Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Electronic Supplementary Information for: A high-throughput computational approach to UV-Vis spectra in protein mutants

Mirko Paulikat,[†] Ricardo A. Mata,[*,†] and Ricard Gelabert[*,‡]

†*Institute of Physical Chemistry, University of Goettingen, Tammannstraße 6, D-37077 Göttingen, Germany*

‡*Departament de Química, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain*

E-mail: rmata@gwdg.de; ricard.gelabert@uab.cat

1 Full List of Mutations

We provide in Table S1 a full list of the runs carried out with our high-throughput approach to the mCardinal system. A selected few have been discussed in the main manuscript. The others are shown here for completeness. The expected ordering (noted in the leftmost column) follows the Scheme we highlighted in the text. Not all mutations were carried out with isosteric residues. This leads to some outliers. The limitations of the method have been highlighted with the example of Y115K.

Table S1: Spectral shifts estimated for a larger palette of single mutants of mCardinal. The results are derived from the small set of structures (100). Results in "Non-Optimized" column correspond to the best rotamer of the mutant produced by PyMol. Results in "Optimized" column are derived selecting the best five rotamers of each structure produced by PyMol, optimizing their structure and retaining only the most stable of them.

Attempted Shift	Mutant	Non-Optimized $\langle\Delta\omega\rangle$ / eV	Optimized $\langle\Delta\omega\rangle$ / eV
Bathochromic	L13N	–0.06	–0.05
	M158E	–0.41	–0.38
	Q41K	–1.08	–1.09
	R195F	–0.50	–0.49
	R195W	–0.49	–0.46
	T141V	–0.04	–0.04
	Y115K	–0.29	–0.34
	Y115N	–0.01	+0.02
	Y115Q	–0.01	–0.00
Isochromic	M158L	–0.02	-0.02
	T141S	–0.09	+0.00
Hypsochromic	L197N	–0.01	+0.08
	L197S	–0.02	+0.06
	L197T	–0.01	+0.03
	M158K	+0.36	+0.38
	M158N	–0.09	+0.06
	M158Q	+0.01	+0.03

2 Applying PT Approach to Crystal Structure

To explore the feasibility of applying the perturbational method directly on a single structure and avoid the most costly part of the calculation, namely the integration of a MD simulation to draw structures from it, we have applied the method directly on the crystal structure deposited on the Protein Data Bank (after completing the structure if missing atoms should be detected). We also present the application of the perturbative method on a more realistic system obtained after solvation of the crystal structure in a cube of water molecules with sides of 76 Å, and addition of ions to ensure electroneutrality. In both cases, optimization of the mutated residue has been carried out. Excitation energy shifts for the best rotamer are presented in Table S2.

Table S2: Spectral shifts estimated for a larger palette of single mutants of mCardinal. The results have been obtained from the most stable rotamer produced by PyMOL on the crystal structure described by the PDB file after completing missing atoms, and from the same structure after solvation and electric neutralization of the cell.

Attempted Shift	Mutant	PDB Structure $\langle\Delta\omega\rangle$ / eV	PDB Structure + Solvation $\langle\Delta\omega\rangle$ / eV
Bathochromic	L13N	–0.03	–0.03
	M158E	–0.80	–0.80
	Q41K	–0.27	–0.27
	R195F	+0.30	–0.32
	R195W	+0.31	–0.33
	T141V	–0.63	–0.61
	Y115K	–0.39	–0.34
	Y115N	+0.05	+0.05
	Y115Q	+0.05	+0.05
Isochromic	M158L	–0.08	–0.08
	T141S	+0.10	–0.10
Hypsochromic	L197N	+0.05	–0.05
	L197S	–0.14	–0.14
	L197T	+0.07	–0.06
	M158K	+1.54	+0.95
	M158N	–0.02	–0.02
	M158Q	+0.10	+0.10

In general, the results agree qualitatively (in the sense of the direction induced by the

mutation on the spectral shift) with those derived from a larger set of structures (q.v. Table S1. Nevertheless, some mutations are predicted opposite effects (e.g. L197S, M158N, M158L) or yield conflicting trends depending on whether the crystal structure or the solvated system are used (e.g. R195F, R195W, L197T). For the systems in which full MD runs were calculated (L13N, Y115K, L197N and M158Q) the agreement was always improved with the perturbative approach.

3 Prediction of Full Absorption Spectra

The procedure described in the manuscript can be used to provide an estimate of the absorption spectra of a given mutant, using only information about the non-mutated protein. To achieve this, it suffices to compute the density difference matrices for all structures between ground and as many excited electronic states as desired, $\{\Delta\mathbf{D}^{0i}\}_{i=1\div N}$. Aftwerwards, Eq 7 should be applied for all structures and each excitation using Eq 9 with the corresponding density difference matrix. This provides the PT estimate of the excitation energies, but the values of the oscillator strengths are also needed to simulate the full spectrum. The computation of the PT-estimate of the oscillator strengths is more involved, so that, in order to provide just an estimate of the spectrum a possibility is to take the value of the oscillator strength of the non-mutated protein, which is available as a side product of the density difference matrix calculation.

We have computed the absorption spectra estimates of mutants L13N, M158Q, Y115K and L198N using this procedure using a set of one thousand structures drawn from the MD simulation of mCardinal. Results are shown in Figure S1. To make meaningful comparisons, we have also built the spectra using the more standard procedure of histogramming excitation energies described in the Computational Details section for the same mutants and the original protein. In these cases, the MD simulations are 10 ns in length and a total of 1,000 structures have been sampled from each. For each of these structures a total of 5 excited states have

been computed using TD-DFT within the Tamm-Dancoff approximation.

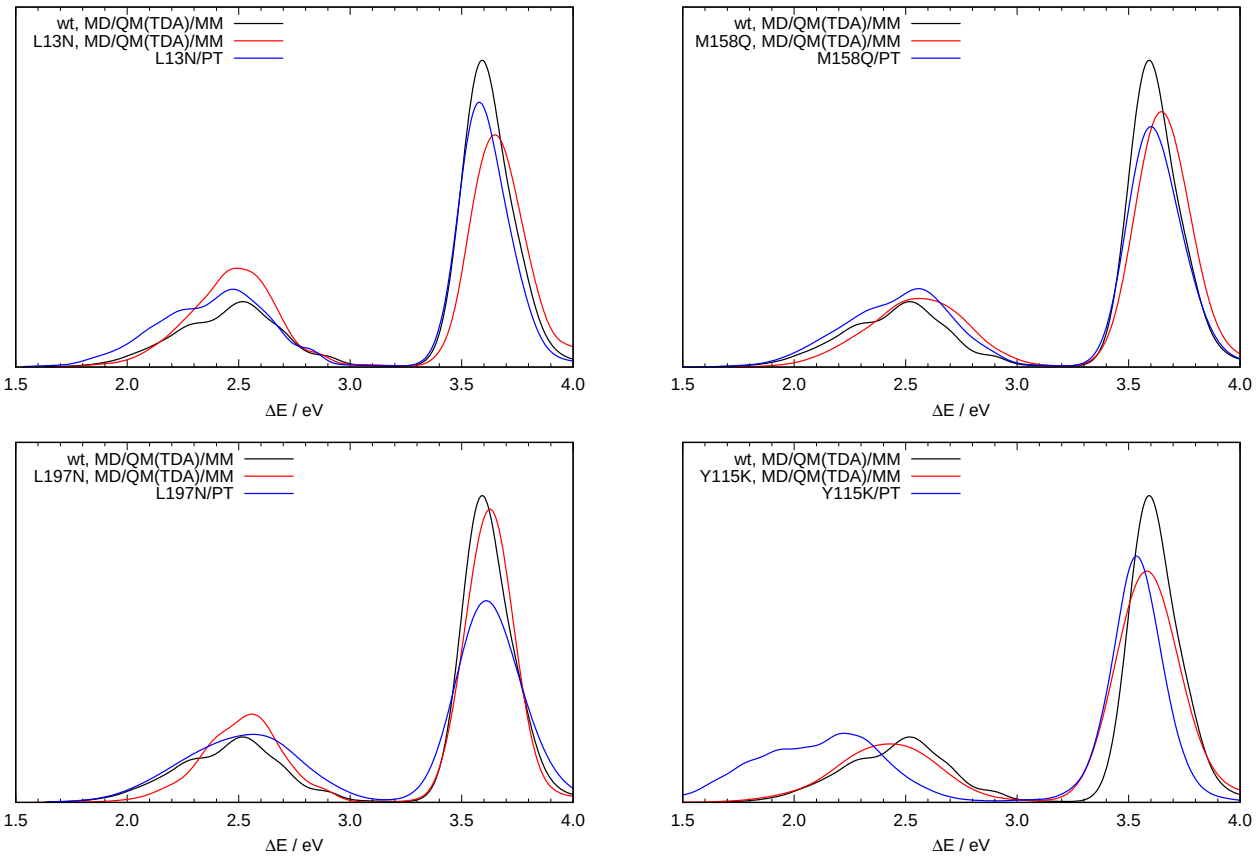

Figure S1: Comparison of MD-based simulation spectra of mutants L13N (top left), M158Q (top right), L197N (bottom left) and Y115K (bottom right). The spectra derived from histogramming of excitation energies are shown in black for mCardinal and in red for each mutant. In blue, the PT estimate of the absorption spectrum of each mutant is shown. In all cases, to atenuate statistical noise all spectra have been convoluted with a Gaussian with σ=0.05 eV

The position of the low energy band is reasonably well reproduced in all cases. Mutants display an absorption peak correctly displaced –hypsochromic or bathochromic– with respect to the non-mutated protein (in black), both for the MD-derived (red) and the PT-derived (blue) versions of the spectrum. As usual, Y115K is predicted to have an absorption peak much more towards lower energies using the perturbative approach, this being caused by the fact that the steric crowding in this case is larger and thus the MD simulation leads to the exploration of more favorable parts of configurational space.

4 Timings

The production runs which were used for the values featured in the table above or those reported in the manuscript were distributed in an heterogeneous computing. We are unable of extracting reliable timing data from the latter. In order to obtain estimates for the relative cost between full and the PT-based calculations, we carried out a small subset of calculations on a dedicated node. All timings discussed here are taken from an 8-core node powered by a E5-2620 v4 processor and equipped with 64GB RAM. Complicating things further, several different proprietary software were used, and one could also question the impact of running parts of the tasks in a parallel fashion, compared to single core or single-node (but making use of all available cores) runs. Our approach was to not use minutes as the measuring unit, but min·core.

All timings are taken as averages from 10-20 runs, depending on the spread. The nomenclature we use is the same as in Eqs. (12)-(15).

- $\langle T_{\mathrm{setup}} \rangle = 3$ min·core
- $\langle T_{\mathrm{QM/MM}} \rangle = 130$ min·core
- $\langle T_{\mathrm{PT}} \rangle = 2$ min·core
- $\langle T_{\mathrm{MD}} \rangle = 15080$ min·core/ns

The T_{setup} value is calculated from the sum of several different operations, including the PYMOL commands for generating a rotamer and optimizing the latter with NAMD. The $T_{\mathrm{QM/MM}}$ is taken as an average from ORCA and Gaussian (both software packages are capable of doing such calculations). The T_{PT} calculation is a very fast step, consisting only of the calculation of nuclear interaction integrals (taken from ORCA). Lastly, the MD has been estimated from a set of NAMD 0.1 ns runs.