Supplementary Materials

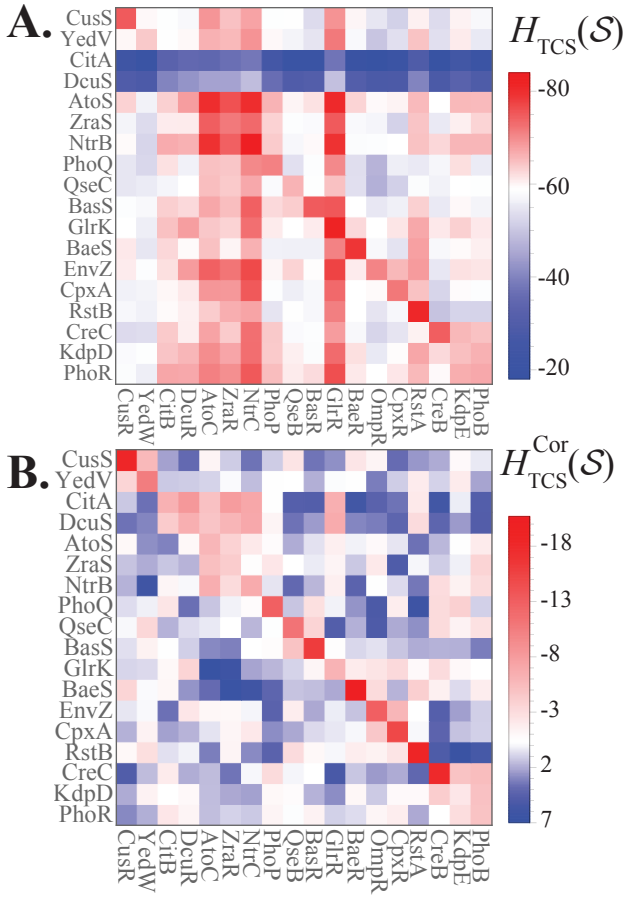

Fig. S1. H_{TCS} **captures partners in *E. coli***. (A) A grid plot of H_{TCS} where all of the cognate HK (y-axis) is plotted against the cognate RR (x-axis). TCS partners encoded adjacent to one another on an operon are referred to as cognate partners. The plot is arranged in such a way that the diagonal reflects the interaction between known partners (Yamamoto, et al. 2005). The strength of H_{TCS} is shown in the legend, where more negative suggests better signal transfer efficiency based on our model (i.e., more favorable interaction). Rows 3 and 4 are consistently unfavorable due to the poor alignment of CitA and DcuS, which contain mostly gaps in the MSA. (B) We can apply an average product correction to the square matrix in Fig. S1A as

$M_{ij} \to M_{ij} - M_{:j}M_{i:}/M_{::}$. Here, ":" denotes an average over the row or column, respectively. We refer to this corrected matrix as $H_{\mathrm{TCS}}^{\mathrm{Cor}}$, which satisfies the property that the sum of all rows and columns equals zero, respectively. This allows for the partner interactions to be clearly visualized along the diagonal. However, there is no physical basis for why the mean of each row should equal the mean of each column. Hence, we do not use any correction to our coevolutionary landscape, H_{TCS}, in our study.

Supplemental Text

Sequence database for HK intra-protein interactions: DHp and ATPase

We searched the Representative Proteomes (RP55) (Chen, et al. 2011) database for HK proteins using the Jackhmmer algorithm on the HMMER web server (Finn, et al. 2011) with default parameters. Once again, we restricted our curated set of HK proteins to ones having the PhoQ domain architecture. We restricted our MSA to a length (number of columns) of $L_{\mathrm{HK}} = 222$ by only including the DHp and ATPase domains. Our remaining dataset consisted of 4,483 non-redundant HK sequences of the form: $S_{\mathrm{HK}} = (A_1, A_2, ..., A_{L-1}, A_L)$ with a total length $L = L_{\mathrm{HK}} = 222$.

Mutational change in coevolutionary fitness landscapes of a kinase-only model

For the HK model (Fig. S2A), we infer a negative, additive fitness landscape with the form of Equation 3 and focus on the statistical couplings between positions in the HK

that are in close proximity in a representative structure of the HK autophosphorylation state (Fig. S2B):

$$H_{\mathrm{HK}}(S_{\mathrm{HK}}) = -\sum_{i=1}^{L_{\mathrm{HK}}-1}\sum_{j=i+1}^{L_{\mathrm{HK}}} J_{ij}(A_i, A_j)\times\Theta(c - r_{ij}) - \sum_{i=1}^{L_{\mathrm{HK}}} h_i(A_i) \qquad \text{(S1)}$$

where Θ is a Heaviside step function, c is the a cutoff distance of 16Å, and r_{ij} is the minimum distance between residues i and j in the representative structure. The mutational changes in our landscape are then computed using $\Delta H_{\mathrm{HK}}(S_{\mathrm{HK}}^{\mathrm{mutant}}) = H_{\mathrm{HK}}(S_{\mathrm{HK}}^{\mathrm{mutant}}) - H_{\mathrm{HK}}(S_{\mathrm{HK}}^{\mathrm{WT}})$.

Construction of a coevolutionary landscape to assess mutational effect on autophosphorylation

We further extend our analysis by examining whether a model based solely on HK (intraprotein) coevolution can identify the functional and non-functional mutational variants of PhoQ. We construct a coevolutionary landscape of HK intraprotein interactions, H_{HK} (Eq. S1), focusing on the DHp and ATPase domains (Fig. S2A). These domains form a binding interface during autophosphorylation (Fig. S2B). To test the fidelity of our inferred model, we plot the top coevolving residue pairs in the HK using the Direct Information (DI) (Weigt, et al. 2009; Morcos, et al. 2011) metric (Fig. S3). We compare the top coevolving pairs of residues with the contacts from an experimentally determined (Mechaly, et al. 2014) autophosphorylation structure of a HK with the same domain architecture as PhoQ, i.e., HAMP-DHp-ATPase. Experimental (Wang, et al. 2013) and predicted (Dago, et al. 2012) autophosphorylation structures of HKs without a

HAMP domain directly connected to the DHp appear to exhibit a different mode of autophosphorylation.

We generate a 2D histogram of ΔH_{HK} and ΔH_{TCS} for the 20^4 PhoQ mutational variants (Fig. S4A). Note that H_{HK} (Eq. S1) and H_{TCS} (Eq. 4) are models of the intraprotein (HK only) and interprotein (HK/RR) coevolution, respectively. Interestingly, the mutational variants as a function of ΔH_{HK} and ΔH_{TCS} are highly correlated with Pearson correlations of 0.72 and 0.75 for the functional and non-functional mutants, respectively (Fig. S4A). It may appear that the correlation between ΔH_{HK} and ΔH_{TCS} reflects the evolutionary constraint to satisfy both HK/RR phosphotransfer and HK autophosphorylation simultaneously. While this possibility exists, we find that ΔH_{HK} is much worse at identifying functional mutations than ΔH_{TCS}, which is quantified by the plot of PPV versus the *N* top mutational variants ranked by ΔH_{HK} (Fig. S5). This suggests that ΔH_{HK} may simply be correlated with ΔH_{TCS} simply by capturing the mutational effects on folding, homodimerization, and/or stability.

Next, we examine the subset mutational variants that limit "cross-talk" (detailed in previous subsection) in a 2D histogram of ΔH_{HK} and ΔH_{TCS} (Fig. S4B). We find that while the remaining functional mutants are clustered around more favorable values of ΔH_{TCS}, they cover a wide range of deleterious mutational changes in ΔH_{HK}. This would suggest that the HK is perhaps more tolerant of mutations that may reduce stability and/or autophosphorylation but more sensitive to mutations that reduce interaction specificity for phosphotransfer or phosphatase activity. Interestingly, it was reported that the response regulator PhoP, is capable of receiving a phosphoryl group from acetyl-

phosphate (Podgornaia and Laub 2015) and thus, mutations that minimally reduce PhoQ autophosphorylation activity but preserving interaction specificity with PhoP may still be identified as functional in experiment.

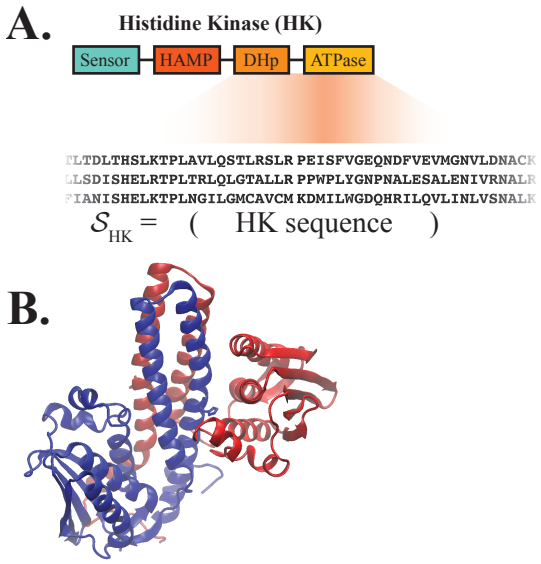

Fig. S2. HK domain interactions of interest. We focus only on HK proteins that have the following domain architecture from N to C terminus: sensor, HAMP, DHp, and ATPase. (A) We also consider a model of the HK sequences, S_{HK}, consisting of the DHp and ATPase domains. (B) The crystal structure of a representative autophosphorylation state is shown for the HAMP-containing HK CpxA in *E. coli (Mechaly, et al. 2014)*. The ATPase domain of one HK protein (red) is bound to the DHp domain of the other HK protein (blue).

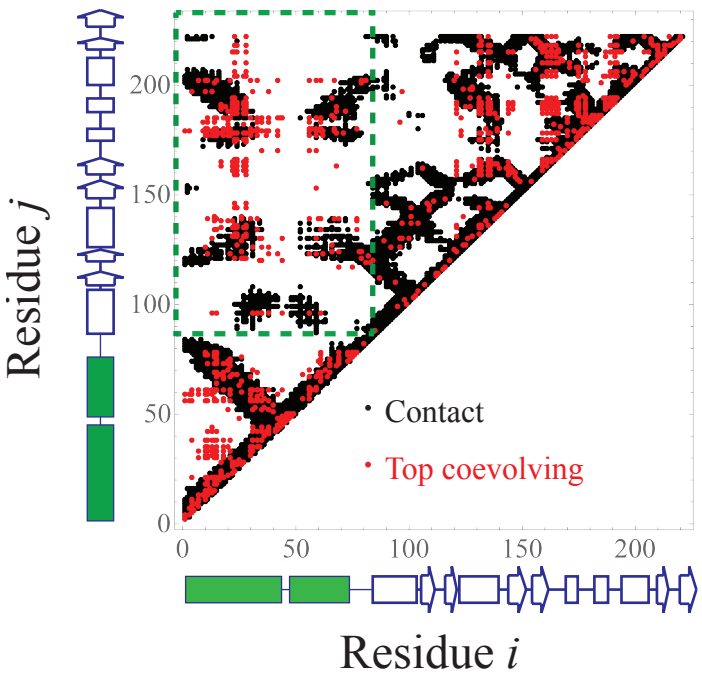

Figure S3. High coevolving pairs of residues in HK correspond to contacts in autophosphorylation state. A contact map is plotted for the crystal structure of the CpxA homodimer (HK) in *E. coli* (PDB: 4BIW(Mechaly, et al. 2014)) in black where each point denotes two residues in the structure that are within 12Å of one another. The secondary structure of the CpxA protein is shown on the x- and y-axes where the solid green color denotes the DHp domain while the ATPase domain is shown in a solid white. The dashed green box in the contact map contains the DHp/ATPase amino acid residue pairs. The 4BIW structure captures the autophosphorylation state of the HK. Since 4BIW is a homodimer, the inter- and intraprotein contacts are projected onto the contact map of a single monomer. Plotted alongside the contact map are the top 1000 coevolving pairs of residues predicted by DCA using the coevolutionary model constructed for the HK alone in red (See Materials and Methods). The Direct Information (DI) metric

is used as a proxy for the amount of coevolution between sites (See Refs: (Weigt, et al. 2009; Morcos, et al. 2011) for more details). The highly coevolved pairs of residues correspond well with the contacts in the autophosphorylation state of the HK.

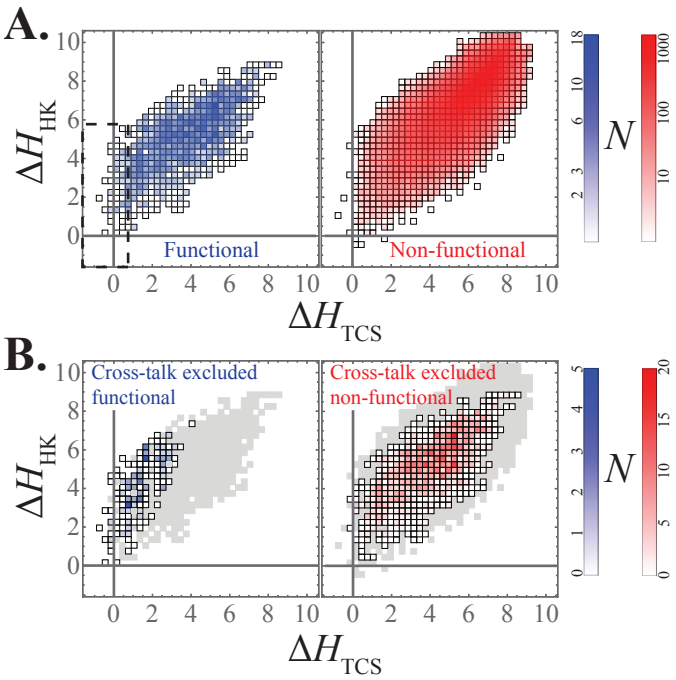

Fig. S4. Effect of mutation on PhoQ autophosphorylation. (A) Considering a 2D histogram of ΔH_{TCS} and ΔH_{HK} for the functional variants (blue) and non-functional variants (red). The two solid lines corresponding to $\Delta H_{TCS} = 0$ and $\Delta H_{HK} = 0$ represent the mutational change with respect to the wild type PhoQ. Additionally, $\Delta H_{TCS} < 0$ and $\Delta H_{HK} < 0$ correspond to mutations that are more favorable than the wild type. The number of variants in each bin is shown on the logarithmic scale by the shade of blue or red, respectively. We observe an enrichment of functional mutants for lower values of both ΔH_{TCS} and ΔH_{HK}. (B) The 2D histogram is plotted for the subset of mutational variants that limit "cross-

talk" by preserving the specificity of the PhoQ/PhoP interaction (labeled Cross-talk excluded). The gray squares denote the parts of the original 2D histogram (panel A) that have been removed.

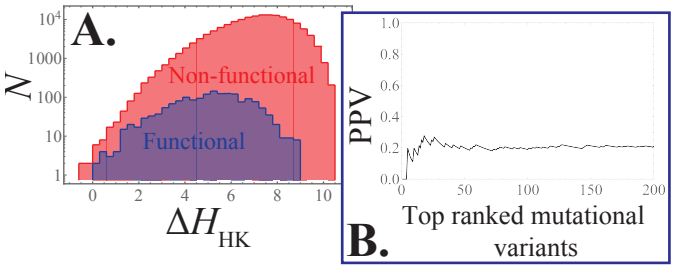

Figure S5. Effect of mutation on the PhoQ autophosphorylation: 1D histogram. (A) A histogram of the mutational change in our coevolutionary landscape, ΔH_{HK} (Eq. S1), is plotted for the functional (blue) and non-functional (red) mutational variants. The color purple shows parts of the plot where the blue and red histograms overlap. By definition, $\Delta H_{\mathrm{HK}} = 0$ corresponds to the mutational change in our landscape with respect to the wild type PhoQ. (B) We plot the positive predictive value, PPV=TP/(TP+FP), as a function of the N mutational variants ranked by ΔH_{HK} from the most to least favorable for the first 200 mutants. Once again, true positives (TP) and false positives (FP) refer to the fraction of mutants that are functional or non-functional, respectively, in the top N ranked variants.

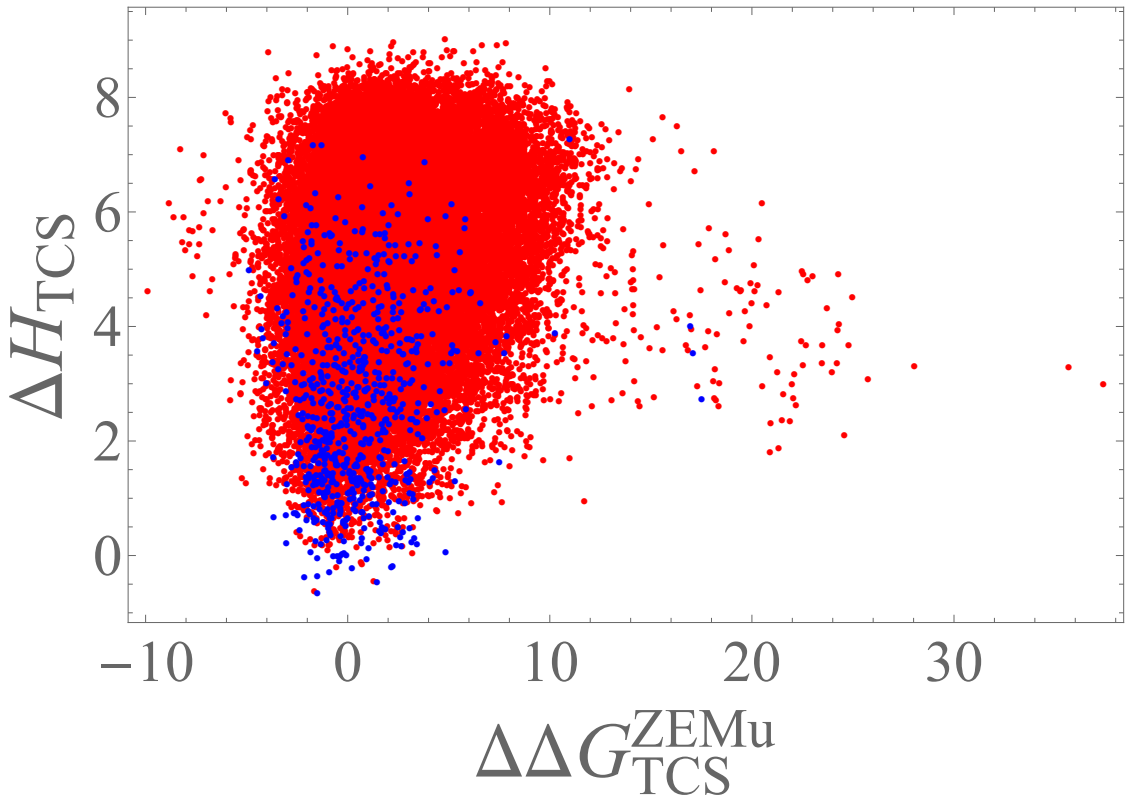

Figure S6. Scatter plot of ΔH_{HK} vs. $\Delta\Delta G_{\mathrm{TCS}}^{\mathrm{ZEMu}}$. A scatter plot of ΔH_{HK} vs. $\Delta\Delta G_{\mathrm{TCS}}^{\mathrm{ZEMu}}$ is plotted for the subset of 42,985 mutational variants explored by ZEMu.

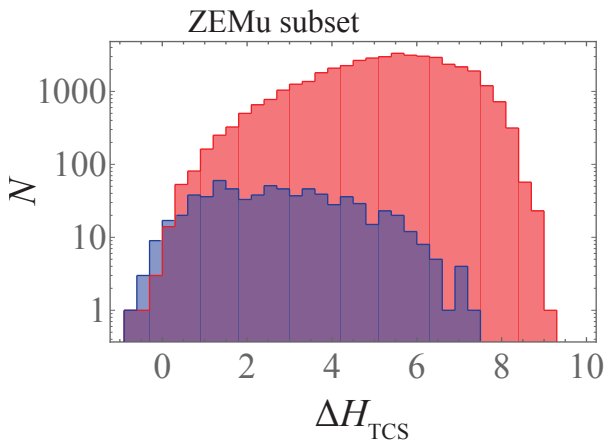

Figure S7. Histogram of mutational change in coevolutionary energy, ΔH_{TCS}, for subset of mutational variants explored by ZEMu calculation. A histogram of the mutational change in our coevolutionary energy, ΔH_{TCS} (Eq. 4), is plotted for the subset of 42,985 mutational variants explored by ZEMu, 702 functional (blue) and 42,283 non-functional (red). The analogous plot for the full dataset of 20^4 mutational variants is in Fig. 2A.

Database of TCS partners, Potts model, and code for calculating Eq. 4 can be obtained from: http://utdallas.edu/~faruckm/PublicationDatasets.html

Chen CM, Natale DA, Finn RD, Huang HZ, Zhang J, Wu CH, Mazumder R. 2011. Representative Proteomes: A Stable, Scalable and Unbiased Proteome Set for Sequence Analysis and Functional Annotation. PLoS One 6.
Dago AE, Schug A, Procaccini A, Hoch JA, Weigt M, Szurmant H. 2012. Structural basis of histidine kinase autophosphorylation deduced by integrating genomics,

molecular dynamics, and mutagenesis. Proceedings of the National Academy of Sciences 109:E1733-E1742.
Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Research 39:W29-W37.
Mechaly AE, Sassoon N, Betton JM, Alzari PM. 2014. Segmental Helical Motions and Dynamical Asymmetry Modulate Histidine Kinase Autophosphorylation. Plos Biology 12:e1001776.
Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, Sander C, Zecchina R, Onuchic JN, Hwa T, Weigt M. 2011. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. Proceedings of the National Academy of Sciences 108:E1293-E1301.
Podgornaia AI, Laub MT. 2015. Pervasive degeneracy and epistasis in a protein-protein interface. Science 347:673-677.
Wang C, Sang JY, Wang JW, Su MY, Downey JS, Wu QG, Wang SD, Cai YF, Xu XZ, Wu J, et al. 2013. Mechanistic Insights Revealed by the Crystal Structure of a Histidine Kinase with Signal Transducer and Sensor Domains. Plos Biology 11.
Weigt M, White RA, Szurmant H, Hoch JA, Hwa T. 2009. Identification of direct residue contacts in protein-protein interaction by message passing. Proc Natl Acad Sci U S A 106:67-72.
Yamamoto K, Hirao K, Oshima T, Aiba H, Utsumi R, Ishihama A. 2005. Functional characterization in vitro of all two-component signal transduction systems from Escherichia coli. Journal of Biological Chemistry 280:1448-1456.