

Photo-processes in biology:

A glimpse from computational chemistry and machine learning

By Debashree Ghosh* and Arpan Choudhury

<https://doi.org/10.51167/acm00043>



Debashree Ghosh

Debashree Ghosh obtained her M.S. degree in Chemical Sciences from IISc in 2005 and PhD degree from Cornell University, USA in 2010. She joined CSIR-NCL as a Senior Scientist in 2012, subsequently moving to Indian Association for the Cultivation of Science. Currently she is a professor at IACS. Her research focuses on the development of polarizable force fields and their hybrid frameworks with excited-state electronic structure methods, geared towards understanding the photo-physics of biological systems. She has been recognized by several awards, including the Annual Medal of the International Academy of Quantum Molecular Science (2021), SERB-POWER fellowship (2022) and Walter Kohn prize awarded by International Centre for Theoretical Physics (2022).



Arpan Choudhury

Arpan Choudhury is a PhD student at the School of Chemical Sciences, Indian Association for the Cultivation of Science (India). He received his bachelor and master degree in Chemistry from University of Calcutta and University of Hyderabad, respectively. His broad research interests include quantum chemistry and machine learning. He is currently working on the applications of machine learning models for a detail-rich understanding of the photo-processes in biological molecules, such as melanin.



Light matter interaction

Light matter interaction is ubiquitous in biology. When we think about photo-processes in biology, we immediately think of processes such as photosynthesis, light harvesting, vision, as well as skin photoprotection, DNA damage etc. [1] [2] Prebiotic chemistry is also rife with examples of light matter interaction. Therefore, light and its interaction with chemical species are crucial in the genesis and continued success of life itself. Researchers have also been interested in understanding these molecular mechanisms to leverage it towards biomimetic solutions (one such important photoprocess and its computational approach is shown in Fig. 1).

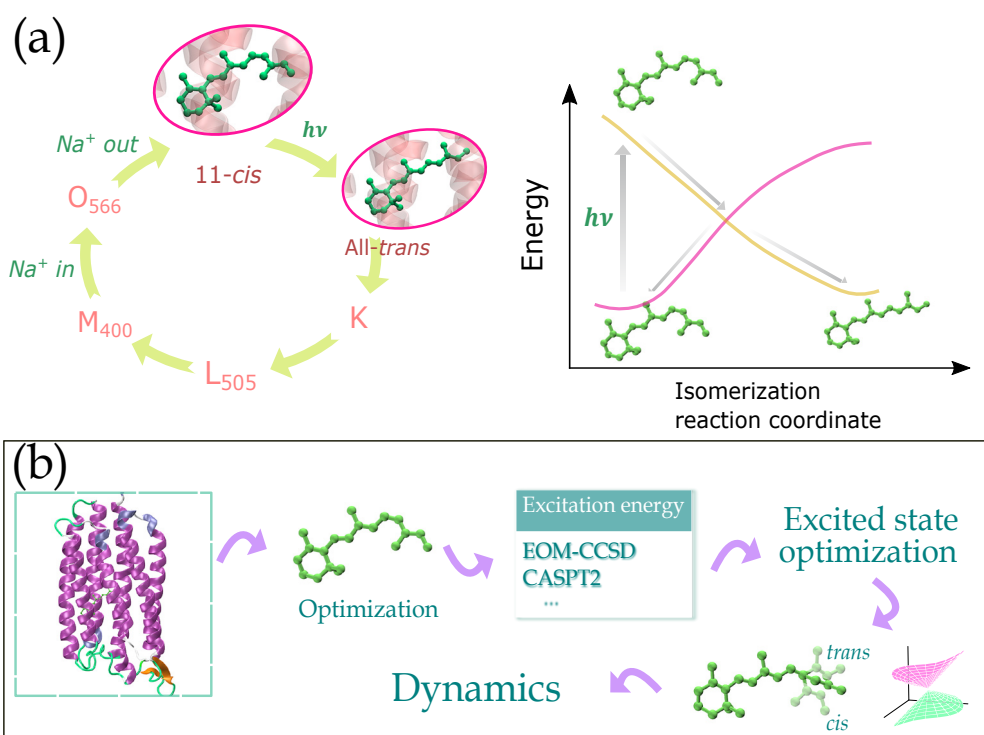
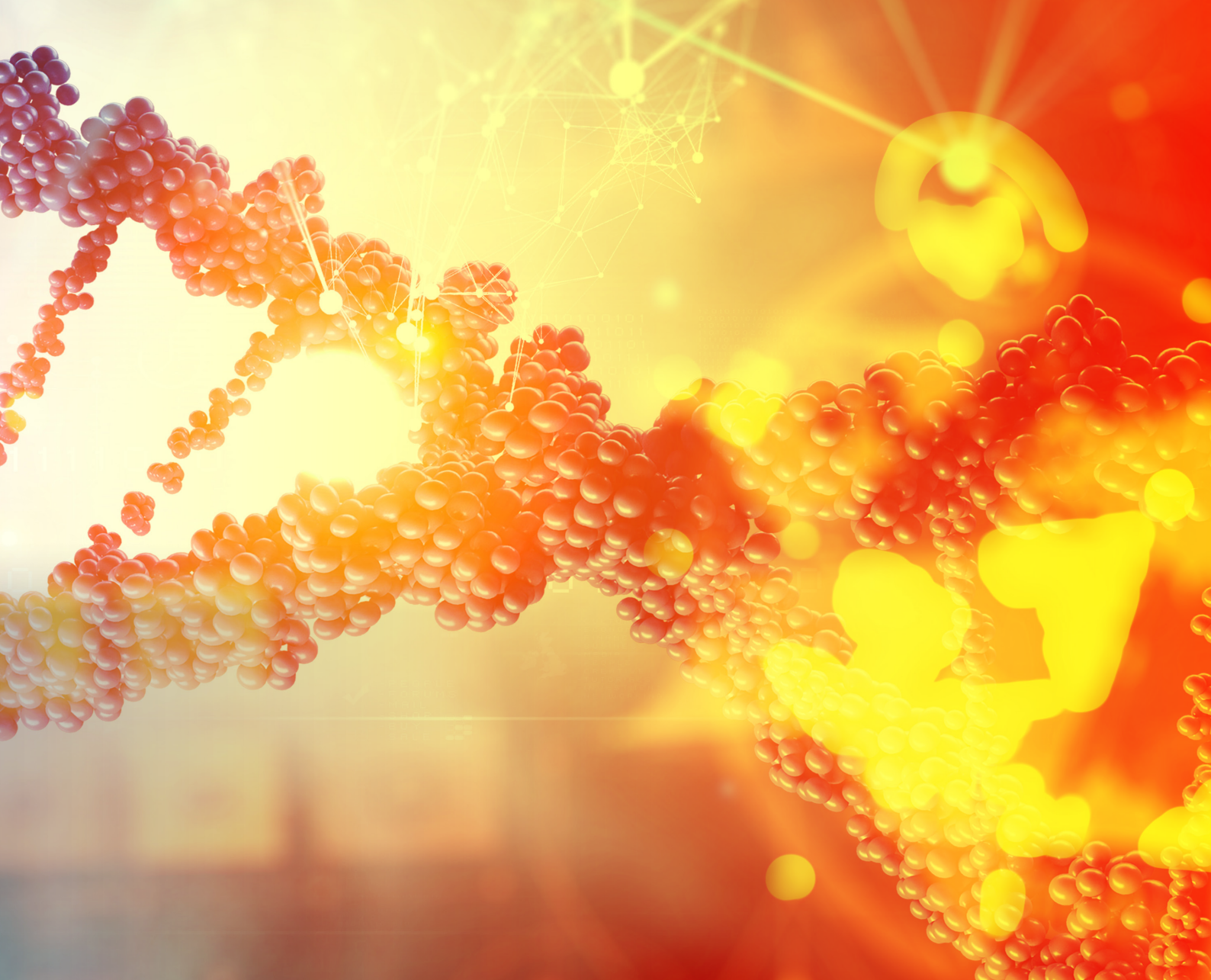


Figure 1: (a) Cycle of processes after light irradiation on light gated channel proteins. (b) The computational approaches to understand the first optically activated step requires an in-depth understanding of the ground and excited states. The steps involved in a typical computation towards that are shown.

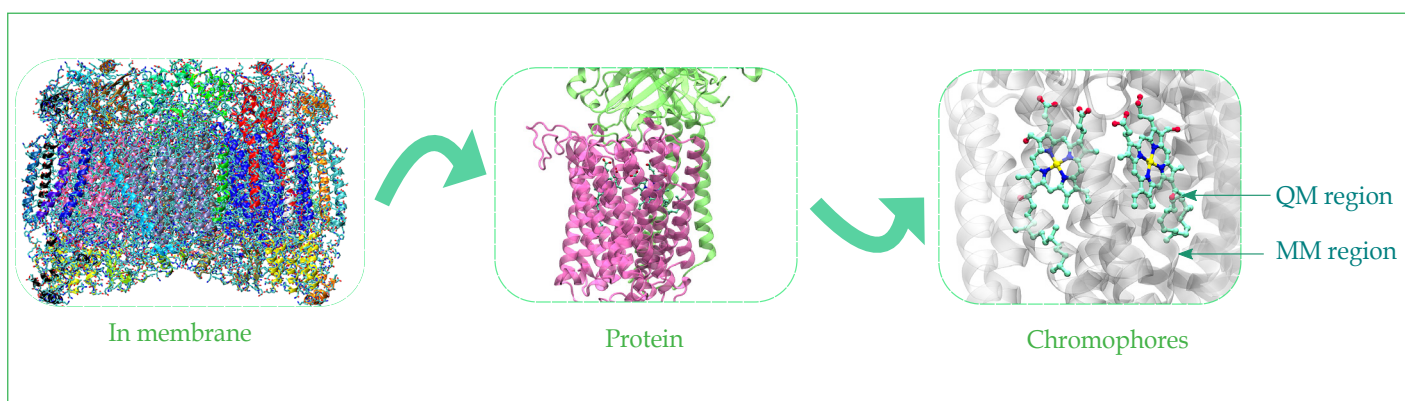


Figure 2: The different length scales that are present in biological systems. Often one requires multiscale modelling to address the different aspects of the problem.

It is evident that light activated chemistry is inherently different from ground state reactivity. Light gives us access to certain spectroscopically allowed excited states of the reactants and subsequently the reactivity under irradiation follows the electronic structure of the reactants and products in these excited states. Since the electronic structure of ground and excited states of the reactants and products vary, their reactivities can therefore, be significantly different with and without irradiation. However, the excited states are short-lived, since all these processes have to compete with radiative decay processes such as fluorescence or phosphorescence etc. That is any successful reaction in excited state has to be completed before the molecule relaxes back via other dissipative processes. Thus, excited state reactions are often extremely fast and that is what makes it so effective in biology. But at the same time, this same reason makes it excruciatingly difficult to probe experimentally. Most of the intermediates thus formed are so short lived that they can seldom be captured and probed successfully.^[3]

Further difficulties can be envisaged from the differential reactivities and electron densities in ground and excited states.^{[4] [5]} Chemists have long relied on empirical rules and understanding of reactivities to understand mechanisms of reactions. However, these rules are based on ground state species and their electronic structures. Since the electronic structure of the excited states of the same species are significantly different, the empirical rules also need to be updated with this knowledge for its successful implementation in light activated processes.

Computational methods to understand excited states

Since experimental understanding of excited state phenomena and reactions are often quite challenging, computational tools are increasingly being used to understand these processes in tandem with experimental

observations. Computational chemistry and biology, augmented by machine learning is quickly becoming a powerful tool. It is a computational chemist's dream to string together the different processes to recreate the many stages of these photo-processes in biology.

It should be noted that the study of the excited state processes in chemistry and biology started with the ability to control pulsed radiation and study the excited state processes at different time scales.^{[6] [7] [8]} However, over the last decade we have witnessed impressive new developments in

the field of computer architecture, machine learning and artificial intelligence. Google DeepMind has started developing a variety of artificial intelligence technology, Alpha Go managed to beat a human being in Go. And of late Alpha fold has predicted the structure of millions of proteins which other technology has struggled to predict over many decades.

In the same line, we have witnessed the increased use of machine learning in chemical and bio-chemical problems.^{[9] [10] [11] [12] [13] [14]} Research in this direction started from the need to predict properties of molecules and materials. However, it has evolved into

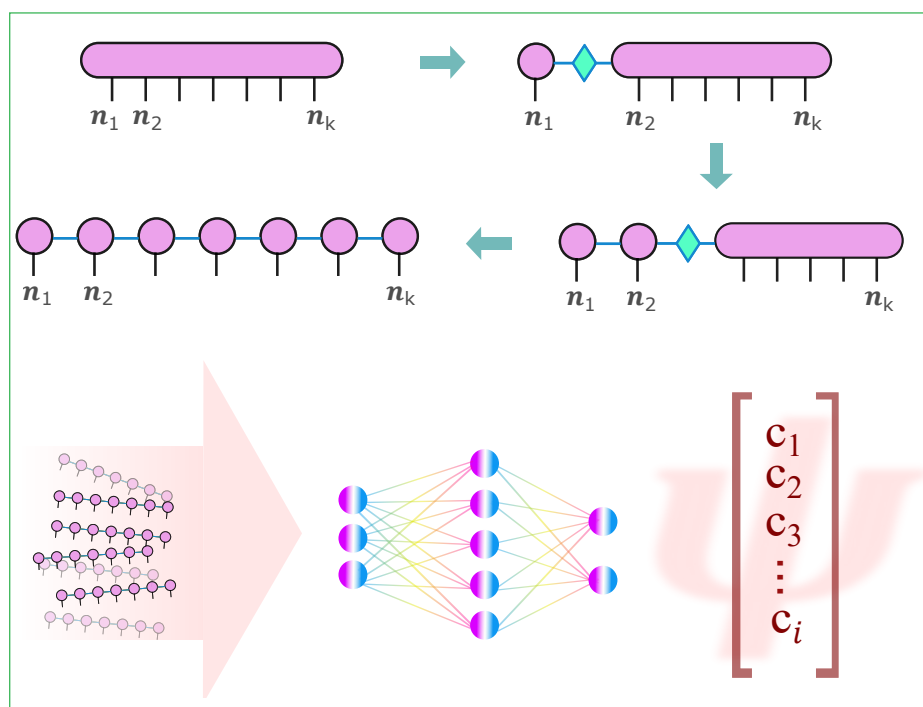


Figure 3: Formation of matrix product state ansatz from an exponentially scaling full configuration interaction wavefunction ansatz. The connection between the two are successive singular value decompositions that enable one to retain only the most important degrees of freedom. Both the full configuration interaction and matrix product state ansatz can be determined using machine learning techniques. The learning process occurs on data degenerated from Monte Carlo approaches.

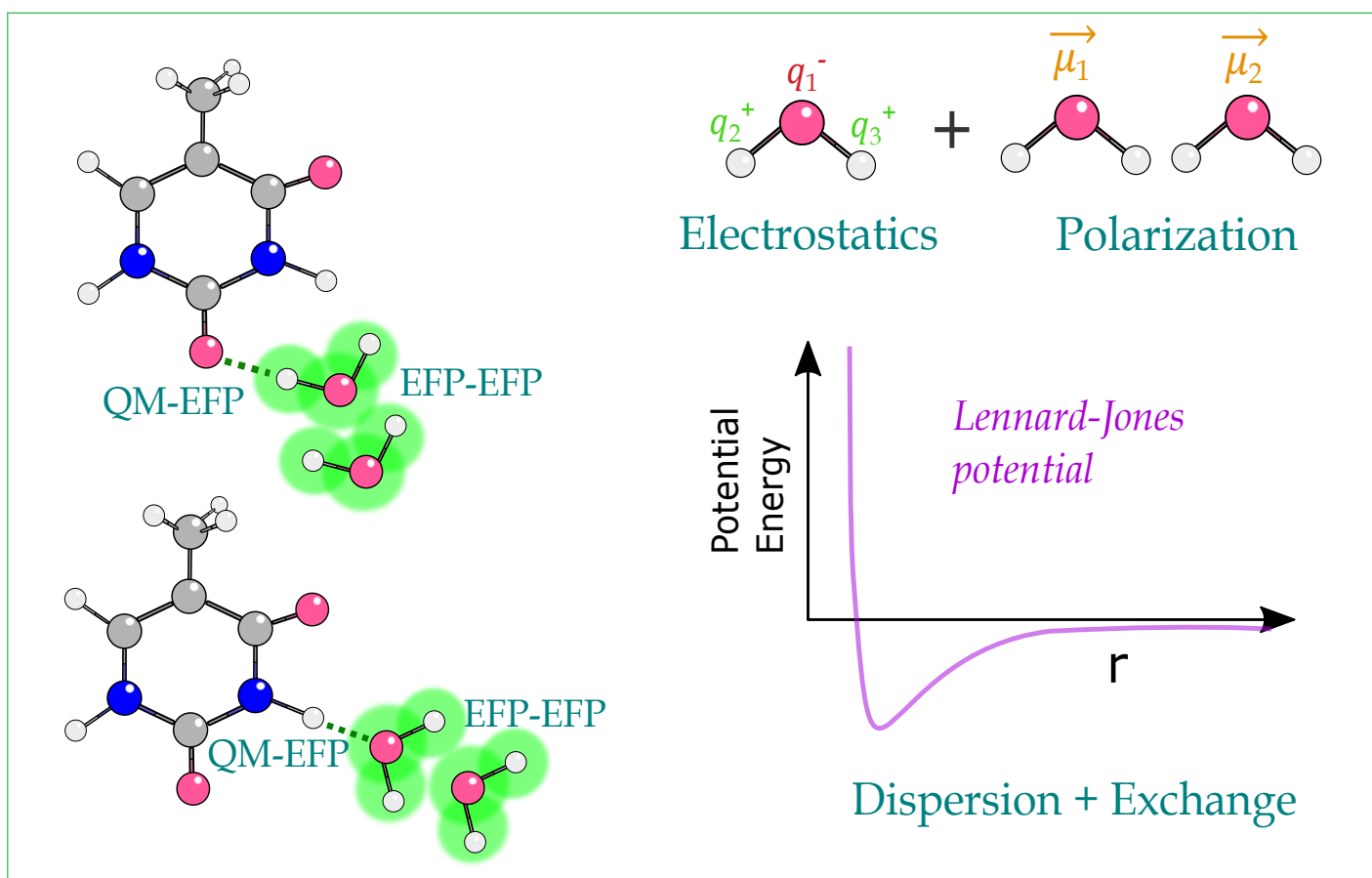


Figure 4: Effective fragment potential can be used as a sophisticated polarizable force field that is calculated a priori ab initio methods. This ab initio approach allows one to develop effective fragment potentials for any molecule with required accuracy which can then be used in a hybrid framework with high-level quantum chemistry methods as detailed in Ref. 26.

the more challenging and certainly more useful inverse design problems.^{[15] [16] [17]} Therefore, it is only natural that we would like to use machine learning and advanced computational techniques to excited state chemical and bio-chemical processes.

Apart from this almost natural progression, excited state chemistry and electronic structure is also an ideal playing field for big data problems. The hallmark of excited state chemistry is that electrons in this regime are invariably strongly correlated with each other. If in a ground state situation, an electron can interact with only the few electrons in its vicinity, i.e., local picture, in the excited state they can interact with a large number of electrons far away from each other. This can alternatively be viewed from an orbital picture. The orbitals in the virtual space (i.e., unoccupied in the ground state) are closer to each other in energy and therefore, an electron when present in those orbitals interact with the other electrons in the neighboring orbitals more strongly. As a result, the bedrock mean field approximations in chemistry becomes qualitatively incorrect.^[18] Therefore, in this regime, one needs to consider the effects of electrons in various configurations, i.e., distributed among different orbitals to understand the excited state

electronic structure. However, the number of such configurations grows exponentially with the size of the system and biological systems are seldom small. Therefore, one can resort to machine learning augmented big data directions to understand such systems.^{[19] [20]} ^[21] Our group is among the few in the world delving in this direction and have applied this to realistic molecular systems (scheme shown in Fig. 3).

Furthermore, due to the size of biological systems a purely quantum mechanical treatment is not affordable. Fortunately, neither is such a treatment necessary. Most biological systems and excited state phenomena have an active site, where the maximal electron density change occurs or bonds are broken and formed (shown in Fig. 2). Therefore, a natural strategy to treat such systems is to use a hybrid quantum mechanical / molecular mechanical (QM/MM) approach.^[22] In this hybrid approach, the active site is treated quantum mechanical and the rest of the system effectively provides a field that changes the energetics of the active site. However, in case of excited state phenomena, where the charge density can be significantly different from the ground state, such a hybrid method requires the consideration of second order

effects, such as polarization. Therefore, we use a sophisticated polarizable method called the effective fragment potential (EFP) to adequately treat such systems.^[23] Hybrid QM/MM methods that use high-level quantum mechanical methods, such as equation of motion coupled cluster with polarizable force fields have been quite effective for most of the biologically relevant excited state processes (shown in Fig. 4).^{[24] [25] [26]}

However, polarizable force fields are typically quite expensive computationally due to the need for self-consistent solution of the charge densities. In order to provide accurate and fast methods, we have developed machine learning based polarizable force fields. Here, physics-based machine learning circumvents the need to self consistently solve the charge densities and the induced dipole moments.^[27]

With the methods developed in the group as well as other high-level methods, our group focuses on a few biologically relevant problems. These include photophysics of skin-pigment melanin, interactions in fluorescent proteins and their analogues, low lying states of conjugated moieties and singlet fission.

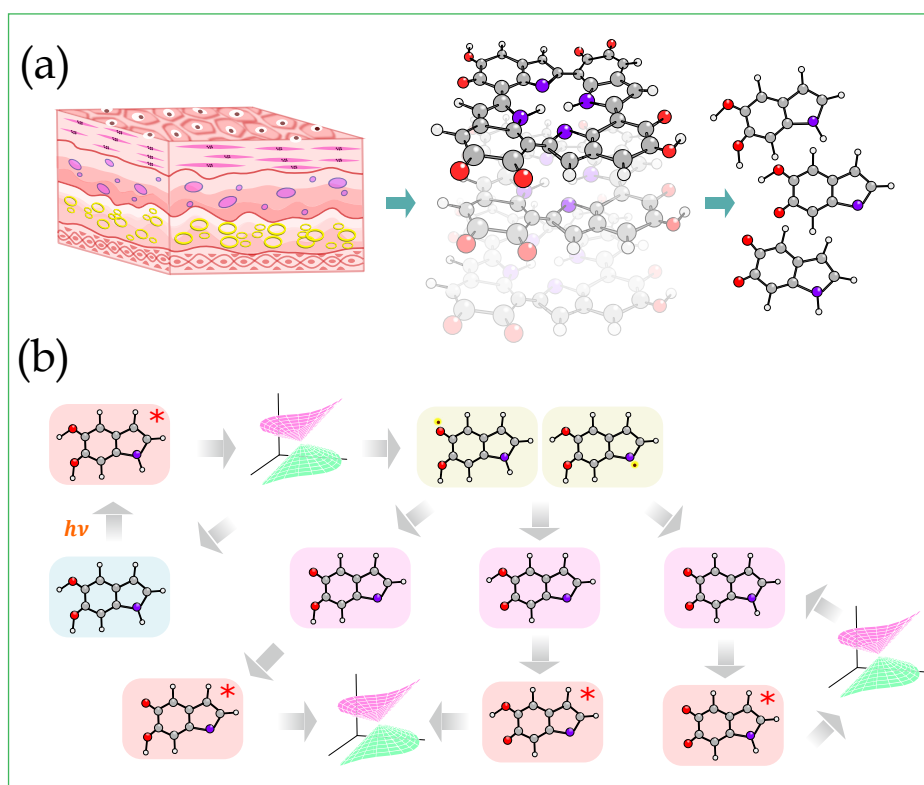


Figure 5: (a) The structure of melanin in skin pigment. It is a heterogeneous polymer containing di-hydroxy indole species. (b) The excited state photoprocesses of the constituents of melanin as elucidated from computational studies. The structural heterogeneity is crucial to its ultrafast photoprocesses.

Photo-physics of melanin

Melanin is the skin pigment present in most animals and human beings (Fig. 5a). It is also present in several plants and are mainly concerned with protection from the sun light. Due to this central property of photo-protection, melanin is capable of absorbing light over the entire UV and visible range. The melanin spectrum is unusually monotonic and extremely broadband.^[28] While this broadband spectrum allows melanin to be an effective absorber of sunlight, this makes it difficult to study the structure and functions of melanin with the usual tools of spectroscopy.

Therefore, computational tools have become one of the important ways to delve into the structure and structure-function correlation of melanin.^[29] There are initial indications of heterogeneity playing a crucial role in the broadband spectra of melanin.^[28] In our group, we have worked on understanding the decay channels and pathways of photoprotection after the initial excitation by UV light.^[30] Using a bottom-up approach of studying each of the monomers, dimers etc. of the system, we ascertain the detailed photoprotection mechanism. We observe that heterogeneity plays an important role in this mechanism. Here, it should be noted that melanin is a heterogeneous polymer

formed from the basic monomer di-hydroxy indole and its oxidized forms. There is also the presence of its carboxylic acid derivatives. The different oxidized forms play a crucial role in the photoprotection pathways. From the study of the critical structures in the excited states, and the low-lying conical intersections (i.e., geometries where there are energetic degeneracies between the ground and excited state), it is noticed that in di-hydroxy indole the major molecular modes are OH bond elongation and out-of-plane puckering modes. The out of plane puckering mode causes exclusive photo-protection and fast quenching of energy. The OH bond elongation mode is therefore, the only pathway for photochemical reaction. However, this photochemical reaction gives rise to one or other of the oxidized forms of melanin itself. On the other hand, the oxidized monomers upon excitation either quenches fast, or interconverts between each other. Therefore, it can be posited that the monomers of melanin interconvert between each other upon excitation and all photochemical processes within monomers still retains in structural heterogeneity (shown in Fig. 5b).

While the importance of structural heterogeneity towards its photoprotection property have been hypothesized from the monomer

studies, it has also been observed in the dimers and higher oligomers.^[31] The effects of small molecules on these pathways have been studied. Here, it should be noted that the success of melanin photo-processes lies in the large number of low energy probable pathways. When a small molecule comes close to the melanin framework, the energetics of each of these pathways change. However, since there are numerous possible pathways, there always exists one or more pathway that is ultrafast in nature. With these computational studies, we have therefore, delved into the many mysteries of melanin photochemistry and have unearthed the reason behind its success.^[32]

Fluorescent protein

Photoreceptor proteins and fluorescent proteins have been studied extensively from a molecular perspective.^[33] There have been several artificially engineered fluorescent proteins as well as new fluorescent proteins in nature that can be leveraged as bio-markers for different properties. In our group, we have been interested in the effect of neighboring amino acids towards tuning the absorption and fluorescence spectra of these fascinating systems.^{[34] [35]}

The chromophore responsible for fluorescence in green fluorescence protein (GFP) is p-hydroxy benzylidene imidazolinone (HBDI). We have studied the substituted HBDIs, especially the fluorine substituted one – DF-HBDI. DF-HBDI is known to intercalate in RNA spinach and senses the presence of G-quadruplex structure in its vicinity (shown in Fig. 6). It was hypothesized that the resultant red shift in the spectra of DF-HBDI was due to the π -stacking interactions between the G-quadruplex and the DF-HBDI molecule. However, using extensive high-level QM/MM calculations we were able to show that while the structure is stabilized due to the π -stacking interactions, the shift in spectra is due predominantly to the long-range electrostatic interactions rather than the short-range π -stacking effects.

Furthermore, the reason why π -stacking is less important in the resultant spectral shift is due to the extreme sensitivity of the wavefunction on the structure which undergoes fluctuation at room temperature. This new understanding brings to scrutiny much of the earlier results on the importance of π -stacking interactions in other fluorescent proteins.^[36]

Polyaromatic hydrocarbons and singlet fission

Low lying states and singlet-triplet gaps of polyaromatic hydrocarbons have long been a fascinating problem that cannot be understood within the confines of single reference

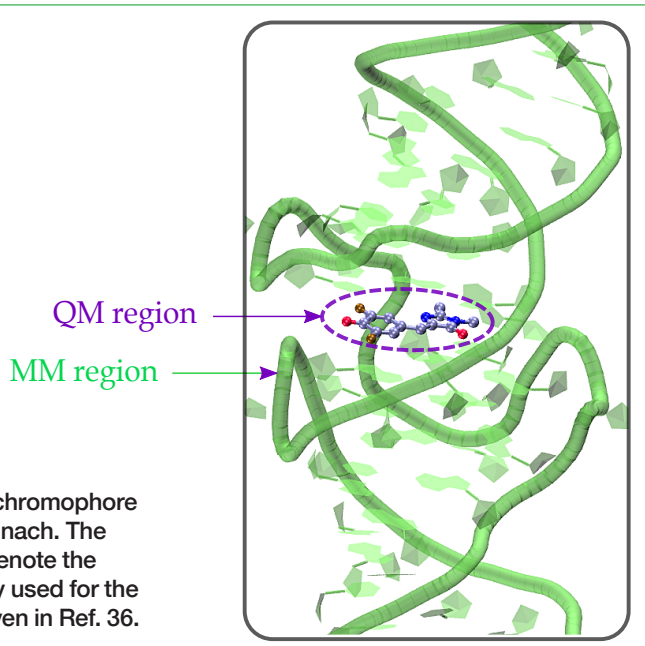


Figure 6: Fluorescent chromophore intercalated in RNA spinach. The QM and MM regions denote the computational strategy used for the hybrid treatment as given in Ref. 36.

molecular orbital theory. The inadequacy of single reference theory arises from the strong correlation between the electrons in their valence space constituted by the π and π^* orbitals. The frontier orbitals can be pseudo-degenerate as the system size increases and that manifests itself in interesting singlet-triplet gaps, triplet stability inversion etc. They have important implications in singlet fission and magnetic materials.

We apply a combined approach of simple model Hamiltonians with density matrix renormalization group^[37] to understand

the singlet-triplet gaps in these systems.^[38]^[39] We have noticed the delicate balance between static and dynamic correlation that is required to adequately handle such systems with quantitative accuracy. Recently, we have seen the effect of spin frustration and topology that is central to the singlet-triplet gaps in these systems, especially in the cases containing odd number of rings. The effect of spin frustration and topology can be leveraged to engineer molecules with different singlet-triplet gaps in a very controlled manner.

The importance of these low-lying states of varied electronic nature have been elucidated in the singlet fission phenomena.^[40] Unlike in case of the acenes, we have noticed that polyenes are even more versatile. In polyenes, many different symmetry states with charge transfer components are crucial to the success of singlet fission phenomena. The greater versatility of the polyenes may be central to the effective singlet fission process in many different geometries and crystal structures.

The way forward

We are entering the realm of combined use of machine learning and traditional computational approaches. Many of the problems stated above can be understood and taken to predictive capabilities with machine learning. For example, the use of machine learning for structure elucidation is becoming extremely popular and can be an invaluable tool for structure determination in melanin. This will be a holy grail of melanin research. Machine learning is also increasingly effective towards prediction of properties and perhaps, more crucially, towards inverse design of systems and materials with the desired properties. Furthermore, more accurate and computationally affordable computational chemistry approaches to treat such excited state phenomena are the need of the day.

References:

- [1] N. Nelson, A. Ben-Shem, *Nature Reviews Molecular Cell Biology*, **5**, 971 (2004)
- [2] D. Gust, T. A. Moore, *Science*, **244**, 4900, 35 (1989)
- [3] A.F. Fidler, V.P. Singh, P.D. Long, P.D. Dahlberg, G.S. Engel, *Nature Communications*, **5**, 3286 (2014)
- [4] L. Vannay, E. Bremond, P. de Silva, C. Corninboeuf, *Chemistry-A European Journal*, **22**, 51, 18442 (2016)
- [5] P.K. Chattaraj, A. Poddar, *The Journal of Physical Chemistry A*, **103**, 43, 8691 (1999)
- [6] R. Berera, R. van Grondelle, J.T.M. Kennis, *Photosynthesis Research*, **101**, 105 (2009)
- [7] L. Nilsson, B. Halle, *Proceedings of the National Academy of Sciences (USA)*, **102**, 39, 13867 (2005)
- [8] M. Weinelt, *Journal of Physics: Condensed Matter*, **14**, R1099 (2002)
- [9] J. Westermayr, M. Gastegger, K.T. Schutt, R.J. Mauer, *The Journal of Chemical Physics*, **154**, 230903, (2021)
- [10] P. M. Pfluger, F. Glorius, *Angewandte Chemie*, **59**, 43, 18860 (2020)
- [11] J.P. Janet, H. Kulik, Machine Learning in Chemistry, American Chemical Society (2020)
- [12] N. Artrith, K.T. Butler, F. Coudert, S. Han, O. Isayev, A. Jain, A. Walsh, *Nature Chemistry*, **13**, 505 (2021)
- [13] P. O. Dral, *The Journal of Physical Chemistry Letters*, **11**, 6, 2336 (2020)
- [14] R. Ramakrishnan, O. A. von Lilienfeld, "Machine Learning, Quantum Chemistry, and Chemical Space," in *Reviews in Computational Chemistry* (2017)
- [15] B. Sanchez-Lengeling, A. Aspuru-Guzik, *Science*, **361**, 6400, 360 (2018)
- [16] B. Sridharan, M. Goel, U. D. Priyakumar, *Chemical Communications*, **58**, 5316 (2022)
- [17] J. G. Freeze, H. R. Kelly, V.S. Batista, *Chemical Reviews*, **119**, 11, 6595 (2019)
- [18] E. Dagotto, *Science*, **309**, 5732, 257 (2005)
- [19] G. Carleo, M. Troyer, *Science*, **355**, 6325, 602 (2017)
- [20] K. T. Schutt, M. Gastegger, A. Tkatchenko, K-R. Muller, R. J. Maurer, *Nature Communications*, **10**, 5024 (2019)
- [21] S. K. Ghosh, M. Rano, D. Ghosh, *The Journal of Chemical Physics*, **154**, 094117 (2021)
- [22] A. Warshel, M. Levitt, *Journal of Molecular Biology*, **103**, 2, 227 (1976)
- [23] M. S. Gordon, M. A. Freitag, P. Bandyopadhyay, J. H. Jensen, V. Kairys, W. J. Stevens, *The Journal of Physical Chemistry A*, **105**, 2, 293 (2001)
- [24] D. Ghosh, D. Kosenkov, V. Novoschi, C. F. Williams, J. M. Herbert, M. S. Gordon, M. W. Schmidt, L. V. Slipchenko, A. I. Krylov, *The Journal of Physical Chemistry A*, **114**, 48, 12739 (2010)
- [25] D. Ghosh, O. Isayev, L. V. Slipchenko, A. I. Krylov, *The Journal of Physical Chemistry A*, **115**, 23, 6028 (2011)
- [26] D. Ghosh, *The Journal of Physical Chemistry A*, **121**, 4, 741 (2017)
- [27] S. Bose, D. Dhawan, S. Nandi, R. R. Sarkar, D. Ghosh, *Physical Chemistry Physical Physics*, **20**, 22987 (2018)
- [28] C-T. Chen, C. Chuang, J. Cao, V. Ball, D. Ruch, M. J. Buehler, *Nature Communications*, **5**, 3859 (2014)
- [29] D. Ghosh, *WIREs: Computational Molecular Science*, **11**, 3, e1505 (2021)
- [30] P. Ghosh, D. Ghosh, *The Journal of Physical Chemistry B*, **121**, 24, 5988 (2017)
- [31] P. Ghosh, D. Ghosh, *Physical Chemistry Chemical Physics*, **21**, 6635 (2019)
- [32] A. Choudhury, D. Ghosh, *Chemical Communications*, **56**, 10481 (2020)
- [33] M.-A. Mroginiski, S. Adam, G. S. Amoyal, A. Barnoy, A.-N. Bondar, et al., *Photochemistry and Photobiology*, **97**, 2, 243 (2020)
- [34] D. Ghosh, A. Golan, A. K. Takahashi, A. I. Krylov, M. Ahmed, *The Journal of Physical Chemistry Letters*, **3**, 1, 97 (2012)
- [35] D. Ghosh, A. Acharya, S. C. Tiwari, A. I. Krylov, *The Journal of Physical Chemistry B*, **116**, 41, 12398 (2012)
- [36] S. Bose, S. Chakrabarty, D. Ghosh, *The Journal of Physical Chemistry B*, **120**, 19, 4410 (2016)
- [37] D. Ghosh, J. Hachmann, T. Yanai, G. K.-L. Chan, *The Journal of Chemical Physics*, **128**, 144117 (2008)
- [38] C. U. Ibeji, D. Ghosh, *Physical Chemistry Chemical Physics*, **17**, 9849 (2015)
- [39] M. Rano, S. K. Ghosh, D. Ghosh, *Chemical Science*, **10**, 9270 (2019)
- [40] S. Santra, J. Ray, D. Ghosh, *The Journal of Physical Chemistry Letters*, **13**, 29, 6800 (2022)