



Red fluorescent proteins engineered from green fluorescent proteins

Hiromi Imamura^{a,1,2} , Shiho Otsubo^{b,1}, Mizuho Nishida^a, Norihiro Takekawa^b , and Katsumi Imada^{b,2}

Edited by Steven Boxer, Stanford University, Stanford, CA; received May 7, 2023; accepted August 31, 2023

Fluorescent proteins (FPs) form a fluorophore through autocatalysis from three consecutive amino acid residues within a polypeptide chain. The two major groups, green FPs (GFPs) and red FPs (RFPs), have distinct fluorophore structures; RFPs have an extended π -conjugation system with an additional double bond. However, due to the low sequence homology between the two groups, amino acid residues essential for determining the different fluorophore structures were unclear. Therefore, engineering a GFP into an RFP has been challenging, and the exact mechanism of how GFPs and RFPs achieve different autocatalytic reactions remained elucidated. Here, we show the conversion of two coral GFPs, AzamiGreen (AG) and mcavGFP, into RFPs by defined mutations. Structural comparison of AG and AzamiRed1.0, an AG-derived RFP, revealed that the mutations triggered drastic rearrangements in the interaction networks between amino acid residues around the fluorophore, suggesting that coordinated multisite mutations are required for the green-to-red conversion. As a result of the structural rearrangements, a cavity suitable for the entry of an oxygen molecule, which is necessary for the double bond formation of the red fluorophores, is created in the proximity of the fluorophore. We also show that a monomeric variant of AzamiRed1.0 can be used for labeling organelles and proteins in mammalian cells. Our results provide a structural basis for understanding the red fluorophore formation mechanism and demonstrate that protein engineering of GFPs is a promising way to create RFPs suitable for fluorescent tags.

fluorescent proteins | protein engineering | X-ray crystallography | live imaging

Aequorea victoria green fluorescent protein (avGFP) and its family proteins (referred hereto simply as fluorescent proteins, FPs) autocatalytically form a fluorophore from three consecutive amino acid residues within the polypeptide chain (1). Due to this unique characteristic of FPs, a transgenic FP can spontaneously form a functional FP without the help of any other genes. Therefore, FPs have become an indispensable and handy tool in modern biological and biomedical research fields as tags for intracellular organelles and proteins or as components of genetically encoded fluorescent biosensors (2, 3). Moreover, FPs have served as model proteins for studying molecular evolution and structure–function relationships (4, 5). The majority of FPs found in nature, including avGFP, emit green fluorescence. On the other hand, there are rarer natural FPs that emit orange to red fluorescence; these include DsRed from *Discosoma* sp. and eqFP578 and eqFP611 from *Entacmaea quadricolor* (Fig. 1A). The fluorophore in a natural FP is formed from three consecutive amino acid residues consisting of an Xaa-Tyr-Gly motif in the center of a helix that penetrates the interior of the characteristic β -barrel structure composed of 11 β -strands. The fluorophore of GFP is generated by cyclization and dehydration of the three-amino-acid main chain and the formation of a double bond between the C α and C β atoms of the Tyr residue (1). In addition to these reactions, the main-chain N-C α bond of the first residue of the amino acid triad is oxidized to form N-acylimine in RFPs (red fluorescent proteins) (6–9) (Fig. 1B). This additional double bond in RFPs extends the fluorophore's π -conjugation system and shifts absorption and emission to longer wavelengths. However, it is still largely unknown which sequence differences and/or geometric rearrangements of amino acids produce this difference in the reaction. Moreover, a binding site for an oxygen molecule that is required for the additional double bond formation in RFP is totally unclear.

In addition to surveys of new FPs in nature, researchers have also made many efforts to create artificial FPs with various properties by engineering natural FPs, including attempts to change the wavelengths of FPs for use in multicolor labeling and biosensor development. FPs showing blue to yellow fluorescence have been developed mainly by engineering avGFP. For example, blue FP (BFP) was created by replacing the Tyr residue that forms the fluorophore of avGFP with His, cyan FP (CFP) with Trp, and ultramarine FP with Phe (12, 13). Yellow FP (YFP) was also created from avGFP by introducing a new Tyr residue for π - π stacking with the fluorophore (14). There are also reports of

Significance

Fluorescent proteins (FPs) are an essential tool for bioimaging and also a model for studying molecular evolution and structure–function relationships. However, despite many attempts to engineer the colors of FPs, the conversion of green FPs (GFPs) into red FPs (RFPs), which requires an additional double bond formation in the fluorophore, has yet to be achieved. Here, we succeeded in engineering coral GFPs to RFPs. The conversion of GFPs into RFPs highlighted the structural rearrangements essential for the red fluorophore formation. In addition, a monomeric variant of the GFP-derived RFP was bright enough for labeling organelles and proteins in mammalian cells. This study provides a structural basis for red fluorophore formation and a strategy to obtain original RFPs.

Author affiliations: ^aDivision of Systemic Life Science, Graduate School of Biostudies, Kyoto University, Kyoto 606-8501, Japan; and ^bDepartment of Macromolecular Science, Graduate School of Science, Osaka University, Toyonaka, Osaka 560-0043, Japan

Author contributions: H.I. and K.I. designed research; H.I., S.O., M.N., N.T., and K.I. performed research; H.I., S.O., N.T., and K.I. analyzed data; and H.I. and K.I. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2023 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹H.I. and S.O. contributed equally to this work.

²To whom correspondence may be addressed. Email: imamura.hiromi.7a@kyoto-u.ac.jp or kimada@chem.sci.osaka-u.ac.jp.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2307687120/-DCSupplemental>.

Published October 23, 2023.

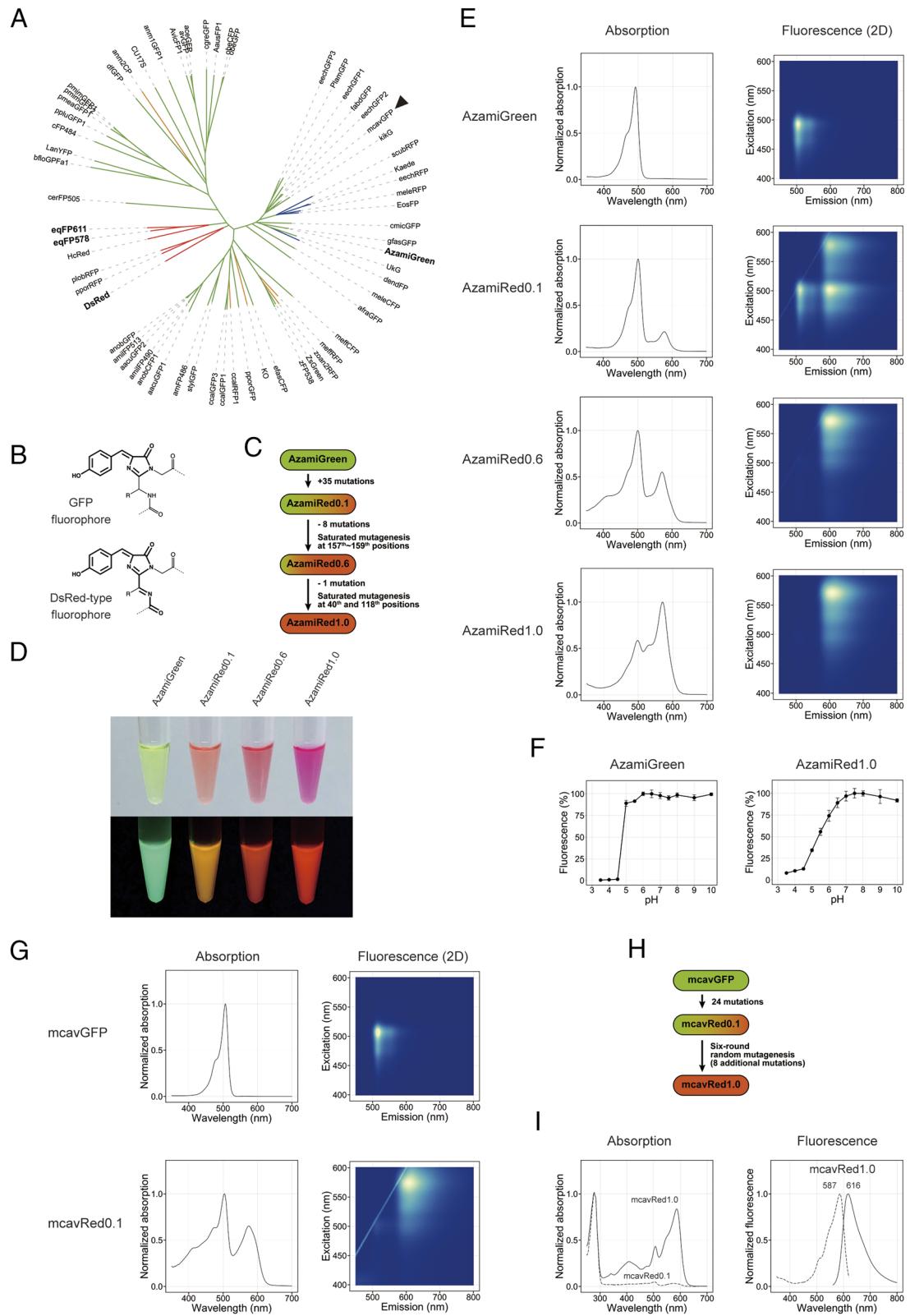


Fig. 1. Conversion of coral GFPs into RFPs. (A) Phylogenetic tree of natural FPs. Green, red, and blue branches indicate green FPs, red/orange FPs, and photoconvertible FPs, respectively. The tree was generated using MAFFT (10) and iTOL (11). (B) Structures of GFP and DsRed-type RFP fluorophores. (C) The process to convert AzamiGreen into RFPs. (D) Photographs of purified AzamiGreen and AzamiRed protein solutions under white light (*Upper*) and UV light (*Lower*). (E) Absorption and two-dimensional fluorescence spectra of AzamiGreen and AzamiRed proteins. Absorption and fluorescence were normalized by peak intensities. (F) pH-sensitivity of AzamiGreen and AzamiRed1.0. Fluorescence intensities of the purified proteins were measured just after diluting with buffers with the indicated pH. (G) Absorption and two-dimensional fluorescence spectra of mcaVGFP and mcaVRed0.1. (H) The process to convert mcaVGFP into RFPs. (I) Improved red fluorophore formation in mcaVRed1.0. Absorption spectra of mcaVRed0.1 and mcaVRed1.0 (*Left*) and fluorescence excitation and emission spectra of mcaVRed1.0 (*Right*).

engineering coral-derived GFP into a photoconvertible FP, which irreversibly changes from GFP to RFP by UV irradiation through cleavage of the main chain and formation of a new double bond that expands the π -conjugation system of the fluorophore (15).

Bright FPs that emit long-wavelength fluorescence have been strongly desired in recent years, especially for their advantages in deep tissue imaging, and many artificial FPs showing orange to deep red fluorescence have been created, mainly from DsRed, eqFP578, and eqFP611 (16–19). However, the fluorescence quantum yields (QYs) of existing long-wavelength FPs, particularly those with fluorescence wavelengths longer than 600 nm, remain lower than those of short-wavelength FPs. This low QY makes long-wavelength FPs relatively dimmer than short-wavelength FPs. Because DsRed, eqFP578, and eqFP611 are evolutionarily close (Fig. 1A), engineered RFPs currently used in research have low sequence diversity, which may limit the freedom to develop new high-performance long-wavelength FPs.

In contrast to the DsRed-type RFPs, GFP sequences found in nature are highly diverse (Fig. 1A). Therefore, if a new lineage of RFP can be artificially extended from GFPs, the sequence diversity of RFPs will be significantly increased, and furthermore, the amino acid residues that play an essential role in the formation of red fluorophores may be identified, which will provide crucial insights into the reaction mechanism of the red fluorophore formation. To date, however, attempts to engineer GFP into FPs consistently emitting orange or longer wavelength light have been challenging.

In this study, we succeeded in creating artificial RFPs with a fluorescence wavelength exceeding 600 nm and a high QY by protein engineering of coral-derived GFPs. A comparison of the three-dimensional structure of the engineered RFP with that of the original GFP revealed changes in the geometric arrangement of amino acid residues associated with red fluorophore formation.

Results

Conversion of Coral GFPs into RFPs. We selected AzamiGreen (AG) (20), a tetrameric bright coral GFP from *Galaxea fascicularis*, with a maximum excitation wavelength of 492 nm and a maximum fluorescence wavelength of 503 nm, as a starting material. AG has 54%, 51%, and 49% amino acid identity with DsRed, eqFP578, and eqFP611, respectively. We hypothesized that highly conserved amino acid residues among the above three RFPs play an essential role in the autocatalytic reaction to form fluorophores with an extended π -conjugation system. We created AG mutants in which amino acid residues highly conserved among the three RFPs but not in AG were replaced with RFP-type amino acids. Among several mutants, we observed weak red fluorescence in a mutant with 35 amino acid substitutions in AG and named this mutant AzamiRed0.1 (AR0.1) (Fig. 1C and D and *SI Appendix*, Figs. S1 and S2). In addition to a strong absorption at 501 nm and a green fluorescent component similar to AG, AR0.1 has an absorption at 577 nm and a red fluorescent component with a maximum at 600 nm (Fig. 1E). When excited around 500 nm, AR0.1 emitted both green and red fluorescence, probably due to Förster resonance energy transfer inside a tetramer. The 35 mutations introduced in AR0.1 may include mutations that do not contribute to the formation of the red fluorophore. Moreover, some of them could adversely affect not only protein folding but even the formation of the red fluorophore. We therefore created mutants of AR0.1 by reverting each of the 35 amino acid mutations to the original AG amino acids and found that some of the new mutants did not show any loss of red fluorescence. Moreover, an H71Y mutant had a reduced green fluorescent component compared with AR0.1.

In addition to the H71Y mutation, we further introduced seven reversion mutations into AR0.1, followed by saturated mutagenesis at residues 157 to 159 to obtain AzamiRed0.6 (AR0.6), which showed increased red fluorescence and reduced green fluorescence (Fig. 1 C–E and *SI Appendix*, Fig. S2). Finally, we obtained AzamiRed1.0 (AR1.0) by further engineering AR0.6, introducing a reversion mutation at residue 195 and saturated mutations at residues 40 and 118 (Fig. 1 C and D and *SI Appendix*, Fig. S2). AR1.0 had significantly reduced absorption at 500 nm compared with its predecessors, with a maximum excitation wavelength at 571 nm and a maximum emission wavelength at 606 nm, and with almost negligible green fluorescence (Fig. 1E). AR1.0 has 29 amino acid substitutions from AG, and the amino acid identity between the two FPs is 87% (*SI Appendix*, Fig. S2). The amino acid triad forming the fluorophore of AR1.0 is Met-Tyr-Gly, the same as in eqFP578, eqFP611, and most engineered RFPs [refer to databases for FPs, such as FPbase (21)]. AR1.0 exhibited somewhat lower stability to acidic pH than AG (Fig. 1F). Molar extinction coefficient of AR1.0 (34,100) was about a half of AG and was much lower than DsRed, eqFP578, and eqFP611 (Table 1). The reason for the low molar extinction coefficient of AR1.0 is unclear. On the other hand, the fluorescence QY of AR1.0 was as much as 0.65 (Table 1), the highest class among FPs having a maximum emission wavelength over 600 nm, surpassing those of cgfmKate2 (0.47) (22) and eqFP611 (0.45) (23). Our results indicate that a FP with a high QY can be developed by engineering natural GFPs.

Next, we examined whether the set of 29 amino acid substitutions in AR1.0 could also transform another GFP into RFP. We introduced these 29 substitutions into a coral GFP from *Montastraea cavernosa* (mcavGFP) (Fig. 1A), which shares 82% sequence identity with AG and approximately 50% identity with DsRed, eqFP578, and eqFP611. Because native mcavGFP already carries 5 out of the 29 amino acid substitutions, 24 amino acid substitutions were introduced (*SI Appendix*, Fig. S3). As expected, the mutant, which we designated as mcavRed0.1, had an absorbance peak at 575 nm and showed dim but apparent red fluorescence (Fig. 1G), indicating that the set of 29 amino acid substitutions confers the ability to form a DsRed-type fluorophore for at least some coral-derived GFPs. The maximum excitation wavelength and the maximum emission wavelength of mcavRed0.1 were 573 nm and 608 nm, respectively. Although mcavRed0.1 showed only negligible green fluorescence, the larger absorbance at 504 nm indicates that it still contains a nonfluorescent GFP-like fluorophore in addition to the red fluorophore. We further performed random mutagenesis to improve the formation of the red fluorophore of mcavRed0.1. After six rounds of random mutagenesis, we obtained mcavRed1.0 with a greatly improved absorption around 580 nm (Fig. 1 H and I and *SI Appendix*, Fig. S3). mcavRed1.0 showed a maximum fluorescence at 616 nm, which is 10 nm longer than AR1.0, and a fluorescence QY of 0.38 (Fig. 1I and Table 1).

Critical Amino Acid Residues for Conversion from GFP to RFP. To investigate which of the 29 amino acid mutations introduced in AG was essential for the green-to-red conversion, we created mutants in which each of the 29 amino acid residues was replaced one by one with an amino acid residue of AG. While many mutations had little or no effect on the formation of the red fluorophore, substitutions at positions 60, 62, 65, 66, 69, 105, 107, 158, or 159 significantly reduced the absorbance of the red fluorophore (*SI Appendix*, Fig. S4), suggesting the critical role of these residues in the formation of the red fluorophore. Moreover, a single amino acid substitution at positions 65, 66, 105, 158, or 159 was accompanied by appearance of green fluorescence

Table 1. Optical properties of FPs

	Absorption (nm)	Excitation (nm)	Emission (nm)	ϵ (M ⁻¹ cm ⁻¹)	Fluorescence QY	Ref.
AG	492	492	505	72,300	0.67	(20)
AR0.1	501 (577)	502 (578)	512 600	11,600 (577 nm)	n.d.	This work
AR0.6	500 (571)	571	601	14,800 (571 nm)	n.d.	This work
AR1.0	571	571	606	34,100	0.65	This work
AR1.1	571	571	608	34,900	0.62	This work
AR1.2	571	571	608	35,200	0.61	This work
AR1.3	571	571	609	32,300	0.61	This work
AR1.4	571	572	603	17,800	n.d.	This work
AR1.5	574	574	609	19,800	0.47	This work
AR1.6	572	571	612	19,200	n.d.	This work
mARs1	560	571	598	27,100	0.26	This work
mcavGFP	506	506	516	83,000	n.d.	This work
mcavRed0.1	504 (575)	573	608 (575 nm)	940	n.d.	This work
mcavRed1.0	586	587	616	24,500	0.38	This work
DsRed		558	583	75,000	0.79	(16)
eqFP578		552	578	102,000	0.54	(24)
eqFP611		559	611	78,000	0.45	(23)
mCherry		587	610	72,000	0.22	(16)
mScarlet-I		569	593	104,000	0.54	(25)

(SI Appendix, Fig. S5). However, it should be noted that none of the 29 single mutants completely lost the red fluorescence.

Next, we successively reverted mutations in AR1.0 to the AG sequence (SI Appendix, Fig. S2). Absorption and fluorescence spectra of AzamiRed1.1 (AR1.1), AzamiRed1.2 (AR1.2), and AzamiRed1.3 (AR1.3), in which four, seven, and nine amino acid residues were reverted, respectively, were quite similar to those of AR1.0 (Fig. 2A and B), indicating that the nine residues substituted in AR1.3 from AR1.0 (Glu6, Arg13, Glu15, Thr39, Lys43, Gly48, Ser82, Leu109, and Gly187) are not essential for forming the red fluorophore (Table 1). Interestingly, the rate of fluorophore formation in cultured mammalian cells was apparently slower in AR1.3 than in AR1.0, AR1.1 and AR1.2 (Fig. 2C). Thr39 and Ser82, which were substituted to create AR1.3 from AR1.2, are located on the surface of the protein and away from the fluorophore, respectively. Therefore, Thr39 and Ser82 are likely to facilitate protein folding rather than the fluorophore formation reaction itself. We further generated AzamiRed1.4 (AR1.4), AzamiRed1.5 (AR1.5), and AzamiRed1.6 (AR1.6) by reverting three, five, and six amino acid residues of AR1.3, respectively. Although the absorption spectra of these mutants were markedly changed compared with AR1.0, these mutants still retained significant red fluorescence (Fig. 2 A and B). There are fourteen amino acid residues in AR1.6 (Pro59, Gln60, Met62, Ser65, Lys66, Ile69, Pro76, Phe87, Gln105, Thr107, Val118, Gly138, Ile158, and Lys159) that are different from AG (SI Appendix, Fig. S2). At least some of these residues must play essential roles in the reaction of the red fluorophore formation in AzamiRed proteins, which we explored by comparing the crystal structure of AR1.0 with that of AG.

Structural Changes Associated with Conversion from GFP to RFP. We determined the crystal structures of AR1.0 and AG at 1.84 and 1.62 Å resolution, respectively. Although our AG

contains three amino acid differences from the AG mutant whose structure is known (PDB ID 6ciu) (26), no significant structural difference was observed between them. The asymmetric unit of the AR1.0 crystal contains six molecules (one and a half tetramers), and that of the AG crystal four (one tetramer). The molecules in each asymmetric unit adopt almost identical structures. The average rmsd values for C α atoms between any two molecules in the asymmetric units are less than 0.335 Å for AR1.0 and 0.191 Å for AG. Both AG and AR1.0 tetramers are similar to the DsRed tetramer in subunit orientation (27). The overall structures of AR1.0 and AG are very similar, with an RMSD for C α atoms less than 0.478 Å between any subunit combinations in the asymmetric units. Most structural differences are localized around the fluorophore, and the significant differences are found in the hydrophilic interaction networks around the fluorophore (SI Appendix, Fig. S6).

The fluorophore of AG formed by Asn62, Tyr63, and Gly64 shows a GFP-type structure, whereas the fluorophore of AR1.0 formed by Met62, Tyr63, and Gly64 has a typical DsRed-type structure (28, 29) (Fig. 3A). The electron density map of AR1.0 clearly shows that the main-chain C of Phe61 and the main-chain N, C α , and C β of Met62 are coplanar with the imidazolinone ring, indicating that the C α of Met62 is an sp² hybrid and the N–C α bond of Met62 is a double bond (Fig. 3A). Thus, the conjugated π -electron system is extended to form an acylimine. The peptide bond preceding the fluorophore (between Phe61 and Met62) shows a cis conformation, as found in other RFPs (27, 29–31). The phenolic ring of the AR1.0 fluorophore is in the same plane as the imidazolinone ring, and therefore, the C α –C β bond of Tyr63 is dehydrogenated to form a double bond.

The phenolic ring adopts a cis conformation, like the DsRed fluorophore (27–29). The phenolic oxygen of the AR1.0 fluorophore

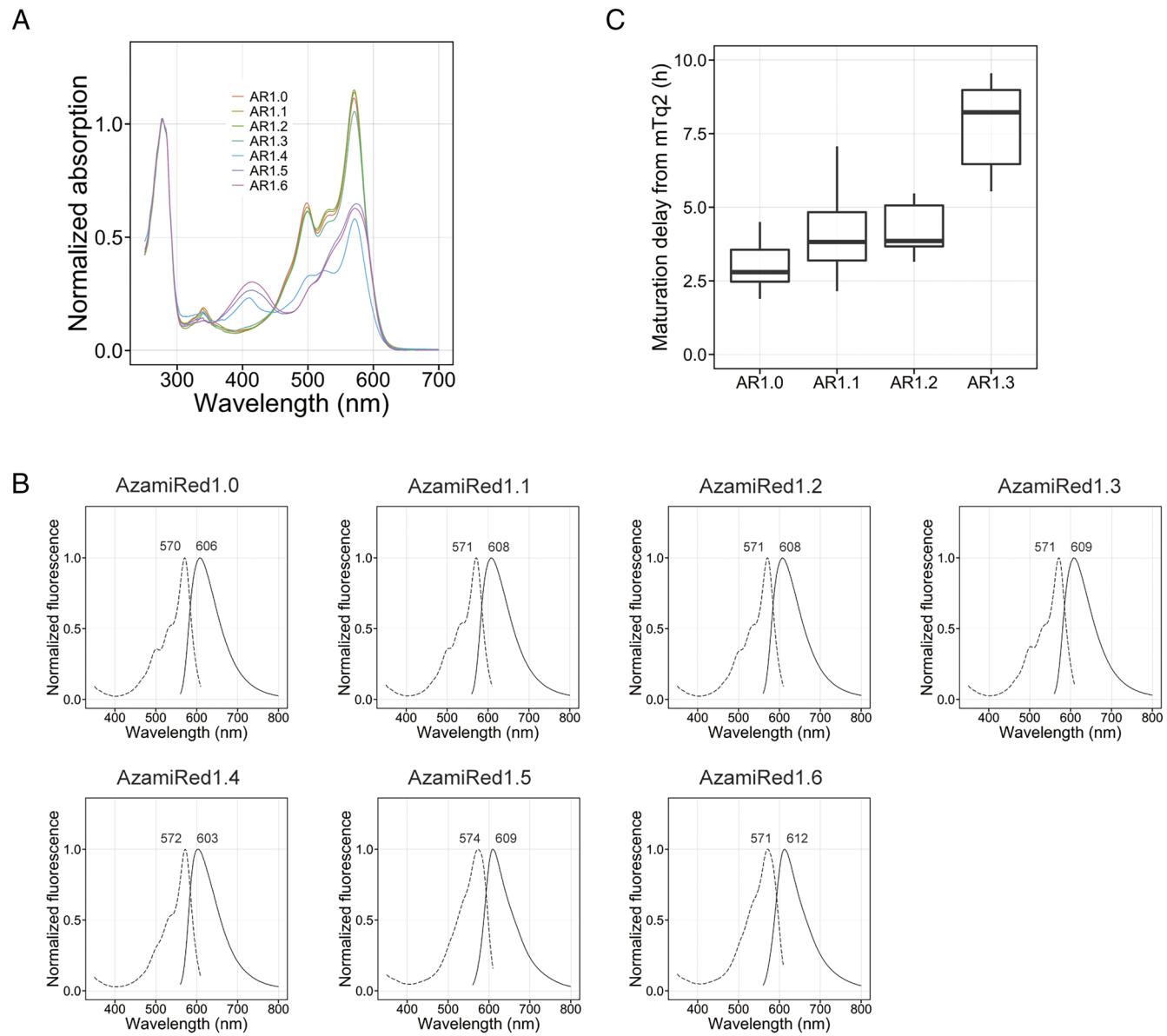


Fig. 2. Spectral characteristics of multiple-reversion variants of AzamiRed1.0. (A) Absorption spectra of AzamiRed1.0 and its variants. The spectra were normalized by intensity at 280 nm. (B) Fluorescence excitation and emission spectra of AzamiRed1.0 and its variants. Fluorescence was normalized by peak intensity. (C) Maturation of AzamiRed1.0 and its variants in cultured mammalian cells. Maturation delay from mTurquoise2 was analyzed for AzamiRed1.0, AzamiRed1.1, AzamiRed1.2, and AzamiRed1.3. Data from 8 cells were used for each FP.

interacts with O γ of Ser142, N ζ of Lys159, and a water molecule in the cleft between β -strand 7 and 10 (Fig. 3B). These atoms form a hydrophilic interaction network with the main-chain carbonyl oxygen atom of Glu140, the nitrogen atoms of Ile195 and Ala160, and a water molecule in the cleft between β 7 and β 8. A similar interaction network is formed in DsRed (PDB ID: 1zgo) (29). The planar phenolate ring is stabilized by π - π stacking with His193 (Fig. 3B). The hydrophilic interaction network and the π - π stacking interaction are conserved in AG, except for Lys159 where the corresponding residue of AG is methionine.

In contrast to the structure around the phenolic oxygen, the hydrophilic interaction network surrounding the imidazolinone ring of AR1.0 differs from those of AG and DsRed (29) (Fig. 3C and *SI Appendix*, Fig. S6) but resembles that of d1eqFP611 (PDB ID: 3E5T), although the phenolic ring of d1eqFP611 adopts a trans-conformation (32). Unlike DsRed, Glu211 in AR1.0 directly interacts with the imidazoline ring nitrogen (N2). Lys66,

Glu144, His193, and Glu211 form a hydrogen bonding network with a water cluster composed of four water molecules below the imidazolinone ring. Glu144 and Glu211 hydrogen bond with N ϵ and N δ of His193, respectively. Lys66 interacts with Glu144 and the water cluster but has no direct interaction with Glu211, unlike DsRed (28). Therefore, the negative charge of Glu211 may be stabilized by the Glu211-His193-Glu144-Lys66 network. Glu144 also interacts with Tyr177 to stabilize the conformation of the side-chain carboxy group. The guanidino group of Arg91 directly interacts with the carbonyl oxygen of the imidazolinone ring. The conformation of the guanidino group of Arg91 is guided by hydrogen bonds with O γ of Thr175, O η of Tyr177, O ϵ of Gln60, and the carbonyl oxygen of Pro59.

Although the hydrogen bonding network formed by Glu211, His193, Glu144, and Tyr177 also exists in AG, no water cluster is found near the imidazolinone ring. Glu211 in AG directly interacts with Arg66, which adopts a unique folded conformation stabilized

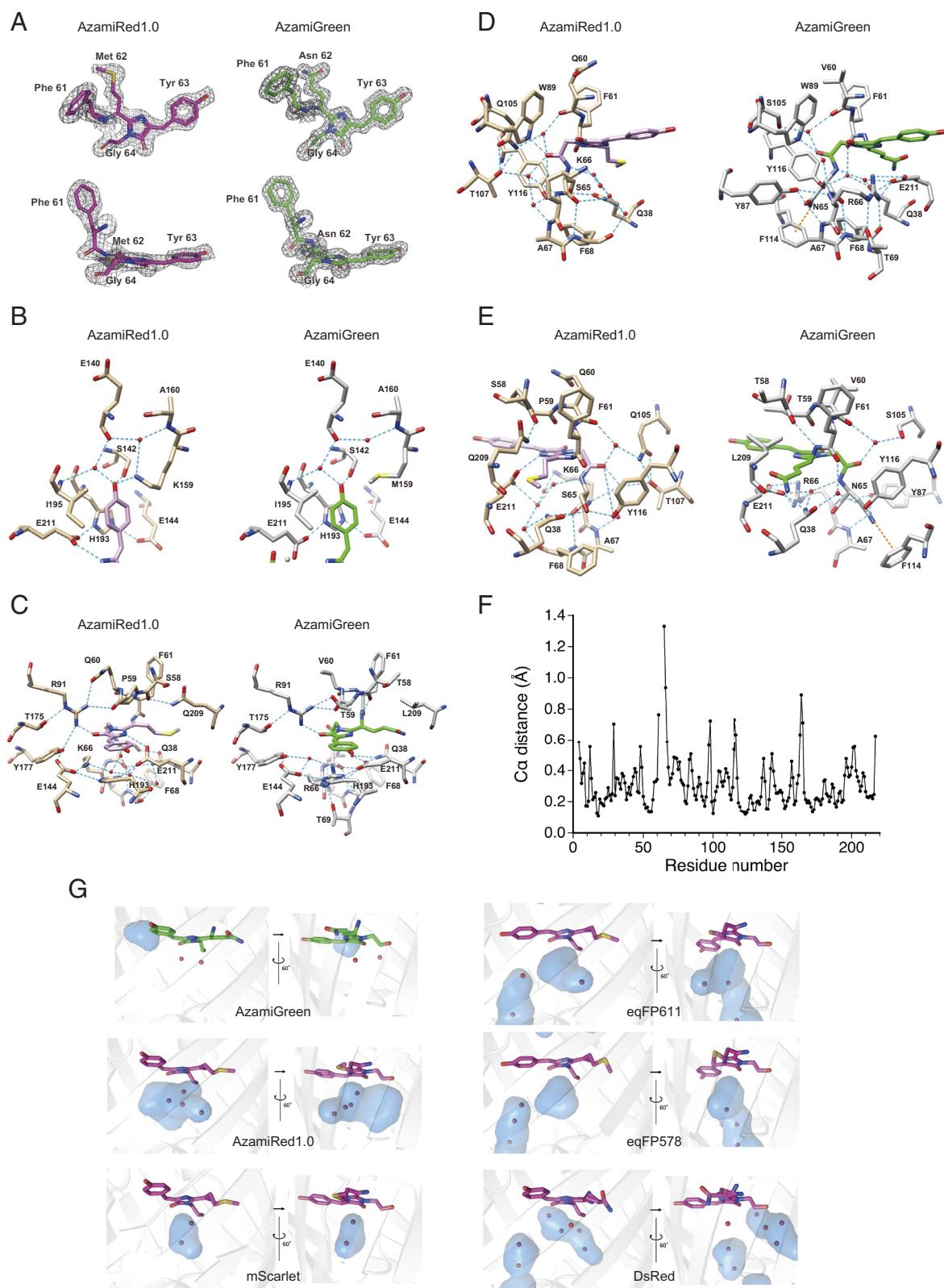


Fig. 3. Structural comparison between AR1.0 and AG. *Left panel*, AR1.0; *Right panel*, AG. The fluorophores in AR1.0 and AG are colored with pink and green, respectively. The nitrogen, oxygen, and sulfur atoms are colored in blue, red, and yellow, respectively. Hydrophilic interactions are indicated by cyan broken lines. (A) Close-up view of the structure of the fluorophore drawn with the 2Fo-Fc electron density map. (B) The hydrophilic interaction network around the phenolic oxygen of the fluorophore. (C) The hydrophilic interaction network surrounding the imidazolone ring. The four water molecules shown by red balls form a water cluster below the fluorophore of AR1.0. (D) The hydrophilic interaction network stabilizing the segment following the fluorophore. (E) The hydrophilic interaction network viewed from the opposite side of D. (F) Comparison of main-chain structures of AR1.0 and AG. The distance between the C α atoms of AR1.0 and AG after least square fitting is plotted against the residue number. (G) Distribution of water and cavities near the fluorophore. Cavities near the fluorophore of AG, AR1.0, mScarlet (PDB ID: 5lk4), d1eqFP611 (PDB ID: 3e5t), eqFP578 (PDB ID: 3pib), and DsRed (PDB ID: 1zgo). The cavities are shown by cyan. The water molecules around the fluorophore are indicated by red balls. The fluorophore is drawn by stick models. The structure of each protein is shown by ribbon models colored with light gray.

by O_y of Thr69 and the carbonyl oxygen of Phe68, but not with the imidazolinone nitrogen (Fig. 3 C–E). Glu211 in AG also interacts with Ne of Gln38 and a water molecule. Thus, the negative charge of Glu211 in AG may be neutralized by these multiple interactions. The orientation of Arg 91 in AG is aligned by hydrogen bonds with O_y and the carbonyl oxygen of Thr59 and O_y of Thr175 to interact with the carbonyl oxygen of the imidazolinone ring.

The segment following the fluorophore (residues from 65 to 67) shows the most prominent main-chain structural difference between AR1.0 and AG (Fig. 3F). The conformation of each segment is maintained by distinct hydrogen bonding networks (Fig. 3 D and E and *SI Appendix*, Fig. S6). The main-chain conformation of the segment in AR1.0 (Ser65-Lys66-Ala67) is stabilized by a hydrogen bonding network formed by the side-chain atoms of Trp89, Gln105, and Thr107, main-chain nitrogen atoms of Lys66 and Ala67, carbonyl oxygen atoms of Gln60 and the fluorophore (Gly 64), and three water molecules. O_y of Ser65 hydrogen bonds with the side chains of Tyr116 and Gln38 (Fig. 3E). Gln38 in AR1.0 also interacts with the main-chain carbonyl oxygen of Ser65 and a water molecule in the water cluster.

On the other hand, the main-chain conformation of the segment in AG (Asn65-Arg66-Ala67) is distorted by the interactions between the guanidino group of Arg66 and its surrounding residues including Phe68, Thr69, and Glu211. The segment conformation is further stabilized by a hydrogen bonding network formed by the side-chain atoms of Trp89 and Ser105, main-chain nitrogen atoms of Arg66 and Ala67, the carbonyl oxygen atoms of Val60 and the fluorophore (Gly 64), and three water molecules (Fig. 3D). The side-chain of Asn65 in AG points in a different direction from Ser65 in AR1.0. N_δ of Asn65 hydrogen bonds with Tyr116 and forms a cation-π interaction with Phe114. The O_δ of Asn65 hydrogen bonds with Tyr87 and Ala67. These interactions also contribute to stabilize the segment conformation (Fig. 3E). As a consequence of these extensive structural rearrangements, a large cavity is created below the imidazolinone ring in AR1.0 that contains four water molecules (Fig. 3G). This cavity is found in most other RFPs and may be essential for the red fluorophore formation (see *Discussion* section).

The trans-cis isomerization of the 61th to 62th peptide bond is an essential step for the red fluorophore formation (27, 28). The main-chain carbonyl oxygen of Phe61 in AG interacts with Gln38 and Tyr116 via a water molecule (Fig. 3E). This interaction may strengthen the trans-conformation of the Phe61-Asn62 peptide bond and disturb the isomerization to the cis-form required for acylimine formation. In contrast, Phe61 in AR1.0 has no hydrogen-bonding partner because the Ser65 side-chain expels the water molecule to form new hydrogen bonds with Gln38 and Tyr116, allowing the trans-cis isomerization in AR1.0 (Fig. 3E). The trans-cis isomerization accompanies the breakage of the main-chain hydrogen bond between the 58th and 62th residues. Loss of the hydrogen bond is compensated by a new hydrogen bond between the carbonyl oxygen of Ser58 and Ne of Gln209 to stabilize the cis-peptide in AR1.0 (Fig. 3E). The contribution of the new hydrogen bond to the red fluorophore formation is supported by the fact that a Q209L mutation reduced the 580 nm absorption (*SI Appendix*, Fig. S4). Similar hydrogen bonding is present in DsRed (27).

Monomerized AzamiRed for Fluorescence Imaging. Multimeric FPs are unsuitable for fluorescent tags because the multimeric nature can affect the subcellular localization and function of the fusion partner protein. AR1.0 is a tetramer, and therefore, we monomerized AR1.0 for versatile use as a fluorescent tag.

We initially introduced the V123T/Y188A/F190K mutation, which was used to monomerize AG (20), into AR1.0, but the fluorescence was almost completely lost. Many GFPs, including AG and avGFP, rarely lose their fluorescence by monomerizing mutations. In contrast, RFPs tend to lose their fluorescence upon monomerization (18, 24, 33). This difference indicates that red fluorophore formation requires strict residue arrangement, and even small conformational changes associated with monomerization prevent the red fluorophore formation. To restore the fluorescence, we performed rounds of random mutagenesis and site-directed saturated mutagenesis on the monomerized AR1.0. During this process, we also eliminated the C-terminal 7 residues. We finally obtained a monomeric variant with red fluorescence, called mARs1, in which a total of 16 residues were substituted from AR1.0, in addition to the deletion of the C-terminal 7 residues (Fig. 4A and *SI Appendix*, Fig. S7A). Most of the substitutions are located on the surface of the protein or apart from the fluorophore (*SI Appendix*, Fig. S7B). The purified mARs1 protein showed a single peak at a position corresponding to a monomer by size exclusion chromatography, even at a very high concentration (Fig. 4B). To confirm whether mARs1 behaves as a monomer inside cells, we conducted the OSER assay (34). The OSER score of mARs1 was as low as that of mCherry, below the monomeric threshold (Fig. 4C). Thus, we concluded that mARs1 has a very low tendency to oligomerize even in the cells. mARs1 exhibited blue-shifted absorption (560 nm) and emission (598 nm) maxima (by about 10 nm) compared with AR1.0, while retaining the same maximum excitation wavelength (571 nm) as AR1.0 (Fig. 4D). The molar extinction coefficient and fluorescence QY of mARs1 were 27,100 M⁻¹ cm⁻¹ and 0.26, respectively (Table 1). The brightness of AR1.0 expressed in cultured mammalian cells exceeded that of mCherry, reaching about 66 % of that of mScarlet-I (Fig. 4E). On the other hand, the brightness of mARs1 expressed in the cells was, reflecting the lower extinction coefficient and QY, approximately 18% and 9% of those of mCherry and mScarlet-I, respectively. Investigation of photostability of AR1.0 and mARs1 in cultured mammalian cells against intense light revealed that both RFPs show complex photobleaching kinetics (*SI Appendix*, Fig. S8). AR1.0 showed two time components in the photobleaching. mARs1 started photobleaching after a short lag, suggesting that photoactivation occurred in mARs1, similar to mOrange2 and mCardinal (35). Finally, we expressed mARs1 fused with proteins or organelle localization sequences in cultured mammalian cells and examined their subcellular localization. The mARs1 fusion proteins visualized intracellular structures such as mitochondria, the nucleus, endoplasmic reticulum, Golgi apparatus, lysosomes, endosomes, plasma membrane, actin fibers, and extending microtubule ends (Fig. 4F and *Movie S1*). Thus, mARs1 functions as well as a standard red fluorescent tag in cells. These results demonstrate that the development of monomeric RFPs suitable for protein and organelle tags can be achieved through the protein engineering of GFPs.

Discussion

Since the cloning of the avGFP gene, various FPs with different emission wavelengths have been artificially produced by protein engineering methods. In contrast, attempts to create orange/red FPs from GFP, including avGFP, have often faced difficulties, although Mishin and colleagues (36) described a mutant avGFP that exhibited a subtle red fluorescent component. Previous molecular evolution analyses demonstrated that the ancestral FPs had

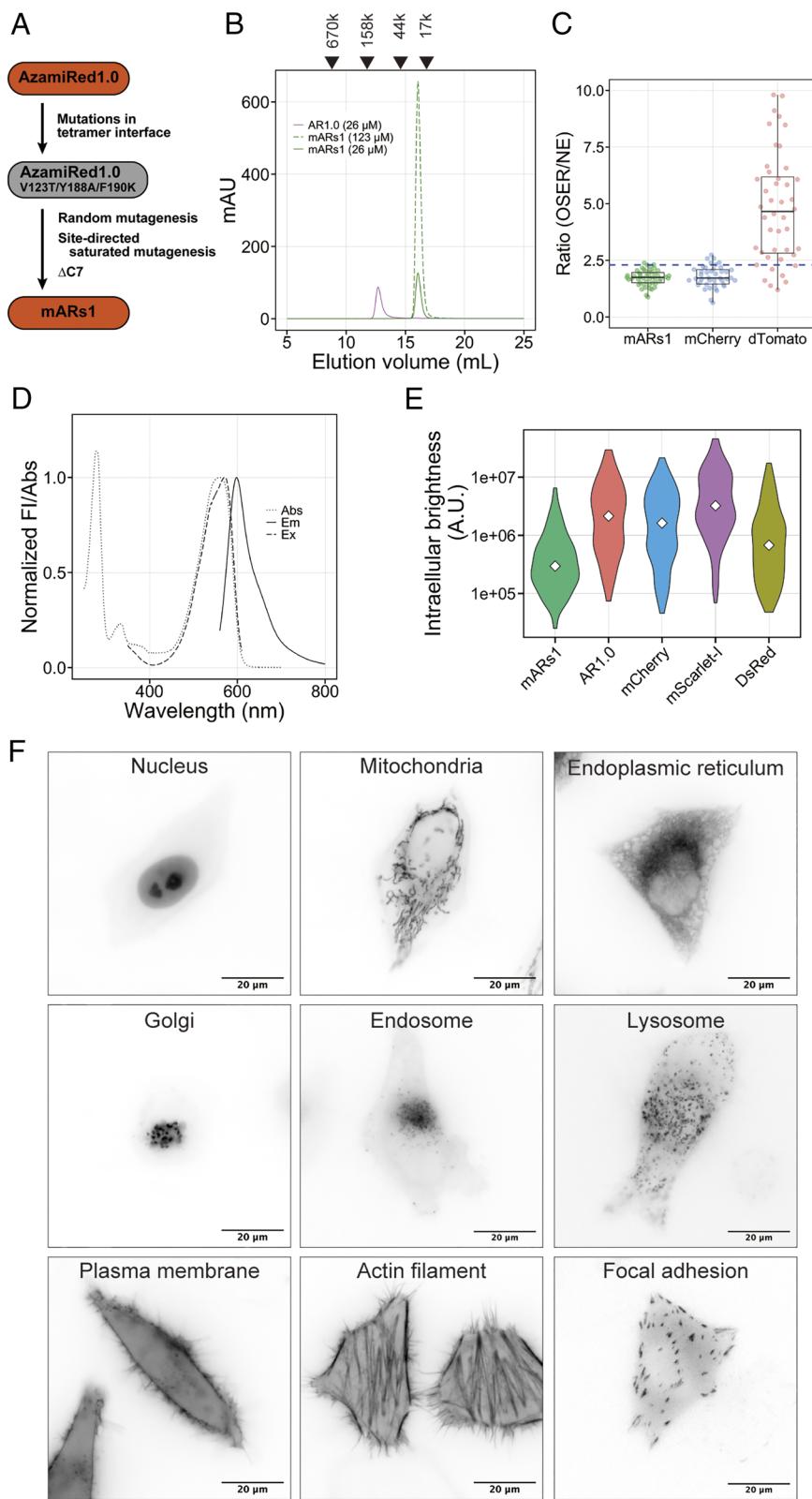


Fig. 4. mARs1, a monomeric mutant of AzamiRed. (A) The process to generate mARs1 from AzamiRed1.0. (B) Oligomeric status of AzamiRed1.0 and mARs1 investigated with gel-filtration chromatography. Elution positions of standard proteins are indicated by arrowheads. (C) Oligomeric properties of mARs1 inside cells investigated by the OSER assay. mCherry and dTomato were used as a control for monomer and dimer, respectively. The number of cells used for measurements was 48 (mARs1), 42 (mCherry), and 44 (dTomato). A dashed horizontal line indicates the threshold of monomeric protein (2.3). (D) Absorption and fluorescence spectra of mARs1. Absorption and fluorescence were normalized by peak intensities. (E) Brightness of FPs expressed in mammalian cells. Cells expressing RFP were imaged with a fluorescence microscope. Integrated fluorescence intensities from each cell were quantified. Data from multiple cells were presented as violin plots (mARs1, 208 cells; AR1.0, 177 cells; mCherry, 224 cells; mScarlet-1, 196 cells; DsRed, 178 cells). White diamonds indicate median values. (F) Visualization of cellular structures using mARs1 in mammalian cells (inverted grayscale). mARs1-tagged proteins were expressed in cultured HeLa cells. The following proteins/tags were fused to mARs1: Nucleus, 3x nuclear localization signal (nls) at the N terminus; mitochondria, 4x mitochondria targeting signal (mts) at the N terminus; endoplasmic reticulum, calreticulin signal at the N terminus and KDEL at the C terminus; Golgi apparatus, Gigantin at the C terminus; late endosome, Rab7 at the C terminus; lysosome, Lamp1 at the N terminus; plasma membrane, Lck(1-7) at the N terminus; actin filament, Lifeact at the N terminus; focal adhesion, paxillin at the C terminus.

green fluorescence, suggesting that RFP evolved from GFP (37, 38). Robust GFP fluorescence despite mutations around the fluorophore also suggests that the ancestor of FPs had a GFP-type fluorophore. Most RFPs, including DsRed, eqFP578, and eqFP611, form an evolutionarily distant group from other FPs, although there are a few exceptional RFPs, such as zoan-2RFP and anm2CP (Fig. 1A). These observations raise a possibility that the reaction leading to DsRed-type fluorophore formation is accomplished only in limited sequence spaces. However, this study clearly shows that the autocatalytic reaction of DsRed-type fluorophore formation is possible in sequence spaces distant from the natural RFPs. We would like to emphasize that the QY of AR1.0 is the highest class among FPs with the maximum fluorescence wavelength exceeding 600 nm. Thus, the further engineering of GFP may be a promising route to develop RFPs with superior properties.

The previously reported 3D structures of GFP and RFP contain many differences unrelated to the formation of fluorophores due to low sequence similarity. The 3D structures of AG and AR1.0, which differ by only 29 residues, provided structural insight into the formation of red fluorophores. First, significant structural changes were almost exclusively confined to regions proximal to the fluorophore, indicating that the environment surrounding the fluorophore almost entirely dictates the autocatalytic reaction to form the red fluorophore. Second, substitution of multiple residues around the fluorophore resulted in a significant reorganization of the hydrophilic networks including hydrogen bonds and electrostatic interactions (*SI Appendix*, Fig. S6), suggesting that conversion of GFP to RFP requires coordinated changes in multiple residues, not a single critical amino acid. The observation that single mutations to amino acid residues involved in the interaction network in AR1.0 largely, but not wholly, prevented the formation of the red fluorophore also supports this idea (*SI Appendix*, Fig. S4). The requirement of coordinated mutations differs from BFP, CFP, and YFP, which could be generated from avGFP essentially by single amino acid substitutions (12, 14).

DsRed-type fluorophores contain a unique double bond between N and C α of the residue corresponding to Met62 of AR1.0 (Figs. 1B and 3A). Since the double bond formation depends on molecular oxygen, a space for oxygen molecules is necessary for the reaction near the fluorophore. AG has only a narrow space for a single water molecule in the vicinity of the 62nd C α and no space to accommodate an oxygen molecule (Fig. 3G and *SI Appendix*, Fig. S6). In contrast, AR1.0 has a large cavity, which is created mainly by R66K and T69I mutations, below the imidazolinone ring. The cavity contains a water cluster consisting of four water molecules. Some RFPs, such as eqFP611, eqFP578, and mScarlet, also have a similar cavity containing multiple water molecules. Interestingly, the crystal structures of eqFP611 (PDB ID: 1UIS) (30) and eqFP578 (PDB ID: 3PIB) (31) contain acetic acid and glycerol in the cavity, which are included in the crystallization solution, respectively. These observations strongly suggest that the cavity below the fluorophore is accessible to the entry of various molecules from outside the protein. Therefore, it is likely that the oxygen molecule required for the double bond formation at the 62nd position is accommodated in the cavity. The cavity of DsRed is, by contrast, slightly farther from the fluorophore and possesses only one water molecule near the fluorophore (Fig. 3G). This structural property might explain the slow maturation of DsRed (33).

The reversion of any one of the amino acid residues at positions 65, 66, 69, 87, 105, and 107 in AR1.0 markedly reduced absorption and fluorescence of the red fluorophore (*SI Appendix*, Figs. S4 and S5). These residues are involved in the hydrogen bonding network stabilizing the conformation of the segment following

the fluorophore (residues 65 to 67). The segment includes the serine residue (Ser65) corresponding to Ser69 of DsRed, which has been implicated in oxidation for the acylimine formation for DsRed (28). AR1.0 and AG show a significant structural difference in the segment as well as the hydrogen bonding network surrounding the segment. Therefore, the segment conformation and the interaction network should be critical for the formation of the red fluorophore.

Previous studies of RFPs demonstrated that proper positioning of the first residue of the fluorophore is important for acylimine formation (28). Gln66, the first residue of the DsRed fluorophore, is fixed by hydrogen bonds with Gln213 and Asn42. The corresponding residues of DsRed Q66M and eqFP611 are methionine but are also fixed by hydrogen bonds (29, 32). Although Met62, the corresponding residue of AR1.0, has no hydrogen bonding with other residues, it is fixed by hydrophobic interactions. The side chain of Met62 is tightly surrounded by the aliphatic side-chain arms of Met40, Gln38, Ile195, Gln209, and Glu211. AR1.0 was produced by introducing the A40M, L195I, and I118V mutations into AR0.6, suggesting the importance of the tight hydrophobic interaction. Moreover, the M40L mutation of AR1.0 slightly decreased the absorbance at 571 nm (*SI Appendix*, Fig. S4). Taken together, the proper positioning of Met62 is important for acylimine formation of AR1.0.

The direct interaction of phenolic oxygen with the positively charged N ζ of Lys159 (Fig. 3B) indicates that the phenolic oxygen is deprotonated, and thus, the fluorophore of AR1.0 is anionic. No such electrostatic interaction exists in AG because the 159th residue of AG is methionine. No significant structural difference was found in the hydrogen bonding network around the phenolic oxygen of AR1.0 and AG other than the electrostatic interaction (Fig. 3B). The K159M mutation in AR1.0 reduced the absorbance at 570 nm and caused the 510 nm emission (*SI Appendix*, Figs. S4 and S5), indicating suppression of the red fluorophore formation. These facts suggest that the acylimine of the AR1.0 fluorophore is formed through an anionic intermediate, and Lys159 stabilizes the negatively charged intermediate (Fig. 3B and *SI Appendix*, Fig. S6).

To find novel RFPs, researchers had to prospect in nature. This work, however, has demonstrated that it is possible to expand the repertoire of novel RFPs by introducing a limited number of mutations into existing GFPs. If the sequence diversity of RFPs can be increased by converting GFP to RFP, it would provide a new basis for the development of long-wavelength FPs with superior properties. Recent advances in *in silico* protein design have made it possible to develop *de novo* proteins with diverse properties, including a novel β -barrel protein that is able to incorporate DFHBI, a fluorogenic compound that resembles the GFP fluorophore (39). Although *de novo* proteins that autocatalytically form fluorophores like FPs have not been reported so far, our results may provide essential structural insights for the future *in silico* design of *de novo* RFP-like proteins.

Materials and Methods

Plasmid Construction and Mutagenesis. cDNAs for AG, AzamiRed0.1, mcavGFP, and mcavRed0.1 were synthesized by GenScript and cloned between Xhol and HindIII sites of the pRSET_B vector (Thermo Scientific) to obtain bacterial expression plasmids for the FPs. Site-directed mutagenesis was performed by a PCR-based method. First, PCR was conducted with two mutagenic primers, which contained mutation(s) and had a 15-base overlap, using a bacterial expression plasmid as a template and KOD One DNA polymerase (TOYOBO). After purification with a spin-column (Promega), the amplified products were circularized using In-Fusion (Clontech), followed by transformation of *Escherichia coli* strain DH5 α .

Random mutagenesis was performed by amplifying the respective FP cDNA using a GeneMorph II random mutagenesis kit (Agilent) using a T7 promoter primer (TAATACGACTCACTATAGGG) and a T7 terminator primer (GCTAGTTATTGCTCAGCGG) according to the manufacturer's protocol. The amplified product was cloned between Xhol and HindIII sites of the pRSET_B vector, followed by transformation of *E. coli* strain JM109(DE3) (Promega). The plates were incubated at 37 °C overnight, then 4 °C for several days. The fluorescence of the colonies was evaluated by microscopy or by eyes. The brightest colonies were selected and cultured to obtain the plasmids for the next round of mutagenesis.

Plasmids for overexpression for crystallization of AG and AzamiRed1.0 were constructed by cloning PCR-amplified DNA for the respective FPs, with incorporated 5' NdeI sites and 3' BamHI sites, into the NdeI-BamHI sites of the pET15b vector (Novagen). The vector includes an N-terminal His₆-tag for affinity purification of expressed proteins.

Plasmids for expression of AR1.0 and mARs1 in the cytoplasm of mammalian cells were constructed by transferring the Xhol-HindIII FP cassette from pRSET_B-AR1.0 and pRSET_B-mARs1, respectively, to the Xhol and HindIII cloning sites in pcDNA3.1(−) vector (Thermo Scientific). To construct plasmids for localization of mARs1 in the nucleus, mitochondrial matrix, plasma membrane, endoplasmic reticulum, and Golgi apparatus, a PCR-amplified fragment carrying mARs1 was inserted between the PinAI and Bsp1407I sites of 3xNLS-mScarlet-I, 4xMTS-mScarlet-I, Lck-mScarlet-I, ER-mScarlet-I, and mScarlet-I-Giantin(40) to obtain 3xNLS-mARs1, 4xMTS-mARs1, Lck-mARs1, ER-mARs1, and mARs1-Giantin, respectively. For expression of mARs1 in late endosomes and focal adhesions, NheI-Xhol and NheI-BglII fragments excised from mARs1-Giantin were inserted between the NheI and Xhol sites of mScarlet-I-Rab7 and between the NheI and BglII sites of mScarlet-I-PXN (40) to obtain mARs1-Rab7 and mARs1-PXN, respectively. To construct plasmids for expression of mARs1 in lysosomes and microtubule tips, a BshTl-Bsp1407I fragment from 3xNLS-mARs1 was cloned into the BshTl and Bsp1407I sites of EB3-mScarlet-I and Lamp1-mScarlet-I (40) to obtain EB3-mARs1 and Lamp1-mARs1, respectively. For expression of mARs1 in actin filaments, a BshI-NotI fragment carrying mARs1 from 3xNLS-mARs1 was inserted into the BshI and NotI sites of LifeAct-mTurquoise2 (41) to obtain LifeAct-mARs1.

Plasmids for investigating the maturation rate in mammalian cells were constructed as follows. Synthetic cDNA for mTurquoise2-P2A (Integrated DNA Technologies, Inc) was inserted between the NheI and Xhol sites of pcDNA3.1(−) using In-Fusion HD to obtain pcDNA-mTurquoise2-P2A-Xhol-HindIII DNA fragments, covering AR1.0, AR1.1, AR1.2, and AR1.3 from the respective pRSET_B-AzamiRed plasmids, were cloned into the Xhol and HindIII sites of pcDNA-mTurquoise2-P2A to obtain pcDNA-mTurquoise2-P2A-AR1.0, pcDNA-mTurquoise2-P2A-AR1.1, pcDNA-mTurquoise2-P2A-AR1.2, and pcDNA-mTurquoise2-P2A-AR1.3.

Plasmids for photobleaching and brightness assay, pcDNA-DsRed, and pcDNA-mCherry were constructed by transferring a cDNA for FPs PCR-amplified from pCAG-DsRed (42) and pmCherry-N1 (Clontech), respectively, to the Xhol and HindIII sites in pcDNA3.1(−) vector. pcDNA-mScarlet-I was synthesized by Genscript.

Plasmids for OSER assay, CytERM-mARs1, and CytERM-mCherry were constructed by transferring a BshTl-Bsp1407I fragment from 3xNLS-mARs1 and pmCherry-N1 into the BshTl and Bsp1407I sites of CytERM-mAvic1 (43), respectively. CytTERM-dTomato was constructed by fusing a dTomato cDNA fragment (synthesized by Eurofins Genomics) into the BshTl and Bsp1407I sites of CytERM-mAvic1.

Protein Expression and Purification of FPs for Spectral Characterization. Colonies of *E. coli* strain JM109(DE3) carrying an FP plasmid were transferred to 50 mL 2 × YT medium and cultured at 24 °C for 2 d with vigorous shaking. Cells were collected by centrifugation, suspended in buffer A [100 mM sodium phosphate (pH 8.0) and 200 mM sodium chloride] containing a protease inhibitor cocktail (Roche), and then disrupted by sonication. Cell-free extracts, obtained by centrifugation at 12,000 × g for 60 min, were applied to a Talon Sepharose columns (Clontech) pre-equilibrated with buffer A. After washing the column with buffer A containing 5 mM imidazole, the his-tagged proteins were eluted by increasing the imidazole concentration to 150 mM. Fractions containing FP were concentrated, followed by application to a Superdex 200 Increase 10/300 GL gel-filtration column (GE Healthcare) pre-equilibrated with buffer B [20 mM Tris-HCl (pH 8.0) and 150 mM sodium chloride]. Fractions containing FP were collected and stored at 4 °C.

Estimation of Molecular Weight by Gel-filtration Chromatography.

Molecular weights of FPs were analyzed by applying a 100-μL purified sample to a gel-filtration column (Superdex200 increase 10/300 GL, GE Healthcare) equilibrated with buffer B. The flow rate was 0.5 mL min^{−1}. Thyroglobulin, γ-globulin, ovalbumin, and myoglobin (Bio-Rad) were used as molecular weight standards.

Absorption and Fluorescence Spectroscopy. The absorption and fluorescence spectra of purified FPs were investigated in buffer B using a V-730bio UV-visible spectrometer and an FP-7300 spectrophotometer (Jasco, Tokyo, Japan), respectively. Molar extinction coefficients of FPs at absorption maximum were calculated using those values at 280 nm, which can be estimated from the numbers of tryptophane and tyrosine residues and their extinction coefficients (5,690 M^{−1} cm^{−1} for tryptophane, and 1,280 M^{−1} cm^{−1} for tyrosine) (44). For estimating the fluorescent QY of FPs, absorption of FPs was measured at 550 nm. Then, the fluorescence emission spectra of FPs, diluted with buffer B, were measured from 555 nm to 800 nm with 550 nm excitation light, using an FP-7300 spectrophotometer (JASCO). The integrated fluorescence intensity was used to calculate the QY of FPs. Rhodamine 101 (Sigma-Aldrich, QY = 0.91) was used as a standard.

To investigate the pH sensitivity of FPs, aliquots of purified samples were diluted 50-fold in 100 mM buffer with a defined pH (citrate, pH 3.5 to 4.5; MES-NaOH, pH 5.0 to 6.5; HEPES-NaOH, pH 7.0 to 8.0; Bicine-NaOH, pH 9.0; CHES-NaOH, pH 10.0). Immediately after dilution, the fluorescence intensity was measured using a microplate reader (SpectraMax M5, Molecular Devices).

Protein Expression and Purification of FPs for X-ray Crystallography.

E. coli strain BL21(DE3), carrying respective pET15b-FP plasmids, was cultured in 2 LLB medium (Lennox; Nacalai Tesque) with ampicillin (50 μg/mL) at 37 °C to an optical density at 600 nm of 0.4 to 0.7. Then, protein expression was induced by the addition of isopropyl-β-D-thiogalactopyranoside (0.5 mM), and the cultivation was continued overnight at 18 °C. The cells were pelleted by centrifugation at 6,700 × g for 10 min, resuspended in buffer B containing cOmplete EDTA-free (Roche) and lysozyme (Wako), and lysed by sonication on ice (for 8 min, with cycles of 1-s pulse and 2-s pause). The sonicated solution was ultracentrifuged at 100,000 × g at 4 °C for 10 min to remove cell debris, membrane, and macromolecular complexes. The supernatant was loaded onto 5 mL of Ni-NTA agarose (Qiagen) filled in an open column at 4 °C, the protein-bound agarose was washed three times with buffer B containing 50 mM imidazole, and the proteins were subsequently eluted with buffer B containing 500 mM imidazole. The eluted protein solution was mixed with thrombin to cleave the N-terminal His₆-tag and dialyzed against buffer B at 4 °C for 16 to 18 h. The dialyzed protein solution was loaded onto a column filled with fresh Ni-NTA agarose, and the flow-through was collected to remove uncleaved proteins. The protein was concentrated using an Amicon Ultra 10 K device (Merck Millipore), loaded on the size-exclusion column (Superdex 200 Increase 10/300 GL [Cytiva]) equilibrated with buffer C [20 mM Tris-HCl (pH 8.0), 200 mM sodium chloride], and eluted using ÄKTA explorer (GE Healthcare). The peak fraction was collected and concentrated using an Amicon Ultra 10 K device (Millipore). The purity of the proteins was monitored by sodium dodecyl sulfate-polyacrylamide gel electrophoresis.

X-ray Crystallography. Initial crystallization screening of AG and AzamiRed1.0 was performed using the following screening kits: Wizard Classic I and II, Wizard Cryo I and II (Rigaku), and Crystal Screen I and II (Hampton Research), and then, the conditions were optimized. The crystallization drops were prepared at 20 °C using the sitting drop vapor diffusion method by mixing 0.5 μL of the protein solution and 0.5 μL of the reservoir solution. The protein concentrations of AG and AzamiRed1.0 were 4.19 mg/mL and 5.98 mg/mL, respectively. Crystals appeared within 2 wk. The crystals used for the X-ray diffraction data collection were grown by using a reservoir solution containing 100 mM HEPES (pH 7.5), 200 mM MgCl₂, and 20% (w/v) PEG-8000 for AG, and 100 mM Tris-HCl (pH 7.0), 20% (w/v) PEG-3350, and 200 mM Ca(OAc)₂ for AzamiRed1.0.

X-ray diffraction data were collected at the synchrotron beamline BL41XU in SPring-8 facility (Sayo-cho, Japan). The crystals were flash-frozen in liquid nitrogen and mounted in a 100 K nitrogen stream for the X-ray diffraction data collection. Prior to cooling, the crystals of AG and AzamiRed1.0 were transferred to a cryoprotecting solution containing 20% glycerol and 80% reservoir solution. The diffraction data were processed with MOSFLM (45) and scaled with AIMLESS (46).

The initial phase was calculated by the molecular replacement method with the Phaser-MR in Phenix (47) program using the structure of AG (PDB ID: 6CIU)

as a search model. The atomic model building and refinement were carried out using *Coot* (48) and *Phenix*, respectively. The statistics of the diffraction data and the refinement are summarized in *SI Appendix, Table S1*. The atomic coordinates and structure factors (codes 8I4J and 8I4K) have been deposited in the protein data bank.

Mammalian Cell Culture and Microscopy. The HeLa cell line was a kind gift from Prof. Shin Yonehara, and cells were grown in Dulbecco's modified Eagle's medium (1 g/L glucose; Nacalai Tesque) supplemented with 10% fetal bovine serum (Nichirei). For investigating the localization of mARs1 constructs, the cells were transfected with the respective mARs1 plasmid using PEI-Max (Polysciences, Inc.) as described previously (49). One day after transfection, the cells were trypsinized and plated on a 35-mm glass-bottom dish (0.16 to 0.19 mm thick; MatTek) coated with type II collagen (Nitta gelatin). The following day, the cells were visualized with a Ti-E inverted microscope (Nikon, Tokyo, Japan) using a Plan Apo 60 \times , 1.40 numerical aperture, dry objective lens (Nikon), and an mCherry filter set (an FF01-562/40 excitation filter, an FF593-Di02 dichroic mirror, and an FF01-641/75 emission filter; Semrock). Cells were illuminated using an LED light source (X-cite Xylis, Excelitas), and fluorescence images of the cells were imaged using a Zyla4.2 scientific CMOS camera (Andor Technologies). The microscope system was controlled by NIS-Elements software (Nikon). Microscope images were processed and analyzed using FIJI software.

To investigate the maturation rate of AzamiRed proteins, HeLa cells cultured in a glass bottom dish with 4 compartments (Greiner Bio-One) were transfected with the pcDNA-mTurquoise2-P2A-AR plasmids using PEI-Max. Then, fluorescence images from the cells were taken every 20 min with a Ti-E inverted microscope (Nikon, Tokyo, Japan) using a Plan Apo 20 \times , 0.75 numerical aperture, dry objective lens (Nikon). Cells were maintained on a microscope at 37 °C with a continuous supply of a 95% air and 5% carbon dioxide mixture by way of a stage-top incubator (Tokai Hit). Cells were illuminated with 440 nm and 555 nm light for mTurquoise2 and AzamiRed, respectively, using a SpectraX Light Engine (Lumencor). For imaging mTurquoise2, an FF458-Di02 dichroic mirror and an FF02-483/32 emission filter (Semrock) were used. For imaging AzamiRed, an FF593-Di03 dichroic mirror and an FF01-593/LP emission filter (Semrock) were used. Fluorescence emissions from cells were imaged using a Zyla4.2 scientific CMOS camera (Andor Technologies). After cell division, the cells start to produce mTurquoise2 and an AzamiRed variant at the same rate

because the two FPs are connected by a P2A sequence. The fluorescence intensity from each cell was plotted with the x axis as elapsed time. A phase with linearly increasing fluorescence intensity was selected for linear fitting, and the time on the x-intercept was defined as the onset of FP maturation. Then, the difference in maturation onset time between mTurquoise2 and an AzamiRed variant was calculated for each cell.

To investigate photobleaching properties and brightness of FPs inside cells, HeLa cells were transfected with a pcDNA3.1(−) vector harboring an FP cDNA using PEI-Max. One day after transfection, the cells were trypsinized and plated on a 35 mm glass-bottom dish (Mattek) coated with Type I-C collagen (Nitta gelatin). The cells were visualized with a Ti-E inverted microscope (Nikon, Tokyo, Japan) using a Plan Apo 20 \times , 0.75 numerical aperture, dry objective lens (Nikon), at 44 h after transfection. An FF01-562/40 excitation filter, an FF593-Di02 dichroic mirror, and an FF01-641/75 emission filter (Semrock) were used. In the photobleaching experiments, the cells were continuously illuminated using an X-cite XYLIS LED light source (EXCELITAS Technologies) without neutral density filters and were imaged every 10 s. The light intensity at the specimen was 3.1 W cm $^{-2}$.

To conduct an OSER assay, HeLa cells cultured on a 35-mm glass-bottom dish were transfected with either CytTERM-mARs1, CytTERM-mCherry, or CytTERM-dTomato plasmid. Images were collected with an SP8 confocal microscope (Leica Microsystems) at 16 to 24 h after transfection. Quantification of fluorescence intensities of OSER structures and nuclear envelopes was performed using FIJI software as described previously (34).

Data, Materials, and Software Availability. Protein structure coordinates data have been deposited in Protein Data Bank 8I4J (50) and 8I4K (51).

ACKNOWLEDGMENTS. We are indebted to Shigehiko Hayashi for valuable discussions and Akira Kakizuka for experimental instruments. We thank the beamline staff at SPring-8 for technical help in data collection and James Alan Hejna for critical assessment of this manuscript. The mScarlet-I plasmids, pLifeAct-mTurquoise2, pCAG-DsRed, and CytTERM-mAvic1 were obtained from Joachim Goedhart, Dorus Gadella, Connie Cepko, and Nathan Shaner, respectively, via Addgene. This work was supported by JSPS KAKENHI Grant Numbers JP20H03221 and JP20H05757 and by Kyoto University Live Imaging Center.

1. R. Y. Tsien, The green fluorescent protein. *Annu. Rev. Biochem.* **67**, 509–544 (1998).
2. N. C. Shaner, G. H. Patterson, M. W. Davidson, Advances in fluorescent protein technology. *J. Cell Sci.* **120**, 4247–4260 (2007).
3. W. B. Frommer, M. W. Davidson, R. E. Campbell, Genetically encoded biosensors based on engineered fluorescent proteins. *Chem. Soc. Rev.* **38**, 2833–2841 (2009).
4. J. Zheng, N. Guo, A. Wagner, Selection enhances protein evolvability by increasing mutational robustness and foldability. *Science* **370**, eabb5962 (2020).
5. E. De Zitter *et al.*, Structure-function dataset reveals environment effects within a fluorescent protein model system. *Angew. Chem. Weinheim Bergstr. Ger.* **133**, 10161–10169 (2021).
6. L. A. Gross, G. S. Baird, R. C. Hoffman, K. K. Baldrige, R. Y. Tsien, The structure of the chromophore within DsRed, a red fluorescent protein from coral. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11990–11995 (2000).
7. S. Pletnev, F. V. Subach, Z. Dauter, A. Wlodawer, V. V. Verkusha, Understanding blue-to-red conversion in monomeric fluorescent timers and hydrolytic degradation of their chromophores. *J. Am. Chem. Soc.* **132**, 2243–2253 (2010).
8. R. L. Strack, D. E. Strongin, L. Mets, B. S. Gluck, R. J. Keenan, Chromophore formation in DsRed occurs by a branched pathway. *J. Am. Chem. Soc.* **132**, 8496–8505 (2010).
9. A. Miyawaki, D. M. Shcherbakova, V. V. Verkusha, Red fluorescent proteins: Chromophore formation and cellular applications. *Curr. Opin. Struct. Biol.* **22**, 679–688 (2012).
10. K. Katoh, J. Rozewicki, K. D. Yamada, MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **20**, 1160–1166 (2019).
11. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
12. R. Heim, D. C. Prasher, R. Y. Tsien, Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 12501–12504 (1994).
13. W. Tomosugi *et al.*, An ultramarine fluorescent protein with increased photostability and pH insensitivity. *Nat. Methods* **6**, 351–353 (2009).
14. M. Ormø *et al.*, Crystal structure of the Aequorea victoria green fluorescent protein. *Science* **273**, 1392–1395 (1996).
15. H. Tsutsui, S. Karasawa, H. Shimizu, N. Nukina, A. Miyawaki, Semi-rational engineering of a coral fluorescent protein into an efficient highlighter. *EMBO Rep.* **6**, 233–238 (2005).
16. N. C. Shaner *et al.*, Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat. Biotechnol.* **22**, 1567–1572 (2004).
17. D. Shcherbo *et al.*, Bright far-red fluorescent protein for whole-body imaging. *Nat. Methods* **4**, 741–746 (2007).
18. S. Kredel *et al.*, mRuby, a bright monomeric red fluorescent protein for labeling of subcellular structures. *PLOS One* **4**, e4391 (2009).
19. M. Z. Lin *et al.*, Autofluorescent proteins with excitation in the optical window for intravital imaging in mammals. *Chem. Biol.* **16**, 1169–1179 (2009).
20. S. Karasawa, T. Araki, M. Yamamoto-Hino, A. Miyawaki, A green-emitting fluorescent protein from Galaxeedae coral and its monomeric version for use in fluorescent labeling. *J. Biol. Chem.* **278**, 34167–34171 (2003).
21. T. J. Lambert, FPbase: A community-editable fluorescent protein database. *Nat. Methods* **16**, 277–278 (2019).
22. T. Suzuki *et al.*, Development of cysteine-free fluorescent proteins for the oxidative environment. *PLoS One* **7**, e37551 (2012).
23. J. Wiedenmann *et al.*, A far-red fluorescent protein with fast maturation and reduced oligomerization tendency from Entacmaea quadricolor (Anthozoa, Actiniaria). *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11646–11651 (2002).
24. E. M. Merzlyak *et al.*, Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nat. Methods* **4**, 555 (2007).
25. D. S. Bindels *et al.*, mScarlet: A bright monomeric red fluorescent protein for cellular imaging. *Nat. Methods* **14**, 53–56 (2017).
26. C. Oi *et al.*, A threonine zipper that mediates protein–protein interactions: Structure and prediction. *Protein Sci.* **27**, 1969–1977 (2018).
27. M. A. Wall, M. Socolich, R. Ranganathan, The structural basis for red fluorescence in the tetrameric GFP homolog DsRed. *Nat. Struct. Biol.* **7**, 1133–1138 (2000).
28. D. Yarborough, R. M. Wachtler, K. Kallio, M. V. Matz, S. J. Remington, Refined crystal structure of DsRed, a red fluorescent protein from coral, at 2.0-Å resolution. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 462–467 (2001).
29. J. L. Tubbs, J. A. Tainer, E. D. Getzoff, 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* **44**, 9833–9840 (2005).
30. J. Petersen *et al.*, The 2.0-Å crystal structure of eqFP611, a far red fluorescent protein from the sea anemone Entacmaea quadricolor. *J. Biol. Chem.* **278**, 44626–44631 (2003).
31. N. V. Pletnev *et al.*, Crystallographic study of red fluorescent protein eqFP578 and its far-red variant Katushka reveals opposite pH-induced isomerization of chromophore. *Protein Sci.* **20**, 1265–1274 (2011).
32. K. Nienhaus, H. Nar, R. Heilker, J. Wiedenmann, G. U. Nienhaus, Trans-cis isomerization is responsible for the red-shifted fluorescence in variants of the red fluorescent protein eqFP611. *J. Am. Chem. Soc.* **130**, 12578–12579 (2008).
33. R. E. Campbell *et al.*, A monomeric red fluorescent protein. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7877–7882 (2002).
34. L. M. Costantini, M. Fossati, M. Francolini, E. L. Snapp, Assessing the tendency of fluorescent proteins to oligomerize under physiologic conditions. *Traffic* **13**, 643–649 (2012).

35. M. Hirano *et al.*, A highly photostable and bright green fluorescent protein. *Nat. Biotechnol.* **40**, 1132–1142 (2022), 10.1038/s41587-022-01278-2.
36. A. S. Mishin *et al.*, The first mutant of the *Aequorea victoria* green fluorescent protein that forms a red chromophore. *Biochemistry* **47**, 4666–4673 (2008).
37. J. A. Ugalde, B. S. W. Chang, M. V. Matz, Evolution of coral pigments recreated. *Science* **305**, 1433–1433 (2004).
38. N. O. Alieva *et al.*, Diversity and evolution of coral fluorescent proteins. *PLoS One* **3**, e2680 (2008).
39. J. Dou *et al.*, De novo design of a fluorescence-activating β -barrel. *Nature* **561**, 485–491 (2018).
40. A. O. Cherkova *et al.*, Robust and bright genetically encoded fluorescent markers for highlighting structures and compartments in mammalian cells. *bioRxiv* [Preprint] (2017). <https://doi.org/10.1101/160374> (Accessed 20 April 2021).
41. J. Goedhart *et al.*, Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. *Nat. Commun.* **3**, 751 (2012).
42. T. Matsuda, C. L. Cepko, Electroporation and RNA interference in the rodent retina *in vivo* and *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 16–22 (2004).
43. G. G. Lambert, H. Depernet, G. Gotthard, D. T. Schultz, *Aequorea's secrets revealed: New fluorescent proteins with unique properties for bioimaging and biosensing*. *PLoS Biol.* **18**, e3000936 (2020).
44. S. C. Gill, P. H. von Hippel, Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319–326 (1989).
45. T. G. G. Battye, L. Kontogiannis, O. Johnson, H. R. Powell, A. G. W. Leslie, iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 271–281 (2011).
46. M. D. Winn *et al.*, Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
47. P. D. Adams *et al.*, PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
48. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
49. G. Morciano *et al.*, Measurement of ATP concentrations in mitochondria of living cells using luminescence and fluorescence approaches. *Methods Cell Biol.* **155**, 199–219 (2020).
50. S. Otsubo, N. Takekawa, H. Imamura, K. Imada, Structure of wild-type Azami Green from *Galaxea fascicularis*. Protein Data Bank. <https://www.rcsb.org/structure/8I4J>. Deposited 19 January 2023.
51. S. Otsubo, N. Takekawa, H. Imamura, K. Imada, Structure of Azami Red 1.0 engineered from Azami Green. Protein Data Bank. <https://www.rcsb.org/structure/8I4K>. Deposited 19 January 2023.