INSTITUT FÜR PHYSIKALISCHE UND THEORETISCHE CHEMIE,
UNIVERSITÄT BONN, GERMANY

MSINDO

Thomas Bredow, Gerald Geudtner and Karl Jug

Version 3.8

February 4, 2020

Preface

MSINDO [1–4] is a semiempirical molecular orbital program for the calculation of molecular and condensed matter properties of systems with first-, second-, third- and fourth-row elements. It is a modification of the original SINDO1 method [5–8]. The current version of MSINDO includes parameters for the following elements: H, Li–F, Na–Cl, K–Br. Calculation of elements Rb–I is now possible, but the parameterization is not yet complete.
The program is parameterized with emphasis on ground state properties, mainly heats of formation, structure data, ionization energies and dipole moments. Ground state properties are calculated on the SCF level, excited states on the CIS level. Full CIS is now available via the Davidson-Liu algorithm. NDDO-type extensions through additional sp-dipol integrals are available for first- and second-row elements.

The program provides the following features:

- Molecular geometries and binding energies
- Bulk and surface simulations with the Cyclic Cluster Model (CCM)
- Vertical excitation energies via the Davidson CIS method with oscillator strengths
- Solvent effects for ground and excited states via the COSMO including gradients
- Dispersion interactions via Grimme's D3BJ correction scheme
- Nudged Elastic Band method for minimum energy path calculations
- Vibration frequencies via numerical second derivatives
- Thermodynamic properties (enthalpy, entropy, heat capacities, zero point energies)
- Born-Oppenheimer molecular dynamics
- Metadynamics techniques, acceleration of reaction processes and free energy barriers
- Møller-Plesset closed-shell energies and gradients
- Molecular electrostatic potentials (MESP)
- Ionization potentials (vertical and ΔSCF) and dipole moments
- Atomic charges (Mulliken and Löwdin) and bond orders, atomic and bond valences

Contents

1 Introduction

The programming language is Fortran77 with Fortran90 extensions (for sure not object oriented). Memory is allocated dynamically for large arrays and statically for small arrays and variables. MSINDO 3.8 can be compiled with most available Fortran90 compilers (e.g. ifort, gfortran). Parallelization is most easily achieved by OPENMP compiler switches on multi-core hardware (compilation with optimization option). The computational performance is improved considerably if optimized BLAS and LAPACK libraries (e.g. within the mkl) are linked together with the MSINDO object files. However, some standard version routines of the BLAS and LAPACK libraries are included in the MSINDO source code (**rblas.F** and **lapack.F**).

Please note the following:

- MSINDO 3.8 has been tested mainly on Linux platforms.
- Some compiler options are available in **MSINDO/3.8/platforms**.
- For installation details see the file **README.msindo**.
- Postprocessing of MD runs: **README.postmd1, README.postmd2**.
- Metadynamics and postprocessing: **README.Vreco**.
- The source includes an automatic routine (MSINDO/msinpar) for the optimization of empirical parameters, if needed.
  For first instructions see **README.msinpar**.

2 Detailed input description

2.1 General input file structure

The input is divided into five sections:

1. Title
2. Keywords
3. Geometry definition
4. Assignment for variables
5. Additional input

Please note the following:

- Sections 1-4 are **unformatted**.
- Each of the sections 2-4 is closed by a line including the string **:END** (case insensitive).
- Optional Section 5 is read in **formatted**.
- Only columns **1-120** of the input file are considered.
- There is **no** distinction between uppercase and lowercase letters.
- Keywords and variables **must** be separated by blanks.

2.2 Title

Any text describing and commenting the calculation. Must occur in the first input line or immediately follow the previous input if several inputs are given in one file.

NOTE: It is possible to use Gaussian inputs for simple MSINDO runs (SCF, geometry optimization). Insert the string **:GAUSS** (lower- or uppercase) in the first line of the Gaussian input. Gaussian job directives will be ignored.

2.3 Keywords

In this section the calculation is controlled by keywords and control parameters. A keyword activates the corresponding program options. Control parameters are assigned values with the equal sign. Integer and real numbers have to be distinguished. Each keyword and control parameter has a default value preset in the program. The ordering of keywords and parameters in the input should have **no** effect on the calculation.

The Keywords are grouped as follows:

- Wave function
- SCF procedure
- Geometry optimization
- Vibration analysis
- Cyclic Cluster Model calculations
- Nudged Elastic Band calculations
- M{o}ller-Plesset perturbation theory
- Molecular dynamics
- Input-Output files for molecular dynamics
- Post-MD tools
- Reaction field calculations
- Embedding procedure
- Symmetry analysis
- Printout options
- Restarting options
- Input check

Please see the table of contents for page numbers of the sections.

2.3.1 Wave function

Keyword	Default	Description	Lit.
RHF	set	Restricted Hartree-Fock calculation for closed-shell systems.	[9]
UHF	not set	Unrestricted Hartree-Fock calculation for open-shell systems (→MULTIP). Possible also for open-shell singlet states, may give results differing from RHF calculations in particular if (→IDEN=1) is used (diradical, anti-ferromagnetic systems). The resulting wave functions are usually spin-contaminated, but the spin densities are qualitatively correct.	[9]
ROHF	not set	Restricted Open-shell Hartree-Fock calculation for open-shell systems (→MULTIP). Original method introduced by Roothaan. Results are identical to RHF for singlet states, but computationally more costly. Gives pure spin states, but possibly incorrect spin densities. The Roothaan factors (→FFAKT, AFAKT, BFAKT) need not be set for non-degenerate high-spin states with filled and half-filled shells.	[10]
NDDO	not set	Additional NDDO-type two-center one-electron and two-electron integrals are calculated. The NDDO parameters for elements H to Cl have been re-optimized and are different from the original INDO parameter set. The changes compared to INDO results are in general moderate, except for subtle effects involving multipole interactions, e.g. surface relaxation of MgO(100). For the other elements NDDO has not been tested yet.	[11]
MULTIP	= 1	Multiplicity of the system. Has to be set for (→ UHF and ROHF) calculations.	
DFTD3	not set	Grimme's D3 correction from 2010; available for RHF and UHF	[12]
VDW	not set	Grimme's D2 correction from 2006; available for RHF and UHF	[13]

1. Wave function (cont.)

Keyword	Default	Description	Lit.
CI	not set	Configuration interaction calculation using singly or doubly excited configurations generated from the SCF ground state determinant. Specification of configurations/determinants and selection of the CI reference state follows in **Section 5**. CI can be used for (→RHF, UHF and ROHF) wave functions. (deprecated feature)	
CIS	not set	Approximation to full singles CI (RHF, UHF, ROHF); the active space given in **section 5** defines the block size for iterative optimization of the selected reference state within all occupied and unoccupied orbitals. (deprecated feature)	
CISD	not set	Approximation to full singles and doubles CI (RHF).	
CIST	not set	Approximation to full singles and selected triples CI (RHF, experimental).	
BODIFF	= 0	Only relevant for (→CI,CIS,DAVIDSONCIS). The N largest changes to atomic charges and interatomic bond orders of the selected CI reference state with respect to the ground state are listed.	
CHARGE	= 0	Net charge of the system. For anions the orbital exponents are changed for a last SCF cycle leading to a new wave function and energy. The procedure is self-consistent. Anion exponents are parameterized only for elements H-Cl.	[18]
FFAKT	= 0.5	Ratio of occupied and total number of open shell spin orbitals (→ROHF). Has to be set only for degenerate states (experts only).	[10]
AFAKT	= 1.0	Roothaan factor used in the (→ROHF) Fock operator. Has to be set only for degenerate states (experts only).	[10]
BFAKT	= 2.0	Roothaan factor used in the (→ROHF) Fock operator. Has to be set only for degenerate states (experts only).	[10]

1. Wave function (cont.)

Keyword	Default	Description	Lit.
OCCB	= 0	Number of doubly occupied orbitals in the (→ROHF) wave function. Has to be set only together with (→AFAKT,BFAKT) (experts only).	
OCCE	= 0	Number of partially occupied orbitals in the (→ROHF) wave function. Has to be set only together with (→AFAKT,BFAKT) (experts only).	
LOCA	not set	Localization procedure by Pipek and Mezey. (→RHF, UHF, and ROHF) wave functions are localized on atoms and bonds with a maximization of atomic charges. Useful for (→CI) calculations in order to reduce the number of determinants of the CI expansion.	[19]
TCSCF	not set	Two configuration SCF procedure for (→RHF). The ground state determinant and an excited determinant generated by a HOMO-LUMO double excitation form the SCF wave function. The MO and CI coefficients are both optimized in the TC-SCF procedure.	[20]
DOFF	= 0	Second row elements (Al-Cl) have an spd standard valence basis set. Setting DOFF=1 eliminates the d functions from the basis set for these atoms. This is not recommended in general because the parameterization of the elements Na - Cl has been performed with d orbitals.	[21]
PON	= 0	MSINDO treats hydrogen bonds by adding 2p functions to the valence basis set of hydrogen atoms (PON=1).	[22]
DONC,DONN,DONO	not set	Additional *d* orbitals on C, N, O (experimental).	
DIRSCF	not set	Direct SCF (RHF and UHF only): saves memory but increases CPU time. Not fully tested.	

1. Wave function (cont.)

Keyword	Default	Description	Lit.
S12APPROX	not set	If a Mulliken population analysis is requested (→MUL), the transformation to the nonorthogonal basis can be performed analytically or numerically with a parameterized approach (in general not recommended). The latter approach is activated by this keyword.	
PARAMET	not set	Has to be set in inputs used for paramerization with MSINPAR; has no effect for standard calculations	

2.3.2 Excited States

Keyword	Default	Description	Lit.
DAVIDSONCIS	not set	Full singles CI; iterative Davidson algorithm is used. **Highly recommended for calculation of optical spectra of open- and closed-shell systems**. Works with the CCM. No additional input is needed in **section 5**	[16]
REFSTATE	= (2,1)	Reference state in Davidson CIS properties calculation; The first value in parentheses denotes the sequence number of the reference state for which properties are calculated, the second is the multiplicity. NOTE: the S0 ground state is represented by (1,1), the first excited triplet state by (1,3).	
SROI	= 4	Number of singlet roots of interest;	
TROI	= 4	Number of triplet roots of interest;	
DVDCYC	= 20	Max. number of iteration cycles in Davidson procedure.	
DVDCONV	1.0E-04	Convergence criterion on the residuals of the Davidson procedure.	
CISSPACE	= 0	Size of the CIS subspace in the pseudo-Davidson procedure. The pseudo-Davidson procedure is only called, if this variable is set. Recommended for large system and machines with low memory. The (→ DVDCYC) variable should be increased, since the number of steps is increased in this procedure. Should be three times larger then (→ SROI, TROI). But the used memory is constant.	
DIAGAPPROX	not set	Use of approximate diagonal elements of the sCIS and CIS Matrix, increases number of Davidson iterations, reduces the calculation time	
CIS_RESTART	not set	Reading the CIS vectors of a previous run (with → CIS_WRITE) from fort.150	
CIS_WRITE	not set	Writing the CIS vectors after a Davidson cycle on fort.150	

2. Excited states (cont.)

Keyword	Default	Description	Lit.
LARGEGUESS	not set	Large starting space of vectors in the Davidson procedure only recommended for large systems, works best in connection with (→ CISSPACE)	
SCALEDCIS	not set	Activates global scaling parameters for Coulomb and exchange integrals in Davidson CIS (similar to the philosophy of SCS-CIS of Grimme)	[14,15]
DIAGAPPROX	not set	Approximative calculation of the diagonal elements of the CIS matrix. Significant savings in CPU time and memory requirements, so far no differences in the final excitation energies detected.	
CISCORR	not set	Includes the empirical d_{ia}^{corr} correction, recommended for the calculation of total symmetric excitations	[15]
CISGRAD	(2)	State to optimize in the geometry optimization NOTE: the the S0 ground state is represented by (1), multiplicity of the state is given by the (→ REFSTATE) keyword. Works with RHF, UHF grounstate wavefunctions and the sCIS approach	[17]
CGS	not set	Using the CGS algorithm for the solution of the CPHF equations. This is much faster than the standard TFQMR algorithm but numerically less stable	
EXCVIB	not set	Calculation of vibronic states of the given electronic excited state (→REFSTATE)	
EADIAB	=0.00	Adiabatic excitation energy in eV, used for vibronic calculations of excited states	
CHARGESTATE	not set	Activates calculation of charge differences of all (→SROI) excited states with respect to the ground state.	

2.3.3 SCF procedure

Keyword	Default	Description	Lit.
NGIV	= 4	Selection of matrix diagonalization routine; diagonalization is the most computer time consuming step in MSINDO, in particular for large systems. Therefore the choice of the optimal routine is crucial for the computational performance. **1:** Unused. This value leads to program termination. **2:** Jacobi procedure; gives the most accurate eigenvectors and eigenvalues, but is by far the slowest procedure. Is included in the MSINDO source code. **3:** F02ABF from the NAG library; only useful if the NAG library is available. Is not included in the MSINDO source code, the object files have to be linked together with NAG (see **chapter 6**). NOTE: The subroutine **diamat.f** has to be modified in order to activate F02ABF. **4:** RS from the EISPACK library; comparable to F02ABF on scalar processors. Standard routine; is included in the MSINDO source code.	

2. SCF procedure (cont.)

Keyword	Default	Description	Lit.
		5: DSYEVX from the LAPACK library (recommended for large systems); only eigenvectors of the occupied MOs are calculated. DSYEVX is therefore 2-3 times faster than F02ABF and RS on scalar processors, but not on vector processors. Considerable speedup with OPENMP compiler options (export OMP_NUM_THREADS=no. of processor cores available). A standard version is included in the MSINDO source code (see **lapack.F**).	
			[26]
		6: Pseudodiagonalization procedure by Stewart. Implemented for (→RHF, UHF and ROHF). Recommended for stable SCF runs, is faster than F02ABF and RS by factors up to 3 if an optimized BLAS library is available (see **chapter 6**). It is an approximation, however, and may lead to numerical inaccuracies. Do not use it together with level shifting (→NAV=5).	

2. SCF procedure (cont.)

Keyword	Default	Description	Lit.
		7: F02FCF from the NAG library; only eigenvectors of the occupied MOs are calculated. Is therefore 2-3 times faster than F02ABF – except on vector processors. Is not included in the MSINDO source code, the object files have to be linked together with NAG (see **chapter 6**). NOTE: The subroutine **diamat.f** has to be modified in order to activate F02FCF.	
		8: DSYEVD from the LAPACK library; in general similar to DSYEVX, but performs better on some platforms. Needs more memory. Considerable speedup with OPENMP compiler options (export OMP_NUM_THREADS=no. of processor cores available). A standard version is included in the MSINDO source code (see **lapack.F**).	
OPTDIAMEM	not set	If set, an the optimal length of temporary vectors used in the diagonalization procedures DSYEVX and DSYEVD are estimated; may lead to a speed-up of 2.	
NAV	= 4	SCF acceleration procedures; the following procedures are available:	

2. SCF procedure (cont.)

Keyword	Default	Description	Lit.
			[23, 24]
		1: Fermi smearing; standard temperature 5000 K. **2, 12:** Strongly damped SCF; will be very slow for normal systems, suitable for some strongly oscillating cases. **3, 13:** Hartree damping; good choice for many ionic systems. **4, 14:** Variable weighting procedure; standard procedure, useful for most systems. **5, 15:** Pongor's level-shifting procedure (→SF1,SF2); last resort for problem cases (oscillating SCF runs). **6, 16:** Zerner's damping procedure; sometimes the best choice for organic molecules. **7:** EDIIS algorithm [30] (only for →CCM, highly recommended for difficult cases)	
		For (→RHF and UHF) calculations on systems with first and second row elements the damping procedures have proved to be reliable. Systems including transition metals, especially open shell systems, very often need the level shifting procedure (NAV=5,(→SF1,SF2)).	
		The procedures 1, 3, 4, and 6 may be combined with Pulay's DIIS procedure by increasing NAV by 10. This is recommended for slow but stable converging SCF runs, not for oscillating SCF runs. DIIS reduces the number of SCF cycles, but probably not the CPU time due to intensive I/O.	[25]

2. SCF procedure (cont.)

Keyword	Default	Description	Lit.
SF1	= 0.8	(Unit Hartree); in the **first** SCF cycle virtual orbitals are shifted by SF1. In (→ROHF) calculations partially occupied orbitals are shifted with respect to doubly occupied orbitals and the virtuals are shifted with respect to the partially occupied orbitals by SF1.	
SF2	= 0.0	(Unit Hartree); level shifter used in the **last** SCF cycle. During the SCF run, the actual shifter is an interpolation between SF1 and SF2 depending on the convergence. SF2 should not be set larger than zero except for badly converging systems. Negative values are not accepted.	
DELEN	1.0E-08	(Unit Hartree) Energy threshold used for the SCF procedure. For systems showing slow convergence larger values may be given. However, results can become quite inaccurate for DELEN > 1.0E-05. In these cases (→NAV,SF1,SF2, or IDEN) should be varied in order to improve SCF convergence.	
IDEN	= 0	Start density for the first SCF cycle. **0:** Core matrix density; an Extended Hückel calculation is performed. In geometry optimization runs the start density for the final run is generated by a new Extended Hückel calculation.	

2. SCF procedure (cont.)

Keyword	Default	Description	Lit.
		1: Atomic start density; only diagonal elements of the density matrix are generated according to atomic orbital populations. Suitable for (→UHF) calculations, in particular UHF singlet states. In geometry optimization runs the start density for the final run is again an atomic density.	
		2: Core matrix density analogous to IDEN=0; In geometry optimization runs the start density for the final run is the optimized density (→IP). This procedure saves computer time, but the final start density may not be reproducible in certain cases. (not implemented for →CCM)	
		3: Atomic start density analogous to IDEN=1; In geometry optimization runs the start density for the final run is the optimized density (→IP) (see comments to IDEN=2).	
		4: Start density is generated analogous to IDEN=0; occupied and unoccupied Hückel MOs (printout: →PRINTOPTS=HUVEC) can be exchanged. The specification of the MOs to be interchanged follows in **Section 5** (→**MO exchanges**). (not implemented for →CCM)	

2. SCF procedure (cont.)

Keyword	Default	Description	Lit.
		6: Start density is generated analogous to IDEN=2; occupied and unoccupied Hückel MOs (printout: →PRINTOPTS=HUVEC) can be exchanged. The specification of the MOs to be interchanged follows in **Section 5** (→**MO exchanges**). (not implemented for →CCM)	
		8: Start density is generated analogous to IDEN=0. The final SCF density matrix (→ROHF: total and open shell density matrices, →UHF: α and β density matrices) is written unformatted to the file **DENSITY**. (not implemented for →CCM)	
		10: Start density (→ROHF: total and open-shell density matrices, →UHF: α and β density matrices) is read from the file **DENSITY**. (not implemented for →CCM)	
		11: Modified atomic start density (similar to IDEN=1); suitable for (→)ROHF calculations (not implemented for →CCM)	

2. SCF procedure (cont.)

Keyword	Default	Description	Lit.
AOCCUPY(N)	9×0.00	AOCCUPY(N)=$(s, p_x, p_y, p_z, d_{z^2}, d_{xz}, d_{yz}, d_{x^2-y^2}, d_{xy})$ sets diagonal elements of the start total density matrix (→RHF,ROHF) and α density matrix (→UHF). Only 2 decimal places are taken into account for each entry. For the meaning of N see (→OCCATOM). Can be combined with all values of (→IDEN).	
BOCCUPY(N)	9×0.00	BOCCUPY(N)=$(s, p_x, p_y, p_z, d_{z^2}, d_{xz}, d_{yz}, d_{x^2-y^2}, d_{xy})$ sets diagonal elements of the start open-shell density matrix (→ROHF) and β density matrix (→UHF). Only 2 decimal places are taken into account for each entry. For the meaning of N see (→OCCATOM). Can be combined with all values of (→IDEN).	
OCCATOM	not set	If OCCATOM is specified, then in (→AOCCUPY(N),BOCCUPY(N)) N is the number of an atom according to the ordering in the input. In open-shell systems, the α density of all atoms not referred to by (→AOCCUPY(N),BOCCUPY(N)) is set to zero. In closed-shell systems, these atoms are unchanged. If OCCATOM is not specified, (→AOCCUPY(N),BOCCUPY(N)) refer to all atoms with atomic number N. All atoms of unreferenced elements are not changed in this case.	
HUCSHIFT(AN)	9×0.00	(unit Hartree). HUCSHIFT(AN)=$(s, p_x, p_y, p_z, d_{z^2}, d_{xz}, d_{yz}, d_{x^2-y^2}, d_{xy})$ shifts the corresponding diagonal elements of all atoms with atomic number AN in the Hückel matrix used in the initial Extended Hückel calculation (→IDEN=0,2). Useful for transition metal compounds in order to take into account ligand-field effects.	

2. SCF procedure (cont.)

Keyword	Default	Description	Lit.
MAXCYC	= 300	Maximum number of SCF cycles for single point calculations.	
IP	= 100	Maximum number of SCF cycles for the re-optimization of the density matrix in geometry optimization runs (→OPT, LSTE). If numerical energy gradients (→ANALY) or second derivatives (→IUPD, LMK) are calculated, a new SCF run is started after each differential geometry change (→RDELTA, WDELTA, DDELTA). IP should be large enough to ensure energy convergence within the preset accuracy (→DELEN). In order to save computer time, IP may be set to small values down to zero; however, numerical first and second derivatives may become quite inaccurate.	
MP2	not set	Møller-Plesset correlation energy for RHF; analytic gradients are available	
DIRECTMP2	not set	Direct MP2; requires less memory	
MP2D	not set	MP2 with only aa-¿ii double excitations; a crude approximation, not recommended	

2.3.4 Geometry optimization

Keyword	Default	Description	Lit.
INTOPT	not set	Activates the geometry optimization for internal coordinates (alternative→CARTOPT). (→LSTE) is set to 100 if it is not explicitly set.	
CARTOPT	not set	Activates the geometry optimization in Cartesian coordinates. This can be done for geometry specifications in both internal or Cartesian (→CARTES) coordinates. If internal coordinates are given, a new Z matrix with optimized coordinates is written to the file **fort.9** if (→NEWZMAT) is specified. See (→ANALY,CARTSLCT,CARTDE2,CARTSDE2,CARTREST).	
LSTE	= 0	Maximum number of optimization steps in one run (→OPT,CARTOPT).	
ABSDR	0.0030	(Unit Bohr); threshold for max. of geometry changes (→OPT,CARTOPT)	
RMSDR	0.0015	(Unit Bohr); threshold for RMS of geometry changes (→OPT,CARTOPT)	
ABSFOR	0.00045	(Unit Hartree/Bohr); threshold for max. gradient (→OPT,CARTOPT)	
RMSFOR: 0.00030	(Unit Hartree/Bohr); threshold for RMS gradient (→OPT,CARTOPT)		
ANALY	set	Analytical gradients for (→RHF,UHF,ROHF) calculations. Not implemented for (→CI,TCSCF,PCM)	

3. Geometry optimization (cont.)

Keyword	Default	Description	Lit.
RDELTA	= 0.001	(Unit Å); increment for changing bond lengths in the calculation of numerical first (→ANALY) and second (→IUPD) energy derivatives. For slow SCF convergence (→DELEN>1.0E-07) or (→CI) wave functions an increase of RDELTA by a factor of 5-10 is recommended.	
WDELTA	= 0.1	(Unit degree); increment for changing bond angles in the calculation of numerical first (→ANALY) and second (→IUPD) energy derivatives. For slow SCF convergence (→DELEN>1.0E-07) or (→CI) wave functions an increase of WDELTA by a factor of 2-5 is recommended.	
DDELTA	= 0.1	(Unit degree); increment for changing dihedral angles in the calculation of numerical first (→ANALY) and second (→IUPD) energy derivatives. For slow SCF convergence (→DELEN>1.0E-07) or (→CI) wave functions an increase of DDELTA by a factor of 5-10 is recommended.	
LMK	= 0	Total number of coupled coordinates for the calculation of nondiagonal elements of the force constant matrix in internal coordinate geometry optimizations. Depending on the sequence of coordinates in **section IV** a full Hessian matrix is calculated for the first LMK coordinates, for all following coordinates only diagonal elements are calculated. For the location of transition structures (→TRASTA) LMK should be equal to the total number of relevant coordinates. Those have to be given first in **section IV**. For the location of minima the coupling of coordinates is important only if there are linear dependencies between internal coordinates. NOTE: The computational effort can be considerably reduced by using analytical derivatives (→ANALY).	[27]

3. Geometry optimization (cont.)

Keyword	Default	Description	Lit.
IUPD	= 0	Code for updated Hessian procedures (only for minima); for IUPD $\neq 0$ the calculation of second derivatives is avoided. IUPD > 0 Update of the Hessian IUPD < 0 Update of the inverse Hessian **±1, ±11:** BFGS update; this option is well tested and recommended especially for large systems. There may be problems for systems with bad SCF behavior. In these cases (→RDELTA,WDELTA,DDELTA) should be increased or (→IP,IDEN,NAV) should be varied. **±2, ±12:** DFP update. **±3, ±13:** Murtagh-Sargent update. **−4, −14:** Greenstadt update. **±10:** Steepest descent optimization (no update). IUPD options modified by ±10 take the unit matrix as starting Hessian. Otherwise the first force constant matrix is calculated numerically.	
RHO	= 0.00	(Unit Hartree); energy threshold for line search. The line search is activated by setting RHO > 0.00. Recommended value is then 0.01.	
GDIIS	not set	Pulay's convergence accelerator; useful to reduce the number of optimization steps if the potential surface is smooth. Recommended for large numbers of coordinates to optimize. May be combined with all values for (→IUPD).	[28]

3. Geometry optimization (cont.)

Keyword	Default	Description	Lit.
CONGDIIS	not set	Modified GDIIS algorithm; especially useful for large molecules. Only available for Cartesian geometry optimization (→CARTOPT). To vary the number of stored error vectors use the keyword VCONGDIIS=n; n≥2	[29]
MORSE	not set	Correction for geometry changes in each optimization cycle; the potential curve is modified by a morse potential. Internal coordinate optimization only.	
TRASTA	not set	Location of transition structures; the calculation of a full Hessian in the basis of all relevant coordinates (→LMK,CARTDE2) is necessary. The implemented algorithm needs good starting geometries.	
FOLLOW	= 1	Search algorithm for transition search by Baker with eigenvector following. Only active in combination with (→TRASTA).	
CARTES	not set	The geometry definition in **Section 3** is given in Cartesian coordinates instead of internal coordinates (which is the default).	
CARTSLCT	not set	Only the Cartesian coordinates of a selected number of atoms will be optimized (→CARTOPT). The specification of these atoms follows in **Section 5**.	
CARTDE2	not set	Activates the explicit calculation of second derivatives during the Cartesian geometry optimization (→CARTOPT). This is very time consuming but often necessary for transition state search (→TRASTA). Alternative: (→CARTSDE2).	
CARTSDE2	not set	May be specified instead of (→CARTDE2). Only the second derivatives of a selected number of atoms will be explicitly calculated. The specification of these atoms follows in **Section 5**.	
CARTREST	not set	Second derivatives will be initially read in from the input file (→CARTDE2,CARTSDE2).	

3. Geometry optimization (cont.)

Keyword	Default	Description	Lit.
DE2OUT	not set	At the end of the Cartesian geometry optimization the second derivatives will be written to the output file (→CARTDE2,CARTSDE2,CARTREST).	
NEWZMAT	not set	Prints out a new Z matrix after Cartesian geometry optimization (→CARTOPT) if the input is given in internal coordinates.	

2.3.5 Cyclic Cluster Model (CCM)

Simulations of perfect and defective crystalline solids and surfaces can be performed with MSINDO using the CCM. The theoretical background and the connection to conventional supercell models is described in [47–49]. In **Section 3** a cluster must be specified that is a translation unit (supercell) of the corresponding system. In addition, dummy atoms **Section 3** must be specified for the definition of translation vectors similar to (→EMBED), see (→VECTA,VECTB,VECTC).

For the CCM, the translation vectors **must** point outside the cluster (different from (→EMBED)!). It is sufficient to specify one of the keywords (→CCM3D,CCM2D,CCM1D) for the translations, since only one symmetric translation shell is needed for the CCM. No occupation numbers (→BESZ) have to be given as for (→EMBED). The cluster is replaced by a cyclic arrangement and the environment of each cluster atom is replaced by a Wigner-Seitz cell as described in the literature [46]. Therefore all translationally equivalent atoms should have the same charge etc. If this is not the case, check your input! For ionic systems it is highly recommended to take into account long-range electrostatic interaction. In MSINDO this is realized by a classical Ewald summation (→EWALD).

Keyword	Default	Description	Lit.
CCM3D	not set	Selection of a bulk calculation using the CCM. Three translation vectors (→VECTA, VECTB, VECTC) must be specified. At present, (→RHF,UHF,ROHF, and DAVIDSONCIS) calculations are possible. Optimization of Cartesian coordinates (→CARTOPT) and lattice parameters (→FULLOPT,LATTICEOPT) is possible	[46]
CCM2D	not set	Selection of a surface calculation using the CCM. Two translation vectors (→VECTA, VECTB) must be specified. Optimization of Cartesian coordinates (→CARTOPT) is possible. Optimization of lattice parameters (→FULLOPT,LATTICEOPT) is not recommended	

10. Cyclic Cluster Model (cont.)

Keyword	Default	Description	Lit.
CCM1D	not set	Selection of a polymer calculation using the CCM. A single translation vector (→VECTA) must be specified. Optimization of Cartesian coordinates (→CARTOPT) is possible. Optimization of lattice parameters (→FULLOPT,LATTICEOPT) is not recommended	
VECTA(A_1, A_2)	(0,0)	Basis vectors for translations of real atoms to generate pseudo atoms.	
VECTB(B_1, B_2)	(0,0)	A_1 and A_2, B_1 and B_2, C_1 and C_2 are numbers of atoms in the sequence of the geometry definition (see **Section 3**),	
VECTC(C_1, C_2)	(0,0)	which are start and end point of the basis vectors. The two atoms in a pair have to be on translationally equivalent positions. It is recommended to use dummy atoms (→**Section 3**) to define the basis vectors. For two-dimensional embedding only two basis vectors are needed (VECTA,VECTB), for 1D embedding VECTA has to be given. No spaces are allowed between keyword, parentheses, commas, and numbers.	
WEIGHTFCT	11	weighting scheme for atoms close to the boundaries of the atomic Wigner-Seitz cells in CCM based on either a projection method (10) or depending on distances **1, 11:** step function **2, 12:** sinus function (**only for orthorhombic cells**) **3, 13:** linear weighting function (recommended)	

10. Cyclic Cluster Model (cont.)

Keyword	Default	Description	Lit.
FULLOPT	not set	Activates the the full cartesian geometry optimization (→CARTOPT) and the optimization of the lattice parameters (→LATTICEOPT). The lengths of the lattice vectors are optimized according to the restrictions of some Bravais lattices (→LATTICEOPT, CUB, TET, ORT). The usual key words for the cartesian geometry optimization (→ANALY, CARTSLCT, LSTE, CARTDE2, CARTSDE2) can be used together with the FULLOPT key word. However, the key word (→CARTOPT) must not be specified together with FULLOPT	
LATTICEOPT	not set	Activates the optimization of the lattice vectors. The optimization is performed within the symmetry of the crystal system (→CUB, TET, HEX, ORT).	
CCMANALY	not set	Analytical gradients for CCM-optimization of lattice parameters	
RRADIUS	calcd.	Cut-off radius in reciprocal space for the Ewald summation applied in (→CCM) calculations. Reasonable values are 4-5. Smaller values will save CPU time, but may lead to inaccurate results.	
DRADIUS	calcd.	Cut-off radius in direct space for the Ewald summation (→EWALD) applied in (→CCM) calculations. The default is that RRADIUS and DRADIUS are internally optimized for each system by MSINDO. The minimum values of RRADIUS and DRADIUS in such an optimization can be specified with (→MRADIUS). Reasonable values are 4-5. Smaller values will save CPU time, but may lead to inaccurate results.	

10. Cyclic Cluster Model (cont.)

Keyword	Default	Description	Lit.
EZINP	not set	Special input for unit cells useful for (→CCM) calculations of bulk systems. Only **Section 3** is needed in this case, and **section IV** is skipped. Cartesian optimization of atom positions is possible (→CARTOPT).	
MADONLY	not set	Only the Madelung energy is printed.	
NOEWALD	not set	Deactivation of the calculation of the Madelung potential for (→CCM) calculations via the Ewald summation technique. Only recommended for nonpolar systems or ionic systems with rapidly converging electrostatic potential, e.g. systems with rocksalt structure.	[34]
NOWSCUPD	not set	No update of the Wigner-Seitz cell (size, included atoms and their weighting factors) during a Cartesian geometry optimization. Recommended in particular for calculations of point defects where there might be convergence problems during optimization (but see →NEWWEIGHT). **Note:** if NOWSCUPD is specified, it is necessary to rerun a second optimization using the optimized coordinates of the first run as start values, and to examine the energy convergence of these runs.	
NEWWEIGHT	not set	**Only for orthorhombic cells**: New weighting scheme for atoms close to the boundaries of the atomic Wigner-Seitz cells in CCM. Instead of a step function a smoother Becke-like weighting function is used. This is the same as WEIGHTFCT=12.	
NOAVERAGING	not set	Averaging of degenerate orbitals is avoided in →CCM calculations. Useful for molecular crystals where intramolecular MO degeneracies occur.	

10. Cyclic Cluster Model (cont.)

Keyword	Default	Description	Lit.
REWALD	not set	File name (less than 14 characters) for reading point charge values and positions of a previous run (e.g. those of a perfect solid, if in the new calculation the formation of point defects in the constant field of the unperturbed crystal is to be modeled). If REWALD is not specified, the Madelung potential of each cluster atom contains the defect. See (→PRINTOPTS=EWALDFIL).	
WEWALD	not set	File name (less than 14 characters) for writing point charge values and positions to be used by a later calculation. See (→REWALD) and (→PRINTOPTS=EWALDFIL).	
MRADIUS	= 5	Minimum values for cut-off radii in both reciprocal and direct space in the optimization of (→RRADIUS,DRADIUS) in the Ewald summation (→EWALD).	
CONCRI	$= 10^{-9}$	(Unit a.u.) Potential convergence criterion for the optimization of (→RRADIUS,DRADIUS) in the Ewald summation (→EWALD).	
CONFAC	$= \frac{\sqrt{\pi}}{\sqrt[3]{V}}$	Convergence factor for the Ewald summation (→EWALD). The default is that CONFAC is calculated from the cell volume of the cyclic cluster. This implies that also for 2D (→CCM) calculations all three lattice vectors have to be specified (→VECTA,VECTB,VECTC).	
WSCATOM	= 0	Print relative positions of all atoms in the WSC of atom WSCATOM; Warning: produces large output! Best together with (→INPCHK).	
LATTICECYC	= 8	Maximum number of optimization steps for the optimization of the lattice vectors (→LATTICEOPT, FULLOPT).	
LATTICEENLIM	10^{-6}	(Unit a.u.) Energy limit for the optimization of the lattice vectors (→LATTICEOPT, FULLOPT).	

10. Cyclic Cluster Model (cont.)

Keyword	Default	Description	Lit.
CUB	not set	All lattice vectors are treated equally during the optimization (→LATTICEOPT, FULLOPT, TET, ORT).	
TET, HEX	not set	The first two vectors (→VECTA, VECTB) are treated equally during the optimization (→LATTICEOPT, FULLOPT, CUB, ORT) the third one (→VECTC) is optimized separately.	
ORT	not set	All lattice vectors (→VECTA, VECTB, VECTC) are optimized independently (→LATTICEOPT, FULLOPT, CUB, TET).	
FULLCYC	= 2	Maximum number of cycles for (→FULLOPT).	
CCESCALEA	not set	The length of vector A (→VECTA) is scaled by this value, before performing the calculation.	
CCESCALEB	not set	The length of vector A (→VECTB) is scaled by this value, before performing the calculation.	
CCESCALEC	not set	The length of vector A (→VECTC) is scaled by this value, before performing the calculation.	

2.3.6 Nudged Elastic Band Method

The keyword NEBCALC activates the calculation of the minimum energy path based on the nudged elastic band (NEB) method between at least two configurations of a group of atoms.

Geometry definition (Cartesian coordinates)

The input structures have to be given in cartesian coordinates. The positions of each atom are specified in one input line. Note that the order of atoms has to be identical in all given structures. Empty lines are ignored. The input ordering is as follows:

$\mathbf{AN(A)}$	$\mathbf{X}_r\mathbf{(A)}$ $\mathbf{Y}_r\mathbf{(A)}$ $\mathbf{Z}_r\mathbf{(A)}$	$[\ \mathbf{X}_{TS}\mathbf{(A)}$ $\mathbf{Y}_{TS}\mathbf{(A)}$ $\mathbf{Z}_{TS}\mathbf{(A)}\ ]$	$\mathbf{X}_p\mathbf{(A)}$ $\mathbf{Y}_p\mathbf{(A)}$ $\mathbf{Z}_p\mathbf{(A)}$
	coordinates of reactand	optional guessed TS-structure(s)	coordinates of product

- **AN(A)** is the integer atomic number or the element symbol of atom **A**.
- $\mathbf{X}_i\mathbf{(A)}$, $\mathbf{Y}_i\mathbf{(A)}$ and $\mathbf{Z}_i\mathbf{(A)}$ are the Cartesian coordinates in Å of the reactant and product structures which will not be optimized. If desired, several guessed transition structures can be inserted between the start and end structure. These will be optimized.
- The input is unformatted.

Keywords

Keyword	Default	Description	Lit.
NEBCALC	not set	Activation of the nudged elastic band calculation	[51]
NEBIMG	5	Number of structures to be created by linear interpolation between each pair of given structures	
NEBSTRCT	not set	Activation of more strict convergence criterion	
NEBMAXCYC	999	Maximum number of optimization cycles	
NEBVARK	not set	Activation of the energy dependence of the spring forces leading to a higher resolution of the reaction path in high energy regions (**recommended**)	[52]
NEBCI	not set	After 5 optimization cycles the structure of the highest energy is turned into a climbing image which will converge to a saddle point. Setting this option is highly **recommended**.	[53]
NEBMTHD	3	Optimization method: 1) steepest descent 2) variation of a molecular dynamics method 3) BFGS	

Keyword	Default	Description	Lit.
		4) conjugated gradients 5) modified steepest descent	
NEBK	25.0	Spring constant in $\frac{\text{eV}}{\text{Å}^2}$ used to control the spacing of the structures	
NEBDEC	0.90	Factor to reduce trust-radius in case the energy has not been lowered by the previous step.	
NEBINC	1.00	Factor to increase "reduced" trust-radius in case the energy has been lowered again.	
NEBTIME	2.067	Time step in au (0.5 fs) used in optimization method 2	
NEBDRMAX	0.5	Max. step width / initial trust-radius in Å.	
NONEBROT	not set	Disactivate temporary rotation of neighboring structures to achieve best agreement of those structures before calculating the tangent.	

2.3.7 Molecular dynamics

Keyword	Default	Description	Lit.
MOLECULAR_DYNAMICS	not set	Activates the Born-Oppenheimer molecular dynamics run.	
NOSHIFTCM	not set	Center of mass of the isolated molecules will not be shifted to the origin at every steps of the molecular dynamics simulation.	
TIMESTEP	= 20.0	Time step in atomic units (1 a.u.≈ 0.024 fs). NOTE: This is **different** from versions up to 3.2.1 where the unit was ps.	
TINITIAL	= 300.0	Initial temperature (in Kelvin) for the molecular dynamics simulation.	
SCALEFRQ	= 10	Frequency at which velocities are scaled to reach a target temperature.	
PRNTFRQ	= 10	It is the frequency at which the informations from the molecular dynamics simulations are printed in the output file.	
TRAJFRQ	= 1	It is the frequency at which the trajectory files **<molecular formula>.molden** and **TRAJECTORY** are written.	
PROCESS	not set	Types of molecular dynamics simulations to be done are mentioned here. The processes which can be mentioned here are H(heating),C(cooling), E(equilibrating), NVE(micro-canonical ensemble run), NVT(canonical ensemble run) and Q(quenching). One can mention one or more process in round brackets separating them by a coma.	

4. Molecular dynamics

Keyword	Default	Description	Lit.
		For example: PROCESS = (H,E,C,NVE). By this definition, user ask the program to perform heating, equilibration, cooling and an NVE ensemble simulation in a sequential manner. NOTE: 1. Heating, cooling and equilibrating are just user friendly definitions. The program do an annealing (by linear scaling of velocities) with out recognizing them separately. 2. A space between the characters and brackets may cause an input error. So please try to avoid it. 3. Maximum number of processes is set to 10 as default. But it can be increased by changing the matrix size in the include file 'mdvar.h'.	
TTARGETS	not set	The target temperatures for the processes are specified here. In a similar way as for the keyword PROCESS, target temperatures are mentioned inside round brackets separating them by a comma. NOTE: For equilibration, NVE and NVT runs target temperatures should not be mentioned. Eg: If processes are defined as PROCESS=(H,E,C), then TTARGETS=(300.0,0.0). This mean that system is heated to 300 K, then equilibrated at that temperature and then cooled down to 0 K.	

4. Molecular dynamics

Keyword	Default	Description	Lit.
STEPS	not set	Number of steps for each process can be mentioned here in the similar input format as above. E.g. for PROCESS=(H,E,C) TTARGETS=(300.0,0.0), STEPS=(200,500,500). It means that system is first heated to target temperature in 200 steps, and then equilibrated for 500 steps and then cooled back to 0 K in 500 steps.	
QFRQ	= 50	The frequency (in steps) at which system should be quenched (if one uses 'Q' as a process).	
ATMFIX	not set	The key word for constrained dynamics. By this key word, one can select the moving atoms during MD run. Number of free atoms and then the input numbers of the free atoms should be mentioned in a similar way as the keyword CARTSLCT in **Section 5** of the input (see Section 2.3.20 of the manual).	
TEMPBATH		Here one can define the method for controlling the temperature during a NVT run. The following options presently exists: = NOSE_HOOVER_CHAINS Uses Nosé-Hoover chain thermostat. It is recommended for a good canonical sampling of the phase-space. = BERENDSEN Uses Berendsen's temperature bath. If no TEMPBATH is mentioned the temperature is controlled by a constant velocity scaling for the NVT run.	

4. Molecular dynamics

Keyword	Default	Description	Lit.
BATHTEMP	not set	Target temperature of the temperature bath or thermostat. It must be mentioned for restarting an NVT simulation.	
TAUT	= 0.4	Coupling constant for the Berendsen's temperature bath in ps.	
NHNC	= 4	Number of Nosé-Hoover chains.	
NSUZUKI	= 5	(To use with Nosé-Hoover chains option) Order of Yoshida-Suzuki integration. One can use 3 or 5 or 7.	
NMULTINT	= 3	(To use with Nosé-Hoover chains option) Number of multiple time steps.	
NOSEFRQ	= 2500.	(To use with Nosé-Hoover chains option) The thermostat frequency in cm^{-1}.	
NOSPEAT	not set	One set of Nosé-Hoover chains for each atoms; better for initial equilibration runs.	
FREEDOM		Determines the number of degrees of freedom. Following are the options: = NOTRAN : (default) No translation of the molecule is assumed. Number of degrees of freedom is calculated as $3N$-3 where N is the number of moving atoms of the system. = NOROT : No rotation of the molecule is assumed. Number of degrees of freedom is calculated as $3N$-6. = FULL : Assumes $3N$ degrees of freedom.	

4. Molecular dynamics

Keyword	Default	Description	Lit.
MD_OPTIONS		Using this keyword, one specifies different options for the molecular dynamics simulations. More than one options can be entered by separating them by a coma: = DATAFORM : The data will be written in a formatted file = NODATA : No data file will be written = TRAJFORM : The trajectories will be written in a formatted file = TRAJALL : Trajectories of all the atoms in the system will be written. If this key word is not mentioned, the trajectories of the moving atoms are only written in the trajectory file. = NOTRAJ : No trajectory file will be written. = MDDEBUG : To debug only the molecular dynamics routines. = NORESTART : No restart files will be written. = SNAPS : To write the snapshots at certain steps which are mentioned by the keyword SNAPAT. The velocities and coordinates will be written in files called snap.vel and snap.cords respectively. This keyword helps to restart the dynamics from these snapshots in case of an accidental crash of the computer during the simulation. See **Section 12** for more details.	

4. Molecular dynamics

Keyword	Default	Description	Lit.
		= EXIT : By this keyword one can stop the molecular dynamics simulation by introducing a file called EXIT in the directory at which the calculation is running.	
MAXCPUT	= Infinity	Maximum CPU time for the MD simulation. Program will exit if the CPU time exceeds the mentioned limit, by writing the restart files.	
SNAPAT	not set	By this keyword one can specify the steps at which the snapshots should be taken. One must specify SNAPS in MD_OPTIONS to activate this process. Steps must be mentioned inside round-brackets separating comma with out any space in between them. For example: SNAPAT=(100,200,300) to write the snapshots at 100 th, 200 th and 300 th steps respectively.	
RESETVEL	not set	The key word will reset the velocities to zero (quenching) during a NVE run, if the temperature will increase more than that specified using RWINDOW. It helps during simulated annealing to get a more relaxed structure from a very arbitrary starting structure.	
RWINDOW	4000.0	The temperature above which the system must be quenched during an NVE run. This will be only activated by the keyword RESETVEL.	

4. Molecular dynamics

Keyword	Default	Description	Lit.
RESTVSCAL	= not set	Using this keyword, the restarting velocities can be scaled to a definite temperature. The temperature in Kelvin should be mentioned after this keyword. In this way velocities read from RESTART.VELO or snap_in.dat can be scaled.	
ASSIGN_MASS	not set	User defined masses for atoms. Atom masses have to be defined in **Section V** of the input (see Section 2.3.20 of the manual).	
META_DYNAMICS	not set	Activates the metadynamics simulation. Collective coordinates and related parameters have to be defined in **Section V** of the input (see Section 2.3.20 of the manual). An extended Lagrangian formalism of metadynamics is implemented. The biasing potential is of the form $V(t, \mathbf{s}) = \sum_{t_i<t} H(t_i) \exp\left\{-\frac{[\mathbf{s}(t)-\mathbf{s}(t_i)]^2}{2[\mathbf{w}(t_i)\delta s(t_i)]^2}\right\}$	[50]
HILL_WIDTH	= 0.05	The width of the above Gaussian bias, δs. (→META_DYNAMICS)	
HILL_HEIGHT	= 0.001	Height of the above Gaussian, H (in a.u.) (→META_DYNAMICS)	
NCV	not set	Number of collective coordinates (→META_DYNAMICS)	
DTMETA	= 50	Frequency (in MD steps) at which biasing potential is updated (→META_DYNAMICS)	
NHILLMAX	= 9999	Maximum number of metadynamics steps in a single run of MD (→META_DYNAMICS)	

4. Molecular dynamics

Keyword	Default	Description	Lit.
ADAPTIVE_DT	-9999.0	An adaptive metadynamics step will be used based on the displacement of the CV from the previous Gaussian center. The minimum displacement is read after "=" sign. Note: if this displacement criteria is not satisfied within DTMETA MD steps, Gaussians will be added.	
META_TART	=TINITIAL	Initial temperature for the collective variables in K; by default it is set to the initial temperature of the nuclei	
COMETT	not set	Activates the temperature control for the collective variables; instantaneous temperature of the collective coordinates will be scaled if they are larger or smaller than the given tolerance META_TTOL from the initial temperature META_TART.	
META_TTOL	=META_TART	The tolerance allowed in the instantaneous temperature of CV (→COMETT)	

2.3.8 Input-Output files for molecular dynamics

- snap_in.dat : This formatted input file should be prepared from snap.vel file for restarting from a snapshot.
- snap.vel : This formatted file contains the velocities of the snapshots. This will be printed as atom number, velocities in x, y and z directions of an atom along with an identification of the step at which the snapshots are written. The velocities must be copied to an another file with the name snap_in.dat. This file must only contains atom number and velocities in x, y and z directions.
- snap.cords : Formatted file contains the coordinates at different steps at which the user is asked to print the snapshots. On restarting from a snapshot, one must change the input coordinates to these coordinates.
- mddata.out : Formatted file. Total MD steps, temperature (K), kinetic energy(a.u.), potential energy(a.u.), binding energy(a.u.), total energy(a.u.) and CPU time (s) are printed.
- mddata.unfmt.out : Unformatted MD data file. It contains the data as mentioned in the case above.
- TRAJECTORY : Unformatted file. MD step, coordinates , and velocities are printed in the following format: steps, Cx, Cy, Cz, Vx, Vy, Vz (where C represents the coordinates and V represents the velocities).
- TRAJECTORY.FMT : Formatted TRAJECTORY file.
- RESTART.DENS : Restart file for Densities (unformatted).
- RESTART.CORD : Restart file for co-ordinates (unformatted).
- RESTART.ACC : Restart file for accumulators (unformatted).
- RESTART.NOSE : Restart file for Nosé-Hoover chain thermostat variables (unformatted).
- RESTART.VELO : Restart file for velocities (unformatted).
- RESTART.META : Restart file for metadynamics
- meta.in : This formatted input file is read during the metadynamics simulation, if it is not restarted. By this file, user defines the scaled collective coordinates for the metadynamics. The format of the this file is as follows.
  First line should be the character which defines the type of the collective coordinate. The

following are the characters for different collective coordinates.

- L - For bond length
- A - For bond angle
- D - For dihedral angle
- C - for coordination number. See (→EXCONS)

If you define L then the next line should be
N1 N2 R S
where N1 and N2 (integers) are the input number of the atoms which defines the bond which is selected as the collective coordinate. Please not that this input number is the number of the atom with out counting the dummy atoms in the input. R (a real number) is the scaled collective coordinate and S is the scaling factor. R is unit less and S should be in Bohr(a.u.).
In a similar way, if you define A or D then the next lines will be
N1 N2 N3 R S
or
N1 N2 N3 N4 R S
For A, the angle is defined as angle N1-N2-N3 with N2 as the center atom. R is unitless and S is in radians. For dihedral angle, the angle is defined between the planes containing N1-N2-N3 and planes containing N2-N3-N4.
If the collective coordinates are average coordination numbers, then the definition is as follows:
C
NA NB
DAB
A1 A2 .... NA B1 B2 .... NB R S
where NA and NB are the number of type A atoms and type B atoms respectively. DAB is the cut-off distance for the bond between type A and type B atoms in Bohr (a.u.). A1 ... NA are the input numbers of type A atoms, followed by B1 ... NB which are the input numbers of type B atoms. R is the scaled average coordination number and S is the scaling factor.
All the collective coordinates can thus be defined one after the another. But it should be noted that when more than one coordination numbers are defined, then they should be in the same order as (→EXCONS) are defined.

- meta_data.out This formatted file contains the data from metadynamics simulation. The format of this file is as follows:
  MD step, all collective variables, velocities of all collective variables, all collective coordinates, total en energy of the collective variable, instantaneous temperature of the collective variables.
- GAUSS.META This unformatted file contains the collective variables during the whole metadynamics simulation. New collective variables are added to this file at every time step of the simulation. The file will be rewinded and read completely during the updating of the forces from the Gaussian terms. To restart the metadynamics, one need this file in the same name.

2.3.9 Post-MD tools :

Informations in the files TRAJECTORY or TRAJECTORY.FMT and mddata.out or mddata.unfmt.out are used to calculate many properties of the system. The program to do this post MD simulations is available in the directory post_md_tools. The following are the available routines:

- **Velocity auto-correlation function; vibrational frequencies**
- **Mean square displacement; diffusion coefficient**
- **Bond length fluctuation**
- **Bond angle distribution**
- **Bond length distribution**
- **Binding energy sampling/distribution**
- **Heat capacity for NVE and NVT ensemble**

2.3.10 Reaction field calculations

For calculations of molecules in solvents both the polarizable continuum model (PCM) [39] and the Conductor-like Screening Model [54] are implemented. In both cases the solvent is characterized by its macroscopic dielectric constant. The solvated molecule is placed in a cavity in the continuous solvent. In the case of the PCM, the shape of the cavity is calculated as an isodensity surface from the electronic properties of the molecule. In contrast, the cavity shape is independed of the electronic structure of the molecule and defined entirely by its geometry in the case of the COSMO. In both cases the dielectric is polarized by the electrostatic potential of the molecule, yet the two methods differ in the way, the polarization is calculated.
While the COSMO is used to calculate the electrostatic contribution to the energy of the solvation only but very efficently, the PCM includes non-electrostatic contributions as well.
A way to include solvent effects on excited states is only implemented for the COSMO. The PCM can only be used for ground state properties.

polarizable continuum model (PCM)

The polarization of the dielectric by the molecular electrostatic potential is calculated using the asymptotic density model ADM. The PCM is used to calculate the electrostatic contribution to the free energy of solvation. For the non-electrostatic contributions more simple approaches are used.

The output contains the free energy of solvation in kcal/mol. In addition to the total value also contributions like electrostatic solvation energy, dispersion energy, repulsion energy and cavity energy are listed individually. Molecular properties in solution are calculated with the PCM derived wave function.
The PCM has been parametrized for the elements H, Li-F and Na-Cl.

Keyword	Default	Description	Lit.
PCM	not set	Polarizable continuum model (PCM) using isodensity surfaces for the cavity surface and the asymptotic density model (ADM) for the required electrostatic potentials.	[39, 40]
DIEKON	= 78.5	Dielectric constant of the solvent; default is the value for water under normal conditions (25 °C, 1 bar). Dielectric constants for common solvents may be taken from the literature.	[41]

8. Reaction field calculations (cont.)

Keyword	Default	Description	Lit.
SCHRIT	= 1.5	(Unit Bohr); intercept length for the Marching Cube Algorithm to generate the isodensity surface. The accuracy of the calculation and the computer time increase with decreasing values of SCHRIT. The default value is a well-balanced compromise between accuracy and computer time and should not be changed.	[42, 43]
GRENZW	= 0.002	(Unit e/Bohr^3); value for the electronic density for calculating the isodensity surface. "Magic number", only experts should change this parameter.	
ITER	set	Iterative method to calculate the surface charge in the PCM; this standard technique should be used for all single point calculations.	[39]
CLS	not set	Surface charges are estimated with the Partial Closure Method; this procedure is faster than the iterative method (→ITER) but less accurate. Recommended for geometry optimizations.	[44]
SOLVENT	none	This keyword defines the solvent for the calculation of non-electrostatic contributions to the solvation energy; after the keyword follows (enclosed in parenthesis) a list of atomic numbers defining all atoms of the solvent molecule, separated by commas. No spaces are allowed between keyword, parenthesis, commas, and numbers.	
DICHTE	= 0.0	(Unit g/cm^3); the macroscopic density of the solvent.	

Conductor-like Screening Model (COSMO)

In the Conductor-like Screening Model (COSMO) the cavity is defined by the geometry of the molecule or periodic cluster. Three different types of cavities (van der Waals type, Klamt-type and solvent accessible surface, see [55]) are available which are all created using the GEPOL algorithm [56].

type of cavity	radius of solvent	scaling factor
van der Waals	0.0	arbitrary
Klamt	≈ 1.3	0.1
solvent accessible surface	≈ 1.3	1.0

To calculate the fluorescence energy of a molecule in solution, two different approaches are available.

Keyword	Default	Description	Lit.
COSMO	not set	Conductor-like Screening Model (COSMO)	[54]
COSMOGRID	(0,3)	Tesselation grids of the cavity. The first number defines the maximum size of the segments, the second the minimum. The larger the number the finer...	
COSMOCIS	1	Approach to calculate the solvent effects on relaxed excited states. 1. single determinant approach 2. Hartree-Fock analogon	
COSMOOLC	not set	Defines the value ”maximum charge of a single segment” when calculating the color of a segment for the povray-output. (→PRINTOPTS=POVRAY) This option does not affect the real segment charges. Its only purpose is to define the color palette to compare colors of povray images of different systems.	
DIEKON/DIELEC	78.5	Dielectric constant of the solvent; default is the value for water under normal conditions	

COSMO (cont.)

Keyword	Default	Description	Lit.
SOLVENT	not set	Predefined dielectric constants and refractive indexes of widely used solvents • WATER: $\epsilon = 78.4$, $\eta = 1.33$ • TOLUENE: $\epsilon = 2.38$, $\eta = 1.4969$ • TRICHLORMETHAN: $\epsilon = 4.89$, $\eta = 1.4459$ • METHANOL: $\epsilon = 32.66$, $\eta = 1.33$ • ACETONITRILE: $\epsilon = 37.5$, $\eta = 1.35$ • ETHANOL: $\epsilon = 24.6$, $\eta = 1.361$ • ACETONE: $\epsilon = 20.7$, $\eta = 1.3561$ • CYCLOHEXAN: $\epsilon = 2.02$, $\eta = 1.43$ • HEXAN: $\epsilon = 1.88$, $\eta = 1.37$ • DIPHENYLETHER: $\epsilon = 3.60$, $\eta = 1.58$ • THF: $\epsilon = 7.58$, $\eta = 1.4072$ • DMF: $\epsilon = 36.71$, $\eta = 1.4305$	
REFRAC	1.33	Refractive index of the solvent; default is the value for water under normal conditions	
RSCAL/RSCALE	0.1	scaling factor of van der Waals spheres to create a cavity of Klamt type	[54]
RSOLVE	0.00	Radius of the solvent in Angstrom; default is zero which corresponds to a van der Waals cavity.	
ZCUTOFF	not set	When modeling a solid-liquid surface by a CCM2D cluster where the lowest layers are held fixed to model the solid, it is possible to model solvation only above the upper layers by defining the atom whose z-coordinate defines the lower limit for solvation.	

2.3.11 Symmetry analysis

The program recognizes most point groups including O_h, T_d and I_h. The symmetry of the molecular framework is analyzed and the irreducible representations of the MO eigenvectors and the normal vectors from the vibration analysis are given (→NVIB).
The calculation of molecular integrals is not affected by the global symmetry because local symmetry is exploited [8].

Keyword	Default	Description	Lit.
NOSYM	not set	Symmetry analysis is deactivated.	
STDO	not set	Before starting the first SCF run the molecule is translated and rotated into standard orientation. A subsequent optimization in internal coordinates (→OPT) is, however, performed in input orientation. Do not combine (→STDO and OPT) with (→IDEN=2 or 3)! An optimization in Cartesian coordinates (→CARTOPT) will be performed in standard orientation.	

2.3.12 Vibration analysis

Setting NVIB > 0 activates the vibration analysis. If the geometry is defined in internal coordinates, the force constant matrix can be calculated in both internal or Cartesian coordinates. For Cartesian inputs an analysis in internal coordinates is not possible and usually makes no sense. In general it is recommended to perform normal mode calculations in Cartesian coordinates, because in most internal geometry definitions there are linear dependencies.

Keyword	Default	Description
NVIB	= 0	Default is to perform no vibration analysis **1:** Diagonal force constant matrix in internal coordinates; is recommended only as a fast approximation. Zero point energies may deviate by several kJ/mol from results obtained with higher values for NVIB, vibration frequencies may change by hundreds of wavenumbers. **2:** Full force constant matrix in internal coordinates; in many cases reliable, but there are problems for ring systems and closed three-dimensional structures. **3:** Full force constant matrix in Cartesian coordinates (non-symmetric geometry changes for numerical calculation of second derivatives); more accurate than NVIB=1,2, but much more time consuming. **4:** Full force constant matrix in Cartesian coordinates of selected atoms (symmetric geometry changes for numerical calculation of second derivatives); the most accurate choice within the limitations of the method. Suitable for calculations of vibrations of adsorbed molecules. The total number and the numbers of the selected atoms has to be specified in **Section 5** (but see (→FULL).
FULL	not set	Select all atoms for a vibration analysis with (→NVIB=4); no additional input is needed in **Section 5**.
THERMO(TMIN,TMAX,TSTEP)	not set	For (→NVIB=4): calculation of thermodynamical functions (H, G, S) in a temperature interval between TMIN and TMAX in steps of TSTEP (all in K).

2.3.13 Embedding procedure

For the simulation of bulk and surface properties with finite cluster models the interaction between the atoms of the internally defined system and atoms from the surrounding has to be included. In this embedding scheme the surrounding atoms are represented by a finite array of pseudo atoms. The positions of the pseudo atoms are generated by translations of cluster atoms. The model cluster must have the correct stoichiometry of the bulk material. Also the translations must generate stoichiometric amounts of pseudo atoms.

Keyword	Default	Description	Lit.
EMBED	not set	Activation of the embedding procedure	[45]
TRANS(n_A, n_B, n_C)	3×0	(Only for →EMBED). Effective translation vector; generated from the basis vectors (→VECTA, VECTB, VECTC) and the TRANS components n_A, n_B, n_C. Each translation vector has to be given explicitly. It is highly recommended to form symmetric shells of pseudo atoms around the cluster, e.g. for cubic systems by giving all combinations ranging from (−a,−a,−a) to (a,a,a). No spaces are allowed between keyword, parentheses, commas, and numbers.	
BESZ(N)	= 0.0	Number of valence electrons for the element with atomic number N; If (→KONSIST) is not specified, the charges of pseudo atoms generated from real atoms with atomic number N are taken as external parameters. Each element occurring in the input has to be assigned a value. This procedure saves computer time because integrals between real atoms and pseudo atoms have to be calculated only once before starting the SCF run. Best choice for BESZ is the averaged atomic occupation number for the particular element from a calculation without embedding. The pseudo atom charges must sum up to zero. No spaces are allowed between keyword, parentheses, and number.	

9. Embedding procedure (cont.)

Keyword	Default	Description	Lit.
APBESZ(N)	= 0.0	Number of valence p electrons for the element with atomic number N which is an adsorbate atom. Used to differentiate between surface atoms and adsorbate molecules.	
ASBESZ(N)	= 0.0	Number of valence s electrons for the element with atomic number N which is an adsorbate atom. Used to differentiate between surface atoms and adsorbate molecules.	
KONSIST	not set	Alternative to fixed atomic occupation numbers (→BESZ); the pseudo atom charges are recalculated in each SCF cycle from the average orbital occupation of corresponding cluster atoms. Needs a lot of computer time because pseudo atom-atom integrals have to be calculated in each SCF cycle.	
NICHTT($n_1, \dots$)	—	Numbers of atoms in the input sequence (→ **Section 3**) which will be excluded from translations; useful e.g. to avoid the translation of adsorbed molecules on a surface so that adsorbate-adsorbate interactions are switched off. Several atoms can be specified with one NICHTT. There may be more than one occurrence of NICHTT specifications in one input. No spaces are allowed between keyword, parentheses, commas, and numbers.	
ANPLUS(N)	= 0	The dummyatom with input number N (→ **Section 3**) is translated and creates pseudo atoms with atomic number ANPLUS(N); If this atomic number is not already included in the input, an occupation number (→BESZ) has to be given. Any dummyatom of the internal input may be chosen for translation. For each dummyatom ANPLUS has to be specified separately. This option may be useful for the simulation of point defects, if the embedding should not contain the defects. No spaces are allowed between keyword, parenthesis, and number.	

9. Embedding procedure (cont.)

Keyword	Default	Description	Lit.
ADSORB($N_1, N_2, \ldots$)	not set	Comma-separated list of atoms that should be treated as adsorbate atoms.	
ADSESZ(AN$_1$, AN$_2$,…)	not set	Comma-separated list of s orbital occupation numbers of adsorbate atoms	
ADPESZ(AN$_1$, AN$_2$,…)	not set	Comma-separated list of p orbital occupation numbers of adsorbate atoms	
ADBESZ(AN$_1$, AN$_2$,…)	not set	Comma-separated list of d orbital occupation numbers of adsorbate atoms	

2.3.14 Output control

Keyword	Default	Description	Lit.
IZUS	= 0	Total number of additional bond lengths, angles and dihedral angles to be printed out; in **Section 5** the appropriate number of additional internal coordinates has to be given.	
NPROP	= 0	Setting NPROP≠0 activates the calculation of electrostatic properties from ADM. For (→RHF) wave functions only. Since the ADM deals with the nonorthogonal basis, (→PRINTOPTS=MUL) has to be set. **1:** Electrostatic potentials in a plane. **2:** Electronic densities in a plane. **3:** Electrostatic potentials on an isodensity surface. The starting points and the basis vectors for the grid or the density value for the isodensity surface have to be specified in **Section 5**.	[31]
NIVO	= 0	Code for the calculation of Improved Virtual Orbitals (Huzinaga's method). Unoccupied orbital energies are corrected through excitations from occupied MOs.	[32]

5. Printout options (cont.)

Keyword	Default	Description	Lit.
		0: All virtual MOs are corrected; excitations only from the highest occupied MO. **1:** The lowest **M1** unoccupied MOs are corrected; excitations from each of the **N1** highest occupied MOs. **2:** The unoccupied MOs **M1** to **M2** are corrected; excitations from the occupied MOs **N1** to **N2**. For NIVO=1,2 the numbers **M1,M2,N1,N2** have to be specified in **Section 5**.	

5. Printout options (cont.)

Keyword	Default	Description	Lit.
PRINTOPTS	STANDARD	Individual output control. Available options, either separated by commas (PRINTOPTS=option1,option2,...,optionN) or given by individual definitions: BMAT G matrix of vibrational analysis in internal coordinates (→NVIB). BONDOR Bond order matrix. Based on a method by Gopinathan and Jug. Based on the density matrix in a local diatomic coordinate system. See also (→VALENC). BONDORDIFF Bond order difference between selected excited state (→DVDSONCIS,REFSTATE) and ground state. BONDORL Bond order orbitals in the local coordinate system. CART Cartesian coordinates in input and standard orientation. CCMWSC CCM Wigner-Seitz cells (→CCM). CHARGEDIFF Löwdin charge difference between selected excited state (→DVDSONCIS,REFSTATE) and ground state. CIMAT CI matrix (→CI). CISPLOT Data table of Gaussian-broadened and oscillatorstrength-weighted excitation energies prepared for **Gnuplot** (file **<molecular formula>.cis**). CIVEC 10 lowest CI eigenvectors (→CI). CIVECFUL All CI eigenvectors (→CI). Not available for (→DAVIDSONCIS) DEBUG: Tracing of the subroutine calling sequence.	[33]

5. Printout options (cont.)

Keyword	Default	Description	Lit.
		DE2CAL Force constants of vibrational analysis (→NVIB).	
		DE2MAT Hessian matrix (geometry optimization) (→CARTDE2,LMK).	
		DE2VAL Eigenvalues of the Hessian matrix (→CARTDE2,LMK).	
		DE2VEC Eigenvectors of the Hessian matrix (→CARTDE2,LMK).	
		DIPTOT Total dipole moment (orthogonal basis).	
		DIPATO Atomic contributions of the dipole moment.	
		DISTAB Interatomic distances.	
		DMADKONST Derivatives of the Madelung constant matrix (→CCM).	[34]
		DOS Density of states for α, β orbitals; projections to s, p, d orbitals. Data written to external files **<molecular formula>.alphados etc.** in a format appropriate for the visualization program **Gnuplot**.	[35]
		EWALDFIL Print data for Ewald summation in (→CCM) read from/written to an external file (→REWALD,WEWALD).	
		EWALDGR Print the convergence factor and the radii for the Ewald summation.	
		EWALDWSC Wigner-Seitz cells for Ewald summation (→CCM).	
		FMAT Print Fock matrix after SCF convergence.	
		FORCES Print force constants of vibrational analysis (→NVIB).	
		FRACCOR Prints the fractional coordinates of the input atoms with respect to the lattice vectors (→CCM).	
		FULLCOR Detailed printing of coordinates and derivatives during geometry optimization (→OPT,CARTOPT).	

5. Printout options (cont.)

Keyword	Default	Description	Lit.
		GMAT GAMMA matrix (coulomb and exchange integrals).	
		HMAT Core Hamiltonian matrix.	
		HUVEC Eigenvalues and eigenvectors of Hückel matrix (→IDEN=0,2).	
		HUEIG Only eigenvalues of Hückel matrix (→IDEN=0,2).	
		INPCCM Generates a new cartesian input (fort.9) for (→CCM) calculations from any CCM-MSINDO input.	
		INPUT Detailed information about input data.	
		IONPOT Vertical ionization potential according to Koopmans' theorem. See (→KATION) for non-vertical ionization energies.	
		IVO Improved virtual orbitals. Method by Huzinaga.	[32]
		KATION The first ionization potential is calculated as energy difference of the neutral system and the positively charged system (→CHARGE,MULTIP). The two inputs must be given in the above order in one file. The KATION option must be specified in the input of the charged system.	
		LCHARG Atomic Löwdin charges. Diagonal elements of the density matrix in the orthogonal basis.	
		MADKONST Madelung constant matrix (→CCM).	
		MADPOT Madelung potential at atom positions (→CCM).	
		MINIMUM: Deletes the default options – minimal ouput for large systems.	
		MOCON Compressed MO vectors. The seven largest atomic contributions to each MO are printed.	

5. Printout options (cont.)

Keyword	Default	Description	Lit.
		MOLDEN Molecular structure (initial structure and at every optimization step), SCF convergence, geometry optimization, vibration frequencies (→NVIB=4), are written to an external file <**molecular formula>.molden** in a format readable by the visualization software **MOLDEN**.	[36]
		MOLDENMO **MOLDEN** input for plotting MOs.	
		MOVEC Full MO vectors.	
		MULAP Atomic Mulliken charges. Based on the orbitals in the non-orthogonal basis after transformation from the orthogonal basis. See (→S12APPROX).	
		MULAOP Atomic Mulliken orbital charges.	
		MULDEN Density matrix in the nonorthogonal basis.	
		MULDIP Dipole moment (nonorthogonal basis).	
		MULOVD Mulliken overlap charge matrix.	
		MULOVDR Reduced Mulliken overlap charges.	
		MULVEC MO vectors in the nonorthogonal basis.	
		NATORB Natural orbitals (→CI).	
		NEIGHBOR=N Print for each atom its N nearest neighbors and the average coordination number for each element. If no value is given for N, the 6 nearest neighbors are printed.	
		OSCIL Oscillator strengths in (→CI) calculations.	
		POVRAY Structure plot in ray-tracer Povray format in file <**molecular formula>.pov**; ball model of Mulliken charges (→PRINTOPTS=MUL) in file <**molecular formula>_charge.pov**; ball model of CIS $S_n - S_0$−charge differences in file <**molecular formula>_cischarge.pov**.	
		RESTDEN Writes out restart file for the density.	

5. Printout options (cont.)

Keyword	Default	Description	Lit.
		PMAT Full alpha and beta density matrices (→UHF,ROHF).	
		SCISPROP Properties of the selected excited singlet state (→DVDSONCIS,REFSTATE) are computed.	
		SCFCYC Energies for each SCF-cycle.	
		SCHAK SCHAKAL input of the initial structure (final structure in optimization runs). The labeling of the atoms now corresponds to the input specification. Formatted data are written to the external file <**molecular formula>.dat**.	[37]
		SHPOT Shell potentials in (→PCM) calculations.	
		SMAT Overlap matrix.	
		SPECTRUM Gnuplot readable data output for plotting CIS excitation spectra in file <**molecular formula>.spectrum** (→DAVIDSONCIS).	
		SPINDEN Atomic spin densities (→UHF,ROHF).	
		SPINDIFF Difference spin density (→DVDSONCIS,REFSTATE).	
		SPINEV Expectation value of the spin operator (→UHF,ROHF).	
		SPINMAT Spin density matrix (→UHF,ROHF).	
		STANDARD: INPUT,CART,SCFCYC,IONP,IVO,LCHARG, SPIN,DIPATO,DIPTOT,OSCIL,WAVEN.	
		SYMELM Symmetry elements.	
		TCISPROP Properties of the selected excited triplet state (→DVDSONCIS,REFSTATE) are computed.	
		TFQMRSTEP Prints residual of each TFQMR iteration (→CISGRAD,EXCVIB).	
		VALENC Interatomic valences based on the interatomic blocks of the density matrix. Always positive and additive, different from bond orders (→BONDOR).	

5. Printout options (cont.)

Keyword	Default	Description	Lit.
		VALRED CI and SCF valence reduction. Useful as criterion for the classification of molecules as diradicals and zwitterions.	[38]
		VALREDA Atomic contributions of the valence reduction.	
		VALREDO Orbital contributions of the valence reduction.	
		VIBVEC Normal coordinates in vibrational analysis (→NVIB).	
		WAVEN Vibrational wave numbers (→NVIB).	
		XCRYSDEN Structure data are written that can be read directly by XCrysDen.	
		XMOL Structure data are written in a format suitable as XMOL input (external file **fort.9**).	
		XYZREST : Writes out restart file for the coordinates.	
		ZEICH 2-dim drawing of the molecule. (deprecated feature)	

2.3.15 Restarting options

You can restart the MSINDO using the keyword MSINDO_RESTART.
The options for the restart must be specified after the above keyword separating an '=' sign.
More than one options can be written by separating them by a comma.

- DENSITY : The density will be restarted. But not completely operational for cyclic cluster method. The densities for restart can be written out specifying RESTDEN in PRINTOPTS if it is not a molecular dynamics run. In the case of a molecular dynamics run, the density restart file will be written automatically.
- COORDINATES : The coordinates can be restarted after a geometry optimization or a molecular dynamics run. But not completely functional for cyclic cluster method. The restart file for the coordinates can be written out by specifying XYZREST in the PRINTOPTS if it is not a molecular dynamics run. In the case of molecular dynamics run, the restart file for the coordinates will be written out automatically.
- VELOCITIES : The velocities can be restarted for a molecular dynamics run.
- SNAPSHOTS : To restart from a snapshot. You have to change in the input coordinates to the coordinates printed in the snapshot file snap.cords for the chosen step. Velocities at this step, which are printed in the file snap.vel must be copied to snap_in.dat. This file will be read for restarting the velocities.
- ACCUMULATORS: The accumulators for statistical calculations will be restarted for a molecular dynamics run.
- NOSE : Nose variables will be restarted for a NVT ensemble molecular dynamics run.
- META : Metadynamics parameters and variables are restarted from a formatted file RESTART.META

2.3.16 Input Check

By specifying the keyword **INPCHK** at any position in **sections II**, the input data given in **sections II-IV** are checked for consistency. No integral or SCF calculation is performed. Recommended for time-intensive calculations or to print the molecular input structure in a format suitable for external software (→SCHAKAL,MOLDEN,XMOL).

Section II is closed with the string **:END** (lower- or uppercase).

2.3.17 Geometry definition (internal coordinates)

The internal input defines the geometry of the system in terms of bond lengths, bond angles, and dihedral angles.

The relative position of each atom with respect to previously defined atoms is defined in a separate line.

The atoms are numbered and specified by their atomic numbers or element symbols. Bond lengths, bond angles, and dihedral angles may be specified as constants or as variables. The geometry optimization (→OPT in **section II**) is possible only for those coordinates given as variables. The first character of the variables has to be one of:

L or [R] for bond lengths,

W or [A] for bond angles, and

D for dihedral angles

The maximum length of variables is 9 characters (plus the sign "−"). Special characters (also the "+") should be avoided. Equal variables will be set to the same values (→ **section IV**) The use of upper or lower case letters makes no difference.

Constants do not have to include the decimal point. The units for lengths and angles are Å and degree, respectively. An integer value of, e.g., 180 is converted to 180.0.

The input is unformatted.

Numbers and variables have to be separated by at least one blank.

Empty lines are ignored.

The definition of the first three atoms is fixed by the program:

		1	**AN(1)**		
	1	**2**	**AN(2)**	$\mathbf{R}_{12}$	
1	**2**	**3**	**AN(3)**	$\mathbf{R}_{23}$	$\mathbf{A}_{123}$

AN is the integer atomic number of the element or the element symbol (upper- or lowercase), $\mathbf{R}_{12}$, $\mathbf{R}_{23}$, and $\mathbf{A}_{123}$ are the corresponding bond lengths and angles, given as (real or integer) constants or variables.

All following atoms have to be specified according to the scheme:

A	**B**	**C**	**D**	**AN(D)**	$\mathbf{R}_{CD}$	$\mathbf{A}_{BCD}$	$\mathbf{D}_{ABCD}$

$\mathbf{R}_{\mathrm{CD}}$, $\mathbf{A}_{\mathrm{BCD}}$, and $\mathbf{D}_{\mathrm{ABCD}}$ may be given as real/integer constants or variables. The dihedral angle $\mathbf{D}_{\mathrm{ABCD}}$ can be visualized by a Newman projection. $\mathbf{D}_{ABCD}$ is the projected angle from **A** to **D** looking in the direction of the **B**-**C** bond. The sign is positive for left rotation (mathematical convention).

In order to define highly symmetric systems it may be better to use the length $\mathbf{R}_{\mathrm{BD}}$ instead of the angle $\mathbf{A}_{\mathrm{BCD}}$ for the definition of atom **D**. The same is possible for the dihedral angle $\mathbf{D}_{\mathrm{ABCD}}$ which can be replaced by the angle $\mathbf{A}_{\mathrm{ACD}}$ or the length $\mathbf{R}_{\mathrm{AD}}$. The replacement is only possible using variables with the appropriate first character. Any other character except the minus sign will be regarded as an input error.

Another possibility for defining symmetric structures is to use dummy atoms. They are characterized by the atomic number zero or the symbol ”X”.

Geometry definition (Cartesian coordinates)

If in **section II** (→CARTES) was specified, the geometry definition has to be given in Cartesian coordinates. The position of each atom is specified in one input line. Empty lines are ignored. The input ordering is as follows:

AN(A) **X(A)** **Y(A)** **Z(A)**

- **AN(A)** is the integer atomic number or the element symbol of atom **A**.
- **X(A)**, **Y(A)** and **Z(A)** are the Cartesian coordinates in Å.
- It is possible to give real constants or variables.
- The values for all variables must be specified in **section IV**.
- The input is unformatted.

The geometry definition (internal or Cartesian) is closed with the string **:END** (lower- or uppercase). **For Cartesian inputs this is different from earlier versions! Be careful with old input files!**

2.3.18 Geometry definition (Unit cell input)

If in **section II** (→EZINP) was specified, the unit cell parameters and fractional coordinates of the atoms in the translational unit of a (→CCM) calculation have to be specified. The following data have to be given in separate lines (unformatted):

1. Unit cell vectors a, b, c (unit Å).
2. Angles between vectors α, β, γ (unit degrees).
3. Total number of atoms N in the unit cell.
4. Diagonal elements of the transformation matrix for a supercell generation (3 integer numbers).
5. Three fractional coordinates and the atomic number for all atoms in the unit cell. One atom per line (total: N lines of input). Allowed are: integer constants, real constants, variables, and fractions of integers, possibly combined with each other; e.g. 1/2-X.

In the following line(s) follows the assignment of values to the variables used above: VARIABLE = VALUE.
After the unit cell input there follows directly **Section 5**. No **:END** must be specified at the end of **Section 3**.

2.3.19 Assignment for variables

If in **Section 3** variables were used to define internal or Cartesian coordinates, they must now be assigned values (exception →EZINP). The assignment is done with the equal sign.
Variable, equal sign and the number should be separated by blanks.
Empty lines will be ignored.
For each variable defined in **Section 3** there must be one assignment.
On the other hand, all variables assigned in this section must match those in **Section 3**.
If no variables were defined in **Section 3**, this section is empty but has to be closed by the string **:END** (lower- or uppercase, see →Input example 1).
If necessary, assignments can be commented out with the character ”#”.
If in **section II** (→OPT) was specified, all coordinates defined by variables are optimized.
If a coordinate shall be excluded from optimization, an asterisk ”*” has to be added at the end of the variable's name (→Input example 11).
Such a distinction is presently not possible for Cartesian coordinates.
Here, a selection of coordinates to be optimized can only be performed with the (→CARTSLCT) keyword and the appropriate input in **Section 5**.

The assignments close with the string **:END** (lower- or uppercase).

2.3.20 Additional input

Depending on the specifications in **section II** additional input has to be given in the last section. This part is partially formatted, empty lines are not allowed and the ordering described below has to be kept.

The following keywords and control parameters require additional input:

1. IZUS > 0 – Additional coordinates
2. IDEN = 4, 6 – MO exchange
3. RHF/ROHF CI – CI with excitations
4. NIVO = 1, 2 – IVO corrections
5. NPROP = 1, 2, 3 – Electrostatic properties
6. CARTSLCT – Cartesian geometry optimization for selected atoms
7. CARTSDE2 – Explicit calculation of second derivatives for selected atoms
8. NVIB = 4 – Vibration analysis for selected atoms

If one item is not given, the next follows without an empty line. Section V and the entire input is closed with the string **END** in columns **1-3**. Otherwise, it is expected that the input of another molecule follows. If the string **END** is not given at the end of an input file, this will be regarded as an input error and the message

```
*** READ ERROR ON INPUT!!!! ***
*** REACHED END OF FILE     ***
```

will appear at the end of the calculation.

1. **Additional coordinates**
   If (→IZUS) was set greater than zero in **section II**, the appropriate number of additional bond lengths, bond angles and dihedral angles to be printed by the program has to be specified.
   Each coordinate is given in a separate line.
   Dihedral angles are defined by the numbers of four atoms, bond angles by three atoms, and bond lengths by two atoms.

The numbers can be specified in free format.
The total number of coordinates must be identical with (→IZUS).

2. **MO exchange**
   If (→IDEN) was set to 4 or 6 in **section II**, the Hückel MOs to be interchanged for the first start density have to be given.
   Only occupied and unoccupied, or, in the ROHF case, doubly and singly occupied or singly occupied and unoccupied MOs can be interchanged.
   In the first line the total number of exchanges is given in (I5) format.
   Then the corresponding number of lines contain the numbers of occupied MOs (first) and unoccupied MOs (second) in (2I5) format.
   For (→UHF) calculations the input has to be repeated for the β MOs.

3. **CI with excitations**

   For (→RHF CI) or (→ROHF CI) the configurations included in the CI wavefunction are given in terms of excitations from the SCF ground state determinant. Single and double excitations are possible. Excitations can be specified individually or by giving active spaces. Both types of specification are possible in one input.
   The first line contains in (2I5,F10.4,2I5) format:

   a) Number of the CI reference state for which properties are calculated.

   b) Multiplicity of the CI reference state.

   c) Total energy of a reference molecule (optional).

   d) Total number of calculated CI eigenvectors (optional); a small value leads to a reduction of CPU time during diagonalization of the CI matrix (→NGIV=5,8).

   e) Code for saving data to external files. If a nonzero number is given, the CI matrix and eigenvectors are stored (Fortran unit 55).

   The numbering of each multiplicity starts with 1 for the lowest state. The multiplicity of the CI reference state may differ from that of the SCF state.
   In the next lines the excitations are specified in (5I5) format. There are the following possibilities:

1	MO1	MO2			single excitation from MO1 into MO2
2	MO1	MO2	MO3	MO4	double excitation from MO1 into MO2 and from MO3 into MO4
111	MOA	MOB	MOC	MOD	all single excitations from occupied MOs MOA-MOB into unoccupied MOs MOC-MOD
112	MOA	MOB	MOC	MOD	all double excitations of type AA→CC from the unoccupied MOs MOA-MOB into unoccupied MOs MOC-MOD
122	MOA	MOB	MOC	MOD	all double excitations of type AA→CD from the unoccupied MOs MOA-MOB into unoccupied MOs MOC-MOD
212	MOA	MOB	MOC	MOD	all double excitations of type AB→CC from the unoccupied MOs MOA-MOB into unoccupied MOs MOC-MOD
222	MOA	MOB	MOC	MOD	all double excitations of type AB→CD from the unoccupied MOs MOA-MOB into unoccupied MOs MOC-MOD
0					End of CI input

4. **IVO corrections**

   For (→NIVO) = 1 one line has to be given including the total number of highest occupied MOs from which electrons are excited, and the total number of lowest unoccupied MOs whose orbital energies are corrected.

   For (→NIVO) = 2 the numbers of four MOs are required: the first two are the lowest and the highest occupied MO from which electrons are excited. The last two are the lowest and the highest unoccupied MO whose orbital energies are corrected.

5. **Electrostatic properties**

   For (→NPROP) = 1 or 2 the first line contains the total number of steps along the two basis vectors which generate the grid (format (2I5)).

   In the next line the x, y, and z coordinates (in Å) for the lower left corner of the grid have to be given in (3F10.5) format. The reference coordinates usually are the input coordinates except if (→STDO) was specified in **section II**.

   The next two lines contain in (3F10.5) format the x, y and z components of the two basis vectors. The step length along the grid is determined by the length of the vectors.

   For (→NPROP) = 3 the step width of the Marching Cube algorithm and the density value for the isodensity surface have to be given in (2F10.5) format. See (→SCHRIT and

GRENZW in **section II**).

6. **Cartesian geometry optimization for selected atoms**
   For (→CARTOPT with CARTSLCT) the first line contains in free format the total number of atoms to be optimized. The next lines contain in free format the input numbers of all atoms to be included in the optimization. If also (→CARTSDE2) was chosen, the first line contains additionally the number of atoms for which the second derivatives have to be explicitly calculated. These atom numbers must be specified at first in the list that follows.

7. **Explicit calculation of second derivatives for selected atoms**
   For (→CARTSDE2 without CARTSLCT) the total number of atoms for which the second derivatives have to be explicitly calculated (full Hessian) must be specified in free format. The first atoms in the list that follows are considered.

8. **Vibration analysis for selected atoms**
   For (→NVIB = 4) the first line contains the total number of atoms for which the vibration analysis will be performed (format (I5)). Exception: see (→FULL). In this analysis the coordinates of all other atoms are fixed, while the position of one atom is changed symmetrically in $\pm x, \pm y, \pm z$ directions. This procedure is repeated for all atoms specified here.
   If a zero is given for the total number, all atoms are taken and no specification is necessary (alternative to (→FULL) in **section II**).
   Otherwise the next lines contain in free format the internal numbers of the atoms to be included in the analysis. The numbering is identical to that of the internal input.

9. **Mass assignment for MD**
   If the keyword (→ASSIGN_MASS) was set in section II, a section with

   ```
   ASSIGN_MASS
   ...
   ATOM N mass (a.u.)
   ...
   END
   ```

   has to be given here.
   Each line has to start with the string "ATOM" followed by the number N of the atom according to the input order, and a mvalue for the mass in a.u.

2.4 Hints and comments

The disk units `7, 9, 10, 15, 19-21, 41, 42, 50-58, 60-63, 70-78` are used in addition to the standard I/O units.

FORTRAN unit `7` (filename **fort.7**) contains a compressed output.

FORTRAN unit `9` (filename **fort.9**) contains a new input with optimized coordinates after a geometry optimization.

FORTRAN unit `10` (filename **fort.10**) contains a copy of the original input.

FORTRAN unit `21` contains (if →NPROP > 0) output for the electrostatic potential or the electronic density, respectively.

The external file **<molecular formula>.molden** contains the molecular structure of the initial input in MOLDEN format.

The external file **<molecular formula>.xyz** contains the molecular structure, the molecular orbitals, geometry optimization, and SCF cycles, in a format suitable for **MOLDEN**.

The external file **<molecular formula>.dat** contains the molecular structure in a format suitable for **SCHAKAL**.

The external file **<molecular formula>.alphados** contains the α DOS in a format suitable for **gnuplot** (similar for β).

The built-in maximum values for the change in lengths, angles, and dihedral angles in each optimization step are:
0.1 Å, 5°, 10° for SCF runs;
0.05 Å, 3°, 5° for CI runs;

The following subroutines are machine dependent:

- F02ABF for matrix diagonalization (from the NAG library)
  called in DIAMAT
- F02FCF for matrix diagonalization (from the NAG library)
  called in DIAMAT
- DSYEVX for matrix diagonalization (from the LAPACK library)
  called in DIAMAT
- DSYEVD for matrix diagonalization (from the LAPACK library)
  called in DIAMAT
- DSYSV for solving systems of linear equations (from the LAPACK library)
  called in RHFDIIS, UHFDIIS, and ROHFDIIS.
- DGEMM,DSYMM,DSYRK,DROT,DCOPY,DAXPY,DSCAL,DSPMV for matrix and vector operations (from the BLAS library)

For the last three subroutines the MSINDO source includes a number of machine dependent files.
The source file **rlapack.F** contains standard codes of all LAPACK and BLAS routines called by MSINDO.
If optimized libraries BLAS or LAPACK are not available, the corresponding source files have to be renamed to **rlapack.f** and the procedure CREX has to be restarted. Similarly, if no DERF or DERFC subroutines are available, the file **derf.F** has to be renamed **derf.f** and the procedure CREX has to be restarted.

The file **include/delimiter.h** contains the most important global field delimiters:

parameter	meaning	dependencies
MV:	Maximum number of atoms + dummy atoms	
MW:	Maximum number of atomic orbitals	
NX:	Maximum number of determinants for CI	
MX:	Maximum number of valence electrons	
MY:	Maximum number of internal coordinates	MY = 3*MV
MZ:	Maximum number of coupled internal coordinates	MZ = MV*(MY+1)/2
AOMAX:	Maximum number of atomic orbitals on one atom	
MAXPKT:	Maximum number of point charges for reaction field calculations	
MMV:	Maximum number of pseudo atoms for embedding	
NTR:	Maximum number of translations for embedding	

External subroutines:

CPU_TIME	Fortran90
DATE_AND_TIME	Fortran90
DAXPY	BLAS level 1
DCOPY	BLAS level 1
DDOT	BLAS level 1
DGEMM	BLAS level 3
DGEMV	BLAS level 2
DGETRF	BLAS level 1
DGETRI	BLAS level 1
DLAMCH	LAPACK
DNRM2	BLAS level 1
DPOTRF	BLAS level 1
DPOTRI	BLAS level 1
DROT	BLAS level 1
DSCAL	BLAS level 1
DSYEV	LAPACK
DSYEVD	LAPACK
DSYEVR	LAPACK
DSYEVX	LAPACK
DSYMM	BLAS level 3
DSYMV	BLAS level 2
DSYRK	BLAS level 3
DSYSV	LAPACK
FLUSH	Fortran90
VDERF	mkl 10.0 and higher
VDERFC	mkl 10.0 and higher

Bibliography

[1] T. Bredow and K. Jug, Theoret. Chem. Acc. **113**, 1 (2005)

[2] B. Ahlswede, K. Jug, J. Comput. Chem. **20**, 563, 572 (1999)

[3] K. Jug, G. Geudtner, T. Homann, J. Comput. Chem. **21**, 974 (2000)

[4] T. Bredow, G. Geudtner, K. Jug, J. Comput. Chem. **22**, 861 (2001)

[5] D.N. Nanda, K. Jug, Theoret. Chim. Acta **57**, 95 (1980)

[6] K. Jug, R. Iffert, J. Schulz, Int. J. Quantum Chem. **32**, 265 (1987)

[7] J. Li, P. Correa de Mello, K. Jug, J. Comput. Chem. **13**, 85 (1992)

[8] K. Jug, T. Bredow, *Methods and Techniques in Computational Chemistry*, METECC-95, Ed. E. Clementi, G. Corongiu, STEF, Cagliari 1995, p. 89

[9] J.A. Pople, D.L. Beveridge, *Approximate Molecular Orbital Theory*, McGraw-Hill Book Company, 1970

[10] C.C.J. Roothaan, Rev. Mod. Phys **32**, 179 (1960)

[11] T. Bredow, K. Jug, MSINDO, in P. v. Ragué Schleyer, H.F. Schaefer III, P.R. Schreiner, W.L. Jorgensen, W. Thiel, R.C. Glen, editors, *Encyclopedia of Computational Chemistry (online edition)*, John Wiley & Sons, Ltd., Chichester, UK (2004).

[12] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, J. Chem. Phys. VOLUME **132**, 154104 (2010)

[13] S. Grimme, J. Comput. Chem. **27**, 1787–1799 (2006)

[14] S. Grimme, J. Chem. Phys. **118**, 9095 (2003)

[15] I. Gadaczek, K. Krause, K.-J. Hintze, and T. Bredow, J. Chem. Theory Comput., **7**, 3675–3678 (2011)

[16] I. Gadaczek, K.-J. Hintze, and T. Bredow, Phys. Chem. Chem. Phys., **14**, 741–750 (2012)

[17] I. Gadaczek, K. Krause, K.-J. Hintze, and T. Bredow, Chem. Theory Comput. **8**, 986–996

(2012)

[18] R. Iffert, K. Jug, Theor. Chim. Acta **72**, 373 (1987)

[19] J. Pipek, P.G. Mezey, J. Chem. Phys. **90**, 4916 (1989)

[20] M. Page, J.W. McIver, Jr., J. Chem. Phys. **79**, 4985 (1983)

[21] K. Jug, M. Krack, Int. J. Quantum Chem. **44**, 517 (1992)

[22] K. Jug, G. Geudtner, J. Comput. Chem. **14**, 639 (1993)

[23] G. Pongor, Chem. Phys. Lett. **24**, 603 (1974)

[24] A.D. Bacon, M.C. Zerner, Theor. Chim. Acta **53**, 21 (1979)

[25] P. Pulay, J. Comput. Chem. **3**, 556 (1982)

[26] J.J.P. Stewart, P. Császár, P. Pulay, J. Comput. Chem. **3**, 227 (1982)

[27] H.B. Schlegel, *Some Practical Suggestions for Optimizing Geometries and Locating Transition States*, in J. Bertrán, I.G. Csizmadia (Eds.), New Theoretical Concepts for Understanding Organic Reactions, NATO ASI Series C, 267 Kluwer, Dordrecht (1989)

[28] P. Császár, P. Pulay, J. Mol. Struct. **114**, 31 (1984)

[29] V. Farkas, H. B. Schlegel, Phys. Chem. Chem. Phys. **4**, 11 (2001)

[30] K. N. Kudin, G. E. Scuseria, E. Cances, J. Chem. Phys. **116**, 8255 (2002).

[31] A.M. Köster, C. Kölle, K. Jug, J. Chem. Phys. **99**, 1224 (1993)
M. Krack, A.M. Köster, K. Jug, J. Comp. Chem. **18**, 301 (1997)

[32] S. Huzinaga, C. Arnau, Phys. Rev. **54**, 1948 (1971)

[33] M.S. Gopinathan, K. Jug, Theoret. Chim. Acta **63**, 497 (1983)

[34] F. Janetzko, T. Bredow, K. Jug, J. Chem. Phys., **116**, 8994 (2002)

[35] http://www.gnuplot.info/

[36] http://cheminf.cmbi.ru.nl/molden/

[37] http://www.krist.uni-freiburg.de/ki/Mitarbeiter/Keller/info.html

[38] K. Jug. A. Poredda, Chem. Phys. Lett. **171**, 394 (1990)

[39] S. Miertus, E. Scrocco, J. Tomasi, Chem. Phys. **55**, 117 (1981)

[40] C. Kölle, K. Jug, J. Comput. Chem., **18**, 1 (1997)

[41] CRC Handbook of Chemistry and Physics, 72nd Edition, CRC Press Boca Raton Ann Arbor Boston (1992)

[42] W.E. Lorensen, H.E. Cline, Comp. Graph. **21**, 163 (1987)

[43] W. Heiden, T. Goetze, J. Brickmann, J. Comput. Chem. **14**, 246 (1993)

[44] E.L. Coitino, J. Tomasi, R. Cammi, J. Comput. Chem. **16**, 20 (1995)

[45] T. Bredow, G. Geudtner, K. Jug, J. Chem. Phys. **105**, 6395 (1996)

[46] T. Bredow, G. Geudtner, K. Jug, J. Comput. Chem. **22**, 89 (2001)

[47] R. A. Evarestov, *Quantum chemistry of solids: The LCAO first principles treatment of crystals*, Springer, Berlin Heidelberg (2007).

[48] T. Bredow, R. A. Evarestov, K. Jug, Phys. Stat. Sol. B **222**, 495 (2000).

[49] R. A. Evarestov, A. V. Bandura, I. I. Tupitsyn, Theoret. Chem. Acc. **137**, 14 (2018).

[50] M. Iannuzzi, A. Laio, M. Parrinello, Phys. Rev. Lett. **90**, 238302 (2003)

[51] H. Jønsson, G. Mills, and K. W. Jacobsen, in B. J. Berne, G. Ciccotti, and D. F. Coker (Eds.), *Classical and Quantum Dynamics in Condensed Phase Simulations*, p. 385, World Scientific, Singapore (1998)

[52] D. R. Alfonso, K. D. Jordan, J. Comput. Chem. **24**, 990 (2003)

[53] G. Henkelman, B. Uberuaga, H. Jønsson, J. Chem. Phys. **113**, 9978 (2000)

[54] A. Klamt, G. Schüürmann, J. Chem. Soc. Perkin Trans. II, 799 (1993)

[55] J. L. Pascual-Ahuir, E. Silla, J. Comput. Chem. **11**, 1047 (1990)

[56] Silla et. al J. Comput. Chem. **12**, 1077 (1991)