

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	Fluorescence-based enzyme assays were collected using a BioTek/Agilent Neo2 plate reader. LC-MS measurements were acquired on an Agilent G6230B TOF instrument and spectra were deconvoluted using BioConfirm version 12.0 software. Circular-dichroism measurements were collected using a Jasco J-1500 instrument. Mammalian-cell fluorescence images were acquired using an IN Cell Analyzer 2500HS controlled with IN Cell Analyzer 2500 HS software version 7.4. X-ray diffraction data were collected at APS NE-CAT 24ID-E and NSLS-II 17-ID-1/17-ID-2 beamlines.
Data analysis	Computational protein design and structural analysis used RFDiffusion2 for Molecular Interfaces (RFD2-MI), ProteinMPNN, AlphaFold 3, Rosetta FastRelax, DSSP, PLACER, Foldseek (https://search.foldseek.com/search), PyRosetta and AMBER24. Crystallographic data were processed using XDS (ver Jun 30, 2024 BUILT=20240723/ 20250119) and Pointless/Aimless (version 0.7.15) in CCP4 (vesrion 8.0.017); structures were solved with Phaser (version 2.8.3), refined with Phenix (version 2.1/2.2) and Refmac (version 5.8.0), built in COOT (version 0.9.8.95) and evaluated using MolProbity (version 4.02). Biochemical kinetic data were analysed using custom Python scripts and nonlinear regression. Mammalian-cell images were background corrected in Fiji and analysed using a custom Jupyter/Python workflow incorporating scikit-image. Code used for biochemical kinetic data and mammalian-cell images analysis is publicly available at Github (https://github.com/hojaec/Cysteine-protease-data-analysis). Gel band intensities were quantified using ImageJ (version 1.54p).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All data supporting the findings of this study are available within the Article, Extended Data and Supplementary Information. Atomic coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 9YNL (Dokki-15 C100A complex), 9YNM (Dokki-15 WT apo), 9YOX (Gawi-09 WT apo), 9YOY (Kal C103A complex), 9YOZ (Kal WT complex) and 9YPO (Kal WT apo). Protein and substrate sequences, LC-MS measurements, crystallographic statistics, structural-similarity analyses and kinetic comparisons are provided in the Supplementary Information.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Reporting on race, ethnicity, or other socially relevant groupings

Population characteristics

Recruitment

Ethics oversight

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size No statistical methods were used to predetermine sample size. Sample sizes were chosen according to the objectives of each analysis and experimental or computational feasibility. For the initial experimental screen, all 69 computationally selected designs were experimentally tested, providing complete coverage of the selected design set. Subsequent biochemical assays were generally performed with three technical replicates per condition to assess measurement variability and the consistency of the measured activities; these replicates assess technical precision rather than variability between independent experiments. Site-saturation mutagenesis experimentally evaluated all 185 variants that passed the pre-specified computational filters, providing complete coverage of the computationally selected variants. PLACER analyses used 50 independently seeded models per design to sample variability in predicted active-site geometry. Molecular dynamics analyses used five independent 500-ns simulation replicas per system to assess whether the observed structural behavior was consistent across independent trajectories, rather than relying on a single simulation. For the mammalian-cell imaging assay, approximately 200 cells were analyzed per condition to characterize cell-to-cell variability and estimate the imaging readout within each experiment. Individual cells were treated as observations within an experiment, rather than as independent biological replicates.

Data exclusions Computational designs were excluded only according to the pre-defined structural and confidence filters described in the Methods. For biochemical experiments, no experimental measurements were excluded unless explicitly described in the analysis procedure. For mammalian-cell image analysis, objects at image boundaries, objects failing the stated size criteria and cells lacking detectable GFP expression were excluded according to the predefined image-analysis workflow.

Replication Protein concentration and all standard curves were measured in triplicated and used on the same day as experiments. Biochemical activity, enzyme kinetic, and catalytic-residue knockout assays measurements were performed in three technical replicates per condition, and the observed trends were reproducible across replicates. Mammalian-cell imaging experiments were performed in 3 independent biological replicates, with consistent results across experiments. PLACER calculations used 50 independently seeded models per design; all 50 calculations per design completed successfully. These models were used to assess variability in predicted active-site geometry. Molecular dynamics simulations used five independent 500-ns trajectories per system; all trajectories completed successfully.

Randomization

Randomization was not applicable to this study. The work did not involve allocation of animals, human participants or biological samples into experimental treatment groups. Proteins were computationally generated and selected according to predefined structural and confidence criteria. Experimental variants were evaluated according to their computational design identities and therefore random allocation would not have been meaningful.

Blinding

Blinding was not performed because it was not relevant to the study design. The experiments involved biochemical characterization of computationally designed proteins with predetermined sequences and identities rather than allocation of participants or organisms into intervention groups. Measurements such as fluorescence, chromatography, LC-MS and X-ray diffraction generated instrument-based quantitative data. The same analysis procedures and predefined criteria were applied across designs.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- | | |
|-------------------------------------|---|
| n/a | Involved in the study |
| ☒ | ☐ Antibodies |
| ☐ | ☒ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology and archaeology |
| ☒ | ☐ Animals and other organisms |
| ☒ | ☐ Clinical data |
| ☒ | ☐ Dual use research of concern |
| ☒ | ☐ Plants |

Methods

- | | |
|-------------------------------------|---|
| n/a | Involved in the study |
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)

HeLa (HeLa CCL-2) were obtained from the American Type Culture Collect (ATCC).

Authentication

Authentication was provided by the vendor, ATCC. The cells we used were consistent with the size and shape of HeLa, and the doubling time and attachment durability were also consistent.

Mycoplasma contamination

The cell lines were not tested for mycoplasma contamination.

Commonly misidentified lines
(See [ICLAC](#) register)

There are no misidentified lines.

Plants

Seed stocks

Not involved in the study.

Novel plant genotypes

Not involved in the study.

Authentication

Not involved in the study.