THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Article

Computational Reconstruction and Analysis of Structural Models of Avian Cryptochrome 4

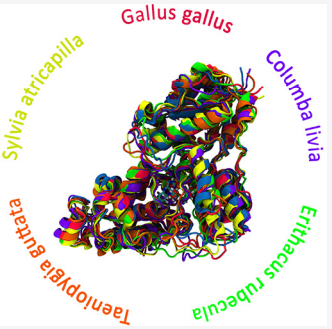
Maja Hanić,* Fabian Schuhmann, Anders Frederiksen, Corinna Langebrake, Georg Manthey, Miriam Liedvogel, Jingjing Xu, Henrik Mouritsen, and Ilia A. Solov'yov*

Cite This: *J. Phys. Chem. B* 2022, 126, 4623−4635

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: A recent study by Xu et al. (*Nature*, **2021**, 594, 535−540) provided strong evidence that cryptochrome 4 (Cry4) is a key protein to endow migratory birds with the magnetic compass sense. The investigation compared the magnetic field response of Cry4 from migratory and nonmigratory bird species and suggested that a difference in magnetic sensitivity could exist. This finding prompted an in-depth investigation into Cry4 protein differences on the structural and dynamic levels. In the present study, the pigeon Cry4 (*Cl*Cry4) crystal structure was used to reconstruct the missing avian Cry4 protein structures via homology modeling for carefully selected bird species. The reconstructed Cry4 structure from European robin, Eurasian blackcap, zebra finch, chicken, and pigeon were subsequently simulated dynamically and analyzed. The studied avian Cry4 structures show flexibility in analogous regions pointing to similar activation mechanisms and/or signaling interaction partners. It can be concluded that the experimentally recorded difference in the magnetic field sensitivity of Cry4 from different birds is unlikely to be due to solely intrinsic dynamics of the proteins but requires additional factors that have not yet been identified.



INTRODUCTION

It has been known for decades that migratory birds use a magnetic compass to orient during their seasonal long-distance journeys.[1,2] Although much progress has been made recently,[1,3,4] it is still unclear how birds sense the geomagnetic field. Observations provide some hints that several bird species seem to have difficulties orienting under red or yellow light[5,6] or when exposed to radiofrequency fields in the MHz range (above 100 kHz).[7−11]

One exciting hypothesis proposed back in 1978 by Schulten *et al.*[12] on how the magnetic compass sense could work relies on quantum mechanics and suggests that a pair of entangled electrons may enable birds to detect the Earth's magnetic field. This hypothesis suggests that specific nonequilibrium states of molecules, called radicals, that have an uneven number of electrons can create a radical-pair. The unpaired electron spins in the radical-pair may be aligned either parallel or antiparallel, and these states are known as triplet and singlet states, respectively. If the radical-pair is brought far outside of thermal equilibrium by light-absorption and exposed to a static magnetic field (e.g., such as the Earth's magnetic field), then the magnetic field can influence the rate of singlet−triplet interconversion, which can subsequently influence the radical-pair reaction outcome.[3,13,14] Therefore, different chemical outcomes could correlate with a change in a geomagnetic field direction that could be neuronally integrated. One essential prerequisite for this hypothesis is that the animal must possess a specific molecular receptor that fulfills the necessary physical and chemical properties to be able to create and host a biologically relevant radical-pair.[3,14−20]

Recently, it has been demonstrated that the cryptochrome 4 (Cry4) protein, a photoreceptor molecule ubiquitously expressed in the bird's body including the eyes, i.e., the assumed sensory organ for light-dependent magnetoreception,[18,19] may be exceptionally well-suited and equipped with the necessary physical and chemical properties to act as a magnetic compass sensor.[4] Figure 1 illustrates the structure of the Cry4 protein from European robin (*Erithacus rubecula*, *Er*Cry4), a migratory bird, and marks a region inside the protein where a light-absorptive cofactor, called flavin adenine dinucleotide (FAD), crucial for Cry4 functioning, is located.[14,16−23]

A pivotal event in Cry4 occurs upon its exposure to blue light, where an electron in the FAD is excited, creating the FAD* state and thus allowing for the creation of spin-correlated radical-pairs through the process of photoexcitation of the electron acceptor molecule.[12] Inside Cry4, the excitation process is followed by a rapid electron-transfer from the neighboring tryptophan residue Trp395 (W_A) to the FAD. A photoinduced electron-transfer involving flavin and tryptophan was studied in a *Arabidopsis thaliana* cryptochrome 1 and in

Received: February 5, 2022
Revised: April 28, 2022
Published: June 15, 2022

https://doi.org/10.1021/acs.jpcb.2c00878
J. Phys. Chem. B 2022, 126, 4623−4635

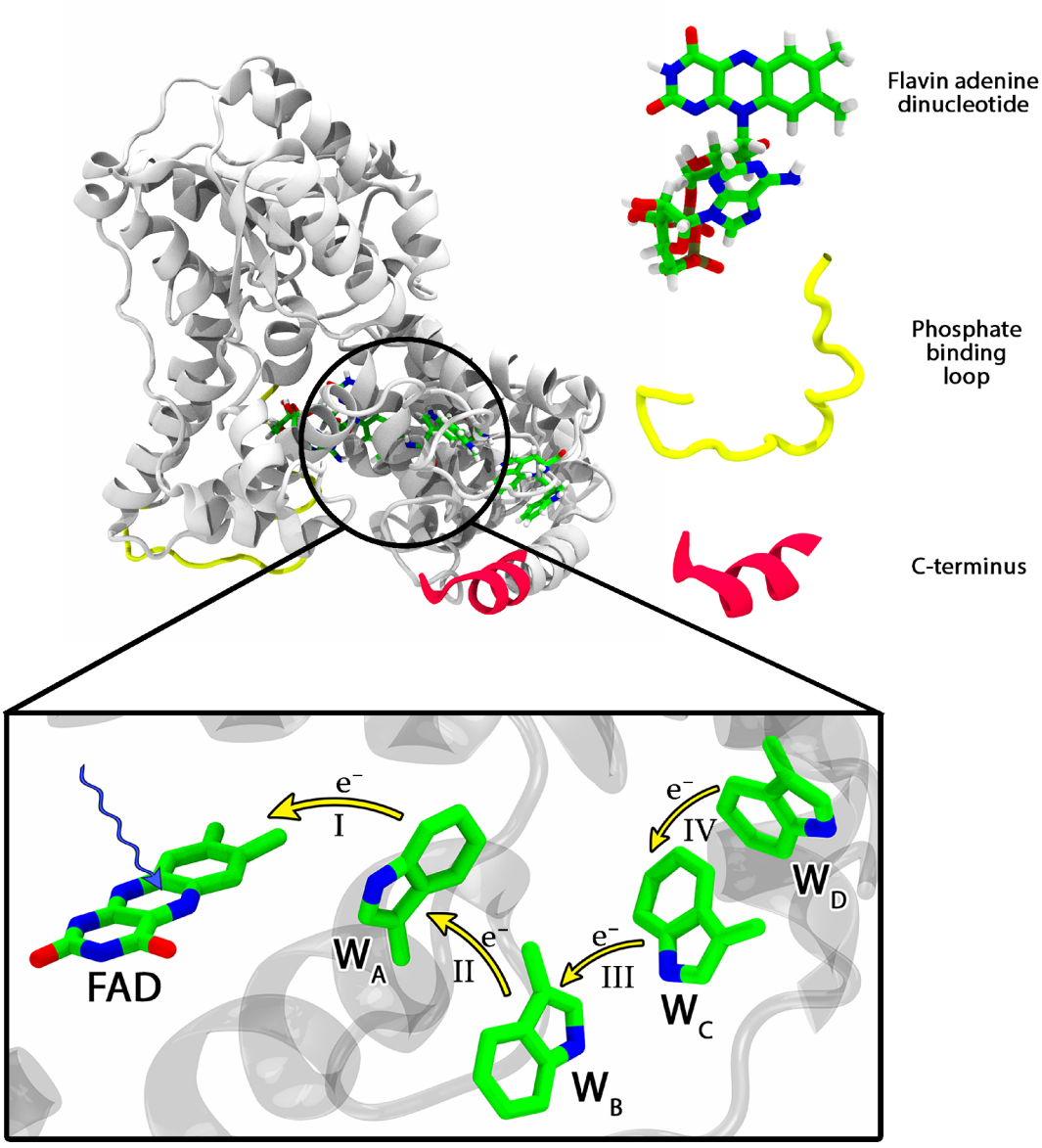


Figure 1. Secondary structure representation of *Er*Cry4 protein highlighting the specific parts of the protein important for functioning as the magnetoreceptive molecule. Upper panel to the right: flavin adenine dinucleotide (FAD), phosphate-binding loop (residues 231−248), and a partial C-terminus (residues 486−496). The circle indicates the approximate position of the FAD binding pocket inside the cryptochrome. The lower part highlights the electron-transfer process occurring between four tryptophans (W_A, W_B, W_C, and W_D) and FAD (only the flavin part of the FAD molecule is shown). The electron-transfer process is initiated upon a blue light excitation of the flavin moiety (shown with a blue arrow).

Escherichia coli CPD photolyase. These studies have demonstrated that a singlet radical-pair is formed upon the ^{1}FAD excitation.[24] The process of a singlet radical-pair formation was shown to have a higher quantum yield compared to other processes, i.e., the intersystem crossing.

The first electron-transfer in Cry4 is followed by subsequent sequential electron-transfers: from W_B (Trp372) to $W_A^{\bullet+}$, from W_C (Trp318) to $W_B^{\bullet+}$, and last from W_D (Trp369) to $W_C^{\bullet+}$ as illustrated in Figure 1. As a result, the [$FAD^{\bullet-}Trp_DH^{\bullet+}$] radical-pair in Cry4 is created, which has been shown to be sensitive to the magnetic field.[4]

Magnetic compass orientation is particularly important for migratory birds, but even some birds that show a sedentary lifestyle seem to be able to sense the magnetic field.[2] Although both migratory and at least some nonmigratory birds seem to be able to sense the magnetic field, they likely differ in their efficiency to do so. This hypothesis was recently supported *in vitro*.[4]

In a recent study by Xu *et al.*,[4] the authors measured changes in the optical absorbance of photoinduced radicals in Cry4 proteins *in vitro* from migratory and nonmigratory bird species exposed to a 30 mT magnetic field. The results suggest that the magnetic sensitivity in Cry4 from a night migratory songbird is enhanced compared to Cry4 from nonmigratory bird species.[4] However, the molecular basis for this apparent difference in magnetic sensitivity between migratory and nonmigratory bird Cry4 is unknown. This calls for a detailed investigation and in-depth comparison of protein structures between migratory and nonmigratory birds' cryptochromes.

Presuming that Cry4 is indeed the key receptor that detects the input from the magnetic field in the eye, the question then is how would that signal be converted to a biological signal inside the cell.[25,26] Such signal transduction is usually

https://doi.org/10.1021/acs.jpcb.2c00878
J. Phys. Chem. B 2022, 126, 4623−4635

connected to the protein's dynamics, which can be investigated using molecular dynamics simulation.

Here the aim is to elucidate possible differences in Cry4 structures between Cry4s from migratory and resident species using computational methods of protein structural reconstruction and molecular dynamics simulation. Both methods could be important for understanding the seemingly different magnetic field sensitivities between Cry4 proteins.

Specifically, the structure and dynamics of Cry4 proteins from five different bird species was investigated: migratory Eurasian blackcap (*Sylvia atricapilla*, *Sa*Cry4) and migratory European robin, *Er*Cry4; nonmigratory zebra finch (*Taeniopygia guttata*, *Tg*Cry4), nonmigratory chicken (*Gallus gallus*, *Gg*Cry4), and nonmigratory pigeon (*Columba livia*, *Cl*Cry4).

METHODS

Protein Sequences of Cry4. To obtain the coding sequences of Cry4 from the five different species, the reference genomes published in the context of the Bird 10,000 Genomes (B10K) project have been used.[27] The Cry4 sequence of *Erithacus rubecula* was used as a query and blasted against the reference genomes of focally selected bird species with default parameters.[28] These reference sequences were compared to individual Sanger sequenced Cry4 isolates[29] to identify sequencing errors and evaluate the reliability of variable sites. For *Sa*Cry4 and *Tg*Cry4, variable sites were checked with individual publicly available whole-genome resequencing data.[30] The most common gene version (allele) in the population was chosen for the final consensus sequence for the identified variable sites. In those cases where no resequencing data of several individuals was available, the conflicting sites (different between the reference genome and Sanger sequenced isolate) were compared to reference genomes of closely related species available through the B10K project.[27] In the case of conflicting sequences at a certain site, the sequence agreeing best with the Cry4 sequences of the closest relatives at that site in the final Cry4 sequence was used; the final sequences as well as their accession numbers are listed in the Supporting Information, see Figures S6 and S7.

Homology Modeling. Currently, the only available crystal structure of an avian Cry4 is *Cl*Cry4[22] (PDB ID: 6PU0), which was successfully used earlier to construct a homology model of *Er*Cry4.[4] In the present study, *Cl*Cry4 was also used as a template to reconstruct the homology models of the other avian cryptochromes, specifically *Tg*Cry4, *Gg*Cry4, and *Sa*Cry4.

Amino acid sequences from the species-specific Cry4s were compared with the reference amino acid sequence of *Cl*Cry4 to establish sequence similarity and sequence identity by utilizing the global alignment method;[31] the results are summarized in Table 1.[32]

Table 1. Summary of Cryptochrome Sequence Alignment with the *Cl*Cry4 Sequence[a]

	*Cl*Cry4	*Er*Cry4	*Gg*Cry4	*Sa*Cry4	*Tg*Cry4
sequence similarity	100%	93.5%	91.7%	92.6%	92.8%
sequence identity	100%	89.7%	85.1%	86.9%	87.3%
gaps	0%	0.4%	0.8%	0.4%	0.4%

[a]The structure of *Cl*Cry4 is resolved through crystallography.[22] A Needelman–Wunsch algorithm[31] was used to compare the protein sequences.

The corresponding atomistic structures of the three avian Cry4s (*Sa*Cry4, *Tg*Cry4, and *Gg*Cry4) were created based on the amino acid sequences using the Swiss-model server for homology modeling.[33,34] The original reports from the Swiss-model server and the quality metrics of the resulting Cry4 structures, as well as the accession codes of the protein sequences are provided in the Supporting Information (see Figures S9–S11). As the crystallized template structure of *Cl*Cry4[22] lacks the C-terminal, all of the structures generated from this template are also lacking this C-terminal end (further details are provided in the Supporting Information).

The algorithmic details of the computational procedure used to generate homology models of Cry4 structures are outlined and described in the workflow diagram in Figure 2. After the protein structures were constructed, they were prepared for atomistic molecular dynamics (MD) simulations. This preparation included determining the protonation states of

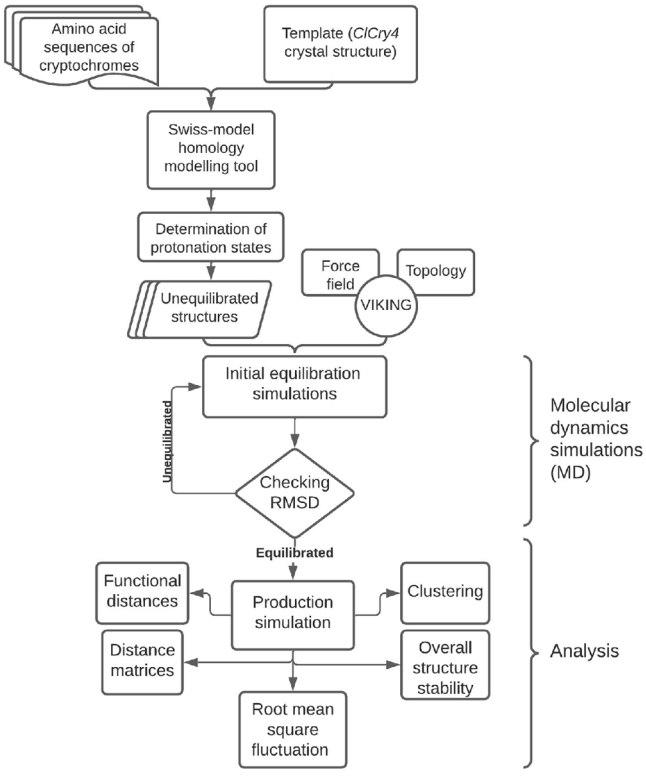


Figure 2. Work-flow diagram summarizing the key steps for studying structure and dynamics of different avian Cry4s. Cry4 sequences[27] were used to reconstruct the corresponding three-dimensional structures with the *Cl*Cry4 structure being used as a molecular template. Next, the Swiss-model homology modeling tool[33,34] was used to obtain the three-dimensional structures of avian cryptochromes. The reconstructed structures were then examined, and the protonation states of different amino acids were determined. This process yielded unequilibrated structures, which were then extensively simulated with VIKING.[35] A quantitative measure called root-mean-square deviation (RMSD) was applied to estimate protein stability. A common method includes comparing the protein conformations throughout the simulation relative to the initial conformation of the MD simulation. RMSD is a quantitative measure that can display the difference between a certain structure and a reference one. Simulations were carried out until the RMSD values of the protein backbone atoms reached some stable equilibrium value. The obtained stable structures were then extensively simulated in several production simulations, which were analyzed.

https://doi.org/10.1021/acs.jpcb.2c00878
J. Phys. Chem. B 2022, 126, 4623−4635

Table 2. Investigated Species for Which the Cry4 Structures Were Studied, Listing the Corresponding Common Name, Protein Name Abbreviation, Duration of MD Trajectories Used for Analysis, Number of Simulation Replicas, and the Source of the Structures[a]

Latin name	common name	protein name	production simulation (ns)	replica	ref	number of atoms
Taeniopygia guttata	Zebra finch	*Tg*Cry4	250	3	this work	91 039
Gallus gallus	Chicken	*Gg*Cry4	250	3	this work	87 927
Columbus livia	Pigeon	*Cl*Cry4	250	3	51	121 097
Sylvia atricapilla	Eurasian blackcap	*Sa*Cry4	250	3	this work	90 263
Erithacus rubecula	European robin	*Er*Cry4	250	3	4	100 518

[a]Simulations of *Cl*Cry4 and *Er*Cry4 were adopted from earlier studies.[4,51]

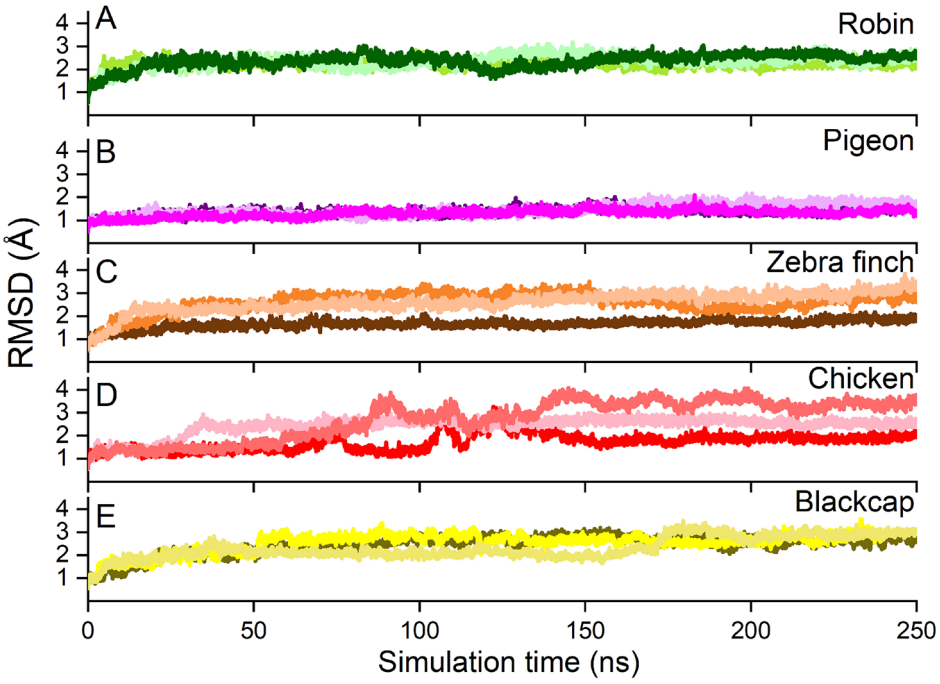


Figure 3. Root-mean-square-deviation (RMSD) of Cry4 backbone atoms, plotted against simulation times in reference to the postequilibration structures. RMSD values shown here indicate that all of the simulated Cry4 models are stable and can be used for analysis. Different color shades indicate results for different MD replica simulations.

the histidine residues, which depends on the surrounding amino acid residues in the protein. With the protonation states determined, the structures were simulated dynamically using the VIKING platform.[35]

Comparing Homology Modeled and Crystal Structure *Cl*Cry4. To allow for direct comparison of the crystallized and homology modeled structure predictions, a diagnostic validation step was added: a homology modeled *Er*Cry4 structure was used as a template for a *Cl*Cry4 amino acid sequence to investigate how well the reversely reconstructed homolology modeled *Cl*Cry4 structure agrees with the crystal structure of *Cl*Cry4.

Molecular Dynamics Simulations. MD simulation is a well-known computational approach to validate protein structure stability and investigate the underlying dynamics.[16,36,37] MD simulations for all avian Cry4 protein structures were performed using the NAMD[38−40] software utilizing the VIKING platform.[35] Three independent replica simulations were carried out for each of the studied protein structures in order to check the reproducibility of the results (see Table 2). The CHARMM36 force field was used for all simulations,[41−47] including CMAP corrections for proteins.[48] All simulations were performed in the dark state of Cry4, and FAD was simulated as fully oxidized using the previously reported parametrization.[4,17,49,50] The technical details regarding the details of MD simulation protocol are given in the Supporting Information.

Clustering Analysis. Clustering analysis was performed with scikit-learn library v1.0.2 in Python 3.8.10.[52] The so-called agglomerative clustering algorithm was employed, which is a type of hierarchical clustering where a hierarchy of clusters is built using the bottom-up approach.[53] For protein structure clustering, the algorithm assumes that each protein structure in a studied data set belongs to an individual cluster, which are then eventually merged into pairs of clusters as the cutoff parameter changes. For each studied Cry4 structure, every 10th MD frame of the concatenated replica simulations was taken into account for clustering. A distance matrix of the root-mean-square deviation (RMSD) between frames of the simulation was calculated employing the rmsd 1.4 module.[54−56] The resulting 1490 × 1490 matrix was used as an input for the clustering method, so that RMSD was used as a distance measure for clustering. The complete linkage option was used for the clustering analysis, meaning that in every clustering step the two clusters with the smallest maximum distance (i.e., ΔRMSD) between structures of the clusters were

https://doi.org/10.1021/acs.jpcb.2c00878
J. Phys. Chem. B 2022, 126, 4623−4635

Table 3. Comparison of the RMSD Values Calculated for Different Cryptochromes Averaged over Last 10 ns of Production Simulation[a]

		ErCry4			ClCry4				TgCry4			GgCry4			SaCry4		
		1	2	3	1	2	3	3*	1	2	3	1	2	3	1	2	3
ErCry4	1	0.000	2.865	3.241	2.906	3.784	2.752	2.602	3.180	3.312	4.371	3.394	3.270	3.856	3.013	3.196	2.989
	2	2.865	0.000	2.617	3.146	3.664	3.233	2.911	3.377	3.694	4.619	3.852	3.762	4.375	3.530	3.654	3.201
	3	3.241	2.617	0.000	3.332	4.026	3.312	3.589	3.725	4.181	5.273	4.518	4.649	5.060	4.297	4.319	3.804
ClCry4	1	2.906	3.146	3.332	0.000	2.725	1.372	3.284	2.597	2.939	3.724	2.716	3.157	3.683	2.958	3.220	3.259
	2	3.784	3.664	4.026	2.725	0.000	2.922	4.186	2.562	2.797	3.309	2.814	3.073	3.055	2.933	3.276	2.750
	3	2.752	3.233	3.312	1.372	2.922	0.000	3.333	2.664	2.889	3.872	2.839	3.325	3.876	3.129	3.267	3.338
	3*	2.602	2.911	3.589	3.284	4.186	3.333	0.000	3.586	3.491	4.065	3.843	3.264	3.997	3.146	3.311	3.448
TgCry4	1	3.180	3.377	3.725	2.597	2.562	2.664	3.586	0.000	2.582	3.347	2.548	3.008	3.205	2.723	3.051	2.797
	2	3.312	3.694	4.181	2.939	2.797	2.889	3.491	2.582	0.000	2.497	2.187	2.209	2.899	2.357	2.432	2.946
	3	4.371	4.619	5.273	3.724	3.309	3.872	4.065	3.347	2.497	0.000	2.732	2.564	3.415	2.778	2.828	3.730
GgCry4	1	3.394	3.852	4.518	2.716	2.814	2.839	3.843	2.548	2.187	2.732	0.000	2.102	3.025	2.439	2.134	3.277
	2	3.270	3.762	4.649	3.157	3.073	3.325	3.264	3.008	2.209	2.564	2.102	0.000	2.528	1.905	2.221	2.892
	3	3.856	4.375	5.060	3.683	3.055	3.876	3.997	3.205	2.899	3.415	3.025	2.528	0.000	2.817	3.326	2.590
SaCry4	1	3.013	3.530	4.297	2.958	2.933	3.129	3.146	2.723	2.357	2.778	2.439	1.905	2.817	0.000	2.603	2.560
	2	3.196	3.654	4.319	3.220	3.276	3.267	3.311	3.051	2.432	2.828	2.134	2.221	3.326	2.603	0.000	3.153
	3	2.989	3.201	3.804	3.259	2.750	3.338	3.448	2.797	2.946	3.730	3.277	2.892	2.590	2.560	3.153	0.000

[a]The numbers 1, 2, and 3 in the column and row headings indicate replica simulations. 3* simulation for the *Cl*Cry4 structure represents the homology model simulation of the pigeon cryptochrome. The green color indicates high structural similarity of the compared structures, and going to red indicates higher RMSD and therefore lower structural similarity between two structures.

merged. The number of clusters was calculated for different ΔRMSD threshold values with an increasing ΔRMSD step size of 0.005 Å, until all considered Cry4 structures happen to belong to a single cluster. The resulting dendrograms were plotted with the Matplotlib library.[57]

RESULTS AND DISCUSSION

This study delivers a unified computational approach for constructing and characterizing cryptochromes from migratory and nonmigratory avian species. Herein, the reconstructed atomistic structures of five avian Cry4 models are presented, and tools to justify the robustness of their structures are provided. By determining possible intraprotein rearrangements and carrying out an extended analysis of inter-residue distances important for protein function, in this work, the differences and similarities in protein structure and dynamics between migratory and nonmigratory birds' Cry4 models are studied.

Protein Stability. The stability of a protein can be probed through analysis of the temporal structure evolution. A common method includes comparing the protein's conformation throughout the simulation to some reference structure, e.g., the initial conformation of the protein prior to the MD production simulation. This comparison can be achieved by analyzing the secondary structure's RMSD. More details about the RMSD calculations are provided in the Supporting Information.

The analyses in Figure 3 show that all of the modeled Cry4 protein structures are well-equilibrated and can be considered stable. The indication of stability can be seen from the time evolution dependency of the RMSD plots that flatten out after a certain simulation time. Somewhat expected are the lower RMSD values noticed in the case of *Cl*Cry4 (Figure 3B, shown in purple), which is the only crystal structure that was simulated. Some deviations are noticeable in the replica simulations of *Gg*Cry4 (Figure 3D, shown in red). However, because the second half of the *Gg*Cry4 simulation does not feature such deviations, the structures in the three replica simulations are still considered to be sufficiently stable. Additionally, Figure S1 shows an RMSD analysis for the different Cry4 simulations, performed without the inclusion of the flexible phosphate-binding loop (residues 231−248). The plots in Figure S1 indicate overall lower values of the RMSD once the flexible motif is excluded. It is important to note that all of the modeled Cry4 structures are largely missing the C-terminal part as it is absent in the *Cl*Cry4 template and therefore cannot be reconstructed reliably (Figure S7 shows what parts of the proteins were reconstructed and which were omitted). It is, therefore, possible that some perturbation of the RMSD values could arise because of the truncated protein structures.

Even though tracking the time evolution of the RMSD can indicate the stability of the protein structures, a comparison of the RMSD value between two protein structures close to the end of the simulation could indicate how varied are the different Cry4 structures that have been simulated dynamically. Table 3 shows an averaged RMSD over the last 10 ns of the simulation. The table suggests that the similarity of the averaged structures is higher when it comes to replica simulations and lower when it comes to different Cry4 species. Furthermore, in the case of *Gg*Cry4, three independent replica simulations yield structures that are reasonably similar to each other, even though Figure 3 may suggest that the final structures differ.

Clustering Analysis. To complement the RMSD analysis of the constructed Cry4 structures and additionally compare their similarity, an agglomerative clustering algorithm was employed. Figure 4 indicates the rapidity of Cry4 structure clustering. The analysis was performed depending on the ΔRMSD threshold parameter. Figure 4 reveals that all of the studied Cry4 structures cluster into a single cluster when ΔRMSD exceeds 4.25 Å. Furthermore, Figure 4 shows that the set of *Cl*Cry4 structures merge into a single cluster at a somewhat lower value of ΔRMSD. This likely happens because *Cl*Cry4 is the only studied crystal structure, while other Cry4s are homology modeled. Nevertheless an RMSD difference of around 1 Å is considered to be small for a protein the size of Cry4.[58] The rate of clustering is the lowest for *Er*Cry4 clusters, meaning that within the *Er*Cry4 ensemble there is a slightly lower degree of similarity between the individual structures taken as MD snapshots; the latter observation goes along with the results shown in Table 3. In contrast, *Cl*Cry4 structures cluster faster, meaning there is a higher degree of similarity

https://doi.org/10.1021/acs.jpcb.2c00878
J. Phys. Chem. B 2022, 126, 4623−4635

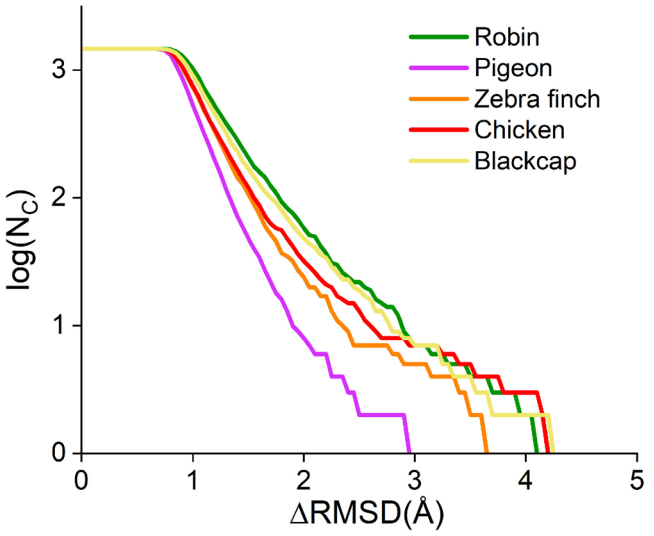


Figure 4. Agglomerative clustering analysis dependence on ΔRMSD performed for the avian Cry4 simulations. N_C indicates the total number of obtained clusters where the considered Cry4 structures are grouped. The three replica simulations for each Cry4 structure were used jointly to provide a unified set of structures analyzed here.

between the structures within that particular data set. It is, however, important to stress that because all the studied Cry4 structures clearly become grouped in a single cluster within ΔRMSD of 4.25 Å one can conclude that the three independent replica simulations yield similar structures that differ in minor details that are likely not critical for the overall structural motifs. The observed clustering trend is also supported by Figure S2, which shows dendrograms of avian Cry4 structures. Figure S2 indicates that clustering for different Cry4 structures happens similarly and the resulting cluster sizes appear to be comparable, revealing that the distribution of different Cry4 structures is uniformly converging.

Protein Rearrangement. Characterization of protein dynamics often provides an important way of defining a protein's function, including its putative signaling role inside a cell. Protein dynamics analysis could reveal certain protein regions which can be essential for protein function and reveal differences between migratory and nonmigratory Cry4 structures, as was also indicated in a recent study by Wang *et al.*,[59] where it was shown that the *Dm*Cry activation mechanism includes an allosteric switch. In order to assess how specific amino acid residues contribute to the overall flexibility of the protein, a root-mean-square fluctuations (RMSF) analysis was performed. RMSF represents the measure for a residue's movement in the protein's structure around its reference position (see the Supporting Information for more details).

Figure 5C shows the result of the RMSF analyses performed for the three 250 ns replica simulations for all studied Cry4s structures. In general, the following regions of interest show a moderate increase in Cry4 flexibility: residues 150−250, residues 400−497, and more specifically, the phosphate-binding loop (231−248) and the C-terminal region (485−495). Increased flexibility of the residues beyond residue 400 could be related to the hyper-flexible C-terminal region. Interestingly, although the C-terminal region is moderately flexible in all of the probed avian cryptochromes, not all demonstrate a hyper-flexible C-terminal behavior when compared between species. Furthermore, the C-terminal in avian cryptochrome could not be accurately modeled in full length because in the *Cl*Cry4 crystal structure,[22] this region was too flexible to be successfully reconstructed and is therefore unavailable for homology modeled structures; Figure S7 in the Supporting Information indicates the missing parts of the C-terminal in the modeled Cry4 structures.

A region of lower flexibility, as seen both from Figure 5B and Table 4, is the so-called tryptophan tetrad region (TTR) which contains all four tryptophans W395 (W_A), W372 (W_B), W318 (W_C), and W369 (W_D), which are essential for the electron-transfer process in Cry4. This region is one of the lowest in flexibility in all of the studied Cry4 models, coinciding well with the proposed mechanism of radical-pair creation, which is highly dependent on the inter-residue distance of the mentioned tryptophan residues.[3] The results in Figure 5C show that apparently, migratory and nonmigratory Cry4s do not exhibit significant differences in their internal dynamics. Even though Figure 5B and Table 4 show that the averaged RMSF values per region can vary between species, the results still indicate that all of the studied Cry4s may share a similar activation mechanism.

A further investigation, in which different radical-pair states of Cry4 from different species would be simulated and analyzed, could show additional differences. However, such a study should be performed separately. Furthermore, it cannot be excluded that the observed similar results might be slightly biased, because all homology modeled structures were established from a single crystal structure of *Cl*Cry.[22] Finally, it is important to note that the most flexible region (the C-terminal (residues 485−495)), in Figure 5C is, in fact, the least preserved region not only among the avian cryptochromes discussed in this paper but also among cryptochromes across the animal kingdom.[61]

Figure 5C indicates that Cry4 from different bird species experiences similar motions. This observation is reflected in both the magnitude of fluctuations and the specific mobile residues. These findings support conserved interaction patterns of Crys from different species as previously reported[62] where it is discussed how interactions with signaling molecules in different regions of cryptochrome could affect its interactions with other proteins. It was concluded that the interaction patterns of cryptochrome from different species could be conserved.

To connect the flexible region in the structures with the uncertainties in the atomistic coordinates, a comparison of RMSF values with the B-values of the *Cl*Cry4 was performed. The results indicate that simulations correlate well with the uncertainties in atomistic coordinates, characterized through the B-values provided for *Cl*Cry4 crystal structure. The details of this comparison are provided in the Supporting Information and Figure S3 in particular.

Specific Distances Related to Function. To elucidate if the creation of radical-pairs could occur efficiently in the modeled cryptochromes, edge-to-edge distances between the FAD cofactor and the W_A, W_B, W_C, and W_D residues as well as the pairwise distances W_A−W_B, W_B−W_C, W_C−W_D, and W_D−Y_{319} were computed. Edge-to-edge distances are important for electron-transfer processes, as shown in Figure 1, that lead to Cry4 activation. With the increase of the inter-residue distance, the electron-transfer could become less efficient and slow down;[3] it is, therefore, important to estimate whether the distances involving FAD and the tryptophan tetrad are

https://doi.org/10.1021/acs.jpcb.2c00878
J. Phys. Chem. B 2022, 126, 4623−4635

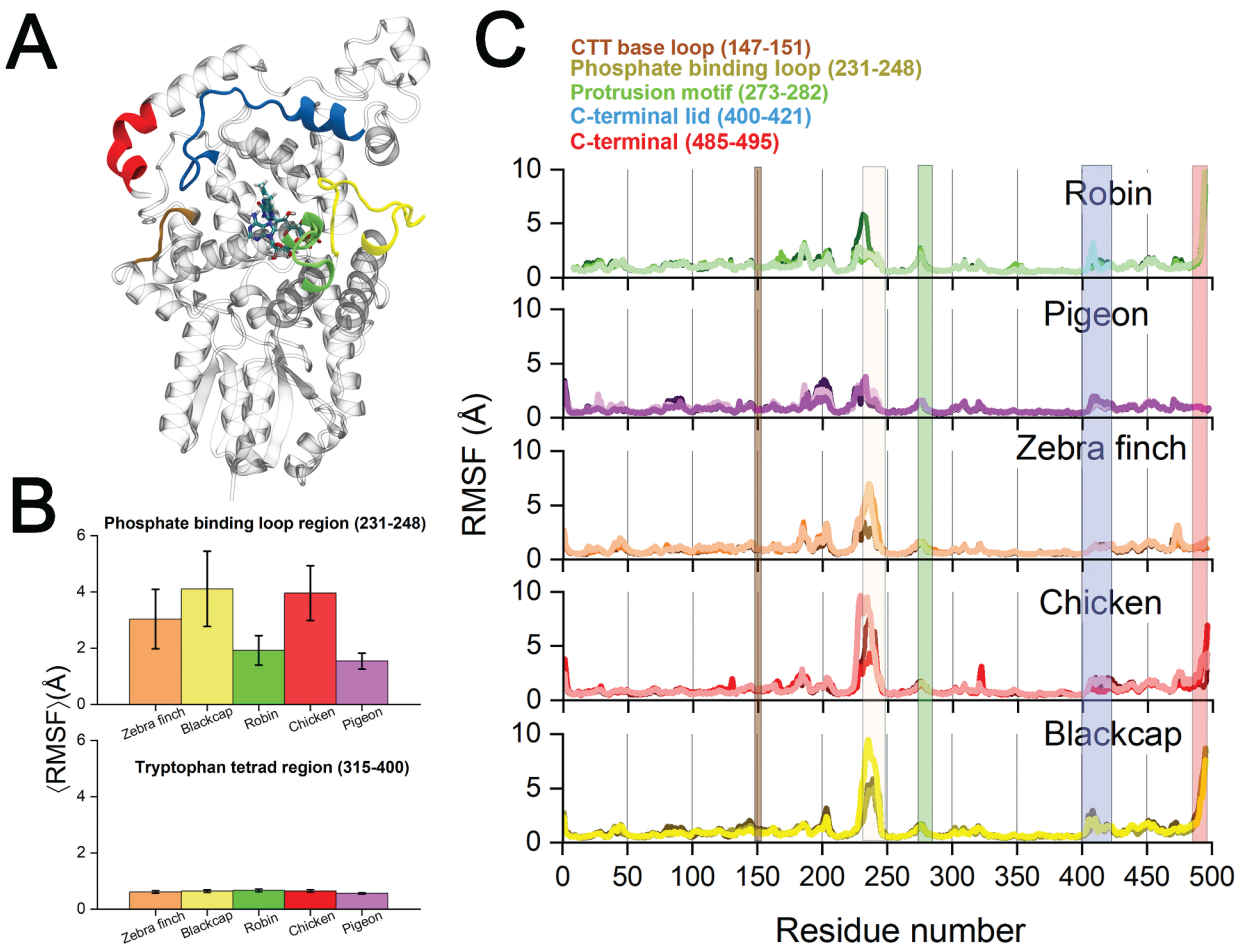


Figure 5. (A) Structure of the *Tg*Cry4 protein with specific regions color-coded (according to the nomenclature inherited from *Drosophila melanogaster*[37,60]): C-terminal shown in red (residues 485−495), C-terminal lid shown in blue (residues 400−421), protrusion motif shown in green (residues 273−282), phosphate-binding loop shown in yellow (residues 231−248), and C-terminal tail (CTT) base loop shown in brown (residues 147−151). (B) Averaged root-mean-square fluctuations (RMSF) values computed for different regions of the five standard Cry4 models for the phosphate-binding loop (top) region and tryptophan tetrad region (bottom). The values are averaged over the results from the three replica MD simulations of *Cl*Cry4. The error bars indicate the standard deviation corresponding to the values shown in Table 4. (C) RMSF values of amino acid residues in the modeled cryptochromes. The fluctuations represent movement of the C_α backbone atoms, around their averaged position. Different color shades indicate different replicas of the MD simulations.

Table 4. Average RMSF Values for Phosphate-Binding Loop (PBL) and Tryptophan Tetrad Region (TTR) Regions of the Proteins[a]

	⟨RMSF⟩ (Å)				
	Zebra finch	Eurasian blackcap	European robin	Chicken	Pigeon
PBL (231−248)	3.03 ± 1.06	4.11 ± 1.34	1.92 ± 0.52	3.96 ± 0.98	1.54 ± 0.28
TTR (315−400)	0.61 ± 0.05	0.64 ± 0.05	0.66 ± 0.05	0.65 ± 0.05	0.56 ± 0.03

[a]The uncertainties correspond standard deviation because the values have been averaged over three simulation replicas.

significantly different between migratory and nonmigratory bird species. The three replica simulations for each of the studied Cry4 models were used for the performed analysis, which includes the distribution of inter-residue distances in the form of a box-plot, shown in Figure 6. Figure 6 indicates that edge-to-edge distances in the five studied Cry4 models are highly similar in all of the studied avian species, which indicates that radical-pair formation could occur similarly in all of them. The definition of the studied distances is given in Figure 1, while their full corresponding name is summarized in Table 5. Even though the observed functional electron-transfer distances are not significantly different across species, it is important to note that the distances could change more profoundly upon Cry4 activation, where the electron-transfer through the protein could affect protein dynamics and influence reorganization of residues involved in radical-pair formation. Such an effect has been observed earlier for *Er*Cry4 and *Cl*Cry4 structures.[37,51] The distance to Tyr319 was also calculated, because of its proposed role as an electron donor to W_D due to its adjacent spatial position to W_D.[4,63]

The distances F_{d3} and F_{d4} between the FAD and W_C and W_D, respectively, introduced in Figure 6, on the other hand, show a larger degree of variation in different Cry4 models. These distances can affect radical-pair recombination in Cry4 after its photoactivation. Therefore, the distances F_{d3} and F_{d4} are specifically important for the spin recombination reaction,

https://doi.org/10.1021/acs.jpcb.2c00878
J. Phys. Chem. B 2022, 126, 4623−4635

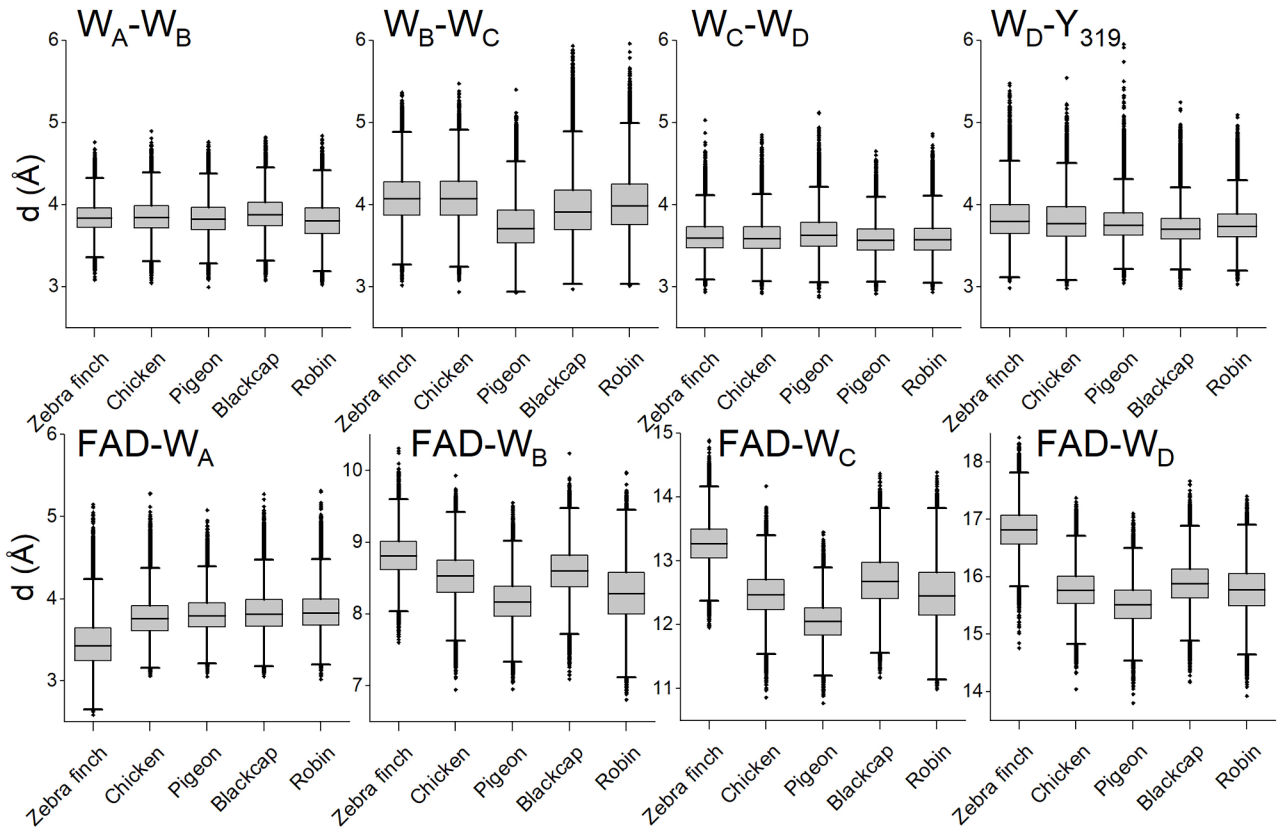


Figure 6. Edge-to-edge distances sampled from MD simulations of the five avian cryptochromes. Hydrogen atoms were not taken into account, and only the flavin part of FAD was considered in the distance calculations. Edge-to-edge distances are important for sequential electron-transfer between FAD and tryptophans (W) in the modeled Cry4 structures. Dark points represent the outlier values, and the full line represents the median line. The gray box shows the middle 50% of the results, and the range given with the straight vertical line shows the interquartile range (IQR) of 1.5 (the range between the 25th and 75th percentile).

Table 5. Nomenclature of the Edge-to-Edge Distances between Different Residues Important for Electron-Transfer Inside Cry4 and the Underlying Magnetic Field Effects

distances in protein	short name
FAD−Trp 395	d1
Trp 395−Trp 372	d2
Trp 372−Trp 318	d3
Trp 318−Trp 369	d4
Trp 369−Tyr 319	d5
FAD−Trp 395	F_{d1}
FAD−Trp 372	F_{d2}
FAD−Trp 318	F_{d3}
FAD−Trp 369	F_{d4}
FAD−Trp 319	F_{d5}

which is crucial for a sensitive magnetic sensor.[13,63−65] It would appear that migratory birds (*Sa*Cry4 and *Er*Cry4) have the F_{d3} and F_{d4} distances in general higher when compared to the values from Cry4 from nonmigratory birds, except for *Tg*Cry4. Here one needs to note that it has been reported that zebra finches can also use a magnetic compass.[66]

Another important characteristic that may affect the magnetic sensitivity of Cry4 is the center-to-center distance between the FAD cofactor and the other residues involved in Cry4 activation, in particular the four tryptophans (W_A, W_B, W_C, and W_D). A pair of radicals needs to be within a certain distance from each other, at approximately 1.5−2.0 nm apart,[3] to have similar energies of singlet and triplet states. If the radical-pairs are too far apart, their formation may not be fast enough, or radical recombination may be too slow to compete with spin relaxation. Figure S8 shows the center-of-mass distances between FAD and each of the four tryptophans (W_A, W_B, W_C, and W_D) and also Tyr 319. Interestingly, the results highlight that there is a clear trend suggesting that Cry4 from migratory birds has similar center-to-center distances when compared to nonmigratory birds.

Overall, the results of Cry4 distance analysis suggest that not only the tryptophans, essential for electron-transfer and radical-pair creation, have been preserved in cryptochromes amino-acid sequences (as seen in Figure S6), but their inter-residue distances, as well as center-to-center distances, are also highly similar. This conclusion strongly suggests that there should likely be other factors like protein reorganization energy or electron-transfer driving force that could affect the response of Cry4 to the external magnetic fields more efficiently than internal structural differences.[20,49]

Inter-residue Distance Analysis. Conformational differences in the dynamics of the protein can further be probed employing the difference distance matrix as seen, for example, in an earlier study.[51] By computing the any-to-any distance for all residues in a protein as measured from the respective backbone C_α atom and averaging these distances over an MD trajectory, a distance matrix was computed for each simulated trajectory. The resulting matrices for two protein structures can then be subtracted element-wise to reach a difference-distance matrix, highlighting rearrangement motifs between the

https://doi.org/10.1021/acs.jpcb.2c00878
J. Phys. Chem. B 2022, 126, 4623−4635

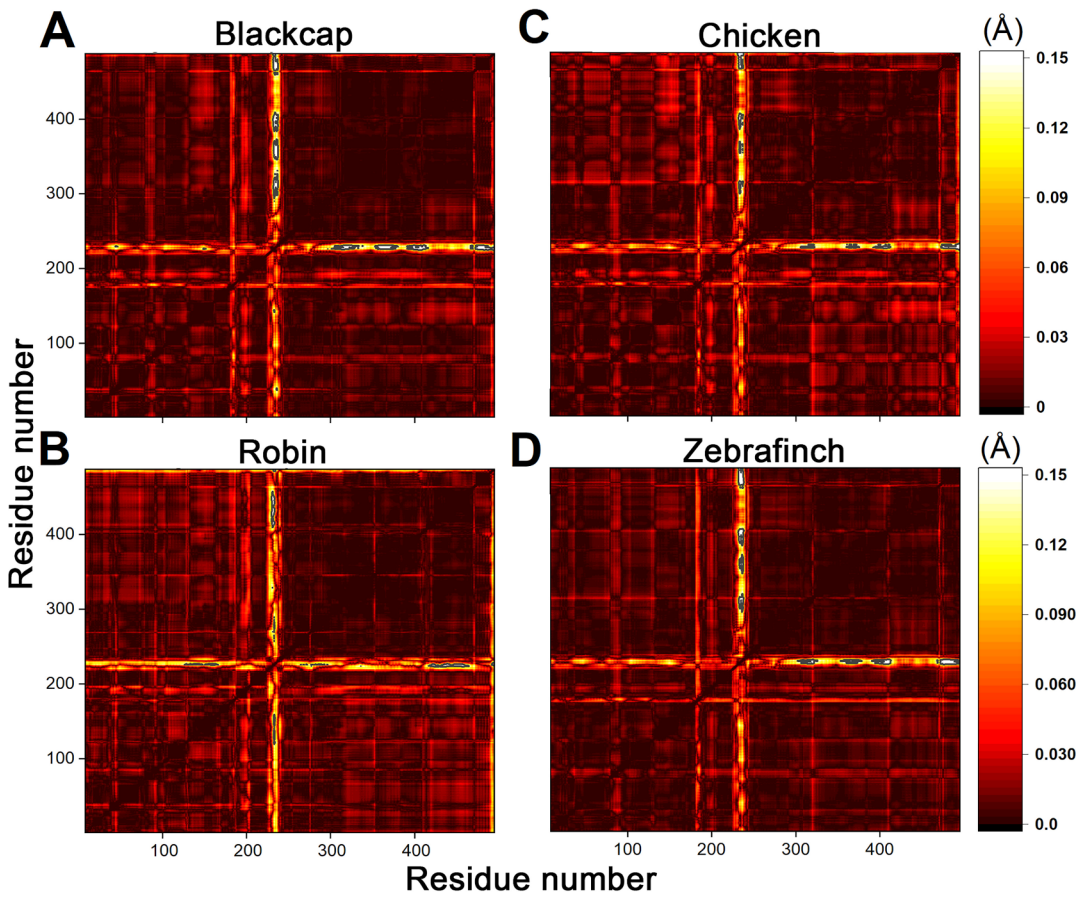


Figure 7. Relative change of the inter-residue distances in different Cry4 models. The contour plots are computed as the absolute inter-residue distance differences relative to those of *Cl*Cry4 (pigeon). All distances were computed as an average value throughout the production simulations of the second replica.

two simulation trajectories, representing different avian Cry4s. The difference distance matrix **A** with entrees a_{ij} is hence given by

$$a_{ij} = \langle \left\| \overrightarrow{R_i} - \overrightarrow{R_j} \right\| \rangle - \langle \left\| \overrightarrow{r_i} - \overrightarrow{r_j} \right\| \rangle \quad (1)$$

where $\vec{R}$ and $\vec{r}$ denote the position of the C_α atoms for the investigated structures, respectively. $\langle \cdot \rangle$ denotes averaging over time.

Figure 7 shows the difference in the distance matrices where the any-to-any distance for all residues in the Cry4 homology models has been compared to those from a simulation of a *Cl*Cry4 crystal structure. At first glance, the four panels in Figure 7 do not differ significantly. However, upon a closer look, it can be noticed that the pattern for *Er*Cry4 is somewhat more delicate compared to the other avian Cry4s. The finer pattern indicates bigger rearrangements that occurred during the simulation.

The difference in the distance maps shown in Figure 7 can be used to obtain an average-distance-per-residue by summing distance deviations for each column in the distance matrix and dividing the obtained value by the total number of residues. The average distance-per-residue plot shows distinctively the residues for which the biggest differences in the structural rearrangements are expected. Figure 8 demonstrates that two similar regions in all Cry4s stand out: the phosphate-binding loop (residues 231−248) and the C-terminal (residues 485−495). Interestingly, in all of the panels, one notices a structural motif prior to the phosphate-binding loop, when two peaks in $|\Delta S|$ indicate increased fluctuation in the proteins. Here the first peak appears to be larger, indicating more notable fluctuation of residues (around residue 200), except the *Er*Cry4 structure in Figure 8C, which compares an average distance change between *Cl*Cry4 and *Er*Cry4.

Looking at inter-residue distances compared to the *Cl*Cry4 structure enables us to temporally compare the similarity of the structures. Such an analysis permits quantification and identification of the difference in Cry4 structures with respect to the *Cl*Cry4 crystal structure. Results indicate that similar motions are present in the modeled structures, indicating similar protein dynamics.

Comparing Crystal and Homology Modeled Structure of *Cl*Cry4. Protein structure determination is a complex problem, and it is often difficult to judge the reliability of the structure predicted through exclusive computational approaches. Even though crystallography-derived protein structures are considered the most accurate representation of the actual protein structure, there are certain caveats with this approach. First, the solved crystal structures are still predicted because there is no way to confirm the atom's exact position because of resolution constraints. Second, the crystal structure presents a solid structure of a protein, as opposed to a dynamic structure that exists in the cell and can be explored with MD simulations.

The uncertainties in atomic positions due to thermal motion can also be accounted for through the B-values given with the

https://doi.org/10.1021/acs.jpcb.2c00878
J. Phys. Chem. B 2022, 126, 4623−4635

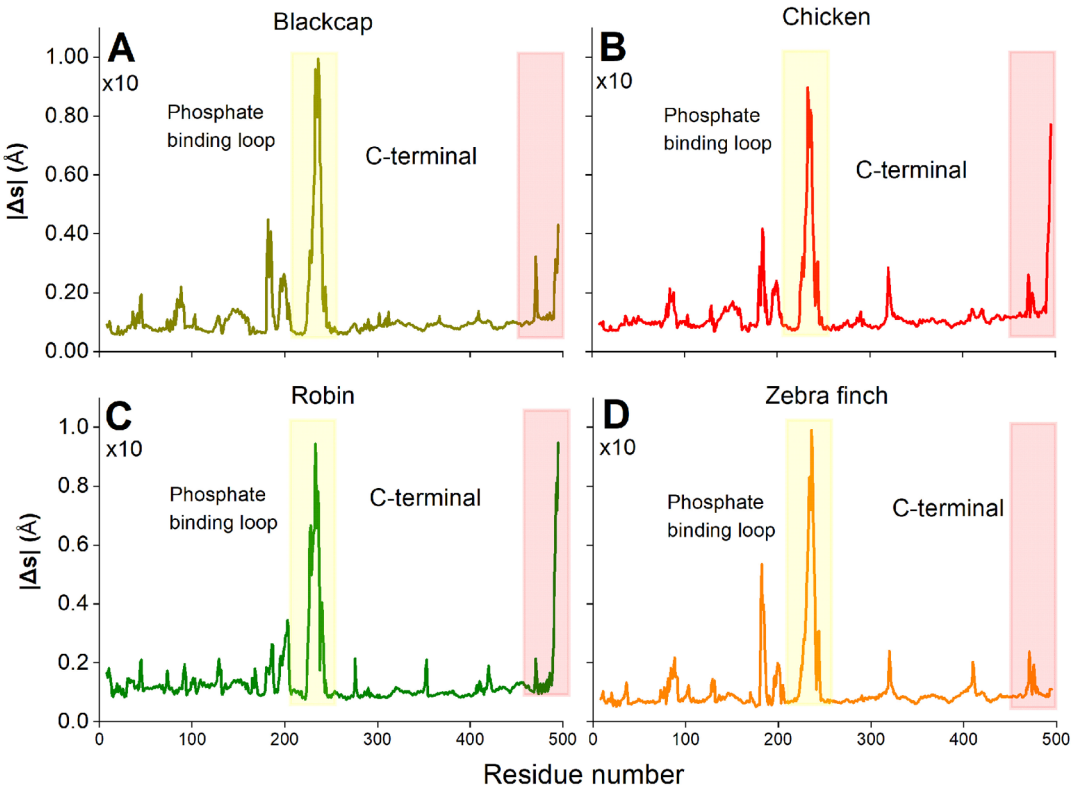


Figure 8. Average deviations of residue displacement in Cry4 structures relative to the corresponding residues in the *Cl*Cry4 structure. The averaging has been performed over the distance deviations shown in Figure 7 and normalizing by the total number of residues in the given protein. Some highly flexible regions in the Cry4 models are indicated by colored bars.

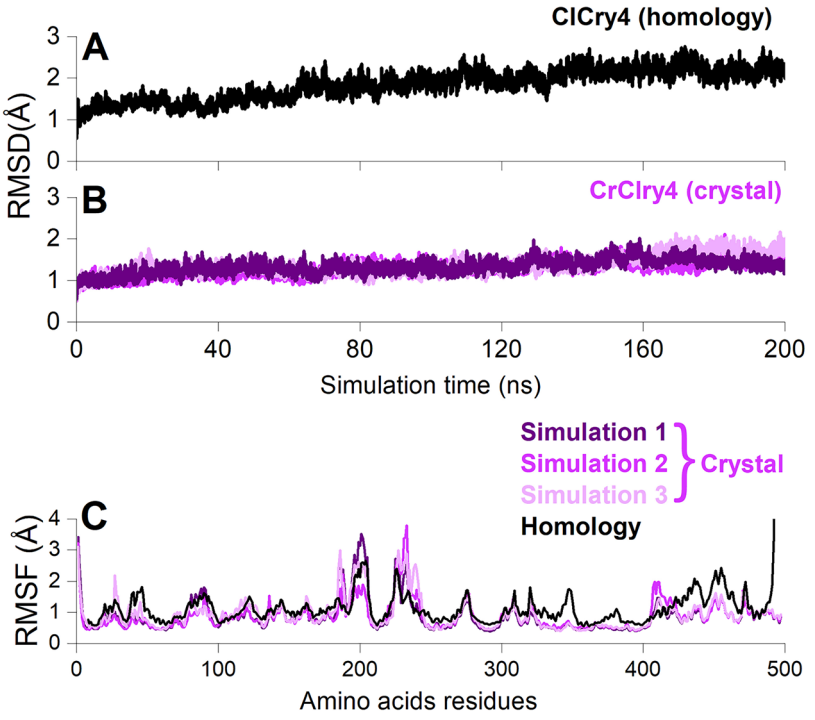


Figure 9. Comparison of RMSD values computed for a homology modeled *Cl*Cry4 structure (A) and simulations of the crystal structure (B). All RMSD values have been calculated for the backbone atoms of the proteins, relative to the structures obtained after the initial equilibration. (C) Comparison of RMSF values of the homology modeled *Cl*Cry4 structure (black) and the crystal *Cl*Cry4 structure (purple).

crystal structure.[22] Higher B-values represent atomic coordinates that have higher uncertainties regarding the position, as shown in Figure S3. Moreover, certain parts of a crystal structure cannot be successfully reconstructed even with the experimental approaches in cases where even atomic coordinates are not accurate enough.[22]

https://doi.org/10.1021/acs.jpcb.2c00878
J. Phys. Chem. B 2022, 126, 4623−4635

To reconcile the possible artifacts of differences between structures derived in different ways, a comparison of their stability and flexibility is needed. A direct comparison of the crystal protein structure of *Cl*Cry4 and an artificially reconstructed *Cl*Cry4 protein could be used to judge on the issue.

Figure 9A,B shows that the *Cl*Cry4 structure based on the *Er*Cry4 homology modeled structure and crystal structure of *Cl*Cry4 do not differ too much in their stability. Flexibility probing which is seen in Figure 9C reveals that a significant part of the structure overlaps in RMSF values with the simulated crystal structure (regions 0−300 and 400−490), and only some relatively minor differences are noted, which could be attributed to the choice of the template structure, which is, in this case, the homology modeled *Er*Cry4 structure.[4] To explain the differences in RMSF for residues 345−450 and 450−470, a comparison of RMSF values computed for the *Er*Cry4 structure with the RMSF values obtained for the homology modeled *Cl*Cry4 has been performed (see Figure S4). The results in Figure S4 indicate that the RMSF values overlap more, which indicates that the choice of the template for homology modeling can influence the RMSF values. Such an influence is, however, expected to be rather minor if the simulation time is long enough and the provided homology model has the possibility to relax. The obtained results thus clearly indicate that homology models can provide reliable protein structures, especially in the cases where amino acid sequence similarity is high (see Table 1 and Figure S6). The approach discussed here is therefore a reasonable strategy to obtain protein structures that are otherwise elusive to the experimental approaches.

To reflect on the recent development in the field of protein reconstruction, a comparison of homology modeled *Er*Cry4 structures and structures made with AlpaFold was performed.[67] Details of this comparison are available in the Supporting Information, where the overall results indicate that the confidence score is low for the equivalent regions of the *Er*Cry4 protein in both methods (see Figure S5). The regions poorly predicted by AlphaFold turn out to be the phosphate-binding loop and the C-terminal tail which have also been identified as highly flexible using RMSF values (see Figure 5C).

CONCLUSION

The comparative approach aimed to characterize and quantify the differences and similarities in the underlying structure and dynamics of various Cry4 structures that could give rise to different magnetic-field responsiveness of the protein *in vitro*.[4] For the investigations represented here, protein structures created by using the homology modeling tools with a follow-up of extensive MD simulations were utilized. The five studied Cry4 models (*Cl*Cry4, *Er*Cry4, *Tg*Cry4, *Sa*Cry4, and *Gg*Cry4) were shown to be stable and to show only small differences between the resulting structures. The five studied protein structures share flexible and stable regions. A comparison of *Cl*Cry4 models derived from a crystal structure and one obtained through homology modeling from *Er*Cry4 also revealed similar flexibility in all regions of the proteins, except the C-terminal.

The results suggest similar activation and signaling mechanisms in the studied avian Cry4 models. The analyses of Cry4 structural elements which could affect the activation of the proteins demonstrated that the electron-transfer distances between the key residues which are important for radical-pair creation are mostly conserved across the studied species, with one important difference in the distance between the FAD cofactor and the W_C (W_D) residues; the variation of these distances could lead to a change in the radical pair recombination reactions in Cry4, which is important for magnetic sensitivity.[3]

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.2c00878.

Further details regarding MD simulation protocols; specific details on RMSD and RMSF calculations; a section which discusses similarity of RMSF values with the B-values of *Cl*Cry4; additional RMSD analysis; dendrograms used to support the clustering analysis; comparison of RMSF values of *Er*Cry4 and homology modeled *Cl*Cry4; amino acid sequence alignments as well as the whole sequences and amino acid sequence accession numbers for the studied cryptochrome structures; additional distance analysis between the important residues inside cryptochromes; the full Swiss-model homology modeling reports for the structures presented in the paper (PDF)

AUTHOR INFORMATION

Corresponding Authors

Maja Hanić − *Department of Physics, Carl von Ossietzky University of Oldenburg, Oldenburg 26129, Germany*; Email: maja.hanic@uni-oldenburg.de

Ilia A. Solov'yov − *Department of Physics, Carl von Ossietzky University of Oldenburg, Oldenburg 26129, Germany; Research Center for Neurosensory Sciences, Carl von Ossietzky University of Oldenburg, Oldenburg 26129, Germany; Center for Nanoscale Dynamics (CENAD), Carl von Ossietzky Universität Oldenburg, Institut für Physik, 26129 Oldenburg, Germany*; orcid.org/0000-0002-8626-145X; Email: ilia.solovyov@uni-oldenburg.de

Authors

Fabian Schuhmann − *Department of Physics, Carl von Ossietzky University of Oldenburg, Oldenburg 26129, Germany*

Anders Frederiksen − *Department of Physics, Carl von Ossietzky University of Oldenburg, Oldenburg 26129, Germany*

Corinna Langebrake − *Institute of Avian Research, Wilhelmshaven 26386, Germany*

Georg Manthey − *Institute of Avian Research, Wilhelmshaven 26386, Germany; Department of Physics, Carl von Ossietzky University of Oldenburg, Oldenburg 26129, Germany*

Miriam Liedvogel − *Institute of Avian Research, Wilhelmshaven 26386, Germany; Department of Biology and Environmental Sciences, Carl von Ossietzky University of Oldenburg, Oldenburg 26129, Germany; MPRG Behavioural Genomics, Max Planck Institute for Evolutionary Biology, Plön 24306, Germany*

Jingjing Xu − *Department of Biology and Environmental Sciences, Carl von Ossietzky University of Oldenburg, Oldenburg 26129, Germany*

https://doi.org/10.1021/acs.jpcb.2c00878
J. Phys. Chem. B 2022, 126, 4623−4635

Henrik Mouritsen − *Department of Biology and Environmental Sciences and Research Center for Neurosensory Sciences, Carl von Ossietzky University of Oldenburg, Oldenburg 26129, Germany*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.2c00878

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank the Volkswagen Foundation (Lichtenberg professorship awarded to IAS), the Deutsche Forschungsgemeinschaft (GRK1885 − Molecular Basis of Sensory Biology and SFB 1372 − Magnetoreception and Navigation in Vertebrates, No. 395940726), the Ministry for Science and Culture of Lower Saxony (Simulations meet experiments on the nanoscale: opening up the quantum world to artificial intelligence (SMART)), and the European Research Council (under the European Union's Horizon 2020 research and innovation programme, Grant Agreement No. 810002, Synergy Grant: "QuantumBirds", awarded to H.M.). Computational resources for the simulations were provided by the CARL Cluster at the Carl-von-Ossietzky University, Oldenburg, supported by the DFG and the Ministry for Science and Culture of Lower Saxony. This work was also supported by the North-German Supercomputing Alliance (HLRN). The authors thank Siu Ying Wong for a fruitful discussion.

■ REFERENCES

(1) Mouritsen, H. Long-distance navigation and magnetoreception in migratory animals. *Nature* **2018**, *558*, 50−59.
(2) Wiltschko, R.; Wiltschko, W. *Magnetic Orientation in Animals*; Springer-Verlag: Berlin, 1995.
(3) Hore, P. J.; Mouritsen, H. The Radical-Pair Mechanism of Magnetoreception. *Annu. Rev. Biophys.* **2016**, *45*, 299−344.
(4) Xu, J.; Jarocha, L. E.; Zollitsch, T.; Konowalczyk, M.; Henbest, K. B.; Richert, S.; Golesworthy, M. J.; Schmidt, J.; Déjean, V.; Sowood, D. J. C.; et al. Magnetic sensitivity of cryptochrome 4 from a migratory songbird. *Nature* **2021**, *594*, 535−540.
(5) Wiltschko, W.; Munro, U.; Ford, H.; Wiltschko, R. Red light disrupts magnetic orientation of migratory birds. *Nature* **1993**, *364*, 525−527.
(6) Wiltschko, R.; Stapput, K.; Thalau, P.; Wiltschko, W. Directional orientation of birds by the magnetic field under different light conditions. *J. R. Soc. Interface* **2010**, *7*, 163−177.
(7) Wiltschko, R.; Thalau, P.; Gehring, D.; Niessner, C.; Ritz, T.; Wiltschko, W. Magnetoreception in birds: the effect of radio-frequency fields. *J. R. Soc. Interface* **2015**, *12*, 20141103.
(8) Hiscock, H. G.; Mouritsen, H.; Manolopoulos, D. E.; Hore, P. Disruption of Magnetic Compass Orientation in Migratory Birds by Radiofrequency Electromagnetic Fields. *Biophys. J.* **2017**, *113*, 1475−1484.
(9) Engels, S.; Schneider, N.-L.; Lefeldt, N.; Hein, C. M.; Zapka, M.; Michalik, A.; Elbers, D.; Kittel, A.; Hore, P. J.; Mouritsen, H. Anthropogenic electromagnetic noise disrupts magnetic compass orientation in a migratory bird. *Nature* **2014**, *509*, 353−356.
(10) Schwarze, S.; Schneider, N.-L.; Reichl, T.; Dreyer, D.; Lefeldt, N.; Engels, S.; Baker, N.; Hore, P. J.; Mouritsen, H. Weak Broadband Electromagnetic Fields are More Disruptive to Magnetic Compass Orientation in a Night-Migratory Songbird (Erithacus rubecula) than Strong Narrow-Band Fields. *Front. Behav. Neurosci.* **2016**, *10*, 55.
(11) Kobylkov, D.; Wynn, J.; Winklhofer, M.; Chetverikova, R.; Xu, J.; Hiscock, H.; Hore, P. J.; Mouritsen, H. Electromagnetic 0.1−100 kHz noise does not disrupt orientation in a night-migrating songbird implying a spin coherence lifetime of less than 10 μs. *J. R. Soc. Interface* **2019**, *16*, 20190716.
(12) Schulten, K.; Swenberg, C.; Weller, A. A Biomagnetic Sensory Mechanism Based on Magnetic Field Modulated Coherent Electron Spin Motion. *Zeitschrift fur Physikalische Chemie* **1978**, *111*, 1−5.
(13) Wong, S. Y.; Frederiksen, A.; Hanić, M.; Schuhmann, F.; Grüning, G.; Hore, P. J.; Solov'yov, I. A. Navigation of migratory songbirds: a quantum magnetic compass sensor. *Neuroforum* **2021**, *27*, 141−150.
(14) Solov'yov, I. A.; Chandler, D. E.; Schulten, K. Magnetic Field Effects in Arabidopsis thaliana Cryptochrome-1. *Biophys. J.* **2007**, *92*, 2711−2726.
(15) Ritz, T.; Adem, S.; Schulten, K. A model for photoreceptor-based magnetoreception in birds. *Biophys. J.* **2000**, *78*, 707−718.
(16) Solov'yov, I. A.; Schulten, K. Reaction Kinetics and Mechanism of Magnetic Field Effects in Cryptochrome. *J. Phys. Chem. B* **2012**, *116*, 1089−1099.
(17) Lüdemann, G.; Solov'yov, I. A.; Kubař, T.; Elstner, M. Solvent driving force ensures fast formation of a persistent and well-separated radical pair in plant cryptochrome. *J. Am. Chem. Soc.* **2015**, *137*, 1147−1156.
(18) Günther, A.; Einwich, A.; Sjulstok, E.; Feederle, R.; Bolte, P.; Koch, K.-W. W.; Solov'yov, I. A.; Mouritsen, H. Double-Cone Localization and Seasonal Expression Pattern Suggest a Role in Magnetoreception for European Robin Cryptochrome 4. *Curr. Biol.* **2018**, *28*, 211−223.e4.
(19) Solov'yov, I. A.; Mouritsen, H.; Schulten, K. Acuity of a Cryptochrome and vision-based magnetoreception system in birds. *Biophys. J.* **2010**, *99*, 40−49.
(20) Solov'yov, I. A.; Domratcheva, T.; Schulten, K. Separation of photo-induced radical pair in cryptochrome to a functionally critical distance. *Sci. Rep* **2015**, *4*, 3845.
(21) Sjulstok, E.; Olsen, J. M. H.; Solov'yov, I. A. Quantifying electron transfer reactions in biological systems: what interactions play the major role? *Sci. Rep.* **2016**, *5*, 18446.
(22) Zoltowski, B. D.; et al. Chemical and Structural Analysis of a Photoactive Vertebrate Cryptochrome from Pigeon. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 19449−19457.
(23) Sjulstok, E.; Solov'yov, I. A. Structural Explanations of Flavin Adenine Dinucleotide Binding in Drosophila melanogaster Cryptochrome. *J. Phys. Chem. Lett.* **2020**, *11*, 3866−3870.
(24) Maeda, K.; Robinson, A. J.; Henbest, K. B.; Hogben, H. J.; Biskup, T.; Ahmad, M.; Schleicher, E.; Weber, S.; Timmel, C. R.; Hore, P. J. Magnetically sensitive light-induced reactions in cryptochrome are consistent with its proposed role as a magnetoreceptor. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 4774−4779.
(25) Wu, H.; Scholten, A.; Einwich, A.; Mouritsen, H.; Koch, K. W. Protein-protein interaction of the putative magnetoreceptor cryptochrome 4 expressed in the avian retina. *Sci. Rep.* **2020**, *10*, 7364.
(26) Hochstoeger, T.; Al Said, T.; Maestre, D.; Walter, F.; Vilceanu, A.; Pedron, M.; Cushion, T. D.; Snider, W.; Nimpf, S.; Nordmann, G. C.; et al. The biophysical, molecular, and anatomical landscape of pigeon CRY4: A candidate light-based quantal magnetosensor. *Sci. Adv.* **2020**, *6*, No. eabb9110.
(27) Feng, S.; Stiller, J.; Deng, Y.; Armstrong, J.; Fang, Q.; Reeve, A. H.; Xie, D.; Chen, G.; Guo, C.; Faircloth, B. C.; et al. Dense sampling of bird diversity increases power of comparative genomics. *Nature* **2020**, *587*, 252−257.
(28) Zhang, Z.; Schwartz, S.; Wagner, L.; Miller, W. A Greedy Algorithm for Aligning DNA Sequences. *Comput. Biol.* **2000**, *7*, 203−214.
(29) Sanger, F.; Nicklen, S.; Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 5463−5467.
(30) Singhal, S.; Leffler, E. M.; Sannareddy, K.; Turner, I.; Venn, O.; Hooper, D. M.; Strand, A. I.; Li, Q.; Raney, B.; Balakrishnan, C. N.; et al. Stable recombination hotspots in birds. *Science* **2015**, *350*, 928−932.

(31) Needleman, S. B.; Wunsch, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **1970**, *48*, 443−453.
(32) Raghava, G. P.; Barton, G. J. Quantification of the variation in percentage identity for protein sequence alignments. *BMC Bioinform* **2006**, *7*, 1−4.
(33) Bordoli, L.; Kiefer, F.; Arnold, K.; Benkert, P.; Battey, J.; Schwede, T. Protein structure homology modeling using SWISS-MODEL workspace. *Nat. Protoc.* **2009**, *4*, 1−13.
(34) Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M. C. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.* **2003**, *31*, 3381−3385.
(35) Korol, V.; Husen, P.; Sjulstok, E.; Nielsen, C.; Friis, I.; Frederiksen, A.; Salo, A. B.; Solov'yov, I. A. Introducing VIKING: A Novel Online Platform for Multiscale Modeling. *ACS Omega* **2020**, *5*, 1254−1260. Paper selected for a cover.
(36) Nielsen, C.; Kattnig, D. R.; Sjulstok, E.; Hore, P. J.; Solov'yov, I. A. Ascorbic acid may not be involved in cryptochrome-based magnetoreception. *J. R. Soc. Interface* **2017**, *14*, 20170657. Paper selected for a cover.
(37) Kattnig, D. R.; Nielsen, C.; Solov'yov, I. A. Molecular dynamics simulations disclose early stages of the photo-Activation of cryptochrome 4. *New J. Phys.* **2018**, *20*, 083018.
(38) Kal, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K.; Kalé, L.; et al. NAMD2: Greater Scalability for Parallel Molecular Dynamics. *J. Comput. Phys.* **1999**, *151*, 283−312.
(39) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(40) Phillips, J. C.; Hardy, D. J.; Maia, J. D.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; et al. Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD. *J. Chem. Phys.* **2020**, *153*, 044130.
(41) Foloppe, N.; MacKerell, A. D. All-Atom Empirical Force Field for Nucleic Acids: I. Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data. *J. Comput. Chem.* **2000**, *21*, 86−104.
(42) Hart, K.; Foloppe, N.; Baker, C. M.; Denning, E. J.; Nilsson, L.; MacKerell, A. D. Optimization of the CHARMM Additive Force Field for DNA: Improved Treatment of the BI/BII Conformational Equilibrium. *J. Chem. Theory Comput.* **2012**, *8*, 348−362.
(43) Huang, J.; MacKerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **2013**, *34*, 2135−2145.
(44) Buck, M.; Bouguet-Bonnet, S.; Pastor, R. W.; MacKerell, A. D., Jr. Importance of the CMAP correction to the CHARMM22 protein force field: dynamics of hen lysozyme. *Biophys. J.* **2006**, *90*, L36−L38.
(45) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ1 and χ2 Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257−3273.
(46) Pavelites, J. J.; Gao, J.; Bash, P. A.; et al. A Molecular Mechanics Force Field for NAD+, NADH, and the Pyrophosphate Groups of Nucleotides. *J. Comput. Chem.* **1997**, *18*, 221−239.
(47) Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; MacKerell, A. D. Impact of 20-Hydroxyl Sampling on the Conformational Properties of RNA: Update of the CHARMM All-Atom Additive Force Field for RNA. *J. Comput. Chem.* **2011**, *32*, 1929−1943.
(48) Mackerell, A. D.; Feig, M.; Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulation. *J. Comput. Chem.* **2004**, *25*, 1400−1415.
(49) Solov'yov, I. A.; Domratcheva, T.; Moughal Shahi, A. R.; Schulten, K. Decrypting Cryptochrome: Revealing the Molecular Identity of the Photoactivation Reaction. *J. Am. Chem. Soc.* **2012**, *134*, 18046−18052.
(50) Sjulstok, E.; Solov'yov, I. A.; Freddolino, P. L. *Methods Enzymol.*, 1st ed.; Elsevier Inc.: London, 2019; Vol. *620*; Chapter 11, pp 277−314.
(51) Schuhmann, F.; Kattnig, D. R.; Solov'yov, I. A. Exploring Post-activation Conformational Changes in Pigeon Cryptochrome 4. *J. Phys. Chem. B* **2021**, *125*, 9652−9659.
(52) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.
(53) Müllner, D. Modern hierarchical, agglomerative clustering algorithms. *arXiv* **2011**, 1109.2378.
(54) Kromann, J. C. *Calculate Root-mean-square deviation (RMSD) of Two Molecules Using Rotation*. 2020; http://github.com/charnley/rmsd, (accessed 2022-04-11).
(55) Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. A* **1976**, *32*, 922−923.
(56) Walker, M. W.; Shao, L.; Volz, R. A. Estimating 3-D location parameters using dual number quaternions. *CVGIP: Image Understanding* **1991**, *54*, 358−367.
(57) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90−95.
(58) Reva, B. A.; Finkelstein, A. V.; Skolnick, J. What is the probability of a chance prediction of a protein structure with an rmsd of 6 å? *Fold. Des.* **1998**, *3*, 141−147.
(59) Wang, Y.; Veglia, G.; Zhong, D.; Gao, J. Activation mechanism of Drosophila cryptochrome through an allosteric switch. *Sci. Adv.* **2021**, *7*, eabg3815.
(60) Zoltowski, B. D.; Vaidya, A. T.; Top, D.; Widom, J.; Young, M. W.; Crane, B. R. Structure of full-length Drosophila cryptochrome. *Nature* **2011**, *480*, 396−399.
(61) Parico, G. C. G.; Partch, C. L. The tail of cryptochromes: an intrinsically disordered cog within the mammalian circadian clock. *Cell Commun. Signal* **2020**, *18*, 182.
(62) Karki, N.; Vergish, S.; Zoltowski, B. D. Cryptochromes: Photochemical and structural insight into magnetoreception. *Protein Sci.* **2021**, *30*, 1521−1534.
(63) Wong, S. Y.; Wei, Y.; Mouritsen, H.; Solov'yov, I. A.; Hore, P. J. Cryptochrome magnetoreception: four tryptophans could be better than three. *J. R. Soc. Interface* **2021**, *18*, 20210601.
(64) Rodgers, C. T.; Hore, P. J. Chemical Magnetoreception in Birds: The Radical Pair Mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 353−360.
(65) Pedersen, J. B.; Nielsen, C.; Solov'yov, I. A. Multiscale description of avian migration: from chemical compass to behaviour modeling. *Sci. Rep.* **2016**, *6*, 36709.
(66) Pinzon-Rodriguez, A.; Muheim, R. Zebra finches have a light-dependent magnetic compass similar to migratory birds. *J. Exp. Biol.* **2017**, *220*, 1202−1209.
(67) Jumper, J.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583−589.

https://doi.org/10.1021/acs.jpcb.2c00878
J. Phys. Chem. B 2022, 126, 4623−4635