

**Basis sets,  
RI and Cholesky methods in MOLCAS**

Roland Lindh  
Dept. of Theoretical Chemistry  
Lund University

# Outline

- Basis sets, some points
- Theoretical background of RI and CD
- Unification
- Accuracy Performance
- RI and CD options in MOLCAS
- Some examples performed with MOLCAS

# Basis sets and auxiliary basis sets

Basis sets (orbital basis):

- kinetic energy,*
- nuclear-electron attraction,*
- electron-electron repulsion*
- etc.*

Auxiliary basis sets (product basis):

- electron-electron repulsion*

# Basis Sets

- Segmented basis sets (STO-3G, 6-31G, etc.)
- Generally contracted basis sets (ANO-type)
- Cartesian vs Real Spherical Harmonics (6-31G family)
- Effective Core Potentials (ECP)
- Pseudo Potentials (PP)

The right basis for the right method!!

# Note on basis sets

- EMSL Basis Set Exchange
- Be careful with the basis sets in relativistic calculations!
- Cartesian or Spherical Harmonic?

The screenshot shows the EMSL Basis Set Exchange website. At the top left is the EMSL Office of Science logo. In the center is the 'BASIS SET EXCHANGE' logo. On the top right, there is a login section with a username field containing 'RolandLindh' and a password field with six asterisks. Below the login fields are buttons for 'Login' and 'Become a Contributor'. A navigation bar below the logo contains 'Basis Set Exchange: v1.2.2' and links for 'Feedback', 'About', 'ReleaseNotes', and 'Help'. On the left side, there is a search box with a dropdown menu set to 'All' and a list of basis set names including '3-21++G', '3-21++G\*', '3-21G', '3-21G\*', '3-21G\* Polarisation', '3-21GSP', '4-22GSP', '4-31G', '6-31++G', '6-31++G\*', '6-31++G\*\*', '6-31+G\*', '6-311++G(2d,2p)', '6-311++G(3d4f,3p4d)', and '6-311++G\*\*'. Below the search box is a 'Search Basis Set Name' input field. In the center, it says 'Total: 376 published basis sets' and displays a periodic table of elements. At the bottom of the periodic table, there is a 'Format:' dropdown set to 'NWChem' and a checked checkbox for 'Optimized General Contractions'. A 'Get Basis Set' button is located below the periodic table.

**Summary:**  
**Primary Developer:**  
**Last Modified:**

**"3-21++G" Basis Set Information**  
VDZD Valence Double Zeta + Diffuse Functions on All Atoms  
N/A  
Mon, 15 Jan 2007 23:47:08 GMT

**Contributor:**  
**Curation Status:**

Dr. David Feller  
published

[More information...](#)  
[User annotations...](#)

When publishing results obtained from use of the Basis Set Exchange (BSE) software and the EMSL Basis Set Library, please cite:

The Role of Databases in Support of Computational Chemistry Calculations  
Feller, D., J. Comp. Chem., 17(13), 1571-1586, 1996.

Basis Set Exchange: A Community Database for Computational Sciences  
Schuchardt, K.L., Didier, B.T., Elsethagen, T., Sun, L., Gurunoorathi, V., Chase, J., Li, J., and Windus, T.L.  
J. Chem. Inf. Model., 47(3), 1045-1052, 2007, doi:10.1021/d600510j.

# The basis set label says it all!

*The Syntax:* Element.Type.Author.Prim.Con.Aux.

Examples (general contraction):

- C.ANO-RCC.Roos.14s9p4d3f2g.8s7p4d3f2g.
- C.ANO-RCC...3s2p1d.
- Generics: C.ANO-RCC-VDZP, H.ANO-RCC-MB,...

Example (segmented contraction):

- C.6-31G..10s4p.3s2p.
- C.6-31G....

# Density Fitting and the Resolution of Identity

The DF method the procedure is of fitting of a “density” with an *external predefined* auxiliary basis.

$$\phi_i \phi_j \approx \widetilde{\phi_i \phi_j} = \sum_K C_{ij}^K \phi_K$$

The fitting coefficients are optimized by minimizing the error of

$$\left( \phi_i \phi_j - \widetilde{\phi_i \phi_j} \middle| \frac{1}{|r_1 - r_2|} \middle| \phi_i \phi_j - \widetilde{\phi_i \phi_j} \right)$$

The final equation expressing the fitting coefficients is

$$C_{ij}^K = \sum_L \langle ij|L \rangle V_{LK}^{-1}$$

and the 2-electron integrals can be expressed as

$$\langle ij|kl \rangle = \sum_{KL} \langle ij|K \rangle V_{KL}^{-1} \langle L|kl \rangle$$



# DF/RI specifics

- “Superfast” Coulombic contributions (e.g. pure DFT in Turbomole)
- Slow Exchange contribution evaluation
- Use of an external predefined 1-center auxiliary basis
- Fixed approximation, accuracy not controlled by single parameter!
- Auxiliary basis small but has to be tailored for each new method and valence basis set!
- Standard error in total energies below 0.01 kcal/mol/atom
- Standard method for “large” calculations (.e.g. in biochemistry)
- analytic RI gradients and Hessians are available

# Cholesky decomposition

Beebe and Linderberg suggested to use the so-called Cholesky decomposition to approximate the two-electron integrals

$$V = LL^T$$



By controlling the accuracy of the decomposition by a single parameter the so-called Cholesky-vectors have a rank which is significantly reduced as compared to the super-matrix, dito computational effort.

In an iterative procedure these recursive formulae are used to derive the Cholesky-vectors

$$L_{JJ} = \left( V_{JJ} - \sum_{K=1}^{J-1} L_{JK}^2 \right)^{1/2}$$

and

$$L_{IJ} = \left( V_{IJ} - \sum_{K=1}^{J-1} L_{IK} L_{JK} \right) / L_{JJ}$$

a single parameter is used to control the accuracy!

# CD specifics

- Considerably reduced computational time as compared to conventional methods
- Accuracy controlled via single parameter
- Accuracy is independent of wf-method
- Methods has been developed for the use of CD in HF, DFT (pure and hybrid), MP2, CC, CASSCF and CASPT2.
- Used typically with an error in the total energy of  $1.0D-6$  -  $1.0D-8$  hartree.
- Discontinious PES and **no** analytic gradients
- Fast Exchange evaluation has been developed (LK approach)

# Unification

Observations:

- both are methods to approximate 2-electron integrals

Beebe and Linderberg are actually rather explicit. They point out that “the CD is equivalent to a particular form of an inner projection procedure”. In this sense it is related to the RI/DF techniques.

To see this take the working equation of the RI/DF approach

$$\langle ij|kl\rangle = \sum_{KL} \langle ij|K\rangle V_{KL}^{-1} \langle L|kl\rangle$$

now transform, via a Gram-Schmidt procedure, the auxiliary basis to an orthogonal basis which will diagonalize the inverse matrix

$$\langle ij|kl\rangle = \sum_{K'} \langle ij|K'\rangle V_{K'K'}^{-1} \langle K'|kl\rangle$$

This finally yield that the Cholesky vectors are expressed as

$$L_{ij,K'} = \langle ij | K' \rangle V_{K'K'}^{-1/2}$$

and Cholesky-like vectors can be derived for the RI/DF approach as

$$R_{ij,K} = \sum_L \langle ij | L \rangle V_{LK}^{-1/2}$$

# Notes

- Both CD and RI/DF are approximations of the complete RI operators
- CD is a numerical approximation based on the full product basis (1- and 2-center auxiliary functions).
- RI/DF approximates by using a fixed externally predefined 1-center auxiliary basis, normally method dependent.
- 2C-CD can not have analytic gradients since with a fixed threshold the auxiliary basis is a function of the geometry.



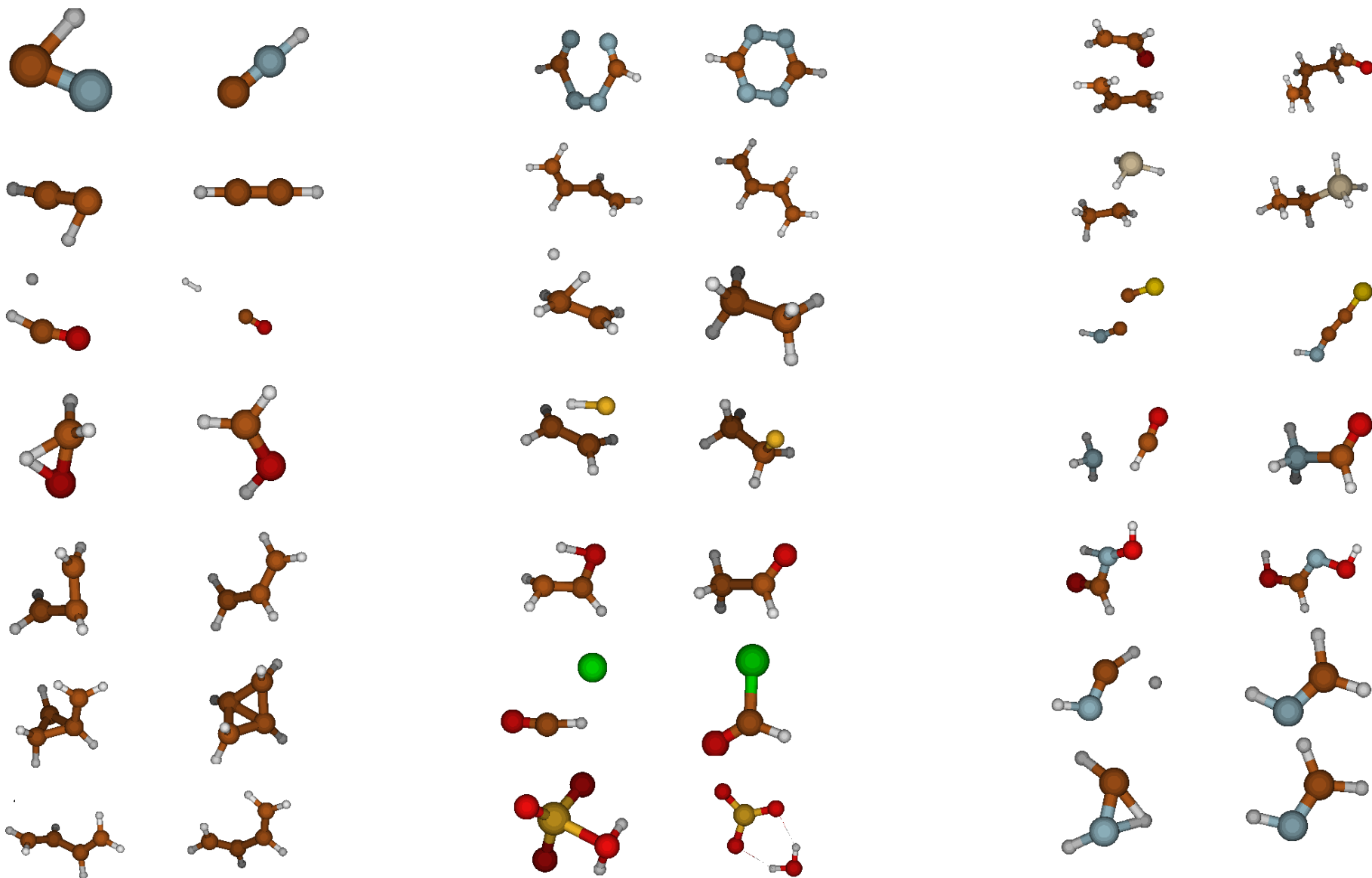
## **What can CD learn from DF/RI?: the 1-center approximation!**

DF/RI has a (rather small) fixed 1-center auxiliary basis. This is key for analytic gradients to be possible.

Could we devise an approximation of the CD in which only the 1-center products are used in the procedure? YES, 1C-CD, but is it accurate?

Use aCD/acCD auxiliary basis sets for general RI/DF calculations.

# Test set: 21 reactions, B3LYP 6-31G structures



# 1-Center CD

vs.

## full CD (SVWN)

basis set /dec. thr.	Type	Total SVWN5 DFT energy		Activation energy	
		RMS error (kcal/mol)	Max. error (kcal/mol)	RMS error (kcal/mol)	Max error (kcal/mol)
SVP	RI-J	0.2879	0.5036	0.0561	0.1439
10 <sup>-4</sup>	1-center	0.2075	0.5425	0.0891	-0.2337
	full	0.1355	0.2597	0.0333	0.0899
10 <sup>-5</sup>	1-center	0.0328	0.0685	0.0140	-0.0350
	full	0.0317	0.0574	0.0147	-0.0364
10 <sup>-6</sup>	1-center	0.0112	0.0307	0.0075	-0.0229
	full	0.0025	0.0076	0.0014	-0.0038

Statistics over 20 reactions.

# 1-Center CD

vs.

## full CD (B3LYP)

basis set /dec. thr.	Type	Total B3LYP DFT energy		Activation energy	
		RMS error (kcal/mol)	Max. error (kcal/mol)	RMS error (kcal/mol)	Max error (kcal/mol)
SVP	RJ-J*	193.03	-491.40	1.4703	4.776
10 <sup>-4</sup>	1-center	0.1807	0.4912	0.0910	-0.2364
	full	0.1152	0.2142	0.0349	-0.0877
10 <sup>-5</sup>	1-center	0.0185	0.0391	0.0120	-0.0273
	full	0.0195	0.0527	0.0148	-0.0351
10 <sup>-6</sup>	1-center	0.0085	-0.0297	0.0071	-0.0203
	full	0.0020	0.0065	0.0014	-0.0038

\*The auxiliary basis is not optimized for hybrid functionals.  
 Statistics over 20 reactions.

# 1-Center CD

vs.

## full CD (MP2)

basis set /dec. thr.	Type	Total MP2 energy		Activation energy	
		RMS error (kcal/mol)	Max. error (kcal/mol)	RMS error (kcal/mol)	Max error (kcal/mol)
SVP	RI-J*				
10 <sup>-4</sup>	1-center	0.0969	-0.2513	0.0914	-0.2068
	full	0.0632	-0.1540	0.0370	-0.0926
10 <sup>-5</sup>	1-center	0.0880	-0.2073	0.0172	0.0292
	full	0.0132	-0.0198	0.0137	-0.0266
10 <sup>-6</sup>	1-center	0.0850	-0.1765	0.0187	0.0382
	full	0.0024	-0.0074	0.0011	-0.0028

Statistics over 17 reactions.

# Observations

- 1C-CD does not degrade the accuracy significantly!
- in the 1C-CD approximation a fixed auxiliary basis set is used, hence we can compute analytic derivatives!
- 1C-CD the decomposition time is 3-4 times faster than the full CD with the same threshold.

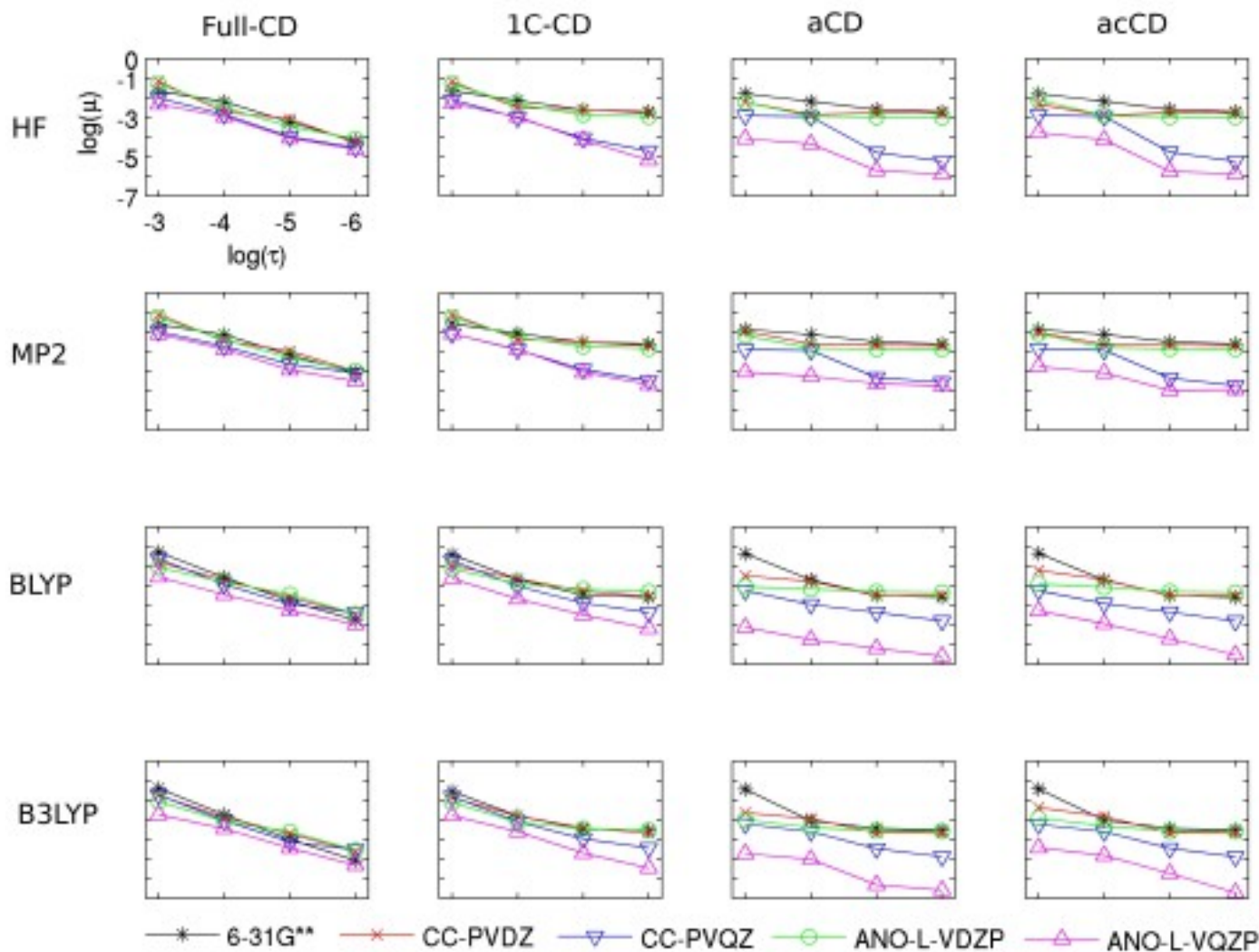
# **What can DF/RI learn from CD?: the Cholesky auxiliary basis sets!**

Given the accuracy of the 1C-CD approach, could 1C-CD be used to design general DF/RI auxiliary basis sets which are method-free?

To investigate this we designed some DF/RI basis sets by using the 1-center product auxiliary basis used by the 1C-CD approach.

The RI/aCD(acCD) approach.

# Accuracy assessment / G2 test suite





# Summary / RI&CD

- the DF/RI, CD, and 1C-CD methods are closely related.
- 1C-CD approximation is equivalent in performance and accuracy to DF/RI
- 1C-CD approach has analytic derivatives
- aCD/acCD approach can be used to derive “method-free” RI auxiliary basis sets
- The better valence basis set the smaller relative size of the aCD auxiliary basis set

# New Implementations

- RI/aCD(acCD) on-the-fly basis set
- Unified code for RI/DF and CD post integral processing.

# MOLCAS keyword

Input to use in the &Seward input section to activate Cholesky and RI options

- **Cholesky**
- **Cholnp; 1-Center; EndCho**
- **RICD; aCD or acCD**
- **RIJ**
- **CDThreshold**

These options will provide significant computational speed up with insignificant loss of accuracy!

# MP2 example

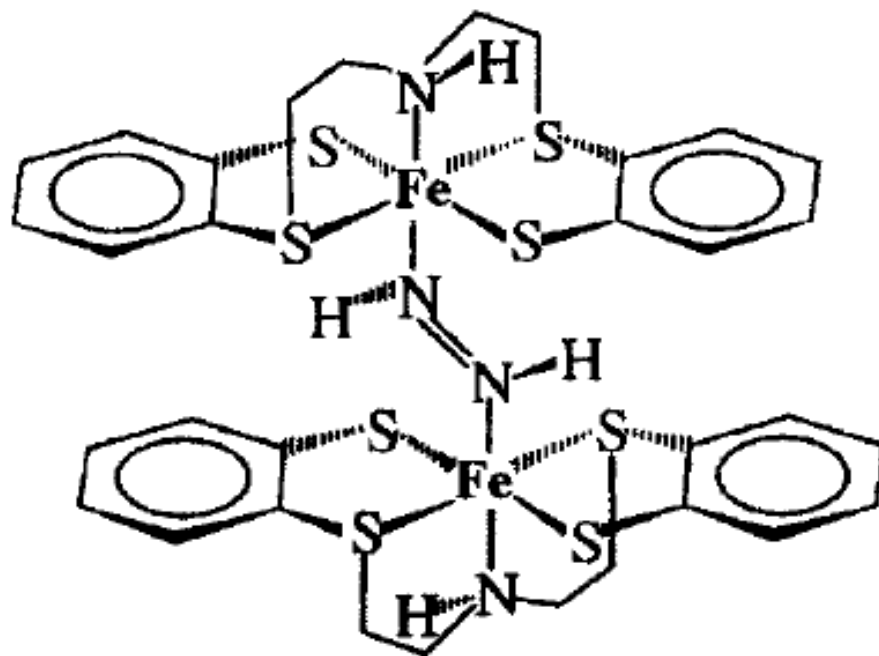


FIG. 6. Schematic structure of the dinuclear diazene iron complex used as sample application for the LK scheme.

MP2/cc-pVTZ 1924 bfn

# CASPT2 example

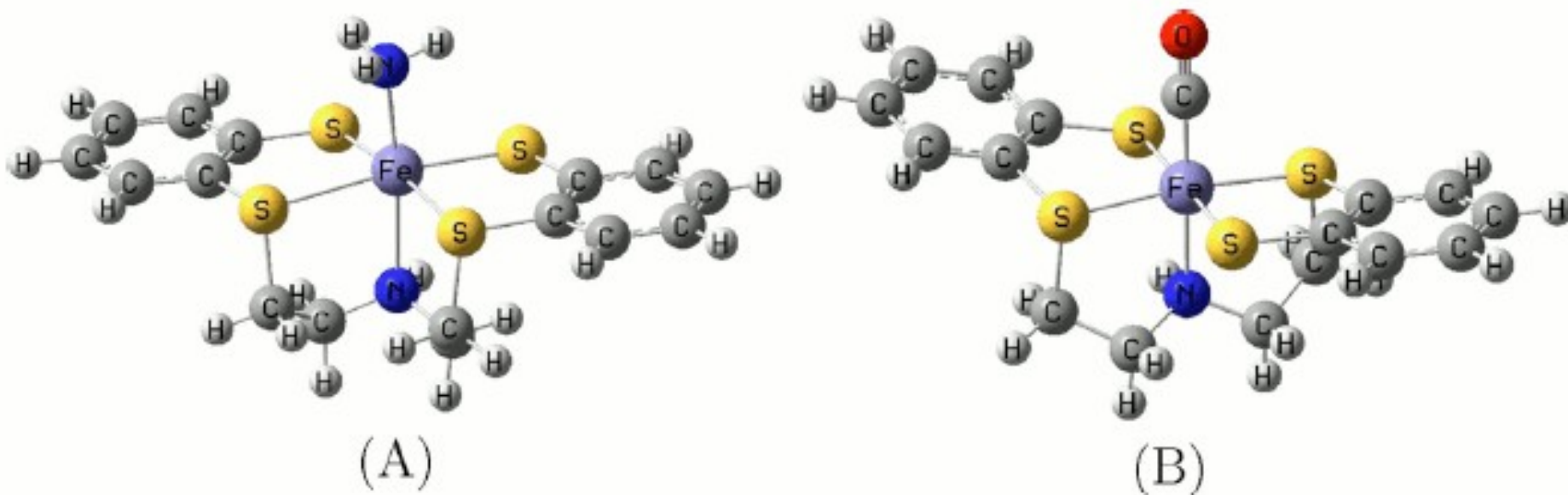


FIG. 1. (Color online) Examples of  $[\text{Fe}(L)(\text{NHS}_4)]$  complexes. (A) The *cis*- $\text{NH}_3$  complex. (B) the *trans*-CO complex.

- CD-CASPT2/CASSCF(14-in-16) (810 bfn no symmetry, 964  $\text{C}_2$  symmetry)