

# Directed Evolution of Near-Infrared Serotonin Nanosensors with Machine Learning-Based Screening

Seonghyeon An <sup>1,†</sup>, Yeongjoo Suh <sup>1,†</sup>, Payam Kelich <sup>2</sup>, Dakyeon Lee <sup>1,3</sup>, Lela Vukovic <sup>2</sup> and Sanghwa Jeong <sup>1,\*</sup>

<sup>1</sup> Department of Biomedical Convergence Engineering, Pusan National University, Yangsan 50612, Republic of Korea  
<sup>2</sup> Department of Chemistry and Biochemistry, University of Texas at El Paso, El Paso, TX 79968, USA  
<sup>3</sup> Department of Chemistry, Pohang University of Science and Technology (POSTECH), Pohang 37673, Republic of Korea  
\* Correspondence: sanghwa.jeong@pusan.ac.kr; Tel.: +82-51-510-8539  
† These authors contributed equally to this work.

a.

Sequence ID	DNA sequence (from 5' to 3')	Predicted $\Delta F/F_0$	Sequence ID	DNA sequence (from 5' to 3')	Predicted $\Delta F/F_0$
1	CCCCCAGCCCCAACACAACCAACTCCCCC	1.533	11	CCCCCAGCCCCAACACCACCAAGCTCCCCC	1.456
2	CCCCCAGCCCTACACAACCAAGCTCCCCC	1.491	12	CCCCCAGCCCCGACACAACCAACTCCCCC	1.451
3	CCCCCAGCCCTACACAACCAACTCCCCC	1.480	13	CCCCCAGCCCTACACAACCAACTCCCCC	1.449
4	CCCCCAGCCCTACACAACCAAGCCCCC	1.479	14	CCCCCAGCAGTACACAACCAACTCCCCC	1.448
5	CCCCCAGCCCTACACAATCCAAGTCCCCC	1.469	15	CCCCCAGCCCATCACAACCAAGCCCCC	1.447
6	CCCCCAGCCCTACCCAACCAACTCCCCC	1.468	16	CCCCCAGCCCTAACACAACCAACTCCCCC	1.444
7	CCCCCAACCCCTACACAACCAACTCCCCC	1.467	17	CCCCCAGCCCATCACAACCAAGCTCCCCC	1.442
8	CCCCCAGGCCCTACACAACCAACTCCCCC	1.466	18	CCCCCAGCCCTACACAACAACCAACTCCCCC	1.441
9	CCCCCAGCCCCAACACCACCACTCCCCC	1.463	19	CCCCCAGCCCCAACACGACCAACTCCCCC	1.440
10	CCCCCAGCCCTAGACAACCAACTCCCCC	1.458	20	CCCCCAGCCCTACACAACCAACCCCCC	1.439

b.

Sequence ID	DNA sequence (from 5' to 3')	Predicted $\Delta F/F_0$	Sequence ID	DNA sequence (from 5' to 3')	Predicted $\Delta F/F_0$
1	CCCCCAGCCCCAACACAACCAACTCCCCC	1.533	11	CCCCCAGCCCCAACACCACCAAGCTCCCCC	1.456
2	CCCCCAGCCCTACACAACCAAGCTCCCCC	1.491	12	CCCCCAGCCCCGACACAACCAACTCCCCC	1.451
3	CCCCCAGCCCTACACAACCAACTCCCCC	1.480	13	CCCCCAGCCCTACACAACCAACTCCCCC	1.449
4	CCCCCAGCCCTACACAACCAAGCCCCC	1.479	14	CCCCCAGCAGTACACAACCAACTCCCCC	1.448
5	CCCCCAGCCCTACACAATCCAAGTCCCCC	1.469	15	CCCCCAGCCCATCACAACCAAGCCCCC	1.447
6	CCCCCAGCCCTACCCAACCAACTCCCCC	1.468	16	CCCCCAGCCCTAACACAACCAACTCCCCC	1.444
7	CCCCCAACCCCTACACAACCAACTCCCCC	1.467	17	CCCCCAGCCCATCACAACCAAGCTCCCCC	1.442
8	CCCCCAGGCCCTACACAACCAACTCCCCC	1.466	18	CCCCCAGCCCTACACAACAACCAACTCCCCC	1.441
9	CCCCCAGCCCCAACACCACCACTCCCCC	1.463	19	CCCCCAGCCCCAACACGACCAACTCCCCC	1.440
10	CCCCCAGCCCTAGACAACCAACTCCCCC	1.458	20	CCCCCAGCCCTACACAACCAACCCCCC	1.439

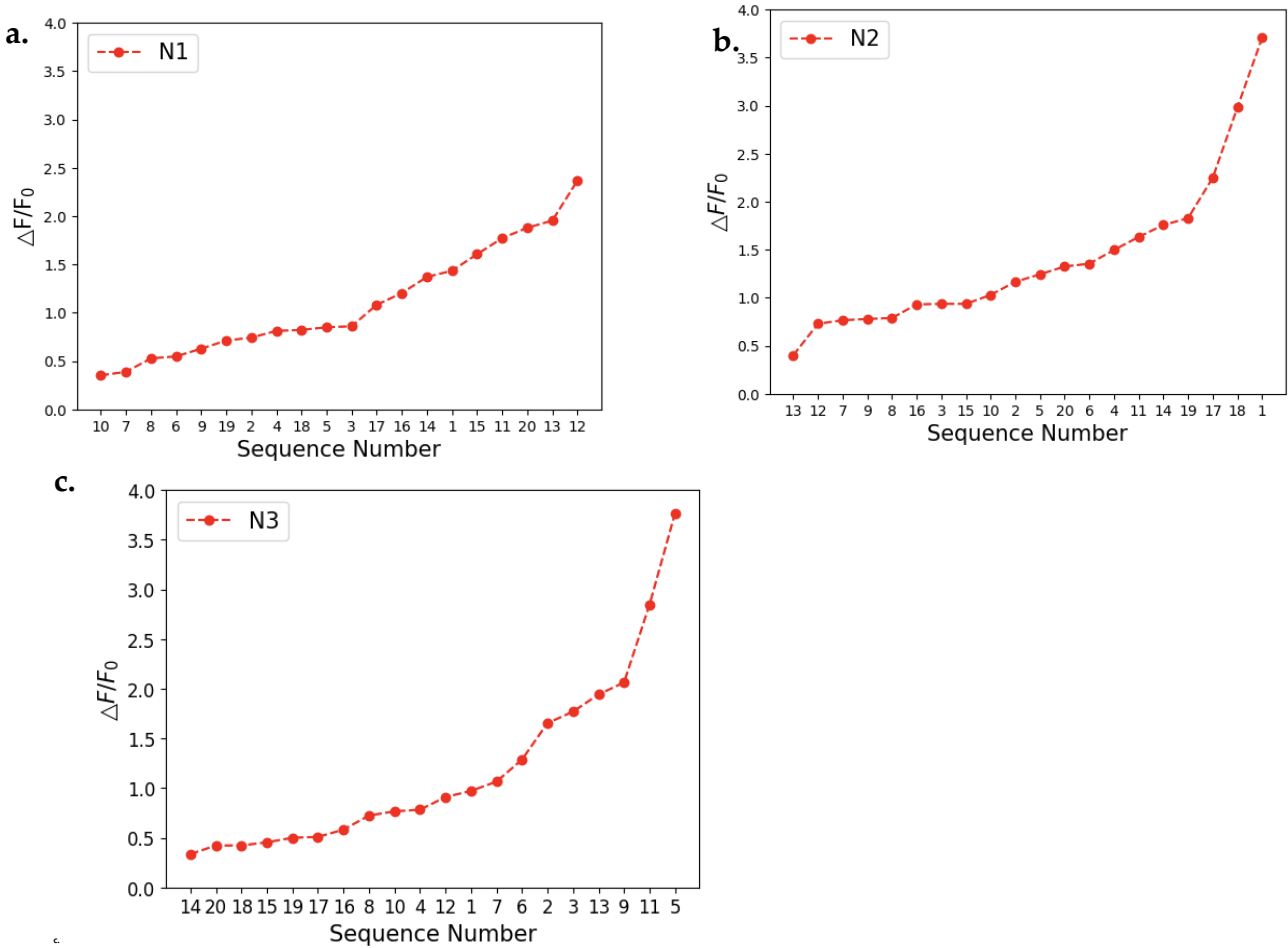
c.

Sequence ID	DNA sequence (from 5' to 3')	Predicted $\Delta F/F_0$	Sequence ID	DNA sequence (from 5' to 3')	Predicted $\Delta F/F_0$
1	CCCCCAGCCCCAACACAACCAACTCCCCC	1.533	11	CCCCCAGCCCCAACACCACCAAGCTCCCCC	1.456
2	CCCCCAGCCCTACACAACCAAGCTCCCCC	1.491	12	CCCCCAGCCCCGACACAACCAACTCCCCC	1.451
3	CCCCCAGCCCTACACAACCAACTCCCCC	1.480	13	CCCCCAGCCCTACACAACCAACTCCCCC	1.449
4	CCCCCAGCCCTACACAACCAAGCCCCC	1.479	14	CCCCCAGCAGTACACAACCAACTCCCCC	1.448
5	CCCCCAGCCCTACACAATCCAAGTCCCCC	1.469	15	CCCCCAGCCCATCACAACCAAGCCCCC	1.447
6	CCCCCAGCCCTACCCAACCAACTCCCCC	1.468	16	CCCCCAGCCCTAACACAACCAACTCCCCC	1.444
7	CCCCCAACCCCTACACAACCAACTCCCCC	1.467	17	CCCCCAGCCCATCACAACCAAGCTCCCCC	1.442
8	CCCCCAGGCCCTACACAACCAACTCCCCC	1.466	18	CCCCCAGCCCTACACAACAACCAACTCCCCC	1.441
9	CCCCCAGCCCCAACACCACCACTCCCCC	1.463	19	CCCCCAGCCCCAACACGACCAACTCCCCC	1.440
10	CCCCCAGCCCTAGACAACCAACTCCCCC	1.458	20	CCCCCAGCCCTACACAACCAACCCCCC	1.439

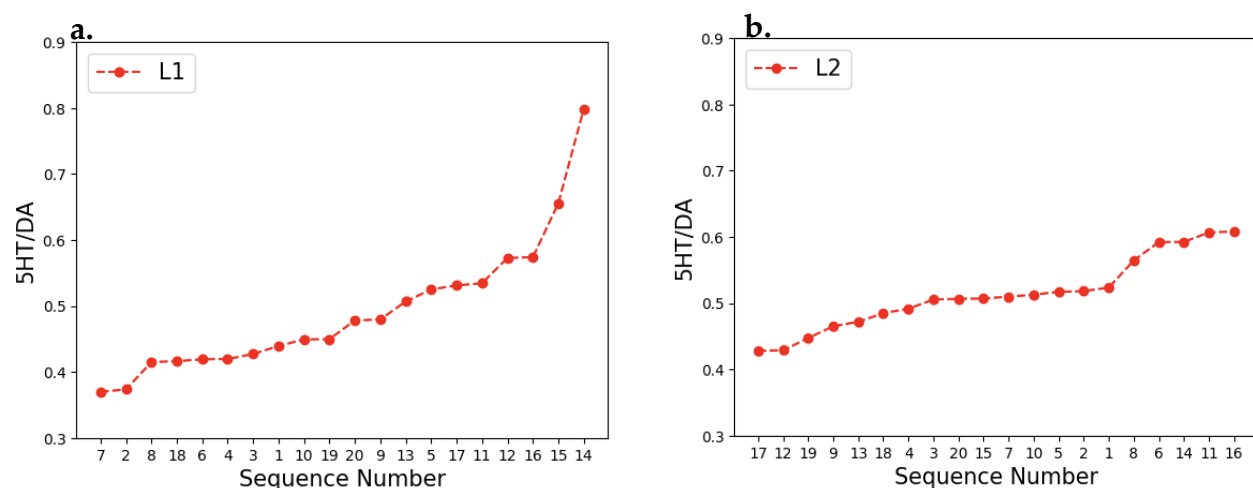
d.

Sequence ID	DNA sequence (from 5' to 3')	Predicted $\Delta F/F_0$	Sequence ID	DNA sequence (from 5' to 3')	Predicted $\Delta F/F_0$
1	CCCCCAACACAAACAACCACTCCCCC	1.498	11	CCCCCAGCACTACCCAACCACTCCCCC	1.395
2	CCCCCAGCACAAACAACCACTCCCCC	1.482	12	CCCCCAACTACACAAGCACTCCCCC	1.394
3	CCCCCAGCACAAACAACCACTCCCCC	1.455	13	CCCCCAACTACACAACCACTCCCCC	1.394
4	CCCCCAACTACTCCAACCACTCCCCC	1.453	14	CCCCCGAAGCTACACAACCACTCCCCC	1.390
5	CCCCCAGCACAAACAAGCACTCCCCC	1.443	15	CCCCCAACTACACGACCACTCCCCC	1.387
6	CCCCCAACGCTACACAACCACTCCCCC	1.422	16	CCCCCAGCGTACACAACCACTCCCCC	1.381
7	CCCCCAACTACACAACCACTCCCCC	1.420	17	CCCCCGGCACTACACAACCACTCCCCC	1.367
8	CCCCCAGCACAAACTACCACTCCCCC	1.411	18	CCCCCAGCACTACACGACCACTCCCCC	1.366
9	CCCCCAGCACAAACCACTCCCCC	1.403	19	CCCCCAGCACAAACCACTCCCCC	1.362
10	CCCCCAACTACACAACCACTCCCCC	1.395	20	CCCCCAGGACTACACCACTCCCCC	1.361

**Figure S1.** The table shows the top 20 sequences in each round of machine learning and the predicted machine learning results. The predicted  $\Delta F/F_0$  is the predicted value of how well the sequence responds to serotonin at 1195 nm. For the original sequence in each round, the positions where 3 base mutations occurred and the results are coloured in red. **(a)** Table for round 1 of sensitivity, selectivity. **(b)** Table for round 2 of sensitivity. **(c)** Table for round 3 of sensitivity. **(d)** Table for round 2 of selectivity.



**Figure S2.** The  $\Delta F/F_0$  values of the top 20 sequences for each round in sensitivity, sorted in ascending order. **(a)** The best performing sequence in round N1 is N1-12, with a  $\Delta F/F_0$  of 2.366. **(b)** The best performing sequence N2-1 with a  $\Delta F/F_0$  of 3.705. The second best performing sequence, N2-18, has a  $\Delta F/F_0$  of 2.981, which is better than N1-12. **(c)** The best performing sequence N3-5 with a  $\Delta F/F_0$  of 3.765.



**Figure S3.** The 5HT/DA values of the top 20 sequences for each round in selectivity, sorted in ascending order. **(a)** The best performing sequence in the L1 round, L1-14, has a value of 0.799, while the second best performing sequence, L1-15, has a 5HT/DA value of 0.655. **(b)** The best performing sequence in the L2 round, L2-16, has a 5HT/DA value of 0.608. The 5HT/DA values of the remaining sequences are distributed between 0.4 and 0.6, and L2-16 has a value smaller than the 5HT/DA value of L1-15, which is 0.655, indicating that it was saturated in the L1 round.