Supplementary Information

Energetic Frustration Dictates Cognate Specificity of Two-Component Signaling System

Yuqing Li[1,2], Peng Li[4], Junhui Nie[4], Jing Zeng[4],
Yingtang Zhou[1,*], Lin Yuan[4,*],
Jin Wang[2,3,*], Zhiqiang Yan[2,*]

[1]Zhejiang Key Laboratory of Petrochemical Environmental Pollution Control, National Engineering Research Center for Marine Aquaculture, Marine Science and Technology College, Zhejiang Ocean University, Zhoushan, Zhejiang, China
[2]Wenzhou Institute, University of Chinese Academy of Sciences, Wenzhou, Zhejiang, China
[3]Department of Chemistry and Physics, State University of New York at Stony Brook, Stony Brook, NY, USA
[4]Institute of Microbiology, Jiangxi Academy of Sciences, Nanchang, Jiangxi, China

*Correspondence: zhouyingtang@zjou.edu.cn, yuanlin2003cn@aliyun.com, jin.wang.1@stonybrook.edu, zqyan@ucas.ac.cn

Supplementary Figures

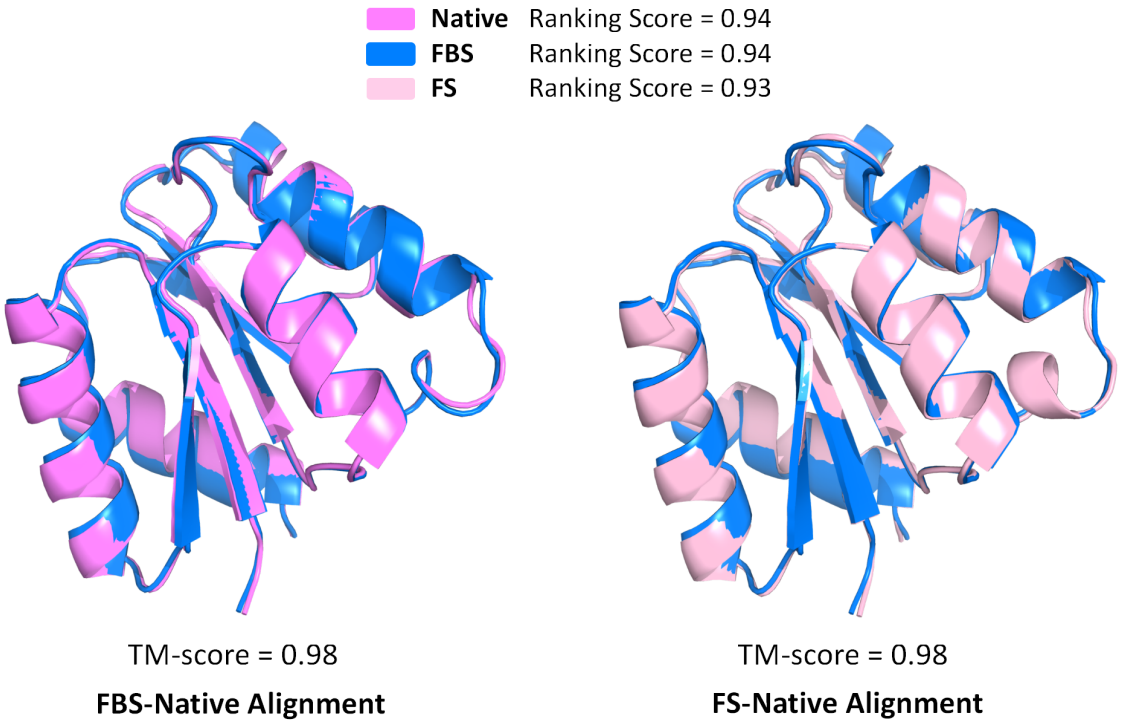

Figure S1: AF3-predicted RR structures for native sequence (NS), folding sequence (FS) and fold-binding sequence (FBS). (Left) Alignment of RR structures predicted with FBS (magenta) and NS (blue). (Right) Alignment of RR structures predicted with FS (pink) and NS (blue). Both FSs and FBSs yield highly accurate structures with average ranking scores (RS) of 0.93 to 0.94, and the TM-score between the predicted structures and the native RR structure is 0.98, indicating that both FSs and FBSs can fold into native-like structure as NS.

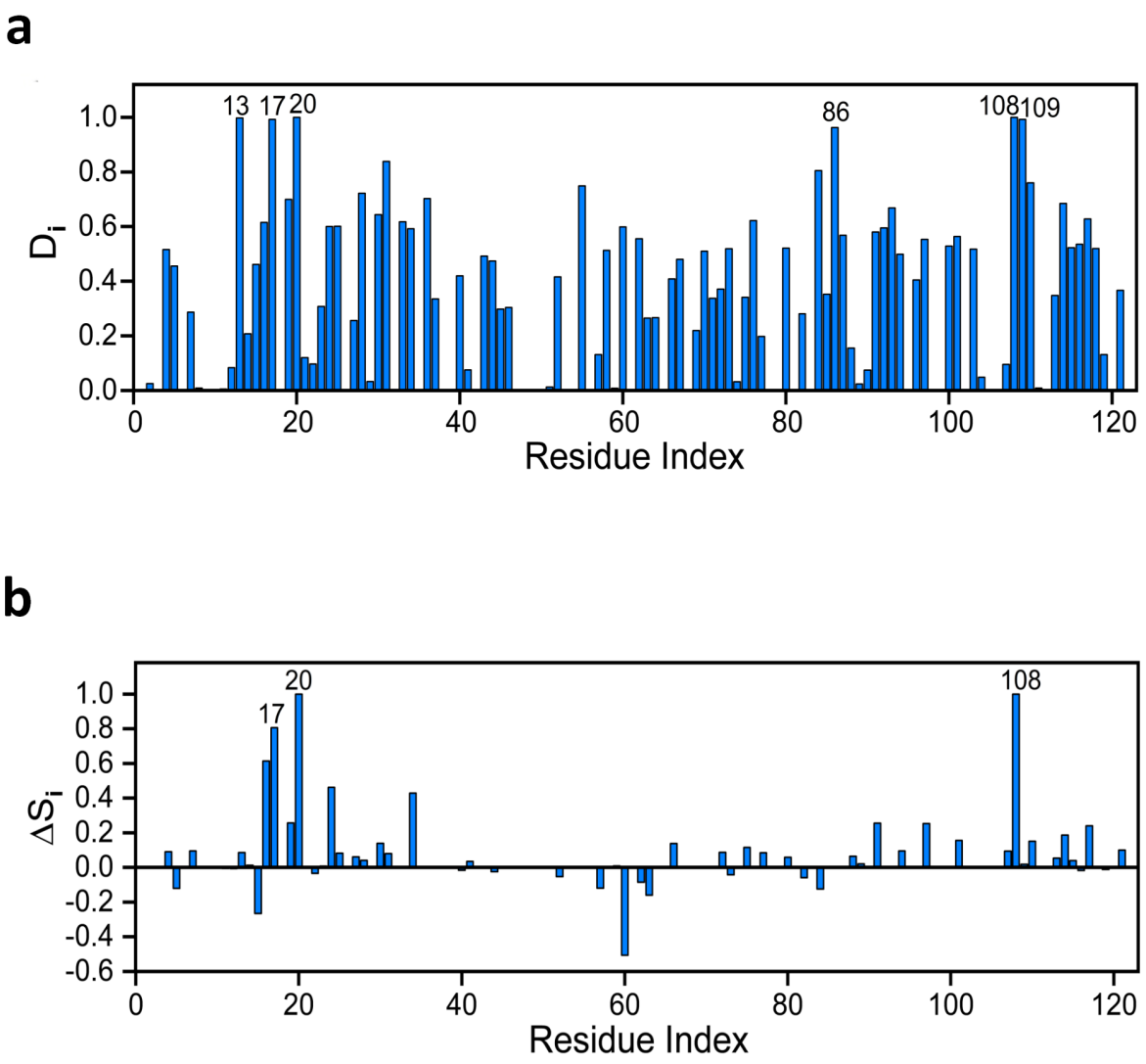

Figure S2: Amino acid similarities among FBSs, FSs and NS. (a) Amino acid dissimilarity between FBSs and FSs, computed as D_i=1-S_i, where S_i is the amino acid similarity at site i between FBSs and FSs. (b) Difference in amino acid similarity to native sequence of RR between FBSs and FSs, computed as ΔS_i=S_i^{FBS}-S_i^{FS}, where S_i^{FBS} and S_i^{FS} are the amino acid similarities of FBSs and FSs to NS, respectively.

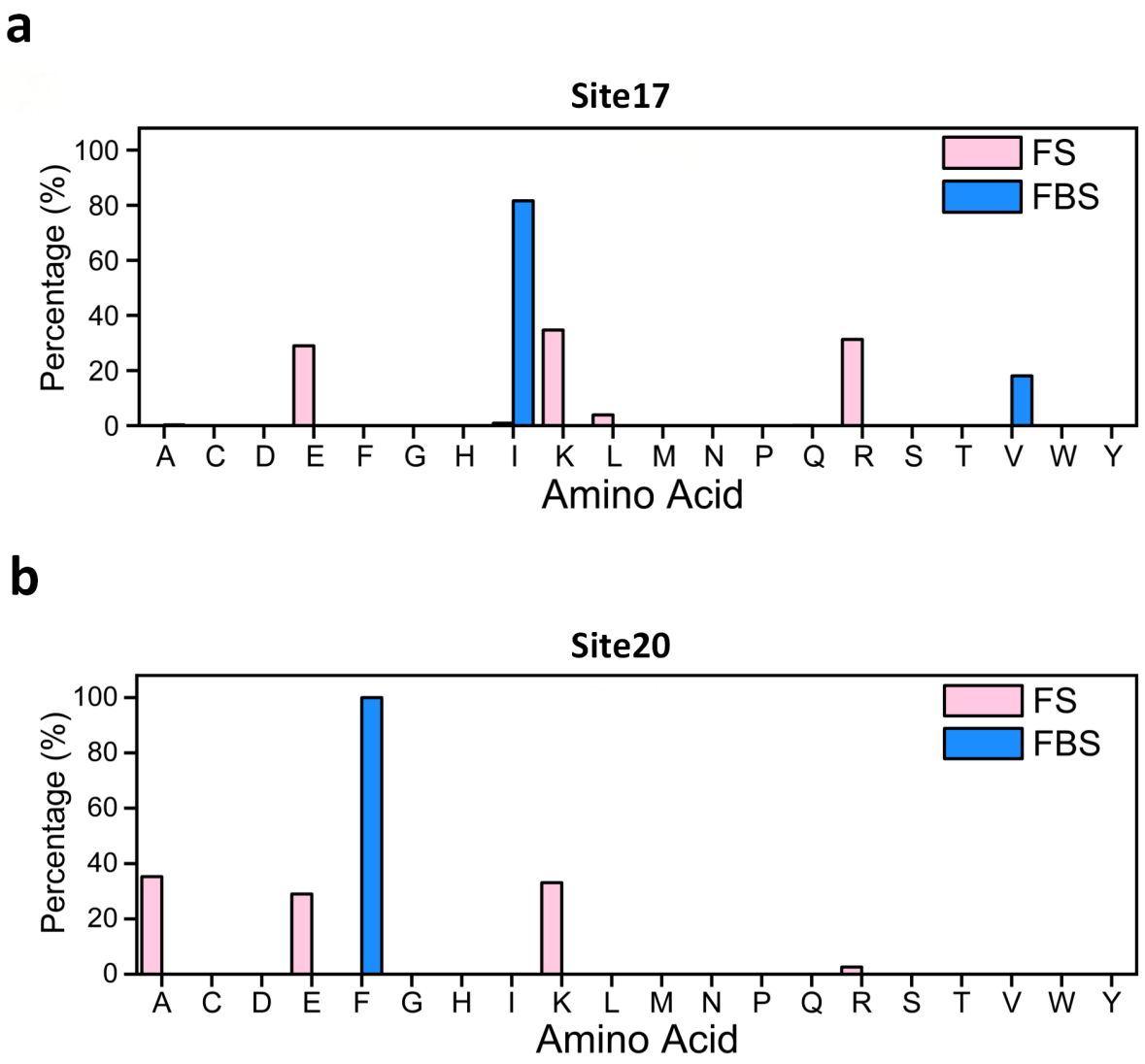

Figure S3: Amino acid preferences of ProteinMPNN-designed FSs and FBSs at Site 17 (a) and Site 20 (b). For FBSs, the most probable amino acids at Sites 17 and 20 are I and F, respectively, which are identical to those in NS; for FSs, however, the most probable amino acids at site 17 and 20 are E/K/R and A/E/K, respectively.

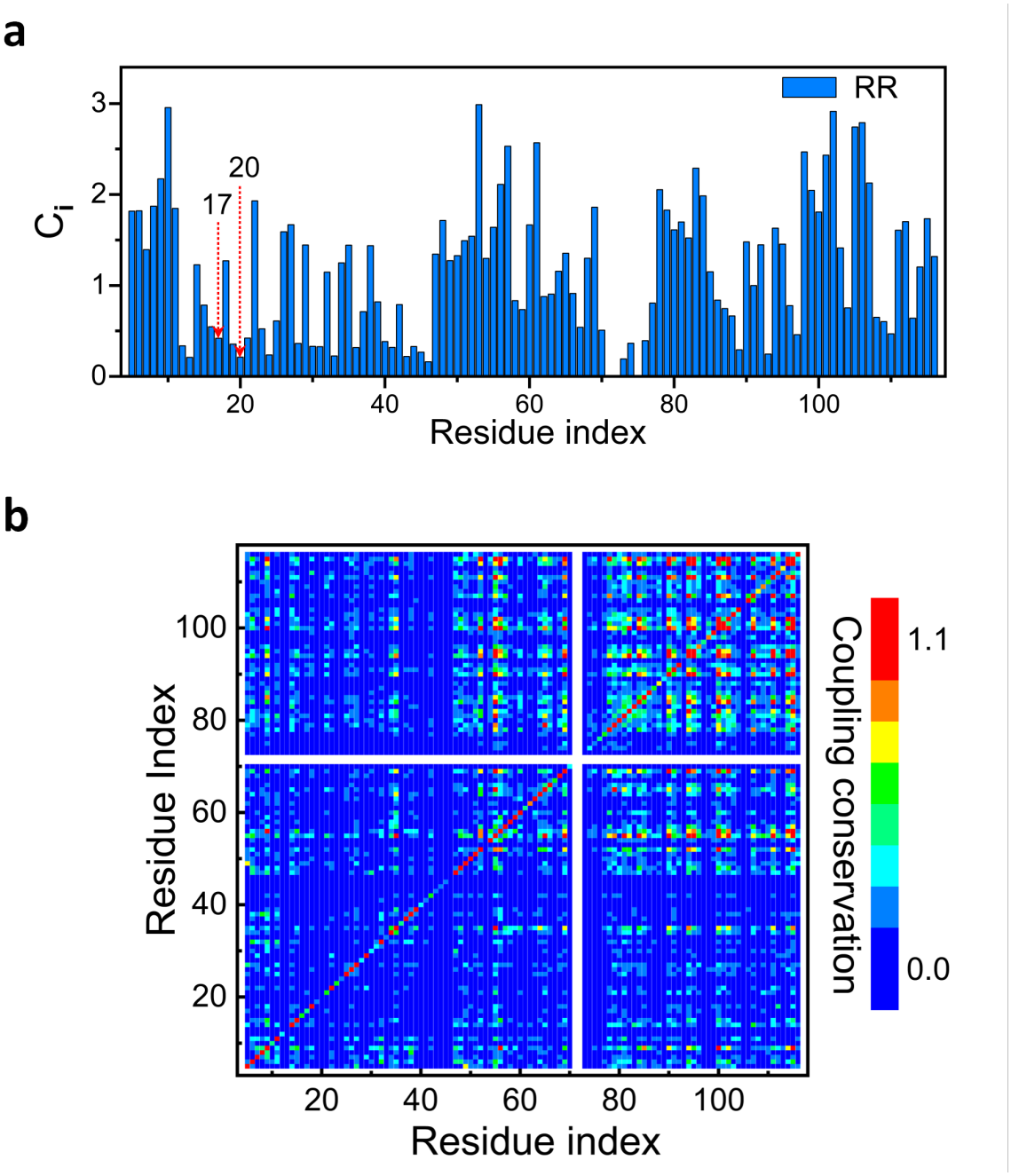

Figure S4: First-order and second-order conservations in homologous multiple sequence alignment (MSA) of RR domain. (a) First-order site conservation (C_i) with two specificity-determining sites labeled. (b) Matrix of second-order coupling conservation with color scaling, highly coupling conservations ($C_{ij} \geq 1.1$) are colored red.

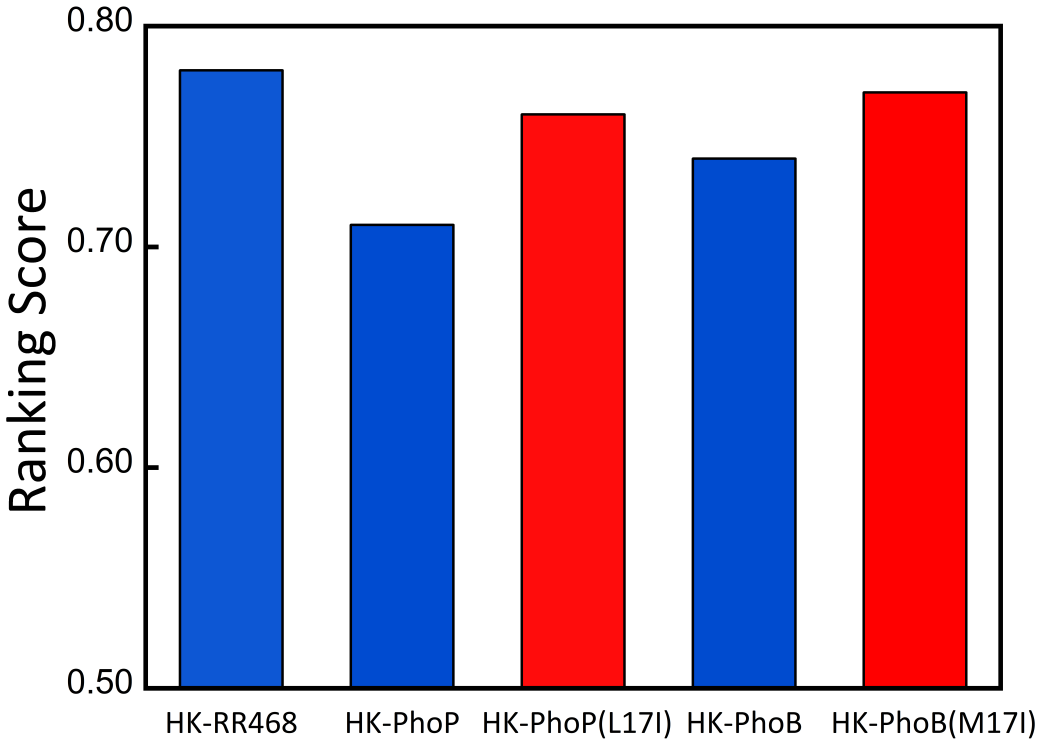

Figure S5: Tuning crosstalk specificity of HK-RR pairings via swapping amino acids at site 17, including L17I swap for HK853-PhoB pairing, and M17I swap for HK853-PhoP pairing.

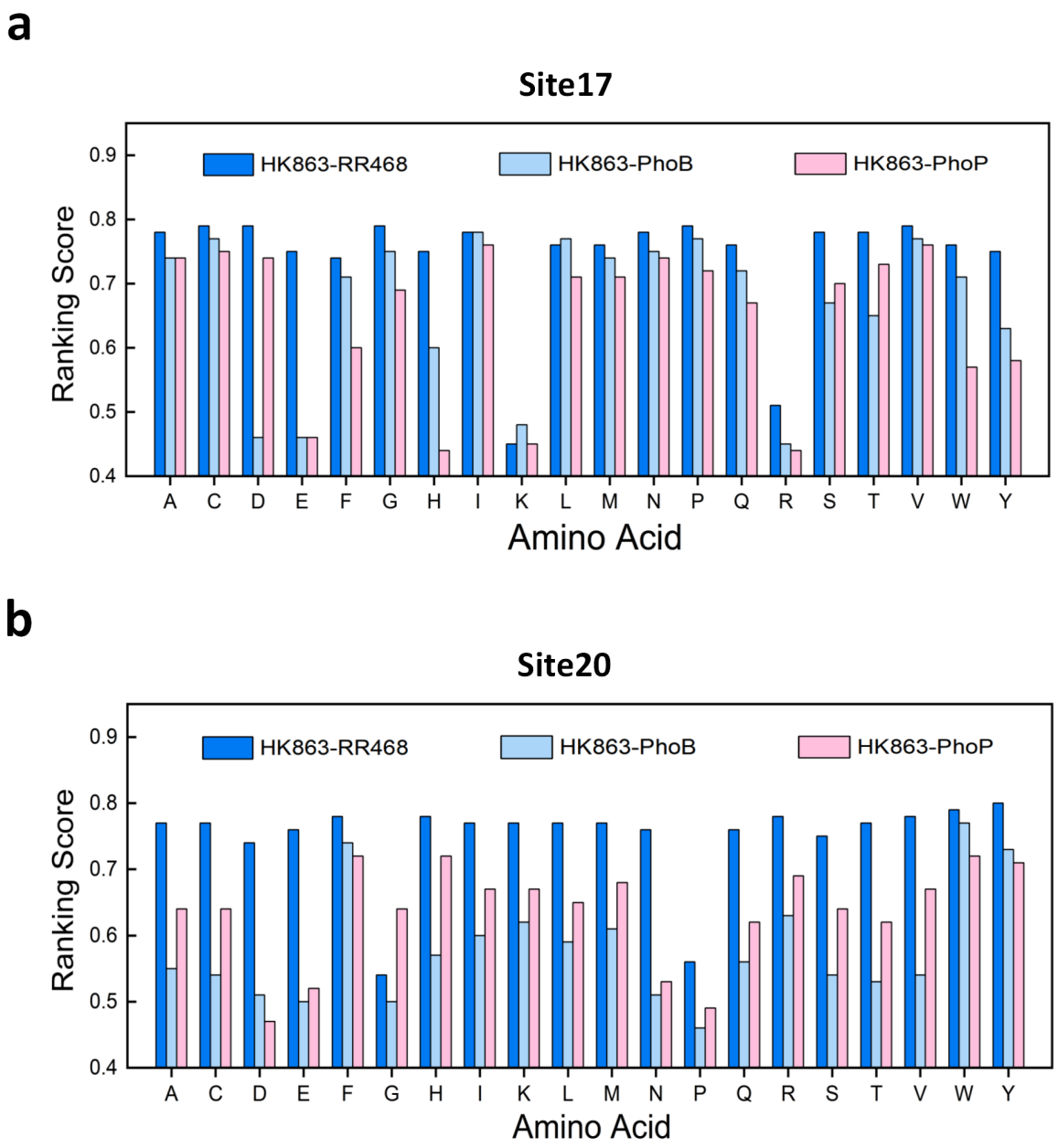

Figure S6: Two categories of substitutions in tuning specificity at site 17 (a) and site 20 (b). The first category not only turn off the cognate specificity of HK853-RR468 pairing but also suppresses the non-cognate specificity of HK853-PhoB or HK853-PhoP pairing. In contrast, the second category retains the cognate specificity of HK853-RR468 pairing and solely suppresses the non-cognate specificity of HK853-PhoB or HK853-PhoP pairing. For example, at site 17, the first category contains K/R which turning off both cognate HK853-RR468 and non-cognate HK853-PhoB pairings, whereas the second category contains D/E which only turns off HK853-PhoB pairing.

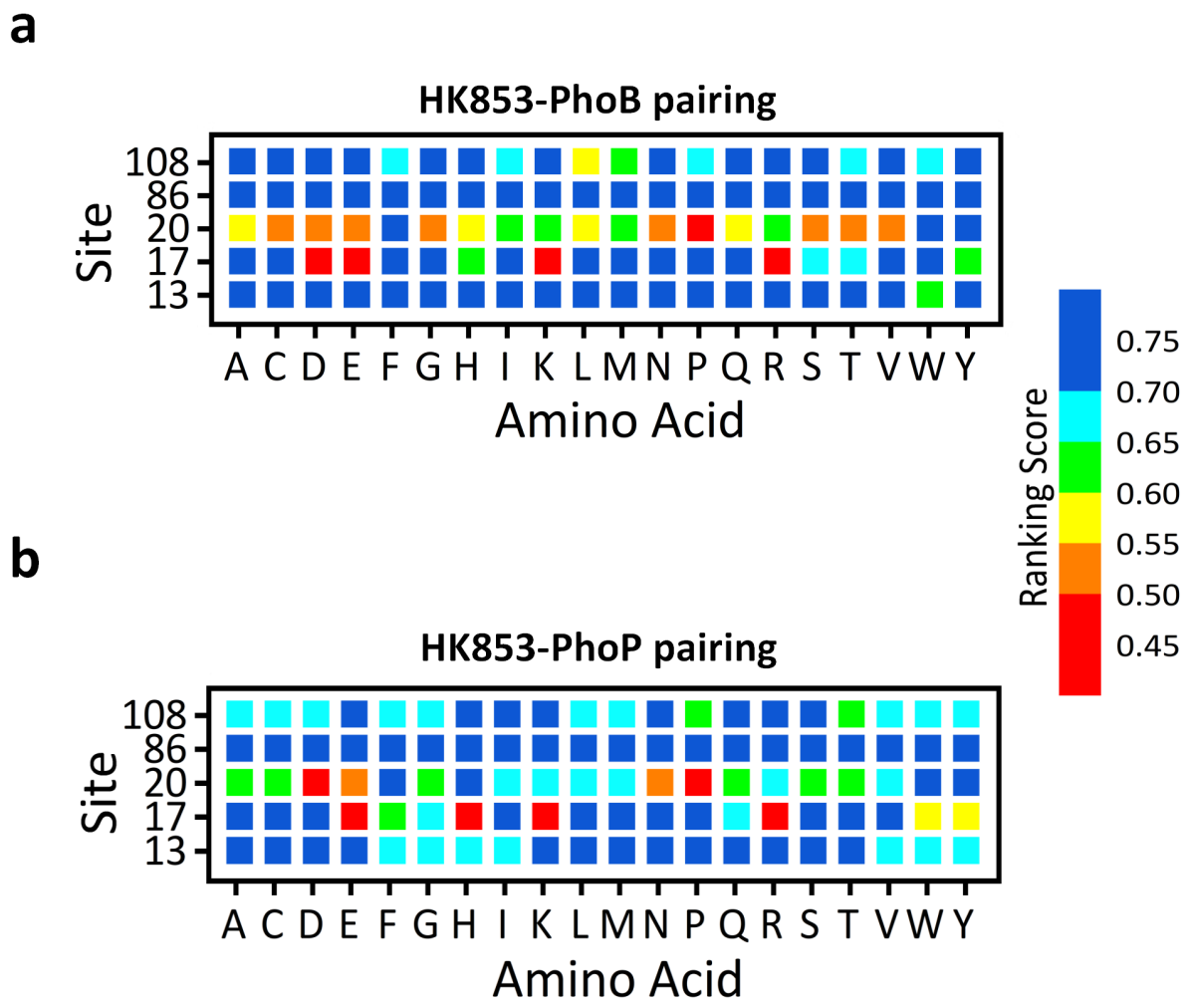

Figure S7: AF3-predicted ranking scores from computational saturation mutagenesis of those RR interface sites (13, 17, 20, 86 and 108) with high local frustration difference and amino acid dissimilarity between FSs and FBSs. Obvious specificity-tuning effects were observed at Sites 17 and 20, whereas only minor tuning effects were observed from saturation mutagenesis of the other sites for the HK853-PhoB and HK853-PhoP pairings.

Supplementary Tables

Table S1: AF3-predicted ranking scores for the wild-type sequences and their variants with single amino acid swaps at site 13, 17, 20, 86 and 108, as well as double swaps of these sites between wild-type NS/FBSs and FSs, the scores reflecting binding specificity reversal of wildtypes are color red.

AF3-predicted ranking score			
Wildtype	**NS**	**FBS**	**FS**
	0.78	0.77	0.44
Single swapping variant	**NS***	**FBS***	**FS***
13	0.78	0.76	0.47
17	**0.45**	**0.50**	0.48
20	0.76	**0.54**	0.45
86	0.79	0.77	0.47
108	0.78	0.76	0.46
Double swapping variant	**NS***	**FBS***	**FS***
13–17	**0.43**	**0.52**	0.45
13–20	0.76	**0.52**	0.44
13–86	0.78	0.77	0.44
13–108	0.78	0.77	0.44
17–20	**0.42**	**0.50**	**0.68**
17–86	**0.45**	**0.45**	0.45
17–108	**0.45**	**0.48**	0.52
20–86	0.76	0.59	0.44
20–108	**0.54**	**0.46**	0.43
86–108	0.79	0.77	0.42

* means corresponding variants of NS, FBS and FS by swapping amino acids

Table S2: Bacterial strains and plasmids used in this study.

Bacterial strains/plasmids	**Description**	**Source**
	E. coli strains	
DH5α	F-Φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rK$^-$, mK$^+$) phoA supE44 λ-thi-1 gyrA96 relA1	Invitrogen
XL1 Blue	recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB lacI^q ZΔM15 Tn10 (Tet^R)]	Stratagene
BTH101	F$^-$, cya-99, araD139, galE15, galK16, rpsL1 (StrR), hsdR2, mcrA1, mcrB1	[1]
	Plasmids	
pKT25	Kan^R, encoding T25 fragment	[2]
pUT18C	Amp^R, encoding T18 fragment	[2]
pKT25-HK	HK was fused to the C termini of adenylate cyclase (T25) in pKT25	This study
pUT18C-RR	RR was fused to the C termini of adenylate cyclase (T18) in pUT18C	This study
pUT18C-phoP	PhoP was fused to the C termini of adenylate cyclase (T18) in pUT18C	This study
pUT18C-phoB	PhoB was fused to the C termini of adenylate cyclase (T18) in pUT18C	This study
pUT18C-RR(I17K)	RR mutant with an I17K substitution (17th isoleucine replaced by lysine)	This study
pUT18C-phoP(L17I)	PhoP mutant with a L17I substitution (17th leucine replaced by isoleucine)	This study
pUT18C-phoB(M17I)	PhoB mutant with a M17I substitution (17th methionine replaced by isoleucine)	This study
pUT18C-RRS1	Designed RR S1 was fused to the C termini of adenylate cyclase (T18) in pUT18C	This study
pUT18C-RRS2	Designed RR S2 was fused to the C termini of adenylate cyclase (T18) in pUT18C	This study
pUT18C-RRS3	Designed RR S3 was fused to the C termini of adenylate cyclase (T18) in pUT18C	This study
pUT18C-RRS4	Designed RR S4 was fused to the C termini of adenylate cyclase (T18) in pUT18C	This study
pUT18C-RRS5	Designed RR S5 was fused to the C termini of adenylate cyclase (T18) in pUT18C	This study
pUT18C-RRS6	Designed RR S6 was fused to the C termini of adenylate cyclase (T18) in pUT18C	This study
pUT18C-RRS7	Designed RR S7 was fused to the C termini of adenylate cyclase (T18) in pUT18C	This study
pUT18C-RRS8	Designed RR S8 was fused to the C termini of adenylate cyclase (T18) in pUT18C	This study
pUT18C-RRS9	Designed RR S9 was fused to the C termini of adenylate cyclase (T18) in pUT18C	This study
pUT18C-RRS10	Designed RR S10 was fused to the C termini of adenylate cyclase (T18) in pUT18C	This study

Table S3: Oligonucleotides used in this study. f and r indicate forward and reverse primer, respectively. The underlined nucleotides indicate the 20-bp homology arms.

Primers	**Sequences (5' → 3')**
HK-f	GCTGCAGGGTCGACTCTAGAGATGGAAGAAATTAAAGAACAGAAC
HK-r	GGATAGGTACCCGGGGATCCTCAGTTATCCTGGCGGTT
RR-f	AGGTCGACTCTAGAGGATCCCATGAGCAAAAAAGTGCTGC
RR-r	ATGAATTCGAGCTCGGTACCTCATTCGTTCAGCAGATGT
phoP-f	AGGTCGACTCTAGAGGATCCCATGAGCCAGAAAGTGCTG
phoP-r	ATGAATTCGAGCTCGGTACCTCAGCGCAGAATCGCTTT
phoB-f	AGGTCGACTCTAGAGGATCCCATGGCGCGCCGCATTCTG
phoB-r	ATGAATTCGAGCTCGGTACCTCAGCGCATCACCGCTTT
RRS1-f	AGGTCGACTCTAGAGGATCCCATGAAGAAGATTTTGTTGGTTG
RRS1-r	ATGAATTCGAGCTCGGTACCTCAGTTCAACAACTTTTCAAC
RRS2-f	AGGTCGACTCTAGAGGATCCCATGAAGTCTGTTTTGTTGGTTG
RRS2-r	ATGAATTCGAGCTCGGTACCTCAGTTCAACAACTTTTCAAC
RRS3-f	AGGTCGACTCTAGAGGATCCCATGAAGTCTATTTTGTTGGTTG
RRS3-r	ATGAATTCGAGCTCGGTACCTCAGTTCAACAATTCCTTAAC
RRS4-f	AGGTCGACTCTAGAGGATCCCATGAAGAAGGTTTTGTTGG
RRS4-r	ATGAATTCGAGCTCGGTACCTCAGTTCAACAATTCCTTAAC
RRS5-f	AGGTCGACTCTAGAGGATCCCATGAAGAAGATTTTGTTGGTTG
RRS5-r	ATGAATTCGAGCTCGGTACCTCATTCCAACAACTTCTTAGC
RRS6-f	AGGTCGACTCTAGAGGATCCCATGAAGAAGGTTTTGTTGG
RRS6-r	ATGAATTCGAGCTCGGTACCTCAGTTCAACAACTTCTTAAC
RRS7-f	AGGTCGACTCTAGAGGATCCCATGAAGAAGGTTTTGGTTG
RRS7-r	ATGAATTCGAGCTCGGTACCTCATTCCAACAACTTCTTAAC
RRS8-f	AGGTCGACTCTAGAGGATCCCATGAAGTCTATTTTGTTGGTTG
RRS8-r	ATGAATTCGAGCTCGGTACCTCAGTTCAACAACTTCTTAACTTC
RRS9-f	AGGTCGACTCTAGAGGATCCCATGAAGTCTATTTTGTTGGTTG
RRS9-r	ATGAATTCGAGCTCGGTACCTCAGTTCAACAACTTCTTAAC
RRS10-f	AGGTCGACTCTAGAGGATCCCATGAAGTCTATTTTGTTGGTTG
RRS10-r	ATGAATTCGAGCTCGGTACCTCAGTTCAACAACTTTTCAAC
RRI17K-f	GCGCAAAAAGGTGAGCTTTAACCTGAAAAAAGAAGGC
RRI17K-r	AAGCTCACCTTTTTGCGCAGCACCGCGC
phoP(L17I)-f	TGTGACCATTCTGAAATATAACCTGGAAACCGCG
phoP(L17I)-r	ATTTCAGAATGGTCACAATGCTATGTTCATCATCC
phoB(M17I)-f	TCGCGAAATTGTGTGCTTTGTGCTGGAACAG
phoB(M17I)-r	AAAGCACACAATTTCGCGAATCGGCGCT

Datasets

Dataset S1.xlsx 10 Top-ranked ProteinMPNN-designed sequences employed for bacterial two-hybrid assays.

Dataset S2.xlsx 5000 ProteinMPNN-designed RR sequences.

Dataset S3.xlsx 4069 HK-RR pairs of native sequences for MSA.

References

[1] Daniel Ladant and Agnes Ullmann. Bordetella pertussis adenylate cyclase: a toxin with multiple talents. *Trends In Microbiology*, 7(4):172–176, 1999.

[2] Gouzel Karimova, Josette Pidoux, Agnes Ullmann, and Daniel Ladant. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proceedings Of The National Academy Of Sciences*, 95(10):5752–5756, 1998.