

RESEARCH ARTICLE

AlphaFold2 and RoseTTAFold predict posttranslational modifications. Chromophore formation in GFP-like proteins

Sophia M. Hartley^{1‡}, Kelly A. Tiernan^{1‡}, Gjina Ahmetaj^{1‡}, Adriana Cretu^{1‡}, Yan Zhuang², Marc Zimmer^{1*}

1 Department of Chemistry, Connecticut College, New London, CT, United States of America, **2** Department of Mathematics and Statistics, Connecticut College, New London, CT, United States of America

‡ SMH and KAT authors contributed equally, first authors, names in alphabetic order. GA and AC authors contributed equally, names in alphabetic order.

* mzim@conncoll.edu



OPEN ACCESS

Citation: Hartley SM, Tiernan KA, Ahmetaj G, Cretu A, Zhuang Y, Zimmer M (2022) AlphaFold2 and RoseTTAFold predict posttranslational modifications. Chromophore formation in GFP-like proteins. *PLoS ONE* 17(6): e0267560. <https://doi.org/10.1371/journal.pone.0267560>

Editor: Maria Gasset, Consejo Superior de Investigaciones Cientificas, SPAIN

Received: February 16, 2022

Accepted: April 12, 2022

Published: June 16, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0267560>

Copyright: © 2022 Hartley 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 paper and its [Supporting Information](#) files.

Abstract

AlphaFold2 and RoseTTAFold are able to predict, based solely on their sequence whether GFP-like proteins will post-translationally form a chromophore (the part of the protein responsible for fluorescence) or not. Their training has not only taught them protein structure and folding, but also chemistry. The structures of 21 sequences of GFP-like fluorescent proteins that will post-translationally form a chromophore and of 23 GFP-like non-fluorescent proteins that do not have the residues required to form a chromophore were determined by AlphaFold2 and RoseTTAFold. The resultant structures were mined for a series of geometric measurements that are crucial to chromophore formation. Statistical analysis of these measurements showed that both programs conclusively distinguished between chromophore forming and non-chromophore forming proteins. A clear distinction between sequences capable of forming a chromophore and those that do not have the residues required for chromophore formation can be obtained by examining a single measurement—the RMSD of the overlap of the central alpha helices of the crystal structure of S65T GFP and the AlphaFold2 determined structure. Only 10 of the 578 GFP-like proteins in the pdb have no chromophore, yet when AlphaFold2 and RoseTTAFold are presented with the sequences of 44 GFP-like proteins that are not in the pdb they fold the proteins in such a way that one can unequivocally distinguish between those that can and cannot form a chromophore.

Introduction

AlphaFold2 [1, 2] and RoseTTAFold [3] are two freely available programs that can predict three-dimensional protein structures from their amino acid sequence with atomic accuracy. Both programs were created by machine learning and the ~180,000 structures in the protein data bank (pdb) [4, 5] were used as an important training set. The three-dimensional structures of many of the protein structures in the pdb have been influenced by the proteins binding ligands and post-translation modifications. Agirre and co-workers have shown that

Funding: The author(s) received no specific funding for this work.

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

AlphaFold2 predicts the folding of glycosylated proteins solely from the amino acid sequence of the proteins because the program was trained on the pdb in which the glycosylated proteins are found in their full or partially glycanated forms [6]. In the absence of its heme cofactor AlphaFold2 will fold a hemoglobin subunit so that it has a cavity complementary with a heme subunit [6, 7] and it will identify iron-sulfur cluster and zinc binding sites [8]. Finally, AlphaFold2 has been shown to generate protein-peptide complex structures without multiple alignments of the peptide fragment [9].

Fluorescent Proteins (FPs) are commonly used molecular tracer molecules. A Nobel Prize was awarded in 2008 for the development of GFP as a “tagging tool in bioscience” [10] and, in 2014, for the use of fluorescent proteins in “the development of super-resolved fluorescence microscopy” [11]. Four books [12–15] and numerous reviews [16–21] have been written about fluorescent proteins. It is their post-translational autocatalytic chromophore formation that makes these genetic tracer molecules so useful. They are all folded into a β -barrel composed of eleven β sheets surrounding the chromophore that is located in a central alpha helix. Some Green Fluorescent Proteins-like (GFP-like) proteins fold into the characteristic β -barrel shape but do not form a chromophore. The training set (pdb) is heavily weighted towards structures with a chromophore, less than 2% of the GFP-like structures in the pdb have no chromophore. There are 578 GFP-like proteins in the pdb, 568 of these structures have a fully formed chromophore, they are the GFP-like fluorescent proteins (GFP-FPs) and 10 have no chromophore.

The commonly accepted mechanism for chromophore formation in GFP-like proteins is shown in Fig 1. In order for the autocatalytic chromophore formation to occur the immature GFP-like protein has to be in the tight-turn conformation (Fig 1I) [22]. This tight turn breaks the canonical i to $i+4$ hydrogen bonding arrangement found in alpha helices [23], forming a kink that removes the intra main-chain hydrogen bonds that are commonly associated with an α -helix (S1 Fig). It is presumed that this aids in chromophore formation because the hydrogen bonds would otherwise have to be broken during maturation. Tyr66, Gly65, Arg96 and Glu222 (GFP numbering) are involved in chromophore formation and are highly conserved in all naturally occurring GFP-like FPs [24, 25]. The central tyrosine of the chromophore, Tyr66, is conserved in all naturally-occurring GFP-like FPs, although any aromatic residue in that position will auto-catalytically form a chromophore [24]. Thus, we have chemical knowledge that can be used to predict whether a GFP-like protein will form a chromophore or not. A GFP-like protein has to have a glycine65, an aromatic residue in position 66 and an arginine96 to form a chromophore. This is chemical knowledge that was not directly provided to AlphaFold2 or RoseTTAFold.

We have used AlphaFold2 and RoseTTAFold to generate the structures of two series of GFP-like sequences; one that can form a chromophore and the other that can't. And have shown that both programs fold the chromophore forming GFP-like sequences into a distinctly different conformation compared to the sequences that cannot form a chromophore. The structures in the training set taught the AI programs chemistry to distinguish which proteins will be post-translationally modified from those that will not. This was highly unexpected.

Materials and methods

Sequence selection

Standard structures. Three GFP-like proteins with known solid-state structures were selected: 1EMA, 2AWJ, 1H4U (pdb codes).

1EMA is the crystal structure of the S65T mutant of GFP. Together with 1GFL [27] it was the first crystal structure of GFP to be solved [28]. It is the structure of a fluorescent protein that has a mature chromophore formed from the 65TYG67 triad of residues (Fig 1V).

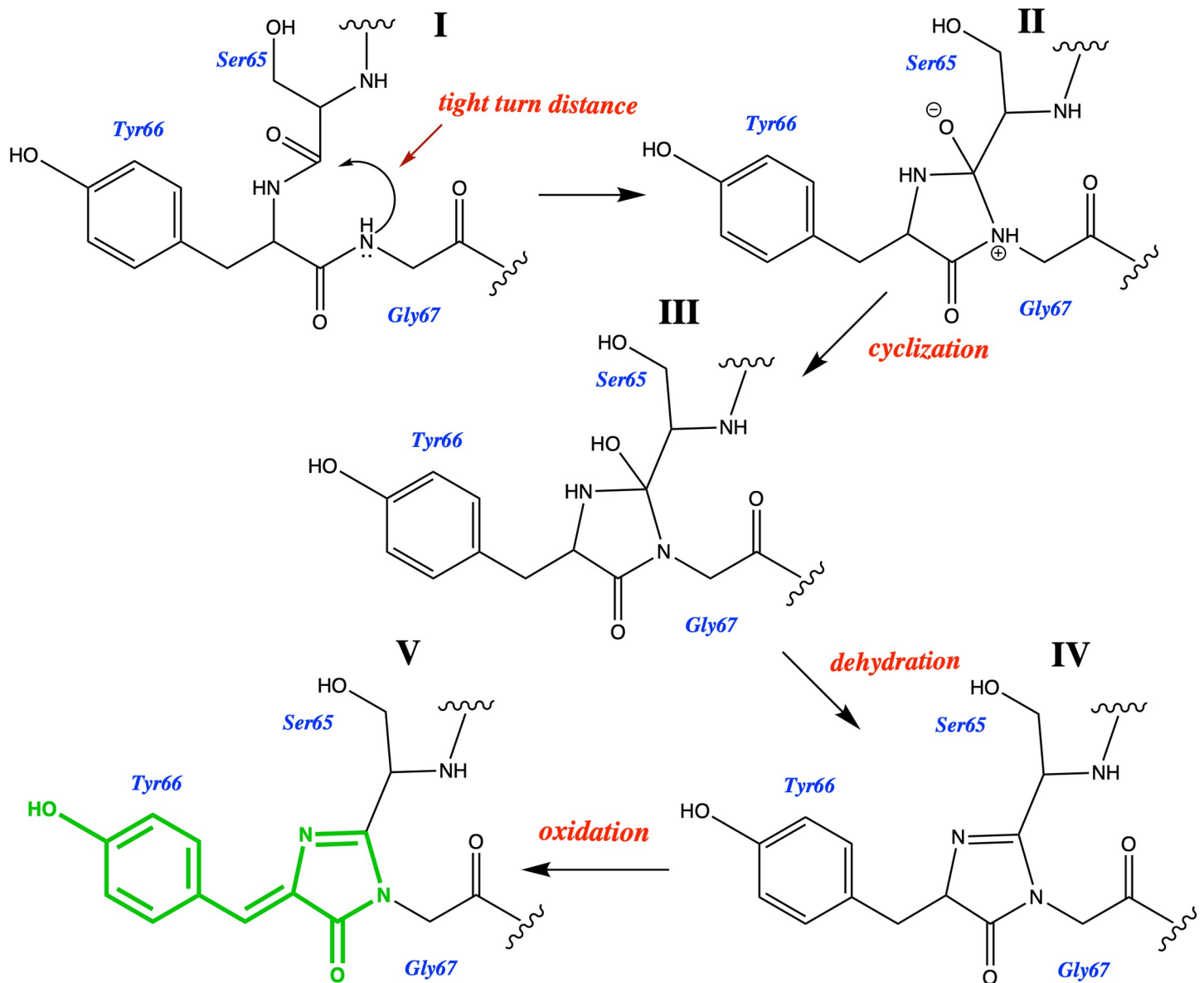


Fig 1. The cyclization-dehydration-oxidation pathway for chromophore formation. [26] Structure I is known as the immature precyclized form of the protein. The tight-turn distance depicted in structure I was measured for all sequences and is presented in [S1 Data](#). Structure V shows the fully formed mature chromophore (green).

<https://doi.org/10.1371/journal.pone.0267560.g001>

2AWJ is an immature precyclized GFP-like protein structure. It is the R96M mutant that takes about 3 months to form the chromophore [29]. Arginine and methionine have similar sizes, allowing us to safely assume that the conformation of immature wild-type GFP and its R96M mutant would be the same. The crystal structure of the R96M mutant is the closest solid-state structure to the immature GFP precyclized structure (Fig 1I).

1H4U is the structure of the G2 domain of mouse nidogen. It is a GFP-like structure with a central alpha helix in an 11 stranded beta barrel. It has no chromophore and cannot form a chromophore [30].

Sequences of chromophore forming GFP-like proteins. Amino acid sequences of GFP-like fluorescent proteins that do not appear in the pdb were obtained from FPbase [31]. The

chromophore forming GFP-like FP sequences were chosen from the FPBase in a way that they spanned as much phylogenetic space possible and had not been crystalized i.e. they did not appear in the pdb. Protein sequences were selected from each FP lineage, starting from the parent protein. If the parent protein's crystal structure was already solved, then the next protein sequence in the lineage closest to the parent was selected. No sequence was used from lineages in which all structures have been solved (DrCBD) and far-red FPs with a tetrapyrroloid chromophores as these FPs are not GFP-like β -barrels.

Sequences of GFP-like proteins that cannot form a chromophore. Sequences of GFP-like non-fluorescent proteins were taken from Haddock's "Non-excitable fluorescent protein orthologs found in ctenophores" paper [32] and a series of nidogen-G2 domains were obtained from a standard protein BLAST search (BLASTP) using the 1H4U FASTA sequence as a protein query (The KAB1270852.1 Nidogen-1 *Camelus dromedarius* sequence was chosen to honor the fact that the Connecticut College mascot is the camel). Only the G2 fragment region of the nidogen sequences were used in the AI alignments and the subsequent analyses. See [S1 Table](#) for sequences studied and their sequence similarity to S65T GFP (1EMA).

AI predictions

RoseTTAFold was accessed via the RosettaCommons web-interface server Robetta [33] and simplified AlphaFold(v2.1.0) was accessed through AlphaFold's Colab notebook [34].

Alignments

MAFFT DASH [35] was used for the multiple alignment and Clustal Omega [36] to find the percent identity matrix of all of the proteins relative to 1EMA, see [S1 Table](#). All residues aligned with residues 60 to 74 of 1EMA were considered part of the central alpha-helix ([S2 Fig](#)).

Measurements

Maestro Version 12.9.123 [37] was used to perform hydrogen bond measurements of the structures obtained from pdb [5], AlphaFold2 and RoseTTAFold. RMSD values for the alpha helical overlaps were obtained by determining the pairwise distances between structures, using the rms displacement after optimal rigid-body superposition between pairs of non-hydrogen backbone atoms of residues highlighted in the alignment shown in [S2 Fig](#) of the supplementary material [38, 39]. All the measurements are presented in [S1 Data](#).

Results and discussion

Structures examined

AlphaFold2 and RoseTTAFold were used to predict the 3-dimensional structure of a series of GFP-like proteins—a set of 3 standard structures with known solid state structures; a group of 21 GFP-like fluorescent proteins that will post-translationally form a chromophore and whose solid state structures have not been determined; as well as a group of 23 GFP-like non-fluorescent proteins that do not have the residues required to form a chromophore and whose solid state structures have not been determined. [S1 Table](#) lists all the sequences and how they were obtained.

How does AI deal with the chromophore?

Since the chromophore formation involves post-translational modification of the 65TYG67 triad we can't expect the AI programs to model the structure of GFP-like FPs with a fully formed chromophore (Structure V in [Fig 1](#)). The central alpha-helical strand (residues 60 to

74) of the structure obtained from the AI predicted folding of the S65T GFP sequence (as found in 1EMA) was overlapped with the pdb structure of 1EMA. Both AlphaFold2 and RoseTTAFold predict a structure closest to the immature precyclized form (structure I in Fig 1, a structure similar to the one obtained if one was to graphically mutate the chromophore back to 65TYG67 and then minimize), a conformation most closely represented by crystal structure of the R96A mutant 2AWJ (RMSD for AF2 = 0.32Å and RF = 0.40Å). The alpha helical overlaps with the regular alpha helix observed in the nidogen G2 domains (RMSD vs 1H4U for AF2 = 1.65Å and RF = 1.68Å) or even vs. its own crystal structure (RMSD vs 1EMA for AF2 = 0.78Å and RF = 0.80Å) are much higher, see [S1 Data](#).

Can AI folding programs predict whether GFP-like proteins will form a chromophore or not?

AlphaFold2 and RoseTTAFold were used to predict the 3-dimensional structures of 21 GFP-like fluorescent proteins that will post-translationally form a chromophore as well as a group of 23 GFP-like non-fluorescent proteins that do not have the residues required to form a chromophore. All sequences folded into GFP-like 11 stranded β -barrels with a central alpha helix. The resultant structures were analyzed to establish whether according to the chromophore forming mechanism discussed in the introduction and shown in Fig 1, they were in conformations geometrically primed for chromophore formation or whether they could not form a chromophore. We overlapped their central alpha helical strands with the strands obtained from the crystal structures of 1EMA (has chromophore), 2AWJ (has no chromophore but is in the right conformation to form one) and 1H4U (has no chromophore and is in the wrong conformation to form chromophore); their tight-turns were measured, as were the i to $i+4$ hydrogen bonding distances in their alpha helices.

Statistical analysis of the AlphaFold2 predictions. From the boxplot, Fig 2, as well as the data presented in [S1 Data](#) and [S2 Table](#), it is apparent that the RMSD for α -helix overlap of the AlphaFold2 predicted structures with 1EMA-crystal is much higher for those that will not form a chromophore than those that do. Moreover, the Welch two sample t-test shows there is a significant difference in the mean values of α -helix overlap with 1EMA-crystal between GFP-like

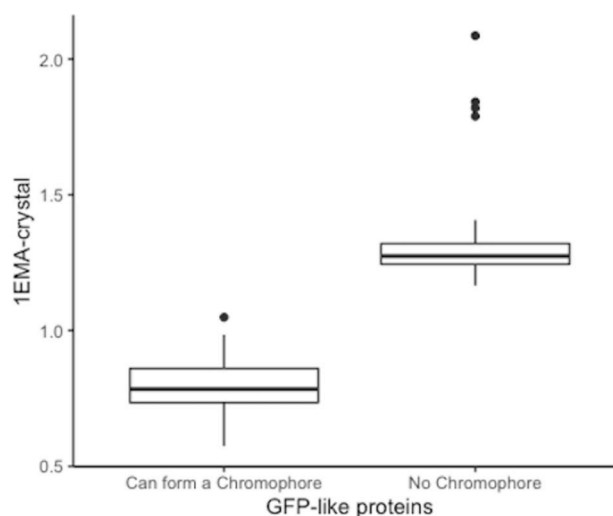


Fig 2. RMSD overlap of the α -helix of the 1EMA-crystal structure with the α -helix of 1EMA as determined by AlphaFold2 for GFP-like proteins that will form a chromophore (SYG) and those that do not (no SYG).

<https://doi.org/10.1371/journal.pone.0267560.g002>

proteins that will form a chromophore and those that do not (t-value = 10.646 and p-value = 7.77×10^{-12}). The average RMSD for GFP-like proteins that will not form a chromophore is 1.372Å while the average value for GFP-like proteins that will form a chromophore is 0.799Å.

For the structures predicted by AlphaFold2 one only needs to compare the predicted alpha helical structure with that of the S65T GFP pdb structure to know whether the GFP-like protein will form a chromophore or not.

A LASSO regression model was built using “can form chromophore” (= 1) and “cannot form chromophore” (= 0) as the response variable with all the measurements including Alpha helix overlap, tight turn distance, as well as H-Bond distances in angstrom. The final model includes the following predictors—the alpha helix overlap with 1EMA-crystal as well as H-bond distances in Angstrom between residues 61–65 (HD2), 62–66 (HD3), 70–74 (HD11). The model shows there is a clear distinction between the sequences capable of chromophore forming and those that don't have the residues required for chromophore formation, see [S3 Fig](#) and [S3 Table](#). A model built on just the measurements of the hydrogen bonding distances is also able to distinguish between chromophore and non-chromophore forming structures.

Statistical analysis of the RoseTTAFold predictions. The RMSD for the α -helix overlap of the RoseTTAFold predicted structures with 1EMA-crystal structures is generally higher for those structures that will not form a chromophore than those that do. However the appropriate boxplot ([S4 Fig](#)) as well as the data presented in [S1 Data](#) and [S4 Table](#), show that while RoseTTAFold can distinguish between posttranslationally modified and non-posttranslationally structures it is not as distinctive as AlphaFold2.

The Welch two sample t-test shows that there is a significant difference in the mean values for alpha helix overlap with 1EMA-crystal between GFP-like proteins that will form a chromophore and those that do not (t-value = 4.699 and p-value = 3.526×10^{-5}). The average value for GFP-like proteins that will not form a chromophore is 1.317 while the average value for GFP-like proteins that will form a chromophore is 1.003.

For data collected using RoseTTaFold, we built a LASSO regression model using “can form chromophore” (= 1) and “cannot form chromophore” (= 0) as the response variable. The selected predictors of final model are: the RMSD of the alpha helix overlap with 1EMA-crystal and 2AWJ-crystal, the tight turn distances, as well as H-bond distances in Angstrom between residues 61 and 65 (HD2), 62–66 (HD3), 63–67 (HD4), 64–68 (HD5), see [S5 Table](#) and [S5 Fig](#). From the prediction results, we can see there is a clear distinction between the sequences capable of forming a chromophore and those that can't. A model built on just the measurements of hydrogen bonding distances is also able to distinguish between chromophore and non-chromophore forming structures.

Conclusions

Only 10 of the 578 GFP-like proteins in the pdb have no chromophore, yet when AlphaFold2 and RoseTTAFold are presented with the sequences of 44 GFP-like proteins that are not in the pdb they fold the proteins in such a way that one can unequivocally distinguish between those that can and cannot form a chromophore. They predict the conformation of the immature protein with a kink in the α -helix (as expected from machine learning vs. memorization) and have used their training set to learn some chemistry, they can distinguish between GFP-like proteins that will form a chromophore and those that do not. We suspect that the pdb training set and multiple sequence alignments enable AlphaFold2 and RoseTTAFold to “think” like chemists and look for the presence of residues equivalent to Arg96, Gly65 and an aromatic residue at position 66 in GFP—the residues required for chromophore formation—and use those to fold the sequence in a chromophore forming or non-chromophore forming conformations.

Supporting information

S1 Fig. Main chain i to $i+4$ “hydrogen bonding” alpha helical interactions measured and presented in [S1 Data](#).

(DOCX)

S2 Fig. Alignment of all sequences used.

(DOCX)

S3 Fig. Predicted possibility that structure forms a chromophore or cannot form a chromophore using LASSO model from [S2 Table](#) on AlphaFold2 data.

(DOCX)

S4 Fig. RMSD overlap of the α -helix of the 1EMA-crystal with the α -helix of 1EMA as determined by RoseTTAFold for GFP-like proteins that will form a chromophore and those that do not.

(DOCX)

S5 Fig. Predicted possibility that structure forms a chromophore or cannot form a chromophore using LASSO model from [S4 Table](#), which is based on RoseTTAFold data.

(DOCX)

S1 Data. RMSD overlap of the α -helix of the 1EMA, 2AWJ, 1H4U crystal structures with the α -helix of the RoseTTAFold and AlphaFold determined structures, tight turn distances of RoseTTAFold and AlphaFold determined structures and the main chain i to $i+4$ “hydrogen bonding” alpha helical interactions. These are the geometric measurements that are crucial to chromophore formation. The structures are divided into two groups, GFP-like proteins that will form a chromophore and those that do not.

(XLSX)

S1 Table. Description of all sequences used in this study [[28–30](#), [32](#), [40–62](#)].

(DOCX)

S2 Table. Summary statistics of the RMSD overlap (in Angstrom) of the α -helix of the 1EMA-crystal with the α -helix of 1EMA as determined by AlphaFold2 for GFP-like proteins that will form a chromophore and those that do not.

(DOCX)

S3 Table. LASSO model results for the alpha helix overlap with 1EMA-crystal as well as H-bond distances in angstrom between residues 61–65 (HD2), 62–66 (HD3), 70–74 (HD11) collected using AlphaFold2.

(DOCX)

S4 Table. Summary statistics of the RMSD overlap (in Angstrom) of the α -helix of the 1EMA-crystal with the α -helix of 1EMA as determined by RoseTTAFold for GFP-like proteins that will form a chromophore and those that do not.

(DOCX)

S5 Table. LASSO model results for the RMSD of the alpha helix overlap with 1EMA and 2AWJ-crystal structures, the tight turn distances, as well as H-bond distances in Angstrom between residues 61 and 65 (HD2), 62–66 (HD3), 63–67 (HD4), 64–68 (HD5) collected using RoseTTAFold simulations.

(DOCX)

Author Contributions

Conceptualization: Marc Zimmer.

Data curation: Sophia M. Hartley, Kelly A. Tiernan, Gjina Ahmetaj, Adriana Cretu.

Formal analysis: Sophia M. Hartley, Kelly A. Tiernan, Gjina Ahmetaj, Adriana Cretu, Yan Zhuang.

Investigation: Sophia M. Hartley, Kelly A. Tiernan, Gjina Ahmetaj, Adriana Cretu, Marc Zimmer.

Methodology: Marc Zimmer.

Project administration: Marc Zimmer.

Resources: Marc Zimmer.

Software: Marc Zimmer.

Supervision: Marc Zimmer.

Validation: Yan Zhuang.

Writing – original draft: Marc Zimmer.

Writing – review & editing: Marc Zimmer.

References

1. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021; 596(7873):583–9. <https://doi.org/10.1038/s41586-021-03819-2> PMID: 34265844
2. Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Židek A, et al. Highly accurate protein structure prediction for the human proteome. *Nature*. 2021; 596(7873):590–6. <https://doi.org/10.1038/s41586-021-03828-1> PMID: 34293799
3. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee Gyu R, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*. 2021; 373(6557):871–6. <https://doi.org/10.1126/science.abj8754> PMID: 34282049
4. Burley SK, Berman HM, Bhikadiya C, Bi C, Chen L, Di Costanzo L, et al. RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Research*. 2018; 47(D1):D464–D74. <https://doi.org/10.1093/nar/gky1004> PMID: 30357411
5. Burley SK, Bhikadiya C, Bi C, Bittrich S, Chen L, Crichlow GV, et al. RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Research*. 2021; 49(D1):D437–D51. <https://doi.org/10.1093/nar/gkaa1038> PMID: 33211854
6. Bagdonas H, Fogarty CA, Fadda E, Agirre J. The case for post-predictional modifications in the AlphaFold Protein Structure Database. *Nature Structural & Molecular Biology*. 2021; 28(11):869–70. <https://doi.org/10.1038/s41594-021-00680-9> PMID: 34716446
7. Hekkelman ML, de Vries Id, Joosten RP, Perrakis A. AlphaFill: enriching the AlphaFold models with ligands and co-factors. *bioRxiv*. 2021:2021.11.26.470110. <https://doi.org/10.1101/2021.11.26.470110>
8. Wehrspan ZJ, McDonnell RT, Elcock AH. Identification of Iron-Sulfur (Fe-S) Cluster and Zinc (Zn) Binding Sites Within Proteomes Predicted by DeepMind's AlphaFold2 Program Dramatically Expands the Metalloproteome. *Journal of Molecular Biology*. 2022; 434(2):167377. <https://doi.org/10.1016/j.jmb.2021.167377> PMID: 34838520
9. Tsaban T, Varga JK, Avraham O, Ben-Aharon Z, Khrumushin A, Schueler-Furman O. Harnessing protein folding neural networks for peptide–protein docking. *Nature Communications*. 2022; 13(1):176. <https://doi.org/10.1038/s41467-021-27838-9> PMID: 35013344
10. The Nobel Prize in Chemistry 2008. Press Release. 2008:<https://www.nobelprize.org/prizes/chemistry/2008/press-release/>.

11. The Nobel Prize in Chemistry 2014. Press Release. 2014:<https://www.nobelprize.org/prizes/chemistry/2014/press-release/>.
12. Zimmer M. *Glowing Genes: A Revolution in Biotechnology*. Amherst, N.Y.: Prometheus Books; 2005.
13. Zimmer M. *Illuminating disease: An introduction to green fluorescent proteins*: Oxford University Press; 2015.
14. Pieribone V, Gruber DF. *Aglow in the dark: the revolutionary science of biofluorescence*. Cambridge, Mass.: Belknap Press of Harvard University Press; 2005.
15. Chalfie M, Kain S, editors. *Green Fluorescent Protein: Properties, Applications, and Protocols*. New York: Wiley-Liss; 1998.
16. Tsien RY. Green Fluorescent Protein. *Annu Rev Biochem*. 1998; 67:509–44. <https://doi.org/10.1146/annurev.biochem.67.1.509> PMID: 9759496
17. Zimmer M. Green fluorescent protein (GFP): Applications, structure, and related photophysical behavior. *Chemical Reviews*. 2002; 102(3):759–81. <https://doi.org/10.1021/cr010142r> PMID: 11890756
18. Day RN, Davidson MW. The fluorescent protein palette: tools for cellular imaging. *Chemical Society Reviews*. 2009; 38(10):2887–921. PubMed PMID: CCC:000270032900008. <https://doi.org/10.1039/b901966a> PMID: 19771335
19. Rodriguez EA, Campbell RE, Lin JY, Lin MZ, Miyawaki A, Palmer AE, et al. The Growing and Glowing Toolbox of Fluorescent and Photoactive Proteins. *Trends Biochem Sci*. 2017; 42(2):111–29. <https://doi.org/10.1016/j.tibs.2016.09.010> PMID: 27814948
20. Stepanenko OV, Stepanenko OV, Kuznetsova IM, Verkhusha VV, Turoverov KK. Chapter Four—Beta-Barrel Scaffold of Fluorescent Proteins: Folding, Stability and Role in Chromophore Formation. In: Jeon KW, editor. *International Review of Cell and Molecular Biology*. 302: Academic Press; 2013. p. 221–78.
21. Nasu Y, Shen Y, Kramer L, Campbell RE. Structure- and mechanism-guided design of single fluorescent protein-based biosensors. *Nature Chemical Biology*. 2021. <https://doi.org/10.1038/s41589-020-00718-x> PMID: 33558715
22. Branchini BR, Nemser AR, Zimmer M. A Computational Analysis of the Unique Protein-Induced Tight Turn That Results in Posttranslational Chromophore Formation in Green Fluorescent Protein. *J Am Chem Soc*. 1998; 120:1–6.
23. Barondeau DP, Kassmann CJ, Tainer JA, Getzoff ED. The case of the missing ring: Radical cleavage of a carbon-carbon bond and implications for GFP chromophore biosynthesis. *Journal of the American Chemical Society*. 2007; 129(11):3118–26. PubMed PMID: ISI:000244896900037. <https://doi.org/10.1021/ja063983u> PMID: 17326633
24. Ong WJH, Alvarez S, Leroux IE, Shahid RS, Samma AA, Peshkepja P, et al. Function and structure of GFP-like proteins in the protein data bank. *Molecular BioSystems*. 2011; 7(4):984–92. PubMed PMID: CCC:000288329300002. <https://doi.org/10.1039/c1mb05012e> PMID: 21298165
25. Lemay NP, Morgan AL, Archer EJ, Dickson LA, Megley CM, Zimmer M. The role of the tight-turn, broken hydrogen bonding, Glu222 and Arg96 in the post-translational green fluorescent protein chromophore formation. *Chemical Physics*. 2008; 348(1–3):152–60. PubMed PMID: ISI:000256737200018. <https://doi.org/10.1016/j.chemphys.2008.02.055> PMID: 19079566
26. Grigorenko BL, Krylov AI, Nemukhin AV. Molecular Modeling Clarifies the Mechanism of Chromophore Maturation in the Green Fluorescent Protein. *Journal of the American Chemical Society*. 2017; 139(30):10239–49. <https://doi.org/10.1021/jacs.7b00676> PMID: 28675933
27. Yang F, Moss LG, Phillips GN. The molecular structure of green fluorescent protein. *Nature Biotechnol*. 1996; 14:1246–51.
28. Ormoe M, Cubitt AB, Kallio K, Gross LA, Tsien RY, Remington SJ. Crystal structure of the *Aequorea victoria* green fluorescent Protein. *Science*. 1996; 273:1392–5. <https://doi.org/10.1126/science.273.5280.1392> PMID: 8703075
29. Wood TI, Barondeau DP, Hitomi C, Kassmann CJ, Tainer JA, Getzoff ED. Defining the role of arginine 96 in green fluorescent protein fluorophore biosynthesis. *Biochemistry*. 2005; 44(49):16211–20. PubMed PMID: ISI:000233898400027. <https://doi.org/10.1021/bi051388j> PMID: 16331981
30. Hopf M, Gohring W, Mann K, Timpl R. Mapping of binding sites for nidogens, fibulin-2, fibronectin and heparin to different IG modules of perlecan. *Journal of Molecular Biology*. 2001; 311(3):529–41. PubMed PMID: ISI:000170509200010. <https://doi.org/10.1006/jmbi.2001.4878> PMID: 11493006
31. Lambert TJ. FPbase: a community-editable fluorescent protein database. *Nature Methods*. 2019; 16(4):277–8. <https://doi.org/10.1038/s41592-019-0352-8> PMID: 30886412
32. Francis WR, Christianson LM, Powers ML, Schnitzler CE, D Haddock SH. Non-excitable fluorescent protein orthologs found in ctenophores. *Bmc Evol Biol*. 2016; 16(1):167. <https://doi.org/10.1186/s12862-016-0738-5> PMID: 27557948

33. Robetta. [January, 2022]. Available from: <https://robetta.bakerlab.org/>.
34. AlphaFoldColab. [January 2022]. Available from: <https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb>.
35. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*. 2019; 20(4):1160–6. <https://doi.org/10.1093/bib/bbx108> PMID: 28968734
36. Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. *Protein Science*. 2018; 27(1):135–45. <https://doi.org/10.1002/pro.3290> PMID: 28884485
37. Maestro. Schroedinger LLC, New York, NY. 2022.
38. Kabsch W. Discussion of Solution for Best Rotation to Relate 2 Sets of Vectors. *Acta Crystallographica Section A*. 1978; 34(SEP):827–8. PubMed PMID: ISI:A1978FR57300032.
39. Kabsch W. Solution for Best Rotation to Relate 2 Sets of Vectors. *Acta Crystallographica Section A*. 1976; 32(SEP1):922–3. PubMed PMID: ISI:A1976CD98100035.
40. Barondeau DP, Kassmann CJ, Tainer JA, Getzoff ED. Understanding GFP chromophore biosynthesis: Controlling backbone cyclization and modifying post-translational chemistry. *Biochemistry*. 2005; 44(6):1960–70. PubMed PMID: ISI:000226969400021. <https://doi.org/10.1021/bi0479205> PMID: 15697221
41. Tubbs JL, Tainer JA, Getzoff ED. Crystallographic structures of *Discosoma* red fluorescent protein with immature and mature chromophores: Linking peptide bond trans-cis isomerization and acylimine formation in chromophore maturation. *Biochemistry*. 2005; 44(29):9833–40. PubMed PMID: ISI:000230628100002. <https://doi.org/10.1021/bi0472907> PMID: 16026155
42. Myšková J, Rybakova O, Brynda J, Khoroshyy P, Bondar A, Lazar J. Directionality of light absorption and emission in representative fluorescent proteins. *Proceedings of the National Academy of Sciences*. 2020; 117(51):32395. <https://doi.org/10.1073/pnas.2017379117> PMID: 33273123
43. Ando R, Flors C, Mizuno H, Hofkens J, Miyawaki A. Highlighted Generation of Fluorescence Signals Using Simultaneous Two-Color Irradiation on Dronpa Mutants. *Biophysical Journal*. 2007; 92(12):L97–L9. <https://doi.org/10.1529/biophysj.107.105882> PMID: 17384059
44. Chudakov DM, Verkhusha VV, Staroverov DB, Souslova EA, Lukyanov S, Lukyanov KA. Photoswitchable cyan fluorescent protein for protein tracking. *Nature Biotechnology*. 2004; 22(11):1435–9. PubMed PMID: ISI:000224960600035. <https://doi.org/10.1038/nbt1025> PMID: 15502815
45. Henderson JN, Remington SJ. Crystal structures and mutational analysis of amFP486, a cyan fluorescent protein from *Anemonia majano*. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102(36):12712–7. PubMed PMID: ISI:000231716700015. <https://doi.org/10.1073/pnas.0502250102> PMID: 16120682
46. Shagin DA, Barsova EV, Yanushevich YG, Fradkov AF, Lukyanov KA, Labas YA, et al. GFP-like proteins as ubiquitous metazoan superfamily: Evolution of functional features and structural complexity. *Molecular Biology and Evolution*. 2004; 21(5):841–50. PubMed PMID: ISI:000221050400006. <https://doi.org/10.1093/molbev/msh079> PMID: 14963095
47. Heim R, Prasher DC, Tsien RY. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc Natl Acad Sci USA*. 1994; 91:12501–4. <https://doi.org/10.1073/pnas.91.26.12501> PMID: 7809066
48. Lambert GG, Depernet H, Gotthard G, Schultz DT, Navizet I, Lambert T, et al. Aequorea's secrets revealed: New fluorescent proteins with unique properties for bioimaging and biosensing. *PLOS Biology*. 2020; 18(11):e3000936. <https://doi.org/10.1371/journal.pbio.3000936> PMID: 33137097
49. Matz MV, Fradkov AF, Labas YA, Savitsky AP, Zaraisky AG, Markelov ML, et al. Fluorescent proteins from nonbioluminescent Anthozoa species. *Nature Biotech*. 1999; 17:969–73. <https://doi.org/10.1038/13657> PMID: 10504696
50. Shimizu A, Shiratori I, Horii K, Waga I. Molecular evolution of versatile derivatives from a GFP-like protein in the marine copepod *Chiridius poppei*. *Plos One*. 2017; 12(7):e0181186. <https://doi.org/10.1371/journal.pone.0181186> PMID: 28700734
51. Labas YA, Gurskaya NG, Yanushevich YG, Fradkov AF, Lukyanov KA, Lukyanov SA, et al. Diversity and evolution of the green fluorescent protein family. *Proceedings of the National Academy of Sciences of the United States of America*. 2002; 99(7):4256–61. <https://doi.org/10.1073/pnas.062552299> PMID: 11929996
52. Shinoda H, Ma Y, Nakashima R, Sakurai K, Matsuda T, Nagai T. Acid-Tolerant Monomeric GFP from *Olin dias formosa*. *Cell Chemical Biology*. 2018; 25(3):330-8.e7. <https://doi.org/10.1016/j.chembiol.2017.12.005> PMID: 29290624
53. Bevis BJ, Glick BS. Rapidly maturing variants of the *Discosoma* red fluorescent protein (DsRed). *Nature Biotechnology*. 2002; 20(11):1159-.

54. Wiedenmann J, Ivanchenko S, Oswald F, Schmitt F, Roecker C, Salih A, et al. EosFP, a fluorescent marker protein with UV-inducible green-to-red fluorescence conversion. *Proceedings of the National Academy of Sciences of the United States of America*. PubMed PMID: ISI:EFIRST040208229. <https://doi.org/10.1073/pnas.0403668101> PMID: 15505211
55. Merzlyak EM, Goedhart J, Shcherbo D, Bulina ME, Shcheglov AS, Fradkov AF, et al. Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nature Methods*. 2007; 4(7):555–7. <https://doi.org/10.1038/nmeth1062> PubMed PMID: ISI:000247648600012. PMID: 17572680
56. Kredel S, Nienhaus K, Oswald F, Wolff M, Ivanchenko S, Cymer F, et al. Optimized and Far-Red-Emitting Variants of Fluorescent Protein eqFP611. *Chemistry & Biology*. 2008; 15(3):224–33. <https://doi.org/10.1016/j.chembiol.2008.02.008> PMID: 18355722
57. Subach OM, Gundorov IS, Yoshimura M, Subach FV, Zhang JH, Gruenwald D, et al. Conversion of Red Fluorescent Protein into a Bright Blue Probe. *Chemistry & Biology*. 2008; 59(10):1116–24. PubMed PMID: CCC:000260362200014. <https://doi.org/10.1016/j.chembiol.2008.08.006> PMID: 18940671
58. Habuchi S, Tsutsui H, Kochaniak AB, Miyawaki A, van Oijen AM. mKikGR, a Monomeric Photoswitchable Fluorescent Protein. *Plos One*. 2008; 3(12):e3944. <https://doi.org/10.1371/journal.pone.0003944> PMID: 19079591
59. Karasawa S, Araki T, Nagai T, Mizuno H, Miyawaki A. Cyan-emitting and orange-emitting fluorescent proteins as a donor/acceptor pair for fluorescence resonance energy transfer. *Biochem J*. 2004; 381:307–12. PubMed PMID: ISI:000222724200035. <https://doi.org/10.1042/BJ20040321> PMID: 15065984
60. Shaner NC, Lambert GG, Chamma A, Ni Y, Cranfill PJ, Baird MA, et al. A bright monomeric green fluorescent protein derived from *Branchiostoma lanceolatum*. *Nature Methods*. 2013; 10(5):407–9. <https://doi.org/10.1038/nmeth.2413> PMID: 23524392
61. Kogure T, Karasawa S, Araki T, Saito K, Kinjo M, Miyawaki A. A fluorescent variant of a protein from the stony coral *Montipora facilitates* dual-color single-laser fluorescence cross-correlation spectroscopy. *Nature Biotechnology*. 2006; 24(5):577–81. PubMed PMID: CCC:000237331300034. <https://doi.org/10.1038/nbt1207> PMID: 16648840
62. Bindels DS, Haarbosch L, van Weeren L, Postma M, Wiese KE, Mastop M, et al. mScarlet: a bright monomeric red fluorescent protein for cellular imaging. *Nature Methods*. 2017; 14(1):53–6. <https://doi.org/10.1038/nmeth.4074> PMID: 27869816