Engineering of soluble bacteriorhodopsin

Andrey Nikolaev[1], Yaroslav Orlov[1], Fedor Tsybrov[1], Elizaveta Kuznetsova[1], Pavel Shishkin[1], Alexander Kuzmin[1], Anatolii Mikhailov[1], Yulia S. Nikolaeva[1], Arina Anuchina[1], Igor Chizhov[1], Oleg Semenov[1], Ivan Kapranov[1], Valentin Borshchevskiy[1], Alina Remeeva[1], Ivan Gushchin[1]

[1] Research Center for Molecular Mechanisms of Aging and Age-Related Diseases, Moscow Institute of Physics and Technology, Dolgoprudny, Russia

* E-mail for correspondence: ivan.gushchin@phystech.edu

Supplementary Information

Data S1-S2

Tables S1-S3

Figures S1-S5

Data S1. Amino acid sequences of engineered soluble BR variants. Structural protein amino acids are shown as capital letters and extra residues (flexible linkers and hexahistidine tags) are shown as small letters. The hexahistidine tags used for metal-affinity purification are underlined.

>NeuroBR_A
MSMRPESKYYEKATEEMEKAYEEFKKKGEGVEDEEAKKFYDLLTEVPRIAYEQYKKILDGEGIEKVEVDGKEIEV
PVARYEDWEKTTPLLLEVLANLVDASEELKEKLISKAKEMITLGKKGALETDPEKRFEYWKKSTEKMNEIIDLLE
NGFKENLSSLKPERKKTYEEARKLTIELWSKYPEIWKKGPLGEGKVPLEETVKQFTELDVSAKVGFGELVLSSEA
IYSgsghhhhhh

>NeuroBR_B
MSSRPESEAYAAATKEMTDAAAKFKEKGKGEKDPEAQKFYELLTKVPAIAAESYQALLDGSGLVRVRVDGKDVEV
PVARYDDWAVTTPLLLEVLALLVNASEELKQKLLSLAQQMIDLGRQGALETDPEKRFEYWDKSTAAMNEIFDLLE
NGFEENLSSLKPERLETYNKLRKMTLELWSQYPEIWRRGPLGKGEVPLAETAARFRELDVSAKVGFGEIVTSSKA
IYSgsghhhhhh

>NeuroBR_C
MADRPEAAALAAATAAMTAAAAAFAAKGAGETDPEAQKFYELATKVPAIAAASYQAMLDGSGIVLVEVDGKEVEV
YVARYDDWKVTTPLLLEILALLVEAKEEVKKELIELANKMIDLGEKGALETEPEKRFEYWEKSTEYMNKIIDILK
NGFEENLESLEPERKETFEKLKEMTIKLWSQYPKIWREGTLGEGKVSLEEEVARFAELDVSAKVGFGDLLLSSKA
IYSgsghhhhhh

Data S2. Nucleotide sequences of BR variants analyzed in this work. Nuclease cleavage sites are underlined. Sequences were designed for cloning into pET-28a(+) plasmid via XbaI (TCTAGA) and BamHI (GGATCC) restriction sites.

>NeuroBR_A

TCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGAGTATGCGTCCTGAGAGCAAATATTACGAAAAAGCAAC
CGAAGAGATGGAAAAAGCCTATGAGGAGTTCAAGAAGAAGGGCGAGGGTGTCGAAGATGAAGAGGCCAAAAAGTTTTATGATC
TGCTGACCGAAGTTCCGCGTATTGCCTATGAACAGTATAAGAAAATTCTGGATGGCGAGGGCATTGAAAAAGTTGAAGTTGAT
GGCAAAGAAATCGAAGTGCCGGTTGCACGTTATGAAGATTGGGAGAAAACCACACCGCTGCTGCTGGAAGTTCTGGCAAATCT
GGTTGATGCAAGCGAAGAACTGAAAGAGAAACTGATTAGCAAAGCCAAAGAGATGATTACTTTAGGTAAAAAGGGTGCACTGG
AAACCGATCCTGAAAAACGCTTTGAATATTGGAAGAAAAGCACCGAGAAAATGAACGAGATTATTGACTTGCTGGAAAACGGC
TTTAAAGAAAATCTGAGCAGCTTGAAGCCCGAACGTAAGAAAACGTATGAAGAAGCACGTAAACTGACCATTGAACTGTGGTC
AAAATACCCGGAAATTTGGAAAAAGGGTCCGTTAGGTGAAGGCAAAGTTCCGCTGGAAGAAACCGTTAAACAGTTTACCGAAC
TGGATGTTAGCGCCAAAGTTGGTTTTGGTGAACTGGTCCTGAGCAGTGAAGCAATTTATAGCGGTAGCGGTCATCATCACCAT
CATCATTAAGGATCC

>NeuroBR_B

TCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGAGCAGCCGTCCTGAAAGCGAAGCCTATGCAGCAGCAAC
CAAAGAAATGACCGATGCAGCCGCTAAATTCAAGGAAAAAGGTAAAGGCGAAAAAGACCCGGAAGCACAGAAATTCTATGAAC
TGCTGACCAAAGTTCCGGCTATTGCAGCAGAAAGCTATCAGGCACTGCTGGATGGTAGCGGTCTGGTTCGTGTTCGTGTTGAT
GGTAAAGATGTTGAAGTTCCGGTTGCACGTTATGATGATTGGGCAGTTACCACACCGCTGCTGCTGGAAGTTCTGGCCCTGCT
GGTTAATGCAAGCGAAGAACTGAAACAGAAACTGCTGAGTCTGGCACAGCAGATGATTGATCTGGGTCGTCAGGGTGCACTGG
AAACCGATCCTGAAAAACGCTTTGAATATTGGGATAAAAGCACCGCAGCCATGAACGAAATTTTTGATCTGCTGGAAAATGGC
TTCGAAGAAAATCTGAGCAGCCTGAAACCGGAACGTCTGGAAACCTATAACAAACTGCGTAAAATGACCCTGGAACTGTGGTC
CCAGTATCCTGAAATTTGGCGTCGTGGTCCGTTAGGTAAAGGTGAAGTGCCGCTGGCAGAAACCGCAGCACGTTTTCGTGAAC
TGGATGTTAGCGCAAAAGTTGGTTTTGGTGAAATTGTTACCAGCAGCAAAGCAATTTATAGCGGTAGTGGTCATCATCATCAC
CATCACTAAGGATCC

>NeuroBR_C

TCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGGCAGATCGTCCTGAAGCTGCGGCACTGGCCGCGGCAAC
CGCCGCAATGACTGCTGCTGCAGCAGCATTTGCAGCCAAAGGTGCCGGTGAAACCGATCCTGAAGCTCAGAAATTCTATGAAC
TGGCAACCAAAGTTCCGGCTATTGCAGCAGCCAGCTATCAGGCAATGCTGGATGGTAGCGGTATTGTTCTGGTTGAAGTTGAT
GGTAAAGAGGTCGAAGTTTACGTTGCACGTTATGATGATTGGAAAGTTACCACACCGCTGCTGCTGGAAATTCTGGCACTGCT
GGTAGAAGCCAAAGAAGAAGTTAAGAAAGAACTGATTGAGCTGGCCAACAAAATGATTGATCTGGGTGAAAAAGGTGCACTGG
AAACCGAACCGGAAAAACGCTTTGAATATTGGGAAAAATCCACCGAGTACATGAACAAGATCATCGACATTCTGAAAAACGGC
TTCGAAGAAAATCTGGAAAGCCTGGAGCCCGAACGTAAGGAAACCTTTGAAAAACTGAAAGAAATGACCATCAAGCTGTGGTC
ACAGTATCCGAAAATTTGGCGTGAAGGTACATTAGGTGAAGGTAAAGTTAGCCTGGAAGAAGAGGTTGCCCGTTTTGCAGAAC
TGGATGTTAGCGCAAAAGTTGGTTTTGGTGATCTGCTGCTGAGTAGCAAAGCAATTTATAGCGGTAGCGGTCATCATCACCAT
CATCATTAAGGATCC

Table S1. Parameters of BR variants studied in this work.

	pLDDT	AlphaFold model RMSD to 7Z09 (Å)	Number of residues (w/o tag)	Molecular weight (with tag, kDa)	Molecular weight (w/o tag)	pI without tag	ε_{280} (M^{-1} cm^{-1})	ε_{280} (mg/mL cm)
Wild type	97.0	0.90	248	-	26.8	-	-	-
NeuroBR_A	96.7	0.62	227	27.5	26.5	5.15	38390	0.715
NeuroBR_B	96.5	0.63	227	26.6	25.6	5.12	33920	0.783
NeuroBR_C	96.3	0.60	227	26.3	25.2	4.67	33920	0.774

Table S2: Diffraction anisotropy information.

Diffraction limits (Å) and corresponding principal axes of the ellipsoid fitted to the diffraction cut-off surface as direction cosines in the orthogonal basis (standard PDB convention), and in terms of reciprocal unit-cell vectors.

	Diffraction limits (Å)	Principal axes in the orthogonal basis	Principal axes reciprocal unit-cell vectors
Diffraction limit #1	1.755	(1.0000, 0.0000, 0.0000)	a*
Diffraction limit #2	1.778	(0.0000, 1.0000, 0.0000)	b*
Diffraction limit #3	2.212	(0.0000, 0.0000, 1.0000)	c*

Eigenvalues of the overall anisotropy tensor on |F|s ($Å^2$), the same eigenvalues after subtraction of the smallest eigenvalue (as used in the anisotropy correction), and corresponding eigenvectors of the overall anisotropy tensor as direction cosines in the orthogonal basis (standard PDB convention), and in terms of reciprocal unit-cell vectors:

	Eigenvalues ($Å^2$)	Eigenvalues used in the anisotropy correction ($Å^2$)	Eigenvectors in the orthogonal basis	Eigenvectors in reciprocal unit-cell vectors
Eigenvalue #1	13.24	0.00	(1.0000, 0.0000, 0.0000)	a*
Eigenvalue #2	16.92	3.68	(0.0000, 1.0000, 0.0000)	b*
Eigenvalue #3	33.57	20.33	(0.0000, 0.0000, 1.0000)	c*

Table S3. Crystallographic data collection and refinement statistics.

PDB ID	9KME
Wavelength, Å	0.979
Resolution range*, Å	36.17 - 1.76 (1.88 - 1.76)
Space group	P 21 21 21
Unit cell, Å °	95.03 96.09 109.93 90 90 90
Total reflections	980346 (39576)
Unique reflections	76128 (3849)
Multiplicity	12.9 (10.3)
Completeness (spherical), %	76.7 (22.2)
Completeness (ellipsoidal), %	95.8 (70.2)
Mean I/σ(I)	15.2 (2.5)
Wilson B-factor*, Å^2	16.5
R-merge	0.127 (0.968)
R-meas	0.132 (1.019)
R-pim	0.036 (0.310)
CC1/2	0.999 (0.655)
Resolution range used in the refinement, Å	36.17 - 1.76, 1.78, 2.21
Reflections used in refinement	70884
Reflections used for R-free	236
R-work	0.1771
R-free	0.2125
Number of non-hydrogen atoms	7944
macromolecules	7423
ligands	20
solvent	501
Protein residues	903
RMS_{bonds}, Å	0.008
RMS_{angles}, °	1.18
Ramachandran favored, %	99
Ramachandran allowed, %	1
Ramachandran outliers, %	0
Rotamer outliers, %	1
Clashscore	11
Average B-factor, Å^2	22.8
macromolecules	22.4
ligands	45.6
solvent	28.3
Number of TLS groups	4

Statistics for the highest-resolution shell are shown in parentheses.

* - anisotropy information is presented in Table S2.

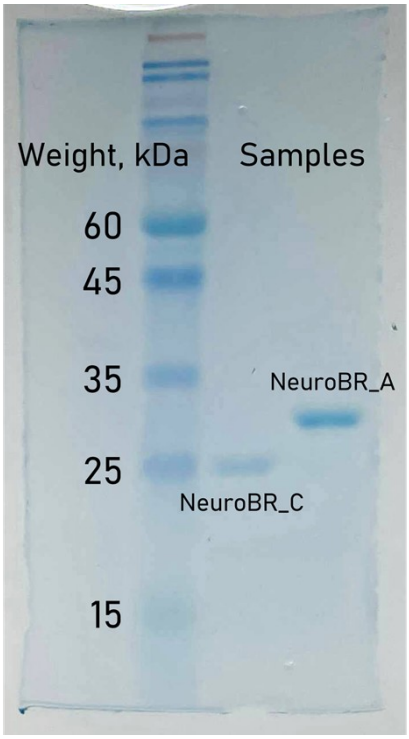

Figure S1. SDS-PAGE of NeuroBR_A and C samples purified as described in the Methods section.

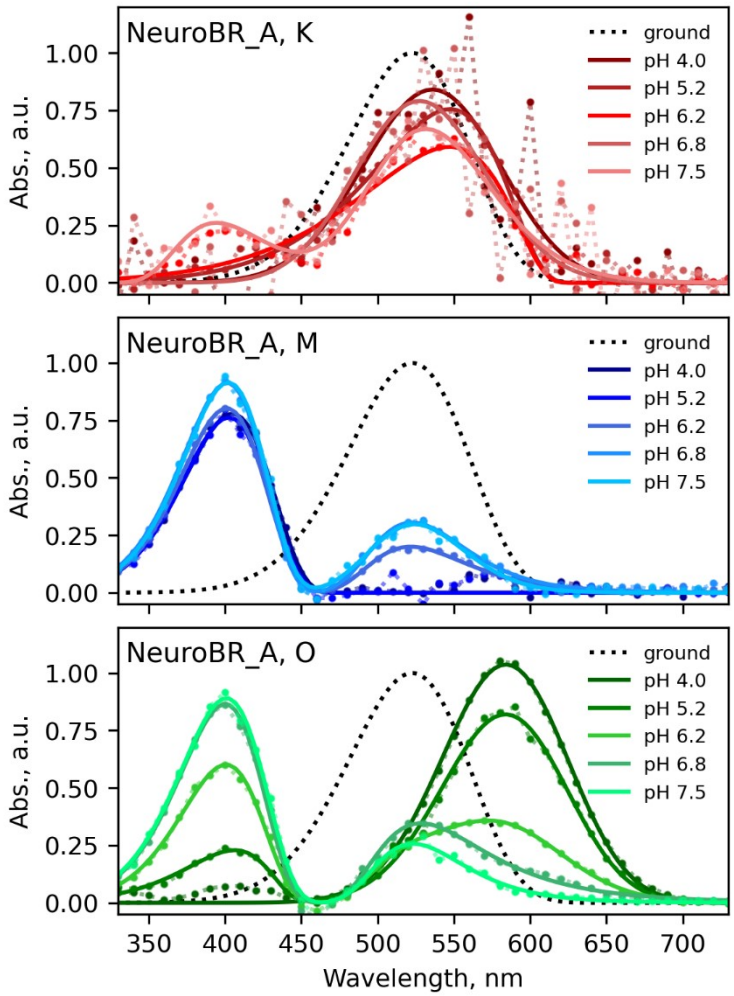

Figure S2. Recovered absorption spectra of NeuroBR_A intermediates at different buffer pH values. The intermediates are designated in accordance with the photocycle at low pH.

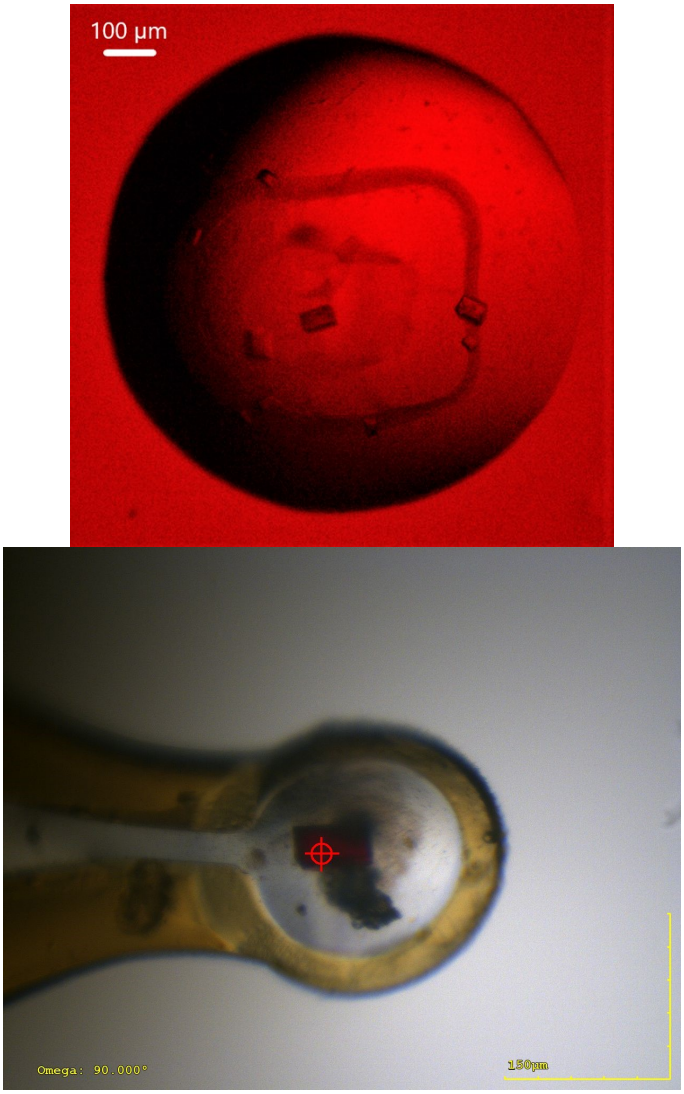

Figure S3. NeuroBR_A crystals. (left) Crystals in the crystallization drop observed via a microscope under red light illumination. (right) Mounted crystal imaged during diffraction data collection at SSRF.

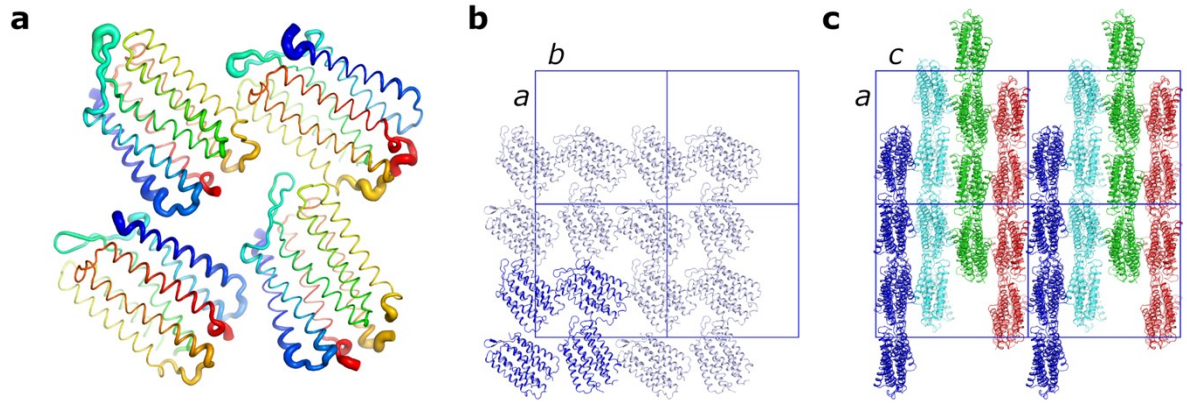

Figure S4. Crystal packing of NeuroBR_A. a, Relative distributions of the B-factor values for the four NeuroBR molecules present in the asymmetric unit cell (ASU). Coloring changes from blue to red from the N-terminus to the C-terminus. **b**, Packing in the *a*-*b* plane. ASU contents of one cell is shown in blue and contents of the adjacent cells is shown in light blue. *a* = 95.03 Å, *b* = 96.09 Å. **c**, Packing in the *a*-*c* plane. Each cell contains 16 copies of NeuroBR, organized as four parallel layers of four NeuroBR molecules. Crystallographic symmetry-related molecules are colored blue, light blue, green and red. *a* = 95.03 Å, *c* = 109.93 Å.

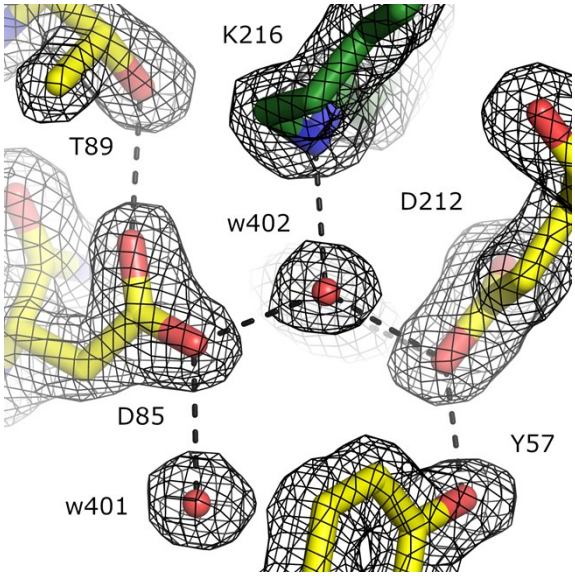

Figure S5. Electron density maps in the retinal Schiff base region of NeuroBR_A.
The weighted $2F_o$-F_c electron density map is contoured at the level of 2 σ.