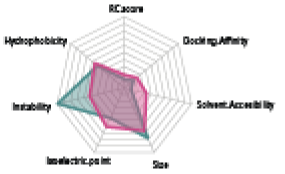
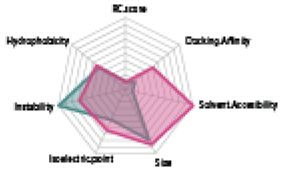
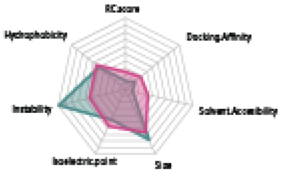
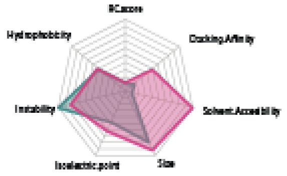
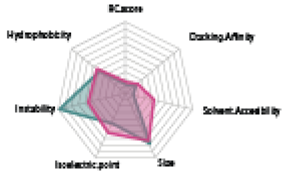
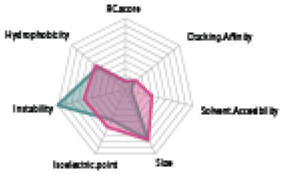
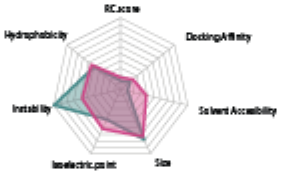
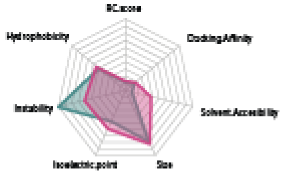
Supplementary Material

Supplementary Text 1:

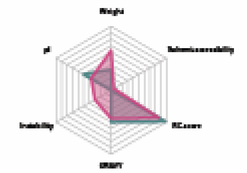
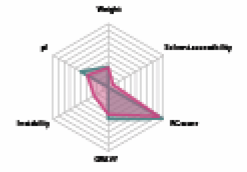
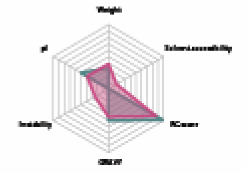
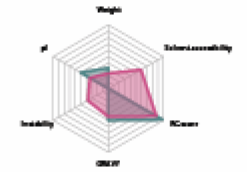
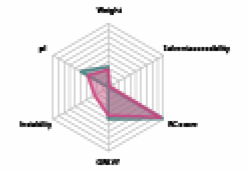
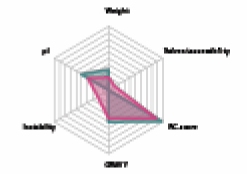
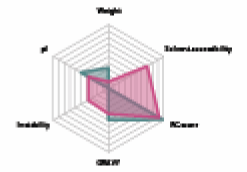
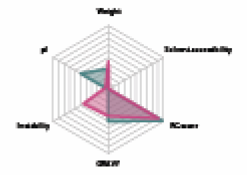
Calculation of *in silico* parameters

Many *in silico* features, including those of *Molecular Weight*, *Theoretical pI*, and *Instability Index* used in this study, are calculated using the ProtParam facility, hosted by expasy. This web-server takes as input only the amino acid sequence and does not require any further user engagement. As these are all calculated based on the amino acid sequence they can all alternatively be calculated with simple scripts in R or Python, as has been done with *Grand Average of Hydropathicity* (see R script in Github repository). The protein tertiary structure prediction was generated by the I-TASSER suite (v5.1), this tool also provides a file detailing the per residue *solvent accessibility.* This can be subset in R to find the accessibility of the active site. If I-TASSER is not used, online tools for solvent accessibility of particular residues exist, such as the GETAREA tool, hosted by the Sealy Center for Structural Biology. The Ramachandran plot is generated on the Saves Server, using the Verify3D utility.

The protein-protein interaction or “Docking” was modelled using a heuristic implementation of the Autodock Vina algorithm, Qvina2, within the MGLTools/Autodock Tools (v1.5.6) interface. The number of potential active sites within a test sequence was calculated using COACH, which is another algorithm within the ITASSER suite.

Test Sequence	F2F Result	F2F Score
Muc1 Diabody 1		6.68
Muc1 Diabody 2		9.48
Muc1 Diabody 3		6.9
Muc1 Diabody 6		9.27
Muc1 Monobody 1		5.79
Muc1 Monobody 2		4.69
Muc1 Monobody 3		5.73
Muc1 Monobody 4		5.04

Supplementary Figure 1: F2F-bridge output for MUC1 test sequences

Test Sequence	F2F Result	F2F Score
Construct 1		8.67
Construct 2		5.61
Construct 3		6.73
Construct 4		12.58
Construct 5		4.92
Construct 6		6.43
Construct 7		14.35
Construct 8		10.68

Supplementary Figure 2: F2F-bridge output for fluorescence test sequences

Supplementary Text 2:

General description of F2F bridge workflow.

1) **Collection of data**

The protein scientist must identify the features considered important for the analysis, from the literature or from experience. The predictive features to be included in the analysis must be converted to the same scale.

2) **Preparation of data**

The data must be stored in a table in the following format:

-Columns must be the predictive features selected for the experiment

-Rows 4 to end must be the unique names of the test sequences to be analysed

-Rows 1 and 2 must be the minimum and maximum values (Should be scaled 1:100 unless impossible)

-Row 3 should contain the user supplied values either taken from the literature or suited to the experimental conditions

3) **Running the programme and creating the data**

The F2F function takes as input a table prepared in the manner described in step 2 and produces both a plot for each sequence and a data frame containing all sequences and their associated F2F-plot score. The script can be called from the linux command line, or executed within R. For high throughput analysis, the option of generating a plot can be disabled. The data frame of scores will be saved to the current working directory.

4) **Database free mode**

As a database of protein test sequences, their OP-scores, and their overall biological performance is ideally required to apply a system of weights to the predictive features used in the plot, an alternative is provided until such a database can be established. A function for feature selection with

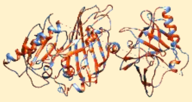
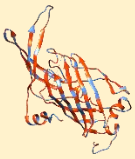
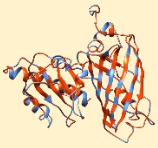
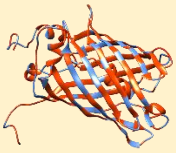
LASSO is provided, and can be used to detect relationships between the input *in silico* data and the overall performance on a subset of the experimental data, and the resulting model can then be applied to the remaining data. The user is not restricted to the LASSO function provided, a variety of tools for feature selection and subsequent model building exist, such as RandomForest which was also implemented in the main manuscript.

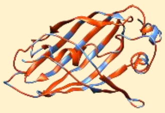
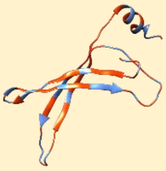
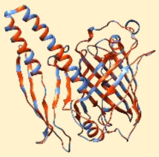
Comprehensive annotated code for F2F bridge can be found at https://github.com/Sidneyw91/F2F-Bridge

Supplementary Text 3:

Fluorescence proteins

Eight synthetic proteins based on the mCerulean Fluorescent Protein, or parts thereof, were generated, corresponding to Supplementary Figure 3.

Fluorescence construct number	3D Model	Construct Description	Wetlab Fluorescence
1		Split mCerulean with modification 1 (docked)	3.06E+08
2		Split mCerulean (docked)	1.05E+09
3		Split mCerulean with chromophore and modification 1	2.91E+07
4		Split mCerulean without chromophore and with modification 1	2.52E+07
5		mCerulean	1.52E+09

6		Split mCerulean with chromophore	3.73E+07
7		Split mCerulean without chromophore	1.88E+07
8		Split mCerulean with modification 2 (docked)	2.85E+08

Protein-related Laboratory Methods

DNA construct design and build: DNA sequences were obtained by reverse translating the amino acid sequences using EMBOSS Backtranseq (https://www.ebi.ac.uk/Tools/st/emboss_backtranseq/). The DNA sequences were codon optimized using IDT codon optimisation tool (https://eu.idtdna.com/codonopt). Each DNA construct was designed with a FLAG-tag and homology arms which were verified for upstream experiments using SnapGene's Gibson Assembly simulator (SnapGene.com).

Gene Block synthesis: Gene blocks for the test constructs were sourced from IDT (Integrated DNA Technologies, Inc) and amplified using corresponding PCR primers. The amplicons were verified using gel electrophoresis (1.5 % agarose) and ImageLab 5.2.1, (Bio Rad Inc) was used for band visualisation.

Primer Design: Primers were designed using Benchling (Benchling.com) to determine appropriate regions for construct amplification, followed by the use of Primer3Plus to test the primer suitability in terms of appropriate Tm as well as the presence of G-C clamps. NEBuilder assembly tool (www.nebuilder.neb.com) was used to design assembly primers for the purpose of facilitating construct insertion into the plasmid during Gibson Assembly. The finalised primers were obtained from IDT.

Competent E. coli: *E.coli* cells were made competent following the protocol described in Cohen et al. 1972. All cells were stored at -80 ˚C and thawed at room temperature. OG176 (Oxford genetics, mammalian expression vector) was used for amplification and expression of the test sequences with luminescence as the overall function and RSFDuet-1 (Novagen, bacterial expression vector) was used for amplification and expression of the test sequences with fluorescence as the overall function. Both the expression plasmids included Kanamycin resistance gene (Kn^{R}). For plasmid amplification, the plasmids were transformed into *E. coli* BL21 by mixing 100 ng plasmid DNA into 30 µL of competent cells. The cells were incubated on ice for 20 min and heat shocked by placing at 42˚C for 45 sec. The cells were then placed on ice for a further two min. The cells are then suspended into 500 µL of LB, 100 µL transformed cells were cultured on LB agar supplemented with 50 µg/mL kanamycin and incubated O/N at 37 ˚C. Select colonies were then grown in 20 mL liquid LB with 30 ng/mL kanamycin O/N.

Plasmid Extraction: After suspension in liquid LB supplemented with 30 ng/mL kanamycin O/N, transformed cells were pelleted by centrifugation at 4000 rpm (2500 x g) for 10 min. Following the instructions of the Monarch Plasmid miniprep kit (New England Biolabs) plasmid DNA was extracted, eluted in 15 µL EB and DNA concentration was quantified with a Nanodrop. The eluted samples were stored at -20 ˚C until further processing.

Restriction Digestion: The plasmids were digested by appropriate restriction enzymes (*NcoI, AflII, NdeI, and AvrII*) with the addition of CutSmart reaction buffer (New England Biolabs) and dH2O, for a total reaction volume 50 µL. The sample was then incubated at 37 ˚C for 1 h, after which time the digestion was confirmed by gel electrophoresis on a 1.5 % agarose gel at 80 V for 90 min. Plasmid DNA was purified using a PCR purification kit (Qiagen) and eluted in 15 µL EB.

Gibson Assembly: The assembly master mix was made up in accordance to the protocols and reagents described by DG Gibson et al 2009. The gene blocks were combined with the plasmid in a DNA concentration ratio of 3:1 in which 72 ng/µL plasmid DNA was incubated in a Gibson Assembly master mix with 225 ng/µL of construct DNA. The mixture was incubated for 1h at 50 ºC followed by transformed into *E. coli* BL21 cells.

Colony PCR: Colony PCR was used to determine the success of the Gibson Assembly and evaluate the transformation of the construct into bacterial cells. In this case, select colonies were added to a PCR master mix containing; 25 µL Q5 polymerase (NEB), 2.5 µL of forward and reverse primers and 20 µL milliQ. Sanger sequencing was then carried out by GATC's light-run service and was verified by aligning with a reference sequence.

Mammalian cell transfection (Luminescence proteins): CHO-K1 (ATCC® CCL-61™) cells were used for luminescence protein production. Turbofect transfection reagent (Cat No: R0532) was used for *in vitro* transfection. Transfection was carried out using manufacturer's protocol and supernatant containing protein collected after 48 h.

Binding assays: 10^8 *Staphylococcus aureus* TCH959 (naturally bearing *clfA)* or 10^6 MCF7 cells (naturally bearing MUC1) were blocked with 5% BSA for 2 h followed by incubation with supernatant containing each test construct. Cells were washed 3 times and resuspended in PBS. Luminescence was measured using Promega GloMax® 96 luminometer.

Fluorescence protein production and bacteria harvesting: Samples were grown overnight in liquid LB with 30 ug/mL kanamycin. 100 ml fresh LB was inoculated with 5 ml of overnight culture. Bacteria were induced with 1 mM Isopropyl ß-D-thiogalactoside at 0.5-0.6. OD. Bacteria were harvested when they reached an OD 0.8. Bacteria were washed and pelleted by centrifugation at 2,500 x g for 10 min. BugBuster lysing buffer supplemented with cOmplete protease inhibitor (Roche) and Lysonase reagent used for bacterial cell lysis according to the manufacturer's protocols. Protein production was confirmed by running an SDS page.

Fluorescence assays: Fluorescence was measured using an Omega Plate Reader (BMG LabTech) and IVIS Lumina II imaging system (Perkin Elmer). Samples were diluted in PBS and transferred to a 96 well plate to measure fluorescence.