Supplementary Information for:

Mapping functional regions of essential bacterial proteins with dominant-negative protein fragments

Andrew Savinov¹,³, Andres Fernandez¹,⁴, and Stanley Fields¹,²,*

¹Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA
²Department of Medicine, University of Washington, Seattle, Washington 98195, USA
³Present address: Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA
⁴Present address: Program in Molecular and Cellular Biology, University of Washington and Fred Hutchinson Cancer Research Center, Seattle, Washington 98195, USA

*Correspondence to: Stanley Fields, ph. 206 616-4522, fields@uw.edu
Supplementary Figures

Figure S1: A random protein fragment assay of DHFR shows that inhibition is reading frame-dependent. (A) Histogram of enrichment (E) by in-frame (red) and out-of-frame (grey) fragments of ≤30 residues derived from the random fragmentation of *E. coli* DHFR. (B) Average enrichment per site due to in-frame fragments covering each residue, across the DHFR sequence (see Materials and Methods). Inhibitory peaks are indicated. Inset: overlay of per-site enrichment of in-frame and out-of-frame fragments. (C) DHFR crystallographic structure (PDB ID 7dfr) with N- and C-termini indicated, overlaid by the known refolding inhibitory fragment (*left*), alongside the C-terminal inhibitory peak (*middle*) and the N-terminal inhibitory peak (*right*) from the random fragment assay.
Figure S2: A protein fragment from interaction site 1 inhibits FtsZ polymerization in vitro. Left: Results from measurements of FtsZ filament polymerization by 90° light scattering (Materials and Methods) in the presence of either a protein fragment from interaction site 1 (FtsZ residues 166-195) or a neutral control fragment (FtsZ residues 34-63). In each reaction, FtsZ was present at 5.06 µM and the peptide was present at 18.4 µM. Error bars: s.e.m. from replicate measurements (N = 4 for the interaction site 1 fragment, N = 3 for the neutral control fragment). The asterisk indicates p < 0.05. Right: Inhibitory fragment map of FtsZ, plotted as in Figure 2, overlaid with the interaction site 1 and neutral control fragments.
Figure S3: A lack of correlation between cellular protein concentration and average susceptibility to fragment-based inhibition in cancer cell line MDA-MB-231. Plot of mean inhibitory effects of 40-residue fragments tiling 48 cancer-related proteins, from Ford et al. (15), as a function of the log of the cellular abundance of the full-length protein, from Lawrence et al. (31). Error bars: s.e.m. A linear fit to the data is shown ($R^2 = 0.005$).
Figure S4: Details of fragment length effects across diverse proteins. (A) – (H) Inhibitory fragment scans of *E. coli* proteins with tiling fragments, at a tile step size of 1 amino acid, comparing different fragment lengths. Data are plotted as in Figure 2, with individual fragment enrichments (E) and their s.e.m. across multiple experiments shown, except that 30-residue fragment data are shown in blue-grey; 20-residue fragment data in green; and 50-residue fragment data in magenta.
Figure S5: Average susceptibility to inhibition by 50-residue fragments remains negatively correlated with parental protein concentration. Mean inhibitory effects (selection enrichment $E$) of 20-, 30-, and 50-residue fragments tiling each cytosolic protein investigated, as a function of the log of the parental protein concentration. Error bars: s.e.m. Not all proteins were tiled with all fragment sizes. 50S ribosomal subunit protein L7/L12 (RplL) is shown with a grey fill color; linear fits to the non-ribosomal protein data are shown for the 30-residue and 50-residue data as dashed lines in corresponding colors. Note that the slopes are very similar. The faded grey dashed line near the 20-residue fragment data is a shifted copy of the trendline for the 30-residue data, suggesting that a similar slope may also be reasonable for 20-residue fragments; however, among the proteins for which 20-residue fragment libraries were generated, a leveling off is evident as the mean $E$ enters the neutral regime ($E \approx 0$), as expected, and there is insufficient sampling of mean $E$ values below zero to fit these data directly.
Figure S6: Parental protein concentration and fragment length effects explain only a fraction of individual fragment-to-fragment variability. (A) Mean inhibitory effects of 30-residue fragments derived from each protein investigated (large blue markers), and corresponding linear fit to the non-ribosomal proteins, as plotted in Figure 3, overlaid with the inhibitory effects (E) of individual fragments of each protein in pale blue. (B) Mean inhibitory effects of 20, 30, and 50-residue fragments of each protein, as in Figure 4B (large markers connected by dashed lines), overlaid with the inhibitory effects (E) of individual fragments (smaller, pale markers). Error bars indicating s.e.m. are included for the protein averages but are smaller than the markers.
Figure S7: Analysis of variance in inhibitory activity of out-of-frame fragments of DHFR. Fraction of total variation in inhibitory activity of out-of-frame protein fragments derived from the folA gene (as in Figure S1) attributable to each property is plotted, based on a nested ANOVA. Properties include protein fragment length and in-frame secondary structure features of the fragment-covered sequence region of the full-length gene; “fragment ID” designates fragment-specific effects not attributable to variation in any of the properties considered.
Figure S8: Comparison of shared fragment enrichments in the two tiling fragment libraries. Enrichment ($E$) of a set of 25-residue fragments of eGFP and RplL shared between the two tiling fragment libraries (Materials and Methods). The red line indicates $y = x$. Error bars: s.e.m. Note that points with lower measurement error tend to be highly reproducible between libraries, and the points farther from the diagonal generally exhibit high measurement error.
Supplementary Tables

Table S1: Results from analysis of variance performed on fragment-based inhibition data.

| Source of Variation                                    | Sum of Squared Errors | Fraction of Total Variation (%) | d.f. | P-value   |
|--------------------------------------------------------|------------------------|---------------------------------|------|-----------|
| Fragment ID                                            | 17371.55               | 71.6946                         | 9921 | 2.97E-93  |
| Fragment length, aa                                     | 2419.80                | 9.9868                          | 40   | 1.46E-227 |
| Parental protein                                        | 1021.78                | 4.2170                          | 9    | 2.30E-191 |
| Fragment contains beta-strand (from structure)         | 173.34                 | 0.7154                          | 1    | 1.88E-97  |
| Fragment contains alpha-helix (from structure)         | 118.07                 | 0.4873                          | 1    | 4.18E-79  |
| Fragment contains turn (from structure)                | 71.92                  | 0.2968                          | 1    | 1.87E-58  |
| Hydrophobicity (Kyte-Doolittle)                         | 169.05                 | 0.6977                          | 1    | 3.39E-96  |
| Relative position in sequence (N-->C)                  | 53.77                  | 0.2219                          | 2    | 6.29E-47  |
| Instability (Guruprasad)                               | 2.54                   | 0.0105                          | 1    | 1.57E-04  |
| Net charge (Lehninger)                                 | 0.56                   | 0.0023                          | 2    | 2.01E-01  |
| Error                                                  | 52.08                  | 0.2150                          | 300  | –         |
Table S2: Table of single-codon indices appended to protein fragments in the tiling library.

| Protein | N-terminal index codon (amino acid), $X_1$ | C-terminal index codon (amino acid), $X_2$ |
|---------|---------------------------------------------|---------------------------------------------|
| DHFR   | gca (A)                                     | ctg (L)                                     |
| GyrA   | gca (A)                                     | agt (S)                                     |
| IleS   | gca (A)                                     | gca (A)                                     |
| FtsZ   | ctg (L)                                     | ctg (L)                                     |
| RpoB   | ctg (L)                                     | gca (A)                                     |
| RplL   | ctg (L)                                     | agt (S)                                     |
| Ssb    | agt (S)                                     | ctg (L)                                     |
| GroEL  | agt (S)                                     | gca (A)                                     |
| GroES  | agt (S)                                     | agt (S)                                     |
| LptG   | tcc (S)                                     | ctg (L)                                     |
Fragment mapping reveals promising target sites for inhibitor development: additional discussion

DHFR is the target of the antibiotic trimethoprim, which binds at the enzymatic active site and inhibits catalysis (51). Our results suggest that inhibition of folding by compounds that bind to the N-terminal beta strand may be another approach to target this important enzyme. Another route would be welcome given the serious issue of trimethoprim resistance (52); due to the distinct mechanisms, a putative folding inhibitor would likely not be subject to the same resistance mutations.

A number of compounds inhibiting bacterial DNA gyrase, including the fluoroquinolones currently employed clinically, function by trapping the enzyme in a DNA cleavage complex (53). Other compounds, such as the aminocoumarins, interfere with ATPase activity (53). Our finding of an inhibitory fragment peak localized to the ends of the C-gate arms of GyrA, which form a homomeric head-to-head complex to close the gate, suggests that compounds targeting these C-gate closure sites would provide another means to inhibit gyrase. Such inhibitors would rely on a distinct binding mode from the aforementioned compounds, which bind at sites at or near the DNA gate or the ATPase site (53, 54). In line with our findings, a monoclonal antibody against *Mycobacterium smegmatis* and *Mycobacterium tuberculosis* GyrA appears to bind at the same location as our fragments and prevent C-gate dimerization. Furthermore, the antibody epitope corresponds to *E. coli* GyrA residues 340-402 (55), fully overlapping with the inhibitory fragment peak centered around residue 386.

Single-stranded DNA binding protein (Ssb) is a desirable antibiotic target because inhibition should interfere with DNA replication and repair. Screening has therefore been performed for Ssb inhibitors (56–58), leading to the identification of numerous compounds; some of these were
shown to interfere with critical interactions of eight conserved C-terminal residues with genome maintenance proteins (57, 58). Unlike the case for FtsZ, we did not observe an inhibitory fragment peak mapping near this interaction-mediating C-terminal tail. Instead, we observed a strong peak mapping to the single alpha-helical element of this protein, suggesting a novel approach for inhibitor development: mimicking the interactions formed by the alpha helix in Ssb dimers to disrupt dimer and tetramer formation.

A further note regarding the transferability of inhibitory sites identified in *E. coli* to other bacteria: The range of bacteria potentially targetable based on these fragment scan results is extensive, given that protein-peptide interactions are based heavily on structural properties. Therefore, more distantly related orthologs retaining the relevant structural features are likely susceptible to the same types of inhibitors, especially those based directly on peptide fragments.

**SI References**

51. D. A. Matthews, *et al.*, Refined crystal structures of Escherichia coli and chicken liver dihydrofolate reductase containing bound trimethoprim. *J. Biol. Chem.* **260**, 381–391 (1985).

52. A. Brolund, M. Sundqvist, G. Kahlmeter, M. Grape, Molecular characterisation of trimethoprim resistance in Escherichia coli and Klebsiella pneumoniae during a two year intervention on trimethoprim use. *PLoS One* **5**, e9233 (2010).

53. T. Khan, *et al.*, DNA gyrase inhibitors: Progress and synthesis of potent compounds as antibacterial agents. *Biomed. Pharmacother.* **103**, 923–938 (2018).

54. B. D. Bax, G. Murshudov, A. Maxwell, T. Germe, DNA Topoisomerase Inhibitors: Trapping a DNA-Cleaving Machine in Motion. *J. Mol. Biol.* **431**, 3427–3449 (2019).

55. U. H. Manjunatha, A. Maxwell, V. Nagaraja, A monoclonal antibody that inhibits mycobacterial DNA gyrase by a novel mechanism. *Nucleic Acids Res.* **33**, 3085–3094 (2005).

56. J. G. Glanzer, *et al.*, Identification of inhibitors for single-stranded DNA-binding proteins in eubacteria. *J. Antimicrob. Chemother.* **71**, 3432–3440 (2016).

57. A. F. Voter, *et al.*, A High-Throughput Screening Strategy to Identify Inhibitors of SSB Protein–Protein Interactions in an Academic Screening Facility. *SLAS Discov.* **23**, 94–101 (2018).

58. D. Lu, D. A. Bernstein, K. A. Satyshur, J. L. Keck, Small-molecule tools for dissecting the roles of SSB/protein interactions in genome maintenance. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 633–638 (2010).