Protein design-scapes generated by microfluidic DNA assembly elucidate domain coupling in the bacterial histidine kinase CpxA

Iain C. Clark[a] [iD], Bruk Mensa[b], Christopher J. Ochs[c], Nathan W. Schmidt[b], Marco Mravic[b] [iD], Francisco J. Quintana[d,e], William F. DeGrado[b,1] [iD], and Adam R. Abate[c,f,1] [iD]

[a]Department of Bioengineering, California Institute for Quantitative Biosciences, University of California, Berkeley, CA 94720; [b]Department of Pharmaceutical Chemistry, University of California, San Francisco, CA, 94158; [c]Department of Bioengineering and Therapeutic Sciences, California Institute for Quantitative Biosciences, University of California, San Francisco, CA 94158; [d]Broad Institute of MIT and Harvard, Cambridge, MA 02142; [e]Ann Romney Center for Neurologic Diseases, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115; and [f]Chan Zuckerberg Biohub, San Francisco, CA 94158

The randomization and screening of combinatorial DNA libraries is a powerful technique for understanding sequence–function relationships and optimizing biosynthetic pathways. Although it can be difficult to predict a priori which sequence combinations encode functional units, it is often possible to omit undesired combinations that inflate library size and screening effort. However, defined library generation is difficult when a complex scan through sequence space is needed. To overcome this challenge, we designed a hybrid valve- and droplet-based microfluidic system that deterministically assembles DNA parts in picoliter droplets, reducing reagent consumption and bias. Using this system, we built a combinatorial library encoding an engineered histidine kinase (HK) based on bacterial CpxA. Our library encodes designed transmembrane (TM) domains that modulate the activity of the cytoplasmic domain of CpxA and variants of the structurally distant "S helix" located near the catalytic domain. We find that the S helix sets a basal activity further modulated by the TM domain. Surprisingly, we also find that a given TM motif can elicit opposing effects on the catalytic activity of different S-helix variants. We conclude that the intervening HAMP domain passively transmits signals and shapes the signaling response depending on subtle changes in neighboring domains. This flexibility engenders a richness in functional outputs as HKs vary in response to changing evolutionary pressures.

droplet microfluidics | rational library design | protein engineering | histidine kinase | signal transduction

Protein engineers and synthetic biologists create and study biological function through the design of nucleic acids, typically assembled from separate coding or noncoding building blocks. In most cases, it is difficult to predict which assembled DNA sequences will yield the desired function. Instead, DNA libraries containing large numbers of variants are screened for the trait of interest. This approach is used to optimize expression of promoter–gene–terminator combinations (1), tune biosynthetic circuits to balance metabolic flux (2, 3), and map specific amino acid changes to protein function (4–6). Molecular methods for producing DNA libraries include random mutagenesis, mutational scanning, and assembly of preexisting sequences. Predefined parts allow control over the location and type of variants constructed, guaranteeing interesting mutations in the library and excluding nonfunctional ones. Together with DNA synthesis, combinatorial assembly eliminates the codon bias of random mutagenesis and facilitates construction of gene circuit and multidomain protein libraries (7–10). Consequently, a suite of molecular methods are now available to assemble predefined DNA parts into combinatorial libraries, including Golden Gate (11), Gibson et al. (12), SLiCE (13), SLIC (14), and BioBrick (15).

Generating libraries in a one-pot assembly of DNA parts is simple and commonplace but has shortcomings. These assembly reactions are subject to bias because all parts are free to react with all others; thus, DNA parts that efficiently combine are overrepresented in the final library. In addition, because all combinations are possible, such approaches generate massive libraries that cannot be thoroughly screened. For a library of m components and n variants per component, n^m possible ordered combinations exist. For instance, even a relatively simple biosynthetic pathway composed of four genes, each with its own promoter and ribosomal binding site (12 parts), and four variants per part, would result in over 16 million unique combinations. If six variants per part are used, a modest number when seeking to enhance a pathway, over 2 billion combinations are possible. For libraries of this size and larger, it can be difficult to assay a significant fraction of members even in pooled high-throughput screens, for example by fluorescence-activated cell sorting (FACS). However, it is likely that most of the sequence space contains nonfunctional variants that neither need to be constructed nor screened. This has driven approaches that limit variant assembly based on intuition or modeling (16), and therefore concentrate experimental effort on subsets most

Significance

DNA variant libraries allow unbiased analysis of sequence–function relationships. However, exhaustive combinatorial assemblies of even a modest number of parts produce libraries too large to practically screen. Library size can be reduced by manually blending parts to remove unwanted combinations, but this yields limited improvements and is difficult to scale. To address this challenge, we built a microfluidic hybrid valve–droplet device that selectively combines and assembles DNA parts into libraries of predefined composition with low bias. We use this system to construct and study a library of engineered histidine kinases, revealing how structural changes are communicated across adjacent domains to modulate kinase activity. Our method creates rationally reduced libraries and has applications in synthetic biology and protein engineering.

likely to be functional (Fig. 1A). Rationally reduced libraries can be constructed with many different assembly reactions that contain different starting parts. However, this manual approach is not scalable with current technology. Moreover, methods to automate DNA synthesis by microfluidics, although promising, have yet to produce large subset libraries at high throughput (17–20). Thus, a system that precisely generates targeted libraries from a complex part list would enable large combinatorial sequence spaces to be screened at a fraction of effort and cost.

To address this challenge, we developed a microfluidic system that can be programmed to create specific combinations of predefined DNA parts at high throughput and with minimal reagent consumption. Our system uses microfluidic valves to combine specific DNA parts, followed by compartmentalization into individual water-in-oil droplets to perform isolated assembly reactions. The system steps through desired DNA combinations sequentially and selectively (Fig. 1B) at about one per second. Less than 150 μL is consumed to produce 5,000 DNA combinations, a 300-fold reduction in reagent consumption compared with automated well plate assembly. The resultant droplet reactors, each of which contains all requisite parts and enzymes to construct a unique variant, are thermally incubated in parallel for molecular assembly. Additionally, droplet compartmentalization suppresses bias since efficient reactions do not compete with less efficient ones, and all have ample time to achieve saturation. Instead of relying on hybridization to colocalize oligonucleotides (21), a process that can be difficult to optimize as library complexity increases, our approach uses programmed microfluidics to create diverse libraries and even libraries from predefined oligonucleotide parts at optimal concentrations. This process can generate libraries of any desired combination from starting parts, and greatly reduces reagent consumption and screening effort when complex combinations of DNA parts are needed.

Using this technique, we constructed a four-part combinatorial library encoding multiple domains of the canonical bacterial histidine kinase (HK) CpxA (22–24). Transmembrane (TM) HKs are multidomain proteins that often relate the binding of an extracellular ligand to the activity of an intracellular kinase. The signal passes along an extended series of homodimeric signal-transducing domains, which communicate via coaxial two- and four-helix bundles along the dimer interface. The signal-transducing domains include a number of modules, such as HAMP and GAF domains and helical linkers, frequently swapped between HK family members (25). Here, we use protein design to investigate how systematic modifications to TM and linker domains modulate the activity of the downstream catalytic domains. Addressing this question is key to understanding how HKs transmit conformational information through multiple linker domains. We replace the TM domain of CpxA with a series of engineered TM dimers, which vary in sequence, structure, and length of the cytoplasmic connector to the N terminus of CpxA's signal-transducing HAMP domain. We also vary the sequence of the signaling helix (S helix) located at the C terminus of the HAMP domain, just prior to the catalytic domain. The sequence of the intervening HAMP domain is held constant, allowing examination of how it responds to and transmits diverse structural signals. We find that the S helix sets the basal rate of activity, which is modulated up or down in a roughly additive manner by changes in the linker, when examined in the population average. In addition, when the variants are examined individually, we find a single TM sequence can elicit opposing effects on the catalytic activity of different S-helix variants—enhancing activity in one variant and inhibiting in another. Thus, the intervening HAMP domain not only passively transmits signals but fundamentally changes the signaling patterns depending on even single-site variants in the neighboring domains. We posit that this flexibility yields a spectrum of functional outputs as HKs vary in response to changing evolutionary pressures, providing a potential explanation for the retention of multiple transmitting domains throughout evolution.

Results

Microfluidic Design and Operation. To enable rational construction of a multidomain protein library, we designed a multivalve microfluidic device to encapsulate specific combinations of presynthesized DNA library parts in 90-μm (382-pL) droplets (Fig. 2A and *SI Appendix*, Fig. S1). The device consists of 38 reagent inlets, each with its own microfluidic membrane valve (26) and an oil inlet for droplet formation. The valves are arranged in two arrays on either side of the microfluidic drop maker (Fig. 2B, valve array). Open valves dispense reagents, which flow into a T-junction drop maker.
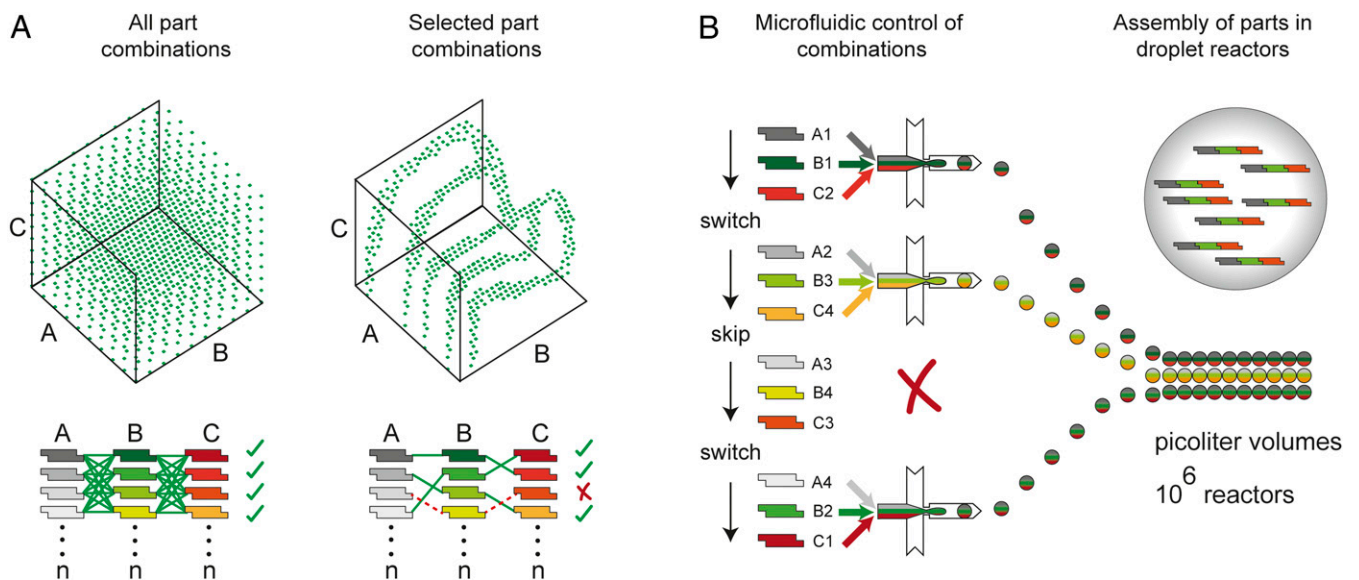
Fig. 1. Generation of DNA libraries of predefined composition. Nonrandom walks through sequence space require libraries with predefined variants. (A) Visual representation of the sequence space. (A, Left) Combinations generated with a one-pot all-by-all approach yield, ideally, all combinations in the sequence space. (A, Right) Combinations generated by selectively combining specific parts allow predetermined walks through the sequence space and reduced screening effort. (B) Microfluidic approach for generating subset libraries. Switching rapidly between inlet channels combines parts and enzymes for assembly inside droplet reactors.

Two valves downstream of the drop maker direct droplets into off-chip waste or collection tubes (Fig. 2C, collection switch). Each fluidic channel is connected via flexible tubing to a single well in a 96-deep-well plate housing DNA parts, enzymes, and buffer. The 96-well plate is pressurized with a manifold (*SI Appendix*, Fig. S1D) that ensures all wells are at equal pressure. Pressurization of the manifold drives flow through serpentine resistor channels (Fig. 2D, pressure balancing), fabricated on a separate layer of the chip, which maintains equivalent flow rates from each inlet channel and, thus, ensures controlled final reagent concentrations. The design is scalable since additional inlets can be added along the length of the central channel. Following the design considerations outlined in *SI Appendix, Materials and Methods*, the device can scale to 100k libraries with minimal changes to design or operation.

Each microfluidic valve can be actuated in <10 ms (Fig. 3 A and B), as measured using a fluorescent dye and droplet cytometer (27), allowing rapid switching between inlets (Movie S1). The device operates on an ~720-ms duty cycle consisting of 500 ms of library collection, 200 ms of channel flushing to waste, with 10 ms to switch valves to the next part combination (Fig. 3 A–C). To start a cycle, the collection valve closes, and the waste valve opens (Movie S2). Four reagent valves open simultaneously, allowing enzyme master mix and DNA parts to flow through the central channel, into the drop maker, and out through the waste valve. After this flush, the collection valve opens, the waste valve closes, and droplets containing the desired library components are collected. To switch to a new part combination, the collection valve closes, the waste valve opens, and the cycle is repeated (Fig. 3D

and Movies S3 and S4). This cycle continues until all desired combinations have been generated. A library of ~10,000 defined combinations can be created in ~2 h. Droplets containing DNA parts and enzymes are incubated off-chip to complete the assembly reaction. Any DNA assembly molecular biology can be used, including those requiring temperature changes, because the droplets are stable to heat and collected in PCR tubes that can be thermocycled.

Rationale for the Multidomain CpxA Library Design. We used our microfluidic library construction method to build a multidomain protein library from four DNA parts, each with eight or nine variants per part (Fig. 4A). This library allows us to explore how diverse structural inputs from a designed TM domain and cytoplasmic linkers influence the kinase activity of the bacterial HK CpxA (28). The TM protein CpxA is a prototypical HK that senses cell-envelope and protein-folding stress in the periplasm (29). The wild-type protein consists of (from the periplasmic to the C-terminal cytoplasmic kinase) 1) a periplasmic sensor domain, 2) an antiparallel four-helix TM domain, 3) a parallel four-helix bundle signal transducing the HAMP domain, 4) a short S-helix domain, and 5) the two conserved catalytic HK domains (Fig. 4B). The signal-transducing domains are conserved and repeated in a modular fashion within many different HKs and combine to yield sensors for diverse molecular and environment cues. It is hypothesized that HKs function by a similar mechanism of transmitting conformational changes across their modular domains to influence kinase domain activity (24, 30–35). Functional diversity is thought to be achieved by fine-tuning HK domain conformational
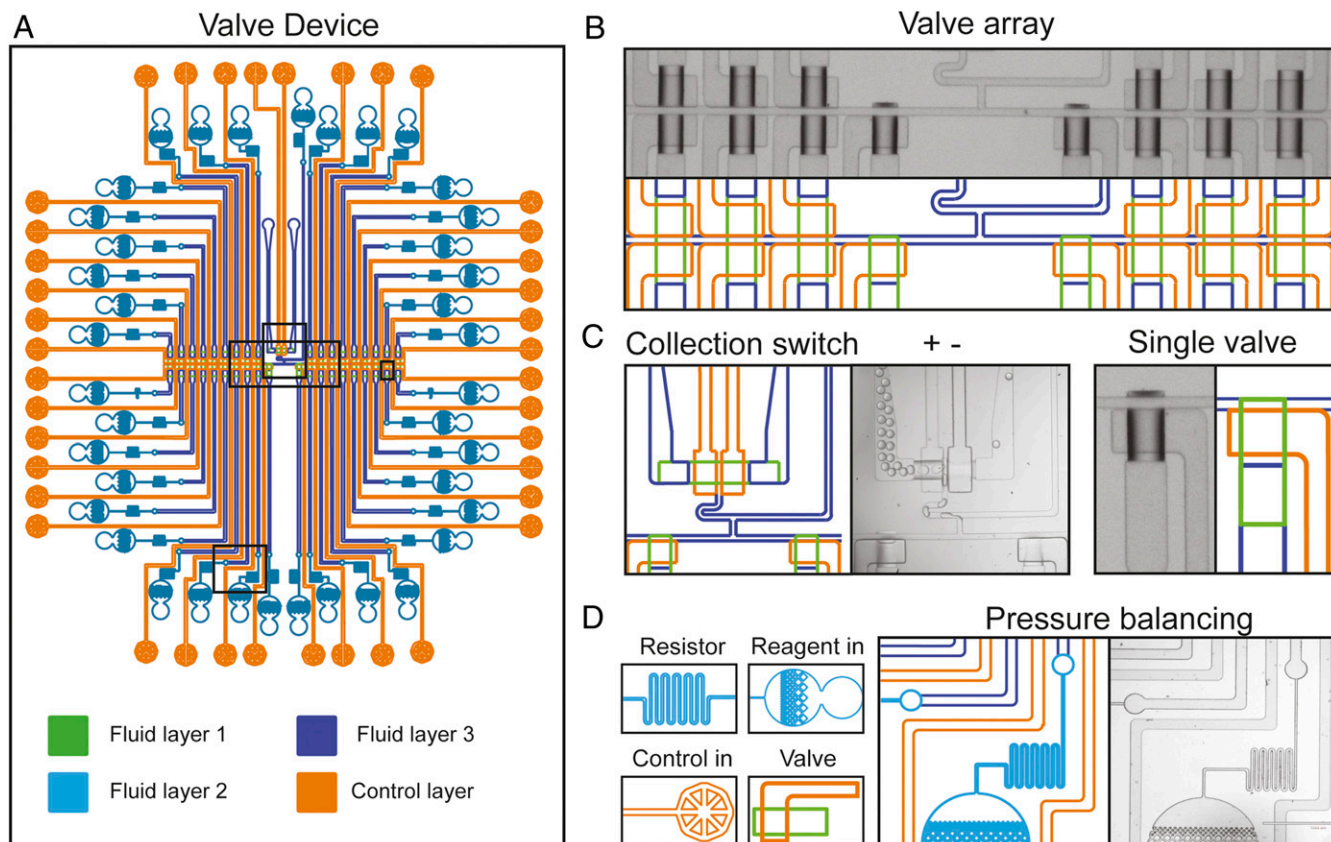
Fig. 2. Hybrid microfluidic device that uses valves and droplets to deterministically combine and assemble DNA parts. (A) Schematic of the device showing fluidic and control layers, inlets, valves, and resistors. (B) Schematics and images of the fabricated device. The valve array contains 38 reagent valves leading to a common channel where reagents are combined and drops are formed. (C) The collection switch is a two-valve system for directing droplets to a collection tube or into waste. (D) Fluid flow rates in inlets with different locations on the main collection channel are equalized by pressure balancing with on-chip resistors.
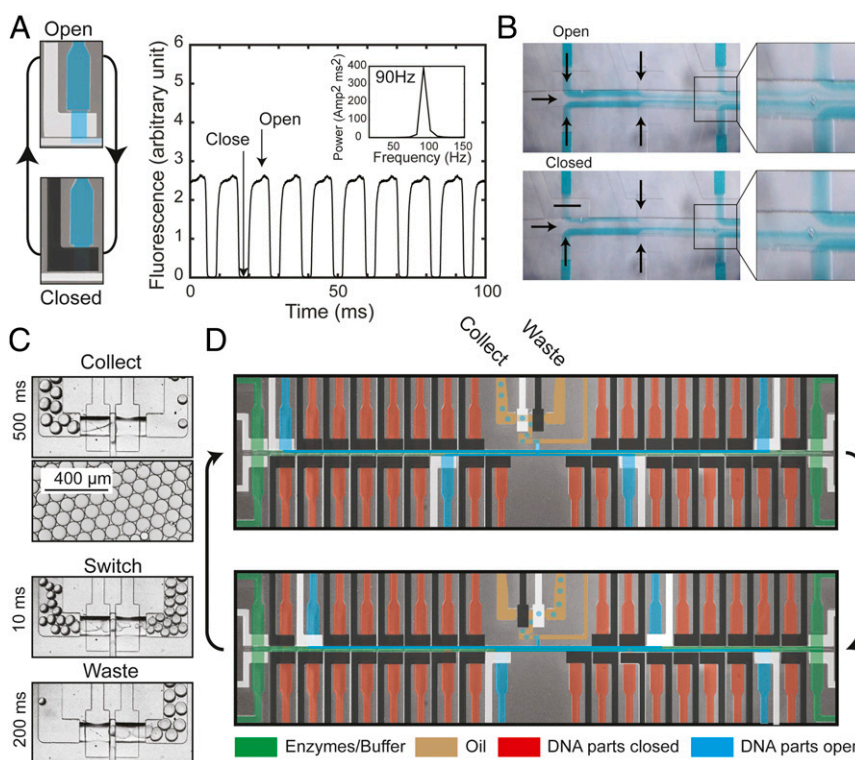
Clark et al.
Protein design-scapes generated by microfluidic DNA assembly elucidate domain coupling in the bacterial histidine kinase CpxA

PNAS | 3 of 10
https://doi.org/10.1073/pnas.2017719118

Fig. 3. Valve array duty cycle. (*A*) Valve opening and closing speed measured by fluorescence intensity. (*B*) Dye-labeled inlets show how reagents are combined in the central channel. Switching a single inlet from open (*Top*) to closed (*Bottom*) eliminates its stream from the central channel. (*C*) Flushing to waste reduces contamination from residual DNA parts from the prior cycle. (*D*) The device cycles between collecting drops (*Top*) and wasting drops (*Bottom*) as part of its duty cycle. Closed valves are colored black and open valves are colored white. Flowing DNA parts are colored blue, and blocked DNA parts are colored red. Enzymatic reagents are colored green. Oil is colored brown.

coupling via specific interdomain protein geometry. Here, we rationally design a family of CpxA variants, which systematically vary in expected interdomain geometry and coupling, to directly test the range of structural features that yield high-kinase activity. The designed components evaluated in this study consist (from the N to C terminus) of 1) an N-terminal maltose-binding protein (MBP) domain and hemagglutinin affinity tag, 2) a set of 81 TM dimers of systematically varying structure, 3) a variable-length alanine linker, 4) the wild-type HAMP domain, 5) a series of eight mutants in the S-helix region, and 6) the wild-type dimerization and histidine phosphotransfer (DHp) and adenosine triphosphate (ATP)-binding domains of the CpxA catalytic domain (Fig. 4 *A* and *B*). The N-terminal MBP tag results in periplasmic location of MBP, and is often used in similar genetic screens to select for expression and proper membrane insertion (36, 37).

The TM helices contain a strong TM helix homodimerization "G-X_3-G" motif, taken from the protein glycophorin A (LILL**G**VMA**G**VIGT) (38, 39). This TM design simplifies the conformational input propagated by the second TM helices of the native CpxA four-helix bundle into a two-helix bundle. The TM helix is constructed by combining two library variants (parts A + B), which allow flexibility in varying the length of the helix and the position of its G-X_3-G motif within the membrane (*SI Appendix*, Fig. S2). Furthermore, a single Trp residue, which prefers to localize near the headgroup region of the bilayer, is scanned through multiple positions of the C-terminal third of the TM helix to influence the location of the helix in the membrane, as in previous studies of HKs (Fig. 4*A*) (40). The helical phase of the last residue of the TM helix is held roughly constant, but the length of the neighboring linker is changed to allow systematic variation of phase. Short linkers are generally present between the TM helices and cytoplasmic domains of HK proteins. We therefore

place zero to seven helix-promoting alanine insertions in a linker (part C of the library; Fig. 4 *A* and *B*) between the TM helix and the cytoplasmic domains of CpxA$_{cyto}$ (HAMP, S-helix, and catalytic domains). Assuming helicity is maintained, each Ala would result in an ~100° phase shift and ~1.5-Å axial translation of the terminal linker residue as it connects to CpxA$_{cyto}$. Finally, in library part D, we introduce a series of eight consecutive single-site leucine substitutions into the S helix of CpxA$_{cyto}$ to alter its basal kinase activity. Evaluating the effects of TM-linker combinations over multiple S-helix variants offers a rigorous test for how conformational information coded by the upstream TM-linker variants couples through the native HAMP domain to the kinase domain. In all, this 5,184-member library consists of 81 TM helices, 8 linkers, and 8 S-helix variants.

Construction and Screening of the Multidomain Protein Library. We used our microfluidic system to generate a full library of all part combinations, as well as a library that has been reduced in size by restricting specific combinations of the TM (parts A and B). We use the Golden Gate assembly (41) of DNA parts with BsaI recognition sites to produce libraries from four DNA parts (A, B, C, and D) (Fig. 4*A* and *SI Appendix*, Table S1). There are nine A parts, nine B parts, eight C parts, and eight D parts. Combinations of parts A and B range from 16 to 28 amino acids, but we hypothesize that some AB combinations, which encode a synthetic transmembrane helix, may not be well-accommodated in the membrane. Therefore, in addition to generating a library from all parts (full library, 5,184 possible variants), we restrict part AB combinations based on the rule 18 AA ≤ [A + B] ≤ 25 AA (subset library, 3,648 possible variants). We sequence each library after on-chip assembly and cloning and compare the diversity and evenness with standard tube assembly. The library is designed such that multiple paired-end 150-bp reads cover the entire ABCD
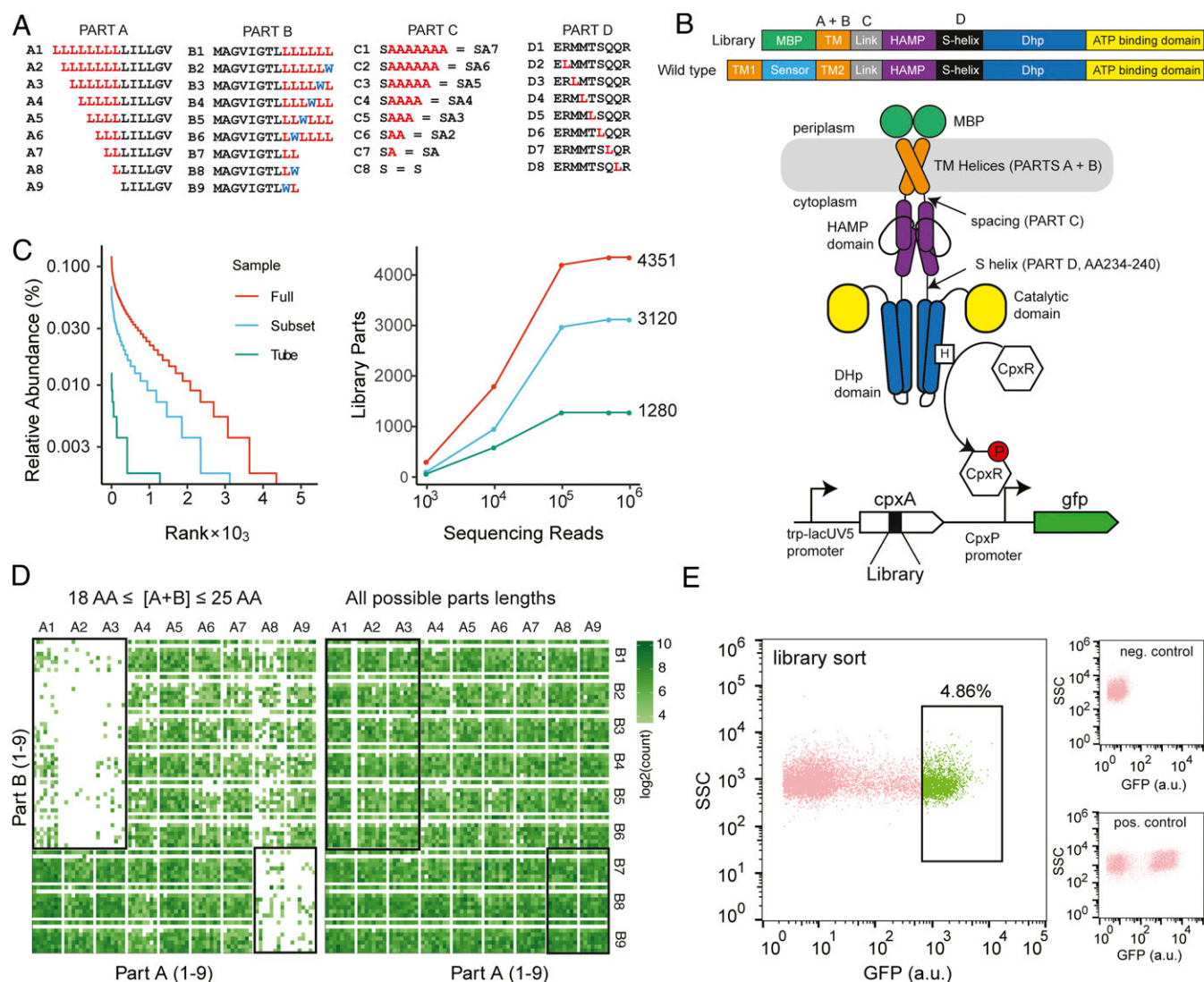
Clark et al.
Protein design-scapes generated by microfluidic DNA assembly elucidate domain
coupling in the bacterial histidine kinase CpxA

Fig. 4. Construction and sequencing of the DNA library demonstrate efficient on-chip library subsetting. (*A*) Sequences of the four parts used to construct the combinatorial DNA library. (*B, Top*) Comparison of the synthetic and wild-type CpxA structures. (*B, Bottom*) The engineered CpxA contains MBP replacing the signal domain, followed by a two-helix transmembrane region encoded by parts A and B. The variable juxtamembrane linker (part C) between the TM and HAMP domain is followed by leucine substitutions in the S helix (part D). CpxA phosphorylates the response regulator CpxR, which activates transcription of a GFP reporter via the *cpxP* promoter. (*C, Left*) Rank-abundance curves comparing the full library generated on-chip (full), full library generated in a tube (tube), and subset library generated on-chip (subset). Microfluidic assembly enhances variant coverage compared with pooled, tube-based assembly. (*C, Right*) Subsampling reads and counting library members quantify diversity and confirm adequate sequencing depth. (*D*) Distribution of library sequences in the subset and full libraries, displayed as log2(read counts). Parts C and D are nested within parts A and B. (*D, Left*) AB combinations constrained by the length of A and B such that 18 AA ≤ [A + B] ≤ 25 AA (subset). (*D, Right*) No restriction is placed on which parts were combined to generate the library (full). (*E*) Sorting of GFP^high reporter cells expressing the CpxA library. Positive (CpxA L243S) and negative (CpxA Q229V) controls are shown.

sequence, allowing us to assess the abundance of library members by next-generation sequencing. Libraries generated in droplets have a higher number of unique members: The on-chip–constructed full (4,351 of 5,184, 84%) and subset (3,120 of 3,648, 86%) libraries have greater coverage and evenness than standard in-tube assembly (1,280 of 5,184, 25%) (Fig. 4C). The subset library has substantially deselected (29.9 to 2.7% of the total library) the undesired part combinations from the final clone library (Fig. 4D, 18 AA ≤ [A + B] ≤ 25 AA, black boxes), demonstrating that our microfluidic system can construct libraries with predefined members.

To experimentally evaluate how these protein domains interact and influence CpxA signal transduction, we designed a pooled assay to screen our engineered protein library (Fig. 4E). On-chip libraries are cloned into a plasmid to create an intact, in-frame, designed *cpxA* protein-coding gene. The plasmid contains a constitutive promoter driving CpxA and a green fluorescent protein (GFP) reporter under the control of the *cpxP* promoter which is activated by CpxR, the cognate response regulator of CpxA. This is a reliable reporter of overall CpxAR activation (42). Libraries are expressed in *Escherichia coli cpxA*::km (JW3882-1; The Coli Genetic Stock Center, Yale University) and the GFP^high population is sorted by FACS, corresponding to ~4.86% of cells (Fig. 4E). The abundance of library members before and after sorting is determined by Illumina sequencing and the enrichment of each ABCD sequence is calculated ([SI Appendix, Materials and Methods](#)).

The Phenotypic Effects of Variations in the A to D Library Components. Our library construction and screening method allows us to evaluate variations in one library part, while either varying the sequence of the other parts or holding them constant. The resulting

Clark et al.
Protein design-scapes generated by microfluidic DNA assembly elucidate domain coupling in the bacterial histidine kinase CpxA

PNAS | 5 of 10
https://doi.org/10.1073/pnas.2017719118

activity and enrichment profiles can be evaluated in the context of different models of conformation coupling. For example, the interplay between the effects of substitutions in the S helix, which set the basal kinase level, and the TM linker informs models for coupling through the HAMP domain to the DHp. At one extreme, the HAMP can be postulated to have just two conformations whose relative energetics change in response to structural transitions of upstream signaling domains (in this case the different variants coded by parts A to C). In this "two-state HAMP model," each TM-linker variant would have the same effect on the activity of each of the S-helix variants, and vice versa. At the other extreme, the "multistate HAMP model," the HAMP might have a more dynamic structure, which continuously varies in response to the TM-linker input conformation (43). In the HAMP multistate model, different inputs from the TM helix linker would be expected to induce differing conformations in the HAMP domain, which might couple differently to mutants of the neighboring S helix. In this scenario, we would see different relative activity patterns for the various S-helix mutants in response to structural changes in upstream domains, arising from the complex coupling of a multistate HAMP to the adjacent S helix (and vice versa).

We first analyze enrichment of the individual members of two library components, averaged over the remaining variants. For
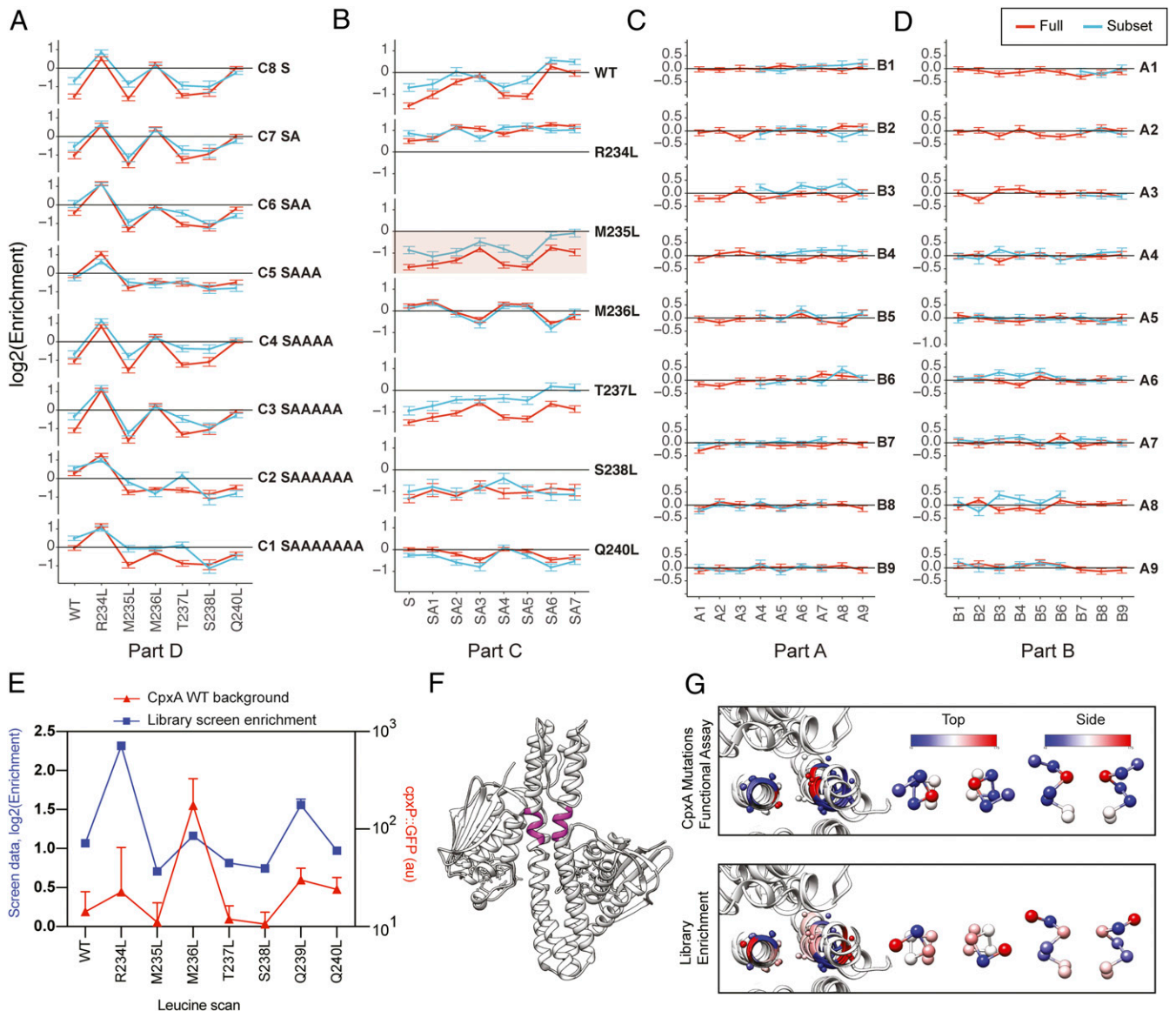

Fig. 5. Expression and screening of the DNA library encoding CpxA identify the S helix as a major determinant of signaling. (*A*) Enrichment as a function of leucine substitution (part D) for each alanine insertion (part C) (averaged over parts A and B). WT, wild type. (*B*) Enrichment as a function of linker alanine insertion (part C) for each Leu substitution (part D) (averaged over parts A and B). (*C*) Enrichment as a function of part A for each part B (averaged over parts C and D). (*D*) Enrichment as a function of part B for each part A (averaged over parts C and D). (*A*–*D*) Enrichment is calculated as the fold change in the normalized abundance of each variant sequence between post and presort. Error bars are SE, calculated as the mean divided by the square root of the sample size. Blue is the subset library and red is the full library. Q239L (full and subset) and 18 AA ≤ [A + B] ≤ 25 AA (subset) were removed from the library because of low counts. (*E*) Comparison of screen (enrichment) data with the functional *cpxP*::GFP assay. For the functional assay, L mutants are made in otherwise wild-type CpxA. (*F*) Location of the S helix in CpxA. (*G*) Mapping of reporter fluorescence (*Top*) and enrichment scores (*Bottom*) on the S helix of aligned CpxA structures (PDB ID codes 4BIV and 4BIU, chains A/B and D/E) shows enriched L variants segregate to the core of the dimer and deenriched variants segregate on the outward face.

example, Fig. 5A shows the enrichment for each of the eight S-helix variants (part D) as a function of the eight linkers (part C), in each case averaged over all possible A and B components. Other pairwise combinations are considered in Fig. 5 to identify which components contribute most to variations in enrichment. When viewed in this context, the most important determinant of the enrichment score is the identity of part D, corresponding to the S helix (Fig. 5A). Given its position between the wild-type HAMP and DHp domains, it is not surprising that S mutants have a large effect on the activity of the kinase. There is also high concordance between the enrichment results from the two libraries (full and subset), which were independently generated, screened, and sequenced. Importantly, the enrichment profiles for S-helix mutants move up and down in concert as the linker length is systematically varied. This matches the result expected in the HAMP two-state model (43–45). To determine the effect of these mutations in full-length native CpxA, we introduce each leucine mutant into wild-type protein and test each mutant's activity using a *cpxP*::GFP reporter strain by flow cytometry. These CpxA mutants display the same pattern of GFP fluorescence as the library enrichment (parts A + B averaged, part C: wild type) (Fig. 5E). The pattern, as a function of the leucine substitution position, is consistent over many replicates (*SI Appendix*, Fig. S3A).

Nonadditive Effects of the Transmembrane and Juxtamembrane Linker on Enrichment and Kinase Activity. While S-helix variants are the dominant contributor to library enrichment and appear to respond additively to substitutions in a manner consistent with the two-state model, in-depth analysis also reveals more subtle deviations from the expectations of such a simple model. The linker length made significant, although less dominant, contributions to library enrichment, as shown in the profiles generated for linkers of differing length while holding the S helix constant and averaging over parts A and B (Fig. 5B). For example, while M235L is deleterious to signaling, the linker sequence modulates activity up or down (Fig. 5B, shaded red). Earlier studies of HKs have shown that when residues are inserted into the helical interdomain linkers of signaling domains the transcriptional activity is modulated by linker length in a sinusoidal manner, with a repeat roughly matching that of the α-helix (46–49). We observe similar sinusoidal results in both libraries, but the phase changes between different S-helix mutants (Fig. 6B). Thus, linker variants that promote HK activity in one S-helix variant inhibit it in another. This finding departs from the expectations of a strict two-state coupling model.

We also examined the effects of varying parts A and B, which together comprise the TM helix. Parts A and B do not appear to significantly affect enrichment when they were examined individually while averaging over parts C and D (Fig. 5 C and D). However, examination of individual datasets shows large variations in the response to A and B variants, depending on the nature of the C and D components (*SI Appendix*, Fig. S3B). To better understand the origin of this variation, we performed hierarchical clustering of the enrichment data, grouping sequences with similar variation with respect to part C (Fig. 6). This allows us to understand how individual members of the A and B components cooperate with the linker Ala insertions to affect activity in the context of different S-helix variants. The periodic effects of alanine insertion observed in the averaged background (Fig. 5B) become even more pronounced in the individual clusters (Fig. 6) and, again, the effects are consistent between the screen replicates (full vs. subset library). Furthermore, we remake A-insertion mutations in full-length native CpxA and test each mutant's activity using the *cpxP*::GFP reporter assay (*SI Appendix*, Fig. S3C). Similar periodicity is observed in this dataset to patterns seen in cluster 15 and in M236L (averaged over A and B) (*SI Appendix*,

Fig. S3D), suggesting that linker effects are reproducible but highly dependent on domain context.

The periodicity and phase of insertional profiles provide important information about the geometry of the helical linkers between signaling domains. The phase relates to the location of the Cα atoms as they wind around an α-helix, while the period relates to the interhelical interaction pattern in a helical dimer. For example, a parallel pair of canonical α-helices has a repeat of 3.6 residues, while a left-handed coiled coil (left-handed crossing near 20°) has a 3.5-residue repeat, and a right-handed glycophorin-like motif (right-handed crossing near 40°) has a period of 4.0. These periods can also be modulated by nonideality of the α-helix, particularly when the registry of the input and output helices being connected is not easily spanned in a helical conformation. To determine the phase and period, we fit the enrichment profiles for the linkers to a sine function. The majority of the highly populated clusters were well-described by a sine wave with repeats within 3 to 4.2 residues (Fig. 6 and Dataset S1), which is within the range seen in helical dimerization motifs, including the dimeric helical linkers in HKs (50). Since clusters contain different S-helix and TM substituents, the range of phases observed is consistent with the observation that linker variants that promote HK activity in one TM/S-helix context often inhibit it in another.

Structure–Function Links Revealed by All-Atom Molecular Dynamics Simulations of the Synthetic Transmembrane Domain. Next, we used molecular dynamics (MD) simulations to determine which structural properties of the TM dimers encoded by parts A and B associate with the different behaviors seen in the functional clustering. The synthetic transmembrane domain built by parts A and B encodes a G-X$_3$-G motif with variable leucine spacers and a shifted C-terminal tryptophan (Fig. 4A). This results in 81 variants with sequence lengths between 16 and 28 residues, which are expected to alter the depth of the crossing motif in the membrane, its crossing angle, and the Cα–Cα distance at the end of the TM region. To structurally define the expected conformations of the 81 TM variants, we ran 80-ns all-atom MD simulations of each TM dimer in 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine lipid bilayers (Fig. 7A) in explicit solvent with CHARMM36 parameters (51) as implemented in GROMACS 2018.

Across all simulations, the TM segments (parts A + B) are generally stable (*SI Appendix*, Fig. S2 B–E) and hold the intended initial dimeric conformation: the canonical G-X$_3$-G–type interaction with close interhelical backbone interactions (3.5 to 4.5 Å) with a right-handed helix–helix crossing angle. The mean rmsd over TM amino acids between final and initial simulation frames is 1.3 Å. For longer TM helices, structural rearrangements accommodate the increased length without significantly disrupting the strong G-X$_3$-G motif interaction. *SI Appendix*, Fig. S2A shows three representative cases of significant deviations from the starting structure for TM sequences that are longer than the canonical 20 residues of single-pass TM helices: The helices widen their crossing angle (*SI Appendix*, Fig. S2 A, *Left*) or adopt helix kink distortions (*SI Appendix*, Fig. S2 A, *Middle* and *Right*) at the N-terminal half of the dimer. These conformations help maintain their apolar side chains embedded in lipid. However, because the structurally stable G-X$_3$-G motif lies close to the cytoplasmic side of the membrane, the grossest structural variation occurs primarily on the outer leaflet of the bilayer, distant from the region that connects to the linker and HAMP domains (*SI Appendix*, Fig. S2A). Therefore, in the future, it would be desirable to design libraries that include a signal-responsive periplasmic sensor domain and remove sequences that encode physically unreasonable extensions of the TM helix as in our subset library.

To allow comparison with phenotypically defined clusters, we computed structural and geometric features likely to influence the conformation and energetics of the downstream domains (Fig. 7B). Specifically, we cluster the 81 models based on the depth of the G-X$_3$-G and the C-terminal end of the helices in the

Clark et al.
Protein design-scapes generated by microfluidic DNA assembly elucidate domain coupling in the bacterial histidine kinase CpxA

PNAS | 7 of 10
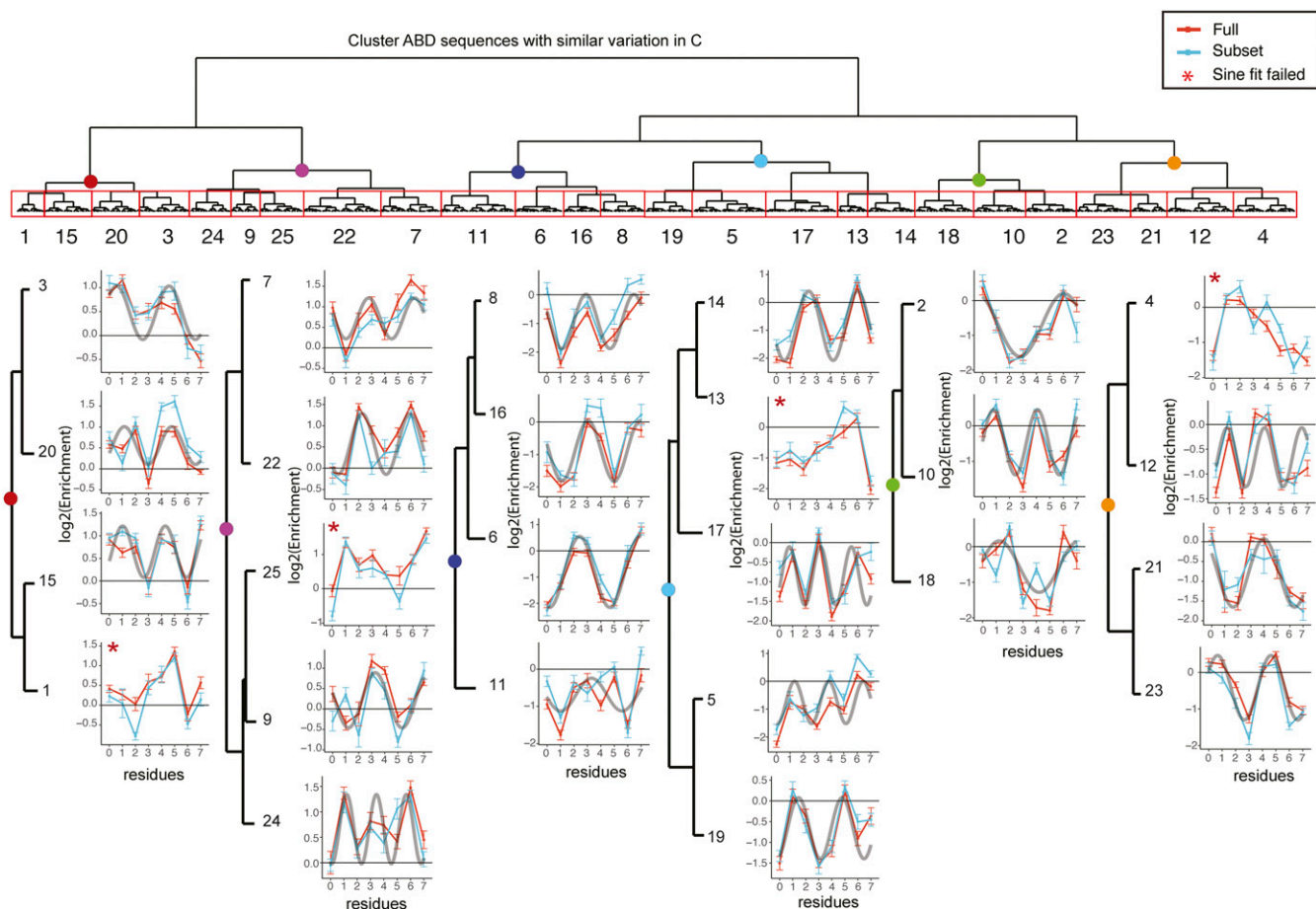https://doi.org/10.1073/pnas.2017719118

Fig. 6. Analysis of variable-length juxtamembrane Ala linkers (part C) in the context of different TM and S-helix domains. ABD sequences with similar variation in part C are clustered and enrichment is plotted as a function of the linker (part C). Sine curves fitted to data are superimposed on each dataset. The period of each sine curve falls within the expected range of 2.7 to 4.2, with few exceptions.

membrane, the interhelical crossing angle, the Cα–Cα distances near the end of the TM region, the helical registry, and the degree to which the dimers deviate from C2 symmetry. The TM-domain geometries cluster into eight structurally distinct groups (*SI Appendix*, Fig. S4A), each expected to pose a unique conformational input to the subsequent juxtamembrane linker (part C) and HAMP domain, and to influence interdomain coupling and kinase activity (52).

We next compared how structurally similar TM domains relate to functionally similar clusters from the screen. We cluster the screen enrichment data based on sequences with similar variation in downstream domains (parts C and D) to find which A + B combinations lead to similar functional outcomes. This procedure leads to phenotypically similar clusters, in which each member is identified by its A and B parts (*SI Appendix*, Fig. S4B). We evaluate the overlap between the individual AB combinations within structural and functional clusters. The resulting plot shows that functionally similar AB sequences correlate with structurally similar ones (Fig. 7C). For example, the two classes of TM helical pairs with the G-X$_3$-G located closest to the cytoplasm (structure classes 7 and 8) associate with only a limited number of phenotypic outcomes (predominantly function classes 3 + 4) based on the co-occurrence of the same AB sequences in the structural and functional cluster parts (*SI Appendix*, Fig. S4C). These findings establish a clear structure–function link between TM geometry and signaling output. It is, however, difficult to determine a single parameter that uniquely controls signaling, in part because

of the interdependence of the geometric features governing the structure. For example, helical ends can be shifted up or down by changing the interhelical crossing angle (scissoring motion) or by altering the depth caused by C-terminal Trp substitutions and Leu insertions. Additional structural studies with the C-terminal domains attached will be required to decipher the relative contributions of these and other features to the activation of the kinase domain.

Discussion

We report a microfluidic system that deterministically combines DNA parts for enzymatic assembly. The microfluidic system can be programmed to generate any library from a set of starting parts, and therefore allows nonrandom walks through sequence space, both in composition and abundance. Because the system uses physical compartmentalization instead of sequence or molecular biology optimization, it is fully compatible with such advances as they arise. The system uses resistors to ensure balanced flow rates, allowing additional part inlets to be added along the main channel to scale synthesis to more complex assemblies. An additional feature of our approach is that it allows biasing of the quantities of each part combination by adjusting the collection times to favor specific library members. If a higher representation of specific part combinations is desired, the device can be programmed to collect a larger number of droplets for those combinations. Thus, our microfluidic construct assembler represents an efficient tool for the generation of precise combinatorial
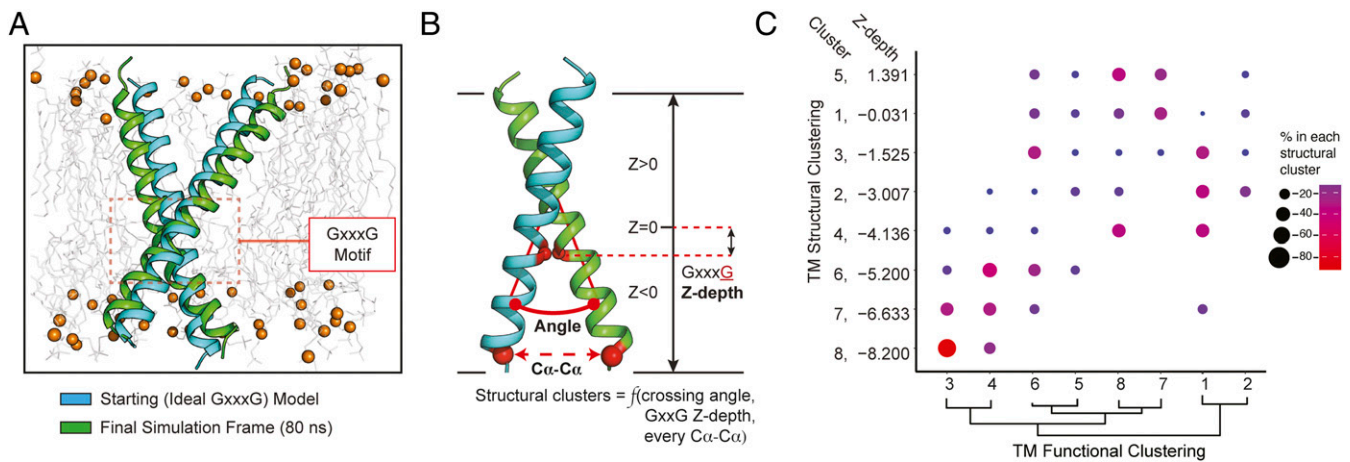
Fig. 7. Structure–function relationships revealed by molecular dynamics simulations. (*A*) MD simulations predict all TM dimer geometries simulated in lipid bilayers, shown as cartoon ribbon main-chain traces (initial model, cyan; final MD frame, green; lipid headgroup phosphates, orange spheres). A sequence with the largest backbone rmsd of the final MD simulation frame (80 ns) versus the initial models is shown as a representative example. (*B*) Visual representation of the TM dimer and structural parameters extracted from MD simulation data. TM-domain structural features are predicted based on MD simulations for each AB combination (also see *SI Appendix*, Fig. S2F). (*C*) TM clustering based on structural features compared with TM clustering based on functional enrichment (screen). Functionally similar AB sequences segregate with structurally similar ones.

libraries with applications in biosynthetic pathway design, protein engineering, and deep mutational scanning.

We apply our microfluidic system to study signal transduction in the prototypical bacterial histidine kinase CpxA. Our microfluidically generated library allows us to ask how geometric "signals" from engineered membrane-spanning domains transmit through HAMP and S-helix domains to a four-helix bundle that serves as the histidine phosphoryl acceptor in the phosphorelay catalytic mechanism of the HKs. Addressing this question is key to understanding not only how HKs signal but why they incorporate multiple linker domains (34, 52–54). What evolutionary advantage is there to retaining and shuffling signal transduction domains and interdomain linkers that have no apparent function in binding or catalysis? They appear to "passively" transmit signals in most proposed mechanisms, but it is hard to understand the conservation of multiple inserted domains rather than a single more concise connection.

To probe the mechanism of conformation coupling at a distance, the library scans Leu substitutions across the S helix, located one full domain (>20 Å) away from the connection between the TM linker and the N terminus of the HAMP. The S-helix variants contribute most to the phenotypic variation in the library, as expected from the S helix's proximity to the catalytic center. This dominant effect was also confirmed in native full-length CpxA (Fig. 5*E*). When enrichment scores are mapped onto available CpxA cytoplasmic structures (Protein Data Bank [PDB] ID codes 4BIV and 4BIU), high-signal variants predominantly fall on the dimeric interface of the two-helix bundle of the S helix while low-signal variants fall on the outside (Fig. 5 *E* and *F*). Substitutions to outward-facing residues near contacts with the ATP-binding domains of the kinase also impacted activity. Together, the Leu substitutions in the S helix served their intended purpose of significantly altering the energetic landscape of the intermediates required in the enzymatic cycle of the catalytic domain (i.e., autophosphorylation, phosphotransfer, and nucleotide exchange) (47), providing a range of basal activities in both the wild-type and engineered CpxA library.

When evaluated over the average of the TM variants, the phenotypes of the S-helix and Ala-linker library members covary in an additive manner. The S helix sets a baseline level of signaling, and the spacer modulates activity around that baseline. This behavior is expected from a simple two-state model, in which the HAMP has two different conformations that can either promote or inhibit kinase activity. In this model, the energetic difference between the two states—but not their structures—is altered by changes in

the sequence and structure of the upstream TM linker. However, by examining the phenotypic variation of each member of the library, we are able to also observe significant nonadditive behavior, in which mutations in one domain can have opposing effects, depending on the sequence of the distant domain. This is seen in the varied patterns of enrichment observed when the juxtamembrane linker is changed in different S-helix backgrounds (Figs. 5*B* and 6). Such behavior is more consistent with the possibility that HAMP domains have multiple conformations, which can vary significantly with respect to the input conformation from the TM domain and linker, and couple differently to the multiple conformational states populated as the kinase transitions through functional states (35, 50, 55, 56). In-depth MD simulations of the TM variants confirm a correlation between structural output from the membrane domain and the corresponding profile of activity of the corresponding variants (Fig. 7). This finding encourages us to speculate on the consistency of our phenotypic screening with structural mechanisms of kinase activation (24, 31–34).

Our results suggest that transducing domains provide opportunities to not only vary the basal levels of HKs but to also more radically change their energy landscapes and signaling patterns in response to evolutionary pressures. This is consistent with other reports documenting multiple continuously varying conformers of variants of HAMP domains (56). It also is consistent with the large structural changes seen in CpxA, as it transitions from a symmetrical resting state, to an asymmetric Michaelis complex, to the intermediate covalent phosphohistidine intermediate, which then binds and transfers the phosphoryl group to its cognate response regulator. Finally, CpxA also has a distinct phosphatase activity, and it is the balance of the phosphatase to the kinase activity that sets the overall transcriptional response. In this view, a multiconformational HAMP domain would engender functionally diverse outcomes to sequence alterations by allowing differential coupling to each of the individual functional structural states of the catalytic core machinery.

Conclusion

We describe a microfluidic system for the construction of rationally reduced DNA libraries. Our approach combines valves with droplet microfluidics to rapidly select and enzymatically assemble predefined DNA parts into constructs. This approach allows construction of libraries with targeted part combinations automatically, at high speed, and with low reagent usage. Thus, our microfluidic construct

Clark et al.
Protein design-scapes generated by microfluidic DNA assembly elucidate domain coupling in the bacterial histidine kinase CpxA

PNAS | 9 of 10
https://doi.org/10.1073/pnas.2017719118

assembler affords a facile way to take efficient walks through sequence space. We used the system to assemble a multidomain protein library of the canonical bacterial sensor kinase CpxA. We show that mutations to the S helix of CpxA globally change kinase activity by directly altering the stability of catalytic states in the kinase domain, which are then energetically modulated by an interplay between sequence features in the TM and linker domain. Thus, we observed additive two-state coupling as the dominant theme but with significant contributions to nonadditive components, consistent with multistate signaling through the HAMP module.

Materials and Methods

Detailed information on the fabrication and operation of the microfluidic system; assembly, sequencing, and screening of the library; sequence and clustering analysis; validation experiments in wild-type backgrounds; and molecular dynamics simulations is found in *SI Appendix, Materials and Methods*. Movies showing operation of the microfluidics are found in Movies S1, S2, S3, and S4. *SI Appendix*, Tables S1 and S2 contain sequences of primers and library oligonucleotides. Datasets S1, S2, and S3 contain sine fits, screening data, and structural clustering results.

1. H. Alper, C. Fischer, E. Nevoigt, G. Stephanopoulos, Tuning genetic control through promoter engineering. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 12678–12683 (2005).
2. M. Jeschek, D. Gerngross, S. Panke, Combinatorial pathway optimization for streamlined metabolic engineering. *Curr. Opin. Biotechnol.* **47**, 142–151 (2017).
3. P. Xu et al., Modular optimization of multi-gene pathways for fatty acids production in *E. coli. Nat. Commun.* **4**, 1409 (2013).
4. L. M. Starita, S. Fields, Deep mutational scanning: A highly parallel method to measure the effects of mutation on protein function. *Cold Spring Harb. Protoc.* **2015**, 711–714 (2015).
5. P. A. Romero, T. M. Tran, A. R. Abate, Dissecting enzyme function with microfluidic-based deep mutational scanning. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7159–7164 (2015).
6. D. M. Fowler, S. Fields, Deep mutational scanning: A new style of protein science. *Nat. Methods* **11**, 801–807 (2014).
7. R. E. Cobb, J. C. Ning, H. Zhao, DNA assembly techniques for next-generation combinatorial biosynthesis of natural products. *J. Ind. Microbiol. Biotechnol.* **41**, 469–477 (2014).
8. C. Neylon, Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: Library construction methods for directed evolution. *Nucleic Acids Res.* **32**, 1448–1459 (2004).
9. C. H. Reynolds et al., Diversity and coverage of structural sublibraries selected using the SAGE and SCA algorithms. *J. Chem. Inf. Comput. Sci.* **41**, 1470–1477 (2001).
10. S. E. Osborne, A. D. Ellington, Nucleic acid selection and the challenge of combinatorial chemistry. *Chem. Rev.* **97**, 349–370 (1997).
11. C. Engler, R. Kandzia, S. Marillonnet, A one pot, one step, precision cloning method with high throughput capability. *PLoS One* **3**, e3647 (2008).
12. D. G. Gibson et al., Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
13. Y. Zhang, U. Werling, W. Edelmann, SLiCE: A novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res.* **40**, e55 (2012).
14. M. Z. Li, S. J. Elledge, Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat. Methods* **4**, 251–256 (2007).
15. R. P. Shetty, D. Endy, T. F. Knight Jr, Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng.* **2**, 5 (2008).
16. M. Jeschek, D. Gerngross, S. Panke, Rationally reduced libraries for combinatorial pathway optimization minimizing experimental effort. *Nat. Commun.* **7**, 11163 (2016).
17. D. S. Kong, P. A. Carr, L. Chen, S. Zhang, J. M. Jacobson, Parallel gene synthesis in a microfluidic device. *Nucleic Acids Res.* **35**, e61 (2007).
18. S. C. C. Shih et al., A versatile microfluidic device for automating synthetic biology. *ACS Synth. Biol.* **4**, 1151–1164 (2015).
19. U. Tangen et al., DNA-library assembly programmed by on-demand nano-liter droplets from a custom microfluidic chip. *Biomicrofluidics* **9**, 044103 (2015).
20. C. J. Ochs, A. R. Abate, Rapid modulation of droplet composition with pincer microvalves. *Lab Chip* **15**, 52–56 (2015).
21. C. Plesa, A. M. Sidore, N. B. Lubock, D. Zhang, S. Kosuri, Multiplexed gene synthesis in emulsions for exploring protein functional landscapes. *Science* **359**, 343–347 (2018).
22. R. F. Weber, P. M. Silverman, The cpx proteins of *Escherichia coli* K12. Structure of the cpxA polypeptide as an inner membrane component. *J. Mol. Biol.* **203**, 467–478 (1988).
23. E. Batchelor, D. Walthers, L. J. Kenney, M. Goulian, The *Escherichia coli* CpxA-CpxR envelope stress response system regulates expression of the porins ompF and ompC. *J. Bacteriol.* **187**, 5723–5731 (2005).
24. A. E. Mechaly, N. Sassoon, J. M. Betton, P. M. Alzari, Segmental helical motions and dynamical asymmetry modulate histidine kinase autophosphorylation. *PLoS Biol.* **12**, e1001776 (2014).
25. T. Krell et al., Bacterial sensor kinases: Diversity in the recognition of environmental signals. *Annu. Rev. Microbiol.* **64**, 539–559 (2010).
26. M. A. Unger, H. P. Chou, T. Thorsen, A. Scherer, S. R. Quake, Monolithic microfabricated valves and pumps by multilayer soft lithography. *Science* **288**, 113–116 (2000).
27. J. J. Agresti et al., Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 4004–4009 (2010).
28. I. Gushchin et al., Mechanism of transmembrane signaling by sensor histidine kinases. *Science* **356**, eaah6345 (2017).
29. T. L. Raivio, T. J. Silhavy, Transduction of envelope stress in *Escherichia coli* by the Cpx two-component system. *J. Bacteriol.* **179**, 7724–7733 (1997).
30. M. P. Bhate, K. S. Molnar, M. Goulian, W. F. DeGrado, Signal transduction in histidine kinases: Insights from new structures. *Structure* **23**, 981–994 (2015).
31. M. Hulko et al., The HAMP domain structure implies helix rotation in transmembrane signaling. *Cell* **126**, 929–940 (2006).
32. H. U. Ferris et al., The mechanisms of HAMP-mediated signaling in transmembrane receptors. *Structure* **19**, 378–385 (2011).
33. M. V. Airola, K. J. Watts, A. M. Bilwes, B. R. Crane, Structure of concatenated HAMP domains provides a mechanism for signal transduction. *Structure* **18**, 436–448 (2010).
34. K. E. Swain, J. J. Falke, Structure of the conserved HAMP domain in an intact, membrane-bound chemoreceptor: A disulfide mapping study. *Biochemistry* **46**, 13684–13695 (2007).
35. J. S. Parkinson, Signaling mechanisms of HAMP domains in chemoreceptors and sensor kinases. *Annu. Rev. Microbiol.* **64**, 101–122 (2010).
36. P.-C. Su, B. W. Berger, A novel assay for assessing juxtamembrane and transmembrane domain interactions important for receptor heterodimerization. *J. Mol. Biol.* **425**, 4652–4658 (2013).
37. P.-C. Su, B. W. Berger, Identifying key juxtamembrane interactions in cell membranes using AraC-based transcriptional reporter assay (AraTM). *J. Biol. Chem.* **287**, 31515–31526 (2012).
38. K. R. MacKenzie, J. H. Prestegard, D. M. Engelman, A transmembrane helix dimer: Structure and implications. *Science* **276**, 131–133 (1997).
39. D. T. Moore, B. W. Berger, W. F. DeGrado, Protein-protein interactions in the membrane: Sequence, structural, and biological motifs. *Structure* **16**, 991–1001 (2008).
40. R. R. Draheim, A. F. Bormans, R. Z. Lai, M. D. Manson, Tryptophan residues flanking the second transmembrane helix (TM2) set the signaling state of the Tar chemoreceptor. *Biochemistry* **44**, 1268–1277 (2005).
41. C. Engler, R. Gruetzner, R. Kandzia, S. Marillonnet, Golden Gate shuffling: A one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS One* **4**, e5553 (2009).
42. P. A. DiGiuseppe, T. J. Silhavy, Signal detection and target gene induction by the CpxRA two-component system. *J. Bacteriol.* **185**, 2432–2440 (2003).
43. V. Stewart, The HAMP signal-conversion domain: Static two-state or dynamic three-state? *Mol. Microbiol.* **91**, 853–857 (2014).
44. L. G. Mondéjar, A. Lupas, A. Schultz, J. E. Schultz, HAMP domain-mediated signal transduction probed with a mycobacterial adenylyl cyclase as a reporter. *J. Biol. Chem.* **287**, 1022–1031 (2012).
45. M. Doebber et al., Salt-driven equilibrium between two conformations in the HAMP domain from *Natronomonas pharaonis*: The language of signal transfer? *J. Biol. Chem.* **283**, 28691–28701 (2008).
46. A. Möglich, R. A. Ayers, K. Moffat, Design and signaling mechanism of light-regulated histidine kinases. *J. Mol. Biol.* **385**, 1433–1444 (2009).
47. V. Stewart, L. L. Chen, The S helix mediates signal transmission as a HAMP domain coiled-coil extension in the NarX nitrate sensor from *Escherichia coli* K-12. *J. Bacteriol.* **192**, 734–745 (2010).
48. K. Winkler, A. Schultz, J. E. Schultz, The S-helix determines the signal in a Tsr receptor/adenylyl cyclase reporter. *J. Biol. Chem.* **287**, 15479–15488 (2012).
49. B. Wang, A. Zhao, R. P. Novick, T. W. Muir, Activation and inhibition of the receptor histidine kinase AgrC occurs through opposite helical transduction motions. *Mol. Cell* **53**, 929–940 (2014).
50. N. W. Schmidt, G. Grigoryan, W. F. DeGrado, The accommodation index measures the perturbation associated with insertions and deletions in coiled-coils: Application to understand signaling in histidine kinases. *Protein Sci.* **26**, 414–435 (2017).
51. J. Huang, A. D. MacKerell Jr, CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013).
52. M. V. Airola et al., HAMP domain conformers that propagate opposite signals in bacterial chemoreceptors. *PLoS Biol.* **11**, e1001479 (2013).
53. H. U. Ferris et al., Mechanism of regulation of receptor histidine kinases. *Structure* **20**, 56–66 (2012).
54. H. U. Ferris, K. Zeth, M. Hulko, S. Dunin-Horkawicz, A. N. Lupas, Axial helix rotation as a mechanism for signal regulation inferred from the crystallographic analysis of the *E. coli* serine chemoreceptor. *J. Struct. Biol.* **186**, 349–356 (2014).
55. L. Zhu, P. G. Bolhuis, J. Vreede, The HAMP signal relay domain adopts multiple conformational states through collective piston and tilt motions. *PLoS Comput. Biol.* **9**, e1002913 (2013).
56. R. Z. Lai, J. S. Parkinson, Functional suppression of HAMP domain signaling defects in the *E. coli* serine chemoreceptor. *J. Mol. Biol.* **426**, 3642–3655 (2014).

10 of 10 | PNAS
https://doi.org/10.1073/pnas.2017719118

Clark et al.
Protein design-scapes generated by microfluidic DNA assembly elucidate domain coupling in the bacterial histidine kinase CpxA