



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

Andrew Savinov<sup>a,1</sup> , Andres Fernandez<sup>a,2</sup> , and Stanley Fields<sup>a,b,3</sup>

Edited by William DeGrado, University of California, San Francisco, CA; received January 4, 2022; accepted May 8, 2022

Massively parallel measurements of dominant-negative inhibition by protein fragments have been used to map protein interaction sites and discover peptide inhibitors. However, the underlying principles governing fragment-based inhibition have thus far remained unclear. Here, we adapted a high-throughput inhibitory fragment assay for use in *Escherichia coli*, applying it to a set of 10 essential proteins. This approach yielded single amino acid resolution maps of inhibitory activity, with peaks localized to functionally important interaction sites, including oligomerization interfaces and folding contacts. Leveraging these data, we performed a systematic analysis to uncover principles of fragment-based inhibition. We determined a robust negative correlation between susceptibility to inhibition and cellular protein concentration, demonstrating that inhibitory fragments likely act primarily by titrating native protein interactions. We also characterized a series of trade-offs related to fragment length, showing that shorter peptides allow higher-resolution mapping but suffer from lower inhibitory activity. We employed an unsupervised statistical analysis to show that the inhibitory activities of protein fragments are largely driven not by generic properties such as charge, hydrophobicity, and secondary structure, but by the more specific characteristics of their bespoke macromolecular interactions. Overall, this work demonstrates fundamental characteristics of inhibitory protein fragment function and provides a foundation for understanding and controlling protein interactions in vivo.

inhibitory peptides | massively parallel measurements | protein interactions | dominant-negative | *E. coli*

Peptides have the capability to act as potent modulators of biological function by binding to proteins at specific sites, thereby altering protein activity. Compared to small molecules, peptides have advantages in specificity, ability to target large and shallow interaction surfaces (1, 2), and genetic encodability. Organisms have leveraged this phenomenon by encoding peptide antimicrobials (3) as well as biological peptides termed small proteins or miniproteins, which frequently regulate larger proteins (4–7). Synthetic peptides can also be engineered for such inhibitory functions, providing potential avenues for novel cancer therapeutics (8), antivirals (9), and antibiotics (10).

Fragments of native protein sequences can reproduce native interactions with binding partners while lacking other functions of the full-length protein. Similar to the case with truncation mutants (11), these binding events allow such fragments to compete with their parental protein for its native interactions and thereby act as dominant-negative inhibitors. While most such examples of dominant-negative inhibition involve intermolecular interactions, protein fragments can also inhibit intramolecular interactions involved in protein folding (12, 13). Given their interference with native binding contacts, inhibitory fragments permit functional mapping of proteins, identifying important interaction sites at sub-gene resolution. DNA sequencing-based methods allow the identification of such fragments in a massively parallel manner (14–16) to map functional domains (14, 15). In these assays, a growth selection is performed on cells carrying a protein fragment library under conditions in which the parental protein is important for growth. Depletion of fragment-encoding sequences from the population, reflecting inhibitory activity, is quantified by performing high-throughput sequencing before and after selection. Peptides derived from such approaches are capable of pulling down their expected interaction partners (15).

Here, we adapted this method to map a diverse set of 10 essential proteins in *Escherichia coli*, reasoning that the abundance of structural and functional information available for this organism should reveal biophysical principles underlying fragment-based inhibition and empower applications of inhibitory fragments as tools to study protein function in vivo. Measurements of *E. coli* growth permitted functional mapping at single residue resolution, revealing the importance of dozens of structural elements that can act as dominant-negative inhibitors. Many inhibitory protein fragments mapped to protein–protein interaction sites,

## Significance

Peptide fragments derived from protein sequences can inhibit interactions of their parental proteins, providing tools to study protein function in vivo. Here, we employed a massively parallel assay to measure inhibition of *Escherichia coli* growth by fragments tiling the sequences of 10 of its essential proteins. We leveraged these data to decipher principles of fragment-based inhibition, demonstrating that parental protein concentration drives activity and characterizing how fragment length interplays with activity and specificity. We employed statistical analysis to parse the roles of biophysical properties in fragment-to-fragment variation, finding that the specific characteristics of each fragment largely drive its inhibitory activity. These results advance our understanding of protein interactions in vivo and have implications for the rational design of peptide inhibitors.

Author contributions: A.S. and S.F. designed research; A.S. and A.F. performed research; A.S. analyzed data; and A.S. and S.F. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2022 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

<sup>1</sup>Present address: Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142.

<sup>2</sup>Present address: Program in Molecular and Cellular Biology, University of Washington and Fred Hutchinson Cancer Research Center, Seattle, Washington 98195.

<sup>3</sup>To whom correspondence may be addressed. Email: fields@uw.edu.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2200124119/-/DCSupplemental>.

Published June 24, 2022.

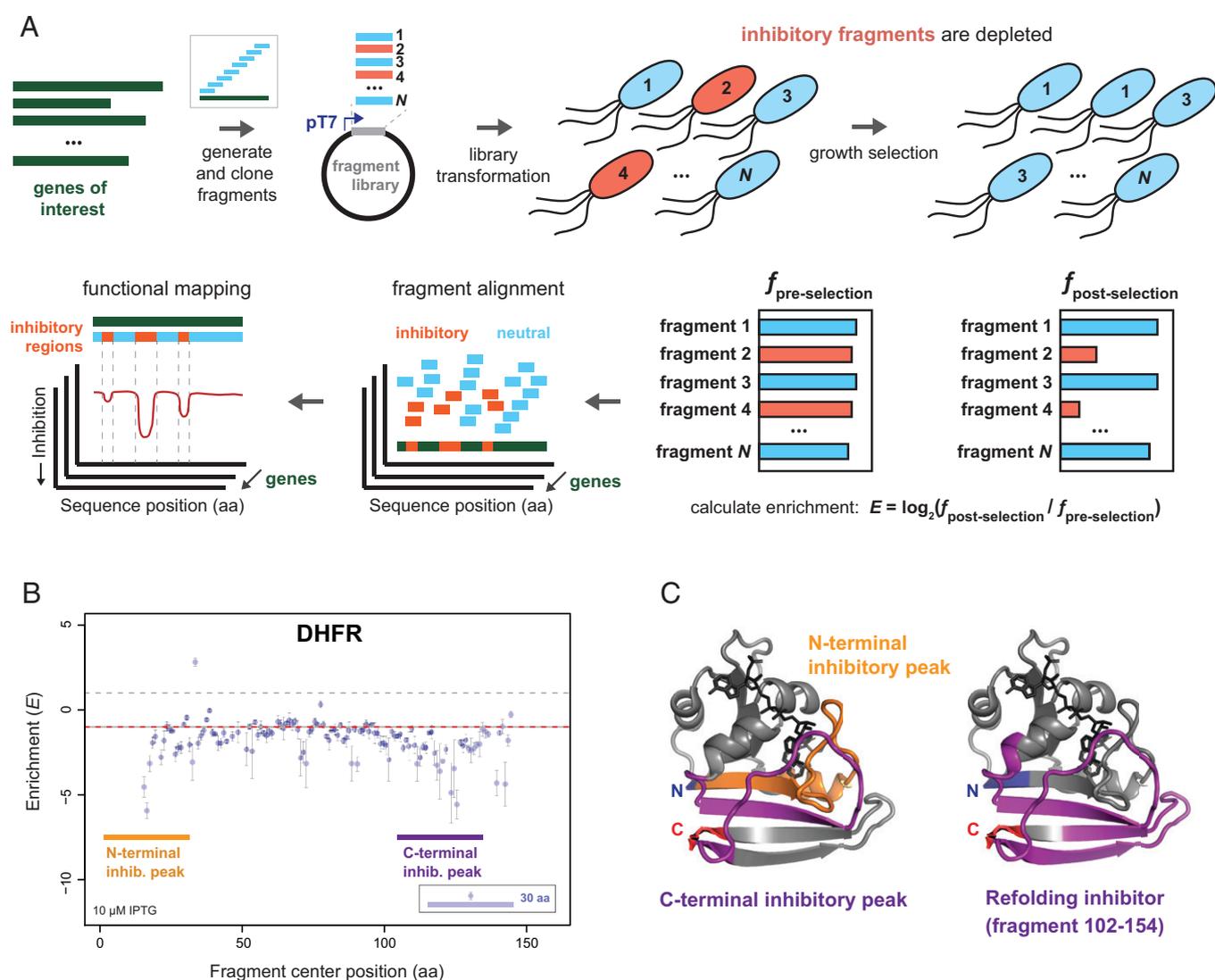
and others mapped to regions that inhibit intramolecular folding interactions. Leveraging the data across diverse proteins, we performed analyses that revealed the importance of fragment length, as well as a strong negative correlation between the cellular concentrations of proteins and their susceptibility to fragment-based inhibition. We also employed ANOVA analysis and AlphaFold modeling to delve into the roles of biophysical properties in the inhibitory activity of fragments. In sum, our results provide key principles for understanding protein fragment inhibition in living cells.

## Results

**A High-Throughput Protein Fragment Assay Recapitulates a Folding-Inhibitory Region of Dihydrofolate Reductase.** To scan for dominant-negative inhibitory fragments of *E. coli* proteins, we developed a high-throughput assay (Fig. 1A and

*Materials and Methods*) analogous to approaches employed in yeast and human cells (14, 15, 17). We combined measurements of growth selection depletion with an alignment of fragments to their parental proteins, generating maps of inhibitory activity as a function of sequence position. To systematically cover all positions of a protein with fragments of tunable length, we array-synthesized libraries of DNA fragments tiling across the coding sequence of each protein with single residue resolution (*Materials and Methods*), similar to the approach of Ford et al. (15).

We sought to establish this approach in *E. coli* using a model protein with a known inhibitory fragment acting by a well-characterized mechanism: dihydrofolate reductase (DHFR), a monomeric protein for which a C-terminal fragment inhibits the refolding of the full-length protein in vitro (12). We reasoned that similar protein fragments might also affect DHFR folding in vivo. We generated an inhibitory fragment map by plotting the inhibitory activity (selection enrichment,  $E$ ) of 30-residue fragments as a



**Fig. 1.** Mapping protein functional regions with dominant-negative inhibitory fragments. (A) Schematic describing the high-throughput inhibitory fragment assay. Fragments generated from genes of interest are cloned into an expression vector and transformed into *E. coli*. A growth selection is performed, and inhibitory fragment depletion from the population is quantified by high-throughput sequencing, allowing determination of the enrichment  $E = \log_2(f_{\text{post-selection}}/f_{\text{pre-selection}})$ , where  $f$  denotes frequency, with  $E < 0$  indicating depletion. Fragment alignment to the parental protein sequence allows identification of inhibitory regions. (B) Tiling inhibitory fragment scan results for *E. coli* DHFR, at a tile step size of 1 amino acid (aa). Enrichment ( $E$ ) is plotted as a function of fragment center position. Points: individual fragment measurements; error bars: SEM from multiple experiments; dashed horizontal lines: guide lines at  $\pm 1E$  ( $\pm$ twofold change in frequency). Colored rectangular markers indicate positions of inhibitory (inhib.) peaks, with width indicating a representative fragment. (C) DHFR crystallographic structure (PDB ID 7dfr) with N- and C-termini indicated, overlaid by the N-terminal and C-terminal inhibitory peaks (Left) and the known refolding-inhibitory fragment (Right) from the work of Hall and Frieden (12). A higher-resolution version of this figure is available on figshare (18).

function of sequence position (Fig. 1B). We observed an inhibitory peak in the same C-terminal region of the protein as the known folding inhibitor (12) (Fig. 1C), suggesting that these fragments also inhibit folding in cells. We additionally observed an N-terminal inhibitory peak, mapping to a region involved in binding dihydrofolate, that might inhibit DHFR via substrate titration (Fig. 1C). We used a Nextera transposase-based random fragment library (14) of the DHFR-encoding *folA* gene to demonstrate coding frame-dependence of inhibition, with out-of-frame fragments predominantly neutral ( $E \approx 0$ ) (SI Appendix, Fig. S1); these results underscore that fragment inhibitory activity is sequence dependent and acts at the protein level.

**Tiling Fragment Scans of Essential *E. coli* Proteins Map Key Interaction Sites.** Given the results with DHFR, we applied high-throughput inhibitory fragment mapping to a set of nine additional bacterial proteins (Table 1), spanning several core processes: DNA replication (DNA gyrase subunit A, GyrA; single-stranded DNA binding protein, Ssb), transcription (RNA polymerase subunit beta, RpoB), translation (isoleucyl tRNA synthetase, IleS; 50S ribosomal protein L7/L12, RplL), protein quality control (GroEL; GroES), cell division (FtsZ), and outer membrane maintenance (lipopolysaccharide transport protein G, LptG). Overall, this assay yielded a raw enrichment signal that delineated inhibitory peaks without requiring coverage-based smoothing (Fig. 2A–G), as employed previously (14, 15). An overview of the results is presented in Tables 1 and 2.

We first considered measurements with tiling fragments of 30 residues, as used with DHFR. The results for FtsZ (Fig. 2A, *Left*) exemplify what can be learned using this approach. FtsZ forms filaments that are essential for bacterial cell division (20); monomers polymerize head-to-tail (21), providing a well-characterized binding interface (Fig. 2A, *Right*). Several binding contacts from either side of this interface yielded inhibitory fragment peaks. The two strongest inhibitory peaks mapped to loop structures with flanking helices, which formed portions of the oligomerization interface that we term sites 1 and 2' (Fig. 2A); inhibitory fragment peaks also mapped to the complementary regions of the adjacent FtsZ monomer, sites 1' and 2 (Fig. 2A). These results suggest that fragments from these regions inhibit filament assembly, similar to the miniprotein MciZ (22); and indeed, we found that a fragment from site 1

inhibited FtsZ polymerization in vitro by ~20% (SI Appendix, Fig. S2). Conversely, no inhibitory peaks were observed mapping to interaction interface regions 3 and 3' (residues 65 to 69 and 1 to 11; Fig. 2A), consistent with a B-factor analysis suggesting that the interactions of these sites are weak (21). Outside of the head-to-tail monomer interface, the fragment scan detected an inhibitory region at the junction of the intrinsically disordered C-terminal linker and the C-terminal-most 15 residues, which enable modulation of FtsZ activity by binding partners such as FtsA, ZipA, and MinC (23). This result demonstrates the utility of inhibitory fragment scanning for functionally mapping intrinsically disordered regions for which no structural information is available.

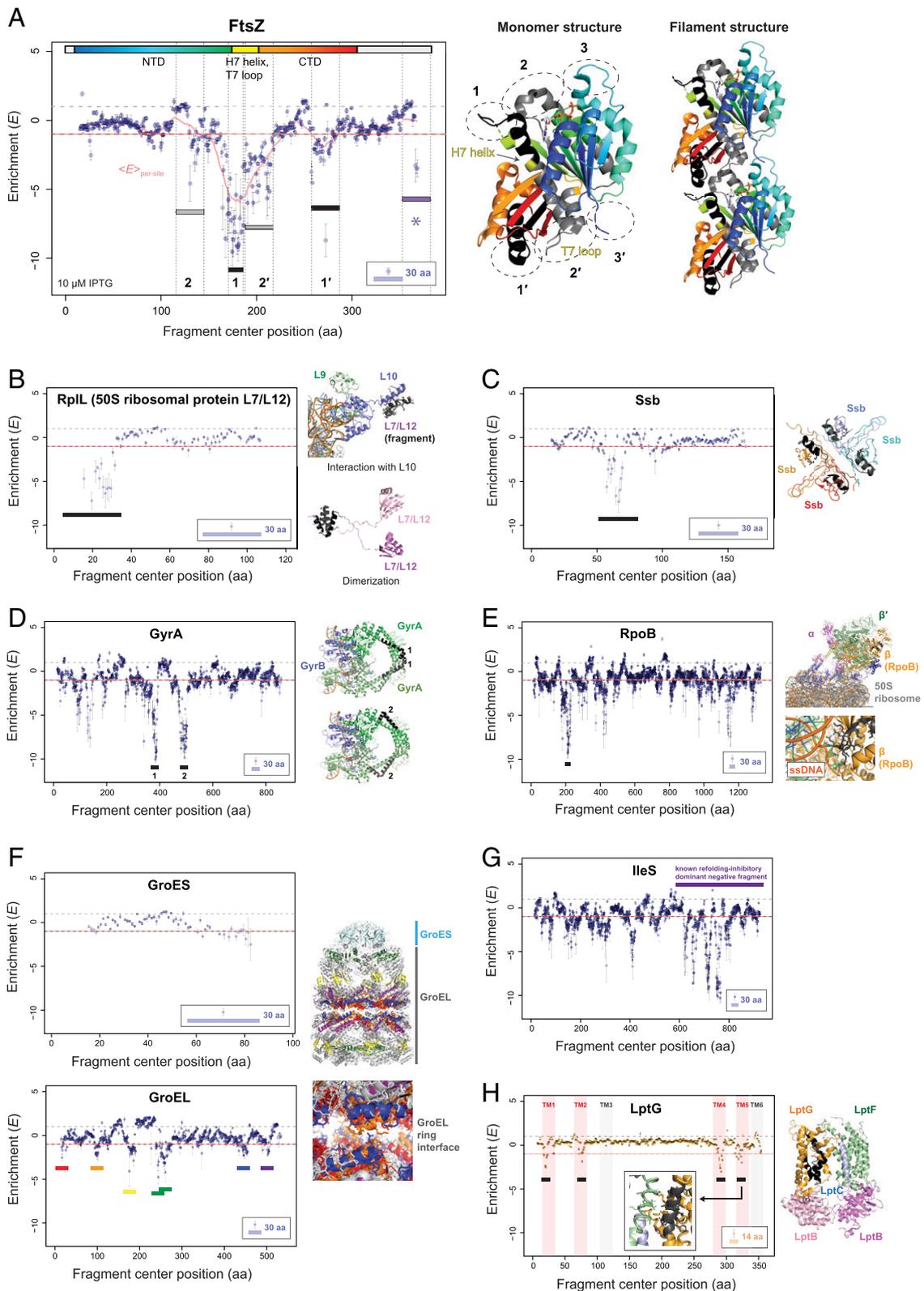
Inhibitory fragments mapped to key interaction sites across the sequences of several additional proteins. For RplL, the single inhibitory peak was localized to an N-terminal region responsible for RplL dimerization and binding to 50S ribosomal protein L10; both interactions are required to form the ribosomal stalk (Fig. 2B). For Ssb, the single inhibitory peak mapped to an alpha-helical region involved in the dimerization of Ssb, which further tetramerizes (Fig. 2C). For GyrA, multiple inhibitory peaks were observed (Fig. 2D), with the strongest two mapping to the C-Gate formed by a GyrA dimer interface; the C-Gate must properly open and close to control the directionality of DNA strand transfer during gyrase activity, with errors resulting in double-stranded DNA (dsDNA) breaks (24). Several peaks were also identified for RpoB, including one localized to a site that interacts with single-stranded DNA (ssDNA) in the context of a transcription-translation “expressome” complex (Fig. 2E; ref. 25). For GroEL, several inhibitory fragment peaks mapped to the interface between the two stacked GroEL heptamers (Fig. 2F, *Right* and *Bottom Left*), suggesting inhibition of proper GroEL/ES complex assembly and chaperone function. On the other hand, we observed no inhibitory peaks in a fragment scan of GroES (Fig. 2F, *Top Left*).

For IleS, the strongest inhibitory peaks clustered in the C-terminal region (Fig. 2G). The structure of *E. coli* IleS has not been determined, but these results are congruent with functional data; the strong C-terminal inhibitory peaks overlap with a large C-terminal fragment (residues 585 to 939) that acts as a folding inhibitor in vitro and a dominant negative in vivo (13). The results of the inhibitory fragment scan suggest that this folding-inhibitory activity is localized more finely in the sequence, with several distinct peaks at residues ~600 to 800.

**Table 1. Overview of inhibitory fragment assay targets, their properties, and overall results**

<i>E. coli</i> gene	Protein	Length (aa)	Cellular abundance (protein molecules/cell)	$\langle E \rangle$ for 30-aa tiling fragments
<i>folA</i>	Dihydrofolate reductase (DHFR)	159	2,515	-1.78
<i>gyrA</i>	DNA gyrase subunit A (GyrA)	875	5,219	-1.21
<i>ileS</i>	Isoleucyl tRNA synthetase (IleS)	938	6,143	-1.26
<i>ftsZ</i>	Cell division protein FtsZ	383	6,750	-1.20
<i>rpoB</i>	RNA polymerase subunit beta (RpoB)	239	16,156	-0.99
<i>rplL</i>	50S ribosomal protein L7/L12 (RplL)	1,342	472,965	-0.84
<i>ssb</i>	Single-stranded DNA binding protein (Ssb)	121	14,444	-0.66
<i>groL</i>	60 kDa chaperonin (GroEL)	178	52,165	-0.28
<i>groS</i>	10 kDa chaperonin (GroES)	548	59,001	-0.21
<i>lptG</i>	Lipopolysaccharide transport protein G (LptG)	97	775	—

Properties of each *E. coli* protein assayed by inhibitory fragment scanning are compiled, together with the average inhibitory effects of tiling 30-residue fragments. Cellular concentrations of proteins are from ref. 19. aa, amino acid.



**Fig. 2.** Inhibitory fragments map interaction interfaces of essential bacterial proteins. Inhibitory fragment scans of *E. coli* proteins with tiling fragments, at a tile step size of 1 amino acid (aa). Enrichment ( $E$ ) is plotted as a function of fragment center position. Points: individual fragment measurements; error bars: SEM from multiple experiments; dashed horizontal lines: guide lines at  $\pm 1E$  ( $\pm$ twofold change in frequency). Colored rectangular markers indicate positions of inhibitory peaks (with width indicating a representative fragment, unless otherwise noted). In protein structures to the right of the plots, locations of inhibitory peaks (generally representative fragments) are overlaid on the structure in a color matching the corresponding peak marker in the plot to the left (generally black). (A–G) 30-residue tiling fragment scans of indicated proteins. (H) A 14-residue tiling fragment scan of LptG. Additional features are present as indicated: (A), salmon line indicates average  $E$  per residue due to all fragments covering that residue; the width of the peak 1 marker indicates the peak width (16 residues); the lavender asterisk marks the inhibitory peak that falls in the C-terminal intrinsically disordered region; and the colored bar at the top of the plot indicates the layout of structural elements in the linear sequence; NTD, N-terminal domain, CTD, C-terminal domain. (H) Shaded regions denote indicated topological domains. PDB IDs for structures: (A) FtsZ: 6unx; (B) RplL: 3j7z (Upper) and 1rq (Lower); (C) Ssb: 1eqq; (D) GyrA: 6rks; (E) RpoB: 6x9q; (F) GroEL-ES: 1aon; (H) LptG: 6mi7. No structure is shown for *E. coli* IleS because it has not yet been determined. A higher-resolution version of this figure is available on figshare (18).

**Table 2. Overview of highlighted inhibitory peaks**

<i>E. coli</i> protein	Example inhibitory peak locations in sequence (aa)	Notes
DHFR	17, 120	Peaks map to a substrate-binding region (17) and a known refolding-inhibitory region (120)
GyrA	386, 493	Strongest two peaks both map to dimeric C-gate, which controls DNA strand transfer directionality; one peak (386) maps to the homodimer contact required for gate closure, the other (493) to the C-gate “hinge” structure
IleS FtsZ	Cluster of peaks at positions 600 to 800 131, 178, 201, 273, 368	Example peaks map within a known refolding-inhibitory region Example peaks map to reciprocal interaction sites for filament formation (131 and 201; 178 and 273) and an intrinsically disordered region implicated in regulatory interactions with other proteins (368)
RpoB	211	Strongest peak maps to a protein region that interacts with ssDNA
RplL	20	Peak maps to a region of RplL (50S L7/L12) involved in homodimerization and interaction with L10, required for ribosome stalk formation
Ssb	67	Peak maps to only region of defined secondary structure, involved in Ssb tetramer formation
GroEL	18, 102, 176, 244/261, 446, 501	Multiple peaks map to the GroEL–GroEL heptameric ring interface (18, 102, 446, and 501)
GroES	—	No inhibitory peaks observed
LptG	21, 77, 292, 324	All four peaks map to specific transmembrane domains (TM1, TM2, TM4, TM5)

Summary of example inhibitory peaks highlighted in the text, for each *E. coli* protein assayed by inhibitory fragment scanning. aa, amino acid.

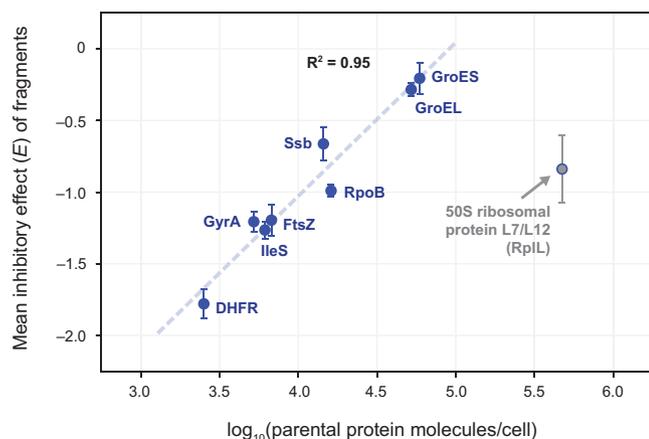
To determine whether membrane proteins are amenable to fragment-based inhibition, we performed an inhibitory fragment scan of LptG, a transmembrane protein involved in lipopolysaccharide transport from the inner to the outer membrane. To better resolve functions of the multiple short cytosolic, transmembrane, and periplasmic domains found throughout the sequence, we scanned LptG with fragments of 14 residues. The fragment scan yielded four inhibitory peaks, which all mapped to transmembrane alpha-helices (TM1, TM2, TM4, or TM5; Fig. 2*H*). It seems unlikely that this inhibition arises from mechanisms generic to hydrophobic peptides (e.g., aggregation), as fragments mapping to the remaining two transmembrane alpha-helices, TM3 and TM6, did not yield inhibitory peaks, although we cannot rule out the possibility that the TM3 and TM6 fragments are less stable. We hypothesize that the inhibitory fragments localize to the membrane and inhibit LptG folding and/or interactions with other membrane proteins. For example, structures of the Lpt complex (26) reveal that LptG TM5 forms a binding interaction with the TM1 helix of LptF. Similarly, LptG TM1 interacts with the transmembrane helix of LptC (26), although the inhibitory fragment peak maps to the center of the TM1 helix rather than the end that directly binds LptC. Membrane localization of these hydrophobic alpha-helical fragments is consistent with posttranslational membrane insertion of short hydrophobic sequences (e.g., an 18-mer helix of YkgR) by the signal recognition particle (SRP) (27). Given the low protein copy number of LptG (Table 1), even limited membrane entry by its inhibitory fragments may perturb function.

Activity of fragments in this assay is a function of both inherent inhibitory activity and steady-state cellular concentration. Transcription and translation levels should be similar due to the common expression system, but fragments may vary in stability. The common N-terminal tag (*Materials and Methods*) should reduce stability differences (28, 29); nonetheless, an absence of inhibitory

activity may in some cases reflect fragment instability, even for an inherently active fragment.

**A Simple Principle Underlies Variable Susceptibility of Proteins to Fragment-Based Inhibition.** We sought to understand the factors driving the varied susceptibility of different proteins to fragment-based inhibition. For example, fragments derived from GroEL and especially GroES had lower average inhibitory activity than fragments derived from the other *E. coli* proteins (Fig. 2*F*, *Left* and Table 1). In contrast, fragments derived from DHFR exhibited a baseline inhibitory effect (Fig. 1*B*), resulting in these fragments exhibiting the highest average inhibitory activity of the proteins tested (Table 1). Many factors could play a role in this differential susceptibility, including localization, folding stability, structural characteristics, number of interaction partners, strength of native binding interactions, and cellular concentration of the parental proteins. This last factor seemed likely to be important, as a given concentration of fragments that bind competitively with native interactions should be less inhibitory when the parental protein is more abundant.

We performed an analysis of the relationship between the average inhibitory effects of 30-residue fragments derived from each protein and the corresponding protein concentration (protein copy number per cell; refs 19, 30). We uncovered a robust negative correlation between parental protein concentration and susceptibility to fragment-based inhibition, observed as a linear relationship ( $R^2 = 0.95$ ) between the log of the protein copy number per cell and the average enrichment  $\langle E \rangle$  of fragments derived from that protein (Fig. 3), with a higher  $\langle E \rangle$  corresponding to lower average inhibitory activity. The slope indicates that a  $\sim 10$ -fold increase in protein concentration is associated with  $\sim 50\%$  less fragment depletion in our assay (an average  $E$  increase of  $\sim 1$ ). Thus, the poor susceptibility of GroEL and GroES to fragment-based inhibition might be explained by their high cellular concentrations, and the



**Fig. 3.** Cellular protein concentration negatively correlates with average susceptibility to fragment-based inhibition. Plot of mean inhibitory effects of 30-residue fragments tiling each of the cytoplasmic proteins included in this study as a function of the log of the cellular concentration of the full-length protein, from Li et al. (19). The 50S ribosomal protein L7/L12 outlier is indicated. Error bars: SEM. A linear fit is shown to the data for the non-ribosomal proteins.

broad sensitivity of DHFR to fragment-based inhibition by its low concentration (Fig. 3 and Table 1). The sole exception to the correlation was the ribosomal protein RplL, fragments of which were much more inhibitory than expected from its cellular concentration (Fig. 3). A possible reason for this exception is that even weak inhibition of ribosome assembly may be amplified in the assay due to the large effect of ribosome levels on growth rates (31). Additionally, RplL fragments may target dynamically assembling pre-ribosomes, which likely exist at much lower concentrations than mature ribosomes.

In sum, these results suggest that protein fragment-based inhibition is driven primarily by fragments acting as competitive inhibitors of native interactions formed by their parental proteins, despite the complexities of the *in vivo* scenario. Additionally, these findings are inconsistent with the idea that the inhibitory effects of fragments are largely due to more generic effects, such as toxicity due to peptide aggregation.

We asked whether the relationship between protein concentration and fragment inhibition could also be observed for fragments tiling cancer-related proteins in a breast cancer cell line (15, 32). However, this correlation was absent from the cancer cell data (SI Appendix, Fig. S3), potentially reflecting important differences from the *E. coli* experiments. First, the *E. coli* proteins we investigated are essential, whereas the essentiality of the cancer-related proteins studied is not well-defined. Second, human cells may employ additional regulatory mechanisms (e.g., certain proteolytic activities) that reduce the impact of inhibitory protein fragments in a sequence-specific manner.

#### Relationship between Fragment Length and Inhibitory Activity.

We sought to understand how the inhibitory effects of protein fragments might depend on fragment length. For example, shorter fragments might allow finer mapping of structural elements, but fragments that are too short might have little effect due to limited binding energy. Conversely, longer fragments might exhibit stronger inhibitory activity by forming larger interaction surfaces or folding more stably in isolation. To investigate these questions, we examined the results of scans performed with fragments of 20 or 50 residues, compared with those performed with 30-residue fragments (Fig. 4 A–C and SI Appendix, Fig. S4).

The results demonstrate several general features of fragment-based inhibition. First, tiling fragments of different lengths generally mapped out the same structural features of the parental proteins (Fig. 4A and SI Appendix, Fig. S4). That the functional maps produced with different fragment lengths are so correlated—along with the finding that average inhibitory effects of fragments of 50 residues also negatively correlated with the concentrations of their parental proteins (SI Appendix, Fig. S5)—further supports the idea that protein fragments typically act in a target- and sequence-specific manner.

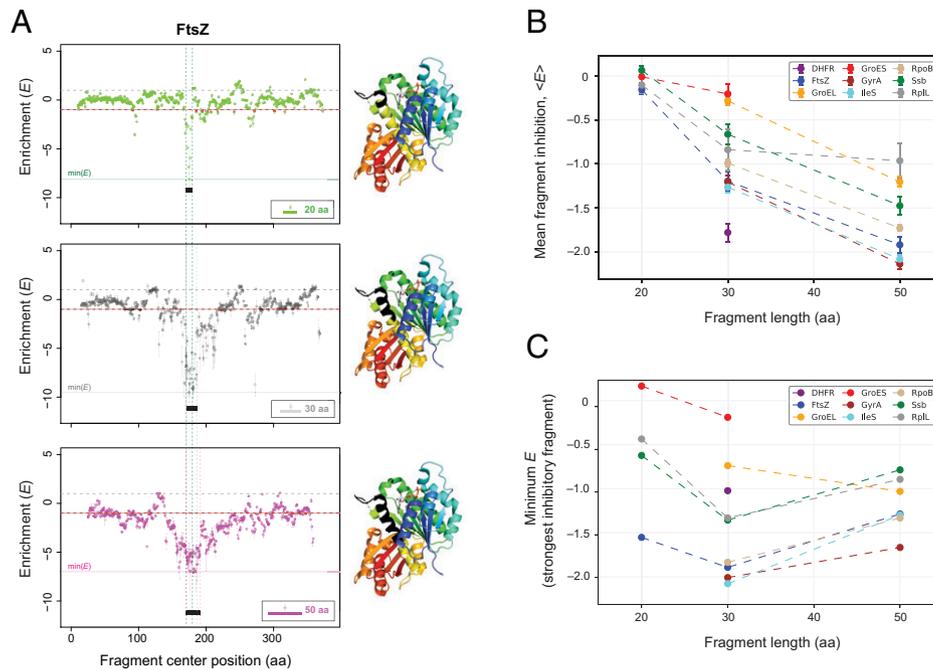
Second, the average inhibitory activity was larger for longer fragments (Fig. 4B and SI Appendix, Fig. S4). Increasing the fragment length from 20 to 30 residues increased the magnitude of inhibitory peaks while maintaining similar levels of background inhibition throughout the sequence (i.e., the baseline was almost unaffected), representing an increase in both inhibitory activity and sequence specificity. However, further increasing the length to 50 residues produced diminishing returns, as these fragments exhibited a marked increase in generic inhibitory activity across the sequence (a baseline shift to lower  $E$ ) and, moreover, a decrease in the strength of many inhibitory peaks relative to the 30-residue fragments (e.g., in FtsZ, RplL, Ssb, GyrA, IleS; Fig. 4A and SI Appendix, Fig. S4 A–D and H), which in almost all cases led to a decreased maximum inhibitory activity (higher minimum  $E$ ) (Fig. 4C). We interpret this result as reflecting a competition between distinct target-specific and nonspecific inhibitory mechanisms (unrelated to perturbation of parental protein interactions) of the longer fragments, leading to reduced inhibition by 50-residue fragments from strong interaction sites. These results suggest that there is a fragment length between 20 and 50 residues that optimizes the balance of inhibitory activity and specificity; a length of 30 residues appears to be near this optimum (Fig. 4C).

Third, shorter fragments allowed for finer-resolution functional mapping, localizing inhibitory peaks to a narrower range of fragment center positions. In the case of FtsZ (Fig. 4A), the 20-residue fragment scan yielded a peak width of just 10 residues for the major inhibitory region (compared to 16 with 30-residue fragments, and 22 with 50-residue fragments), containing only one turn of the H7 alpha-helix but still encompassing the loop of monomer interaction site I (Fig. 4A, Middle). However, the weaker inhibitory activity on average also meant that some peaks observed with 30-residue fragments were less evident in the 20-residue fragment scan.

In sum, there are trade-offs between fragment length and important functional parameters that include inhibitory strength, specificity, and resolution of functional mapping. Shorter protein fragments allow higher-resolution mapping but are associated with weaker inhibitory activity. Longer fragments can provide greater sequence-specific activity, but only to a point, with nonspecific background inhibition and reduced inhibitory peak magnitudes becoming evident by a fragment size of 50 residues in *E. coli*.

#### The Bulk of the Variability in the Inhibitory Effects of Individual Fragments Is Case Specific.

The effects of parental protein concentration and fragment length explain only a portion of the individual fragment-to-fragment variability (SI Appendix, Fig. S6). Indeed, the strength of inhibition attainable at any specific site in a protein sequence must depend on unique features of each protein's binding interactions. Hence, we sought to identify common features of individual fragments that might explain which sites in protein sequences are susceptible to fragment-based inhibition. Using analysis of variance (ANOVA), we approached this question across all protein targets investigated and several



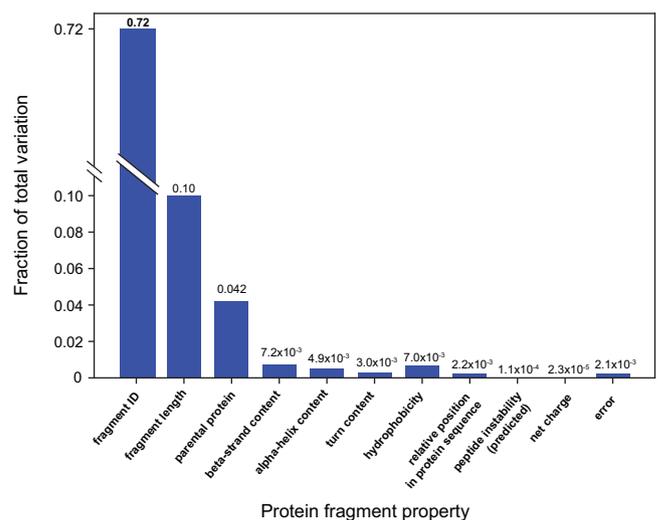
**Fig. 4.** Inhibitory activity depends on fragment center length. (A) Inhibitory fragment scans of FtsZ with tiling fragments, at a tile step size of one amino acid (aa), comparing different fragment lengths. Data are plotted as in Fig. 2, with individual fragment measurements and their SEM across multiple experiments shown. A comparison of 20- (Top), 30- (Middle), and 50-residue (Bottom) fragment scans is displayed. For each plot, the width of inhibitory peak 1 is indicated. Structures of FtsZ monomers to the right are overlaid with the width of the corresponding inhibitory peak in black. (B) Average inhibitory activity ( $E$ ) of fragments tiling each cytosolic protein plotted as a function of fragment length. (C) Strongest inhibitory fragment among fragments tiling each protein, as a function of fragment length. A higher-resolution version of this figure is available on figshare (18).

physicochemical and structural properties. In particular, we determined the percentage of total variability in the inhibitory effects of protein fragments that is attributable to hydrophobicity, charge, secondary structure features, fragment length, parental protein, relative position in the protein sequence, predicted peptide stability, and fragment-specific effects not explained by these other properties (see *Materials and Methods* and *SI Appendix*, Table S1).

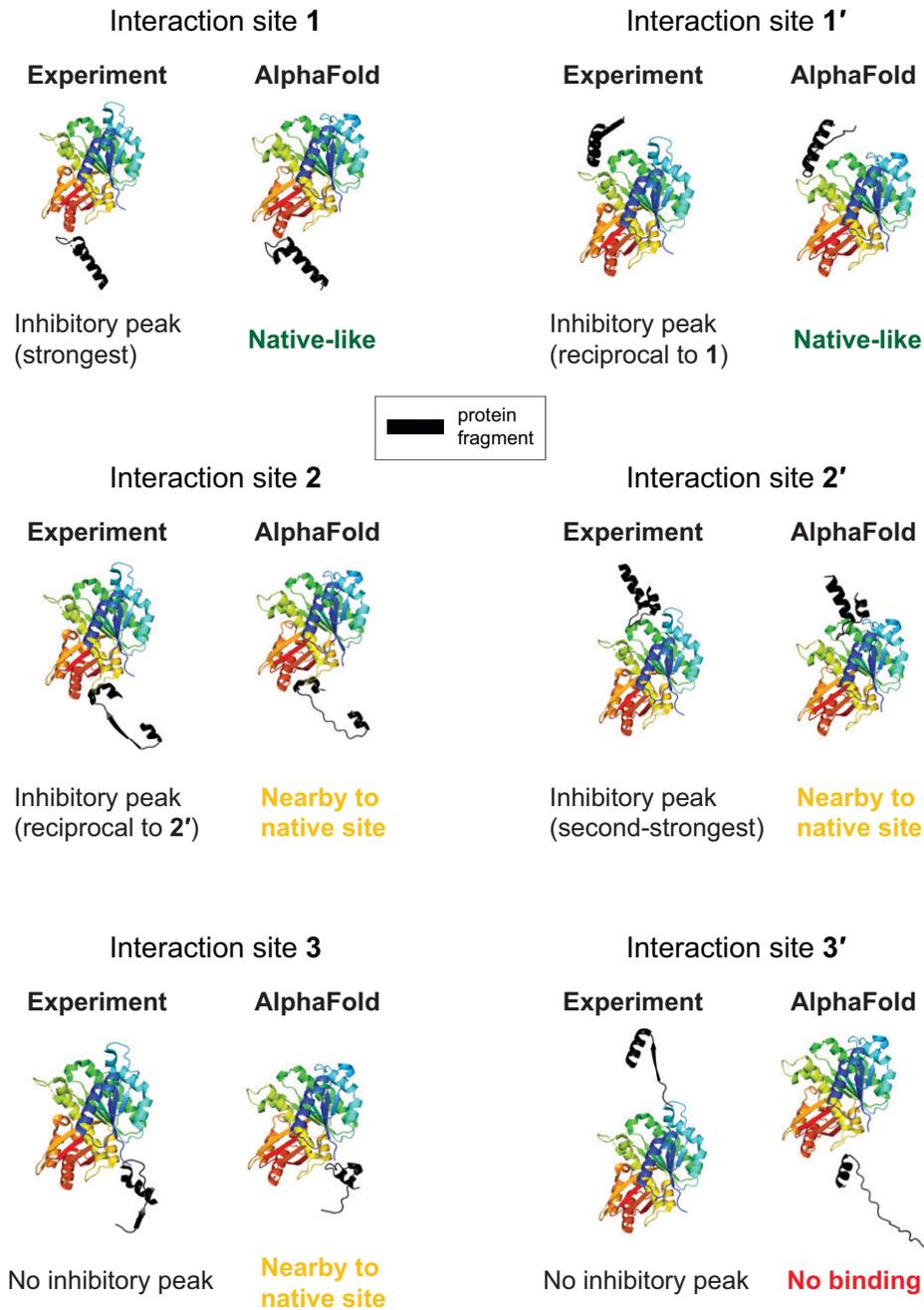
We found that a striking  $\sim 72\%$  of the total variability was attributable to fragment-to-fragment variation, outside of that attributable to any of the properties considered (Fig. 5). The properties that contributed the most to variability were fragment length, accounting for  $\sim 10\%$  of the total variation, and parental protein, accounting for  $\sim 4.2\%$ , matching the overall trends in fragment function. All other properties considered each explained less than 1% of the variation. Added together, secondary structure features of the fragment-covered portion of the protein structure explained  $\sim 1.5\%$  of the variation:  $\sim 0.3$  to  $0.7\%$  each for alpha-helical, beta-sheet, or turn content. Hydrophobicity had a comparable effect to secondary structure ( $\sim 0.7\%$  of the variation), and charge had an even smaller effect, well below that attributable to error, providing additional evidence against a generic inhibitory mechanism of aggregation due to charge and/or hydrophobicity, similar to the results of Ford et al. (15). Relative position of the fragment in the protein sequence and predicted fragment stability (33) were both negligible for explaining variability, each being associated with a fraction of the variation similar to or below the contribution of error, suggesting that stability differences are not a significant factor for inhibitory activity. The results of this analysis speak to the diversity of inhibitory function seen among protein fragments, mirroring the vast diversity of possible folds and interactions of proteins that provide the likely basis for most inhibitory activity.

Inhibition by random peptides not derived from protein sequences might have simpler unifying features, reflecting nonspecific

inhibitory mechanisms. To search for such features, we performed a similar analysis of out-of-frame fragments of *E. coli* DHFR (*SI Appendix*, Fig. S7). Only fragment length explained a substantial fraction of the variability ( $\sim 16\%$ ), with the majority ( $\sim 77\%$ ) explained by none of the properties considered, similar to the in-frame fragment results (Fig. 5). These results suggest that even nonspecific inhibition by random peptides depends strongly on specific sequence and structural features. There may also be a weak charge-driven effect ( $\sim 1.6\%$  of variability) specific to random peptides.



**Fig. 5.** Generic properties are not the main drivers of protein-fragment-based inhibition. Fraction of total variation attributable to each property is plotted, based on a nested ANOVA. Properties include fragment length, parental protein, and secondary structure features of the fragment-covered sequence region of the full-length protein; "fragment ID" designates fragment-specific effects not attributable to variation in any of the properties considered.



**Fig. 6.** Computational modeling of FtsZ protein–peptide complexes suggests that AlphaFold can predict strong inhibitory peaks. For each of the major interaction sites of FtsZ filament formation, 1, 1', 2, 2', 3, and 3', a side-by-side comparison is shown between 1) the 30-residue protein fragment representing the site in question, bound to the adjacent monomer, from the experimentally determined structure of the FtsZ filament (*Left*; from PDB ID 6unx) and 2) the AlphaFold-predicted structure of the corresponding protein–peptide complex (*Right*). FtsZ monomers are colored from blue (N terminus) to red (C terminus), as in Fig. 2, and protein fragments are colored black. Experimental results from the protein fragment scanning experiments are noted below the crystallographic structures, and the predicted binding mode (or lack thereof) is noted below the predicted structures. A higher-resolution version of this figure is available on figshare (18).

**AlphaFold Predictions of Protein–Peptide Interactions Provide a Complementary Approach.** Our ability to predict protein fragments that inhibit native interactions might be improved by employing computational models of protein–peptide interactions. In particular, predicted modes of peptide binding and the extent to which they are native-like may correlate with the experimentally measured inhibitory strength of peptides from different binding interfaces. If so, peptide and interface properties used by the model to make these predictions might be extracted to determine design principles for inhibitory peptides. Given that, in the case of protein fragments, the major inhibitory mechanism appears to be the mimicry of native interactions, we reasoned that

machine learning-based structure prediction algorithms trained on native folds and sequences would be a good choice of model.

We performed AlphaFold modeling of protein–peptide interactions (ref. 34; *Materials and Methods*) using the FtsZ results as a case study. This modeling predicted that representative protein fragments from the strongest inhibitory peak (site 1) and its reciprocal peak (site 1') both form native-like interactions with FtsZ (Fig. 6, *Top Right*). Fragments from the second-strongest inhibitory peak (site 2') and reciprocal site 2 were predicted to form interactions with FtsZ that were nonnative but nearby to their native interaction sites in the crystal structure, and still along the monomer interaction interface (Fig. 6, *Middle Right*). Finally, a

fragment representing site 3', which yielded no inhibitory peak, was predicted to make only tenuous (and nonnative) interactions with FtsZ (Fig. 6, *Bottom Right*), and a fragment from the similarly noninhibitory site 3 was predicted to form substantial but nonnative interactions, again along the interaction interface (Fig. 6, *Bottom Right*). These results thus show an approximate association between the potency of inhibitory protein fragment peaks and AlphaFold predictions of native-like binding for this case study, suggesting that these predictions indeed contain latent information on features underlying peptide-based inhibition that might be extracted to determine unifying properties of inhibitory fragments and enable fragment design.

An important caveat is that the predicted structures of these complexes could be influenced by memorization of the FtsZ structure by AlphaFold, due to the presence of multiple FtsZ orthologs (though not *E. coli* FtsZ itself) in the training set (34). However, the training set does not include protein complexes, only single polypeptide chains (34–36). Thus, while there may be an influence of memorization, the interactions between these peptides and FtsZ are not drawn from the training set and likely reflect learned predictions. Consistent with not strictly using memorization, AlphaFold was unable to predict the 3' interaction site (Fig. 6, *Bottom Right*), whereas it has predicted novel protein–peptide interactions through learning (35).

## Discussion

In this work, we uncovered numerous inhibitory fragment peaks localized to protein interaction sites, indicating that these sites are likely susceptible to perturbation by competitive binding. As the parental proteins are essential, these peak regions constitute promising target sites for the development of antimicrobials, potentially based on the protein fragments themselves (15, 16). Furthermore, these proteins are extremely well conserved in sequence and structure across many bacterial strains and species, including human pathogens such as *Shigella*, *Klebsiella*, and *Salmonella* (37), thus making inhibition-prone sites derived from *E. coli* fragment scans likely to be directly translatable to other strains.

In the case of FtsZ, our results suggest that the 1, 1', 2, and 2' interaction sites (Fig. 2A) employed in filament formation are good targets, but that the 3 and 3' interaction sites are not. Although no approved antibiotic targets FtsZ, multiple inhibitors thought to target the cleft between the N- and C-terminal domains, the guanosine triphosphate (GTP)-binding pocket, and the T7 loop (site 2') have been reported (38), and the small protein MciZ binds sites 1' and 2' simultaneously (22). The fragment scans of DHFR, GyrA, and Ssb similarly yielded both known antimicrobial binding sites and potential novel ones (*SI Appendix, SI Text*).

The protein fragments characterized here bear a striking resemblance to biologically encoded miniproteins (39, 40). Miniproteins consist of  $\leq 50$  residues and are frequently composed of simple structural motifs, such as individual alpha-helices (6), similar to many of the inhibitory protein fragments we report. These natively produced peptides commonly regulate proteins by binding to specific sites, analogous to the mapping of inhibitory fragments to interaction surfaces. The functional similarities are in some cases direct. Fragments of *E. coli* FtsZ mapping to the 1–1' and 2–2' interaction site pairs tended to be inhibitory (Fig. 2A), presumably by competitively inhibiting the corresponding monomer–monomer interactions. In *Bacillus subtilis*, the MciZ miniprotein performs this function by binding in a manner comparable to the inferred binding modes of interaction site 1 and 2 fragments, occluding the

1' and 2' (T7 loop) interaction sites and thereby inhibiting filament assembly and cell division (22). The lack of inhibitory peaks mapping to the 3 and 3' interaction sites suggests that these are poorer sites for competitive inhibition, consistent with the MciZ binding target.

Hydrophobic alpha-helical miniproteins, often just long enough to traverse the membrane, are common in bacteria (4, 6, 41), and structures of such miniproteins bound to their membrane protein targets have been determined (e.g., ref. 42). These miniproteins regulate their membrane protein targets (e.g., AcrZ modulates antibiotic export by the AcrAB-TolC efflux pump (6, 43)) and can readily enter the inner membrane through a mechanism mediated by posttranslational binding to SRP (27).

This work highlights important considerations for future applications of high-throughput inhibitory fragment assays. Parental proteins must be selected carefully based on cellular concentration; abundant proteins would require substantially higher levels of peptide to inhibit their interactions. Protein fragment length is another crucial variable that must be controlled; for example, inhibition by fragments mapping to different sequence regions can be directly compared only if they are of the same size. In applications for which the primary goal is peptide inhibitor development, longer fragments (here  $\sim 30$ -residue) are preferred as these provide more robust inhibitory activities. In studies geared toward functional mapping of interaction sites, scanning with multiple peptide fragment lengths (e.g., 14 to 30 residues) might be best, as shorter fragments yielded higher-resolution maps while longer fragments more readily identified weaker peaks. Fragments that are too long (here  $\sim 50$  residues) should generally be avoided due to increased background inhibition and loss of site-specific inhibition.

Given its effectiveness in cells of *Saccharomyces cerevisiae*, *Homo sapiens*, and now *E. coli*, inhibitory fragment scanning should be generalizable to essentially any genetically tractable species, enabling comparative analyses of fragment inhibition across orthologous proteins. Total proteome-wide fragment mapping is also within reach, though currently limited to a lower resolution than the measurements in this work. This approach also holds promise for comprehensively measuring the relative importance of interaction sites under multiple environmental settings and drug treatments, enabling the elucidation of condition-specific roles of interactions across the proteome.

## Materials and Methods

**Tiling Fragment Library Construction.** Tiling oligonucleotides were designed to cover each parental protein in each fragment size of interest with a 3-bp (one codon) step size. DNA oligonucleotides encoding these fragments were array-synthesized by Twist. Oligo sequences were centered around an exact match to the desired region of endogenous parental protein sequence, with the following features added in the flanking sequences of each fragment: 1) flanking sequences for Gibson assembly into the desired plasmid vector; 2) gene-specific pairs of 3nt indices at the 5' and 3' ends of the fragment, which uniquely identified the parental protein and allowed selective PCR amplification from the library (*SI Appendix, Table S2*); and 3) a stop codon at the 3' end of the fragment (downstream of the gene-specific 3-bp index). Array-synthesized ssDNA libraries were amplified into dsDNA by qPCR following manufacturer's recommendations. The resulting dsDNA libraries were cloned into the multiple cloning site of the pET-9a expression vector (Novagen) by Gibson assembly (NEBuilder HiFi DNA Assembly kit, NEB). Fragment-encoding sequences were expressed under control of a T7 promoter in a shared exogenous sequence context provided by the vector. The pET-9a vector adds an N-terminal T7 tag peptide and short linker to expressed polypeptides, so the resulting sequence of each protein fragment had the following structure: MASMTGGQQMGRGS-X<sub>1</sub>(fragment)-X<sub>2</sub>\*, where the T7 tag sequence is underlined and X<sub>1</sub> and X<sub>2</sub> are serine, alanine, or leucine, encoded by the gene-specific 3-bp indices.

Two tiling fragment-encoding libraries were constructed in this manner: one containing 20- and 30-residue scans of the cytosolic proteins investigated (Library 1) and another containing 50-residue scans of cytosolic proteins and 14-residue scans of the membrane protein LptG (Library 2). Each library contained ~6,000 fragments. Thus, Library 2 contained a larger proportion of longer protein fragments, which are more inhibitory on average. This compositional difference between the libraries might produce systematic differences in growth rates that affect the relative inhibitory peak depths with 50-residue fragments compared to 20- and 30-residue fragments (Fig. 4C). However, a set of 25-residue fragments derived from RplL and enhanced green fluorescent protein that were common to both libraries generally gave highly similar results (SI Appendix, Fig. S8), yielding an average enrichment score difference of 0.08 (frequency change fold-difference of 1.06) with an SD of 0.57 (frequency change fold-difference range of 0.71 to 1.58).

Following Gibson assembly, tiling fragment libraries were transformed into ElectroMax *E. coli* cells (Thermo Fisher Scientific). Plasmids were isolated from the transformed ElectroMax cells (Qiaprep Spin Miniprep Kit, Qiagen), and assembled library composition was confirmed by high-throughput sequencing. These purified plasmid libraries served as the starting point for subsequent selection experiments.

**Massively Parallel Measurements of Dominant-Negative Inhibition by Protein Fragments.** The inhibitory activity of protein fragments was measured in a high-throughput manner via growth selection experiments, similarly to previous studies (14, 15). Tiling fragment-encoding plasmid libraries were transformed into highly electrocompetent *E. coli* BL21(DE3) cells (Sigma-Aldrich), yielding transformants at >200-fold coverage of the library size. After 1 h in recovery medium, each 1 mL of transformed cells was transferred to 50 mL Luria-Bertani (LB) medium + 50 µg/mL kanamycin + 10 µM isopropyl β-d-1-thiogalactopyranoside (IPTG) to begin the growth selection. These cultures were grown at 37 °C with shaking at 220 rpm. Multiple replicate experiments, entailing independent transformations and growth selections, were performed for each library ( $n = 4$  for Library 2 experiments with 20- and 30-residue fragments, and  $n = 2$  for Library 1 experiments with 14- and 50-residue fragments). Cells were grown to a final optical density at 600 nm of 0.6 to 0.9 and harvested by centrifugation; this constituted the end point of the selection. Plasmids were isolated from each sample by miniprep (Qiagen). The selection end point samples and the plasmid library input were prepared for sequencing as follows. PCR amplification ( $\leq 12$  cycles) was used to extract fragment-coding sequences from the plasmid library and add sequencing adapters and sample-identifying indices. High-throughput sequencing was performed using Illumina's NextSeq 550 platform. Paired-end sequencing was employed to uniquely identify fragments by their 5' and 3' ends. Sequencing was performed with ~6 million reads per sample (1,000-fold coverage).

**Fragment Assay Data Analysis.** The identity (parental protein, sequence location, and orientation) of each fragment was determined by aligning each read pair to the set of gene sequences included in the library using Bowtie2. Fragment counts in each sample were determined, and fragments with insufficient read depth in the input sample were filtered out of the dataset ( $< 50$  reads for Library 1 and  $< 100$  reads for Library 2). Fragment sequences that completely dropped out in the growth selection were assigned pseudocounts of 0.5 in the corresponding output samples. Fragment frequencies in each sample were then determined from the counts of each fragment in each sample, divided by the total counts of all fragment sequences in the sample. The enrichment ( $E$ ) of each fragment was determined from its growth selection input and output frequencies as  $E = \log_2(f_{\text{post-selection}}/f_{\text{pre-selection}})$ , where  $f$  denotes frequency.  $E < 0$  therefore indicates depletion. Mean and SEM values of  $E$  were calculated for each fragment based on multiple replicate experiments. Fragment maps were generated by plotting  $E$  for each fragment tiling a protein as a function of the fragment's position along the parental protein sequence. Average and maximum inhibitory effects of fragments derived from a protein were determined by calculating the mean and minimum of  $E$ , respectively, for the appropriate subsets of fragments. Parental protein structures overlaid with representative fragments or widths of inhibitory peaks were generated from protein structures found in the RSCB Protein Data Bank (PDB) using Pymol.

Cellular concentrations of parental proteins were retrieved from the data of Li and colleagues (19), as maintained by the EcoCyc database (30); the protein

concentrations determined under rich media conditions (Neidhardt EZ rich defined medium) were used. Physicochemical properties of protein fragments were determined using the Peptides package in R (44). In particular, Peptides was employed to calculate the peptide properties hydrophobicity [Kyte-Doolittle scale (45)]; charge [Lehninger  $pK_a$  scale (46), at pH 7]; and instability [Guruprasad instability index (33)]. For the ANOVA analysis, each fragment was classified as "hydrophobic" if the Kyte-Doolittle hydrophobicity charge was  $> 0$  and "hydrophilic" otherwise; as "neutral" if the Lehninger  $pK_a$ -based charge was in the range  $(-1, 1)$ , "negative" if the charge was less than  $-1$ , and "positive" if the charge was  $> 1$ ; and as "unstable" if the Guruprasad instability index was  $\geq 40$  [following the standard definition (33)] and "stable" otherwise. Alpha-helical, beta-strand, and turn content of protein fragments in the full-length protein context was determined as follows. Secondary structure annotations were retrieved from each parental protein's Uniprot entry (47), and each protein fragment was annotated as containing each type of secondary structure content if the sequence region covered by the fragment overlapped with the corresponding secondary structure feature; otherwise, the fragment was annotated as lacking that type of secondary structure.

$N$ -way nested ANOVA was performed using the built-in "anovan" function in Matlab. The following parameters were included in the analysis: fragment length; parental protein; hydrophobicity, charge, and predicted instability, classified as described above; fragment alpha-helical, beta-strand, and turn content (a binary True/False classification for each); and fragment relative position in the parental protein sequence (classified as "N-terminal" for fragments sourced from the first third of the sequence, "central" for fragments sourced from the central third, and "C-terminal" for fragments sourced from the C-terminal third); finally, fragment identity (a unique tag assigned to each fragment based on parental protein, sequence start and stop sites, and orientation) was nested under all other parameters. The measurable variable was fragment enrichment.  $E$ . ANOVA was performed on the combined data from both tiling libraries, excluding only fragments of IleS, due to a lack of structural information that prevented assignment of fragment secondary structure content.

**DHFR Random Fragment Library Experiments.** The random fragment library of DHFR was generated as previously described (14). Briefly, the *folA* gene was PCR-amplified out of the *E. coli* genome and fragmented using the Nextera transposase (Illumina). The resulting gene fragments were amplified by PCR to add sequence adapters for Gibson assembly (as well as stop codons in all three frames) and cloned into the pET-9a vector backbone at the multiple cloning site. Growth selection experiments were performed as described for the tiling fragment libraries, except 100 µM IPTG was used. High-throughput sequencing of input and output library samples and subsequent analyses to determine fragment localization, frequencies, and enrichment were performed as for the tiling libraries. Average per-site enrichment was calculated by determining the average  $E$  of all fragments whose sequence covered a given position in the parental protein sequence.

**FtsZ Filament Polymerization Experiments.** FtsZ filament polymerization was monitored by 90° scattering of 350 nm light, as previously described (22, 48), a measurement shown to correlate with MciZ-induced filament length reduction as revealed by electron microscopy (22) and to faithfully report the critical monomer concentration for polymerization (48). Scattering measurements were performed on a SpectraMax M5 instrument in fluorimetry mode with excitation and emission wavelengths set to 350 nm, using a quartz cuvette with a path length of 1.5 mm. Experiments were conducted using purified *E. coli* FtsZ protein obtained from Cytoskeleton and chemically synthesized FtsZ protein fragment peptides (T7-tagged identically to the constructs expressed in cells, as described in Tiling Fragment Library Construction;  $> 97\%$  purity) obtained from the Swanson Biotechnology Center Biopolymers & Proteomics Core Facility at the Massachusetts Institute of Technology. FtsZ (final concentration 5.06 µM) was incubated for 5 to 10 min at room temperature with either a FtsZ site 1 protein fragment (residues 166 to 195) or a neutral control fragment, FtsZ residues 35 to 54 (final peptide concentration 18.4 µM) in MMK buffer (final composition 50 mM MES, pH 6.5, 10 mM MgCl<sub>2</sub>, and 50 mM KCl), then mixed with GTP or adenosine 5'-triphosphate (ATP) (final concentration 2 mM). Light scattering was then monitored over time. Steady-state values of light scattering achieved within 5 to 20 min were recorded. Scattering in mock reactions with ATP added was

subtracted from scattering in reactions with GTP added to determine  $\Delta$  Light scattering (GTP – ATP) =  $\Delta$  Light scattering (filament polymerization).

#### AlphaFold Computational Predictions of Peptide–Protein Complexes.

Peptide–protein complexes were modeled using AlphaFold2 (34), employing the AlphaFold2\_complexes notebook hosted on the Google Colab servers (49, 50). Input query sequences for each model were 1) the full-length sequence of *E. coli* BL21(DE3) FtsZ and 2) the sequence of a 30-residue fragment of FtsZ representing an inhibitory peak or other sequence region of interest. Default settings were used for other parameters: num\_models = 5, msa\_mode = MMseqS4, pair\_msa = False. The structures reported in the paper correspond to output model 1 from each prediction.

**Data Availability.** Source data have been deposited in figshare (10.6084/m9.figshare.19617648) ([https://figshare.com/articles/dataset/Source\\_Data\\_for\\_Savinov\\_et\\_al\\_2022\\_protein\\_fragment\\_mapping/19617648](https://figshare.com/articles/dataset/Source_Data_for_Savinov_et_al_2022_protein_fragment_mapping/19617648)) (18). All other data are included in the manuscript and/or *SI Appendix*.

1. J. A. Wells, C. L. McClendon, Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **450**, 1001–1009 (2007).
2. A. J. Quartararo *et al.*, Ultra-large chemical libraries for the discovery of high-affinity peptide binders. *Nat. Commun.* **11**, 3183 (2020).
3. G. Upert, A. Luther, D. Obrecht, P. Emert, Emerging peptide antibiotics with therapeutic potential. *Med. Drug Discov.* **9**, 100078 (2021).
4. E. C. Hobbs, F. Fontaine, X. Yin, G. Storz, An expanding universe of small proteins. *Curr. Opin. Microbiol.* **14**, 167–173 (2011).
5. J. Weaver, F. Mohammad, A. R. Buskirk, G. Storz, Identifying small proteins by ribosome profiling with stalled initiation complexes. *MBio* **10**, e02819-18 (2019).
6. M. R. Hemm, J. Weaver, G. Storz, *Escherichia coli* small proteome. *EcoSal Plus* **9**, 1–16 (2020).
7. J. Chen *et al.*, Pervasive functional translation of noncanonical human open reading frames. *Science* **367**, 1140–1146 (2020).
8. D. A. Silva *et al.*, De novo design of potent and selective mimics of IL-2 and IL-15. *Nature* **565**, 186–191 (2019).
9. A. Chevalier *et al.*, Massively parallel de novo protein design for targeted therapeutics. *Nature* **550**, 74–79 (2017).
10. M. D. T. Torres, S. Sothiselvam, T. K. Lu, C. de la Fuente-Nunez, Peptide design principles for antimicrobial applications. *J. Mol. Biol.* **431**, 3547–3567 (2019).
11. I. Herskowitz, Functional inactivation of genes by dominant negative mutations. *Nature* **329**, 219–222 (1987).
12. J. G. Hall, C. Frieden, Protein fragments as probes in the study of protein folding mechanisms: Differential effects of dihydrofolate reductase fragments on the refolding of the intact protein. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 3060–3064 (1989).
13. J. E. A. Michaels, P. Schimmel, K. Shiba, W. T. Miller, Dominant negative inhibition by fragments of a monomeric enzyme. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 14452–14455 (1996).
14. M. W. Dorrity, C. Queitsch, S. Fields, High-throughput identification of dominant negative polypeptides in yeast. *Nat. Methods* **16**, 413–416 (2019).
15. K. M. Ford *et al.*, Peptide-tiling screens of cancer drivers reveal oncogenic protein domains and associated peptide inhibitors. *Cell Syst.* **12**, 716–732.e7 (2021).
16. S. Nim *et al.*, Pooled screening for antiproliferative inhibitors of protein-protein interactions. *Nat. Chem. Biol.* **12**, 275–281 (2016).
17. A. Savinov, F. P. Roth, Seeds of their own destruction: Dominant-negative peptide screening yields functional insight and therapeutic leads. *Cell Syst.* **12**, 691–693 (2021).
18. A. Savinov, A. Fernandez, S. Fields, Source Data for Savinov *et al.* 2022: protein fragment mapping. figshare. [https://figshare.com/articles/dataset/Source\\_Data\\_for\\_Savinov\\_et\\_al\\_2022\\_protein\\_fragment\\_mapping/19617648](https://figshare.com/articles/dataset/Source_Data_for_Savinov_et_al_2022_protein_fragment_mapping/19617648). Deposited 20 April 2022.
19. G. W. Li, D. Burkhardt, C. Gross, J. S. Weissman, Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* **157**, 624–635 (2014).
20. S. Du, J. Lutkenhaus, At the heart of bacterial cytokinesis: The Z ring. *Trends Microbiol.* **27**, 781–791 (2019).
21. M. A. Schumacher, T. Ohashi, L. Corbin, H. P. Erickson, High-resolution crystal structures of *Escherichia coli* FtsZ bound to GDP and GTP. *Acta Crystallogr. F Struct. Biol. Commun.* **76**, 94–102 (2020).
22. A. W. Bisson-Filho *et al.*, FtsZ filament capping by MciZ, a developmental regulator of bacterial division. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E2130–E2138 (2015).
23. P. J. Buske, P. A. Levin, A flexible C-terminal linker is required for proper FtsZ assembly in vitro and cytokinetic ring formation in vivo. *Mol. Microbiol.* **89**, 249–263 (2013).
24. M. G. Rudolph, D. Klostermeier, Mapping the spectrum of conformational states of the DNA- and C-gates in *Bacillus subtilis* gyrase. *J. Mol. Biol.* **425**, 2632–2640 (2013).
25. C. Wang *et al.*, Structural basis of transcription-translation coupling. *Science* **369**, 1359–1365 (2020).
26. Y. Li, B. J. Orlando, M. Liao, Structural basis of lipopolysaccharide extraction by the LptB<sub>2</sub>FGC complex. *Nature* **567**, 486–490 (2019).
27. R. Steinberg *et al.*, Posttranslational insertion of small membrane proteins by the bacterial signal recognition particle. *PLoS Biol.* **18**, e3000874 (2020).
28. J. W. Tobias, T. E. Shrader, G. Rocap, A. Varshavsky, The N-end rule in bacteria. *Science* **254**, 1374–1377 (1991).
29. A. Varshavsky, N-degron and C-degron pathways of protein degradation. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 358–366 (2019).
30. I. M. Keseler *et al.*, EcoCyc: a comprehensive database of *Escherichia coli* biology. *Nucleic Acids Res.* **39**, D583–D590 (2011).
31. N. M. Belliveau *et al.*, Fundamental limits on the rate of bacterial growth and their influence on proteomic composition. *Cell Syst.* **12**, 924–944.e2 (2021).
32. R. T. Lawrence *et al.*, The proteomic landscape of triple-negative breast cancer. *Cell Rep.* **11**, 630–644 (2015).
33. K. Guruprasad, B. V. B. Reddy, M. W. Pandit, Correlation between stability of a protein and its dipeptide composition: A novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Eng.* **4**, 155–161 (1990).
34. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
35. T. Tsaban *et al.*, Harnessing protein folding neural networks for peptide-protein docking. *Nat. Commun.* **13**, 176 (2022).
36. R. Evans *et al.*, Protein complex prediction with AlphaFold-Multimer. *bioRxiv* [Preprint] (2021). <https://doi.org/10.1101/2021.10.04.463034>. Accessed 1 April 2022.
37. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
38. K. D. Kusuma, M. Payne, A. T. Ung, A. L. Bottomley, E. J. Harry, FtsZ as an antibacterial target: Status and guidelines for progressing this avenue. *ACS Infect. Dis.* **5**, 1279–1294 (2019).
39. G. Storz, Y. I. Wolf, K. S. Ramamurthi, Small proteins can no longer be ignored. *Annu. Rev. Biochem.* **83**, 753–777 (2014).
40. D. Schlesinger, S. J. Elsässer, Revisiting sORFs: Overcoming challenges to identify and characterize functional microproteins. *FEBS J.* **289**, 53–74 (2022).
41. F. Fontaine, R. T. Fuchs, G. Storz, Membrane localization of small proteins in *Escherichia coli*. *J. Biol. Chem.* **286**, 32464–32474 (2011).
42. C. S. Huang, B. P. Pedersen, D. L. Stokes, Crystal structure of the potassium-importing KdpFABC membrane complex. *Nature* **546**, 681–685 (2017).
43. D. Du *et al.*, Structure of the AcrAB-TolC multidrug efflux pump. *Nature* **509**, 512–515 (2014).
44. D. Osorio, P. Rondón-Villarreal, R. Torres, Peptides: A package for data mining of antimicrobial peptides. *R J.* **7**, 4–14 (2015).
45. J. Kyte, R. F. Doolittle, A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).
46. D. L. Nelson, M. M. Cox, *Lehninger: Principles of Biochemistry* (W. H. Freeman, ed. 4, 2004).
47. A. Bateman *et al.*; UniProt Consortium, UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49** (D1), D480–D489 (2021).
48. A. Mukherjee, J. Lutkenhaus, Analysis of FtsZ assembly by light scattering and determination of the role of divalent metal cations. *J. Bacteriol.* **181**, 823–832 (1999).
49. S. Ovchinnikov, AlphaFold2\_complexes notebook (2021). [https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2\\_complexes.ipynb](https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2_complexes.ipynb). Accessed 1 January 2022.
50. M. Mirdita, S. Ovchinnikov, M. Steinegger, ColabFold – Making protein folding accessible to all. *bioRxiv* [Preprint] (2021). <https://doi.org/10.1101/2021.08.15.456425>. Accessed 1 April 2022.

Author affiliations: <sup>a</sup>Department of Genome Sciences, University of Washington, Seattle, WA 98195; and <sup>b</sup>Department of Medicine, University of Washington, Seattle, WA 98195