Structure-based rational design of an enhanced fluorogen-activating protein for fluorogens based on GFP chromophore

Marina Goncharuk

Institute of Bioorganic Chemistry, Russian Academy of Sciences

Nadezhda Baleeva

Institute of Bioorganic Chemistry, Russian Academy of Sciences

Dmitry Nolde

Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry

Alexey Gavrikov

Institute of Bioorganic Chemistry, Russian Academy of Sciences

Alexander Mishin

Institute of Bioorganic Chemistry, Russian Academy of Sciences

Andrey Sosorev

Shemyakin-Ovchinnikov Institute of bioorganic chemistry of the Russian Academy of Sciences

<https://orcid.org/0000-0001-9343-5502>

Alexander Arseniev

Russian Academy of Sciences

Sergey Goncharuk

Institute of Bioorganic Chemistry <https://orcid.org/0000-0002-0263-6462>

Roman Efremov

Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry RAS

Konstantin Mineev (✉ mineev@nmr.ru)

Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry <https://orcid.org/0000-0002-2418-9421>

Mikhail Baranov

Institute of Bioorganic Chemistry, Russian Academy of Sciences

Article

Keywords:

Posted Date: January 21st, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1188331/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Communications Biology on July 15th, 2022. See the published version at <https://doi.org/10.1038/s42003-022-03662-9>.

Abstract

“Fluorescence-Activating and absorption-Shifting Tag” (FAST) is a well-studied fluorogen-activating protein with high brightness and low size, able to activate a wide range of fluorogens. This makes FAST a promising target for both protein and fluorogen optimization. Here, we describe the structure-based rational design of the enhanced FAST mutants, optimized for the **N871b** fluorogen. Using the spatial structure of the FAST/**N871b** complex, NMR relaxation analysis, and computer simulations, we identified the mobile regions in the complex and suggested mutations that could stabilize both the protein and the ligand. Two of our mutants appeared brighter than the wild-type FAST, and these mutants provided up to 35% enhancement for several other fluorogens of similar structure, both *in vitro* and *in vivo*. Analysis of the mutants by NMR reveals that brighter mutants demonstrate the highest stability and lowest length of intermolecular H-bonds. Computer simulations provide the structural basis for such a stabilization.

Introduction

Fluorogen-activating proteins (FAPs) represent an important class of instruments in modern molecular biology. Unlike the fluorescent proteins (FPs), they do not contain a chromophore and become active only upon interaction with special low-molecular-weight compounds, fluorogens. Fluorogens themselves are usually dim and become bright only after their binding to FAP. In contrast to FPs, FAPs do not need time and oxygen to mature and are in general smaller, which decreases their chance to distort the behavior of the object under investigation. Thus, FAP is a very convenient genetically encoded fluorescent label that can be switched on and off on demand. Moreover, while the performance of a given FP may be altered only via the protein mutagenesis, the optic properties of a FAP may be additionally changed by the variation of the fluorogen structure, which broadens substantially the possible range of emission and absorption maxima.

Out of all FAPs, the so-called “Fluorescence-Activating and absorption-Shifting Tag” (FAST) received a lot of attention during the past several years^{1–3}. The protein complexes with fluorogens are relatively bright, comparable to the brightness of some FPs^{4–6}. Besides, it can bind a vast class of chromophores and has a rather low size of ~14 kDa, which makes it a promising target for both protein and fluorogen optimization. In particular, dozens of FAST ligands with various absorption/emission properties have been proposed^{7–11}, several FAST mutants, optimized for different ligands have been found^{10–14} and a split-construct of FAST was designed¹⁵. The latest studies report the FAST mutants, redFAST, and greenFAST, applicable for the orthogonal fluorescent labeling¹³, and a promiscuous variant of FAST, pFAST, capable of efficient binding of several ligands of different color¹⁰.

Almost all of the FAST improvements made to date, have been achieved via the random mutagenesis/directed evolution approach. This technique is highly efficient since it allows covering a vast variety of mutants in a single experiment, and it is actually evident, looking at the results of the recent works in the area¹⁰. On the other hand, the techniques used for the selection of the enhanced clones are not ideal, and some favorable mutations may be lost in the experiments of such kind.

Furthermore, in most cases, the mutagenesis results in the protein variant, which is optimal for a particular fluorogen or a group of compounds, and has poor performance with other possible ligands¹³. Random mutagenesis of FAST for each out of dozens of ligands may be too time-consuming. The structure-driven rational design may be considered as an alternative to the random mutagenesis, however, this pathway was hindered for FAST until recently, when our group published the first spatial structure of the protein in the apostate and in complex with the ligand **N871b**¹⁶. Using the spatial structure, we managed to design the shorter variant of FAST, nanoFAST, and find an appropriate fluorogen. Here, we utilize the spatial structure, NMR relaxation analysis, and molecular dynamics simulations to rationally design an improved variant of FAST, optimized for the **N871b** fluorogen.

Results

Identification of "hot spots" in FAST/N871b complex

For rational mutagenesis, one needs first to identify the "hot spots" in the protein structure. It is well known that the brightness of fluorophores depends on the extinction coefficient and the quantum yield of fluorescence. On the one hand, the brightness of FAPs is typically only slightly dependent on the dissociation constant^{10–13}, therefore it is hard to predict the effect of mutations, considering only the free energy of protein/ligand interactions. On the other hand, the quantum yield depends on the probability of nonradiative transitions, and, therefore, on the overall mobility of the fluorogen inside the cavity of the FAP. It is quite difficult to analyze the mobility of a fluorogen, however, one can easily measure the mobility of the overall complex, and try to minimize it, fixing the fluorogen. Here, we measured the NMR relaxation parameters of ¹⁵N nuclei, to assess the mobility of the FAST backbone in a complex with **N871b**. In addition several molecular dynamics (MD) simulations were carried out to investigate the flexibility of protein side chains (Fig. S1). The structure of the complex was obtained previously by NMR¹⁶, **N871b** was selected as a parent fluorogen for the family of compounds, designed for the multicolor labeling⁹. Analysis of the relaxation parameters with a model-free approach reveals three major mobile regions in close proximity to the ligand - loops 98-102, 70-73, and 52-53 (Fig. 1). There are several other high-mobility sites in the protein, including the N- and C-termini, helices H1, H2, and H5, loops 86-88, 112-116, and a row of residues in the β -strand (Q41, I31, W119, M109). Results of MD simulations are in good agreement with NMR data and reveal the fast sidechain motions in the regions 50-54, 68-74, and 97-102. In close proximity to the ligand, we observed the motions of R52, P68, I96, T98, and S99 sidechains. Thus, we searched for the possible mutations in these three regions that could stabilize the complex.

Design of the enhanced mutants

It is noteworthy that **N871b** is substantially different from the initially proposed FAST ligands. The latter are based on 4-hydroxy-benzylidene-rhodanine¹, while **N871b** contains an additional double bond and a pyridine moiety. This pyridine ring exits the cavity of FAST and is not in tight contact with the protein.

Moreover, the pyridine group demonstrates the elevated root-mean-square fluctuations of the atomic coordinates during MD simulations, which may be ascribed to the high propensity of its random flips. On the other hand, due to the peculiar charge distribution (Fig. 2), this group is capable of various stabilizing interactions: stacking, π -cation, π -anion and hydrophobic. Therefore, we decided to take advantage of these facts and to search for the FAST residues both in the abovementioned dynamic "hot spots" of the protein and in close proximity to the **N871b** pyridine group. The first and obvious candidate is R52. The residue is in direct contact with pyridine of **N871b**, but the interaction could be unfavorable. We mutated it to the aromatic amino acids (R52Y and R52F), hydrophobic (R52L), negatively charged (R52E and R52D), positively charged (R52K) and neutral residue (R52A). Following similar principles and taking into account the charge distribution over **N871b**, we mutated S99 (S99K, S99R, and S99E), P68 (P68K, P68R), and D65 (D65R and D65K, D65R/R52E). Next, we took several mutations, which proved favorable for other fluorogens^{12,14}, namely, F62L, P68T, V107I, P73S. Finally, we considered the possibility to change the structure of the loop 98-102 and close the binding pocket by mutating P97 (P97T, P97T/T98G). This led us to 21 mutants, summarized in table S1.

Fluorogenic properties of the FAST mutants

As a next step, we synthesized in *E.coli* and purified all the mutant variants, and studied their properties in complex with **N871b** fluorogen (Fig. 3, S2-S10, S21 and Tables 1, S2). Out of 21 tested proteins, only two performed better than the wild-type - R52Y and R52L. Mutations increased both the extinction and the quantum yield and provided ~ 20-25% higher overall brightness. Several mutants appeared stabilizing and decreased the dissociation constant but did not improve the brightness of the complex (all the S99 mutants, P73S, P68K, D65K). Finally, several suggested mutations completely destroyed the fluorogen-activating ability of FAST with respect to **N871b** (P97T and R52D). Therefore, we can conclude that the initial suggestion of the prospective mutation sites was correct, R52, D65, P68, and S99 mutations either improved the brightness or improved the binding affinity of FAST. However, it is not sufficient to find a mutation that would decrease the free energy of the FAST/**N871b** complex. Apparently, it is also necessary to reduce the conformational lability of the chromophore, imposing the restrictions on its particular degrees of freedom that determine the spectral properties. These restrictions are quite difficult to predict, which explains that only one of selected mutagenesis sites provided a reasonable brightness improvement.

To find out whether the R52Y and R52L mutations can improve the performance of FAST in complex with other fluorogens, we tested three structural analogs of **N871b**, previously proposed for multicolor labeling: **MCL.1b** (emission at 579 nm), **MCL.1f** (593 nm) and **MCL.3a** (639 nm)⁹ (Table 1, Fig. 3, S11-S18). The classical FAST ligand **HMBR** was also added to this set as a sample compound with drastically different structure. Brightness of all three **N871b** analogs was increased in complexes with both mutants in comparison to the parent protein. The greatest enhancement was observed for **MCL.1b** and **MCL.1f** in complexes with FAST-R52Y (25 and 27%, respectively) and for **MCL.3a** in a complex with FAST-R52L (34%). In contrast, **HMBR** was much less bright in complexes with the mutant proteins, the brightness

decreased by 36 and 47% for R52L and R52Y, and the stability of the complex decreased by one order of magnitude. Thus, the effects of R52Y and R52L mutations in FAST are ligand-specific and provide the substantial enhancement for several arylidene-imidazolones based fluorogens, applicable for the multicolor fluorescent labeling. With this regard we suggest naming the novel variants of FAST as (multicolor) mcFAST-Y and mcFAST-L.

Next, we compared the novel FAST variants and original FAST with various fluorogens in live-cell experiments. We used a TagBFP2 fluorescent protein as an internal reference, which was included in the H2B-TagBFP2-FAST construct, that allowed us to estimate the relative brightness (Fig S25). As clearly seen in figure 4, the values of relative brightness of all novel pairs were in good agreement with *in vitro* data.

Some discrepancies probably arose because the wavelength of the exciting light in microscopy did not always correspond to the maximum absorption for which the EC was measured and the relative brightness was calculated. However, mcFAST-Y turned out to be significantly brighter in pair with fluorogens **N871b**, and **MCL.1b**, while mcFAST-L is brighter with fluorogen **MCL.3a**.

Table 1
Optical properties of **HMBR**, **N871b**, **MCL.1b**, **MCL.1f** and **MCL.3a** in complexes with FAST mutants.

Chr.	Mutant	$K_D, \mu\text{M}^a$	$\epsilon, \text{M}^{-1}\cdot\text{cm}^{-1}^b$	FQY, % ^c	Brightness
HMBR	FAST	0.13±0.02	45000±680	31±0.6	13900±480
	R52L	0.48±0.11	34500±520	26±0.6	8900±340
	R52Y	1.34±0.11	34000±510	22±0.1	7350±140
N871b 609 nm	FAST	0.33±0.01	27000±410	26±1.4	7000±480
	R52L	0.27±0.01	30000±450	29±1.7	8750±640
	R52Y	0.24±0.02	29000±440	29±1.5	8350±560
MCL.1b 582 nm	FAST	0.20±0.04	34000±510	31±2.2	10650±920
	R52L	0.44±0.20	37500±560	29±0.6	11000±390
	R52Y	0.27±0.06	35000±530	38±1.4	13350±690
MCL.1f 595 nm	FAST	0.27±0.04	47000±710	32±0.6	14800±500
	R52L	0.39±0.10	50000±750	33±1.7	16500±1100
	R52Y	0.31±0.05	48500±730	39±1.3	18850±910
MCL.3a 639 nm	FAST	0.08±0.01	56000±840	14±0.7	7900±510
	R52L	0.06±0.01	61000±920	17±0.6	10550±530
	R52Y	0.11±0.02	50500±760	17±1.4	8300±810
a – represented as mean ± SD (n = 3);					

b – represented as result of single measurement ± the precision of the measuring instruments (weighing and pipetting errors);

c – fluorescence quantum yield, represented as mean ± SD (n = 9, except **HMBR-R52Y**, where n = 3).

Structural determinants of fluorogen-activating ability of FAST mutants

The obtained set of mutants with varied brightness could be utilized to find out the structural and dynamic features of the complex that determine the performance of FAST protein. Solving the spatial structure and measuring the backbone mobility for all 21 FAST variants is too expensive and time-consuming, however, some important features could be found even in 1D NMR spectra. Namely, the downfield region of ¹H spectrum of the FAST/**N871b** complex contains three protons that take part in intermolecular hydrogen bonding - Y42 H η , E46 H ϵ 2 (with the oxygen of phenolic ring), and H ϵ 1 W94

(with the oxygen of imidazolone ring) (Fig. 5a,S19, Table S3). Thus, we characterized the hydrogen bonding in the studied FAST/**N871b** complexes, using the proton chemical shift as an indicator of the H-bond length¹⁷, and normalized intensity of the signal as an indicator of the H-bond stability. The correlation analysis reveals that the overall brightness of the complex demonstrates a correlation with the chemical shift of W94 ($r=0.57$, $p<0.01$) and a weak correlation with the chemical shift of E46 ($r=0.40$, $p<0.1$). The greater the chemical shift, the shorter the corresponding H-bond, the brighter the complex. Correlation between the "stability" of three hydrogen bonds (sum of the normalized intensities of three observed signals) and brightness is also weak: $r=0.36$, $p<0.2$. However, we observe three highly stable mutants: R52L, R52Y, and R52F, two of which provide substantial improvement. Therefore, high H-bond stability is a good indicator of the potentially high brightness of the FAST/**N871b** complex. A more detailed analysis with separate treatment of various contributions reveals the reliable correlations between the W94 chemical shift and the quantum yield, as well as between the stability of E46 and Y42 H-bonds and the extinction coefficient (Fig. 5e). To summarize, the brightest mutants correspond to the complexes with the shortest intermolecular H-bonds formed by W94 side chains and the highest stability of intermolecular H-bonds formed by E46 and Y42.

The structural basis of the enhancement could be analyzed with MD simulations as well. For this purpose, we ran MD simulations with the parent and several mutant FAST proteins, starting from the NMR structure (Fig. 6). In the parent protein, the sidechain of R52 tends to form a π -cation interaction with the pyridine moiety of **N871b**. The same behavior is observed in the case of R52Y mutant: the tyrosine sidechain forms the stacking interaction with the pyridine ring. However, the R52 sidechain has many degrees of freedom, which makes it mobile and allows the conformational diversity (Fig. 6, S20). In contrast, tyrosine seems to have only two possible modes of stacking, which restricts the motions of Y52 in R52Y mutant and this may be considered as a source of **N871b** stabilization and fluorescence enhancement. The behavior of leucine at position 52 is radically different. This sidechain does not interact with the pyridine group of the ligand - it is tightly packed together with the aliphatic chains of V57 and V66. These hydrophobic interactions restrict the mobility of L52 and it is locked mostly in the single χ_1/χ_2 conformation (Fig. S20). Thus, the source of the enhancement in the case of R52L is the overall stabilization of the protein core.

Discussion

In the present work, we suggest a rational approach to the modification of FAST protein for the particular fluorogen **N871b**. The approach relies on the analysis of protein dynamics with NMR and computer simulations. Such an analysis provides the "hot spots" that may be improved, aiming to freeze the motions of both protein and fluorogen. According to our data, mobile regions include the N-terminal two-helical domain (1-26), which becomes unstructured in the apoprotein of FAST¹⁶, five loops: 50-54, 70-74, 85-89, 97-103, and 111-117. Slow μ s-ms motions are also observed in the helix H5 (Fig. 1). In case these "hot spots" are common for various ligands, the FAST optimization strategy should include the mutations that decrease mobility in the specified regions. To test whether the found regions represent the common

mutagenesis sites for different ligands, we analyzed the known mutations in FAST that were previously reported to enhance the brightness or ligand-binding efficacy of the protein. To date, 58 FAST mutants were published as the final or intermediate result of optimization, including the iFAST, rFAST, gFAST, frFAST, oFAST, tFAST, and pFAST¹⁰⁻¹³. These 58 proteins contained in total 253 mutations, 97 substitutions were unique and were located at 54 different sites (Table S4). The largest fraction of this set includes 26 mutations that affect the N-terminal domain or its interactions with the FAST core, 14 mutations affect the residues that are in direct contact with the ligand molecule and 15 mutations are located in helix5. Finally, 23 substitutions are located in the loops that were found mobile in the current study. Therefore, the vast majority of known mutations (64 of 97) are located in the "hot spot" regions, 54 out of 58 mutant proteins contained the mutations in the "hot spots", which supports the general character of our findings (Fig. 7).

Using the proposed approach, we suggested 17 mutants, 2 of which (mcFAST-Y and mcFAST-L) revealed the improved brightness, corresponding to the success rate of ~12%. These mutations appeared beneficial not only for **N871b**, but also for several similar compounds with various emission maxima. The enhancement (20-34%) is comparable to the effect observed in the case of iFAST_{V107I} (10-30% for various fluorogens), which was also found by rational design based on the structure of PYP protein, parent to FAST¹². Compared to directed evolution, rational design has its advantages and drawbacks. Fast screening approaches, used for the selection of clones, are not very precise. Therefore, if the gain of function is not substantial, it may be left unnoticed and the enhanced variant would be dismissed. Besides, the brightness in the living cells may depend on the fluorogen binding constant and the expression level of the protein, which may result in the selection of variants that are not brighter, when the protein is purified and analyzed *in vitro*. This was clearly demonstrated in the case of three FAST mutants that were selected in massive screening as the brightest¹⁴, but were not superior to FAST, when isolated¹². On the other hand, using the rational approach it is hard to predict and construct the multipoint mutants, such as pFAST¹⁰, which provided the 2-3 fold enhancement relative to FAST for some fluorogens. Thus, today, rational design may be considered as a feasible alternative to the directed evolution, when the FAP is well adapted to the fluorogen and only a fine-tuning is required.

Conclusion

To conclude, here we report a successful attempt to rationally design an enhanced fluorogen-activating protein mutant for a particular fluorogen. Using the spatial structure of the FAST/**N871b** complex, NMR relaxation parameters, and MD-simulations, we identified the mobile regions in the complex and rationally suggested 17 mutations that could stabilize both the protein and the pyridine group of **N871b**. Two of our mutants appeared superior to the wild-type FAST, while none of the tested mutations, that were shown favorable for FAST in complex with other fluorogens, provided any improvement for **N871b**. The mutants selected for **N871b** performed well for several other fluorogens of similar structure and did not work with conventional FAST fluorogen **HMBR**. Analysis of the set of mutants by NMR reveals that brighter mutants demonstrate the highest stability and lowest length of intermolecular H-bonds and MD simulations

provide the structural basis for such a stabilization. According to the literature, most of the known FAST mutations that were found in numerous directed evolution studies, are located in the mobile regions that are reported here. This suggests the general character of our findings and provides the basis for the rational combination of mutations that stabilize the protein and mutations that affect the ligand binding.

Methods

General

All solid chromophores were dissolved in DMSO (Sigma Aldrich, “for molecular biology” grade, #cat D8418) in 5 mM concentration, and stored in dark place at -20°C for no more than 3 months.

UV-VIS spectra were recorded on a Varian Cary 100 spectrophotometer. Fluorescence excitation and emission spectra were recorded on Agilent Cary Eclipse fluorescence spectrophotometer.

Absorption and emission maxima are determined as is from raw data, without using of any approximation procedure.

Protein production

All genetic constructs based on pEV expression vector were produced by Cloning Facility (Russia). The sequence encoding His6-tag was added at the 3' end of target protein genes.

The proteins were produced and purified similar to the protocol described earlier for the FAST protein¹⁶. In brief, the BL21(DE3) *E.coli* strain and M9 minimal salts medium were used for protein production. In order to obtain the ¹⁵N-labelled protein samples, ¹⁵NH₄Cl was implemented. Cells were grown at 37°C, 250 rpm (300 mL of cell culture per 2l Ehrlenmeyer flask) in a shaker incubator (New Brunswick Innova 44R). Protein expression was induced at OD₆₀₀ ~ 0.6 using 0.25 mM of isopropyl β-d-1-thiogalactopyranoside (IPTG). After 4 hours of cultivation cells were harvested and stored at -20°C. Cells were lysed by sonication (Bandelin Sonoplus, GM 2200, titanium tapered tip KE76) on ice followed by target protein purification using Ni²⁺ Chelating Sepharose High Performance (GE Healthcare) and overnight dialysis (12 kDa MWCO) against the 1xPBS buffer at room temperature.

NMR spectroscopy.

NMR spectra for the relaxation analysis were recorded at pH 7.0, 20 mM NaPi, 20 mM NaCl. The concentration of protein was 0.85 mM, the concentration of **N871b** was taken equal to 0.89 mM. The experiments were run at 25 °C using the Bruker Avance 700 MHz spectrometer. 15N longitudinal (T1), transverse (T2) relaxation rates were measured using the pseudo-3D HSQC-based experiments with varied relaxation delay¹⁸. Heteronuclear equilibrium 1H,15N-NOE magnitudes were obtained using the 1H presaturation for 3s during the recycling delay. Reference and NOE spectra were recorded in the

interleaved mode. Relaxation parameters were analyzed using the model-free approach as implemented in the TENSOR2 software, assuming the isotropic rotation¹⁹.

To measure the ¹H spectra, FAST or its mutants in PBS buffer pH 7.4 were mixed with **N871b** at 10% molar excess of fluorogen. The resulting concentrations of protein were 50-300 μ M. NMR spectra were recorded at 25 °C using the Bruker Avance III 800 MHz NMR spectrometer, equipped with a triple-resonance cryoprobe. ¹H chemical shifts were referenced with respect to the water signal, chemical shift of water protons was taken equal to 4.80 ppm at 25 °C. Relative intensities of ¹H signals were calculated as $\frac{1}{n}$ of a ratio between the integral of the signal of interest and of an alone-standing methyl group at -0.66 ppm. The "stability" of an intermolecular H-bond is calculated as a ratio between the relative intensity of the signal of interest and the maximal relative intensity, obtained for this signal in the whole set of tested mutants. The summary stability of three intermolecular H-bonds is calculated as a sum of three individual stabilities and lies in the range {0;3}.

Computer simulations

Five starting structures of the complex (models 1, 5, 9, 13, 17) were chosen from pdb entry 7AVA¹⁶. Protein mutations were made using the mutagenesis wizard of the PyMOL Molecular Graphics System (Schrödinger LLC). MD simulations were carried out in Gromacs version 2020.4 software²⁰. All-atom Amber99sb-ildn²¹ forcefield was used for protein, tip3p model for water and parameters for **N871b** were developed in this work. Equilibrium geometry (minimum energy conformation) of **N871b** and energy profiles for five dihedral angles (χ_1 - χ_5 , Fig. S23) were calculated using DFT (B3LYP functional, 6-31G(d,p) basis set) in GAMESS software^{22,23}. Molecular electrostatic potential for **N871b** was visualized from these calculations using Jmol program.

Atom types, bond and angles parameters were assigned in Ambertools 20²⁷. Partial atomic charges were calculated by averaging DFT charges of chemically equivalent atoms. Since energy profiles have clear minima in planar conformations (Fig. S24) we had to exclude some nonbonded interactions that shift the minimum position. Excluded nonbonded atom pairs are shown in Fig S23. Parameters for dihedral angles χ_2 , χ_3 , and χ_5 were calculated by fitting an energy barrier between two minima in GAMESS and Gromacs energies with proper periodic dihedral angle terms. Dihedral angle energy barriers for angles χ_1 , and χ_4 are extremely large and prevent flips between two minima. To correctly describe these dihedral angle parameters around their minima we used Ryckaert-Bellemans dihedral potential. Ryckaert-Bellemans parameters were calculated by fitting GAMESS and Gromacs energies (Fig. S24).

Starting structures were placed in the center of the dodecahedron box at least 1.2 nm from the box edge, solvated with water. Na⁺ ions were added to neutralize the system charge. All simulations were performed with periodic boundary conditions, Verlet cutoff scheme, plain cutoff of 1.4 nm for van-der-Waals and particle-mesh Ewald method for Coulomb interactions. Simulation included 4 stages: energy minimization in double precision, heating from 5 to 300 K during 1 ns with 1 fs timestep using V-rescale thermostat and fixed protein, equilibration during 50 ns with timestep 2 fs using V-rescale thermostat and

Parrinello-Rahman barostat, production run 200 ns with the same parameters. Lincs algorithm was employed to constrain the bonds with H-atoms in MD simulations except for the heating phase. RMSD from starting structures (Fig S25) showed that 50 ns is obviously enough for system equilibration.

Measurement of dissociation constants

The experiments were performed on the Tecan Infinite 200 Pro M Nano dual mode plate reader at 25°C. The affinity constants were determined by spectrofluorometric titration of protein in single concentration (0.10 µM) using various chromophore concentrations. Fitting was performed using Origin 8.6 software. Titration curves are presented in Supporting Information.

Measurement of extinction coefficients

The chromophores solutions were mixed with a protein solution in PBS buffer (pH 7.4, #cat E404-200TABS, Amresco). The final concentration of chromophores (together in free form and in complex) for all experiments was 5 µM. Proteins were added in such amount that led to almost complete transfer of the chromophore into the complex ($\alpha \geq 97\%$). The final protein concentration (together in free form and in complex) for each complex was calculated using the equation:

$$[Pr] = \frac{K_D \times (\alpha \times [Chr])}{[Chr] - (\alpha \times [Chr])} + (\alpha \times [Chr]) \quad (1)$$

where, K_D – dissociation constant, $[Chr]$ – final chromophore concentration.

The molar extinction coefficient was calculated by the formula:

$$\varepsilon = \frac{A}{cl} \quad (2)$$

where A is the absorbance intensity at maxima, c is the molar concentration of complexes, l is the pathlength.

The data were obtained from a single measurement. The errors were determined from the precision of the measuring instruments (weighing and pipetting errors).

Measurement of quantum yields

Fluorescence quantum yields were calculated according to the procedure described in the literature²⁴ with the use of Rhodamine 101 (for **N871b**, **MCL.1b** and **MCL.1f**), Fluorescein (for **HMBR**) and Oxazine 1 (for **MCL.3a**) as standards. The final concentration of chromophores (together in free form and in complex) was 5 µM for absorption spectra registration and 1.67, 0.5, and 0.167 µM for emission spectra registration, (except R52E-**N871b**, R52E/D65R-**N871b** and R52Y-**HMBR** pairs, where only 0.5 µM chromophore concentration was used for emission spectra registration). Proteins were added in such amount that in all cases lead $\alpha \geq 97\%$ (see above).

The quantum yield was calculated by the formula:

$$\Phi_x = \Phi_{st} \times \frac{F_x}{F_{st}} \times \frac{f_{st}}{f_x} \times \frac{n_x^2}{n_{st}^2} (3)$$

where F is the area under the emission peak, f is the absorption factor (see below), n is the refractive index of the solvent, Φ is the quantum yield, the subscript x corresponds to the complexes, the subscript st – for standards.

$$f = 1 - 10^{-A} (4)$$

where A is absorbance at the excitation wavelength.

Molecular cloning

Plasmids for mammalian expression of H2B-TagBFP fusions with FAST and FAST mutants (R52L, R52Y) were assembled using Golden Gate cloning following MoClo standard²⁵. Each transcriptional unit for mammalian expression consisted of the CMV promoter, coding sequence for the fusion protein, and the SV40 terminator. All Golden Gate cloning reactions were performed in the T4 ligase buffer (ThermoFisher, USA) supplied with 10 U of T4 ligase, 20 U of either BsaI or BpiI (ThermoFisher, USA), and 100 ng of DNA of each DNA fragment. Golden Gate reactions were performed with the following cycling conditions: 30 cycles between 37°C and 16°C (90 sec at 37°C, 180 sec at 16°C).

Cell culture and transient transfection

HeLa cells were grown in Dulbecco's modification of Eagle's medium (DMEM) (PanEco) supplied with 50 U/ml penicillin and 50 µg/ml streptomycin (PanEco), 2 mM L-glutamine (PanEco), and 10% fetal bovine serum (HyClone, Thermo Scientific) at 37°C and 5% CO₂. Cells were plated and grown in 35-mm glass-bottom dishes (SPL). For transient transfection, FuGENE HD reagent (Promega) was used. Imaging was performed after 24-26h after transient transfection. Immediately before imaging DMEM was replaced with Hanks Buffer (PanEco) supplemented with 20 mM HEPES (Sigma).

Fluorescent microscopy

Widefield fluorescence microscopy of live HeLa cells was performed with Leica DMI6000b inverted microscope equipped with HCX PL FLUOTAR L 20x0.40 CORR lens, COOLLED pE-300 light source, Andor Zyla 5.5 sCMOS camera (Andor), and using Blue, mCherry, and CY5 filter cubes. Illumination powers in blue, red and far-red channel were 0.47 W/cm², 0.8 W/cm², and 0.2 W/cm² respectively. Cells were imaged before and after chromophore addition to the final concentration of 20 µM for each chromophore.

Relative brightness calculation

For evaluation of relative brightness of FAST mutants with chromophores, we selected regions of interests (ROI) in cell nuclei. Then the target signal in red of the far-red channel in each ROI was divided

by the signal in the channel of the TagBFP fluorescent protein. Obtained values of the relative brightness were used to draw boxplots. To obtain p-values two-tailed t-test was used.

Declarations

Acknowledgments

The work was supported by the Russian Science Foundation grant # 18-73-10105. MD simulations were supported by the HSE University Basic Research Program and carried out with the use of computational facilities of the Supercomputer Center “Polytechnical” at the St. Petersburg Polytechnic University and IACP FEB RAS Shared Resource Center “Far Eastern Computing Resource” equipment (<https://cc.dvo.ru>).

Author Contributions

K.S.M., M.S.B. and R.G.E. designed the experiments, M.V.G. and S.A.G. synthesized the recombinant proteins, K.S.M. performed the NMR experiments, N.S.B. measured the fluorescence efficiency of the mutant proteins, A.Y.S. did the QM calculations, D.E.N. performed the MD simulations, A.S.G. and A.S.M performed the living cell transfection and fluorescence microscopy, K.S.M., M.S.B., R.G.E. and A.S.A. supervised the project, M.S.B. acquired the funding, K.S.M. and M.S.B. wrote the paper with assistance from all the authors.

Competing Interests

Authors declare no competing interests.

Data availability

All the data are available on request to the corresponding authors

References

1. Plamont, M.-A., et al.: Small fluorescence-activating and absorption-shifting tag for tunable protein imaging in vivo. *PNAS* **113**, 497–502 (2016)
2. Broch, F., Gautier, A.: Illuminating Cellular Biochemistry: Fluorogenic Chemogenetic Biosensors for Biological Imaging. *ChemPlusChem* **85**, 1487–1497 (2020)
3. Aissa, H.B., Gautier, A.: Engineering Glowing Chemogenetic Hybrids for Spying on Cells. *Eur. J. Org. Chem.* **2020**, 5637–5646 (2020)
4. Chudakov, D.M., Matz, M.V., Lukyanov, S., Lukyanov, K.A.: Fluorescent Proteins and Their Applications in Imaging Living Cells and Tissues. *Physiol. Rev.* **90**, 1103–1163 (2010)

5. Rodriguez, E.A., et al.: The Growing and Glowing Toolbox of Fluorescent and Photoactive Proteins. *Trends Biochem. Sci.* **42**, 111–129 (2017)
6. Mishin, A.S., Belousov, V.V., Solntsev, K.M., Lukyanov, K.A.: Novel uses of fluorescent proteins. *Curr. Opin. Chem. Biol.* **27**, 1–9 (2015)
7. Chen, C., et al.: Developing Bright Green Fluorescent Protein (GFP)-like Fluorogens for Live-Cell Imaging with Nonpolar Protein–Chromophore Interactions. *Chemistry – A European Journal* **27**, 8946–8950 (2021)
8. Povarova, N.V., et al.: Red-Shifted Substrates for FAST Fluorogen-Activating Protein Based on the GFP-Like Chromophores. *Chemistry – A European Journal* **25**, 9592–9596 (2019)
9. Myasnyanko, I.N., et al.: Color Tuning of Fluorogens for FAST Fluorogen-Activating Protein. *Chemistry – A European Journal* **27**, 3986–3990 (2021)
10. Benaissa, H., et al.: An engineered multifunctional protein tag for advanced fluorescence imaging. *Nat. Comm.* **12**, 6989 (2021)
11. Li, C., et al.: A Far-Red Emitting Fluorescent Chemogenetic Reporter for In Vivo Molecular Imaging. *Angew. Chem. Int. Ed.* **59**, 17917–17923 (2020)
12. Tebo, A.G., Pimenta, F.M., Zhang, Y., Gautier, A.: Improved Chemical-Genetic Fluorescent Markers for Live Cell Microscopy. *Biochemistry* **57**, 5648–5653 (2018)
13. Tebo, A.G., et al.: Orthogonal fluorescent chemogenetic reporters for multicolor imaging. *Nat Chem Biol* **17**, 30–38 (2021)
14. Emanuel, G., Moffitt, J.R., Zhuang, X.: High-throughput, image-based screening of pooled genetic-variant libraries. *Nat Methods* **14**, 1159–1162 (2017)
15. Tebo, A.G., Gautier, A.: A split fluorescent reporter with rapid and reversible complementation. *Nat Commun* **10**, 2822 (2019)
16. Mineev, K.S., et al.: NanoFAST: structure-based design of a small fluorogen-activating protein with only 98 amino acids. *Chem. Sci.* **12**, 6719–6725 (2021)
17. Siskos, M.G., Tzakos, A.G., Gerothanassis, I.P.: Accurate ab initio calculations of O–H···O and O–H···[−]O proton chemical shifts: towards elucidation of the nature of the hydrogen bond and prediction of hydrogen bond distances. *Org. Biomol. Chem.* **13**, 8852–8868 (2015)
18. Farrow, N.A., et al.: Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ¹⁵N NMR relaxation. *Biochemistry* **33**, 5984–6003 (1994)
19. Dosset, P., Hus, J.C., Blackledge, M., Marion, D.: Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J. Biomol. NMR* **16**, 23–28 (2000)
20. Abraham, M.J., et al.: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2 GROMACS**, 19–25 (2015)
21. Lindorff-Larsen, K., et al.: Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function, and Bioinformatics* NA-NA (2010) doi:10.1002/prot.22711
22. *Theory and applications of computational chemistry: the first forty years.* (Elsevier, 2005)

23. Schmidt, M.W., et al.: General atomic and molecular electronic structure system. *J. Comput. Chem.* **14**, 1347–1363 (1993)
24. Würth, C., Grabolle, M., Pauli, J., Spieles, M., Resch-Genger, U.: Relative and absolute determination of fluorescence quantum yields of transparent samples. *Nat Protoc* **8**, 1535–1550 (2013)
25. Engler, C., Marillonnet, S.: Generation of Families of Construct Variants Using Golden Gate Shuffling. In: Lu, C., Browse, J., Wallis, J.G. (eds.) *cDNA Libraries*, vol. 729, pp. 167–181. Humana Press (2011)
26. Case, D.A., et al., Amber 2021, in Amber 2021 (University of California, 2021)

Figures

Figure 1

Dynamics of FAST/N871b complex. At the top - the NMR-derived spatial structure of FAST/**N871b** is colored according to the generalized order parameters (S^2) of NH bonds (pink and red, ps-ns motions) and exchange contributions to the transverse relaxation (R_{ex} , blue, us-ms motions). At the bottom - the resulting FAST/**N871b** structure after 100 ns MD simulations is colored according to the magnitudes of atom B-factors. Blue and red correspond to relatively stable and mobile regions, respectively.

A

Figure 2

A - Residues, selected for mutagenesis in the current study (shown with letters and numbers). **B** - electrostatic potential distribution over the contact surface of **N871b** according to DFT calculations. Red zones correspond to the negative potential, blue - to the positive one.

Figure 3

The absorption and emission spectra of various fluorogens in complexes with FAST protein and its mutants. Absorption spectra are normalized to the protein/ligand complex concentration and represented in the molar absorptivity coefficient scale. Emission spectra are normalized to the FQY.

Figure 4

Comparison of relative brightness of FAST mutants with different chromophores in live HeLa Kyoto cells transiently transfected with H2B-TagBFP-FAST(or its mutants) constructs. Whiskers show standard deviation, horizontal lines indicate median values, and circles indicate outliers. n = 29 for each box. Stars indicate the p-values: * p<0.05, ** p<0.01, *** p<0.001, ns p>0.05.

Figure 5

A - Overlay of ^1H NMR spectra of FAST/**N871b** complexes, recorded for the wild-type protein and its three mutants: R52Y, R52L and P68T. Spectra were obtained at 25 °C, molecular brightness of the complex is indicated. B - correlation between the chemical shift of $\text{H}_{\epsilon 1}$ W94 and brightness of FAST/**N871b** complex. C,D - correlations between the NMR-derived stability of three intermolecular H-bonds and FAST/**N871b** brightness (C) or molar extinction coefficient (ϵ , D). E - Pearson correlation coefficients, obtained for the chemical shifts and relative intensities of signals, corresponding to intermolecular H-bond protons of FAST/**N871b** complexes, and either quantum yield (Q.Y.) or molar extinction coefficient. Bars are colored according to the statistical significance of the correlation.

Figure 6

11 structures of FAST/**N871b** complex (FAST, R52Y and R52L mutants), selected from the last 10 ns of the MD simulation in 1 ns interval. **N871b** is shown in magenta and the sidechain of residue 52 is shown by orange sticks. Structures are superimposed over the backbone atoms of the β -sheet residues.

Figure 7

A - structure of the FAST/**N871b** complex. Sidechains of residues that were mutated in any of the enhanced FAST variants are shown by sticks and colored with respect to the presumed role of the mutation. Ligand is shown in black. **B** - occurrences of mutations at certain positions in 58 reported enhanced FAST variants. Magenta corresponds to the residues, directly involved in the ligand binding; blue - to the residues that support the packing of the N-terminal domain; cyan - to the residues in the helix H5, red - to the residues in mobile loops. Gray denotes the mutations that do not fall in any of the listed groups.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [mcFASTYLSupplementaryv1.pdf](#)
- [nrreportingsummarycommsbio220084T.pdf](#)