Lecture 3: Quantum Mechanics

Junmei Wang

Department of Pharmacology, University of Texas Southwestern Medical Center at Dallas

Junmei.wang@utsouthwestern.edu

Project 1: Drug Lead Identification

1. Select a protein system
2. Perform 2D-similarity search
    - Download a sdf file of the ligand from RCSB Protein Databank (http://www.rcsb.org/pdb/home/home.do)
    - Similarity searches with OpenBabel
3. Prepare necessary files for molecular docking with AutoDock vina
    - Ligand.pdbqt
    - Receptor.pdbqt
    - Configure.txt

Protein Class	PDB Code	$-\log k_d$	Resolution
Neuraminidase	2QWG	8.4	1.8
DHFR	1DHF	7.4	2.3
L-arabinose	1ABE	6.52	1.7
Thrombin	1A5G	10.15	2.06
Human oxresin receptor 1	4ZJ8	~10	2.75

Project 1: Drug Lead Identification- Continued

4. Docking screening
    - Submit jobs from BioHPC

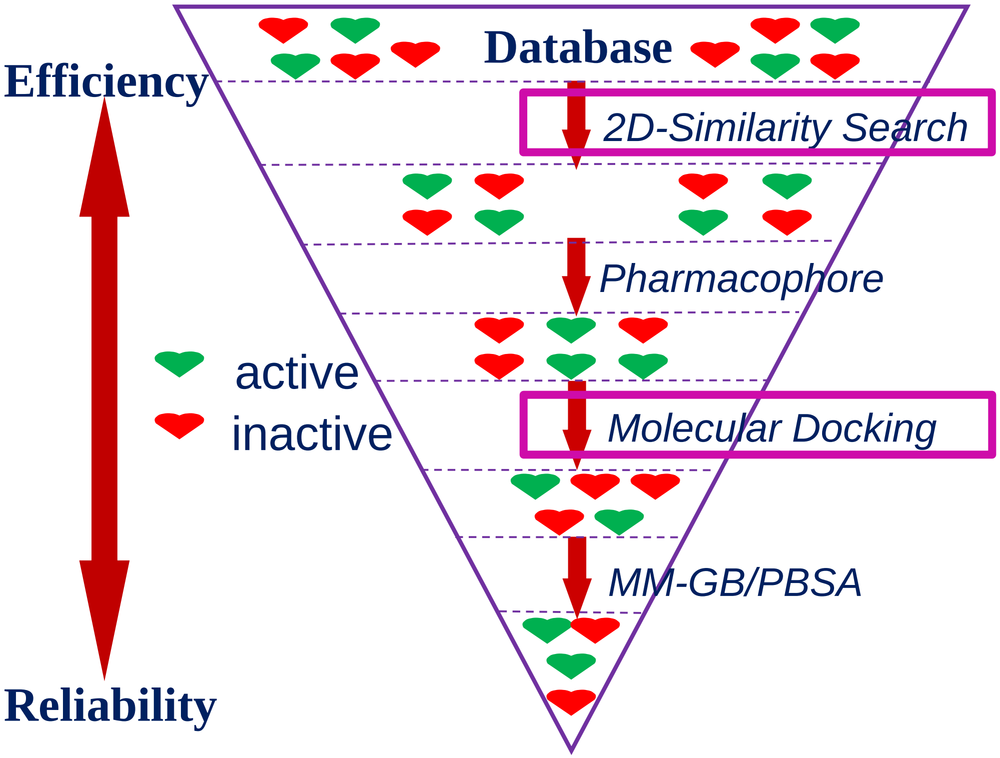
    Login: ssh –Y jwang@nucleus.biohpc.swmed.edu

    sftp jwang@nucleus.biohpc.swmed.edu

5. Analyze docking results
    - Prioritize compounds based on docking scores
    - Examine docking poses with PyMOL

6. Write report
    - Hit rate (HR)
    - Enrichment factor (EF)
    - Enrichment curve

Lead Identification Through Virtual Screening Using A Set of Hierarchical Filters



$$HR = \frac{m}{M}$$

$$EF = HR \times \frac{N}{n} = \frac{mN}{Mn}$$

HR : hit rate

EF : enrichment factor

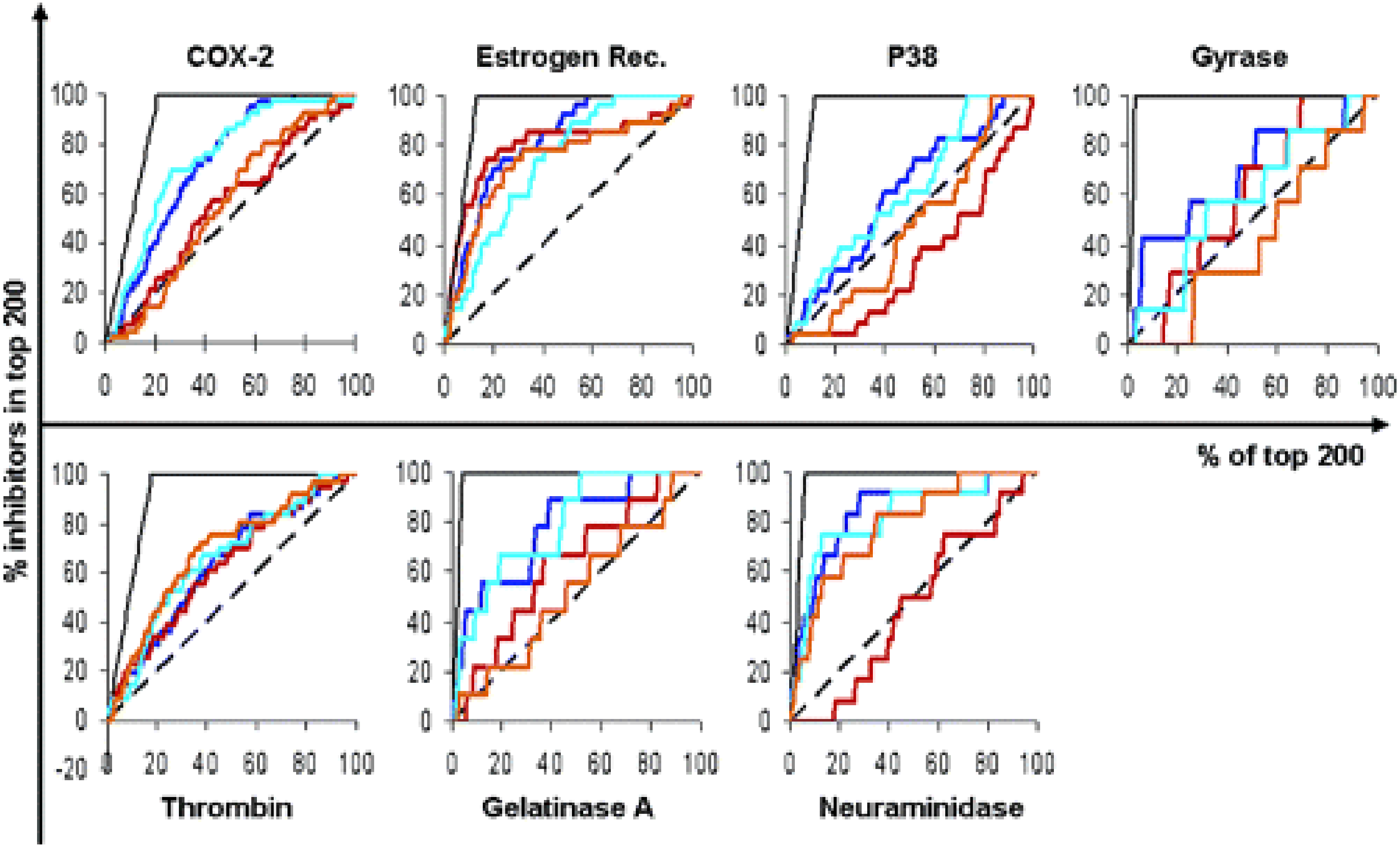
N – total number of molecules in the library

n – number of hits

M – total number of known inhibitors

m – number of known inhibitors recognized as hits

Enrichment Curves



random selection (black, dashed)
ideal performance (black, solid)performance

Example: Autodock

- Autodock uses pre-calculated affinity maps for each atom type in the substrate molecule, usually C, N, O and H, plus an electrostatic map
- These grids include energetic contributions from all the usual sources

$$\begin{aligned} \Delta G \quad &= \quad G_1 \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + G_2 \sum_{i,j} \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} + E_{hbond} \right) \\ &+ \quad G_3 \sum_{i,j} \frac{q_i q_j}{\epsilon(r) r_{ij}} + G_4 \Delta G_{tor} + G_5 \sum_{i,j} S_i V_j e^{-r_{ij}^2/2\sigma^2} \end{aligned}$$

Stouten Pairwise Atomic Solvation Parameters

Favorable for C, A ; Unfavorable for O, N

Proportional to the absolute value of partial charges

AutoDock Vina Scoring Function

Combination of knowledge-based and empirical approach

$$\Delta G_{binding} = \Delta G_{gauss} + \Delta G_{repulsion} + \Delta G_{hbond} + \Delta G_{hydrophobic} + \Delta G_{tors}$$

ΔG_{gauss}

Attractive term for dispersion, two gaussian functions

$\Delta G_{repulsion}$

Square of the distance if closer than a threshold value

ΔG_{hbond}

Ramp function - also used for interactions with metal ions

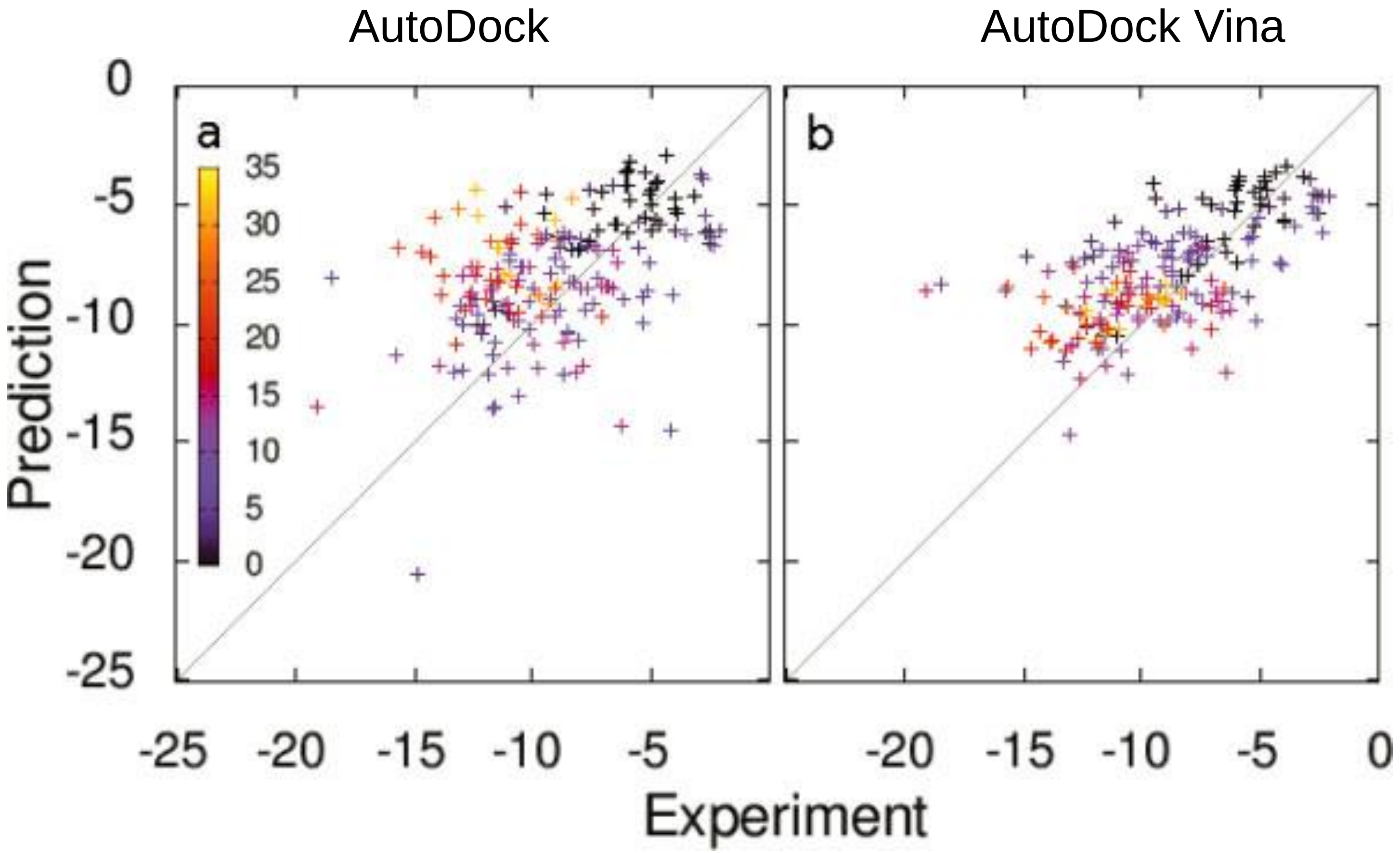
$\Delta G_{hydrophobic}$

Ramp function

ΔG_{tors}

Proportional to the number of rotatable bonds

Performance of Two Scoring Functions



AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading, J. Comput. Chem., 31, 455-461, 2009

Introduction to Computational Quantum Chemistry

Quantum Mechanics

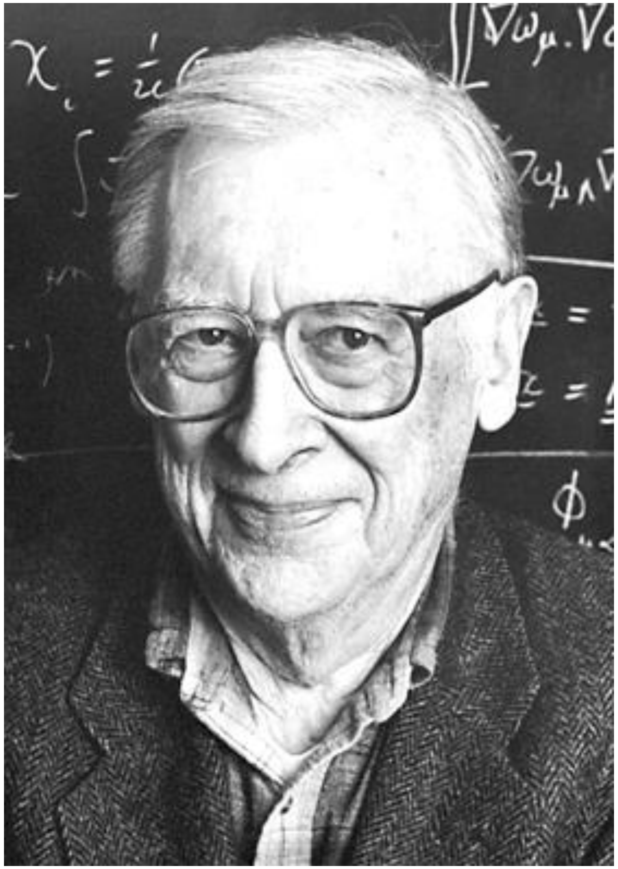
- **Quantum Mechanics**: $\hat{H}\Psi = E\,\Psi$

The Solvay Conference 1927

Quantum Mechanics

Nobel Prize of Chemistry Winner 1998

Walter Kohn



John Pople

The 1998 Nobel Prize in Chemistry was awarded to Walter Kohn “for his development of the density functional theory” and John Pople “for his development of computational methods in quantum chemistry”.

What’s it Good For?

- Computational quantum chemistry is a rapidly growing field in chemistry.
  - Computers are getting faster.
  - Algorithims and programs are maturing.
- Some of the almost limitless properties that can be calculated with computational chemistry are:
  - Equilibrium and transition-state structures
  - dipole and quadrapole moments and polarizabilities
  - Vibrational frequencies, IR and Raman Spectra
  - NMR spectra
  - Electronic excitations and UV spectra
  - Reaction rates and cross sections
  - thermochemical data

Motivation

- Schrödinger Equation can only be solved exactly for simple systems.
  - Rigid Rotor, Harmonic Oscillator, Particle in a Box, Hydrogen Atom
- For more complex systems (i.e. many electron atoms/molecules) we need to make some simplifying assumptions/approximations and solve it numerically.
- However, it is still possible to get very accurate results (and also get very crummy results).
  - In general, the "cost" of the calculation increases with the accuracy of the calculation and the size of the system.

Getting into the theory...

- Three parts to solving the Schrödinger equation for molecules:
  - Born-Oppenheimer Approximation
    - Leads to the idea of a potential energy surface
  - The expansion of the many-electron wave function in terms of Slater determinants.
    - Often called the "Method"
  - Representation of Slater determinants by molecular orbitals, which are linear combinations of atomic-like-orbital functions.
    - The basis set

The Born-Oppenheimer Approximation

Time Independent Schrödinger Equation

- We’ll be solving the Time-Independent Schrödinger Equation

$$\hat{H}\psi = E\psi$$

Hamiltonian operator

$$\hat{H} = \hat{T} + \hat{V}$$

Your book writes this as: $\hat{H} = \hat{E}_{kinetic} + \hat{E}_{potential}$

For Many electron atoms/molecules:

$$\hat{H} = \underbrace{-\frac{\hbar^2}{2}\sum_{\alpha}\frac{1}{m_\alpha}\nabla_\alpha^2 - \frac{\hbar^2}{2m_e}\sum_{i}\nabla_i^2}_{\hat{T}} + \underbrace{\sum_{\alpha}\sum_{\alpha>\beta}\frac{Z_\alpha Z_\beta e'^2}{r_{\alpha\beta}} - \sum_{\alpha}\sum_{i}\frac{Z_\alpha e'^2}{r_{i\alpha}} + \sum_{j}\sum_{i>j}\frac{e'^2}{r_{ij}}}_{\hat{V}}$$

Nuclei kinetic energy

Electron kinetic energy

Nuclear-Nuclear repulsion

Nuclear-electron attraction

Electron-electron repulsion

$$\hat{T}\quad\left(\frac{-\hbar^2}{2m}\frac{d^2}{dx^2}\right)$$

Laplacian

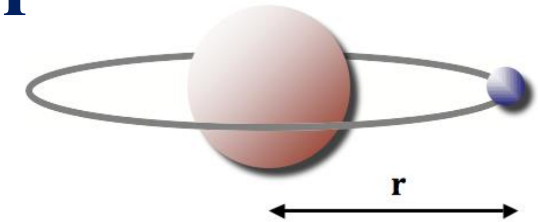
$\hat{V}$

where: $\nabla_i^2 \equiv {}^{\partial^2}\!/_{\partial x^2} + {}^{\partial^2}\!/_{\partial y^2} + {}^{\partial^2}\!/_{\partial z^2}$

Hydrogen-Like Atom

Schrodinger equation and solution of hydrogen-like atoms

The Born-Oppenheimer Approximation



- The wave-function of the many-electron molecule is a function of electron and nuclear coordinates: $\psi(R,r)$ (R=nuclear coords, r=electron coords).
- The motions of the electrons and nuclei are coupled.
- However, the nuclei are much heavier than the electrons
  - $m_p \approx 2000\ m_e$
- And consequently nuclei move ***much*** more slowly than do the electrons ($E=1/2mv^2$). To the electrons the nuclei appear fixed.
- Born-Oppenheimer Approximation: to a high degree of accuracy we can separate electron and nuclear motion:

$$\psi(R,r) = \psi_{el}(r;R)\ \psi_N(R)$$

Electronic Schrödinger Equation

- Now we can solve the electronic part of the Schrödinger equation separately.

$$\hat{H}_{el}\psi_{el}(r;R) = E_{el}\psi_{el}(r;R)$$

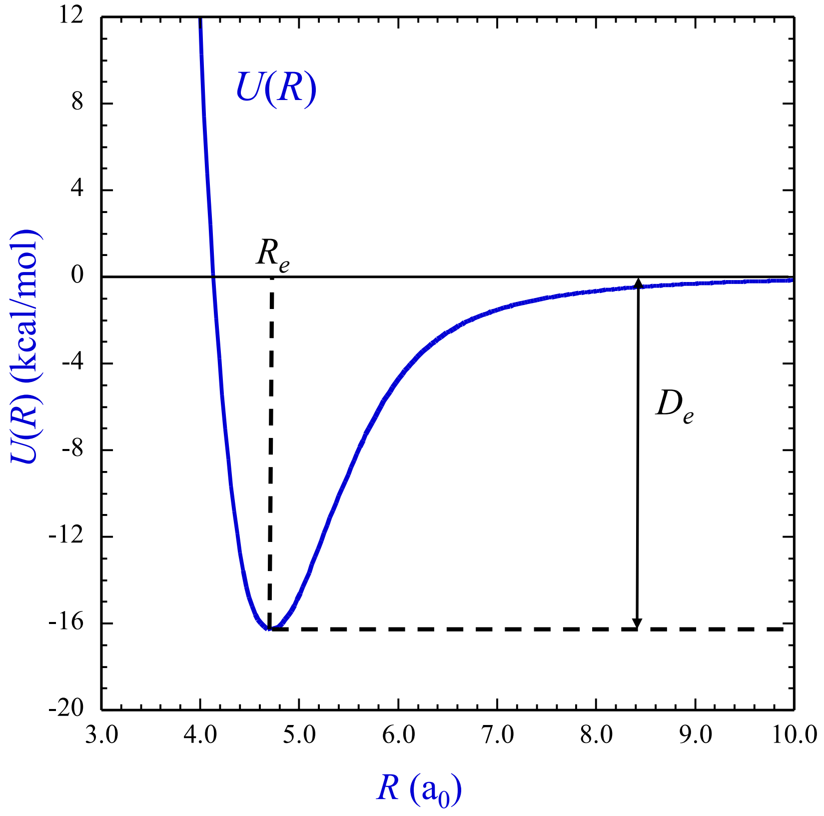
$$\hat{H}_{el} = -\frac{\hbar^2}{2m_e}\sum_i \nabla_i^2 - \sum_\alpha \sum_i \frac{Z_\alpha e'^2}{r_{i\alpha}} + \sum_j \sum_{i>j} \frac{e'^2}{r_{ij}}$$

- BO approximation leads to the idea of a potential energy surface.

$$U(R) = E_{el} + V_{NN}$$

$$V_{NN} = \sum_\alpha \sum_{\alpha>\beta} \frac{Z_\alpha Z_\beta e'^2}{r_{\alpha\beta}}$$

Diatomic Potential Energy Surface (HgBr)



Atomic unit of length
1 bohr = 1 a_0 = 0.529177 Å

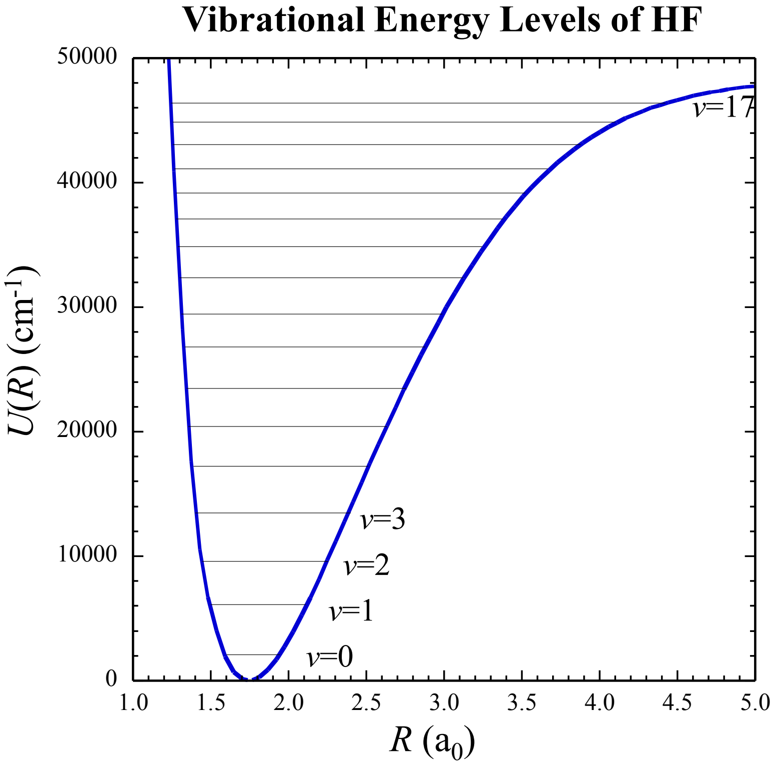
Nuclear Schrödinger Equation

- Once we have the Potential Energy Surface (PES) we can solve the nuclear Schrödinger equation.

$$\hat{H}_N \psi_N(R) = E_N \psi_N(R)$$

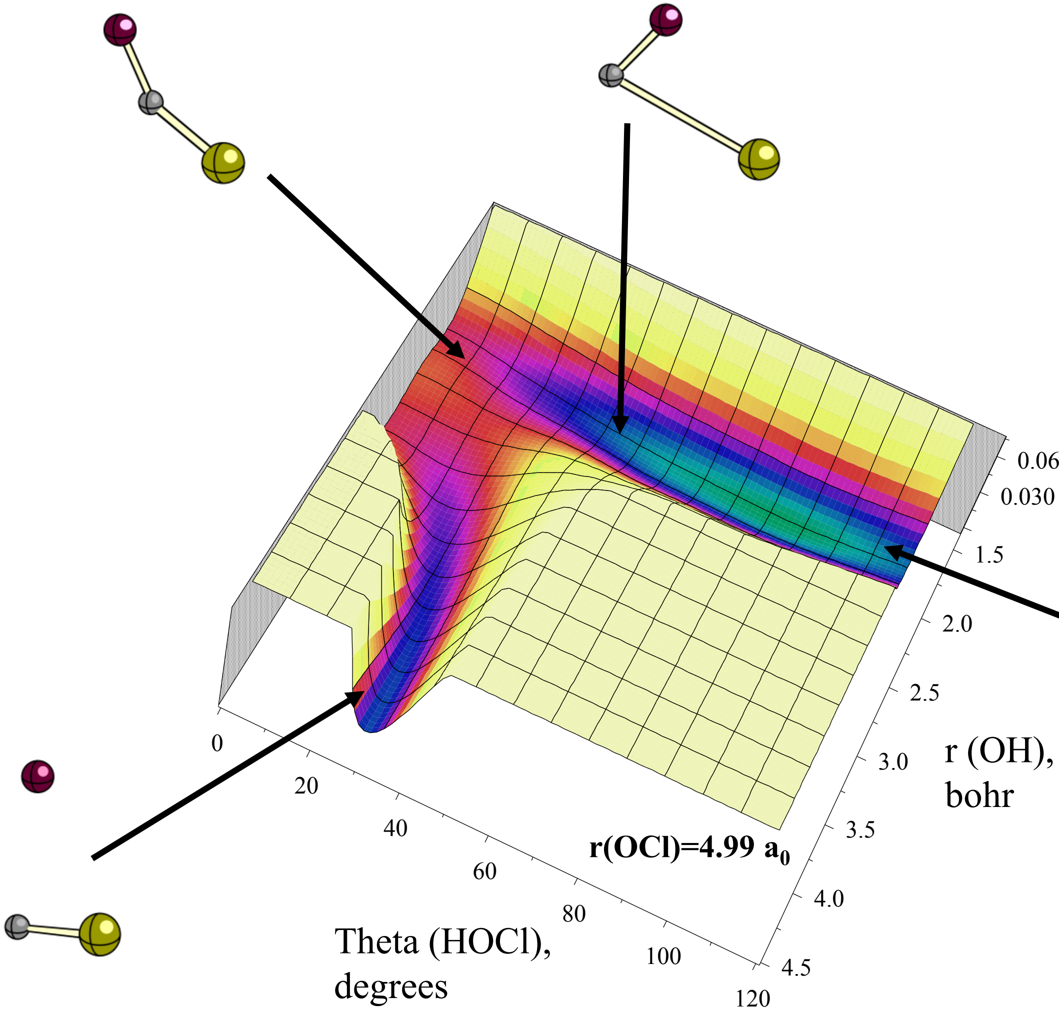
$$\hat{H}_N = -\frac{\hbar^2}{2}\sum_{\alpha}\frac{1}{m_\alpha}\nabla_\alpha^2 + U(R)$$

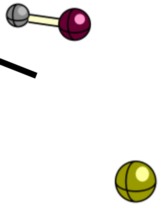
- Solution of the nuclear SE allow us to determine a large variety of molecular properties. An example are vibrational energy levels.

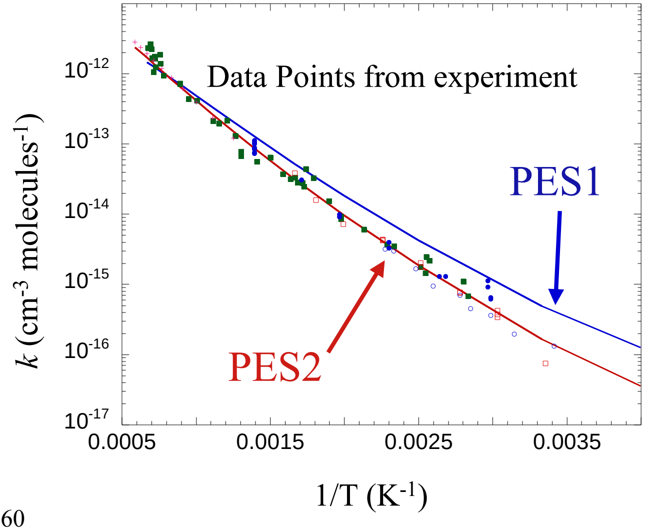


Polyatomic Potential Energy Surfaces

O + HCl → OH + Cl



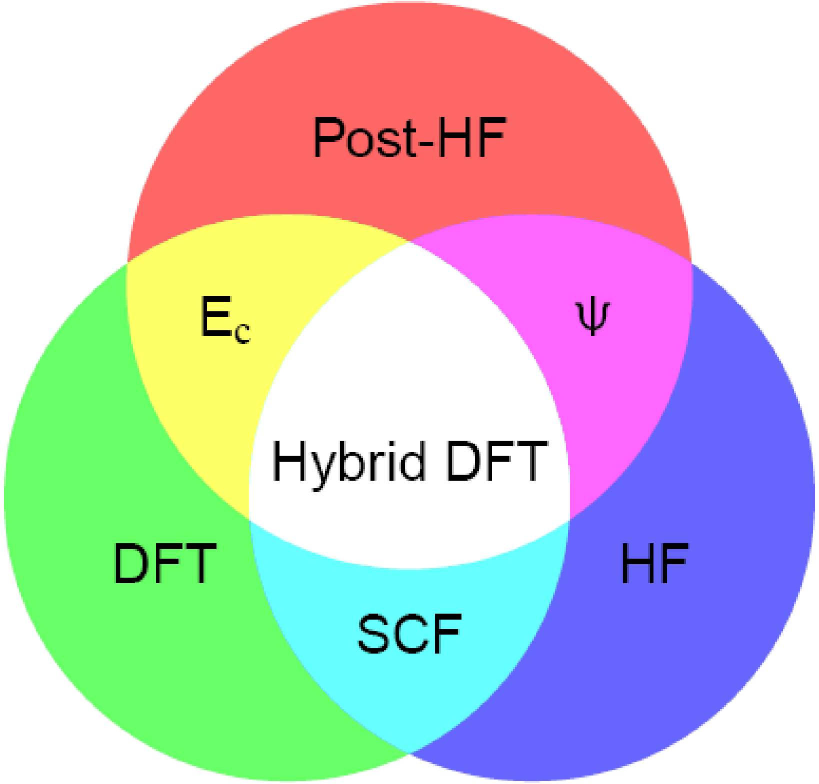




Thermal Rate Coefficients

- We can only look at cuts/slices
  - 3n-6 degrees of freedom
- Minima and Transition states
- Minimum energy path
  - Like following a stream-bed

The Method



So how do we solve Electronic S.E.?

- For systems involving more than 1 electron, still isn't possible to solve it exactly.

$$\hat{H}_{el} = -\frac{\hbar^2}{2m_e}\sum_i \nabla_i^2 - \sum_\alpha \sum_i \frac{Z_\alpha e'^2}{r_{i\alpha}} + \boxed{\sum_j \sum_{i>j} \frac{e'^2}{r_{ij}}}$$

  - The electron-electron interaction is the culprit

Approximating ψ: *The Method*

- After the B-O approximation, the next important approximation is the expansion of ψ in a basis of Slater determinants:

$$\psi_{el} = \sum_i d_i \Phi_i = d_0 \Phi_0 + d_1 \Phi_1 + d_2 \Phi_2 + \ldots$$

- Slater Determinant:

$$\Phi_0 = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1 \alpha(1) & \phi_1 \beta(1) & \phi_2 \alpha(1) \ldots & \phi_M \beta(1) \\ \phi_1 \alpha(2) & \phi_1 \beta(2) & \phi_2 \alpha(2) \ldots & \phi_M \beta(2) \\ \ldots & \ldots & \ldots & \ldots \\ \phi_1 \alpha(N) & \phi_1 \beta(N) & \phi_2 \alpha(N) \ldots & \phi_M \beta(N) \end{vmatrix}$$

  - α/β are spin-functions (spin-up/spin-down)
  - ϕ_i are spatial functions (molecular orbitals
  - ϕ_i α and ϕ_i β are called spin-orbitals
  - Slater determinant gives proper anti-symmetry (Pauli Principle)

Hartree-Fock Approximation

- Think of Slater determinants as configurations.
  - Ex: Neon:
    - Ground-state electron configuration $1s^2 2s^2 2p^6$ this would be Φ_0
    - Φ_1 might be $1s^2 2s^2 2p^5 3s^1$
- If we had a complete set of Φ_i's the expansion would be exact (not feasible).
- **Hartree-Fock (HF) Approximation**: Use 1 determinant, Φ_0.
  - A variational method (energy for approximate ψ will always be higher than energy of the true ψ)
  - Uses **self-consistent field** (SCF) procedure
  - Finds the optimal set of molecular orbitals for Φ_0
  - Each electron only sees average repulsion of the remaining electrons (no instantaneous interactions).

Accuracy of Hartree-Fock Calculations

- Hartree-Fock wavefunctions typically recover ~99% of the total electronic energy.
  - total energy of O-atom ≈ -75.00 E_h (1 Hartree= 1 E_h = 2625.5 kJ/mol).
  - 1 % of total energy is 0.7500 E_h or ~1969 kJ/mol
  - With more electrons this gets worse. Total energy of S atom ≈ -472.88 E_h (1% of energy is 12415 kJ/mol)
- Fortunately for the Hartree-Fock method (and all Quantum Chemists) chemistry is primarily interested in energy differences, not total energies. Hartree-Fock calculations usually provide at least qualitative accuracy in this respect.
  - Bond lengths, bond angles, vibrational force constants, thermochemistry, ... can generally be predicted qualitatively with HF theory.

Spectroscopic Constants of CO (Total E_e≈-300,000 kJ/mol)

	R_e (Å)	ω_e (cm^{-1})	D_e (KJ/mol)
HF/cc-pV6Z	1.10	2427	185
Experiment	1.13	2170	260
% Error	2.7%	11.8%	28.8%

Electron Correlation

- Electron Correlation: Difference between energy calculated with exact wave-function and energy from using Hartree-Fock wavefunction.

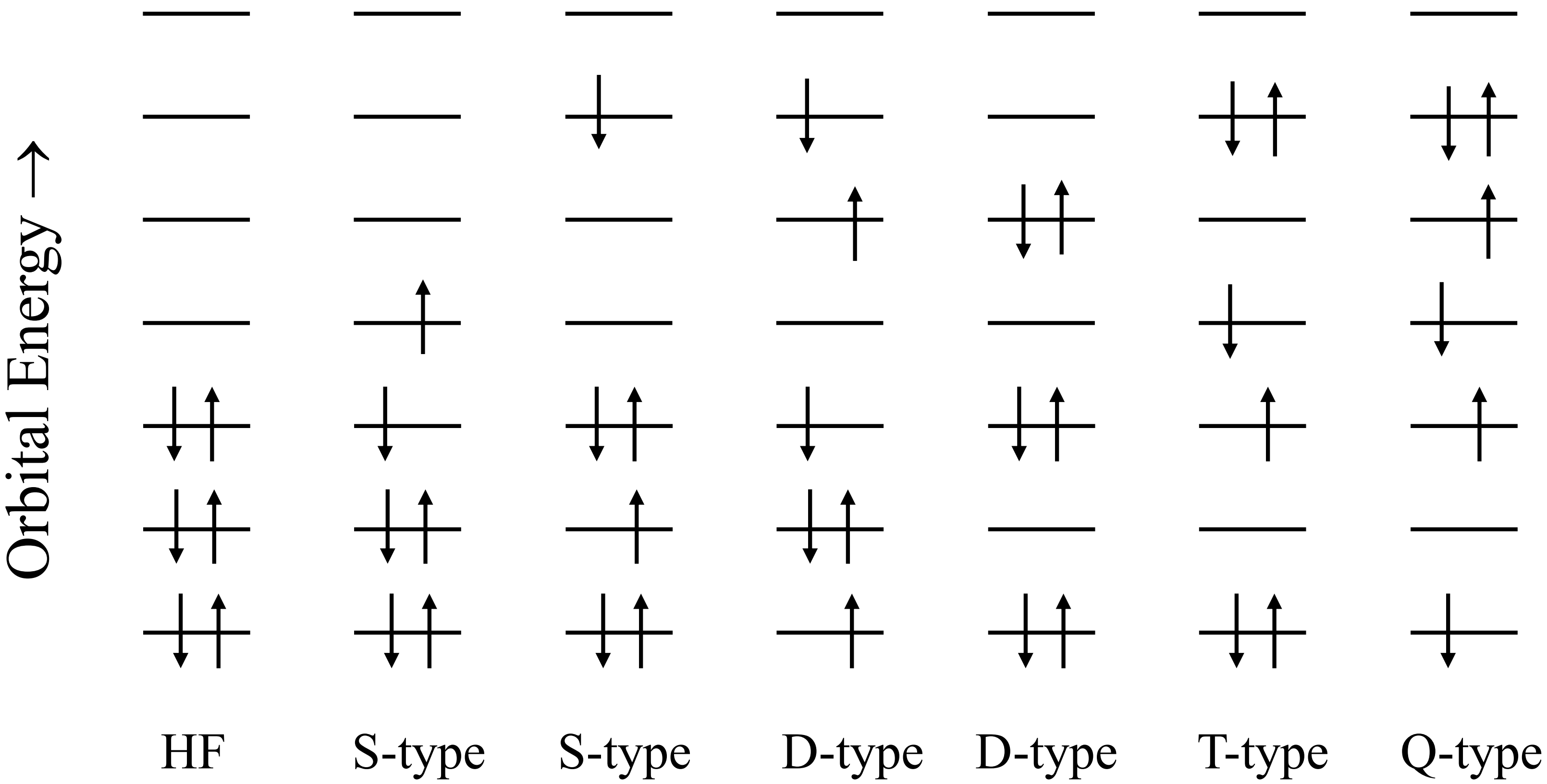
$$E_{corr} = E_{exact} - E_{HF}$$

- Accounts for the neglect of instantaneous electron-electron interactions of Hartree-Fock method.
- In general, we get correlation energy by adding additional Slater determinants to our expansion of ψ.

$$\psi_{el} = d_0 \Phi_{HF} + \sum_{i=1} d_i \Phi_i$$

- Hartree-Fock wavefunction is often used as our starting point.
- Additional Slater determinants are often called “excited.”
  - Mental picture of orbitals and electron configurations must be abandoned.
- Different correlation methods differ in how they choose which Φ_i to include and in how they calculate the coefficients, d_i.

Excited Slater Determinants



Configuration Interaction

- Write ψ as a linear combination of Slater Determinants and calculate the expansion coefficients such that the energy is minimized.

$$\psi_{el} = d_0\Phi_{HF} + \sum_{i=1} d_i\Phi_i$$

- Linear variational principle: *no matter what wave function is used, the energy is always equal to or greater than the true energy.*
- If we include all excited Φ_i we will have a full-CI, and an exact solution for the given basis set we are using.
- Full-CI calculations are generally not computationally feasible, so we must truncate the number of Φ_i in some way.
- CISD: Configuration interaction with single- and double-excitations.
  - Include all determinants of S- and D- type.
- MRCI: Multi-reference configuration interaction
- CI methods can be very accurate, but require long (and therefore expensive) expansions.
  - hundreds of thousands, millions, or more

Møller-Plesset Perturbation Theory

- Perturbation methods, like Møller-Plesset (MP) perturbation theory, assume that the problem we'd like to solve (correlated ψ and E) differ only slightly from a problem we've already solved (HF ψ and E).
- The energy is calculated to various orders of approximation.
  - Second order MP2; Third order MP3; Fourth order MP4...
  - Computational cost increases strongly with each successive order.
  - At infinite order the energy should be equal to the exact solution of the S.E. (for the given basis set). However, there is no guarantee the series is actually convergent.
  - In general only MP2 is recommended
- MP2 ~ including all single and double excitations

Coupled Cluster (CC) Theory

- An exponential operator is used in constructing the expansion of determinants.
- Leads to accurate and compact wave function expansions yielding accurate electronic energies.
- Common Variants:
  - CCSD: singles and doubles CC
  - CCSD(T): CCSD with approximate treatment of triple excitations. This method, when used with large basis sets, can generally provide highly accurate results. With this method, it is often possible to get thermochemistry within chemical accuracy, 1 kcal/mol (4.184 kJ/mol)

Frozen Core Approximation

- In general, only the valence orbitals are involved in chemical bonding.
- The core orbitals don't change much when atoms are involved in molecules than when the atoms are free.
- So, most electronic structure calculations only correlate the valence electrons. The *core* orbitals are kept *frozen*.
  - i.e., 2*s* and 2*p* electrons of Oxygen would be correlated, and the 1*s* electrons would not be correlated.

Density Functional Theory

- The methods we've been discussing can be grouped together under the heading "Wave function methods."
  - They all calculate energies/properties by calculating/improving upon the wave function.
- Density Functional Theory (DFT) instead solves for the electron density.
  - Generally computational cost is similar to the cost of HF calculations.
  - Most DFT methods involve some empirical parameterization.
  - Generally lacks the systematics that characterize wave function methods.
  - Often the best choice when dealing with very large molecules (proteins, large organic molecules...)

Basis Set

Basis Set Approximation: LCAO-MO

$$\Phi_0 = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1\alpha(1) & \phi_1\beta(1) & \phi_2\alpha(1)\ldots & \phi_M\beta(1) \\ \phi_1\alpha(2) & \phi_1\beta(2) & \phi_2\alpha(2)\ldots & \phi_M\beta(2) \\ \cdots & \cdots & \cdots & \cdots \\ \phi_1\alpha(N) & \phi_1\beta(N) & \phi_2\alpha(N)\ldots & \phi_M\beta(N) \end{vmatrix}$$

- Slater determinants are built from molecular orbitals, but how do we define these orbitals?
- We do another expansion: Linear Combination of Atomic Orbitals-Molecular Orbitals (LCAO-MO)

$$\phi_i = \sum_k^M c_{ki}\chi_k$$

  - Molecular orbital coefs, c_{ki}, determined in SCF procedure
  - The basis functions, χ_i, are atom-centered functions that mimic solutions of the H-atom (*s* orbitals, *p* orbitals,...)
- The larger the expansion the more accurate and expensive the calculations become.

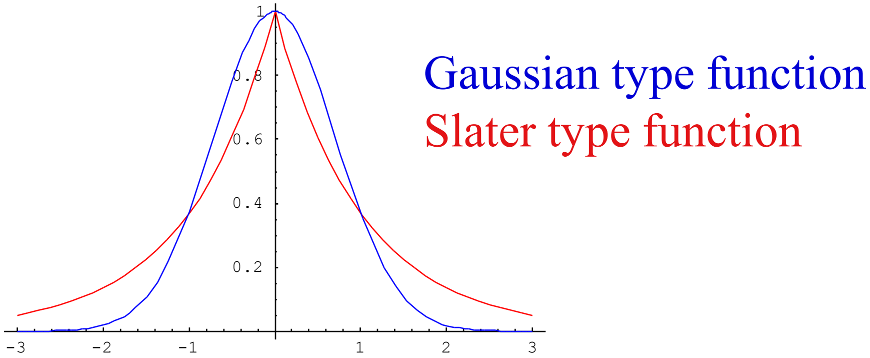
Gaussian Type Orbitals

- The radial dependence of the H-atom solutions are Slater type functions:

$$\chi_{\alpha,n,l,m}(r,\theta,\varphi) = N Y_{l,m}(\theta,\varphi) r^{n-1} \boxed{e^{-\alpha r}}$$

- Most electronic structure theory calculations (what we've been talking about) use Gaussian type functions because they are computationally much more efficient.

$$\chi_{\alpha,l_x,l_y,l_z}(x,y,z) = N x^{l_x} y^{l_y} z^{l_z} \boxed{e^{-\alpha r^2}}$$



- $l_x + l_y + l_z = l$ and determines type of orbitals (l=1 is a *p*...)
- χ's can be single Gaussian functions (primitives) or themselves be linear combinations of Gaussian functions (contracted).

Pople-style basis sets

- Named for Prof. John Pople who won the Nobel Prize in Chemistry for his work in quantum chemistry (1998).
- Notation: 6-31G

Use 6 primitives contracted to a single contracted-Gaussian to describe inner (core) electrons (1*s* in C)

Use 2 functions to describe valence orbitals (2s, 2p in C). One is a contracted-Gaussian composed of 3 primitives, the second is a single primitive.

6-311G Use 3 functions to describe valence orbitals...

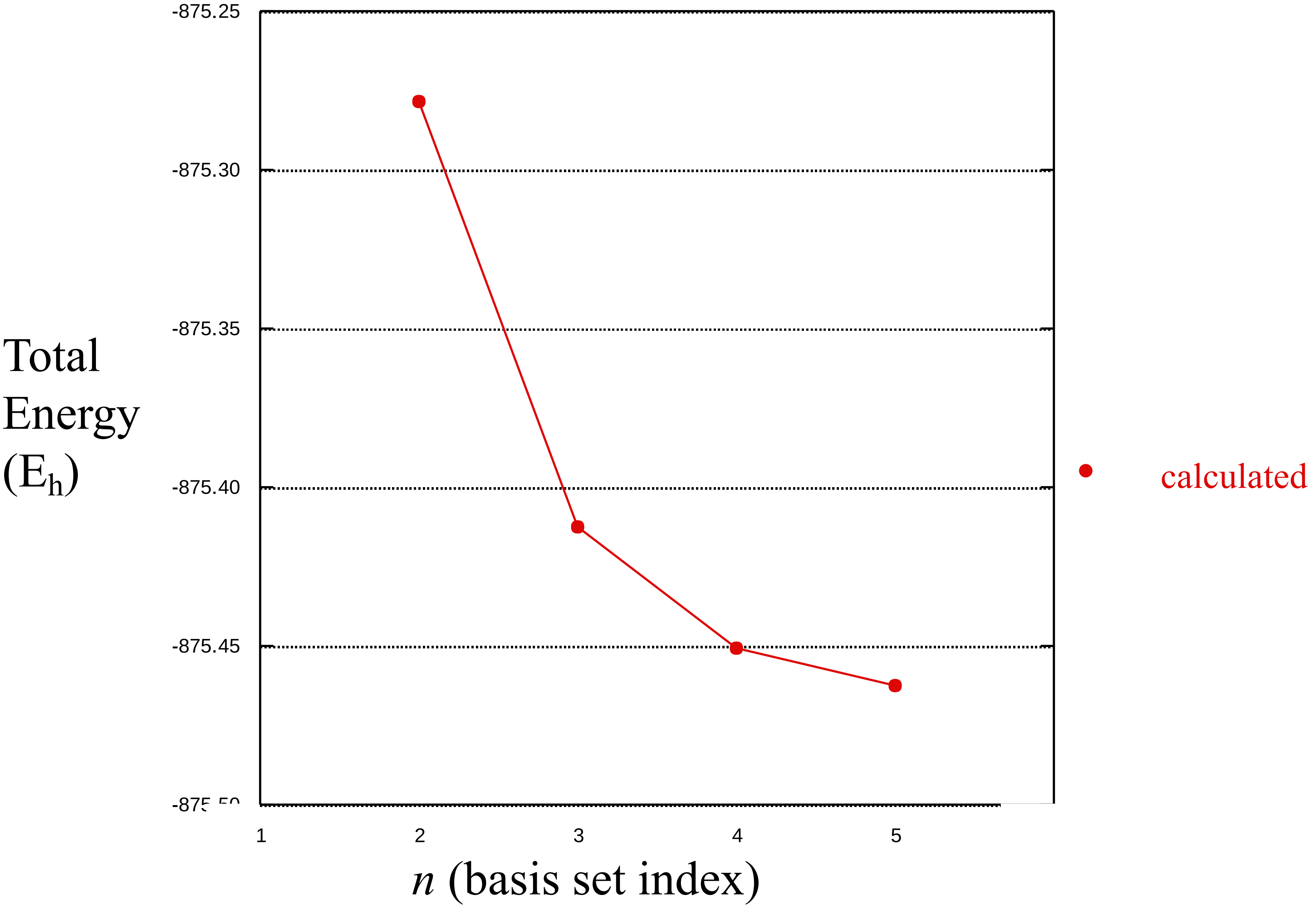
6-31G* Add functions of ang. momentum type 1 greater than occupied in bonding atoms (For N we'd add a d)

6-31G(d) Same as 6-31G* for 2nd and 3rd row atoms

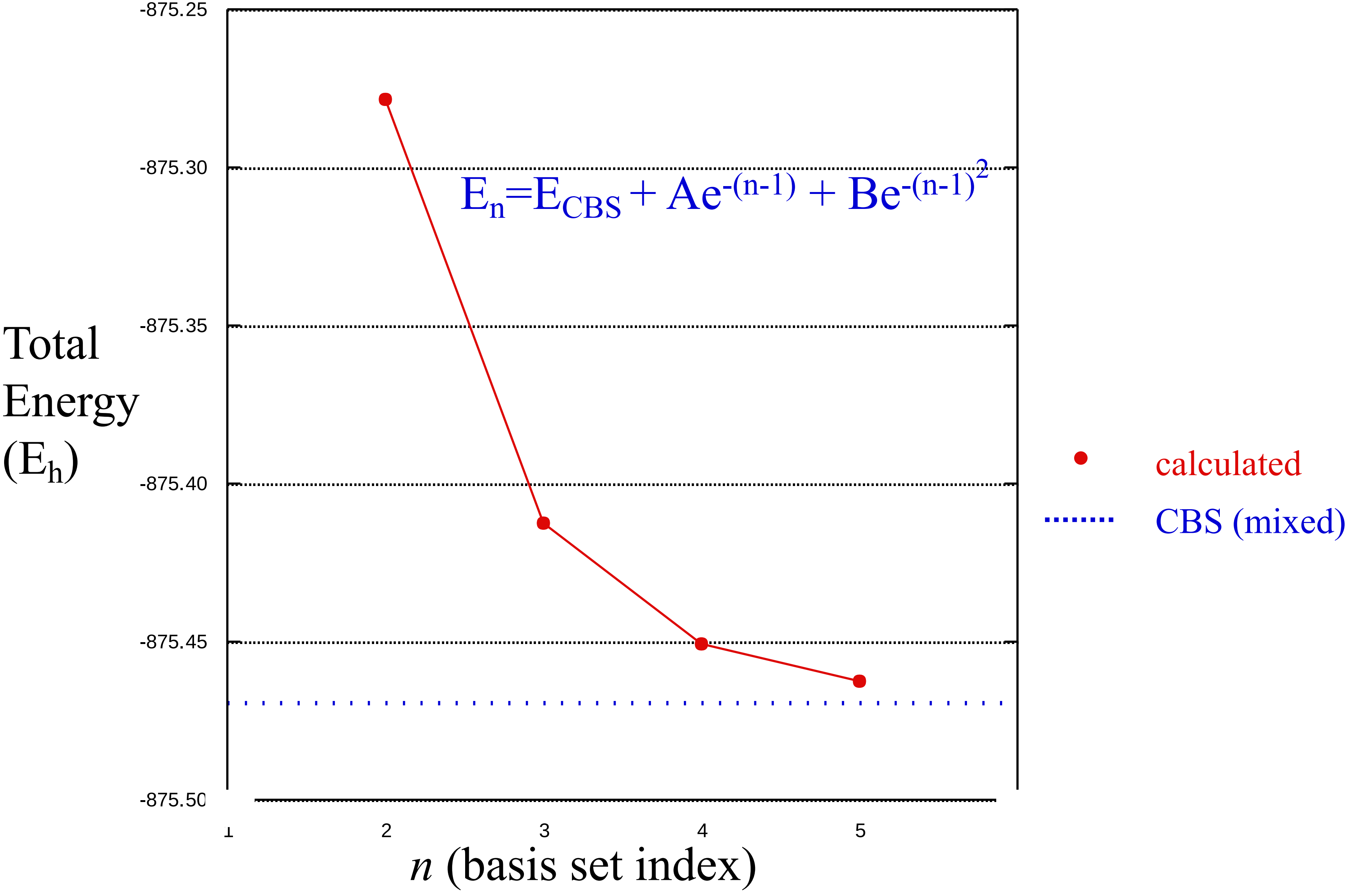
Correlation-Consistent Basis Sets

- Designed such that they have the unique property of forming a *systematically convergent* set.
- Calculations with a series of correlation consistent (cc) basis sets can lead to accurate estimates of the Complete Basis Set (CBS) limit.
- Notation: cc-pV*n*Z
  - correlation consistent polarized valence *n*-zeta
    - *n* = D, T, Q, 5,... (double, triple, quadruple, quintuple, ...)
      - double zeta-use 2 Gaussians to describe valence orbitals; triple zeta-use 3 Gaussians...
  - aug-cc-pV*n*Z: add an extra diffuse function of each angular momentum type
- Relation between Pople and cc basis sets
  - cc-pVDZ ≈ 6-31G(d,p)
  - cc-pVTZ ≈ 6-311G(2df,2pd)

Basis set convergence for the BrCl total energy: CCSD(T)/aug-cc-pV*n*Z

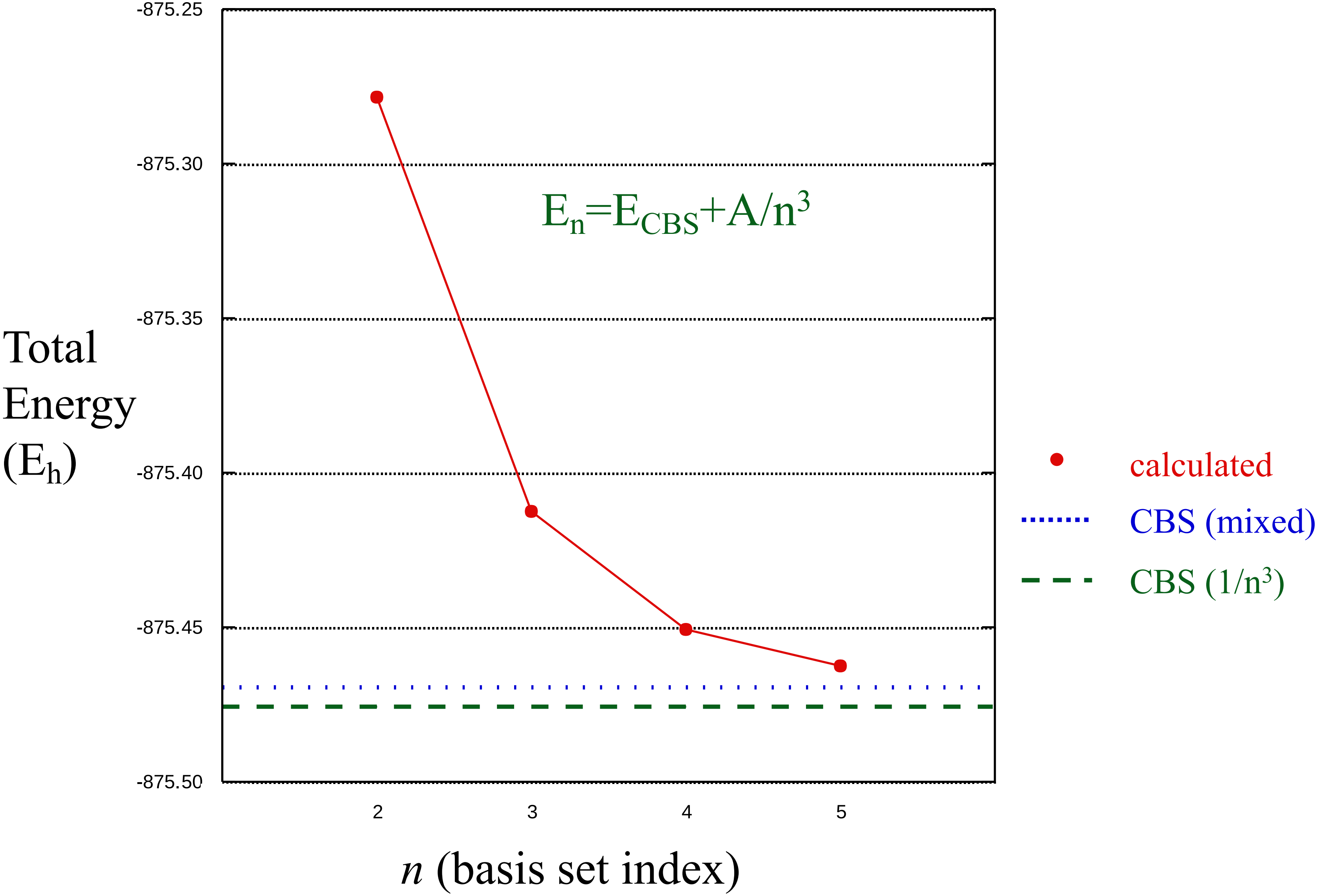


Basis set convergence for the BrCl total energy: CCSD(T)/aug-cc-pV*n*Z

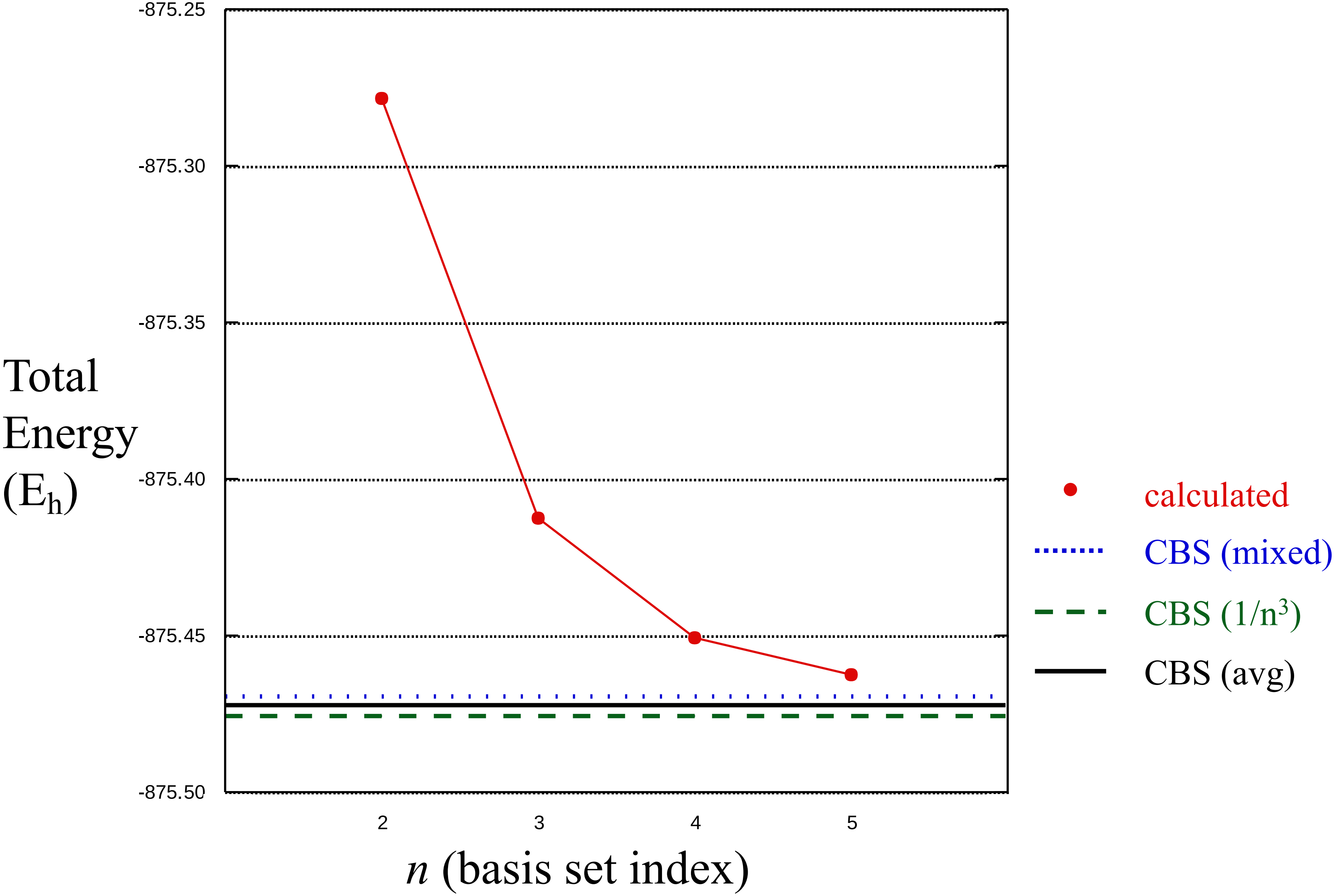


$$E_n = E_{CBS} + Ae^{-(n-1)} + Be^{-(n-1)^2}$$

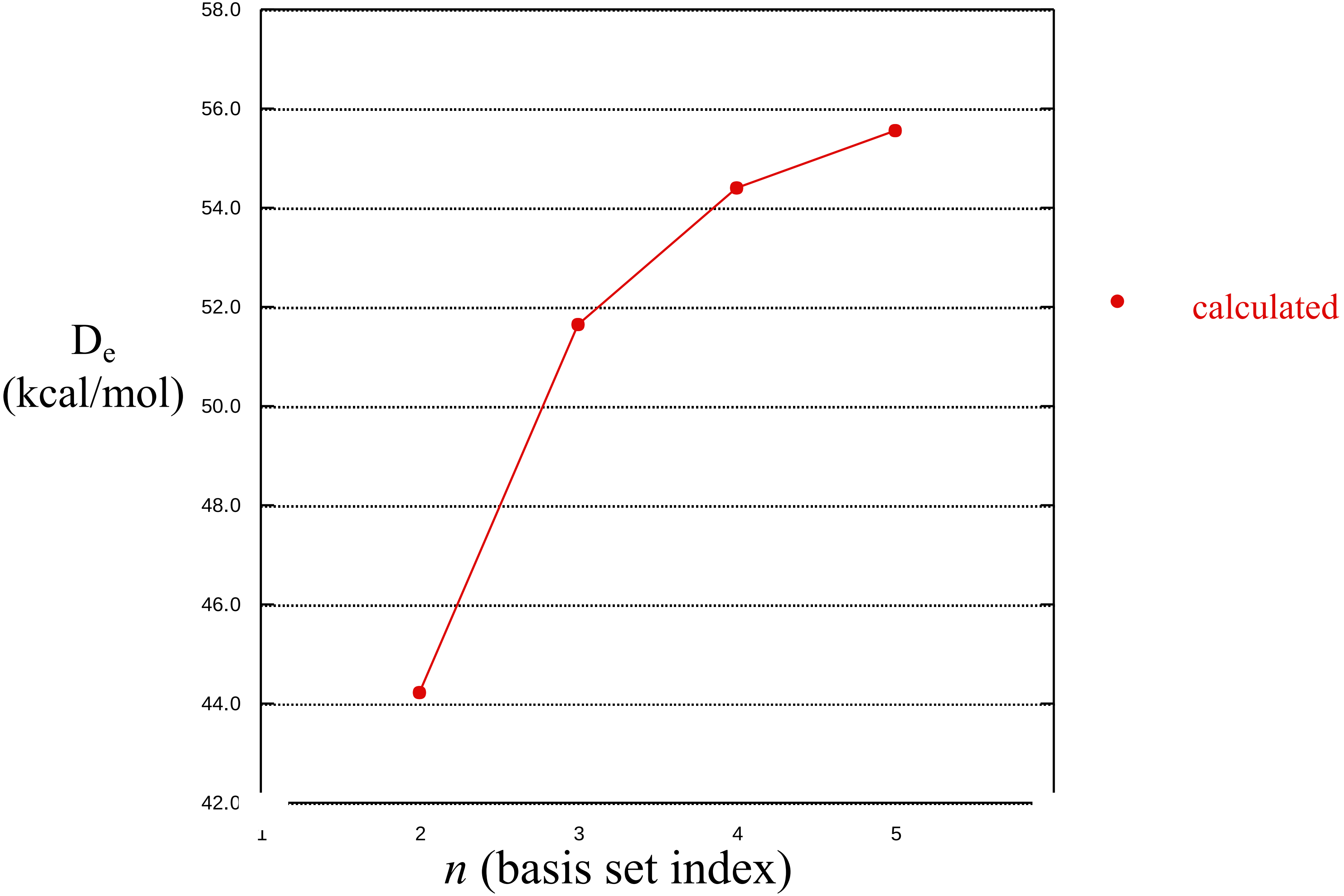
Basis set convergence for the BrCl total energy: CCSD(T)/aug-cc-pV*n*Z



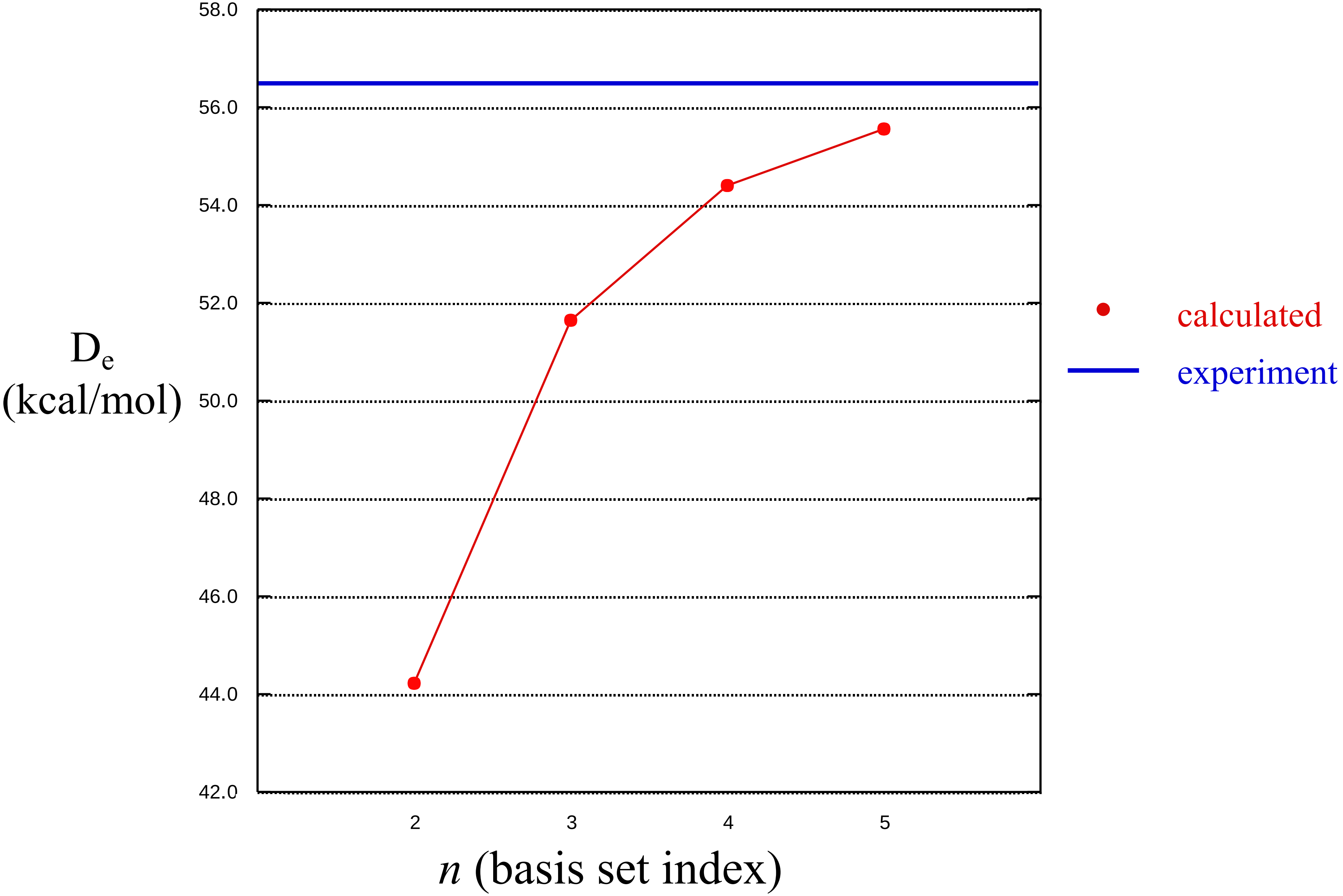
Basis set convergence for the BrCl total energy: CCSD(T)/aug-cc-pV*n*Z



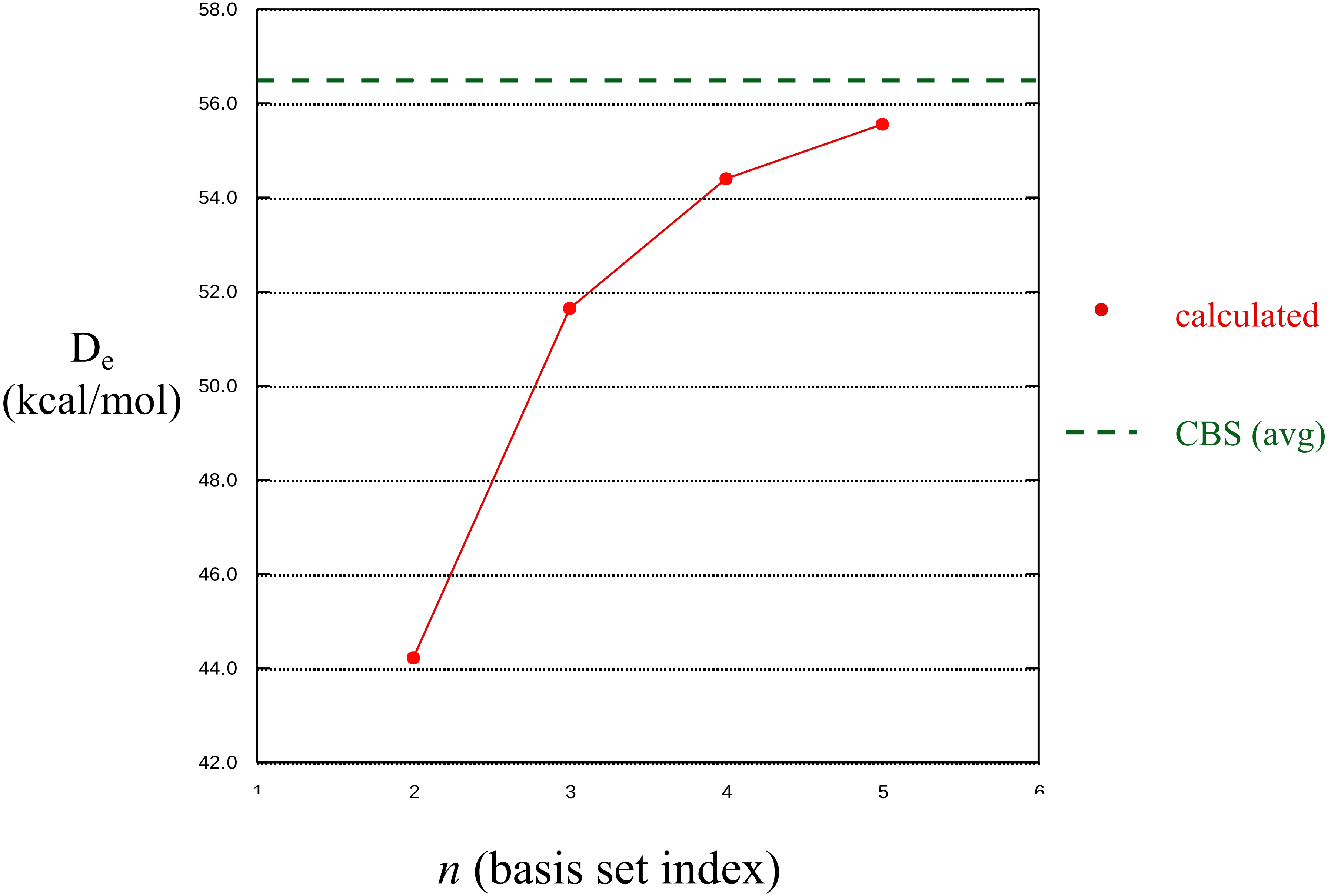
Basis set convergence for the **BrCl** D_e: CCSD(T)/aug-cc-pV*n*Z



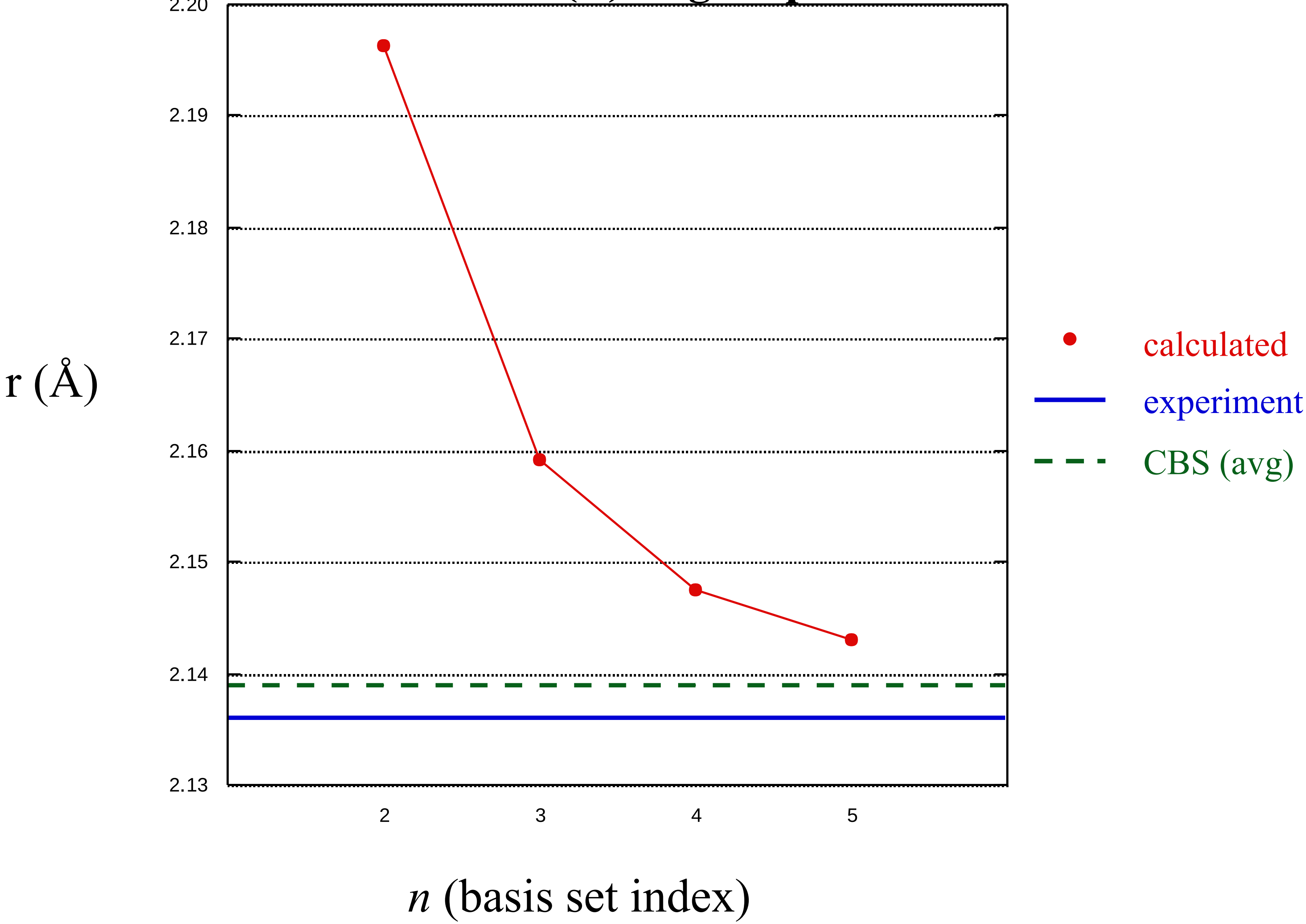
Basis set convergence for the BrCl D_e: CCSD(T)/aug-cc-pVnZ



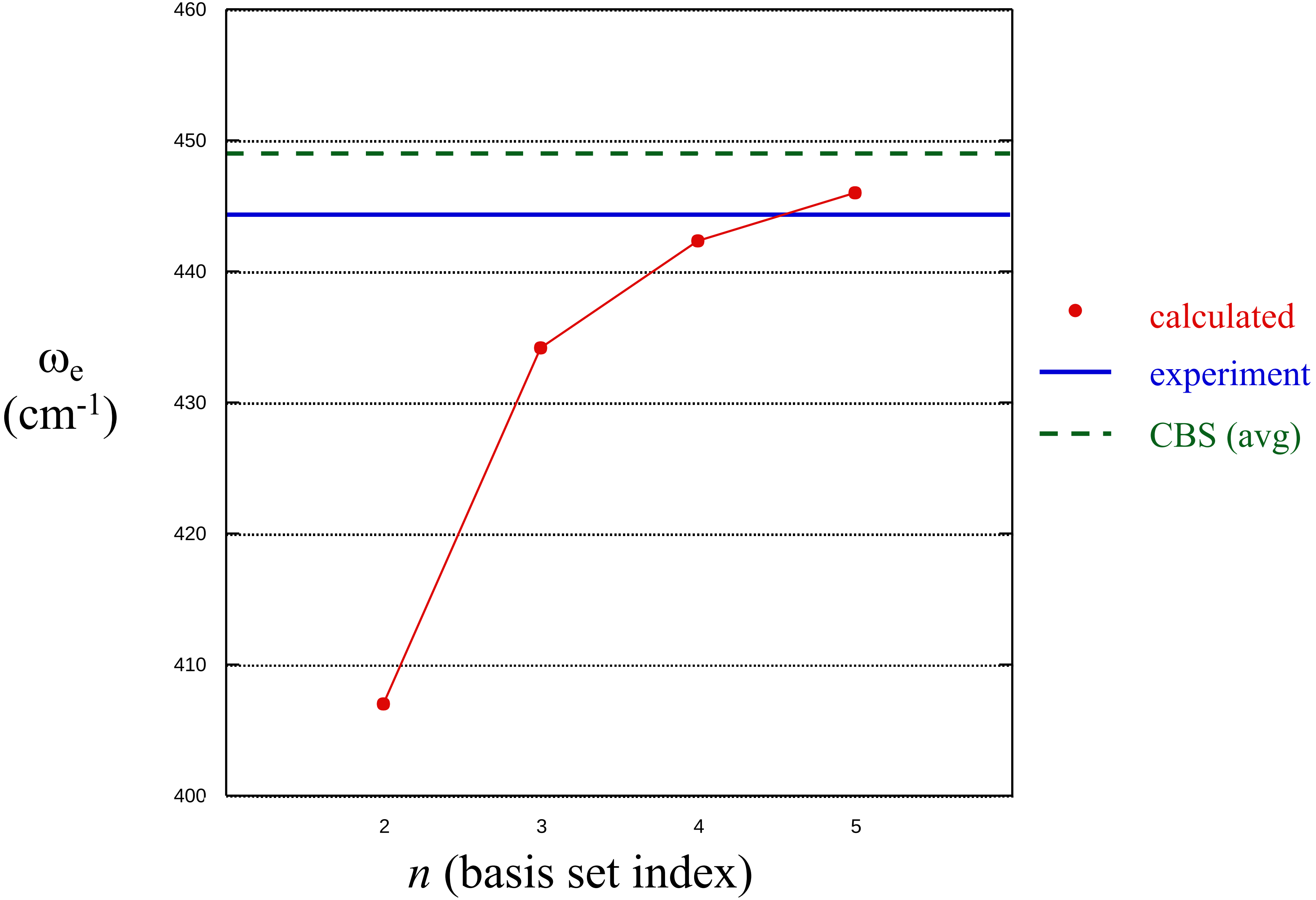
Basis set convergence for the **BrCl** D_e: CCSD(T)/aug-cc-pV*n*Z

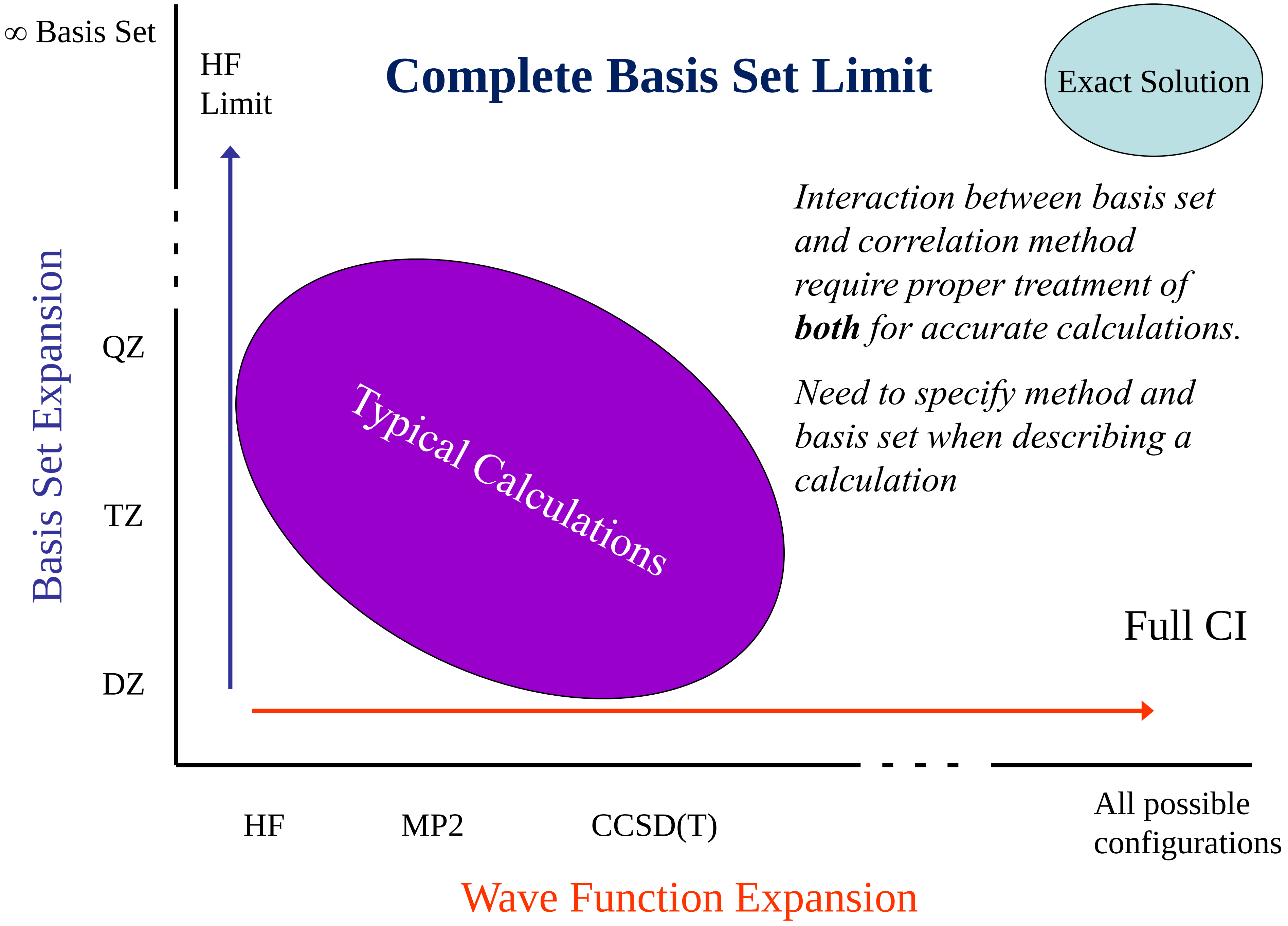


Basis set convergence for the BrCl bond length: CCSD(T)/aug-cc-pV*n*Z



Basis set convergence for the BrCl ω_e: CCSD(T)/aug-cc-pV*n*Z





∞ Basis Set

HF
Limit

Complete Basis Set Limit

Exact Solution

Typical Calculations

QZ

TZ

DZ

Basis Set Expansion

Full CI

HF

MP2

CCSD(T)

All possible
configurations

Wave Function Expansion

Interaction between basis set and correlation method require proper treatment of ***both*** *for accurate calculations.*

Need to specify method and basis set when describing a calculation

Computational Cost

Method	Scaling of Cost
HF	$M^2 - M^3$
MP2	M^5
CCSD	M^6
CCSD(T)	M^7

- Why not use best available correlation method with the largest available basis set?
  - A MP2 calculation would be 100x more expensive than HF calculation with same basis set.
  - A CCSD(T) calculation would be 10^4x more expensive than HF calculation with same basis set.
  - Tripling basis set size would increase MP2 calculation 243x (3^5).
  - Increasing the molecule size 2x (say ethane→butane) would increase a CCSD(T) calculation 128x (2^7).

High accuracy possible

- Despite all these approximations highly accurate results are still possible.

CCSD(T) Atomization Energies for Various Molecules

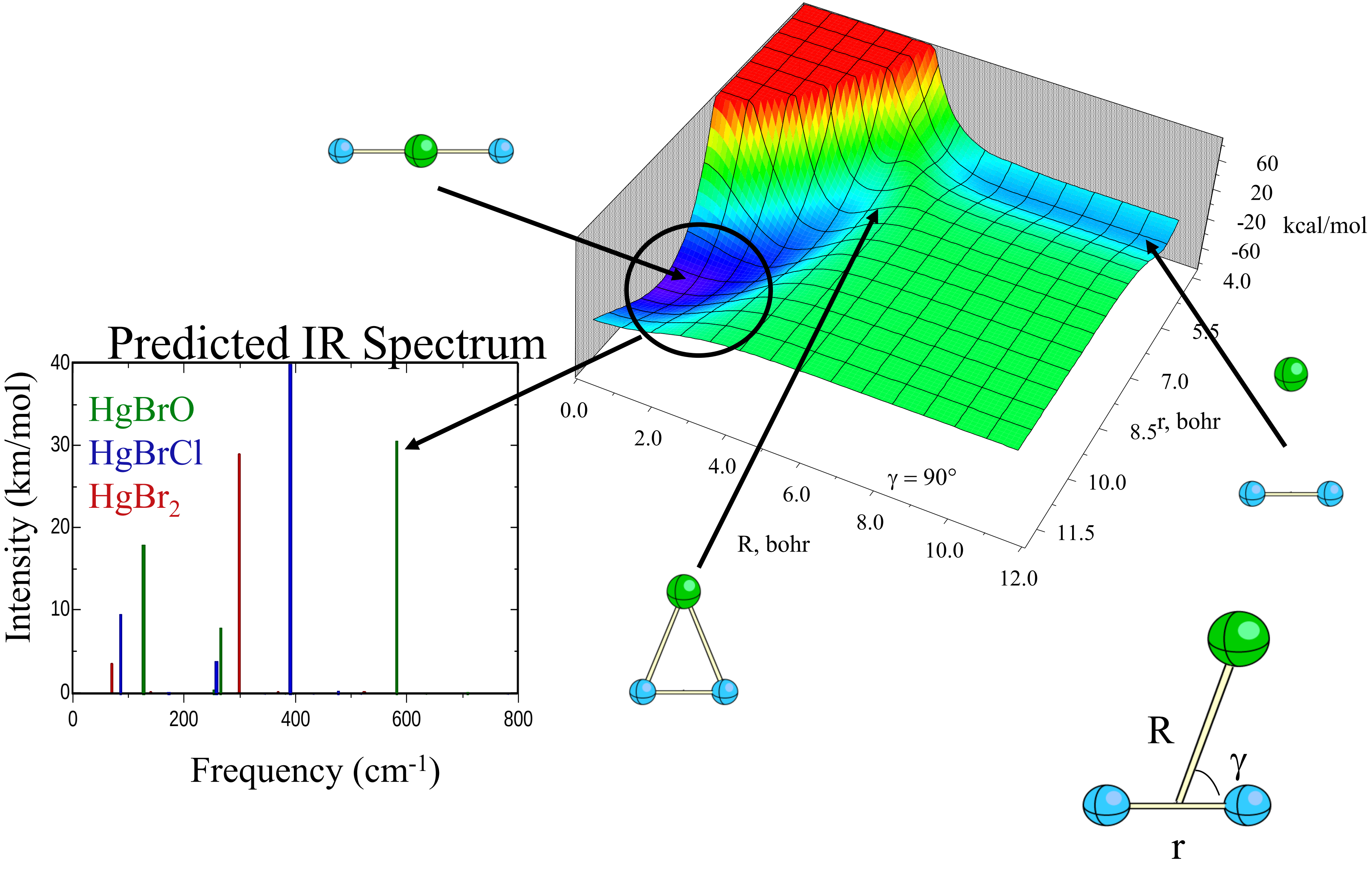
Molecule	CBS ΣD_e	ΔE_{ZPVE}	ΔE_{CV}	ΔE_{SR}	Atomic SO	Theoretical ΣD_0 (0K)	Experimental ΣD_0 (0K)
NH_2 ($^2B_1$)	181.8	-12.0	0.4	-0.2	0.0	170.0	170.0±0.3
H_3CSH ($^1A'$)	473.5	-28.6	1.5	-0.6	-0.6	445.2	445.1
SO_2 ($^1A_1$)	257.6	-3.9	1.0	-0.9	-1.0	253.7	254.0±0.2
Na_2 ($^1\Sigma_g^+$)	16.8	-0.2	0.3	0.0	0.0	16.9	16.8±0.3
BrCl ($^1\Sigma^+$)	56.6	-0.6	0.4	-0.4	-4.4	51.6	51.5±0.3
CH_3I ($^1\Sigma^+$)	372.3	-22.4	3.7	-0.9	-7.3	345.4	344.8±0.3
Si_2H_6 ($^1A_{1g}$)	536.1	-30.5	0.0	-1.1	-0.9	503.6	500.1

Atomization energies are notoriously difficult to calculate.

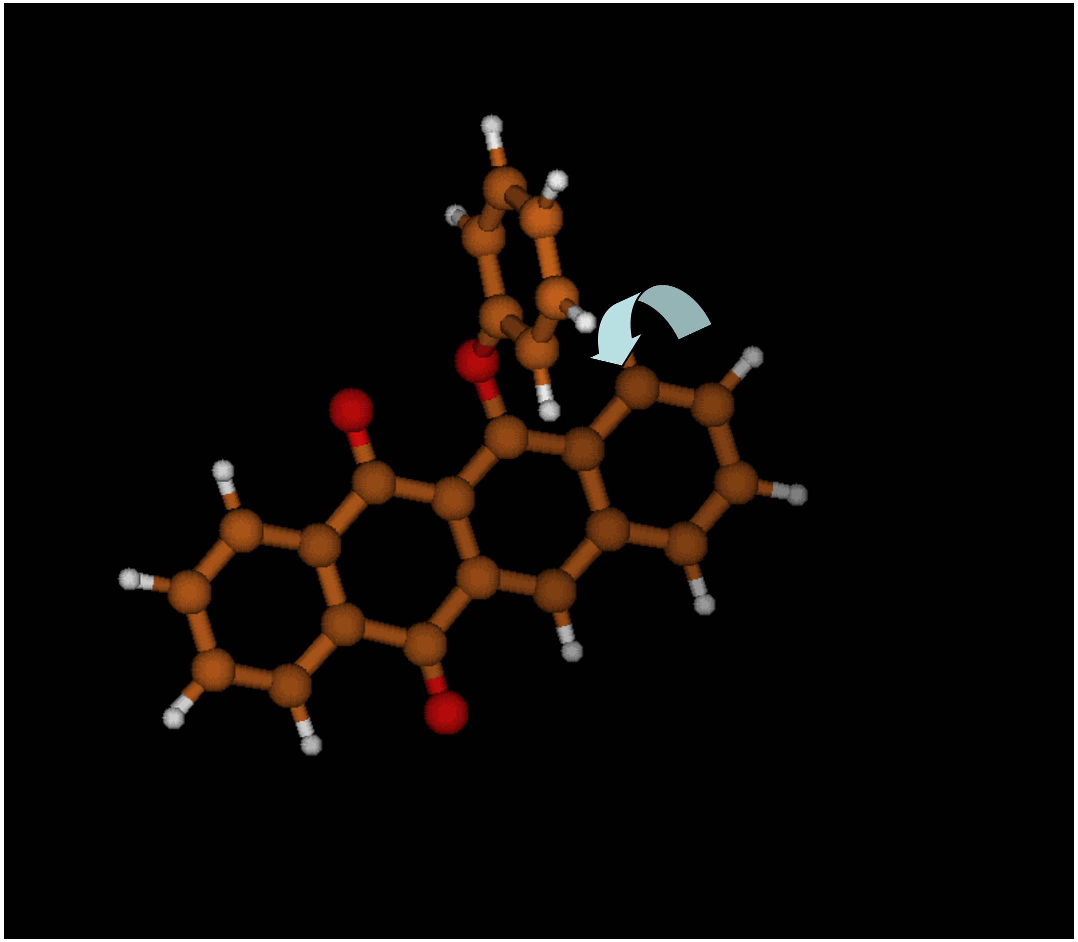
E_{ZPVE}: Zero-point energy correction
E_{CV}: core/valence correlation
E_{SR}: molecular/atomic scalar relativistic correlation
E_{SO}: atomic spin-orbit scalar relativistic correlation

Applications

Dynamics and Spectroscopy of the reactions of Hg and Halogens



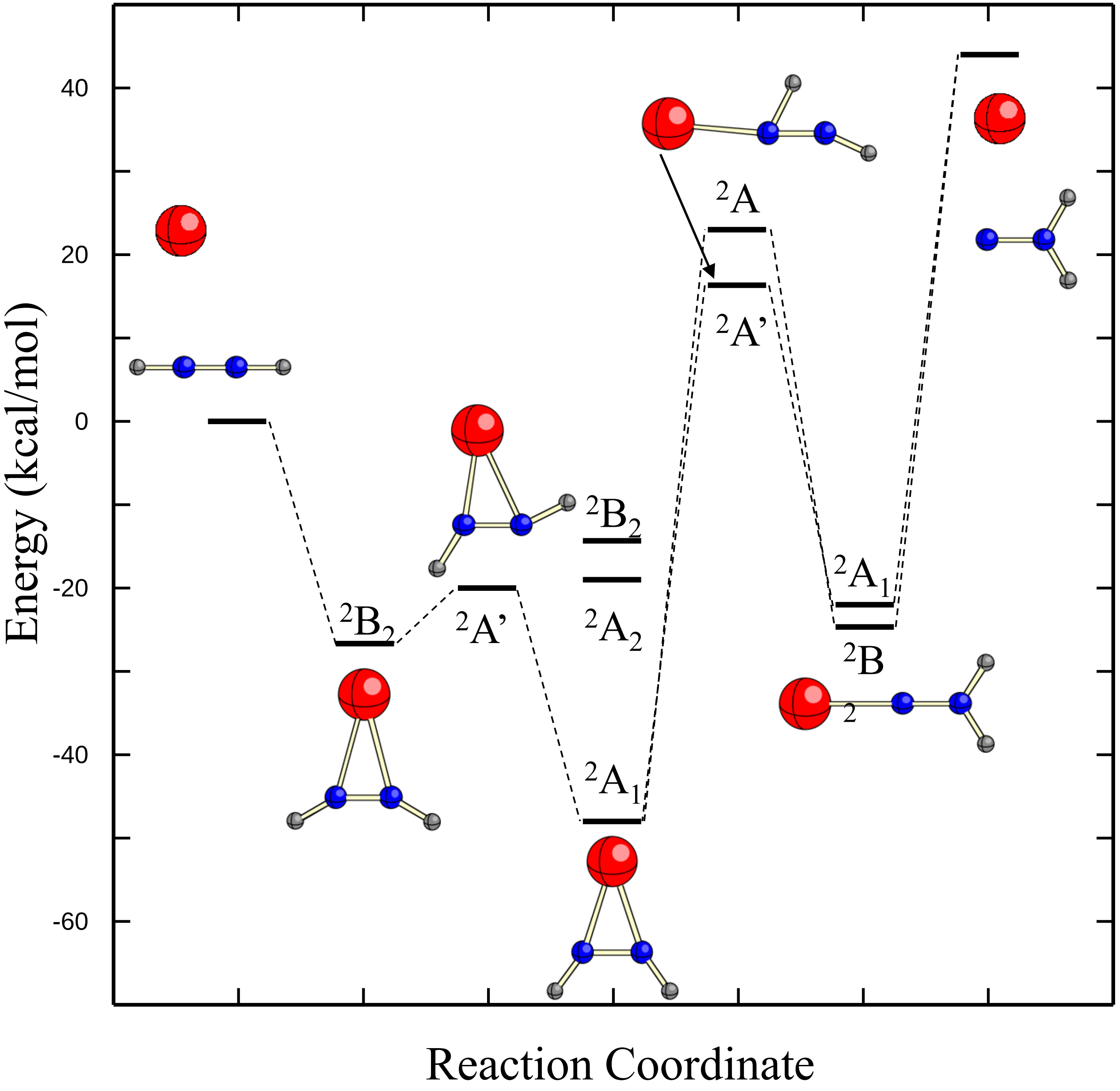
Materials Science Applications



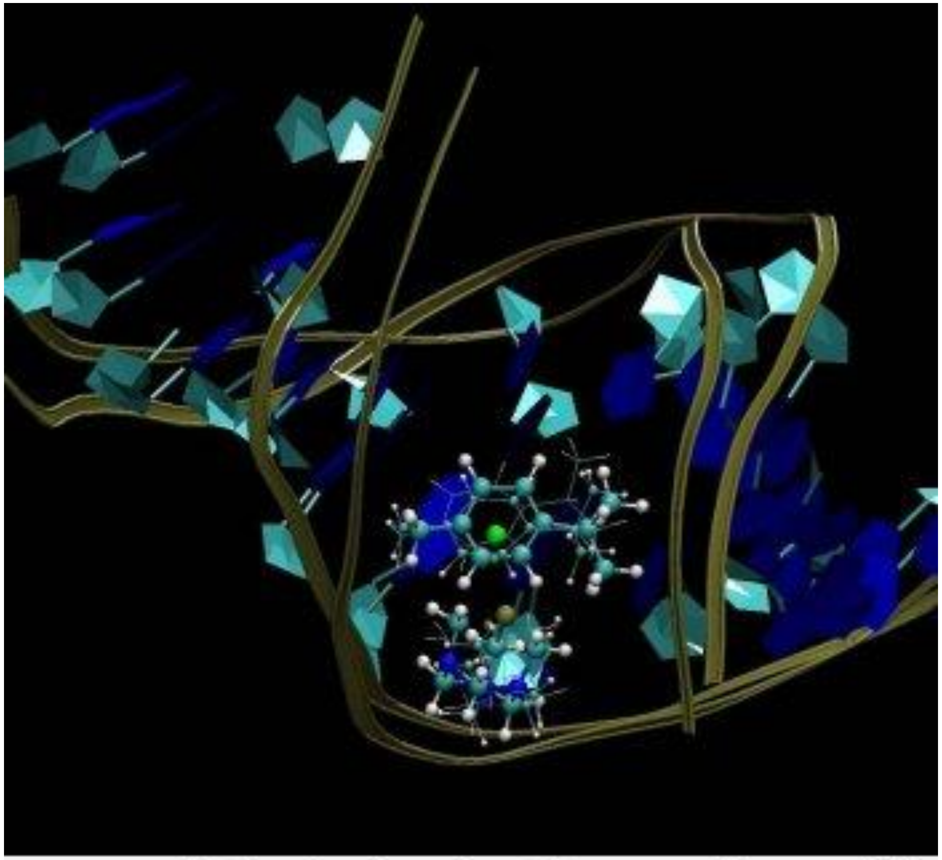
Potential photo-switch

- A **photo-switch** is a sensor that detects the presence or change of light.
- Photo switches are one type of molecular machines, a class of molecules that can be switched between at least two distinct thermodynamically stable forms by the application of an external stimulus.
- Development of such devices is crucial in the framework of the field of nanotechnology

Yttrium (Y) catalyzed rearrangement of acetylene



Biochemistry applications



Ruthenium-based anti-cancer drug on DNA

Laboratory of Computational Chemistry and Biochemistry
Institute of Chemical Sciences and Engineering
Swiss Federal Institute of Technology EPF Lausanne
Group Röthlisberger

Get your paper and pencil ready...

- There exist a large number of software packages capable of performing electronic structure calculations.
  - MOLPRO, GAMESS, COLUMBUS, NWCHEM, MOLFDIR, ACESII, GAUSSIAN, ...
  - The different programs have various advantages and capabilities.
- In this class we will be using the Gaussian program package.
  - Broad capabilities
  - Relatively easy for non-experts to get started with
  - Probably most widely used
- Gaussview, which is a GUI that interfaces with Gaussian, can be applied in building molecules and viewing output.

Caution!

- Different choices of methods and basis sets can yield a large variation in results.
- It is important to know the errors associated with and limitations of different computational approaches.
- This is important when doing your own calculations, and when evaluating the calculations of others.

- Don’t just accept the numbers the computer spits out at face value!

Conclusion

- Born-Oppenheimer Approximation
  - Separate electronic motion from nuclear motion and solve the electronic and nuclear S.E. separately.
- Expansion of the many electron wave function: "The Method"
  - Represent wave function as linear combination of Slater determinants.
  - More Slater determinants (in principle) yield more accurate results, but more expensive calculations.
- Expansion of molecular orbitals: "The Basis Set"