Second-harmonic generation in GFP-like proteins

Supporting information

Inge Asselberghs,[1] Cristina Flors,[1] Lara Ferrighi,[2] Edith Botek,[2] Benoît Champagne,[2] Hideaki Mizuno,[3] Ryoko Ando,[3] Atsushi Miyawaki,[3] Johan Hofkens,[1] Mark Van der Auweraer,[1] Koen Clays[1].

[1]Department of Chemistry and Institute for Nanoscale Physics and Chemistry (INPAC), Katholieke Universiteit Leuven, Celestijnenlaan 200D and F, B-3001 Leuven

[2]Facultés Universitaires Notre Dame de la Paix, Namur, Belgium.

[3]Laboratory for Cell Function and Dynamics, Brain Science Institute, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198 Japan.

Expanded theoretical results

The GFP has a complex structure, in particular when considering the residues forming its surrounding. Since the GFP properties depend on the structure of both the chromophore and the surrounding residues, a careful optimization strategy is needed. Thus, the structures have been successively enlarged to include the side chains and the residues around the chromophore (**Chart SI1**), in particular, the residues involved in the proton transfer from the Tyr66 to Ser65 residues. This transfer takes place mainly through the hydrogen bonding network involving W22, Ser205, and Glu222, but the overall process also leads to changes in other residues including His148, Asn146, and Thr203 (**Chart SI2**). In particular, the last one switches conformation with either a carbonyl or a hydroxyl group facing the chromophore depending on the protonation state of the chromophore itself.[1] Furthermore, the other residues in the surrounding also participate in the stabilization of the network and permit the protein to adopt two stable forms.

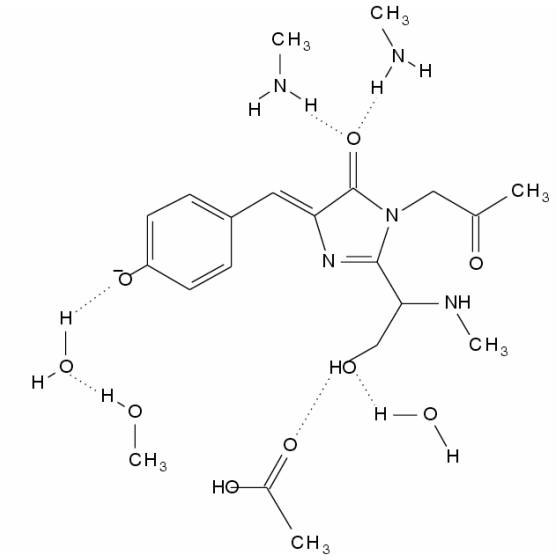
The more complex model **E** (see **Chart SI1**) is built from model **D** by adding the Arg96 and Gln94 represented by methylamine molecules while model **F** is also obtained from model **D** by including the His148 residue, which stabilizes by H-bond the phenolate. Like for models **G** and **H**, only the deprotonated form of model **F** is considered. In model **G**, one also considers the stabilization of the phenolate by Thr203. Finally, models **E** and **G** are combined to obtain model **H** (built from model **D** by adding the His148, Thr203, Arg96, and Gln94 residues), the most elaborated one considered here. Although some approximations are needed to ensure computational feasibility the successive model compounds aim at defining the proper structure to represent the GFP active domain and its properties, the first hyperpolarizability β as well as related linear optical properties.

Comparing models **B** and **A** for the deprotonated form, the electron delocalization along the path linking the O_1 and O_2 atoms increases. Indeed, the double bonds become larger, the single bonds get smaller so that the bond length alternation (BLA) decreases (see **Table SI1**). The BLA further decreases by going from model **B** to model **C**. On the other hand, the inclusion of the hydrogen bonding network (in models **D**, …) leads to the opposite behaviors associated with a reduction of the delocalization of the O atom lone pair into the benzene ring and therefore a reduction of the

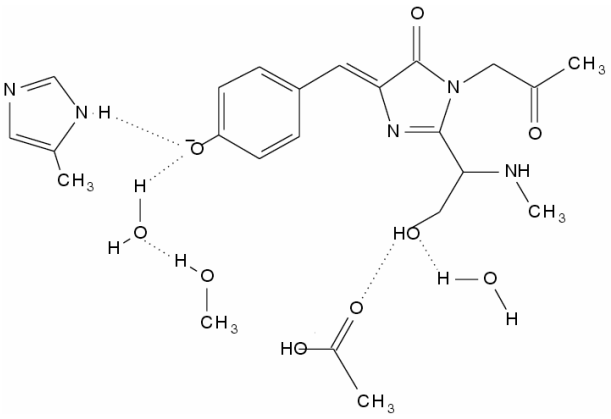
quinonoïd character. Similar stabilization of the benzenoid form by surrounding molecules was already evidenced by Laino *et al* when a water molecule is connected to Tyr66.[2] The BLA of the deprotonated form results thus from the balance between the residues which stabilize the 6- and 5-membered rings of the chromophore. For the protonated species, the changes in geometry are smaller, this explains why we have limited our investigation to models **A**-**E**.

Before considering the NLO properties, the absorption spectra of the successive GFP active site models were simulated. These are sketched in **Figure SI1** for the different pairs of protonated and deprotonated forms whereas the essential data are listed in **Table SI2**. The calculated spectra can be compared with experimental data in **Figure SI2**. For both forms, the linear absorption spectra change little when including more and more surrounding residues. The comparison between models **A** and **B** demonstrate the importance of considering the Gly67 and Ser65 residues, which lead to a reduction of the excitation energy by about 0.1 eV. On the other hand, models **B**-**D** present comparable absorption spectra. Going to model **E** is accompanied by bathochromatic shifts (8 and 9 nm for the protonated and deprotonated forms, respectively). Like for the electron delocalization and BLA, going from **E** to **F** and **G** results in the opposite behavior, *i.e.* a hypsochromatic shift associated with the H-bond stabilization of the phenolate. This stabilization is also associated with an increase of intensity (f). This follows the trend reported by Laino *et al*. and the experimental observation that mutants without one of the hydrogen bonds on the phenolate give red-shifted absorptions.[2] Although theory reproduces the bathochromatic shift accompanying the deprotonation, using the most elaborated model, it overestimates the excitation energies, by 0.20 eV for the protonated species and by 0.50 eV for the deprotonated one. Moreover, the oscillator strength (and the associated transition dipole moment) is larger for the deprotonated form than for the protonated species, no matter which chromophore model is considered. This is in agreement with experimental observation and earlier reports.[3] Similar results were also obtained for the ordering of the excitation energies and the oscillator strengths when employing the Configuration Interaction Singles (CIS) scheme, either *ab initio* (6-311G** basis set) or semi-empirically (ZINDO), or at the CIS(D) level of approximation for the excitation energies.

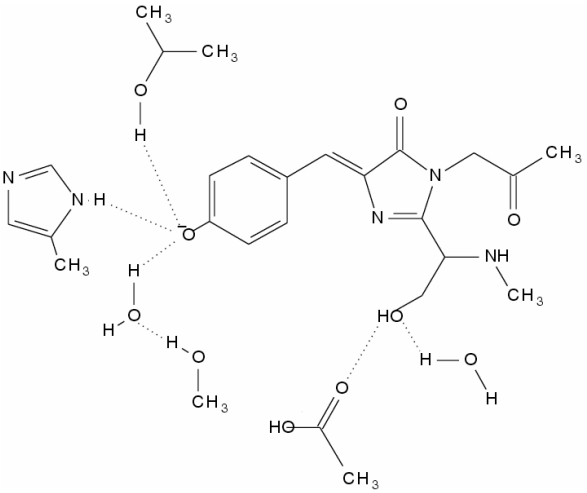
A representation of the HOMO and LUMO orbitals, responsible for the main peaks in the absorption spectra of models **A** are given in **Figure SI3**. The major absorption band of the protonated species is dominated by a HOMO to LUMO transition. On the other hand, for the deprotonated species, the dominant character is HOMO to LUMO for model **A** and HOMO to LUMO + 1 for model **B**. Indeed, the LUMO level of the deprotonated form of model **B** is located on the CH_2-CO-R chain attached to the imidazolone ring while the LUMO + 1 presents the same pattern as LUMO of model **A**. For model **C**-**H**, HOMO to LUMO is again the dominant characteristic of the lowest-energy absorption band and their shapes are similar to those of Model **A**.



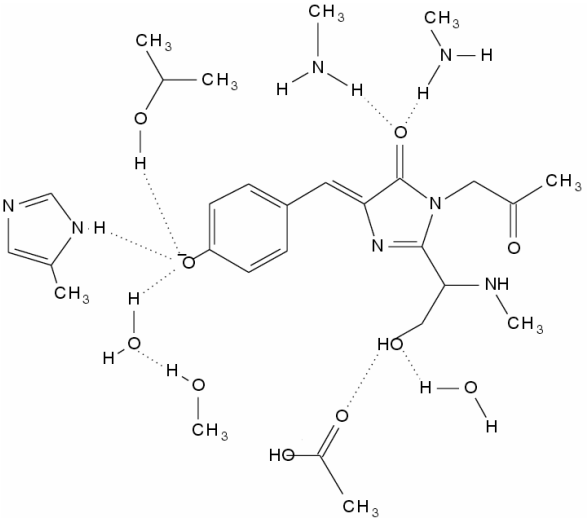
E



F



G



H

Chart SI1: Sketch of the next set of successive models (**E**, **F**, **G**, and **H**) of the GFP chromophore and its surrounding.

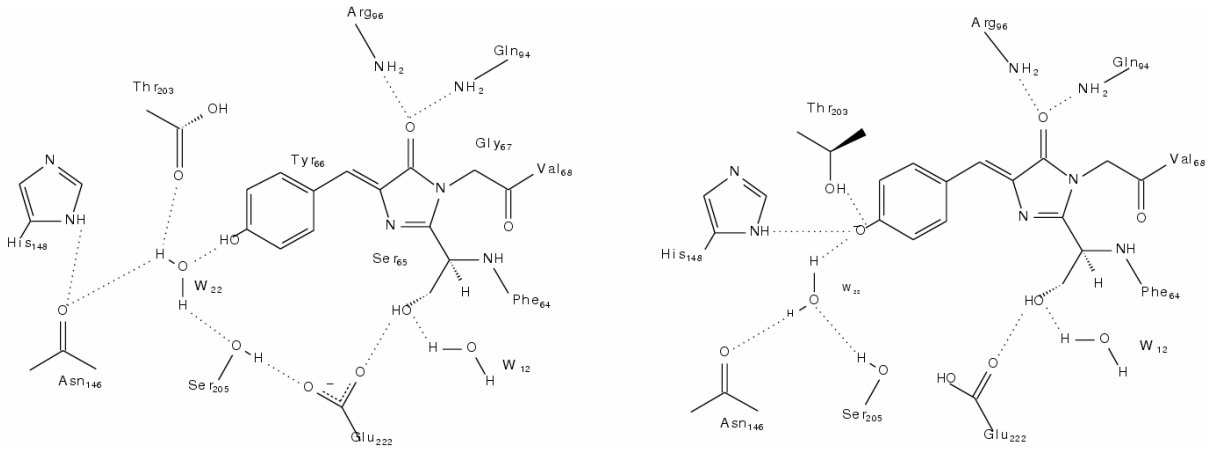


Chart SI2: Sketch of the different residues involved in the proton transfer between the protonated (left) and deprotonated (right) forms.

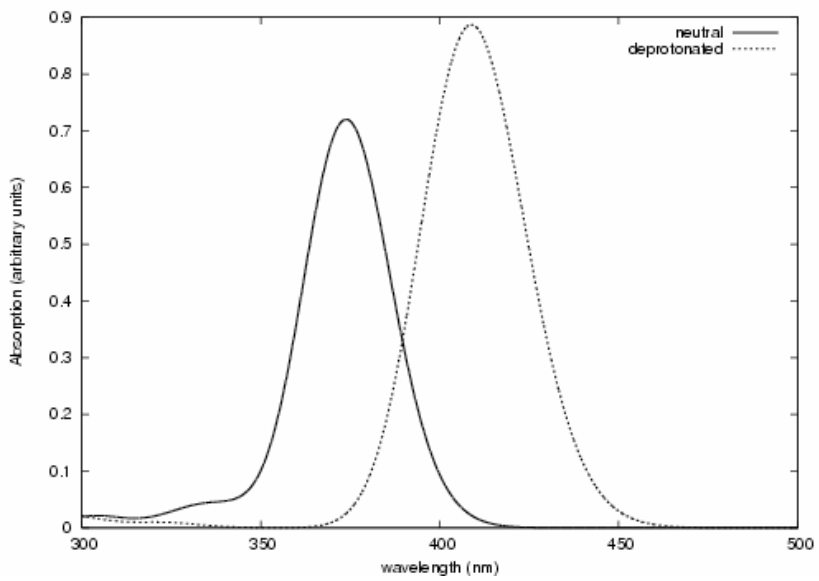


Figure SI1. TDDFT/B3LYP/6-311G** absorption spectra of models E.

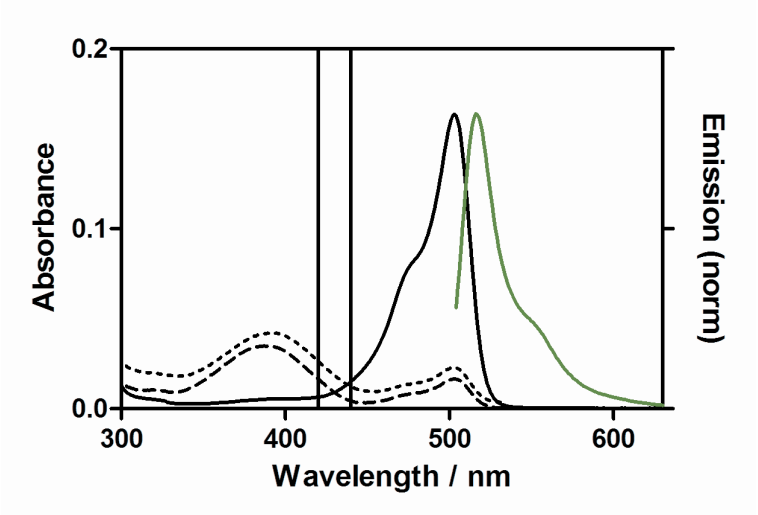


Figure SI2. Absorption spectra of Dronpa at pH 8 before photoconversion (deprotonated) (—), after photoconversion (protonated) (---), and at pH 5 (···). Emission spectrum at pH 8 before photoconversion (λ_{ex} = 488 nm, green). Vertical lines represent the second-harmonic of the wavelengths used in the experiments.

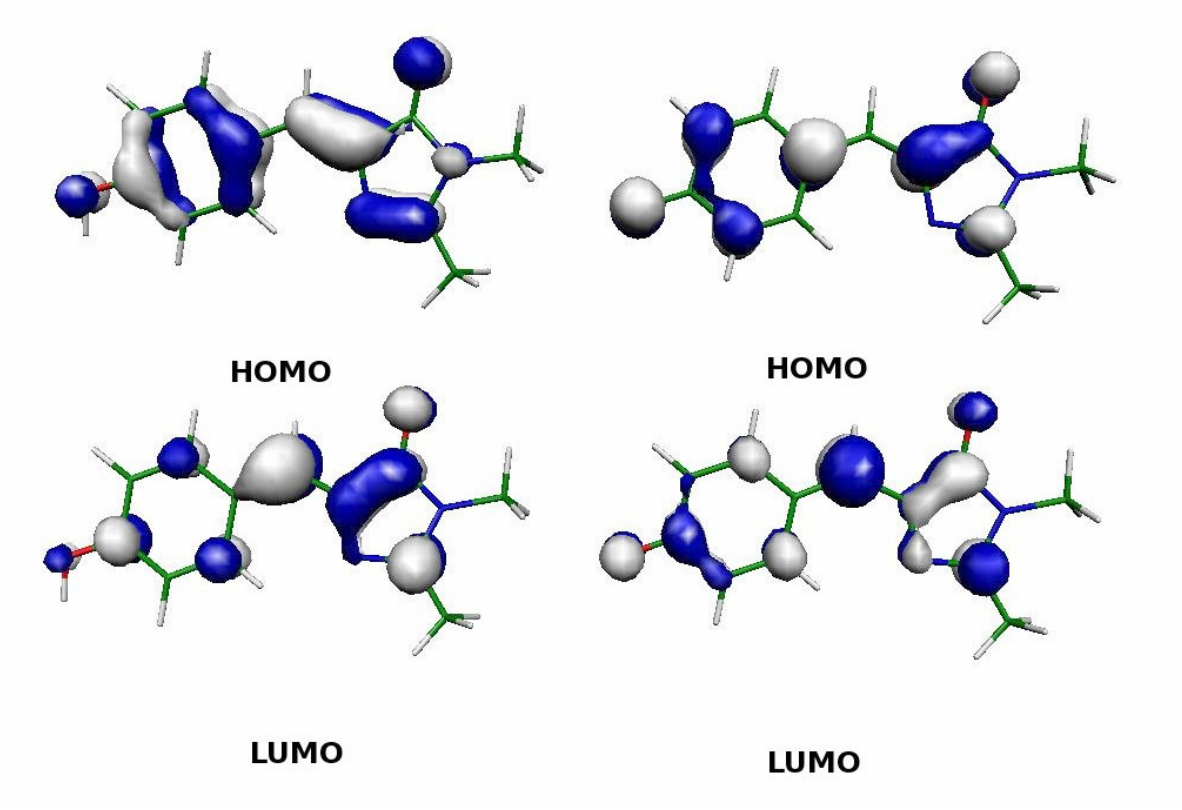


Figure SI3. Frontiers orbitals characterizing the dominant low-energy absorption band of the protonated (left) and deprotonated (right) forms of model **A** of the GFP chromophore.

Table SI1

	A		B		C		D		E		F	G	H	Exp
Bond	Proton.	Dep.	Proton.	Dep.	Proton.	Dep.	Proton.	Dep.	Proton.	Dep.	Dep.	Dep.	Dep.	
O1-C1	1.362	1.247	1.359	1.245	1.357	1.242	1.344	1.266	1.341	1.263	1.283	1.298	1.295	1.385
C1-C2	1.398	1.458	1.398	1.459	1.399	1.461	1.405	1.443	1.406	1.445	1.432	1.426	1.428	1.374
C2-C3	1.384	1.365	1.383	1.363	1.383	1.361	1.383	1.370	1.382	1.369	1.375	1.377	1.376	1.385
C3-C4	1.410	1.432	1.411	1.434	1.411	1.437	1.411	1.426	1.413	1.428	1.420	1.418	1.419	1.394
C4-C5	1.410	1.434	1.410	1.436	1.410	1.438	1.413	1.426	1.414	1.428	1.422	1.420	1.422	1.401
C5-C6	1.385	1.365	1.385	1.363	1.384	1.361	1.382	1.373	1.389	1.370	1.376	1.397	1.377	1.375
C6-C1	1.400	1.460	1.400	1.462	1.402	1.463	1.407	1.449	1.408	1.450	1.437	1.431	1.432	1.379
C4-C7	1.448	1.407	1.446	1.402	1.444	1.397	1.443	1.417	1.440	1.414	1.425	1.430	1.427	1.457
C7-C8	1.355	1.386	1.356	1.391	1.357	1.396	1.358	1.377	1.361	1.381	1.370	1.367	1.371	1.366
C8-C9	1.488	1.457	1.483	1.449	1.480	1.443	1.483	1.461	1.478	1.453	1.467	1.471	1.464	1.476
C9-N2	1.405	1.418	1.411	1.425	1.402	1.412	1.405	1.415	1.395	1.402	1.412	1.411	1.399	1.380
N2-C10	1.390	1.390	1.393	1.392	1.389	1.386	1.405	1.399	1.409	1.399	1.400	1.400	1.400	1.369
C10-N1	1.298	1.298	1.300	1.302	1.302	1.305	1.290	1.294	1.289	1.297	1.293	1.293	1.296	1.315
N1-C8	1.402	1.403	1.401	1.400	1.402	1.399	1.401	1.402	1.400	1.400	1.403	1.403	1.400	1.418
C9-O2	1.215	1.232	1.214	1.232	1.221	1.244	1.220	1.227	1.228	1.242	1.224	1.223	1.235	1.227
N2-R1	1.450	1.441	1.441	1.432	1.457	1.452	1.437	1.432	1.439	1.445	1.433	1.434	1.447	-
C10-R2	1.490	1.498	1.506	1.504	1.511	1.507	1.505	1.503	1.505	1.512	1.504	1.504	1.513	1.470
BLA	0.093	0.021	0.090	0.011	0.084	0.001	0.085	0.040	0.079	0.033	0.055	0.063	0.056	0.081

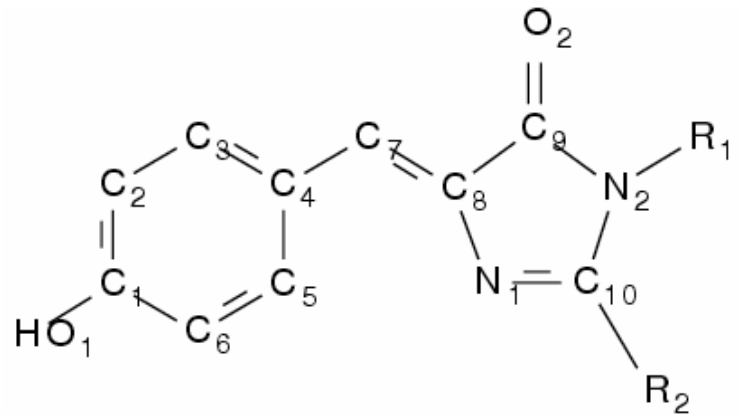


Table SI1: B3LYP/6-311G** optimized geometrical parameters for the successive models in comparison with the X-Ray structure of Model A. BLA is defined as the difference between the C_4-C_7 and C_7-C_8 bond lengths.

Table SI2. Main characteristics of the low-energy excited states of the protonated and deprotonated forms of the GFP chromophore: comparison between the successive models **F**-**H**. The excitation energies ΔE are given in eV, the wavelengths λ in nm, H = HOMO, and L = LUMO, f is the oscillator strength.

	protonated			**deprotonated**		
	ΔE (λ)	f	dominant character	ΔE (λ)	f	dominant character
Model **A**	3.52 (352)	0.690	H → L	3.16 (392)	0.972	H → L
	4.21 (294)	0.079	H - 2 → L	4.30 (288)	0.04	H → L + 2
Model **B**	3.39 (366)	0.510	H → L	3.09 (401)	0.979	H → L + 1
	3.58 (346)	0.251	H -1 → L	3.55 (348)	0.039	H → L + 2
Model **C**	3.39 (365)	0.692	H → L	3.10 (400)	1.040	H → L
	3.63 (342)	0.029	H - 1 → L			
Model **D**	3.38 (367)	0.742	H → L	3.10 (400)	0.939	H → L
	3.51 (354)	0.014	H - 2 → L			
Model **E**	3.31 (375)	0.472	H → L	3.03 (409)	0.658	H → L
	3.34 (371)	0 .255	H - 2 → L	3.05 (407)	0.231	H → L + 1
Model **F**				3.13 (396)	1.030	H → L
Model **G**				3.12 (397)	0.0573	H → L + 1
				3.14 (394)	0.9816	H → L
Model **H**				3.05 (407)	0.9768	H → L

(1) Kummer, A. D.; Wiehler, J.; Rehaber, H.; Kompa, C.; Steipe, B.; Michel-Beyerle, M. E. *J. Phys. Chem. B* **2000**, *104*, 4791-4798.

(2) Laino, T.; Nifosi, R.; Tozzini, V. *Chem. Phys.* **2004**, *298*, 17-28.

(3) Brejc, K.; Sixma, T. K.; Kitts, P. A.; Kain, S. R.; Tsien, R. Y.; Ormo, M.; Remington, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 2306-2311.

Full details of reference 45 in main text:

Gaussian 03, Revision C.02, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian, Inc., Wallingford CT, 2004.