doi:10.1038/nature11500

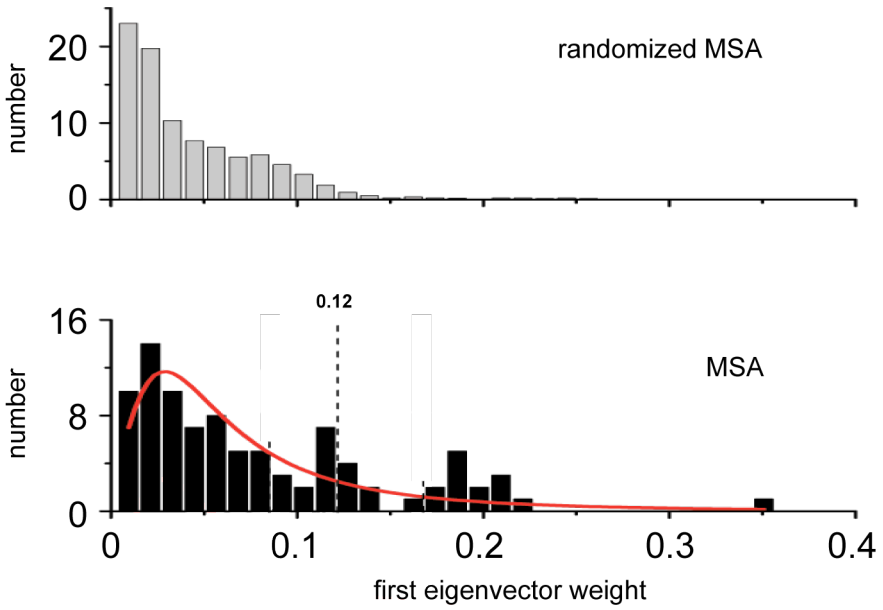

Figure S1: The sector in the PDZ domain family. Sectors are defined as groups of statistically coevolving amino acid positions in a protein family and are identified by the pattern of contribution to the top significant eigenmodes of the SCA correlation matrix [1]. The PDZ family shows a pattern consistent with a single sector; accordingly positions comprising this sector are identified by weights on the first eigenmode. The histograms show the pattern of contribution of PDZ positions towards the top eigenvector of the SCA matrix for either the actual sequence alignment (bottom panel) or for 100 trials of randomized alignments (top panel). The red curve is an empirical fit to a lognormal distribution, and the sector is defined by a cutoff that includes 20% of residues in the tail based on the cumulative density function from the fit. This cutoff selects residues outside of expectation for randomized alignments and provides a simple basis for sector definition. Importantly, the conclusions of this work are robust to the precise cutoff used for sector definition (Fig. S6).

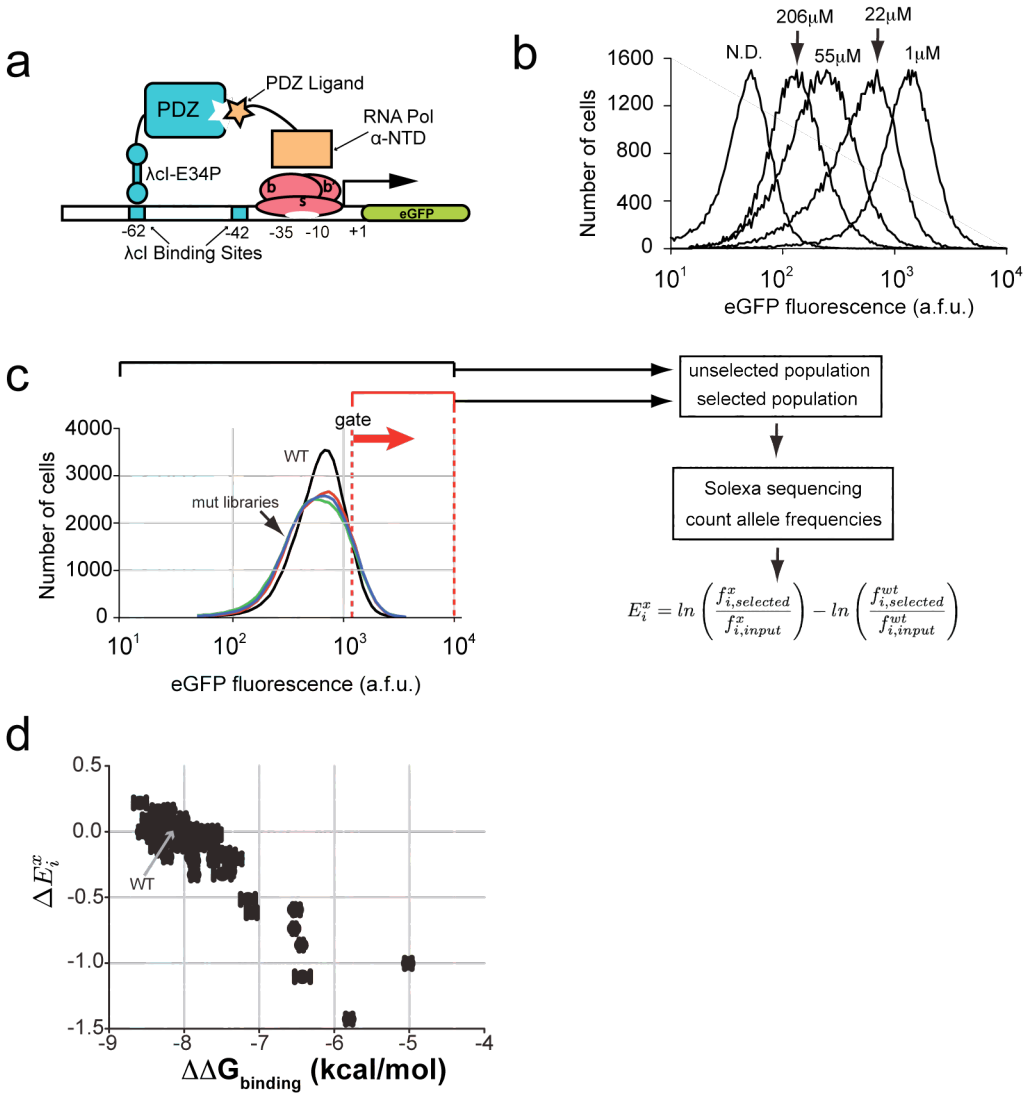

Figure S2: A quantitative high-throughput mutagenesis system for functional analysis of PDZ domains. **a**, the core component of the system is a bacterial two-hybrid assay in which transcription of GFP is a function of binding between the PDZ domain of interest (fused to a variant of the phage λ-cI DNA binding domain) and a target peptide ligand (fused to the N-terminal domain of the *E. coli* RNA polymerase α-subunit)[2,3,6]. **b**, Flow cytometry of bacteria carrying PSD95^{pdz3} variants with a range of functional effects. The data show that the steady-state distribution of GFP production is a function of affinity for the target ligand. **c**, Fluorescence-activated cell sorting (FACS) of bacteria carrying either wild-type (black) or libraries of all possible single point mutants of PSD95^{pdz3} (1578 variants total, colored). Sorting the population of bacteria showing expression of GFP higher than a specific threshold and Solexa-based deep sequencing of the sorted and unsorted distributions and counting of variant frequencies permits calculation of the functional cost of every mutation x at each position i relative to wild-type (ΔE_i^x). **d**, Comparison of ΔE_i^x with ligand binding energy (measured by fluorescence polarization, see methods) for a subset of 86 PSD95^{pdz3} mutants shows that the B2H/sequencing assay is an excellent reporter of PDZ function. Error bars for binding energy represent standard deviations of at least three measurements, and for ΔE_i^x , the range over two independent experiments.

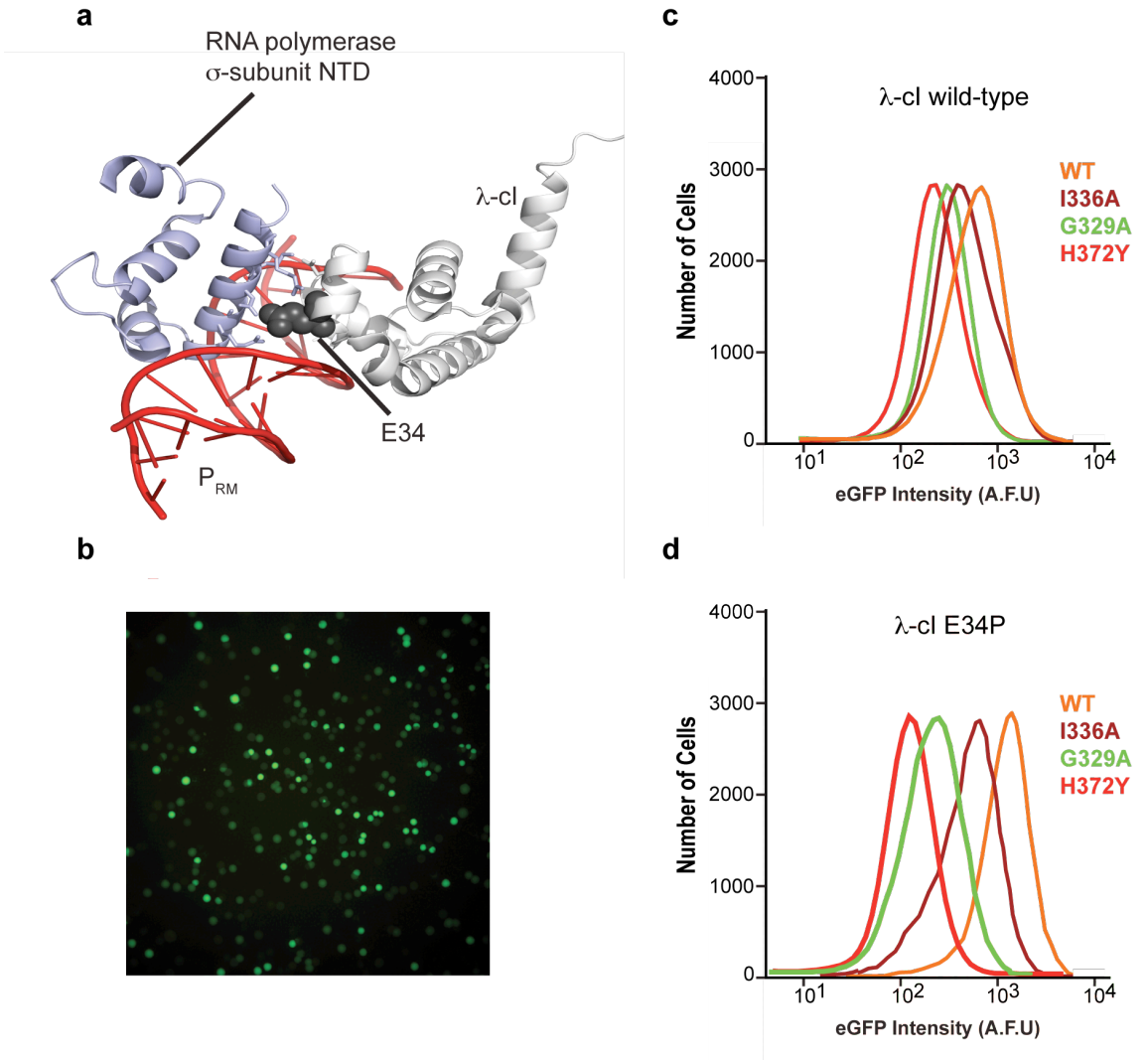

Figure S3: The expanded dynamic range of λ-cI E34P in the bacterial 2-hybrid assay and the role of λ-cI E34 in the ternary complex with PRM and the RNA polymerase σ-subunit. **a**, The crystal structure of the λ-repressor in complex with the RNA polymerase σ-subunit on the PRM promoter shows the interaction of λ-cI E34 with the polymerase[5]. Mutation of this amino acid disrupts the RNA polymerase α-subunit-independent activation of the polymerase by λ-cI[4,6]. Figure created from PDB 1RIO. **b**, A library of all amino acids at position 34 was screened in a colony format of the B2H assay, and E34P was identified as the mutation that produced minimal activation GFP expression in the absence of interacting proteins and maximal expression in the presence of a high-affinity interaction. **c-d**, The range of GFP expression for PSD95^{pdz3} mutants spanning the range of affinities sampled by the single-mutant library increases significantly in the E34P background relative to wild-type λ-cI. The mean GFP of PSD95^{pdz3} wild-type is 8.5-fold higher than that of H372Y in the E34P background, relative to 2.5-fold in the wild-type background, resulting in a significant increase in the dynamic range and sensitivity of the assay.

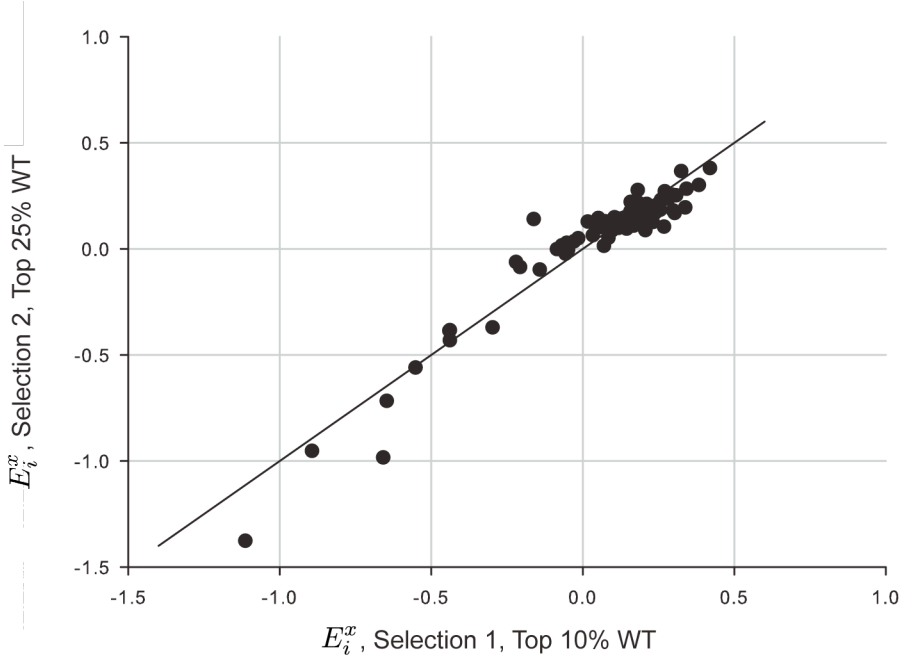

Figure S4: The correlation of the functional cost of mutations (ΔE_i^x) for two independent experiments with different FACS selection thresholds. For the library of 86 single mutations in PSD95^{pdz3}, B2H/FACS/Solexa sequencing experiments were carried out for two FACS gate thresholds, comprising the top 10% and top 25% of the GFP distribution for the wild-type PSD95^{pdz3} domain. The data show strong correlation between the two gates, demonstrating the reproducibility of the experiment across independent trials and the robustness of E_i^x to the choice of gate. As a result, values for both gates were averaged for each mutant. In this figure, the data show $E_i^x = \log\left[f_i^{x,sel} / f_i^{x,unsel} \right]$, the log ratios of the frequency of finding each variant in each gated population relative to the unselected population, not normalized for the corresponding wild-type values.

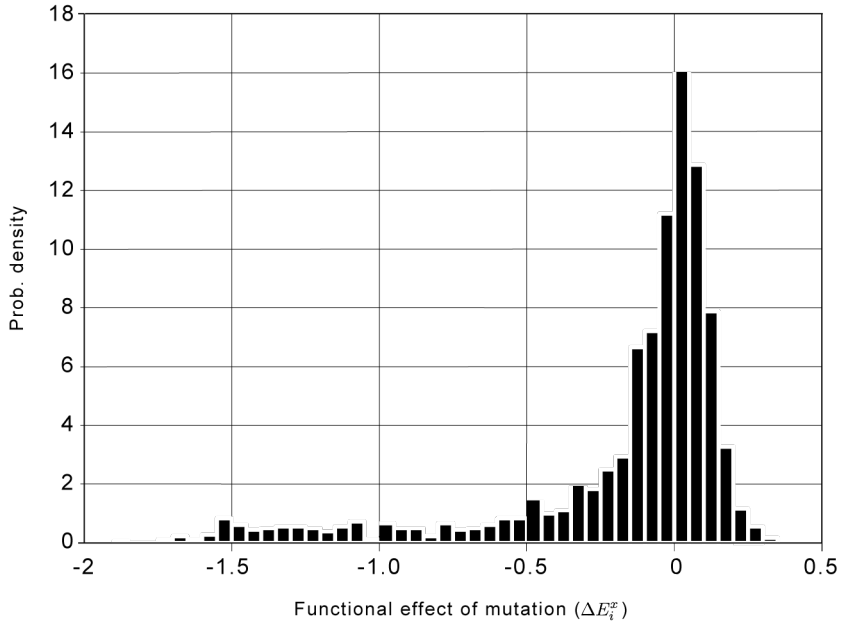

Figure S5: The distribution of mutational effects of complete single mutagenesis in PSD95^{pdz3}. The histogram of functional effects (ΔE_i^x) of all single amino acid mutations in PSD95^{pdz3} (Fig. 2a). The data show that the vast majority of mutations show functional effects centered around zero, highlighting the general tolerance of the protein to perturbation. A distinct subset of mutations show a long tail of destabilizing effects on ligand binding, showing that a small subset of mutations can significantly influence PDZ function.

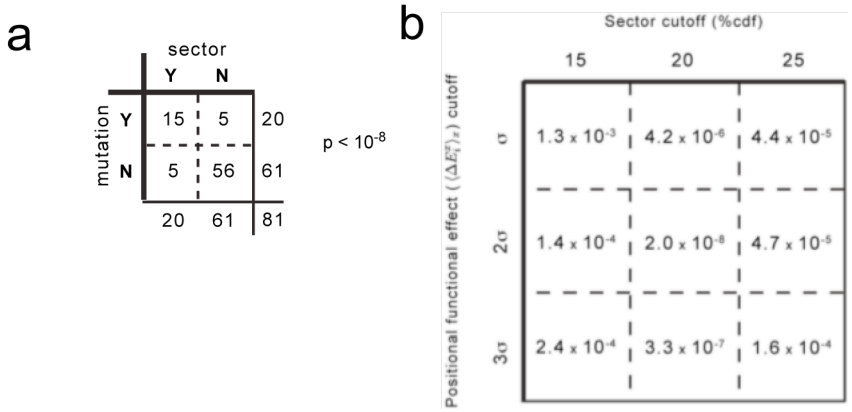

Figure S6: Statistical correlation between the sector definition and mutational sensitivity (by Fisher Exact Test). a, A 2 X 2 contingency matrix showing the relationship between sector positions in the PDZ family and mutationally significant residues in PSD95^{pdz3}; the Fisher Exact Test rejects the null hypothesis that these classifications are independent (p-value < 2 X 10^{-8}). **b,** Fisher Exact p-values calculated for three different significance cutoffs for the mutation data (1σ, 2σ, and 3σ), and for three different cutoffs for sector definition (75, 80, and 85 percent of the cumulative density function (cdf) of positional contributions to the sector (see Fig. S1). Overall, this analysis shows that sector definition and mutational significance are highly correlated and that this association is not strongly dependent on cutoffs used for either category.

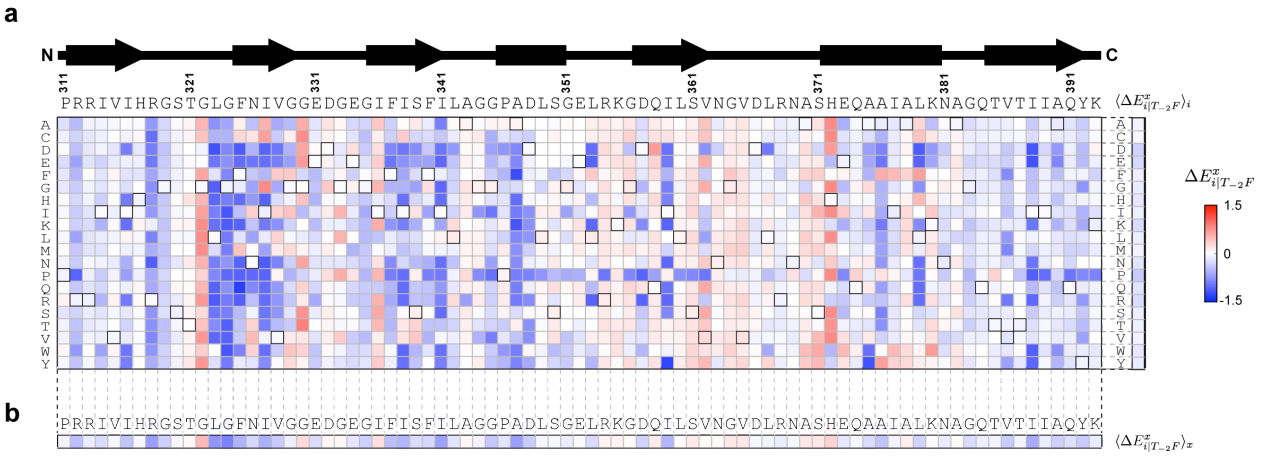

Figure S7: Complete single mutagenesis in PSD95^{pdz3} in the context of the Class II T$_{-2}$F peptide. a, The full data matrix showing the functional cost of every amino acid x (rows) at each position i (columns) relative to wild-type for PSD95^{pdz3} binding the T$_{-2}$F peptide ($\Delta E^{x}_{i|T_{-2}F}$). The data are shown colorimetrically, with blue representing loss of function (lower affinity) and red representing gain of function (higher affinity), and the wild-type amino acid at each position is indicated by the bold outline. The sequence of the wild-type domain is shown above with the secondary structure pattern. The average of these data taken over each row is shown at right, the mean functional cost of each amino acid substitution considered over all positions ($\left\langle \Delta E^{x}_{i|T_{-2}F} \right\rangle_i$). **b**, The average of the data taken over each column gives the mean functional cost of all amino acid substitutions at each position ($\left\langle \Delta E^{x}_{i|T_{-2}F} \right\rangle_x$).

II. Data

Supplementary data are provided as a MATLAB workspace. The variable "notes", reproduced here, contains information about the content of the workspace.

```
% Supplementary Data for McLaughlin et al.
% "The spatial architecture of protein function and adapatation"
%
% The variables in the workspace contain the data presented in this paper.
% There are three structure variables:
%
% (1) test_mutations: This structure contains the binding free energies (by
% fluorescence polarization...see methods) for the sub-library of 85 point
% mutations in PSD95pdz3. The data are referred to the text and in Fig. S2
% and S4.  The fields contained within this struture are:
%
%       delG_muts: [86x2 double]
%         delG_WT: [-8.2061 0.0506]
%     delG_Labels: 'kcal/mol standard_deviation'
%       mutations: {86x1 cell}
%
% where delG_muts and delG_WT give the equilbrium binding free energies for
% mutant and wild type proteins, delG_labels give the column labels for
% these data, and mutations give the row labels for each point mutant.
%
%
% (2) single_mutagenesis_cript:  This structure give the B2H/sequencing
% data for all amino acids at each positions in response to binding the
% wild-type CRIPT peptide ligand.  These data comprise Fig. 2a.  The fileds
% contained are:
%
%    data: [20x83 double]
%      aa: 'ACDEFGHIKLMNPQRSTVWY'
%     pos: [1x83 double]
%
% where data contains the delE(x,i) values for each amino acid x at each
% position i, aa contains the order of amino acids in the rows of data, and
% pos contains the positions labels for the columns of data.
%
%
% (3) single_mutagenesis_Tm2F:  This structure gives the B2H/sequencing
% data for all amino acids at each positions in response to binding the T
% minus 2 F variant peptide ligand, and in the data shown in Fig. S7. The
% epistasis data shown in Fig. 4 are derived by subtracting
% single_mutagenesis_Tm2F from single_mutagenesis_cript...the difference in
% the effect of each mutations at each position for the wild-type and Tm2F
% ligands. The variables are as in item 2 above.
```

III. References

1 Halabi, N., Rivoire, O., Leibler, S. & Ranganathan, R. Protein Sectors: Evolutionary Units of Three-Dimensional Structure. *Cell* **138**, 774-786 (2009).

2 Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res* **25**, 1203-1210, doi:gka167 [pii] (1997).

3 Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335-338, doi:10.1038/35002125 (2000).

4 Bushman, F. D., Shang, C. & Ptashne, M. A single glutamic acid residue plays a key role in the transcriptional activation function of lambda repressor. *Cell* **58**, 1163-1171, doi:0092-8674(89)90514-X [pii] (1989).

5 Jain, D., Nickels, B. E., Sun, L., Hochschild, A. & Darst, S. A. Structure of a ternary transcription activation complex. *Mol Cell* **13**, 45-53, doi:S1097276503004830 [pii] (2004).

6 Whipple, F. W., Ptashne, M. & Hochschild, A. The activation defect of a lambda-cI positive control mutant. *Journal of Molecular Biology* **265**, 261-265 (1997).