

Accurate quantum chemistry calculations for chromophores in photoactive proteins

Emanuele Coccia¹, Daniele Varsano² and Leonardo Guidoni¹

¹Dipartimento di Scienze Fisiche e Chimiche, Università degli Studi dell'Aquila,
L'Aquila (Italy)

²Centro S3, CNR Istituto di Nanoscienze, Modena (Italy)

PRACE Scientific Conference 2013

- 1 Quantum Monte Carlo: why and how
- 2 Mechanism of vision: Rhodopsin

- 1 Quantum Monte Carlo: why and how
- 2 Mechanism of vision: Rhodopsin
- 3 Peridinin in PCP complex

- 1 Quantum Monte Carlo: why and how
- 2 Mechanism of vision: Rhodopsin
- 3 Peridinin in PCP complex
- 4 Conclusions

- 1 Quantum Monte Carlo: why and how
- 2 Mechanism of vision: Rhodopsin
- 3 Peridinin in PCP complex
- 4 Conclusions

- Computational Biophysics, Biochemistry and Chemistry
- People from:
 - University of L'Aquila, L'Aquila, Italy (Prof. Leonardo Guidoni)
 - "Sapienza" - University of Rome, Rome, Italy
 - Centro S3, CNR Istituto di Nanoscienze, Modena, Italy
- Research interests:
 - Geometry, electronic structure and energy transfer in photosynthetic systems by AIMD and classical MD
 - Quantum Monte Carlo in chemistry: methods and applications (TurboRVB code)
 - Electronic excited states of biomolecules
 - Molecular vibrations
- MultiscaleChemBio: 5-year "IDEAS" research project supported by the European Research Council
- <http://bio.phys.uniroma1.it>

Quantum Monte Carlo: why and how

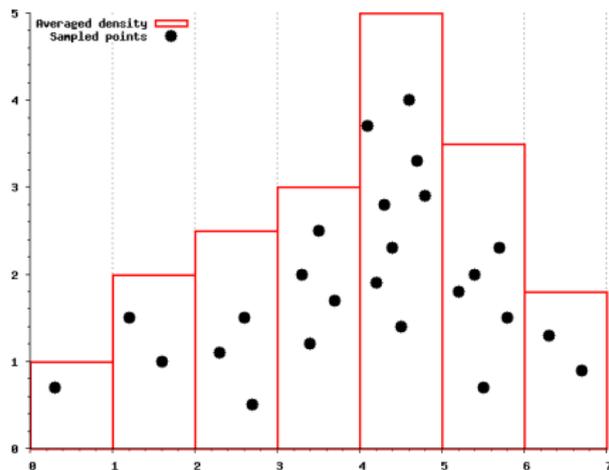
Why Quantum Monte Carlo

- Explicit correlation
- Conjugated systems
- Weak interactions
- Reaction barriers
- ...



	DFT	post-HF	QMC
Scaling	N^3	$N^{5,7,10}$	$N^{3,4}$
Pros	Large systems Plane wave codes	Very accurate	Very accurate Intrinsically parallel
Cons	Many systems are still a challenge for XC functionals	Not applicable to large systems	Large prefactor Stochastic error No "standards"

$$E_{VMC} = \min_{\{\alpha_j\}} \frac{\langle \Psi_T(\{\alpha_j\}) | \hat{H} | \Psi_T(\{\alpha_j\}) \rangle}{\langle \Psi_T(\{\alpha_j\}) | \Psi_T(\{\alpha_j\}) \rangle} \geq E_0$$

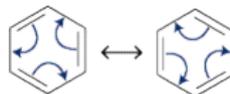


- Integration in the $3N$ variational space by stochastic methods
- With M sampling points the stochastic error $\epsilon \propto \frac{1}{\sqrt{M}}$
- ϵ independent of $N!$
- M points randomly drawn from $\Pi \equiv \frac{|\Psi_T|^2}{\int d\mathbf{r} |\Psi_T|^2}$
- Choice of the trial wave function Ψ_T

VMC optimization

$$\Psi_T(\mathbf{r}, \mathbf{R}) = D(\mathbf{r}, \mathbf{R}) \times J(\mathbf{r}, \mathbf{R})$$

- Determinantal part (AGP)



$$\psi_{AGP} = \hat{A} [\Phi_G(\mathbf{r}_1^\uparrow; \mathbf{r}_1^\downarrow) \Phi_G(\mathbf{r}_2^\uparrow; \mathbf{r}_2^\downarrow) \cdots \Phi_G(\mathbf{r}_{N/2}^\uparrow; \mathbf{r}_{N/2}^\downarrow)]$$

$$\Phi_G = \phi_G(\mathbf{r}_i, \mathbf{r}_j) \frac{1}{\sqrt{2}} (|\uparrow\rangle_i |\downarrow\rangle_j - |\uparrow\rangle_j |\downarrow\rangle_i)$$

$$\phi_G(\mathbf{r}_i, \mathbf{r}_j) = \sum_{a,b=1}^M \sum_{\mu\sigma, \nu_b} \lambda_{\mu\sigma\nu_b} \psi_{\mu\sigma}(\mathbf{r}_i) \psi_{\nu_b}(\mathbf{r}_j)$$

- Jastrow term

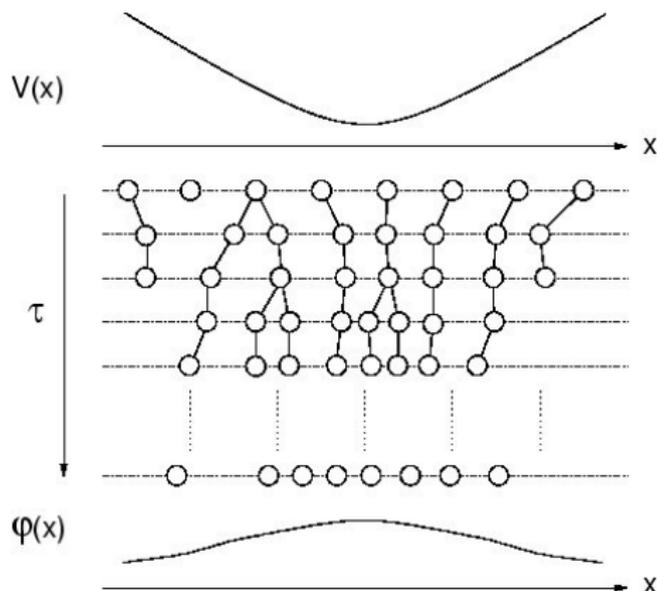
$$J(\mathbf{r}, \mathbf{R}) = J_1(\mathbf{r}, \mathbf{R}) \times J_2(\mathbf{r}, \mathbf{R}) \times J_{3/4}(\mathbf{r}, \mathbf{R})$$

- Ψ_T optimization: Stochastic evaluation of gradients S. Sorella, M. Casula and D. Rocca, *JCP*, **127**, 014105 (2007)
- Geometry optimization: Adjoint Algorithmic Differentiation S. Sorella and L. Capriotti, *JCP*, **133**, 234111 (2010)

- M. Barborini, S. Sorella and L. Guidoni, *JCTC*, **8**, 1260 (2012) (triplet state C_2H_4)
- EC, O. Chernomor, M. Barborini, S. Sorella and L. Guidoni, *JCTC*, **8**, 1952 (2012) (electrical properties HCCH)
- EC and L. Guidoni, *JCC*, **33**, 2332 (2012) (Retinal Minimal Model $C_5H_6NH_2^+$)
- A. Zen, D. Zelyazov and L. Guidoni, *JCTC*, **8**, 4204 (2012) (molecular vibrations)
- M. Barborini and L. Guidoni, *JCP*, **137**, 224309 (2012) (reaction pathways)
- EC, D. Varsano and L. Guidoni, *JCTC*, **9**, 8 (2013) (Rhodopsin)
- A. Zen, Y. Luo, S. Sorella and L. Guidoni, submitted on *JCTC* (molecular vibrations)
- EC, D. Varsano and L. Guidoni, in preparation (gas phase peridinin)

QMC: basic algorithm

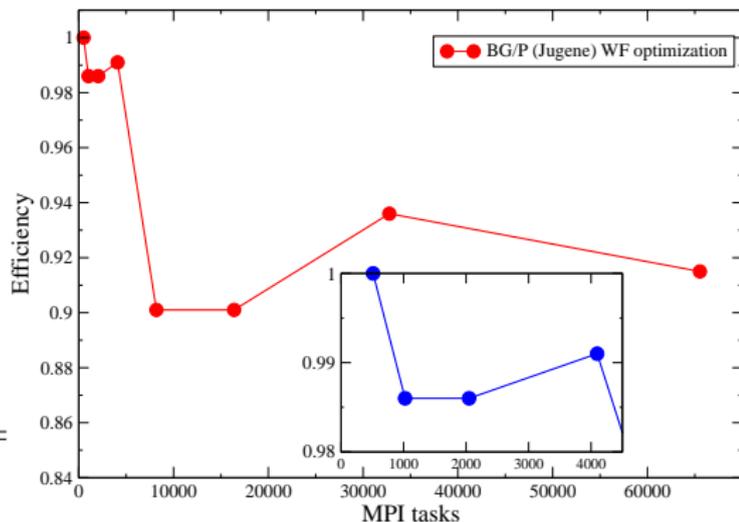
- 0) choose initial parameters β^{old} (1 sampling point per core)
- 1) select new parameters $\{\beta^{new}\} = \{\beta^{old}\} + \{\zeta\}$
- 2) apply Metropolis to accept or reject the move
- 3) update E_{VMC} (collecting data)
- 4) until the energy no longer diminishes



QMC on BlueGene

- QMC embarrassing parallel algorithm
- Suitable for BlueGene architectures
- Wave function optimization of Retinal
- Pure MPI runs
- Efficiency = $T^*/T(\text{MPI tasks})$, where $T^*=T(512)$
- Total time in *weak scaling* regime
- Calculations on Jugene (BG/P)

MPI tasks	time (s)	Efficiency
512	1602	-
1024	1624	0.986
2048	1624	0.986
4096	1617	0.991
8192	1779	0.901
16384	1769	0.901
32768	1712	0.936
65536	1751	0.915



List of grants for QMC

- *Quantum Monte Carlo methods for biological systems* ([Jugene](#), [Preparatory Access](#))
- *Protein effects on the structural and optical properties of biological chromophores: Quantum Monte Carlo/Molecular Mechanics calculations on Rhodopsin and Light Harvesting Complexes* ([Jugene](#), [PRACE Tier-0 Regular access](#))*
- *QMC-MEP - Reaction pathways by Quantum Monte Carlo: from benchmarks to biochemistry* ([Curie](#), [PRACE Tier-0 Regular Access](#))
- *RHODQMC - Energy storage in the first step of vision explored by Quantum Monte Carlo/Molecular Mechanics calculations* ([Juqueen](#), [PRACE Tier-0 Regular Access](#))
- *Fully Correlated Molecular Electric Properties by Quantum Monte Carlo* ([Fermi](#), [National Grant](#))
- *Rhodopsin environmental effects on the Retinal ground state structure: a Quantum Monte Carlo study* ([Fermi](#), [National Grant](#))
- *Quantum Monte Carlo polarizability of long polyacetylene chains* ([Fermi](#), [National Grant](#))

*Present results

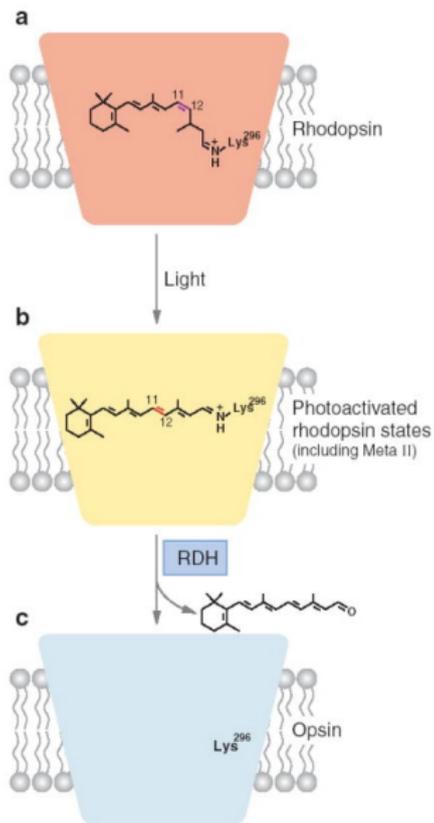
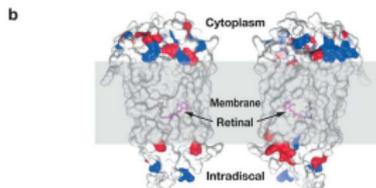
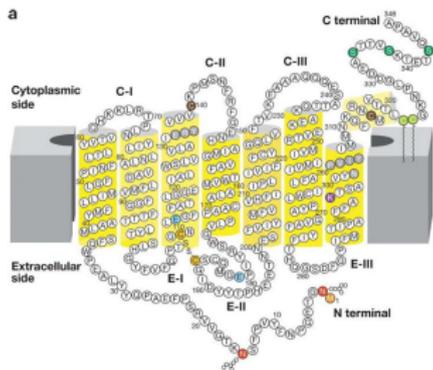
Why QMC in (photo)chemistry

- Interaction with light fundamental for many biological processes ([photosynthesis](#), [vision](#) etc.)
- Role of **conjugated chromophores**
- Accuracy in simulating absorption: structure + excited states
- Methods with [favorable scaling \(\$N^3\$ \)](#) could be not accurate enough for the description of conjugated chromophores
- System-size prevents the use of correlated post-HF approaches ([bad scaling and not parallel](#))
- QMC as optimal candidate for such kind of molecules (**hundreds of electrons**)
- Accuracy comparable with that of CCSD methods
- Explicit dynamical electronic correlation

Mechanism of vision: Rhodopsin

Retinal: an overview

- Chromophore in light-detecting proteins
- Rhodopsin in the retina of vertebrates
- Very fast isomerization (~ 200 fs, faster than in solution)



K. Palczewski, *Annu. Rev. Biochem.*, **75**, 743 (2006)

VMC validation on $C_5H_6NH_2^+$

- $\Psi_T(\mathbf{r}, \mathbf{R}) = D(\mathbf{r}, \mathbf{R}) \times J(\mathbf{r}, \mathbf{R})$

- VMC1** AGP from cc-pVDZ (without d orbitals)

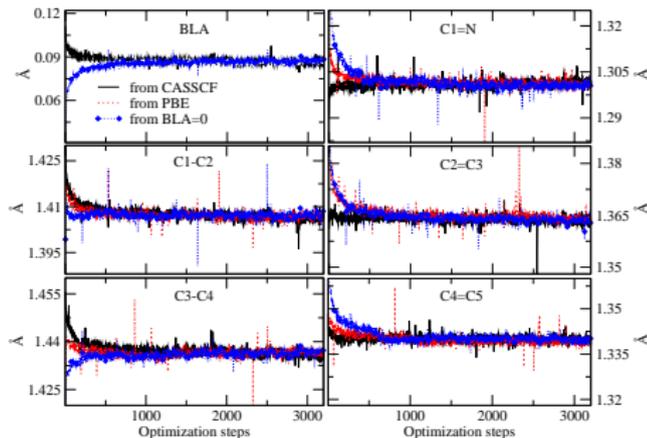
- $C, N = (4s4p)/(2s2p)$
- $H = (3s1p)/(2s1p)$

- VMC2** AGP from cc-pVDZ

- $C, N = (4s4p1d)/(2s2p1d)$
- $H = (3s1p)/(2s1p)$

- Same Jastrow J_3 for all AGPs

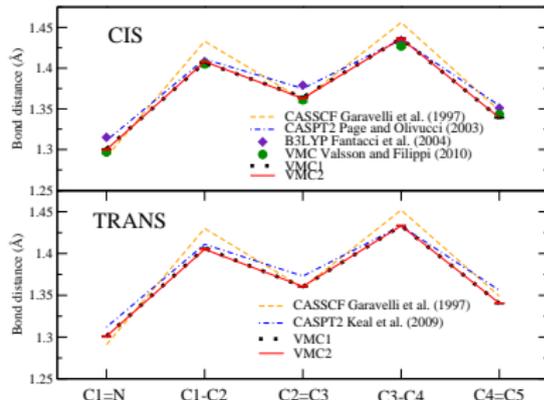
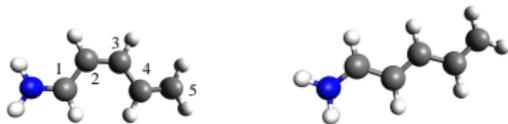
- $C, N = (3s2p)/(2s1p)$
- $H = (2s1p)/(1s1p)$



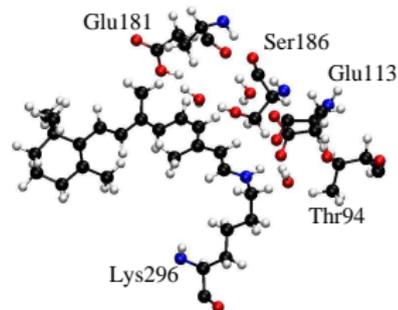
All parameters optimized!

Average structure from the equilibrium points

- $\langle \text{BLA} \rangle = \frac{\sum N_s \text{ single}}{N_s} - \frac{\sum N_d \text{ double}}{N_d}$

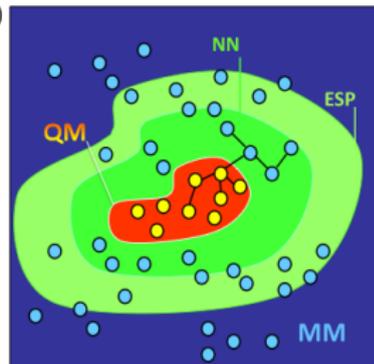


- VMC/MM: starting from 1HZX structure
- Our model: full protein (chain A), water, membrane (*n*-octane)
- QM/MM DFT (BLYP) annealing of the full system
- Glu181 negatively charged
- His211, Asp83 and Glu122 taken neutral
- MM: Amber/parm99 force field (TIP3P for water, OPLS for *n*-octane)
- Three cavity waters close to RPSB
- Full RPSB at VMC level (**VMC1**)

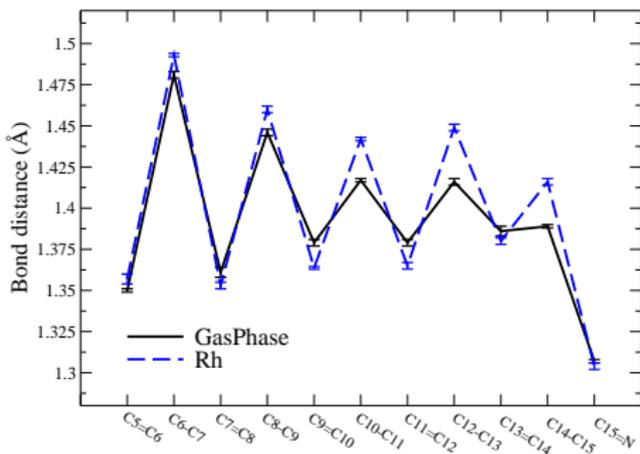


$$\begin{aligned}
 H &= H_{\text{QMC}} + H_{\text{MM}} + H_{\text{QMC/MM}} \\
 H_{\text{QMC/MM}} &= H_B + H_{\text{NB}} \\
 H_{\text{NB}} &= \sum_{i \in \text{MM}, j \in \text{QMC}} E_{\text{vdW}}(\mathbf{R}_{ij}) + \sum_i q_i \int d\mathbf{r} \frac{\rho(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_i|} v_i(|\mathbf{r} - \mathbf{R}_i|) \\
 &+ \sum_{i \in \text{MM}, j \in \text{QMC}} \frac{q_i z_j}{\mathbf{R}_{ij}} \\
 H_B &= \sum_{i \in \text{MM}, j \in \text{QMC}} [E_{\text{angles}} + E_{\text{dihedrals}}]
 \end{aligned}$$

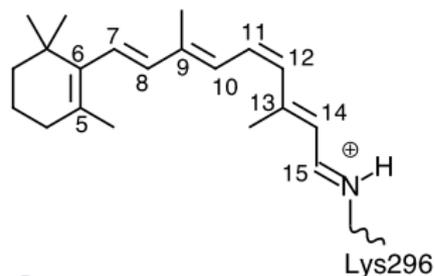
- MM atoms **fixed** → VMC geometry opt in classical field
- Classical field from CPMD (A. Laio *et al.*, *JCP*, **116**, 6941 (2002))
- Standard $E_{\text{angles}} + E_{\text{dihedrals}}$



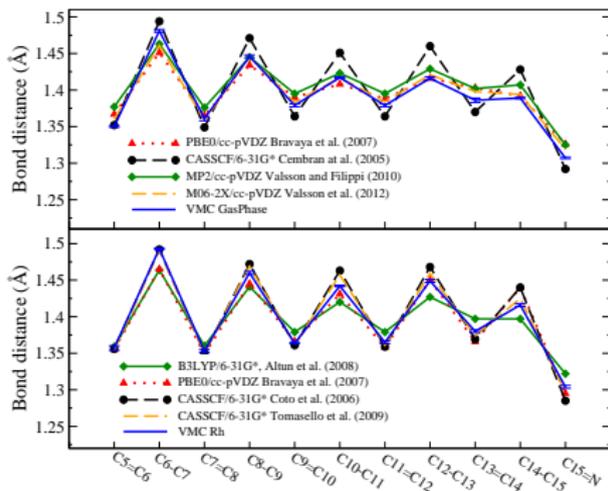
VMC on RPSB



	$\langle \text{BLA} \rangle$ (Å)	$\phi(\text{C}_5\text{-C}_6\text{-C}_7\text{-C}_8)$ ($^\circ$)
B3LYP	0.033	-33.5
PBE0	0.038	-39.5
MP2	0.044	-40.5
M06-2X	0.051	-38.0
CAM-B3LYP	0.053	-44.1
VMC	0.059(3)	-42(1)
CASSCF	0.101	-68.8
VMC Rh	0.088(3)	-43(1)

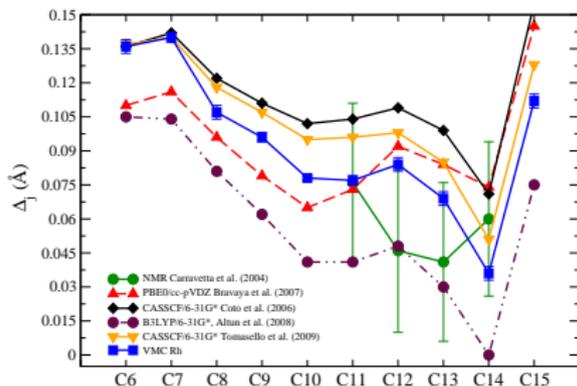


- Retinal Protonated Schiff Base (RPSB)
- $\text{C}_{21}\text{H}_{32}\text{N}^+$ (54 atoms, 120 valence electrons)



EC, D. Varsano and L. Guidoni, *JCTC*, **9**, 8 (2013)

- "Local" BLA: $\Delta_j = |\mathbf{R}_{j-1,j} - \mathbf{R}_{j,j+1}|$



- Blue shift due to the protein
- B3LYP: 0.28 (S_1) and 0.24 (S_2) eV
- Red shift due to geometrical effects
- B3LYP: 0.21 (S_1) and 0.16 (S_2) eV

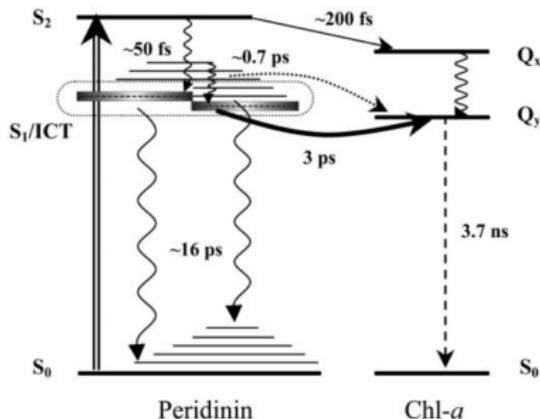
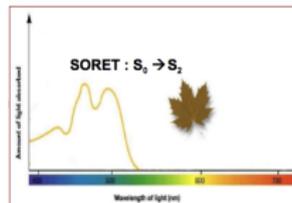
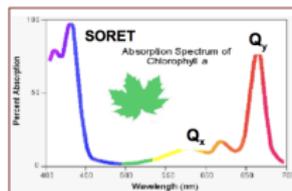
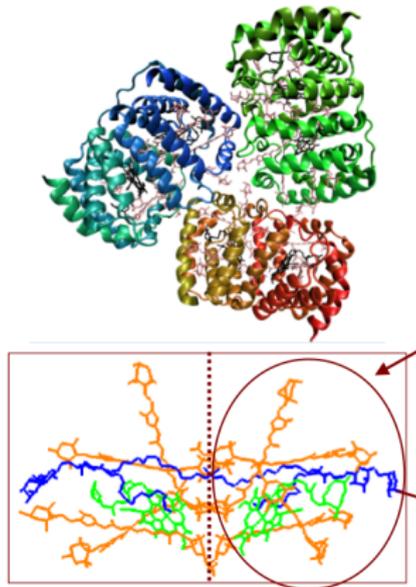
- TDDFT excited states S_1 and S_2 , 6-311+G* basis set

	Gas Phase (eV (nm) f)	Dist (eV (nm) f)	Rh (eV (nm) f)
S_1 BLYP	1.97 (629) 0.56	1.73 (717) 0.40	2.17 (571) 0.54
S_1 B3LYP	2.26 (549) 0.95	2.05 (605) 0.66	2.54 (488) 1.00
S_1 CAM-B3LYP	2.56 (484) 1.49	2.49 (498) 1.19	2.89 (429) 1.44
S_1 Expt.	2.03-2.34 (530-610)	-	2.48(1) (500(2))
S_2 BLYP	2.76 (449) 0.88	2.62 (473) 0.66	2.91 (426) 0.81
S_2 B3LYP	3.12 (397) 0.80	2.96 (419) 0.76	3.36 (369) 0.52
S_2 CAM-B3LYP	3.69 (336) 0.38	3.57 (347) 0.38	4.19 (296) 0.24
S_2 Expt.	3.18 (390)	-	3.27(1) (380(2))

Peridinin in PCP complex

Peridinin - Chlorophyll a - Protein (PCP)

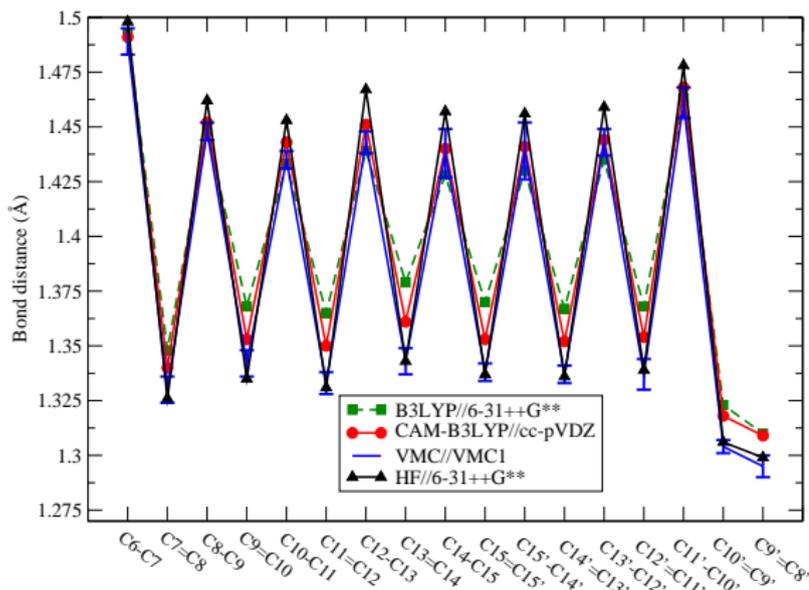
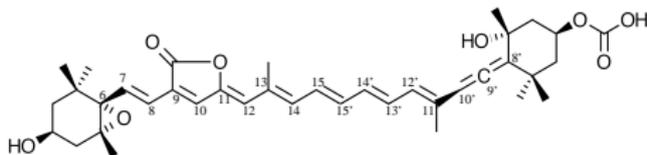
- Light-Harvesting complex (water soluble)
- Protein trimer from photosynthetic marine dinoflagellate *Amphidinium carterae*
- Two symmetric domains in each monomer
- Ratio 4/1 Peridinin/Chlorophyll a



E. Hofmann *et al.*, *Science*, **272**, 1788 (1996); K. Zigmantas *et al.*, *PNAS*, **99**, 16760 (2002)

PID: ground state structure

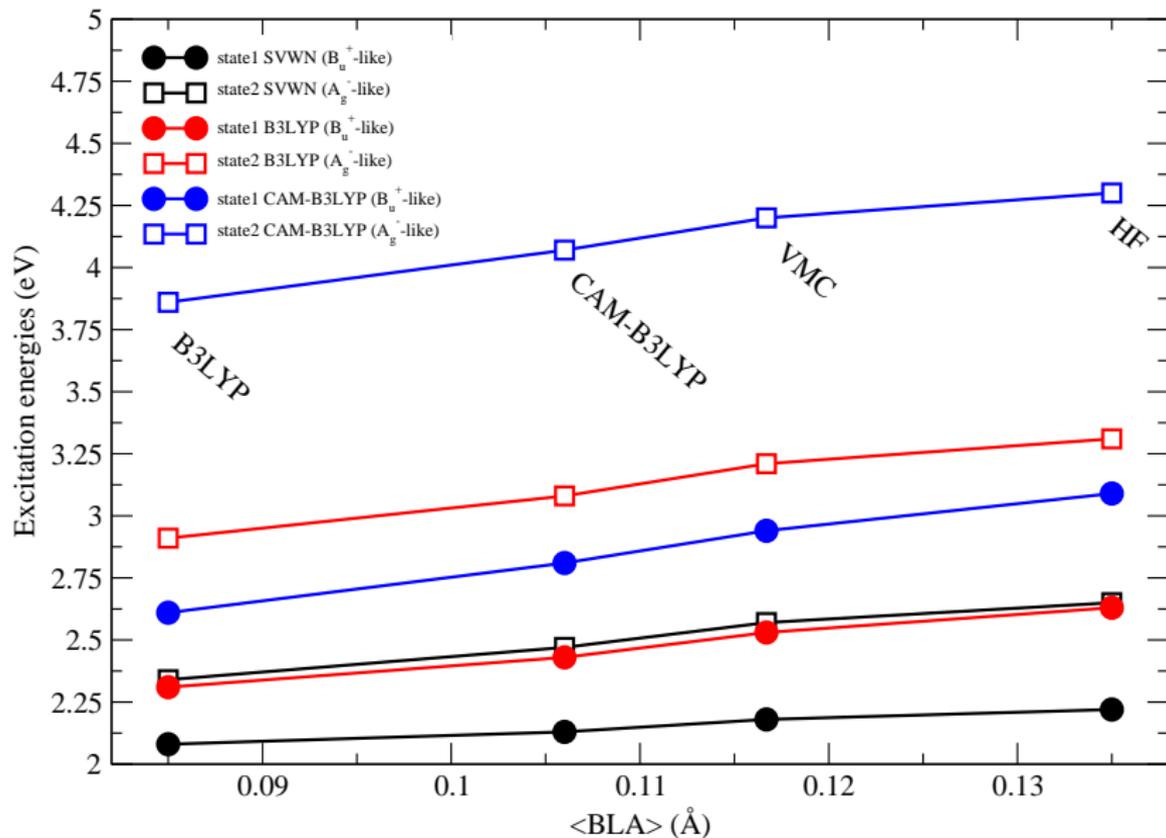
- Gas phase PID
- 6-31++G** for HF and B3LYP
- cc-pVDZ for CAM-B3LYP
- VMC1 for VMC



- $C_{39}H_{50}O_7$
- 96 atoms
- 248 valence electrons
- Highly substituted carotenoid

	(BLA) (Å)
B3LYP	0.085
CAM-B3LYP	0.106
VMC	0.1167(53)
HF	0.135

PID: TDDFT excited states

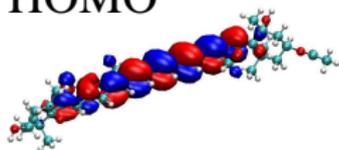


PID: TDDFT excited states

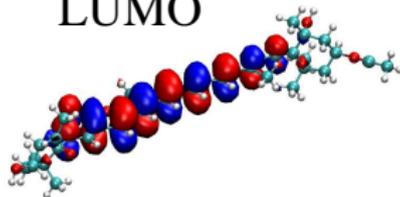
- Energies in eV
- Oscillator strengths in *italics>*
- Expt: 2.56 eV (bright) and 2.0-2.3 eV (dark)
- Wrong level ordering
- VMC + TD-B3LYP → good matching with expts for bright

Geometry	B3LYP	CAM	VMC	HF
TD-SVWN				
state1	2.08 (2.61)	2.13 (2.20)	2.19 (2.08)	2.23 (1.85)
state2	2.34 (0.61)	2.47 (0.81)	2.58 (0.80)	2.65 (0.91)
TD-B3LYP				
state1	2.31 (3.29)	2.44 (3.03)	2.53 (2.94)	2.63 (2.79)
state2	2.91 (0.31)	3.09 (0.48)	3.21 (0.54)	3.32 (0.62)
TD-CAM				
state1	2.61 (3.62)	2.82 (3.55)	2.94 (3.53)	3.10 (3.51)
state2	3.86 (0.09)	4.07 (0.02)	4.20 (0.00)	4.30 (0.03)

HOMO



LUMO



Geometry	B3LYP	CAM	VMC	HF
TD-B3LYP				
state1	H → L 0.71	H → L 0.70	H → L 0.70	H → L 0.70
state2	H-1 → L 0.54			
	H → L+1 0.45	H → L+1 0.44	H → L+1 0.44	H → L+1 0.44

EC, D. Varsano and L. Guidoni, in preparation

- QMC as mature technique applied to **biomolecules**
- Accurate, fully correlated ground state equilibrium structures
- **RPSB:**
 - Basis set study on the minimal model $C_5H_6NH_2^+$
 - First QMC/MM calculations
- **PID:**
 - Large chromophores: gas phase peridinin (preliminary study)
 - EOM-CCSD and MBPT calculations
- **QMC for chromophores:**
 - Dynamical correlation for geometry optimization of **biomolecules** (100-250 valence electrons)
 - Structural effects crucial in the spectral tuning of **RPSB** and **PID** absorption spectrum
 - Very good agreement with experimental data
 - Reasonable real time with HPC

Acknowledgements

- Prof. L. Guidoni
- Dr. Daniele Varsano
- CBBC group
- Prof. S. Sorella (ISAS, Trieste)
(TurboRVB)



Wavefunction optimization

- Statistical uncertainty
- Stochastic Reconfiguration technique

$$\begin{aligned}\alpha'_k &= \alpha_k + \delta\alpha_k \\ \delta\alpha_k &= \Delta t \sum_{k'} s_{k,k'}^{-1} f_{k'} \\ \Delta t &> 0 \\ f_{k'} &= -\frac{\partial E}{\partial \alpha_{k'}} \\ s_{k,k'} &= \langle O_k O_{k'} \rangle - \langle O_k \rangle \langle O_{k'} \rangle \\ O_k(x) &= \partial_{\alpha_k} \ln |\langle x | \Psi_T \rangle|\end{aligned}$$

- $s_{k,k'}$ to accelerate convergence
- Regularization of $s_{k,k}$

$$s_{k,k} \rightarrow s_{k,k}(1 + \epsilon)$$

$$E_{VMC}^{OPT} = \min_{\{\bar{\alpha}, \bar{\mathbf{R}}\}} E_{VMC} [\bar{\mathbf{R}}; \Psi_T(\{\bar{\alpha}, \bar{\mathbf{R}}\})]$$

$$\mathbf{F}_\alpha(\bar{\mathbf{R}}) = -\nabla_{\mathbf{R}_\alpha} E_{VMC} (\{\bar{\mathbf{R}}; \bar{\alpha}(\bar{\mathbf{R}})\})$$

- Finite difference approach

$$\mathbf{F}_\alpha(\bar{\mathbf{R}}) = - \lim_{\Delta \mathbf{R}_\alpha \rightarrow 0} \frac{E_{VMC}(\{\bar{\mathbf{R}}'; \bar{\alpha}(\bar{\mathbf{R}}')\}) + E_{VMC}(\{\bar{\mathbf{R}}; \bar{\alpha}(\bar{\mathbf{R}})\})}{\Delta \mathbf{R}_\alpha}$$

- where $\bar{\mathbf{R}}' = \bar{\mathbf{R}} + \Delta \mathbf{R}_\alpha$
- QMC energies affected by a stochastic error that propagates in the calculation of forces, increasing when $\Delta \mathbf{R}_\alpha \rightarrow 0$
- Finite difference approach usually coupled with the correlated sampling technique

Structural optimization (II)

- Local energy $E_L = \frac{\hat{H}\Psi_T}{\Psi_T}$
- Analytical derivatives

$$\mathbf{F}_\alpha(\bar{\mathbf{R}}) = -\frac{\partial}{\partial \mathbf{R}_\alpha} E_{VMC}(\{\bar{\mathbf{R}}; \bar{\alpha}(\bar{\mathbf{R}})\}) - \frac{\partial}{\partial \bar{\alpha}(\bar{\mathbf{R}})} E_{VMC}(\{\bar{\mathbf{R}}; \bar{\alpha}(\bar{\mathbf{R}})\}) \cdot \frac{d\bar{\alpha}(\bar{\mathbf{R}})}{d\mathbf{R}_\alpha}$$

- Second term $\frac{\partial E_{VMC}}{\partial \bar{\alpha}(\bar{\mathbf{R}})} = 0$ at minimum

$$\mathbf{F}_\alpha(\bar{\mathbf{R}}) = -\left\langle \frac{dE_L(\bar{\mathbf{x}})}{d\mathbf{R}_\alpha} \right\rangle_{\Pi(\bar{\mathbf{x}})} + 2 \left\{ \langle E_L(\bar{\mathbf{x}}) \rangle_{\Pi(\bar{\mathbf{x}})} \left\langle \frac{d \ln[\Psi_T(\bar{\mathbf{x}})]}{d\mathbf{R}_\alpha} \right\rangle_{\Pi(\bar{\mathbf{x}})} - \left\langle E_L \frac{d \ln[\Psi_T(\bar{\mathbf{x}})]}{d\mathbf{R}_\alpha} \right\rangle_{\Pi(\bar{\mathbf{x}})} \right\} = \mathbf{F}_\alpha^{H-F}(\bar{\mathbf{R}}) + \mathbf{F}_\alpha^P(\bar{\mathbf{R}})$$

- $\mathbf{F}_\alpha^{H-F}(\bar{\mathbf{R}}) \rightarrow$ Hellmann-Feynman term
- $\mathbf{F}_\alpha^P(\bar{\mathbf{R}}) \rightarrow$ Pulay term

Structural optimization (III)

- Space Warp Coordinate transformation reduces the variance on forces
- Each ionic displacement $\Delta \mathbf{R}_\alpha$ is followed by the translation of the electronic positions around the nuclei

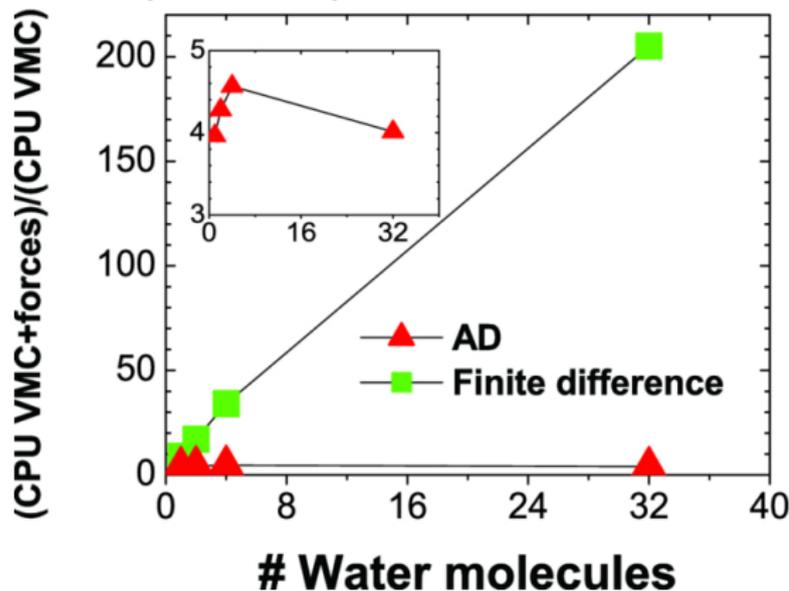
$$\begin{cases} \mathbf{r}'_i = \mathbf{r}_i + \Delta \mathbf{R}_\alpha \omega_\alpha(\mathbf{r}_i) \\ \omega_\alpha(\mathbf{r}_i) = \frac{F(r_{i\alpha})}{\sum_{b=1}^M F(r_{ib})} \end{cases}$$

- $F(r_{i\alpha}) = \frac{1}{r_{i\alpha}^4}$ with $r_{i\alpha} = |\mathbf{r}_i - \mathbf{R}_\alpha|$
- Unbounded variance near the nodal surface \rightarrow reweighting methods
- Guiding function $\Pi_\epsilon(\bar{\mathbf{x}}) = |\Psi_G(\bar{\mathbf{x}})|^2$

$$\Psi_G(\bar{\mathbf{x}}) = \frac{R_\epsilon(\bar{\mathbf{x}})}{R(\bar{\mathbf{x}})} \Psi_T(\bar{\mathbf{x}})$$

Adjoint Algorithm differentiation

- Derivative written using the chain rule as the propagation of the derivatives of simpler functions (polynomials, cosines and sines...etc)
- Following the chain rule intermediate results stored in memory and used to calculate other derivatives sharing the same intermediate values
- The computational overload for calculating forces does not have any linear dependence on the system size
- Optimizing wave functions and geometries of large molecular systems



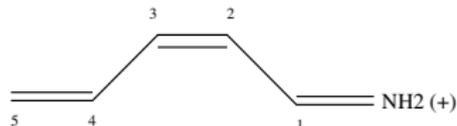
S. Sorella and L. Capriotti, *JCP*, **133**, 234111 (2010)

- 72 racks (9x8) - 73728 nodes (294912 cores)
- Rack: 2 midplanes a 16 nodeboards (4096 cores)
- 2304 Nodeboards a 32 compute nodes (128 cores)
- Overall peak performance: 1 Petaflops
- Main memory: 144 TB I/O Nodes: 600 (Connected to FORCE10 Switch)
- Compute Card/Processor: Power PC 450, 32-bit, 850 MHz, 4-way SMP L3 Cache: shared, 8 MB
- Networks:
 - Three-dimensional torus (compute nodes),
 - bandwidth per link: 425 MB/s (total: 5.1 GB/s), hardware latency: 100ns - 800ns
 - Global tree and collective (compute nodes, I/O nodes),
 - bandwidth per link: 850 MB/s (total 1.7 GB/s)
 - External: 10 GigE / Functional network (I/O Nodes)
- June 2009: 3 (Europe: 1)
Nov. 2009: 4 (Europe: 1)
June 2010: 5 (Europe: 1)
Nov. 2010: 9 (Europe: 2)

VMC validation on $C_5H_6NH_2^+$

CIS	C1=N	C1-C2	C2=C3	C3-C4	C4=C5	BLA
VMC1	1.3008(1)	1.4068(2)	1.3643(1)	1.4364(1)	1.3390(1)	0.0869(1)
VMC2	1.2999(1)	1.4079(1)	1.3629(2)	1.4364(2)	1.3392(1)	0.0881(3)
VMC	1.297(2)	1.405(3)	1.361(3)	1.427(2)	1.343(1)	0.082(3)
CASSCF	1.291	1.433	1.361	1.456	1.348	0.111
CASPT2	1.312	1.422	1.381	1.446	1.362	0.082
MP2	1.311	1.421	1.379	1.444	1.359	0.083
B3LYP	1.315	1.408	1.379	1.434	1.351	0.073
TRANS	C1=N	C1-C2	C2=C3	C3-C4	C4=C5	BLA
VMC1	1.3012(1)	1.4063(2)	1.3604(3)	1.4333(1)	1.3404(1)	0.0858(2)
VMC2	1.3010(2)	1.4051(2)	1.3607(3)	1.4331(1)	1.3406(1)	0.0850(1)
CASSCF	1.291	1.430	1.359	1.452	1.349	0.108
CASPT2	1.312	1.411	1.373	1.433	1.356	0.075

CIS	C2-C1=N	C3=C2-C1	C4-C3=C2	C5=C4-C3
VMC1	123.83(1)	122.60(1)	128.70(1)	119.86(1)
VMC2	123.54(1)	122.81(2)	128.75(1)	119.85(1)
VMC	123.9(2)	123.5(2)	128.9(1)	120.3(1)
CASSCF	123.0	123.6	128.6	121.4
CASPT2	123.1	122.9	128.5	120.1
MP2	123.4	122.9	128.7	119.8
B3LYP	124.0	123.8	128.9	120.5
TRANS	C2-C1=N	C3=C2-C1	C4-C3=C2	C5=C4-C3
VMC1	124.54(2)	119.34(1)	124.28(1)	120.88(1)
VMC2	124.50(2)	119.20(2)	124.25(1)	120.83(1)
CASSCF	124.0	120.1	124.2	122.1
CASPT2	124.1	119.4	124.2	121.0



- On a CASSCF geometry

ΔE (kcal/mol)	CASPT2	CASSCF	VMC1	VMC2	VMC3	VMC4
CIS/TRANS	3.5	3.39	4.27(8)	3.52(8)	3.85(8)	3.68(7)
CI/TRANS	54.3	59.6	55.28(9)	54.34(9)	54.71(7)	55.50(8)
μ TRANS (D)	-	-	4.58(1)	4.88(1)	4.64(1)	4.78(1)
μ CIS (D)	-	-	4.33(1)	4.36(1)	4.16(1)	4.28(1)

- VMC3 AGP from aug-cc-pVDZ

- C,N = (6s5p2d)/(3s3p2d)

- H = (5s2p)/(3s2p)

- VMC4 AGP from aug-cc-pVDZ (* for STOs)

- C,N = (4s2s*3p2p*2d*)/(3s3p2d)

- H = (3s2s*2p*)/(3s2p)

- Dipole and isomerization energies from VMC1 and VMC2 structures

- Convergence in AGP basis

- Reliable results even with VMC1

μ (D) VMC1	VMC1	VMC2	VMC3	VMC4
TRANS	4.21(1)	4.28(1)	4.33(1)	4.63(1)
CIS	3.92(1)	3.97(1)	3.98(1)	3.93(1)
μ (D) VMC2	VMC1	VMC2	VMC3	VMC4
TRANS	-	4.25(1)	4.33(1)	4.34(1)
CIS	-	4.01(1)	3.94(1)	3.92(1)

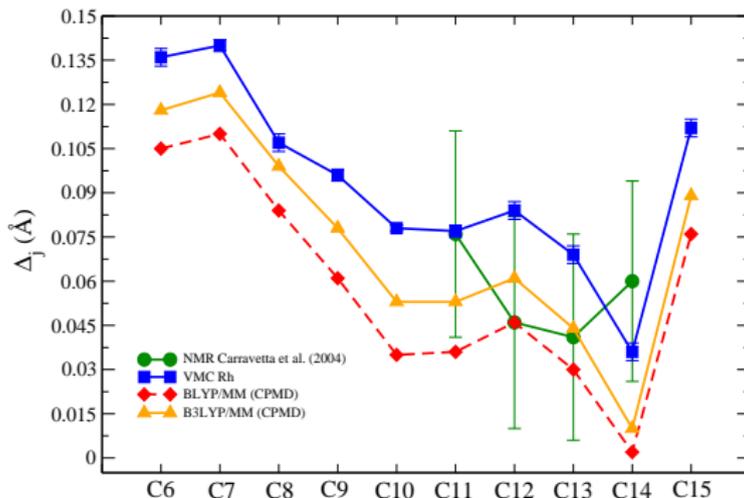
ΔE (kcal/mol)	VMC1	VMC2	VMC3	VMC4
VMC1	4.09(8)	3.57(8)	3.70(7)	3.68(7)
VMC2	-	3.80(8)	3.75(7)	3.66(7)

VMC on RPSB

- Relaxation from X-ray structure

	X-ray 1HZX	BLYP relaxed
N(Lys296) - O1(Glu113)	3.915	4.022
N(Lys296) - O2(Glu113)	3.597	2.713
C12 - O1(Glu181)	6.551	7.085
C12 - O2(Glu181)	4.438	5.104
N(Lys296) - O(Ser186)	4.202	4.330
N(Lys296) - O(Thr94)	4.986	5.169

- Same electrostatic coupling



Gas Phase

Excitation	ΔE (eV, (nm))	Kohn-Sham transition components
B3LYP		
S ₁	2.26 (549)	0.69 HOMO→LUMO
S ₂	3.12 (397)	0.66 HOMO-1→LUMO
BLYP		
S ₁	1.97 (629)	0.50 HOMO→LUMO
S ₂	2.76 (449)	0.46 HOMO-1→LUMO
CAM-B3LYP		
S ₁	2.56 (484)	0.69 HOMO→LUMO
S ₂	3.69 (336)	0.61 HOMO-1→LUMO

Rhodopsin

Excitation	ΔE (eV, (nm))	Kohn-Sham transition components
B3LYP		
S ₁	2.54 (488)	0.69 HOMO→LUMO
S ₂	3.36 (369)	0.67 HOMO-1→LUMO
BLYP		
S ₁	2.17 (571)	0.61 HOMO→LUMO
S ₂	2.91 (426)	0.56 HOMO-1→LUMO
CAM-B3LYP		
S ₁	2.89 (429)	0.69 HOMO→LUMO
S ₂	4.19 (296)	0.67 HOMO-1→LUMO

- Excited states on gas phase structures

eV	PBE0	MP2	M06-2X	CAM-B3LYP
B3LYP				
S ₁	2.28	2.21	2.25	2.26
S ₂	3.09	3.05	3.12	3.09
BLYP				
S ₁	2.06	1.97	2.00	1.96
S ₂	2.73	2.71	2.77	2.75
CAM-B3LYP				
S ₁	2.45	2.43	2.49	2.54
S ₂	3.66	3.59	3.66	3.64

- Ground state Ψ_{CC}^g
- CCSD $\rightarrow \mathbf{T}_1$ and \mathbf{T}_2 (N^6)

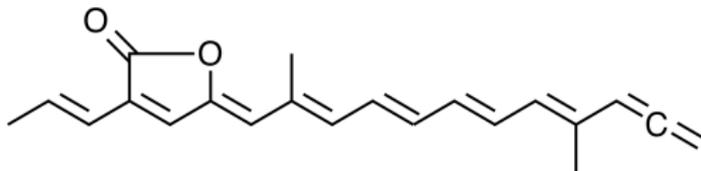
$$\begin{aligned}\Psi_{CC}^g &= e^{\mathbf{T}} \phi_0 \\ e^{\mathbf{T}} &= 1 + \mathbf{T} + \frac{1}{2} \mathbf{T}^2 + \frac{1}{6} \mathbf{T}^3 + \dots = \sum_{k=0}^{\infty} \frac{1}{k!} \mathbf{T}^k \\ \mathbf{T} &= \mathbf{T}_1 + \mathbf{T}_2 + \mathbf{T}_3 + \dots \mathbf{T}_N \\ \mathbf{T}_1 \phi_0 &= \sum_i^{occ} \sum_a^{vir} t_i^a \phi_i^a \\ \mathbf{T}_2 \phi_0 &= \sum_{i < j}^{occ} \sum_{a < b}^{vir} t_{ij}^{ab} \phi_{ij}^{ab}\end{aligned}$$

- Excited State Ψ_{CC}^x
- Exact for two-electron systems
- Single-reference method

$$\begin{aligned}\Psi_{CC}^x &= \mathbf{R} \Psi_{CC}^g \\ \mathbf{R} &= \mathbf{R}_1 + \mathbf{R}_2 + \mathbf{R}_3 + \dots \mathbf{R}_N \\ \mathbf{H}(\mathbf{R} e^{\mathbf{T}} \phi_0) &= E(\mathbf{R} e^{\mathbf{T}} \phi_0) \\ e^{-\mathbf{T}} \mathbf{H} e^{\mathbf{T}} (\mathbf{R} \phi_0) &= E(\mathbf{R} \phi_0) \\ \bar{\mathbf{H}} &\equiv e^{-\mathbf{T}} \mathbf{H} e^{\mathbf{T}}\end{aligned}$$

PID1: TDDFT excited states

- Reduced model (PID1) $C_{21}H_{22}O_2$
- 45 atoms and 118 valence electrons



Geometry	B3LYP	VMC
TD-SVWN		
state1	2.18 (1.92)	2.28 (1.62)
state2	2.40 (0.88)	2.65 (1.01)
TD-B3LYP		
state1	2.40 (2.78)	2.62 (2.50)
state2	2.95 (0.27)	3.25 (0.44)
TD-CAM		
state1	2.69 (3.10)	3.02 (3.03)
state2	3.89 (0.09)	4.26 (0.02)

Geometry	B3LYP	VMC
TD-B3LYP		
state1	H → L 0.71	H → L 0.70
state2	H-1 → L 0.52	H-1 → L 0.53
	H → L+1 -0.46	H → L+1 0.45

- Representative model → expensive EOM-CCSD and MBPT calculations

PID1: EOM-CCSD approach

- Energy window for active space
- Mixing between σ and π orbitals
- Excitation energies with respect to MP2 ground state
- No clear convergence
- Bright state: $H \rightarrow L$ dominant (same as in TDDFT)
- Pseudo-dark state:
 - $H-1 \rightarrow L$
 - $H \rightarrow L+1$
 - $H^2 \rightarrow L^2$ (double excitation, missing in TDDFT)

16+16	B3LYP	VMC
3-21G	1.90 (0.06)	2.66 (0.48)
	2.47 (2.99)	3.02 (2.75)
6-31G	2.01 (0.09)	2.75 (0.71)
	2.50 (3.04)	3.07 (2.53)
6-311G	2.95 (2.39)	3.46 (3.29)
	3.22 (1.68)	3.98 (0.79)
D95	2.15 (0.22)	2.85 (1.54)
	2.50 (3.08)	3.14 (1.85)
cc-pVDZ	2.89 (2.53)	3.39 (3.37)
	3.12 (1.34)	3.88 (0.52)
cc-pVTZ	3.14 (3.51)	3.63 (3.73)
	3.52 (0.63)	4.29 (0.39)

8+8	B3LYP	VMC
3-21G	2.73 (0.90)	3.45 (3.00)
	3.01 (2.98)	3.91 (1.05)
6-31G	2.92 (1.85)	3.43 (2.98)
	3.20 (2.12)	3.93 (1.00)
6-311G	3.27 (2.71)	3.77 (3.32)
	3.62 (1.47)	4.38 (0.83)
D95	2.77 (1.19)	3.46 (3.06)
	3.03 (2.64)	3.95 (0.93)
cc-pVDZ	3.28 (3.36)	3.77 (3.66)
	3.70 (0.85)	4.47 (0.52)
cc-pVTZ	3.47 (3.65)	3.94 (3.79)
	3.92 (0.58)	4.70 (0.39)

32+32	B3LYP	VMC
3-21G	1.03 (0.12)	1.74 (0.78)
	1.51 (2.69)	2.07 (2.20)
6-31G	1.23 (0.17)	1.90 (1.07)
	1.64 (2.71)	2.22 (1.92)
6-311G	1.68 (0.24)	2.35 (1.38)
	2.04 (2.86)	2.65 (1.81)
D95	1.96 (0.68)	2.55 (2.12)
	2.24 (2.60)	2.92 (1.23)
cc-pVDZ	1.84 (0.64)	2.44 (2.33)
	2.09 (2.64)	2.76 (1.01)
cc-pVTZ	-	-
	-	-

EC, D. Varsano and L. Guidoni, in preparation