SUPPLEMENTARY INFORMATION
Modern Quantum Chemistry with [Open]Molcas

Francesco Aquilante,[1] Jochen Autschbach,[2] Alberto Baiardi,[3] Stefano Battaglia,[4] Veniamin A. Borin,[5] Liviu F. Chibotaru,[6] Irene Conti,[7] Luca De Vico,[8] Mickaël Delcey,[9] Ignacio Fdez. Galván,[4] Nicolas Ferré,[10] Leon Freitag,[3] Marco Garavelli,[7] Xuejun Gong,[11] Stefan Knecht,[3] Ernst D. Larsson,[12] Roland Lindh,[4] Marcus Lundberg,[9] Per Åke Malmqvist,[12] Artur Nenov,[13] Jesper Norell,[14] Michael Odelius,[14] Massimo Olivucci,[8,15] Thomas B. Pedersen,[16] Laura Pedraza-González,[8] Quan M. Phung,[17] Kristine Pierloot,[6] Markus Reiher,[3] Igor Schapiro,[5] Javier Segarra-Martí,[18] Francesco Segatta,[7] Luis Seijo,[19] Saumik Sen,[5] Dumitru-Claudiu Sergentu,[2] Christopher J. Stein,[3] Liviu Ungur,[11] Morgane Vacher,[20] Alessio Valentini,[21] and Valera Veryazov[12]

[1)] *Theory and Simulation of Materials (THEOS) and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*
[2)] *Department of Chemistry, University at Buffalo, Buffalo, NY 14260-3000, USA*
[3)] *ETH Zurich, Lab. f. Phys. Chem., Vladimir-Prelog-Weg 2, 8093 Zurich, Switzerland*
[4)] *Department of Chemistry – BMC, Uppsala University, P.O. Box 576, SE-751 23 Uppsala, Sweden*
[5)] *Fritz Haber Center for Molecular Dynamics Research, Institute of Chemistry, The Hebrew University of Jerusalem, Jerusalem 9190401, Israel*
[6)] *Department of Chemistry, KU Leuven, Celestijnenlaan 200F, 3001 Leuven, Belgium*
[7)] *Dipartimento di Chimica Industriale "Toso Montanari", Università di Bologna, Viale del Risorgimento 4, Bologna, I-40136 Italy*
[8)] *Dipartimento di Biotecnologie, Chimica e Farmacia, Università degli Studi di Siena, via Aldo Moro 2, 53100, Siena, Italy*
[9)] *Department of Chemistry – Ångström Laboratory, Uppsala University, SE-751 21 Uppsala, Sweden*
[10)] *Aix-Marseille Univ, CNRS, Institut Chimie Radicalaire, Marseille, France*
[11)] *Department of Chemistry, University of Singapore, 3 Science Drive 3, 117543 Singapore*
[12)] *Division of Theoretical Chemistry, Lund University, POB 124, Lund 22100, Sweden*
[13)] *Dipartimento di Chimica Industriale Toso Montanari, Universitá di Bologna, Viale del Risorgimento 4, 40136 Bologna, Italy*
[14)] *Department of Physics, AlbaNova University Center, Stockholm University, SE-106 91 Stockholm, Sweden*
[15)] *Department of Chemistry, Bowling Green State University, Bowling Green, Ohio 43403, United States of America*
[16)] *Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, University of Oslo, P.O. Box 1033 Blindern, N-0315 Oslo, Norway*
[17)] *Institute of Transformative Bio-Molecules (WPI-ITbM), Nagoya University, Chikusa, Nagoya 464-8602, Japan*
[18)] *Department of Chemistry, Molecular Sciences Research Hub, Imperial College London, White City Campus, 80 Wood Lane, London W12 0BZ, UK*
[19)] *Departamento de Química, Instituto Universitario de Ciencia de Materiales Nicolás Cabrera, and Condensed Matter Physics Center (IFIMAC), Universidad Autónoma de Madrid, 28049 Madrid, Spain*
[20)] *Laboratoire CEISAM - UMR CNRS 6230, Université de Nantes, 44300 Nantes, France*
[21)] *Theoretical Physical Chemistry, Research Unit MolSys, Université de Liège, Allée du 6 Août, 11, 4000 Liège, Belgium*

The supplementary information contains several sections and illustrates the short description of features, described in the main text.

This section contains ready to use examples.

- S1. Overview of main computational codes
  - Short description of computational modules
  - Flowchart
  - Description of Molcas input language (EMIL)
- S2. Frenkel excitonic Hamiltonian
- S3. BCHL excitation energies
- S4. X-ray absorption calculations
- S5. ARM protocol
- S6. Examples of DMRG calculations

- S7. Tools to select active space
- S8. Excited states of lanthanides in solids
- S9. Example of magnetic properties calculation
- S10. Molecular dynamics calculations
- S11. RASSCF/RASPT2/RASSI methodology for heavy elements
- S12. SINGLE_ANISO calculations
- S13. Example of applying the semi-*ab initio* approach: fragmentation + POLY_ANISO

S1. OVERVIEW OF MAIN COMPUTATIONAL CODES IN MOLCAS

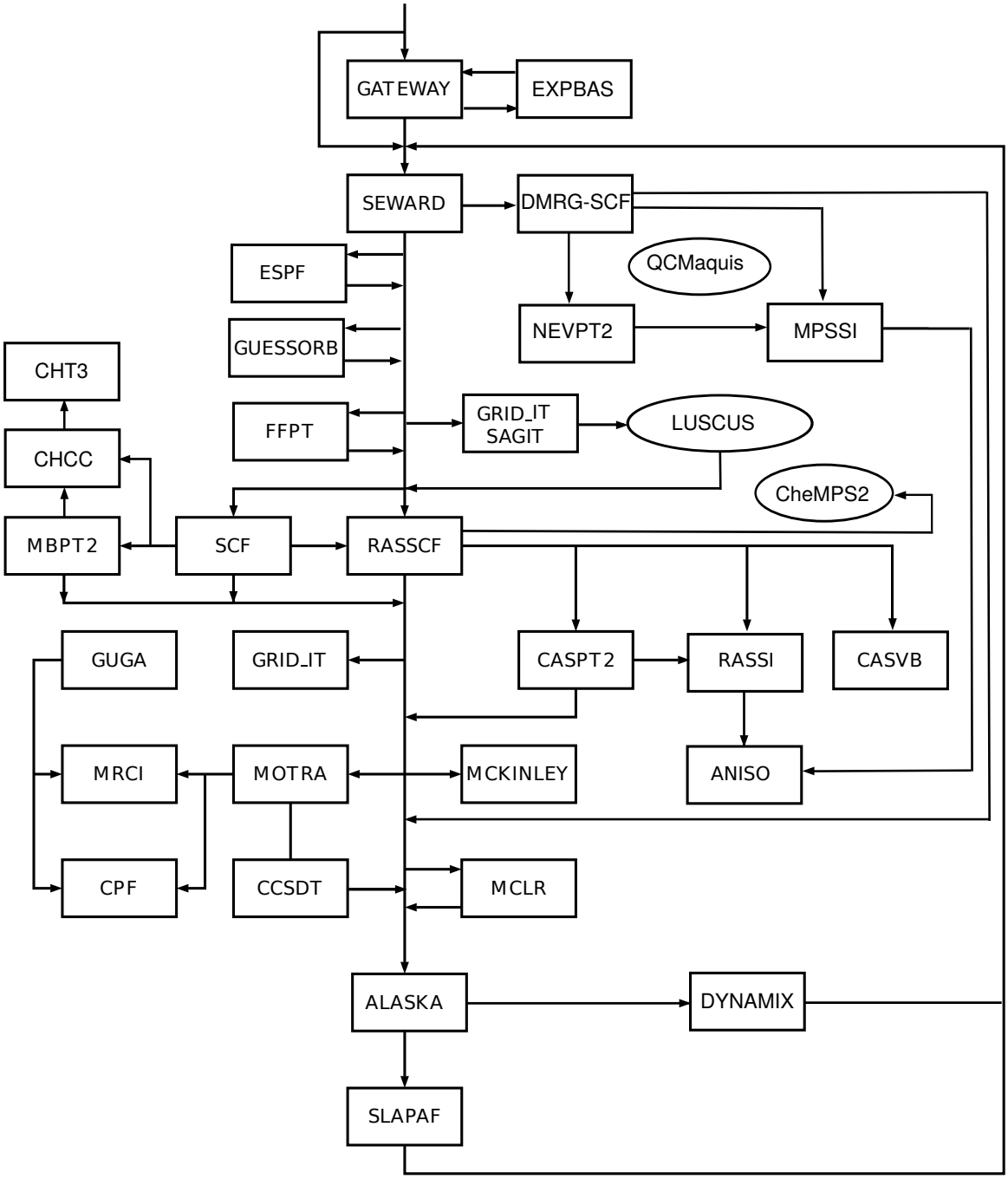

FIG. S1. Flowchart of Molcas modules

Compilation of Molcas requires Fortran and C compiler. The recent version of the code is only supports 64-bit arithmetic. The regular testing of the code is performed at Linux and Mac OS, however, the Molcas code can be compiled virtually under any operating system, including MS Windows (using Cygwin or MinGW), AIX, FreeBSD. Very little modifications are needed to run Molcas at a new environment. Molcas can be compiled with the following Fortran compilers:

- gfortran (and clones: g95, llvm)
- Intel ifort
- Sun Sunstudio
- NVIDIA Portland compiler
- NAG fortran compiler

A short description of the main computational codes in [Open]Molcas (in alphabetic order):

- &ALASKA

- *Description:* ALASKA computes energy gradients for supported wavefunction models.
- *Capabilities:* Derivative integrals are computed on the fly and contracted with density matrices to produce the gradient. Analytical gradients for variational methods (SCF, CASSCF, RASSCF) are immediate. Some non-variational methods have analytical gradients available too (MP2, SA-CASSCF). For other methods the program resorts to numerical differentiation (SA-RASSCF, CASPT2). Analytical gradients are also available with density fitting. Analytical nonadiabatic couplings are available for SA-CASSCF.
- *Limitations:* Gradients are currently not available for spin–orbit coupled states.
- *Performance:* Both analytical and numerical differentiation modes are parallelized.
- *Computational bottlenecks:* For large SA-CASSCF the bottleneck lies often in solving the linear-response equations (MCLR).
- *References:*[1–3]

• &CASPT2

- *Description:* CASPT2 computes dynamic correlation on top of multiconfigurational wave functions.
- *Capabilities:* Can handle one or several CASSCF or RASSCF reference functions.
- *Limitations:* As for the RASSCF program.
- *Performance:* 2.3 million CSFs, 406 basis functions, 2 states: 70 minutes
- *Computational bottlenecks:* Code is parallel. Code depends on DGEMM and Eigenvectors solvers
- *References:*[4,5]

Open Molcas–CheMPS2 interface

- *Description:* [Open]Molcas–CheMPS2 interface is an interface between [Open]Molcas (&RASSCF and &CASPT2) and the CheMPS2 library.
- *Capabilities:* It computes DMRG-CASSCF and DMRG-CASPT2.
- *Limitations:* In DMRG-CASSCF, up to CAS(40,40) for general active spaces and CAS(100,100) and for one-dimensional active spaces. In DMRG-CASPT2, up to CAS(30,30).
- *Performance:* The performance of calculations primarily depends on the performance of the CheMPS2 library. See CheMPS2 references.
- *Computational bottlenecks:* In DMRG-CASPT2 calculations, the calculation of F.4-RDM is the bottleneck.
- *References:*[6–9]

• &DMRGSCF

- *Description:* DMRGSCF activates a program to perform density matrix renormalization group self-consistent-field (DMRG-SCF) calculations
- *Capabilities:* It acts as a front-end to DMRG codes (currently, only for QCMaquis) to perform DMRG-CI and DMRG-SCF calculations.
- *Limitations:* DMRG-CI can be performed with up to 70 orbitals if enough memory is present (1 TB would be required)
- *Performance:* 20 orbitals, $m = 1000$, 5 sweeps – 1-4 hours.
- *Computational bottlenecks:* Code is parallelized with OpenMP. Bottleneck – Jacobi-Davidson diagonalization and, for very large active spaces, saving of the matrix-product-state (MPS) and of the boundaries in memory (for 60-70 orbitals, 1 TB RAM might be needed).
- *References:*[10,11]

• &DYNAMIX

- *Description:* DYNAMIX is a module to perform molecular dynamics (MD) simulations using the velocity Verlet algorithm. It requires atomic coordinates and velocities as input.
- *Capabilities:* The nonadiabatic transitions between different electronic states can be taken into account using a deterministic algorithm or a Tully surface hopping algorithm.
- *Limitations:* MD simulations are available with any methods for which gradients (numerical or analytical) are available.
- *Performance:* It depends on the choice of the electronic structure method
- *Computational bottlenecks:* The cost of single MD step is determined by a single-point gradient calculation.
- *References:*[12,13]

• &MPSSI

- *Description:* MPSSI is a module to perform state-interaction (SI) calculations for orthogonal and non-orthogonal DMRG input wavefunctions of arbitrary spin symmetry (obtained with QCMaquis).

- *Capabilities:* It can calculate one- and two-particle transition density matrix elements as well as spin-(transition-)density matrix elements for orthogonal and non-orthogonal DMRG input wavefunctions. In the latter case, a transformation of the MPS wavefunction to biorthonormal basis is carried out, hence the name MPSSI. MPSSI supports the calculation of effective Hamiltonians as well as the computation of one- and two-electron properties including spin-orbit (SO) matrix elements. The resulting SO-coupled wavefunctions can, for example, serve as input wave functions for the SINGLE_ANISO and POLY_ANISO module, respectively.
- *Limitations:* Currently, only SI calculations within C1 point group symmetry orbitals are possible. The CAS spaces for the input wave functions have to be of equal dimension (in terms of # electrons and # of spatial orbitals).
- *Computational bottlenecks:* The transformation of the MPS to biorthonormal basis requires twice the evaluation of *exact* matrix-product operator (MPO) $\times$ MPS products for each spatial orbital. The bond dimension m of the resulting MPS is the product of the bond dimension of the MPO (m_{MPO}) and that of the MPS (m_{MPS}). Hence, at each step, an additional compression of the product MPS is required which scales approximately with m^3_{MPO}.
- *References:*[14]

- &NEVPT2
  - *Description:* NEVPT2 is a module to perform n-electron valence state second-order perturbation theory (NEVPT2) calculations
  - *Capabilities:* It can perform NEVPT2 calculations based on a DMRG reference wavefunction obtained with QCMaquis. Quasi-degenerate NEVPT2 (QD-NEVPT2) calculations are also supported. Contains a module for a separate massively parallel evaluation of the four-particle density matrix (4-RDM).
  - *Limitations:* Currently, on a single node calculations with up to 12 active orbitals and in massively parallel 4-RDM evaluation mode, calculations with up to 22 orbitals are possible.
  - *Computational bottlenecks:* The 4-RDM is evaluated exactly, which scales as L^8 with L as the number of active orbitals.
  - *References:*[15–19]

- &POLY_ANISO
  - *Description:* POLY_ANISO allows to use the semi-*ab initio* approach for the simulation of the magnetism of poly-nuclear compounds using *ab initio* results of single-site fragment calculations (i.e. RASSCF/-CASPT2/RASSI/SINGLE_ANISO).
  - *Capabilities:* Simulation of coupled spectra, parameters of g-tensor, D tensor, magnetic exchange. Thermodynamic magnetic properties: magnetic susceptibility, molar magnetization, magnetization torque, Zeeman splitting.
  - *Limitations:* Typical calculations are routinely done up to 5000 coupled functions. Many large sized arrays allocated inside the code depend on this number in a quadratic way.
  - *Performance:* The required memory is allocated dynamically and depends (mostly) on the number of the coupled states included in the calculation. The computational time is similar to the SINGLE_ANISO code.
  - *Computational bottlenecks:* Computation of the powder magnetization is the slowest. The user is advised to adjust the averaging grid and field points by MAVE and HINT keys, or to invoke the computation of powder magnetization only when it is necessary. Time is spent mostly in ZHPEV and ZGEMM operations.
  - *References:*[20,21]

- &RASSCF
  - *Description:* RASSCF is a multiconfigurational Self-Consistent Field program.
  - *Capabilities:* It computes CASSCF, RASSCF, GASSCF and CASVB wavefunctions. The code can be interfaced to DMRG codes (QCMaquis and ChemPS2).
  - *Limitations:* for CASSCF wavefunction - 16 electrons at 16 orbitals, for RASSCF up to 40 correlated orbitals in RAS1/RAS3 (up to double excitations) and 8-12 orbitals in RAS2 space
  - *Performance:* 2.2 million CSFs, 90 basis functions: 31 minutes using 10 cores.
  - *Computational bottlenecks:* Code is parallel. Code depends on DGEMM operations
  - *References:*[22,23]

- &RASSI
  - *Description:* RASSI computes matrix elements, expectation and transition values of various operators over multiconfigurational wave functions.

  - *Capabilities:* Wave functions can be of CASSCF or RASSCF type and do not need to be orthonormal. Energies and properties can be computed over orthonormal, non-interacting linear combinations of the input wave functions. Hamiltonian operators can be complemented with spin-orbit operators of AMFI type from SEWARD, yielding energies and properties for individual spin states. Dyson amplitudes can also be obtained.
  - *Limitations:* Similar as for RASSCF program.
  - *Performance:* 2.2 million CSFs, 90 basis functions, 3x3 states: 11 hours
  - *Computational bottlenecks:* Code is parallel. Code depends on DGEMM operations
  - *References:*[24,25]

- &SEWARD
  - *Description:* SEWARD computes one- and two-electron integrals
  - *Capabilities:* Many type of one-electron integrals are supported. Most integrals are computed analytically. Auxiliary basis sets (density fitting, RI) and Cholesky decomposition are available for two-electron integrals. Arbitrary angular momentum functions are supported (up to $l = 15$ by default). The code is optimized for general contraction.
  - *Limitations:* The generated integrals have to be stored on disk, although an integral-direct implementation is available for SCF. Only the electron repulsion operator is implemented for two-electron integrals.
  - *Performance:* Two-electron integral generation is parallelized. Timing for 292 basis functions (772 primitives), with RICD: 27 seconds on single core. Timing for 1076 basis functions (2702 primitives), with RICD: 7.2 hours on a single core.
  - *Computational bottlenecks:* If not using density fitting or Cholesky decomposition, writing to disk the two-electron integrals can be a significant slowdown.
  - *References:*[26–28]

- &SINGLE_ANISO
  - *Description:* SINGLE_ANISO computes magnetic properties arising from one magnetic center using the results of a successful CASSCF(CASPT2)RASSI spin-orbit calculation
  - *Capabilities:* Computation of the parameters of pseudospin Hamiltonians: g and D-tensors for any dimension of the pseudospin, crystal field parameters for lanthanides and transition metal compounds, temperature dependence of the magnetic susceptibility, temperature and field dependent molar magnetization, etc. on the basis of a successful CASSCF(CASPT2)RASSI spin-orbit calculation
  - *Limitations:* There are no real limitations.
  - *Performance:* The code does not use much resources. The required memory is allocated dynamically and depends (mostly) on the number of the spin-orbit states included in the RASSI calculation. As example, the SINGLE_ANISO step of a calculation on Tb^{3+} free ion CAS=(8 ele. in 7 orb.) involving 1225 spin free roots (leading to 3003 spin-orbit states after spin-orbit interaction is considered in RASSI), takes about 30 minutes on a single core for the computation of g-tensor, parameters of the crystal field, zero-field magnetic susceptibility and field-dependent magnetic susceptibility at 301 temperature points.
  - *Computational bottlenecks:* Most of the functions are very fast, usually done in seconds. Computation of the powder magnetization is the slowest step since it needs averaging over directions and field strength points. The user can adjust the averaging grid and field points by MAVE and HINT keys. Time is spent mostly in ZHPEV and ZGEMM operations.
  - *References:*[29,30]

- &SLAPAF
  - *Description:* SLAPAF is the geometry optimization driver.
  - *Capabilities:* It can optimize minima or saddle points (transition states) on a potential energy surface. The optimization can be subject to symmetry or arbitrary constraints. Crossing points between electronic states of different or the same spin can also be located. Intrinsic reaction coordinate and minimum energy paths can be computed automatically as a series or constrained optimizations.
  - *Limitations:* Gradients (analytical or numerical) are required from the electronic structure method. Collective optimization such as the nudged elastic band method are not yet supported.
  - *Performance:* Usually not more than a few seconds.
  - *Computational bottlenecks:* The electronic structure calculation (energy and gradient) is the bottleneck.
  - *References:*[3,31–35]

S1.A. EMIL language

Input standards for any quantum chemical program is a complicated matter. A program should get information in different formats, such as coordinates of atoms, basis set data, initial starting point, thresholds, details of calculations, etc. The order of calculations is also might vary dependently on the computational task.

Since 2000 Molcas uses a specially designed input language, called EMIL (Enhanced Molcas Input Language), which has elements of a simple computer language. In fact, an input written on EMIL can be considered as a simple prototype for shell programming, oriented towards solving quantum chemical problems. The special efforts in EMIL are related to a parallel execution of Molcas.

The interpretation of EMIL commands can be performed by several codes, including moclas.plx and pymolcas.

Here is the list of the elements of EMIL:

- *&MODULE* : a command to run module
- >> : EMIL command (similar to a UNIX command)
- KEY or KEY=VALUE : an input for current module

An input is preprocessed before the execution: any variable started from a $ sign is replaced to a value; any variable started from a @ sign is replaced by a set of commands, defined in alias-like library; several signs (;, =) can be used as a separators. That allows to create very compact inputs. User can comment parts of input, by using C and/or C++ style comments.

```
 &SCF        /* run SCF code */
   CHARGE    // keyword for SCF
   1         // value for keyword CHARGE
 &GRID_IT    /* run GRID_IT code */
   ALL       // keyword for GRID_IT

/* Shorter form */
 &SCF; CHARGE=1; &GRID_IT

/* use predefined custom aliaces */

  @SCF1; @PLOT
```

EMIL commands also can be used to combine several files together, as it shown in the input below:

```
// Insert a file include.src into input
>>INCLUDE include.inp
// Extract a part of the current file into a
    separate file
>>FILE extra.inp
/* the content of
   extra.inp */
>>EOF
```

An important concept of EMIL language, which makes many UNIX commands looks similar, but operates differently is related to the directories, used in the calculation. The directory named CurrDir is the current directory, where the input file is located. By default, the same directory will be used to collect the important files after the calculation (the behavior can be overwritten by OutputDir). Another directory, called WorkDir, is used to store all temporary files. In parallel execution of Molcas, WorkDir is a common name for all scratch areas, which can reside at different computers. Thus, for instance, a copy command, executed in parallel environment, is executed for each WorkDir. Since all EMIL command are executed from WorkDir, there is no need to specify it.

For most common situations, EMIL provides own set of commands, listed below:

```
General Unix-like commands

SHELL : execute a command in serial
EXEC  : execute a command in parallel
ECHO  : similar to ECHO command in DOS (with value
    ON and OFF)
EXPORT: export environment and propagate it to
    slaves
EVAL  : evaluate a value

Operations with files

COPY  : similar to cp command in serial
LINK  : similar to ln command in serial
CLONE : distribute a file to all slaves' scratch
    directories
COLLECT: collect a file from all slaves' scratch
    directories
RM    : delete a file

Control over the execution

DO WHILE: start a loop, finished by special
    condition, e.g. convergence
DO GEO : start a loop for geometry optimization
    with user's defined internal coordinates
FOREACH : start a loop with predefined set of values
IF      : conditional execution, based on return
    code, existence of a file, or a value

GOTO    : jump operator to LABEL
ENDIF   : IF statement terminator
ENDDO   : Loop terminator
EXIT    : exit
```

Finally, some examples for simple input scenario:

```
/* DFT calculation of water molecule */

 &GATEWAY
    Coord = Water.xyz
    Basis = STO-3G
 &SEWARD
 &SCF
    KSDFT = B3LYP
```

```
/* UHF calculation of H2 dissociation */

>> EXPORT R=0.5
>> FOREACH STEP (1..100)
>> EVAL DIST=$R+$STEP/10
  &GATEWAY
    Coord
     2
       H2 molecule
     H 0 0 0
     H $DIST 0 0
    BASIS=ANO-S-VDZP
    GROUP=C1
   &SEWARD
   &SCF
    UHF
    SCRAMBLE = 0.1
>> ENDDO <<
```

S2. FRENKEL EXCITONIC HAMILTONIAN

The general expression of the Frenkel Excitonic Hamiltonian (comprising sites and site-site coupling terms) can be connected to the microscopic Hamiltonian of a system of interacting molecules. In the Born-Oppenheimer approximation, the Hamiltonian of a two-chromophore system composed by molecules having respectively $\mathcal{N}_I$ fixed nuclei (described by the positions R and charges Z) and n_I electrons (positions r, momenta p, and charge of -1), with $I = A, B$, is given by:

$$\begin{aligned}\hat{H} = &\left[\sum_{i}^{n_1}\frac{p_i^2}{2} - \sum_{i\bar{i}}\frac{Z_{\bar{i}}}{|\boldsymbol{r}_i - \boldsymbol{R}_{\bar{i}}|} + \sum_{i,\bar{i}>i}\frac{1}{r_{i\bar{i}}}\right]_{:=H_A} \\ &+ \left[\sum_{j}^{n_2}\frac{p_j^2}{2} - \sum_{j\bar{j}}\frac{Z_{\bar{j}}}{|\boldsymbol{r}_j - \boldsymbol{R}_{\bar{j}}|} + \sum_{j,\bar{j}>j}\frac{1}{r_{j\bar{j}}}\right]_{:=H_B} \\ &+ \left[\sum_{ij}\frac{1}{r_{ij}} - \sum_{ij}\frac{Z_i}{|\boldsymbol{R}_i - \boldsymbol{r}_j|} - \sum_{ij}\frac{Z_j}{|\boldsymbol{r}_i - \boldsymbol{R}_j|} + \sum_{ij}\frac{Z_i Z_j}{R_{ij}}\right]_{:=V_{AB}} \end{aligned} \quad (1)$$

where he indexes i and $\bar{i}$ (j and $\bar{j}$) run over either electrons or nuclei of molecule A (B). The first two terms (labeled as H_A and H_B, respectively) are responsible for the quantum mechanical properties of the isolated sites, while the last term (V_{AB}) is the coupling term between sites, responsible of the mixing of the local site properties. This is the sum of different contributions, namely, the electron-electron part, the nuclear-electron part and the nuclear-nuclear part.

We are interested in the evaluation of the terms $\langle a_A b_B | V_{AB} | a'_A b'_B \rangle$, where a, a' run over the states of the monomer A, and b, b' run over the states of the monomer B (and $|a_A b_B\rangle$ is a dimer state with molecule A in state a and molecule B in state b). The various components of V_{AB} in eq. (1) can be rewritten[36] as follows

$$\begin{aligned}\langle a_A b_B | V_{AB} | a'_A b'_B \rangle = &\int d\boldsymbol{r} d\boldsymbol{r}' \rho^A_{aa'}(\boldsymbol{r}) \frac{1}{|\boldsymbol{r} - \boldsymbol{r}'|} \rho^B_{bb'}(\boldsymbol{r}') \\ &- \delta_{aa'} \int d\boldsymbol{r}' \sum_{i\in A} \frac{\rho^B_{bb'}(\boldsymbol{r}') Z_i}{|\boldsymbol{R}_i - \boldsymbol{r}'|} \\ &- \delta_{bb'} \int d\boldsymbol{r} \sum_{j\in B} \frac{\rho^A_{aa'}(\boldsymbol{r}) Z_j}{|\boldsymbol{r} - \boldsymbol{R}_j|} \\ &+ \delta_{aa'}\delta_{bb'} \sum_{ij} \frac{Z_i Z_j}{R_{ij}} \end{aligned} \quad (2)$$

where ρ^A_{aa} is the electronic charge density of the electronic state a of molecule A, and $\rho^A_{aa'}(\boldsymbol{r})$, with $a \neq a'$, is the so called transition charge density related to the transition between the states a and a' of molecule A, and similarly for B. These (transition) densities can be expressed in terms of the atomic orbitals $\chi_\mu(\boldsymbol{r})$ and the (transition) density matrix between states a and a' expressed in the atomic orbital basis, e.g. $D^{aa'}_{\mu\nu}$, so that:

$$\rho^A_{aa'}(\boldsymbol{r}) = \sum_{\mu\nu} D^{aa'}_{\mu\nu} \chi_\mu(\boldsymbol{r}) \chi_\nu(\boldsymbol{r}) \quad (3)$$

In many cases of interest, the most important contribution is the one that involves transition densities on both sites ($a \neq a'$ and $b \neq b'$). This term for example enters to describe excitation energy transfer processes in which an excited molecule returns to the ground state while exciting a nearby molecule (e.g., the Förster energy transfer mechanism).

The implementation in Molcas of Eq. (2) uses extensively the Cholesky infrastructure for the first term of the equation, removing completely the computational bottleneck of the excitonic coupling calculation. The following tensor intermediates (W) involving the Cholesky vectors ($\mathcal{L}$) are computed and stored in a manner transparent to the user:

$$\vec{W}^{aa'} = \sum_{\mu\nu} \vec{\mathcal{L}}_{\mu\nu} D^{aa'}_{\mu\nu} \ , \quad (4)$$

where each vector $\vec{W}$ has the same length as the length of the Cholesky vectors.

S2.A. Input example: dimer of acrolein

Note the use of the keywords MONA and MONB for the separate runs of &RASSI, as well as the actual couplings calculation keyword EXCI in the second &RASSI input. The use of the keyword GHOST in &CASPT2 is highly recommended to avoid overhead in the PT2 calculation due to the use of ghost atoms.

```
***** SYSTEM A ******

 &GATEWAY
```

```
coord=acrolein-A.xyz
coord=acrolein-B.xyz
bsse=2
Basis=ANO-RCC-VDZP
NoMove
Group=Nosym
ricd

 &SEWARD

>>> COPY $Project.RunFile AUXRFIL

***** SYSTEM B ******

 &GATEWAY
coord=acrolein-A.xyz
coord=acrolein-B.xyz
bsse=1
Basis=ANO-RCC-VDZP
NoMove
Group=Nosym
ricd

 &SEWARD
fake ricd

 &RASSCF
nactel = 6 0  0
inactive = 12
ras2 = 5
ciroot = 5 5  1

 &CASPT2
multistate = 5 1 2  3  4  5
GHOSTorbitals = 0.1

 &RASSI
Nr  of  Job = 1 5;  1  2  3  4  5
TRD1
MONB

>>> COPY $Project.RunFile AUXRFIL


***** SYSTEM A ******

 &GATEWAY
coord=acrolein-A.xyz
coord=acrolein-B.xyz
bsse=2
Basis=ANO-RCC-VDZP
NoMove
Group=Nosym
ricd

 &SEWARD
fake ricd

 &RASSCF
nactel = 6 0  0
inactive = 12
ras2 = 5
ciroot = 4 4  1
fileorb = $Project.GssOrb

 &CASPT2
multistate = 4 1 2  3  4
GHOSTorbitals = 0.1

 &RASSI
Nr  of  Job = 1 4;  1  2  3  4
TRD1
MONA
EXCItonic couplings
```

S3. HOW TO COMPUTE BCHL EXCITATION ENERGIES

S3.A. First approximate wavefunction input

Assuming that one has the Cartesian coordinates of a BChl molecule in a file called `BChl-coord.xyz` (see the Supporting Information of Ref. 37 for sets of such coordinates), the first input (Figure S2) to be executed computes the Hartree-Fock energy of it, and generates a first approximate wavefunction.

```
>>> export MOLCAS_MOLDEN=ON
>>> COPY $CurrDir/BChl-coord.xyz .

 &Gateway
Title
BChl molecule
COORD = BChl-coord.xyz
BASIS = ANO-RCC-VDZP
Group = NOSYMM
RICD

 &Seward

 &SCF
CHARGE = 0
```

FIG. S2. Input to generate a SCF wavefunction.

The obtained wavefunction has 169 doubly occupied molecular orbitals (MO).

S3.B. Second approximate wavefunction input

To speed up the calculation, the first 51 MOs, corresponding to the heavy atoms $1s$ and Mg $2s$, $2p_x$, $2p_y$, $2p_z$ core orbitals, can be kept frozen. This leaves 107 inactive MOs, followed by 13 fully occupied MOs to be included in RAS1. Thus, the second, approximated wavefunction is produced by a RASSCF calculation that employs 13 MOs in RAS1, 0 MOs in RAS2, 12 MOs in RAS3, 2 holes in RAS1, 2 excitations in RAS3, and a 2-roots state averaging procedure.

The starting wavefunction is read from the file `$Project.InpOrb`, which is a copy of the previously obtained `$Project.ScfOrb`. In the following example (Figure S3), 5 MOs are exchanged through the `ALTER` keyword, following visual inspection of the previously obtained `$Project.scf.molden` file. Furthermore, 2 RASSCF calculations are performed one after the other, again to speed up the overall procedure. The first RASSCF input has a relatively large level shift value of 0.5, while the second has a more production-quality level shift value of 0.1. Finally, slightly more loose than default threshold values are employed for the convergence of the wavefunction.

S3.C. Final RASSCF wavefunction input

Finally, the active space orbitals are re-ordered according to the average of their occupation numbers in both roots. This is achieved by manually modifying the `$Project.RasOrb` file, changing the order of the orbitals, and saving the new file as the `InpOrb` file for the next calculation. The final active space is built by including 11 MOs in RAS1, 4 MOs in RAS2, 10 MOs in RAS3, 3 allowed holes in RAS1 and 3 allowed excitations in RAS3, still employing a 2 roots state averaging procedure. The corresponding input is reported in Figure S4.

S3.D. MS-RASPT2 and RASSI inputs

Once the final wavefunction is obtained, the corresponding `JobIph` file is copied as input for the final 3 calculations. The first 2 calculations (due to the inputs shown in Figures S5 and S6) can be run simultaneously, in a "job-farming" manner. Each calculation computes the RASPT2 energy of only 1 root, and its Effective Hamiltonian Coupling terms to the other. In both cases, the highest in energy 369 virtual MOs are excluded using the `DELETED` keyword, so as to reduce the calculation time.

Once both calculations are done, it is possible to run the MS-RASPT2 calculation (Figure S7), which takes as input the computed couplings. `RASSI` computes transition dipole moment (and corresponding oscillator strength and components) and transition quadrupole moment (and corresponding oscillator strength and components) for the S_0 to S_1 electronic excitation.

```
>>> export MOLCAS_MOLDEN=ON
>>> COPY $CurrDir/BChl-coord.xyz .
>>> COPY $CurrDir/$Project.InpOrb INPORB

 &Gateway
Title
BChl molecule
COORD = BChl-coord.xyz
BASIS = ANO-RCC-VDZP
Group = NOSYMM
RICD

 &Seward

 &RASSCF
LUMORB
SPIN = 1
FROZEN = 51
INACTIVE = 107
RAS1 = 13
RAS2 = 0
RAS3 = 12
NACTEL = 26 2 2
LEVSHIFT = 0.5
CIROOT = 2 2 1
ALTER = 5
1 164 154
1 163 153
1 162 151
1 161 149
1 160 147
THRS = 0.100E-03 0.100E+00 0.500E-03

 &RASSCF
JOBIPH
CIRESTART
SPIN = 1
FROZEN = 51
INACTIVE = 107
RAS1 = 13
RAS2 = 0
RAS3 = 12
NACTEL = 26 2 2
LEVSHIFT=0.1
CIROOT = 2 2 1
THRS = 0.100E-03 0.100E+00 0.500E-03
```

FIG. S3. Input to generate a singles-and-doubles RASSCF wavefunction.

```
>>> export MOLCAS_MOLDEN=ON
>>> COPY $CurrDir/BChl-coord.xyz .
>>> COPY $CurrDir/$Project.InpOrb INPORB

 &Gateway
Title
BChl molecule
COORD = BChl-coord.xyz
BASIS = ANO-RCC-VDZP
Group = NOSYMM
RICD

 &Seward

 &RASSCF
LUMORB
SPIN = 1
FROZEN = 51
INACTIVE = 107
RAS1 = 11
RAS2 = 4
RAS3 = 10
NACTEL = 26 3 3
LEVSHIFT=0.1
CIROOT = 2 2 1
THRS = 0.100E-03 0.100E+00 0.500E-03
```

FIG. S4. Input to generate the final RASSCF wavefunction.

```
/**** Input for Root 1 ****/

>>> export MOLCAS_MOLDEN=ON
>>> COPY $CurrDir/BChl-coord.xyz .
>>> COPY $CurrDir/$Project.JobOld JOBIPH

 &Gateway
Title
BChl molecule
COORD = BChl-coord.xyz
BASIS = ANO-RCC-VDZP
Group = NOSYMM
RICD

 &Seward

 &CASPT2
DELETED = 369
IMAGINARY = 0.1
MULTISTATE = 2 1 2
ONLY root = 1
```

FIG. S5. Input to compute the RASPT2 energy of the ground state and its contribution to the effective Hamiltonian couplings.

```
/**** Input for Root 2 ****/

>>> export MOLCAS_MOLDEN=ON
>>> COPY $CurrDir/BChl-coord.xyz .
>>> COPY $CurrDir/$Project.JobOld JOBIPH

 &Gateway
Title
BChl molecule
COORD = BChl-coord.xyz
BASIS = ANO-RCC-VDZP
Group = NOSYMM
RICD

 &Seward

 &CASPT2
DELETED = 369
IMAGINARY = 0.1
MULTISTATE = 2 1 2
ONLY root = 2
```

FIG. S6. Input to compute the RASPT2 energy of the first electronically excited state and its contribution to the effective Hamiltonian couplings.

```
>>> export MOLCAS_MOLDEN=ON
>>> COPY $CurrDir/BChl-coord.xyz .
>>> COPY $CurrDir/$Project.JobOld JOBIPH

 &Gateway
Title
BChl molecule
COORD = BChl-coord.xyz
BASIS = ANO-RCC-VDZP
Group = NOSYMM
RICD

 &Seward

 &CASPT2
DELETED = 369
IMAGINARY = 0.1
MULTISTATE = 2 1 2
EFFEctive Hamiltonian Couplings
-2257.99494868    0.00030943
    0.00093786 -2257.93315282

>>COPY $Project.JobMix JOB001

 &RASSI
EJOB
Property
9
'Mltpl 1' 1 'Mltpl 1' 2 'Mltpl 1' 3
'Mltpl 2' 1 'Mltpl 2' 2 'Mltpl 2' 3
'Mltpl 2' 4 'Mltpl 2' 5 'Mltpl 2' 6
MEIN
MEES
```

FIG. S7. Input to compute the MS-RASPT2 energy of the system, transition dipole moments and their components and transition quadrupole moments and their components for the computed S_0 to S_1 electronic excitation.

S4. X-RAY ABSORPTION CALCULATIONS & FROM RASSI STATES TO NBO

A set of [Open]Molcas input examples for calculating the Np $M_{4,5}$-edge X-ray absorption near edge structure (XANES) spectrum of neptunyl (NpO_2^{2+}) with RAS-SO, as well as a Jupyter Notebook example, that can be used to generate NBO FILE47 inputs from RASSI states.

S4.A. Np $M_{4,5}$-edge XANES in NpO_2^{2+}

1) A SEWARD calculation is performed (Figure S8) and the scratch directory is saved for further use. 2) Upon wish, a state-averaged RASSCF calculation can be performed for the seven NpO_2^{2+} ungerade, spin-doublet states with a minimal active space (Np^{VI}-$5f^1$, RASSCF(1e, 7o), see Figure S9), or a larger one, starting either from the guess orbitals produced by SEWARD or from some restricted KS-DFT or HF orbitals converged for NpO_2^{3+}. The RASSCF(1e, 7o) orbitals can be saved in the home directory or elsewhere for further use. 3) The orbitals produced by the previous RASSCF(1e, 7o) calculation can be inspected with LUSCUS (via the '.lus' file) and a new orbital file (e.g. 'orbs.GvOrb') can be generated with the Np $3d^10$ shell added to RAS1. The valence, Np-centered, nonbonding ϕ and δ orbitals can be added to RAS2. The antibonding π^* and σ^* orbitals can be added to RAS3. First, the valence seven spin-doublet states are calculated with an active space comprising 11 electrons and 12 orbitals ($3d^{10}$ plus $5f^1$) as shown in Figure S10, 4) the 200 core-hole spin-doublet states are calculated as shown in Figure S11, and 5) the 90 core-hole spin-quartet states are calculated as shown in Figure S12. 6) The RASSI calculation is performed (see Figure S13) to introduce spin-orbit coupling and calculate transition dipoles (and additionally other multipoles). The SONORB keyword is used to obtain orbital files (with the extension '.SOOrb.x') to be used in the production of NBO FILE47 inputs by *eXatomic*. In the sample input below, the $Project.SOOrb.1 and $Project.SOOrb.2 orbital files will be produced for the ground state Kramers pair (RASSI states 1 and 2). Such orbital files can be obtained for any RASSI state of interest as soon as the state indexes are correctly given. 7) The Np $M_{4,5}$-edge XANES can be generated from the RASSI output with any software / scripts of choice.

S4.B. Running eXatomic

Instructions for downloading and installing *eXatomic* can be found at: *https://github.com/exa-analytics/exatomic.git*. The ingredients needed to generate NBO FILE47 input files from RASSI states are 1) the orbital file produced with the RASSI 'SONORB' keyword for the RASSI state of interest, 2) the SEWARD output file containing the basis set information (exponents, contraction coefficients), and 3) some '.h5' file containing the overlap matrix (could be the $Project.rassi.h5 file for example). *eXatomic* currently does not make use of symmetry. If symmetry was used in the calculations, the 'SOOrb.x' orbital files can be de-symmetrized with the EXPBAS module of MOLCAS/OpenMolcas. A Jupyter notebook example showing how *eXatomic* can be used to generate NBO FILE47 files from RASSI states is shown in Figure S14. It also illustrates how to plot NLMOs of interest from NBO FILE39 files.

```
&SEWARD
Title=NEPTUNYL, seward calculation
High Cholesky
BSSHOW
R02E00
Symmetry=XYZ
Basis set
Np.ANO-RCC-VTZP
Np1     0.00000000     0.00000000     0.00000000
    angstroms
end of basis
Basis set
O.ANO-RCC-VTZP
O1      0.00000000     0.00000000     1.75000000
    angstroms
End of basis
Angmom=0.0 0.0 0.0
AMFI

* run a simple restricted KS-DFT calculation
* (or a RHF one if wished) for neptunyl(3+)
* to obtain guess orbitals that can be used
* in a subsequent CASSCF calculation

&SCF
Charge=3
KSDFT=PBE


* Run GRID_IT to generate a $Project.lus file,
* containing the ScfOrbs produced above,
* that can be handled with LUSCUS.

>> COPY $Project.ScfOrb INPORB

&GRID_IT
All
```

FIG. S8. Input for a SEWARD and a KS-DFT calculation on NpO_2^{2+}.

```
/* the non-expert user can prepare a set of input
    orbitals (INPORB)
* from the $Project.ScfOrbs with the LUSCUS program
    and use it
* here together with the TypeIndex keyword.
    Otherwise, the CAS input
* should be specified manually: Inactive=25 27;
    RAS2= 0 7 */

>> COPY $CurrDir/orbs.GssOrb INPORB

&RASSCF
Linear
Spin=2
Symmetry=2
nActEl=1 0 0

/* the non-expert user would use the TypeIndex
    keyword to read
* the orbital information defined during the
* preparation of the INPORBs */

Typeindex
CIRoots=7 7 1
Iter=100 100
LUMORB
ORBListing=all
ORBAppear=compact

>> COPY $Project.RasOrb
    $CurrDir/$Project-cas17-7d.RasOrb
>> COPY $Project.JobIph
    $CurrDir/$Project-cas17-7d.JobIph

>> COPY $Project.RasOrb INPORB

&GRID_IT
All
```

FIG. S9. Input for a CASSCF(1e,7o) calculation for the $5f^1$ valence spin-doublet states of NpO_2^{2+}.

```
>> COPY $CurrDir/orbs.GvOrb INPORB

&RASSCF
Linear
Spin=2
Symmetry=2
nActEl=11 1 1
Typeindex
CIRoots=7 7 1
Iter=100 100
LUMORB
ORBListing=all
ORBAppear=compact

* Supersymmetry is used to confine
* the Np 3d core-orbitals to RAS1

SupSym=1; 5 4 5 6 7 8; 0

>> COPY $Project.RasOrb
    $CurrDir/$Project-cas1112-ud.RasOrb
>> COPY $Project.JobIph
    $CurrDir/$Project-cas1112-ud.JobIph
```

FIG. S10. Input for a CASSCF(11e,12o) calculation for the $3d^{10}5f^1$ valence spin-doublet states of NpO_2^{2+}.

```
>> COPY $CurrDir/orbs.GvOrb INPORB

&RASSCF
Linear
Spin=2
Symmetry=1
nActEl=11 1 1
Typeindex
CIRoots=200 200 1
Iter=100 100
LUMORB
ORBListing=all
ORBAppear=compact
SupSym=1; 5 4 5 6 7 8; 0

>> COPY $Project.RasOrb
    $CurrDir/$Project-cas1112-gd.RasOrb
>> COPY $Project.JobIph
    $CurrDir/$Project-cas1112-gd.JobIph
```

FIG. S11. Input for a CASSCF(11e,12o) calculation for the $3d^95f^2$ spin-doublet core-excited states of NpO_2^{2+}.

```
>> COPY $CurrDir/orbs.GvOrb INPORB

&RASSCF
Linear
Spin=4
Symmetry=1
nActEl=11 1 1
Typeindex
CIRoots=90 90 1
Iter=100 100
LUMORB
ORBListing=all
ORBAppear=compact
SupSym=1; 5 4 5 6 7 8; 0

>> COPY $Project.RasOrb
    $CurrDir/$Project-cas1112-gq.RasOrb
>> COPY $Project.JobIph
    $CurrDir/$Project-cas1112-gq.JobIph
```

FIG. S12. Input for a CASSCF(11e,12o) calculation for the $3d^9 5f^2$ spin-quartet core-excited states of NpO_2^{2+}.

```
>> COPY $CurrDir/$Project-cas1112-ud.JobIph JOB001
>> COPY $CurrDir/$Project-cas1112-gd.JobIph JOB002
>> COPY $CurrDir/$Project-cas1112-gq.JobIph JOB003

&RASSI
NROF=3 all
SPINORBIT
SONORB=2; 1; 2
PROPerties=13
'AngMom' 1
'AngMom' 2
'AngMom' 3
'Mltpl 0' 1
'MltPl 1' 1
'MltPl 1' 2
'MltPl 1' 3
'MltPl 2' 1
'MltPl 2' 2
'MltPl 2' 3
'MltPl 2' 4
'MltPl 2' 5
'MltPl 2' 6
HEFF
MEES
MESO
PRPR
```

FIG. S13. Input for a RASSI (RASSCF+SOC) calculation NpO_2^{2+} using the wavefunctions converged for the $3d^{10}5f^1$ valence states and for the $3d^9 5f^2$ core-excited states.

```
In[1]:
import exa
import exatomic
import pandas as pd
import numpy as np
import os
from exatomic import UniverseWidget as UW
from exatomic.algorithms.orbital import
    add_molecular_orbitals
from exatomic import molcas
adir =
    "/Users/clausserg/Documents/From_RUSH/neptunyl/"
In[2]:
uni = exatomic.molcas.Output(adir +
    "neptunyl-seward.out")
# in this example we use RASSI GS orbitals
# desymmetrized from Ci symmetry with EXPBAS.
uni.add_orb(adir + "neptunyl.SOOrb.1.DeSymOrb")
# overlap (C1 symmetry) is needed to obtain NBO
    file47 input
uni.overlap = molcas.HDF(adir +
    "neptunyl.h5").overlap
uni = uni.to_universe()
In[3]:
from exatomic import nbo
nbo = exatomic.nbo.Input.from_universe(uni)
In[4]:
nbo.write(adir + "neptunyl.SOOrb.1.FILE47")
In[5]:
# plot nbo/nlmos from file37/file39 files
uni =
exatomic.molcas.Output(adir +
    'neptunyl-seward.out').
                    to_universe()
orbs = exatomic.nbo.output.MOMatrix(adir +
     'neptunyl.SOOrb.1.FILE39')
orbs.parse_momatrix(len(uni.basis_set_order.index))
uni.momatrix = orbs.momatrix
uni.orbital = orbs.orbital
In[6]:
# plot nlmos of interest, replace the nlmo1,
# nlmo2 etc below with the NLMO indexes of interest.
uni.add_molecular_orbitals(vector=[nlmo1, nlmo2,
    etc],
        field_params={'rmin':-15, 'rmax':15,
            'nr':80})
UW(uni)
```

FIG. S14. Example of a Jupyter notebook showing how *eXatomic* can be used to produce NBO FILE47 input files from RASSI states.

S5. *a*-ARM PROTOCOL

The *a*-ARM protocol is constituted by Phase I, *input file generator* and Phase II, *QM/MM model generator*, as shown in Figure S15. In the first step of Phase I, a user-provided, rhodopsin structure in pdb format is elaborated, so as to choose a suitable chain, residue conformations and presence of water molecules. In the second step, the program Fpocket[38] is used to locate the residues surrounding the chromophore, and thus constituting its cavity. If requested, the protocol can generate variants by exploiting the backbone-dependent rotamer library implemented in the software SCWRL4.[39] During step 3, the protocol assigns protonation states to the ionizable residues, based on the results provided by the program PROPKA,[40] and a user-provided pH value. Following the spatial distribution of charged residues, counterions are placed around the protein in step 4, thus arriving at the final ARM input.

Phase II takes the previously generated input to perform the actual calculations, as described in the main text. After two preliminary steps based on using the programs Dowser[41] (step 1) and Gromacs[42] (step 2), the protocol performs the actual QM/MM calculations employing [Open]Molcas and Tinker, as detailed in the main text. Further details can also be found in ref. 43.

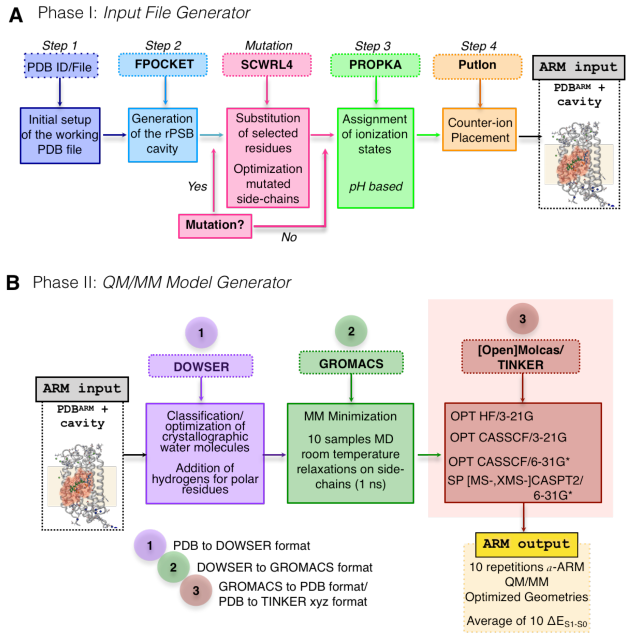

FIG. S15. General workflow of the *a*-ARM protocol for the generation of QM/MM models of wild-type and mutant rhodopsins. The *a*-ARM protocol comprises two phases: (A) *input file preparation phase*, (B) *QM/MM model generator phase*.

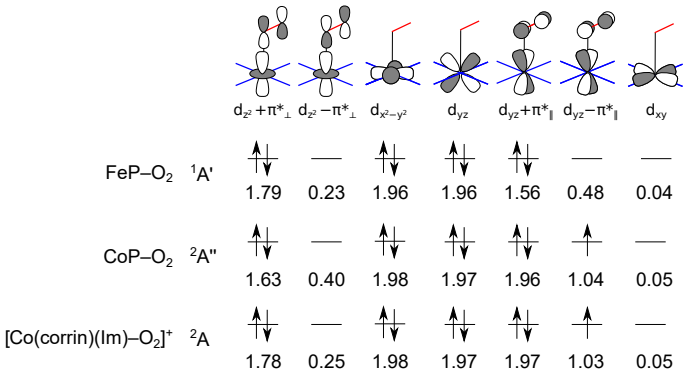

FIG. S16. Electron configurations and DMRG natural orbital occupation numbers (NOONs) of ^{1}[FeP–O_2], ^{2}[CoP–O_2], and ^{2}[Co(corrin)(Im)–O_2]$^+$

S6. ELECTRONIC STRUCTURE AND ENERGETIC PROPERTIES CALCULATED WITH DMRG-CASPT2 – COMPUTATIONAL DETAILS

All geometries were obtained using the BP86 functional,[44,45] Stuttgart effective-core potentials (ECP) and the associated basis set ECP-10-MDF[46] for the metal atom, and the 6-31G(d) basis set for the ligand atoms. The calculations were done with the TURBOMOLE v.7.3 program package.[47] Only the ground states of the metal macrocycles and the adducts were considered, i.e. ^{3}FeP, ^{2}CoP, ^{2}[Co(corrin)(Im)]$^+$, ^{1}[FeP–O_2], ^{2}[CoP–O_2], and ^{2}[Co(corrin)(Im)–O_2]$^+$. The electron configurations of the adducts are shown in Figure S16.

In DMRG calculations, the chosen active spaces were based on a previous work.[48] The active space of M–O_2 consists of 19 active orbitals: five (predominantly) metal 3d orbitals; two pairs of orbitals describing the O_2 π bond, denoted as $\pi_\perp$–$\pi^*_\perp$ (perpendicular to the macrocycle) and $\pi_\parallel$–$\pi^*_\parallel$ (parallel with the macrocycle); one pair of orbitals describing the O_2 σ bond, denoted as σ_{2p}–σ^*_{2p}; one σ_{M-N} orbital describing the covalent interaction between the metal center and the four N ligand atoms; and 7 'double-shell' orbitals of either metal and O_2. For MP, we used an active space consisting of 11 orbitals: five metal 3d orbitals, five 'double-shell' metal 3d′ orbitals, and the σ_{M-N} orbital. In [Co(corrin)(Im)]$^+$, one 3d′ is removed since it rotates out of the active space. For O_2, we used a CAS(8,10): one σ_{2p}–σ^*_{2p} pair, two π_{2p}–π^*_{2p} pairs, and four 'double-shell' π_{3p} and π^*_{3p} orbitals.

Standard DMRG-CASPT2[m = 2000] (with m number of renormalized states) calculations were done with the ANO-RCC type basis sets: [7s6p5d3f2g1h] for the metal atom;[49] [4s3p2d1f] for C, N, and O;[50] and [3s1p] for H.[51] Cholesky decomposition technique[52] with a threshold of 10^{-6} au were employed. Scalar relativistic effects were included using a standard second-order Douglas-Kroll-Hess (DKH) Hamiltonian.[53–55]

In the PT2 treatment, we used the standard ionization potential electron affinity (IPEA) shift[56] of 0.25 au. and

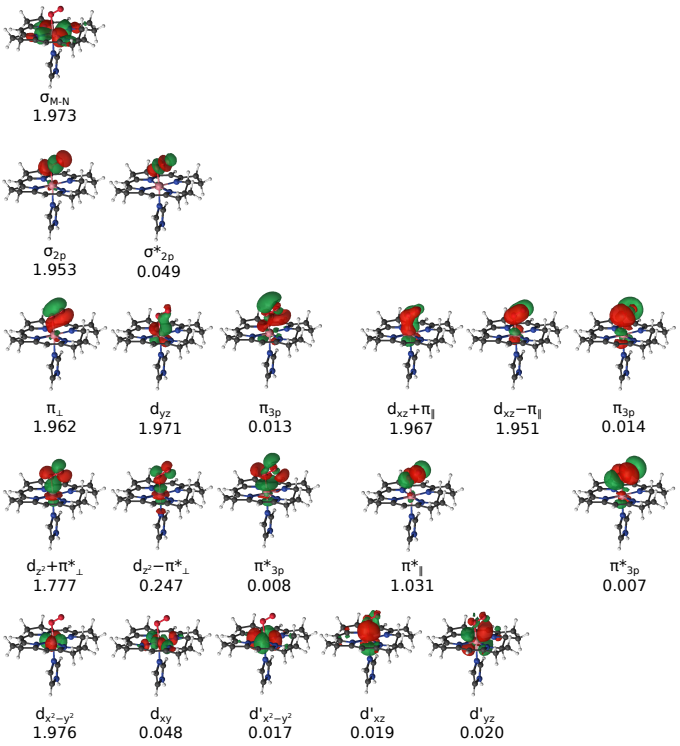

FIG. S17. Active orbitals and their natural orbital occupation numbers (NOONs) in ^{2}[Co(corrin)(Im)–O_2]$^+$

TABLE S1. 3s3p correlation contributions to the binding of O_2 to MP and [Co(corrin)(Im)]$^+$ (in kcal/mol), calculated with UKS-UCCSD and UKS-UCCSD(T). UKS = UB3LYP. Coupled-cluster calculations were done with Orca 4.0.[59]

	FeP–O_2	CoP–O_2	[Co(corrin)(Im)–O_2]$^+$
UCCSD	−0.7	−0.2	−0.1
UCCSD(T)	0.1	0.3	0.3

an imaginary shift[57] of 0.1 au. Only valence electrons were correlated. The semicore electrons (3s and 3p electrons) of the metal atom were kept frozen since we found that their contributions to the relative energies are limited (calculated at the UKS-UCCSD(T) level of theory, see Table S1). We note that in the previous study[48] a significantly larger 3s3p correlation effect was obtained for FeP–O_2 with ROHF-RCCSD(T). The difference between both results may be explained by the strong multiconfigurational character of the FeP–O_2 ground state, reducing the reliability of the ROHF-RCCSD(T) approach.[58]

The binding enthalpies of O_2 to MP and [Co(corrin)(Im)]$^+$ include the following contributions: the binding energy at the DMRG-CASPT2 level, counterpoise correction (CP) for the basis set superposition error, and thermal correction calculated at the DFT level (Table S2).

The nature of the M–O_2 bond is analyzed by the

TABLE S2. Calculated standard binding enthalpy $\Delta H^\circ_{\mathrm{DMRG-CASPT2}}$ (in kcal/mol) of O_2 to MP (M = Fe, Co) and [Co(corrin)(Im)]$^+$

	FeP–O_2	CoP–O_2	[Co(corrin)(Im)–O_2]$^+$
$\Delta E_{\mathrm{DMRG-CASPT2}}$	13.60	9.41	16.00
CP correction	−4.67	−4.51	−5.43
Thermal correction	−0.76	−0.78	−1.22
$\Delta H^\circ_{\mathrm{DMRG-CASPT2}}$	8.17	4.11	9.34

TABLE S3. Diradical character Y (in %) of MP–O_2 (M = Fe, Co) and [Co(corrin)(Im)–O_2]$^+$.

	FeP–O_2	CoP–O_2	[Co(corrin)(Im)–O_2]$^+$
$Y_\perp$	22	39	23
$Y_\parallel$	46	-	-

DMRG natural orbital occupation numbers (NOONs) and the diradical character Y_i of a bond i:

$Y_i = (1 - \frac{n^+ - n^-}{2}) \cdot 100\%$

with n^+ and n^- the NOON of the bonding and antibonding molecular orbitals, respectively. Two values of diradical character were calculated: $Y_\parallel$ related to the π bond d_{xz}–$\pi^*_\parallel$ and $Y_\perp$ related to the σ bond d_{z^2}–$\pi^*_\perp$. The large values of diradical character shown in Table S3 indicate that the wave function of FeP–O_2 is a mixture of the dominant Weiss $^2Fe^{III}$–$^2O_2^-$ structure and the McClure-Goddard $^3Fe^{II}$–$^3O_2$ configuration. Similarly, CoP–O_2 and [Co(corrin)(Im)–O_2]$^+$ can be described as a mixture of a major $^1Co^{III}$–$^2O_2^-$ and a minor $^2Co^{II}$–$^3O_2$.

S7. SELECTION OF ACTIVE SPACE WITH EXPBAS UTILITY

S7.A. Introduction

Selecting active spaces in Molcas can be made simple by using ANO basis sets and Molcas' EXPBAS module (written by B.O. Roos). EXPBAS uses the knowledge about the structure of ANO-type basis set and transforms an orbital file, computed with a small basis set into one, corresponding to a larger basis set. There is a twofold advantage to doing this while selecting active spaces. First, the virtual orbitals are far more localised and "chemical" when using a minimal basis, allowing easier identification of relevant orbitals; second, the computational time is much reduced, meaning that the obtained active space can be diagnosed much quicker than when using a large basis set. This example will work with ANO-RCC, but any ANO-type basis set could be used in a similar fashion, for instance the upcoming ANO-R

Fe	-0.000000	0.000000	0.000000
N	-1.412573	-1.412573	0.000000
C	3.437277	-0.000000	0.000000
C	-0.000000	-3.437277	0.000000
C	2.504407	-3.473828	0.000000
C	3.473828	-2.504407	0.000000
C	1.233048	-2.789799	0.000000
C	2.789799	-1.233048	0.000000
H	4.531268	0.000000	0.000000
H	0.000000	-4.531268	0.000000
H	4.557963	-2.625070	0.000000
H	2.625070	-4.557963	0.000000
N	-1.412573	1.412573	0.000000
N	1.412573	-1.412573	0.000000
N	1.412573	1.412573	0.000000
C	-3.437277	0.000000	0.000000
C	0.000000	3.437277	0.000000
C	2.504407	3.473828	0.000000
C	-2.504407	-3.473828	0.000000
C	-2.504407	3.473828	0.000000
C	3.473828	2.504407	0.000000
C	-3.473828	-2.504407	0.000000
C	-3.473828	2.504407	0.000000
C	1.233048	2.789799	0.000000
C	-1.233048	-2.789799	0.000000
C	-1.233048	2.789799	0.000000
C	2.789799	1.233048	0.000000
C	-2.789799	-1.233048	0.000000
C	-2.789799	1.233048	0.000000
H	-4.531268	0.000000	0.000000
H	0.000000	4.531268	0.000000
H	4.557963	2.625070	0.000000
H	-4.557963	-2.625070	0.000000
H	-4.557963	2.625070	0.000000
H	2.625070	4.557963	0.000000
H	-2.625070	-4.557963	0.000000
H	-2.625070	4.557963	0.000000

FIG. S18. Cartesian coordinates of FeP $^3A_{2g}$

Fe	-0.035477	0.001436	0.000000
O	-0.005914	-1.808768	0.000000
O	1.107650	-2.438912	0.000000
N	1.362796	0.250373	1.409282
N	1.362796	0.250373	-1.409282
N	-1.463257	0.160510	1.408574
N	-1.463257	0.160510	-1.408574
C	1.181755	0.225776	2.784175
C	1.181755	0.225776	-2.784175
C	2.739197	0.307341	1.231533
C	2.739197	0.307341	-1.231533
C	3.421596	0.333840	2.503499
C	3.421596	0.333840	-2.503499
C	2.452072	0.282953	3.469865
C	2.452072	0.282953	-3.469865
C	-2.841076	0.165785	1.231129
C	-2.841076	0.165785	-1.231129
C	-3.523986	0.171659	2.503559
C	-3.523986	0.171659	-2.503559
C	-2.553811	0.162097	3.470407
C	-2.553811	0.162097	-3.470407
C	-1.282189	0.154233	2.785129
C	-1.282189	0.154233	-2.785129
C	-3.489024	0.165274	0.000000
C	-0.049760	0.168345	3.430778
C	-0.049760	0.168345	-3.430778
C	3.385639	0.330261	0.000000
H	-0.048987	0.157159	4.524646
H	-0.048987	0.157159	-4.524646
H	-2.671491	0.167468	4.554832
H	-2.671491	0.167468	-4.554832
H	-4.607891	0.187748	2.623715
H	-4.607891	0.187748	-2.623715
H	-4.582888	0.166326	0.000000
H	4.478863	0.368105	0.000000
H	4.504139	0.388212	2.624084
H	4.504139	0.388212	-2.624084
H	2.569176	0.284253	4.554361
H	2.569176	0.284253	-4.554361

FIG. S19. Cartesian coordinates of FeP–O_2 $^1A'$

basis set.

This text will explain a useful strategy for using EXPBAS in Molcas, with $[Co(H_2O)(OH)_3]$ as a model system. For the purpose of this example, it will be assumed that the $[Co(OH)_3]$-subsystem is a minimal cluster model of a Co_3O_4 surface where the hydrogens are both terminators as well as ensuring charge neutrality, as the Co has a formal oxidation state of +III. The H_2O molecule represents an adsorbed water which is interacting with this surface. As we are interested in the Co-H_2O interaction, we make the initial assumption that we only need an active space on the Co-H_2O subsystem and that we can partition the basis set, using ANO-RCC-VQZP on Co and H_2O while we can be satisfied with -VDZP on the OH^--ligands in order to reduce the total amount of basis functions.

In order to obtain the targeted active space we will have to do (at least) three different calculations, 1) an initial HF/DFT calculation in a small basis to obtain some orbitals from which to select the active space, 2) a CASSCF calculation in the small basis to ensure that the active space is stable and 3) a final CASSCF calculation in the full basis following the use of EXPBAS.

S7.B. Selecting the smallest reasonable basis set

Since the whole point of using EXPBAS is to start with an as small basis as possible, the first thing to do when selecting active spaces using this module is to think about what type of active space you are after. If the active space of interest is formed purely from bonding and antibonding orbitals, not much thinking has to be done since such an active space can be described directly by a minimal basis and one can thus just set 'basis=ANO-RCC-MB' in GATEWAY and everything will work out fine. If there is reason to suspect that lone-pair electron-

Co	0.000000	-0.000000	0.000000
N	1.404637	-1.404637	0.000000
C	0.000000	3.432564	0.000000
C	3.432564	0.000000	0.000000
C	3.467401	2.499876	0.000000
C	2.499876	3.467401	0.000000
C	2.782453	1.228012	0.000000
C	1.228012	2.782453	0.000000
H	0.000000	4.526328	0.000000
H	4.526328	0.000000	0.000000
H	2.615892	4.551928	0.000000
H	4.551928	2.615892	0.000000
N	-1.404637	-1.404637	0.000000
N	1.404637	1.404637	0.000000
N	-1.404637	1.404637	0.000000
C	0.000000	-3.432564	0.000000
C	-3.432564	0.000000	0.000000
C	-3.467401	2.499876	0.000000
C	3.467401	-2.499876	0.000000
C	-3.467401	-2.499876	0.000000
C	-2.499876	3.467401	0.000000
C	2.499876	-3.467401	0.000000
C	-2.499876	-3.467401	0.000000
C	-2.782453	1.228012	0.000000
C	2.782453	-1.228012	0.000000
C	-2.782453	-1.228012	0.000000
C	-1.228012	2.782453	0.000000
C	1.228012	-2.782453	0.000000
C	-1.228012	-2.782453	0.000000
H	0.000000	-4.526328	0.000000
H	-4.526328	-0.000000	0.000000
H	-2.615892	4.551928	0.000000
H	2.615892	-4.551928	0.000000
H	-2.615892	-4.551928	0.000000
H	-4.551928	2.615892	0.000000
H	4.551928	-2.615892	0.000000
H	-4.551928	-2.615892	0.000000

FIG. S20. Cartesian coordinates of CoP $^2A_{1g}$

Co	-0.018387	0.000000	0.027299
O	-0.005998	0.000000	1.936660
O	1.105153	0.000000	2.546691
N	1.374254	1.404584	-0.162897
N	1.374254	-1.404584	-0.162897
N	-1.432265	1.403943	-0.076688
N	-1.432265	-1.403943	-0.076688
C	1.195751	2.776517	-0.255010
C	1.195751	-2.776517	-0.255010
C	2.746798	1.228896	-0.086935
C	2.746798	-1.228896	-0.086935
C	3.433398	2.498998	-0.144624
C	3.433398	-2.498998	-0.144624
C	2.468122	3.461548	-0.265265
C	2.468122	-3.461548	-0.265265
C	-2.801901	1.228352	0.052796
C	-2.801901	-1.228352	0.052796
C	-3.490425	2.498259	0.020042
C	-3.490425	-2.498259	0.020042
C	-2.531066	3.460935	-0.140241
C	-2.531066	-3.460935	-0.140241
C	-1.259418	2.776332	-0.183046
C	-1.259418	-2.776332	-0.183046
C	-3.446195	0.000000	0.145834
C	-0.033802	3.424025	-0.285492
C	-0.033802	-3.424025	-0.285492
C	3.393109	0.000000	-0.015364
H	-0.036174	4.515142	-0.360073
H	-0.036174	-4.515142	-0.360073
H	-2.652544	4.541924	-0.219287
H	-2.652544	-4.541924	-0.219287
H	-4.572246	2.615588	0.093855
H	-4.572246	-2.615588	0.093855
H	-4.535236	0.000000	0.246607
H	4.484879	0.000000	0.049400
H	4.517164	2.616735	-0.112094
H	4.517164	-2.616735	-0.112094
H	2.585691	4.543006	-0.343605
H	2.585691	-4.543006	-0.343605

FIG. S21. Cartesian coordinates of CoP–O_2 $^2A'$

s/nonbonding orbitals are of interest, however, then a minimal basis will not suffice, as such orbitals often require a secondary shell in order to be properly correlated.

In the model that we are considering here, we are targeting an active space which can describe the interaction between Co and H_2O. A simple guess for such an active space would be to include the 3*d*-orbitals of Co and the valence orbitals of H_2O formed from oxygens 2*p*-orbitals. This would correspond to 8 orbitals, of which only two are bonding orbitals, as one of the 2*p*-orbitals is nonbonding in H_2O. To properly correlate the 3*d*-orbitals and the H_2O nonbonding orbital, we would thus need to add a secondary orbital shell on Co and the oxygen in H_2O, giving a total active space of (16,12).

The brute-force approach to include the additional set of orbitals would be to simply employ ANO-RCC-VDZ since that basis contains all necessary orbitals. Such a strategy does, however, generate a lot of unnecessary virtual orbitals which, due to the increased number of basis functions, become fairly delocalised, making it hard to identify the actual orbitals of interest. A more elegant strategy is to remember that in Molcas, basis set names such as ANO-RCC-XXXX are actually alias' used to simplify input. For example, ANO-RCC-MB is an alias for ANO-RCC...4s3p1d and -VDZ for ...5s4p2d for Co. The input will accept any number of combinations of contracted functions, so rather than using the full -VDZ contraction, we can set the Co basis set to be ANO-RCC...4s3p2d instead, meaning that we are only adding the basis functions which we think are necessary to correlate the Co 3*d*-orbitals, namely the Co 4*d*-orbitals. Similarly, for the nonbonding H_2O-orbital which, in essence, is an oxygen 2*p*-orbital, we can just expand the basis set of the oxygen in H_2O to ANO-RCC...2s2p to add 3*p*-orbitals to it. Since we have more than one oxygen in the complex and we only care about adding an addi-

Co	-0.065808	0.133912	-0.378993
N	1.813143	0.136651	-0.378723
N	-0.111162	2.072740	-0.264643
N	-1.985391	-0.055311	-0.560758
N	0.187375	-1.655826	-0.901181
C	2.480933	-1.192479	-0.394705
C	3.912154	-0.884416	-0.880713
C	4.120363	0.587906	-0.441660
C	2.697345	1.117340	-0.380750
C	2.314465	2.486385	-0.317222
C	1.001059	2.907800	-0.275241
C	0.582749	4.368651	-0.278798
C	-0.948202	4.305916	-0.109634
C	-1.248623	2.823734	-0.226130
C	-2.544632	2.309356	-0.289631
C	-2.882300	0.966132	-0.463409
C	-4.312965	0.480463	-0.598550
C	-4.163687	-1.053976	-0.628761
C	-2.659875	-1.262289	-0.703117
C	-2.062021	-2.494555	-0.884652
C	-0.656020	-2.663085	-1.032362
C	0.068160	-3.950277	-1.394350
C	1.544189	-3.602593	-1.066698
C	1.570742	-2.073276	-1.251470
N	-0.629859	-1.035260	3.772328
C	0.212726	0.042605	3.986861
C	-0.830340	-1.157453	2.426986
N	-0.159344	-0.211485	1.770023
C	0.494945	0.541282	2.732135
H	1.742496	-1.813772	-2.317750
H	-1.495644	4.890548	-0.867846
H	4.755614	1.167998	-1.131432
H	-1.453110	-1.925032	1.972012
H	1.769357	-3.857290	-0.015007
H	-4.676448	-1.523831	-1.484147
H	-3.368347	3.027216	-0.239121
H	4.664780	-1.570926	-0.463798
H	3.956307	-0.958102	-1.982392
H	4.588041	0.655704	0.559873
H	3.100292	3.245931	-0.337277
H	1.088528	4.936067	0.519957
H	0.871111	4.840351	-1.235299
H	-1.276135	4.680993	0.877351
H	-4.941369	0.844116	0.231976
H	-4.760705	0.873107	-1.529650
H	-4.571164	-1.526102	0.283490
H	-2.705884	-3.372371	-0.984843
H	-0.070420	-4.171020	-2.470799
H	-0.317280	-4.820986	-0.837665
H	2.266424	-4.128804	-1.709429
H	2.481601	-1.583809	0.643787
H	0.522776	0.346566	4.983310
H	1.120037	1.387331	2.455073
H	-1.030982	-1.632478	4.493449

FIG. S22. Cartesian coordinates of $[Co(corrin)(Im)]^+$ ^2A

Co	-0.077104	0.040558	-0.226013
N	1.808130	0.029192	-0.134257
N	-0.127689	1.975628	0.025795
N	-2.015321	-0.166405	-0.312714
N	0.179846	-1.786813	-0.616449
C	2.468228	-1.300855	-0.097912
C	3.902933	-1.001079	-0.580866
C	4.105642	0.483660	-0.178270
C	2.681192	1.012573	-0.139164
C	2.295267	2.384512	-0.082516
C	0.983763	2.806354	-0.003498
C	0.558384	4.264675	0.029809
C	-0.966953	4.191241	0.248162
C	-1.265890	2.710864	0.104273
C	-2.562403	2.191478	0.046669
C	-2.902058	0.853158	-0.163810
C	-4.336684	0.368487	-0.261708
C	-4.188439	-1.155207	-0.454015
C	-2.683487	-1.369737	-0.475289
C	-2.082600	-2.605015	-0.625473
C	-0.670838	-2.784557	-0.725036
C	0.054340	-4.089480	-1.017358
C	1.528544	-3.734364	-0.685238
C	1.565903	-2.213980	-0.932107
N	-0.616843	-1.110307	3.847142
C	0.247984	-0.049926	4.060285
C	-0.835630	-1.222594	2.505864
N	-0.152077	-0.282736	1.850415
C	0.528280	0.453592	2.807908
H	1.741279	-1.985715	-2.001933
H	-1.541760	4.789313	-0.478221
H	4.738285	1.048342	-0.882996
H	-1.475836	-1.975048	2.050519
H	1.740102	-3.949291	0.378293
H	-4.641625	-1.510230	-1.395161
H	-3.386933	2.904449	0.134185
H	4.654133	-1.674255	-0.140516
H	3.951280	-1.100971	-1.679470
H	4.569806	0.581925	0.822185
H	3.080828	3.143539	-0.114620
H	1.088335	4.825839	0.816876
H	0.810297	4.746441	-0.931793
H	-1.263668	4.540524	1.254240
H	-4.904567	0.636858	0.646065
H	-4.849153	0.855088	-1.109989
H	-4.657529	-1.733681	0.360684
H	-2.725129	-3.482855	-0.731577
H	-0.071719	-4.349251	-2.086291
H	-0.344061	-4.935151	-0.432133
H	2.254113	-4.289284	-1.299083
H	2.462601	-1.655304	0.954448
H	0.574294	0.242151	5.054928
H	1.165991	1.288969	2.529538
H	-1.023076	-1.704304	4.568085
O	-0.114191	0.480368	-2.141899
O	0.804156	0.028869	-2.910774

FIG. S23. Cartesian coordinates of $[Co(corrin)(Im)–O_2]^+$ ^2A

tional set of functions on one of them, we can label the atoms by adding *_LABEL* in the xyz-file, as illustrated in **figure S24**, to make sure that we are only adding more orbitals where needed. Adding *_LABEL* to the xyz-file is a part of Molcas' extension to the xyz-file format

which allows for a simple partitioning of basis sets. In **figure S24**, the expansion to ANO-RCC-VQZP/-VDZP is anticipated by labeling the atoms which will be of -VDZP quality by *_MB*.

```
 10
File generated by luscus version 0.8.6 beta
 CO     0.408972000     0.000003000    -0.000041000
 O     -1.630028000     0.000019840     0.000020980
 H     -2.110029522     0.788290319    -0.264229262
 H     -2.315334131    -0.637383577     0.213731978
 O_MB   0.994383000    -1.320566000    -0.999132000
 O_MB   0.994418000     1.525504000    -0.644165000
 O_MB   0.994473000    -0.204942000     1.643120000
 H_MB   1.924311000     1.435453000    -0.950073000
 H_MB   1.923035000    -1.542611000    -0.765324000
 H_MB   1.923873000     0.106686000     1.717068000
```

FIG. S24. XYZ coordinates for $[Co(H_2O)(OH)_3]$. A Molcas specific extension for element names is used in this example.

S7.C. The first HF/DFT calculation

In $[Co(H_2O)(OH)_3]$ we have a Co(III), which is a d^6 species. There are therefore three spin-states which may be relevant for this complex; singlet, triplet and quintet. Regardless of which of the spin-states are of actual interest in the study, we recommend to start with optimising the wavefunction in the highest possible spin. This is because it will drastically reduce the number of configuration state functions, meaning that an active space can be found quicker. So in this example, we want to begin by finding an active space for the quintet state. The obtained high-spin CASSCF wavefunction can then be used as input for the lower spin-states. When doing so, one should keep in mind that the active space in lower spin states may not be identical to the high-spin active space, so some tweaking may become necessary.

For the starting HF/DFT wavefunction, there are three possibilities for generating it. The first is to do an unrestricted HF/DFT calculation using Molcas' SCF module, for which an average natural orbital file is produced containing a single set of MOs obtained by averaging over the alpha and beta orbitals (the *$Project.UnaOrb* file). Rather than an unrestricted HF/DFT calculation, a restricted open-shell HF calculation can be performed using the RASSCF module and setting the spin, active electrons and active orbitals in such a way that only one single CSF is allowed. Finally, the complex can be forced to be a closed-shell singlet for the HF/DFT calculation, not bothering about the spin state until the CASSCF calculation. Which strategy to pursue is mostly a matter of taste, although if there is reason to suspect that a DFT optimisation will provide a better set of starting orbitals than HF, either the first or the third strategy will have to be employed since Molcas does not allow for restricted open-shell DFT optimisations.

For the purpose of this example, the second strategy was employed by setting the RASSCF keywords 'spin=5', 'nactel=4' and 'ras2=4', with the full input given in **figure S25**. From this calculation, the targeted (16,12)-active space could easily be identified, with the canonical ROHF orbitals of the selected active space being given in **figure S26**.

```
/* 1: Generate the job information using GATEWAY. */

&GATEWAY
  coord=$CurrDir/$Project.xyz
  basis=ANO-RCC-MB, Co.ANO-RCC...4s3p2d,
      O.ANO-RCC...2s2p
  group=C1
  RICD

/* 2: Compute integrals using SEWARD. */

&SEWARD

/* 3: Set up a ROHF caculation using the RASSCF
    module. With 4 electrons in 4 orbitals and a
    quintet spin state, only one possible CSF
    exists which corresponds to a single Slater
    determinant with all orbital empty or doubly
    occupied except for 4 singly occupied
    orbitals. */

&RASSCF
  spin=5
  charge=0
  nactel=4
  ras2=4

/* 4: Generate a .lus file for orbital
    visualtisation using GRID_IT. The keyword
    "all" specifies that all orbitals should be
    visualised, which is beneficial when
    identifying the active space. Since a small
    basis is used, the resulting orbital file will
    not be obnoxiously large. */

&GRID_IT
  all
```

FIG. S25. Input file for ROHF calculation.

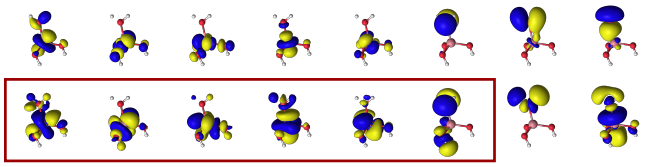

FIG. S26. Canonical orbitals of $[Co(H_2O)(OH)_3]$ from a ROHF calculation. The orbitals in the red box are the orbitals that would not appear, had only ANO-RCC-MB been used.

S7.D. Running the large basis CASSCF

After optimising the active space identified from the HF/DFT calculation with CASSCF and making sure that the desired active space has been obtained, EXPBAS can be used to expand the wavefunction into the larger basis set. When using EXPBAS, the *$Project.RunFile* file from both the small and the large basis set calculations are needed in the Molcas work directory as 'RUNFIL1' and 'RUNFIL2' respectively. Further, the RasOrb file containing the CASSCF wavefunction optimised in the smaller basis is also needed. An example of how to write the input is given in **figure S27**. As can be seen from **figures S28** and **S29**, displaying the small and large basis natural CASSCF orbitals respectively, the active space was stable after expanding the wavefunction into the larger basis.

```
/* 1: Collect information about the small basis set
     job using GATEWAY. This is done in order to
     create the RUNFIL1 file necessary for EXPBAS
     with EMIL's copy command. This file could have
     been stored during the previous small basis
     CASSCF run by using the EMIL command '> copy
     $Project.RunFile $CurrDir */

&GATEWAY
  coord=$CurrDir/$Project.xyz
  basis=ANO-RCC-MB, Co.ANO-RCC...4s3p2d,
      O.ANO-RCC...2s2p
  group=C1
  RICD
> copy $Project.RunFile RUNFIL1

/* 2: A second call to GATEWAY with the large basis
     set. Notice that the RunFile must explicitly
     be copied to RUNFIL2. */

&GATEWAY
  coord=$CurrDir/$Project.xyz
  basis=ANO-RCC-VQZP, O_MB.ANO-RCC-VDZP,
      H_MB.ANO-RCC-VDZP
  group=C1
  RICD
> copy $Project.RunFile RUNFIL2

/* 3: Call to EXPBAS with the converged CASSCF
     wavefunction from the small basis. In this
     example, the wavefunction was stored as
     $Project.MB.InpOrb in $CurrDir. */

&EXPBAS
  fileorb=$CurrDir/$Project.MB.InpOrb

/* 4: Compute the integrals of the large basis set
     using SEWARD. */

&SEWARD

/* 5: Start the CASSCF wavefunction optimisation
     with the RASSCF module. Notice that
     $Project.ExpOrb generated by EXBAS has to be
     given explicitly as input orbital file. */

&RASSCF
  fileorb=$CurrDir/$Project.ExpOrb
  spin=5
  charge=0
  nactel=12
  ras2=16

/* 6: Generate a .lus file for orbital
     visualisation using GRID_IT. Default is to
     only visualise the active orbitals. */

&GRID_IT
```

FIG. S27. Input demonstrating the usage of EXPBAS utility.

An alternative approach for finding the active space would be to simply specify the desired number of active electrons, orbitals and the spin and allow the optimisation algorithm to find the most stable active space by starting with guess orbitals contained in the *$Project.GuessOrb* file. These orbitals are generated by Molcas' GuessOrb program which is called by SEWARD. So what is the point of going through the hassle of doing EXPBAS? Simply put, a CASSCF optimisation starting from the orbitals in the *$Project.GuessOrb* file can easily lead to active spaces that either are local minima or that are not localised in the relevant part of the complex. As an example, starting with such orbitals for $[Co(H_2O)(OH)_3]$ results in the active space visualised in **figure S30**. The obtained active space contains some of the desired *d*-orbitals, but none of the H_2O orbitals. Rather, the optimisation algorithm decided that the π-system of the surrounding OH^--groups was more important to correlate. This active space would be useless for describing the Co-H_2O interaction. Additionally, the active space obtained by this procedure corresponds to a local minimum, as the energy of the EXPBAS-obtained active space is 9 mE_h, or 0.25 eV, lower in energy. Of course, this does not guarantee that the EXPBAS-obtained active space is the global minimum active space, but the chemistry associated with this active space is certainly of more interest.

As an additional bonus, in this particular example the total time that Molcas was run was 46 min 40s using the EXPBAS procedure, while the run time for the guess orbital based calculation was 1 h 48 min, running MOLCAS 8.4 with 8 MPI cores on the Tetralith cluster in NSC Linköping, Sweden; compiled with intel compilers and using sequentially threaded MKL as a BLAS library.. This is largely due to the much better input provided for the large basis CASSCF calculation, with the EXPBAS calculation converging in 21 interations, as compared to 74 iterations with the guess orbital based calculation. While the speed benefit is not that important for a small model such as the one used here, this can have a big impact on larger structures.

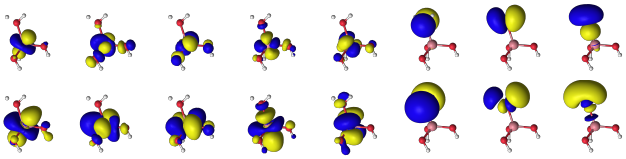

FIG. S28. Natural CASSCF orbitals of $[Co(H_2O)(OH)_3]$ from the small basis calculation.

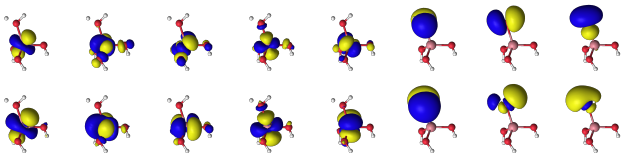

FIG. S29. Natural CASSCF orbitals of $[Co(H_2O)(OH)_3]$ from the large basis calculation

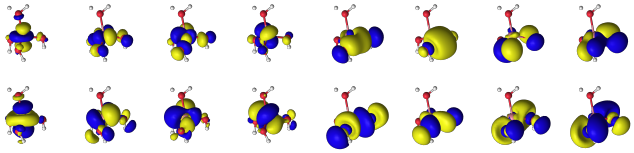

FIG. S30. Natural CASSCF orbitals of $[Co(H_2O)(OH)_3]$ when starting from the guess orbitals with the large basis.

S7.E. Concluding statements

In this example, one strategy for obtaining active spaces using Molcas' EXPBAS module has been discussed for a $[Co(H_2O)(OH)_3]$ model system. The main advantage of using EXPBAS lies in the reduction of basis functions used in the active space identification stage of a multiconfigurational study. Reducing the number of basis function both produces virtual orbitals which are easier to identify, as well as reduces the computational time spent when optimising the active space. Conversely though, the lowest number of required calculations for obtaining the active space is three, one initial HF/DFT calculation, one small basis XASSCF (X=C,R) calculation and one large basis XASSCF calculation. Starting from guess orbitals obtained through SEWARD or a HF/DFT in the large basis would require one or two calculations respectively, if the correct active space is immediately identified. This is often not the case, however, giving the EXPBAS procedure a competitive edge as it often produces the desired active space on the first attempt due to the careful build up of the active space.

S8. $4f^7$ AND $4f^65d^1$ ENERGY LEVELS OF Eu^{2+} IN CaF_2

Here we summarize calculations of the $4f^7$ and $4f^65d^1$ manifolds of Eu^{2+} doped in CaF_2.[60] Full details on how to perform calculations like this can be followed in the tutorial "Performing ab initio calculations on complex manifolds of excited states of lanthanides in solids".[61]

In CaF_2:Eu^{2+}, Eu^{2+} substitutes for Ca^{2+} in an 8-fold O_h cubic coordination. The calculations are done in D_{2h} symmetry an the wave functions are identified and assigned to the O_h group. An AIMP embedded-cluster Hamiltonian of CaF_2:EuF_8 is used, which is made of

an AIMP embedding potential of the solid host CaF_2, added to the second-order relativistic Douglas-Kroll-Hess (DKH) Hamiltonian of the cluster EuF_8. A two-step calculation is performed.

In the first, spin-orbit coupling free, step, SA-RASSCF/MS-CASPT2 calculations are done. The $(25s22p15d11f4g2h)[9s8p5d4f3g2h]$ (quadruple-zeta with polarization quality) relativistic ANO-RCC basis set is used for Europium and $(14s9p4d)[5s4p3d]$ for Fluorine. The outermost $3s$ and $3p$ orbitals of the neighboring Ca^{2+} ions are used as orthogonalization functions, contracted as $[1s1p]$. Five s-type Gaussian functions were also added to the cluster bases at the next six neighboring interstitial sites in order to be able to represent impurity-trapped-exciton states.

A restricted active space made of 7 electrons in 20 molecular orbitals of main character Eu-4f, Eu-5d, Eu-6s, and Eu-5f is used. We compute the following terms and used the following state-averages for the molecular orbital optimizations: For states in the $4f^7$ configuration, 1 $^8A_{1u}$ term; 2 $^6A_{1u}$, 2 $^6A_{2u}$, and 4 ^6E_u; and 6 $^6T_{1u}$ and 6 $^6T_{2u}$ terms. For the $4f^65d$ and $4f^66s$ configurations, 1 $^8A_{1g}$, 2 $^8A_{2g}$, and 3 ^8E_g; 6 $^8T_{1g}$ and 5 $^8T_{2g}$; 39 $^6A_{1g}$, 36 $^6A_{2g}$, and 75 ^6E_g; and 108 $^6T_{1g}$ and 111 $^6T_{2g}$ terms. In all cases, all possible occupations were allowed in the Eu-4f shells and up to 4 electrons were allowed in the Eu-5d, Eu-6s, and Eu-5f shells (we refer to this calculations as SA-RASSCF($4f/5d6s5f-4e$)); this restricted active space should be fairly close to a complete 7-electron active space where all possible occupations of the 20 molecular orbitals would be allowed. Inclusion of the Eu 5f-shells in the active space is needed for a better account of the large radial correlation effects in the 4f-shell, which strongly affect interconfigurational transitions also in the first half of the lanthanide series.

In the MS-RASPT2 calculations we correlated all cluster valence electrons except those with main character Eu-4d. We used the standard IPEA value (0.25 au) and an imaginary level shift of 0.15 to avoid intruder states. Symmetry analyses of the SA-RASSCF multiconfigurational reference wave functions allowed us to perform MS-RASPT2 calculations over actual spin and O_h symmetry blocks, which avoids symmetry breakings at the MS-RASPT2 level of calculations and beyond. Small spikes or irregularities in the potential energy surfaces were found coinciding with sharp avoided crossings between some high energy lying spin sextet excited states of the $4f^65dt^1_{2g}$ and $4f^6ITE^1_{a_{1g}}$ configurations, which did not disappear using different and reasonable values of MS-RASPT2 parameters. These high energy sextet states were excluded from the subsequent RASSI-SO calculations.

The resulting potential energy curves vs. Eu-F distance in this spin-orbit-coupling free step are shown in the left panels of Fig. S31.

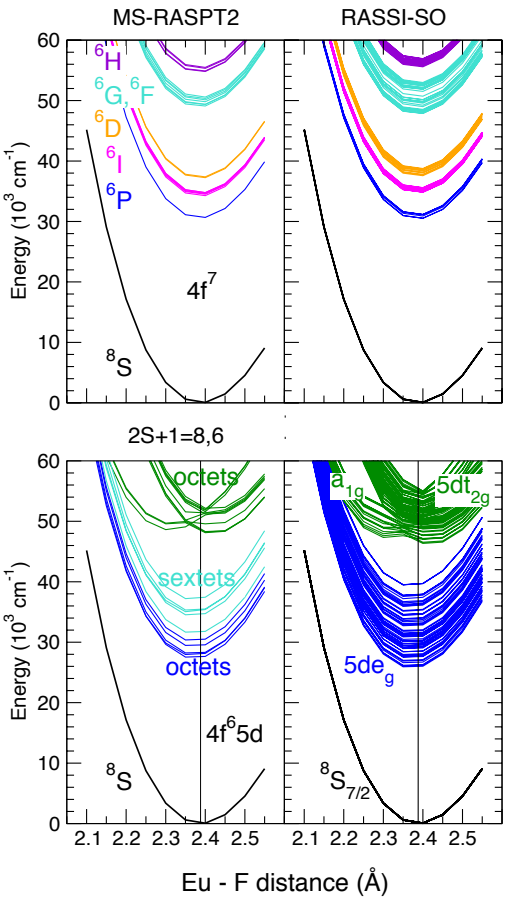

FIG. S31. Potential energy curves of the $4f^7$ (top) and $4f^65d$ (bottom) configurations of Eu^{2+} in CaF_2 without (left) and with (right) spin-orbit coupling.

In the second step, RASSI-SO calculations are done with the full DKH Hamiltonian (including spin-orbit coupling) in the atomic mean-field integrals approximation (AMFI). In these, we chose the transformed RASSCF wave functions (first-order wave functions of the MS-RASPT2 method) as a many-electron basis and the MS-RASPT2 energies as shifting factors. We allow the coupling of all MS-RASPT2 states of spin multiplicity $2S+1 \geq 6$. The potential energy curves are shown in the right panels of Fig. S31.

The bond lengths, vibrational frequencies, and minimum-to-minimum energies of the RASSI-SO potential energy curves are used to simulate absorption spectra. The result for CaF_2:Eu^{2+} is shown in the upper left panel of Fig. **??**. In the Figure, we include the equivalent results for Eu^{2+} in the series CaF_2, SrF_2, BaF_2 and CaS, SrS, BaS, as well as their comparison with experimental absorption spectra.

S9. MAGNETIC PROPERTIES OF AN ACTINIDE COMPLEX CALCULATED WITH MPSSI – COMPUTATIONAL DETAILS

The structure of the hexakisamidouranium(V) complex, $[U(NH_2)_6]^{-1}$ was optimized within a density functional theory (DFT) approach based on the BP86 functional,[44,45] and making use of a Stuttgart effective-core potential (ECP)[62] and the associated basis set def-TZVP basis set for the uranium atom as well as corresponding def-TZVPP basis sets for the ligand atoms[63]. The optimization was done with the TURBOMOLE v.7.0.2 program package[47] with the size of the numerical grid within the DFT module set to m5. Cartesian coordinates of the optimized complex are compiled in Table S4.

TABLE S4. Cartesian coordinates of $[U(NH_2)_6]^{-1}$

U	0.0000000	0.0000000	0.0000000
N	-0.0000000	2.2670895	0.0000000
N	0.0000000	0.0000000	2.2670895
N	-2.2670895	0.0000000	0.0000000
N	-0.0000000	-2.2670895	0.0000000
N	0.0000000	0.0000000	-2.2670895
N	2.2670895	-0.0000000	0.0000000
H	2.8561785	0.0000000	-0.8332511
H	-0.8332511	2.8561785	0.0000000
H	0.8332511	2.8561785	0.0000000
H	-0.0000000	0.8332511	2.8561785
H	0.0000000	-0.8332511	2.8561785
H	-2.8561785	0.0000000	-0.8332511
H	-2.8561785	0.0000000	0.8332511
H	0.8332511	-2.8561785	0.0000000
H	-0.8332511	-2.8561785	0.0000000
H	0.0000000	-0.8332511	-2.8561785
H	-0.0000000	0.8332511	-2.8561785
H	2.8561785	0.0000000	0.8332511

Two types of active spaces were considered in the DMRG calculations. The minimal active orbital space for this U(V) complex comprises all $5f$ orbitals of the U center, resulting in a CAS(1,7). In addition, the latter active orbital space was further enlarged by including the predominantly U $6p$ orbitals (see also Fig. 4 in the main text) leading to a CAS(7,10).

The number of renormalized states m was set to $m = 1024$ in the DMRG-SCF and ensuing MPSSI calculations which is sufficiently large to yield identical results in comparison with corresponding CASSCF/CASSI calculations. All [Open]Molcas calculations were carried out with the ANO-RCC type basis sets[49–51] using a TZVP contraction type in combination with Cholesky decomposition[52] of the two-electron repulsion integrals with a threshold of 10^{-6} au were employed. Scalar relativistic effects were taken into account by means of a standard second-order Douglas-Kroll-Hess Hamiltonian.[53–55]

The multi-state CASPT2 calculations were carried out with a zero ionization potential electron affinity shift[56] while employing otherwise the default settings of the multi-state CASPT2 approach in [Open]Molcas. In the SINGLE_ANISO calculations which take the spin-orbit coupled output states of the preceding standard MPSSI calculation as input functions, the computation of the temperature-dependent magnetic susceptibility $\chi(T)$ was performed within a temperature range of $T = 1 - 300$K using a step size of 1K and applying an external magnetic field of 0.7 Tesla.

S10. MOLECULAR DYNAMICS

The input file below is an example of a QM/MM molecular dynamics simulation with trajectory surface hopping. We used this input to study the excited state reactivity of the green-absorbing Proteorhodopsin (GPR). The QM/MM simulation requires the use of a classical force field which is done by the TINKER program. Therefore the tinker keyword is placed in the GATEWAY section of the input file. The ESPF module is used to account for the interaction between the QM region with the electrostatic potential of the MM region. The full retinal Schiff base moiety is in the QM region and described at the CASSCF(12,12)/6-31G* level of theory. Two roots are used for the state averaged CASSCF calculation and the excited state MD is initialized in the second root (first excited state). The input of the DYNAMIX module specifies 400 steps with timestep of 1 fs. The number of steps can be decided based on the reaction of interest. In this example 400 MD steps are sufficient to observe the surface hopping event. Each step of this simulation takes about 40-45 minutes, depending on the hardware. Further specification in the DYNAMIX module include the choice of the integrator for the equations of motion, namely the Velocity-Verlet algorithm. The surface hopping algorithm was selected to be Tullys Fewest Switches with decoherence correction of 0.1 Hartree. The results of this dynamics run are presented in Fig. 5 of the manuscript. Please note that this input also requires the forcefield parameters set, TINKER key file with QM and MM subsystem information and the geometry in TINKER xyz format.

```
/* Nonadiabatic MD example */
&Gateway
```

```
 tinker            /* QM/MM via Tinker interface */
 group = NoSym
 basis = 6-31G*

>> COPY $Project.JobIph $Project.JobOld
>> LINK -FORCE $Project.JobOld JOBOLD

>> FOREACH A in (1 .. 400) /* 400 steps of MD */

&Seward
 Title = Nonadiabatic MD

&Espf
 External = tinker
 Lamorok

&Rasscf
 Jobiph
 Charge = +1
 Nactel = 12 0 0
 Ras2 = 12
 Ciroot = 2 2 1
 Mdrlxr = 2   /* Note: MDRLXR, not RLXR */

>> COPY $Project.JobIph $Project.JobOld

&Surfacehop   /* To take surface hops into account
    */
 Tully        /* Use the Tully algorithm for
     surface hop */
 Decoherence = 0.1 /* Must be used with Tully.
     Includes a decoherence
           correction in the ppulation density
               matrix */

&Alaska

&Dynamix   /* Compute MD */
 Velver       /* Use Velocity-Verlet algorithm */
 Velo = 1     /* Read initial velocities from file
     $Project.velocity.xyz */
 Dt = 41.3    /* Timestep in a.u. of time. Here -
     1fs */
 Thermo = 1   /* Scale velocities in order to keep
     total energy constant */

>> COPY $Project.xyz $CurrDir/$Project.$A.xyz
          /* Saves coordinates at every step */
>> COPY $Project.velocity.xyz
    $CurrDir/$Project.velocity.$A.xyz
          /* Saves velocities at every step */

>>ENDDO       /* End of the FOREACH loop */
```

S11. CONSIDERATIONS ON THE APPLICATION OF THE RASSCF / RASPT2 / RASSI METHODOLOGY FOR THE ACCURATE DESCRIPTION OF ELECTRONIC STRUCTURE OF COMPOUNDS CONTAINING LANTHANIDE AND/OR TRANSITION METAL IONS

The computational strategy based on RASSCF / RASPT2 / RASSI is able to tackle difficult near-degenerate problems, while the lanthanides and transition metal compounds are perfect examples of this kind, given their partly filled and relatively isolated f/d shell. A few considerations are given here as advice for readers who wish to try running such calculations for the evaluation of electronic structure and magnetic properties of such compounds:

- The number of roots optimized at the RASSCF level is important. The minimal number of roots needed to describe qualitatively correct the energy spectra and properties of a lanthanide compound includes the states originating from the ground atomic term L. Including less states than the ground term's orbital multiplicity $2L + 1$ into the spin–orbit mixing may result in artifacts and even erroneous results. Ideally, quantitative description is obtained when all roots arising from the f^n shell (d^n for transition metals) are included in the spin–orbit mixing in RASSI. The effect of the spin–orbit mixing of the excited states is noticed in a shift of the saturation values of magnetic susceptibility and magnetization (see e.g. Table 1.4 in[64]), and also plays a role for the g-tensor of the ground and excited states.

- Electron correlation is important for quantitative description of the crystal-field splitting and magnetic properties. As shown in[30], the account of electron correlation effects by either enlarging the active space of the CASSCF method and/or by multiconfigurational second-order perturbation theory (CASPT2), significantly improves the comparison between calculated and measured energy spectra and magnetic properties of the low-lying states.

- Enlarging the active space of the RASSCF method beyond the f^n shell poses significant computational difficulties. Their origin lies in the necessity to optimize a large amount of roots (needed by the spin–orbit interaction), while using a large variational space. Another issue is the nature of the wave functions of the excited roots. Typically, the f character of low-lying states is dominant, while the excited ligand field states of f^n character are usually intermixed with the charge transfer states to/from additional orbitals, of $f^{n\pm1}L^{\mp1}$ kind. State-average CASSCF optimization of such mixture of different roots poses not only convergence issues, but also, the obtained results may be qualitatively worse compared to minimal ac-

tive space calculation. Following[30], part of dynamical correlation may be included at RASSCF level, by enlarging the active space for lanthanides to include $5s^2$, $5p^6$, $5d^0$, $6s^0$, $6p^0$ shells and, eventually, the double shell—$5f^0$, all localized on the metal site. To overcome the difficulties of finding the right empty orbitals with dominant metal character, a new keyword (IVO) has been added to the RASSCF module. This keyword enables the diagonalization of the subspace of the core Hamiltonian spanned by the empty orbitals, delivering thus a set of highly localized empty orbitals. The subsequent RASSCF calculation will typically have a straightforward convergence, provided the active space is enlarged by complete atomic shells. Nevertheless, it is recommended to inspect the active orbitals by, e.g. Luscus[65].

- The electrostatic effects of the remaining crystal, i.e. the role of the charged ions in the crystal (Madelung potential) needs to be accounted as well, at least by point charge model (SEWARD / XFIELD).

S12. AB INITIO CALCULATION OF MAGNETIC PROPERTIES OF A MONO-NUCLEAR COMPOUND / FRAGMENT

S12.A. Standard input for a CASSCF/RASSI calculation on a $[CoCl_4]^{2-}$ dianion

In this section we give brief explanation of the main input and output sections of the SINGLE_ANISO program.

```
&SEWARD
ANGM
-1.03554860   0.70324574   0.00000000 Angstrom
AMFI
DOUG
RICD
SDIP
CDTH = 1.0d-6
Coord
5

 Co1 -1.03554860 0.70324574 0.00000000
 Cl2 -3.18554860 0.70327223 0.00000000
 Cl3 -0.31886953 1.71675612 1.75546798
 Cl4 -0.31890654 -1.32380240 0.00000000
 Cl5 -0.31886953 1.71675612 -1.75546798
Group=NoSYMM
Basis=ANO-RCC-VDZ

&RASSCF
SPIN = 4
NACTEL = 7 0 0
INAC = 45
RAS2 = 5
CIROOT = 10 10 1

>>COPY $Project.JobIph JOB001
>>COPY $Project.RasOrb INPORB

&RASSCF
LUMORB
SPIN = 2
NACTEL = 7 0 0
INAC = 45
RAS2 = 5
CIROOT = 40 40 1

>>COPY $Project.JobIph JOB002
>>COPY $Project.RasOrb INPORB

&RASSI
MEES
MESO
Prop = 6
AngMom 1
AngMom 2
AngMom 3
MLTPL1 1
MLTPL1 2
MLTPL1 3
SpinOrbit
NR  OF  JOBIPHS
2  10 40
1 2 3 4 5 6 7 8 9 10
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
    21 22 23 24 25 26 27 28 29 30 31 32 33 34 35
    36 37 38 39 40

&SINGLE_ANISO
MLTP
  8
  2 2 2 2 2 2 2 2
UBAR
TINT = 0 300 301
CRYS
Co =  2
HINT = 0 7.0 71
TMAG
  20  0.5 1.0 1.5 1.8 2.0 2.2 2.5 3.0 3.5 4.0 5.0
      6.0 7.0 8.0 9.0 10.0 15.0 20.0 30.0 40.0
XFIE =  0.1
PLOT

>>COPY ANISOINPUT  $CurrDir/$Project.aniso
```

The input provided above is a basic one. The following options need to be considered: a) larger basis set; b) larger active space; c) dynamic correlation e.g. by CASPT2. However, the above input runs smoothly within a few seconds on any machine and is a good example for testing and learning.

A few comments about the above input:

- SEWARD: The keywords AMFI and ANGM in the GATEWAY/SEWARD section are mandatory for the computation of spin-orbit coupling and magnetic properties. The origin of ANGM is usually taken the position of the main magnetic center

(because the magnetic f or d orbitals are mainly localized on the metal site).

- SEWARD: The keywords RICD helps speeding up much the computation of integrals and subsequent calculations and is therefore recommended.

- RASSCF: It is recommended to optimize first the states with the largest spin by using the minimal active space. Larger active space calculations may be carried out using the optimal orbitals obtained in a minimal active space.

- RASSCF: It is recommended to optimize all states which have the same origin (e.g. all states arising from $3d^7$ configuration in above example in a single state-average RASSCF calculation. Optimizing each of the 10 states of S=3/2 in ten separate calculations is prone to convergence issues and to less accurate final results, where the ligand-field energies are important. Charge transfer states need to be optimized separately from the ligand-field states, if one wishes to achieve high accuracy.)

- RASSI: It is recommended to include all roots originating from a given electron configuration d^n or f^n in the spin-orbit mixing. The above calculation indeed mixes all states of the $3d^7$ configuration.

- RASSI: Keyword SPIN and MEES are mandatory. The first one will enable the spin-orbit interaction, while the second one will ensure the computation of one-electron integrals over the RASSCF states. Keyword PROP and the subsequent labels request explicitly the computation of angular momentum (ANGMOM) and electric dipole momentum integrals (MLTPL1). AMFI integrals will be done by default. The keyword NROF allows selecting which RASSCF roots are included into the spin-orbit coupling calculation.

- SINGLE_ANISO: Keyword MLTP asks the computation of the g-tensor for the eight low-lying Kramers doublets. TINT and HINT specify the intervals for temperature and applied magnetic field, respectively. Keyword CRYS enables the computation of the parameters of the ligand-field for the ground atomic term ^{4}F of the Co^{2+} ion, using the *ab initio* results. Keyword XFIE allows computation of the magnetic susceptibility at the indicated field strength 0.1 Tesla. Keyword TMAG asks the code to compute molar magnetization at the indicated temperatures. Keyword PLOT will ensure generation of several data files (*.dat), several gnuplot scripts (*.plt) and will attempt running the linux utility `gnuplot` in order to produce corresponding images of the computed properties (*.png or *.eps). Keyword UBAR computes the blocking barrier of a molecular magnet.

Below is a standard input for computing the $DyCl_3$ molecule.

```
&SEWARD
ANGM
0.00000000 0.00000000 0.00000000 Angstrom
AMFI
DOUG
RICD
SDIP
CDTH = 1.0d-6

Coord
4

 Dy1  0.00000000 0.00000000 0.00000000
 Cl2 -2.23434554 1.29000000 0.00000000
 Cl3  2.23434554 1.29000000 0.00000000
 Cl4  0.00000000 -2.58000000 0.00000000
Group=NoSYMM
Basis=ANO-RCC-VDZ


&RASSCF
SPIN = 6
NACTEL = 9 0 0
INAC = 54
RAS2 = 7
CIROOT = 21 21 1

>>COPY $Project.JobIph JOB001
>>COPY $Project.RasOrb INPORB

&RASSCF
LUMORB
SPIN = 4
NACTEL = 9 0 0
INAC = 54
RAS2 = 7
CIROOT = 224 224 1

>>COPY $Project.JobIph JOB002

&RASSCF
LUMORB
SPIN = 2
NACTEL = 9 0 0
INAC = 54
RAS2 = 7
CIROOT = 490 490 1
End Of Input
>>COPY $Project.JobIph JOB003

/* We mix in RASSI only a subset of the CASSCF
    states available in the above JOBIPH files.
    The RASSI is able to mix all available spin
    states, but the computation time will increase
    quadratically with the number of states
    included in RASSI. */

&RASSI
MEES
MESO
Prop = 6
```

```
AngMom 1
AngMom 2
AngMom 3
MLTPL1 1
MLTPL1 2
MLTPL1 3
SpinOrbit
NR  OF  JOBIPHS
3 21 128 130
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
    21
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
    21 22 23 24 25 26 27 28 29 30 31 32 33 34 35
    36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
    51 52 53 54 55 56 57 58 59 60 61 62 63 64 65
    66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
    81 82 83 84 85 86 87 88 89 90 91 92 93 94 95
    96 97 98 99 100 101 102 103 104 105 106 107
    108 109 110 111 112 113 114 115 116 117 118
    119 120 121 122 123 124 125 126 127 128
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
    21 22 23 24 25 26 27 28 29 30 31 32 33 34 35
    36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
    51 52 53 54 55 56 57 58 59 60 61 62 63 64 65
    66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
    81 82 83 84 85 86 87 88 89 90 91 92 93 94 95
    96 97 98 99 100 101 102 103 104 105 106 107
    108 109 110 111 112 113 114 115 116 117 118
    119 120 121 122 123 124 125 126 127 128 129 130

&SINGLE_ANISO
MLTP
  8
  2 2 2 2 2 2 2 2
UBAR
TINT = 0 300 301
CRYS
Dy
HINT =  0 7.0 71
TMAG
  20  0.5 1.0 1.5 1.8 2.0 2.2 2.5 3.0 3.5 4.0 5.0
      6.0 7.0 8.0 9.0 10.0 15.0 20.0 30.0 40.0
XFIE =  0.1
PLOT

>>COPY ANISOINPUT  $CurrDir/$Project.aniso
```

S13. SEMI-AB INITIO APPROACH: FRAGMENT CALCULATIONS + POLY_ANISO – COMPUTATIONAL DETAILS

We give here a practical example how to apply fragment calculations + POLY_ANISO approach for the investigation of the magnetic properties of a model binuclear Co_2 compound. Expanding to larger nuclearity compounds is straightforward.

The original coordinates of the investigated molecule are given in Table S5. Figure S33 shows this structure.

The first step is to find the local electronic and magnetic properties originating from each Co^{2+} site. To this end, we prepare two fragments, both having exactly the same number of atoms as the original ion, with the exception that one of the Co sites is computationally replaced by diamagnetic (closed shell) Zn^{2+} ion.

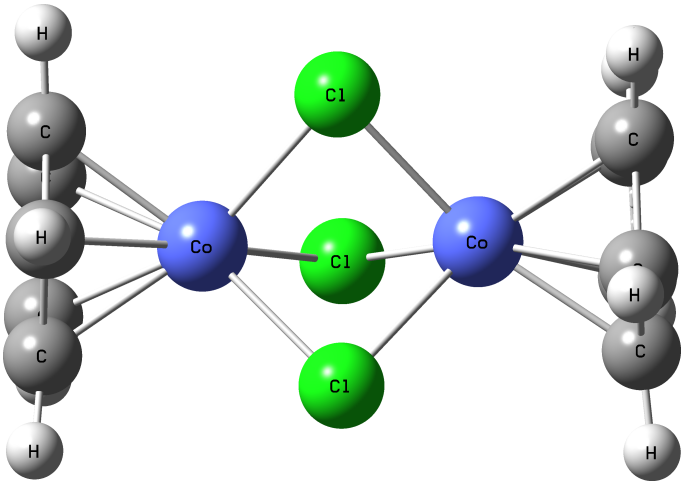

FIG. S33. Structure of the model dinuclear Co_2 anion. Atomic coordinates are listed in Table S5.

S13.A. Ab initio calculation for sites Co1 and Co2

Here we show the simplest *ab initio* calculation for a Co^{2+} compound, by employing only CASSCF and RASSI methods. The calculations shown here take roughly 3-5 min. on any laptop or workstation. Realistic calculations would require larger basis sets (at least VTZP quality), larger active space (e.g. CAS(7in10) for accounting the double shell on Co^{2+} ion, or larger, as discussed in other sections of this article), employing the CASPT2 method for accounting the dynamical correlation, including the 40 RASSCF states with S=1/2 into the spin-orbit mixing in RASSI.

```
/*************************************************
/*  CASSCF/RASSI on site 1 in Co2: fragment Co1-Zn2

&SEWARD
 ANGM = 0.0 0.0 0.0
 AMFI
 DOUG
 RICD
 SDIP
 CDTH = 1.0d-6

Coord
25

 Co1  4.09498600 12.65466600 5.12278000
 Zn2  2.81659800 10.06605400 5.12730000
 Cl3  3.47332600 11.35777800 6.85548500
 Cl4  2.09929800 12.03681800 4.21876000
 Cl5  4.79794000 10.67873800 4.18862600
 C6   3.91327000 14.57482500 5.72546000
```

```
 C7   5.07529600 13.97066100 6.34019400
 C8   5.92968000 13.55497500 5.34727800
 C9   5.34468200 13.90611300 4.05302300
 C10  4.10455000 14.53351500 4.27902800
 C11  2.81659800 8.21302400 4.29409500
 C12  3.06366800 8.15105800 5.68025900
 C13  1.94468000 8.65194700 6.37334100
 C14  1.02653600 9.07537800 5.34878500
 C15  1.57168400 8.79137000 4.10425100
 H16  1.80736716 8.70198995 7.43331312
 H17  0.10426562 9.57030937 5.57095674
 H18  1.08510667 8.92457607 3.16064163
 H19  3.42814710 7.80439337 3.51696650
 H20  3.94166838 7.71265073 6.10665600
 H21  6.85980907 13.04232975 5.47748841
 H22  5.25583944 13.90130108 7.39256906
 H23  3.07500736 15.02082667 6.21871272
 H24  3.44197584 14.96839946 3.56016015
 H25  5.79289037 13.70130007 3.10325412
Group=NoSymm
Basis=ANO-RCC-VDZ

&RASSCF
  SPIN  = 4
  NACTEL = 7 0 0
    INAC = 86
    RAS2 = 5
  CIROOT = 10 10 1

>>COPY $Project.JobIph JOB001
&RASSI
MEES
Prop =6
AngMom 1
AngMom 2
AngMom 3
MLTPL1 1
MLTPL1 2
MLTPL1 3
HEFF
NR  OF  JOBIPHS
1 10
1 2 3 4 5 6 7 8 9 10
SpinOrbit

&SINGLE_ANISO
MLTP
  4
   2  2  2  2
UBAR
TINT = 0 300 301
CRYS
Co =  2
QUAX = 3
 1.0 0.0 0.0
 0.0 1.0 0.0
 0.0 0.0 1.0
HINT =  0 7.0 71
TMAG =  3  0.5 1.0 2.2
MAVE =  1 20
XFIE =  0.1

>>COPY ANISOINPUT $CurrDir/site1.aniso
```

```
/************************************************
/*  CASSCF/RASSI on site 2 in Co2: fragment Zn1-Co2

&SEWARD
 ANGM = 0.0 0.0 0.0
 AMFI
 DOUG
 RICD
 SDIP
 CDTH = 1.0d-6

Coord
25

 Zn1  4.09498600 12.65466600 5.12278000
 Co2  2.81659800 10.06605400 5.12730000
 Cl3  3.47332600 11.35777800 6.85548500
 Cl4  2.09929800 12.03681800 4.21876000
 Cl5  4.79794000 10.67873800 4.18862600
 C6   3.91327000 14.57482500 5.72546000
 C7   5.07529600 13.97066100 6.34019400
 C8   5.92968000 13.55497500 5.34727800
 C9   5.34468200 13.90611300 4.05302300
 C10  4.10455000 14.53351500 4.27902800
 C11  2.81659800 8.21302400 4.29409500
 C12  3.06366800 8.15105800 5.68025900
 C13  1.94468000 8.65194700 6.37334100
 C14  1.02653600 9.07537800 5.34878500
 C15  1.57168400 8.79137000 4.10425100
 H16  1.80736716 8.70198995 7.43331312
 H17  0.10426562 9.57030937 5.57095674
 H18  1.08510667 8.92457607 3.16064163
 H19  3.42814710 7.80439337 3.51696650
 H20  3.94166838 7.71265073 6.10665600
 H21  6.85980907 13.04232975 5.47748841
 H22  5.25583944 13.90130108 7.39256906
 H23  3.07500736 15.02082667 6.21871272
 H24  3.44197584 14.96839946 3.56016015
 H25  5.79289037 13.70130007 3.10325412
Group=NoSYMM
Basis=ANO-RCC-VDZ

&RASSCF
  SPIN  = 4
  NACTEL = 7 0 0
    INAC = 86
    RAS2 = 5
  CIROOT = 10 10 1

>>COPY $Project.JobIph JOB001

&RASSI
MEES
Prop =  6
AngMom 1
AngMom 2
AngMom 3
MLTPL1 1
MLTPL1 2
MLTPL1 3
HEFF
NR OF JOBIPHS
1 10
```

```
1 2 3 4 5 6 7 8 9 10
SpinOrbit

&SINGLE_ANISO
MLTP
  4
   2  2  2  2
UBAR
TINT =  0 300 301
CRYS
Co =  2
QUAX = 3
 1.0 0.0 0.0
 0.0 1.0 0.0
 0.0 0.0 1.0
HINT =  0 7.0 71
TMAG =  3  0.5 1.0 2.2
MAVE = 1 20
XFIE =  0.1

>>COPY ANISOINPUT $CurrDir/site2.aniso
```

At the end of each calculation we obtain one file ANISOINPUT for each site which we save in our local folder $FileDir for later usage.

S13.B. Extraction of the Co1-Co2 magnetic exchange interaction using the available experimental magnetic data.

Semi-*ab initio* simulation of the coupled properties of fragment Co1-Co2 using variable J as a parameter.

```
/*****************************************
   copy the ab initio results obtained for
   each site in $WorkDir: */

>>COPY $CurrDir/site1.aniso aniso_1.input
>>COPY $CurrDir/site2.aniso aniso_2.input

/*****************************************/
// loop over an array of parameters
// grep the standard deviation (ST.DEV) from the
// final output
// J is a variable parameter defining the
// magnetic interaction.
// Values of J are defined in parenthesis.

>> FOREACH J in
    (-6.0,-5.0,-4.5,-4.0,-3.0,-2.0,-1.0,0.0)

&POLY_ANISO
 NNEQ
   2  T  F
   1  1
   2  2
 PAIR
 1
 1  2  $J
 COOR
 4.09498600 12.65466600 5.12278000
 2.81659800 10.06605400 5.12730000
 XFIE
 0.5
 TEXP
  24
  4.04  0.06089855
  8.08  0.58081879
 12.12  1.42846878
 16.16  2.19618547
 20.20  2.80632688
 24.24  3.28333320
 28.28  3.66148978
 32.32  3.96772238
 36.36  4.22112986
 40.40  4.43502587
 44.44  4.61875490
 48.48  4.77896930
 52.52  4.92048122
 56.56  5.04682795
 60.60  5.16065095
 64.64  5.26395372
 68.68  5.35827991
 72.72  5.44483782
 76.76  5.52458863
 80.80  5.59830933
 84.84  5.66663813
 88.88  5.73010735
 92.92  5.78916756
 96.96  5.84420526

>> ENDDO
```

TABLE S5. Cartesian coordinates of $[Co_2Cl_3(C_5H_5)_2]^-$ anion

Co	4.09498600	12.65466600	5.12278000
Co	2.81659800	10.06605400	5.12730000
Cl	3.47332600	11.35777800	6.85548500
Cl	2.09929800	12.03681800	4.21876000
Cl	4.79794000	10.67873800	4.18862600
C	3.91327000	14.57482500	5.72546000
C	5.07529600	13.97066100	6.34019400
C	5.92968000	13.55497500	5.34727800
C	5.34468200	13.90611300	4.05302300
C	4.10455000	14.53351500	4.27902800
C	2.81659800	8.21302400	4.29409500
C	3.06366800	8.15105800	5.68025900
C	1.94468000	8.65194700	6.37334100
C	1.02653600	9.07537800	5.34878500
C	1.57168400	8.79137000	4.10425100
H	1.80736716	8.70198995	7.43331312
H	0.10426562	9.57030937	5.57095674
H	1.08510667	8.92457607	3.16064163
H	3.42814710	7.80439337	3.51696650
H	3.94166838	7.71265073	6.10665600
H	6.85980907	13.04232975	5.47748841
H	5.25583944	13.90130108	7.39256906
H	3.07500736	15.02082667	6.21871272
H	3.44197584	14.96839946	3.56016015
H	5.79289037	13.70130007	3.10325412

REFERENCES

[1] R. Lindh, "The reduced multiplication scheme of the Rys–Gauss quadrature for 1st order integral derivatives," Theor. Chim. Acta **85**, 423–440 (1993).

[2] J. Stålring, A. Bernhardsson, and R. Lindh, "Analytical gradients of a state average mcscf state and a state average diagnostic," Mol. Phys. **99**, 103–114 (2001).

[3] I. Fdez. Galván, M. G. Delcey, T. B. Pedersen, F. Aquilante, and R. Lindh, "Analytical state-average complete-active-space self-consistent field nonadiabatic coupling vectors: Implementation with density-fitted two-electron integrals and application to conical intersections," J. Chem. Theory Comput. **12**, 3636–3653 (2016).

[4] P. Å. Malmqvist and B. O. Roos, "The CASSCF state interaction method," Chem. Phys. Letters **155**, 189–194 (1989).

[5] P.-Å. Malmqvist, K. Pierloot, A. R. M. Shahi, C. J. Cramer, and L. Gagliardi, "The restricted active space followed by second-order perturbation theory method: Theory and application to the study of CuO_2 and Cu_2O_2 systems," J. Chem. Phys. **128**, 204109 (2008).

[6] S. Wouters, W. Poelmans, P. W. Ayers, and D. Van Neck, "Chemps2: a free open-source spin-adapted implementation of the density matrix renormalization group for ab initio quantum chemistry," Comput. Phys. Commun. **185**, 1501–1514 (2014).

[7] S. Wouters and D. Van Neck, "The density matrix renormalization group for ab initio quantum chemistry," Eur. Phys. J. D **68**, 272 (2014).

[8] S. Wouters, T. Bogaerts, P. Van Der Voort, V. Van Speybroeck, and D. Van Neck, "Communication: Dmrg-scf study of the singlet, triplet, and quintet states of oxo-mn(salen)," J. Chem. Phys. **140**, 241103 (2014).

[9] S. Wouters, V. Van Speybroeck, and D. Van Neck, "Dmrg-caspt2 study of the longitudinal static second hyperpolarizability of all-trans polyenes," J. Chem. Phys. **145**, 054120 (2016).

[10] S. Keller, M. Dolfi, M. Troyer, and M. Reiher, "An efficient matrix product operator representation of the quantum-chemical hamiltonian," J. Chem. Phys. **143**, 244118 (2015).

[11] S. Keller and M. Reiher, "Spin-adapted matrix product states and operators," J. Chem. Phys. **144**, 134101 (2016).

[12] I. Schapiro, M. N. Ryazantsev, L. M. Frutos, N. Ferré, R. Lindh, and M. Olivucci, "The Ultrafast Photoisomerizations of Rhodopsin and Bathorhodopsin Are Modulated by Bond Length Alternation and HOOP Driven Electronic Effects," J. Am. Chem. Soc. **133**, 3354–3364 (2011).

[13] F. Aquilante, J. Autschbach, R. K. Carlson, L. F. Chibotaru, M. G. Delcey, L. De Vico, I. Fdez. Galván, N. Ferré, L. M. Frutos, L. Gagliardi, M. Garavelli, A. Giussani, C. E. Hoyer, G. Li Manni, H. Lischka, D. Ma, P. Å. Malmqvist, T. Müller, A. Nenov, M. Olivucci, T. B. Pedersen, D. Peng, F. Plasser, B. Pritchard, M. Reiher, I. Rivalta, I. Schapiro, J. Segarra-Martí, M. Stenrup, D. G. Truhlar, L. Ungur, A. Valentini, S. Vancoillie, V. Veryazov, V. P. Vysotskiy, O. Weingart, F. Zapata, and R. Lindh, "Molcas 8: New Capabilities for Multiconfigurational Quantum Chemical Calculations Across the Periodic Table," J. Comput. Chem. **37**, 506–541 (2016).

[14] S. Knecht, S. Keller, J. Autschbach, and M. Reiher, "A nonorthogonal state-interaction approach for matrix product state wave functions," J. Chem. Theory Comput. **12**, 5881–5894 (2016).

[15] C. Angeli, R. Cimiraglia, S. Evangelisti, T. Leininger, and J.-P. Malrieu, "Introduction of n-electron valence states for multireference perturbation theory," J. Chem. Phys. **114**, 10252 (2001).

[16] C. Angeli, R. Cimiraglia, and J.-P. Malrieu, "N-electron valence state perturbation theory: A fast implementation of the strongly contracted variant," Chem. Phys. Lett. **350**, 297–305 (2001).

[17] C. Angeli, R. Cimiraglia, and J.-P. Malrieu, "N-electron valence state perturbation theory: A spinless formulation and an efficient implementation of the strongly contracted and of the partially contracted variants," J. Chem. Phys. **117**, 9138–9153 (2002).

[18] C. Angeli, S. Borini, M. Cestari, and R. Cimiraglia, "A quasidegenerate formulation of the second order n-electron valence state perturbation theory approach," J. Chem. Phys. **121**, 4043–4049 (2004).

[19] L. Freitag, S. Knecht, C. Angeli, and M. Reiher, "Multireference perturbation theory with Cholesky decomposition for the density matrix renormalization group," J. Chem. Theory Comput. **13**, 451–459 (2017).

[20] L. Ungur and L. F. Chibotaru, "Computational modelling of the magnetic properties of lanthanide compounds," in *Lanthanides and Actinides in Molecular Magnetism* (John Wiley and Sons, Ltd, 2015) Chap. 6, pp. 153–184.

[21] L. F. Chibotaru, L. Ungur, and A. Soncini, "The origin of nonmagnetic kramers doublets in the ground state of dysprosium triangles: Evidence for a toroidal magnetic moment," Angewandte Chemie International Edition **47**, 4126–4129 (2008).

[22] B. O. Roos, "The complete active space self-consistent field method and its applications in electronic structure calculations," in *Advances in Chemical Physics; Ab Initio Methods in Quantum Chemistry - II*, edited by K. P. Lawley (John Wiley & Sons Ltd., Chichester, England, 1987) Chap. 69, p. 399.

[23] P.-Å. Malmqvist, A. Rendell, and B. O. Roos, "The restricted active space self-consistent-field method, implemented with a split graph unitary group approach," J. Phys. Chem. **94**, 5477–5482 (1990).

[24] P.-Å. Malmqvist, "Calculation of transformation density matrices by nonunitary orbital transformations," Int. J. Quantum Chem. **30**, 479 (1986).

[25] P.-Å. Malmqvist, B. O. Roos, and B. Schimmelpfennig, "The Restricted Active Space (RAS) State Interaction Approach with Spin-Orbit Coupling," Chem. Phys. Letters **357**, 230–240 (2002).

[26] R. Lindh, U. Ryu, and B. Liu, "The reduced multiplication scheme of the rys quadrature and new recurrence relations for auxiliary function based two-electron integral evaluation," The Journal of chemical physics **95**, 5889–5897 (1991).

[27] T. B. Pedersen, F. Aquilante, and R. Lindh, "Density fitting with auxiliary basis sets from cholesky decompositions," Theor. Chem. Acc. **124**, 1–10 (2009).

[28] F. Aquilante, M. G. Delcey, T. B. Pedersen, I. Fdez. Galván, and R. Lindh, "Inner projection techniques for the low-cost handling of two-electron integrals in quantum chemistry," Mol. Phys. **115**, 2052–2064 (2017).

[29] L. F. Chibotaru and L. Ungur, "Ab initio calculation of anisotropic magnetic properties of complexes. I. Unique definition of pseudospin Hamiltonians and their derivation," Journal of Chemical Physics **137**, 064112 (2012).

[30] L. Ungur and L. F. Chibotaru, "Ab initio crystal field for lanthanides," Chemistry – A European Journal **23**, 3708–3718 (2017).

[31] E. Besalú and J. M. Bofill, "On the automatic restricted-step rational-function-optimization method," Theor. Chem. Acc. **100**, 265–274 (1998).

[32] R. Lindh, A. Bernhardsson, G. Karlström, and P.-Å. Malmqvist, "On the use of a Hessian model function in molecular geometry optimizations," Chem. Phys. Letters **241**, 423–428 (1995).

[33] R. Lindh, A. Bernhardsson, and M. Schütz, "Force-constant weighted redundant coordinates in molecular geometry optimizations," Chem. Phys. Letters **303**, 567–575 (1999).

[34] L. De Vico, M. Olivucci, and R. Lindh, "New General Tools for Constrained Geometry Optimizations," J. Chem. Theory Comput. **1**, 1029–1037 (2005).

[35] M. Stenrup, R. Lindh, and I. Fdez. Galván, "Constrained numerical gradients and composite gradients: Practical tools for geometry optimization and potential energy surface navigation," J. Comput. Chem. **36**, 1698–1708 (2015).

[36] V. May and O. Kühn, *Charge and Energy Transfer Dynamics in Molecular Systems* (Wiley-VCH, 2011).

[37] A. Anda, T. Hansen, and L. De Vico, "Multireference Excita-

tion Energies for Bacteriochlorophylls A within Light Harvesting System 2," J. Chem. Theory Comput. **12**, 1305–1313 (2016).
[38]V. Le Guilloux, P. Schmidtke, and P. Tuffery, "Fpocket: An open source platform for ligand pocket detection," Bioinformatics **10**, 168–179 (2009).
[39]G. G. Krivov, M. V. Shapovalov, and R. L. Dunbrack, "Improved prediction of protein side-chain conformations with scwrl4," Proteins: Struct., Funct., Bioinf. **77**, 778–795 (2009).
[40]M. H. Olsson, C. R. Søndergaard, M. Rostkowski, and J. H. Jensen, "Propka3: Consistent treatment of internal and surface residues in empirical pk_a predictions," J. Chem. Theory Comput. **7**, 525–537 (2011).
[41]L. Zhang and J. Hermans, "Hydrophilicity of cavities in proteins," Proteins: Structure, Function, and Bioinformatics **24**, 433–438 (1996).
[42]S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. Van Der Spoel, B. Hess, and E. Lindahl, "Gromacs 4.5: A high-throughput and highly parallel open source molecular simulation toolkit," Bioinformatics **29**, 845–854 (2013).
[43]L. Pedraza-González, L. De Vico, M. d. C. Marín, F. Fanelli, and M. Olivucci, "*a*-ARM: Automatic Rhodopsin Modeling with Chromophore Cavity Generation, Ionization State Selection and External Counter-ion Placement," J. Chem. Theory Comput. **15**, 3134–3152 (2019).
[44]A. D. Becke, "Density-functional exchange-energy approximation with correct asymptotic behavior," Phys. Rev. A **38**, 3098 (1988).
[45]J. P. Perdew and Y. Wang, Phys. Rev. B **33**, 8822 (1986).
[46]M. Dolg, U. Wedig, H. Stoll, and H. Preuss, J. Chem. Phys. **86**, 866 (1987).
[47]R. Ahlrichs, M. Bar, M. Häser, H. Horn, and C. Kolmel, "Electronic structure calculations on workstation computers: The program system turbomole," Chem. Phys. Lett. **162**, 165–169 (1989).
[48]Q. M. Phung and K. Pierloot, "Electronic structure of N-bridged high-valent Diiron-Oxo," Chem. Eur. J. **25**, 12491–12496 (2019).
[49]B. O. Roos, R. Lindh, P.-Å. Malmqvist, V. Veryazov, and P.-O. Widmark, "New relativistic ano basis sets for transition metal atoms," J. Phys. Chem. A **109**, 6575–6579 (2005).
[50]B. O. Roos, R. Lindh, P.-A. Malmqvist, V. Veryazov, and P.-O. Widmark, "Main group atoms and dimers studied with a new relativistic ANO basis set," J. Phys. Chem. A **108**, 2851 (2004).
[51]P.-O. Widmark, P.-Å. Malmqvist, and B. O. Roos, Theor. Chem. Acc. **77**, 291 (1990).
[52]F. Aquilante, P.-Å. Malmqvist, T. B. Pedersen, A. Ghosh, and B. O. Roos, "Cholesky decomposition-based multiconfiguration second-order perturbation theory (cd-caspt2): Application to the spin-state energetics of co-iii(diiminato)(nph)," J. Chem. Theory Comput. **4**, 694 (2008).
[53]B. A. Hess, "Relativistic electronic-structure calculations employing a 2-component no-pair formalism with external-field projection operators," Phys. Rev. A **33**, 3742–3748 (1986).
[54]M. Reiher and A. Wolf, "Exact decoupling of the Dirac Hamiltonian. I. General Theory," J. Chem. Phys. **121**, 2037–2047 (2004).
[55]M. Reiher and A. Wolf, "Exact decoupling of the Dirac Hamiltonian. II. The generalized Douglas–Kroll–Hess transformation up to arbitrary order," J. Chem. Phys. **121**, 10945–10956 (2004).
[56]G. Ghigo, B. O. Roos, and P.-Å. Malmqvist, "A modified definition of the zeroth-order hamiltonian in multiconfigurational perturbation theory (caspt2)," Chem. Phys. Lett. **396**, 142–149 (2004).
[57]N. Forsberg and P.-Å. Malmqvist, "Multiconfiguration perturbation theory with imaginary level shift," Chem. Phys.Lett. **274**, 196 – 204 (1997).
[58]M. Feldt, Q. M. Phung, K. Pierloot, R. A. Mata, and J. N. Harvey, "Limits of coupled-cluster calculations for non-heme iron complexes," J. Chem. Theory. Comput. **15**, 922–937 (2019).
[59]F. Neese, "Software update: the orca program system, version 4.0," Wiley Interdiscip. Rev. Comput. Mol. Sci. **8**, e1327 (2018).
[60]J. Joos, P. S. Smet, L. Seijo, and Z. Barandiarán, "Insights on the complexity of the excited states of Eu-doped luminescent materials," Inorg. Chem. Front. **submitted** (2019).
[61]Z. Barandiarán and L. Seijo, "Performing ab initio calculations on complex manifolds of excited states of lanthanides in solids," (2016), `https://sites.google.com/site/howtodoabinitioln/`, Last accessed on 2019-12-11.
[62]W. Küchle, M. Dolg, H. Stoll, and H. Preuss, "Energyadjusted pseudopotentials for the actinides. parameter sets and test calculations for thorium and thorium monoxide," J. Chem. Phys. **100**, 7535 (1994).
[63]K. Eichkorn, F. Weigend, O. Treutler, and R. Ahlrichs, "Auxiliary basis sets for main row atoms and transition metals and their use to approximate Coulomb potentials," Theor. Chem. Acc. **97**, 119–124 (1997).
[64]L. Ungur, "Introduction to the electronic structure, luminescence, and magnetism of lanthanides," in *Lanthanide-Based Multifunctional Materials*, Advanced Nanomaterials, edited by P. Martin-Ramos and M. R. Silva (Elsevier, 2018) pp. 1 – 58.
[65]G. Kovačević and V. Veryazov, "Luscus: molecular viewer and editor for MOLCAS," Journal of Cheminformatics **7**, 16 (2015).