**Additional file 1:**

**Analyzing** **the genetic characteristics of a tryptophan-overproducing *Escherichia coli***

Dongqin Ding1,2,3, Danyang Bai1,2,3, Jinlong Li1,4, Zhitao Mao5, Yaru Zhu1,2,3, Pi Liu1,4, Jianping Lin1,4,6, Hongwu Ma5 and Dawei Zhang1,2,3\*

\*Correspondence: [zhang\_dw@tib.cas.cn](mailto:zhang_dw@tib.cas.cn)

1 Tianjin Institutes of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin, 300308, P.R. China

2 Key Laboratories of Systems Microbial Biotechnology, Chinese Academy of Sciences, Tianjin, 300308, P. R. China

3 National Technology Innovation Center of Synthetic Biology, Tianjin, 300308, P. R. China

4 Biodesign Center, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin, 300308, P. R. China

5 Biodesign Center, Key Laboratory of Systems Microbial Biotechnology, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin, 300308, China

6 College of Pharmacy, Nankai University, Haihe Education Park, 38 Tongyan Road, Tianjin 300353, P. R. China

**Simulation Methodology**

All atom molecular dynamics simulations have been performed using AMBER18 molecular dynamics package1**.** The bonded and non-bonded description of the interactions between the various atoms has been described using the AMBER18 force fields which include the ff14SBforce field parameters with hydrogen mass repartitioning (HMR) method2. The ANTECHAMBER module and GAFF2 with AM1-BCC charges3 are used to obtain force field parameters for ligands. Initially, we performed a series of energy minimization steps to eliminate any bad contacts in the initially built structures. During the minimization, protein (@CA,O,N,C) were restrained with harmonic force constants 20 kcal/mol. The minimization step involves 5000 steps steepest descent followed by 5000 steps of conjugate gradient method. After the energy minimization, the system was slowly heated up to 300 K in 100 ps MD using 1 fs integration time step, while restraining the solute with 20 kcal/mol harmonic force constant. After this, we performed 5ns NPT equilibration of the structures with no harmonic restraints. Finally, 500 ns NVT production simulations were performed at 300 K with 4 fs integration time step. We have implemented periodic boundary condition across the system using a TIP3P water box4. We used Particle Mesh Ewald (PME) techniques integrated with AMBER package to account for the long range part of the electrostatic interactions5. During the dynamics, all the bonds involving hydrogen are restrained using the SHAKE algorithm6. Langevin thermostat with collision frequency of 1 ps-1 is used to maintain the constant temperature while the pressure was controlled by anisotropic Monte-Carlo barostat7. The accelerated GPU version of PMEMD8 was performed on NVIDIA GeForce Series cards. We have employed CPPTRAJ9functionality of AMBERTOOLs1 to perform various analyses on the equilibrium MD simulation trajectories. The images and graphics of the structures shown here were generated using the software PyMOL10.

**Fig. S1**



Fig. S1 Schematic diagram of the principle of the L-trp biosensor. L-trp biosensor is composed of *tnaC*, which encodes the leader sequence of the *tnaCAB* operon, fused upstream to the reporter GFPmut2. Rho factor can access the Rho-dependent terminator on *tnaCAB* and thus terminating its continued transcription, for the absence of L-trp resulting in the release of the TnaC leader peptide from the tRNA and the ribosome. Access by Rho is blocked, allowing the transcription of *tnaCAB* continues in the presence of L-trp. Based on this mechanism, *tnaAB* was replaced by GFPmut2 so that the biosensor is induced by L-trp, that means in the absence of L-trp, the GFPmut2 transcription is blocked, and in the presence of L-trp, transcription of GFPmut2 continues so that the expressed fluorescent protein can be detected.

**Fig. S2**



Fig. S2 Response of the biosensor towards external Ala-Trp characterized by multimode microplate reader. Experiments were conducted in triplicate and measurements are represented as the means ± s.d.

**Fig. S3**



Fig. S3 Fluorescence microscopy images of cells carrying the L-trp sensor plasmid fed with 2 mM Ala-Trp dipeptide (lower row) and cells without dipeptide (upper row). Experiments were conducted in triplicate and measurements are represented as the means ± s.d.

**Fig. S4**



Fig. S4 Extracellular concentration of L-trp produced by the four recombinant strains and fluorescence values of the four defined strains carrying the L-trp sensor plasmid. Experiments were conducted in triplicate and measurements are represented as the means ± s.d.

**Fig. S5**



Fig. S5 The lethality curve of strains mutagenized with atmospheric and room-temperature plasma (ARTP) for different times. Experiments were conducted in triplicate and measurements are represented as the means ± s.d.

**Fig. S6**



Fig. S6 The color of the fermentation broth of strains TPD5-RM3 and TPD5 after 38 h of fermentation. Experiments were conducted in triplicate and measurements are represented as the means ± s.d.

**Fig. S7**



Fig. S7 The cell state of strains TPD5 and TPD5-RM3 at different fermentation stages was observed by optical microscope. Experiments were conducted in triplicate and measurements are represented as the means ± s.d.

**Fig. S8**



Fig. S8 The correlation of gene expression level between strain TPD5-RM3 and strain TPD5.

**Table S1**—Strains and plasmids used in this study

|  |  |  |
| --- | --- | --- |
| **Strain or plasmid** | **Relevant characteristics** | **Sources of reference** |
| **Strains** |  |  |
| DH5α | F−∆*lac*U169(Ø80d *lac*Z∆M15) *sup*E44 *hsd*R17 *rec*A1 *gyr*A96 *end*A1 *thi*-1 *rel*A1 | Invitrogen |
| TPD1 | W3110 derivative, overexpressing *aroG* and *trpEDCBA*; | This work |
| TPD2 | TPD1 derivative, overexpressing *ppsA*, *trpR* gene knockout; | This work |
| TPD3 | L-trp overproducing strain KW. It was derived from *E. coli* W3110, which was initially subjected to multiple rounds of random mutagenesis; | (Chen et al., 2018) |
| TPD4 | TPD3 derivative, overexpressing *aroG* and *trpEDCBA* on plasmid PH5a, overexpressing *ppsA* on the genome, *trpR* gene knockout; | This work |
| TPD5 | Derived from TPD4 by random mutation; | This work |
| TPD5-RM1 | Derived from TPD5 by back mutation of acrR; | This work |
| TPD5-RM2 | Derived from TPD5 by back mutation of aroG; | This work |
| TPD5-RM3 | Derived from TPD5 by back mutation of rpoS; | This work |
| TPD5-RM4 | Derived from TPD5 by back mutation of puuP; | This work |
| TPD5-RM5 | Derived from TPD5 by back mutation of arsB; | This work |
| TPD5-RM6 | Derived from TPD5 by back mutation of ebgA; | This work |
| TPD5-RM7 | Derived from TPD5 by back mutation of glpQ; | This work |
| TPD5-RM8 | Derived from TPD5 by back mutation of dinD; | This work |
| TPD5-RM9 | Derived from TPD5 by back mutation of mlc; | This work |
| TPD5-RM10 | Derived from TPD5 by back mutation of trpE; | This work |
| TPD5-RM11 | Derived from TPD5 by back mutation of rpoB; | This work |
| TPD5-RM12 | Derived from TPD5 by back mutation of oppF; | This work |
| TPD5-RM13 | Derived from TPD5 by back mutation of entE; | This work |
| TPD5-RM14 | Derived from TPD5 by back mutation of gabD; | This work |
| TPD5-RM15 | Derived from TPD5 by back mutation of flhD; | This work |
| TPD5-RM16 | Derived from TPD5 by back mutation of argH; | This work |
| TPD5-RM17 | Derived from TPD5 by back mutation of ybdK; | This work |
| TPD5-RM18 | Derived from TPD5 by back mutation of garL; | This work |
| TPD5-RM19 | Derived from TPD5 by back mutation of acnA; | This work |
| TPD5-RM20 | Derived from TPD5 by back mutation of dicA; | This work |
| TPD5-RM21 | Derived from TPD5 by back mutation of oppA; | This work |
| TPD5-RM22 | Derived from TPD5 by back mutation of polA; | This work |
| TPD5-RM23 | Derived from TPD5 by back mutation of imp; | This work |
| TPD5-RM24 | Derived from TPD5 by back mutation of lacI; | This work |
| TPD5-RM25 | Derived from TPD5 by back mutation of ydeI; | This work |
| TPD5-RM26 | Derived from TPD5 by back mutation of ygiQ; | This work |
| TPD5-RM27 | Derived from TPD5 by back mutation of tnaA; | This work |
| TPD5-RM28 | Derived from TPD5 by back mutation of ylbE; | This work |
| **Plasmids** |  |  |
| PH5a | Plasmid used for gene overexpression, Tetr | Lab stock |
| PH5a-D1 | PH5a derivative, carrying *aroG* and *trpEDCBA* genes, Tetr | This work |
| CRISPR/Cas9 | Used for deletion or regulation of genes, Ampr | (Zhao et al., 2016) |
| Cas9-ppsA | Cas9 derivative containing promoter Trc, Ampr | This work |
| Cas9-trpR | Cas9 derivative for knocking out trpR, Ampr | This work |
| ptrc99a | Cloning vector, carries *bla* and Trc promoter, Ampr | Lab stock |
| Trp biosensor | ptrc99a derivative, containing regulatory element and GFPmut2, Ampr | This work |

**Table S2**—Primers used in this study

|  |  |
| --- | --- |
| **Primer** | **Sequence (5’→3’)** |
| aroG-F | CGCATCCGACAATTAAACCTTACCCGCGACGCGCTTTTA |
| aroG-R | TGGCAACACTGGAACAGACATGAATTATCAGAACGACGA |
| PH5a-M-F | CGTCGTTCTGATAATTCATGTCTGTTCCAGTGTTGCCAT |
| PH5a-M-R | AGCGGCGACGCGCAGTTAATCCCACAGCCGCCAGTTCCG |
| trpEDCBA-F | GGAACTGGCGGCTGTGGGATTAACTGCGCGTCGCCGCTT |
| trpEDCBA-R | ACAAAATTAGAGAATAACAATGCAAACACAAAAACCGAC |
| PH5a-ver-F | TCGGTTTTTGTGTTTGCATTGTTATTCTCTAATTTTGTT |
| PH5a-ver-R | AAAAGCGCGTCGCGGGTAAGGTTTAATTGTCGGATGCGC |
| ppsA-up-F | GAATCCATGGGCCTGTTGAAAGCATAAATTAAAAACG |
| ppsA-up-R | TTAAACAAAATTATTGGGGAATTGTTATCCGCTCACAATTCCACACATTATACGAGCCGATGATTAATTGTCAACGAACAATCCTTTTGTGATA |
| ppsA-down-F | ATTCCCCAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGTCCAACAATGGCTCGTCACCGCTCGTGCTTTGGTATAACCAAC |
| ppsA-down-R | TCCAAGCTTCCATTCAGAAGGGAGTGTCGATAATCC |
| ppsA-ver-F | TCGACACTCCCTTCTGAATGGAAGCTTGGATTCTC |
| ppsA-ver-R | AATTTATGCTTTCAACAGGCCCATGGATTCTTC |
| ppsA-N20-F | TAGCCAACAATGGCTCGTCACCGCGTTTTAGAGCTAGAAATAGC |
| ppsA-N20-R | CGCGGTGACGAGCCATTGTTGGCTAAGATCTGACTCCATAA |
| trpR-up-F | AATCCATGGGCCTGTAGCAGCTTATAACGCCGGA |
| trpR-up-R | ATCAGGCCTACAAAAAATATGTCGCCATTGTTAGC |
| trpR-down-F | CAATGGCGACATATTTTTTGTAGGCCTGATAAGAC |
| trpR-down-R | CCAAGCTTCCATTCATGGTCCCGTGATGTCGCGT |
| trpR-ver-F | ACATCACGGGACCATGAATGGAAGCTTGGATTCTC |
| trpR-ver-R | GCGTTATAAGCTGCTACAGGCCCATGGATTCTTC |
| trpR-N20-F | GCCAGATGAGCGCGAAGCGTGTTTTAGAGCTAGAAATAGC |
| trpR-N20-R | ACGCTTCGCGCTCATCTGGCGCTAAGATCTGACTCCATAA |
| tnaC-F | gaaacagaccatggaattcGTGTCTTGCGAGGATAAGTG |
| tnaC-R | TCTTCTCCTTTACTCATCGGTTCAGGGAGATGTTTAAAG |
| GFP-tnaC-F | TAAACATCTCCCTGAACCGATGAGTAAAGGAGAAGAACT |
| GFP-tnaC-R | CGCCAAAACAGCCAAGCTTCTATTTGTATAGTTCATCCA |
| ptrc99a-ver-F | GGATGAACTATACAAATAGAAGCTTGGCTGTTTTGGCGG |
| ptrc99a-ver-R | ACTTATCCTCGCAAGACACGAATTCCATGGTCTGTTTCC |

**Table S3**—Statistics of genome sequencinga

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sample name** | **Seq. type** | **Total reads** | **Good readb** | **Total bases** | **Q20c** | **Result** |
| TPD5 | DNA | 11,942,244 | 11,942,244 | 1,194,224,400 | 0.919 | qualified |

aThe genomes of E. coli TPD5 were sequenced using 454 pyrosequencing and aligned to the reference E. coli W3110 (NCBI reference NC.xxx).

bGood reads refers to the number of Reads that are machine-filtered to remove low quality bases.

cQ20=bases of Q>=20 / all bases of sequencing

**Table S4**—Substitutions in ORF

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Gene name** | **Reference sequence** | **Observed sequence** | **Amino acid changing** | **Position** |
| imp | G | A | E288K | 56248 |
| lacI | TTGGG | T | - | 366513 |
| acrR | G | A | D139N | 486175 |
| ylbE | A | AG | - | 547831 |
| entE | G | A | A407T | 626511 |
| ybdK | C | T | T235M | 606680 |
| aroG | C | T | S211F | 786264 |
| oppA | A | T | N271Y | 1301992 |
| oppF | T | G | S325A | 1306736 |
| trpE | C | T | A63V | 1322758 |
| trpE | G | A | T425T | 1321513 |
| acnA | A | G | S522G | 1337394 |
| puuP | A | G | Y110C | 1358859 |
| ydeI | AT | A | - | 1626064 |
| dicA | G | A | E123K | 1648300 |
| mlc | GC | G | - | 1667884 |
| flhD | G | T | V84F | 1977948 |
| glpQ | C | T | R101C | 2350711 |
| gabD | A | C | T130P | 2791660 |
| rpoS | C | T | Q33\* | 2867455 |
| ygiQ | CT | C | - | 3157654 |
| ebgA | G | A | G111D | 3222964 |
| garL | G | T | A190S | 3272845 |
| dinD | C | T | P7S | 3822637 |
| tnaA | AG | A | - | 3889092 |
| asrB | G | A | E303K | 3990574 |
| polA | G | A | G763S | 4049252 |
| argH | G | A | V393M | 4158026 |
| rpoB | C | A | S917Y | 4183994 |

Asterisks (\*) indicate nonsense mutations. The minus (-) signs stand for insertion or deletion.

**Table S5**—Mutations in Non-ORF

|  |  |  |  |
| --- | --- | --- | --- |
| **Mutation startpoint in genome** | **Mutation endpoint in genome** | **Reference sequence** | **Observed sequence** |
| 366351 | 366351 | A | T |
| 684504 | 684504 | G | A |
| 987152 | 987152 | G | T |
| 1992019 | 1992019 | C | T |
| 2470659 | 2470659 | G | A |
| 2543565 | 2543565 | C | CATCAT |
| 3534456 | 3534456 | G | T |
| 4168571 | 4168571 | G | A |
| 4296380 | 4296380 | A | ACG |

### Reference

(1) D.A. Case, I. Y. B.-S., S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden,, R.E. Duke, D. G., M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, Y. Huang,, S. Izadi, A. K., T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J., Mermelstein, K. M. M., Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R., Qi, D. R. R., A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. Salomon-, Ferrer, J. S., R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman, and (2018), A., University of California, San Francisco. (2018).

(2) Hopkins, C. W., Le Grand, S., Walker, R. C., and Roitberg, A. E. (2015) Long-Time-Step molecular dynamics through hydrogen mass repartitioning. *J Chem. Theory. Comput.* *11*, 1864-1874.

(3) Jakalian, A., Bush, B. L., Jack, D. B., and Bayly, C. I. (2000) Fast, efficient generation of high-quality atomic Charges. AM1-BCC model: I. Method. *J Comput Chem* *21*, 132-146.

(4) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* *79*, 926-935.

(5) Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., Debolt, S., Ferguson, D., Seibel, G., and Kollman, P. (1995) Amber, a package of computer-programs for applying molecular mechanics, normal-mode analysis, molecular-dynamics and free-energy calculations to simulate the structural and energetic properties of molecules. *Comput Phys Commun* *91*, 1-41.

(6) Andersen, H. C. (1983) Rattle: A “Velocity” version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.* *52*, 24-34.

(7) Chow, K. H., and Ferguson, D. M. (1995) Isothermal isobaric molecular-dynamics simulations with monte-carlo volume sampling. *Comput Phys Commun* *91*, 283-289.

(8) Le Grand, S., Götz, A. W., and Walker, R. C. (2013) SPFP: Speed without compromise—A mixed precision model for GPU accelerated molecular dynamics simulations. *Comput Phys Commun* *184*, 374-380.

(9) Roe, D. R., and Cheatham, T. E., 3rd. (2013) PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J Chem. Theory. Comput.* *9*, 3084-3095.

(10) The PyMOL molecular graphics system, Version 2.0 Schrödinger, LLC.