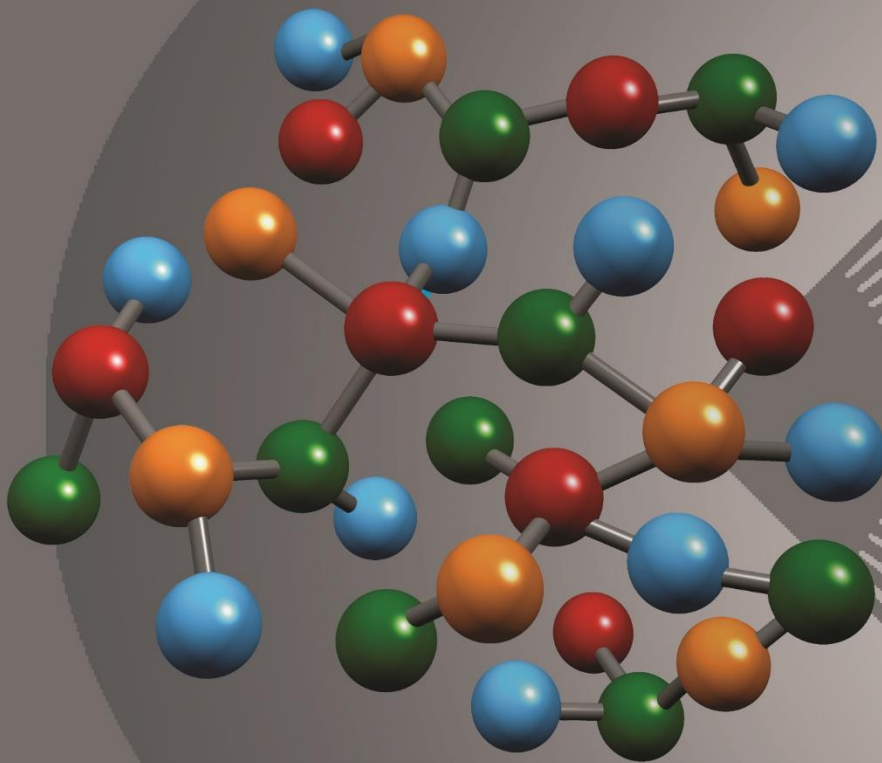


# 19

USER MANUAL



**EMPIRE**™

## Impressum

### Copyright

© 2019 by Cepas InSilico GmbH  
Waldstrasse 15  
90587 Obermichelbach  
[www.ceposinsilico.de](http://www.ceposinsilico.de)

### Manual

Timothy Clark

### Layout

[www.eh-bitartist.de](http://www.eh-bitartist.de)



EMPIRE™: a Parallel Semiempirical Molecular Orbital Program

# TABLE OF CONTENTS

<b>1 INTRODUCTION</b>	<b>4</b>
1.1 New in Empire19	5
<b>2 INPUT FORMATS</b>	<b>6</b>
2.1 .xyz and .xmol files	6
2.2 .sdf files	7
2.3 .dat files	8
<b>3 PROGRAM OPTIONS</b>	<b>10</b>
<b>4 PERIODIC CALCULATIONS</b>	<b>14</b>
<b>5 MOLECULAR DYNAMICS</b>	<b>15</b>
<b>6 PROGRAM OUTPUT</b>	<b>16</b>
<b>7 ADDITIONAL FILE FORMATS</b>	<b>22</b>
7.1 EMPIRE™ archive file	22
7.2 HDF file format	23
7.3 Binary wavefunction file	24
(deprecated, replaced by the HDF5 file)	
7.4 SDF output file	25
<b>8 GLOSSARY ENTRIES</b>	<b>28</b>
<b>9 SUPPORT</b>	<b>30</b>
9.1 Contact	30
9.2 Cepos InSilico GmbH	30
<b>10 LIST OF TABLES</b>	<b>31</b>
<b>11 LIST OF FIGURES</b>	<b>32</b>
<b>12 REFERENCES</b>	<b>33</b>

# 1 INTRODUCTION

EMPIRE™ is a completely new **NDDO**-based semiempirical molecular orbital (MO) program designed for moderately (the single-node version) or massively (the cluster version) parallel single-point calculations, geometry optimizations, normal-mode calculations or MD simulations for open- and closed-shell molecules and excited states within the **Restricted and Unrestricted Hartree-Fock** (RHF and UHF) **self-consistent field** (SCF) formalisms and with **configuration interaction**. Because of its radical new design, EMPIRE™ uses different algorithms for many steps in the calculation to the established traditional programs and can therefore provide impressive parallel performance, robustness, reliability, and simplicity of operation. A further important advantage of EMPIRE™ is that it is able to calculate large molecules with ease (the size of the calculations is only restricted by the hardware available). For instance, one of the development tests for EMPIRE™ was to perform a single-point calculation on a molecule containing more than 76,000 atoms on 1,024 CPUs with high parallel efficiency. The current version of EMPIRE™ does not use **linear-scaling techniques**. It performs the full NDDO-based SCF calculations without approximations. It thus provides a “gold standard” against which all other programs can be measured.

EMPIRE™ currently only performs full geometry optimizations using Cartesian coordinates with a modified version of **Baker's EF-algorithm**.<sup>[1]</sup> Later versions will provide other optimization options.

## Citing EMPIRE™:

*EMPIRE19*, M. Hennemann, J. T. Margraf, B. Meyer and T. Clark, Cepos InSilico GmbH, Obermichelbach, 2019. <http://www.ceposinsilico.de/products/empire.htm>

*EMPIRE: A highly parallel semiempirical molecular orbital program: 1: Self-Consistent Field Calculations*, M. Hennemann and T. Clark, *J. Mol. Model.* **2014**, *20*, 2331 (11 pages, <https://doi.org/10.1007/s00894-014-2331-4>).

*EMPIRE: A highly parallel semiempirical molecular orbital program: 2: Periodic Boundary Conditions*, J. T. Margraf, M. Hennemann, B. Meyer and T. Clark, *J. Mol. Model.*, **2015**, *21*, 144 (7 pages, <https://doi.org/10.1007/s00894-015-2692-3>).

## 1.1 New in Empire19

EMPIRE19 contains several updates, extensions and improvements relative to the previous version. These are listed below:

### New Hamiltonians:

PM6	Stewart's PM6 Hamiltonian is now available ( <a href="#">Hamil=PM6</a> )
RM1	The RM1 Hamiltonian has been implemented ( <a href="#">Hamil=RM1</a> )
hpCADD	The hpCADD Hamiltonian is available for single-point calculations. This Hamiltonian requires the specification of force-field-like atom types, as specified in the original literature (see below) and is not available for optimizations, frequency calculations or molecular dynamics. ( <a href="#">Hamil=hpCADD</a> )
MNDO-F	MNDO extended with the Feynman dispersion correction ( <a href="#">Hamil=MNDO-F</a> )

**Numerical Frequencies** Numerical frequencies can now be calculated within the harmonic approximation.

**Faster gradients** The algorithm used to calculate the first derivatives of the energy has been improved. This gives better performance for optimizations, frequency calculations and molecular dynamics.

**Excited States** Configuration-interaction (CI) calculations are now available for calculating excited states in non-periodic systems.

## 2 INPUT FORMATS

EMPIRE™ can use the following input file formats:

### 2.1 .xyz and .xmol files

Input files with the extension **.xyz** or **.xmol** are the simplest form of EMPIRE™ input. An example for adamantane is shown in **Figure 1**:

```

26
adamantane
C      -2.804713   -0.419657   -0.756331   0.000000
C      -2.178158   0.472538   -1.823698   0.000000
C      -0.686676   0.168118   -1.926677   0.000000
C      -0.016348   0.435007   -0.582434   0.000000
C      -0.645057   -0.457079   0.483778   0.000000
C      -2.136662   -0.154334   0.589405   0.000000
C      -2.371296   1.936304   -1.439390   0.000000
C      -1.703934   2.204223   -0.093869   0.000000
C      -0.212905   1.898529   -0.198672   0.000000
C      -2.330995   1.309943   0.971428   0.000000
H      -2.675998   -1.495723   -1.038190   0.000000
H      -3.902259   -0.209884   -0.682892   0.000000
H      -2.670214   0.276113   -2.813001   0.000000
H      -0.221869   0.810667   -2.717242   0.000000
H      -0.537509   -0.902562   -2.218751   0.000000
H      1.080739    0.211523   -0.659300   0.000000
H      -0.150309   -0.274141   1.471527   0.000000
H      -0.494211   -1.533564   0.215065   0.000000
H      -2.597946   -0.811138   1.373628   0.000000
H      -3.464266   2.170659   -1.374134   0.000000
H      -1.921655   2.596929   -2.224091   0.000000
H      -1.847649   3.280820   0.188371   0.000000
H      0.286344    2.103433   0.782564   0.000000
H      0.256433    2.559787   -0.970729   0.000000
H      -3.423926   1.536538   1.061529   0.000000
H      -1.852892    1.509796   1.964158   0.000000
  
```

**Figure 1** .xyz (.xmol) input for adamantane

The input consists simply of the first line, which contains only the number of atoms in the molecule, a second, title line and then one line per atom containing the element symbol, x-, y- and z-coordinates and an optional (in this case fictitious) net atomic charge. The input is format-free. There may or may not be empty lines after the specification of the atoms.

Note that this file format does not allow the user to specify a molecular charge, which must then be specified (if it is not zero) via the command-line argument **charge=n** (see below).

## 2.2 .sdf files

The **structure-data (SD) file format version V2000** can be used as an input file with the extension **.sdf**. **Figure 2** shows an example, again for adamantane.

```
adamantane
XXEMPIRE1909131910553D 1 1.00000 0.00000 0
EMPIRE input file
26 28 0 0 0 0 1 v2000
-0.7632 -0.7127 1.2463 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.7626 -0.7166 1.2484 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1.2726 0.7213 1.2450 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.7680 1.4401 -0.0025 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-0.7578 1.4420 -0.0031 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-1.2713 0.0054 -0.0002 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1.2692 -1.4378 0.0030 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.7632 -0.7212 -1.2452 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1.2740 0.7164 -1.2465 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-0.7625 -0.7172 -1.2438 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-1.1421 -0.1945 2.1639 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-1.1447 -1.7653 1.2602 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1.1351 -1.2449 2.1657 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
2.3922 0.7272 1.2602 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.9105 1.2535 2.1614 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1.1445 2.4973 -0.0044 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-1.1358 1.9782 -0.9106 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-1.1366 1.9819 0.9019 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-2.3936 0.0081 -0.0006 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.9068 -2.4973 0.0049 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
2.3888 -1.4545 0.0033 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1.1354 -1.2528 -2.1607 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.9144 1.2454 -2.1657 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
2.3936 0.7214 -1.2596 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-1.1442 -1.7698 -1.2533 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-1.1416 -0.2028 -2.1634 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 2 1 0 0 0 0
2 3 1 0 0 0 0
3 4 1 0 0 0 0
4 5 1 0 0 0 0
1 6 1 0 0 0 0
5 6 1 0 0 0 0
2 7 1 0 0 0 0
7 8 1 0 0 0 0
4 9 1 0 0 0 0
8 9 1 0 0 0 0
6 10 1 0 0 0 0
8 10 1 0 0 0 0
1 11 1 0 0 0 0
1 12 1 0 0 0 0
2 13 1 0 0 0 0
3 14 1 0 0 0 0
3 15 1 0 0 0 0
4 16 1 0 0 0 0
5 17 1 0 0 0 0
5 18 1 0 0 0 0
6 19 1 0 0 0 0
7 20 1 0 0 0 0
7 21 1 0 0 0 0
8 22 1 0 0 0 0
9 23 1 0 0 0 0
9 24 1 0 0 0 0
10 25 1 0 0 0 0
10 26 1 0 0 0 0
M END
$$$$
```

**Figure 2** .sdf input for adamantane (the format is **structure-data file V2000**)

If the **.sdf** format is used, EMPIRE™ can calculate the molecular charge from the **formal charges** of the atoms. The charge can be derived (in order of preference) from:

1. The **<TOTAL COULSON CHARGE>** floating point number
2. The **<TOTAL CHARGE>** integer
3. Summed atomic charges from **M CHG**
4. Summed atomic charges from the atoms block

The bond definitions have no effect on EMPIRE™ except that they are echoed to the output unmodified. The molecule is terminated by the **\$\$\$\$** line. EMPIRE™ can read **.sdf** files with multiple molecules. The next input simply starts after the **\$\$\$\$** line.

## 2.3 .dat files

The **.dat** file format is similar to that used by **Mopac™** and **Vamp™** and is shown in **Figure 3**.

```
Hamil=AM1 Charge=0 Calc=Opt
adamantane
EMPIRE input file
C -2.81969539 1 -0.42849875 1 -0.75975579 1
C -2.18365402 1 0.47148130 1 -1.83477804 1
C -0.67864902 1 0.15981271 1 -1.94047398 1
C 0.00064190 1 0.43403192 1 -0.58652920 1
C -0.63342139 1 -0.46715388 1 0.48888876 1
C -2.13878008 1 -0.15942829 1 0.59514446 1
C -2.37988782 1 1.94829175 1 -1.44582725 1
C -1.70312377 1 2.21849294 1 -0.08896003 1
C -0.19846430 1 1.91055875 1 -0.19423207 1
C -2.33962326 1 1.31635509 1 0.98489736 1
H -2.67775461 1 -1.52266577 1 -1.04655070 1
H -3.93313548 1 -0.19703042 1 -0.68055019 1
H -2.68385018 1 0.27386442 1 -2.83994038 1
H -0.20925217 1 0.82113255 1 -2.74167472 1
H -0.53628795 1 -0.93588502 1 -2.22115130 1
H 1.11503337 1 0.20722999 1 -0.66585440 1
H -0.13323818 1 -0.26884374 1 1.49392100 1
H -0.48872273 1 -1.56086722 1 0.20174181 1
H -2.60574060 1 -0.82763989 1 1.39203911 1
H -3.49424775 1 2.17514548 1 -1.36620759 1
H -1.91233800 1 2.61641532 1 -2.24245014 1
H -1.84860379 1 3.31220683 1 0.19778976 1
H 0.30042996 1 2.10661307 1 0.81188282 1
H 0.27181706 1 2.58013872 1 -0.98801924 1
H -3.45462600 1 1.54188055 1 1.05910740 1
H -1.84331880 1 1.51579233 1 1.99162802 1
```

**Figure 3** .dat input for an AM1 optimization of adamantane





The first line contains the program options, which are defined in **Section 3** below. Note that the standard Mopac™ or Vamp™ options will not be recognized, only the correct EMPIRE™ options. The next two lines contain the molecular title and a comment, followed by one line per atom, which consists of the atomic symbol, x-, y- and z-coordinates, each followed by a **1** or **0** to indicate whether the coordinate should be optimized. The **1** or **0** optimization flags indicate whether the coordinate will be optimized or not. The molecule is terminated by a blank line.

## 3 PROGRAM OPTIONS

EMPIRE™ options are specified either in the first line of the `.dat` file input or in the command line when the program is executed. The following table details the available options.

Table 1 EMPIRE™ program options

Option	Details	
<code>-h</code>	Do not do a calculation, but print the list of options as a “help”	
<code>Charge=&lt;i&gt;</code>	Set the total molecular charge to <code>&lt;i&gt;</code>	
<code>Mult=&lt;n&gt;</code>	Set the Multiplicity to <code>&lt;n&gt;</code>	
<code>Hamil=&lt;s&gt;</code>	Set the semiempirical <b>Hamiltonian</b> to be used. The Hamiltonian files provided with EMPIRE™ allow the following values for <code>&lt;s&gt;</code> :	
	<code>AM1</code>	Use the AM1 [2] <code>charge=n</code> Hamiltonian (this is the default if no Hamiltonian is defined)
	<code>AM1-JS</code>	Use the AM1 [2f] Hamiltonian with reparametrized Na and Mg parameters
	<code>AM1S</code>	Use the AM1* [3] Hamiltonian
	<code>hpCADD</code>	Use the hpCADD [4] Hamiltonian
	<code>MNDO</code>	Use the MNDO [2f, 5] Hamiltonian
	<code>MNDO-JS</code>	Use the MNDO Hamiltonian with reparametrized Mg, K and Ca parameters [2f, 5]
	<code>MNDOD</code>	Use the MNDO/d [6] Hamiltonian
	<code>MNDO-F</code>	Use the MNDO-F Hamiltonian [2f, 7]
	<code>PM3</code>	Use the PM3 [8] Hamiltonian
	<code>PM6</code>	Use the PM6 [9] Hamiltonian
	<code>RM1</code>	Use the RM1 [2d] Hamiltonian
<code>Form=&lt;s&gt;</code>	Defines the formalism	
	<code>RHF</code>	Restricted Hartree-Fock (default for <b>Mult=1</b> )
	<code>UHF</code>	Unrestricted Hartree-Fock (default for <b>Mult&gt;1</b> )
<code>Calc=&lt;s&gt;</code>	Defines the type of calculation to be performed	
	<code>SPT</code>	Perform a single-point calculation (this is the default if no calculation is defined)
	<code>OPT</code>	Perform a full geometry optimization in Cartesian coordinates
	<code>FREQ</code>	Perform a frequency calculation
	<code>MD</code>	Perform a molecular dynamics simulation
<b>Initial Guess</b>		
<code>Guess=&lt;s&gt;</code>	Initial guess Eigenvectors	
	<code>Hueckel</code>	Extended Hückel calculation (default for larger molecules)
	<code>Density</code>	via initial density matrix (default for small molecules)
	<code>*.vwf</code>	Read the Eigenvectors from the specified <code>.vwf</code> file
	<code>*.aux</code>	Read the Eigenvectors from the specified <code>.aux</code> file
<code>Randomize=&lt;f&gt;</code>	Randomize the initial guess matrix (Hückel or density) (default: 10 <sup>-3</sup> )	

<b>InitDiag</b>	Do an initial full diagonalization (this is useful when reading the Eigenvectors from file, or to obtain a better UHF initial guess)	
<b>MixMOs</b> [= -h, +l]	Mix the HOMO-h and LUMO+l orbitals of an UHF initial guess (Default: 0,0 => HOMO,LUMO ; for " <b>Mult=1</b> ")	
<b>Input Options</b>		
<b>Center</b>	Center the molecular structure	
<b>Center</b> =<n>	Center the molecular structure on atom <n>	
<b>Center</b> =<x>, <y>, <z>	Center the molecular structure on point x,y,z	
<b>Orient</b>	Orient the molecular structure according to its principal axes	
<b>Project</b>	Project atoms into the central lattice cell	
<b>Output Options</b>		
<b>HDF</b> [=<n>]	Write an HDF5 wavefunction file: <Basename>_e.h5 [add a structure every <n> MD steps]	
<b>VWF</b>	Write a vamp wavefunction file: <Basename>_e.vwf (deprecated; please use HDF instead)	
<b>XYZ</b> [=<n>]	Write an XYZ trajectory file: <Basename>_e.xyz [add a structure every <n> OPT/MD steps]	
<b>DAT</b>	Force output of a restart DAT file: <Basename>_e.dat	
<b>SDF</b>	Write an MDL structure-data file: <Basename>_e.sdf (SDF output is restricted to less than 1000 atoms; Atom-type symbols are replaced by element symbols)	
<b>SDFBonds</b> =<s>	Type of bonds written to the output SDF	
	<b>Keep</b>	do not alter bonds and bond types
	<b>Check</b>	check if the connectivity changed (default)
	<b>Aryl</b>	check connectivity and adjust aryl bond types to 4
	<b>Calc</b>	calculate bond types from bond orders (Bond type 4 is used for aromatic bonds)
<b>Calculation Details</b>		
<b>MaxTime</b> =<n>	Define the maximum CPU-time (default : 365 days). Default units are seconds (e.g. 365 is interpreted as 365s), times can also be given in minutes (e.g. 30m), hours (e.g. 24h) or days (e.g. 7d)	
<b>FullDiag</b> =<n>	Do a full diagonalization every <n> OPT/MD steps (default = 1)	
<b>Geometry Optimization</b>		
<b>MaxOpt</b> =<n>	Define the maximum number of optimization cycles (default = 10,000)	
<b>ConvergeG</b> =<f>	Set the <b>gradient norm criterion</b> for the geometry optimization (default = 0.4 kcal mol <sup>-1</sup> Å <sup>-1</sup> )	
<b>Molecular Dynamics</b>		
<b>MDsteps</b> =<n>	Number of MD steps (default=100)	
<b>MDsteps</b> =<n1, n2>	Use two numbers <n1,n2> to specify the starting and ending step (used for restarting a MD calculation)	
<b>TimeStep</b> =<f>	MD time step in fs (default = 0.5)	
<b>Temp</b> =<f>	MD target temperature in K (default = 300.0)	
<b>Ensemble</b> =<s>	MD Ensemble	
	<b>NVT</b>	Constant number of particles, volume and temperature (default)
	<b>NVE</b>	Constant number of particles, volume and energy
<b>Thermostat</b> =<s>	Thermostat type	

	<b>Berendsen</b>	Berendsen thermostat (default)
	<b>Langevin</b>	Langevin thermostat
<b>RelaxTime=&lt;f&gt;</b>	Thermostat relaxation time in fs (default = 100.0)	
<b>ReadVelocities</b>	Read initial velocities from the DAT file (for restarting an MD calculation)	
<b>FixCOM=&lt;n&gt;</b>	Remove center of mass translation and rotation every <n> steps (default = 10000; only used for the Berendsen thermostat)	
<b>GuessAll</b>	New initial guess at each MD step (Default for NVE calculations)	
<b>Self-Consistent Field</b>		
<b>MaxSCF=&lt;n&gt;</b>	Define the maximum number of SCF cycles (default = 1,000)	
<b>ConvergeE=&lt;f&gt;</b>	Set the energy convergence limit for the SCF (default = $10^{-4}$ kcal mol <sup>-1</sup> )	
<b>ConvergeR=&lt;f&gt;</b>	Set the maximum off-diagonal <b>CFC element</b> for SCF convergence (default = $10^{-4}$ )	
<b>CheckP</b>	Use the <b>convergence on the density matrix</b> (RSS(Density)) to test for SCF convergence, instead of max(CFC). This option is not the default in EMPIRE™ and is not suitable for very large molecules. The default technique in EMPIRE™ is to test for the maximum rotation angle in the <b>pseudodiagonalization procedure</b> . <sup>[10]</sup> This is equivalent to testing the density directly and far more suitable for large molecules.	
<b>ConvergeP=&lt;f&gt;</b>	Set the convergence limit for RSS(Density) (default = $10^{-6}$ ). This option implies <b>CheckP</b> .	
<b>IMemory=&lt;n&gt;</b>	Use up to <n> MB/thread of memory for storing integrals (default = 1024; Unlimited for periodic calculations)	
<b>RecalcH</b>	Calculate the <b>one-electron matrix</b> on the fly. This option saves memory and may be necessary for very large calculations. The default is to save the one-electron matrix, which is faster.	
<b>FullRot</b>	Do pseudorotations between all combinations of orbitals, including occupied-occupied and virtual-virtual rotations, rather than just virtual-occupied (This can help if the pseudodiagonalization fails).	
<b>MinBandGap=&lt;f&gt;</b>	Use <b>dynamic level shifting</b> to ensure that the <b>band gap</b> used for the SCF does not fall below <f> (default = 1.0; disabled when set to $\leq 0.0$ )	
<b>RCutoff=&lt;f&gt;</b>	Defines the distance cutoff for the overlap / resonance integrals (default = 20.0)	
<b>PDcutoff=&lt;f&gt;</b>	Defines the cutoff value for CFC at which rotations are included in the pseudodiagonalization (default = $0.04 \times$ the maximum CFC value)	
<b>EiCutoff=&lt;f&gt;</b>	Defines the maximum value of CFC below which the <b>Eigenvalues</b> will not be recalculated in the SCF cycle. The default is <f>= 0.0, which means that the Eigenvalues are always calculated. Setting <b>EiCutoff</b> to higher values will make calculations faster but may slow down or prevent SCF convergence.	
<b>Eigensolver=&lt;s&gt;</b>	Preferred Eigensolver	
	<b>DSYEVR</b>	DSYEVR or PDSYEVR (Default)
	<b>DSYEVD</b>	DSYEVD or PDSYEVD
	<b>ELPA1</b>	ELPA one-stage
	<b>ELPA2</b>	ELPA two-stage
<b>Embedding Point Charges</b>		

Embedding point charges can be specified as charged dummy atoms after the real atoms in the .dat or .xyz input file using the atom type "Q":

	XYZ: <b>Q X-Coord. Y-Coord. Z-Coord. Charge</b> DAT: <b>Q X-Coord. Y-Coord. Z-Coord. Charge</b> Embedded point charges cannot be optimized, and the optimization flags in the .dat file will be ignored.
--	--

### Periodic Boundary Conditions [11]

Up to three translation vectors can be specified at the end of the .dat or .xyz file input using the atom type "TV". If lattice optimization is requested via the optimization flags in the .dat file input, the first and second translation vector must be aligned with the x-axis and the xy-plane, respectively:

	<b>TV Vector1-X 0/1 0.0 0 0.0 0</b> <b>TV Vector2-X 0/1 Vector2-Y 0/1 0.0 0</b> <b>TV Vector3-X 0/1 Vector3-Y 0/1 Vector3-Z 0/1</b>
--	---

**ScreeningR=<f>** Simple screening radius beta in Ångström (default = 30.0)

**DampingR=<f>** Integral damping radius in Bohr (default = 10.0)

**DampingC=<f>** Integral damping coefficient (default = 0.25)

### Natural Orbitals

**NOs** Calculate natural orbitals

### Configuration Interaction

<b>CI=*</b>	Type of CI calculation:	
	<b>Full</b> <b>PECI</b> <b>1 or S</b> <b>2 or SD</b> <b>3 or SDT</b> <b>4 or SDTQ</b> .....	Full CI Paired Electron CI [12] Singles (CISD) Singles and doubles (CISD) Singles, doubles and triples (CISDT) Singles, doubles, triples and quadruples ... and so on ...

**AS=<t>[,\*,\*]** Active space, defined by <t>:

	<b>ALL</b> <b>MO,&lt;n1&gt;,&lt;n2&gt;</b> <b>OV,&lt;n1&gt;,&lt;n2&gt;</b> <b>UNO,&lt;f&gt;[,&lt;f&gt;]</b>	All orbitals (Default for RHF) From orbital n1 to orbital n2 n1 occupied and n2 virtual orbitals UNO occupation numbers in UNO-CI [13] The default for UHF is "UNO,0.02,1.98"
--	--	---

**State=<n>** State to be calculated (default = 1; ground state)

**Osc[=<f>]** Calculate oscillator strengths for the excitations [up to <f> eV (default = 5.0)]

<n> positive number, <i> integer, <f> floating-point number, <s> text string.

## 4 PERIODIC CALCULATIONS

EMPIRE™ performs calculations with periodic boundary conditions if translation vectors (atom type **TV**) are specified. The user can supply up to three translation vectors in line format at the bottom of the input file (.xyz and .dat formats are supported). In .dat files, optimization flags (1 or 0) must be given for each lattice coordinate, just as for the atomic coordinates (see examples below).

We recommend performing geometry optimizations of atomic coordinates and lattice vectors separately, by using the optimization flags in the .dat input format. A robust procedure is to first optimize the atomic coordinates and then the lattice vectors. This can be repeated, as necessary, to obtain a fully relaxed structure. In troublesome cases, it may be helpful to increase the SCF convergence criterion during the lattice optimization runs.

Example .xyz:

```

72
ZnO EMPIRE Input
Zn -0.00000 1.87595 0.00000
O -0.00000 1.87595 1.99161
Zn 1.62462 0.93797 2.60273
O 1.62462 0.93797 4.59433
[...]
Zn 3.24928 7.50387 5.20545
O 3.24928 7.50387 7.19705
Zn 4.87390 6.56590 7.80818
O 4.87390 6.56590 9.79978
TV 9.70780 0.00000 0.00000
TV 0.00000 8.44184 0.00000
TV 0.00000 0.00000 10.41090
    
```

**Figure 4** .xyz input example for a three-dimensional periodic calculation

Example .dat:

```

Hamil=AM1 Calc=Opt
ZnO EMPIRE Input
Optimization of lattice vectors at fixed atomic coordinates
Zn 0.00000 0 1.87595 0 0.00000 0
O 0.00000 0 1.87595 0 1.99161 0
Zn 1.62462 0 0.93797 0 2.60273 0
O 1.62462 0 0.93797 0 4.59433 0
[...]
Zn 3.24928 0 7.50387 0 5.20545 0
O 3.24928 0 7.50387 0 7.19705 0
Zn 4.87390 0 6.56590 0 7.80818 0
O 4.87390 0 6.56590 0 9.79978 0
TV 9.70780 1 0.00000 0 0.00000 0
TV 0.00000 0 8.44184 1 0.00000 0
TV 0.00000 0 0.00000 0 10.41090 1
    
```

**Figure 5** .dat input example for a three-dimensional periodic calculation

## 5 MOLECULAR DYNAMICS

The keyword **CALC=MD** specifies that a Born-Oppenheimer molecular dynamics simulation is performed. Currently non-periodic and periodic simulations can be performed in the NVT and NVE ensembles. In NVT simulations, the temperature is regulated via the Berendsen thermostat. The simulation parameters are set via the keywords **TimeStep**, **MDsteps**, **Temp**, **RelaxTime** and **Ensemble=[<s>]**. Times are given in fs, temperatures in K.

If **ReadVelocities** is specified, the initial atomic velocities are read from the input file, where they should be specified after each coordinate (see example). This feature is especially useful for restarting simulations. If **ReadVelocities** is not specified, the velocities are initialized randomly to the requested temperature.

Please note that the optimization flags also work in MD simulations. This can be used to constrain the geometry of the system partially.

```
Calc=MD TimeStep=0.5 RelaxTime=400 MDsteps=1000 ReadVelocities
H2O EMPIRE Input
Molecular Dynamics Simulation
H 8.9125 1 -10.6523 1 -7.2520 1 -0.0023 0.0023 0.0095
H 7.5429 1 -11.3054 1 -7.0816 1 0.0235 -0.0240 -0.0017
O 8.4326 1 -11.3652 1 -6.8007 1 0.0027 -0.0001 -0.0070
```

**Figure 6** .dat input example for a molecular dynamics simulation

## 6 PROGRAM OUTPUT

The EMPIRE™ output for an adamantane single point (using the input shown in [Figure 1](#) and the command `empire.exe adamantane.xmol hamil=AM1`) is shown in [Figure 7](#).

```

EEE M M PPP III RRR EEE
E MM MM P P I R R E
EE M M M PPP I RRR EE
E M M M P I R R E
EEE M M P III R R EEE

EMPIRE19 OMP (Revision 2633)

(c) Matthias Hennemann, Johannes Margraf, David Whitley, Tim Clark*

References:
=====
1) Matthias Hennemann, Timothy Clark,
"EMPIRE: a highly parallel semiempirical molecular orbital program:
1: self-consistent field calculations",
J Mol Model (2014) 20:2331 (11 pages).
DOI 10.1007/s00894-014-2331-4
2) Johannes T. Margraf, Matthias Hennemann, Bernd Meyer, Timothy Clark,
"EMPIRE: a highly parallel semiempirical molecular orbital program:
2: periodic boundary conditions",
J Mol Model (2015) 21:144 (7 pages).
DOI 10.1007/s00894-015-2692-3

<> Computing resources

Number of cores = 8
MPI tasks = 1
OMP threads = 8
OMP max threads = 8
OMP num procs = 16

<> Options

Input file = Adamatane.xyz

Molecular Charge Charge = 0
Multiplicity Mult = 1

Hamiltonian Hamil = AM1
Formalism Form = RHF
Calculation Type Calc = SPT

Initial Guess Guess = DEFAULT
Randomize initial guess Randomize = 1.0E-03
Initial Diagonalization InitDiag = F
Mix initial UHF orbitals MixMOs = F

Simple screening radius ScreeningR = 40.000 Angstrom
Integral damping radius DampingR = 10.000 Bohr
Integral damping coefficient DampingC = 0.250

Maximum computation time MaxTime = 31536000 seconds
Maximum optimization steps MaxOPT = 10000
Maximum SCF Iterations MaxSCF = 1000

Converge gradient norm ConvergeG = 4.0E-01
Converge delta (HoF) ConvergeE = 1.0E-04 kcal/mol
Converge max (CFC) ConvergeR = 1.0E-04

Dynamic level shifting MinBandGap = 1.0E+00
Resonance integrals cutoff RCutoff = 2.0E+01
Pseudodiagonalization cutoff PDCutoff = 4.0E-02
Eigenvalues cutoff EiCutoff = 1.0E-02

Do full diagonalizations FullDiag = 1
Memory reserved for integrals IMemory = 1024 MB/thread

```







**Figure 8** shows the EMPIRE™ output for the same input file but using the PM3 Hamiltonian for a full geometry optimization (`empire.exe adamantane.xmol hamil=PM3 calc=OPT`)

```

EEE M M PPP III RRR EEE
E MM MM P P I R R E
EE M M M PPP I RRR EE
E M M M P I R R E
EEE M M P III R R EEE
    
```

EMPIRE19 OMP (Revision 2633)

(c) Matthias Hennemann, Johannes Margraf, David Whitley, Tim Clark\*

References:

=====

- 1) Matthias Hennemann, Timothy Clark,  
 "EMPIRE: a highly parallel semiempirical molecular orbital program:  
 1: self-consistent field calculations",  
 J Mol Model (2014) 20:2331 (11 pages).  
 DOI 10.1007/s00894-014-2331-4
- 2) Johannes T. Margraf, Matthias Hennemann, Bernd Meyer, Timothy Clark,  
 "EMPIRE: a highly parallel semiempirical molecular orbital program:  
 2: periodic boundary conditions",  
 J Mol Model (2015) 21:144 (7 pages).  
 DOI 10.1007/s00894-015-2692-3

<> Computing resources

```

Number of cores = 8
MPI tasks      = 1
OMP threads    = 8
OMP max threads = 8
OMP num procs  = 16
    
```

<> Options

```

Input file                      = Adamatane-PM3-OPT.dat

Molecular Charge                Charge = 0
Multiplicity                    Mult = 1

Hamiltonian                    Hamil = PM3
Formalism                      Form = RHF
Calculation Type               Calc = OPT

Initial Guess                   Guess = DEFAULT
Randomize initial guess        Randomize = 1.0E-03
Initial Diagonalization        InitDiag = F
Mix initial UHF orbitals       MixMOs = F

Simple screening radius         ScreeningR = 40.000 Angstrom
Integral damping radius        DampingR = 10.000 Bohr
Integral damping coefficient    DampingC = 0.250

Maximum computation time       MaxTime = 31536000 seconds
Maximum optimization steps     MaxOPT = 10000
Maximum SCF Iterations         MaxSCF = 1000

Converge gradient norm        ConvergeG = 4.0E-01
Converge delta (HoF)          ConvergeE = 1.0E-04 kcal/mol
Converge max (CFC)            ConvergeR = 1.0E-04

Dynamic level shifting         MinBandGap = 1.0E+00
Resonance integrals cutoff     RCutoff = 2.0E+01
Pseudodiagonalization cutoff  PDCutoff = 4.0E-02
Eigenvalues cutoff             EiCutoff = 1.0E-02

Do full diagonalizations       FullDiag = 1
Memory reserved for integrals  IMemory = 1024 MB/thread
Recalculate Hcore              RecalcH = F
Full rotations                 FullRot = F

Preferred Eigensolver          Eigensolver = DSYEVD
    
```



```
<> Parameters

/usr/local/cepos/cepos19/etc/PM3.par

Hamiltonian = PM3
Reference   = J.J.P. Stewart, J. Mol. Model., 10, 155-164 (2004).

H      J.J.P. Stewart, J. Comp. Chem., 10, 209-220 (1989).
Li     E. Anders, R. Koch, P. Freunsch, J. Comput. Chem., 14, 1301-1312 (1993).
Be     J.J.P. Stewart, J. Comp. Chem., 12, 320-341 (1991).
B      J.J.P. Stewart, J. Mol. Model., 10, 155-164 (2004).
C      J.J.P. Stewart, J. Comp. Chem., 10, 209-220 (1989).
N      J.J.P. Stewart, J. Comp. Chem., 10, 209-220 (1989).
O      J.J.P. Stewart, J. Comp. Chem., 10, 209-220 (1989).
F      J.J.P. Stewart, J. Comp. Chem., 10, 209-220 (1989).
Na     J.J.P. Stewart, J. Mol. Model., 10, 155-164 (2004).
Mg     J.J.P. Stewart, J. Comp. Chem., 12, 320-341 (1991).
Al     J.J.P. Stewart, J. Comp. Chem., 10, 209-220 (1989).
Si     J.J.P. Stewart, J. Comp. Chem., 10, 209-220 (1989).
P      J.J.P. Stewart, J. Comp. Chem., 10, 209-220 (1989).
S      J.J.P. Stewart, J. Comp. Chem., 10, 209-220 (1989).
Cl     J.J.P. Stewart, J. Comp. Chem., 10, 209-220 (1989).
K      J.J.P. Stewart, J. Mol. Model., 10, 155-164 (2004).
Ca     J.J.P. Stewart, J. Mol. Model., 10, 155-164 (2004).
Zn     J.J.P. Stewart, J. Comp. Chem., 12, 320-341 (1991).
Ga     J.J.P. Stewart, J. Comp. Chem., 12, 320-341 (1991).
Ge     J.J.P. Stewart, J. Comp. Chem., 12, 320-341 (1991).
As     J.J.P. Stewart, J. Comp. Chem., 12, 320-341 (1991).
Se     J.J.P. Stewart, J. Comp. Chem., 12, 320-341 (1991).
Br     J.J.P. Stewart, J. Comp. Chem., 10, 209-220 (1989).
Rb     J.J.P. Stewart, J. Mol. Model., 10, 155-164 (2004).
Sr     J.J.P. Stewart, J. Mol. Model., 10, 155-164 (2004).
Cd     J.J.P. Stewart, J. Comp. Chem., 12, 320-341 (1991).
In     J.J.P. Stewart, J. Comp. Chem., 12, 320-341 (1991).
Sn     J.J.P. Stewart, J. Comp. Chem., 12, 320-341 (1991).
Sb     J.J.P. Stewart, J. Comp. Chem., 12, 320-341 (1991).
Te     J.J.P. Stewart, J. Comp. Chem., 12, 320-341 (1991).
I      J.J.P. Stewart, J. Comp. Chem., 10, 209-220 (1989).
Cs     J.J.P. Stewart, J. Mol. Model., 10, 155-164 (2004).
Ba     J.J.P. Stewart, J. Mol. Model., 10, 155-164 (2004).
Hg     J.J.P. Stewart, J. Comp. Chem., 12, 320-341 (1991).
Tl     J.J.P. Stewart, J. Comp. Chem., 12, 320-341 (1991).
Pb     J.J.P. Stewart, J. Comp. Chem., 12, 320-341 (1991).
Bi     J.J.P. Stewart, J. Comp. Chem., 12, 320-341 (1991).

<><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><>
<> Structure

adamantane
Molecular Formula = C10H16
Molecular Mass    = 136.236
Charge            = 0
Multiplicity      = 1

Number of atoms           = 26
Number of electrons       = 56
Number of orbitals        = 56
Number of occupied orbitals = 28

<> Extended-Hueckel initial guess

<> Geometry optimization

Step  Cycles  Time  Time left    Gnrm  Heat of Formation  NIMAG    Lowest Eigenvalues
1      14    0.0  31536000.0   49.101  -33.2804          0   176.64  179.44  179.51
2       8    0.0  31536000.0   43.348  -33.8359          0   173.38  176.65  179.53
3       7    0.0  31536000.0  10.411  -34.5161          0   172.26  176.64  179.43
4       6    0.0  31536000.0   9.013  -34.5484          0   164.33  176.69  179.32
5       4    0.0  31536000.0   2.180  -34.5779          0   158.35  176.52  179.31
6       4    0.0  31536000.0   1.529  -34.5815          0   135.20  174.97  178.39
7       4    0.0  31535999.9   0.683  -34.5827          0   139.50  172.00  178.32
8       2    0.0  31535999.9   0.374  -34.5829

<> Geometry optimized using EF

Heat of formation           = -34.582928 kcal/mol
Electronic energy           = -9099.493460 eV
Core-core repulsion         = 7666.266642 eV
Ionization potential        = 10.851278 eV
Electron affinity           = -3.573067 eV
Computation time            = 0.043 seconds
```





## 7 ADDITIONAL FILE FORMATS

EMPIRE™ writes the output from the input file `<filename>.xyz`, `<filename>.xmol`, `<filename>.dat` or `<filename>.sdf` to the output file `<filename>_e.log`. It also generates an archive file `<filename>_e.arc` and if requested a binary wavefunction file `<filename>.vwf`. EMPIRE™ calculations that use the SDF format as input also write an output SDF file with the name `<filename>_e.sdf`. The formats of these files are defined below:

### 7.1 EMPIRE™ archive file

The archive file written by EMPIRE™ for the calculation whose output is shown in [Figure 8](#) is shown below in [Figure 9](#).

```

*** Summary of PM3 Calculation ***
*** EMPIRE'19 (Revision 2633) ***

<><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><>

adamantane
Molecular Formula = C10H16
Charge = 0
Multiplicity = 1

<> Geometry optimization
<> Optimization completed

Heat of formation = -34.582928 kcal/mol
Electronic energy = -9099.493460 eV
Core-core repulsion = 7666.266642 eV
No. of filled levels = 28
Ionization potential = 10.851278 eV
Electron affinity = -3.573067 eV
SCF calculations = 8
Computation time = 0.043 seconds

<> Final geometry obtained

26
adamantane
C -2.809329 -0.424015 -0.757182 -0.093948
C -2.180468 0.470810 -1.827842 -0.072729
C -0.684018 0.165731 -1.930469 -0.093953
C -0.011811 0.433569 -0.581715 -0.072791
C -0.642760 -0.461337 0.487531 -0.093941
C -2.139207 -0.157369 -0.592802 -0.072725
C -2.374871 1.939353 -1.442639 -0.093919
C -1.705442 2.208650 -0.092959 -0.072767
C -0.209496 1.901944 -0.197449 -0.093992
C -2.334449 1.311651 0.975827 -0.093958
H -2.699568 -1.489852 -1.035982 0.050977
H -3.897758 -0.232753 -0.688877 0.050963
H -2.669039 0.275576 -2.810638 0.060805
H -0.219397 0.787521 -2.719927 0.050903
H -0.527534 -0.887261 -2.234524 0.050974
H 1.078143 0.211813 -0.657886 0.060826
H -0.148676 -0.297684 1.464606 0.050988
H -0.485014 -1.527936 0.236098 0.050960
H -2.597914 -0.809758 1.371951 0.060776
H -3.453682 2.182709 -1.390678 0.050959
H -1.946258 2.599559 -2.221271 0.050966
H -1.848128 3.278255 0.187158 0.060833
H 0.293937 2.117490 0.764705 0.050953
H 0.265504 2.561497 -0.949235 0.050911
H -3.412600 1.540087 1.082108 0.050987
H -1.876616 1.514231 1.963268 0.050944
  
```

Figure 9 The EMPIRE™ archive file for a PM3 geometry optimization of adamantane

The calculation results are summarized followed by a table of the Cartesian coordinates of the optimized geometry and the Coulson net atomic charges in `.xyz` format.



## 7.2 HDF file format

HDF5 is a binary file format supported by the HDF Group ([www.hdfgroup.org](http://www.hdfgroup.org)). It was designed for storing hierarchical data structures containing large numerical arrays. The overall structure of an HDF5 file is a tree with nodes called *Groups*. Each group may contain other groups, *Attributes* and *Datasets*. Attributes are intended for small data items and datasets store n-dimensional arrays. The hierarchy is based at the *root* group, denoted “/”, and objects are located via paths, e.g. “/group1/group2/datasetD”. This mimics the structure of a Unix file system, with groups in place of directories, and attributes and datasets in place of files. The conventional file extension for HDF5 files is “.h5”.

A cross-platform visual tool, HDFView, for browsing and editing HDF5 files, is available from the HDF Group site. Alternatively, the HDF5 distribution contains a command-line utility program, h5dump, which converts a binary HDF5 file into a readable Data Definition Language (DDL) format. The usage is simply “h5dump <file>.h5” and the output is written to the standard output channel. The “-H” option restricts the output to the header information (without the data), allowing the overall structure of the file to be seen more clearly.

Internally, HDF5 stores arrays of rank > 1 in row-major (C-style) order. This has implications for files that will be accessed by programs written in both Fortran and C-style languages (i.e. C/C++ and Java). When reading and writing matrices from Fortran, the conversion from the column-major memory storage used by Fortran to the row-major storage in the HDF5 file is transparent and can be ignored. However, a C/C++ or Java program will see the transpose of a matrix written by a Fortran program, and vice-versa. This applies, in particular, to h5dump (written in C) and HDFView (written in Java), which display the transpose of all matrices in HDF5 files written by EMPIRE™ (written in Fortran).

The overall structure of an HDF5 Wavefunction file is shown in **Table 2**. All HDF5 object names are in uppercase with no spaces, with multiple words broken by underscores where appropriate.

**Table 2** HDF5 wavefunction file structure.

Group / (root)			
	Attribute CH5TYPE		
	Attribute CH5VERSION		
	Attribute PROGRAM		
	Group PROGRAM_OPTIONS		
	Group MOLECULES		
		Group MOLECULE_ID	
			Group IDENTITY
			Group STRUCTURE
			Group WAVEFUNCTION
			Group HAMILTONIAN

The root group contains three attributes:

- **CH5TYPE** A string defining the file format.
- **CH5VERSION** A compound data type comprising three integers representing a triplet version number (majnum, minnum, relnum).
- **PROGRAM** A string describing the program used to generate the file.

The **PROGRAM\_OPTIONS** group contains a set of attributes recording the command-line options used to create the file.

The **MOLECULES** group contains a list of molecule groups labelled by a **MOLECULE\_ID** "1", "2", "3", etc. The data for each molecule is then contained in the following groups:

- **IDENTITY** Molecule identifiers (Name, ID etc.).
- **STRUCTURE** Molecular structure (atomic numbers, charges and coordinates).
- **WAVEFUNCTION** The wavefunction Eigenvalues and Eigenvectors.
- **HAMILTONIAN** The Hamiltonian parameters for the atom types in the molecule.

Precise details of the entries in the HDF5 Wavefunction file are available on request.

## 7.3 Binary wavefunction file (deprecated, replaced by the HDF5 file)

The binary wavefunction file contains all the information necessary to process the results of the EMPIRE™ calculation further. Its contents are:

**Table 3** Contents of binary wavefunction file

Definition	Type
First comment line from the input	character(len=80)
Title line from the input	character(len=80)
Hamiltonian	character(len=6) "AM1 ", "AM1* ", "MNDO ", "MNDO/c", "MNDO/d", "PM3 " or "PM6 "
Formalism	character(len=3) "RHF"
Number of atoms	integer
Number of orbitals (Norbs)	integer
Number of doubly occupied orbitals	integer
Number of singly occupied orbitals	integer = 0
Charge on the molecule	integer
Heat of formation in kcal mol <sup>-1</sup>	double precision





Energy of the HOMO (eV)	double precision
Energy of the LUMO (eV)	double precision
x, y and z-components of the dipole moment (Debye)	double precision(1:3)
For each atom:	
Atomic number	integer
Cartesian coordinates	double precision(1:3)
Number of atomic orbitals	integer
Principal quantum number (s and p)	integer
Principal quantum number (d)	integer
Slater exponents ( $\zeta_s$ , $\zeta_p$ , $\zeta_d$ )	double precision(1:3)
Overlap integrals ( $\beta_s$ , $\beta_p$ , $\beta_d$ )	double precision(1:3)
Multipole parameters (dd and qq)	double precision(1:2)
End atoms	
Eigenvalues (eV)	Double precision(1:Norbs)
Eigenvectors	Double precision(1:Norbs2)

## 7.4 SDF output file

The output file for a PM3 optimization of adamantane using an [sdf](#) file as input is shown in [Figure 10](#).

<pre>adamantane XXEMPIRE1909131915043D 1 1.00000 0.00000 0 EMPIRE input file 26 28 0 0 0 0 1 V2000 -2.8093 -0.4240 -0.7572 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -2.1805 0.4708 -1.8278 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -0.6840 0.1657 -1.9305 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -0.0118 0.4335 -0.5817 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -0.6428 -0.4613 0.4875 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -2.1392 -0.1574 0.5928 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -2.3749 1.9393 -1.4426 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -1.7054 2.2086 -0.0930 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -0.2095 1.9019 -0.1975 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -2.3345 1.3116 0.9758 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -2.6996 -1.4899 -1.0360 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -3.8978 -0.2328 -0.6889 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -2.6691 0.2756 -2.8106 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -0.2194 0.7875 -2.7199 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -0.5275 -0.8873 -2.2345 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1.0781 0.2118 -0.6579 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -0.1487 -0.2977 1.4646 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -0.4850 -1.5280 0.2361 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -2.5979 -0.8098 1.3720 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -3.4537 2.1827 -1.3907 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -1.9463 2.5995 -2.2213 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -1.8481 3.2782 -0.1871 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.2939 2.1175 0.7647 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.2655 2.5615 -0.9492 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -3.4126 1.5401 1.0821 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -1.8766 1.5142 1.9633 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 2 1 0 0 0 0 2 3 1 0 0 0 0 3 4 1 0 0 0 0 4 5 1 0 0 0 0 1 6 1 0 0 0 0 5 6 1 0 0 0 0 2 7 1 0 0 0 0 7 8 1 0 0 0 0 4 9 1 0 0 0 0 8 9 1 0 0 0 0 6 10 1 0 0 0 0 8 10 1 0 0 0 0 1 11 1 0 0 0 0 1 12 1 0 0 0 0 2 13 1 0 0 0 0 3 14 1 0 0 0 0 3 15 1 0 0 0 0 4 16 1 0 0 0 0 5 17 1 0 0 0 0 5 18 1 0 0 0 0 6 19 1 0 0 0 0 7 20 1 0 0 0 0 7 21 1 0 0 0 0 8 22 1 0 0 0 0 9 23 1 0 0 0 0 9 24 1 0 0 0 0 10 25 1 0 0 0 0</pre>	<pre>&gt; &lt;DENSITY MATRIX ELEMENTS&gt; [EMPIRE'19] 1.1649010188 -0.0092583420 0.9872339531 -0.0083019506 -0.0238931812 0.9886293361 -0.0019812092 -0.0068350971 0.0103237086 0.9531831276 1.1777105113 -0.0080399042 0.9605157843 -0.0031983539 0.0024487034 0.9553894998 -0.0161502128 0.0122265007 0.0049011145 0.9791142179 1.1649248277 0.0058208759 0.9553549254 -0.0041064693 0.0083778716 1.0090064991 -0.0102998017 -0.0084285160 -0.0128602285 0.9646666694 1.177183682 0.0179370449 0.9847411612 -0.0036464207 -0.0061621752 0.9557184320 -0.0012551957 -0.0021265316 0.0004076401 0.9546134968 1.1648772333 0.0061558476 0.9560518443 -0.0085525610 0.0037882462 0.9876984859 0.0069051712 0.0121476912 0.0255867880 0.9853135972 1.177339678 -0.0075098695 0.9597838382 -0.0107283085 0.0076593669 0.9652797763 0.0127848620 -0.0091381454 -0.0129502522 0.9699266451 1.1648882854 -0.0061817242 1.0000063387 0.0085166236 0.0081813898 0.9603948040 -0.0068959272 -0.0217592664 -0.0122709343 0.9686300688 1.1776978226 -0.0024018687 0.9549764303 0.0176050059 -0.0038486331 0.9836402544 0.0046024292 -0.0010093177 0.0076364538 0.9564526403 1.1649382199 0.0091709167 0.9579887718 0.0082581031 0.0065308208 0.9604895385 0.0020138107 0.0026471599 -0.0140810950 1.0105757961 1.1649176996 -0.0059207234 1.0014526107 0.0040402132 -0.0029211938 0.9518668554 0.0103332584 0.0222004429 0.0032740892 0.9757216968  &gt; &lt;CHARGE ON HYDROGENS&gt; [EMPIRE'19] 0.0509776740 0.0509635078 0.0608052760 0.0509028691 0.0509739160 0.0608261410 0.0509880032 0.0509601503 0.0607755786 0.0509593785 0.0509654858 0.0608320087 0.0509530866 0.0509114040 0.0509867651 0.0509438014  &gt; &lt;ORBITAL VECTORS&gt; [EMPIRE'19] RHF calculation 28 56 28 Nhom0, Nao, Nocc, Nunocc Nr. of atomic orbitals of each atom 4 4 4 4 4 4 4 4 4 4 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 Orbital information from occ. to unocc. -44.608 -0.266 -0.047 -0.042 -0.010 -0.301 -0.026 -0.010 -0.053 -0.266 0.030 -0.021 -0.052 -0.301 0.059 -0.012 -0.004 -0.266 0.031 -0.043 0.035 -0.301 -0.025 -0.035 0.042 -0.266 -0.031 0.043 -0.035 -0.301 -0.008 0.058 0.015 -0.266 0.047 0.042 0.010 -0.266 -0.030 0.021 0.052 -0.104 -0.104 -0.105 -0.104 -0.104 -0.105 -0.104 -0.104 -0.105 -0.104 -0.104 -0.105 -0.104 -0.104 -0.104 -0.104  Further orbitals not shown here  5.334 0.082 0.256 0.085 0.082 0.051 -0.093 0.140 -0.259 0.049 -0.008 -0.106 0.223 -0.013 0.000 0.250 0.177 0.001 -0.012 -0.167 -0.201 0.013 0.068 0.151 0.258 -0.001 -0.025 0.104 0.147 -0.051 0.182 -0.245 0.046 -0.082 0.145 0.239 -0.038 -0.049 -0.192 0.153 0.016 0.021 0.136 -0.198 0.143 -0.050 0.050 0.153 -0.152</pre>
---	--



```
10 26 1 0 0 0 0
M END

> <CALCULATION> [EMPIRE'19]
8
Geometry optimized using EF

> <HAMILTONIAN> [EMPIRE'19]
PM3

> <MOPACBASICS> [EMPIRE'19]
-34.582931 -10.851285 3.573069 0.000170 0.000084 -
0.000038

> <ENERGY> [EMPIRE'19]
-9099.494285 7666.267466 -1433.226818

> <COULSON CHARGES> [EMPIRE'19]
-0.0939474355 -0.0727300133 -0.0939529215 -0.0727914581 -0.0939411607
-0.0727242238 -0.0939194968 -0.0727671476 -0.0939923263 -0.0939588625
0.0509776740 0.0509635078 0.0608052760 0.0509028691 0.0509739160
0.0608261410 0.0509880032 0.0509601503 0.0607755786 0.0509593785
0.0509654858 0.0608320087 0.0509530866 0.0509114040 0.0509867651
0.0509438014

> <TOTAL COULSON CHARGE> [EMPIRE'19]
0.000000000

-0.052 -0.039 0.038 0.200 -0.023 -0.134 -0.142 0.049
5.335
0.001 0.124 -0.094 -0.159 -0.038 0.017 0.270 0.145
-0.003 -0.134 -0.181 -0.017 0.035 0.190 0.139 -0.197
0.095 -0.142 0.207 -0.163 0.039 -0.296 -0.010 0.085
-0.095 -0.155 0.199 -0.161 -0.036 -0.098 -0.127 -0.263
-0.001 -0.075 0.038 0.212 0.003 0.188 0.113 0.050
-0.106 0.108 0.148 0.109 -0.115 -0.139 0.089 0.093
-0.151 -0.096 -0.086 0.142 -0.116 0.114 0.113 -0.107
5.486
-0.025 -0.190 -0.170 -0.041 0.089 0.131 0.053 0.263
-0.025 0.121 -0.084 -0.213 0.089 -0.292 0.059 0.020
-0.025 0.127 -0.175 0.142 0.089 0.122 0.174 -0.208
-0.025 -0.127 0.176 -0.141 0.089 0.039 -0.286 -0.075
-0.025 0.191 0.170 0.041 -0.025 -0.121 0.084 0.212
-0.089 -0.088 0.168 -0.089 -0.089 0.168 -0.089 -0.088
0.167 -0.089 -0.088 0.168 -0.089 -0.088 -0.089 -0.088

> <CALCULATION STATUS> [EMPIRE'19]
SUCCESS
$$$$
```

Figure 10 The EMPIRE™ output .sdf file for a PM3 geometry optimization of adamantane



All EMPIRE™-specific tags use the suffix (**EMPIRE' 19**). These are:

Table 4 EMPIRE™ specific tags

Tag	Contents	
>>CALCULATION>	1 Single point calculation	<n> is the number of optimization cycles (always 1 for a single point)
	<n> Geometry optimized using EF	
>>HAMILTONIAN>	AM1, AM1*, MNDO, MNDOD, PM3, PM6, RM1 ...	
>>MOPACBASICS>	1. Heat of formation 2. HOMO energy (eV) 3. LUMO energy (eV) x-, y-, and z-components of the dipole moment (Debye)	Format: 6F13.6
>>CHARGE>	Molecular charge as integer	
>>ENERGY>	1. Electronic energy 2. Core-core repulsion 3. Total energy	All in eV, format 3F20.6
>>COULSON CHARGES>	Coulson net atomic charges (one floating point per atom)	Format: 5F14.10
>>TOTAL COULSON CHARGE>	Sum of the <b>Coulson charges</b> (floating point)	Format: F14.10
>>DENSITY MATRIX ELEMENTS>	The one-atom blocks of the density matrix printed as lower triangles for each atom (10 elements (2 lines) for s,p-atoms and 45 elements (9 lines) for s,p,d-atoms. Hydrogen blocks are not included.	Format: 5F14.10
>>CHARGE ON HYDROGENS>	The <b>Coulson charges</b> of the hydrogen atoms (this block is redundant as the charges are also contained in the ">>COULSON CHARGES>" block).	Format: 5F14.10
>>ORBITAL VECTORS>	Block for the molecular orbitals Eigenvectors	
RHF calculation	1. Number of the HOMO ( <b>Nhomo</b> ) 2. Number of orbitals ( <b>Na0</b> ) 3. Number of occupied orbitals ( <b>Nocc</b> ) 4. Number of virtual orbitals ( <b>Nunocc</b> )	Format: 4(i4,1x),
	Nr. of atomic orbitals of each atom	Format: 20i4
	Orbital information from occ. to unocc.	<b>For each molecular orbital:</b> The Eigenvalue (eV) The MO coefficients
>>CALCULATION STATUS>	SUCCESS	

## 8 GLOSSARY ENTRIES

### CFC elements

“CFC elements” are the calculated interaction matrix elements between virtual and occupied orbitals that must be made zero in the SCF procedure. They are the basis for the pseudodiagonalization procedure.[10]

### Configuration interaction

In configuration interaction calculations, the molecular orbitals of the reference wavefunction (either be the RHF ground state or the UHF natural orbitals in EMPIRE™) are used to construct microstates by exciting electrons into virtual orbitals. These microstates are allowed to mix to give new wavefunctions for ground and excited states.

### Convergence on the density matrix

At SCF convergence, the density matrix should not change between iterations. This can be tested by calculating the maximum difference of any density-matrix element between cycles. This is the usual procedure in SCF calculations but is impractical for very large systems, so that by default, EMPIRE™ uses the CFC criterion, which is essentially equivalent to checking the convergence on the density matrix.

### Coulson charge

Adapted from T.Clark, in Chemoinformatics – From Data to Knowledge (Ed.: J. Gasteiger), WILEY-VCH, Weinheim, **2003**, pp. 947-976.

The number of electrons assigned to an atom is expressed as the sum of the populations of the atomic orbitals (AOs) centered at its nucleus. The Coulson analysis assumes that the orbitals are orthogonal, which leads to the very simple expression that the electronic population of  $P_i$  of atom  $i$  is given by the following equation:

$$P_i = \sum_{j=1}^{N_{occ}} \sum_{k=i_{first}}^{i_{last}} n_j c_{j,k}^2$$

Where  $N_{occ}$  is the number of occupied molecular orbitals (MOs),  $i_{first}$  and  $i_{last}$  are the first and last atomic orbitals centered on atom  $i$ , respectively,  $n_j$  is the occupancy number of the molecular orbital  $j$ , and  $c_{i,k}$  is the coefficient of AO  $k$  in MO  $j$ . The net atomic charge is simply the sum of the electronic population of the atom and its nuclear charge.



## Dynamic level shifting

When solving the Hartree-Fock-Roothan equations problems concerning the convergence of the calculations sometimes arise. As a rule, this is connected with two problems: the slow convergence of iteration process and the occurrence of oscillations. These oscillations can be suppressed and convergence obtained using the dynamic level shift method, in which the energies of virtual orbitals are artificially increased.

The dynamic "level shift" method for improving the convergence of the SCF procedure, A. V. Mitin, Journal of Computational Chemistry, 1988, 9, 107-110.

## Gradient norm criterion

The gradient norm is the square root of the sum of the squares of the energy gradients with respect to geometric distortions. The gradient norm should be zero at a minimum or transition state. In practice, the optimization stops when a sufficiently small value is reached.

## Linear-scaling techniques

The computational effort scales linearly with the size (number of atoms) of the system for these techniques.

## One-electron matrix

The one-electron matrix contains the electron-nucleus interactions. It is used to form the Fock matrix, which is diagonalized in the SCF procedure. The one-electron matrix does not contain electron-electron interactions.

## Self-consistent field

An iterative procedure used to obtain the converged wavefunction starting from guessed molecular orbitals (the initial guess). The procedure is judged to be converged when neither the energy nor the wavefunction change significantly from iteration to iteration.

# 9 SUPPORT

## 9.1 Contact

Questions regarding EMPIRE™ should be sent directly to:

**[support@ceposinsilico.com](mailto:support@ceposinsilico.com)**

## 9.2 Cepas InSilico GmbH

Waldstrasse 15  
90587 Obermichelbach  
Germany

**[support@ceposinsilico.com](mailto:support@ceposinsilico.com)**

Tel. +49 (0)9131 9704910

Fax. +49 (0)9131 9704911

**[www.ceposinsilico.com/contact](http://www.ceposinsilico.com/contact)**



# 10 LIST OF TABLES

Table 1	EMPIRE™ program options .....	10
Table 2	HDF5 wavefunction file structure. ....	23
Table 3	Contents of binary wavefunction file .....	24
Table 4	EMPIRE™ specific tags .....	27



# 11 LIST OF FIGURES

Figure 1	.xyz (.xmol) input for adamantane .....	6
Figure 2	.sdf input for adamantane (the format is structure-data file V2000).....	7
Figure 3	.dat input for an AM1 optimization of adamantane.....	8
Figure 4	.xyz input example for a three-dimensional periodic calculation .....	14
Figure 5	.dat input example for a three-dimensional periodic calculation .....	14
Figure 6	.dat input example for a molecular dynamics simulation .....	15
Figure 7	The EMPIRE™ output for an AM1 single-point calculation on adamantane .	18
Figure 8	The EMPIRE™ output for a PM3 geometry optimization of adamantane .....	21
Figure 9	The EMPIRE™ archive file for a PM3 geometry optimization of adamantane .. .....	22
Figure 10	The EMPIRE™ output .sdf file for a PM3 geometry optimization of adamantane .....	26





## 12 REFERENCES

- [1] J. Baker. An algorithm for the location of transition states, *Journal of Computational Chemistry* **1986**, *7*, 385-395. doi:10.1002/jcc.540070402
- [2] a) M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, J. J. P. Stewart. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model, *J Am Chem Soc* **1985**, *107*, 3902-3909. doi:10.1021/ja00299a024; b) M. J. S. Dewar, C. Jie, J. Yu. SAM1; The first of a new series of general purpose quantum mechanical molecular models, *Tetrahedron* **1993**, *49*, 5003-5038. doi:10.1016/S0040-4020(01)81868-8; c) R. O. Freire, G. B. Rocha, A. M. Simas. Sparkle Model for the Calculation of Lanthanide Complexes: AM1 Parameters for Eu(III), Gd(III), and Tb(III)†, *Inorganic Chemistry* **2005**, *44*, 3299-3310. doi:10.1021/ic048530+; d) G. B. Rocha, R. O. Freire, A. M. Simas, J. J. P. Stewart. RM1: A reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I, *Journal of Computational Chemistry* **2006**, *27*, 1101-1111. doi:10.1002/jcc.20425; e) A. R. Leach, *Molecular Modelling: Principles and Applications*, Pearson Education Limited, **2001**; f) J. J. Stewart. Optimization of parameters for semiempirical methods IV: extension of MNDO, AM1, and PM3 to more main group elements, *Journal of Molecular Modeling* **2004**, *10*, 155-164. doi:10.1007/s00894-004-0183-z
- [3] a) P. Winget, A. C. Horn, C. Selçuki, B. Martin, T. Clark. AM1\* parameters for phosphorus, sulfur and chlorine, *Journal of Molecular Modeling* **2003**, *9*, 408-414. doi:10.1007/s00894-003-0156-7; b) P. Winget, T. Clark. AM1\* parameters for aluminum, silicon, titanium and zirconium, *Journal of Molecular Modeling* **2005**, *11*, 439-456. doi:10.1007/s00894-005-0236-y; c) H. Kayi, T. Clark. AM1\* parameters for copper and zinc, *Journal of Molecular Modeling* **2007**, *13*, 965-979. doi:10.1007/s00894-007-0214-7; d) H. Kayi, T. Clark. AM1\* parameters for vanadium and chromium, *Journal of Molecular Modeling* **2009**, *15*, 1253-1269. doi:10.1007/s00894-009-0489-y; e) H. Kayi, T. Clark. AM1\* parameters for bromine and iodine, *Journal of Molecular Modeling* **2009**, *15*, 295-308. doi:10.1007/s00894-008-0419-4; f) H. Kayi, T. Clark. AM1\* parameters for manganese and iron, *Journal of Molecular Modeling* **2010**, *16*, 1109-1126. doi:10.1007/s00894-009-0614-y; g) H. Kayi, T. Clark. AM1\* parameters for cobalt and nickel, *Journal of Molecular Modeling* **2010**, *16*, 29-47. doi:10.1007/s00894-009-0503-4; h) H. Kayi, T. Clark. AM1\* parameters for palladium and silver, *Journal of Molecular Modeling* **2011**, *17*, 2585-2600. doi:10.1007/s00894-010-0940-0
- [4] H. B. Thomas, M. Hennemann, P. Kibies, F. Hoffgaard, S. Gussregen, G. Hessler, S. M. Kast, T. Clark. The hpCADD NDDO Hamiltonian: Parametrization, *J Chem Inf Model* **2017**, *57*, 1907-1922. doi:10.1021/acs.jcim.7b00080
- [5] M. J. S. Dewar, W. Thiel. Ground states of molecules. 38. The MNDO method. Approximations and parameters, *J Am Chem Soc* **1977**, *99*, 4899-4907; 4907-4917. doi:10.1021/ja00457a004
- [6] a) W. Thiel, A. A. Voityuk. Extension of MNDO to d Orbitals: Parameters and Results for the Second-Row Elements and for the Zinc Group, *The Journal of Physical Chemistry* **1996**, *100*, 616-626. doi:10.1021/jp952148o; b) W. Thiel. Perspectives on Semiempirical Molecular Orbital Theory, in *Advances in Chemical Physics*, John Wiley & Sons, Inc., **2007**, pp. 703-757. doi:10.1002/9780470141526.ch10
- [7] a) J. J. P. Stewart. Optimization of parameters for semiempirical methods I. Method, *Journal of Computational Chemistry* **1989**, *10*, 209-220. doi:10.1002/jcc.540100208; b) J. J. P. Stewart. Optimization of parameters for semiempirical methods II. Applications, *Journal of Computational Chemistry* **1989**, *10*, 221-264. doi:10.1002/jcc.540100209
- [8] a) M. Kriebel, K. Weber, T. Clark. A Feynman dispersion correction: a proof of principle for MNDO, *J Mol Model* **2018**, *24*, 338. doi:10.1007/s00894-018-3874-6; b) M. Kriebel, A. Heßelmann, M. Hennemann, T. Clark. The Feynman dispersion correction for MNDO extended to F, Cl, Br and I, *Journal of Molecular Modeling* **2019**, *25*, 156. doi:10.1007/s00894-019-4038-z; c) M. Kriebel, A. Heßelmann, M. Hennemann, T. Clark. Correction to: The



- Feynman dispersion correction for MNDO extended to F, Cl, Br and I, *Journal of Molecular Modeling* **2019**, *25*, 257. doi:10.1007/s00894-019-4142-0
- [9] J. J. P. Stewart. Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements, *Journal of molecular modeling* **2007**, *13*, 1173-1213. doi:10.1007/s00894-007-0233-4
- [10] J. J. P. Stewart, P. Császár, P. Pulay. Fast semiempirical calculations, *Journal of Computational Chemistry* **1982**, *3*, 227-228. doi:10.1002/jcc.540030214
- [11] J. T. Margraf, M. Hennemann, B. Meyer, T. Clark. EMPIRE: A highly parallel semiempirical molecular orbital program: 2: Periodic Boundary Conditions, *Journal of Molecular Modeling* **2015**, *21*, 144. doi:10.1007/s00894-015-2692-3
- [12] T. Clark, J. Chandrasekhar. NDDO-Based CI Methods for the Prediction of Electronic Spectra and Sum-Over-States Molecular Hyperpolarization, *Israel Journal of Chemistry* **1993**, *33*, 435-448. doi:10.1002/ijch.199300050
- [13] P. O. Dral, T. Clark. Semiempirical UNO-CAS and UNO-CI: Method and Applications in Nanoelectronics, *The Journal of Physical Chemistry A* **2011**, *115*, 11303-11312. doi:10.1021/jp204939x