Equation-of-motion coupled cluster study of the vertical excitation spectra of cytosine adducts.

Tomasz Kus, Victor F. Lotrich and Rodney J. Bartlett

Collaborators: Mark Ponton(Aces QC), Erik Deumens(QTP), Norbert Flocke(QTP), Ajith Perera(QTP)
Introduction

• Equation-of-motion (EOM) theory
• Implementation: ACESIII
• Application: H and OH adducts of cytosine
• Performance
Coupled-cluster methods

- Standard tools used in highly accurate calculations of ground and excited state calculations.
- High scaling (CCSD:N$^6$) has previously limited the applicability of such calculations.
- Parallel computing has allowed these calculations to be performed, many problems can be reinvestigated now.
EOM-CCSD

- Direct method: A spectrum of electronic states obtained in a single calculation.
- Computationally ‘simple’ (at least conceptually!)
- Excited state wave functions generated from an RHF CC ground state are spin adapted.
Some fundamentals of the EOM-CCSD method.

\[ |\Psi_x \rangle = \hat{R} |\Psi_g \rangle \]  \hspace{1cm} (1)

where \( \hat{R} \) is a linear CI like excitation operator

\[ \hat{R} = \hat{R}_1 + \hat{R}_2 + \ldots \]  \hspace{1cm} (2)

\[ \hat{R}_n = (\frac{1}{n!})^2 \sum r_{ijk \ldots}^a b^i c^j k \ldots \]  \hspace{1cm} (3)

The ground state wavefunction (CC) is,

\[ |\Psi_g \rangle = e^{T} |\phi_0 \rangle \]  \hspace{1cm} (4)

where \( |\phi_0 \rangle \) is the single determinant referance wavefunction (SCF usually). Thus,

\[ \hat{H} e^{T} |\phi_0 \rangle = E e^{T} |\phi_0 \rangle \]  \hspace{1cm} (5)

leading to the eigenvalue problem,

\[ (\hat{H} - E) \hat{R} |\phi_0 \rangle = 0 \]  \hspace{1cm} (6)

where the effective Hamiltonian \( \hat{H} = e^{-T} H e^T \).
Steps required in EOM-CCSD

- SCF
- 2-electron integral transformation
- CCSD
- CI singles
- HBAR
- EOM
- Single determinant reference
- Transformed integrals
- Ground state wavefunction
- Initial excited state guess
- Effective Hamiltonian
- Excited states
Traditional Design

code

control

compute

communication

disk input output

hardware
ACESIII Design

- **Control**
- **Compute**
- **Communication**
- **Disk I/O**

- **Hardware**
- **Code**
ACESIII design

High level
- concepts
- Data structures
- algorithms
- Super instruction
  Assembly language
  SIAL

Low level
- communication
- Input/output
  Super instruction
  Processor
  SIP (xaces3)

Problem
Performance
input
output
SIAL (Super Instruction Assembly Language)

- **Key features**
  - Index segmentation
  - Data blocking
  - Task isolation

- **Advantageous**
  - Flexibility
  - Tune ability: Fast optimization
  - New methods implemented in reduced time
  - Portable
Cytosine

- One of five main nitrogenous bases used in storing and transporting information within a cell.
- Can be found as part of DNA, RNA, or as part of a nucleotide.
- Attacked by OH radicals (main species for DNA damage).
- Understanding the basic chemistry of OH radicals and DNA bases is an important step in characterizing the potential damage on DNA.
Geometry of cytosine

![Diagram of cytosine molecule with labels for atoms and hydrogens.]

- Atoms labeled as 5, +H, or OH
- Atoms labeled as 6, +H or OH
- Atoms labeled as 3, +H
Some computational details

- Geometries were optimized at the MBPT(2) level using basis sets ranging from 6-31G to aug-cc-pvtz (PBS structures adequate).
- UHF reference functions were used in all calculations.
- Core electrons were not correlated.
- The C5-OH, C6-OH, C5-H, C6-H, and N3-H adducts were considered.
CCSD energies (Hartree) and energy differences (kJ/mol) for the adducts of cytosine. The MCSCF results are from M. Krauss and R. Osman, J. Phys. Chem. A, 101, 33 (1997).

<table>
<thead>
<tr>
<th>Adduct</th>
<th>CCSD [Hartree]</th>
<th>Diff₁ [kJ/mol]</th>
<th>Diff₂ [kJ/mol]</th>
</tr>
</thead>
<tbody>
<tr>
<td>C6-OH</td>
<td>-469.63198</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>C5-OH</td>
<td>-469.63175</td>
<td>0.61(19)</td>
<td>----</td>
</tr>
<tr>
<td>N3-H</td>
<td>-394.54380</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>C5-H</td>
<td>-394.53733</td>
<td>16.98(24)</td>
<td>----</td>
</tr>
<tr>
<td>C6-H</td>
<td>-394.53167</td>
<td>31.85(26)</td>
<td>14.86(2)</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Adduct</th>
<th>Method/BASIS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>C6-OH</td>
<td>EOM/PBS</td>
<td>2.63</td>
<td>2.92</td>
<td>3.39</td>
<td>3.64</td>
<td>4.62</td>
</tr>
<tr>
<td></td>
<td>EOM/aug-cc-pvtz</td>
<td>2.67</td>
<td>2.92</td>
<td>3.45</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>MCSCF/DZd</td>
<td>2.92</td>
<td>2.97</td>
<td>—</td>
<td>3.51</td>
<td>4.52</td>
</tr>
<tr>
<td></td>
<td>MCSCF/DZ</td>
<td>2.69</td>
<td>2.82</td>
<td>—</td>
<td>3.17</td>
<td>4.41</td>
</tr>
<tr>
<td>C5-OH</td>
<td>EOM/PBS</td>
<td>3.71</td>
<td>4.00</td>
<td>4.32</td>
<td>4.47</td>
<td>4.50</td>
</tr>
<tr>
<td></td>
<td>MCSCF/DZ</td>
<td>3.64</td>
<td>3.93</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C5-H</td>
<td>EOM/PBS</td>
<td>3.56</td>
<td>4.30</td>
<td>4.51</td>
<td>4.66</td>
<td>4.73</td>
</tr>
<tr>
<td></td>
<td>MCSCF/DZ</td>
<td>3.92</td>
<td>4.14</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C6-H</td>
<td>EOM/PBS</td>
<td>2.87</td>
<td>3.05</td>
<td>3.33</td>
<td>3.79</td>
<td>5.23</td>
</tr>
<tr>
<td></td>
<td>MCSCF/DZ</td>
<td>2.69</td>
<td>2.89</td>
<td>3.13</td>
<td>4.04</td>
<td>—</td>
</tr>
<tr>
<td>N3-H</td>
<td>EOM/PBS</td>
<td>2.94</td>
<td>2.97</td>
<td>3.39</td>
<td>4.03</td>
<td>4.27</td>
</tr>
<tr>
<td></td>
<td>EOM/aug-cc-pvtz</td>
<td>2.98</td>
<td>3.14</td>
<td>3.56</td>
<td>4.14</td>
<td>4.15</td>
</tr>
<tr>
<td></td>
<td>MCSCF/DZ</td>
<td>3.17</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

The C6-OH adduct

• The lowest energy species among the OH radicals: 0.3kJ/mol lower in energy than the C5-OH radical and 21.3kJ/mol lower than the C4-OH radical. Yan Ju Ji et. al. Journal of Molecular Structure: Theochem 723 (2005) 123-129

• Consistent with our energy difference of 0.61kJ/mol.
Vertical excitation energies\[\text{eV}\] for the C6-OH adduct of cytosine.

<table>
<thead>
<tr>
<th>Method/Basis</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>EOM/PBS</td>
<td>2.63</td>
<td>2.92</td>
<td>3.39</td>
<td>3.64</td>
<td>4.62</td>
</tr>
<tr>
<td>EOM/aug-cc-pvtz</td>
<td>2.67</td>
<td>2.92</td>
<td>3.45</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>MCSCF/DZd</td>
<td>2.92</td>
<td>2.97</td>
<td>-----</td>
<td>3.51</td>
<td>4.52</td>
</tr>
<tr>
<td>MCSCF/DZ</td>
<td>2.69</td>
<td>2.82</td>
<td>-----</td>
<td>3.17</td>
<td>4.41</td>
</tr>
</tbody>
</table>
Performance: Scaling

- C5-OH adduct
- PBS basis (270 AO’s)
- Number of alpha/beta correlated occupied orbitals: 34/33
- Reference: UHF
- Point group symmetry: $C_1$
- MO codes used in post HF calculations
Time for one CCSD–EOM iteration

- Ideal
- Actual

Super scaling region

Normal scaling region

Time [sec]

Number of processors

1899

4.4 minutes

10^3

32

256
Performance: Timings

- N3-H adduct
- Aug-cc-pvtz basis (506 AO’s)
- 30/29 alpha/beta correlated occupied orbitals
- UHF reference
- $C_1$ point group symmetry
- 256 processors used (3/1 worker/server)
Time in minutes for calculation of the various modules required in the EOM-CCSD calculation

- SCF/iter
- Transformation
- CIS
- Hbar
- CCSD/iter
- EOM/iter
- Grand Total

- $< 1$ (42 iterations)
- 12
- 39
- 88 (IO intense)
- 11 (18 iterations)
- 10.5 (146 iterations)
- 1893