the 3D structure of the complex with their favorite molecule viewer.

```
Pd 0.00000 0.00000 0.00000
N 2.18381 0.00000 0.00000
P -0.19353 2.33087 0.00000
Si -2.09382 -0.72920 0.40993
Cl -3.11786 -1.66043 -1.19030
Cl -3.49847 0.77293 1.00006
Cl -2.26795 -2.09058 2.02296
C 1.00751 3.35326 -0.90266
C 2.19320 2.92863 -1.63738
C 2.55933 1.49397 -1.90948
N 3.04680 0.78384 -0.70880
C 4.30216 0.71267 -0.18548
C 4.25805 -0.16196 0.88628
C 2.91893 -0.57760 0.96569
Xx 0.74788 -1.69468 -1.67891
C 1.00986 -2.06361 -1.18981
C 0.48590 -1.32574 -2.16801
H 0.42486 -2.82737 -0.67948
H 2.04440 -1.93825 -0.86750
H 1.06712 -0.55943 -2.68284
H -0.54392 -1.46716 -2.49409
C -0.02313 3.09172 1.69751
C -1.80681 3.04317 -0.56800
C 1.06326 4.78500 -0.83878
C 2.90531 4.10963 -2.00851
H 1.82015 0.84892 -1.41063
C 2.56571 1.10455 -3.38246
H 5.13150 1.27389 -0.59600
H 5.08293 -0.44584 1.52605
C 2.28008 -1.58274 1.88382
Fe 1.04114 4.20972 -2.75447
H 3.29128 1.69565 -3.95722
H 2.82573 0.03883 -3.48724
H 1.57260 1.26129 -3.82619
C 2.23262 5.27491 -1.51096
H 3.82950 4.13635 -2.57654
H 2.53722 6.31008 -1.62736
H 0.35288 5.41923 -0.32204
C -0.36634 3.33949 -3.93243
C -0.80398 4.62382 -3.46691
C 0.14944 5.60144 -3.90508
C 1.17457 4.92275 -4.64218
C 0.85084 3.52607 -4.66695
H 1.38331 2.75840 -5.21594
H -0.88905 2.39926 -3.80335
H -1.70862 4.82843 -2.90699
```

```
H 0.10418 6.66814 -3.70967
H 2.03573 5.38561 -5.11158
C -2.52881 2.31699 -1.52071
C -3.74190 2.80302 -2.01422
C -4.24163 4.02394 -1.55340
C -3.52216 4.75609 -0.60549
C -2.30668 4.26891 -0.11529
H -2.14465 1.36750 -1.87729
H -4.29443 2.23258 -2.75379
H -5.18572 4.40286 -1.93263
H -3.90891 5.70542 -0.24710
H -1.76468 4.85404 0.61704
C -1.11030 3.12908 2.57588
C -0.97634 3.68572 3.85061
C 0.25426 4.20707 4.25852
C 1.35020 4.15676 3.39277
C 1.21425 3.58782 2.12359
H -2.06730 2.72549 2.27375
H -1.82678 3.71355 4.52356
H 0.35921 4.64813 5.24539
H 2.30826 4.55806 3.70733
H 2.08227 3.52803 1.47710
C 2.71278 -2.91472 1.84328
C 2.09262 -3.86832 2.66014
C 1.02394 -3.51226 3.48850
C 0.62539 -2.17333 3.54762
C 1.26348 -1.20208 2.77018
C 3.85077 -3.34061 0.90256
H 2.44054 -4.89548 2.65117
C 0.28665 -4.56067 4.32981
H -0.18696 -1.88618 4.20708
C 0.83254 0.25922 2.93922
H 4.81150 -3.02160 1.33267
H 3.72469 -2.87842 -0.08835
H 3.87528 -4.43158 0.76399
H 1.55468 0.95128 2.48757
H 0.75353 0.50536 4.00860
H -0.15276 0.39937 2.47265
H -0.79950 -4.40165 4.25848
H 0.51050 -5.58360 3.99174
H 0.59686 -4.45719 5.38063
Xx 1.88038 4.09028 -1.37966
Xx 0.20091 4.40271 -4.12271
```

Step 4. Z-matrix and constraints

This simulation will be a linear transit simulation where the reaction coordinate will be the distance from the Pd center to the midpoint of the complexed ethene molecule. In order to use midpoint of the ethene molecule as the reaction coordinate we must use a dummy atom to define the midpoint and a few constraints to maintain the dummy atom at the midpoint. The dummy atom is atom number 15 and the two carbon atoms of the ethene moiety are atoms 16 and 17. Each of the two carbons of the ethene will be 'bonded' to the midpoint using the same (free) bond distance variable 'B15'. This will ensure that the midpoint dummy atom is always equidistant to each ethene carbon. To ensure that the midpoint dummy atom always lies along the C-C bond vector the dihedral variable 'D14' will be constrained to 180 degrees. Finally, to prevent C-C bond distance of olefinic group used to model the ferrocenyl ligand to revert to its natural bond length of approximately 1.34 Ang, the C-C distance (B8) will be constrained to 1.45 Ang. This is

50

the distance found in the C-C bond distance in the ferrocenyl ligand. One might be concerned about the internal coordinate definition of the cyclopentadienyl rings of the ferrocenyl ligand. Since the ferrocenyl ligand is part of the MM region, the Z-matrix will be used only to construct the initial geometry. From there the molecular mechanics code takes over and where the optimization is done in Cartesian coordinates. For MM atoms it is important that care be taken when defining the connection table.

```
ATOMS internal
 Pd 0 0 0 0 0 0
 N 1 0 0 B1 0 0
 P 1 2 0 B2 A1 0
 Si 1 2 3 B3 A2 D1
 Cl 4 1 2 B4 A3 D2
 Cl 4 1 5 B5 A4 D3
 Cl 4 1 5 B6 A5 D4
 C 3 1 2 B7 A6 D5
 C 8 3 1 B8 A7 D6
 C 9 8 3 B9 A8 D7
 N 10 9 8 B10 A9 D8
 C 11 2 1 B11 A10 D9
 C 12 11 2 B12 A11 D10
 C 2 1 11 B13 A12 D11
 XX 1 2 3 B14 A13 D12
 C 15 1 2 B15 A14 D13
 C 15 1 16 B15 A15 D14
 H 16 15 1 B17 A16 D15
 H 16 17 18 B18 A17 D16
 H 17 16 18 B19 A18 D17
 H 17 16 18 B20 A19 D18
 C 3 1 8 B21 A20 D19
 C 3 1 8 B22 A21 D20
 C 8 3 9 B23 A22 D21
 C 9 8 10 B24 A23 D22
 H 10 9 8 B25 A24 D23
 C 10 9 26 B26 A25 D24
 H 12 11 13 B27 A26 D25
 H 13 12 11 B28 A27 D26
 C 14 2 1 B29 A28 D27
 Fe 24 8 3 B30 A29 D28
 H 27 10 9 B31 A30 D29
 H 27 10 32 B32 A31 D30
 H 27 10 32 B33 A32 D31
 C 25 9 8 B34 A33 D32
 H 25 9 35 B35 A34 D33
 H 35 25 9 B36 A35 D34
 H 24 8 31 B37 A36 D35
 C 31 24 8 B38 A37 D36
 C 39 31 24 B39 A38 D37
 C 40 39 31 B40 A39 D38
 C 41 40 39 B41 A40 D39
 C 39 31 40 B42 A41 D40
 H 43 39 31 B43 A42 D41
 H 39 31 40 B44 A43 D42
 H 40 39 41 B45 A44 D43
 H 41 40 42 B46 A45 D44
 H 42 41 40 B47 A46 D45
 C 23 3 1 B48 A47 D46
 C 49 23 3 B49 A48 D47
```

```
  C 50 49 23 B50 A49 D48
  C 51 50 49 B51 A50 D49
  C 52 51 50 B52 A51 D50
  H 49 23 50 B53 A52 D51
  H 50 49 51 B54 A53 D52
  H 51 50 52 B55 A54 D53
  H 52 51 53 B56 A55 D54
  H 53 52 51 B57 A56 D55
  C 22 3 1 B58 A57 D56
  C 59 22 3 B59 A58 D57
  C 60 59 22 B60 A59 D58
  C 61 60 59 B61 A60 D59
  C 62 61 60 B62 A61 D60
  H 59 22 60 B63 A62 D61
  H 60 59 61 B64 A63 D62
  H 61 60 62 B65 A64 D63
  H 62 61 63 B66 A65 D64
  H 63 62 61 B67 A66 D65
  C 30 14 2 B68 A67 D66
  C 69 30 14 B69 A68 D67
  C 70 69 30 B70 A69 D68
  C 71 70 69 B71 A70 D69
  C 72 71 70 B72 A71 D70
  C 69 30 70 B73 A72 D71
  H 70 69 71 B74 A73 D72
  C 71 70 72 B75 A74 D73
  H 72 71 73 B76 A75 D74
  C 73 72 71 B77 A76 D75
  H 74 69 30 B78 A77 D76
  H 74 69 79 B79 A78 D77
  H 74 69 79 B80 A79 D78
  H 78 73 72 B81 A80 D79
  H 78 73 82 B82 A81 D80
  H 78 73 82 B83 A82 D81
  H 76 71 70 B84 A83 D82
  H 76 71 85 B85 A84 D83
  H 76 71 85 B86 A85 D84
  XX 24 8 31 B87 A86 D85
  XX 41 40 42 B88 A87 D86
END

GEOVAR
 B1=2.18381
 B2=2.33889
 B3=2.25474
 B4=2.11579
 B5=2.13955
 B6=2.11791
 B7=1.81730
 B8=1.45807 F
 B9=1.50544
 B10=1.47768
 B11=1.36193
 B12=1.38405
 B13=1.34409
 B14=2.50000 5.000
 B15=0.66631
```

```
B17=1.08903
B18=1.09082
B19=1.09092
B20=1.08943
B21=1.86801
B22=1.85275
B23=1.43426
B24=1.42814
B25=1.10060
B26=1.52360
B27=1.08226
B28=1.08182
B29=1.50380
B30=2.00032
B31=1.09827
B32=1.10198
B33=1.09896
B34=1.43456
B35=1.08513
B36=1.08532
B37=1.08347
B38=2.03122
B39=1.43448
B40=1.43413
B41=1.43347
B42=1.43383
B43=1.08361
B44=1.08347
B45=1.08341
B46=1.08540
B47=1.08451
B48=1.39867
B49=1.39691
B50=1.39740
B51=1.39722
B52=1.39822
B53=1.08455
B54=1.08520
B55=1.08569
B56=1.08594
B57=1.08280
B58=1.39816
B59=1.39740
B60=1.39734
B61=1.39754
B62=1.39750
B63=1.08166
B64=1.08484
B65=1.08604
B66=1.08530
B67=1.08396
B68=1.40109
B69=1.40042
B70=1.39823
B71=1.39824
B72=1.39818
B73=1.53667
```

```
B74=1.08452
B75=1.53315
B76=1.08501
B77=1.53287
B78=1.09990
B79=1.10064
B80=1.10001
B81=1.09746
B82=1.10018
B83=1.09915
B84=1.10005
B85=1.10036
B86=1.10053
B87=1.20119
B88=1.21941
A1=94.7463
A2=158.2223
A3=117.0192
A4=115.6423
A5=115.0054
A6=120.3960
A7=128.6001
A8=124.4982
A9=113.0381
A10=110.6944
A11=107.5291
A12=123.1566
A13=72.5934
A14=90.0001
A15=90.0001
A16=121.4060
A17=121.7463
A18=121.7273
A19=121.2990
A20=113.4758
A21=117.0770
A22=124.4140
A23=107.1936
A24=108.2564
A25=114.7923
A26=122.2418
A27=126.8321
A28=119.5733
A29=70.7204
A30=111.7865
A31=109.8825
A32=110.4946
A33=110.3514
A34=125.5260
A35=127.2980
A36=125.4579
A37=133.3349
A38=68.8766
A39=107.9605
A40=108.0463
A41=69.7028
A42=125.7664
```

```
A43=129.5676
A44=126.2672
A45=125.9595
A46=126.0404
A47=117.3031
A48=120.5489
A49=119.8703
A50=119.8287
A51=120.1922
A52=119.7123
A53=120.0017
A54=120.0478
A55=119.9075
A56=118.9447
A57=120.7197
A58=120.6090
A59=119.9563
A60=119.7586
A61=120.0640
A62=120.1660
A63=120.0615
A64=120.1253
A65=119.9602
A66=119.0914
A67=119.1156
A68=119.5849
A69=120.7057
A70=119.1275
A71=120.7701
A72=120.6568
A73=119.8697
A74=121.2330
A75=119.5074
A76=118.2033
A77=109.0914
A78=110.4834
A79=111.6228
A80=111.7639
A81=109.9169
A82=109.0716
A83=109.9009
A84=111.6715
A85=108.9122
A86=55.1829
A87=54.0180
D1=-150.6568
D2=-100.8347
D3=-119.5809
D4=123.4778
D5=35.1598
D6=-2.9958
D7=7.2171
D8=-72.9107
D9=164.5027
D10=0.9220
D11=163.0063
D12=135.2677
```

```
D13=65.6662
D14=180.0000 F
D15=90.0000
D16=179.5031
D17=-179.9904
D18=-0.4168
D19=-117.3520
D20=124.7005
D21=-169.6021
D22=-179.2713
D23=-3.3919
D24=119.3500
D25=179.9891
D26=-179.9514
D27=17.4102
D28=-129.2972
D29=60.6432
D30=120.3390
D31=-120.3043
D32=1.2969
D33=179.4155
D34=179.5397
D35=123.6746
D36=68.3005
D37=62.0893
D38=58.4426
D39=0.0736
D40=119.7199
D41=127.1496
D42=-120.1862
D43=178.4186
D44=-179.8093
D45=179.5019
D46=-30.2454
D47=-178.8402
D48=-0.0598
D49=0.3012
D50=-0.1575
D51=-179.8476
D52=-179.9824
D53=179.8951
D54=-179.8926
D55=179.8489
D56=-81.7949
D57=179.3002
D58=-0.2287
D59=-0.6445
D60=0.0510
D61=-179.9960
D62=179.9913
D63=-179.8879
D64=-179.8700
D65=-178.0390
D66=114.6178
D67=-177.7647
D68=1.7082
D69=-3.4460
```

```
D70=1.3564
D71=179.4189
D72=179.8505
D73=-179.9397
D74=-179.8352
D75=-176.6101
D76=76.9134
D77=-120.1615
D78=119.9038
D79=165.3767
D80=-119.6911
D81=121.2928
D82=-135.8162
D83=120.3122
D84=-119.4211
D85=-58.7928
D86=0.0762
END
```

Step 5. Connection Table, MM force field atom-types and Force Field Modification

In order to construct a molecular mechanics potential, the program needs to know the connectivity of the molecular system and the molecular mechanics force field atom-type designations. In this example we are using the Tripos or Sybyl force field. The Tripos force field does not support either Pd or ferrocenyl ligands, so we need to modify the standard force field file to handle these groups. Modification of a molecular mechanics force field without re-parameterization of the force field may not always be appropriate. However, in this case sort of 'ad hoc' additions to the Tripos force field can be justified. For Pd, all of the principle interactions will be contained within the QM region and only weak non-bonded interactions involving Pd will be approximated by the molecular mechanics potential. The ferrocenyl ligand is assumed to act as a spectator ligand and therefore it is adequate to simply attain the approximate structure of the complex with the molecular mechanics potential.

In the Tripos force field, the nitrogen atoms of the pyrazole ring should be assigned the 'N_2' atom-type; the P atom of the phosphine should be assigned the 'P_3' atom-type. The Cl, H and Si atoms are given the 'Cl', 'H', and 'Si' atom types respectively. The carbon atoms of the phenyl substituents are given the 'C_ar' atom-type, while the sp^3 hybridized carbon atoms are given the 'C_3' atom-type.

Connections involving the dummy atom defining the midpoint of the ethene molecule are really not needed since this atom is contained within the QM region.

For the ferrocenyl ligand the ferrocene force field of Bosnich and coworkers will be used. Four new MM atom types will be introduced, C_cp, H_cp and CEN, representing the carbon, hydrogen and centroid of the cyclopentadienyl rings, respectively and Fe. In the connection table, the C_cp atoms will be bonded to the centroid and not the Fe center. The only two bonds made to the Fe center will be to the (two) central dummy atoms of the Cp rings. In making a connection between the C_cp atom and the centroid, a direct bond will is made to a QM atom and a MM atom. A warning will be issued during the run but as long as the 'WARNING_LEVEL' flag is set to 1 the job will continue. In this case the link bond between the C_cp atom and the centroid does not need to be mediated by a capping atom. This bond is used only for the construction of the MM potential for the ferrocenyl ligand. Special bond stretching, bending, torsion and out-of-plane potentials need to be added to the force field file. For the most part these parameters are taken from the Bosnich Ferrocene force field. For example for the bond stretches, the following potentials need to be added to the force field.

```
# Parameters added for Pd - ethene complex
C_cp   C_cp   1      1400.00      1.434      From the Bosnich ferrocene force field.
C_cp   H_cp   1      692.00       1.085      Bosnich
```

```
CEN   Fe    1      600.00      1.617     Bosnich
CEN   C_cp  1      600.00      1.220     Bosnich
```

The force field file is simply a text file and so the above section needs to be added to the 'BONDS' key block between the two '=======' separator lines. Bond potentials need to be defined between the centroid of the Cp rings with the Fe center and the carbon atoms.

For the angle and torsion terms, the additions are somewhat more complex. The following angle potential terms need to be introduced.

```
# Parameters added for Pd - ethene complex
C_cp  C_cp  C_cp  1      78.80      126.0     Bosnich
C_cp  C_cp  H_cp  1      78.80      126.0     Bosnich
CEN   C_cp  C_cp  1      0.00       0.0       no potential
CEN   C_cp  H_cp  1      0.00       0.0       no potential
C_cp  CEN   C_cp  1      0.00       0.0       no potential
C_cp  CEN   Fe    1      100.00     90.0      Bosnich
CEN   Fe    CEN   1      100.00     180.0     Bosnich
CEN   C_cp  P_3   1      0.00       0.0       no potential
CEN   C_cp  C_3   1      0.00       0.0       no potential
```

Any angle potentials involving the Cp centroid and any atoms outside of the ferrocenyl ligand have been set to zero since the centroid was only a construct for the optimization of the ferrocenyl ligand. For the torsions, the following potentials have been added to the standard Tripos force field.

```
# Parameters added for Pd - ethene complex
P_3   C_cp  C_cp  C_cp  2    2.0000    -2.0    Sybyl *-C_ar-C_ar-* aromatic bond
P_3   C_cp  C_cp  H_cp  2    2.0000    -2.0    Sybyl *-C_ar-C_ar-* aromatic bond
C_3   C_cp  C_cp  H_cp  2    2.0000    -2.0    Sybyl *-C_ar-C_ar-* aromatic bond
H_cp  C_cp  C_cp  H_cp  2    2.0000    -2.0    Sybyl *-C_ar-C_ar-* aromatic bond
*     C_cp  C_cp  C_cp  2    2.3500    -2.0    same as SYBYL * C_ar C_ar C_ar
*     Fe    CEN   *     0    0.0000    -2.0    no potential involving centroid
*     C_cp  CEN   *     0    0.0000    0.0     no potential involving centroid
*     C_cp  C_cp  CEN   0    0.0000    0.0     no potential involving centroid
Pd    P_3   C_cp  CEN   0    0.0000    0.0     no potential involving centroid
CEN   C_cp  P_3   C_ar  0    0.0000    0.0     no potential involving centroid
N_2   C_3   C_cp  CEN   0    0.0000    0.0     no potential involving centroid
CEN   C_cp  C_3   H     0    0.0000    0.0     no potential involving centroid
CEN   C_cp  C_3   C_3   0    0.0000    0.0     no potential involving centroid
```

Here any torsional potentials involving the Centroid atom of the ferrocenyl ligand were set to zero. Torsional potentials involving atoms outside of the ferrocenyl ligand and having the C_cp-C_cp atoms central atom pair, these potentials were equated with those of the Tripos ' *-C_ar C_ar - * ' torsional potentials. Again, these somewhat arbitrary choices for the MM potentials involving the ferrocenyl ligand are justified by the fact that the ferrocenyl ligand acts only as a spectator group.

The van der Waals parameters used for the five new atoms types, Pd, Fe, CEN, C_cp and H_cp were taken from either existing Tripos van der Waals parameters of similar atom-types or they were taken from Rappe's UFF (Universal Force Field). They are given below with their origins provided in the 'NOTES' column.

```
# Parameters added for Pd - ethene complex
C_cp  0.1070   3.4000   12.00     same as Tripos C_ar
Fe    0.0130   2.9120   12.00     UFF92 Fe6+2
Pd    0.0480   2.8990   12.00     UFF92 Pd4+2
CEN   0.0000   1.0000   12.00     zero
H_cp  0.0420   3.0000   12.00     same as Tripos H
```

Now that the addition of the new MM potentials and atom-types has been discussed, the 'MM_CONNECTION_TABLE' subkey block is given below. In practice, one typically constructs the input first, and then runs the program to see what force field potentials/parameters are missing. If any force field parameters are missing in the force field file, the ADF QM/MM program will print all missing potentials that need to be defined in the force field and then stop.

```
MM_CONNECTION_TABLE
 1 Pd QM 2 3 4 0 0 0
 2 N_2 QM 1 11 14 0 0 0
 3 P_3 QM 1 8 22 23 0 0
 4 Si QM 1 5 6 7 0 0
 5 Cl QM 4 0 0 0 0 0
 6 Cl QM 4 0 0 0 0 0
 7 Cl QM 4 0 0 0 0 0
 8 C_cp QM 3 9 24 88 0 0
 9 C_cp QM 8 10 25 88 0 0
 10 C_3 QM 9 11 26 27 0 0
 11 N_2 QM 2 10 12 0 0 0
 12 C_ar QM 11 13 28 0 0 0
 13 C_ar QM 12 14 29 0 0 0
 14 C_ar QM 2 13 30 0 0 0
 15 XX QM 16 17 1 0 0 0
 16 C_2 QM 15 17 18 19 0 0
 17 C_2 QM 15 16 20 21 0 0
 18 H QM 16 0 0 0 0 0
 19 H QM 16 0 0 0 0 0
 20 H QM 17 0 0 0 0 0
 21 H QM 17 0 0 0 0 0
 22 C_ar LI 3 59 63 0 0 0
 23 C_ar LI 3 49 53 0 0 0
 24 C_cp LI 8 35 38 88 0 0
 25 C_cp LI 9 35 36 88 0 0
 26 H QM 10 0 0 0 0 0
 27 C_3 LI 10 32 33 34 0 0
 28 H QM 12 0 0 0 0 0
 29 H QM 13 0 0 0 0 0
 30 C_ar LI 14 69 73 0 0 0
 31 Fe MM 88 89 0 0 0 0
 32 H MM 27 0 0 0 0 0
 33 H MM 27 0 0 0 0 0
 34 H MM 27 0 0 0 0 0
 35 C_cp MM 24 25 37 88 0 0
 36 H_cp MM 25 0 0 0 0 0
 37 H_cp MM 35 0 0 0 0 0
 38 H_cp MM 24 0 0 0 0 0
 39 C_cp MM 40 43 45 89 0 0
 40 C_cp MM 39 41 46 89 0 0
 41 C_cp MM 40 42 47 89 0 0
 42 C_cp MM 41 43 48 89 0 0
 43 C_cp MM 39 42 44 89 0 0
 44 H_cp MM 43 0 0 0 0 0
 45 H_cp MM 39 0 0 0 0 0
 46 H_cp MM 40 0 0 0 0 0
 47 H_cp MM 41 0 0 0 0 0
 48 H_cp MM 42 0 0 0 0 0
 49 C_ar MM 23 50 54 0 0 0
 50 C_ar MM 49 51 55 0 0 0
```

```
  51 C_ar MM 50 52 56 0 0 0
  52 C_ar MM 51 53 57 0 0 0
  53 C_ar MM 23 52 58 0 0 0
  54 H MM 49 0 0 0 0 0
  55 H MM 50 0 0 0 0 0
  56 H MM 51 0 0 0 0 0
  57 H MM 52 0 0 0 0 0
  58 H MM 53 0 0 0 0 0
  59 C_ar MM 22 60 64 0 0 0
  60 C_ar MM 59 61 65 0 0 0
  61 C_ar MM 60 62 66 0 0 0
  62 C_ar MM 61 63 67 0 0 0
  63 C_ar MM 22 62 68 0 0 0
  64 H MM 59 0 0 0 0 0
  65 H MM 60 0 0 0 0 0
  66 H MM 61 0 0 0 0 0
  67 H MM 62 0 0 0 0 0
  68 H MM 63 0 0 0 0 0
  69 C_ar MM 30 70 74 0 0 0
  70 C_ar MM 69 71 75 0 0 0
  71 C_ar MM 70 72 76 0 0 0
  72 C_ar MM 71 73 77 0 0 0
  73 C_ar MM 30 72 78 0 0 0
  74 C_3 MM 69 79 80 81 0 0
  75 H MM 70 0 0 0 0 0
  76 C_3 MM 71 85 86 87 0 0
  77 H MM 72 0 0 0 0 0
  78 C_3 MM 73 82 83 84 0 0
  79 H MM 74 0 0 0 0 0
  80 H MM 74 0 0 0 0 0
  81 H MM 74 0 0 0 0 0
  82 H MM 78 0 0 0 0 0
  83 H MM 78 0 0 0 0 0
  84 H MM 78 0 0 0 0 0
  85 H MM 76 0 0 0 0 0
  86 H MM 76 0 0 0 0 0
  87 H MM 76 0 0 0 0 0
  88 CEN MM 8 9 24 25 35 31
  89 CEN MM 39 40 41 42 43 31
SUBEND
```

Step 6. LINK_BONDS

In this example there are 6 link bonds as depicted in Figure 5-3a. The link bond parameters a for each of these bonds will be determined by comparing bond lengths in the X-ray structure of a similar bis-trichlorosilyl Pd complex with that of the calculated pure QM gas-phase structure of the QM model system. As shown in Figure 5-3b all link bonds will be capped with Hydrogen atoms. Although the MM connection table defines direct bonds between atoms 8 and 9 (C Cp atoms) with the centroid of the Cp rings, they are not mediated by capping atoms. In other words the Cp centroid is defined as a MM atom-type not a LI atom type. Therefore, no link parameters are necessary for those two bonds.

```
LINK_BONDS
:: ---------- ------ ---
:: atoms alpha dummy
:: ---------- ------ ---
 22 - 3 1.2990 H
```

```
 23 - 3 1.2990 H
 24 - 8 1.3200 H
 25 - 9 1.3200 H
 27 - 10 1.3710 H
 30 - 14 1.3800 H
:: ---------- ------ ---
SUBEND
```

Step 7. CHARGES

In this example, the Pd center has a formal positive charge. In order to obtain the proper electronic structure in the calculation of the QM model system, we must define in the ADF input a formal positive charge with the CHARGE keyword.

```
CHARGE 1  {note this is in the main ADF input}
```

The original Tripos force field was parameterized without explicit electrostatic terms. Thus, we will use this convention and turn off the electrostatic coupling between the QM and MM regions using the ELSTAT_COUPLING_MODEL keyword in the QMMM key block.

```
ELSTAT_COUPLING_MODEL=0
```

Although this may seem like a dubious choice, experience with organometallic complexes has shown that this is a good approximation. For other types of molecular systems, namely amino and nucleic acids it is not a good choice to turn off the electrostatic coupling between the QM and MM regions. Furthermore, force fields designed for this biochemical species almost always have charges included in the parameterization process.

Step 9. Putting it all together: The whole ADF QM/MM input.

The whole ADF QM/MM input for the sample system is given below. Some comments are provided in bold. Additionally, some lengthy sections have been omitted that have already been given in full above.

```
TITLE Complex force field Example of Pd+-Ethene complex.
NOPRINT SFO
Fragments
 Pd T21.Pd.3d.rel
 C  T21.C.1s.rel
 Si T21.Si.2p.rel
 Cl T21.Cl.2p.rel
 H  T21.H.rel
 N  T21.N.1s.rel
 P  T21.P.2p.rel
End

RELATIVISTIC SCALAR
COREPOTENTIALS ADF.t12 &
 Pd 1
 P  2
 Si 3
 Cl 4
 N  5
 C  6
 H  7
END
```

```
SYMMETRY NOSYM

CHARGE 1 CHARGE is defined from the QM model system

GEOMETRY
 LINEAR TRANSIT 4
 ITERATIONS 2
 HESSUPD BFGS
 CONVERGE E=2.0E-3 GRAD=0.002
 DIIS N=5 OK=0.005 CYC=2
END

XC
 LDA VWN
 GGA Becke Perdew
End

Integration 3.0 3.0

SCF
 Iterations 60
 Converge 1.0E-06 1.0E-06
 Mixing 0.20
 DIIS N=10 OK=0.500 CX=5.00 CXX=25.00 BFAC=0.00
 LShift 0.00
End

QMMM
 FORCE_FIELD_FILE sybyl.ff
 OUTPUT_LEVEL=1
 WARNING_LEVEL=-1
 ELSTAT_COUPLING_MODEL=0
 MM_CONNECTION_TABLE
    SAME AS IN ABOVE
 SUBEND
 LINK_BONDS
:: ---------- ------ ---
:: atoms alpha dummy
:: ---------- ------ ---
   22 - 3 1.2990 H
   23 - 3 1.2990 H
   24 - 8 1.3200 H
   25 - 9 1.3200 H
   27 - 10 1.3710 H
   30 - 14 1.3800 H
:: ---------- ------ ---
 SUBEND
END

ATOMS internal
 SAME AS IN ABOVE
END

GEOVAR
 SAME AS IN ABOVE
END
```

```
END INPUT
```

2.5 Examples in $ADFHOME/examples/adf

Shown here are QM/MM examples that are stored in the subdirectories under $ADFHOME/examples/adf, where $ADFHOME is the main directory of the ADF package. Note that the examples described here are also used to check that the program has been installed correctly, thus that technically the QM/MM functionality is working.

QMMM_Butane: Basic QMMM Illustration

Sample directory: adf/QMMM_Butane/

This example is a simple illustration of the QM/MM functionality: half of the butane molecule is treated quantum-mechanically, the other half by molecular mechanics.

```
$ADFBIN/adf << eor
Title  BUTANE in Z-matrix input
```

(Omitted in this printout: the usual specifications of fragments, symmetry, integration accuracy, -)

```
QMMM
   FORCEFIELD_FILE  $ADFRESOURCES/ForceFields/amber95.ff
   RESTART_FILE mm.restart
   OUTPUT_LEVEL=2
   WARNING_LEVEL=2
   ELSTAT_COUPLING_MODEL=0

   LINKS
     1 - 4        1.38000        H
   SUBEND

   MM_CONNECTION_TABLE
      1 CT   QM     2    3    4    5
      2 HC   QM    1
      3 HC   QM    1
      4 CT   LI    1    9   13   14
      5 CT   QM    1    6    7    8
      6 HC   QM    5
      7 HC   QM    5
      8 HC   QM    5
      9 CT   MM    4   10   11   12
     10 HC   MM    9
     11 HC   MM    9
     12 HC   MM    9
     13 HC   MM    4
     14 HC   MM    4
   SUBEND

End

Atoms    Internal
```

```
   C   0   0   0            0              0              0
   H   1   0   0            B1             0              0
   H   1   2   0            B2             A1             0
   C   1   2   3            B3             A2             D1
   C   1   2   3            B4             A3             D2
   H   5   1   2            B5             A4             D3
   H   5   1   6            B6             A5             D4
   H   5   1   6            B7             A6             D5
   C   4   1   2            B8             A7             D6
   H   9   4   1            B9             A8             D7
   H   9   4   10           B10            A9             D8
   H   9   4   10           B11            A10            D9
   H   4   1   9            B12            A11            D10
   H   4   1   9            B13            A12            D11
   End

   GeoVar
   ....
```

In the QMMM key block, the MM connection table identifies the atoms as belonging to either the QM (quantum mechanics) part, or the MM (molecular mechanics) part, or to the set of LI (link) atoms, which define the connection between the QM and the MM regions. Order and numbering are one-to-one with the list under the Atoms key.

The Link atom, part of the MM section of the system, is associated with a *capping atom*, in the QM part of the system. The Links subkey block specifies for each LI atom defined under the MM_Connection_Table subkey block the chemical type of the replacing capping atom (here: H). On the same line we find the ratio of the QM atom LI atom distance to the QM atom capping atom distance (here: 1.38), and the numbers (1 and 4) of the involved QM atom and LI atom.

The other subkeys in the QM key block are simple subkeys. The specify the file with the force field parameters to be used in the MM subsystem, the (restart) file to write MM data to, print and warning levels and a code for the electrostatic coupling model to use. See the rest of the QM/MM manual for a detailed discussion of all options.

The calculation is a simple geometry optimization (the Geometry key is not displayed here, but is contained in the full input). This consists of a repeated two-step process. At the first step, the MM system is kept frozen, the SCF equations are solved for the QM system, where potentials resulting from the MM system are included, and gradients on the QM atoms are computed from the SCF solution. At the second step, the QM system's geometry is updated and then kept frozen while the MM system's geometry is optimized (converged) for that particular QM configuration. And so on, until the whole combined system is self-consistently converged.

QMMM_CYT

Sample directory: adf/QMMM_CYT/

See the rest of the QM/MM manual , where this case is used as a 'walk through' for the QMMM feature.

It is a more or less straightforward application of QM/MM to geometry optimization (Cytocine). In the Atoms block all atoms are listed (QM as well as MM). All QM/MM aspects, such as which atoms belong to the QM core and which are to be treated by the approximate MM method, are found in the QMMM key block, and its various subkey blocks. The remainder of the input file is not different from what it would be in a non-QM/MM run.

The standard amber95 force field is used, which is located in the database of the ADF distribution.

```
$ADFBIN/adf << eor
Title  CYT amber95 - Cartesian Geometry Optimization

Fragments
  C    t21.C
  H    t21.H
End

Charge  0  0

Atoms Cartesian
   1 C        1.94807    3.58290   -0.58162
   2 C        1.94191    3.61595    1.09448
   3 H        1.69949    4.49893   -1.05273
   4 H        2.99455    3.17964   -0.86304
   5 C        0.94659    2.40054   -0.92364
   6 N       -1.74397   -3.46417    0.31178
   7 C       -1.00720   -2.20758    0.33536
   8 C       -1.66928   -1.00652    0.31001
   9 C       -0.92847    0.25653    0.34895
  10 N        0.43971    0.26735    0.38232
  11 N        0.36409   -2.20477    0.28992
  12 C        1.09714   -0.95413    0.22469
  13 H       -2.89781   -3.50815    0.31746
  14 H       -1.21484   -4.49217    0.31721
  15 H       -2.80940   -0.93497    0.30550
  16 H       -1.55324    1.21497    0.33885
  17 C        1.23309    1.44017    0.30994
  18 O        2.58277   -1.01636    0.23914
  19 H        2.37276    1.25557    0.29984
  20 O        1.02358    2.43085    1.50880
  21 H        1.17136    1.95097   -1.87367
  22 H       -0.10600    2.77333   -0.80348
  23 H        1.62170    4.54039    1.51392
  24 H        2.99608    3.28749    1.41345
End

QMMM
   FORCEFIELD_FILE   $ADFRESOURCES/ForceFields/amber95.ff
   RESTART_FILE mm.restart
   OUTPUT_LEVEL=1
   WARNING_LEVEL=2
   ELSTAT_COUPLING_MODEL=1

   LINK_BONDS
     1 - 5       1.38000       H
     1 - 2       1.38030       H
   SUBEND

   MM_CONNECTION_TABLE
     1    CT    QM     2    3    4    5
     2    CT    LI     1   20   23   24
     3    HC    QM     1
     4    HC    QM     1
     5    CT    LI     1   17   21   22
```

```
    6    N2     MM      7  13  14
    7    CA     MM      6   8  11
    8    CM     MM      7   9  15
    9    CM     MM      8  10  16
   10    N*     MM      9  12  17
   11    NC     MM      7  12
   12    C      MM     10  11  18
   13    H      MM      6
   14    H      MM      6
   15    HA     MM      8
   16    H4     MM      9
   17    CT     MM      5  10  19  20
   18    O      MM     12
   19    H2     MM     17
   20    OS     MM      2  17
   21    HC     MM      5
   22    HC     MM      5
   23    H1     MM      2
   24    H1     MM      2
  SUBEND

  CHARGES
   1   0.0   CT
   2   0.0   CT
   3   0.0   HC
   4   0.0   HC
   5   0.0   CT
   6  -0.9530  N2
   7   0.8185  CA
   8  -0.5215  CM
   9   0.0053  CM
  10  -0.0484  N*
  11  -0.7584  NC
  12   0.7538  C
  13   0.4234  H
  14   0.4234  H
  15   0.1928  HA
  16   0.1958  H4
  17   0.0066  CT
  18  -0.6252  O
  19   0.2902  H2
  20  -0.2033  OS
  21   0.0000  HC
  22   0.0000  HC
  23   0.0000  H1
  24   0.0000  H1
  SUBEND

END

Geometry
  Iterations   20
  Converge    E=1.0E-3   Grad=0.0005
  Step        Rad=0.3    Angle=5.0
End

XC
```

```
  LDA  VWN
  GGA  PostSCF Becke Perdew
End


Integration  3.0

SCF
  Iterations  60
  Converge  1.0E-06  1.0E-6
  Mixing  0.20
  DIIS  N=10  OK=0.500  CX=5.00  CXX=25.00  BFAC=0.00
End


End Input
eor
```

QMMM_Surface: Ziegler-Natta catalysis

Sample directory: adf/QMMM_Surface/

This is an example of a Ziegler-Natta type catalytic system: a TiCl complex embedded in a MgCl surface with two organic substrates also attached to the surface. To make the computation faster, the QM/MM approach is applied. The QM part includes only the active site and a piece of the MgCl surface.

The computation is formally a geometry optimization, but to keep the sample doable in a reasonable time the sample performs only one geometry update step. In the optimization, all of the MgCl surface atoms are frozen.

The standard force field has been modified to accommodate this calculation. The modified force field file is part of the sample run script. In this modified file, bonds are defined between Mg-Cl atoms in the MM connection table. This results in some torsions where the atoms are collinear. To rectify this problem, the torsional potentials for these atoms are set to potential type '0' (no potential).

There are no capping atoms mediating the bonds between the QM and MM regions because the boundary goes through the MgCl surface, which is ionically bound.

```
cat << eor > champ_de_force.ff
YBYL/TRIPOS FORCE FIELD FILE FOR ADF QM/MM
MODIFIED WITH UFF1.01 FOR Si Mg Ti Cl
L. Petitjean 15.11.1999
************************************************************************
```

(Most of the contents of the modified force field file is omitted here. You quickly get the difference with the standard sybyl force field file in the ADF database by running a UNIX *diff* on the two files.

```
   =================================
eor


$ADFBIN/adf  << eor
Title  ADF-QMMM in a surface study
NoPrint SFO, Frag, Functions
```

```
! keywords for calculation methods and optimization
XC
 GGA     BLYP
End

Geometry
 Optim          Cartesian Selected
 Iterations     1
 HessUpd        BFGS
 Converge       e=1e-4 grad=1e-3 rad=1e-2
 Step           rad=0.15
END
```

The 'Iterations 1' subkey specification in the Geometry block specifies that only one step in the optimization is carried out.

```
Integration   3.0 3.0

SCF
 Iterations  250
 Converge    1E-6 1E-6
 Mixing      0.2
 DIIS        N=10 OK=0.5 cyc=5 CX=5.0 BFAC=0
End


! keywords for molecule specification
Charge 0 0

Atoms Cartesian
    1 Mg    x1    y1    z1
```

(all other atoms in the Atoms block omitted here)

```
End

GeoVar
   x1=.00000 F
   y1=.00000 F
   z1=.00000 F
   x2=.00000 F
   y2=1.72129 F
   z2=1.82068 F
   x3=.00000 F
   y3=.00000 F
   z3=-3.64100 F
   x4=.00000 F
   y4=-1.72130 F
   z4=-1.82068 F
   x5=.00000 F
   y5=1.72130 F
   z5=-1.82032 F
   x6=.00000 F
   y6=1.72130 F
   z6=-5.46132 F
   x7=2.53903
```

```
    y7=.03004
    z7=-3.50645
    x8=2.50628
    y8=-.07048
    z8=-.10022
    x9=2.63009
    y9=3.50093
    z9=-3.02634
...
```

Many of the coordinates have a 'F' after their initial value specification under Geovar, indicating that these coordinates will be kept frozen during optimization.

The remaining initial value specifications are omitted here.

```
END
QMMM
    OPTIMIZE
      MAX_STEPS 3000
      MAX_GRADIENT 0.01
      METHOD BFGS
      PRINT_CYCLES 100
    SUBEND

    FORCE_FIELD_FILE champ_de_force.ff
```

The local file 'champ_de_force.ff' is used as force field file. Of course, this is the file we've just set up in the run script.

```
    OUTPUT_LEVEL=1
    WARNING_LEVEL=1
    ELSTAT_COUPLING_MODEL=1

    MM_CONNECTION_TABLE
        1   Mg   QM    2    4    5    8    58    60
...
```

Contents of the MM_Connection_Table block is omitted.

```
    SUBEND
    CHARGES
      1    .957
      2   -.608
      3   1.017
      4   -.411
      5   -.561
...
```

Initial charges are specified for (all) the atoms. Whether or not the charges on the QM (and LI) atoms are used depends on the type of electrostatic coupling between the QM and MM system. See the rest of the QM/MM manual for details.

```
    SUBEND
END

Fragments
 Ti t21.Ti
```

```
 Cl t21.Cl
 Mg t21.Mg
 C  t21.C
 H  t21.H
End


End Input
eor
```

3 pdb2adf: transform PDB file to QM/MM input file

The pdb2adf utility program was written by Marcel Swart.

3.1 Overview

General description

Starting from the adf2005.01 version the utility pdb2adf is available in the official release. Previously this utility could be found on the contributed software page. Starting from adf2008.01 there is support for the NEWQMMM subkey if the environment variable SCM_PDB2ADF is set to NEW.

The pdb2adf utility was written to read a PDB file, which contains the atomic coordinates of a protein structure, and transform it into an ADF inputfile, particularly for use with QM/MM calculations. Starting from the current release it can also be used for setting up a solvent shell around a solute molecule.

The PDB files are generally used for protein structures, and are formatted according to certain rules, see: http://www.wwpdb.org/docs.html, and the part about the official PDB format below.

For every residue/molecule present in the PDB file, there should be a fragment file available, either in the general ADF library ($ADFRESOURCES/pdb2adf directory), or in the local directory where the pdb2adf program is being called. Fragment files in the local directory take higher priority than those in the general ADF library. The fragment files are formatted, based loosely on AMBER parameter files, and contain information about the residues; e.g., the atoms present, with their general and forcefield atomnames, atomic charges, connections to other atoms for creating their positions when not found on the PDB file, etc.; see part about fragment files below. Available in the ADF library are fragment files for amino acid residues, including those at the N- or C-terminal residue, three solvents (water, methanol, chloroform), some ions that are present frequently in protein structures (copper, fluoride), etc.

Also present in the ADF library are solvent box files that can be used to place a layer of solvents surrounding the protein, or a solute. Available are the three solvents mentioned above.

After reading the PDB and corresponding fragment files, the program tries to figure out which atoms are missing, and will add those; it uses the information provided on the fragment files to do so. For certain amino acid residues, there are several protonation states possible, e.g. histidine can be protonated at the N-delta position, at the N-epsilon position, or on both. The default option is to choose the fully charged option for aspartate (Asp), glutamate (Glu), lysine (Lys) residues, and decide for each histidine (His) and cysteine (Cys) residue individually what the protonation state should be. In those individual cases, the distances of neighboring molecules/residues are given that may help determine the protonation state. See the protein example below.

After all that is setup properly, a list is given with residue names/numbers, from which you can choose those that should be placed in the QM system; afterwards, for each of the selected QM residues, a choice should be made where to cut-off the QM part. The most appropriate point to cut-off seems to be at the C-alpha position, except when dealing with a proline (Pro). The latter residue is cyclic, e.g. the sidechain is connected to the C-alpha carbon ! For that residue, it may be better to include the C-alpha, H-alpha, and backbone carbonyl group of the preceding residue in the QM part.

The program will try to use to replace the ".pdb" extension of the PDB file by ".pdb2adf" for the ADF inputfile to be made; for convenience, the program also writes out an ".p2a.pdb" file with the complete system as it being made by the program. This file can then be visualized by conventional viewer programs (such as iMol, VMD, Molekel, ADFview) for visual inspection if everything has been carried out correctly.

Given below are two examples, one for the application of a protein, the other how to set up a solvent shell run.

Things to notice

- The current QM/MM implementation in ADF is limited to a total of 1000 QMMM atoms; currently, a new implementation is underway that is more flexible, and that doesn't have this limit. This new implementation is available with the NEWQMMM subkey, work in progress.
- The NEWQMMM format is used if the environment variable SCM_PDB2ADF is set to NEW.
- The pdb2adf program uses AMBER parameter files, and is setup to work with the AMBER force field, version AMBER95, which is designed for and works well for biosystems.
- For questions, remarks, contact: support@scm.com.

Official PDB format

Columns	Data Type	Field	Definition
1 - 6	Record name	'ATOM' or 'HETATM'	
7 - 11	Integer	serial	Atom serial number.
13 - 16	Atom	name	Atom name.
17	Character	altLoc	Alternate location indicator.
18 - 20	Residue name	resName	Residue name.
22	Character	chainID	Chain identifier.
23 - 26	Integer	resSeq	Residue sequence number.
27	AChar	iCode	Code for insertion of residues.
31 - 38	Real(8.3)	x	Orthogonal coordinates for X in Angstroms.
39 - 46	Real(8.3)	y	Orthogonal coordinates for Y in Angstroms.
47 - 54	Real(8.3)	z	Orthogonal coordinates for Z in Angstroms.
55 - 60	Real(6.2)	occupancy	Occupancy.
61 - 66	Real(6.2)	tempFactor	Temperature factor.
73 - 76	LString(4)	segID	Segment identifier, left-justified.
77 - 78	LString(2)	element	Element symbol, right-justified.
79 - 80	LString(2)	charge	Charge on the atom.

Typical examples from PDB-files:

```
          1         2         3         4         5         6         7         8
 12345678901234567890123456789012345678901234567890123456789012345678901234567890

 ATOM     76  O   GLY   A9       6.671  55.354  35.873  1.00 14.75      A
 ATOM     77  N   ASN   A10      6.876  53.257  36.629  1.00 16.09      A
 ATOM     62  O   GLY A   9      6.791  55.214  35.719  1.00 15.61      4AZU 153
 ATOM     63  N   ASN A  10      6.892  53.135  36.555  1.00 12.64      4AZU 154
```

The pdb2adf utility is flexible, and should be able to read most PDB files, even those with incomplete or erroneous line formats. From every ATOM/HETATM line, it tries to read:

- atom number
- atom name

- residuename
- chain identifier
- residue number
- X,Y,Z coordinates

Hints for proper formatting:

- always group together atoms that belong to one residue
- always give the atom name on columns 13-16
- when specifying a chain-id use only letters (or a blank)

Contents of fragment file

Given below is the contents of the fragment file for water. The first line is a comment line, the only important parameter is the NOCONNECT keyword, which indicates that the program should not try to make any connections to other residues/molecules. Then follow three lines, that define the orientation in space of the residue; they are not used for general fragments, but are relevant and important for amino acid residues and DNA nucleotides. Finally, for each atom in the molecule, there should be a line with its number in the fragment; its name to be used in PDB files; the AMBER forcefield atomtype; a dummy atomname; connections and coordinates (bond, angle, dihedral angle) to other atoms in the molecule that can be used to give the position of the atom if it is not present in the PDB file; the atomic charge; and after the exclamation mark (!) the connections to other atoms in this fragment, or other fragments in case of amino acid residues/DNA nucleotides. The current version does not use the latter connections yet, but the next version will probably use them.

```
HOH   Water molecule   NOCONNECT
   1   DUMM   DU    M      0   0   0      0.0000      0.0000      0.0000
   2   DUMM   DU    M      1   0   0      1.4490      0.0000      0.0000
   3   DUMM   DU    M      2   1   0      1.5220    111.1000      0.0000
   4   O      OW    O      0   0   0      0.0000      0.0000      0.0000   -0.8340  !  5  6
   5   H1     HW    H      4   0   0      0.9572      0.0000      0.0000    0.4170  !  4
   6   H2     HW    H      4   5   0      0.9572    104.5200      0.0000    0.4170  !  4
```

Contents of solvent box files

The first line is a comment line, followed by a line with the total number of atoms in the solvent box and the dimensions of the box (in Angstroms); then for each atom in the box, the atom name, which must match the PDB atomname, and the Cartesian coordinates, again in Angstroms.

3.2 Usage of pdb2adf

Short description

The program works interactively, and should be straightforwardly to use. However, for some of the stages in the output a short description is given below.

```
                    P D B 2 A D F - program
                         version 2005.01
                    Written by: Marcel Swart, 2005

                 This program uses AMBER parameter files
                    see: http://amber.scripps.edu
```

```
Do you want a logfile to be written (Y/n) ?
```

This option exists to create a logfile of what pdb2adf does. However, it should normally be used only for debugging purposes.

```
Ignoring atom on line:
ATOM    974   OH LYS A 128     -10.073 42.775 15.690 1.00 38.79    5AZU1065
```

This is a warning that the atom on that particular line is ignored, should normally occur only few times (less than ten). Depends also on how well the PDB file follows the PDB format rules.

```
Data Processed:
    Nat:       2519
    Nmol:       196
 NChains:         1
```

Information about what has been read on the PDB file: the total number of atoms (Nat), number of molecules/residues (Nmol) and number of protein chains (Nchains).

```
Please wait, making connection tables
```

At this point, the connections between the atoms are being made by looking at atom distances. It may take a while, depending on the size of the system.

```
Do you want to make separate files for each chain (Y/n) ?
```

You have the option to make different inputfiles for different protein chains, but you can also make one inputfile for all of them together.

```
Found the following terminal amino acid residues : (C-term) 128 (N-term) 1
Do you want to use these as terminal residues (Y/n) ?
```

Info is given about the C- and N-terminal residue of each chain. Reported for making sure they are chosen correctly. Note, if the C- and N-terminal residues are connected (rarely the case probably), enter N here.

```
Multiple AMBER options for HIS :
  0          Decide every time differently
  1   HID   Histidine Delta Hydrogen
  2   HIE   Histidine Epsilon Hydrogen
  3   HIP   Histidine E & D Hydrogens

Suggested option: 0
```

For a number of residues (His, Glu, Asp, Lys and Cys) there is more than one option available in the AMBER95 force field, depending on the protonation state (His, Glu, Asp and Lys) or the existence of a sulphur bridge/connection to a metal atom (Cys). The default is to choose a different option for the His and Cys residues, and use one option for Glu, Asp and Lys (fully charged). However, if wanted you can make a choice for all residues.

```
Multiple AMBER options for CYS   3 (    3) :
  1   CYS   Cysteine (SH)
  2   CYM   Deprotonated Cysteine (S-)
  3   CYX   Cystine (S-S bridge)

  Connections and Nearest Atoms for SG CYS   3 SG  ( P2A #   41 PDB#   20 )
    Dist  P2A Nr  PDB Nr  Label              Near    Dist  P2A Nr  PDB Nr  Label
1   1.82      38      19  CB  CYS    3  CB      1    3.79    2382     980  O   HOH  151  O
2   2.02     461     193  SG  CYS   26  SG      2    3.80      22       0  HC  GLN    2
                                               3    4.04    2391     983  O   HOH  154  O
                                               4    4.15     509     206  O   GLN   28  O
                                               5    4.18     522       0  HA  PHE   29
Suggestion: 3
```

The options for Cys3 are given, with information about the atoms bonded to the SG sulphur atom (on the left), as well as the closest five non-bonded atoms (on the right). This information may help you decide which choice to make for this particular residue. Also given (on the bottom) is the suggested choice, which is based, in this case, on the presence of a sulphur bridge.

```
Multiple AMBER options for HIS   46 (   46) :
  1   HID   Histidine Delta Hydrogen
  2   HIE   Histidine Epsilon Hydrogen
  3   HIP   Histidine E & D Hydrogens

 Connections and Nearest Atoms for ND HIS   46 ND1 ( P2A #  844 PDB#  347 )
     Dist  P2A Nr  PDB Nr  Label              Near    Dist  P2A Nr  PDB Nr  Label
1   1.37    843     346  CG  HIS   46  CG     1      2.62   2166      0  H1  MET  121
2   1.33    846     349  CE  HIS   46  CE1    2      3.23   2080    863  ND  HIS  117  ND1
3   2.04   2318     959  CU  CU   130  CU     3 HB   3.33   2163    900  S   MET  121  SD
                                             4      3.40   2164    901  CT  MET  121  CE
                                             5      3.57   2082    865  CE  HIS  117  CE1

 Connections and Nearest Atoms for NE HIS   46 NE2 ( P2A #  848 PDB#  350 )
     Dist  P2A Nr  PDB Nr  Label              Near    Dist  P2A Nr  PDB Nr  Label
1   1.32    846     349  CE  HIS   46  CE1    1 HB   2.70    162     67  O   ASN   10  O
2   1.37    850     348  CD  HIS   46  CD2    2      2.83    814      0  H1  MET   44
                                             3      3.23   2166      0  H1  MET  121
                                             4      3.52    822    332  O   MET   44  O
                                             5      3.74    813    334  CT  MET   44  CG
Suggestion: 2
```

For His residues, the information is given for both the delta- and the epsilon nitrogen atoms. Also indicated (by HB) is the presence of a hydrogen bond with another atom. The definition used here is that two atoms are hydrogen bonded if they are both non-carbon/non-hydrogen atoms, and the distance between them is less than the sum of the van der Waals radii of the atoms. It is a simple definition, but seems to be effective. In this case, as the N(delta) is bonded to copper, the proton should be attached to the N(epsilon).

```
Making choice for which molecules should be QM, which MM
```

Now we come to the part where the division in the QM and MM systems is made.

```
Residues belonging to chain  0
Option  Molecule    Option  Molecule     Option  Molecule    Option  Molecule     Option  Molecule
   1: ALA    1         28: GLN   28         55: ASP   55        82: ALA   82         109: ALA  109
   2: GLN    2         29: PHE   29         56: LYS   56        83: HIS   83         110: TYR  110
etc
```

All molecules/residues belonging to chain 0 are given, with an option number.

```
Give option number of molecules to be put in QM region (or 'c' to continue):
Note: by specifying a negative number a molecule is removed from the QM region
```

Here you are asked to enter the option numbers of the residues you want to put in the QM system.

```
Putting GLY   45 in QM region
Putting HIS   46 in QM region
```

In this case, Gly45 and His46 have been put in the QM system.

```
Make a choice for the QM/MM treatment of GLY   45
 0:  Put completely in QM region
 1:  Cut off at C-alpha (put NH in QM region, CO in MM region)
 2:  Cut off at C-alpha (put NH in MM region, CO in QM region)
 3:  Cut off at C-alpha (put NH and CO in MM region)
 4:  Cut off at C-alpha (put NH and CO in QM region,  sidechain in MM region)
 5:  Put only part of sidechain in QM region

Suggestion: 2
Give choice:
```

A choice should be made for where to cut-off the QM system. Normally this is done at the C(alpha) position, and you should simply choose the Suggestion.

```
Solvent molecules (SOL/HOH) belonging to this chain:
   1    2    3    4    5    6    7    8    9   10   11   12   13   14   15   16   17   18   19
20
  21   22   23   24   25   26   27   28   29   30   31   32   33   34   35   36   37   38   39
40
  41   42   43   44   45   46   47   48   49   50   51   52   53   54   55   56   57   58   59
60
  61   62   63   64   65   66

Give the number of the molecule to be put in QM region (or 'c' to continue):
```

Also water molecules can be put in the QM system.

```
Box Shape options:
 1  Spherical box
 2  Cubic box
Make a choice:
```

Type of box to be used.

```
Maximum atomic distance (Angs) from center        25.62
Give boxsize (def.:    28.62 Angs)
```

Size of box to be used to put a layer of solvent molecules around the system. Max. dist. is the maximal distance of any protein atom from the center of mass of the protein. Usually you should choose a boxsize at least 6 Angstrom larger (so at least two solvent molecules are surrounding the system).

```
Using BOXSIZE value of   30.0000
Adding atoms for box    1  Added (Box):     0 (Total):     0  Excl. (1):   648  Excl. (2):     0
Adding atoms for box    2  Added (Box):     9 (Total):     9  Excl. (1):   639  Excl. (2):     0
...
Adding atoms for box   63  Added (Box):     3 (Total): 7635  Excl. (1):   645  Excl. (2):     0
Adding atoms for box   64  Added (Box):     0 (Total): 7635  Excl. (1):   648  Excl. (2):     0
Writing inputfile for chain   1
```

A total amount of 7635 atoms (2545 water molecules) has been added.

```
Inputfile(s) written, everything processed, work has been done.
Thank you for using the PDB2ADF program.


================================
Normal ending of PDB2ADF program
================================
```

ADF inputfile(s) have been written, the PDB-file has been processes. Everything is done.

An example on protein structure

The idea of this example is to make an adf-input file using a PDB of an azurin (1DYZ.pdb). The result of this example should be that in the adf-input file the active site of azurin (Figure 1) is in the QM part, and the rest of the protein is in the MM part, and that the solvent water is added (in a box), which is also in the MM part.
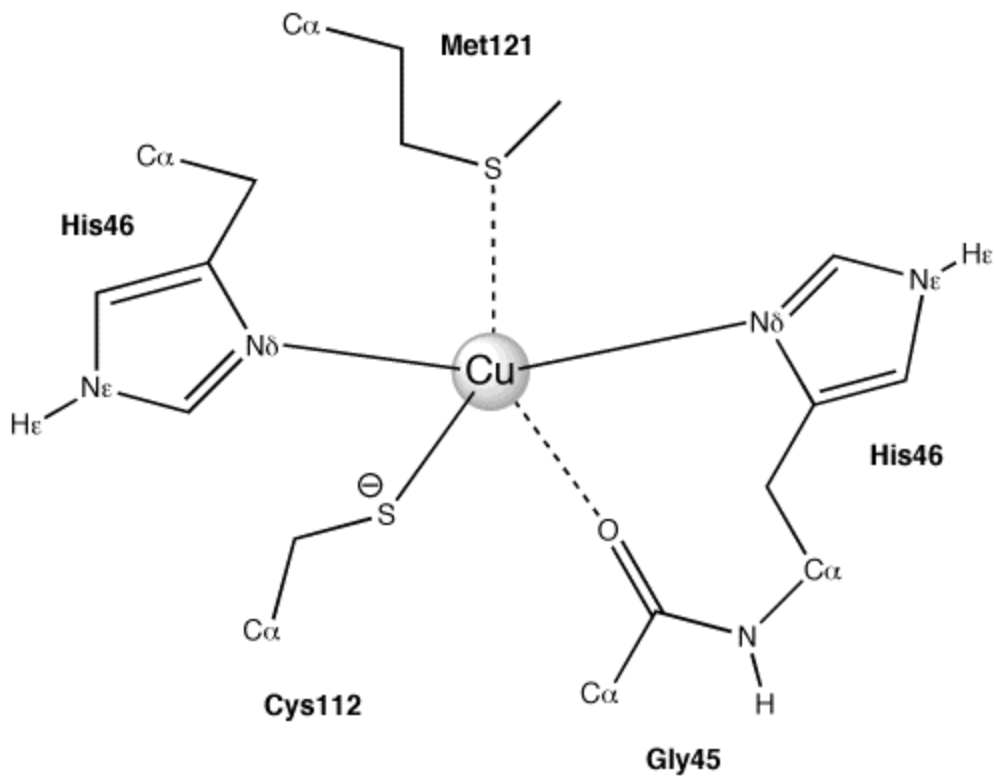
Figure 1: the active site of azurin

Usage of pdb2adf

The program works interactively. Given below in **bold** are the parts that the user has to type. In cases where the user agrees with the suggestion given by the program, the user can press the `Enter` key indicated with **Enter**.

```
              P D B 2 A D F - program
                  version 2005.01
              Written by: Marcel Swart, 2005

            This program uses AMBER parameter files
                see: http://amber.scripps.edu
```

Do you want a logfile to be written (Y/n) ?

Enter

Please give name of PDB-file

1DYZ.pdb

 read fragments

Data Processed:
     Nat: 2519

```
   Nmol:        196
 NChains:         1

Please wait, making connection tables
Now finding nearby atoms
Assigning chain ID to all residues
Completing residues for which only option is available

Found the following terminal amino acid residues : (C-term)   129 (N-term)
1
Do you want to use these as terminal residues (Y/n) ?
```

Enter

```
Refinding nearby atoms (including atoms added in residue completion)

Multiple AMBER options for HIS :
  0        Decide every time differently
  1   HID  Histidine Delta Hydrogen
  2   HIE  Histidine Epsilon Hydrogen
  3   HIP  Histidine E & D Hydrogens

Suggested option: 0
```

Enter

```
Using 0: Decide every time differently

Multiple AMBER options for GLU :
  0        Decide every time differently
  1   GLU  Glutamic acid (COO-)
  2   GLH  Neutral Glutamic acid (COOH)

Suggested option: 1
```

Enter

```
Using 17   GLU   Glutamic acid (COO-)

Multiple AMBER options for ASP :
  0        Decide every time differently
  1   ASP  Aspartic acid (COO-)
  2   ASH  Neutral Aspartatic acid (COOH)

Suggested option: 1
```

Enter

```
Using 18   ASP   Aspartic acid (COO-)

Multiple AMBER options for LYS :
  0        Decide every time differently
  1   LYS  Charged Lysine (NH3+)
  2   LYN  Neutral Lysine (NH2)
```

```
Suggested option: 1


Enter


Using 19   LYS   Charged Lysine (NH3+)

Multiple AMBER options for CYS :
  0          Decide every time differently
  1   CYS   Cysteine (SH)
  2   CYM   Deprotonated Cysteine (S-)
  3   CYX   Cystine (S-S bridge)


Suggested option: 0


Enter


Using 0: Decide every time differently
- - - - - - - - - - - - - - - - - - - - - - - - - - - -
         Making Choices for Chain      0
- - - - - - - - - - - - - - - - - - - - - - - - - - - -

Multiple AMBER options for CYS    3 (    3) :
  1   CYS   Cysteine (SH)
  2   CYM   Deprotonated Cysteine (S-)
  3   CYX   Cystine (S-S bridge)

  Connections and Nearest Atoms for SG CYS    3 SG  ( P2A #   41 PDB#   20 )
     Dist  P2A Nr  PDB Nr  Label            Near    Dist  P2A Nr  PDB Nr  Label
1   1.82     38      19  CB  CYS    3  CB     1   3.79   2382     980  O   HOH  151  O
2   2.02    461     193  SG  CYS   26  SG     2   3.80     22       0  HC  GLN    2
                                             3   4.04   2391     983  O   HOH  154  O
                                             4   4.15    509     206  O   GLN   28  O
                                             5   4.18    522       0  HA  PHE   29
Suggestion: 3


Enter


Multiple AMBER options for CYS   26 (   26) :
  1   CYS   Cysteine (SH)
  2   CYM   Deprotonated Cysteine (S-)
  3   CYX   Cystine (S-S bridge)

  Connections and Nearest Atoms for SG CYS   26 SG  ( P2A #  461 PDB#  193 )
     Dist  P2A Nr  PDB Nr  Label            Near    Dist  P2A Nr  PDB Nr  Label
1   1.82    458     192  CB  CYS   26  CB     1   3.41    522       0  HA  PHE   29
2   2.02     41      20  SG  CYS    3  SG     2   3.43    411     168  O   ASP   23  O
                                             3   3.60   2322     960  O   HOH  131  O
                                             4   3.91    403     169  CB  ASP   23  CB
                                             5   4.15    387       0  HC  VAL   22
Suggestion: 3


Enter


Multiple AMBER options for HIS   32 (   32) :
  1   HID   Histidine Delta Hydrogen
  2   HIE   Histidine Epsilon Hydrogen
  3   HIP   Histidine E & D Hydrogens

  Connections and Nearest Atoms for ND HIS   32 ND1 ( P2A #  581 PDB#  244 )
     Dist  P2A Nr  PDB Nr  Label            Near    Dist  P2A Nr  PDB Nr  Label
1   1.39    580     243  CG  HIS   32  CG     1   3.41    545       0  HC  THR   30
2   1.33    583     246  CE  HIS   32  CE1    2   3.43     76      33  O   ALA    5  O
                                             3   3.58     90      40  OH  THR    6  OG1
                                             4   3.99     91       0  HO  THR    6
                                             5   4.17     68       0  H   ALA    5

  Connections and Nearest Atoms for NE HIS   32 NE2 ( P2A #  585 PDB#  247 )
     Dist  P2A Nr  PDB Nr  Label            Near    Dist  P2A Nr  PDB Nr  Label
1   1.31    583     246  CE  HIS   32  CE1    1   2.86    544       0  HC  THR   30
2   1.37    587     245  CD  HIS   32  CD2    2   3.00    545       0  HC  THR   30
                                             3   3.14   1677       0  HO  SER   94
                                             4   3.42    542     229  CT  THR   30  CG2
```

```
                                                        5      3.65     1676      688  OH  SER   94  OG
Suggestion: 1
```

3

```
Multiple AMBER options for HIS   35 (   35) :
  1   HID   Histidine Delta Hydrogen
  2   HIE   Histidine Epsilon Hydrogen
  3   HIP   Histidine E & D Hydrogens

  Connections and Nearest Atoms for ND HIS   35 ND1 ( P2A #  649 PDB#  271 )
     Dist  P2A Nr  PDB Nr  Label                 Near   Dist  P2A Nr  PDB Nr  Label
1    1.38     648     270  CG  HIS   35  CG       1      2.46     682       0  H   GLY   37
2    1.32     651     273  CE  HIS   35  CE1      2      2.69    1604       0  H1  GLY   89
                                                  3      3.31     681     282  N   GLY   37  N
                                                  4      3.56    1602     653  CT  GLY   89  CA
                                                  5      3.67     152       0  H1  ASN   10

  Connections and Nearest Atoms for NE HIS   35 NE2 ( P2A #  653 PDB#  274 )
     Dist  P2A Nr  PDB Nr  Label                 Near   Dist  P2A Nr  PDB Nr  Label
1    1.33     651     273  CE  HIS   35  CE1      1 HB   2.91     822     332  O   MET   44  O
2    1.37     655     272  CD  HIS   35  CD2      2      3.24     814       0  H1  MET   44
                                                  3      3.24     850     348  CD  HIS   46  CD2
                                                  4      3.34    1593       0  H1  GLY   88
                                                  5      3.75     848     350  NE  HIS   46  NE2
Suggestion: 2
```

3

```
Multiple AMBER options for HIS   46 (   46) :
  1   HID   Histidine Delta Hydrogen
  2   HIE   Histidine Epsilon Hydrogen
  3   HIP   Histidine E & D Hydrogens

  Connections and Nearest Atoms for ND HIS   46 ND1 ( P2A #  844 PDB#  347 )
     Dist  P2A Nr  PDB Nr  Label                 Near   Dist  P2A Nr  PDB Nr  Label
1    1.37     843     346  CG  HIS   46  CG       1      2.62    2166       0  H1  MET  121
2    1.33     846     349  CE  HIS   46  CE1      2      3.23    2080     863  ND  HIS  117  ND1
3    2.04    2318     959  CU  CU   130  CU       3 HB   3.33    2163     900  S   MET  121  SD
                                                  4      3.40    2164     901  CT  MET  121  CE
                                                  5      3.57    2082     865  CE  HIS  117  CE1

  Connections and Nearest Atoms for NE HIS   46 NE2 ( P2A #  848 PDB#  350 )
     Dist  P2A Nr  PDB Nr  Label                 Near   Dist  P2A Nr  PDB Nr  Label
1    1.32     846     349  CE  HIS   46  CE1      1 HB   2.70     162      67  O   ASN   10  O
2    1.37     850     348  CD  HIS   46  CD2      2      2.83     814       0  H1  MET   44
                                                  3      3.23    2166       0  H1  MET  121
                                                  4      3.52     822     332  O   MET   44  O
                                                  5      3.74     813     334  CT  MET   44  CG
Suggestion: 2
```

Enter

```
Multiple AMBER options for HIS   83 (   83) :
  1   HID   Histidine Delta Hydrogen
  2   HIE   Histidine Epsilon Hydrogen
  3   HIP   Histidine E & D Hydrogens

  Connections and Nearest Atoms for ND HIS   83 ND1 ( P2A # 1494 PDB#  613 )
     Dist  P2A Nr  PDB Nr  Label                 Near   Dist  P2A Nr  PDB Nr  Label
1    1.39    1493     612  CG  HIS   83  CG       1      2.67    1317       0  HC  VAL   73
2    1.33    1496     615  CE  HIS   83  CE1      2      3.63    1315     542  CT  VAL   73  CG2
                                                  3      3.74    1310       0  HC  VAL   73
                                                  4      3.82    1316       0  HC  VAL   73
                                                  5      3.86    1313       0  HC  VAL   73

  Connections and Nearest Atoms for NE HIS   83 NE2 ( P2A # 1498 PDB#  616 )
     Dist  P2A Nr  PDB Nr  Label                 Near   Dist  P2A Nr  PDB Nr  Label
1    1.32    1496     615  CE  HIS   83  CE1      1      3.09    1313       0  HC  VAL   73
2    1.38    1500     614  CD  HIS   83  CD2      2      3.44    1317       0  HC  VAL   73
                                                  3      3.88    2385     981  O   HOH  152  O
                                                  4      3.93    1311     541  CT  VAL   73  CG1
                                                  5      4.03    1309     540  CT  VAL   73  CB
Suggestion: 2
```

3

```
Multiple AMBER options for CYS  112 (  112) :
  1   CYS   Cysteine (SH)
  2   CYM   Deprotonated Cysteine (S-)
```

```
    3   CYX   Cystine (S-S bridge)

  Connections and Nearest Atoms for SG CYS  112 SG  ( P2A # 2001 PDB#  828 )
      Dist  P2A Nr  PDB Nr  Label               Near   Dist  P2A Nr  PDB Nr  Label
1   1.82   1998     827  CB  CYS  112  CB        1   2.53    858      0  H   ASN   47
2   2.14   2318     959  CU  CU   130  CU        2   2.65   2023      0  H   PHE  114
                                                 3   3.00   2028      0  HC  PHE  114
                                                 4   3.29    868      0  H   ASN   47
                                                 5   3.39   2027      0  HC  PHE  114
Suggestion: 2
```

Enter

```
Multiple AMBER options for HIS  117 (  117) :
    1   HID   Histidine Delta Hydrogen
    2   HIE   Histidine Epsilon Hydrogen
    3   HIP   Histidine E & D Hydrogens

  Connections and Nearest Atoms for ND HIS  117 ND1 ( P2A # 2080 PDB#  863 )
      Dist  P2A Nr  PDB Nr  Label               Near   Dist  P2A Nr  PDB Nr  Label
1   1.37   2079     862  CG  HIS  117  CG        1   2.82   2028      0  HC  PHE  114
2   1.34   2082     865  CE  HIS  117  CE1       2   3.23    844    347  ND  HIS   46  ND1
3   1.99   2318     959  CU  CU   130  CU        3   3.26   2031      0  HA  PHE  114
                                                 4   3.27    832    340  O   GLY   45  O
                                                 5   3.43    846    349  CE  HIS   46  CE1

  Connections and Nearest Atoms for NE HIS  117 NE2 ( P2A # 2084 PDB#  866 )
      Dist  P2A Nr  PDB Nr  Label               Near   Dist  P2A Nr  PDB Nr  Label
1   1.31   2082     865  CE  HIS  117  CE1       1   2.57    209      0  H1  MET   13
2   1.37   2086     864  CD  HIS  117  CD2       2   2.65   2031      0  HA  PHE  114
                                                 3 HB 2.74   2406    988  O   HOH  159  O
                                                 4   3.34   2030    841  CA  PHE  114  CD1
                                                 5   3.41    204      0  H1  MET   13
Suggestion: 2
```

Enter

```
   - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
           Making Choices for Chain      1
   - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Completing residues with multiple options available, and solvent molecules

Checking positions of newly added atoms


Making choice for which molecules should be QM, which MM


Residues belonging to chain  0
Option  Molecule    Option  Molecule    Option  Molecule    Option  Molecule    Option  Molecule
    1: ALA    1       28: GLN   28       55: ASP   55       82: ALA   82      109: ALA  109
    2: GLN    2       29: PHE   29       56: LYS   56       83: HIS   83      110: TYR  110
    3: CYS    3       30: THR   30       57: GLN   57       84: THR   84      111: PHE  111
    4: GLU    4       31: MET   31       58: ALA   58       85: LYS   85      112: CYS  112
    5: ALA    5       32: HIS   32       59: VAL   59       86: VAL   86      113: SER  113
    6: THR    6       33: LEU   33       60: ALA   60       87: ILE   87      114: PHE  114
    7: VAL    7       34: LYS   34       61: THR   61       88: GLY   88      115: PRO  115
    8: GLU    8       35: HIS   35       62: ASP   62       89: GLY   89      116: GLY  116
    9: SER    9       36: VAL   36       63: GLY   63       90: GLY   90      117: HIS  117
   10: ASN   10       37: GLY   37       64: MET   64       91: GLU   91      118: TRP  118
   11: ASP   11       38: LYS   38       65: GLY   65       92: SER   92      119: ALA  119
   12: ALA   12       39: MET   39       66: ALA   66       93: ASP   93      120: MET  120
   13: MET   13       40: ALA   40       67: GLY   67       94: SER   94      121: MET  121
   14: GLN   14       41: LYS   41       68: LEU   68       95: VAL   95      122: LYS  122
   15: TYR   15       42: VAL   42       69: ALA   69       96: THR   96      123: GLY  123
   16: ASN   16       43: ALA   43       70: GLN   70       97: PHE   97      124: THR  124
   17: VAL   17       44: MET   44       71: ASP   71       98: ASP   98      125: LEU  125
   18: LYS   18       45: GLY   45       72: TYR   72       99: VAL   99      126: LYS  126
   19: GLU   19       46: HIS   46       73: VAL   73      100: SER  100      127: LEU  127
   20: ILE   20       47: ASN   47       74: LYS   74      101: LYS  101      128: GLY  128
   21: VAL   21       48: LEU   48       75: ALA   75      102: ILE  102      129: SER  129
   22: VAL   22       49: VAL   49       76: GLY   76      103: ALA  103      130: CU   130
   23: ASP   23       50: LEU   50       77: ASP   77      104: ALA  104
   24: LYS   24       51: THR   51       78: THR   78      105: GLY  105
   25: SER   25       52: LYS   52       79: ARG   79      106: GLU  106
   26: CYS   26       53: ASP   53       80: VAL   80      107: ASN  107
   27: LYS   27       54: ALA   54       81: ILE   81      108: TYR  108

Give option number of molecules to be put in QM region (or 'c' to continue):
Note: by specifying a negative number a molecule is removed from the QM region
```

45 46 112 117 121 130

```
Putting GLY   45 in QM region
Putting HIS   46 in QM region
Putting CYS  112 in QM region
Putting HIS  117 in QM region
Putting MET  121 in QM region
Putting CU   130 in QM region


Give option number of molecules to be put in QM region (or 'c' to continue):
Note: by specifying a negative number a molecule is removed from the QM region
```

c

```
Make a choice for the QM/MM treatment of GLY   45
 0:  Put completely in QM region
 1:  Cut off at C-alpha (put NH in QM region, CO in MM region)
 2:  Cut off at C-alpha (put NH in MM region, CO in QM region)
 3:  Cut off at C-alpha (put NH and CO in MM region)
 4:  Cut off at C-alpha (put NH and CO in QM region,  sidechain in MM
region)
 5:  Put only part of sidechain in QM region

Suggestion: 2
Give choice:
```

Enter

```
Make a choice for the QM/MM treatment of HIS   46
 0:  Put completely in QM region
 1:  Cut off at C-alpha (put NH in QM region, CO in MM region)
 2:  Cut off at C-alpha (put NH in MM region, CO in QM region)
 3:  Cut off at C-alpha (put NH and CO in MM region)
 4:  Cut off at C-alpha (put NH and CO in QM region,  sidechain in MM region)
 5:  Put only part of sidechain in QM region

Suggestion: 1
Give choice:
```

Enter

```
Make a choice for the QM/MM treatment of CYS  112
 0:  Put completely in QM region
 1:  Cut off at C-alpha (put NH in QM region, CO in MM region)
 2:  Cut off at C-alpha (put NH in MM region, CO in QM region)
 3:  Cut off at C-alpha (put NH and CO in MM region)
 4:  Cut off at C-alpha (put NH and CO in QM region,  sidechain in MM region)
 5:  Put only part of sidechain in QM region

Suggestion: 3
Give choice:
```

Enter

```
Make a choice for the QM/MM treatment of HIS  117
 0:  Put completely in QM region
 1:  Cut off at C-alpha (put NH in QM region, CO in MM region)
 2:  Cut off at C-alpha (put NH in MM region, CO in QM region)
 3:  Cut off at C-alpha (put NH and CO in MM region)
 4:  Cut off at C-alpha (put NH and CO in QM region,  sidechain in MM region)
 5:  Put only part of sidechain in QM region

Suggestion: 3
Give choice:
```

Enter

```
Make a choice for the QM/MM treatment of MET  121
 0:  Put completely in QM region
 1:  Cut off at C-alpha (put NH in QM region, CO in MM region)
 2:  Cut off at C-alpha (put NH in MM region, CO in QM region)
 3:  Cut off at C-alpha (put NH and CO in MM region)
 4:  Cut off at C-alpha (put NH and CO in QM region,  sidechain in MM region)
 5:  Put only part of sidechain in QM region

Suggestion: 3
Give choice:
```

Enter

```
Make a choice for the QM/MM treatment of CU    130
 0:  Put completely in QM region
 1:  Put only part of molecule in QM region

Suggestion: 0
Give choice:
```

Enter

```
Total formal charge on molecule CU    130      2.0000

Solvent molecules (SOL/HOH) belonging to this chain:
   1    2    3    4    5    6    7    8    9   10   11   12   13   14   15   16   17   18   19
20
  21   22   23   24   25   26   27   28   29   30   31   32   33   34   35   36   37   38   39
40
  41   42   43   44   45   46   47   48   49   50   51   52   53   54   55   56   57   58   59
60
  61   62   63   64   65   66

Give the number of the molecule to be put in QM region (or 'c' to continue):
```

c

```
Residues belonging to chain  1

Do you want to add solvent to your system (Y/n) ?
```

Enter

```
Solvent (box) available:
```

```
   1:   HOH      HOH  Water molecule
   2:   MOH      MOH  Methanol molecule
   3:   CHL      CHL  Chloroform molecule

1


Reading contents of solvent box
p2abox.HOH

Box Shape options:
 1  Spherical box
 2  Cubic box
Make a choice:

1


Writing inputfile for chain   0

Using total charge   1.0 and total spin   1.0

Maximum atomic distance (Angs) from center        25.62
Give boxsize (def.:    28.62 Angs)

30.0

Using BOXSIZE value of   30.0000
Adding atoms for box    1  Added (Box):     0 (Total):      0  Excl. (1):   648  Excl. (2):     0
Adding atoms for box    2  Added (Box):     9 (Total):      9  Excl. (1):   639  Excl. (2):     0
Adding atoms for box    3  Added (Box):     3 (Total):     12  Excl. (1):   645  Excl. (2):     0
Adding atoms for box    4  Added (Box):     0 (Total):     12  Excl. (1):   648  Excl. (2):     0
Adding atoms for box    5  Added (Box):     6 (Total):     18  Excl. (1):   642  Excl. (2):     0
Adding atoms for box    6  Added (Box):   228 (Total):    246  Excl. (1):   420  Excl. (2):     0
Adding atoms for box    7  Added (Box):   219 (Total):    465  Excl. (1):   429  Excl. (2):     0
Adding atoms for box    8  Added (Box):     9 (Total):    474  Excl. (1):   639  Excl. (2):     0
Adding atoms for box    9  Added (Box):     0 (Total):    474  Excl. (1):   648  Excl. (2):     0
Adding atoms for box   10  Added (Box):   225 (Total):    699  Excl. (1):   423  Excl. (2):     0
Adding atoms for box   11  Added (Box):   216 (Total):    915  Excl. (1):   432  Excl. (2):     0
Adding atoms for box   12  Added (Box):     6 (Total):    921  Excl. (1):   642  Excl. (2):     0
Adding atoms for box   13  Added (Box):     0 (Total):    921  Excl. (1):   648  Excl. (2):     0
Adding atoms for box   14  Added (Box):     6 (Total):    927  Excl. (1):   642  Excl. (2):     0
Adding atoms for box   15  Added (Box):    12 (Total):    939  Excl. (1):   636  Excl. (2):     0
Adding atoms for box   16  Added (Box):     0 (Total):    939  Excl. (1):   648  Excl. (2):     0
Adding atoms for box   17  Added (Box):    12 (Total):    951  Excl. (1):   636  Excl. (2):     0
Adding atoms for box   18  Added (Box):   210 (Total):   1161  Excl. (1):   438  Excl. (2):     0
Adding atoms for box   19  Added (Box):   219 (Total):   1380  Excl. (1):   429  Excl. (2):     0
Adding atoms for box   20  Added (Box):     3 (Total):   1383  Excl. (1):   645  Excl. (2):     0
Adding atoms for box   21  Added (Box):   216 (Total):   1599  Excl. (1):   417  Excl. (2):    15
Adding atoms for box   22  Added (Box):   381 (Total):   1980  Excl. (1):     3  Excl. (2):   264
Adding atoms for box   23  Added (Box):   261 (Total):   2241  Excl. (1):     3  Excl. (2):   384
Adding atoms for box   24  Added (Box):   183 (Total):   2424  Excl. (1):   423  Excl. (2):    42
Adding atoms for box   25  Added (Box):   189 (Total):   2613  Excl. (1):   426  Excl. (2):    33
Adding atoms for box   26  Added (Box):   186 (Total):   2799  Excl. (1):     3  Excl. (2):   459
Adding atoms for box   27  Added (Box):   351 (Total):   3150  Excl. (1):     3  Excl. (2):   294
Adding atoms for box   28  Added (Box):   222 (Total):   3372  Excl. (1):   420  Excl. (2):     6
Adding atoms for box   29  Added (Box):     9 (Total):   3381  Excl. (1):   639  Excl. (2):     0
Adding atoms for box   30  Added (Box):   162 (Total):   3543  Excl. (1):   429  Excl. (2):    57
Adding atoms for box   31  Added (Box):   219 (Total):   3762  Excl. (1):   426  Excl. (2):     3
Adding atoms for box   32  Added (Box):     6 (Total):   3768  Excl. (1):   642  Excl. (2):     0
Adding atoms for box   33  Added (Box):     6 (Total):   3774  Excl. (1):   642  Excl. (2):     0
Adding atoms for box   34  Added (Box):   219 (Total):   3993  Excl. (1):   426  Excl. (2):     3
Adding atoms for box   35  Added (Box):   216 (Total):   4209  Excl. (1):   432  Excl. (2):     0
Adding atoms for box   36  Added (Box):     6 (Total):   4215  Excl. (1):   642  Excl. (2):     0
Adding atoms for box   37  Added (Box):   219 (Total):   4434  Excl. (1):   426  Excl. (2):     3
Adding atoms for box   38  Added (Box):   279 (Total):   4713  Excl. (1):     6  Excl. (2):   363
Adding atoms for box   39  Added (Box):   231 (Total):   4944  Excl. (1):     0  Excl. (2):   417
Adding atoms for box   40  Added (Box):   195 (Total):   5139  Excl. (1):   432  Excl. (2):    21
Adding atoms for box   41  Added (Box):   231 (Total):   5370  Excl. (1):   414  Excl. (2):     3
Adding atoms for box   42  Added (Box):   324 (Total):   5694  Excl. (1):     0  Excl. (2):   324
Adding atoms for box   43  Added (Box):   408 (Total):   6102  Excl. (1):     6  Excl. (2):   234
Adding atoms for box   44  Added (Box):   204 (Total):   6306  Excl. (1):   435  Excl. (2):     9
Adding atoms for box   45  Added (Box):     6 (Total):   6312  Excl. (1):   642  Excl. (2):     0
Adding atoms for box   46  Added (Box):   177 (Total):   6489  Excl. (1):   435  Excl. (2):    36
Adding atoms for box   47  Added (Box):   219 (Total):   6708  Excl. (1):   429  Excl. (2):     0
```

```
Adding atoms for box    48  Added (Box):    6 (Total):  6714  Excl. (1):   642  Excl. (2):    0
Adding atoms for box    49  Added (Box):    0 (Total):  6714  Excl. (1):   648  Excl. (2):    0
Adding atoms for box    50  Added (Box):    3 (Total):  6717  Excl. (1):   645  Excl. (2):    0
Adding atoms for box    51  Added (Box):    6 (Total):  6723  Excl. (1):   642  Excl. (2):    0
Adding atoms for box    52  Added (Box):    0 (Total):  6723  Excl. (1):   648  Excl. (2):    0
Adding atoms for box    53  Added (Box):    9 (Total):  6732  Excl. (1):   639  Excl. (2):    0
Adding atoms for box    54  Added (Box):  222 (Total):  6954  Excl. (1):   426  Excl. (2):    0
Adding atoms for box    55  Added (Box):  213 (Total):  7167  Excl. (1):   426  Excl. (2):    9
Adding atoms for box    56  Added (Box):    6 (Total):  7173  Excl. (1):   642  Excl. (2):    0
Adding atoms for box    57  Added (Box):    3 (Total):  7176  Excl. (1):   645  Excl. (2):    0
Adding atoms for box    58  Added (Box):  219 (Total):  7395  Excl. (1):   423  Excl. (2):    6
Adding atoms for box    59  Added (Box):  219 (Total):  7614  Excl. (1):   429  Excl. (2):    0
Adding atoms for box    60  Added (Box):    6 (Total):  7620  Excl. (1):   642  Excl. (2):    0
Adding atoms for box    61  Added (Box):    0 (Total):  7620  Excl. (1):   648  Excl. (2):    0
Adding atoms for box    62  Added (Box):   12 (Total):  7632  Excl. (1):   636  Excl. (2):    0
Adding atoms for box    63  Added (Box):    3 (Total):  7635  Excl. (1):   645  Excl. (2):    0
Adding atoms for box    64  Added (Box):    0 (Total):  7635  Excl. (1):   648  Excl. (2):    0
Writing inputfile for chain    1
There are no atoms in this chain, ignoring it

Inputfile(s) written, everything processed, work has been done.
Thank you for using the PDB2ADF program.

==============================
Normal ending of PDB2ADF program
==============================
```

Contents of the 1DYZ.pdb2adf file generated by pdb2adf

The file is not given completely, since it contains more than 9000 atoms.

```
#! /bin/sh

$ADFBIN/adf << eor
TITLE QM/MM calculation setup by pdb2adf: M.Swart, 2005

Symmetry NOSYM

EPRINT
  SFO NOEIG NOOVL
END

XC
 GGA Becke-Perdew
END

GEOMETRY
 CONVERGE grad=1.0e-3 rad=1.0e-1
END

BASIS
 type TZP
 core small
END

SCF
 Converge 1.0e-5 1.0e-5
 Iterations 99
END

INTEGRATION 5.0 5.0 5.0

CHARGE   1.0  1.0
UNRESTRICTED
```

```
ATOMS
    1  C      11.3760     8.5410     29.7530
    2  H      10.9114     9.3322     30.3413
    3  H      12.4602     8.6423     29.8009
    4  C      10.9630     8.7450     28.3090
    5  O      10.8510     7.7910     27.5300
    6  N      10.6890     9.9800     27.9260
    7  H      10.7572    10.7382     28.5898
    8  C      10.2900    10.2500     26.5530
    9  H      10.5517     9.3991     25.9240
   10  C       8.7770    10.5120     26.4440
   11  H       8.5050    11.3473     27.0893
   12  H       8.5229    10.7532     25.4118
   13  C       7.9110     9.3590     26.8430
   14  N       8.0710     8.0910     26.3490
   15  C       7.1230     7.3010     26.8370
   16  H       7.0894     6.2496     26.5516
   17  N       6.3580     8.0230     27.6330
   18  H       5.5568     7.6742     28.1395
   19  C       6.8210     9.3110     27.6620
   20  H       6.3141    10.0588     28.2719
   21  C      11.0290     8.8020     20.9600
   22  H      11.3902     9.8061     21.1823
   23  C      10.0620     8.3640     22.0630
   24  H       9.2477     9.0845     22.1402
   25  H       9.6557     7.3817     21.8218
   26  S      10.8340     8.2410     23.7100
   27  C      10.1650     3.3080     22.4340
   28  H       9.2929     2.7403     22.7584
   29  C      10.1750     4.6030     23.2620
   30  H      11.1220     5.1220     23.1143
   31  H       9.3551     5.2459     22.9418
   32  C      10.0160     4.3980     24.7440
   33  N       9.7040     5.4090     25.6080
   34  C       9.6570     4.9300     26.8540
   35  H       9.4228     5.5952     27.6851
   36  N       9.9280     3.6450     26.8000
   37  H       9.9617     3.0260     27.5974
   38  C      10.1580     3.2710     25.4990
   39  H      10.3982     2.2340     25.2644
   40  C       6.0350     6.2800     19.5280
   41  H       4.9702     6.5113     19.5559
   42  C       6.6730     6.7710     20.8330
   43  H       7.7511     6.6157     20.7919
   44  H       6.4641     7.8329     20.9631
   45  C       6.1560     6.0500     22.0720
   46  H       5.0693     6.1257     22.1101
   47  H       6.4453     5.0000     22.0292
   48  S       6.7760     6.6970     23.6140
   49  C       6.0690     8.3070     23.6050
   50  H       4.9825     8.2271     23.5709
   51  H       6.3654     8.8396     24.5086
   52  H       6.4202     8.8537     22.7299
   53  CU      9.5640     7.3450     25.1750
   54  N      10.9860     7.2480     30.2860
   55  C      10.9790    11.4950     26.0450
   56  N      10.3490     8.8040     19.6720
```

```
    57 C      12.2010      7.8400     20.8870
    58 N      11.3800      2.5190     22.6730
    59 C       9.8250      3.6710     20.9970
    60 N       6.2160      4.8460     19.4030
    61 C       6.6330      7.0040     18.3530
    62 N      -1.1930     25.6890     17.1840
    63 H      -0.3133     25.1929     17.1970
    64 H      -1.3738     25.1438     18.0148
    65 H      -1.5170     24.8559     16.7138
    66 C      -1.4820     27.1340     16.8960
    67 H      -2.1350     27.2082     16.0264
    68 C      -2.1950     27.7860     18.0880
    69 H      -1.5602     27.7210     18.9717
    70 H      -2.3971     28.8331     17.8627
    71 H      -3.1350     27.2677     18.2776
    72 C      -0.1820     27.8790     16.5880
    73 O       0.8890     27.4920     17.0690
    74 N      -0.2890     28.9420     15.7940
    75 H      -1.1936     29.2105     15.4339
    76 C       0.8750     29.7430     15.4220
    77 H       0.5616     30.5606     14.7728
    78 C       1.5270     30.3290     16.6860
...
  9746 O      31.1328     34.4612     22.6903
  9747 H      31.8908     34.5740     22.1167
  9748 H      30.6706     35.2981     22.6446
END

QMMM
  FORCE_FIELD_FILE $ADFRESOURCES/ForceFields/amber95.ff
  RESTART_FILE mm.restart  ! old style restart
  OUTPUT_LEVEL=1
  WARNING_LEVEL=0
  ! ------------------
  ! for AddRemove model
  ! ------------------
  PARTITION=5
  ELSTAT_COUPLING_MODEL=4 ! for AddRemove model

  OPTIMIZE
    METHOD CONJGRAD
    MAX_STEPS 5000
    PRINT_CYCLES 10
    MM_NOTCONVERGED 0
  SUBEND

  LINK_BONDS
      1 -     54   1.3320 H H1
      8 -     55   1.3861 H H1
     21 -     56   1.3362 H H1
     21 -     57   1.3927 H H1
     27 -     58   1.3471 H H1
     27 -     59   1.3951 H H1
     40 -     60   1.3310 H H1
     40 -     61   1.3799 H H1
  SUBEND
```

```
MM_CONNECTION_TABLE
    1 CT   QM    54     4     2     3
    2 H1   QM     1
    3 H1   QM     1
    4 C    QM     1     5     6
    5 O    QM     4
    6 N    QM     8     4     7
    7 H    QM     6
    8 CT   QM     6    10    55     9
    9 H1   QM     8
   10 CT   QM     8    13    11    12
   11 HC   QM    10
   12 HC   QM    10
   13 CC   QM    10    14    19
   14 NB   QM    13    15    53
   15 CR   QM    14    17    16
   16 H5   QM    15
   17 NA   QM    15    19    18
   18 H    QM    17
   19 CW   QM    13    17    20
   20 H4   QM    19
   21 CT   QM    56    23    57    22
   22 H1   QM    21
   23 CT   QM    21    26    24    25
   24 H1   QM    23
   25 H1   QM    23
   26 SH   QM    23    53
   27 CT   QM    58    29    59    28
   28 H1   QM    27
   29 CT   QM    27    32    30    31
   30 HC   QM    29
   31 HC   QM    29
   32 CC   QM    29    33    38
   33 NB   QM    32    34    53
   34 CR   QM    33    36    35
   35 H5   QM    34
   36 NA   QM    34    38    37
   37 H    QM    36
   38 CW   QM    32    36    39
   39 H4   QM    38
   40 CT   QM    60    42    61    41
   41 H1   QM    40
   42 CT   QM    40    45    43    44
   43 HC   QM    42
   44 HC   QM    42
   45 CT   QM    42    48    46    47
   46 H1   QM    45
   47 H1   QM    45
   48 S    QM    45    49
   49 CT   QM    48    50    51    52
   50 H1   QM    49
   51 H1   QM    49
   52 H1   QM    49
   53 CU   QM    14    26    33
   54 N    LI     1   748   750
   55 C    LI     8   751   752
   56 N    LI    21  1683  1685
```

```
    57 C    LI    21  1686  1687
    58 N    LI    27  1737  1739
    59 C    LI    27  1740  1741
    60 N    LI    40  1790  1792
    61 C    LI    40  1793  1794
    62 N3   MM    66    63    64    65
    63 H    MM    62
    64 H    MM    62
    65 H    MM    62
    66 CT   MM    62    68    72    67
    67 HP   MM    66
    68 CT   MM    66    69    70    71
    69 HC   MM    68
    70 HC   MM    68
    71 HC   MM    68
    72 C    MM    66    73    74
    73 O    MM    72
    74 N    MM    76    72    75
    75 H    MM    74
    76 CT   MM    74    78    89    77
    77 H1   MM    76
    78 CT   MM    76    81    79    80
....
  9746 OW   MM  9747  9748
  9747 HW   MM  9746
  9748 HW   MM  9746
  SUBEND

  CHARGES
       1  -0.0252
       2   0.0698
       3   0.0698
       4   0.5973
       5  -0.5679
       6  -0.4157
       7   0.2719
       8  -0.0581
       9   0.1360
      10  -0.0074
      11   0.0367
      12   0.0367
      13   0.1868
      14  -0.5432
      15   0.1635
      16   0.1435
      17  -0.2795
      18   0.3339
      19  -0.2207
      20   0.1862
      21   0.0350
      22   0.0480
      23  -0.7360
      24   0.2440
      25   0.2440
      26  -0.7360
      27  -0.0581
      28   0.1360
```

```
29   -0.0074
30    0.0367
31    0.0367
32    0.1868
33   -0.5432
34    0.1635
35    0.1435
36   -0.2795
37    0.3339
38   -0.2207
39    0.1862
40   -0.0237
41    0.0880
42    0.0342
43    0.0241
44    0.0241
45    0.0018
46    0.0440
47    0.0440
48   -0.2737
49   -0.0536
50    0.0684
51    0.0684
52    0.0684
53    2.0000
54   -0.4157
55    0.5973
56   -0.4630
57    0.6160
58   -0.4157
59    0.5973
60   -0.4157
61    0.5973
62    0.1414
63    0.1997
64    0.1997
65    0.1997
66    0.0962
67    0.0889
68   -0.0597
69    0.0300
70    0.0300
71    0.0300
72    0.6163
73   -0.5722
74   -0.4157
75    0.2719
76   -0.0031
77    0.0850
78   -0.0036
....
9746   -0.8340
9747    0.4170
9748    0.4170
   SUBEND
END
```

```
COMMENT
  Atom  ADF ID  p2a ID  Amber   Mol   Nr   Amber   pdb ID   atom
   C       1     828      CT     GLY   45    GLY     338      CA
   H       2     829      H1     GLY   45    GLY       0      HA2
   H       3     830      H1     GLY   45    GLY       0      HA3
   C       4     831      C      GLY   45    GLY     339      C
   O       5     832      O      GLY   45    GLY     340      O
   N       6     836      N      HIS   46    HIE     341      N
   H       7     837      H      HIS   46    HIE       0      H
   C       8     838      CT     HIS   46    HIE     342      CA
   H       9     839      H1     HIS   46    HIE       0      HA
   C      10     840      CT     HIS   46    HIE     345      CB
   H      11     841      HC     HIS   46    HIE       0      HB2
   H      12     842      HC     HIS   46    HIE       0      HB3
   C      13     843      CC     HIS   46    HIE     346      CG
   N      14     844      NB     HIS   46    HIE     347      ND1
   C      15     846      CR     HIS   46    HIE     349      CE1
   C      15     846      CR     HIS   46    HIE     349      CE1
   H      16     847      H5     HIS   46    HIE       0      HE1
   N      17     848      NA     HIS   46    HIE     350      NE2
   H      18     849      H      HIS   46    HIE       0      HE2
   C      19     850      CW     HIS   46    HIE     348      CD2
   H      20     851      H4     HIS   46    HIE       0      HD2
   C      21    1996      CT     CYS  112    CYM     824      CA
   H      22    1997      H1     CYS  112    CYM       0      HA
   C      23    1998      CT     CYS  112    CYM     827      CB
   H      24    1999      H1     CYS  112    CYM       0      HB3
   H      25    2000      H1     CYS  112    CYM       0      HB2
   S      26    2001      SH     CYS  112    CYM     828      SG
   C      27    2074      CT     HIS  117    HIE     858      CA
   H      28    2075      H1     HIS  117    HIE       0      HA
   C      29    2076      CT     HIS  117    HIE     861      CB
   H      30    2077      HC     HIS  117    HIE       0      HB2
   H      31    2078      HC     HIS  117    HIE       0      HB3
   C      32    2079      CC     HIS  117    HIE     862      CG
   N      33    2080      NB     HIS  117    HIE     863      ND1
   C      34    2082      CR     HIS  117    HIE     865      CE1
   H      35    2083      H5     HIS  117    HIE       0      HE1
   N      36    2084      NA     HIS  117    HIE     866      NE2
   H      37    2085      H      HIS  117    HIE       0      HE2
   C      38    2086      CW     HIS  117    HIE     864      CD2
   H      39    2087      H4     HIS  117    HIE       0      HD2
   C      40    2155      CT     MET  121    MET     895      CA
   H      41    2156      H1     MET  121    MET       0      HA
   C      42    2157      CT     MET  121    MET     898      CB
   H      43    2158      HC     MET  121    MET       0      HB2
   H      44    2159      HC     MET  121    MET       0      HB3
   C      45    2160      CT     MET  121    MET     899      CG
   H      46    2161      H1     MET  121    MET       0      HG2
   H      47    2162      H1     MET  121    MET       0      HG3
   S      48    2163      S      MET  121    MET     900      SD
   C      49    2164      CT     MET  121    MET     901      CE
   H      50    2165      H1     MET  121    MET       0      HE1
   H      51    2166      H1     MET  121    MET       0      HE2
   H      52    2167      H1     MET  121    MET       0      HE3
   CU     53    2318      CU     CU   130              959      CU
   N      54     826      N      GLY   45    GLY     337      N
```

```
     C         55      852    C     HIS    46     HIE       343      C
     N         56     1994    N     CYS   112     CYM       823      N
     C         57     2003    C     CYS   112     CYM       825      C
     N         58     2072    N     HIS   117     HIE       857      N
     C         59     2088    C     HIS   117     HIE       859      C
     N         60     2153    N     MET   121     MET       894      N
     C         61     2168    C     MET   121     MET       896      C
     N         62        1    N3    ALA     1     ALA         1      N
     H         63        2    H     ALA     1     ALA         0      H1
     H         64       11    H     ALA     1     ALA         0      H2
     H         65       12    H     ALA     1     ALA         0      H3
     C         66        3    CT    ALA     1     ALA         2      CA
     H         67        4    HP    ALA     1     ALA         0      HA
     C         68        5    CT    ALA     1     ALA         5      CB
     H         69        6    HC    ALA     1     ALA         0      HB1
     H         70        7    HC    ALA     1     ALA         0      HB2
     H         71        8    HC    ALA     1     ALA         0      HB3
     C         72        9    C     ALA     1     ALA         3      C
     O         73       10    O     ALA     1     ALA         4      O
     N         74       14    N     GLN     2     GLN         6      N
     H         75       15    H     GLN     2     GLN         0      H
     C         76       16    CT    GLN     2     GLN         7      CA
     H         77       17    H1    GLN     2     GLN         0      HA
     C         78       18    CT    GLN     2     GLN        10      CB
....
     O       2111     2517    OW    HOH   196     SOL      1025      O
     H       2112     2518    HW    HOH   196     SOL         0      H1
     H       2113     2519    HW    HOH   196     SOL         0      H2
END

ENDINPUT
eor
```

An example on solvent shell run

The idea of this example is to make an adf-input file using a PDB file of water (hoh.pdb), in the solvent methanol. The water molecule in the adf-input file should be in the QM part, and the solvent methanol (in a box) is in MM part.

Contents of the hoh.pdb file

```
TITLE      PDB-FILE CORRESPONDING TO pdb2adf-GENERATED ADF-INPUTFILE
REMARK     Written by M. Swart, March 2005
HETATM    1 H1    HOH     1        1.716  26.282  11.239  1.00  0.00        1DYZ H
HETATM    2 O     HOH     1        2.439  25.795  11.634  1.00  0.00        1DYZ O
HETATM    3 H2    HOH     1        3.140  26.440  11.729  1.00  0.00        1DYZ H
END
```

Usage of pdb2adf

The program works interactively. Given below in **bold** are the parts that the user has to type. In cases where the user agrees with the suggestion given by the program, the user can press the Enter key indicated with **Enter**.

```
                         P D B 2 A D F - program
                            version 2005.01
                      Written by: Marcel Swart, 2005

                    This program uses AMBER parameter files
                        see: http://amber.scripps.edu
```

Do you want a logfile to be written (Y/n) ?

Enter

Please give name of PDB-file

hoh.pdb

 read fragments

Data Processed:
     Nat: 3
     Nmol: 1
 NChains: 0

Please wait, making connection tables
Now finding nearby atoms
Assigning chain ID to all residues
Completing residues for which only option is available

Refinding nearby atoms (including atoms added in residue completion)

 -
            Making Choices for Chain   0
 -

Completing residues with multiple options available, and solvent molecules

Checking positions of newly added atoms

Making choice for which molecules should be QM, which MM

Residues belonging to chain 0

Solvent molecules (SOL/HOH) belonging to this chain:
     1

Give the number of the molecule to be put in QM region (or 'c' to continue):

1

Putting HOH   1 in QM region

Give the number of the molecule to be put in QM region (or 'c' to continue):

c

```
Do you want to add solvent to your system (Y/n) ?
```

Enter

```
Solvent (box) available:
   1:  HOH     HOH  Water molecule
   2:  MOH     MOH  Methanol molecule
   3:  CHL     CHL  Chloroform molecule
```

2

```
Reading contents of solvent box
p2abox.MOH

Box Shape options:
 1  Spherical box
 2  Cubic box
Make a choice:
```

1

```
Writing inputfile for chain   0

Using total charge   0.0 and total spin   0.0

Maximum atomic distance (Angs) from center        0.92
Give boxsize (def.:    15.00 Angs)
```

14.0

```
Using BOXSIZE value of  14.0000
Adding atoms for box    1  Added (Box):    84 (Total):    84  Excl. (1):   660  Excl. (2):     6
Adding atoms for box    2  Added (Box):   102 (Total):   186  Excl. (1):   642  Excl. (2):     6
Adding atoms for box    3  Added (Box):   102 (Total):   288  Excl. (1):   642  Excl. (2):     6
Adding atoms for box    4  Added (Box):   108 (Total):   396  Excl. (1):   642  Excl. (2):     0
Adding atoms for box    5  Added (Box):   120 (Total):   516  Excl. (1):   630  Excl. (2):     0
Adding atoms for box    6  Added (Box):    96 (Total):   612  Excl. (1):   654  Excl. (2):     0
Adding atoms for box    7  Added (Box):   108 (Total):   720  Excl. (1):   642  Excl. (2):     0
Adding atoms for box    8  Added (Box):   102 (Total):   822  Excl. (1):   642  Excl. (2):     6

Inputfile(s) written, everything processed, work has been done.
Thank you for using the PDB2ADF program.

================================
Normal ending of PDB2ADF program
================================
```

Contents of the hoh.pdb2adf file generated by pdb2adf

The file is not given completely, since it contains more than 800 atoms.

```
#! /bin/sh

$ADFBIN/adf << eor
TITLE QM/MM calculation setup by pdb2adf: M.Swart, 2005

Symmetry NOSYM

EPRINT
  SFO NOEIG NOOVL
END
```

```
XC
 GGA Becke-Perdew
END

GEOMETRY
 CONVERGE grad=1.0e-3 rad=1.0e-1
END

BASIS
 type TZP
 core small
END

SCF
 Converge 1.0e-5 1.0e-5
 Iterations 99
END

INTEGRATION 5.0 5.0 5.0

CHARGE   0.0

ATOMS
     1 O      2.4390    25.7950    11.6340
     2 H      1.7160    26.2820    11.2390
     3 H      3.1400    26.4400    11.7290
     4 C    -10.0667    22.2493    11.7437
     5 H    -10.2077    21.5053    10.9597
     6 H    -10.5047    21.8683    12.6667
     7 H    -10.5167    23.2103    11.4977
     8 O     -8.7387    22.3983    12.0617
     9 H     -8.3007    22.6943    11.2607
    10 C     -0.2827    19.0253     2.2847
    11 H     -0.5357    18.2063     2.9567
    12 H      0.7633    19.2913     2.4407
    13 H     -0.9267    19.8753     2.5107
    14 O     -0.4997    18.6373     0.9467
    15 H      0.1123    17.9313     0.7287
....
   823 H      5.4711    27.9401    19.5645
   824 O      5.5611    28.7181    17.7095
   825 H      5.2631    27.8621    17.3935
END

QMMM
  FORCE_FIELD_FILE $ADFRESOURCES/ForceFields/amber95.ff
  RESTART_FILE mm.restart  ! old style restart
  OUTPUT_LEVEL=1
  WARNING_LEVEL=0
  ! -------------------
  ! for AddRemove model
  ! -------------------
  PARTITION=5
  ELSTAT_COUPLING_MODEL=4 ! for AddRemove model

  OPTIMIZE
```

```
        METHOD CONJGRAD
        MAX_STEPS 5000
        PRINT_CYCLES 10
        MM_NOTCONVERGED 0
      SUBEND

    MM_CONNECTION_TABLE
          1 OW    QM      2      3
          2 HW    QM      1
          3 HW    QM      1
          4 CT    MM      5      6      7      8
          5 H1    MM      4
          6 H1    MM      4
          7 H1    MM      4
          8 OH    MM      4      9
          9 HO    MM      8
         10 CT    MM     11     12     13     14
         11 H1    MM     10
         12 H1    MM     10
         13 H1    MM     10
         14 OH    MM     10     15
         15 HO    MM     14
....
        823 H1    MM    820
        824 OH    MM    820    825
        825 HO    MM    824
      SUBEND

    CHARGES
          1  -0.8340
          2   0.4170
          3   0.4170
          4   0.1166
          5   0.0372
          6   0.0372
          7   0.0372
          8  -0.6497
          9   0.4215
         10   0.1166
         11   0.0372
         12   0.0372
         13   0.0372
         14  -0.6497
         15   0.4215
...
        823   0.0372
        824  -0.6497
        825   0.4215
      SUBEND
    END

    COMMENT
      Atom  ADF ID  p2a ID  Amber    Mol   Nr   Amber    pdb ID    atom
       O        1      1     OW     HOH    1    SOL         2      O
       H        2      2     HW     HOH    1    SOL         1      H1
       H        3      3     HW     HOH    1    SOL         3      H2
    END
```

```
ENDINPUT
eor
```

3.3 Example in $ADFHOME/examples/adf

Shown here ia a pdb2adf example that is stored in the subdirectories under $ADFHOME/examples/adf, where $ADFHOME is the main directory of the ADF package.

pdb2adf: transforms a PDB file in a QM/MM adf-input file

Sample directory: adf/pdb2adf/

This example shows how to use the utiliy pdb2adf, which creates an ADF input file from a PDB file, for a subsequent QM/MM calculation using ADF.

First create the PDB file that can be used in this example.

```
cat << eor > chymotrypsin.pdb
HEADER    COMPLEX (SERINE PROTEASE/INHIBITOR)     12-MAR-97    1AFQ
TITLE     CRYSTAL STRUCTURE OF BOVINE GAMMA-CHYMOTRYPSIN COMPLEXED
TITLE    2 WITH A SYNTHETIC INHIBITOR
REMARK
REMARK     Adaptation of original PDB file by M. Swart, March 2005
REMARK     only coordinates of GAMMA-CHYMOTRYPSIN are kept;
REMARK     rest has been deleted.
REMARK
ATOM      1  N   CYS A   1      13.717  20.021  22.754  1.00 13.46      PROA N
ATOM      2  CA  CYS A   1      14.211  18.932  23.617  1.00 13.34      PROA C
ATOM      3  C   CYS A   1      13.597  19.033  25.005  1.00 13.34      PROA C
...
ATOM     68  CD2 LEU A  10       9.768  11.681  39.555  1.00 27.46      PROA C
ATOM     69  OXT LEU A  10       6.329  11.066  42.743  1.00 27.55      PROA O
TER      70      LEU A  10
END
eor
```

Then run the pdf2adf program to create ADF inputfile

```
$ADFBIN/pdb2adf << eor

chymotrypsin.pdb



4 5
c


Y
1
1
17.5
eor
```

The program works interactively. The input described here are answers to the questions that were asked interactively. In cases where the user agrees with the suggestion given by the program, the user can press the Enter key, which is shown here with an empty line.

The questions asked can be found in the output file, and are repeated here. The `Enter` key or empty line is indicated here with **Enter**.

```
Do you want a logfile to be written (Y/n) ?
```

Enter

```
Please give name of PDB-file
```

chymotrypsin.pdb

> ```
> Found the following terminal amino acid residues : (C-term)    10
> (N-term)      1
> Do you want to use these as terminal residues (Y/n) ?
> ```

Enter

```
Multiple AMBER options for CYS :
  0         Decide every time differently
  1   CYS   Cysteine (SH)
  2   CYM   Deprotonated Cysteine (S-)
  3   CYX   Cystine (S-S bridge)


Suggested option: 0
```

Enter

```
Multiple AMBER options for CYS    1 (    1) :
  1   CYS   Cysteine (SH)
  2   CYX   Cystine (S-S bridge)

  Connections and Nearest Atoms for SG CYS    1 SG  ( P2A #    8 PDB#    6 )
     Dist P2A Nr  PDB Nr  Label              Near    Dist  P2A Nr  PDB Nr  Label
1    1.83      5       5 CB CYS    1 CB        1     5.58      19       0 H1  GLY    2
                                              2     6.06      36       0 HC  VAL    3
                                              3     6.09      26       0 H   VAL    3
                                              4     6.47      25      11 N   VAL    3 N
                                              5     7.15      35      17 CT  VAL    3 CG2
Suggestion: 1
```

Enter

```
Option  Molecule    Option  Molecule    Option  Molecule    Option  Molecule    Option  Molecule
   1: CYS    1         4: PRO    4         7: GLN    7        10: LEU   10
   2: GLY    2         5: ALA    5         8: PRO    8
   3: VAL    3         6: ILE    6         9: VAL    9

Give option number of molecules to be put in QM region (or 'c' to continue):
Note: by specifying a negative number a molecule is removed from the QM region
```

4 5

```
Give option number of molecules to be put in QM region (or 'c' to continue):
Note: by specifying a negative number a molecule is removed from the QM region
```

c

```
Make a choice for the QM/MM treatment of PRO    4
 0:  Put completely in QM region
 1:  Cut off at C-alpha (put NH in QM region, CO in MM region)
 2:  Cut off at C-alpha (put NH in MM region, CO in QM region)
 3:  Cut off at C-alpha (put NH and CO in MM region)
 4:  Cut off at C-alpha (put NH and CO in QM region,  sidechain in MM region)
 5:  Put only part of sidechain in QM region
```

```
Suggestion: 2
...
Give choice:
```

Enter

```
Make a choice for the QM/MM treatment of ALA    5
 0:  Put completely in QM region
 1:  Cut off at C-alpha (put NH in QM region, CO in MM region)
 2:  Cut off at C-alpha (put NH in MM region, CO in QM region)
 3:  Cut off at C-alpha (put NH and CO in MM region)
 4:  Cut off at C-alpha (put NH and CO in QM region,  sidechain in MM region)
 5:  Put only part of sidechain in QM region

Suggestion: 1
Give choice:
```

Enter

```
Do you want to add solvent to your system (Y/n) ?
```

Y

```
Solvent (box) available:
    1:  HOH     HOH  Water molecule
    2:  MOH     MOH  Methanol molecule
    3:  CHL     CHL  Chloroform molecule
```

1

```
Make a choice:
```

1

```
Give boxsize (def.:    16.71 Angs)
```

17.5

References

1. T.K. Woo, L. Cavallo and T. Ziegler, *Implementation of the IMOMM methodology for performing combined QM/MM molecular dynamics simulations and frequency calculations.* Theoretical Chemistry Accounts **100**, 307 (1998)

2. F. Maseras and K. Morokuma, *IMOMM: A new integrated ab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states.* Journal of Computational Chemistry **16**, 1170 (1995)

3. M. Swart, *AddRemove: A new link model for use in QM/MM studies.* International Journal of Quantum Chemistry **91**, 177 (2003)

4. W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell and P.A. Kollman, *A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules.* Journal of the American Chemical Society **117**, 5179 (1995)

5. M. Clark, R.D. Cramer III and N. Van Opdenbosch, *Validation of the general purpose tripos 5.2 force field.* Journal of Computational Chemistry **10**, 982 (1989)

6. U.C. Singh and P.A. Kollman, *A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the CH_3Cl + Cl^- exchange reaction and gas phase protonation of polyethers.* Journal of Computational Chemistry **7**, 718 (1986)

7. M. Swart, P.Th. van Duijnen and J.G. Snijders, *A charge analysis derived from an atomic multipole expansion.* Journal of Computational Chemistry **22**, 79 (2001)

Appendix A. List of Publications Using ADF QM/MM

T.K. Woo, L. Cavallo and T. Ziegler, *Implementation of the IMOMM methodology for performing combined QM/MM molecular dynamics simulations and frequency calculations.* Theoretical Chemistry Accounts **100**, 307 (1998)

L. Cavallo, T.K. Woo and T. Ziegler, *A Combined QM/MM Study of Ligand Substitution Enthalpies in the $L_2Fe(CO_3)$, $RuCpL_2Cl$ and $RuCp^*L_2Cl$ Systems.* Canadian Journal of Chemistry **76**, 1457 (1998)

L. Deng, T.K. Woo, L. Cavallo, P.M. Margl and T. Ziegler, *The Role of Bulky Substituents in Brookhart-type Ni(II) Diimine Polymerization Catalysts: A Combined Molecular Mechanics and Density Functional Study.* Journal of the American Chemical Society **119**, 6177 (1997)

L. Deng, T. Ziegler, T.K. Woo, P.M. Margl and L. Fan, *Computer Design of Living Olefin Polymerization Catalyst: A combined Density Functional Theory and Molecular Mechanics Study on Polymerization of Ethylene by Chelating Diamide Complexes of Titanium, Zirconium and Hafnium.* Organometallics **17**, 3240 (1998)

P.M. Margl, T.K. Woo and T. Ziegler, *Potential Catalyst Deactivation Reaction in Homogeneous Ziegler-Natta Polymerization of Olefins: Formation of an Allyl Intermediate.* Organometallics **17**, 4997 (1998)

T.K. Woo and T. Ziegler, *The influence of electronic and steric factors on chain branching in ethylene polymerization by Brookhart-type Ni(II) diimine catalysts: a combined density functional theory and molecular mechanics study.* Journal of Organometallic Chemistry **591**, 204 (1999)

T.K. Woo, P.E. Blöchl and T. Ziegler, *Monomer Capture in Brookhart's Ni(II) Diimine Olefin Polymerization Catalyst: Static and Dynamic Quantum Mechanics/Molecular Mechanics Study.* Journal of Physical Chemistry A **104**, 121 (2000)

P. Margl, L. Deng and T. Ziegler, *Cobalt(II) Imino Pyridine Assisted Ethylene Polymerization: A Quantum-Mechanical/Molecular-Mechanical Density Functional Theory Investigation.* Organometallics **18**, 5701 (1999)

L. Deng, P. Margl and T. Ziegler, *Mechanistic Aspects of Ethylene Polymerization by Iron(II)-Bisimine Pyridine Catalysts: A Combined Density Functional Theory and Molecular Mechanics Study.* Journal of the American Chemical Society **121**, 6479 (1999)

G. Moscardi, F. Piemontesi and L. Resconi, *Propene Polymerization with the Isospecific, Highly Regiospecific rac-Me$_2$C(3-t-Bu-1-Ind)$_2$ZrCl$_2$/MAO Catalyst. 1. Influence of Hydrogen on Initiation and Propagation: Experimental Detection and Theoretical Investigation of 2,1 Propene Insertion into the Zr-H Bond.* Organometallics **18**, 5264 (1999)

G. Guerra, P. Longo, P. Corradini and L. Cavallo, *(E)-(Z) Selectivity in 2-Butene Copolymerization by Group 4 Metallocenes. A Combined Density Functional Theory and Molecular Mechanics Study.* Journal of the American Chemical Society **121**, 8651 (1999)

G. Milano, G. Guerra, C. Pellecchia and L. Cavallo, *Mechanism of Unlink Stereoselectivity in 1-Alkene Primary Insertions: Syndiospecific Propene Polymerization by Brookhart-Type Nickel(II0 Catalysts.* Organometallics **19**; 1343 (2000)

Keywords