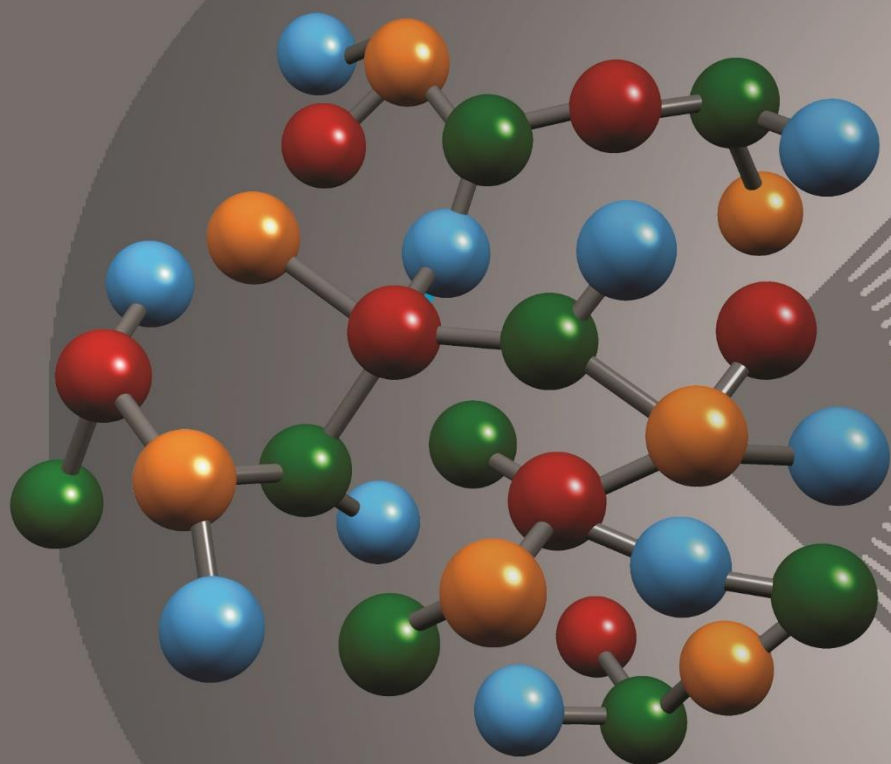


20

USER MANUAL



EMPIRE™

Impressum

Copyright

© 2020 by Cepas InSilico GmbH
Waldstrasse 15
90587 Obermichelbach
www.ceposinsilico.com

Manual

Timothy Clark

Layout

www.eh-bitartist.de





TABLE OF CONTENTS

1 INTRODUCTION	4
1.1 New in Empire20	5
2 INPUT FORMATS	6
2.1 .xyz and .xmol files	6
2.2 .sdf files	7
2.3 .dat files	8
3 PROGRAM OPTIONS	10
4 TRANSITION-STATE OPTIMIZATIONS	14
5 SCRF IMPLICIT SOLVENT MODEL	15
6 CONFIGURATION INTERACTION	17
7 PERIODIC CALCULATIONS	18
8 MOLECULAR DYNAMICS	19
9 PROGRAM OUTPUT	20
10 ADDITIONAL FILE FORMATS	26
10.1 EMPIRE™ archive file	26
10.2 HDF file format	27
10.3 Binary wavefunction file (deprecated, replaced by the HDF5 file)	28
10.4 SDF output file	29
11 GLOSSARY ENTRIES	32
12 SUPPORT	34
12.1 Contact	34
12.2 Cepas InSilico GmbH	34
13 LIST OF TABLES	35
14 LIST OF FIGURES	36
15 REFERENCES	37

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 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**.^[1] Later versions will provide other optimization options.

Citing EMPIRE™:

EMPIRE20, M. Hennemann, J. T. Margraf, B. Meyer and T. Clark, Cepas InSilico GmbH, Obermichelbach, 2020. <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>).

EMPIRE: A highly parallel semiempirical molecular orbital program: 3: Born-Oppenheimer molecular dynamics, J. T. Margraf, M. Hennemann and T. Clark, *J. Mol. Model.*, **2020**, *26*, 43. (9 pages, <https://doi.org/10.1007/s00894-020-4293-z>).

1.1 New in Empire20

EMPIRE20 contains several updates, extensions and improvements relative to the previous version. These are listed below:

Excited States	Configuration-interaction (CI) calculations are now available using the UHF Natural Orbital approach for both closed and open-shell systems (up to triplet reference states)
Transition-state optimization	EMPIRE20™ features a particularly reliable optimization algorithm for transition states.
Solvent simulation	A polarized continuum model (PCM) implicit solvent model has been implemented in EMPIRE20™ for ground states.
Atomic multipole analysis	The electrostatic properties of the molecule can now be obtained as an array of atom-centered monopoles, dipoles and quadrupoles.
Localized Molecular Orbitals	Localized orbitals can now be calculated and printed to the SDF or EH5 files. However, the current algorithm scales very poorly with molecular size and should therefore only be used for small molecules.
HDF5 Initial Guess	EMPIRE20™ can read the initial guess from EMPIRE™ _e.h5 files in addition to .vwf and .aux files.

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      -2.597509   -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. 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=<i></code>	Set the total molecular charge to <i>	
<code>Mult=<n></code>	Set the Multiplicity to <n>	
<code>Hamil=<s></code>	Set the semiempirical Hamiltonian to be used. The Hamiltonian files provided with EMPIRE™ allow the following values for <s>:	
	<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=<s></code>	Defines the formalism	
	<code>RHF</code>	Restricted Hartree-Fock (default for Mult=1)
	<code>UHF</code>	Unrestricted Hartree-Fock (default for Mult>1)
<code>Calc=<s></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. This is equivalent to <code>CALC=OPT, FREQ</code>
	<code>SPT, FREQ</code>	Perform a frequency calculation on the input structure without any optimization
	<code>OPT, FREQ</code>	First optimize the structure and then perform a frequency calculation
	<code>TSOPT, FREQ</code>	First perform a transition-state optimization and then a frequency calculation
<code>MD</code>	Perform a molecular dynamics simulation	
<code>TSOPT</code>	Optimize a transition state	
Initial Guess		
<code>Guess=<s></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>*.h5</code>	Read the Eigenvectors from the specified EMPIRE™ .h5 file
	<code>*.vwf</code>	Read the Eigenvectors from the specified .vwf file
<code>*.aux</code>	Read the Eigenvectors from the specified .aux file	
<code>Randomize=<f></code>	Randomize the initial guess matrix (Hückel or density) (default: 10 ⁻³)	
<code>InitDiag</code>	Do an initial full diagonalization (this is useful when reading the Eigenvectors from file, or to obtain a better UHF initial guess)	
<code>MixMOs [= -h, +l]</code>	Mix the HOMO-h and LUMO+l orbitals of an UHF initial guess	

Option	Details	
	(Default: 0,0 => HOMO,LUMO ; for "Mult=1")	
Input Options		
Center	Center the molecular structure	
Center=<n>	Center the molecular structure on atom <n>	
Center=<x>,<y>,<z>	Center the molecular structure on point x,y,z	
Orient	Orient the molecular structure according to its principal axes	
Project	Project atoms into the central lattice cell	
Output Options		
HDF [=<n>]	Write an HDF5 wavefunction file: <Basename>_e.h5 [add a structure every <n> MD steps]	
VWF	Write a vamp wavefunction file: <Basename>_e.vwf (deprecated; please use HDF instead)	
XYZ [=<n>]	Write an XYZ trajectory file: <Basename>_e.xyz [add a structure every <n> OPT/MD steps]	
DAT	Force output of a restart DAT file: <Basename>_e.dat	
SDF	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)	
SDFBonds=<s>	Type of bonds written to the output SDF	
	Keep	Do not alter bonds and bond types
	Check	Check if the connectivity changed (default)
	Ary1	Check connectivity and adjust aryl bond types to 4
	Calc	Calculate bond types from bond orders (bond type 4 is used for aromatic bonds)
Calculation Details		
MaxTime=<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)	
FullDiag=<n>	Do a full diagonalization every <n> OPT/MD steps (default = 1)	
Geometry Optimization (apply to both OPT and TSOPT)		
MaxOpt=<n>	Define the maximum number of optimization cycles (default = 10,000)	
ConvergeG=<f>	Set the gradient norm criterion for the geometry optimization (default = 0.4 kcal mol ⁻¹ Å ⁻¹)	
Molecular Dynamics		
MDsteps=<n>	Number of MD steps (default=100)	
MDsteps=<n1 , n2>	Use two numbers <n1 , n2> to specify the starting and ending step (used for restarting a MD calculation)	
TimeStep=<f>	MD time step in fs (default = 0.5)	
Temp=<f>	MD target temperature in K (default = 300.0)	
Ensemble=<s>	MD Ensemble	
	NVT	Constant number of particles, volume and temperature (default)
	NVE	Constant number of particles, volume and energy
Thermostat=<s>	Thermostat type	
	Berendsen	Berendsen thermostat (default)
	Langevin	Langevin thermostat
RelaxTime=<f>	Thermostat relaxation time in fs (default = 100.0)	
ReadVelocities	Read initial velocities from the DAT file (for restarting an MD calculation)	
FixCOM=<n>	Remove center of mass translation and rotation every <n> steps (default = 10000; only used for the Berendsen thermostat)	

Option	Details	
GuessAll	New initial guess at each MD step (Default for NVE calculations)	
Self-Consistent Field		
MaxSCF=<n>	Define the maximum number of SCF cycles (default = 1,000)	
ConvergeE=<f>	Set the energy convergence limit for the SCF (default = 10^{-4} kcal mol ⁻¹)	
ConvergeR=<f>	Set the maximum off-diagonal CFC element for SCF convergence (default = 10^{-4})	
CheckP	Use the convergence on the density matrix (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 pseudodiagonalization procedure . ^[10] This is equivalent to testing the density directly and far more suitable for large molecules.	
ConvergeP=<f>	Set the convergence limit for RSS(Density) (default = 10^{-6}). This option implies CheckP .	
IMemory=<n>	Use up to <n> MB/thread of memory for storing integrals (default = 1024; Unlimited for periodic calculations)	
RecalcH	Calculate the one-electron matrix 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.	
FullRot	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).	
MinBandGap=<f>	Use dynamic level shifting to ensure that the band gap used for the SCF does not fall below <f> (default = 1.0; disabled when set to <= 0.0)	
RCutoff=<f>	Defines the distance cutoff for the overlap / resonance integrals (default = 20.0)	
PDcutoff=<f>	Defines the cutoff value for CFC at which rotations are included in the pseudodiagonalization (default = 0.04 x the maximum CFC value)	
EiCutoff=<f>	Defines the maximum value of CFC below which the Eigenvalues will not be recalculated in the SCF cycle. The default is <f>= 0.0, which means that the Eigenvalues are always calculated. Setting EiCutoff to higher values will make calculations faster but may slow down or prevent SCF convergence.	
Eigensolver=<s>	Preferred Eigensolver	
	DSYEVR	DSYEVR or PDSYEVR (Default)
	DSYEVD	DSYEVD or PDSYEVD
	ELPA1	ELPA one-stage
	ELPA2	ELPA two-stage
Embedding Point Charges		
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: Q X-Coord. Y-Coord. Z-Coord. Charge DAT: Q X-Coord. Y-Coord. Z-Coord. Charge Embedded point charges cannot be optimized, and the optimization flags in the .dat file will be ignored.	
Self-Consistent Reaction-Field Calculations [11]		
SCRf=<s>	<s> is the solvent. A list of available solvents and acceptable names for them is given in the SCRf_solvents.par file and in Table 2 below.	

Option	Details	
Periodic Boundary Conditions [12]		
Up to three translation vectors can be specified at the end of the <code>.dat</code> or <code>.xyz</code> file input using the atom type "TV". If lattice optimization is requested via the optimization flags in the <code>.dat</code> file input, the first and second translation vector must be aligned with the x-axis and the xy-plane, respectively:		
	<pre>TV Vector1-X 0/1 0.0 0 0.0 0 TV Vector2-X 0/1 Vector2-Y 0/1 0.0 0 TV Vector3-X 0/1 Vector3-Y 0/1 Vector3-Z 0/1</pre>	
<code>ScreeningR=<f></code>	Simple screening radius beta in Ångström (default = 30.0)	
<code>DampingR=<f></code>	Integral damping radius in Bohr (default = 10.0)	
<code>DampingC=<f></code>	Integral damping coefficient (default = 0.25)	
Natural Orbitals		
<code>NOs</code>	Calculate natural orbitals	
Atomic Multipole Analysis		
<code>AtMult</code>	Calculate atomic multipoles up to quadrupole	
Localized Molecular Orbitals		
<code>Local</code>	Localize the molecular orbitals and print to the <code>.sdf</code> or <code>.eh5</code> file, whichever has been specified.	
Configuration Interaction		
<code>CI=*</code>	Type of CI calculation:	
	<pre>Full PECI 1 or S 2 or SD 3 or SDT 4 or SDTQ</pre>	Full CI Paired Electron CI [13] Singles (CISD) Singles and doubles (CISD) Singles, doubles and triples (CISDT) Singles, doubles, triples and quadruples ... and so on ...
<code>AS=<t>[,*,*]</code>	Active space, defined by <t>:	
	<pre>ALL MO,<n1>,<n2> OV,<n1>,<n2> UNO,<f1>[,<f2>]</pre>	All orbitals (Default for RHF) From orbital n1 to orbital n2. For UHF reference wavefunctions, this option selects UNOs n1 to n2 n1 occupied and n2 virtual orbitals UNO occupation numbers in UNO-CI [14] (The default for UHF is "UNO,0.02,1.98")
<code>REF=<t></code>	Reference wavefunction for UNO-CI calculations, defined by <t>:	
	<pre>AUTO MULT</pre>	Use the multiplicity derived from the UNO occupation numbers Use the multiplicity defined in the <code>MULT=<n></code> option (default)
<code>State=<n></code>	State to be calculated (default = 1; ground state)	
<code>Osc[=<f>]</code>	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 TRANSITION-STATE OPTIMIZATIONS

Transition states can now be optimized using the option **TSOPT**. EMPIRE20™ uses a specialized update of the NS01A algorithm [15] to optimize transition states based on an initially calculated force-constant (Hessian) matrix. Although the algorithm is very robust, it does need a starting geometry quite close to the transition state in order to converge properly. It is often adequate to construct a geometry in which the relevant parameters (bond lengths, angles, torsion angles etc.) are set to approximate their expected values in the transition state. This is, for instance, the case with the following input for a Diels-Alder transition state:

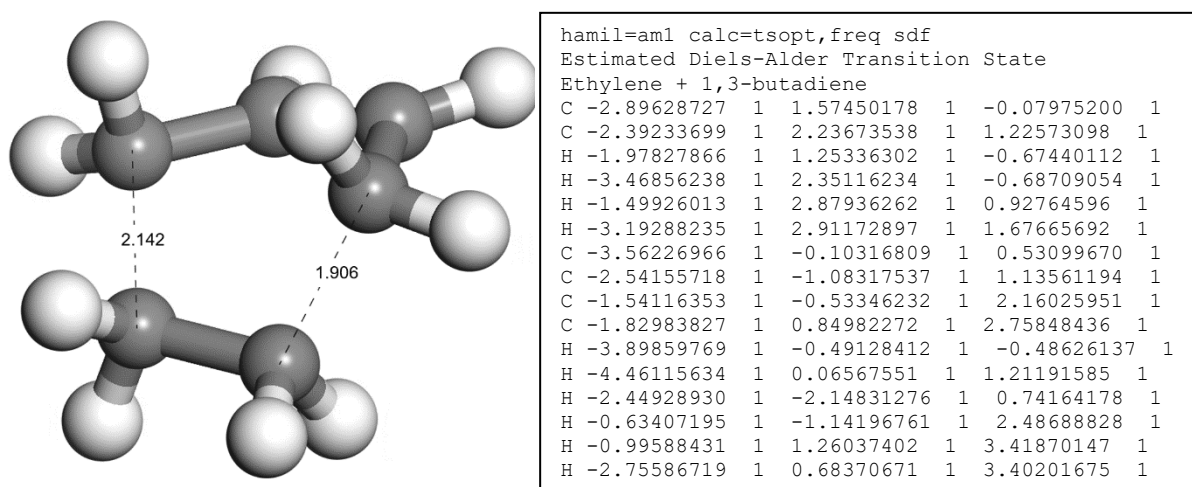


Figure 4 Estimated starting structure and .dat input for an AM1 optimization of the transition state for the Diels-Alder addition of ethylene to cis-1,3-butadiene.

The starting geometry was constructed simply by stretching the two bonds formed in the Diels-Alder reaction in the product cyclohexene in an interactive modeling program. Clearly, the structures of the ethylene and 1,3-butadiene moieties are far from optimal but the input shown in Figure 4 gives a successful transition-state optimization in 100 cycles to give a structure with a single imaginary frequency at $939i\text{ cm}^{-1}$. The optimized transition state is almost symmetrical:

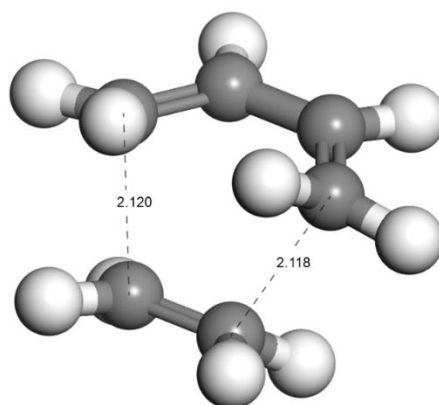


Figure 5 The optimized transition-state structure obtained from the above calculation.

5 SCRf IMPLICIT SOLVENT MODEL

The self-consistent reaction field (SCRf, or polarizable continuum model, PCM) implicit solvation model [11] is available for ground-state single-point calculations in EMPIRE20™. Parameters for the available solvents are given in the `SCRf_solvents.par` file, which is ASCII and therefore editable, so that the user can add solvents as required. Table 2 shows the solvents and data available:

Table 2 EMPIRE™ Solvent parameters from the default `SCRf-Solvents.par` file

Solvent	Input Names ^a	ϵ^b	$n^D c$	Volume ^d	Diameter ^e	TCE ^f	Ionization potential (au)
n-Hexane	N-HEXANE HEXANE	1.890	1.3751	131.600	5.920	1.361	0.374
n-Hexadecane	N-HEXADECANE HEXADECANE C16H34	2.050	1.4345	294.099	8.600	1.250	0.476
Carbon tetrachloride	CARBON_TETRACHLORIDE CCL4	2.238	1.4601	97.088	5.380	1.229	0.421
Benzene	BENZENE C6H6	2.284	1.5011	89.420	5.260	1.213	0.339
Toluene	TOLUENE C6H5CH3	2.379	1.4961	105.000	4.750	1.250	0.383
Cyclohexane	CYCLOHEXANE	3.100	1.4262	91.470	3.080	1.001	0.363
Diethyl ether	DIETHYL_ETHER ETHER ET2O	4.335	1.4115	104.720	4.740	1.654	0.355
Chloroform	CHLOROFORM CHCL3	4.806	1.4459	80.670	5.430	1.260	0.420
THF	Tetrahydrofuran TETRAHYDROFURAN thf	7.580	1.4072	61.711	5.400	1.001	0.345
Dichloromethane	DICHLOROMETHANE METHYLENE_CHLORIDE CH2CL2	9.080	1.4242	64.500	4.760	1.391	0.417
1-Octanol	1-OCTANOL OCTANOL	10.300	1.4295	160.800	7.000	8.280	0.368
Acetone	ACETONE 2-PROPANONE	20.700	1.3588	74.044	4.790	1.430	0.356
Ethanol	ETHANOL ETOH C2H5OH	24.300	1.3611	58.700	4.360	1.103	0.386
Methanol	METHANOL MEOH CH3OH	32.630	1.3288	40.700	3.710	0.998	0.398
Benzonitrile	BENZONITRILE C6H5CN PHCN	26.000	1.5279	103.100	5.450	1.200	0.358
Acetonitrile	ACETONITRILE CH3CN	36.640	1.3440	52.869	4.520	1.368	0.448
Dimethyl sulfoxide	DIMETHYL_SULFOXIDE DMSO	46.450	1.4790	71.327	4.910	0.928	0.331
Water	WATER H2O	78.540	1.3330	18.070	2.800	2.570	0.463
Pyridine	PYRIDINE C6H4N	12.300	1.5095	79.860	4.880	1.122	0.355
Nitrobenzene	NITROBENZENE C6H5NO2	34.820	1.5562	107.160	5.510	0.830	0.364
Test (Eps = 1000)	TEST	1000	1.6000	75.000	5.000	1.000	0.400

^a Solvent names recognized by EMPIRE20™. Up to three alternative names can be defined in the `SCRf_solvents.par` file and the existing ones modified if required. Names are case sensitive. ^b The static dielectric constant of the solvent. ^c The solvent refractive index for the sodium d-line. ^d The solvent molecular volume in Å³. ^e The solvent diameter in Å. ^f The thermal expansion coefficient of the solvent.



Currently, SCRF calculations with EMPIRE20™ are limited to ground-state single points. The SCRF treatment is important for calculations of proteins without explicit solvent. “Gas phase” EMPIRE™ calculations converge very slowly because of purely inductive charge-migration through the protein during the SCF iterations. [16] There are two possible solutions to this problem. The first, which has been used up to EMPIRE19™ was to include a layer of explicit solvent molecules around the protein. Using **SCRF=WATER** in EMPIRE20™ generally leads to fast convergence without needing to include specific solvent molecules.

6 CONFIGURATION INTERACTION

EMPIRE20™ includes an expanded configuration-interaction capability compared to EMPIRE19™. Configuration-interaction calculations for open-shell species up to triplet multiplicity are now available. The active-space option allows several options to define the molecular orbitals to be used in the CI calculation. The default option is to include all UNOs with occupation numbers between 1.98 and 0.02. These limits can be adjusted using the `AS=UNO,<f1>,<f2>` option, which sets the lower limit to <f1> and the upper to <f2>. Alternatively, the `AS=UNO(<n1>,<n2>)` option defines the active space to include molecular orbitals <n1> to <n2>, inclusive.

UNO-CI is currently only available for CIS and full CI for all multiplicities and PEI for singlets.

The reference wavefunction for the CI calculation can be chosen to have the multiplicity defined by the `Mult=<n>` option (default) or the multiplicity suggested by the occupation of the UHF natural orbitals (assuming single electrons to be spin-up).

7 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 6 .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 7 .dat input example for a three-dimensional periodic calculation

8 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 8 .dat input example for a molecular dynamics simulation

9 PROGRAM OUTPUT

The EMPIRE™ output for an adamantane single point (using the input shown in [Figure 1](#) and the command `empire.exe adamantane.xyz hamil=AM1`) is shown in [Figure 9](#).

```

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 P I R R E
EEE M M P III R R EEE

EMPIRE'20 MPI (Revision 2734)

(c) Matthias Hennemann, Johannes Margraf, David Whitley, Bernd Meyer, 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
3) Johannes T. Margraf, Matthias Hennemann, Timothy Clark,
"EMPIRE: A highly parallel semiempirical molecular orbital program:
3: Born-Oppenheimer molecular dynamics.", J Mol Model (2020) 46:23 (9 pages).
DOI: 10.1007/s00894-020-4293-z

<> Computing resources

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

<> Options

Input file = adamantane.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

```


Figure 10 shows the EMPIRE™ output for the same input file but using the PM3 Hamiltonian for a full geometry optimization (`empire.exe adamantane.xyz 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 P I R R E
EEE M M P III R R EEE

EMPIRE'20 MPI (Revision 2734)

(c) Matthias Hennemann, Johannes Margraf, David Whitley, Bernd Meyer, 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
3) Johannes T. Margraf, Matthias Hennemann, Timothy Clark,
   "EMPIRE: A highly parallel semiempirical molecular orbital program:
   3: Born-Oppenheimer molecular dynamics.", J Mol Model (2020) 46:23 (9 pages).
   DOI: 10.1007/s00894-020-4293-z

<> Computing resources

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

<> Options

Input file = adamantane.xyz

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
    
```


10.2 HDF file format

HDF5 is a binary file format supported by the HDF Group (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 3**. All HDF5 object names are in uppercase with no spaces, with multiple words broken by underscores where appropriate.

Table 3 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.

10.3 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 4 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 ⁻¹	double precision



<pre> 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 8 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 > <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] 0.0000000000 </pre>	<p style="color: cyan;">Further orbitals not shown here</p> <pre> 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 -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 \$\$\$\$ </pre>
--	--

Figure 12 The EMPIRE™ output .sdf file for a PM3 geometry optimization of adamantane

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

Table 5 EMPIRE™ specific tags

Tag	Contents	
><CALCULATION>	1 Single point calculation <n> Geometry optimized using EF	<n> is the number of optimization cycles (always 1 for a single point)
><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)	
><TOTAL COULSON CHARGE>	Sum of the Coulson charges (floating point)	
><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.	
><CHARGE ON HYDROGENS>	The Coulson charges of the hydrogen atoms (this block is redundant as the charges are also contained in the "><COULSON CHARGES>" block).	
><ORBITAL VECTORS> and ><LOCALIZED ORBITALS>	Block for the canonical or localized molecular-orbital eigenvectors	



Tag	Contents	
RHF calculation	<ol style="list-style-type: none">1. Number of the HOMO (Nhomo)2. Number of orbitals (Nao)3. Number of occupied orbitals (Nocc)4. Number of virtual orbitals (Nunocc)	Format: 4(i4,1x),
Nr. of atomic orbitals of each atom	The number of basis functions (atomic orbitals) per atom; s-only = 1; s,p = 4, s,p,d = 9.	Format: 20i4
Orbital information from occ. to unocc.	For each molecular orbital:	
	The Eigenvalue (eV)	Format: F9.3
	The MO coefficients	Format: F6.3,7F7.3
><CALCULATION STATUS>	SUCCESS	

11 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.

12 SUPPORT

12.1 Contact

Questions regarding EMPIRE™ should be sent directly to:

support@ceposinsilico.com

12.2 Cepos InSilico GmbH

Waldstrasse 15
90587 Obermichelbach
Germany

support@ceposinsilico.com
Tel. +49 (0)9131 9704910
Fax. +49 (0)9131 9704911

www.ceposinsilico.com/contact



13 LIST OF TABLES

Table 1	EMPIRE™ program options	10
Table 2	EMPIRE™ Solvent parameters from the default SCRF-Solvents.par file	15
Table 3	HDF5 wavefunction file structure.	27
Table 4	Contents of binary wavefunction file	28
Table 5	EMPIRE™ specific tags	30



14 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	Estimated starting structure and .dat input for an AM1 optimization of the transition state for the Diels-Alder addition of ethylene to cis-1,3-butadiene.	14
Figure 5	The optimized transition-state structure obtained from the above calculation.	14
Figure 6	.xyz input example for a three-dimensional periodic calculation	18
Figure 7	.dat input example for a three-dimensional periodic calculation.....	18
Figure 8	.dat input example for a molecular dynamics simulation	19
Figure 9	The EMPIRE™ output for an AM1 single-point calculation on adamantane.....	22
Figure 10	The EMPIRE™ output for a PM3 geometry optimization of adamantane.	25
Figure 11	The EMPIRE™ archive file for a PM3 geometry optimization of adamantane.....	26
Figure 12	The EMPIRE™ output .sdf file for a PM3 geometry optimization of adamantane.....	30



15 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. <http://dx.doi.org/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. <http://dx.doi.org/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. <http://www.sciencedirect.com/science/article/pii/S0040402001818688>; 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+. <http://dx.doi.org/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. <http://dx.doi.org/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. <http://www.ncbi.nlm.nih.gov/pubmed/14997367>
- [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. <http://dx.doi.org/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. <http://dx.doi.org/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. <http://dx.doi.org/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. <http://dx.doi.org/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. <http://dx.doi.org/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. <http://dx.doi.org/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. <http://dx.doi.org/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. <http://dx.doi.org/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. <https://www.ncbi.nlm.nih.gov/pubmed/28700231>
- [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. <http://dx.doi.org/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. <http://dx.doi.org/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.



- <http://dx.doi.org/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. <http://dx.doi.org/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. <https://www.ncbi.nlm.nih.gov/pubmed/30443727>; 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. <https://doi.org/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. <https://doi.org/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. <https://www.ncbi.nlm.nih.gov/pubmed/17828561>, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2039871/>
- [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. <http://dx.doi.org/10.1002/jcc.540030214>
- [11] G. Rauhut, T. Clark, T. Steinke. A numerical self-consistent reaction field (SCRf) model for ground and excited states in NDDO-based methods, *Journal of the American Chemical Society* **1993**, *115*, 9174-9181. doi:10.1021/ja00073a036. <https://doi.org/10.1021/ja00073a036>
- [12] 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. <http://www.ncbi.nlm.nih.gov/pubmed/25983105>
- [13] 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. <https://doi.org/10.1002/ijch.199300050>
- [14] 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. <https://doi.org/10.1021/jp204939x>
- [15] M. J. D. Powell. A Fortran subroutine for solving systems of nonlinear algebraic equations, *Atomic Energy Research Establishment, Harwell, England (UK)* **1968**, AERE-R-5947. <https://www.osti.gov/biblio/4772677-fortran-subroutine-solving-systems-nonlinear-algebraic-equations>
- [16] C. R. Wick, M. Hennemann, J. J. Stewart, T. Clark. Self-consistent field convergence for proteins: a comparison of full and localized-molecular-orbital schemes, *Journal of Molecular Modeling* **2014**, *20*, 2159. doi:10.1007/s00894-014-2159-y. <http://www.ncbi.nlm.nih.gov/pubmed/24573500>