

# Mutants of Monomeric Red Fluorescent Protein mRFP1 at Residue 66: Structure Modeling by Molecular Dynamics and Search for Correlations with Spectral Properties

E. E. Khrameeva<sup>1</sup>, V. L. Drutsa<sup>2</sup>, E. P. Vrzheschch<sup>1</sup>, D. V. Dmitrienko<sup>3</sup>, and P. V. Vrzheschch<sup>1,3\*</sup>

<sup>1</sup>Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, 119992 Moscow, Russia; fax: (495) 939-4218; E-mail: peter@genebee.msu.ru

<sup>2</sup>Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, 119992 Moscow, Russia; fax: (495) 938-2323; E-mail: drutsa@genebee.msu.ru

<sup>3</sup>International Biotechnological Center, Lomonosov Moscow State University, 119992 Moscow, Russia; fax: (495) 939-5022

Received February 4, 2008

Revision received March 17, 2008

**Abstract**—To study the interrelation between the spectral and structural properties of fluorescent proteins, structures of mutants of monomeric red fluorescent protein mRFP1 with all possible point mutations of Glu66 (except replacement by Pro) were simulated by molecular dynamics. A global search for correlations between geometrical structure parameters and some spectral characteristics (absorption maximum wavelength, integral extinction coefficient at the absorption maximum, excitation maximum wavelength, emission maximum wavelength, and quantum yield) was performed for the chromophore and its 6 Å environment in mRFP1, Q66A, Q66L, Q66S, Q66C, Q66H, and Q66N. The correlation coefficients (0.81–0.87) were maximal for torsion angles in phenolic and imidazolidine rings as well as for torsion angles in the regions of connection between these rings and chromophore attachment to β-barrel. The data can be used to predict the spectral properties of fluorescent proteins based on their structures and to reveal promising positions for directed mutagenesis.

DOI: 10.1134/S0006297908100040

**Key words:** fluorescent proteins, mRFP1, mutants at residue 66, spectral characteristics, structure, correlations, molecular dynamics

Fluorescent proteins are widely used as markers for visualization of processes in intact biological systems. The family of fluorescent proteins includes proteins of various representatives of Coelenterata that are able to absorb light and emit in the visible spectral range. All fluorescent proteins have similar structure represented by 11-strand β-barrel with α-helix inside, which contains a chromophore in the middle. The chromophore is a heterogroup formed by an autocatalytic posttranslational reaction between three adjacent amino acid residues. The presence of the chromophore imparts color to the protein and induces its possible fluorescence [1].

Fluorescence of proteins of this family is influenced by external conditions: pH [2, 3], temperature [1], and ionic content of the medium [4, 5], but structures of the chromophore and its nearest environment [1, 6] are the key factors. The representatives of this family in maturation rate, stability, spectral properties (absorption and fluorescence spectra, fluorescence quantum yield, photobleaching, etc.) are described in [1, 7]. Theoretical, experimental, and computational studies are carried out to investigate the mechanism of fluorescence of these proteins [8–11].

Green fluorescent protein (GFP) was discovered in the early 1960s [12], but its active study began only after cloning of the GFP gene in 1992 [13] and demonstration of its heterologous expression in other organisms. In 1999, another family of colored proteins including the red protein DsRed was cloned from corals [14]. However, DsRed is a tetramer, and its use as a fluorescent marker is limited because of possible effects of its large molecular mass

**Abbreviations:** DsRed) red fluorescent protein drFP583 from coral *Discosoma* sp.; GFP) green fluorescent protein from *Aequorea victoria*; MD) molecular dynamics; mRFP1) monomeric red fluorescent protein, mutant of DsRed; Q66X) mutants of mRFP1 with replacement of residue 66 by residue X.

\* To whom correspondence should be addressed.

on functioning of the protein under investigation or its toxic effect on cells.

The first monomer derived from DsRed was mRFP1; for this purpose, 33 mutations were made in DsRed [15]. Red emission is also typical of the fluorescent protein mRFP1, so it is of special interest for minimization of the background signal [16]. Although mRFP1 is already widely used in cell biology, attempts to find new forms of red fluorescent proteins in order to improve their properties (quantum yield, brightness, photo- and pH stability, etc.) are performed.

However, the interrelation between spectral and structural properties of fluorescent proteins necessary for development of these studies is rather unclear. A method for prediction of properties of fluorescent proteins based on their structure, that is, a general algorithm for design of proteins with pre-defined properties, is still not developed. This problem might be solved by analysis of properties of mutant proteins with point mutations and subsequent molecular dynamics (MD) simulations of their structures *in silico* in order to find correlations between structural parameters and spectral properties.

In this work we simulated the structures of mRFP1 mutants with 18 point mutations of Glu66 by molecular dynamics, obtained mutant proteins in which Glu66 is replaced by an Ala, Leu, Ser, Cys, His, or Asp residue, and searched for interrelations between structural characteristics of the modeled proteins and spectral properties of the mutants.

## MATERIALS AND METHODS

In this study, the GROMACS program package was used for MD simulations [17, 18].

**Force field.** For MD calculations of protein trajectory in the selected force field, interaction potential parameters of all atoms in a protein molecule should be set. In the case of fluorescent proteins, it is necessary to have such parameters also for atoms of the chromophore incorporated into the protein chain as an amino acid residue. The extended force field OPLS-AA/DsRed1 was selected for MD calculations [19] because the chromophore of red fluorescent protein DsRed is parameterized in the OPLS-AA/DsRed1 force field; this allows MD simulations of all proteins containing DsRed chromophore, including simulations of mRFP1 mutants.

Chromophores of 18 mRFP1 mutants (except X = Pro) are similar to the DsRed chromophore, excluding the site replaced as a result of mutation of residue 66. That is why a new parameterization was unnecessary for these chromophores, and parameters of corresponding amino acid residues were used as parameters for side chain atoms of residue 66. For this purpose, in the force field file \*.rtp (\*.rtp = residue topology parameter file) with topology parameters of amino acid residues, the

atom types, new chemical bonds, valent and torsion angles, atom charges, and parameters of van der Waals interactions were recorded for each parameterized chromophore. The force field file \*.hdb (\*.hdb = hydrogen data base) was also edited: in this file, each hydrogen atom is assigned one of nine possible parameter sets according to atom type [20]. Thus, chromophores of 18 mRFP1 mutants were parameterized in the OPLS-AA/DsRed1 force field.

**Specification of the starting structure.** Since mRFP1 is a DsRed mutant, it is reasonable to suppose that the 3D structure of the mRFP1 monomer is analogous to that of DsRed tetramer subunit. Thus, the structure of the DsRed subunit (PDB ID: 1ZGO) was used as the starting structure for MD simulation of mRFP1; the structure file was retrieved from the Protein Data Bank site (<http://www.pdb.org>) [21, 22]. This file contains X-ray coordinates of the non-hydrogen atoms of the DsRed tetramer [23].

To build up the mRFP1 structure model, the PDB file was modified by 33 mutations using the Swiss-PdbViewer program [24]; thus, PDB files for each of the mRFP1 mutants with corresponding substitutions at residue 66 were created. After energy minimization, these models were used for MD simulations.

**MD simulations.** The mdrun program was used for energy minimization [25, 26]. The energy minimum was reached using the Steepest Descent (SD) minimization method with the following parameters: step (Emstep), 0.01 nm; maximal energy gradient (Emtol), 2000 kJ/(mol·nm) as the criterion for procedure termination. The energy minimization procedure yielded a coordinate file that was then used for MD simulation with limited mobility of protein atoms. Parameters for mdrun program were the following: integration step, 2 fsec; total calculation time, 10 psec; Berendsen thermostat [26]; temperature (ref\_t), 300 K; thermostat characteristic time (tau\_t), 0.1 psec. MD simulation with the limited mobility of peptide chain atoms yielded the coordinate file which was used for MD simulation of system trajectory at 300 K. Parameters for the mdrun program were the following: integration step, 2 fsec; Berendsen thermostat; temperature (ref\_t), 300 K; thermostat characteristic time (tau\_t), 0.1 psec; Parinello–Rachman isotropic barostat; barostat characteristic time (tau\_p), 0.5 psec; pressure (ref\_p), 1.0 bar; compressibility,  $4.5 \cdot 10^{-5}$ . As a result, MD trajectories for mRFP1 and its mutants, which contained coordinates and velocities of each atom after each 500 fsec, were obtained.

MD simulations for mRFP1 and all its possible mutants at residue 66 except for proline, for which chromophore parameterization is still absent, were performed as described above [19]. Moreover, formation of the chromophore in the proline mutant at residue 66 is hardly possible; this is supported by complete absence of data on the existence of such proline mutants in the family of flu-

orescent proteins. For further analysis, we used MD trajectories corresponding to mRFP1 and its six mutants (Q66A, Q66N, Q66H, Q66C, Q66S, and Q66L), for which spectral properties were determined (absorption maximum wavelength, integral extinction coefficient at the absorption maximum, fluorescence emission maximum wavelength, and quantum yield).

**Analysis of MD trajectories.** Root mean square deviations (RMSD) of atom positions from those in the starting structure were calculated for MD trajectories using the *g\_rms* program [20]. This program fits the present structure corresponding to the time point  $t$  with the starting structure corresponding to  $t = 0$  by the least squares method; then RMSD is calculated by the equation:

$$RMSD(t) = \left[ \frac{1}{M} \sum_{i=1}^N m_i \left\| \vec{r}_i(0) - \vec{r}_i(t) \right\|^2 \right]^{\frac{1}{2}},$$

where  $\vec{r}_i(t)$  is the radius-vector of atom  $i$  at time point  $t$ ,  $\vec{r}_i(0) - \vec{r}_i(t)$  is the distance between positions of atom  $i$  at  $t = 0$  and at time point  $t$ ,  $m_i$  is mass of atom  $i$ ,  $N$  is the total number of atoms in the molecule, and

$$M = \sum_{i=1}^N m_i \text{ is the total mass of all atoms.}$$

Based on RMSD values for each studied protein, extended sections of the trajectory with the maximally unchanged RMSD were chosen. Average protein structures were built up via these sections of the MD trajectories using the *g\_covar* program [20]. These average structures were superimposed using Swiss-PdbViewer (superposition via all residues) [24].

Based on the same sections of the MD trajectories that were used for building up the average structures, we calculated distributions of bond distances and valent and torsion angles and determined their most probable values using the *g\_bond* and *g\_angle* programs of the GRO-MACS package [20].

**Correlation analysis.** For all bond distances and valent and torsion angles considered here, correlations between the most probable values of these parameters and experimental spectral data for corresponding mutant proteins were analyzed.

These correlations were obtained by calculations of the Pearson correlation coefficient applying the formula for normally distributed variables:

$$r_{xy} = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}},$$

where  $x_i$  are values taken in sampling X,  $y_i$  are values taken in sampling Y,  $\bar{x}$  is the average value in sampling X, and  $\bar{y}$  is the average value in sampling Y.

**Preparation of mutant proteins and evaluation of their spectral properties.** Mutants Q66A, Q66N, Q66H, Q66C, Q66S, and Q66L of mRFP1 protein were obtained by expression of the cloned mutant mRFP1 genes with subsequent isolation on Ni-NTA agarose [27]. Fluorescence and absorption spectra were recorded by a Cary Eclipse (Varian Inc., USA) in K-phosphate buffer (67 mM  $K_2HPO_4$ , 67 mM  $NaH_2PO_4$ , pH 7.5) at 25°C. Integral extinction coefficients were determined as a ratio of protein absorption and its total concentration. Fluorescence quantum yields were determined by comparative method [28] with rhodamine 6G as a standard (quantum yield in ethanol 0.95 [29]).

## RESULTS AND DISCUSSION

**Calculation of root mean square deviations.** Changes in root mean square deviations were plotted for all calculated MD trajectories, and they appeared to be similar for mRFP1 and all its 18 mutants. One of these plots is presented in Fig. 1 (Q66N mutant). RMSD change depending on MD calculation time is similar for mRFP1 and all its 18 mutants: RMSD increases up to 2 nsec, and then the growth becomes slower (protein structure changes become negligible). There were trajectory sections from 2 to 5 nsec, which were used for building up average protein structures and searching for bond distance and valent and torsion angle values.

**Superposition of average structures.** Average structures of mRFP1, Q66A, Q66N, Q66H, Q66C, Q66S, and Q66L were calculated via MD trajectory sections from 2 to 5 nsec. For comparison, the structures were superimposed. It appeared that chromophores remained flat, changes in their form were minimal, and displacements of other residues were negligible.

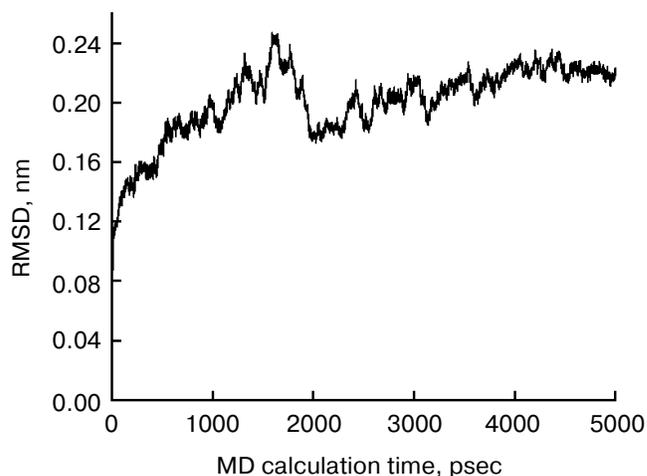
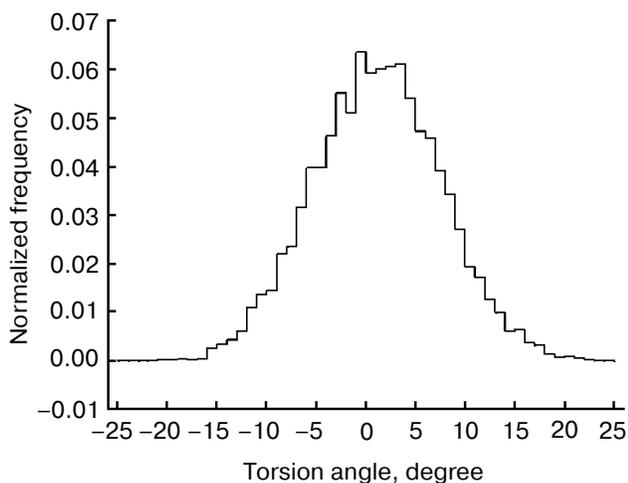


Fig. 1. Dynamics of RMSD of Q66N MD structure.



**Fig. 2.** Distribution of CZ\_CRQ66\_CE2\_CRQ66\_CD2\_CRQ66\_CG2\_CRQ66 torsion angle of mutant protein Q66N via MD trajectory.

**Calculation of most probable values of bond distances and bond and torsion angles.** Based on MD trajectories, distributions of all bond distance and valent and torsion angle values for mRFP1, Q66A, Q66N, Q66H, Q66C, Q66S, and Q66L were plotted. Analysis of distributions showed that they can be approximated as normal. For example, the distribution presented in Fig. 2 was obtained for torsion angle CZ\_CRQ66\_CE2\_CRQ66\_CD2\_CRQ66\_CG2\_CRQ66 in the Q66N chromophore.

For the case presented in Fig. 2, the expectation (average) value of the torsion angle is  $0.66^\circ$ . This is calculated by the following equation:

$$E\xi = \sum_{k=1}^{360} x_k p_k,$$

where  $x_k$  is angle (in this work step is  $1^\circ$ ) and  $p_k$  is normalized frequency defined via MD trajectory (taken as probability). As can be seen from Fig. 2, torsion angle of  $0.66^\circ$  is less probable than, for example,  $-1^\circ$  or  $3^\circ$ . That is why it would be improper to use average bond distance or angle values for correlation analysis, and in this work, the latter was performed using the most probable values (in this case  $-1^\circ$ ).

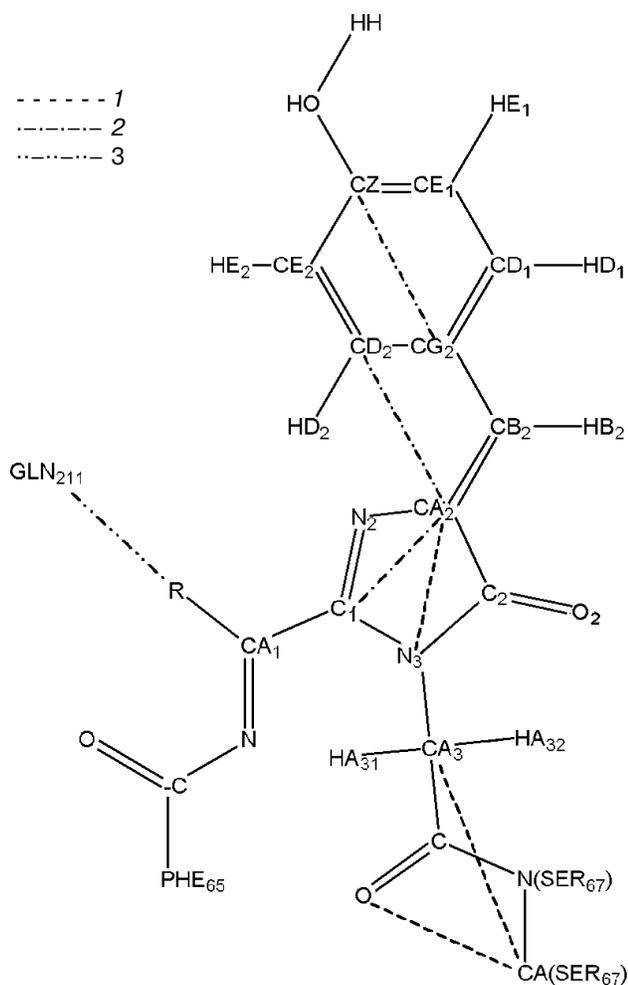
**Correlation between the most probable values of bond distances, valent and torsion angles, and spectral properties of proteins.** As it appeared, the most probable values of bond distances and angles are similar for various mutants (the module of values difference is much less than the distribution half-width), so no significant correlations were found. That is why only correlations for torsion angles were further analyzed. For torsion angles in the chromophore and residues forming an envelope around it (6 Å environment), the following correlations were found.

**A. Correlation between the torsion angles in the chromophore and spectral properties.** The values of six torsion angles in chromophore (C1\_CRQ66\_N3\_CRQ66\_C2\_CRQ66\_CA2\_CRQ66, CA2\_CRQ66\_N2\_CRQ66\_C1\_CRQ66\_N3\_CRQ66, O\_CRQ66\_C\_CRQ66\_N\_SER67\_CA\_SER67, CA3\_CRQ66\_C\_CRQ66\_N\_SER67\_CA\_SER67, CA2\_CRQ66\_CB2\_CRQ66\_CG2\_CRQ66\_CD2\_CRQ66, and CG2\_CRQ66\_CD2\_CRQ66\_CE2\_CRQ66\_CZ\_CRQ66) correlate with the following spectral properties of mRFP1, Q66A, Q66N, Q66H, Q66C, Q66S, and Q66L proteins: absorption maximum wavelength, integral extinction coefficient of the absorption maximum, fluorescence emission maximum wavelength, and quantum yield (Table 1, correlation coefficients not less than 0.80). As shown in Fig. 3, the highest correlation coefficients (0.81-0.87, Table 1) were found for torsion angles in phenolic and imidazolidine rings and also for torsion angles in the region of connection between these

**Table 1.** Correlation coefficients between torsion angles in the chromophore and spectral properties ( $\lambda_{\text{Abs. max}}$ , absorption maximum wavelength;  $\epsilon$ , integral extinction coefficient at the absorption maximum;  $\lambda_{\text{Ex. max}}$ , excitation maximum wavelength;  $\lambda_{\text{Em. max}}$ , fluorescence emission maximum wavelength;  $\psi$ , quantum yield)

Torsion angle	$\lambda_{\text{Abs. max}}$	$\epsilon$	$\lambda_{\text{Ex. max}}$	$\lambda_{\text{Em. max}}$	$\psi$
C1_CRQ66_N3_CRQ66_C2_CRQ66_CA2_CRQ66	-0.44	0.86	-0.31	-0.52	0.81
CA2_CRQ66_N2_CRQ66_C1_CRQ66_N3_CRQ66	-0.83	0.69	-0.74	-0.85	0.6
O_CRQ66_C_CRQ66_N_SER67_CA_SER67	0.43	-0.86	0.3	0.59	-0.42
CA3_CRQ66_C_CRQ66_N_SER67_CA_SER67	0.21	-0.81	0.14	0.44	-0.37
CA2_CRQ66_CB2_CRQ66_CG2_CRQ66_CD2_CRQ66	-0.57	0.61	-0.48	-0.57	0.83
CG2_CRQ66_CD2_CRQ66_CE2_CRQ66_CZ_CRQ66	-0.32	0.84	-0.21	-0.45	0.87

Note: Only torsion angles with the module of correlation coefficient with at least one of the spectral characteristics not less than 0.8 are given. Torsion angle notations correspond with those in Fig. 3.



**Fig. 3.** Structure of the mRFP1 protein chromophore. Lines connect atoms 1 and 4 of torsion angles whose values to the maximal degree correlate with the quantum yield value (1) or other spectral characteristic value (2); 3) a hydrogen bond between the side chain (R) of the substituting amino acid and Gln211.

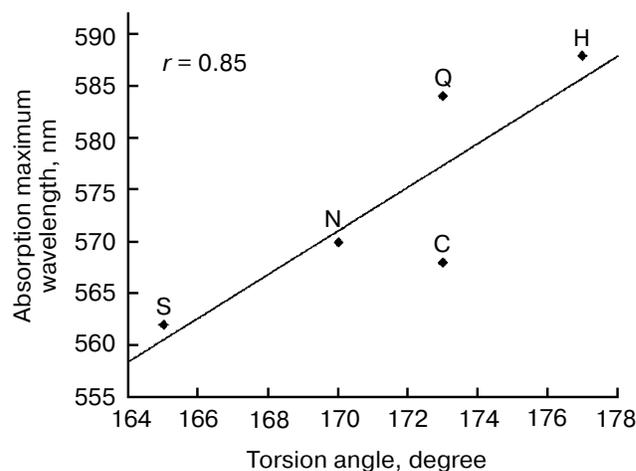
rings and chromophore attachment to  $\beta$ -barrel (via SER67).

**B. Correlations between the torsion angles in the 6 Å environment of the chromophore and spectral properties.** The torsion angles of 43 amino acid residues in the 6 Å environment of the chromophore correlate with spectral properties of proteins mRFP1, Q66A, Q66N, Q66H, Q66C, Q66S, and Q66L (correlation coefficients with at least one spectral property not less than 0.80 (Table 2)). The residues presented in Table 2 are situated in the nearest neighborhood of the chromophore, and the change in their position is probably related with the changes in chromophore structure due to mutations at residue 66.

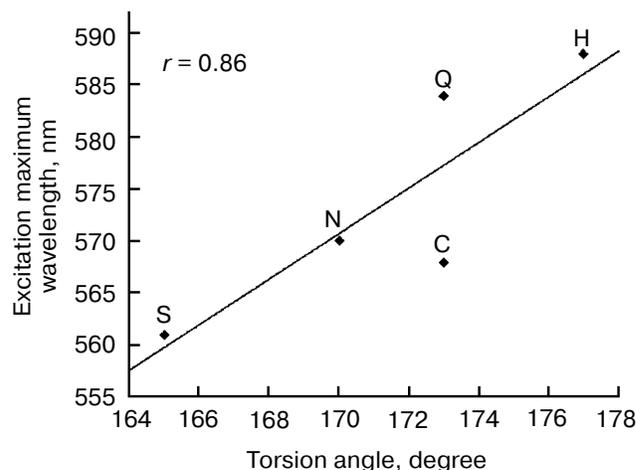
We also found correlations of CA\_GLN211\_C\_GLN211\_N\_TYR212\_CA\_TYR212 torsion angle in the main chain of the GLN211 residue with absorption, excitation, and emission wavelength maxima. For all mutants

the correlation coefficients were 0.74, 0.65, and 0.78, respectively. However, if alanine and leucine mutant proteins are excluded as aliphatic, for the remaining polar replacing residues correlation coefficients increase to 0.85, 0.86, and 0.80, respectively (Figs. 4-6).

**C. Correlations between torsion angles in the chromophore and those in the 6 Å environment of the chromophore.** We analyzed correlations between the torsion angles in the chromophore, which correlate with spectral properties, and those in the 6 Å environment of the chromophore, which also correlate with spectral properties (Tables 1 and 2, respectively). The torsion angles in 25 amino acid residues in the 6 Å environment of the chromophore were found to correlate with at least one torsion



**Fig. 4.** Correlation between torsion angle CA\_GLN211\_C\_GLN211\_N\_TYR212\_CA\_TYR212 and absorption maximum wavelength for mutants at residue 66.



**Fig. 5.** Correlation between torsion angle CA\_GLN211\_C\_GLN211\_N\_TYR212\_CA\_TYR212 and excitation maximum wavelength for mutants at residue 66.

angle of the six considered angles in the chromophore (correlation coefficient not less than 0.80) (Table 3).

**D. Correlations between the volumes of replacing amino acid residues and spectral properties.** For proteins mRFP1, Q66A, Q66N, Q66H, Q66C, Q66S, and Q66L we found correlations between fluorescence excitation and emission wavelength maxima and the volume of the substituting amino acid residues. Red shift of excitation and emission wavelength maxima is observed with increase in amino acid residue volume determined according to [30] (Fig. 7a). For all mutants, excitation maximum wavelengths correlate with residue 66 volume with coefficient 0.67, whereas emission maximum wavelengths correlate with coefficient 0.75. However, if alanine and leucine proteins are excluded from the group of mutants as those with aliphatic substituting residue, for the remaining proteins with polar substituting amino acid residues correlation coefficients increase to 0.99 and 0.93 for excitation and emission maxima, respectively.

For mutants of green fluorescent protein GFP at residue 65, which corresponds with residue 66 of mRFP1 (as the first position of the chromophore), we analyzed analogous correlations using literature data. The wild type GFP contains serine at position 65, and mutants with serine replaced by alanine, cysteine, and threonine were obtained in [11]. If excitation and emission maxima determined in [11] are plotted versus substituting amino acid residue volume [30], correlation coefficients appear to be 0.92 and 0.71, respectively (Fig. 7b).

Analysis of the MD simulations showed that the spectral properties of the proteins correlate with torsion angles in the chromophore and in residues of its 6 Å environment that seem to influence chromophore structure distortions to the maximal extent.

**A. Correlations between torsion angles in the chromophore and spectral properties.** The presence of a relatively large central cavity containing the chromophore was shown for GFP by molecular mechanics and MD

**Table 2.** Correlation coefficients of torsion angles in the chromophore environment and spectral properties

Torsion angle	$\lambda_{\text{Abs. max}}$	$\epsilon$	$\lambda_{\text{Ex. max}}$	$\lambda_{\text{Em. max}}$	$\psi$
1	2	3	4	5	6
N_PHE14_CA_PHE14_CB_PHE14_CG_PHE14	-0.9	0.13	-0.91	-0.84	0.26
CA_VAL16_C_VAL16_N_ARG17_CA_ARG17	-0.81	0.11	-0.74	-0.65	-0.1
CA_GLY33_C_GLY33_N_GLU34_CA_GLU34	0.67	-0.57	0.64	0.8	-0.57
CB_THR43_CA_THR43_C_THR43_N_ALA44	0.89	0.07	0.94	0.73	0.1
N_LEU46_CA_LEU46_CB_LEU46_CG_LEU46	-0.82	0.11	-0.77	-0.66	0.18
CA_LEU61_CB_LEU61_CG_LEU61_CD1_LEU61	0.95	-0.31	0.88	0.79	-0.2
CB_GLN64_CA_GLN64_C_GLN64_O_GLN64	0.33	-0.72	0.26	0.62	-0.92
CB_GLN64_CA_GLN64_C_GLN64_N_PHE65	0.36	-0.63	0.3	0.66	-0.8
N_GLN64_CA_GLN64_C_GLN64_N_PHE65	0.33	-0.48	0.29	0.57	-0.81
N_GLN64_CA_GLN64_CB_GLN64_CG_GLN64	-0.53	0.9	-0.46	-0.78	0.9
C_GLN64_CA_GLN64_CB_GLN64_CG_GLN64	-0.46	0.74	-0.38	-0.72	0.83
CA_PHE65_CB_PHE65_CG_PHE65_CD1_PHE65	0.79	-0.51	0.69	0.8	-0.49
CD1_PHE65_CE1_PHE65_CZ_PHE65_CE2_PHE65	0.81	-0.6	0.77	0.92	-0.32
CE1_PHE65_CD1_PHE65_CG_PHE65_CD2_PHE65	-0.88	0.56	-0.79	-0.8	0.22
CB_SER67_CA_SER67_C_SER67_O_SER67	-0.33	0.85	-0.17	-0.51	0.52
CB_SER67_CA_SER67_C_SER67_N_LYS68	-0.32	0.88	-0.16	-0.52	0.56
N_SER67_CA_SER67_C_SER67_O_SER67	-0.31	0.87	-0.15	-0.5	0.54
N_SER67_CA_SER67_C_SER67_N_LYS68	-0.3	0.88	-0.15	-0.52	0.56
C_SER67_N_LYS68_CA_LYS68_CB_LYS68	0.15	-0.86	0	0.39	-0.5
C_SER67_N_LYS68_CA_LYS68_C_LYS68	0.13	-0.86	-0.01	0.37	-0.49
CA_LYS68_C_LYS68_N_ALA69_CA_ALA69	0.5	-0.97	0.41	0.72	-0.81
O_LYS68_C_LYS68_N_ALA69_CA_ALA69	0.5	-0.97	0.41	0.73	-0.81
N_ALA69_CA_ALA69_C_ALA69_O_ALA69	-0.74	0.18	-0.81	-0.72	-0.04

**Table 2.** (Contd.)

1	2	3	4	5	6
N_ALA69_CA_ALA69_C_ALA69_N_TYR70	-0.9	0.38	-0.89	-0.84	0.2
CA_PHE89_C_PHE89_N_LYS90_CA_LYS90	0.87	-0.33	0.85	0.89	-0.11
O_PHE89_C_PHE89_N_LYS90_CA_LYS90	0.82	0.06	0.81	0.69	-0.09
N_PHE89_CA_PHE89_CB_PHE89_CG_PHE89	0.73	0	0.82	0.76	-0.08
CE2_PHE89_CD2_PHE89_CG_PHE89_CD1_PHE89	0.5	-0.97	0.41	0.72	-0.8
CB_TRP91_CA_TRP91_C_TRP91_N_GLU92	0.82	-0.49	0.77	0.74	-0.55
CE3_TRP91_CD2_TRP91_CE2_TRP91_NE1_TRP91	-0.9	0.34	-0.9	-0.82	0.06
CE3_TRP91_CD2_TRP91_CG_TRP91_CB_TRP91	0.02	-0.66	0.01	0.39	-0.83
C_ARG93_CA_ARG93_CB_ARG93_CG_ARG93	0.8	-0.38	0.75	0.85	-0.28
C_THR106_N_GLN107_CA_GLN107_C_GLN107	0.56	-0.83	0.44	0.71	-0.69
CD2_TYR118_CE2_TYR118_CZ_TYR118_OH_TYR118	-0.59	0.46	-0.53	-0.67	0.8
CB_VAL120_CA_VAL120_C_VAL120_N_LYS121	-0.07	0.51	0.03	-0.24	0.85
N_VAL120_CA_VAL120_CB_VAL120_CG1_VAL120	0.03	0.88	0.17	-0.21	0.86
O_TRP141_C_TRP141_N_GLU142_CA_GLU142	-0.75	0.37	-0.79	-0.85	0.15
CE3_TRP141_CD2_TRP141_CG_TRP141_CD1_TRP141	0.81	-0.6	0.77	0.92	-0.32
CB_GLU142_CG_GLU142_CD_GLU142_OE2_GLU142	0.79	-0.35	0.76	0.88	-0.28
CB_SER144_CA_SER144_C_SER144_N_THR145	0.46	-0.61	0.32	0.57	-0.81
C_SER144_N_THR145_CA_THR145_CB_THR145	-0.52	0.5	-0.44	-0.63	0.81
CB_GLU146_CG_GLU146_CD_GLU146_OE1_GLU146	-0.82	0.16	-0.84	-0.63	-0.07
C_GLU174_N_VAL175_CA_VAL175_CB_VAL175	-0.11	-0.56	-0.18	0.01	-0.82
C_GLU174_N_VAL175_CA_VAL175_C_VAL175	0.07	-0.52	0.02	0.18	-0.86
CA_VAL_175_C_VAL175_N_LYS176_CA_LYS176	0.15	0.76	0.2	-0.15	0.82
C_THR178_N_TYR179_CA_TYR179_CB_TYR179	-0.07	-0.78	-0.14	0.19	-0.87
C_TYR179_CA_TYR179_CB_TYR179_CG_TYR179	0.75	-0.37	0.78	0.84	-0.17
CA_TYR179_C_TYR179_N_MET180_CA_MET180	-0.81	0.61	-0.77	-0.92	0.33
CD1_TYR179_CE1_TYR179_CZ_TYR179_OH_TYR179	0.72	0.27	0.8	0.62	0.24
CE1_TYR179_CD1_TYR179_CG_TYR179_CD2_TYR179	0.05	0.47	0.1	-0.16	0.89
CB_ILE195_CA_ILE195_C_ILE195_O_ILE195	-0.69	0.15	-0.8	-0.8	0.26
CA_LEU197_C_LEU197_N_ASP198_CA_ASP198	-0.6	0.86	-0.48	-0.8	0.63
CA_ASP198_C_ASP198_N_ILE199_CA_ILE199	0.01	-0.81	-0.14	0.31	-0.72
C_ASP198_CA_ASP198_CB_ASP198_CG_ASP198	-0.7	0.44	-0.7	-0.8	0.26
CA_ASP198_CB_ASP198_CG_ASP198_OD1_ASP198	0.01	-0.86	-0.08	0.32	-0.6
CA_TYR212_CB_TYR212_CG_TYR212_CD2_TYR212	0.84	-0.09	0.79	0.63	-0.09
O_GLN211_C_GLN211_N_TYR212_CA_TYR212	0.87	-0.16	0.84	0.75	-0.34
CD2_TYR212_CE2_TYR212_CZ_TYR212_CE1_TYR212	0.81	-0.6	0.77	0.92	-0.32
C_TYR212_N_GLU213_CA_GLU213_CB_GLU213	-0.73	0.49	-0.71	-0.81	0.3
O_GLU213_C_GLU213_N_ARG214_CA_ARG214	-0.77	-0.21	-0.8	-0.6	-0.18
CA_ARG214_C_ARG214_N_ALA215_CA_ALA215	0.73	-0.56	0.64	0.82	-0.41
CA_ARG214_CB_ARG214_CG_ARG214_CD_ARG214	0.21	-0.87	0.06	0.43	-0.5
CB_ARG214_CG_ARG214_CD_ARG214_NE_ARG214	-0.8	-0.15	-0.89	-0.7	-0.02

Note: Torsion angles for which correlation coefficients with at least one spectral property are not less than 0.8 are presented. Notations correspond with those in Table 1.

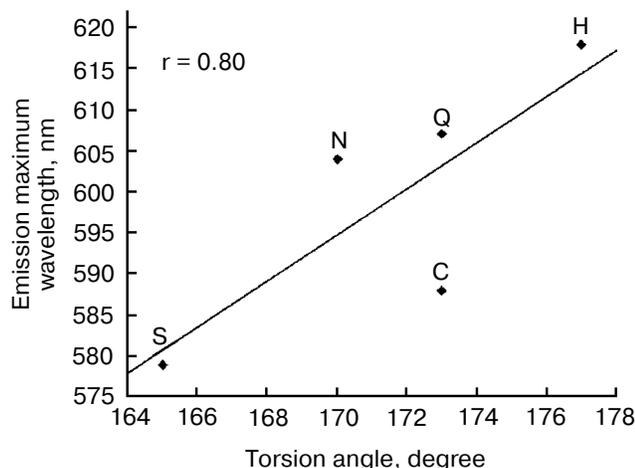


Fig. 6. Correlation between torsion angle CA\_GLN211\_C\_GLN211\_N\_TYR212\_CA\_TYR212 and emission maximum wavelength for mutants at residue 66.

[31]. However, its form is not complementary to a planar chromophore. The protein causes a certain tension in the chromophore structure when it is planar, and the presence of a planar chromophore in GFP can be rationalized only by delocalization of  $\pi$  electrons. The protein environment of GFP gives some rotational freedom to the chromophore, especially in the hula-twist direction (cooperative change in  $\tau$  and  $\varphi$  angles (Fig. 8) resulting in rotation of the methene bridge between phenolic and imidazolidine rings of the chromophore in the vertical plane perpendicular to the plane of the figure) and via torsion angle  $\varphi$  [16].

Since mRFP1 is homologous to GFP and its chromophore is also a system of conjugated  $\pi$  bonds, all the

characteristics mentioned above are also true for mRFP1. Deviations of torsion angles presented in Table 1 from 0 and  $180^\circ$  indicate that the chromophore structure becomes less planar, and this probably results in changes in the spectral properties of the corresponding proteins. This can be supported by the high correlation coefficients between spectral characteristics and torsion angles in the chromophore (Table 1).

**B. Correlations between torsion angles in the  $6 \text{ \AA}$  environment of the chromophore and spectral properties.** The structure of the fluorescent protein is presented by 11-strand  $\beta$ -barrel with  $\alpha$ -helix inside, which contains a chromophore in its middle part. The integrity of the  $\beta$ -barrel is maintained by multiple hydrogen bonds with energy high enough (2-5 kcal/mol per bond) for  $\beta$ -barrel to become a rather rigid system. Such rigid  $\beta$ -barrel seems to influence the conformation of the chromophore, which is exactly planar in vacuum [19], but its structure is more tense within the protein (torsion angle values vary from 0 and  $180^\circ$ ).

In the chromophore environment, 43 amino acid residues are found with torsion angles correlating with spectral properties of the proteins. However, the torsion angles of most (25) residues were shown to correlate with chromophore torsion angles sensitive to spectral properties of the protein. We suggest that distortion in chromophore geometry is the cause of the change in protein spectral properties, and structural changes in the chromophore environment can affect chromophore geometry (supported by correlations presented in Table 3). Then it is natural to suggest that changes in the chromophore environment can indirectly cause changes in spectral properties, which is supported by the correlations found (Table 2).

It should be noted that only 63 torsion angle values correlate with spectral properties; these 63 angles are in

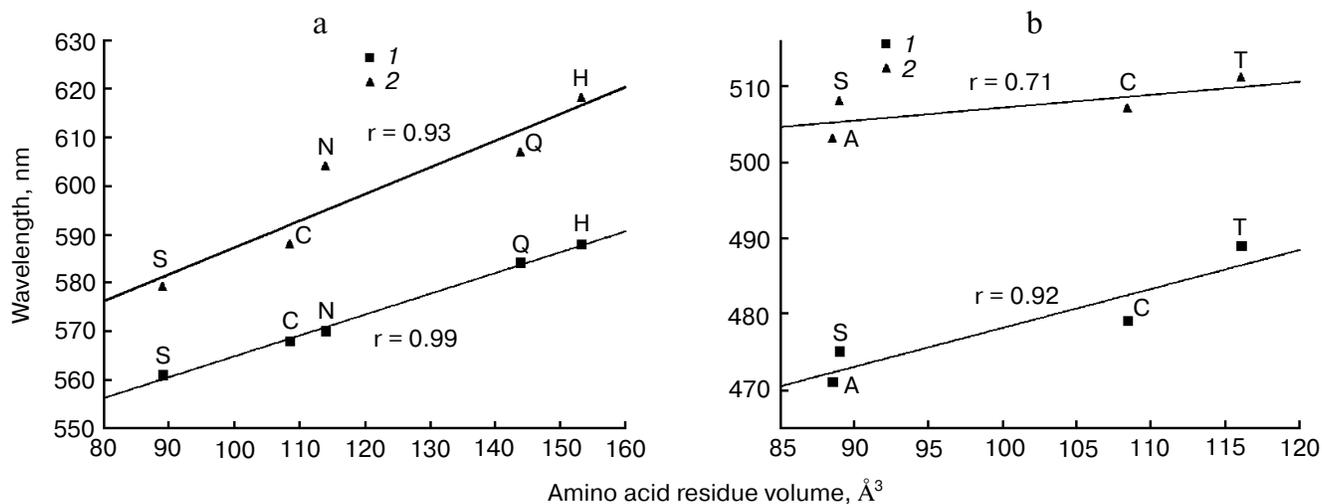


Fig. 7. Correlation between the volume of the amino acid residue at the first position of the chromophore and fluorescence excitation (1) and emission (2) wavelength maxima for proteins: a) mRFP1 (substitution at residue 66); b) GFP (substitution at residue 65).

**Table 3.** Correlation coefficients between the torsion angles in the chromophore and in its environment

Torsion angle	1	2	3	4	5	6
CA_GLY33_C_GLY33_N_GLU34_CA_GLU34	-0.37	-0.88	0.44	-0.16	-0.5	-0.67
N_GLN64_CA_GLN64_CB_GLN64_CG_GLN64	0.83	0.76	-0.69	0.29	0.74	0.79
CA_PHE65_CB_PHE65_CG_PHE65_CD1_PHE65	-0.46	-0.91	0.49	0.04	-0.59	-0.58
CE1_PHE65_CD1_PHE65_CG_PHE65_CD2_PHE65	0.58	0.81	-0.74	0.33	0.47	0.38
CB_SER67_CA_SER67_C_SER67_O_SER67	0.59	0.64	-0.91	0.61	0.3	0.61
CB_SER67_CA_SER67_C_SER67_N_LYS68	-0.51	-0.4	0.9	-0.75	-0.11	-0.5
N_SER67_CA_SER67_C_SER67_O_SER67	0.62	0.64	-0.92	0.61	0.32	0.63
N_SER67_CA_SER67_C_SER67_N_LYS68	0.6	0.63	-0.91	0.61	0.31	0.62
C_SER67_N_LYS68_CA_LYS68_CB_LYS68	-0.55	-0.5	0.91	-0.71	-0.19	-0.57
C_SER67_N_LYS68_CA_LYS68_C_LYS68	-0.55	-0.48	0.91	-0.75	-0.18	-0.57
CA_LYS68_C_LYS68_N_ALA69_CA_ALA69	-0.84	-0.77	0.83	-0.53	-0.65	-0.81
O_LYS68_C_LYS68_N_ALA69_CA_ALA69	0.8	0.6	-0.91	0.72	0.44	0.61
CB_TRP91_CA_TRP91_C_TRP91_N_GLU92	-0.78	-0.73	0.37	0.12	-0.87	-0.52
CD2_TYR118_CE2_TYR118_CZ_TYR118_OH_TYR118	-0.4	0.65	-0.42	0.01	0.88	-0.25
C_THR106_N_GLN107_CA_GLN107_C_GLN107	-0.86	-0.64	0.81	-0.27	-0.66	-0.56
N_VAL120_CA_VAL120_CB_VAL120_CG1_VAL120	0.76	0.47	-0.57	0.42	0.58	0.88
C_SER144_N_THR145_CA_THR145_CB_THR145	0.67	0.6	-0.25	-0.38	0.84	0.56
C_GLU174_N_VAL175_CA_VAL175_CB_VAL175	0.67	-0.28	-0.01	0.1	-0.72	0.93
C_GLU174_N_VAL175_CA_VAL175_C_VAL175	0.73	-0.3	0.17	0.07	-0.83	0.76
CA_VAL175_C_VAL175_N_LYS176_CA_LYS176	-0.74	0.28	0.44	-0.53	0.46	-0.88
C_THR178_N_TYR179_CA_TYR179_CB_TYR179	0.87	-0.3	-0.34	0.51	-0.6	0.87
CE1_TYR179_CD1_TYR179_CG_TYR179_CD2_TYR179	0.65	-0.17	-0.26	0.24	-0.48	1
CA_LEU197_C_LEU197_N_ASP198_CA_ASP198	0.65	0.81	-0.89	0.47	0.47	0.64
CA_ASP198_CB_ASP198_CG_ASP198_OD1_ASP198	-0.64	-0.29	0.83	-0.72	-0.23	-0.51
C_TYR212_N_GLU213_CA_GLU213_CB_GLU213	0.23	0.83	-0.55	0.33	0.26	0.46
CA_ARG214_CB_ARG214_CG_ARG214_CD_ARG214	0.55	-0.54	-0.7	0.86	-0.22	0.34

Note: Torsion angles for which correlation coefficient module with at least one torsion angle in the chromophore (which, in turn, correlate with spectral properties) is not less than 0.8 are presented. Torsion angles in chromophore:

- 1) C1\_CRQ66\_N3\_CRQ66\_C2\_CRQ66\_CA2\_CRQ66;
- 2) CA2\_CRQ66\_N2\_CRQ66\_C1\_CRQ66\_N3\_CRQ66;
- 3) O\_CRQ66\_C\_CRQ66\_N\_SER67\_CA\_SER67;
- 4) CA3\_CRQ66\_C\_CRQ66\_N\_SER67\_CA\_SER67;
- 5) CA2\_CRQ66\_CB2\_CRQ66\_CG2\_CRQ66\_CD2\_CRQ66;
- 6) CG2\_CRQ66\_CD2\_CRQ66\_CE2\_CRQ66\_CZ\_CRQ66.

43 amino acid residues (total number of residues in the 6 Å environment of the chromophore is 72, and 582 torsion angles are related to them). The search for amino acid residues with torsion angles in the chromophore environment highly and/or multiply correlating with spectral properties is necessary for directed mutagenesis, because mutations at these positions might significantly effect chromophore geometry and, therefore, spectral properties of the proteins.

Consideration of the average structures calculated by MD simulations shows that a residue with a polar side chain at position 66 in the chromophore causes formation

of a hydrogen bond between the chromophore and the side chain of Gln211 (Fig. 3). Hydrogen bond formation requires mutual approach of corresponding atoms, and through this approach the less rigid structure will be deformed to a higher extent than a more rigid one, which is energetically more favorable. That is why this hydrogen bond is probably formed not at the cost of change in position of Gln211, a constituent of rigid  $\beta$ -barrel, but due to change in the chromophore conformation.

MD simulations showed that torsion angles in the chromophore of mutant proteins with polar side chain at position 66 really change. Therefore, modifications in

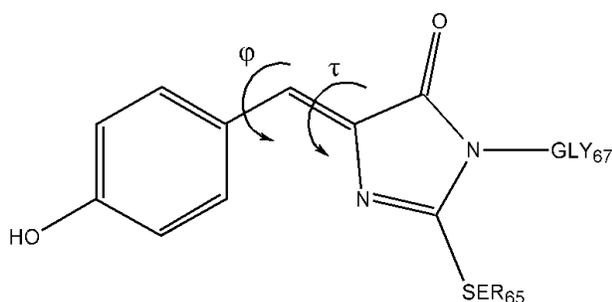


Fig. 8. Structure of GFP chromophore (4-(*p*-hydroxybenzylidene)-imidazolid-5-one).  $\tau$  and  $\phi$ , torsion angles.

chromophore geometry as a result of formation of a new hydrogen bond might be the cause of changes in spectral properties of corresponding proteins, which is supported by correlations (Table 1). This interrelation is also supported by correlation of torsion angle CA\_GLN211\_C\_GLN211\_N\_TYR212\_CA\_TYR212 in the main chain of Gln211 residue (6 Å environment of the chromophore) with absorption, excitation, and emission wavelength maxima (0.74, 0.65, and 0.78 for all mutants, and 0.85, 0.86 and 0.80 for proteins with polar amino acid residues, respectively).

*C. Correlations between the volumes of substituting amino acid residues and spectral properties.* As shown for all mutants, fluorescence excitation and emission wavelength maxima correlate with the residue volume at position 66 (correlation coefficients 0.67 and 0.75, respectively). However, for proteins with polar amino acid residues the corresponding correlation coefficients rise to 0.99 and 0.93. This effect might be rationalized as follows: dependent on the side chain type of residue 66, hydrogen bonds between polar amino acids and Gln211 might be formed; this is indirectly supported in the case of aliphatic replacements by alanine and leucine at residue 66: in these cases hydrogen bonds cannot be formed, and the absorption and excitation wavelength maxima are practically equal in spite of different volume of the replacing residue. For mutants with polar substitutions at residue 66, such hydrogen bonds are probably formed and induce changes in geometry of the system of conjugated  $\pi$ -bonds, which may induce changes in spectral properties depending on the volume of the substituting amino acid residue. For green fluorescent protein GFP, we also found sensitivity of spectral properties of the fluorescent protein to the volume of the substituting amino acid residue (Fig. 7b).

In conclusion, spectral properties of fluorescent protein mRFP1 depend on chromophore geometry, which, in turn, is highly sensitive to hydrogen bonding between side chains of residue 66 in the chromophore and Gln211 in the chromophore environment. That is why position 211 in the gene of fluorescent protein mRFP1 is the most promising for direct mutagenesis.

This work was financially supported by the Interdisciplinary Scientific project of Lomonosov Moscow State University (No. 2/07, 2007).

## REFERENCES

1. Tsien, R. Y. (1998) *Annu. Rev. Biochem.*, **67**, 509-544.
2. Verkhusha, V. V., Akovbian, N. A., Efremenko, E. N., Varfolomeyev, S. D., and Vrzheschch, P. V. (2001) *Biochemistry (Moscow)*, **66**, 1342-1351.
3. Vrzheschch, P. V., Akovbian, N. A., Varfolomeyev, S. D., and Verkhusha, V. V. (2000) *FEBS Lett.*, **29**, 203-208.
4. Miyawaki, A., Llopis, J., Heim, R., McCaffery, J. M., Adams, J. A., Ikura, M., and Tsien, R. Y. (1997) *Nature*, **28**, 882-887.
5. Romoser, V. A., Hinkle, P. M., and Persechini, A. (1997) *J. Biol. Chem.*, **16**, 13270-13274.
6. Terskikh, A., Fradkov, A., Ermakova, G., Zaraisky, A., Tan, P., Kajava, A. V., Zhao, X., Lukyanov, S., Matz, M., Kim, S., Weissman, I., and Siebert, P. (2000) *Science*, **290**, 1585-1588.
7. Shaner, N. C., Campbell, R. E., Steinbach, P. A., Giepmans, B. N., Palmer, A. E., and Tsien, R. Y. (2004) *Nat. Biotechnol.*, **22**, 1567-1572.
8. Jung, G., Wiehler, J., and Zumbusch, A. (2005) *Biophys. J.*, **88**, 1932-1947.
9. Weber, W., Helms, V., McCammon, J. A., and Langhoff, P. W. (1999) *Proc. Natl. Acad. Sci. USA*, **25**, 6177-6182.
10. Reuter, N., Lin, H., and Thiel, W. (2002) *J. Phys. Chem. B*, **106**, 6310-6321.
11. Heim, R., Cubitt, A., and Tsien, R. Y. (1995) *Nature*, **23**, 663-664.
12. Shimomura, O., Johnson, F. H., and Saiga, Y. (1962) *J. Cell. Comp. Physiol.*, **59**, 223-239.
13. Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G., and Cormier, M. J. (1992) *Gene*, **111**, 229-233.
14. Matz, M. V., Fradkov, A. F., Labas, Y. A., Savitsky, A. P., Zaraisky, A. G., Markelov, M. L., and Lukyanov, S. A. (1999) *Nature Biotech.*, **17**, 969-973.
15. Zhang, J., Campbell, R. E., Ting, A. Y., and Tsien, R. Y. (2002) *Nature Rev. Mol. Cell Biol.*, **3**, 906-918.
16. Zubova, N. N., Bulavina, A. Yu., and Savitsky, A. P. (2003) *Uspekhi Biol. Khim.*, **43**, 163-224.
17. Lindahl, E., Hess, B., and van der Spoel, D. (2001) *J. Mol. Model.*, **7**, 306-317.
18. Van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. C. (2005) *J. Comput. Chem.*, **26**, 1701-1718.
19. Dmitrienko, D. V., Vrzheschch, E. P., Druitsa, V. L., and Vrzheschch, P. V. (2006) *Biochemistry (Moscow)*, **71**, 1133-1152.
20. Van der Spoel, D., Lindahl, E., Hess, B., van Buuren, A. R., Apol, E., Meulenhoff, P. J., Tieleman, D. P., Sijbers, A. L. T. M., Feenstra, K. A., van Drunen, R., and Berendsen, H. J. C. (2004) [www.gromacs.org](http://www.gromacs.org), Gromacs User Manual version 3.2.
21. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) *J. Mol. Biol.*, **112**, 535-542.

22. Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., Feng, Z., Gilliland, G. L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J. D., and Zardecki, C. (2002) *Acta Crystallogr. D*, **58**, 899-907.
23. Tubbs, J. L., Tainer, J. A., and Getzoff, E. D. (2005) *Biochemistry*, **44**, 9833-9840.
24. Guex, N., and Peitsch, M. C. (1997) *Electrophoresis*, **18**, 2714-2723.
25. Berendsen, H. J. C., van der Spoel, D., and van Drunen, R. (1995) *Comp. Phys. Comm.*, **91**, 43-56.
26. Berendsen, H. J. C., Postma, J. P. M., DiNola, A., and Haak, J. R. (1984) *J. Chem. Phys.*, **81**, 3684-3690.
27. Drutsa, V. L., Kaberdin, V. R., Koroleva, O. N., and Shilov, I. A. (1991) *Bioorg. Khim.*, **17**, 1487-1493.
28. Fletcher, A. N. (1969) *Photochem. Photobiol.*, **9**, 439-444.
29. Kubin, R. F., and Fletcher, A. N. (1982) *J. Luminescence*, **27**, 455-462.
30. Zamyatin, A. A. (1972) *Progr. Biophys. Mol. Biol.*, **24**, 107-123.
31. Chen, M. C., Lambert, C. R., Urgitis, J. D., and Zimmer, M. (2001) *Chem. Phys.*, **270**, 157-164.