

RESEARCH ARTICLE

Short-wavelength-sensitive 2 (Sws2) visual photopigment models combined with atomistic molecular simulations to predict spectral peaks of absorbance

Dharmeshkumar Patel¹, Jonathan E. Barnes², Wayne I. L. Davies^{3,4,5,6,7}, Deborah L. Stenkamp^{8,9}, Jagdish Suresh Patel^{1,8*}

1 Institute for Modeling Collaboration and Innovation (IMCI), University of Idaho, Moscow, ID, United States of America, **2** Department of Physics, University of Idaho, Moscow, ID, United States of America, **3** Umeå Centre for Molecular Medicine (UCMM), Umeå University, Umeå, Sweden, **4** School of Biological Sciences, University of Western Australia, Perth, WA, Australia, **5** The Oceans Graduate School, University of Western Australia, Perth, WA, Australia, **6** The Oceans Institute, University of Western Australia, Perth, WA, Australia, **7** Lions Eye Institute, University of Western Australia, Perth, WA, Australia, **8** Department of Biological Sciences, University of Idaho, Moscow, ID, United States of America, **9** Institute for Bioinformatics and Evolutionary Biology, University of Idaho, Moscow, ID, United States of America

* thejagdishpatel@gmail.com

OPEN ACCESS

Citation: Patel D, Barnes JE, Davies WIL, Stenkamp DL, Patel JS (2020) Short-wavelength-sensitive 2 (Sws2) visual photopigment models combined with atomistic molecular simulations to predict spectral peaks of absorbance. PLoS Comput Biol 16(10): e1008212. <https://doi.org/10.1371/journal.pcbi.1008212>

Editor: Alexander MacKerell, University of Maryland School of Pharmacy, UNITED STATES

Received: July 28, 2020

Accepted: September 21, 2020

Published: October 21, 2020

Copyright: © 2020 Patel et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: This research was supported by the Center for Modeling Complex Interactions (CMCI) sponsored by the NIGMS under award number NIH P20 GM104420, through a Pilot Grant to JSP that supported JSP, DP, and JB. JSP and JB were also supported in part by National Science Foundation EPSCoR Track-II grant under award

Abstract

For many species, vision is one of the most important sensory modalities for mediating essential tasks that include navigation, predation and foraging, predator avoidance, and numerous social behaviors. The vertebrate visual process begins when photons of the light interact with rod and cone photoreceptors that are present in the neural retina. Vertebrate visual photopigments are housed within these photoreceptor cells and are sensitive to a wide range of wavelengths that peak within the light spectrum, the latter of which is a function of the type of chromophore used and how it interacts with specific amino acid residues found within the opsin protein sequence. Minor differences in the amino acid sequences of the opsins are known to lead to large differences in the spectral peak of absorbance (i.e. the λ_{\max} value). In our prior studies, we developed a new approach that combined homology modeling and molecular dynamics simulations to gather structural information associated with chromophore conformation, then used it to generate statistical models for the accurate prediction of λ_{\max} values for photopigments derived from Rh1 and Rh2 amino acid sequences. In the present study, we test our novel approach to predict the λ_{\max} of phylogenetically distant Sws2 cone opsins. To build a model that can predict the λ_{\max} using our approach presented in our prior studies, we selected a spectrally-diverse set of 11 teleost Sws2 photopigments for which both amino acid sequence information and experimentally measured λ_{\max} values are known. The final first-order regression model, consisting of three terms associated with chromophore conformation, was sufficient to predict the λ_{\max} of Sws2 photopigments with high accuracy. This study further highlights the breadth of our approach in reliably predicting λ_{\max} values of Sws2 cone photopigments, evolutionary-more distant from template bovine RH1, and provided mechanistic insights into the role of known spectral tuning sites.

number OIA1736253, and DLS was also supported in part by NIH R01 EY012146 and NSF DEB 1638567. Computer resources were provided by the Institute for Bioinformatics and Evolutionary Studies Computational Resources Core sponsored by the National Institutes of Health (NIH P30 GM103324). This research also made use of the computational resources provided by the high-performance computing center at Idaho National Laboratory, which is supported by the Office of Nuclear Energy of the U.S. DOE and the Nuclear Science User Facilities under Contract No. DE-AC07-05ID14517. WILD was supported by the Australian Research Council (ARC) in the form of a Future Fellowship (FT110100176) and a Discovery Project grant (DP140102117), and is currently supported by a JC Kempe Memorial Scholarship from the Kempe Foundation, Sweden. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

In vertebrates, color vision depends on the complement of cone visual photopigments that have different spectral peaks of absorbance (λ_{\max}) within the cone population. Together, the type of chromophore and the amino acid sequence of the opsin protein directly affect the λ_{\max} value. To understand this relationship further at a structural level, we previously developed a new molecular modeling approach to study Rh1 and Rh2 opsin classes by combining homology modeling, molecular dynamics simulations to extract structural parameters of chromophore conformations and statistical modeling. Here, we used this novel modeling approach to accurately predict the λ_{\max} values for teleost Sws2 photopigments. Such a genome-to-phenome approach for predicting visual pigment function will be of great interest to evolutionary biologists, vision scientists, and molecular modelers, to better understand the diversity and mechanisms of sensory function. Moreover, it will pave the way for novel strategies to forward engineer visual pigments suitable for optogenetics applications.

Introduction

For many animals, vision is a critical sensory modality that facilitates essential tasks that include navigation, predation and foraging, predator avoidance, and numerous social behaviors. In vertebrates, vision is initiated when photons enter the eye and interact with rod and cone photoreceptors found within the neural retina. Specifically, light is detected by photosensitive visual photopigments that are housed within the folded membrane of photoreceptor outer segments. Upon the absorbance of photons, these photopigments activate a specific phototransduction cascade that results in a change in membrane potential. Once hyperpolarized, visual photoreceptors cause a change in neurotransmitter release that propagates neural signals through other retinal neurons that ultimately lead the visual centers of the brain [1–3].

A vertebrate visual photopigment consists of a transmembrane opsin protein that is covalently linked to a vitamin A-derived chromophore. Indeed, it is the interaction of the chromophore with specific amino acids of the protein sequence of the opsin that results in a broad array of different spectral peaks of absorbance (i.e. the λ_{\max} value) [1–6]. In the visual system, the predominant chromophores are either based on 11-*cis* retinal (i.e. rhodopsins or vitamin- A_1 photopigments) or 11-*cis*-3,4-didehydro retinal (i.e. porphyropsins or vitamin- A_2 photopigments) [1–3,7]. Throughout vertebrate evolution, the associated visual opsin genes and their gene products have diverged into five classes: these comprise a long-wavelength-sensitive (LWS) opsin gene class, two short-wavelength-sensitive (SWS) opsin gene classes (SWS1 and SWS2), and two medium-wavelength-sensitive (*mws*) opsin gene classes called rhodopsin-like 1 (RH1) and rhodopsin-like 2 (RH2), respectively [1–3,8]. In general, RH1 opsins form the highly sensitive visual photopigments (rod opsins) of rod photoreceptors, which typically mediate dim light or scotopic vision by maximally detecting wavelengths at around 500 nm [1–3], although λ_{\max} values can shift to shorter wavelengths (e.g. 470–490 nm in some deep-sea fishes [9–11]). The remaining four opsin gene classes (LWS, SWS1, SWS2 and RH2) form the visual photopigments expressed in cone photoreceptors that mediate bright light or photopic vision. Color vision is possible when at least two cone photopigments with distinct λ_{\max} values and overlapping spectral absorbance profiles are present within distinct cone populations, thus providing differential input to other retinal neurons [12]. Cone opsins of the SWS1 class typically produce photopigments with λ_{\max} values between 360 nm and 450 nm (i.e.

perceived as ultraviolet (UV) to violet parts of the light spectrum), SWS2 photopigments present with λ_{\max} values of ~400–470 nm (perceived as blue), RH2 opsins produce photopigments with λ_{\max} values of ~480–530 nm (perceived as blue-green), whereas LWS photopigments have λ_{\max} values that are largely sensitive to a range of wavelengths from 500–570 nm (perceived as red) [1–3,8]. These spectral ranges are based on photopigments that possess a vitamin-A₁-derived chromophore (i.e. rhodopsins); whereas for porphyropsins, the presence of a retinal chromophore based on vitamin-A₂ shifts the λ_{\max} value towards longer wavelengths (e.g. up to ~620 nm [1–3,7,13], a spectral property that is more pronounced the longer the wavelength [14].

There is a great deal of diversity within each class of vertebrate opsins, since visual photopigment proteins are under strong natural selection [15–17]. This diversity is particularly striking within the teleost fishes, where many of the opsin genes have been tandemly or otherwise replicated, followed by subfunctionalization or neofunctionalization to generate new photopigments with distinct λ_{\max} values [10,18–21]. For example, many fish genomes harbor two or three copies of the *sws2* cone opsin gene [22], the subject of the present study. In some cases, only minor changes in the amino acid sequence have resulted in major changes in the λ_{\max} value, as was demonstrated for the A269T (Amino acid numbering in this study is based upon bovine RH1 sequence numbering; S1 Fig) change in the *Sws2a* opsin of the spotted flounder (*Verasper variegatus*) that resulted in a photopigment that is sensitive to longer wavelengths (i.e. λ_{\max} = 485 nm instead of 466 nm) [23].

There is great interest in understanding the evolutionary, as well as the molecular mechanisms, that underlie the diversity of visual photopigments and their spectral peak absorbances. In general, there are two main experimental techniques for defining the λ_{\max} value of visual photopigments, namely microspectrophotometry, which analyzes the spectral profile of photoreceptors directly, but only with fresh retinas, and spectral tuning site substitutions cannot be studied in isolation [24] or *in vitro* regeneration [2,23,25–27]. The latter is a popular, yet highly specialized and labor-intensive, approach that has been used to deduce the spectral properties of photopigments in isolation from wildtype and inferred ancestral protein sequences, followed by reconstitution with chromophore (usually 11-*cis* retinal) and experimental measurement of absorbance from 200–800 nm [25,28–31]. This technique, when combined with site-directed mutagenesis, has illuminated the contributions of specific amino acid substitutions to shifts in the λ_{\max} value [2,25,26,30]. Although success of such studies has permitted the accurate prediction of the spectral peak of absorbance for some opsin classes (specifically, LWS [32] and UV-sensitive SWS1 [26] photopigments), such efforts are far from sufficient to understand pigment function that allows prediction of the λ_{\max} value based entirely upon the amino acid sequence [6,33]; this is particularly the case for both SWS2 and RH2 photopigments, where experimental interventions are frequently employed. One of the long-term goals of our studies is to develop computational tools that result in straightforward, genome-to-phenome, predictive pipelines, as is the case for the application of spectral modeling and atomistic molecular simulations for both the Rh1 and Rh2 classes of visual photopigments [10,34].

The λ_{\max} value of any functional photopigment in its inactive form is determined by the conformation adopted by the chromophore in the dark state, a function that is dependent upon the shape and composition of the retinal binding pocket, as well as the counterions that stabilize the Schiff base linkage of the chromophore to lysine (K) 296 of the opsin protein [35–37]. Therefore, our aim of generating genome-to-phenome pipelines for predicting the λ_{\max} values of visual photopigments from their amino acid sequences, has been to increase an understanding of chromophore conformation via atomistic molecular simulations and to use this structural information to generate predictive models [10,34]. The approach includes: 1) building homology models for classes of visual photopigments, using the solved crystal

structure of bovine rod opsin (RH1) as a template [38]; 2) carrying out atomistic molecular dynamics (MD) simulations using homology models of photopigments with experimentally-measured λ_{\max} values; 3) identifying structural features of the chromophore and opsin that are correlated with a particular λ_{\max} value; and 4) using these features to generate a statistical model that can in turn be used to predict λ_{\max} values of other photopigments. This approach was successful for predicting λ_{\max} values of teleost Rh1 photopigments [10] and a closely-related Rh2 class of teleost cone photopigments [34]. Notably, this approach also revealed structural features of the Rh1 protein, namely the presence vs. the absence of a C111-C188 disulfide bridge that powerfully predicts λ_{\max} values >475 nm when present vs. λ_{\max} values <475 nm when absent [10].

In this present study, we test the hypothesis that the approach outlined above and in our previous publications [10,34], can also successfully predict λ_{\max} values of a class of cone photopigments that are phylogenetically-distinct from the known bovine RH1 template, namely the teleost Sws2 class. The SWS2 opsins are more divergent from RH1 than RH2 vs. RH1 [8,21] and are known to be notoriously difficult when attempting to successfully predict λ_{\max} values from the amino acid sequence alone [1,2]. Furthermore, teleost Sws2 opsins display only ~48–51% amino acid sequence identity to the bovine rod opsin sequence (Table 1). By contrast, our previous studies tested this approach for teleost Rh1 photopigments, with ~49–83% identity to bovine rod opsin [10], and Rh2 cone pigments, with ~63–72% identity to bovine rod opsin [34]. Here we also test the hypothesis that this approach will work for a class of photopigments with a broader, and more short-wavelength-shifted range of λ_{\max} values (i.e. 397–485 nm; Table 1) than the Rh1 (444–519 nm) or Rh2 (467–528 nm) photopigments used in our prior studies [10,34].

In this study, we show that this approach was highly successful at predicting spectral peaks of absorbance values of 11 teleost Sws2 opsins for which sequences and λ_{\max} values are known (Fig 1, Table 1). We identified three parameters of chromophore conformation that together accurately predict the λ_{\max} value. Furthermore, we discuss these results in the context of known amino acid substitutions that likely contribute to divergent λ_{\max} values [22]. These studies, therefore, not only provide a valuable extension of our prior work, but also guidance and strategic directions for the improvement of functionally-predictive photopigment modeling.

Table 1. Selected Sws2 opsin sequences used for homology modeling.

| Fish [Reference] | Target Sws2 Sequences | UniProt Accession Number | Sequence Identity (%) vs. Bovine RH1 | Experimentally Measured Spectral Peak of Absorbance (nm) |
|------------------------|-----------------------|--------------------------|--------------------------------------|--|
| Spotted flounder [23] | Sws2a | A0A1L7P076 | 51.27 | 485.4 |
| | Sws2b | A0A1L7P082 | 49.30 | 415.8 |
| Slime flounder [23] | Sws2a | A0A1L7P074 | 50.70 | 451.9 |
| Olive flounder [23] | Sws2a | D7RP09 | 49.86 | 465.6 |
| Guppy [39] | Sws2a | A0A140JTJ4 | 48.02 | 438 |
| | Sws2b | A0A140JTJ5 | 48.31 | 408 |
| Medaka [20] | Sws2a | Q2L6A3 | 50.00 | 439 |
| | Sws2b | Q2L6A2 | 47.75 | 405 |
| Bluefin killifish [40] | Sws2a | Q7T2U6 | 50.28 | 448 |
| | Sws2b | Q7T2U7 | 49.30 | 397 |
| Zebrafish [1] | Sws2 | Q9W6A8 | 47.47 | 416 |

Sequence UniProt accession numbers, percentage sequence identity compared to the bovine RH1 protein sequence, and experimentally measured spectral peaks of absorbance are indicated in third, fourth and fifth columns, respectively.

<https://doi.org/10.1371/journal.pcbi.1008212.t001>

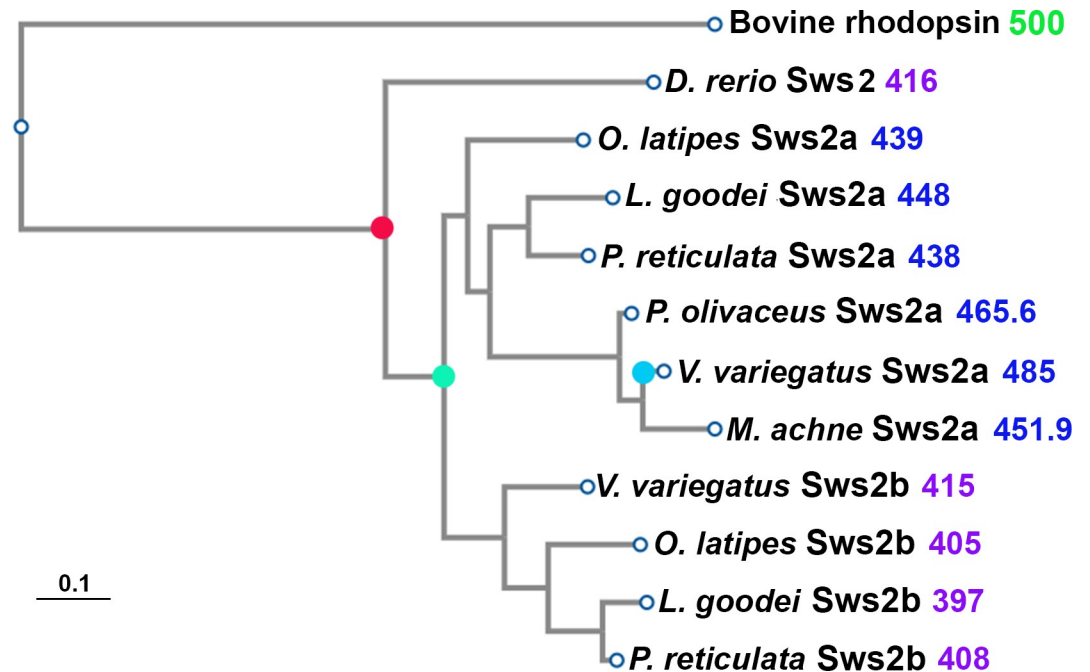


Fig 1. Evolutionary relationships of teleost Sws2 opsin proteins used in the simulations, as inferred by PhyML. Red filled circle indicates the speciation event that occurred prior to the duplication (green filled circle) of the *sws2* opsin genes in teleosts [18]. The duplication generated the *sws2a* clade, which encode photopigments with λ_{\max} values that are shifted to longer wavelengths, and the *sws2b* clade, which encode photopigments with short-wavelength-shifted λ_{\max} values. The blue filled circle indicates the amino acid substitution A269T (with numbering standardized to the bovine rod opsin sequence) that is likely to be the spectral tuning site important for a further shift of the λ_{\max} value of *V. variegatus* Sws2a to longer wavelengths [23]. Experimentally measured λ_{\max} values (in nm) are indicated next to the name of each opsin and are color-coded for λ_{\max} values that are <430 nm (violet) or >430 nm (blue).

<https://doi.org/10.1371/journal.pcbi.1008212.g001>

Results

To develop a model to predict spectral peaks of absorbance (i.e. λ_{\max} values) from a diverse set of teleost Sws2 cone opsins, the following amino acid sequences were selected: two from *V. variegatus* (spotted flounder) [23], one from *Microstomus achne* (slime flounder) [23], one from *Paralichthys olivaceus* (olive flounder) [23], two from *Poecilia reticulata* (guppy) [39], two from *Oryzias latipes* (medaka) [20], two from *Lucania goodei* (bluefin killifish) [40], and one from *Danio rerio* (zebrafish) [1]. These Sws2 opsin amino acid sequences show a wide range of experimentally measured λ_{\max} values that range from 397 nm to 485 nm (Table 1). Such functional divergence probably evolved as primary *sws2* gene sequences mutated and were positively conserved, which likely led to distinct conformations of the Sws2-associated chromophore, 11-*cis* retinal, in the dark state (Fig 2).

We used a similar protocol that was shown to be successful for the prediction of Rh1 and Rh2 λ_{\max} values [10,34], through building homology models using bovine rod opsin (RH1) as the template [41]. RH1 opsins are primarily present in vertebrate rods [8] and are the only mammalian visual pigment class where the protein structure has been experimentally determined [38,41]. Therefore, this, and in particular the bovine rod opsin protein structure, is the only template available for accurate homology modeling of vertebrate cone opsins. The sequence identity of the bovine RH1 template compared to teleost Rh1 rod opsins and Rh2 cone opsins ranged from 49–83% and 63–72%, respectively; which proved to be more than sufficient for the reliable modeling and accurate prediction of λ_{\max} values [10,34]. In the present study, we test whether this approach is also reliable for predicting λ_{\max} values of more

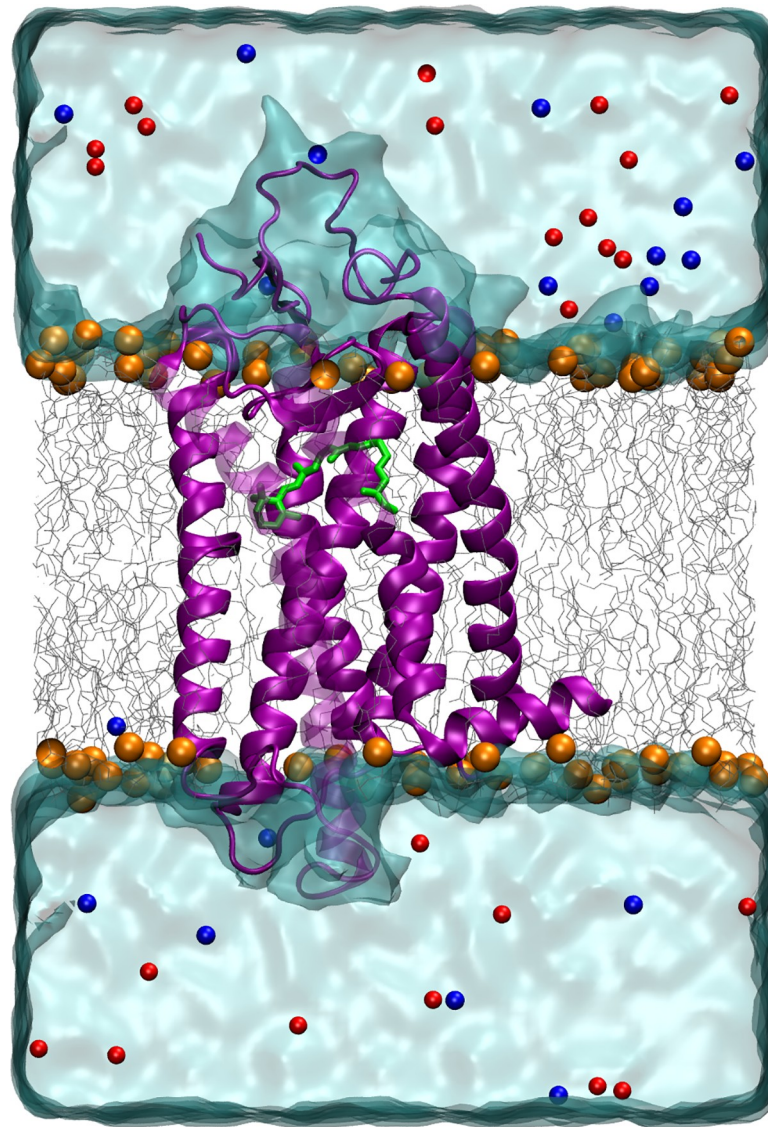


Fig 2. A representative 3D structure of Sws2 cone opsin (λ_{\max} values <430 nm) homology structure (violet) with the chromophore (green) bound covalently to K296 of the opsin protein. It is inserted in a phospholipid bilayer (gray, carbon atoms; orange, phosphorus atoms) and surrounded by water molecules (light blue). Blue and red spheres indicate positive and negative counter ions, respectively.

<https://doi.org/10.1371/journal.pcbi.1008212.g002>

evolutionarily-distant Sws2 cone opsins (Fig 1) that have ~48–51% sequence identity to the bovine RH1 template (Table 1). Furthermore, the selected Sws2 cone photopigments display a broader range and blue-wavelength-shifted λ_{\max} values (i.e. 397–485 nm) compared to either Rh2 cone photopigments or rod opsins presented in prior teleost studies: 467–528 nm for Rh2 [34] and 444–519 nm for Rh1 [10] compared to the bovine RH1 template with a λ_{\max} value at 498 nm [9]. Therefore, this present study also tests the capacity of our prior approach of combining homology modeling and MD simulations to accurately predict λ_{\max} values for an opsin class that is difficult to predict from just the amino acid sequences, that are short-wavelength-shifted compared to the bovine template, and cover 88 nm of the electromagnetic spectrum.

Using the sequence information of 11 teleost Sws2 opsins, we built homology models based on the bovine rod photopigment (RH1 opsin + 11-*cis* retinal chromophore) structure as a

template (Protein Data Bank (PDB) ID: 1U19) [41]. These homology structures with the chromophore attached covalently to K296 within the binding pocket were placed in lipid bilayers and water models (Fig 2). Each of these systems were then subjected to 100 ns classical MD [42] simulations using the protocol described in the Methods section (S1 Movie).

We analyzed the MD simulations for all 11 visual photopigments to understand the dynamics and to identify structural features associated with the chromophore and attached lysine residue (Fig 3A) that could potentially be used to explain differences in the λ_{\max} values of the representative Sws2 photopigments used in this study. To understand the dynamics of the chromophore within the opsin binding pocket, we visualized the conformations of the chromophore seen in violet- (λ_{\max} values <430 nm) vs. blue-sensitive (λ_{\max} values >430 nm) photopigments (Fig 3B), where we observed a relatively compact cluster of chromophore conformations for blue-sensitive photopigments compared to violet-sensitive photopigments. This difference was more evident for the β -ionone ring and two methyl groups present at positions C9 and C13. In our previous study, the area under the curve (AUC) of root mean square

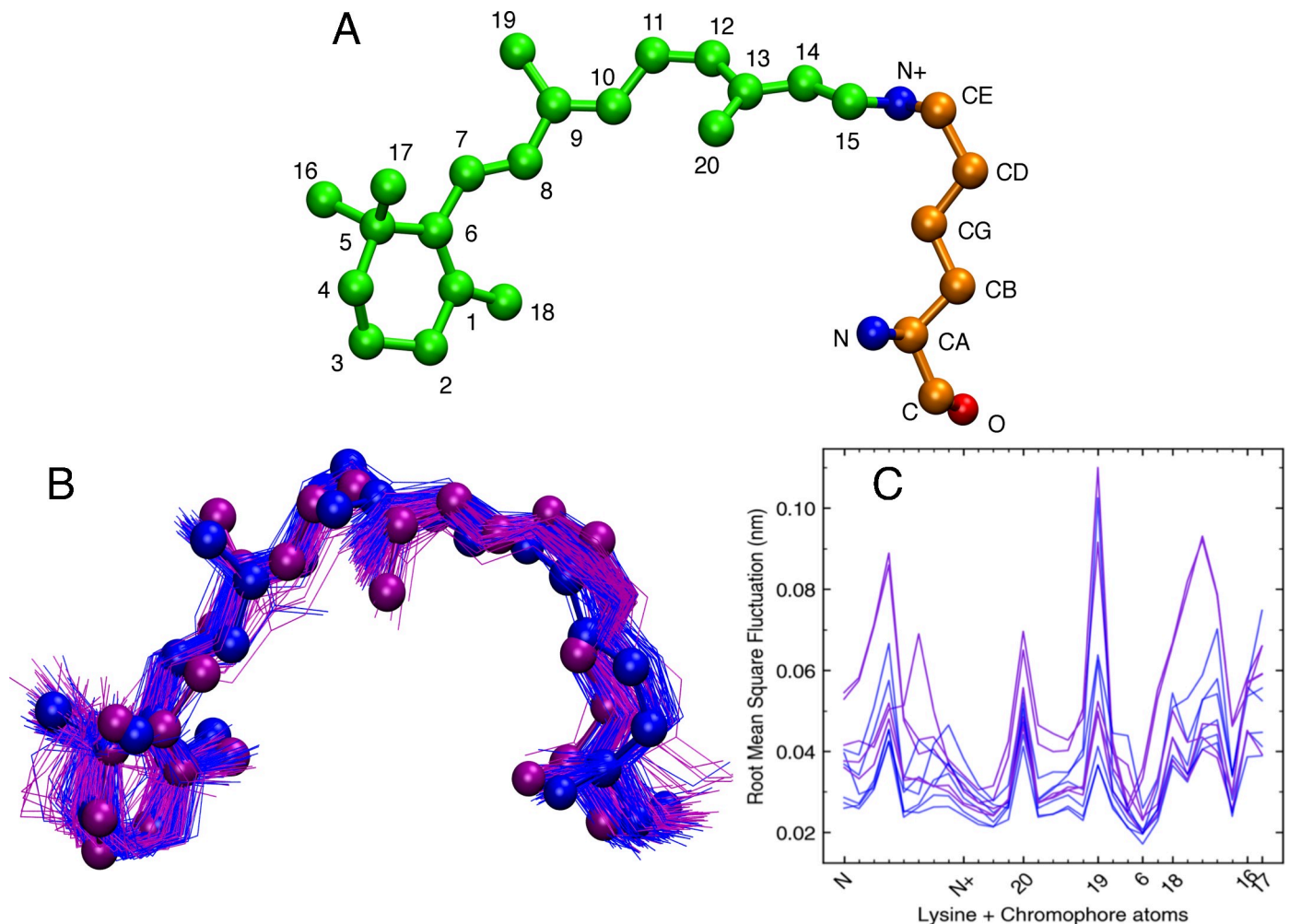


Fig 3. Conformations and fluctuations of 11-*cis* retinal chromophore and attached lysine in Sws2 photopigments. A) 3D orientation of 11-*cis* retinal linked to K296 of the opsin protein. B) Superposition of 11-*cis* retinal conformations from MD simulations trajectories. C) Root mean square fluctuation (RMSF) of 11-*cis* retinal linked to K296 (LYS+RET). The horizontal axis represents atoms of the LYS+RET. Blue- vs. violet-colored ball-and-stick conformations are those associated with Sws2 photopigments with λ_{\max} values >430 nm vs. λ_{\max} values <430 nm, respectively.

<https://doi.org/10.1371/journal.pcbi.1008212.g003>

fluctuations (RMSF) served as an additional feature of the chromophore that helped to predict the λ_{\max} values of Rh2 cone visual photopigments [34]. We, therefore, first calculated RMSF values of all the heavy atoms of the chromophore and the linked lysine residue (LYS+RET) (Fig 3C) for each Sws2 photopigment as follows:

$$RMSF_{(V)} = \sqrt{\frac{1}{T} \sum_{t=1}^T (v_t - \bar{v})^2},$$

where T is the total number of molecular dynamics trajectory frames (V) and then calculated AUC of RMSF values (AUC RMSF_(LYS+RET)) for each Sws2 photopigment. The atoms within the blue-sensitive photopigments clearly show lower RMSF values compared to the violet-sensitive photopigments (Fig 3C), suggesting that this chromophore feature may also be useful in predicting of λ_{\max} values. Interestingly, this outcome is the opposite to what one would anticipate based upon our prior study, in which lower values of AUC RMSF_(LYS+RET) were associated with photopigments with shorter wavelength λ_{\max} values [34].

The dynamic nature of the chromophore during MD simulations suggests that we may use the geometric angles, dihedrals and AUC RMSF_(LYS+RET) parameters to differentiate between blue-sensitive and violet-sensitive Sws2 photopigments. For each photopigment, we examined a total of 19 angles (15 Torsion Angles and 4 Geometric Angles) (S2 Fig, Fig 3A) formed by the heavy atoms of the lysine residue at position 296 of the opsin covalently linked to 11-*cis* retinal. The model parameters that showed a reasonable correlation to experimental λ_{\max} values were the median values of Torsions 3, 9, 10, 11, 12, as well as Geometric Angles 1 and 3, from a total of 19 examined angles (S3 Fig). The standard model selection procedure was then used to determine the simplest linear regression model that best fitted to the shortlist of parameters showing a reasonable correlation to experimental λ_{\max} values. From our model selection procedure, we found the simplest model for the 11 Sws2 photopigments examined contained three terms: the median values of Torsion 3 (C15–C14–C13–C20), Torsion 12 (C19–C9–C8–C7), and Angle 3 (C3–C7–C8) (Fig 4). AUC RMSF_(LYS+RET) values were not identified by the model selection procedure as being predictively useful.

The full model is explicitly given by:

$$\begin{aligned} & \text{Predicted } \lambda_{\max} \text{ value} \\ & = 2677.5348 - (17.052 \times \text{Angle 3}) + (5.1634 \times \text{Torsion 3}) + (2.3642 \times \text{Torsion 12})(1) \end{aligned}$$

The larger values of Angle 3 predicted a spectral shift towards shorter wavelengths (i.e. a violet shift), while larger values of Torsion 3 and Torsion 12 predicted a shift towards longer wavelengths (i.e. a blue shift). Fig 5 shows empirically determined λ_{\max} values vs. the model predicted values for each Sws2 photopigment analyzed, where our full model highly correlates with experimental data ($R^2 = 0.95$). The error in prediction (error = |pred-exp|) ranged from

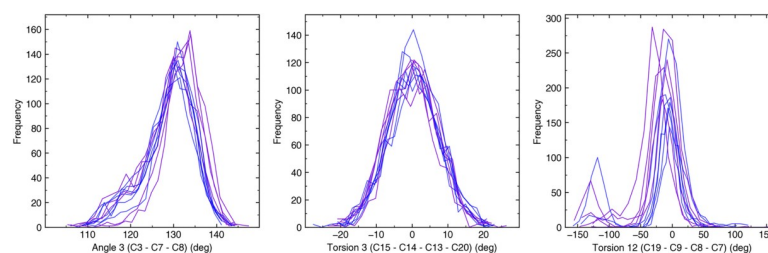


Fig 4. Frequency distribution of Angle 3, Torsion 3 and Torsion 12 observed in each opsin simulation. Blue and violet colors correspond to Sws2 photopigments with λ_{\max} values >430 nm vs. λ_{\max} values <430 nm, respectively.

<https://doi.org/10.1371/journal.pcbi.1008212.g004>

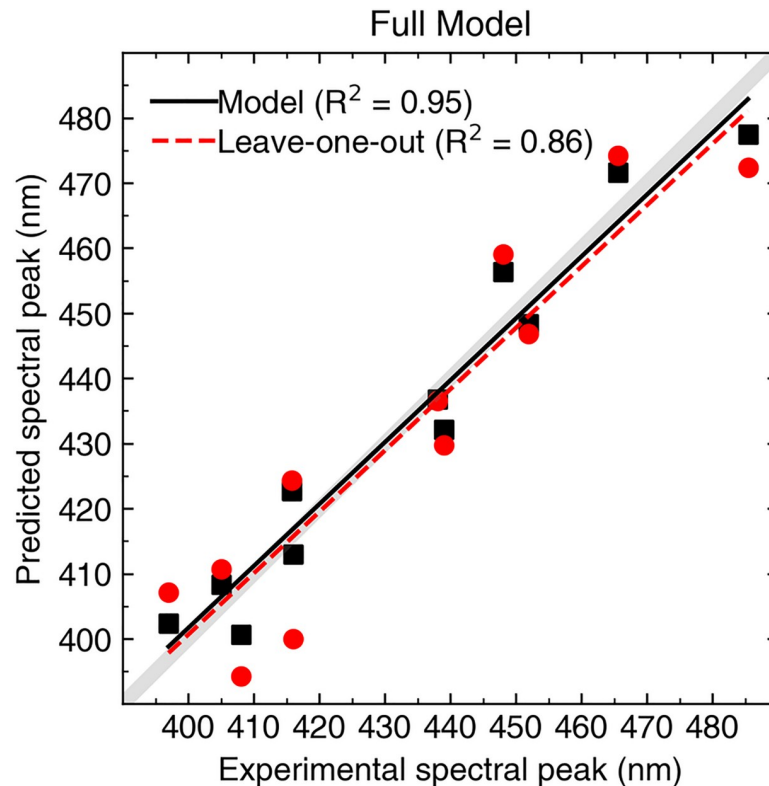


Fig 5. Experimental spectral peaks of absorbance (λ_{\max}) compared to predicted λ_{\max} values by the full model Eq 1 outlined in the main text for all 11 Sws2 photopigments analyzed. Gray lines indicate a perfect (100%) correlation. Solid black lines and black symbols represent the linear relationships between model-predicted and the experimental λ_{\max} values, whereas dashed red lines and red symbols show linear relationships between “leave-one-out” predictions and experimental λ_{\max} values. Corresponding correlation coefficients for both approaches are indicated.

<https://doi.org/10.1371/journal.pcbi.1008212.g005>

1.21 nm to 8.35 nm with an average error of 5.44 nm. To further test our statistical model, we carried out a “leave-one-out” analysis, where each Sws2 photopigment was removed from the regression analysis to obtain the coefficients for a model using Angle 3, Torsion 3 and Torsion 12 parameters, and then the λ_{\max} value of the removed photopigment was predicted based upon the new linear model. The correlation of the individual predictions based upon only 10 pigments was reduced, but it was still highly acceptable ($R^2 = 0.86$) (Fig 5). The lower correlation coefficient derived from the “leave-one-out” approach is largely explained by the less accurate prediction of the λ_{\max} value for the zebrafish Sws2 photopigment (i.e. experimental λ_{\max} value at 416 nm vs. “leave-one-out” predicted λ_{\max} value at 400 nm). This less accurate prediction is likely due to Torsion 3, which has a relatively higher median value compared to Torsion 12 with a relatively lower median value for zebrafish compared to the median values of Torsion 3 and Torsion 12 of Sws2 pigments from other species. Nevertheless, our results indicate that the statistical model derived from MD simulations of predicted Sws2 visual photopigment structures has the power to predict their λ_{\max} values with reasonable accuracy.

To further test this approach and to simulate performance at predicting unknown λ_{\max} values for Sws2 photopigments, we performed a “leave-one-out” approach at a species level. For each species where both Sws2a and Sws2b opsins are represented in the dataset (Table 1), a subset was generated where their Sws2 chromophore parameters were removed. A new model based on the newly generated subset was chosen using the full model selection procedure detailed in the Methods section. This new best-fit model was then used to predict the spectral

peaks of absorbance for the omitted opsins. In this study, there are four species with both Sws2a and Sws2b photopigments (Table 1), namely the medaka, the bluefin killifish, the guppy, and the spotted flounder. For these four species, the following models were determined:

i. With the medaka Sws2 photopigment removed:

$$\lambda_{max} = 2761.5524 + 2.405 \times Torsion\ 12 - 17.64 \times Angle\ 3 \quad (2)$$

ii. With the spotted flounder Sws2 photopigment removed:

$$\lambda_{max} = 4201.6725 - 1.4975 \times Torsion\ 10 - 24.7764 \times Angle\ 1 - 3.2922 \times Angle\ 3 \quad (3)$$

iii. With the bluefin killifish Sws2 photopigment removed:

$$\lambda_{max} = 2706.2512 - 3.1214 \times Torsion\ 10 - 13.8215 \times Angle\ 3 \quad (4)$$

iv. With the guppy Sws2 photopigment removed:

$$\lambda_{max} = 5624.1521 + 3.6078 \times Torsion\ 9 - 1.8923 \times Torsion\ 10 - 11.0413 \times Torsion\ 11 + 16.0527 \times Torsion\ 12 - 20.9014 \times Angle\ 3 \quad (5)$$

Performance of the “leave-one-species-out” models varied with R^2 values of 0.84 to 0.93 for the full model and 0.76 to 0.94 for the “leave-one-photopigment-out” approach (S4 Fig). This indicates that our approach generates models with good predictive power for new Sws2 sequences with unknown λ_{max} values that are not included in the model selection process. However, the drop in the performance when leaving out certain members can be attributed phylogenetic relationship among the chosen sequences and to the fact that we are leaving out two sequences from a relatively small dataset of 11 Sws2 sequences and leaving only nine opsin sequences to build the model using the shortlisted seven structural parameters. Angle 3 was present in all models (Eqs 2–5), as well as in the full model (Eq 1), and had the largest weighting within each approach, with the exception of the model where the spotted flounder Sws2 modeling experiment was omitted. In this latter model, Angle 1 was also present and had the largest weighting. In general, these findings imply that Angle 3 (C3–C7–C8) is a significant predictor of λ_{max} values for Sws2 photopigments.

To explain the mechanistic role of 11 putative spectral tuning sites at positions 41, 47, 94, 97, 99, 109, 116, 168, 183, 269, 299 (S1 Fig) revealed by comparative analyses of Sws2a vs. Sws2b sequences [22,23] on the conformation of the chromophore, as predicted by Cortesi et al. (2015) [22], we carried out MD trajectory analyses using visual molecular dynamics (VMD) via visual inspection. It should be noted that except for positions 94 and 269, the other nine putative tuning sites are located far from the chromophore binding pocket. Specifically, position 94 is located in close proximity to the heavy atoms (CE, N+) (Fig 3A) of the K296 residue that is covalently bound to the chromophore, so this site was analyzed in detail. In teleost Sws2b photopigments, position 94 is occupied by a cysteine (C) residue and its sidechain invades the space next to atoms (CE, N+) of K296, resulting in reduced space available for fluctuations of dihedrals/angles of the chromophore. By contrast, position 94 in Sws2a photopigments is occupied by threonine (T), alanine (A) or glycine (G). The T94 residue interacts with S186 and affects the atoms (CE, N+) of the chromophore bound to K296, while A94 and G94

lack large sidechains, thus providing space for the fluctuation of these atoms (CE, N+). To understand the influence of the spectral tuning site at position 94 on the fluctuation of these heavy atoms, we analyzed the dihedrals and geometric bond angles formed by (CE, N+) atoms of K296. We found Angle 2 to be clearly distinguishable for Sws2a compared to Sws2b photopigments (S5 Fig), such that shorter wavelength-shifted (i.e. violet) Sws2b photopigments have wider distributions of Angle 2 compared to longer wavelength-shifted (i.e. blue) Sws2a photopigments.

The spectral tuning site at position 269, as described by Kasagi et al. (2018) [23], is occupied by alanine in all of the Sws2 opsin proteins analyzed in this study, except for *V. variegatus* Sws2a, in which this site is occupied by threonine that results in a shift of the λ_{\max} value to longer wavelengths. Analysis of MD trajectories reveals that position 269 is located within the chromophore binding pocket and directly interacts with the β -ionone ring of the chromophore. The methyl group of the A269 side chain forms favorable hydrophobic contacts with the β -ionone ring of the chromophore, but with an A269T substitution, the hydroxyl group of threonine results in the β -ionone ring of the chromophore being positioned more distantly. Among all the 19 structural parameters analyzed that are associated with the chromophore, only Torsions 11 and 12 display spatial relationships such that they could be affected by known tuning sites (see S1 Fig for positions of known tuning sites). Interestingly, Torsion 12 is affected by A269T and was indeed useful for predicting λ_{\max} values as can be seen in the combined full model.

Discussion

We have developed a new model for the prediction of the spectral peaks of absorbance for teleost Sws2 cone photopigments, with high accuracy over a wide range of λ_{\max} values (i.e. 397–485 nm). Our approach is based on a predictive model as described in our previous studies [10,34]. In the present study, this approach required Sws2 opsin protein sequence data as input and a known template photopigment structure to build the homology models. MD simulations were then performed on these opsin homology models, with parameters that describe the conformational change in both opsin structure and the covalently-bound chromophore being extracted. From this, a statistical model was built. Similar to our previous studies, our current approach revealed that the structural features of the chromophore and its lysine attachment site play important roles in the prediction of λ_{\max} values. In fact, the final predictive model consisted of three terms associated with chromophore conformation. This simple first-order regression model was found to be sufficient to estimate the λ_{\max} values of Sws2 photopigments with high accuracy. This study further highlights the versatility of our approach in reliably predicting the λ_{\max} values of evolutionarily more distant Sws2 cone opsin sequences, with ~48–51% sequence identity to the bovine RH1 template.

A number of molecular and evolutionary approaches have been used in the field of visual neuroscience to predict visual photopigment λ_{\max} values. One common strategy is the application of site-directed mutagenesis to opsin sequences derived directly found to be expressed in a particular extant species, where amino acid substitutions are made followed by measuring the spectral characteristics to identify potential contributions of specific amino acid residues to any observed spectral shifts [2,23,25–27,30]. A similar strategy is to infer the amino acid sequence of the ancestral opsin sequence within a clade, followed by the same technique to experimentally determine λ_{\max} values [9,25]. Whereas the latter approach generally involves investigating multiple amino acid substitutions that may or may not be directly related to spectral tuning, the former technique frequently only studies single residue differences. Nonetheless, these methods are frequently used together and have complemented any comparative

analyses that preliminarily identify residues that are likely to influence the λ_{\max} value of a particular photopigment (e.g. Cortesi et al. (2015) [22]; reviewed by Shichida and Matsuyama (2009) [43]). For example, such an approach identified that a spectral shift of the λ_{\max} value of teleost Rh2 photopigments, specifically with sensitivity in the green region of the visible spectrum to blue, was largely due to an E122Q substitution [24,25]. Apart from some vertebrate LWS photopigments [32] and UV-sensitive SWS1 photopigments [26,33], the prediction of other cone opsin λ_{\max} values (i.e. SWS2 and RH2 photopigment classes) by manual interrogation of the amino acid sequences alone is extremely difficult and often inaccurate [1,2]. Experimental spectral analyzes by MSP and/or *in vitro* regeneration of photopigments are not always viable options due to the financial limitations and the demand for specialist technical expertise. As such, alternative methods of accurately predicting λ_{\max} values to understand photopigment function and ecological adaptation are critical. Our alternative molecular modeling-based approach is, by contrast, simple, accurate and efficient, and does not require site by site substitutions followed by *in vitro* experimentation or complex quantum calculations. Previously, we successfully used a similar approach presented in this study to accurately predicted a broad range of λ_{\max} values (467–528 nm) for teleost Rh2 cone photopigments [34] and rod (Rh1) photopigments (444–519 nm) [10]. With this current investigation, our approach now provides accurate predictions of λ_{\max} values for Sws2 photopigments from 397–485 nm (i.e. those sensitive to violet vs. blue wavelengths). Not only is our approach useful for accurately predicting λ_{\max} values using known spectral tuning sites, it may be used to identify and assess the spectral effects of putative unknown residues on the spectral properties of photopigments. It should be noted, however, that the predictions made using our modeling approach results in λ_{\max} values that are based on rhodopsins that utilize a vitamin-A₁-derived chromophore. This is also the case for experimental approaches using *in vitro* regeneration protocols [2,23,25–27,30]. In some vertebrates (e.g. lampreys [31,44,45], many freshwater teleosts [46], lungfishes [47,48], the green anole lizard *Anolis carolinensis* [49,50]; reviewed in [1–3,51]), the visual system is based on porphyropsins that incorporate a vitamin-A₂-derived chromophore or a combination of both rhodopsins and porphyropsins. Therefore, within the context of biological relevance, the predicted λ_{\max} values that result using our approach may have to be converted, where appropriate, to longer wavelengths to account for the possession of a vitamin-A₂-derived chromophore in native photopigments. This is easily conducted by using a number of rhodopsin-to-porphyropsin transformation algorithms, such as those by Loew and Dartnell [52], Harosi [53], and Whitmore and Bowmaker [14].

One of the key elements of our predictive model is Angle 3 (C3–C7–C8), which is the dominant parameter in predicting the λ_{\max} values of Sws2 photopigments. In contrast, the dominant parameters for the λ_{\max} predictive models in Rh2 and RH1 class photopigments were AUC RMSF_(LYS+RET) and presence and absence of disulfide bridge, respectively [10,34]. This suggests that the structural features associated with spectral tuning are visual photopigment class specific. Specifically, our results show that larger values of Angle 3 lead to greater shorter wavelength shifts of λ_{\max} values to the violet region of the visible spectrum. Furthermore, Angle 3 is an important parameter when the “leave-one-out” approach was applied at a species level for predicting the spectral peaks of absorbance for unknown opsins, suggesting this parameter is broadly predictive. Other parameters affecting the prediction of λ_{\max} values are Torsion 3 and Torsion 12, but the magnitude of their effects is reduced in comparison with that of Angle 3. Nonetheless, like Angle 3, larger values of Torsion 3 and Torsion 12 also short-wavelength-shifted the λ_{\max} value. Interestingly, Torsion 12 was the only element of the “three-term” model that is likely to be directly influenced by a known or suspected tuning site (i.e. residue 269). These results explain the possible mechanism that causes the observed long-wavelength shift in the λ_{\max} value of the spotted flounder Sws2a photopigment. With the

exception of this specific example, other residues affecting Angle 3, Torsion 3, and Torsion 12 cannot be pinpointed to any one particular known or suspected tuning site. Instead, it is likely that multiple tuning sites collectively, either directly or indirectly, influence these chromophore structural features. This finding underscores the power of the homology modeling/MD approach, which takes into consideration the collective influence of the entire amino acid sequence to predict λ_{\max} values rather than the sole specific contributions of individual sites. Although predictive and useful for explaining the influence of the known tuning sites, our genome to phenome approach does not provide chemical-physics mechanism to explain the shift in the spectral peaks of absorbance and identification of opsin-charge determinants affecting the chromophore. Such questions would be best answered via thorough QM (quantum mechanics)/MM (molecular mechanics) investigation for each Sws2 sequence [54]. Nonetheless, we believe our study can provide a reasonable starting geometry to carry out such QM/MM simulations.

Within the chromophore of the *D. rerio* (zebrafish) Sws2 photopigment, Torsion 3 has a relatively higher median value compared to the median value range of Torsion 3 of Sws2 photopigments found in other species, while Torsion 12 has a relatively lower median value compared to the median value range of Torsion 12 of other Sws2 photopigments. These “out-of-range” median values of Torsion 3 and Torsion 12 lead to a less accurate prediction of the λ_{\max} value for the *D. rerio* Sws2 photopigment. Based on the evolutionary relationships of teleost Sws2 opsin protein sequences examined in this study, it appears that *D. rerio* (and maybe other cyprinids) diverged from the main *sws2* opsin clade before *sws2* duplicated into *sws2a* and *sws2b* subclasses. It is possible, therefore, that the zebrafish Sws2 opsin protein holds the retinal chromophore in a distinct conformation vs. the other Sws2 proteins investigated in this study, but which also results in a λ_{\max} value that is similar to that exhibited by the Sws2b photopigment subclass. Thus, teleost Sws2 photopigments may have independently evolved more than one opsin-chromophore conformation strategy for attaining short-wavelength shifts of the λ_{\max} value to the violet region of the visible spectrum. Such knowledge means that the model presented here and its predictive power might be improved in the future if a more diverse set of Sws2 photopigments is used to develop a more all-encompassing Sws2 model and/or to develop distinctive models for some selected phylogenetic groups.

Comparative analyzes of Sws2a vs. Sws2b photopigments have revealed candidate spectral tuning sites that could potentially explain the sensitivity to violet vs. blue wavelengths for Sws2b and Sws2s subgroup, respectively [22]. Thus, MD simulations can serve as a valuable, complementary tool to understand the contributions of candidate tuning sites to spectral shifts in the λ_{\max} value. From MD trajectory analysis, Angle 2 was identified as a potential parameter that is affected by the presence of different residues at spectral tuning site 94. However, we note that as Angle 2 did not display a significant linear correlation with the actual λ_{\max} value, Angle 2 was not considered as a candidate for the model selection procedure; as such, Angle 2 did not appear in the predictive “three-term” model.

In conclusion, we have successfully tested our previously studied molecular modeling approach combining homology modeling, MD simulations and structural information of chromophore conformations and to accurately predicted the λ_{\max} of Sws2 opsins. In future studies, we plan to consider additional features of each parameter (e.g. narrow vs. broad distribution) other than simple linear correlations, and other structural features such as distance between certain atoms and the functional group of the chromophore or the opsin tuning sites, which may further improve the predictive power of the resulting models. We will also expand our approach to extensively study the more phylogenetically distant classes of opsins, with large number of opsins included in the dataset. Finally, we aim to develop model(s) using a dataset of individual distinct classes of opsins to generate an online web platform for accurately predicting the λ_{\max} values for any unknown vertebrate photopigment.

Methods

Phylogenetic analysis

The alignment and phylogenetic reconstructions were performed using the function "build" of ETE3 v3.1.1 [55] as implemented on the GenomeNet (<https://www.genome.jp/tools/ete/>). The multiple sequence alignment was provided as input file. ML tree was inferred using PhyML v20160115 ran with model and parameters:--alpha e--pinv e--f m--o tlr--bootstrap -2--nclases 4 [56]. Branch supports are the Chi2-based parametric values return by the approximate likelihood ratio test.

Homology modeling

Eleven teleost Sws2 cone opsin amino acid sequences (Table 1), namely *V. variegatus* (spotted flounder) [23], *M. achne* (slime flounder) [23], *P. olivaceus* (olive flounder) [23], *P. reticulata* (guppy) [39], *O. latipes* (Japanese rice fish; medaka) [20], *L. goodei* (bluefin killifish) [40], and *D. rerio* (zebrafish) [1], were downloaded from the UniProt database (<https://www.uniprot.org/>). These were selected because the corresponding spectral peaks of absorbance for their Sws2 photopigments (when reconstituted with a 11-*cis* retinal chromophore) have been experimentally measured. Collectively, these photopigments exhibit a wide range of λ_{\max} values from 397–485 nm. An experimental 3D structure of a Sws2 cone photopigment is not available; hence, to build a homology model of 11 Sws2 photopigments, a template search was carried out using SWISS-MODEL (<https://swissmodel.expasy.org/>). The closest homologue (~50% sequence identity) with a high-quality 3D structure was found to be that of the bovine rod opsin (RH1). A high-resolution crystallographic structure of the bovine RH1 photopigment (PDB ID 1U19, 2.2 Å) [38], which lacks mutations and has an 11-*cis* retinal chromophore covalently bound within its binding pocket, was downloaded from the Protein Data Bank. 3D coordinates of the bovine rod opsin structure were then used to build the homology models of 11 teleost Sws2 cone opsin sequences. The structure prediction wizard from the PRIME module of the Schrödinger suite was used for building a homology model for each protein sequence [57,58]. The non-templated loops were refined using the refine loops module of PRIME and a generated model structure was validated by generating a Ramachandran plot by analyzing acceptable phi-psi regions of residues. The final homology model was modified to remove the intracellular unstructured coil (~25 residues) region towards the carboxyl-terminus to prevent it from crossing the periodic boundaries during the molecular dynamics (MD) simulation.

Molecular dynamics (MD) simulation

All 11 Sws2 photopigment homology models were subjected to atomistic MD simulations using our protocol for input file generation and the system setup for MD simulations reported in our previous studies [10,34]. Briefly, the homology models of each Sws2 opsin sequence with the chromophore bound covalently to the lysine residue in the binding pocket (K296) were uploaded to the CHARMM-GUI webserver (<http://charmm-gui.org>). Each system was placed in lipid bilayers and hydrated using a hexagonal solvent box with a 15 Å TIP3P water layer. The charge of the system was neutralized with 150 mM NaCl. The CHARMM36m force-field [32] parameters were selected for all the components of the systems. After minimization and short equilibration simulations with harmonic restraints, the systems were subjected to 100 ns atomistic MD simulations. The production simulations were performed under an NPT ensemble for 100 ns using a Parrinello-Rahman barostat [59] with semi-isotropic pressure coupling and a Nosè-Hoover thermostat [60]. During the production MD simulations,

snapshots were saved every 10 ps. GROMACS-2018.3 [61] was used for all 11 MD simulations. The visualization and analysis of MD trajectories were carried out using Visual Molecular Dynamics package [62]. GROMACS trajectory analysis tools were used to analyze Root Mean Square Fluctuations (RMSF) and 19 different internal degrees of freedom of the chromophore (i.e. torsion angles and geometric bond angles) (S1 File).

Quantification and statistical analysis

To determine the best linear regression model for predicting Sws2 λ_{\max} values, and to avoid overfitting (i.e. a model with the same number of terms as the number of datapoints it is being fit to) we chose a shortlist of parameters to use in model selection based on the linear correlation of each of the 19 median angles and the experimental λ_{\max} (S2 Fig). These were then ranked by r^2 (i.e. coefficient of determination) values and the top seven performing angles were used, exception being taken with angles which did not show true linear correlation through visual inspection (S3 Fig). The resulting model selection parameters included the median values of Torsions 3, 9, 10, 11, 12, and Geometric Angles 1 and 3 (S2 File). From our list of 19 parameters, a shortlist was generated that composed of the medians of Torsion 3, Torsion 9, Torsion 10, Torsion 11, Torsion 12, and geometric Angles 1 and 3. A model selection procedure was then performed using the `regsubsets` function of the “leaps” R library (<https://cran.r-project.org/web/packages/leaps/leaps.pdf>). This evaluated all possible model subsets using the shortlisted angles and ranked them according to their Bayesian Information Criterion (BIC) value [63], which quantifies the explanatory power of a model with a penalty for the number of terms included. The resulting best-fit model was further evaluated via a “leave-one-out” procedure by reweighting the parameters of the best-fit model after removing each photopigment iteratively. For each Sws2 photopigment stimulation, the parameters of the best-fit model were reweighted with the given Sws2 photopigment data removed. This reweighted model was subsequently used to predict the spectral peak of absorbance for the omitted Sws2 opsin sequence.

Additionally, to further validate our approach and to simulate performance at predicting the spectral peaks of absorbance for Sws2 photopigments with unknown λ_{\max} values, a systemic “leave-one-species-out” process was also performed. For a given species, both Sws2a and Sws2b opsin sequences were removed from the dataset. A new model was chosen using the full model selection procedure detailed above. Since removing two opsins results in a smaller sample of data, overfitting (i.e. a model with the same number of terms as the number of datapoints it is being fit to) is a possible risk factor. To mitigate this, the generated model space was explored within +2 BIC value of the best BIC ranked value. The resulting model was then used to predict the spectral peak of absorbance for the removed Sws2a and Sws2b photopigments.

Supporting information

S1 Fig. Sequence alignment of Sws2 opsins used in the simulations, together with the bovine rod opsin template. Red arrows indicate amino acid positions (using numbering standardized to bovine rod opsin) highlighted by Cortesi et al. (2015) [22] that were identified as likely to have contributed to the functional diversification of teleost Sws2 opsin sequences, particularly those that distinguish Sws2b photopigments (violet-sensitive) from the more long-wavelength-shifted Sws2a photopigments (blue-sensitive). The blue arrow indicates position 269, that is likely to be the spectral tuning site that mediates a shift in the λ_{\max} value of *V. variegatus* Sws2a to 485 nm [23]. A large orange asterisk (*) depicts a conserved lysine (K) residue at position 296 that is required for the formation of the Schiff-base linkage to the retinal chromophore. The gray bars below each alignment indicate a quality score, such that lower scores

correspond to greater amino acid variability in the column. All species and photopigment subclasses are followed by experimentally measured λ_{\max} values.

(TIF)

S2 Fig. Frequency distribution of all analyzed torsions and geometric angles observed for each photopigment simulation. Each panel represents the frequency distribution of an individual torsions and angles associated with photopigment. Blue and violet colors indicate the relative spectral peaks of absorbance of each Sws2 photopigment, corresponding to λ_{\max} values >430 nm vs. λ_{\max} values <430 nm, respectively.

(TIF)

S3 Fig. Linear correlation between all analyzed parameters and experimental spectral peaks of absorbance. Each panel shows the relationship between the given parameter and the experimental spectral peaks of absorbance. Correlation in terms of r^2 is given and the corresponding best-fit line is shown in black.

(TIF)

S4 Fig. Experimental spectral peaks of absorbance (i.e. the λ_{\max} value) compared to predicted λ_{\max} values generated by the model resulting from “leave-one-system-out” analysis. The prediction for the removed Sws2a and Sws2b photopigments are indicated by blue and violet squares, respectively. Black squares correspond to the Sws2 photopigments used in the model selection procedure with the black line indicating the R^2 for the model. Red circles and dashed line correspond to the leave-one-out analysis for single photopigments included in the model. The grey line corresponds to a 1:1 relationship.

(TIF)

S5 Fig. Frequency distribution of Angle 2 observed in Sws2a and Sws2b of each species used in the simulation. Each panel represents the frequency distribution of Angle 2 observed in individual species with Sws2a and Sws2b (*L. goodei*, *O. latipes*, *P. reticulata*, *V. variegatus*). Sws2a and Sws2b photopigments are indicated by blue and violet colors, respectively.

(TIF)

S1 Movie. Molecular dynamics simulations of Sws2a opsin with 11-*cis* retinal from *V. variegatus* (spotted flounder) embedded in the lipid bilayer and water.

(MP4)

S1 File. Time series data for all analyzed torsion and geometric angles extracted from each photopigment MD simulation.

(ZIP)

S2 File. Median values associated with all analyzed angles used for statistical analysis.

Median values associated with all 19 analyzed parameters are provided for each Sws2 sequence with the experimental spectral peak of absorbance. Rows are highlighted based on absorbance, with blue and violet corresponding to λ_{\max} values >430 nm vs. λ_{\max} values <430 nm, respectively. Columns X and Y give the predicted λ_{\max} for the full model and leave-one-out models respectively with Z and AA showing the corresponding absolute value for the error.

(XLSX)

Acknowledgments

We thank Dr. F. Marty Ytreberg for constructive feedback on an earlier version of the manuscript.

Author Contributions

Conceptualization: Wayne I. L. Davies, Deborah L. Stenkamp, Jagdish Suresh Patel.

Formal analysis: Dharmeshkumar Patel, Jonathan E. Barnes, Deborah L. Stenkamp.

Funding acquisition: Wayne I. L. Davies, Deborah L. Stenkamp, Jagdish Suresh Patel.

Investigation: Dharmeshkumar Patel, Jonathan E. Barnes.

Methodology: Dharmeshkumar Patel, Jonathan E. Barnes, Jagdish Suresh Patel.

Project administration: Jagdish Suresh Patel.

Software: Dharmeshkumar Patel, Jonathan E. Barnes.

Supervision: Wayne I. L. Davies, Deborah L. Stenkamp, Jagdish Suresh Patel.

Validation: Dharmeshkumar Patel, Jonathan E. Barnes.

Visualization: Dharmeshkumar Patel.

Writing – original draft: Dharmeshkumar Patel, Jonathan E. Barnes, Wayne I. L. Davies, Deborah L. Stenkamp, Jagdish Suresh Patel.

Writing – review & editing: Dharmeshkumar Patel, Jonathan E. Barnes, Wayne I. L. Davies, Deborah L. Stenkamp, Jagdish Suresh Patel.

References

1. Yokoyama S, Blow NS, Radlwimmer FB. Molecular evolution of color vision of zebra finch. *Gene*. 2000; 259:17–24. [https://doi.org/10.1016/S0378-1119\(00\)00435-2](https://doi.org/10.1016/S0378-1119(00)00435-2) PMID: 11163957
2. Davies WIL, Wilkie SE, Cowing JA, Hankins MW, Hunt DM. Anion sensitivity and spectral tuning of middle- and long-wavelength-sensitive (MWS/LWS) visual pigments. *Cell Mol Life Sci*. 2012; 69:2455–2464. <https://doi.org/10.1007/s00018-012-0934-4> PMID: 22349213
3. Bowmaker JK. Evolution of vertebrate visual pigments. *Vision Res*. 2008; 48:2022–2041. <https://doi.org/10.1016/j.visres.2008.03.025> PMID: 18590925
4. Hárosi FI. An analysis of two spectral properties of vertebrate visual pigments. *Vision Research*. 1994; 34:1359–1367. [https://doi.org/10.1016/0042-6989\(94\)90134-1](https://doi.org/10.1016/0042-6989(94)90134-1) PMID: 8023444
5. Parry JWL, Bowmaker JK. Visual pigment reconstitution in intact goldfish retina using synthetic retinal-dehyde isomers. *Vision Research*. 2000; 40:2241–2247. [https://doi.org/10.1016/S0042-6989\(00\)00101-2](https://doi.org/10.1016/S0042-6989(00)00101-2) PMID: 10927111
6. Wang W, Geiger JH, Borhan B. The photochemical determinants of color vision. *BioEssays*. 2014; 36:65–74. <https://doi.org/10.1002/bies.201300094> PMID: 24323922
7. Enright JM, Toomey MB, Sato S, Temple SE, Allen JR, Fujiwara R, et al. Cyp27c1 Red-Shifts the Spectral Sensitivity of Photoreceptors by Converting Vitamin A1 into A2. *Current Biology*. 2015; 25:3048–3057. <https://doi.org/10.1016/j.cub.2015.10.018> PMID: 26549260
8. Lagman D, Ocampo Daza D, Widmark J, Abalo XM, Sundström G, Larhammar D. The vertebrate ancestral repertoire of visual opsins, transducin alpha subunits and oxytocin/vasopressin receptors was established by duplication of their shared genomic region in the two rounds of early vertebrate genome duplications. *BMC Evolutionary Biology*. 2013; 13:238. <https://doi.org/10.1186/1471-2148-13-238> PMID: 24180662
9. Yokoyama S, Tada T, Zhang H, Britt L. Elucidation of phenotypic adaptations: Molecular analyses of dim-light vision proteins in vertebrates. *Proceedings of the National Academy of Sciences*. 2008; 105:13480–13485. <https://doi.org/10.1073/pnas.0802426105> PMID: 18768804
10. Musilova Z, Cortesi F, Matschiner M, Davies WIL, Patel JS, Stieb SM, et al. Vision using multiple distinct rod opsins in deep-sea fishes. *Science*. 2019; 364:588–592. <https://doi.org/10.1126/science.aav4632> PMID: 31073066
11. Hunt DM, Dulai KS, Partridge JC, Cottrill P, Bowmaker JK. The molecular basis for spectral tuning of rod visual pigments in deep-sea fish. *J Exp Biol*. 2001; 204:3333–3344. PMID: 11606607
12. Lythgoe JN. *The Ecology of Vision*. Clarendon Press; 1979.

13. Wald G. THE PORPHYROPSIN VISUAL SYSTEM. *J Gen Physiol.* 1939; 22:775–794. <https://doi.org/10.1085/jgp.22.6.775> PMID: 19873132
14. Whitmore AV, Bowmaker JK. Seasonal variation in cone sensitivity and short-wave absorbing visual pigments in the rudd *Scardinius erythrophthalmus*. *J Comp Physiol A.* 1989; 166:103–115. <https://doi.org/10.1007/BF00190215>
15. Marques DA, Taylor JS, Jones FC, Palma FD, Kingsley DM, Reimchen TE. Convergent evolution of SWS2 opsin facilitates adaptive radiation of threespine stickleback into different light environments. *PLOS Biology.* 2017; 15: e2001627. <https://doi.org/10.1371/journal.pbio.2001627> PMID: 28399148
16. Bloch NI. Evolution of opsin expression in birds driven by sexual selection and habitat. *Proceedings of the Royal Society B: Biological Sciences.* 2015; 282:20142321. <https://doi.org/10.1098/rspb.2014.2321> PMID: 25429020
17. Härer A, Torres-Dowdall J, Meyer A. Rapid adaptation to a novel light environment: The importance of ontogeny and phenotypic plasticity in shaping the visual system of Nicaraguan Midas cichlid fish (*Amphilophus citrinellus* spp.). *Molecular Ecology.* 2017; 26:5582–5593. <https://doi.org/10.1111/mec.14289> PMID: 28792657
18. Lin J-J, Wang F-Y, Li W-H, Wang T-Y. The rises and falls of opsin genes in 59 ray-finned fish genomes and their implications for environmental adaptation. *Scientific Reports.* 2017; 7:15568. <https://doi.org/10.1038/s41598-017-15868-7> PMID: 29138475
19. Chinen A, Hamaoka T, Yamada Y, Kawamura S. Gene Duplication and Spectral Diversification of Cone Visual Pigments of Zebrafish. *Genetics.* 2003; 163:663–675. PMID: 12618404
20. Matsumoto Y, Fukamachi S, Mitani H, Kawamura S. Functional characterization of visual opsin repertoire in Medaka (*Oryzias latipes*). *Gene.* 2006; 371: 268–278. <https://doi.org/10.1016/j.gene.2005.12.005> PMID: 16460888
21. Hofmann CM, Carleton KL. Gene duplication and differential gene expression play an important role in the diversification of visual pigments in fish. *Integr Comp Biol.* 2009; 49:630–643. <https://doi.org/10.1093/icb/079> PMID: 21665846
22. Cortesi F, Musilová Z, Stieb SM, Hart NS, Siebeck UE, Malmström M, et al. Ancestral duplications and highly dynamic opsin gene evolution in percomorph fishes. *Proc Natl Acad Sci USA.* 2015; 112:1493–1498. <https://doi.org/10.1073/pnas.1417803112> PMID: 25548152
23. Kasagi S, Mizusawa K, Takahashi A. Green-shifting of SWS2A opsin sensitivity and loss of function of RH2-A opsin in flounders, genus *Verasper*. *Ecology and Evolution.* 2018; 8:1399–1410. <https://doi.org/10.1002/ece3.3745> PMID: 29375806
24. Govardovskii VI, Lychakov DV. [Some quantitative characteristics of the frog retinal rod outer segments]. *Tsitologija.* 1975; 27:524–529. PMID: 1202680
25. Chinen A, Matsumoto Y, Kawamura S. Reconstitution of Ancestral Green Visual Pigments of Zebrafish and Molecular Mechanism of Their Spectral Differentiation. *Mol Biol Evol.* 2005; 22:1001–1010. <https://doi.org/10.1093/molbev/msi086> PMID: 15647516
26. Cowing JA, Poopalasundaram S, Wilkie SE, Robinson PR, Bowmaker JK, Hunt DM. The molecular mechanism for the spectral shifts between vertebrate ultraviolet- and violet-sensitive cone visual pigments. *Biochem J.* 2002; 367:129–135. <https://doi.org/10.1042/BJ20020483> PMID: 12099889
27. Matsumoto Y, Hiramatsu C, Matsushita Y, Ozawa N, Ashino R, Nakata M, et al. Evolutionary renovation of L/M opsin polymorphism confers a fruit discrimination advantage to ateline New World monkeys. *Molecular Ecology.* 2014; 23:1799–1812. <https://doi.org/10.1111/mec.12703> PMID: 24612406
28. Davies WL, Cowing JA, Bowmaker JK, Carvalho LS, Gower DJ, Hunt DM. Shedding Light on Serpent Sight: The Visual Pigments of Henophidian Snakes. *J Neurosci.* 2009; 29:7519–7525. <https://doi.org/10.1523/JNEUROSCI.0517-09.2009> PMID: 19515920
29. Davies WL, Carvalho LS, Cowing JA, Beazley LD, Hunt DM, Arrese CA. Visual pigments of the platypus: A novel route to mammalian colour vision. *Current Biology.* 2007; 17:R161–R163. <https://doi.org/10.1016/j.cub.2007.01.037> PMID: 17339011
30. Davies WL, Carvalho LS, Tay B-H, Brenner S, Hunt DM, Venkatesh B. Into the blue: Gene duplication and loss underlie color vision adaptations in a deep-sea chimaera, the elephant shark *Callorhynchus milii*. *Genome Res.* 2009 [cited 16 Jul 2020]. <https://doi.org/10.1101/gr.084509.108> PMID: 19196633
31. Davies WL, Cowing JA, Carvalho LS, Potter IC, Trezise AEO, Hunt DM, et al. Functional characterization, tuning, and regulation of visual pigment gene expression in an anadromous lamprey. *The FASEB Journal.* 2007; 21:2713–2724. <https://doi.org/10.1096/fj.06-8057com> PMID: 17463225
32. Yokoyama S, Radlwimmer FB. The molecular genetics and evolution of red and green color vision in vertebrates. *Genetics.* 2001; 158:1697–1710. PMID: 11545071
33. Yokoyama S, Radlwimmer FB, Kawamura S. Regeneration of ultraviolet pigments of vertebrates. *FEBS Lett.* 1998; 423:155–158. [https://doi.org/10.1016/s0014-5793\(98\)00086-6](https://doi.org/10.1016/s0014-5793(98)00086-6) PMID: 9512349

34. Patel JS, Brown CJ, Ytreberg FM, Stenkamp DL. Predicting peak spectral sensitivities of vertebrate cone visual pigments using atomistic molecular simulations. *PLOS Computational Biology*. 2018; 14: e1005974. <https://doi.org/10.1371/journal.pcbi.1005974> PMID: 29364888
35. Rajput J, Rahbek DB, Andersen LH, Hirshfeld A, Sheves M, Altoè P, et al. Probing and Modeling the Absorption of Retinal Protein Chromophores in Vacuo. *Angewandte Chemie International Edition*. 2010; 49:1790–1793. <https://doi.org/10.1002/anie.200905061> PMID: 20104555
36. Sekharan S, Sugihara M, Buss V. Origin of Spectral Tuning in Rhodopsin—It Is Not the Binding Pocket. *Angewandte Chemie International Edition*. 2007; 46:269–271. <https://doi.org/10.1002/anie.200603306> PMID: 17120281
37. Kloppmann E, Becker T, Ullmann GM. Electrostatic potential at the retinal of three archaeal rhodopsins: Implications for their different absorption spectra. *Proteins: Structure, Function, and Bioinformatics*. 2005; 61:953–965. <https://doi.org/10.1002/prot.20744> PMID: 16247786
38. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, et al. Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor. *Science*. 2000; 289:739–745. <https://doi.org/10.1126/science.289.5480.739> PMID: 10926528
39. Kawamura S, Kasagi S, Kasai D, Tezuka A, Shoji A, Takahashi A, et al. Spectral sensitivity of guppy visual pigments reconstituted in vitro to resolve association of opsins with cone cell types. *Vision Research*. 2016; 127:67–73. <https://doi.org/10.1016/j.visres.2016.06.013> PMID: 27476645
40. Yokoyama S, Takenaka N, Blow N. A novel spectral tuning in the short wavelength-sensitive (SWS1 and SWS2) pigments of bluefin killifish (*Lucania goodei*). *Gene*. 2007; 396:196–202. <https://doi.org/10.1016/j.gene.2007.03.019> PMID: 17498892
41. Okada T, Sugihara M, Bondar A-N, Elstner M, Entel P, Buss V. The Retinal Conformation and its Environment in Rhodopsin in Light of a New 2.2Å Crystal Structure††This paper is dedicated to Dr Yoshimasa Kyogoku. *Journal of Molecular Biology*. 2004; 342:571–583. <https://doi.org/10.1016/j.jmb.2004.07.044> PMID: 15327956
42. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. *Nat Struct Mol Biol*. 2002; 9:646–652. <https://doi.org/10.1038/nsb0902-646> PMID: 12198485
43. Shichida Y, Matsuyama T. Evolution of opsins and phototransduction. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2009; 364:2881–2895. <https://doi.org/10.1098/rstb.2009.0051> PMID: 19720651
44. Collin SP, Hart NS, Shand J, Potter IC. Morphology and spectral absorption characteristics of retinal photoreceptors in the southern hemisphere lamprey (*Geotria australis*). *Vis Neurosci*. 2003; 20:119–130. <https://doi.org/10.1017/s0952523803202030> PMID: 12916734
45. Hárosi FI, Kleinschmidt J. Visual pigments in the sea lamprey, *Petromyzon marinus*. *Visual Neuroscience*. 1993; 10:711–715. <https://doi.org/10.1017/s0952523800005411> PMID: 8338808
46. Toyama M, Hironaka M, Yamahama Y, Horiguchi H, Tsukada O, Uto N, et al. Presence of rhodopsin and porphyropsin in the eyes of 164 fishes, representing marine, diadromous, coastal and freshwater species—a qualitative and comparative study. *Photochem Photobiol*. 2008; 84:996–1002. <https://doi.org/10.1111/j.1751-1097.2008.00344.x> PMID: 18422881
47. Morphology, characterization, and distribution of retinal photoreceptors in the Australian lungfish *Neoceratodus forsteri* (Krefft, 1870)—Bailes—2006—*Journal of Comparative Neurology*—Wiley Online Library. [cited 19 Jul 2020]. Available: <https://onlinelibrary.wiley.com/doi/full/10.1002/cne.20809>
48. Bailes HJ, Davies WL, Trezise AE, Collin SP. Visual pigments in a living fossil, the Australian lungfish *Neoceratodus forsteri*. *BMC Evolutionary Biology*. 2007; 7:200. <https://doi.org/10.1186/1471-2148-7-200> PMID: 17961206
49. Provencio I, Loew ER, Foster RG. Vitamin A2-based visual pigments in fully terrestrial vertebrates. *Vision Res*. 1992; 32:2201–2208. [https://doi.org/10.1016/0042-6989\(92\)90084-v](https://doi.org/10.1016/0042-6989(92)90084-v) PMID: 1287997
50. Loew ER, Fleishman LJ, Foster RG, Provencio I. Visual pigments and oil droplets in diurnal lizards: a comparative study of Caribbean anoles. *Journal of Experimental Biology*. 2002; 205:927–938. PMID: 11916989
51. Katti C, Stacey-Solis M, Coronel-Rojas NA, Davies WIL. The Diversity and Adaptive Evolution of Visual Photopigments in Reptiles. *Front Ecol Evol*. 2019; 7. <https://doi.org/10.3389/fevo.2019.00352>
52. Loew ER, Dartnall HJA. Vitamin A1/A2-based visual pigment mixtures in cones of the rudd. *Vision Research*. 1976; 16:891–896. [https://doi.org/10.1016/0042-6989\(76\)90217-0](https://doi.org/10.1016/0042-6989(76)90217-0) PMID: 948877
53. Harosi FI. Ultraviolet- and violet-absorbing vertebrate visual pigments: dichroic and bleaching properties. *The visual system*. 1985:41–55
54. Rajamani R, Lin YL, Gao J. The opsin shift and mechanism of spectral tuning in rhodopsin. *J Comput Chem*. 2011; 32(5):854–865. <https://doi.org/10.1002/jcc.21663> PMID: 20941732

55. Huerta-Cepas J, Serra F, Bork P. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol Biol Evol.* 2016; 33:1635–1638. <https://doi.org/10.1093/molbev/msw046> PMID: 26921390
56. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst Biol.* 2010; 59:307–321. <https://doi.org/10.1093/sysbio/syq010> PMID: 20525638
57. Jacobson MP, Pincus DL, Rapp CS, Day TJJ, Honig B, Shaw DE, et al. A hierarchical approach to all-atom protein loop prediction. *Proteins: Structure, Function, and Bioinformatics.* 2004; 55:351–367. <https://doi.org/10.1002/prot.10613> PMID: 15048827
58. Jacobson MP, Friesner RA, Xiang Z, Honig B. On the Role of the Crystal Environment in Determining Protein Side-chain Conformations. *Journal of Molecular Biology.* 2002; 320:597–608. [https://doi.org/10.1016/s0022-2836\(02\)00470-9](https://doi.org/10.1016/s0022-2836(02)00470-9) PMID: 12096912
59. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics.* 1981; 52:7182–7190.
60. Evans DJ, Holian BL. The Nose–Hoover thermostat. *The Journal of Chemical Physics.* 1998; 83:4069. <https://doi.org/10.1063/1.449071>
61. Berendsen H, Van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark A. GROMACS: fast, flexible, and free. *J Comput Chem.* 2005; 26:1701–1718. <https://doi.org/10.1002/jcc.20291> PMID: 16211538
62. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics.* 1996; 14:33–38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5) PMID: 8744570
63. Schwarz G. Estimating the Dimension of a Model. *Ann Statist.* 1978; 6:461–464. <https://doi.org/10.1214/aos/1176344136>