



(51) International Patent Classification:

C07K 14/00 (2006.01) *C12N 15/63* (2006.01)
C12N 15/09 (2006.01) *G01N 21/64* (2006.01)

(21) International Application Number:

PCT/US2015/025540

(22) International Filing Date:

13 April 2015 (13.04.2015)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/979,436 14 April 2014 (14.04.2014) US

(71) Applicant: PRESIDENT AND FELLOWS OF HARVARD COLLEGE [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US).

(72) Inventors: ZHUANG, Xiaowei; 5 Luongo Farm Lane, Lexington, MA 02421 (US). WANG, Siyuan; 28 Fernald Drive, Apt. #31, Cambridge, MA 02138 (US). MOFFITT, Jeffrey, R.; 175 Beacon Street, Apt. #108, Somerville, MA 0243 (US).

(74) Agent: CHEN, Tani; Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210-2206 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: PHOTOCONVERTIBLE FLUORESCENT PROTEINS

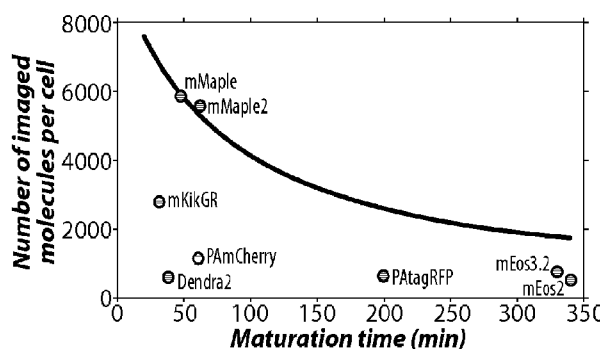


Fig. 8

(57) Abstract: The present invention generally relates to fluorescent proteins, for various uses such as for super-resolution microscopy. In one set of embodiments, the protein is mMAPLE2 or mMAPLE3. In another set of embodiments, the protein comprises two or substitutions in mMAPLE. In addition, in certain embodiments, the present invention is generally directed to systems and methods for making or using such proteins, kits involving such proteins, nucleic acids encoding such proteins, vectors or plasmids comprising such proteins or nucleic acids, or cells or organisms comprising such proteins or nucleic acids, as well as fragments, homologs, analogs, derivatives, or functionally equivalent compositions.

PHOTOCONVERTIBLE FLUORESCENT PROTEINS

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Patent Application Serial
5 No. 61/979,436, filed April 14, 2014, entitled "Photoconvertible Fluorescent Proteins,"
by Zhuang, *et al.*, incorporated herein by reference.

GOVERNMENT FUNDING

This invention was made with government support under Grant Nos. GM096450
and GM068518 awarded by National Institute of Health. The government has certain
10 rights in the invention.

FIELD

The present invention generally relates to fluorescent proteins, for various uses
such as for super-resolution microscopy.

BACKGROUND

15 Photoactivatable fluorescent proteins (PAFPs) are important probes, for example,
for super-resolution microscopy, which allows the spatial organization of proteins in
living cells to be probed with sub-diffraction-limit resolution. Super-resolution
microscopy, in general, is defined as optical microscopy at resolutions that exceed the
resolution limit set by the diffraction limit of light. Recent advances in super-resolution
20 microscopy include stochastic optical reconstruction microscopy (STORM), near-field
scanning optical microscopy (NSOM), stimulated emission depletion (STED), ground
state depletion microscopy (GSD), reversible saturable optical linear fluorescence
transition (RESOLFT), saturated structured-illumination microscopy (SSIM), and
photoactivated localization microscopy (PALM).

25

SUMMARY

The present invention generally relates to fluorescent proteins, for various uses
such as for super-resolution microscopy. The subject matter of the present invention
involves, in some cases, interrelated products, alternative solutions to a particular
problem, and/or a plurality of different uses of one or more systems and/or articles.

30

In one set of embodiments, the present invention is directed to a composition
comprising mMAPLE2. In another set of embodiments, the present invention is directed
to a composition comprising mMAPLE3. In another set of embodiments, the present
invention is directed to a composition comprising SEQ ID NO: 2. In another set of

embodiments, the present invention is directed to a composition comprising positions 2-237 of SEQ ID NO: 2.

In another set of embodiments, the present invention is directed to a composition comprising SEQ ID NO: 1 with the provisos that position 111 is any amino acid residue except I and position 198 is any amino acid residue except Y. In another set of
5 embodiments, the present invention is directed to a composition comprising positions 2-237 of SEQ ID NO: 1 with the provisos that position 111 is any amino acid residue except I and position 198 is any amino acid residue except Y.

In another set of embodiments, the present invention is directed to a composition
10 comprising positions 17-225 of SEQ ID NO: 2. In another set of embodiments, the present invention is directed to a composition comprising positions 17-225 of SEQ ID NO: 1 with the provisos that position 111 is any amino acid residue except I and position 198 is any amino acid residue except Y.

In another set of embodiments, the present invention is directed to a composition
15 comprising SEQ ID NO: 3. In another set of embodiments, the present invention is directed to a composition comprising positions 2-237 of SEQ ID NO: 3.

In another set of embodiments, the present invention is directed to a composition comprising SEQ ID NO: 1 with the provisos that position 82 is any amino acid residue except E, position 83 is any amino acid residue except D, position 111 is any amino acid
20 residue except I, position 197 is any amino acid residue except D, and position 198 is any amino acid residue except Y.

In another set of embodiments, the present invention is directed to a composition comprising positions 2-237 of SEQ ID NO: 1 with the provisos that position 82 is any amino acid residue except E, position 83 is any amino acid residue except D, position
25 111 is any amino acid residue except I, position 197 is any amino acid residue except D, and position 198 is any amino acid residue except Y.

In another set of embodiments, the present invention is directed to a composition comprising positions 17-225 of SEQ ID NO: 3.

In one set of embodiments, the present invention is generally directed to a nucleic
30 acid, at least a portion of which encodes the amino acid sequence of mMAPLE2. In another set of embodiments, the present invention is generally directed to a nucleic acid, at least a portion of which encodes the amino acid sequence of mMAPLE3. In another set of embodiments, the present invention is generally directed to a nucleic acid, at least

a portion of which encodes the amino acid sequence of SEQ ID NO: 2. In another set of embodiments, the present invention is generally directed to a nucleic acid, at least a portion of which encodes positions 2-237 of the amino acid sequence of SEQ ID NO: 2.

In another set of embodiments, the present invention is generally directed to a nucleic acid, at least a portion of which encodes SEQ ID NO: 1 with the provisos that position 111 is any amino acid residue except I and position 198 is any amino acid residue except Y. In another set of embodiments, the present invention is generally directed to a nucleic acid, at least a portion of which encodes positions 2-237 of SEQ ID NO: 1 with the provisos that position 111 is any amino acid residue except I and position 198 is any amino acid residue except Y.

In another set of embodiments, the present invention is generally directed to a nucleic acid, at least a portion of which encodes the amino acid sequence of SEQ ID NO: 3. In another set of embodiments, the present invention is generally directed to a nucleic acid, at least a portion of which encodes positions 2-237 of the amino acid sequence of SEQ ID NO: 3.

In another set of embodiments, the present invention is generally directed to a nucleic acid, at least a portion of which encodes SEQ ID NO: 1 with the provisos that position 82 is any amino acid residue except E, position 83 is any amino acid residue except D, position 111 is any amino acid residue except I, position 197 is any amino acid residue except D, and position 198 is any amino acid residue except Y.

In another set of embodiments, the present invention is generally directed to a nucleic acid, at least a portion of which encodes positions 2-237 of SEQ ID NO: 1 with the provisos that position 82 is any amino acid residue except E, position 83 is any amino acid residue except D, position 111 is any amino acid residue except I, position 197 is any amino acid residue except D, and position 198 is any amino acid residue except Y.

In another set of embodiments, the present invention is generally directed to a nucleic acid encoding an amino acid sequence that has a sequence identity of at least about 90% with SEQ ID NO: 2. In another set of embodiments, the present invention is generally directed to a nucleic acid encoding an amino acid sequence that has a sequence identity of at least about 95% with SEQ ID NO: 2. In another set of embodiments, the present invention is generally directed to a nucleic acid encoding an amino acid sequence that has a sequence identity of at least about 90% with SEQ ID NO: 3. In another set of

embodiments, the present invention is generally directed to a nucleic acid encoding an amino acid sequence that has a sequence identity of at least about 95% with SEQ ID NO: 3.

5 In another aspect, the present invention encompasses methods of making one or more of the embodiments described herein, for example, proteins such as mMAPLE2 and/or mMAPLE3. In still another aspect, the present invention encompasses methods of using one or more of the embodiments described herein, for example, proteins such as mMAPLE2 and/or mMAPLE3.

10 Other advantages and novel features of the present invention will become apparent from the following detailed description of various non-limiting embodiments of the invention when considered in conjunction with the accompanying figures. In cases where the present specification and a document incorporated by reference include conflicting and/or inconsistent disclosure, the present specification shall control. If two or more documents incorporated by reference include conflicting and/or inconsistent
15 disclosure with respect to each other, then the document having the later effective date shall control.

BRIEF DESCRIPTION OF THE DRAWINGS

Non-limiting embodiments of the present invention will be described by way of example with reference to the accompanying figures, which are schematic and are not
20 intended to be drawn to scale. In the figures, each identical or nearly identical component illustrated is typically represented by a single numeral. For purposes of clarity, not every component is labeled in every figure, nor is every component of each embodiment of the invention shown where illustration is not necessary to allow those of ordinary skill in the art to understand the invention. In the figures:

25 Figs. 1A-1D illustrates measurements of certain PAFPs in accordance with one embodiment of the invention;

Figs. 2A-2F illustrates measurements of various PAFPs in accordance with another embodiment of the invention;

30 Figs. 3A-3D illustrates dimerization of certain PAFPs in yet another embodiment of the invention;

Figs. 4A-4B illustrates signaling efficiency of certain PAFPs in still another embodiment of the invention;

Figs. 5A-5G illustrates certain proteins in accordance with various embodiments of the invention;

Figs. 6A-6D illustrates blinking number distributions of certain PAFPs in another embodiment of the invention;

5 Figs. 7A-7B illustrates maturation half-life in yet another embodiment of the invention;

Fig. 8 illustrates signaling efficiency versus maturation half-life in still another embodiment of the invention; and

10 Figs. 9A-9C illustrates excitation and emission spectra in accordance with certain embodiments of the invention.

BRIEF DESCRIPTION OF THE SEQUENCES

SEQ ID NO: 1 is mMAPLE having a sequence:

1	MVSKGEETIM	10
11	SVIKPDMKIK	20
21	LRMEGNVNGH	30
31	AFVIEGEGSG	40
41	KPFEGIQTID	50
51	LEVKEGAPLP	60
61	FAYDILTTAF	70
71	HYGNRVFTKY	80
81	PEDIPDYFKQ	90
91	SFPEGYSWER	100
101	SMTYEDGGIC	110
111	IATNDITMEE	120
121	DSFINKIHFK	130
131	GTNFPPNGPV	140
141	MQKRTVGWEV	150
151	STEKMYVRDG	160
161	VLKGDVKMKL	170
171	LLKGGSHYRC	180
181	DFRTTYKVKQ	190
191	KAVKLPDYHF	200
201	VDHRIEILSH	210
211	DKDYNKVKLY	220
221	EHAVARNSTD	230
231	SMDELYK	237

SEQ ID NO: 2 is mMAPLE2 having a sequence:

1	MVSKGEETIM	10
11	SVIKPDMKIK	20
21	LRMEGNVNGH	30
31	AFVIEGEGSG	40
41	KPFEGIQITID	50
51	LEVKEGAPLP	60
61	FAYDILTTAF	70
71	HYGNRVFTKY	80
81	PEDIPDYFKQ	90
91	SFPEGYSWER	100
101	SMTYEDGGIC	110
111	NATNDITMEE	120
121	DSFINKIHFK	130
131	GTNFPPNGPV	140
141	MQKRTVGWEV	150
151	STEKMYVRDG	160
161	VLKGDVKMKL	170
171	LLKGGSHYRC	180
181	DFRTTYKVKQ	190
191	KAVKLPAHF	200
201	VDHRIEILSH	210
211	DKDYNKVKLY	220
221	EHAVARNSTD	230
231	SMDELYK	237

SEQ ID NO: 3 is mMAPLE3 having a sequence:

1	MVSKGEETIM	10
11	SVIKPDMKIK	20
21	LRMEGNVNGH	30
31	AFVIEGEGSG	40
41	KPFEGIQITID	50
51	LEVKEGAPLP	60
61	FAYDILTTAF	70
71	HYGNRVFTKY	80
81	PRKIPDYFKQ	90
91	SFPEGYSWER	100
101	SMTYEDGGIC	110
111	NATNDITMEE	120
121	DSFINKIHFK	130
131	GTNFPPNGPV	140
141	MQKRTVGWEV	150
151	STEKMYVRDG	160
161	VLKGDVKMKL	170
171	LLKGGSHYRC	180
181	DFRTTYKVKQ	190
191	KAVKLPAHF	200
201	VDHRIEILSH	210
211	DKDYNKVKLY	220
221	EHAVARNSTD	230
231	SMDELYK	237

5 SEQ ID NO: 4 has a sequence DPPVAT.

SEQ ID NO: 5 has a sequence SGGGGSK.

SEQ ID NO: 6 has a sequence CHHHHHHGSG.

SEQ ID NO: 7 has a sequence GSGHHHHHH.

DETAILED DESCRIPTION

The present invention generally relates to fluorescent proteins, for various uses
5 such as for super-resolution microscopy. In one set of embodiments, the protein is
mMAPLE2 or mMAPLE3. In another set of embodiments, the protein comprises two or
substitutions in mMAPLE. In addition, in certain embodiments, the present invention is
generally directed to systems and methods for making or using such proteins, kits
involving such proteins, nucleic acids encoding such proteins, vectors or plasmids
10 comprising such proteins or nucleic acids, or cells or organisms comprising such proteins
or nucleic acids, as well as fragments, homologs, analogs, derivatives, or functionally
equivalent compositions.

In one aspect, the present invention is generally directed to proteins that are
generally related to mMAPLE (SEQ ID NO: 1), such as mMAPLE2 (SEQ ID NO: 2)
15 and mMAPLE3 (SEQ ID NO: 3). Other proteins of the invention related to mMAPLE
are discussed in more detail below. It should be noted that many of the proteins
discussed herein are not naturally-occurring.

In one set of embodiments, such proteins may be fluorescent, or may be useful of
microscopy. For example, the proteins may be useful in super-resolution microscopy
20 techniques such as PALM or STORM, or in other techniques including those described
herein. In one set of embodiments, such proteins may be photoactivatable,
photoswitchable, or photoconvertible. An entity is “activatable” if it can be activated
from a state not capable of emitting light (e.g., at a specific wavelength) to a state
capable of emitting light (e.g., at that wavelength). The entity may or may not be able to
25 be deactivated, e.g., by using deactivation light or other techniques for deactivating light.
An entity is “switchable” if it can be switched between two or more different states, one
of which is capable of emitting light (e.g., at a specific wavelength). In the other state(s),
the entity may emit no light, or emit light at a different wavelength. For instance, an
entity can be “activated” to a first state able to produce light having a desired
30 wavelength, and “deactivated” to a second state not able to produce light of the same
wavelength.

If the entity is activatable using light, then the entity is a “photoactivatable”
entity. Similarly, if the entity is switchable using light in combination or not in

combination with other techniques, then the entity is a “photoswitchable” entity. For instance, a photoswitchable entity may be switched between different light-emitting or non-emitting states by incident light of different wavelengths. Typically, a “switchable” entity can be identified by one of ordinary skill in the art by determining conditions
5 under which an entity in a first state can emit light when exposed to an excitation wavelength, switching the entity from the first state to the second state, e.g., upon exposure to light of a switching wavelength, then showing that the entity, while in the second state, can no longer emit light (or emits light at a reduced intensity) or emits light at a different wavelength when exposed to the excitation wavelength.

10 In some embodiments, the switchable entity includes a first, light-emitting portion (e.g., a fluorophore), and a second portion that activates or “switches” the first portion. For example, upon exposure to light, the second portion of the switchable entity may activate the first portion, causing the first portion to emit light.

In some cases, the entity can be reversibly switched between the two or more
15 states, e.g., upon exposure to the proper stimuli. For example, a first stimulus (e.g., a first wavelength of light) may be used to activate the switchable entity, while a second stimulus (e.g., a second wavelength of light or light with the first wavelength) may be used to deactivate the switchable entity, for instance, to a non-emitting state. Any suitable method may be used to activate the entity. For example, in one embodiment,
20 incident light of a suitable wavelength may be used to activate the entity to be able to emit light, and the entity can then emit light when excited by an excitation light. Thus, the photoswitchable entity can be switched between different light-emitting or non-emitting states by incident light.

In one set of embodiments, an emissive entity in a sample is an entity such as an
25 activatable entity, a switchable entity, a photoactivatable entity, or a photoswitchable entity. Examples of such entities are discussed herein. In some cases, more than one type of emissive entity may be present in a sample.

As mentioned, in some embodiments, the present invention is generally directed to proteins such as mMAPLE2 or mMAPLE3. In one embodiment, the protein
30 comprises a sequence SEQ ID NO: 2 or SEQ ID NO: 3. In some cases, the protein may also contain fragments or portions of SEQ ID NO: 2 or SEQ ID NO: 3. For instance, in one embodiment, the starting amino acid (M) may be absent, such that the protein comprises positions 2-237 of SEQ ID NO: 2 or SEQ ID NO: 3. In addition, one or both

of the head and tail fragments of the protein may be removed, e.g., without substantially affecting the structural integrity of the protein. For instance, in one embodiment, positions 1-16 may be removed and/or replaced with other sequences. In another embodiment, positions 226-237 may be removed and/or replaced with other sequences.

- 5 In some cases, both may be removed and/or replaced with other sequences (e.g., keeping positions 17-225).

In one set of embodiments, the present invention is generally directed to a protein that is based on mMAPLE, but contains one or more modifications. In some cases, the protein may have 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20
10 modifications with respect to mMAPLE (SEQ ID NO: 1).

For instance, in one set of embodiments, the protein contains modifications in at least positions 111 and 198 with respect to mMAPLE. The protein may also contain modifications in other positions as well. For instance, the protein may contain modifications in one or more of positions 82, 83, or 197.

- 15 In one set of embodiments, the protein contains modifications at any one or more of positions 82, 83, 111, 197, or 198 of mMAPLE. In certain embodiments, one or more of these modifications is directed to changing an amino acid from one category (i.e., of basic, acidic, polar, or nonpolar amino acid residues) to a different category. Without wishing to be bound by any theory, it is believed that these amino acids facilitate the
20 dimerization of mMAPLE to itself. Accordingly, it is believed that changing the amino acids at one or more of these locations may reduce or inhibit the ability of mMAPLE to dimerize. The basic amino acids are H, K, or R. The acidic acids are D or E. The polar amino acids are S, T, Y, Q, N, C, G. The nonpolar amino acids are F, L, I, M, V, P, A, W.

- 25 Thus, in one embodiment, position 111 and/or position 198 may be altered. For instance, in some cases, position 111 may be altered to N, or position 198 may be altered to A. If both of these substitutions are made relative to mMAPLE (i.e., I111N and Y198A), then the resulting protein is mMAPLE2. However, it should be understood that both of these substitutions need not be made, and that other amino acid residues can be
30 used instead (e.g., position 111 may be substituted with any basic, polar, or acidic amino acid residue, and/or position 198 may be substituted with any basic, acidic, or nonpolar amino acid residue).

In some embodiments, 1, 2, 3, 4, or all 5 of positions 82, 83, 111, 197, and 198 may be modified. As mentioned, in some cases, for each of these positions, the amino acid residue may independently be altered to an amino acid residue. Thus, for example, position 82 may be altered to R, position 83 may be altered to K, position 111 may be altered to N, position 197 may be altered to K, and/or position 198 may be altered to A. In some cases, the amino acid residue may be substituted with an amino acid residue of a different category (i.e., of basic, acidic, polar, or nonpolar amino acid residues). In some cases, position 82 may be substituted with a basic, polar, or nonpolar amino acid; position 83 may be substituted with a basic, polar, or nonpolar amino acid; position 111 may be substituted with a basic, polar, or acidic amino acid; position 197 may be substituted with a basic, polar, or nonpolar amino acid; and/or position 198 may be substituted with a basic, acidic, or nonpolar amino acid.

It should be understood that any of the substitutions described above may also be made on fragments, derivatives, etc. of mMAPLE, or any of the other proteins discussed herein. For instance, in one embodiment, a protein may comprise positions 2-237 of mMAPLE with one or more of these substitutions; a protein may comprise positions 17-225 of mMAPLE with one or more of these substitutions, a protein may comprise positions 1-225 or 2-225 of mMAPLE with one or more of these substitutions, a protein may comprise positions 17-237 of mMAPLE with one or more of these substitutions, etc. Thus, for instance, any of these proteins or protein fragments may contain one or more substitutions in positions 82, 83, 111, 197, and 198, as discussed above.

Other substitutions are also possible, e.g., in addition to and/or instead of those described above. In another aspect, the present invention is directed to a protein having a sequence having at least 80% sequence identity with either mMAPLE2 or mMAPLE3. In other cases, the protein may have at least 85%, at least 90%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity with either mMAPLE2 or mMAPLE3. It should be understood that these sequences do not encompass mMAPLE itself.

As used herein, an "amino acid" is given its ordinary meaning as used in the field of biochemistry. A series of isolated amino acids may be connected to form a peptide or a protein by reaction of the $-NH_2$ of one amino acid with the $-COOH$ of another amino acid to form a peptide bond ($-CO-NH-$). The "natural amino acids," as used herein, are the 20 amino acids commonly found in nature, typically in the L-isomer, i.e., alanine

(“Ala” or “A”), arginine (“Arg” or “R”), asparagine (“Asn” or “N”), aspartic acid (“Asp” or “D”), cysteine (“Cys” or “C”), glutamine (“Gln” or “Q”), glutamic acid (“Glu” or “E”), glycine (“Gly” or “G”), histidine (“His” or “H”), isoleucine (“Ile” or “I”), leucine (“Leu” or “L”), lysine (“Lys” or “K”), methionine (“Met” or “M”), phenylalanine (“Phe” or “F”), proline (“Pro” or “P”), serine (“Ser” or “S”), threonine (“Thr” or “T”), tryptophan (“Trp” or “W”), tyrosine (“Tyr” or “Y”), and valine (“Val” or “V”). In some cases, an amino acid that can be either asparagine or aspartic acid is referred to as “Asx” or “B,” while an amino acid that can be either glutamine or glutamic acid is referred to as “Glx” or “Z.” Some amino acids are basic, e.g., H, K, or R. Some amino acids are acidic, e.g., D or E. Some amino acids are polar, e.g., S, T, Y, Q, N, C, G. Some amino acids are nonpolar, e.g., F, L, I, M, V, P, A, W.

In another aspect, the present invention is generally directed to nucleic acids encoding any of the proteins described above, or elsewhere herein. Those of ordinary skill in the art will be able to readily identify nucleic acids that encode any of these proteins, e.g., by using the Genetic Code to identify the sequence of amino acids encoded by a given nucleic acid sequence. The nucleic acid may be, for example, DNA, cDNA, RNA, etc. In some cases, the nucleic acid may be present as an isolated nucleic acid, or the nucleic acid may be incorporated into a cell, e.g., to produce the corresponding protein. For instance, the nucleic acid may be one that can be transfected into a cell. In one embodiment, the nucleic acid is present in a vector or a plasmid.

Accordingly, in certain aspects, various proteins and nucleic acids such as those discussed herein may be produced using standard molecular biology methods generally known to those of ordinary skill in the art (*see, e.g., Green and Sambrook, Molecular Cloning, A Laboratory Manual*, 2012, Cold Spring Harbor Press).

For example, in some embodiments, a nucleic acid may be included within a vector. A vector is a nucleic acid (*e.g., DNA or cDNA*) used as a vehicle to artificially carry genetic material (*e.g., an engineered nucleic acid*) into another cell where, for example, it can be replicated and/or expressed. In some embodiments, a vector is an episomal vector (*see, e.g., Van Craenenbroeck K. et al. Eur. J. Biochem.* 267, 5665, 2000, incorporated by reference herein). A non-limiting example of a vector is a plasmid. Plasmids are double-stranded generally circular DNA sequences that are capable of automatically replicating in a host cell. Plasmid vectors typically contain an origin of replication that allows for semi-independent replication of the plasmid in the

host and also the transgene insert. Plasmids may have more features, including, for example, a “multiple cloning site,” which includes nucleotide overhangs for insertion of a nucleic acid insert, and multiple restriction enzyme consensus sites to either side of the insert. Another non-limiting example of a vector is a viral vector.

5 Nucleic acids as discussed herein may comprise, in some embodiments, promoters operably linked to a nucleotide sequence. A “promoter” is a control region of a nucleic acid at which initiation and rate of transcription of the remainder of a nucleic acid are controlled. A promoter may also contain sub-regions at which regulatory proteins and molecules may bind, such as RNA polymerase and other transcription
10 factors. Promoters may be constitutive, inducible, activatable, repressible, tissue-specific or any combination thereof.

In some embodiments, a nucleic acid sequence may be positioned under the control of a recombinant or heterologous promoter, which refers to a promoter that is not normally associated with the nucleotide sequence in its natural environment. Such
15 promoters may include promoters of other genes; promoters isolated from any other prokaryotic cell; and synthetic promoters that are not “naturally occurring” such as, for example, those that contain different elements of different transcriptional regulatory regions and/or mutations that alter expression through methods of genetic engineering that are known in the art. In addition to producing nucleotide sequences of promoters
20 synthetically, sequences may be produced using recombinant cloning and/or nucleic acid amplification technology, including polymerase chain reaction (PCR). Those of ordinary skill in the art will be aware of suitable promoters for a variety of cell types. Many such promoters can be readily obtained commercially.

A promoter may drive expression or drive transcription of the nucleic acid
25 sequence that it regulates. A promoter is considered to be “operably linked” when it is in a correct functional location and orientation in relation to the nucleotide sequence it regulates to control (“drive”) transcriptional initiation and/or expression of that sequence.

A promoter may be classified in some cases as strong or weak according to its
30 affinity for RNA polymerase (and/or sigma factor); this is related to how closely the promoter sequence resembles the ideal consensus sequence for the polymerase. The strength of a promoter may depend on whether initiation of transcription occurs at that promoter with high or low frequency. Different promoters with different strengths may

be used to construct nucleic acids with different levels of gene/protein expression (*e.g.*, the level of expression initiated from a weak promoter is lower than the level of expression initiated from a strong promoter).

In some embodiments, a nucleic acid may be codon-optimized, for example, for expression in human cells or other types of cells, such as *E. coli*. Codon optimization is a technique to maximize the protein expression in living organism by increasing the translational efficiency of gene of interest by transforming a DNA sequence of nucleotides of one species into a DNA sequence of nucleotides of another species. Methods of codon optimization are well-known to those of ordinary skill in the art.

10 Nucleic acids such as those described herein can be expressed in a variety of cell types. In some embodiments, engineered constructs are expressed in mammalian cells. For example, in some embodiments, engineered constructs are expressed in human cells, primate cells (*e.g.*, vero cells), rat cells (*e.g.*, GH3 cells, OC23 cells) or mouse cells (*e.g.*, MC3T3 cells). There are a variety of human cell lines, including, without limitation,
15 HEK cells, HeLa cells, cancer cells from the National Cancer Institute's 60 cancer cell lines (NCI60), DU145 (prostate cancer) cells, Lncap (prostate cancer) cells, MCF-7 (breast cancer) cells, MDA-MB-438 (breast cancer) cells, PC3 (prostate cancer) cells, T47D (breast cancer) cells, THP-1 (acute myeloid leukemia) cells, U87 (glioblastoma) cells, SHSY5Y human neuroblastoma cells (cloned from a myeloma) and Saos-2 (bone
20 cancer) cells. In some embodiments, engineered constructs are expressed in human embryonic kidney (HEK) cells (*e.g.*, HEK 293 or HEK 293T cells). In some embodiments, engineered constructs are expressed in bacterial cells, yeast cells, insect cells or other types of cells.

Cells, in some embodiments, can be modified. A modified cell is a cell that
25 contains an exogenous nucleic acid or a nucleic acid that does not occur in nature, such as those nucleic acids described herein. In some embodiments, a modified cell contains an independently replicating nucleic acid (*e.g.*, an engineered nucleic acid present on an episomal vector). In some embodiments, a modified cell is produced by introducing a foreign or exogenous nucleic acid into a cell. A nucleic acid may be introduced into a
30 cell by conventional methods, such as, for example, electroporation (*see, e.g.*, Heiser W.C. *Transcription Factor Protocols: Methods in Molecular Biology*TM 2000; 130: 117-134), chemical (*e.g.*, calcium phosphate or lipid) transfection (*see, e.g.*, Lewis W.H., *et al.*, *Somatic Cell Genet.* 1980 May; 6(3): 333-47; Chen C., *et al.*, *Mol Cell Biol.* 1987

August; 7(8): 2745–2752), fusion with bacterial protoplasts containing recombinant plasmids (*see, e.g.,* Schaffner W. *Proc Natl Acad Sci USA*. 1980 Apr; 77(4): 2163-7), or microinjection of purified DNA directly into the nucleus of the cell (*see, e.g.,* Capecchi M.R. *Cell*. 1980 Nov; 22(2 Pt 2): 479-88). Those of ordinary skill in the art will be aware of various systems and methods for introducing a non-native nucleic acid construct into a cell in order to cause the cell to produce a protein of interest.

In some embodiments, a cell is modified to overexpress a protein of interest (*e.g.,* via introducing or modifying a promoter or other regulatory element near the endogenous gene that encodes the protein of interest to increase its expression level). In some embodiments, a cell is modified by mutagenesis. In some embodiments, a cell is modified by introducing a recombinant nucleic acid into the cell in order to produce a genetic change of interest (*e.g.,* via insertion or homologous recombination). Those of ordinary skill in the art will be aware of suitable techniques to cause a cell to produce a protein of interest, *e.g.,* through overexpression or other well-known techniques. In some cases, the cell may be the cell of an organism, *e.g.,* a human or a non-human mammal, a bacterial cell (*e.g., E. coli*), or the like.

Nucleic acids such as those discussed herein may be transiently expressed or stably expressed in various embodiments. “Transient cell expression” refers to expression by a cell of a nucleic acid that is not integrated into the nuclear genome of the cell. By comparison, “stable cell expression” refers to expression by a cell of a nucleic acid that remains in the nuclear genome of the cell and its daughter cells. Typically, to achieve stable cell expression, a cell is co-transfected with a marker gene and an exogenous nucleic acid (*e.g.,* engineered nucleic acid) that is intended for stable expression in the cell. The marker gene gives the cell some selectable advantage (*e.g.,* resistance to a toxin, antibiotic, or other factor). Few transfected cells will, by chance, have integrated the exogenous nucleic acid into their genome. If a toxin, for example, is then added to the cell culture, only those few cells with a toxin-resistant marker gene integrated into their genomes will be able to proliferate, while other cells will die. After applying this selective pressure for a period of time, only the cells with a stable transfection remain and can be cultured further. Examples of marker genes and selection agents for use in accordance with the present disclosure include, without limitation, dihydrofolate reductase with methotrexate, glutamine synthetase with methionine sulfoximine, hygromycin phosphotransferase with hygromycin, puromycin N-

acetyltransferase with puromycin, and neomycin phosphotransferase with Geneticin, also known as G418. Other marker genes/selection agents known to those of ordinary skill in the art are also contemplated in other embodiments. Expression of nucleic acids in transiently-transfected and/or stably-transfected cells may be constitutive or inducible.

5 Inducible promoters for use as provided herein are described above.

Mammalian cells modified to comprise a nucleic acid as discussed herein (e.g., encoding a protein as discussed herein) may be cultured (e.g., maintained in cell culture) using conventional mammalian cell culture methods known to those of ordinary skill in the art (*see, e.g., Phelan M.C. Curr Protoc Cell Biol.* 2007 Sep; Chapter 1: Unit 1.1, incorporated by reference herein). For example, cells may be grown and maintained at 10 an appropriate temperature and gas mixture (e.g., 37 °C, 5% CO₂ for mammalian cells) in a cell incubator. Culture conditions may vary for each cell type. For example, cell growth media may vary in pH, glucose concentration, growth factors, and the presence of other nutrients. Growth factors used to supplement media are often derived from the 15 serum of animal blood, such as fetal bovine serum (FBS), bovine calf serum, equine serum and/or porcine serum. In some embodiments, culture media used as provided herein may be commercially available and/or well-described (*see, e.g., Birch J. R., R.G. Spier (Ed.) Encyclopedia of Cell Technology, Wiley. 411-424, 2000; Keen M. J. Cytotechnology 17: 125-132, 1995; Zang, et al. Bio/Technology. 13: 389-392, 1995*). In 20 some embodiments, chemically defined media is used.

In certain aspects of the invention, images of a sample may be obtained using stochastic imaging techniques. In many stochastic imaging techniques, various entities are activated and emit light at different times and imaged; typically the entities are activated in a random or “stochastic” manner. For example, a statistical or “stochastic” 25 subset of the entities within a sample can be activated from a state not capable of emitting light at a specific wavelength to a state capable of emitting light at that wavelength. Some or all of the activated entities may be imaged (e.g., upon excitation of the activated entities), and this process repeated, each time activating another statistical or “stochastic” subset of the entities. Optionally, the entities are deactivated (for 30 example, spontaneously, or by causing the deactivation, for instance, with suitable deactivation light). Repeating this process any suitable number of times allows an image of the sample to be built up using the statistical or “stochastic” subset of the activated emissive entities activated each time. Higher resolutions may be achieved in some cases

because the emissive entities are not all simultaneously activated, making it easier to resolve closely positioned emissive entities. Non-limiting examples of stochastic imaging techniques which may be used include stochastic optical reconstruction microscopy (STORM), single-molecule localization microscopy (SMLM), spectral
5 precision distance microscopy (SPDM), super-resolution optical fluctuation imaging (SOFI), photoactivated localization microscopy (PALM), and fluorescence photoactivation localization microscopy (FPALM). In addition, in some cases, techniques such as those discussed herein can be combined with other 3-dimensional techniques for determining the position of entities within a sample; see, e.g.,
10 International Patent Application No. PCT/US2008/013915, filed December 19, 2008, entitled "Sub-diffraction Limit Image Resolution in Three Dimensions," by Zhuang, *et al.*, incorporated herein by reference in its entirety.

As used herein, the term "light" generally refers to electromagnetic radiation, having any suitable wavelength (or equivalently, frequency). For instance, in some
15 embodiments, the light may include wavelengths in the optical or visual range (for example, having a wavelength of between about 380 nm and about 750 nm, i.e., "visible light"), infrared wavelengths (for example, having a wavelength of between about 700 micrometers and 1000 nm), ultraviolet wavelengths (for example, having a wavelength of between about 400 nm and about 10 nm), or the like. In certain cases, more than one
20 type of entity may be used, e.g., entities that are chemically different or distinct, for example, structurally. However, in other cases, the entities are chemically identical or at least substantially chemically identical.

In certain embodiments, the resolution of the entities in the images can be, for instance, on the order of 1 micrometer or less, as described herein. In some cases, the
25 resolution of an entity may be determined to be less than the wavelength of the light emitted by the entity, and in some cases, less than half the wavelength of the light emitted by the entity. For example, if the emitted light is visible light, the resolution may be determined to be less than about 700 nm. In some cases, two (or more) entities can be resolved even if separated by a distance of less than about 500 nm, less than about 300
30 nm, less than about 200 nm, less than about 100 nm, less than about 80 nm, less than about 60 nm, less than about 50 nm, or less than about 40 nm. In some cases, two or more entities separated by a distance of less than about 35 nm, less than about 30 nm,

less than about 25 nm, less than about 20 nm, less than about 15 nm, less than about 10 nm, or less than about 5 nm can be resolved using embodiments of the present invention.

One non-limiting example is stochastic optical reconstruction microscopy (STORM). See, e.g., U.S. Patent No. 7,838,302, issued November 23, 2010, entitled

5 “Sub-Diffraction Limit Image Resolution and Other Imaging Techniques,” by Zhuang, *et al.*, incorporated herein by reference in its entirety. In STORM, incident light is applied to emissive entities within a sample in a sample region to activate the entities, where the incident light has an intensity and/or frequency that is able to cause a statistical subset of the plurality of emissive entities to become activated from a state not capable of emitting

10 light (e.g., at a specific wavelength) to a state capable of emitting light (e.g., at that wavelength). Once activated, the emissive entities may spontaneously emit light, and/or excitation light may be applied to the activated emissive entities to cause these entities to emit light. The excitation light may be of the same or different wavelength as the activation light. The emitted light can be collected or acquired, e.g., in one, two, or more

15 objectives as previously discussed. In certain embodiments the positions of the entities can be determined in two or three dimensions from their images. In some cases, the excitation light is also able to subsequently deactivate the statistical subset of the plurality of emissive entities, and/or the entities may be deactivated via other suitable techniques (e.g., by applying deactivation light, by applying heat, by waiting a suitable

20 period of time, etc.). This process repeated as needed, each time with a statistically different subset of the plurality of emissive entities to emit light. In this way, a stochastic image of some or all of the emissive entities within a sample may be produced, e.g., from the determined positions of the entities. In addition, as discussed herein, various image processing techniques, such as noise reduction and/or x , y and/or z

25 position determination can be performed on the acquired images.

In some cases, incident light having a sufficiently weak intensity may be applied to a plurality of entities such that only a subset or fraction of the entities within the incident light are activated, e.g., on a stochastic or random basis. The amount of activation can be any suitable fraction, e.g., less than about 0.01%, less than about

30 0.03%, less than about 0.05%, less than about 0.1%, less than about 0.3%, less than about 0.5%, less than about 1%, less than about 3%, less than about 5%, less than about 10%, less than less than about 15%, less than about 20%, less than about 25%, less than about 30%, less than about 35%, less than about 40%, less than about 45%, less than

about 50%, less than about 55%, less than about 60%, less than about 65%, less than about 70%, less than about 75%, less than about 80%, less than about 85%, less than about 90%, or less than about 95% of the entities may be activated, depending on the application. For example, by appropriately choosing the intensity of the incident light, a
5 sparse subset of the entities may be activated such that at least some of them are optically resolvable from each other and their positions can be determined. In some embodiments, the activation of the subset of the entities can be synchronized by applying a short duration of incident light. Iterative activation cycles may allow the positions of all of the entities, or a substantial fraction of the entities, to be determined. In some cases, an
10 image with sub-diffraction limit resolution can be constructed using this information.

Multiple locations on a sample can each be analyzed to determine the entities within those locations. For example, a sample may contain a plurality of various entities, some of which are at distances of separation that are less than the wavelength of the light emitted by the entities or below the diffraction limit of the emitted light. Different
15 locations within the sample may be determined (e.g., as different pixels within an image), and each of those locations independently analyzed to determine the entity or entities present within those locations. In some cases, the entities within each location are determined to resolutions that are less than the wavelength of the light emitted by the entities or below the diffraction limit of the emitted light, as previously discussed.

20 In some cases, the activation light and deactivation light have the same wavelength. In some cases, the activation light and deactivation light have different wavelengths. In some cases, the activation light and excitation light have the same wavelength. In some cases, the activation light and excitation light have different wavelengths. In some cases, the excitation light and deactivation light have the same
25 wavelength. In some cases, the excitation light and deactivation light have different wavelengths. In some cases, the activation light, excitation light and deactivation light all have the same wavelength.

The light may be monochromatic (e.g., produced using a laser) or polychromatic. In another embodiment, the entity may be activated upon stimulation by electric fields
30 and/or magnetic fields. In other embodiments, the entity may be activated upon exposure to a suitable chemical environment, e.g., by adjusting the pH, or inducing a reversible chemical reaction involving the entity, etc. Similarly, any suitable method may be used to deactivate the entity, and the methods of activating and deactivating the

entity need not be the same. For instance, the entity may be deactivated upon exposure to incident light of a suitable wavelength, or the entity may be deactivated by waiting a sufficient time.

In one set of embodiments, the switchable entity can be immobilized, e.g., covalently, with respect to a binding partner, i.e., a molecule that can undergo binding with a particular analyte. Binding partners include specific, semi-specific, and non-specific binding partners as known to those of ordinary skill in the art. The term “specifically binds,” when referring to a binding partner (e.g., protein, nucleic acid, antibody, etc.), refers to a reaction that is determinative of the presence and/or identity of one or other member of the binding pair in a mixture of heterogeneous molecules (e.g., proteins and other biologics). Thus, for example, in the case of a receptor/ligand binding pair, the ligand would specifically and/or preferentially select its receptor from a complex mixture of molecules, or vice versa. Other examples include, but are not limited to, an enzyme would specifically bind to its substrate, a nucleic acid would specifically bind to its complement, an antibody would specifically bind to its antigen. The binding may be by one or more of a variety of mechanisms including, but not limited to ionic interactions, and/or covalent interactions, and/or hydrophobic interactions, and/or van der Waals interactions, etc. By immobilizing a switchable entity with respect to the binding partner of a target molecule or structure (e.g., DNA or a protein within a cell), the switchable entity can be used for various determination or imaging purposes. For example, a switchable entity having an amine-reactive group may be reacted with a binding partner comprising amines, for example, antibodies, proteins or enzymes.

In some embodiments, more than one switchable entity may be used, and the entities may be the same or different. In some cases, the light emitted by a first entity and the light emitted by a second entity have the same wavelength. The entities may be activated at different times and the light from each entity may be determined separately. This allows the location of the two entities to be determined separately and, in some cases, the two entities may be spatially resolved, even at distances of separation that are less than the wavelength of the light emitted by the entities or below the diffraction limit of the emitted light (i.e., “sub-diffraction limit” resolutions). In certain instances, the light emitted by a first entity and the light emitted by a second entity have different wavelengths (for example, if the first entity and the second entity are chemically

different, and/or are located in different environments). The entities may be spatially resolved even at distances of separation that are less than the wavelength of the light emitted by the entities or below the diffraction limit of the emitted light. In certain instances, the light emitted by a first entity and the light emitted by a second entity have
5 substantially the same wavelengths, but the two entities may be activated by light of different wavelengths and the light from each entity may be determined separately. The entities may be spatially resolved even at distances of separation that are less than the wavelength of the light emitted by the entities, or below the diffraction limit of the emitted light.

10 In some cases, the entities may be independently switchable, i.e., the first entity may be activated to emit light without activating a second entity. For example, if the entities are different, the methods of activating each of the first and second entities may be different (e.g., the entities may each be activated using incident light of different wavelengths). As another non-limiting example, if the entities are substantially the
15 same, a sufficiently weak intensity of light may be applied to the entities such that only a subset or fraction of the entities within the incident light are activated, i.e., on a stochastic or random basis. Specific intensities for activation can be determined by those of ordinary skill in the art using no more than routine skill. By appropriately choosing the intensity of the incident light, the first entity may be activated without activating the
20 second entity. The entities may be spatially resolved even at distances of separation that are less than the wavelength of the light emitted by the entities, or below the diffraction limit of the emitted light. As another non-limiting example, the sample to be imaged may comprise a plurality of entities, some of which are substantially identical and some of which are substantially different. In this case, one or more of the above methods may
25 be applied to independently switch the entities. The entities may be spatially resolved even at distances of separation that are less than the wavelength of the light emitted by the entities, or below the diffraction limit of the emitted light.

In some cases, incident light having a sufficiently weak intensity may be applied to a plurality of entities such that only a subset or fraction of the entities within the
30 incident light are activated, e.g., on a stochastic or random basis. The amount of activation may be any suitable fraction, e.g., about 0.1%, about 0.3%, about 0.5%, about 1%, about 3%, about 5%, about 10%, about 15%, about 20%, about 25%, about 30%, about 35%, about 40%, about 45%, about 50%, about 55%, about 60%, about 65%,

about 70%, about 75%, about 80%, about 85%, about 90%, or about 95% of the entities may be activated, depending on the application. For example, by appropriately choosing the intensity of the incident light, a sparse subset of the entities may be activated such that at least some of them are optically resolvable from each other and their positions can
5 be determined. In some embodiments, the activation of the subset of the entities can be synchronized by applying a short duration of the incident light. Iterative activation cycles may allow the positions of all of the entities, or a substantial fraction of the entities, to be determined. In some cases, an image with sub-diffraction limit resolution can be constructed using this information.

10 In some embodiments, a microscope may be configured so to collect light emitted by the switchable entities while minimizing light from other sources of fluorescence (e.g., "background noise"). In certain cases, imaging geometry such as, but not limited to, a total-internal-reflection geometry, a spinning-disc confocal geometry, a scanning confocal geometry, an epi-fluorescence geometry, an epi-fluorescence geometry with an
15 oblique incidence angle, etc., may be used for sample excitation. In some embodiments, as previously discussed, a thin layer or plane of the sample is exposed to excitation light, which may reduce excitation of fluorescence outside of the sample plane. A high numerical aperture lens may be used to gather the light emitted by the sample. The light may be processed, for example, using filters to remove excitation light, resulting in the
20 collection of emission light from the sample. In some cases, the magnification factor at which the image is collected can be optimized, for example, when the edge length of each pixel of the image corresponds to the length of a standard deviation of a diffraction limited spot in the image.

In some embodiments of the invention, the switchable entities may also be
25 resolved as a function of time. For example, two or more entities may be observed at various time points to determine a time-varying process, for example, a chemical reaction, cell behavior, binding of a protein or enzyme, etc. Thus, in one embodiment, the positions of two or more entities may be determined at a first point of time (e.g., as described herein), and at any number of subsequent points of time. As a specific
30 example, if two or more entities are immobilized relative to a common entity, the common entity may then be determined as a function of time, for example, time-varying processes such as movement of the common entity, structural and/or configurational changes of the common entity, reactions involving the common entity, or the like. The

time-resolved imaging may be facilitated in some cases since a switchable entity can be switched for multiple cycles, with each cycle giving one data point of the position of the entity.

In some cases, one or more light sources may be time-modulated (e.g., by shutters, acoustic optical modulators, or the like). Thus, a light source may be one that is activatable and deactivatable in a programmed or a periodic fashion. In one embodiment, more than one light source may be used, e.g., which may be used to illuminate a sample with different wavelengths or colors. For instance, the light sources may emanate light at different frequencies, and/or color-filtering devices, such as optical filters or the like, may be used to modify light coming from the light sources such that different wavelengths or colors illuminate a sample.

Various image-processing techniques may also be used to facilitate determination of the entities. For example, drift correction or noise filters may be used. Generally, in drift correction, a fixed point is identified (for instance, as a fiduciary marker, e.g., a fluorescent particle may be immobilized to a substrate), and movements of the fixed point (i.e., due to mechanical drift) are used to correct the determined positions of the switchable entities. In another example method for drift correction, the correlation function between images acquired in different imaging frames or activation frames can be calculated and used for drift correction. In some embodiments, the drift may be less than about 1000 nm/min, less than about 500 nm/min, less than about 300 nm/min, less than about 100 nm/min, less than about 50 nm/min, less than about 30 nm/min, less than about 20 nm/min, less than about 10 nm/min, or less than 5 nm/min. Such drift may be achieved, for example, in a microscope having a translation stage mounted for x-y positioning of the sample slide with respect to the microscope objective. The slide may be immobilized with respect to the translation stage using a suitable restraining mechanism, for example, spring loaded clips. In addition, a buffer layer may be mounted between the stage and the microscope slide. The buffer layer may further restrain drift of the slide with respect to the translation stage, for example, by preventing slippage of the slide in some fashion. The buffer layer, in one embodiment, is a rubber or polymeric film, for instance, a silicone rubber film.

Accordingly, one embodiment of the invention is directed to a device comprising a translation stage, a restraining mechanism (e.g., a spring loaded clip) attached to the translation stage able to immobilize a slide, and optionally, a buffer layer (e.g., a silicone

rubber film) positioned such that a slide restrained by the restraining mechanism contacts the buffer layer. To stabilize the microscope focus during data acquisition, a “focus lock” device may be used in some cases. As a non-limiting example, to achieve focus lock, a laser beam may be reflected from the substrate holding the sample and the
5 reflected light may be directed onto a position-sensitive detector, for example, a quadrant photodiode. In some cases, the position of the reflected laser, which may be sensitive to the distance between the substrate and the objective, may be fed back to a z-positioning stage, for example a piezoelectric stage, to correct for focus drift. The device may also include, for example, a spatial light modulator and/or a polarizer, e.g., as discussed
10 herein.

Another aspect of the invention is directed to a computer-implemented method. For instance, a computer and/or an automated system may be provided that is able to automatically and/or repetitively perform any of the methods described herein. In some cases, a computer may be used to control excitation of the switchable entities and the
15 acquisition of images of the switchable entities. In one set of embodiments, a sample may be excited using light having various wavelengths and/or intensities, and the sequence of the wavelengths of light used to excite the sample may be correlated, using a computer, to the images acquired of the sample containing the switchable entities. For instance, the computer may apply light having various wavelengths and/or intensities to a
20 sample to yield different average numbers of activated switchable elements in each region of interest (e.g., one activated entity per location, two activated entities per location, etc.). In some cases, this information may be used to construct an image of the switchable entities, in some cases at sub-diffraction limit resolutions, as noted above.

Still other embodiments of the invention are generally directed to a system able to
25 perform one or more of the embodiments described herein. For example, the system may include a microscope, a device for activating and/or switching the entities to produce light having a desired wavelength (e.g., a laser or other light source), a device for determining the light emitted by the entities (e.g., a camera, which may include color-filtering devices, such as optical filters), and a computer for determining the spatial
30 positions of the two or more entities.

In other aspects of the invention, the systems and methods described herein may also be combined with other imaging techniques known to those of ordinary skill in the art, such as high-resolution fluorescence *in situ* hybridization (FISH) or

immunofluorescence imaging, live cell imaging, confocal imaging, epi-fluorescence imaging, total internal reflection fluorescence imaging, etc. In one set of embodiments, an existing microscope (e.g., a commercially-available microscope) may be modified using components such as discussed herein, e.g., to acquire images and/or to determine
5 the positions of emissive entities, as discussed herein.

U.S. Pat. No. 7,838,302, issued November 23, 2010, entitled “Sub-Diffraction Limit Image Resolution and Other Imaging Techniques,” by Zhuang, *et al.*; International Patent Application No. PCT/US2008/013915, filed December 19, 2008, entitled “Sub-Diffraction Limit Image Resolution in Three Dimensions,” by Zhuang, *et al.*, published
10 as WO 2009/085218 on July 9, 2009; or International Patent Application No. PCT/US2012/069138, filed December 15, 2012, entitled “High Resolution Dual-Objective Microscopy,” by Zhuang, *et al.*, published as WO 2013/090360 on June 20, 2013, are each incorporated herein by reference in its entirety. In addition, U.S. Provisional Patent Application Serial No. 61/979,436, filed April 14, 2014, entitled
15 “Photoconvertible Fluorescent Proteins,” by Zhuang, *et al.*, is incorporated herein by reference in its entirety.

The following examples are intended to illustrate certain embodiments of the present invention, but do not exemplify the full scope of the invention.

EXAMPLE

20 Photoactivatable fluorescent proteins (PAFPs) have been widely used for super-resolution imaging based on the switching and localization of single molecules. Several properties of PAFPs may influence the quality of the super-resolution images. These properties include (i) the number of photons emitted per switching cycle, which affects the localization precision of individual molecules, (ii) the ratio of the on- and off-
25 switching rate constants, which limits the achievable localization density, (iii) the dimerization tendency, which could cause undesired aggregation of target proteins, and (iv) the signaling efficiency, which determines the fraction of target-PAFP fusion proteins that is detectable in a cell. In this example, these properties were evaluated for twelve commonly used PAFPs. Notably, none of the existing PAFPs provided optimal
30 performance in all four criteria, particularly in the signaling efficiency and dimerization tendency. The PAFPs with low dimerization tendencies exhibited low signaling efficiencies, whereas mMaple showed the highest signaling efficiency but also a high dimerization tendency. To address this limitation, two new PAFPs based on mMaple,

mMaple2 and mMaple3, were engineered. These proteins exhibited substantially reduced or undetectable dimerization tendencies as compared to mMaple but maintained the high signaling efficiency of mMaple. These proteins provided photon numbers and on-off switching rate ratios that are comparable to the best achieved values among PAFPs.

5 Photoactivated localization microscopy (PALM), Stochastic optical reconstruction microscopy (STORM), and related imaging methods take advantage of photoswitching and imaging of single molecules to circumvent the diffraction limit of spatial resolution in light microscopy. In these methods, only a subset of the fluorescent labels in the sample is switched on at any given time such that the positions of individual
10 fluorophores can be localized from their images with high precision. Iteration of this process allows numerous fluorescent labels to be localized and an image with sub-diffraction-limit resolution to be reconstructed from the fluorophore localizations. Fluorescent proteins that can be activated from dark to fluorescent or converted from one color to another are widely used for such imaging approaches. Although photoactivatable
15 fluorescent proteins (PAFPs) are generally dimmer than photoswitchable dyes and hence give lower image resolution, the ease and high specificity of labeling protein targets in living cells with fluorescent proteins makes PAFPs highly appealing probes for imaging the dynamics of cellular structures.

For single-molecule-based super-resolution imaging methods, several properties
20 of PAFPs may be important for the image quality. Four such properties are as follows. i) The photon budget, defined as the average number of photons emitted in each switching event. Given that the error of localizing an individual fluorophore approximately scales with the inverse square root of the number of detected photons, a higher photon budget may lead to higher localization precision and hence higher image resolution. ii) The on-
25 off switching rate ratio (on-off ratio), defined as the ratio between the on-switching (activation) and off-switching or photobleaching rates under the illumination of the imaging light only. Even in the absence of activation light, the imaging light itself can also switching on the PAFPs, albeit at a low rate. Thus the ratio between the on-switching and off-switching rates under this condition determines the lower bound of the
30 fraction of PAFP molecules in the on state at any given time. The presence of activation light would increase this fraction. When the product of this fraction and the density of fluorescent labels reaches ~ 1 fluorophore per diffraction-limited volume, it becomes difficult to resolve and precisely localize the activated fluorophores. Hence the on-off

ratio limits the density of fluorescent labels that can be localized, which in turn affects the effective image resolution based on the Nyquist sampling theorem. iii) The dimerization tendency. Many PAFPs have a weak tendency to form dimers, this could even be true for the PAFP that are reported as being monomeric. When these proteins are fused to target proteins that also tend to polymerize, they may cause undesired aggregation of the target proteins and distort the native distribution of the protein of interest. iv) The signaling efficiency, defined as the ratio between the number of detectable PAFP-fusion molecules per cell and the expression level of the fusion protein. Fluorescent proteins do not necessarily fold with 100% efficiency. Among the folded molecules, not all of them will become mature at the time of imaging. Among the matured PAFP molecules, only a subset can be photoactivated and imaged. Because of these deficiencies, the number of fusion molecules detected could be substantially lower than the expression level of the fusion protein. PAFPs with higher signaling efficiencies will lead to higher localization densities for a given target protein, which will in turn increase the effective image resolution.

This example measured the above properties of twelve commonly used PAFPs, including PAGFP, Dendra2, mEos2, mEos3.2, tdEos, mKikGR, PAmCherry, PAtagRFP, mMaple, PSCFP2, Dronpa, and mGeosM. From this screen, it was found that none of these PAFPs was optimal in all four criteria described above. For example, PAtagRFP and mEos3.2 exhibited the highest photon budgets among PAFPs, excellent on-off ratios, and undetectable dimerization tendencies, but showed poor signaling efficiencies. Alternatively, mMaple provided excellent signaling efficiency and on-off ratio with a photon budget nearly equal to those of PAtagRFP and mEos3.2, but had a substantial dimerization tendency. To address this limitation, two new PAFPs were developed in this example based on mMaple that exhibited substantially reduced or undetectable dimerization tendencies while maintaining the high signaling efficiency, high photon budget and low on-off ratio of mMaple. These new PAFPs may substantially facilitate super-resolution imaging of cellular structures.

Photon budget of photoactivatable fluorescent proteins. To evaluate the properties of PAFPs under conditions similar to typical super-resolution imaging experiments, each PAFP was fused to various target proteins and expressed these fusion proteins in either mammalian cells or bacteria. In order to measure the photon budget, the PAFPs were fused to the mammalian focal adhesion protein Zyxin, transiently

transfected BS-C-1 cells with the fusion constructs, and imaged the cells using the super-resolution imaging mode in which individual activated proteins were imaged. The distributions of the photon numbers detected per activation event were determined for all twelve PAFPs. Four example distributions, for mEos3.2, mMaple, PSCFP2 and PAGFP, are shown in Fig. 1. The mean photon numbers determined from these distributions are listed in Table 1.

Figure 1. Photon number measurements of PAFPs. The histograms are example photon number distributions of mEos3.2 (Fig. 1A), mMaple (Fig. 1B), PSCFP2 (Fig. 1C), and PAGFP (Fig. 1D) measured by imaging individual Zyxin-PAFP fusion proteins in live BS-C-1 cells. The mean photon numbers are indicated.

Table 1

PAFP	Pre/post-activation emission wavelength (nm) ^a	Photon number	On-off switching rate ratio	ClpP clustering ^b	Number of localizations per cell ^c	Maturation time (min) ^d
Dendra2	507 / 573	686	4.2×10^{-6}	-	1810	38
mEos2	519 / 584	745	2.9×10^{-6}	+	1290	340
mEos3.2	516 / 580	809	2.6×10^{-6}	-	1950	330
tdEos	516 / 581	774	3.2×10^{-6}	-	1800	330
mKikGR	515 / 591	599	4.1×10^{-6}	+	3800	31
PAmCherry	---- / 595	706	7.8×10^{-6}	+	4200	61
PAtagRFP	---- / 595	906	5.7×10^{-6}	-	760	200
mMaple	505 / 583	798	1.9×10^{-6}	+	24000	48
mMaple2	506 / 582	783	1.0×10^{-6}	+	21000	62
mMaple3	506 / 583	675	6.2×10^{-7}	-	12300	49
PAGFP	---- / 517	313	1.3×10^{-3}	-		<10
PSCFP2	468 / 511	223	8.1×10^{-6}	+		
Dronpa	---- / 517	262	5.8×10^{-3}	-		25
mGeosM	---- / 514	248	4.9×10^{-4}	+		<10

Properties of PAFPs. A 405 nm laser was used for photo-activation. The photon number and on-off switching rate ratio were measured in live BS-C-1 cells. The ClpP clustering, number of localizations per cell and maturation time were measured in live *E. coli* cells. ^aThe mMaple2 and mMaple3 emission wavelengths were measured in this work with purified proteins. ^b“+” indicates that ClpP-PAFP exhibits clustered distributions in at least a subset of cells, whereas “-” indicates that ClpP-PAFP does not

exhibit clustered distributions in any cells. ^cThe number of HU-PAFP localizations per *E. coli* cell. ^dMaturation time is defined as the half-life of the immature state.

Fewer photons were detected from the green PAFPs (PAGFP, PSCFP2, Dronpa, mGeosM) than from the red ones (Dendra2, mEos2, mEos3.2, tdEos, mKikGR, PAmCherry, PAtagRFP, mMaple). However, within the same color group, the difference in photon budget was less than two-fold.

In addition to the above listed fluorophores, rsFastLime and rsEGFP were also imaged. Both of these proteins gave relatively low photon budget (< 60 photons per switching event), which would lead to relatively poor localization precision. It is, however, worth noting that these proteins are excellent choices for a different mode of super-resolution imaging (RESOLFT) due to the large number of switching cycles that they exhibit before photobleaching.

On-off ratio of photoactivatable fluorescent proteins. To determine the on-off ratio, the rates for switching on and switching off (or photobleaching) the PAFPs were measured in the presence of imaging light only. By definition, the on-switching rate is the increment in probability of the on-switching events per unit time. To measure this quantity, the Zyxin-PAFP expressing cells were imaged in the super-resolution mode for a short period of time without any activation light (with imaging light only). The samples were then imaged to completion with an additional activation light at 405 nm. The ratio of the total number of activation events accumulated by a certain time during the period without activation light over the total number of activation events by the end of the imaging process was determined. The slope of this cumulative activation probability against time then gave the on-switching rate (Fig. 2A, B, Table 2). The off-switching rate was determined from the inverse of the mean lifetime of the on state of each PAFP in the presence of the imaging light (Fig. 2C, D, Table 2). The ratios between the on- and off-switching rates were determined for all twelve PAFPs (Table 1). Because both on- and off-switching rates scale linearly with the illumination intensity, the on-off ratio should be independent of the light intensity. The on-off ratios were generally very small (10^{-5} - 10^{-6}) except for PAGFP, Dronpa and mGeosM, which were $\sim 10^{-3}$.

Table 2.

PAFP	On-switching rate (frame ⁻¹)	Mean lifetime (frame)
Dendra2	5.91×10^{-8}	0.71
mEos2	4.14×10^{-8}	0.69
mEos3.2	3.50×10^{-8}	0.74
tdEos	4.70×10^{-8}	0.68
mKikGR	6.01×10^{-8}	0.68
PAmCherry	8.21×10^{-8}	0.96
PAtagRFP	4.95×10^{-6}	1.1
mMaple	2.54×10^{-6}	0.74
mMaple2	1.50×10^{-5}	0.67
mMaple3	9.23×10^{-7}	0.67
PAGFP	1.88×10^{-3}	0.70
PSCFP2	1.49×10^{-2}	0.54
Dronpa	9.66×10^{-4}	0.60
mGeosM	8.00×10^{-4}	0.61

On-switching rate and mean lifetime of PAFPs. These data were measured in live BS-C-1 cells transiently transfected with zyxin::PAFP plasmids. Each frame
 5 corresponds to 16 ms.

From the super-resolution images of Zyxin, it is evident that the PAFP with a smaller on-off ratio gave images with a higher localization density and hence higher image quality (Fig. 2E, F). This correlation can be understood based on the Nyquist
 10 sampling theorem, which suggests that the final resolution of an image is at least twice the average distance between the localized probes.

Dimerization tendency of photoactivatable fluorescent proteins. Dimerization or oligomerization of fluorescent proteins may cause undesired aggregation of the target proteins. Here, instead of measuring the dimerization affinity of PAFPs *in vitro*, whether
 15 the PAFPs could cause aggregation of a target protein was directly probed. In this approach, the fluorescent proteins are fused to the *E. coli* protease ClpP, which itself oligomerizes to form a tetradecameric complex. It has been suggested that ClpP proteins tend to aggregate and form a single visible punctum in *E. coli* when fused to a
 fluorescent protein with a substantial dimerization tendency, while fusion to a
 20 fluorescent protein with a weak or no dimerization tendency tends to display a diffuse localization pattern. *E. coli* codon-optimized PAFP sequences were fused to the chromosomal copy of *clpP*, and it was reported whether cells exhibit the single-punctum phenotype (Table 1, Fig. 3A). Fusion with mKikGR, mGeosM, mMaple, PAmCherry, PSCFP2, or mEos2 all led ClpP to form a single punctum in at least a subset of cells,

suggesting that these PAFPs exhibit appreciable dimerization tendency, whereas Dendra2, mEos3.2, tdEos, PAtagRFP, PAGFP and Dronpa did not show high enough dimerization tendency to drive detectable ClpP aggregation (Table 1 and Fig. 3A). As a cautionary note, because different target proteins have different expression levels and intrinsic oligomerization or polymerization tendencies, PAFPs that cause appreciable aggregation of ClpP may not necessarily cause aggregation of other target proteins, and vice versa.

Figure 2. On-off switching rate ratio measurements of PAFPs. Sample data are presented for PSCFP2 (Figs. 2A, 2C, 2E) and PAGFP (Figs. 2B, 2D, 2F). (Figs. 2A-2B) Cumulative on-switching probability as a function of time without activation light. The slope of the line gives the on-switching rate. (Figs. 2C-2D) Distribution of the on-state lifetime. Bars represent measured data, which were fitted with a binned exponential

function (dots) $P_n = \int_{n-1}^n \frac{1}{m} e^{-\frac{x}{m}} dx$, to take into account that lifetimes are rounded to the next larger integer. m is the mean lifetime, the inverse of which gives the off-switching rate. The on-off switching rate ratio is defined as the on-rate divided by the off-rate. Each frame corresponds to 16 ms. (Figs. 2E-2F) Representative super-resolution images of Zyxin-PSCFP2 and Zyxin-PAGFP in live BS-C-1 cells.

To further illustrate the potential effect of PAFP dimerization on target proteins, these were imaged: *E. coli* nucleoid-associated protein H-NS, *E. coli* chemotactic receptor Tar, and mammalian intermediate filament protein Vimentin fused to different PAFPs. H-NS appears as a few large and discrete clusters in cells when fused to mEos2, a previously reported monomeric version of the Eos fluorescent protein. Similar results were observed when H-NS was fused to another reported monomeric PAFP, PAmCherry. Given the residual dimerization tendency of mEos2 and PAmCherry detected by the ClpP assay (Fig. 3A), the study of H-NS using other PAFPs was extended. Notably, H-NS generally appeared as large and discrete clusters in cells when fused to PAFPs with appreciable dimerization tendency detected by the ClpP assay, such as mEos2, PAmCherry and mMaple. H-NS also appeared as large clusters when fused to tdEos, even though tdEos did not show sufficient dimerization to drive ClpP aggregation. On the other hand, fusion of H-NS to Dendra2, mEos3.2 and PAtagRFP, all without detectable dimerization tendency by the ClpP assay, appeared as much finer clusters that were more dispersed in the nucleoids (Fig. 3B). it was also observed the more dispersed

phenotype with immunofluorescence imaging of H-NS, but did not present this comparison here because fixation of the bacterial nucleoid is in general discouraged. It is known that H-NS oligomerizes/clusters even without fusion to fluorescent proteins and this intrinsic H-NS clustering effect brings gene loci in different chromosome regions into spatial proximity in wild type cells. However, fusion with some of the PAFPs may have exaggerated the H-NS clustering effect. Without prior knowledge of the dimerization tendency of these previously designated “monomeric” PAFPs, the large clusters of H-NS-mEos2 and H-NS-PAmCherry were earlier interpreted as an intrinsic structure of H-NS. Results here suggest that this interpretation might be incorrect because the clustering of H-NS is likely enhanced by the under-appreciated dimerization effect of mEos2 and PAmCherry, though it is formally possible that the more monomeric PAFPs might have disrupted intrinsic H-NS clustering by discouraging PAFP self-interactions.

As another example, Tar preferentially clustered at cell poles when fused to mEos2 and mKikGR (Fig. 3C), but spread out more evenly along the envelope of *E. coli* when fused to the more monomeric mEos3.2 (Fig. 3C), suggesting that the preferential cell pole distribution might have been exaggerated by the dimerization of mEos2 and mKikGR. Such aggregation effect could also apply to eukaryotic systems. For example, fusion to mEos2 and mKikGR caused Vimentin filaments to cluster into thick bundles in mammalian cells, whereas Vimentin-mEos3.2 appeared as thin filaments that were similar to immunofluorescence images of Vimentin in untransfected cells (Fig 3D).

Figure 3. Dimerization tendency of PAFPs and its effect on the distributions of target proteins. (Fig. 3A) Phase contrast (left) and conventional fluorescent images (right) of live *E. coli* cells expressing ClpP-PAFP fusions. PAFPs with substantial dimerization tendencies (mEos2, mMaple) result in the formation of ClpP puncta. PAFPs with little to no dimerization tendencies (mEos3.2, PAtagRFP) produce a diffusive ClpP distribution. (Fig. 3B) Overlaid super-resolution and phase contrast (gray) images of live *E. coli* cells expressing H-NS-PAFP fusions. (Fig. 3C) Overlaid super-resolution and phase contrast images of fixed *E. coli* cells expression Tar-PAFP fusions. (Fig. 3D) Super-resolution images of fixed Cos-7 cells expressing Vimentin-PAFP fusions or fixed untransfected Cos-7 cells with immunofluorescent labeling of Vimentin (IM). (Scale bars: 1000 nm.)

Signaling efficiency of photoactivatable fluorescent proteins. Even when the PAFP is fused to the target protein at its endogenous chromosomal locus, the number of detectable fluorescent proteins does not necessarily reflect the expression level of the fusion protein. This deficiency not only prevents quantitative analysis of the target protein, but may also reduce the resolution of the image by effectively decreasing the labeling density. Here the relative signaling efficiency of different PAFPs was compared by determining the number of single-molecule localizations per cell of the PAFPs fused to a common target protein under endogenous expression. the PAFPs were fused to the *E. coli* gene *hupA*, which encodes a subunit of the nucleoid-associated protein HU, at its endogenous chromosomal locus and determined the total number of HU-PAFP localizations per cell (Fig. 4 and Table 1). The four green PAFPs were too dim to be detected as single molecules in *E. coli*, while the eight red PAFPs showed similar photon budget to those measured in mammalian cells (data not shown). Among the red PAFPs, mMaple provided by far the largest number of HU localizations, which was 6 – 32 fold higher than those of other HU-PAFP fusion proteins (Fig. 4).

Figure 4. Signaling efficiency comparison among PAFPs. (Fig. 4A) Super-resolution images of live *E. coli* cells expressing HU-PAFP fusions overlaid with phase contrast images (gray). Super-resolution images were acquired till all the PAFP molecules in the field of view were bleached. (Fig. 4B) Numbers of observed localizations per cell for different HU-PAFP fusions. (Error bars represent standard errors of the means).

The difference in the number of HU-PAFP localizations for different PAFPs may be attributed to multiple reasons. One of the possible reasons could be the difference in the number of blinking (switching) events per fluorophore. the average number of blinking events were measured in fixed HU-PAFP expressing cells (Table 3). The average number of blinking events spanned only a small range from 1.7 to 3.3, which are similar to the results derived from purified PAFP *in vitro* (Fig. 6). The localization numbers divided by the average numbers of blinking events yielded the numbers of imaged molecules per cell (Table 3), which is still much bigger in the case of HU-mMaple expressing cells (4 – 36 fold higher than in other HU-PAFP expressing cells).

Figure 6. Blinking number distributions of mEos2, mMaple, mMaple2 and mMaple3 from *in vitro* measurements. The percentages of mEos2, mMaple, mMaple2 and mMaple3 that exhibit more than 1 emission bursts are 56%, 57%, 61%, and 60%

respectively. The average numbers of blinking (burst) events of the mEos2, mMaple, mMaple2 and mMaple3 molecules are 2.9, 2.8, 3.3 and 3.1, respectively.

A second possible reason for the difference in localization numbers could be the difference in HU-PAFP expression levels. However, quantitative western blot experiments showed that most HU-PAFP constructs are expressed at similar levels (Table 3), except that fusion with tdEos led to a large decrease in expression. the number of imaged molecules per cell was divided by the number of expressed fusion proteins per cell to calculate the percentage of PAFP imaged (Table 3). The percentage for mMaple imaged was still 5 – 22-fold higher than the other PAFPs. Therefore, neither the number of blinking events nor the expression level is a major contributing factor to the difference in the observed localization numbers.

Table 3.

PAFP	Average number of blinking events ^a	Number of imaged molecules per cell ^b	Expression level by quantitative western (molecules per cell) ^c	Percentage of PAFP imaged ^d
Dendra2	1.7	1060	27000	3.9%
mEos2	2.8	460	47000	1.0%
mEos3.2	3.0	650	38000	1.7%
tdEos	3.3	270	6600	4.0%
mKikGR	1.7	2200	82000	2.7%
PAmCherry	1.9	2200	61000	3.6%
PAtagRFP	1.7	450	50000	0.89%
mMaple	2.5	9700	49000	20%
mMaple2	2.7	7700	42000	18%
mMaple3	2.8	4400	42000	10%

15

Signaling efficiency of PAFPs in *E. coli* adjusted for the number of blinking events per fluorophore and the fusion protein expression level. ^aThe average number of blinking events per molecule measured in fixed *E. coli* cells expressing HU-PAFP. ^bThe number of imaged molecules per cell, calculated by dividing the number of localizations (Table 1) by the average number of blinking events. For tdEos, the result is further divided by two due to the presence of two fluorophores per molecule. ^cThe number of expressed HU-PAFP molecules measured with anti-HU antibody in quantitative western blot experiments. ^dThe percentage of PAFP imaged, calculated by dividing the number of imaged molecules per cell by the number of expressed molecules per cell. mMaple gives

the highest signaling efficiency of 20%. The 20% value is roughly consistent with the observations that ~60% maturable mMaple have matured under the imaging conditions in *E. coli*, ~70% of mature mMaple can be converted from the green form to the red form, and ~80% of the converted molecules are bright enough to be localized by the analysis software. Based on the above observations, approximately one third of mMaple molecules are expected to be localized. The left-over discrepancy could be due to that not all PAFPs are folded into the fluorescent conformation.

Another potential reason for the difference in localization numbers is the difference in the fluorescent protein maturation time. Because *E. coli* has a relatively short doubling time under these growth conditions, a substantial fraction of the PAFP molecules may not have sufficient time to mature and become fluorescent. To test this idea, the *in vivo* maturation time of several PAFPs was measured by using kanamycin to block protein synthesis in *E. coli* cells expressing HU-PAFP, and then recording the increase in cellular fluorescence due to PAFP maturation (Fig. 7, Table 1). Indeed, compared to some of the PAFPs with low signaling efficiencies, mMaple has a substantially faster maturation time. However, the difference in maturation time was substantially smaller than and hence not sufficient to account for the observed difference in the localization numbers (see below and Fig. 8).

Figure 7. Maturation half-life measurements. *E. coli* cells expressing HU-mEos2 (Fig. 7A) or HU-mMaple (Fig. 7B) were immobilized on a coverslip and treated with kanamycin to inhibit protein synthesis. Further increase in cellular fluorescence was caused by maturation of existing fluorophores. The average fluorescence per cell pixel was plotted as a function of time (dots) and fit to an exponential curve (black line) to get the maturation half-life.

Figure 8. PAFP signaling efficiency versus maturation half-life. The numbers of imaged molecules per cell after adjustment for the number of blinking events per molecule are plotted against the maturation half-lives. The numbers of molecules in the figure are from *E. coli* cells grown at 32 °C to match the temperature used for maturation half-life measurements. As a result the numbers are different from those reported in Table 3 (37 °C). The black line indicates the expected number of observable molecules at each half-life value if the maturation time of the fluorophore is the only variable causing the difference in the signaling efficiency. This line was derived from the number of

imaged mMaple molecules and Eq. 5 (see below), assuming equal expression rate. The discrepancy between the line and the measured dots shows that the difference in maturation half-life cannot fully account for the difference in signaling efficiency.

Other potential contributing factors include the fraction of properly folded PAFP at equilibrium, the fraction of folded PAFP that can ultimately mature, the fraction of folded and mature PAFP that can be photoactivated, and the fraction of activated PAFP that are sufficiently bright to image. Given that different red PAFPs tend to have similar photoactivation efficiencies and photon budgets (Table 1), the last two factors are less likely explanations for the observed range of signaling efficiencies. It is also possible that the signaling efficiency depends on the imaging conditions, among other factors. More experiments are required to test these possibilities.

Finally, the signaling efficiency of the PAFPs was measured in mammalian cells by transient transfection of A549 cells with the *zyxin::PAFP* constructs. For fast data collection, conventional fluorescent microscopy was used and cells were imaged till the PAFPs were fully bleached. The number of imaged molecules in each cell was estimated by dividing the total photon count from the cell by the product of the average photon number per switching event (Table 1) and the average number of blinking events for each PAFP (Table 4). The results indicate that mMaple again offers much higher signaling efficiency (7-70 fold higher) than the other red PAFPs, even in mammalian cells. Among the green PAFPs, PSCFP2 offers a comparable signaling efficiency to that of mMaple (Table 4).

New PAFPs with high signaling efficiency and low dimerization tendency. The above results show that mMaple has a much higher signaling efficiency than all of the other tested PAFPs. However, the dimerization tendency of mMaple could lead to aggregation effects on the target proteins. On the other hand, Dendra2, mEos3.2, and PAtagRFP exhibit undetectable dimerization tendency, but have low signaling efficiency. It is thus desirable to develop a new PAFP that has both high signaling efficiency and low dimerization tendency. To this end, two new PAFPs were engineered by introducing point mutations into mMaple designed to destabilize the dimerization of this protein.

Two mutations make mEos3.2 more monomeric than mEos2: I102N and Y189A. Based on a sequence alignment between mEos2 and mMaple, comparable mutations were made, I111N and Y198A, in mMaple. The properties of this new protein were

tested. The protein was termed mMaple2. The properties were tested by fusing it to Zyxin, ClpP, and HupA and performing measurements on the photon budget, on-off ratio, dimerization tendency and signaling efficiency as described above. mMaple2 exhibited a similar photon budget, on-off ratio and signaling efficiency as those of mMaple (Fig. 5A, B, C, Table 1, 3, 4). ClpP-mMaple2 proteins still formed puncta in some cells. However, the percentage of cells with puncta was significantly reduced in comparison to ClpP-mMaple, suggesting a lower dimerization tendency of mMaple2 (Fig. 5D).

Figure 5. mMaple2 and mMaple3 exhibit both high signaling efficiency and low dimerization tendency. (Figs. 5A-5D) Photon budget (Fig. 5A), on-off switching rate ratio (Fig. 5B), signaling efficiency (Fig. 5C) and dimerization tendency (Fig. 5D) of mMaple2 and mMaple3 in comparison with mMaple. Insets in (Fig. 5D): Sample fluorescent images of ClpP-mMaple2 and ClpP-mMaple3 expressing *E. coli*. Error bars are standard errors of the means and are too small to be visualized in (Fig. 5A). (Fig. 5E) H-NS appears more spread out in *E. coli* cells when fused to mMaple3 in comparison to the mMaple and mMaple2 fusion proteins. (Fig. 5F) Tar-mMaple2 and Tar-mMaple3 appear more distributed along the cell envelope and less concentrated at the polar caps than Tar-mMaple. (Fig. 5G) Vimentin-mMaple2 and Vimentin-mMaple3 are less bundled than Vimentin-mMaple. (Scale bars: 1000 nm.)

Table 4.

PAFP	Average number of blinking events ^a	Number of imaged molecules per cell (millions) ^b
Dendra2	2.3	2.4
mEos2	2.2	1.4
mEos3.2	2.4	1.9
tdEos	3.8	0.30
mKikGR	1.6	1.2
PAmCherry	1.5	3.2
PAtagRFP	1.6	0.78
mMaple	2.4	21
mMaple2	2.2	16
mMaple3	2.0	16
PAGFP	1.8	2.0
PSCFP2	1.6	22
Dronpa	1.8	9.6
mGeosM	2.6	6.0

Signaling efficiency of PAFPs in mammalian cells. fixed A549 cells transfected with the *zyxin::PAFP* plasmids were imaged. ^aThe average number of blinking events per molecule measured in fixed A549 cells expressing Zyxin-PAFP. ^bThe number of imaged molecules per cell, defined as the total photon count measured from the cell until cells were fully photobleached, divided by the mean photon budget of each blinking event (Table 1) and then by the average number of blinking events. For tdEos, the result was further divided by 2.

10

To further reduce the residual dimerization tendency of mMaple2, a series of mutations were screened to residues that are predicted to be solvent-exposed near residues 111 and 198. Certain charged residues were switched to the opposite charge. One of the derivatives, with mutations E82R, D83K and D197K, exhibited undetectable dimerization tendency when fused to ClpP (Fig. 5D, Table 1). This derivative, termed mMaple3, again had a similar photon budget to that of mMaple and even a lower on-off ratio when compared to mMaple (Fig. 5A, B, Table 1). Its signaling efficiency was only moderately reduced from those of mMaple or mMaple2 (Fig. 5C, Table 1, 3, 4).

Aggregation effects of mMaple2 and mMaple3 were tested on H-NS, Tar and Vimentin. H-NS-mMaple3 showed substantially more spread out H-NS distributions

20

than H-NS-mMaple, though H-NS-mMaple2 still appeared as a few discrete clusters in each cell (Fig. 5E). Tar-mMaple2 and Tar-mMaple3 were substantially less concentrated at the cell poles than Tar-mMaple (Fig. 5F). Vimentin-mMaple2 and Vimentin-mMaple3 filaments were much less bundled than Vimentin-mMaple filaments (Fig. 5G). All of these observations are consistent with the reduced dimerization tendency of mMaple2 and the undetectable dimerization tendency of mMaple3.

The excitation and emission spectra of mMaple2 and mMaple3 were similar to those of mMaple (Table 1, Fig. 9).

Figure 9. Excitation and emission spectra of mMaple, mMaple2 and mMaple3 before and after photo-conversion. The dashed lines are excitation spectra. The solid lines are emission spectra. The left lines are measurements before photo-conversion. The right lines are measurements after photo-conversion. The spectra are nearly identical among the three PAFPs.

In these experiments, four properties of PAFPs were characterized that are important for super-resolution imaging based on single-molecule switching and localization. First, the photon budgets are not substantially different for different PAFPs within the same color group, but the red PAFPs provide substantially more photons (600-900 photons per activation event) than the green PAFPs (200-300 photons). Even the red PAFPs' photon budgets are substantially lower than those of the bright photoswitchable/photoactivatable dyes (several thousand to one million photons) (7, 32). Given that the localization precision scales approximately linearly with the inverse square root of the photon numbers, the PAFPs thus give substantially lower localization precision than photoswitchable dyes. Second, the on-off ratios are also not substantially different for different PAFPs within the same color group, but the red PAFPs tend to show much lower on-off ratio ($10^{-5} - 10^{-6}$) than the green PAFPs ($\sim 10^{-3}$). PSCFP2 is a noticeable exception with an on-off ratio similar to the red PAFPs. The on-off ratios of the red PAFPs are substantially lower than those of the popularly used photoswitchable dyes ($10^{-3} - 10^{-4}$), and hence PAFPs can provide substantially higher localization density. Third, natural fluorescent proteins tend to dimerize or tetramerize. Although mutant fluorescent proteins have been made to reduce dimerization, many of the so-called monomeric fluorescent proteins still have some residual dimerization tendency and thus can cause undesired aggregation and mislocalization of the target proteins to which they are fused. Among the twelve PAFPs tested here, the majority appeared to exhibit a

substantial propensity towards aggregation or bundling of target proteins. The exceptions are Dendra2, mEos3.2, PAtagRFP, PAGFP and Dronpa, which did not show a detectable aggregation effect for the target proteins tested here. It is important to note, though, that these results do not imply that such aggregation will not happen when these PAFPs are fused to other target proteins with strong intrinsic clustering or polymerization tendency. As another cautionary note, it is also possible that monomeric fluorescent proteins that discourage self-interactions may disrupt natural oligomerization of the target proteins. Thus it is a good practice to verify the spatial organization derived from PAFPs with alternative approaches, such as immunohistochemistry or by using SNAP/CLIP/HALO or short peptide tags to label proteins with dyes. The fourth property that was probed is the signaling efficiency, which determines the number of detectable molecules of the PAFP fusion protein for a given target protein. When the signaling efficiency is low, the number of localizations is not sufficient to map out the fine organization of the target protein. Notably, among the eight red PAFPs tested, mMaple has the highest signaling efficiency, which is about one order of magnitude higher than the other PAFPs. The downside of mMaple is, however, its relatively high dimerization tendency.

To overcome this problem, two new PAFPs, mMaple2 and mMaple3 were developed, which largely maintained the superb signaling efficiency of mMaple but exhibited substantially reduced dimerization tendency. In particular, mMaple3 exhibited no detectable dimerization tendency. The photon numbers and on-off ratios of both are similar to those of mMaple, which are among the best for PAFPs.

mMaple3 provides excellent performance in all four key properties described here, including high signaling efficiency, low dimerization tendency, high photon budget and low on-off ratio. In the cases that the localization number provided by mMaple3 is sub-optimal, one should also consider labeling with mMaple2, which will give substantially higher localization numbers, but it is important to check whether fusion with mMaple2 has led to undesired aggregation effect. The large number of localizations provided by mMaple2 and mMaple3 will not only facilitate mapping out the fine spatial distribution of the target protein, but will also allow the localizations to be divided into more snapshots to facilitate time-lapsed imaging of the dynamics of cellular structures. When localization number is less of a concern, mEos3.2 and PAtagRFP are excellent choices for their high photon budget and undetectable dimerization tendency.

The four properties measured in this work are four key properties of PAFPs to consider when imaging the spatial organization of a target protein with single-molecule-based super-resolution imaging. However, they are not the only important properties to consider when other experimental requirements need to be taken into account. For example, spectral properties are essential for multi-color imaging. For two-color imaging, green PAFPs are particularly useful (when paired with a red PAFP) even though they generally give lower photon numbers and higher on-off ratios. Dark-to-fluorescent PAFPs are also more favorable than color-changing PAFPs as the former take a smaller spectral space. The number of switching cycles is another important parameter to consider. For quantifying the stoichiometry of the target proteins, it is desirable to have PAFPs that can be switched on only once though almost all PAFPs tend to blink more than once. On the other hand, for mapping of protein spatial organization, a larger number of switching cycles per molecule would allow the target structure to be sampled multiple times, which can improve the image quality. This property is also advantageous when it comes to time-lapsed imaging for probing dynamics. For a different super-resolution imaging method, RESOLFT, reversibly switching fluorescent proteins with a large number of switching cycles (such as rsEGFP) may be required.

New plasmids were constructed with PCR and isothermal assembly. *E. coli* chromosomal insertions were created by lambda RED recombination. *E. coli* cells were grown to exponential phase in M9 minimal media before imaging. Mammalian cell lines were transfected with purified plasmids using nucleofection or lipofection, and incubated for 24-26 hours before imaging. Phase-contrast, super-resolution or conventional fluorescence images were collected on an Olympus IX-71 inverted microscope with 405-, 488-, and 561-nm laser lines. Images were analyzed with custom-written software. See below for details.

Plasmids and strains. All new plasmids in this work were cloned with PCR and isothermal assembly. All of the *E. coli* chromosomal DNA constructs were cloned with lambda RED recombination. The *zyxin::PAFP* fusions were cloned with a Clontech pEYFP-N1 backbone by replacing the *eyfp* with *PAFP* sequences and inserting *zyxin* as well as a six-amino-acid linker (DPPVAT) (SEQ ID NO: 4) to the 5' end of *PAFP* sequences. The other *vimentin::PAFP* plasmids were cloned based on the same design. For *E. coli* protein fusions, *E. coli* codon-optimized *PAFP* sequences with a seven-amino-acid linker (SGGGGSK) (SEQ ID NO: 5) at the 5' ends were inserted into

plasmid pBAD18, followed by a chloramphenicol resistance marker (*cam*). There are also two *frt* sites flanking the *cam* region. The *linker::PAFP-frt-cam-frt* sequences were then fused to the chromosomal copies of *E. coli* genes *clpP*, *hns*, or *hupA* in Strain BW25993. The *cam* markers in the *clpP* fusion constructs were later removed by Flp-*frt* recombination. The plasmids expressing Tar-PAFPs were designed with a pTrcHis2a (Life Technologies) backbone and direct fusion of *PAFP* sequences to the 3' end of *tar*. The plasmids for PAFP protein purification were cloned by inserting *PAFP* sequences into a pET-3a backbone (EMD Millipore), with an extra 10-amino-acid sequence after the start codon. The extra sequence codes for 6 histidines and a linker (CHHHHHHGSG) (SEQ ID NO: 6). The plasmid for HupA protein purification was cloned by inserting into pET-3a the *hupA* coding sequence, followed by a 3-amino-acid linker and 6 histidine codons at the 3' end (GSGHHHHHH) (SEQ ID NO: 7). All newly constructed plasmids and chromosomal insertions have been verified by sequencing.

Imaging sample preparation. To measure the photon budget and on-off ratio of PAFPs, BS-C-1 cells grown at 37 °C in a 5% CO₂ atmosphere in Eagle Modified Minimum Essential Medium (ATCC) supplemented with 10% fetal bovine serum (Invitrogen) were transfected with pZyxin-PAFP plasmids using Amaxa cell line nucleofector kit V (Lonza) according to the manufacturer's protocol, and plated in LabTek II 8-well chambered coverglass (Nunc). 24 hrs after transfection, live cells were washed with Dulbecco's Phosphate-buffered Saline (DPBS) and imaged.

To characterize the dimerization tendency and signaling efficiency of PAFPs in live bacteria, *E. coli* strains expressing the ClpP-PAFP, H-NS-PAFP or HU-PAFP fusions were inoculated into lysogeny broth (LB), and grown to saturation overnight while shaking at 37 °C (unless otherwise stated). The LB cultures were then diluted 2000 times into M9 minimal media supplemented with 0.4% glucose, MEM amino acids and vitamins (Invitrogen), and further grown at 37 °C for ~5 hrs to an OD_{600nm} of 0.15 (unless otherwise stated). 1 mL of each culture was concentrated 40 times by centrifugation and plated on an agarose gel pad, made with 3% SeaPlaque GTG agarose (Lonza) melted in the M9 minimal media above. (For ClpP imaging, the cell cultures were treated with 20 micrograms/ml chloramphenicol for 24 hrs at room temperature before plating on agarose pads to allow sufficient fluorophore maturation.) Coverslips pre-cleaned in a plasma sterilizer were then placed on the pads. VALAP was used to seal

the samples before microscopy. Tar-PAFP cell samples were prepared using known techniques.

To probe the dimerization effect of PAFPs in mammalian cells, Cos7 cells grown at 37 °C in a 5% CO₂ atmosphere in Dulbecco's Modified Eagle Medium (Gibco) supplemented with 10% fetal bovine serum and MEM non-essential amino acids (Gibco) were transfected with the pVimentin-PAFP plasmids using Lipofectamine 2000 (Life Technologies) according to the manufacturer's protocol. 24 hrs after transfection, cells were fixed with 2% paraformaldehyde in PHEM buffer (25 mM HEPES, 10 mM EGTA, 60 mM PIPES, 2 mM MgCl₂, pH 6.9) for 15 minutes, rinsed three times in PHEM buffer, and imaged in DPBS.

To measure the signaling efficiency of PAFPs in mammalian cells, A549 cells grown in the same condition as Cos7 were transfected with pZyxin-PAFP plasmids using Lipofectamine 2000 according to the manufacturer's protocol. A549 cells were used for their higher transfection efficiency as compared to BS-C-1. 26 hrs after transfection, cells were fixed with 2% paraformaldehyde in PHEM buffer for 15 minutes. Cells were then rinsed three times in PHEM buffer, and imaged in DPBS.

Optical setup and imaging conditions. An Olympus IX-71 inverted microscope was used with a 100x UPlanSApo, NA 1.40, oil immersion phase objective, and an active sample stabilization system. A 405 nm laser (CUBE 405-50C, Coherent) was used for photo-activation. A 488 nm laser (Sapphire 488-200CW, Coherent) was used to excite and image the green PAFPs. A 561 nm laser (Sapphire 561-200CW) was used to excite and image the red PAFPs. The power densities at the sample, with epi-fluorescence illumination, were 0-1 kW/cm² at 405 nm, 390 W/cm² at 488 nm, and 1.8 kW/cm² at 561 nm. The illumination was further adjusted to near-TIRF (for bacteria) or TIRF (for mammalian cells) configurations, which increased the power density by another factor of 3-4. On the emission path, a Di01-R561 (Semrock) dichroic mirror and an FF01-617/73 (Semrock) band pass emitter for the red PAFPs was used, or a ZT488RDC (Chroma) dichroic mirror and an ET525/50m (Chroma) band pass emitter for the green PAFPs. Movies of fluorescence images were acquired at 60 Hz. In most cases, super-resolution imaging was done with simultaneous illumination of the activation and excitation light, with gradual increase in the activation light power. For the on-off ratio measurements, super-resolution imaging of the Zyxin-PAFP samples was performed first without activation for 30 sec, and then gradually increased the activation

power till the samples were fully imaged. For conventional fluorescence imaging, samples were simultaneously illuminated with the activation and excitation light.

Image analysis. The image processing procedure of the super-resolution images used known techniques. Conventional fluorescence images of ClpP-PAFP were obtained
5 by averaging 120 frames of the 60-Hz movie. Other analyses, including cell segregation, punctum detection, statistical calculation, and curve fitting, were performed with MATLAB R2012a.

Maturation half-life measurements. Single colonies of the bacterial strains expressing HupA-PAFP fusions were inoculated into MOPS rich defined medium
10 (Teknova EZ Rich; M2105) supplemented with 0.2% w/v glucose and 34 micrograms/mL chloramphenicol and grown to saturation overnight while shaking at 32 °C. This culture was diluted 1 in 200 into fresh medium and grown for 2 hours at 32 °C. Cells were harvested at an OD600 of 0.2-0.3. A single channel of a fluidics chamber (Ibidi μ -Slide VI 0.4; 80606) was coated with 1% polyethylenimine (PEI, Aldrich) for 15
15 minutes at room temperature then washed twice with fresh medium. Freshly harvested cells were added to the channel, and the chamber was spun at 200G for 30 s to press the cells flat against the surface. The chamber was then washed five times with MOPS minimal defined medium (Teknova M2106) supplemented with 0.2% w/v glucose. This medium was also supplemented with 500 microgram/mL kanamycin to completely
20 inhibit protein synthesis.

Cells were imaged on the same microscope used for STORM imaging and were maintained at 32-33 °C by heating the objective to 33.0 °C with an objective heater (Biopetechs Objective Heater System). For each maturation time point, an initial phase contrast image of each 40x40 micrometer field of view was collected, then the field of
25 view was exposed to 5 s of $\sim 60 \text{ W/cm}^2$ (1.1 mW across the field of view) of 405 nm light to activate the PAFPs. A conventional fluorescence image was then collected using 561 nm for the red PAFPs and 488 nm for the green PAFPs. Each field of view was only imaged once to avoid complications due to partial activation or photobleaching. Cells were identified via the phase contrast images and the average fluorescence per cell pixel
30 was calculated as a function of time. This fluorescence was fit to determine the half-life for the maturation for each fluorophore.

Protein purification and spectral measurements. Rosseta 2(DE3) competent *E. coli* cells (EMD Millipore) were transformed with pET3a-6xHis-PAFP plasmids, and

colony purified. 15 mL overnight culture of each transformant was diluted into 1 L of LB+ampicillin (50 micrograms/mL), and incubated at 37 °C with shaking till the OD_{600nm} reached 0.8. Next 0.8 mM of isopropyl beta-D-1-thiogalactopyranoside (IPTG) was added, and the culture was further incubated at 37 °C with shaking for 5 hrs.

5 Cells were harvested with centrifugation, and resuspended in 15 mL of native wash buffer (50 mM Tris, pH 8.0 / 200 mM NaCl / 5 mM Imidazole / 10 mM βME / 0.5 mM benzamidine) with half of a tablet of protease inhibitor (Roche). Cells were lysed by sonication and the debris was spun down by centrifugation. 4mL of Ni-NTA agarose (Qiagen) were washed in sterile water by centrifugation, applied to the supernatant of

10 cell lysate, and incubated at 4 °C for >1 hr. The Ni-NTA particles were then washed with 15 mL of native wash buffer by centrifugation, and transferred into a chromatography column (Bio-Rad, Poly-Prep). The particle matrix was washed with an additional 40 mL of native wash buffer by gravity flow, washed 1 final time with 1 mL of elution buffer (100 mM NaCl / 300 mM Imidazole pH 7.0 / 10 mM beta-ME), and then eluted in 1 mL

15 fractions with elution buffer. The fractions containing PAFP were combined and dialyzed overnight in dialysis buffer (20 mM Tris, pH 7.5 / 100 mM NaCl / 1 mM EDTA / 10 mM beta-ME). The excitation and emission spectra of the purified proteins were measured with a Cary Eclipse fluorescence spectrophotometer. Photo-conversion of the purified proteins was achieved by shining the 405-nm imaging laser on the protein

20 solution.

Blinking characterization. To characterize fluorophore blinking *in vitro*, biotinylated PAFP molecules were sparsely immobilized on a streptavidin coated coverslip in a flow chamber. Purified PAFPs were biotinylated with ImmunoProbe Biotinylation Kit (Sigma-Aldrich) according to the manufacturer's protocol. Flow

25 chambers and streptavidin coated coverslips were prepared. The biotinylated proteins were diluted in phosphate buffered saline (PBS) before being added into the chamber. After 1 min, the chamber was washed with PBS to remove unbound molecules. Immobilized PAFPs were imaged with the same microscope and settings as in previous super-resolution imaging. The density of immobilized molecules was kept low such that

30 individual molecules were well resolved from each other. The numbers of blinking events were extracted with custom written programs.

To measure the number of blinking events *in situ*, fixed *E. coli* cells expressing HU-PAFP or fixed A549 cells expressing Zyxin-PAFP were imaged with the super-

resolution settings until the majority of fluorophores were bleached and the unbleached fluorophores left were well-separated in space. From this point, the samples were treated as sparsely immobilized PAFPs as described above. For *E. coli* fixation, cells were resuspended in 4% paraformaldehyde in PBS, incubated for 15 min at room temperature, and washed once with PBS. The fixed *E. coli* cells were imaged on agarose pads made with PBS.

Quantitative western blot. HupA protein solution used as concentration standard were prepared by transforming BL21(DE3)pLysS competent cells (Invitrogen) with the pET3a-hupA-6xHis plasmid. Protein purification procedure was the same as that described in the PAFP purification section. The size and purity of the protein were verified by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and SYPRO Red staining (Invitrogen). The concentration was determined with a bovine serum albumin standard.

For the western blot, *E. coli* cells with *hupA::PAFP* fusions were grown under the same condition as in the imaging experiments. The cell concentrations were estimated by optical density measurements and LB agar plating. 1 mL of cells were resuspended in 97 microliters of T200 buffer (50 mM Tris, pH 8 / 200 mM NaCl) plus 1 microliters of TURBO DNase (Life Technologies) and 2 microliters of Ready-Lyse Lysozyme (Epicentre). Cells were lysed by three freeze-thaw cycles in liquid nitrogen and 37 °C water bath. The HupA concentration standard or the lysed cell content solutions were mixed with 2x Laemmli Sample Buffer (BioRad) with a 1:1 ratio, heat-denatured, separated by a 4-15% Tris-HCl Ready Gel PAGE gel (BioRad), and transferred to a Hybond-P membrane (GE Healthcare). The membrane was blocked with 5% milk (Nestle) in DPBS plus 0.1% Tween 20 (PBST) for 1 hr at room temperature. Next, polyclonal anti-HU primary antibody (provided by Professor Jon Kaguni) was diluted 1:30,000 into the 5% milk in PBST and added to the membrane. After overnight incubation at 4 °C, the membrane was washed three times in PBST (7 min each). Then the secondary antibody -- ECL HRP-linked donkey-anti-rabbit IgG (GE Healthcare) was diluted 1:10,000 into 5% milk in PBST, added to the membrane, and incubated for 45 min at room temperature. After three more washes in PBST (7 min each), chemiluminescence signal was generated by Lumigen TMA-6 Reagent, and detected with an AlphaInnotech ChemiImager.

Immunofluorescence sample preparation. For the immunofluorescence imaging of Vimentin, untransfected Cos7 cells were fixed as in the PAFP imaging experiments. Then cells were blocked with a blocking buffer (DPBS with 3% BSA, 0.5% Triton-X-100) for 15 min, and treated with ~1 microgram/mL mouse anti-Vimentin primary
 5 antibody (Invitrogen) in the blocking buffer for 30 min at room temperature. The cells were washed three times (10 min each) with a wash buffer (DPBS with 0.2% BSA and 0.1% Triton-X-100), and then treated with ~2.5 microgram /mL donkey anti-mouse secondary antibody conjugated with Alexa 647 (Jackson ImmunoResearch) in the
 10 blocking buffer for 30 min at room temperature. Afterwards, cells were washed three times (10 min each) with the wash buffer and one more time with DPBS, and post-fixed with 3% paraformaldehyde and 0.1% glutaraldehyde for 10 min at room temperature. After three more washes with DPBS, the cells were imaged in the imaging buffer (DPBS with 1% β ME, 0.8% glucose, 1 mg/mL glucose oxidase and 40 microgram /mL catalase).

An analytic model describing the dependence of the concentration of mature
 15 PAFPs on the maturation half-life. If I is the cellular concentration of immature PAFP, M is the cellular concentration of mature PAFP, E is the expression rate of new, immature PAFP, km is the maturation rate, and kd is the dilution rate due to cell growth, the dynamics of I and M can be described with the following differential equations:

$$\frac{dI}{dt} = E - k_m I - k_d I; \quad [1]$$

$$\frac{dM}{dt} = k_m I - k_d M. \quad [2]$$

20 If it is assumed that the cellular concentration of PAFPs are stable, in other words

$$\frac{dI}{dt} = 0 \quad \text{and} \quad \frac{dM}{dt} = 0, \text{ this yields:}$$

$$M = \frac{Ek_m}{k_d(k_d + k_m)}. \quad [3]$$

km is inversely proportional to the PAFP maturation half-life (τm). kd is inversely
 25 proportional to the doubling time (τd) of *E. coli* under these growth condition.

The fraction of PAFP that has matured is:

$$\frac{M}{M+I} = \frac{k_m}{k_d+k_m} = \frac{\tau_d}{\tau_d+\tau_m}. \quad [4]$$

For example, the maturation half-life of mMaple at 32 °C was measured to be 48 min and the doubling time of *E. coli* at 32 °C in the M9 media was determined to be 74 min. So the mature fraction for mMaple is 0.61.

- 5 The expected ratio of mature fluorophores between two different PAFPs can be written as:

$$\frac{M_1}{M_2} = \frac{E_1(\tau_{m2}+\tau_d)}{E_2(\tau_{m1}+\tau_d)}. \quad [5]$$

- For example, the maturation half-life of mEos2 and mMaple at 32 °C are 340 and 48 min respectively. The expression levels of the two are similar (Table 3). The expected ratio of
 10 the cellular concentrations of the mature fluorophores should be $M_{mEos2} / M_{mMaple} = 0.29$. However, the measured ratio of the numbers of imaged mEos2 versus mMaple molecules per cell is 0.08 (Fig. 8). This indicates that the difference in PAFP maturation half-life is not sufficient to account for the difference in signaling efficiency.

- When PAFPs are expressed in mammalian cells, the mature fraction is expected
 15 to be close to 1 due to the very long doubling time of the cells (~22 hrs for A549). As a result, the signaling efficiency of the slowly maturing PAFPs (e.g. mEos2) relative to that of the fast-maturing mMaple should improve in the mammalian cells. However, the improvement may not be large. If it is assumed the maturation half-lives of the PAFPs in mammalian cells are similar to those measured in *E. coli*, the expected improvement of
 20 the signaling efficiency of mEos2 relative to that of mMaple should only be ~2.8-fold. Moreover, because maturation time may depend on the cell environment, protein target, incubation temperature, etc., the observed amount of improvement may differ. Indeed, the observed improvement for mEos2 relative to mMaple was only 1.4-fold when comparing mammalian cell results to the results obtained with *E. coli* cells (Table 3, 4).
 25 This change is not big enough to alter the conclusion that mMaple and its derivatives offer much higher signaling efficiency than the other tested red PAFPs even for mammalian cells.

- While several embodiments of the present invention have been described and
 30 illustrated herein, those of ordinary skill in the art will readily envision a variety of other

means and/or structures for performing the functions and/or obtaining the results and/or one or more of the advantages described herein, and each of such variations and/or modifications is deemed to be within the scope of the present invention. More generally, those skilled in the art will readily appreciate that all parameters, dimensions, materials, and configurations described herein are meant to be exemplary and that the actual parameters, dimensions, materials, and/or configurations will depend upon the specific application or applications for which the teachings of the present invention is/are used. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. It is, therefore, to be understood that the foregoing embodiments are presented by way of example only and that, within the scope of the appended claims and equivalents thereto, the invention may be practiced otherwise than as specifically described and claimed. The present invention is directed to each individual feature, system, article, material, kit, and/or method described herein. In addition, any combination of two or more such features, systems, articles, materials, kits, and/or methods, if such features, systems, articles, materials, kits, and/or methods are not mutually inconsistent, is included within the scope of the present invention.

All definitions, as defined and used herein, should be understood to control over dictionary definitions, definitions in documents incorporated by reference, and/or ordinary meanings of the defined terms.

The indefinite articles “a” and “an,” as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean “at least one.”

The phrase “and/or,” as used herein in the specification and in the claims, should be understood to mean “either or both” of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with “and/or” should be construed in the same fashion, i.e., “one or more” of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the “and/or” clause, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, a reference to “A and/or B”, when used in conjunction with open-ended language such as “comprising” can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other

than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

As used herein in the specification and in the claims, “or” should be understood to have the same meaning as “and/or” as defined above. For example, when separating
5 items in a list, “or” or “and/or” shall be interpreted as being inclusive, i.e., the inclusion of at least one, but also including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as “only one of” or “exactly one of,” or, when used in the claims, “consisting of,” will refer to the inclusion of exactly one element of a number or list of elements. In general,
10 the term “or” as used herein shall only be interpreted as indicating exclusive alternatives (i.e. “one or the other but not both”) when preceded by terms of exclusivity, such as “either,” “one of,” “only one of,” or “exactly one of.” “Consisting essentially of,” when used in the claims, shall have its ordinary meaning as used in the field of patent law.

As used herein in the specification and in the claims, the phrase “at least one,” in
15 reference to a list of one or more elements, should be understood to mean at least one element selected from any one or more of the elements in the list of elements, but not necessarily including at least one of each and every element specifically listed within the list of elements and not excluding any combinations of elements in the list of elements. This definition also allows that elements may optionally be present other than the
20 elements specifically identified within the list of elements to which the phrase “at least one” refers, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, “at least one of A and B” (or, equivalently, “at least one of A or B,” or, equivalently “at least one of A and/or B”) can refer, in one embodiment, to at least one, optionally including more than one, A, with no B present (and optionally
25 including elements other than B); in another embodiment, to at least one, optionally including more than one, B, with no A present (and optionally including elements other than A); in yet another embodiment, to at least one, optionally including more than one, A, and at least one, optionally including more than one, B (and optionally including other elements); etc.

30 It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one step or act, the order of the steps or acts of the method is not necessarily limited to the order in which the steps or acts of the method are recited.

- In the claims, as well as in the specification above, all transitional phrases such as “comprising,” “including,” “carrying,” “having,” “containing,” “involving,” “holding,” “composed of,” and the like are to be understood to be open-ended, i.e., to mean including but not limited to. Only the transitional phrases “consisting of” and
- 5 “consisting essentially of” shall be closed or semi-closed transitional phrases, respectively, as set forth in the United States Patent Office Manual of Patent Examining Procedures, Section 2111.03.

CLAIMS

What is claimed is:

1. A composition comprising mMAPLE2.
- 5 2. A composition comprising mMAPLE3.
3. A composition comprising SEQ ID NO: 2.
4. A composition comprising positions 2-237 of SEQ ID NO: 2.
- 10 5. A composition comprising SEQ ID NO: 1 with the provisos that position 111 is any amino acid residue except I and position 198 is any amino acid residue except Y.
6. The composition of claim 5, wherein position 111 is a basic amino acid.
- 15 7. The composition of claim 5, wherein position 111 is a polar amino acid.
8. The composition of claim 5, wherein position 111 is an acidic amino acid.
- 20 9. The composition of claim 5, wherein position 111 is N.
10. The composition of claim 5, wherein position 111 is H, K, or R.
11. The composition of claim 5, wherein position 111 is S, T, Y, Q, N, C, or G.
- 25 12. The composition of claim 5, wherein position 111 is D or E.
13. The composition of any one of claims 5-12, wherein position 198 is a basic amino acid.
- 30 14. The composition of any one of claims 5-12, wherein position 198 is an acidic amino acid.

15. The composition of any one of claims 5-12, wherein position 198 is a nonpolar amino acid.
16. The composition of any one of claims 5-12, wherein position 198 is A.
17. The composition of any one of claims 5-12, wherein position 198 is H, K, or R.
18. The composition of any one of claims 5-12, wherein position 198 is D or E.
19. The composition of any one of claims 5-12, wherein position 198 is F, L, I, M, V, P, A, or W.
20. A composition comprising positions 2-237 of SEQ ID NO: 1 with the provisos that position 111 is any amino acid residue except I and position 198 is any amino acid residue except Y.
21. The composition of claim 20, wherein position 111 is a basic amino acid.
22. The composition of claim 20, wherein position 111 is a polar amino acid.
23. The composition of claim 20, wherein position 111 is an acidic amino acid.
24. The composition of claim 20, wherein position 111 is N.
25. The composition of claim 20, wherein position 111 is H, K, or R.
26. The composition of claim 20, wherein position 111 is S, T, Y, Q, N, C, or G.
27. The composition of claim 20, wherein position 111 is D or E.
28. The composition of any one of claims 20-27, wherein position 198 is A.
29. The composition of any one of claims 20-27, wherein position 198 is a basic amino acid.

30. The composition of any one of claims 20-27, wherein position 198 is an acidic amino acid.
- 5 31. The composition of any one of claims 20-27, wherein position 198 is a nonpolar amino acid.
32. The composition of any one of claims 20-27, wherein position 198 is H, K, or R.
- 10 33. The composition of any one of claims 20-27, wherein position 198 is D or E.
34. The composition of any one of claims 20-27, wherein position 198 is F, L, I, M, V, P, A, or W.
- 15 35. A composition comprising positions 17-225 of SEQ ID NO: 2.
36. A composition comprising positions 17-225 of SEQ ID NO: 1 with the provisos that position 111 is any amino acid residue except I and position 198 is any amino acid residue except Y.
- 20 37. The composition of claim 36, wherein position 111 is a basic amino acid.
38. The composition of claim 36, wherein position 111 is a polar amino acid.
- 25 39. The composition of claim 36, wherein position 111 is an acidic amino acid.
40. The composition of claim 36, wherein position 111 is H, K, or R.
41. The composition of claim 36, wherein position 111 is S, T, Y, Q, N, C, or G.
- 30 42. The composition of claim 36, wherein position 111 is D or E.
43. The composition of claim 36, wherein position 111 is N.

44. The composition of any one of claims 36-43, wherein position 198 is a basic amino acid.
- 5 45. The composition of any one of claims 36-43, wherein position 198 is an acidic amino acid.
46. The composition of any one of claims 36-43, wherein position 198 is a nonpolar amino acid.
- 10 47. The composition of any one of claims 36-43, wherein position 198 is A.
48. The composition of any one of claims 36-43, wherein position 198 is H, K, or R.
49. The composition of any one of claims 36-43, wherein position 198 is D or E.
- 15 50. The composition of any one of claims 36-43, wherein position 198 is F, L, I, M, V, P, A, or W.
51. A composition comprising SEQ ID NO: 3.
- 20 52. A composition comprising positions 2-237 of SEQ ID NO: 3.
53. A composition comprising SEQ ID NO: 1 with the provisos that position 82 is any amino acid residue except E, position 83 is any amino acid residue except D, position 111 is any amino acid residue except I, position 197 is any amino acid residue except D, and position 198 is any amino acid residue except Y.
- 25 54. The composition of claim 53, wherein position 82 is a basic amino acid.
- 30 55. The composition of claim 53, wherein position 82 is a polar amino acid.
56. The composition of claim 53, wherein position 82 is a nonpolar amino acid.
57. The composition of claim 53, wherein position 82 is R.

58. The composition of claim 53, wherein position 82 is H, K, or R.
59. The composition of claim 53, wherein position 82 is S, T, Y, Q, N, C, or G.
- 5 60. The composition of claim 53, wherein position 82 is F, L, I, M, V, P, A, or W.
61. The composition of any one of claims 53-60, wherein position 83 is a basic amino acid.
- 10 62. The composition of any one of claims 53-60, wherein position 83 is a polar amino acid.
63. The composition of any one of claims 53-60, wherein position 83 is a nonpolar amino acid.
- 15 64. The composition of any one of claims 53-60, wherein position 83 is K.
65. The composition of any one of claims 53-60, wherein position 83 is H, K, or R.
- 20 66. The composition of any one of claims 53-60, wherein position 83 is S, T, Y, Q, N, C, or G.
67. The composition of any one of claims 53-60, wherein position 83 is F, L, I, M, V, P, A, or W.
- 25 68. The composition of any one of claims 53-67, wherein position 111 is a basic amino acid.
69. The composition of any one of claims 53-67, wherein position 111 is a polar amino acid.
- 30 70. The composition of any one of claims 53-67, wherein position 111 is an acidic amino acid.

71. The composition of any one of claims 53-67, wherein position 111 is N.
72. The composition of any one of claims 53-67, wherein position 111 is H, K, or R.
- 5 73. The composition of any one of claims 53-67, wherein position 111 is S, T, Y, Q, N, C, or G.
74. The composition of any one of claims 53-67, wherein position 111 is D or E.
- 10 75. The composition of any one of claims 53-74, wherein position 197 is a basic amino acid.
76. The composition of any one of claims 53-74, wherein position 197 is a polar amino
- 15 acid.
77. The composition of any one of claims 53-74, wherein position 197 is a nonpolar amino acid.
- 20 78. The composition of any one of claims 53-74, wherein position 197 is K.
79. The composition of any one of claims 53-74, wherein position 197 is H, K, or R.
80. The composition of any one of claims 53-74, wherein position 197 is S, T, Y, Q, N, C,
- 25 or G.
81. The composition of any one of claims 53-74, wherein position 197 is F, L, I, M, V, P, A, or W.
- 30 82. The composition of any one of claims 53-81, wherein position 198 is a basic amino acid.
83. The composition of any one of claims 53-81, wherein position 198 is an acidic amino acid.

84. The composition of any one of claims 53-81, wherein position 198 is a nonpolar amino acid.
- 5 85. The composition of any one of claims 53-81, wherein position 198 is A.
86. The composition of any one of claims 53-81, wherein position 198 is H, K, or R.
87. The composition of any one of claims 53-81, wherein position 198 is D or E.
- 10 88. The composition of any one of claims 53-81, wherein position 198 is F, L, I, M, V, P, A, or W.
89. A composition comprising positions 2-237 of SEQ ID NO: 1 with the provisos that
15 position 82 is any amino acid residue except E, position 83 is any amino acid residue except D, position 111 is any amino acid residue except I, position 197 is any amino acid residue except D, and position 198 is any amino acid residue except Y.
90. The composition of claim 89, wherein position 82 is a basic amino acid.
- 20 91. The composition of claim 89, wherein position 82 is a polar amino acid.
92. The composition of claim 89, wherein position 82 is a nonpolar amino acid.
- 25 93. The composition of claim 89, wherein position 82 is R.
94. The composition of claim 89, wherein position 82 is H, K, or R.
95. The composition of claim 89, wherein position 82 is S, T, Y, Q, N, C, or G.
- 30 96. The composition of claim 89, wherein position 82 is F, L, I, M, V, P, A, or W.
97. The composition of any one of claims 89-96, wherein position 83 is a basic amino acid.

98. The composition of any one of claims 89-96, wherein position 83 is a polar amino acid.
- 5 99. The composition of any one of claims 89-96, wherein position 83 is a nonpolar amino acid.
100. The composition of any one of claims 89-96, wherein position 83 is K.
- 10 101. The composition of any one of claims 89-96, wherein position 83 is H, K, or R.
102. The composition of any one of claims 89-96, wherein position 83 is S, T, Y, Q, N, C, or G.
- 15 103. The composition of any one of claims 89-96, wherein position 83 is F, L, I, M, V, P, A, or W.
104. The composition of any one of claims 89-103, wherein position 111 is a basic amino acid.
- 20 105. The composition of any one of claims 89-103, wherein position 111 is a polar amino acid.
106. The composition of any one of claims 89-103, wherein position 111 is an acidic amino acid.
- 25 107. The composition of any one of claims 89-103, wherein position 111 is N.
108. The composition of any one of claims 89-103, wherein position 111 is H, K, or R.
- 30 109. The composition of any one of claims 89-103, wherein position 111 is S, T, Y, Q, N, C, or G.
110. The composition of any one of claims 89-103, wherein position 111 is D or E.

111. The composition of any one of claims 89-110, wherein position 197 is a basic amino acid.
- 5 112. The composition of any one of claims 89-110, wherein position 197 is a polar amino acid.
113. The composition of any one of claims 89-110, wherein position 197 is a nonpolar amino acid.
- 10 114. The composition of any one of claims 89-110, wherein position 197 is K.
115. The composition of any one of claims 89-110, wherein position 197 is H, K, or R.
- 15 116. The composition of any one of claims 89-110, wherein position 197 is S, T, Y, Q, N, C, or G.
117. The composition of any one of claims 89-110, wherein position 197 is F, L, I, M, V, P, A, or W.
- 20 118. The composition of any one of claims 89-117, wherein position 198 is a basic amino acid.
119. The composition of any one of claims 89-117, wherein position 198 is an acidic amino acid.
- 25 120. The composition of any one of claims 89-117, wherein position 198 is a nonpolar amino acid.
- 30 121. The composition of any one of claims 89-117, wherein position 198 is A.
122. The composition of any one of claims 89-117, wherein position 198 is H, K, or R.
123. The composition of any one of claims 89-117, wherein position 198 is D or E.

124. The composition of any one of claims 89-117, wherein position 198 is F, L, I, M, V, P, A, or W.
- 5 125. A composition comprising positions 17-225 of SEQ ID NO: 3.
126. The composition of claim 125, wherein position 82 is a basic amino acid.
127. The composition of claim 125, wherein position 82 is a polar amino acid.
- 10 128. The composition of claim 125, wherein position 82 is a nonpolar amino acid.
129. The composition of claim 125, wherein position 82 is R.
- 15 130. The composition of claim 125, wherein position 82 is H, K, or R.
131. The composition of claim 125, wherein position 82 is S, T, Y, Q, N, C, or G.
132. The composition of claim 125, wherein position 82 is F, L, I, M, V, P, A, or W.
- 20 133. The composition of any one of claims 125-132, wherein position 83 is a basic amino acid.
134. The composition of any one of claims 125-132, wherein position 83 is a polar amino acid.
- 25 135. The composition of any one of claims 125-132, wherein position 83 is a nonpolar amino acid.
- 30 136. The composition of any one of claims 125-132, wherein position 83 is K.
137. The composition of any one of claims 125-132, wherein position 83 is H, K, or R.

138. The composition of any one of claims 125-132, wherein position 83 is S, T, Y, Q, N, C, or G.
- 5 139. The composition of any one of claims 125-132, wherein position 83 is F, L, I, M, V, P, A, or W.
140. The composition of any one of claims 125-139, wherein position 111 is a basic amino acid.
- 10 141. The composition of any one of claims 125-139, wherein position 111 is a polar amino acid.
142. The composition of any one of claims 125-139, wherein position 111 is an acidic amino acid.
- 15 143. The composition of any one of claims 125-139, wherein position 111 is N.
144. The composition of any one of claims 125-139, wherein position 111 is H, K, or R.
- 20 145. The composition of any one of claims 125-139, wherein position 111 is S, T, Y, Q, N, C, or G.
146. The composition of any one of claims 125-139, wherein position 111 is D or E.
- 25 147. The composition of any one of claims 125-146, wherein position 197 is a basic amino acid.
148. The composition of any one of claims 125-146, wherein position 197 is a polar amino acid.
- 30 149. The composition of any one of claims 125-146, wherein position 197 is a nonpolar amino acid.
150. The composition of any one of claims 125-146, wherein position 197 is K.

151. The composition of any one of claims 125-146, wherein position 197 is H, K, or R.
- 5 152. The composition of any one of claims 125-146, wherein position 197 is S, T, Y, Q, N, C, or G.
153. The composition of any one of claims 125-146, wherein position 197 is F, L, I, M, V, P, A, or W.
- 10 154. The composition of any one of claims 125-153, wherein position 198 is a basic amino acid.
155. The composition of any one of claims 125-153, wherein position 198 is an acidic amino acid.
- 15 156. The composition of any one of claims 125-153, wherein position 198 is a nonpolar amino acid.
157. The composition of any one of claims 125-153, wherein position 198 is A.
- 20 158. The composition of any one of claims 125-153, wherein position 198 is H, K, or R.
159. The composition of any one of claims 125-153, wherein position 198 is D or E.
- 25 160. The composition of any one of claims 125-153, wherein position 198 is F, L, I, M, V, P, A, or W.
161. A composition comprising a protein having a sequence identity of at least about 90% with SEQ ID NO: 2.
- 30 162. A composition comprising a protein having a sequence identity of at least about 95% with SEQ ID NO: 2.

163. A composition comprising a protein having a sequence identity of at least about 90% with SEQ ID NO: 3.
- 5 164. A composition comprising a protein having a sequence identity of at least about 95% with SEQ ID NO: 3.
165. A kit, comprising a composition of any one of claims 1-64.
- 10 166. A nucleic acid, at least a portion of which encodes the amino acid sequence of mMAPLE2.
167. A nucleic acid, at least a portion of which encodes the amino acid sequence of mMAPLE3.
- 15 168. A nucleic acid, at least a portion of which encodes the amino acid sequence of SEQ ID NO: 2.
169. A nucleic acid, at least a portion of which encodes positions 2-237 of the amino acid sequence of SEQ ID NO: 2.
- 20 170. A nucleic acid, at least a portion of which encodes SEQ ID NO: 1 with the provisos that position 111 is any amino acid residue except I and position 198 is any amino acid residue except Y.
- 25 171. A nucleic acid, at least a portion of which encodes positions 2-237 of SEQ ID NO: 1 with the provisos that position 111 is any amino acid residue except I and position 198 is any amino acid residue except Y.
- 30 172. A nucleic acid, at least a portion of which encodes the amino acid sequence of SEQ ID NO: 3.
173. A nucleic acid, at least a portion of which encodes positions 2-237 of the amino acid sequence of SEQ ID NO: 3.

174. A nucleic acid, at least a portion of which encodes SEQ ID NO: 1 with the provisos that position 82 is any amino acid residue except E, position 83 is any amino acid residue except D, position 111 is any amino acid residue except I, position 197 is any amino acid residue except D, and position 198 is any amino acid residue except Y.

5

175. A nucleic acid, at least a portion of which encodes positions 2-237 of SEQ ID NO: 1 with the provisos that position 82 is any amino acid residue except E, position 83 is any amino acid residue except D, position 111 is any amino acid residue except I, position 197 is any amino acid residue except D, and position 198 is any amino acid residue except Y.

10

176. A nucleic acid encoding an amino acid sequence that has a sequence identity of at least about 90% with SEQ ID NO: 2.

15 177. A nucleic acid encoding an amino acid sequence that has a sequence identity of at least about 95% with SEQ ID NO: 2.

178. A nucleic acid encoding an amino acid sequence that has a sequence identity of at least about 90% with SEQ ID NO: 3.

20

179. A nucleic acid encoding an amino acid sequence that has a sequence identity of at least about 95% with SEQ ID NO: 3.

180. A vector comprising a nucleic acid of any one of claims 166-179.

25

181. The vector of claim 180 wherein the vector is a plasmid.

182. The vector of claim 180 wherein the nucleic acid is DNA.

30 183. The vector of claim 180 wherein the nucleic acid is cDNA.

184. The vector of claim 180 wherein the nucleic acid is RNA.

185. A cell comprising a nucleic acid of any one of claims 166-179.

186. An organism comprising a nucleic acid of any one of claims 166-179.

1/7

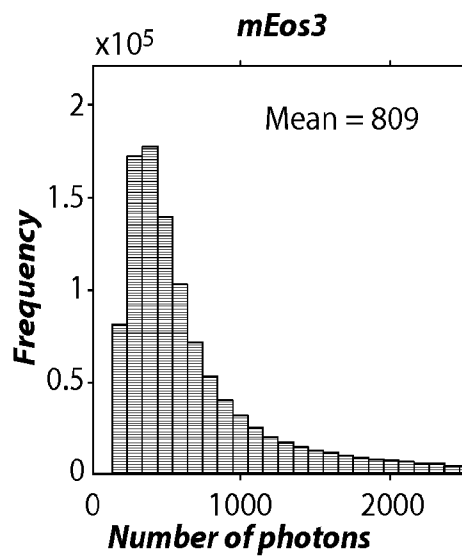


Fig. 1A

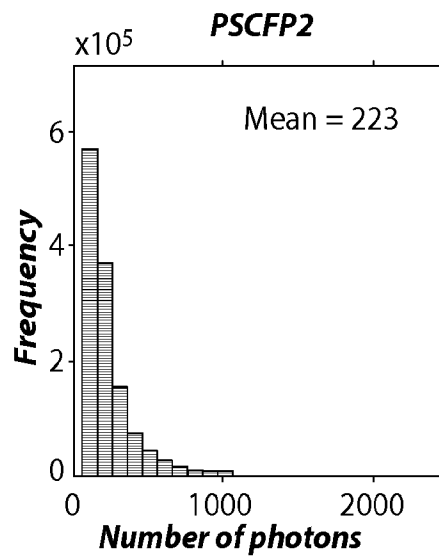


Fig. 1C

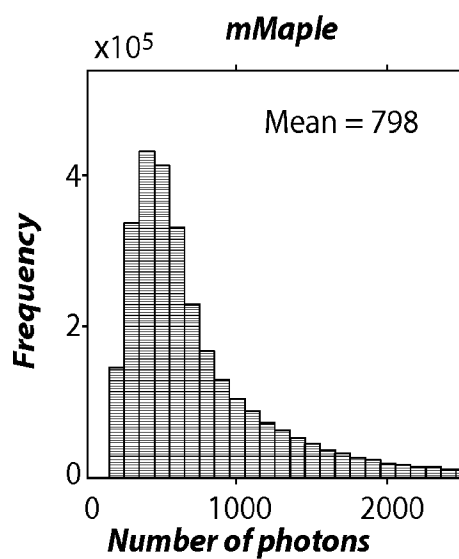


Fig. 1B

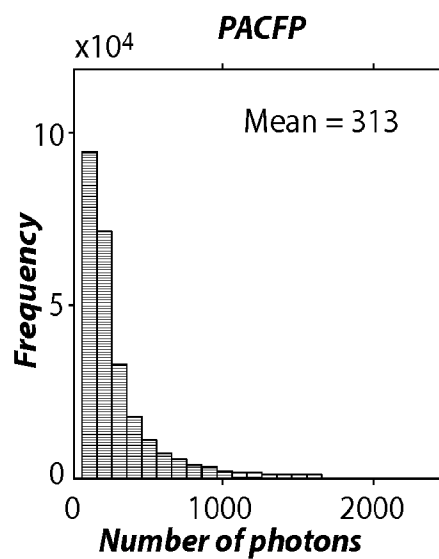


Fig. 1D

2/7

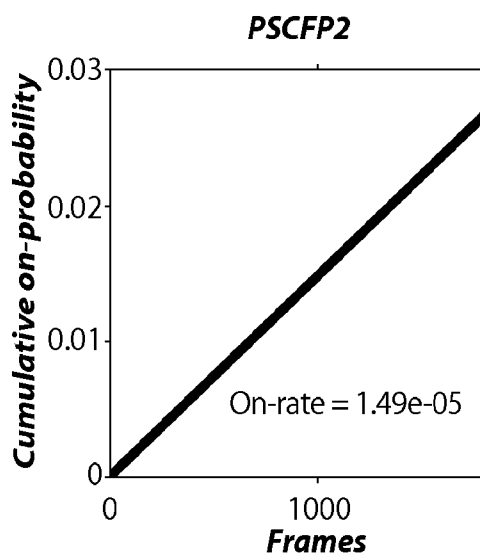


Fig. 2A

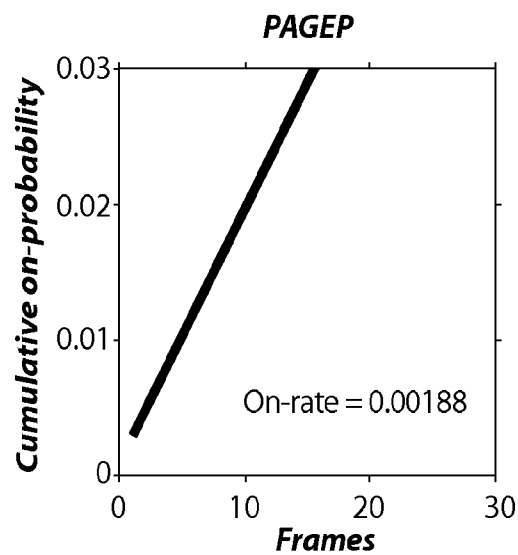


Fig. 2B

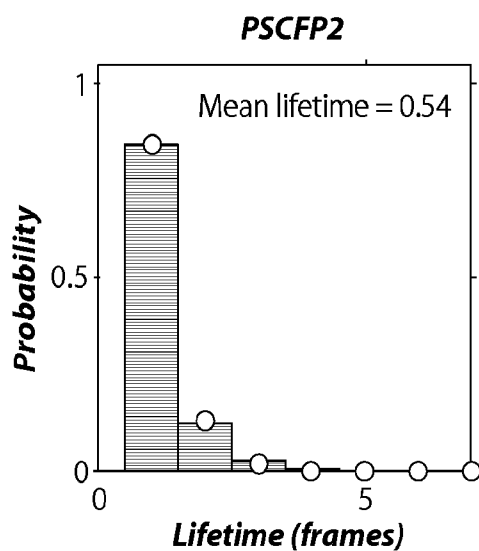


Fig. 2C

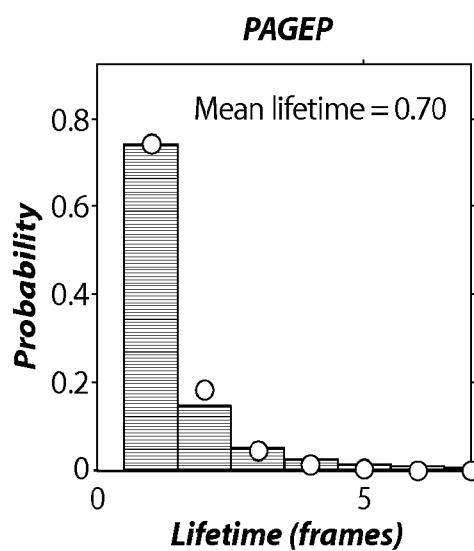


Fig. 2D



Fig. 2E



Fig. 2F

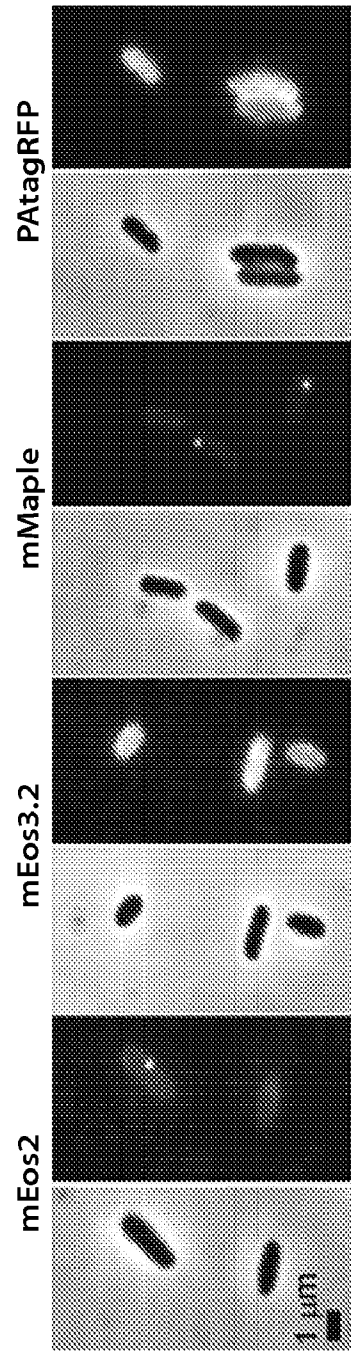


Figure 3A

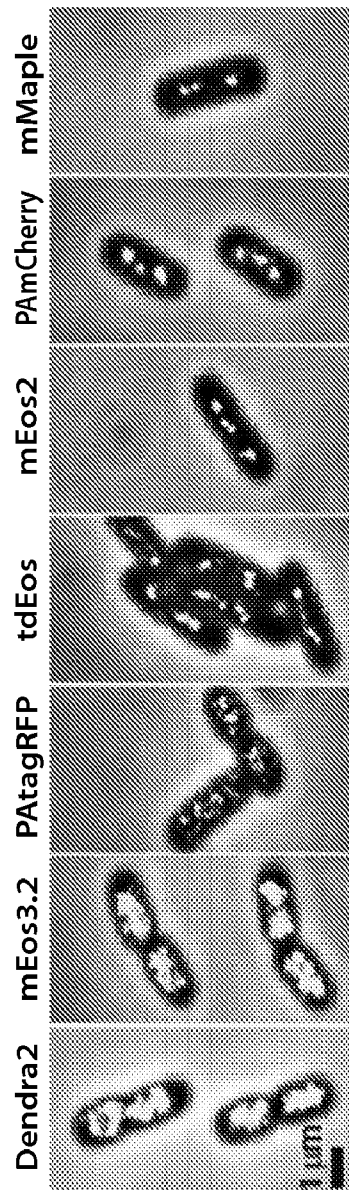


Figure 3B

4/7



Fig. 3C

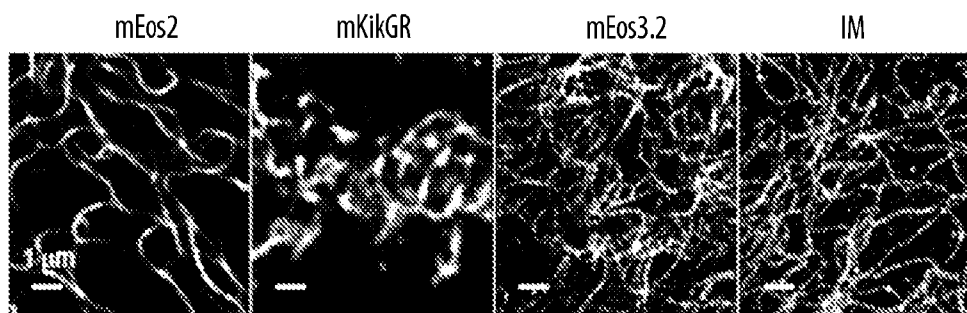


Fig. 3D

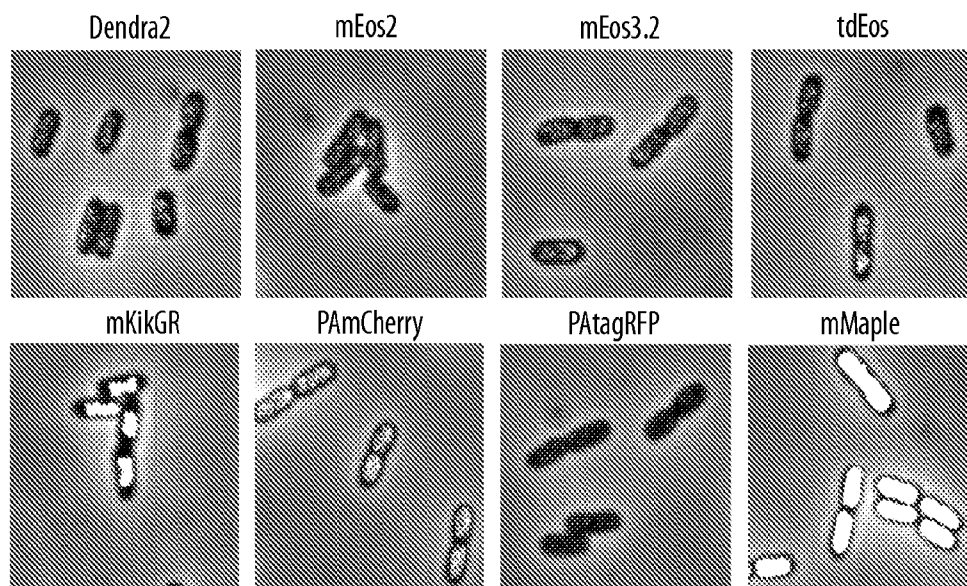


Fig. 4A

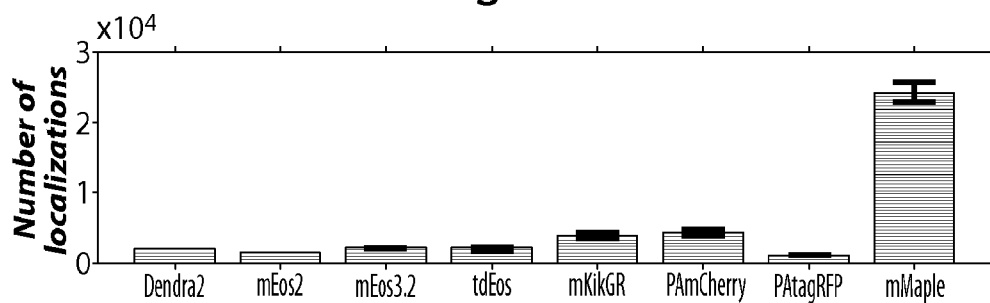


Fig. 4B

5/7

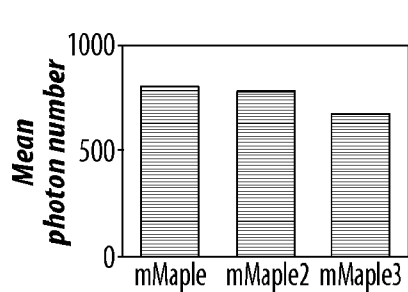


Fig. 5A

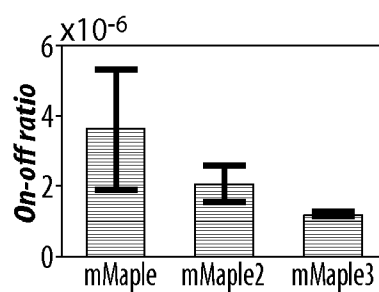


Fig. 5B

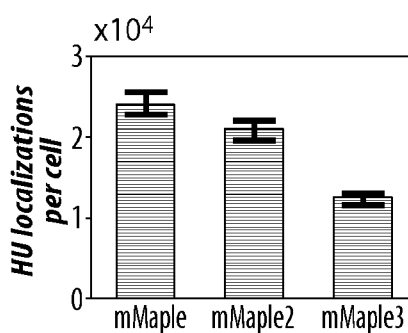


Fig. 5C

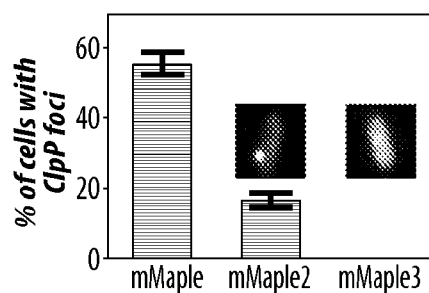


Fig. 5D

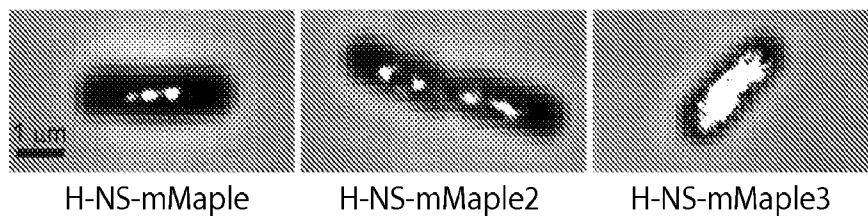


Fig. 5E

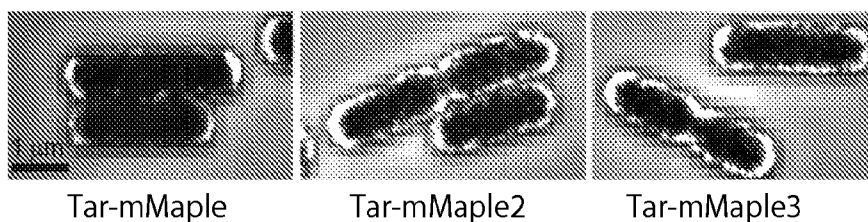


Fig. 5F

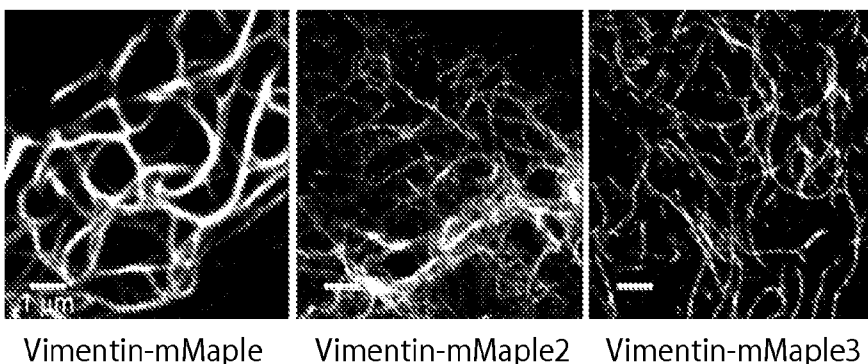


Fig. 5G

6/7

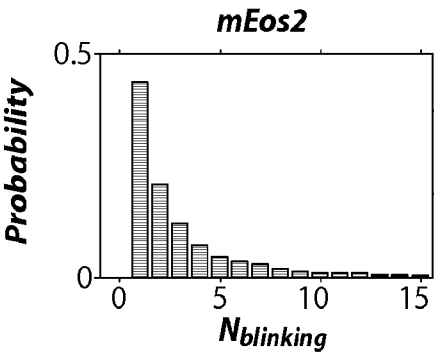


Fig. 6A

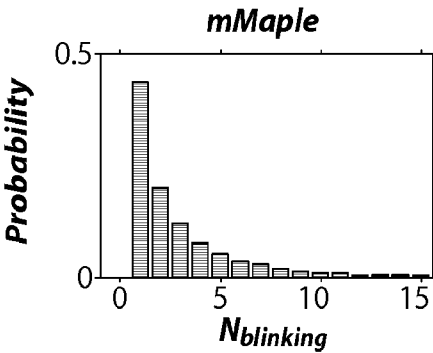


Fig. 6B

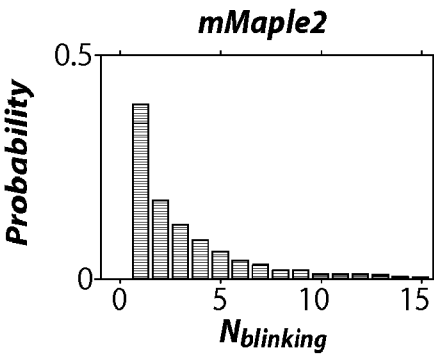


Fig. 6C

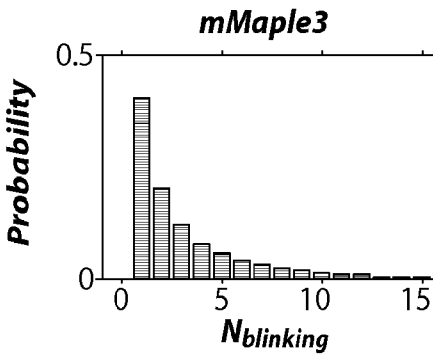


Fig. 6D

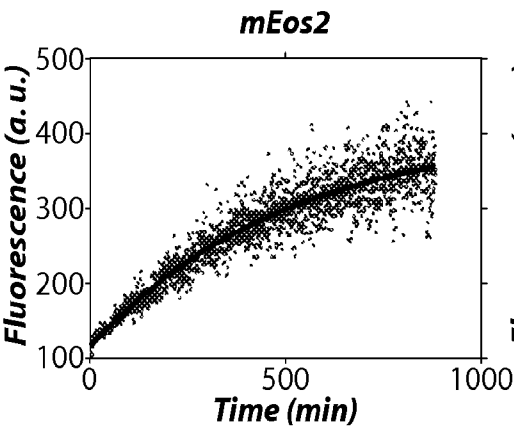


Fig. 7A

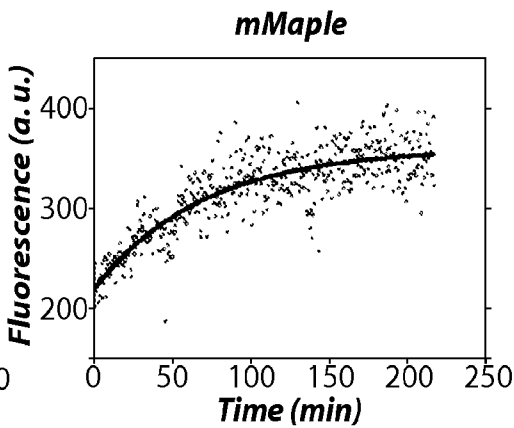


Fig. 7B

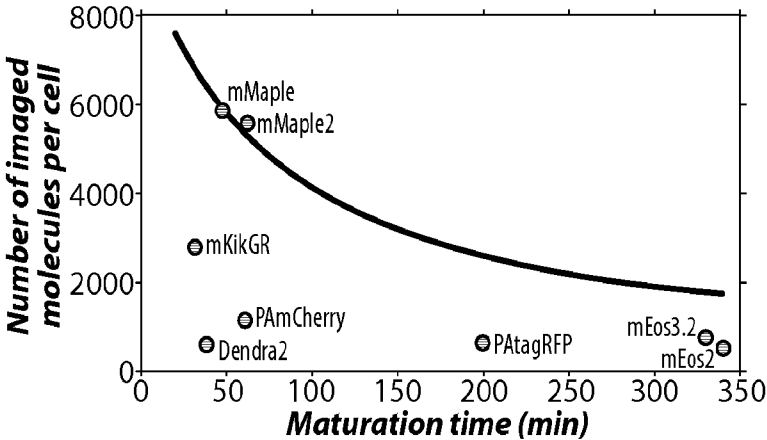


Fig. 8

7/7

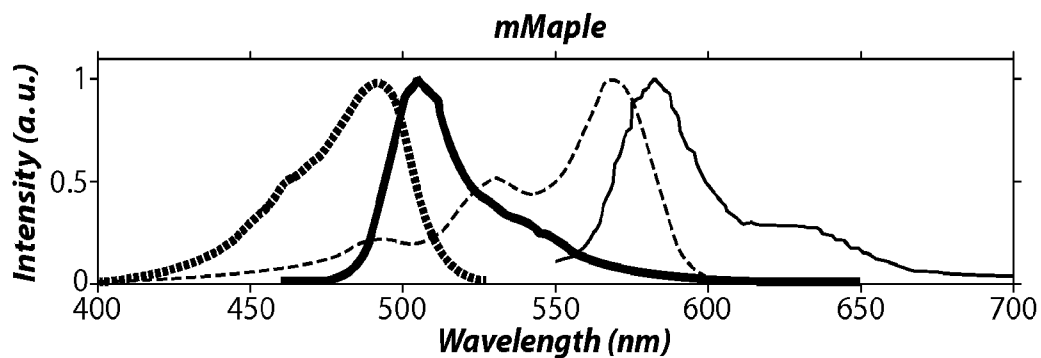


Fig. 9A

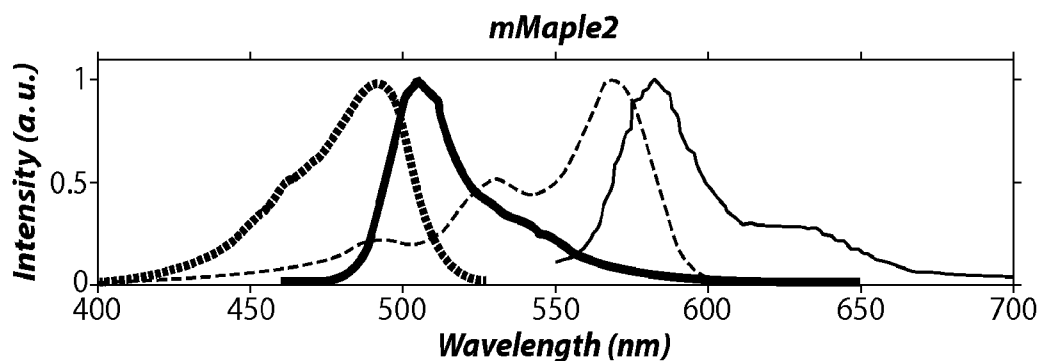


Fig. 9B

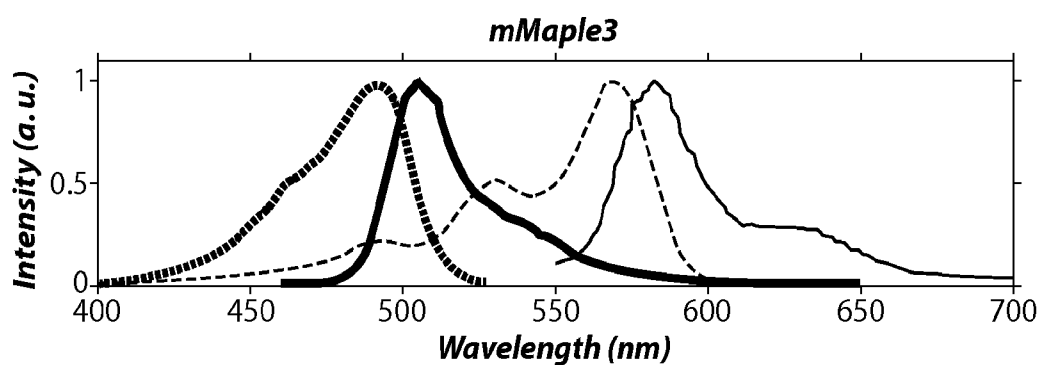


Fig. 9C

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2015/025540**A. CLASSIFICATION OF SUBJECT MATTER****C07K 14/00(2006.01)i, C12N 15/09(2006.01)i, C12N 15/63(2006.01)i, G01N 21/64(2006.01)i**

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C07K 14/00; C12N 15/09; C12N 15/63; G01N 21/64

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean utility models and applications for utility models

Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS(KIPO internal) & Keywords:mMAPLE, photoactivatable fluorescent protein (PAFP), mutation, position 82, position 83, position 111, position 197, position 198

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HOI et al., `Highlightable Ca <sup>2+</sup> indicators for live cell imaging` Journal of the American Chemical Society, Vol. 135, No. 1, pp. 46-49 (Epub. 20 December 2012) See abstract; pages 46-48; figure 1; pages S1-S5; and figures S2-S4.	161-164, 176-186
A		1-67, 89-103 , 125-139, 166-175
A	MCEVOY, `Nanoscale spatial organization of prokaryotic cells studied by super-resolution optical microscopy` Doctoral Dissertation, University of California, Berkeley (2012) See pages 107-109; figures 5-2-5-7, 5-10-5-12; and tables 5-1, 5-8.	1-67, 89-103 , 125-139, 161-164 , 166-186
A	ZHANG et al., `Rational design of true monomeric and bright photoactivatable fluorescent proteins` Nature Methods, Vol. 9, No. 7, pp. 727-729 (2012) See abstract; pages 727-728; figures 1-2; table 1; supplementary figures 1-17; and supplementary tables 1-7.	1-67, 89-103 , 125-139, 161-164 , 166-186
A	DURISIC et al., `Single-molecule evaluation of fluorescent protein photoactivation efficiency using an in vivo nanotemplate` Nature Methods, Vol. 11, No. 2, pp. 156-162 (February 2014) See the whole document.	1-67, 89-103 , 125-139, 161-164 , 166-186

☒ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

25 June 2015 (25.06.2015)

Date of mailing of the international search report

25 June 2015 (25.06.2015)

Name and mailing address of the ISA/KR

International Application Division
Korean Intellectual Property Office
189 Cheongsa-ro, Seo-gu, Daejeon Metropolitan City, 302-701,
Republic of Korea

Facsimile No. +82-42-472-7140

Authorized officer

KIM, Seung Beom

Telephone No. +82-42-481-3371



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2015/025540

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	NCBI, GenBank accession no. ABW74902.1 (25 March 2008) See the whole document.	1-67, 89-103 , 125-139, 161-164 , 166-186
PX	WANG et al., "Characterization and development of photoactivatable fluorescent proteins for single-molecule-based superresolution imaging," PNAS, Vol. 111, No. 23, pp. 8452-8457 (Epub. 27 May 2014) See abstract; pages 8452-8456; figures 1-5; table 1; SI Methods; figures S1-S4; and tables S1-S4.	1-67, 89-103 , 125-139, 161-164 , 166-186
PA	US 2015-0037812 A1 (HOWARD HUGHES MEDICAL INSTITUTE) 05 February 2015 See the whole document.	1-67, 89-103 , 125-139, 161-164 , 166-186

INTERNATIONAL SEARCH REPORTInternational application No.
PCT/US2015/025540**Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☒ Claims Nos.: 68-88,104-124,140-160,165
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of any additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2015/025540

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2015-0037812 A1	05/02/2015	None	