Discovery of novel chemoeffectors and rational design of *Escherichia coli* chemoreceptor specificity

Shuangyu Bi\(^{a,b}\), Daqi Yu\(^{a,b}\), Guangwei Si\(^{b,c}\), Chunxiong Luo\(^{b,c}\), Tongqing Li\(^{a,b}\), Qi Ouyang\(^{b,c,d}\), Vladimir Jakovljevic*, Victor Sourjik\(^{e,f}\), Yuhai Tu\(^{b,g}\), and Luhua Lai\(^{a,b,d,1}\)

*Beijing National Laboratory for Molecular Sciences, State Key Laboratory for Structural Chemistry of Unstable and Stable Species, College of Chemistry and Molecular Engineering, Center for Quantitative Biology, Academy for Advanced Interdisciplinary Studies, College of Physics, and Peking University, Beijing 100871, China; Zentrum für Molekulare Biologie der Universität Heidelberg, D-69120 Heidelberg, Germany; Max Planck Institute for Terrestrial Microbiology, D-35043 Marburg, Germany; and IBM T. J. Watson Research Center, Yorktown Heights, NY 10598

Edited* by Howard C. Berg, Harvard University, Cambridge, MA, and approved September 10, 2013 (received for review April 11, 2013)

Bacterial chemoreceptors mediate chemotactic responses to diverse stimuli. Here, by using an integrated in silico, in vitro, and in vivo approach, we screened a large compound library and found eight novel chemoeffectors for the *Escherichia coli* chemoreceptor Tar. Six of the eight new Tar binding compounds induce attractant responses, and two of them function as antagonists that can bind Tar without inducing downstream signaling. Comparison between the antagonist and attractant binding patterns suggests that the key interactions for chemotaxis signaling are mediated by the hydrophobic bonds formed between a donor group in the attractant and the main-chain carbonyls (Y149 and/or Q152) on the α4 helix of Tar. This molecular insight for signaling is verified by converting an antagonist to an attractant when introducing an N-H group into the antagonist to restore the hydrogen bond. Similar signal triggering effect by an O-H group is also confirmed. Our study suggests that the Tar chemoreceptor binding pocket may be separated into two functional regions: region I mainly contributes to binding and region II contributes to both binding and signaling. This scenario of binding and signaling suggests that Tar may be rationally designed to respond to a nonnative ligand by altering key residues in region I to strengthen binding with the novel ligand while maintaining the key interactions in region II for signaling. Following this strategy, we have successfully redesigned Tar to respond to L-arginine, a basic amino acid that does not have chemotactic effect for WT Tar, by two site-specific mutations (R69E and R73E).

Two-component signaling pathways are ubiquitous in bacteria. They enable the cells to recognize and respond to different environmental stimuli (1). The control network of bacterial chemotaxis uses such a two-component system to sense the extracellular chemoeffector concentrations (2, 3). Chemoreceptors are the key upstream sensory components in the chemotaxis signaling pathway. They directly interact with specific extracellular chemoeffectors and transfer environmental information to the downstream response regulator, which ultimately controls the cell’s motility (4, 5).

Tar is one of the major chemoreceptors found in *Escherichia coli* and *Salmonella enterica* serovar Typhimurium (6). Attractant and repellent molecules that can induce chemotactic responses of the cells by interacting with Tar were studied (7). In addition to these two types of chemoeffectors, antagonist molecules that can directly bind to chemoreceptors without generating chemotactic responses should also exist. For example, antagonists for the sensor kinase TodS were found in the TodS/TodT two-component system (8). However, so far, antagonist molecules that function by directly binding to *E. coli* chemoreceptors have not been reported.

Much progress has been made in understanding the structural basis of chemoreceptor signaling. Crystal structures show that each monomer (in the Tar homodimer) contains a four-helix bundle (helices α1–α4) structure, of which the α1 and α4 helices span the membrane to form transmembrane domains (TM1 and TM2) (9–12). Increasing evidence verifies the piston model (13–20), which suggests that attractant binding to the Tar periplasmic domain generates a subtle piston-like sliding movement of the α4 helix relative to the α1 helix in one monomer within the chemoreceptor homodimer. This modest conformational change can be transduced over long distances to the cytoplasm. The attractant binding pocket of Tar has also been well characterized (12, 21–24). Among the key residues that bind with the ligand, some are on the signaling monomer (the monomer with the α4 sliding movement upon signaling) and others are on the nonsignaling monomer. This raises the question of the relationship between binding and signaling, that is, does binding always lead to signaling?

Tar is highly selective toward its chemoeffectors with the highest sensitivity for its native attractant L-aspartate (Asp), and lower sensitivities for several other amino acids (6, 25). Redesigning chemoreceptors to recognize and respond to nonnative ligands is highly desirable with potential applications in bioengineering and biotechnology. So far, however, successful change (or improvement) in chemoreceptor specificity were all carried out by genetic screening (i.e., directed evolution) (26). Structure-based rational design of receptor specificity remains highly challenging as the key molecular features for attractant binding and signaling are not fully understood.

In this paper, we report our work in trying to address these related questions. By using a combination of in vitro, in vivo, and in silico methods, we discovered several new chemoeffectors for Tar, including two antagonists. Comparing the molecular binding patterns of the attractants and the newly discovered antagonists suggests that the ligand–Tar interaction can be separated into two groups. The first group of interactions (type I) mainly

**Significance**

Chemotaxis is a universal phenomenon whereby motile cells, like motile bacteria, navigate by following chemical gradients in their environment. Bacterial chemoreceptors can bind with specific chemoeffectors and transfer environmental signals to the cell. However, the molecular mechanisms for chemoeffector binding and signaling are not fully understood, and rational design of bacteria to respond to new chemicals has been challenging. In this study, by using a combined experimental and computational approach, we discovered novel antagonists and attractants for the *Escherichia coli* chemoreceptor Tar. The interaction differences of the novel antagonists and attractants with Tar provide clues to alter Tar–ligand specificity. Based on these understandings, *E. coli* strain was successfully engineered to sense L-arginine, a ligand unrecognized before.

Author contributions: Y.T. and L.L. designed research; S.B., D.Y., G.S., C.L., T.L., Q.O., V.J., and V.S. performed research; S.B., D.Y., G.S., C.L., Y.T., and L.L. analyzed data; and S.B., Y.T., and L.L. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

1To whom correspondence may be addressed. E-mail: yuhai@us.ibm.com or lhlai@pku.edu.cn.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1306811110/-/DCSupplemental.
stabilizes ligand binding. The second group of interactions (type II), which includes the hydrogen bonds formed between attractant and the main-chain carbonyls of the α4 helix, contributes to binding and signaling. These two types of interactions do not occur sequentially; they need to act together in concert to induce attractant signaling. Ligands that bind to Tar with only the type I interaction function as antagonists. Based on these molecular insights on binding and signaling, an E. coli mutant strain that responds to L-arginine (Arg) was designed successfully by introducing only two site-specific mutations.

Results

Screening for Potential Novel Chemoeffectors by Molecular Docking. Molecular docking calculations were used to virtually screen compounds that may function as novel chemoeffectors for the E. coli Tar receptor. As no complex crystal structure is available for the periplasmic domain of E. coli Tar with Asp, we built a complex structure model based on the holo structure of Tar from Salmonella with Asp (Protein Data Bank ID code 1VLT) (12) and used it in virtual screening. Because of the limited size of the Asp binding pocket, compounds with molecular weight less than 300 Da were used (from the Available Chemical Directory, Elsevier MDL) in the docking study (27). The top 10,000 compounds with the lowest estimated binding free energies were manually inspected. Eighty typical compounds were purchased for the subsequent experimental study.

Binding Affinity Measurements by Isothermal Titration Calorimetry. Isothermal titration calorimetry (ITC) is a technique for quantitatively measuring protein–ligand binding thermodynamic parameters. It is becoming widely used in fragment-based drug discovery to study low-affinity binding of fragments with target (refs. 28 and 29; www.gelifesciences.com/microcal). We measured the binding affinities of these 80 compounds to the purified E. coli Tar periplasmic domain by using ITC. The measured $K_d$ of α-methyl-DL-aspartate (AMA; a known attractant for Tar) with the periplasmic domain (compound 1 in Table 1; see also SI Appendix, Fig. S1A) agrees with previously published $K_d$ values obtained from competition centrifugation assay (6). Eight compounds exhibited significant binding to the Tar periplasmic domain (compounds 2–9 in Table 1; see also SI Appendix, Table S1 and Fig. S1 B–I).

Novel Chemoeffectors Identified by Microfluidic Experiments. We measured the responses of E. coli cells to these 80 compounds, especially those that exhibited Tar binding, by using a specially designed microfluidic device in which the chemotaxis responses to several chemicals could be measured in parallel (Fig. 1A). We first measured the chemotactic responses of E. coli strains RP437 (WT) and UU1624 (possessing only the Tar chemoreceptor) to each of the 80 compounds. Cells labeled with GFP were placed in the central hole, and different test compounds (or different concentrations of the same compound) were introduced in the peripheral holes. For a true attractant such as

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Abbreviation</th>
<th>Structure</th>
<th>$K_a$, mM</th>
<th>EC50, μM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>α-Methyl-DL-aspartate</td>
<td>AMA</td>
<td><img src="image1.png" alt="Image" /></td>
<td>0.559 ± 0.036</td>
<td>0.54 ± 0.04</td>
</tr>
<tr>
<td>2</td>
<td>(±)-α-Amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid</td>
<td>AMPA</td>
<td><img src="image2.png" alt="Image" /></td>
<td>4.2 ± 0.2</td>
<td>2.3 ± 0.4</td>
</tr>
<tr>
<td>3</td>
<td>Formimino-L-aspartate</td>
<td>FIA</td>
<td><img src="image3.png" alt="Image" /></td>
<td>33 ± 2</td>
<td>4.5 ± 0.1</td>
</tr>
<tr>
<td>4</td>
<td>Guanidinosuccinic acid</td>
<td>GSA</td>
<td><img src="image4.png" alt="Image" /></td>
<td>70 ± 4</td>
<td>13.7 ± 1.4</td>
</tr>
<tr>
<td>5</td>
<td>N-methyl-L-aspartate</td>
<td>NMA</td>
<td><img src="image5.png" alt="Image" /></td>
<td>26 ± 3</td>
<td>18.2 ± 0.8</td>
</tr>
<tr>
<td>6</td>
<td>N-formyl-L-aspartate</td>
<td>NFA</td>
<td><img src="image6.png" alt="Image" /></td>
<td>69 ± 6</td>
<td>109 ± 7</td>
</tr>
<tr>
<td>7</td>
<td>(2-Imino-4-oxo-thiazolidin-5-yl)-acetic acid</td>
<td>IOTA</td>
<td><img src="image7.png" alt="Image" /></td>
<td>75 ± 4</td>
<td>268 ± 38</td>
</tr>
<tr>
<td>8</td>
<td>cis-1,2-cyclohexane-dicarboxylic acid</td>
<td>CHDCA</td>
<td><img src="image8.png" alt="Image" /></td>
<td>18 ± 1</td>
<td>—†</td>
</tr>
<tr>
<td>9</td>
<td>Phthalic acid</td>
<td>PA</td>
<td><img src="image9.png" alt="Image" /></td>
<td>56 ± 2</td>
<td>—†</td>
</tr>
<tr>
<td>10</td>
<td>cis-(2R,3S)-2,3-piperidine dicarboxylic acid</td>
<td>cis-PDA</td>
<td><img src="image10.png" alt="Image" /></td>
<td>99 ± 7</td>
<td>222 ± 31</td>
</tr>
<tr>
<td>11</td>
<td>L-malic acid</td>
<td>LMA</td>
<td><img src="image11.png" alt="Image" /></td>
<td>69 ± 4</td>
<td>ND</td>
</tr>
</tbody>
</table>

ND, not determined.

*The attractant concentration eliciting the half-maximum response in FRET measurements.
†No. 8 and 9 compounds induce no chemotactic response in cells upon binding to Tar.
‡No.10 compound was tested as an analog of the no. 8 compound CHDCA.
§No.11 compound was tested as an analog of Asp.
AMA, which serves as the positive control, the cells sensed its gradient established by diffusion, moved through region 1 of the microchannel, and finally accumulated in region 2 (analysis region). Compared with the blank buffer, high fluorescence intensity in region 2, as in the case for AMA (Fig. 1B), indicated strong chemotactic response. Thus, the fluorescence intensity in region 2 indicated whether and how strong the corresponding test compound behaved as an attractant.

Five compounds, guanidinosuccinic acid (GSA), AMPA, formimino-L-aspartate (FIA), N-methyl-L-aspartate (NMA), and N-formyl-L-aspartate (NFA) behaved as attractants for Tar, with a compound source concentration of 10 mM (Figs. 1B and 2A and B). Another compound, (2-imino-4-oxo-thiazolidin-5-yl)-acetic acid (IOTA) showed attractant function at a compound source concentration of 100 mM (Fig. 2A and B). Purity analysis ruled out possible contamination of Asp in the compound samples (SI Appendix, Fig. S2). All these six novel attractants showed binding to Tar in the ITC experiments. N-methyl-DL-aspartate was previously reported as an attractant to a mutant strain of E. coli (26). By using the highly sensitive microfluidic device, we found that NMA is an attractant specific for Tar. We also measured E. coli responses to different source concentrations of these novel attractants. The responses of cells were concentration dependent (SI Appendix, Fig. S3 A–F).

To test whether these new compounds may also be sensed by other chemoreceptors, we used the mutant strain RP2361, which has all the chemoreceptors except Tar. As shown in Fig. 2C, RP2361 did not respond to any of the new compounds, showing that the new attractants work exclusively through Tar.

However, not all of the novel Tar binding molecules induced a chemotactic response. Compounds cis-1,2-cyclohexanedicarboxylic acid (CHDCA) and phthalic acid (PA), both binding to Tar, did not attract cells even when the source concentration was as high as 0.1 M (SI Appendix, Fig. S3 G and H). These two compounds have no repellent effect on Tar either, as verified by using another recently reported microfluidic device (30) optimized for detecting repellents as well as attractants. The discovery of these “futile” binding molecules provides a useful probe to understand the relation ship of chemoeffector binding and chemotaxis signaling.

In Vivo Responses to Novel Chemoeffectors Measured by FRET. To monitor the intracellular kinase activities, we measured cell intracellular response by using an in vivo assay based on FRET between CheY-YFP and CheZ-CFP (31). The Tar-only strain that expresses the CheY-YFP/CheZ-CFP pair and Tar receptor was stimulated by stepwise addition or removal of novel attractants at various concentrations (SI Appendix, Fig. S4 A–G). Following the same procedure reported in previous works (31, 32), the dose–response curves for each new attractant were measured and are shown in Fig. 3A. The EC50 value (i.e., the attractant concentration eliciting the half-maximum response) for each attractant was calculated. AMA was used as a control. Note that the Tar-only strain has more sensitive responses compared with the WT strain possessing all the chemoreceptors. For AMA, the EC50 value obtained here is 0.54 μM (for the Tar-only strain), whereas that from the previous study using the WT strain was 2.2 μM (33). The EC50 values for all of the novel attractants are in the micromolar or submillimolar range (Table 1), which are much lower than the Kd measured by ITC. Although we do not have an exact explanation for the large difference between the EC50 values and the Kd values, it is worth noting that the Kd value characterizes only the in vitro binding affinity of the ligand to the purified Tar periplasmic domain whereas the EC50 values depend on the in vivo kinase activity of the intact receptor complex, and they can be influenced by receptor clustering, which can further amplify signals resulting from ligand binding (31, 34). As a control, the Tsar-only strain (expressing the Tsar receptor, which is the chimera receptor replacing the periplasmic domain of Tar with the periplasmic domain of Tsr) showed no FRET response to the novel attractants, confirming that the attractant response was mediated by the Tar periplasmic domain (SI Appendix, Fig. S4H).

We also measured the responses of the Tar-only strain to CHDCA and PA. No FRET signal changes were observed, which further verified that neither of them are attractants or repellents (Fig. 3B and SI Appendix, Fig. S4I).

Antagonists Compete with Attractants for Binding. To understand the function of the antagonists, we used CHDCA to compete with AMA for Tar binding in several assays. By using ITC, we measured AMA binding with the purified Tar periplasmic domain in the presence of different concentrations of CHDCA. The measurements showed that the heat released from the binding of AMA to Tar was reduced depending on the CHDCA concentration (Fig. 4A). We also measured the intracellular kinase response to AMA in the presence of CHDCA by using FRET. With 1 mM background CHDCA, the dose–response curve of AMA shifted to higher concentrations, and the apparent EC50 increased from 0.5 μM to 1.2 μM (Fig. 4B and SI Appendix, Fig. S4I). Finally, E. coli cells were tested for their responses to AMA gradients with and without 1 mM uniform CHDCA by using the microfluidic assay. The mean speed and the mean angular speed of UU1624 cells that swimming in the blank buffer and in the ambient 1 mM CHDCA

Fig. 2. Responses of different E. coli strains to the novel attractants. The fluorescence intensities in region 2 emitted by different strains responding to novel attractants were normalized against the fluorescence intensity of cells responding to blank buffer. The relative intensities in region 2 emitted by RP437 (A) and UU1624 (B) are significantly stronger than the blank buffer, but those emitted by RP2361 (C) are similar to the blank buffer (mean ± SEM; n ≥ 2). As a control, the responses of RP437 and UU1624 to AMA, as well as RP2361 to -serine (Glu), were also measured. The source concentrations were 1 mM for AMA and Ser; 10 mM for AMPA, GSA, FIA, NMA, and NFA; and 100 mM for IOTA. The compound concentration range for cells in region 1 was 0% to 50% of the source concentration (*Significance at P < 0.05 vs. blank buffer by one-way ANOVA).
we tested the effect of an CHDCA analog, cis-PDA, on the binding affinity of Tar in the periplasmic domain measured by FRET. Cell responses to attractant AMA and repellent NiCl₂ were measured as the controls.

were almost the same during the experimental period, indicating that 1 mM CHDCA has little effect on the vitality and motility of cells (SI Appendix, Table S2). We found that cells showed weaker chemotactic responses to the AMA gradient with 1 mM CHDCA than without it (Fig. 4C). Higher concentrations of CHDCA were not investigated because of possible changes in the intracellular pH. All these results show that CHDCA functions as an antagonist by competitively binding to the attractant binding site in the Tar receptor.

As shown in Fig. 5, molecular docking analysis predicted that CHDCA interacts with the Tar residues R64, R69, and R73', key residues reported for Asp binding (21, 24). To confirm these interactions, we made mutations R64A, R69D, and R73'A in Tar and measured their binding affinities with CHDCA by ITC. These mutations degraded Tar’s binding affinity to CHDCA, with $K_d$ values of 133 ± 35 mM for R64A and >200 mM for R69'D and R73'A, confirming the critical roles of these residues in CHDCA binding.

Converting an Antagonist to an Attractant. To understand the key molecular features that differentiate attractants from antagonists, we compared their chemical structures and interaction patterns with Tar. All the attractant molecules possess at least one N-H group placed at a similar position as in Asp, whereas the two antagonist compounds do not. The N-H groups in the attractant molecules all form hydrogen bonds with the main-chain carbonyl group placed at a similar position as in Asp, whereas the two positive charged residues to negative ones may produce Tar and measured their binding affinities with CHDCA by ITC. Previous work reported that LMA is an attractant for E. coli and Salmonella (6, 35), although the detailed molecular interactions are not clear. Our ITC study shows that LMA can bind to the Tar periplasmic domain (Table 1 and SI Appendix, Table S1 and Fig. S1A). Microfluidic experiments show that only the strains possessing functional Tar (RP437 and UU1624) are attracted by LMA (SI Appendix, Fig. S5 F–H).

Taken together, our study indicates that a hydrogen bond donor group at the appropriate positions (near Y149 and Q152 of Tar) serves as the signaling trigger in attractant molecules.

Rational Design of Tar Specificity. Based on the differences between attractants and antagonists analyzed here, we can classify the interactions between the receptor and chemoeffectors into two types: (i) those that contribute mainly to binding, and (ii) those that contribute to binding and signaling. For the molecules we tested, the interactions of the chemoeffector’s carboxyl groups with the binding pocket residues, especially R64, R69’, and R73’ (region I of the binding pocket), are type I interactions, whereas those of the chemoeffector’s N-H or O-H groups with the main-chain carbonyls of Y149 and/or Q152 on the α4 helix (region II of the binding pocket) are type II interactions. Both types of interactions contribute to the binding free energy, but type II interactions directly trigger downstream signaling. These two types of interactions act together in concert to induce attractant signaling. Based on these understandings, the binding pocket of Tar can be rationally designed to sense novel ligands by keeping interactions with region II while changing region I to form new interactions with the novel ligands. As Tar has the highest sensitivity to the acidic amino acid Asp, as an example of specificity engineering, we redesigned Tar to recognize Arg, a basic amino acid that cannot be sensed by WT Tar (6, 25) (SI Appendix, Fig. S6A). As the amino group in Arg can form hydrogen bonds with the region II of the Tar binding pocket, the main redesigning task is to make the region I in Tar favorably interact with Arg side chain. The R69’ and R73’ residues in the WT Tar binding pocket interact favorably with Asp. Changing these two positive charged residues to negative ones may produce Tar mutants that can bind to Arg. Following this strategy, we prepared...
three Tar mutants—R69′E, R73′E, and R69′ER73′E—and tested their responses to Arg.

Because we need to test response to only a single ligand, a simpler microfluidic device developed previously (30) was used to detect the responses of Tar mutant strains to Arg gradients (as detailed in SI Appendix). We found that the mutant strain R69′ER73′E in the observation channel could sense the gradient of Arg and accumulate in the analysis region (Fig. 6A). The relative fluorescent intensity in the analysis region increased with time, and the response was concentration dependent (Fig. 6B). The mutant strains R69′E and R73′E cannot sense Arg, probably because the remaining R69′ or R73′ destabilizes Arg side chain binding. In addition to Arg, the mutant strain R69′ER73′E can still sense Asp, although with a much lower ability (SI Appendix, Fig. S6B), and does not sense the other 18 types of amino acids. The mutant receptor R69′E, R73′E, and R69′ER73′E have similar expression level with WT Tar (SI Appendix, Fig. S6C and D). Molecular docking analysis showed that the side chain of Arg interacts with the side chains of E69′ and E73′, and its amino group forms hydrogen bond with main-chain carbonyl group of Y149 (Fig. 6C). These results demonstrate the validity of our rational design strategy.

Summary and Discussion

In this paper, we combined computational and experimental methods to search for novel chemoeffectors for Tar. Virtual screening, an approach that has been used extensively in the drug discovery process (36, 37), was used to search for Tar binding compounds from a large chemical library. Six of the eight Tar binding compounds were attractants. The other two compounds behaved as antagonists. Borrok et al. identified an antagonist of the periplasmic glucose (or galactose) binding protein, which can inhibit glucose chemotaxis in E. coli via the chemoreceptor Trg (38). In the present study, we identified the first examples of antagonists that function by directly binding to the E. coli chemoreceptor. As chemotaxis is an important virulence factor for pathogenic bacteria, inhibiting chemotaxis of pathogenic bacteria is a potent therapeutic strategy to prevent or cure disease (39, 40). Discovering antagonists that directly interact with chemoreceptors may provide useful clues for antagonist design to inhibit the chemotaxis of pathogenic bacteria.

Our combined computational and experimental approaches can be extended to discover novel chemoeffectors for other chemoreceptors.

Comparison between attractants and the antagonist molecules discovered here provided new molecular insights in chemoreceptor signaling. Previous work provided clues about the possible role of Asp amino group in chemotaxis signaling (10, 14, 15, 35, 41). Our study indicated explicitly that the interactions between the N-H group of chemoeffectors with carbonyls of Y149 and/or Q152 are crucial for signaling. Furthermore, we showed that the N-H group is not unique in serving as the signaling “trigger,” as other hydrogen bond donor groups such as the O-H group also work. Based on our study, the molecular interactions of a chemoeffector with Tar can be divided into two types: type I for binding and type II for binding and signaling. It would be interesting to study whether this molecular picture holds true for other chemoreceptors.

The structure-based rational design approach has been widely used in protein engineering and drug design. However, for chemoreceptor specificity change, rational design was difficult, as ligands designed to bind to Tar may not necessarily induce the right conformational changes that lead to chemotaxis (26). In this work, rational design of Tar to sense Arg was carried out by keeping the type II interactions intact while making changes to enhance/enable the type I interactions. By using this strategy, we designed a Tar variant that can positively sense Arg by only mutating two residues in the binding pocket. The successful Tar specificity redesign verifies our molecular understanding of the binding and signaling process in Tar. Although further optimization design or mutagenesis selection is needed to increase its sensitivity, our study demonstrated the proof of principle for rational design of chemoreceptor specificity, which can be extended to redesign specificities of other chemoreceptors (such as Tsr) or Tar to other nonnative chemoeffectors.

Materials and Methods

Strains and Plasmids. Information regarding the genotypes, phenotypes, and sources of the bacterial strains and plasmids used in this study are listed in SI Appendix, Table S3.

Virtual Screening. We modeled the complex structure of the E. coli Tar periplasmic domain with Asp based on the crystal structure of Salmonella Tsr (SI Appendix, SI Materials and Methods). The AutoDock program (version 4.0.1) was used for the docking screening (42). Molecules with molecular weight <300 Da in the MDL Available Chemical Directory were selected for the docking study (149,063 molecules).

Fig. 6. Rational design of Tar to chemotaxis to Arg. (A) Responses of the mutant strain R69′ER73′E to a gradient of Arg recorded by fluorescence microscopic images at different times (Scale bar, 100 μm.). The Arg gradient is from 0 to 4 mM across the observation channel. The response is characterized by the fluorescence intensity in the analysis region (yellow rectangle) of the observation channel. (B) Responses of the mutant strain R69′ER73′E to different Arg gradients. The concentrations given are the maximum Arg concentrations at the right end of the observation channel (the Arg concentration at the left end is 0 mM). The fluorescent intensities in the analysis region relative to the initial intensity were plotted as a function of time for different Arg gradients. (C) Interactions of Arg with the binding pocket of Tar variant R69′ER73′E predicted by molecular docking.
ITC Measurements. ITC was performed at 25 °C on a MicroCal ITC200 calorimeter (GE Healthcare) to measure the binding affinities of compounds with the purified WT Tar ligand binding domain and those of CHDCA with the mutant proteins. All ITC data sets were analyzed by using the Origin software package supplied by MicroCal. See SI Appendix for details.

Microfluidic Experiments. Details of the design, fabrication, and calibration of the device are shown in Fig. 1A and SI Appendix. For novel chemoeffector receptor selection, compound solution was loaded into individual peripheral holes and stable linear gradients were generated. The prepared E. coli cells expressing GFP proteins were loaded into the central hole. One to two hours after loading the cells, the responses of different compounds (or different concentrations of the same compound) were observed by using an inverted microscope at 30 °C. The fluorescence signals were measured to quantify cell densities in region 2. For the antagonist function detection, the test device and control device were filled with 1 mM CHDCA and blank buffer, respectively. Differences between the chemostatic responses of cells to the AAM gradient with or without CHDCA were observed and analyzed according to the accumulated intensities in region 2. All data were analyzed by using ImageJ (National Institutes of Health [NIH]).

FRET Measurements. The cell preparation and FRET measurement were performed as described before (31, 32). Cells of the Tar-only strain or the Tar-only strain were stimulated with chemoeffectors of interest. The fluorescence signals were recorded in the cyan and yellow channels. Data were analyzed as previously described and fit to a Hill model to obtain EC50 for each novel attractant (31).

Rational Design of Tar. Tar mutants R69E, R73E, and R69E73E were generated by site-directed mutagenesis by using the plasmid pPD12 encoding WT Tar as the template. The plasmid pPD12 or its mutants together with the GFP encoding plasmid pCM18 were transformed into E. coli strains (31). The responses of E. coli strains expressing WT Tar, R69E, R73E, or R69E73E mutant Tar proteins to the gradients of arginine were measured by using a previously described microfluidic device (30) (SI Appendix). The images were captured every 5 s for 90 min.

Other detailed experimental and computational procedures are described in SI Appendix, SI Materials and Methods.

ACKNOWLEDGMENTS. We thank Prof. J. S. Parkinson (University of Utah) for providing the RP437, UU1250, UU1624, and RP2361 strains and the plc113 plasmid; Prof. M. Golian (University of Pennsylvania) for providing the pD10 and pD12 plasmid; and Prof. S. Molin (Technical University of Denmark) for providing the pCM18 plasmid. This work was supported in part by National Natural Science Foundation of China 21173013 and 11021463, Ministry of Science and Technology of China 2009CB18500, and NIH Grants GM081747 (to Y.T.) and GM082938 (to V.J. and V.S.).
**Supplementary Information for**

“Discovery of Novel Chemoeffectors and Rational Design of *Escherichia coli* Chemoreceptor Specificity”

Shuangyu Bi\(^a,b\), Daqi Yu\(^a,b\), Guangwei Si\(^b,c\), Chunxiong Luo\(^b,c\), Tongqing Li\(^a,b\),
Qi Ouyang\(^b,c,d\), Vladimir Jakovljevic\(^e\), Victor Sourjik\(^e,f\), Yuhai Tu\(^b,g,1\),
and Luhua Lai\(^a,b,d,1\)

\(^a\) BNLMS, State Key Laboratory for Structural Chemistry of Unstable and Stable Species, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China.

\(^b\) Center for Quantitative Biology, AAIS, Peking University, Beijing 100871, China.

\(^c\) College of Physics, Peking University, Beijing 100871, China.

\(^d\) Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China.

\(^e\) Zentrum für Molekulare Biologie der Universität Heidelberg, DKFZ-ZMBH Alliance, Im Neuenheimer Feld 282, 69120 Heidelberg, Germany.

\(^f\) Max Planck Institute for Terrestrial Microbiology, D-35043 Marburg, Germany.

\(^1\) Max Planck Institute for Terrestrial Microbiology, D-35043 Marburg, Germany.

\(^g\) IBM T. J. Watson Research Center, Yorktown Heights, New York 10598, USA.

\(^1\) Corresponding authors:

Luhua Lai
College of Chemistry and Molecular Engineering, Peking University
Beijing 100871, China.
E-mail: lhlai@pku.edu.cn; Tel: 86-10-62757486

Yuhai Tu
IBM T. J. Watson Research Center, Yorktown Heights
New York 10598, USA
E-mail: yuhai@us.ibm.com; Tel: (914)945-2762
SI Results

**Binding affinity measurements by ITC.** ITC was performed at 25 °C on a MicroCal ITC200 calorimeter (GE Healthcare) to measure the binding affinities of compounds with the purified Tar periplasmic domain. Titrations were carried out in a buffer of 200 mM phosphate buffered saline, 200 mM NaCl. The results of ITC are shown in Table S1 and Fig. S1. As most of the compounds bind weakly, we followed the guidelines for measuring low affinity ligand binding using ITC (1-2, www.gelifesciences.com/microcal). The highest possible concentrations of the protein and compounds permitted by solubility were used while keeping a constant pH. Eight of the eleven titrations reached over 80% receptor saturation and three of them were around 70%. The resulting \( c \) values (\( c = nK_d[M_0] \), \( M_0 \) is the concentration of proteins in the cell, \( n \) is the number of sites) were between 0.3 and 0.005, which were above the recommended lowest \( c \) value in ITC studies (1).

**Novel chemoeffector identified by microfluidic experiments.** We discovered six attractants using microfluidic experiments. Purity analysis eliminated possible contamination of Asp in the compound samples (Fig. S2). We measured the responses of *E. coli* RP437 cells to different source concentrations of novel attractants (Fig. S3A-F). CHDCA and PA, both of which bind with Tar, did not attract cells even at high source concentration of 0.1 M (Fig. S3G-H).

**FRET measurement of intracellular response to novel chemoeffectors.** As
shown in Fig. S4A-G, the Tar-only strain that expresses the wild-type Tar receptor, and the CheY-YFP/CheZ-CFP FRET pair were stimulated by stepwise addition or removal of attractants at indicated concentration. Upon addition of attractants, the FRET signal (the ratio of YFP/CFP) decreased, reflecting lowered kinase activity. The Tsar-only strain, which expresses the chimera receptor of Tar and Tsr had no FRET response when stimulated by the novel attractants (Fig. S4H). PA cannot induce the change of FRET signals, as shown in Fig. S4I. The repellents, such as nickel ion, have opposite effects to attractants (Fig. S4I).

The futile binders act as antagonists that compete with attractants for binding. To make sure the antagonist does not affect the cell’s swimming speed, the mean speed and mean angular speed of UU1624 swimming in the blank buffer and ambient 1 mM CHDCA were measured following previous study (3). Cells were tracked by Image J (National Institutes of Health). Data were analyzed according to previous method (3). We found that the mean speed and mean angular speed of UU1624 cells swimming in the blank buffer and in the ambient 1 mM CHDCA were almost the same during the experimental period, indicating that 1 mM CHDCA have little effect on the vitality and motility of cells (Table S2). We measured the influence of CHDCA to the intracellular response to the steps of AMA using FRET. We observed that 1 mM CHDCA could influence the FRET response of Tar-only strain to AMA. The difference of YFP/CFP change could be observed when adding AMA with 1 mM CHDCA together, as shown in Fig. S4J.
Converting an antagonist to an attractant. We measured the responses of *E. coli* cells to multiple concentration gradients of *cis*(2*R*, 3*S*)-2,3-piperidine dicarboxylic acid (*cis*-PDA) using the microfluidic device (Fig.1A) as well as the FRET measurement. *Cis*-PDA can attract cells possessing functional Tar (Fig. S5A-E). We also measured the responses of different *E. coli* strains to multiple concentration gradients of L-malic acid (LMA) using the microfluidic device (Fig.1A). LMA can attract cells possessing functional Tar (Fig. S5F-H).

Rational design of Tar for novel chemotaxis specificity. We redesigned chemoreceptor Tar to recognize L-arginine, a basic amino acid that cannot be sensed by the wild-type Tar (Fig. S6A). The Tar mutant R69'ER73'E can sense L-arginine as an attractant. Besides L-arginine, R69'ER73'E showed weak attractant response to L-aspartate, weaker than its response to arginine (Fig. S6B). We have also verified that the mutant receptor R69'E, R73'E and R69'ER73'E have similar expression level with the wild-type Tar by using Western Blot (Fig. S6C and D).

SI Materials and Methods

Strains, plasmids, and materials. Information regarding the genotypes, phenotypes, and sources of the bacterial strains and plasmids used in this study are listed in Table S3. Guanidinosuccinic acid, (±)-α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid, formimino-L-aspartate, and N-formyl-L-aspartate were purchased from Sigma Aldrich. N-methyl-L-aspartate was purchased from Acros Organics. *cis* 1,2-cyclohexanedicarboxylic acid and phthalic acid were purchased
from Alfa Aesar. (2-imino-4-oxo-thiazolidin-5-yl)-acetic acid was purchased from Matrix Scientific. cis-(2R,3S)-2,3-piperidine dicarboxylic acid was purchased from Beijing Repharma Co., Ltd. L-malic acid was purchased from J&K Scientific Ltd.

**Virtual screening for novel chemoeffectors.** As far as we know, there are at least three available crystal structures (4-6) for the Tar periplasmic domain with ligands in *Salmonella* and *E. coli*. Previous studies (6) showed little differences among these structure candidates in the binding interactions with aspartate. Currently, no crystal structure is available for the *E. coli* periplasmic domain of Tar with the ligand Asp bound. The available apo structure (7) or the pseudoligand-bound structure (8) does not reflect the specific Asp binding-induced conformational changes. The sequences between the periplasmic domain of Tar in *Salmonella* and *E. coli* share 66% identity with no gaps. We chose the newly published Tar receptor structure from *Salmonella* (PDB code: 1VLT) (6) as the template to build the *E. coli* receptor structure for the virtual screening. The sequence of the Tar periplasmic domain in *Salmonella* was mutated to that in *E. coli* using Scap (9), a program for side chain conformation prediction and residue mutation. Ninety among 284 residues were mutated in total. The mutated residues were all farther than 5 Å away from the binding pocket. The mutated structure was then optimized in CHARMM c33b1 (10). The ligands and water molecules were removed in the next steps. We modeled the structure of the *E. coli* Tar periplasmic domain based on the crystal structure of *Salmonella* Tar. The AutoDock program (version 4.0.1) was used for the virtual screening by docking (11). Molecules with molecular weight < 300 Da in the MDL
ACD were selected for the docking study (149,063 molecules). This molecular weight limit was set based on the size of the aspartate binding pocket. The top 10,000 compounds with the lowest estimated binding free energies lower than -5.5 kcal mol\(^{-1}\) were selected. Eighty compounds were selected manually and purchased for experimental studies.

**Clone and mutagenesis of the periplasmic domain of Tar.** The plasmid pMDL101 was constructed to express the *E. coli* Tar periplasmic domain. The coding sequence of residues 32-188 was amplified using PCR reaction from the pLC113, a plasmid encodes wild-type full length tar. Two oligonucleotides were used for the PCR reaction, introducing restriction sites *Nde*I at the 5’-end, *Bam*HI and a stop codon TGA at the 3’-end. The amplified fragment was digested and ligated with pET-28a (His-Tag containing expression vector; Novagen) to create plasmid pMDL101. Mutants of the periplasmic domain of Tar, R64A, R69’D, and R73’A, were generated by Muta-direct™ site-directed mutagenesis kit (SBS Genetech). The plasmid pMDL101 was the template for the mutagenesis. All mutants were verified by DNA sequencing.

**Expression and purification of wild-type and mutant Tar periplasmic domain.** The plasmid pMDL101 and the mutants were transferred into *E. coli* BL21 (DE3) to express the target proteins. Cells with the plasmids were inoculated at 37 °C in Luria Bertani (LB) medium supplemented with 30 μg ml\(^{-1}\) kanamycin. When OD\(_{600}\) value reached 0.6-0.8, 0.5 mM isopropyl β-D-thiogalactopyranoside (IPTG) was
added to induce the expression of target proteins. The induction time for the expression of the wild-type periplasmic domain and the mutant R64A was 6 hours at 26 °C, whereas that for the R73’A, and R69’D mutants was 8 hours at 18 °C. Cells were lysed by sonication in the sonication buffer (50 mM Tris-HCl, pH 8.0 (12), 200 mM NaCl, 2 mM PMSF, 10 mM iminazole). Cell debris was pelleted by centrifugation and the supernatant was applied to 5 ml HisTrap™ HP column (GE Healthcare) equilibrated with buffer A (50 mM Tris-HCl, pH 8.0, 200 mM NaCl, and 10 mM iminazole). A liner gradient of buffer B (50 mM Tris-HCl, pH 8.0, 200 mM NaCl, 500 mM iminazole) was applied to eluted the target proteins. The peak fraction containing target proteins from the HisTrap™ HP column was applied to a 120 ml Sephacryl S-200 HR (GE Healthcare) equilibrated with buffer C (50 mM Tris-HCl, pH 8.0, 200 mM NaCl). Proteins were eluted with the same buffer and analyzed by SDS-PAGE.

**Design, fabrication, and calibration of the microfluidic device for the chemoeffector selection**

**Design of the device.** A specially designed (unpublished previously) microfluidic device was used in this study. Kim et al. recently reported a similar design (13); however, our design is more suitable for large-scale screens for novel chemoeffectors. A schematic representation of the device is shown in Fig. 1A. It consists of a central hole (diameter, 5.0 mm) with 12 circular peripheral holes (diameter, 3.0 mm) around it, connected by means of 12 microchannels to the central hole. Each microchannel is divided into three regions. Regions 1 and 3 have the same dimensions: length 1.5 mm,
width 50 μm, and height 5 μm. Region 2, also termed the analysis region, is 500 μm in length, 200 μm in width, and 25 μm in height. In our design, similar to previously published work (13), agarose gel was used to avoid the convection current and allow the diffusion of small molecules to generate stable linear concentration gradients. But one of the new advantages is that, in our design, the agarose plug can be constructed in either Region 1 or Region 3, so the concentration gradients can be generated either from the peripheral holes to the central hole or from the central hole to the peripheral holes. The two directions of concentration gradients broaden the applications of the device. When agarose plugs were constructed in Region 3, the peripheral holes are the sources for compounds, and the central hole is used as the cells source. Compounds diffused from the peripheral holes to the central hole along the microchannels. This process enabled us to investigate the responses of cells with the same conditions to different chemicals or different chemical concentrations at the same time. When agarose plugs were constructed in Region 1, the central hole is filled with attractant solution, and the peripheral holes are filled with cells of the same or different conditions. Compounds diffused from the central hole to the peripheral holes. This process allowed us to observe the responses of cells under different environmental conditions to the same chemical concentration gradient simultaneously.

**Fabrication of the device.** Standard soft lithography procedure (14) was used to fabricate microfluidic devices. The well prepared silicon master with the features described above was used to make the mold of microfluidic devices. Polydimethylsiloxane (PDMS, RTV615 044-Pail Kit, crosslinking agent: silicone
potting compound ~1:7, Momentive Specialty Chemicals Inc.) was poured on the master, cured at 75 °C and peeled off. Holes were punched into the patterned PDMS at the positions of the central hole and peripheral holes using cutting tips with tip diameters of 5.0 mm and 3.0 mm (Harris Uni-Core™). The microfluidic devices were bonded to clean microscope cover classes (Fisher Scientific) after treated with oxygen plasma for 1 min in a plasmacleaner (Harrick Plasma) to create hydrophilic devices. Then agarose plugs can be constructed. For the novel chemoeffectors screening, 3% agarose solution incubated at 75 °C was loaded into each peripheral hole at the room temperature. Agarose solution flew into Region 3 and solidified at the entrance of Region 2 (analysis region). For the antagonist function detection, 3% agarose was loaded into the central hole. The agarose solution can flow along Region 1 and stop at the entrance of Region 2. The reasons for the agarose stopped at the interface of Region 2 and Region 1 (Region 3) are described in another study (15). Minimal salt buffer (also termed blank buffer; 10 mM PBS, 0.1 mM EDTA, 0.01 mM L-methionine, 10 mM sodium DL-lactate, pH 7) was loaded into the central hole and the peripheral holes to fill the device with buffer.

**Calibration of the device.** After the device was well fabricated, fluorescein solution was loaded in the peripheral hole to the final concentration of 100 μM and let it diffuse in wet environment for 15 hours. Then the fluorescence signals in the microchannel were observed using a Nikon Ti-E inverted microscope (Nikon Instruments) with a QuantEM512SC CCD (Roper Scientific). Fluorescence images were recorded using 10× objective lenses. As shown in Fig. S7A, the concentration
gradient of fluorescein is linear in Region 1 and Region 3 at the steady state. The chemical concentration in the microchannel can be described by the one dimensional diffusion equation. At the steady state, the concentration gradient \( \nabla C = C_0/l \) is linear, where \( C_0 \) is the source concentration of compound, and \( l \) is the length of the microchannel (total length of Region 1 and Region 3, 3 mm here). The fluorescent signal in Region 2 is stronger than Region 1 and Region 3, because the height of Region 2 is five times larger. The gradient in Region 2 is not obvious, theoretical explanations are as follows: The equation for the diffusion of compounds is, 
\[
\Delta Q' = -D(\frac{dc}{dx})s\Delta t ,
\]
where \( s \) is the cross-sectional area, \( D \) is the diffusion coefficient. At the steady state, \( \Delta Q' \) is the same along microchannel. So in Region 1 (Region 3) and Region 2, 
\[
\Delta Q' = -D(\frac{dc}{dx})s_1\Delta t = -D(\frac{dc}{dx})s_2\Delta t ,
\]
\[
(\frac{dc}{dx})_1 / (\frac{dc}{dx})_2 = s_2 / s_1 ,
\]
where \( s_1, s_2 \) is the cross-sectional area of Region 1 (Region 3) and Region 2 respectively. That is, the gradient is in inverse proportion to \( s \). The cross-sectional area of Region 1 (Region 3) is 19 times larger than that of Region 2, so the gradient in Region 2 is only 1/20 of the gradient in Region 1 (Region 3). Because the concentration change is very small in Region 2, we can estimate that the concentration in Region 2 is almost half of the source concentration of compound, here 50 \( \mu \)M of fluorescein. If we do not consider Region 2, the time \( t_1 \) for the concentration gradient to reach a steady state can be approximated 
\[
t_1 = \frac{l^2}{D},
\]
where \( D \) is the diffusion coefficient of compounds. It is assumed that the agarose gel has the same diffusion coefficient as water. According to this theoretical estimation, the time for fluorescein to diffuse through the microchannel is about \( l^2/D = 5 \) h, where \( l = 3 \)
mm (total length of Region 1 and Region 3), \( D = 500 \ \mu m^2 \ s^{-1} \). This is only a simplified method to estimate \( t_1 \), because the different dimension of Region 2 will influence the diffusion time. But 15 hours is sufficient to generate stable linear gradient, as seen in Fig. S7B and C. The linear concentration gradient is stable during the period of experiment, as shown in Fig. S7B and C. The stable time \( t_2 \) for the linear concentration gradient can be estimated by the equation \( \Delta cV = Ds_1t_2dc/dl \), where \( V \) is the volume of the central hole or the peripheral hole, \( s_1 \) is the cross-sectional area of Region 1 or Region 3. If the compounds diffuse along the direction from the peripheral hole to the central hole, at the steady state, the concentration in the central hole is zero. Assuming that after \( t_2 \), the concentration in the central hole changes from zero to \( C_0/10,000 \), the equation can be rewrote as \( t_2 = V_{\text{center}}l/10,000Ds_1 = 26.1 \ \text{h} \), where \( V_{\text{center}} = \pi r^2h \approx 39.25 \ \text{mm}^3 \ (h \approx 2 \ \text{mm}) \), \( l = 3 \ \text{mm} \), \( D = 500 \ \mu m^2 \ s^{-1} \), \( s_1 = 250 \ \mu m^2 \). If the compounds diffuse along the direction from the central hole to the peripheral hole, the time for the concentration in the peripheral hole changes from zero to \( C_0/10,000 \) is \( t_2 = V_{\text{around}}l/10,000Ds_1 = 6.52 \ \text{h} \), where \( V_{\text{around}} \approx 14.13 \ \text{mm}^3 \). So, the concentration gradients are very stable during the entire period of experiments. This design can have good function even if the concentration changes from zero to \( C_0/10 \). The time \( t_2' \) for the concentration changes from zero to \( C_0/10 \) is 1,000 times longer than \( t_2 \). So, the device keeps good function during very long period of time. Since a linear gradient was established, the compound concentration range in Region 1 was about 0–50% of the source concentration in the peripheral hole.

**Cell and compound preparation for the microfluidic experiments**
**select novel chemoeffectors.** Single colonies of *E. coli* strains RP437, UU1624, and RP2361 expressing GFP proteins were grown at 30 °C overnight in Tryptone Broth medium (TB, 10 g L⁻¹ tryptone and 5 g L⁻¹ NaCl) supplemented with 100 μg ml⁻¹ ampicillin. The grown cultures were then diluted with 100 times by fresh TB medium containing antibiotics and grown at 30 °C until OD₆₀₀ had reached ~0.3. Cells were harvested by centrifugation at 3,000 rpm for 5 min at room temperature. The supernatant was discarded and the pelleted cells were washed twice with minimal salt buffer to remove remaining TB medium (16-17). Finally, cells were resuspended in minimal salt buffer. All the compounds used in the microfluidic experiments were dissolved in minimal salt buffer (pH7).

**Microfluidic experiments to detect cell responses to L-arginine.**

The cell preparation was similar as described above, except that the *E. coli* strains UU1250 expressing wild-type or mutant Tar receptor were grown in TB medium supplemented with 100 μg ml⁻¹ ampicillin, 30 μg ml⁻¹ chloramphenicol and 500 μM IPTG. The responses of *E. coli* strains expressing wild-type Tar, R69’E, R73’E or R69’ER73’E mutant Tar to the gradients of arginine were measured by microfluidics. Since we only need to detect response to a single ligand (Arg), we did not use the multi-channel radial design developed for parallel compound screening shown in Fig. 1A. Instead, a previously reported simpler microfluidic device was used (15). This microfluidic device is more suitable for detecting the responses of different strains to the same compound simultaneously. The design, fabrication, and calibration of this microfluidic are described in detail in (15). Experiments were operated as described in
In short, a sink side pore and a source side pore are connected by an observation channel and agarose gel channels. The prepared *E. coli* cells were added in the sink side pores of the microfluidic device. It takes roughly one hour for the cells to diffuse into the observation channel and reach a steady state. Then, we added compound solutions in the source side pores. The compound will diffuse into the observation channel and establish a concentration gradient gradually. After adding the compound, images were recorded to detect the cell fluorescent intensities in the observation channel. The images were captured every 5 min for 90 min. The responses of cells were characterized by the fluorescence intensities in the analysis region (yellow rectangle) of the observation channel (Fig. 6A). Data were analyzed by Image J.

**Expression level analysis using Western Blot.** The expression level of mutant Tar receptors expressed from pPD12 derivatives were measured in UU1250 using Western Blot. Strains were grown and suspended in minimal salt buffer as described above. Cells were lysed by boiling and subjected to electrophoretic separation using SDS-PAGE. Proteins were transferred from the gel to the nitrocellulose membrane, treated by rabbit polyclonal anti-Tar antibody and detected by goat anti-rabbit IgG (AP) secondary antibody. Intensity profiles in individual lanes were analyzed using Image J. The relative amounts of Tar proteins in different lanes were compared by using a chromosomally encoded protein as the internal standard. The expression level of Tar mutant R69’E, R73’E or R69’ER73’E was compared with that of wild-type Tar (expressed from pPD12)
**Statistical analysis.** Statistical analysis was done using GraphPad Prism 5.04 (GraphPad Software). For statistical comparison, one-way ANOVA followed by *Dunnett's* test, or a Student’s two-tailed unpaired *t*-test were used. *P* < 0.05 was considered significant.

**SI References**


**Fig. S1.** Binding affinity measurements by ITC. Binding affinity of Tar periplasmic domain with (A) AMA, (B) AMPA, (C) FIA, (D) GSA, (E) NMA, (F) NFA, (G) IOTA, (H) CHDCA, (I) PA, (J) cis-PDA, and (K) LMA.

**Fig. S2.** Results of ESI$^+$ experiments for the novel attractants (A) GSA, (B) FIA, (C) NMA, and (D) NFA. The molecular weight for L-aspartate should be: $M = 133.10$, $M + H = 134.10$, $M + Na^+ = 156.10$. There is no peak at those three positions, so the four attractants do not contain aspartate contamination.
Fig. S3. Responses of *E. coli* cells to different source concentrations of novel attractants, CHDCA and PA. RP437 responses to different source concentrations of (A) AMA, (B) GSA, (C) FIA, (D) NMA, (E) NFA and (F) AMPA were measured in
the analysis region (Region 2) (mean ± SD, n = 2). CHDCA and PA cannot attract (G) RP437 and (H) UU1624 even if the source concentrations were in the order of 0.1 M. The fluorescent intensities in the analysis region for these two compounds were similar with that in the blank buffer (mean ± SD, n = 2). The cell concentration and exposure time were determined by each experiment.
**Fig. S4.** Intracellular responses of *E. coli* to novel chemoeffectors observed by FRET. The Tar-only strain that expresses CheY-YFP and CheZ-CFP pair were stimulated by stepwise addition or removal of attractants (A) AMA, (B) AMPA, (C) FIA, (D) GSA, (E) NMA, (F) NFA, (G) IOTA. (H) The Tsar-only strain cannot response to novel attractants. (I) PA cannot induce the change of FRET signals. (J) FRET measurement of the influence of CHDCA on response to steps of AMA.
Fig. S5. Responses of *E. coli* to *cis*-PDA and LMA. Relative fluorescence intensities in the analysis region emitted by stratins (A) RP437, (B) UU1624, and (C) RP2361 responding to different source concentrations of *cis*-PDA (mean ± SEM, n = 3). Cell responses to buffer were set to one. Strains expressing a functional Tar receptor are attracted by *cis*-PDA in the microfluidic experiments. (D) FRET measurement, plotted as a change in YFP/CFP ratio, to stepwise addition or
subsequent removal of cis-PDA. (E) Dose-response curve for cis-PDA calculated from (D). Relative fluorescence intensities in the analysis region emitted by strains (F) RP437, (G) UU1624, and (H) RP2361 in response to different source concentrations of LMA were recorded (mean ± SEM, n = 2). Strains expressing functional Tar receptors are attracted by LMA. Asterisks in (A)-(B), and (F)-(H) indicate statistical significance (P < 0.05) compared to the blank buffer by one-way ANOVA.
**Fig. S6.** (A) The *E. coli* strain with wild-type Tar cannot sense L-arginine as an attractant. (B) The Tar mutant R69’ER73’E has stronger chemotaxis response to L-arginine than L-aspartate. (C) The expression level of wild-type and mutant Tar R69’E, R73’E and R69’ER73’E determined by Western Blot. A chromosomally encoded protein was used as the internal standard. UU1250 with pPD10 plasmid was the control. (D) R69’E, R73’E and R69’ER73’E have similar expression level with the
wild-type Tar (WT). The concentrations given in (A) and (B) are the maximum concentrations at the right end of the observation channel.

**Fig. S7.** Microfluidic device calibration. (A) The profile of the concentration gradient of fluorescein imaged in the microchannel after diffusing for 15 hours. The source concentration for the fluorescein is 100 µM. After 15 hours, the concentration gradient maintains linear and stable during the period of experiments. The coordinate of X-axis (B) -700-0 and (C) 0-700 correspond to the coordinate signed in (A).
<table>
<thead>
<tr>
<th></th>
<th>$[M_0]^*$</th>
<th>$[X_0]^{†}$</th>
<th>$K_a$</th>
<th>$K_d$</th>
<th>$c^{‡}$</th>
<th>Saturation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(mM)</td>
<td>(mM)</td>
<td>(M$^{-1}$)</td>
<td>(mM)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMA</td>
<td>0.34</td>
<td>10</td>
<td>1790 ± 116</td>
<td>0.559 ± 0.036</td>
<td>0.304</td>
<td>95</td>
</tr>
<tr>
<td>AMPA</td>
<td>0.3</td>
<td>20</td>
<td>237 ± 14</td>
<td>4.2 ± 0.2</td>
<td>0.036</td>
<td>83</td>
</tr>
<tr>
<td>FIA</td>
<td>1</td>
<td>350</td>
<td>30.1 ± 1.8</td>
<td>33 ± 2</td>
<td>0.015</td>
<td>91</td>
</tr>
<tr>
<td>GSA</td>
<td>0.85</td>
<td>300</td>
<td>14.2 ± 0.8</td>
<td>70 ± 4</td>
<td>0.006</td>
<td>81</td>
</tr>
<tr>
<td>NMA</td>
<td>0.5</td>
<td>200</td>
<td>37.8 ± 3.8</td>
<td>26 ± 3</td>
<td>0.009</td>
<td>88</td>
</tr>
<tr>
<td>NFA</td>
<td>1</td>
<td>300</td>
<td>14.4 ± 1.3</td>
<td>69 ± 6</td>
<td>0.007</td>
<td>81</td>
</tr>
<tr>
<td>IOTA</td>
<td>1</td>
<td>200</td>
<td>13.3 ± 0.7</td>
<td>75 ± 4</td>
<td>0.007</td>
<td>73</td>
</tr>
<tr>
<td>CHDCA</td>
<td>0.5</td>
<td>110</td>
<td>54.8 ± 3.2</td>
<td>18 ± 1</td>
<td>0.014</td>
<td>86</td>
</tr>
<tr>
<td>PA</td>
<td>1</td>
<td>250</td>
<td>17.7 ± 0.6</td>
<td>56 ± 2</td>
<td>0.009</td>
<td>82</td>
</tr>
<tr>
<td>cis-PDA</td>
<td>1</td>
<td>200</td>
<td>10.1 ± 0.7</td>
<td>99 ± 7</td>
<td>0.005</td>
<td>67</td>
</tr>
<tr>
<td>LMA</td>
<td>1</td>
<td>200</td>
<td>14.4 ± 0.9</td>
<td>69 ± 4</td>
<td>0.007</td>
<td>74</td>
</tr>
</tbody>
</table>

* $[M_0]$ is the concentration of proteins
† $[X_0]$ is the concentration of compounds
‡ $c = nK_a[M_0]$, $n = 0.5$
**Table S2** Mean speed and mean angular speed analysis of UU1624 cells swimming in the ambient blank buffer and 1mM CHDCA, pH 7

<table>
<thead>
<tr>
<th>Swimming time</th>
<th>5 min</th>
<th>60 min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swimming medium</td>
<td>Blank buffer</td>
<td>1mM CHDCA</td>
</tr>
<tr>
<td>Number of cells tracked</td>
<td>26</td>
<td>31</td>
</tr>
<tr>
<td>Tracking time (s)</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Mean speed (μm s(^{-1})) *,†</td>
<td>17 ± 6</td>
<td>17 ± 7</td>
</tr>
<tr>
<td>Mean angular speed (deg frame(^{-1})) *,†</td>
<td>53 ± 16</td>
<td>51 ± 15</td>
</tr>
</tbody>
</table>

* mean ± SD
† Frame interval: 0.07 s
Table S3 Strains and plasmids used in this study

<table>
<thead>
<tr>
<th></th>
<th>Genotype or phenotype</th>
<th>Description</th>
<th>Source or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasmids</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pLC113</td>
<td>tar Cam&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Expresses <em>E. coli</em> wild-type full-length Tar receptor</td>
<td>(18)</td>
</tr>
<tr>
<td>pMDL101</td>
<td><em>tar</em> (33-188) Kan&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Expresses <em>E. coli</em> wild-type Tar periplasmic domain, N-terminal His-tag</td>
<td>This study</td>
</tr>
<tr>
<td>GFP plasmid</td>
<td>gfpmut2 Amp&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Expresses GFP proteins</td>
<td>(19)</td>
</tr>
<tr>
<td>pCM18</td>
<td>gfpmut3 Cam&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Expresses GFP proteins</td>
<td>(20)</td>
</tr>
<tr>
<td>pVS88</td>
<td>cheZ-ecfp / cheY-yfp Amp&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Expresses FRET pair CheY-YFP and CheZ-CFP</td>
<td>(21)</td>
</tr>
<tr>
<td>pVS1092</td>
<td><em>tar</em> [QEQE] Cam&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Expresses <em>E. coli</em> Tar receptor</td>
<td>(22)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expresses Tsar receptor, the chimera receptor replacing the periplasmic</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>domain of Tar with the periplasmic domain of Tsr</td>
<td></td>
</tr>
<tr>
<td>pVS1252</td>
<td>tsar Cam&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Expresses <em>E. coli</em> wild-type full-length Tar</td>
<td>(23)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tar receptor. Used for mutagenesis for Tar rational design. <em>tar</em> was</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ligated into pPD10.</td>
<td></td>
</tr>
<tr>
<td>pPD12</td>
<td>tar Amp&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Expression plasmid</td>
<td>(23)</td>
</tr>
<tr>
<td>E. coli strains</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BL21 (DE3)</td>
<td>F&lt;sup&gt;−&lt;/sup&gt;, <em>ompT</em> hsdS&lt;sub&gt;B&lt;/sub&gt; (r&lt;sub&gt;B&lt;/sub&gt;&lt;sup&gt;−&lt;/sup&gt; m&lt;sub&gt;B&lt;/sub&gt;&lt;sup&gt;−&lt;/sup&gt;) gal dcm (DE3)</td>
<td>Periplasmic domain of Tar expression and purification strain</td>
<td>Novagen, Germany</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RP437</td>
<td>thr-1leuB6 his-4 metF59 eda-50 rpsL136</td>
<td>Wild-type <em>E. coli</em> strain</td>
<td>(24)</td>
</tr>
<tr>
<td>Strain</td>
<td>Mutation Description</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>UU1250</td>
<td>$\Delta_{aer}1\Delta_{(tar-tap)}5201\Delta_{tsr}-7028$</td>
<td>The strain lacks all five chemoreceptors Aer, Tar, Tsr, Trg, and Tap</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\Delta_{trg}100\ ygjG::Gm\ zbd::Tn5$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\Delta_{aer}1\Delta_{tap}-3654$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UU1624</td>
<td>$\Delta_{tsr}-7028\Delta_{trg}100\ ygjG::Gm\ zbd::Tn5$</td>
<td>The strain possesses only the Tar chemoreceptor</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\Delta_{aer}1\Delta_{tap}-3654$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RP2361</td>
<td>$\Delta_{tar}3862$</td>
<td>The strain lacks the chemoreceptor Tar</td>
<td></td>
</tr>
<tr>
<td>VS181</td>
<td>$\Delta_{(cheY\ cheZ)} \Delta_{tsr} \Delta_{(tar\ tap)}$</td>
<td>The strain lacks all five chemoreceptors Aer, Tar, Tsr, Trg, Tap, as well as CheY and CheZ.</td>
<td></td>
</tr>
<tr>
<td>Tar-only strain</td>
<td>VS181 tar</td>
<td>VS181 with pVS88 and pVS1092. Used in FRET measurement.</td>
<td></td>
</tr>
<tr>
<td>Tsar-only strain</td>
<td>VS181 tsar</td>
<td>VS181 with pVS88 and pVS1252. Used in FRET measurement.</td>
<td></td>
</tr>
</tbody>
</table>