Supplementary Information for

# *De novo* design of metalloproteases for targeted amyloid-β cleavage

Yannan Qu[1, 2, †], Chentong Wang[1, 2, †], Hongli Zhu[1,2], Yanjun Wang[1,2], Longxing Cao[1,2*]

1. State Key Laboratory of Gene Expression and Zhejiang Key Laboratory of Structural Biology, School of Life Sciences, Westlake University, Hangzhou, Zhejiang, China.

2. Artificial Intelligence Drug Design Core Laboratory, Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang, China.

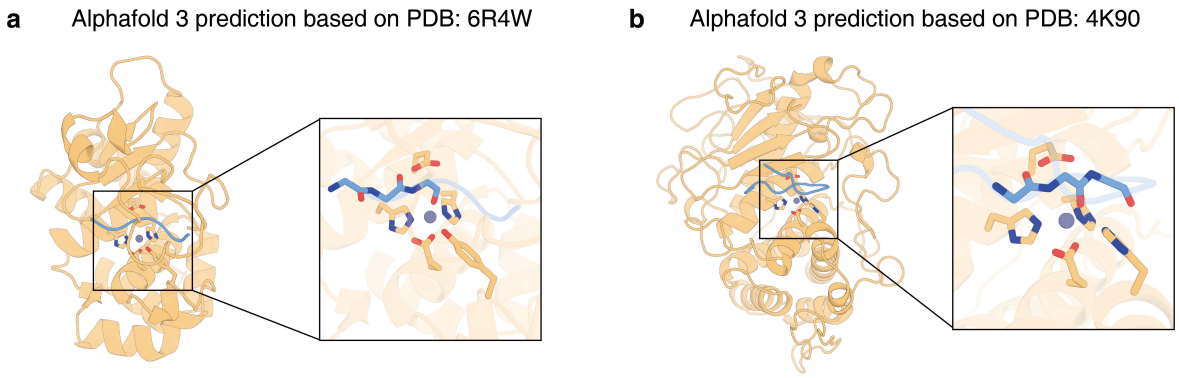
*Corresponding author. E-mail: caolongxing@westlake.edu.cn

† These authors are co-first authors.

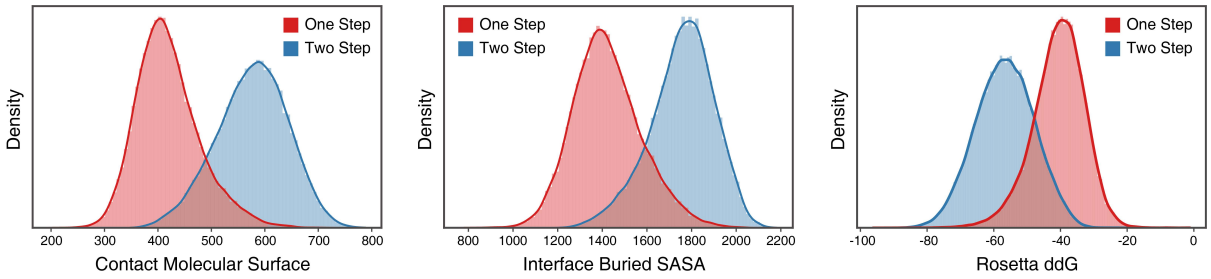
**The PDF file includes:**

Supplementary Figures S1 to S22

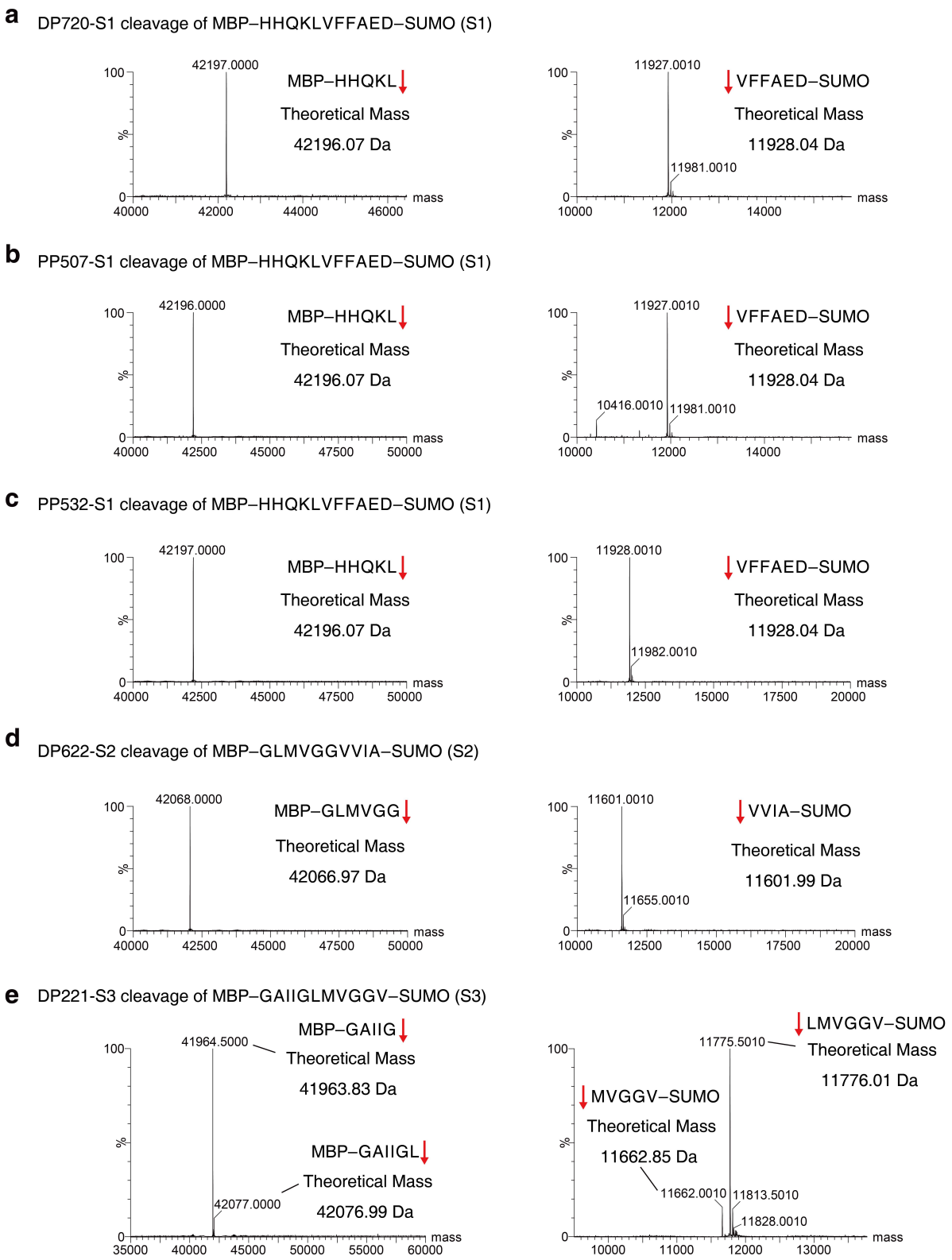
Supplementary Tables S1 to S4




**Supplementary Fig. S1 | AlphaFold3 predictions of native metalloprotease-substrate complexes.** Structural models of PPEP-1 (PDB: 6R4W) (**a**) and Fungalysin (PDB: 4K90) (**b**) in complex with their native substrate peptides were generated using AlphaFold3. The overall architectures are shown with their respective native substrates bound. Insets provide close-up views of the active sites, highlighting the atomic interactions among the catalytic residues, the divalent zinc ion, and the scissile peptide bond.

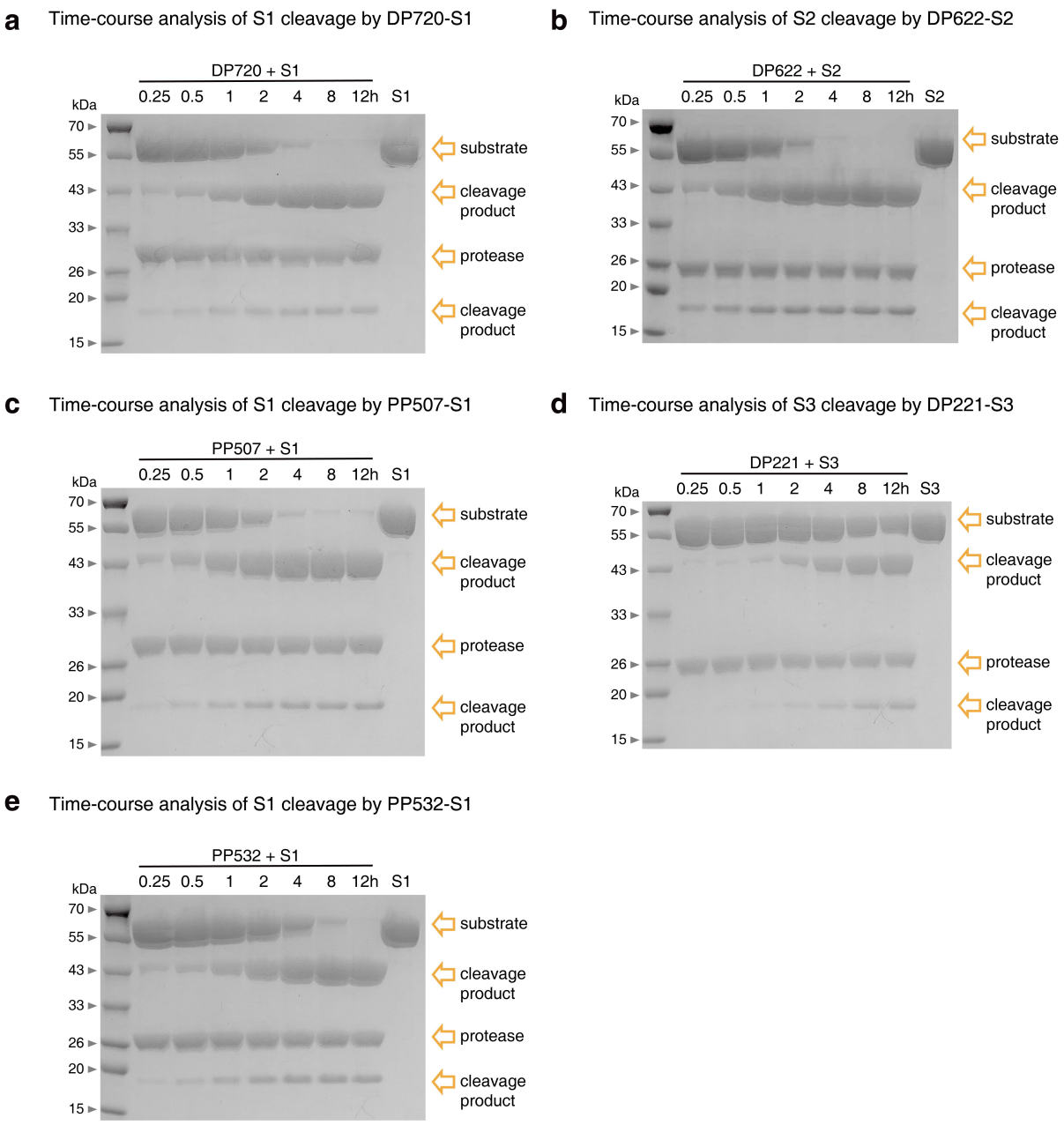



**Supplementary Fig. S2 | Quantitative evaluation of interface packing in one-step versus two-step design strategies.** Computational metrics for scaffold-target complexes generated via the one-step baseline and the two-step encapsulation strategy were calculated and compared. Distributions of Contact Molecular Surface, Interface Buried Solvent-Accessible Surface Area (SASA), and Rosetta binding energy (ddG) are shown for both design populations. The comparative analysis demonstrates that the two-step strategy yields systematically larger buried surface areas, increased surface contacts, and lower (more favorable) binding energies.

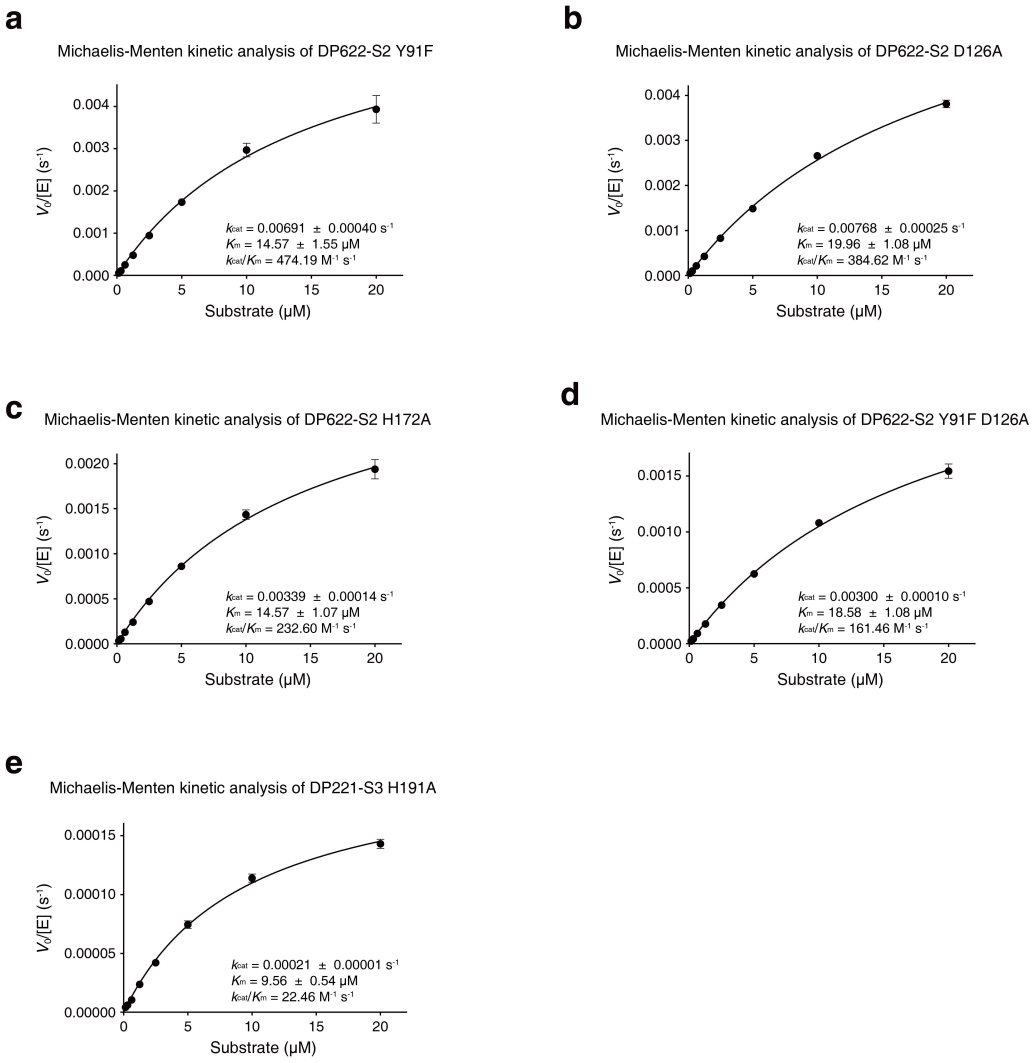



**Supplementary Fig. S3 | Mass spectrometric validation of pre-defined substrate cleavage by designed enzymes.** Mass spectrometry analysis of proteolytic fragments was used to verify the accuracy of the cleavage sites. The cleavage sites for the designed proteases (**a**, DP720-S1;

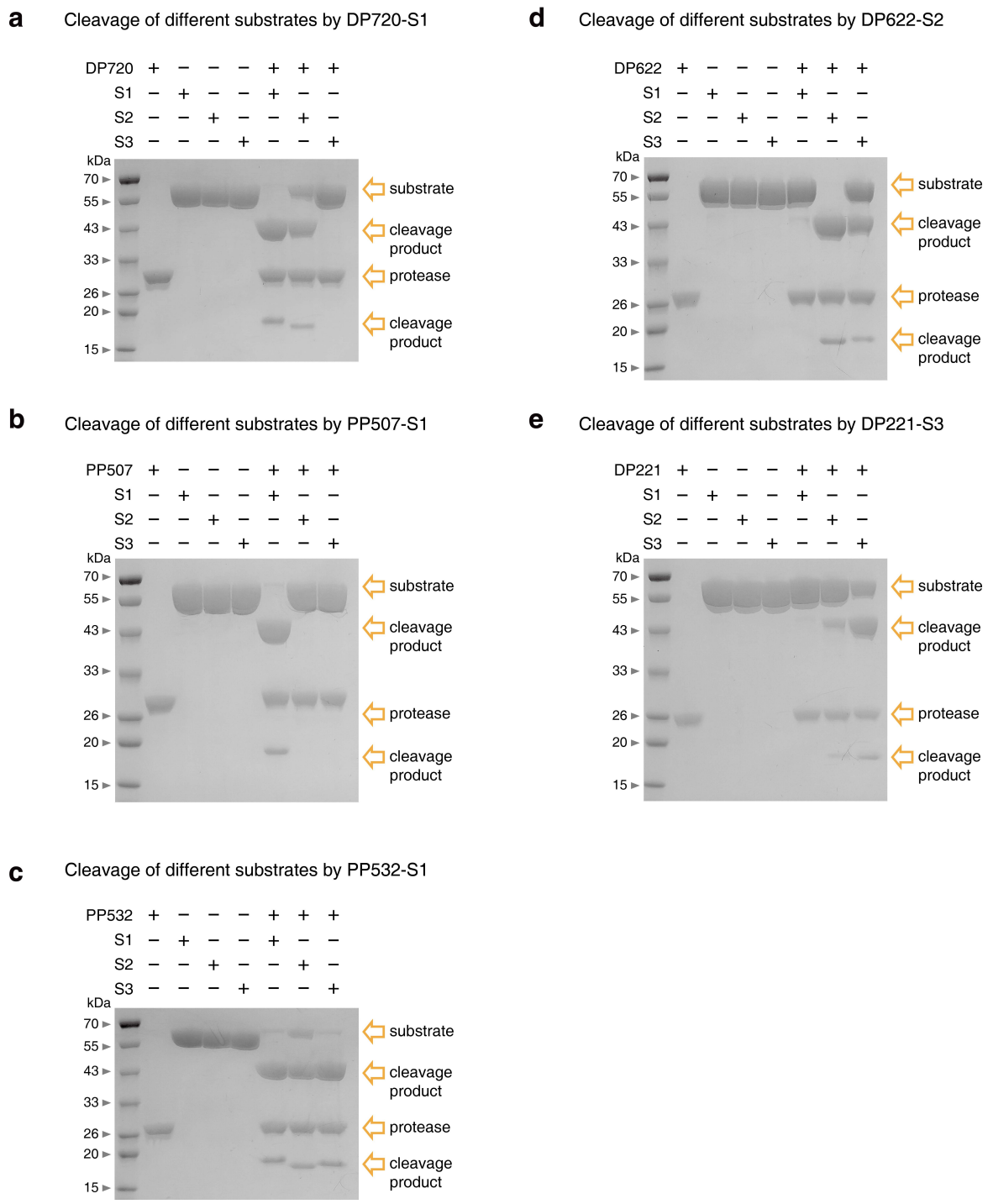
**b**, PP507-S1; **c**, PP532-S1; **d**, DP622-S2; **e**, DP221-S3) were mapped by measuring the molecular weights of the resulting peptide fragments.



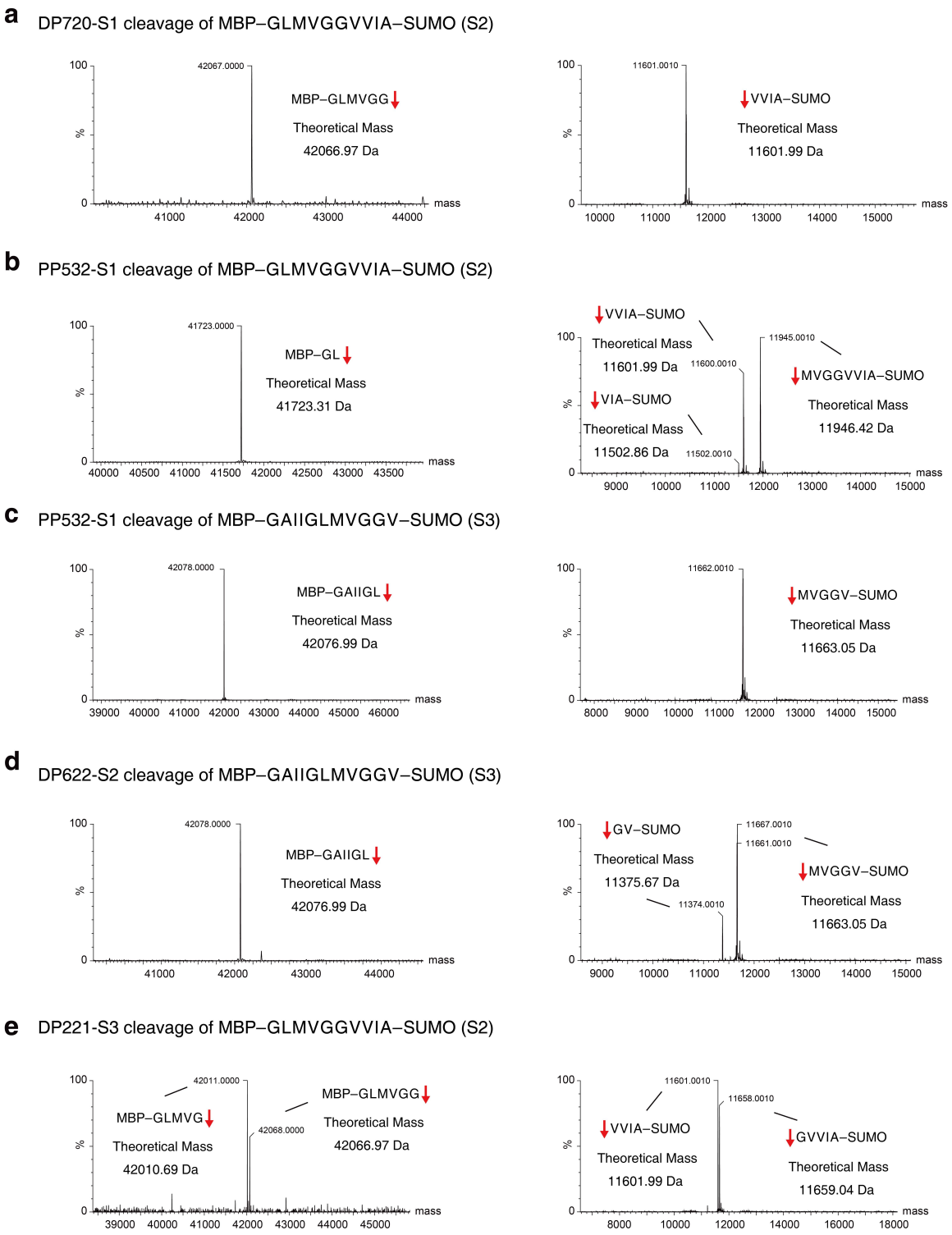

**Supplementary Fig. S4 | Time-course SDS-PAGE analysis of proteolytic activity.** The proteolytic activity of the designed proteases (**a**, DP720-S1; **b**, DP622-S2; **c**, PP507-S1; **d**, DP221-S3; **e**, PP532-S1) was monitored using SDS-PAGE at the indicated time intervals. The gel profiles show the progressive depletion of the MBP-Aβ_segment-SUMO substrate and the corresponding appearance of cleavage products over the course of the reaction.




**Supplementary Fig. S5 | Kinetic characterization of designed protease variants.** Michaelis-Menten kinetics were determined for DP622-S2 Y91F (**a**), D126A (**b**), H172A (**c**), Y91F/D126A (**d**), and DP221-S3 H191A (**e**) to assess the functional impact of mutations of the oxyanion hole and second coordination sphere residues.

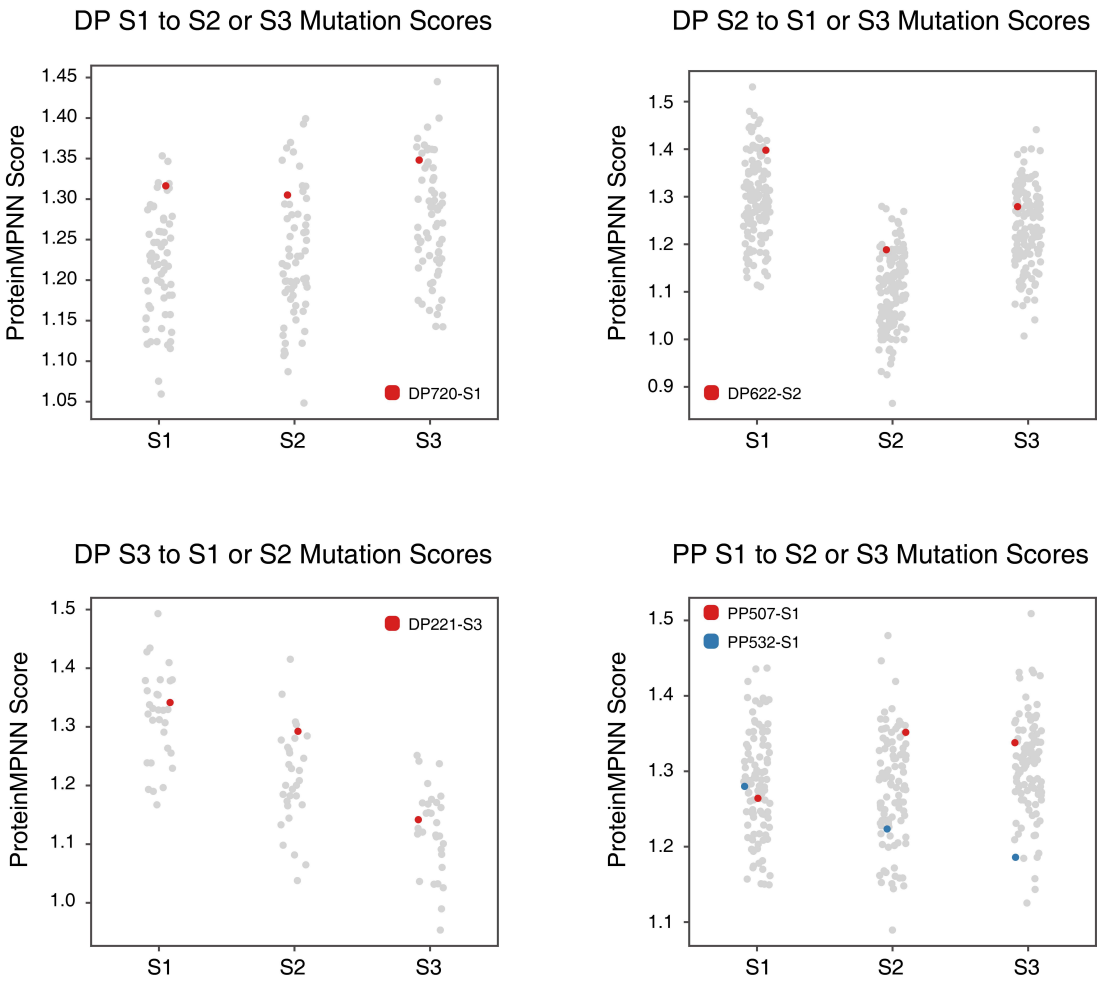



**Supplementary Fig. S6 | Proteolytic substrate selectivity of designed proteases.** The substrate recognition profiles of five designed proteases (**a**, DP720-S1; **b**, PP507-S1; **c**, PP532-S1; **d**, DP622-S2; **e**, DP221-S3) were systematically evaluated using the three structurally distinct MBP-Aβ_segment-SUMO fusion constructs (Substrates 1, 2, and 3).



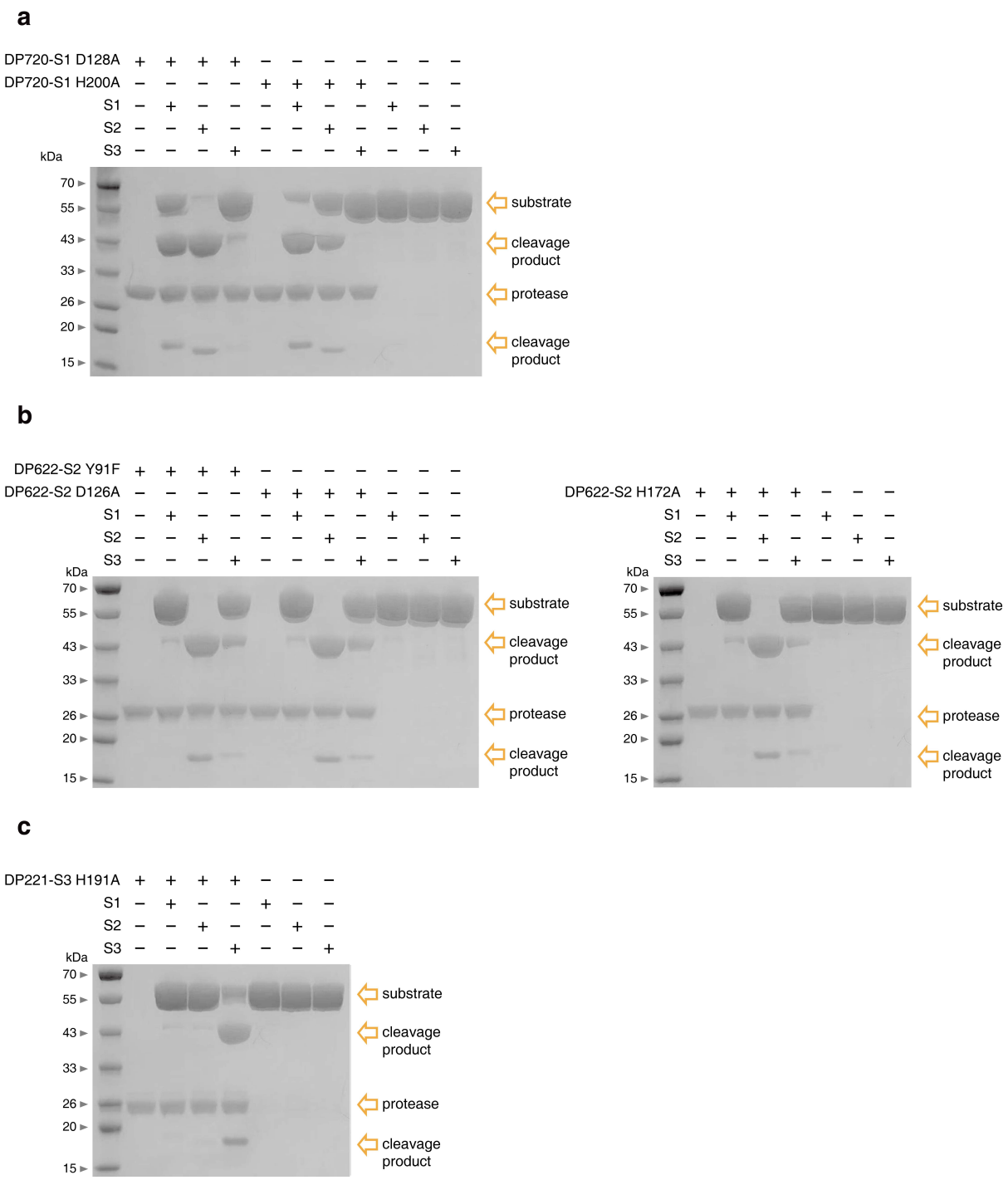

**Supplementary Fig. S7 | Mass spectrometric validation of non-intended substrates cleavage by designed enzymes.** Mass spectrometry analysis of proteolytic fragments was employed to determine the exact cleavage locations on non-intended substrates. The distinct cleavage sites generated by the designed proteases (**a**, DP720-S1 against S2; **b**, PP532-S1

against S2; **c**, PP532-S1 against S3; **d**, DP622-S2 against S3; **e**, DP221-S3 against S2) were mapped by measuring the molecular weights of the resulting fragments.

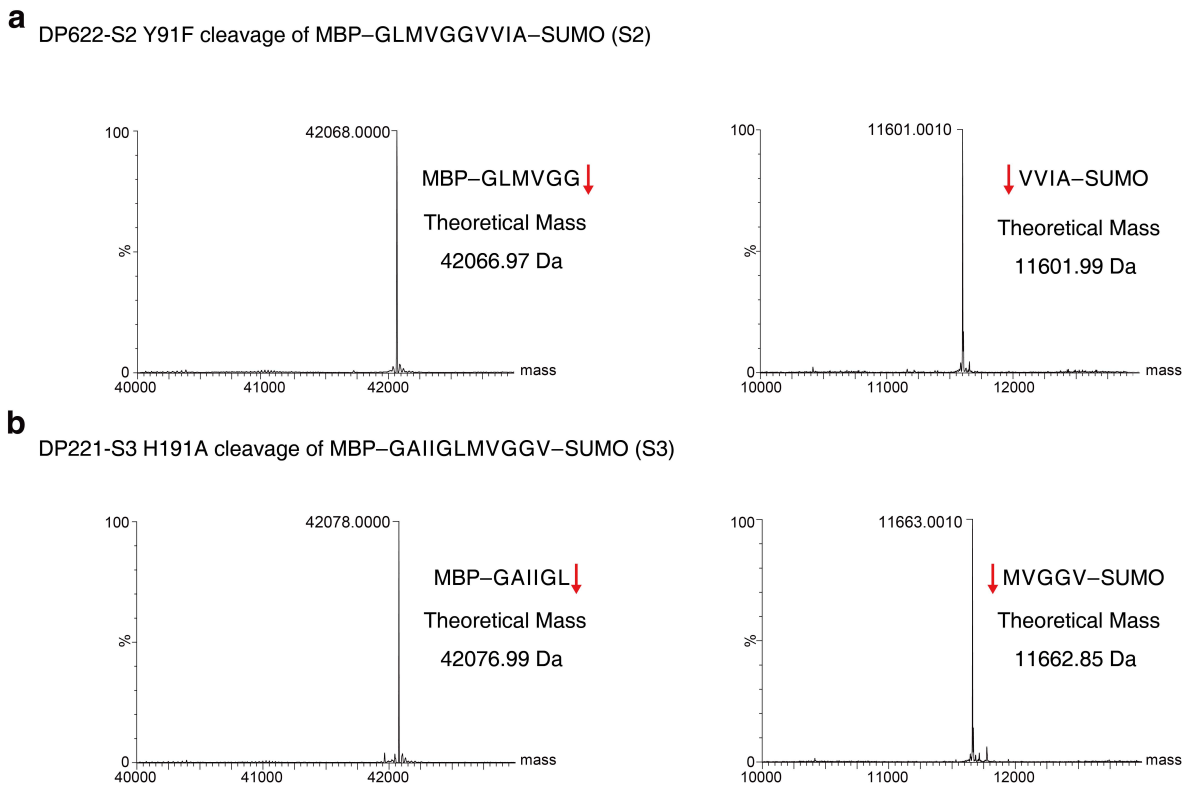



**Supplementary Fig. S8 | Computational assessment of substrate substitution on designed proteases.** ProteinMPNN sequence scores were calculated for all selected designs from four independent batches (DP-S1, DP-S2, DP-S3, and PP-S1) after substituting the target with two other substrates. The X axis represents the substrate sequences, and the Y axis indicates the calculated ProteinMPNN scores. Highlighted points on the plot denote the specific enzymes successfully obtained through experimental screening. Retrospective analysis reveals that these computational preferences correlate well with the experimental profiles. Specifically, DP720-S1 scores for S1 and S2 are comparable and lower than S3, mirroring the cleavage assays. For DP622-S2 and DP221-S3, scores for intended targets are significantly lower than for S1, explaining their primary selectivity; however, these static scores do not capture their observed

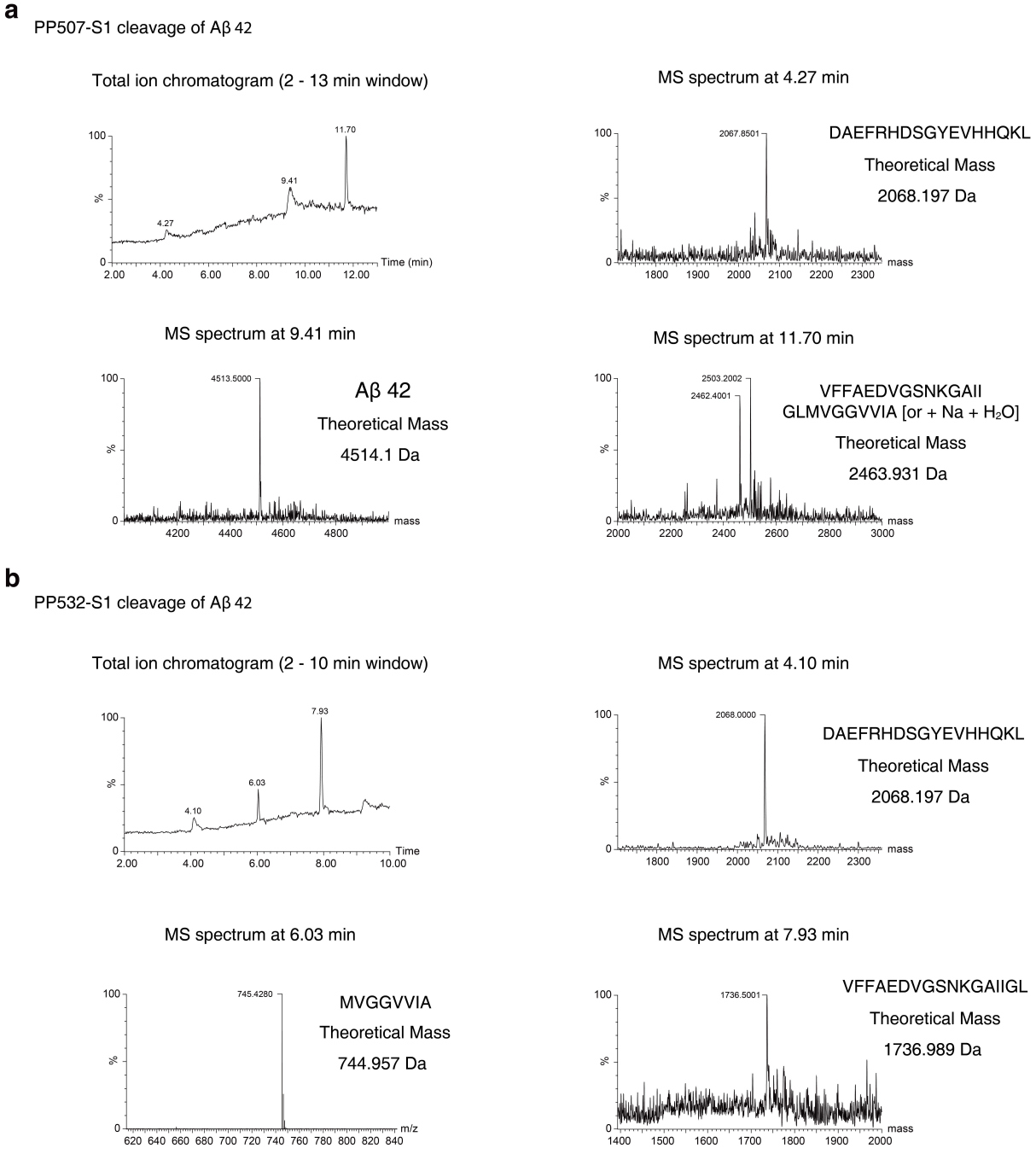
cross reactivity with S3 and S2, respectively. Within the PP series, PP507-S1 scores best for S1, correctly reflecting its target selectivity, whereas PP532-S1 scores worse for S1 than alternative substrates, capturing its substrate promiscuity. Overall, these results demonstrate the utility of ProteinMPNN for predicting broad cross reactivity trends, while the inability to capture specific nuances presumably arises.




**Supplementary Fig. S9 | Proteolytic substrate selectivity of the designed proteases mutants.** Substrate recognition and cleavage site fidelity were characterized for the protease mutants (**a**, DP720-S1; **b**, DP622-S2; **c**, DP221-S3). The impact of site directed mutagenesis on substrate selectivity was assessed using a panel of intended and unintended substrates.

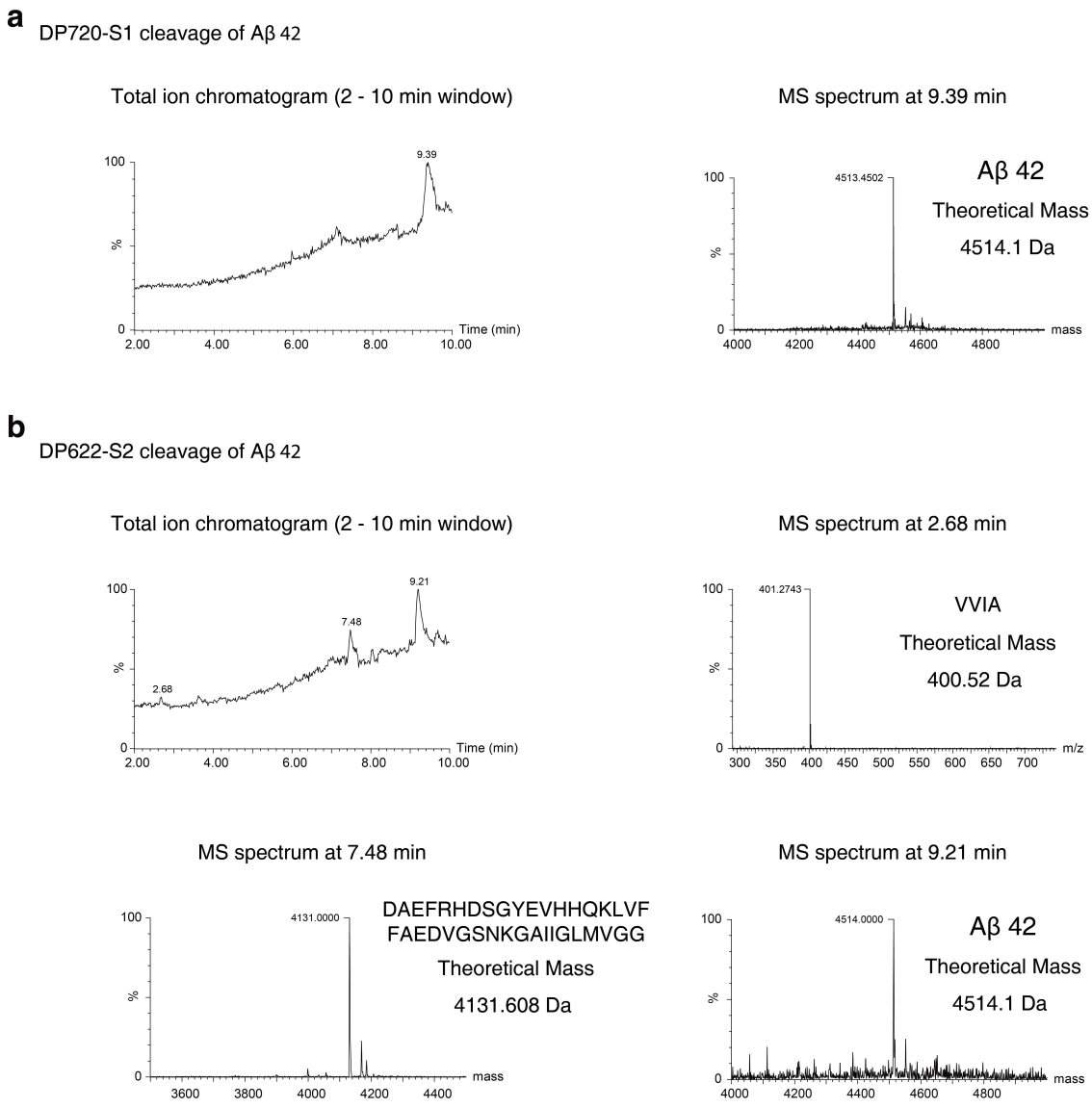



**Supplementary Fig. S10 | Mass spectrometric identification of cleavage sites for DP622-S2 Y91F and DP221-S3 H191A. a** DP622-S2 Y91F retained the designed cleavage site. **b** DP221-S3 H191A exhibited a clear shift in cleavage specificity, with the scissile bond moving one residue toward the C-terminus — from the G-L peptide bond to the L-M peptide bond.



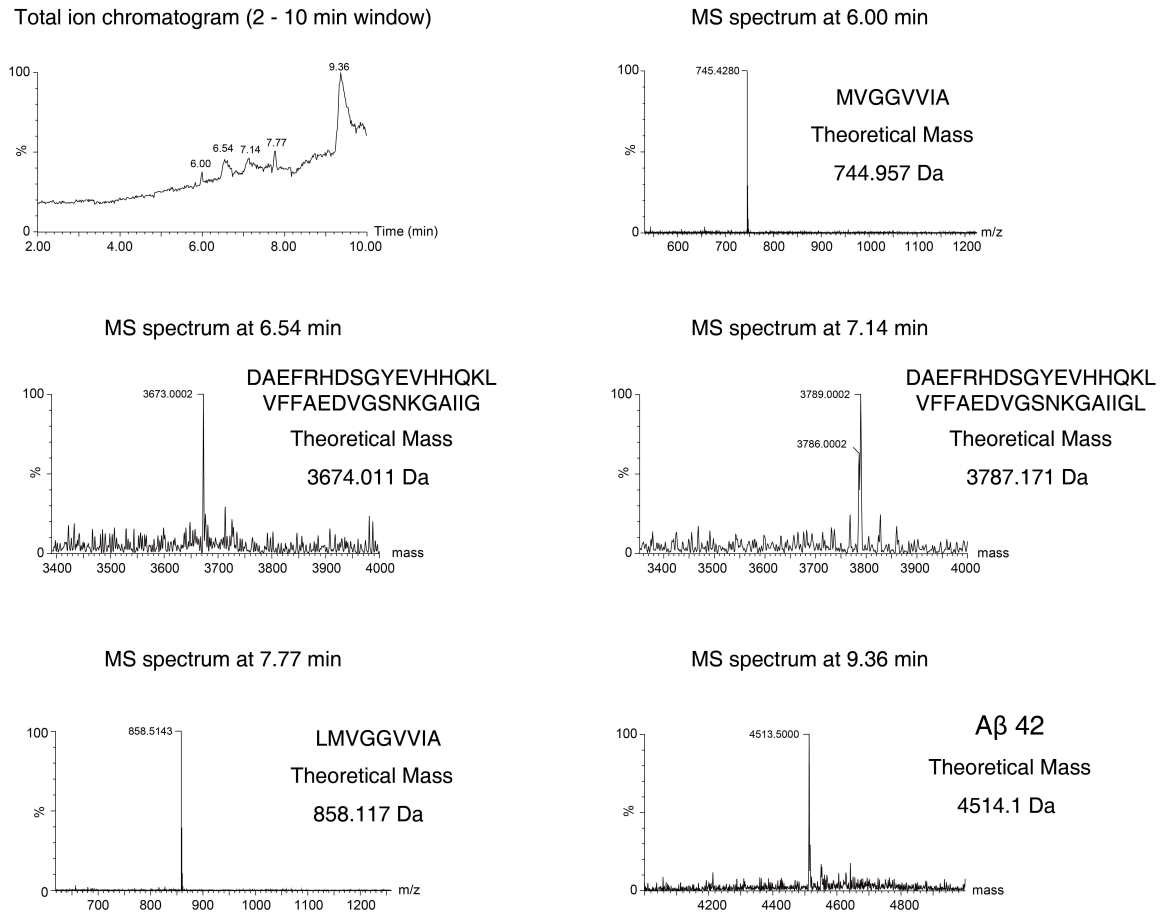

**Supplementary Fig. S11 | Mass spectrometric validation of Aβ42 peptide degradation.** LC-MS analysis of Aβ42 proteolysis by PP507-S1 (**a**) and PP532-S1 (**b**). Total ion chromatograms (TICs) show the elution profiles of the cleavage products, accompanied by mass spectra (m/z or deconvoluted mass) for the annotated peaks. PP507-S1 yields fragments corresponding to a

single intended cleavage site, while PP532-S1 generates fragments from multiple cleavage positions.

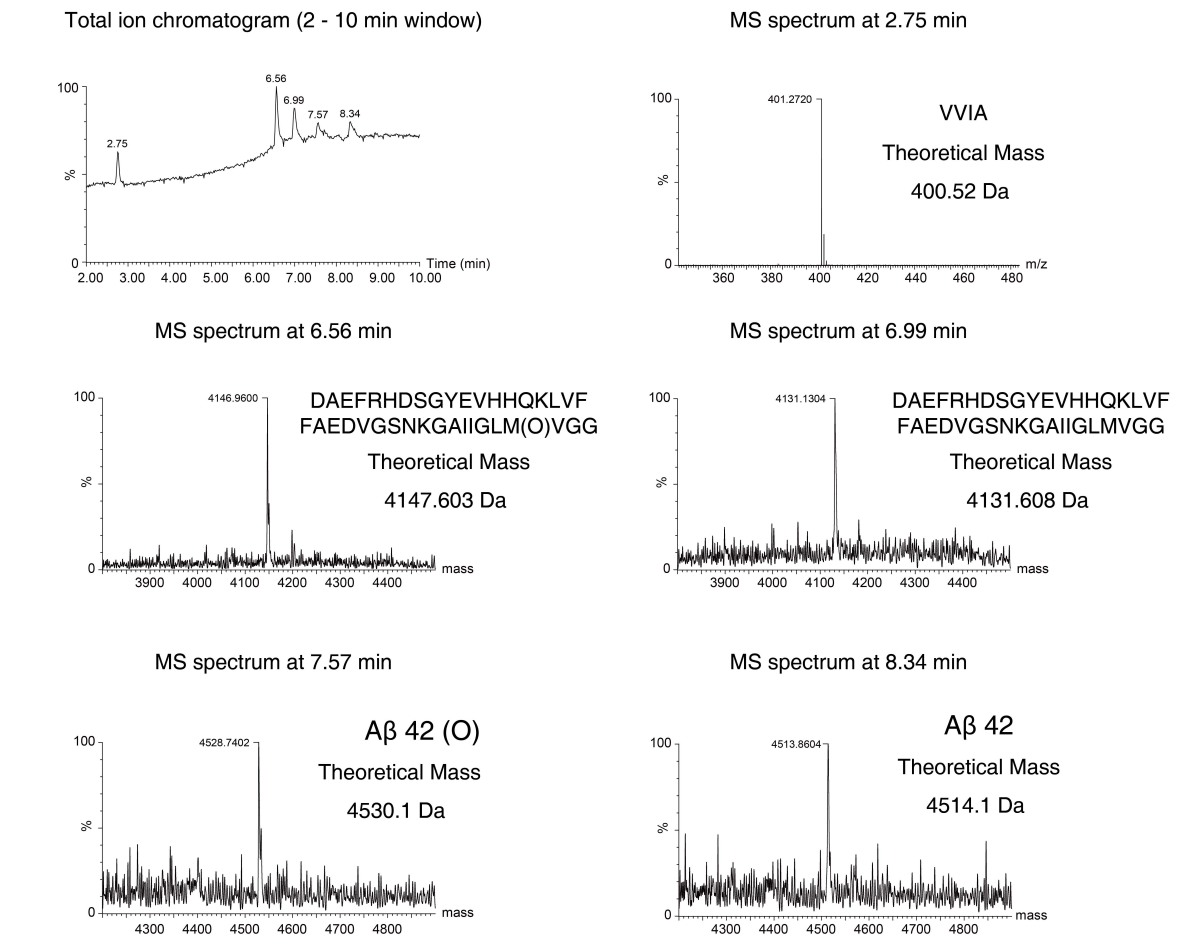



**Supplementary Fig. S12 | Mass spectrometric validation of Aβ42 peptide degradation.** LC-MS analysis of Aβ42 proteolysis by DP720-S1 (**a**) and DP622-S2 (**b**). TICs and mass spectra show no cleavage of Aβ42 by DP720-S1, whereas DP622-S2 cleaves precisely at the intended site.

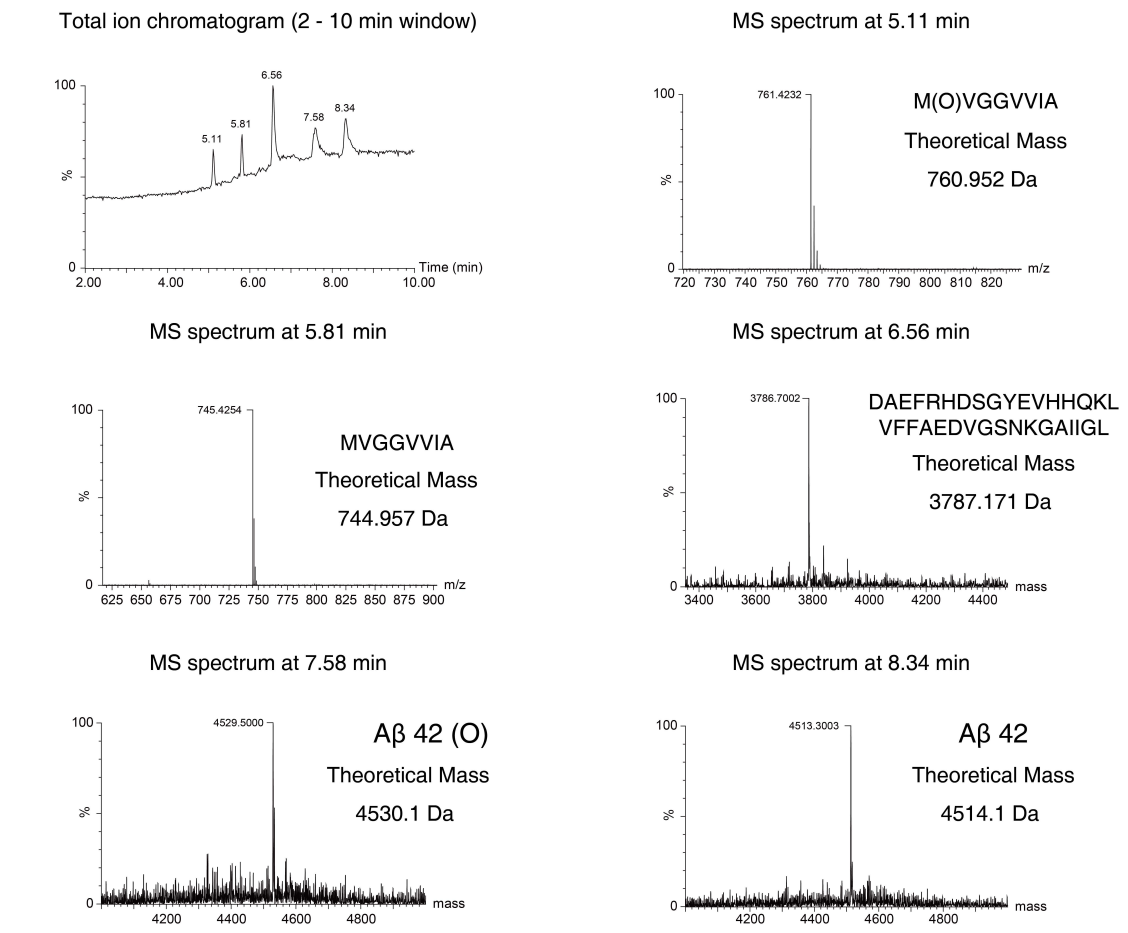
DP221-S3 cleavage of Aβ 42



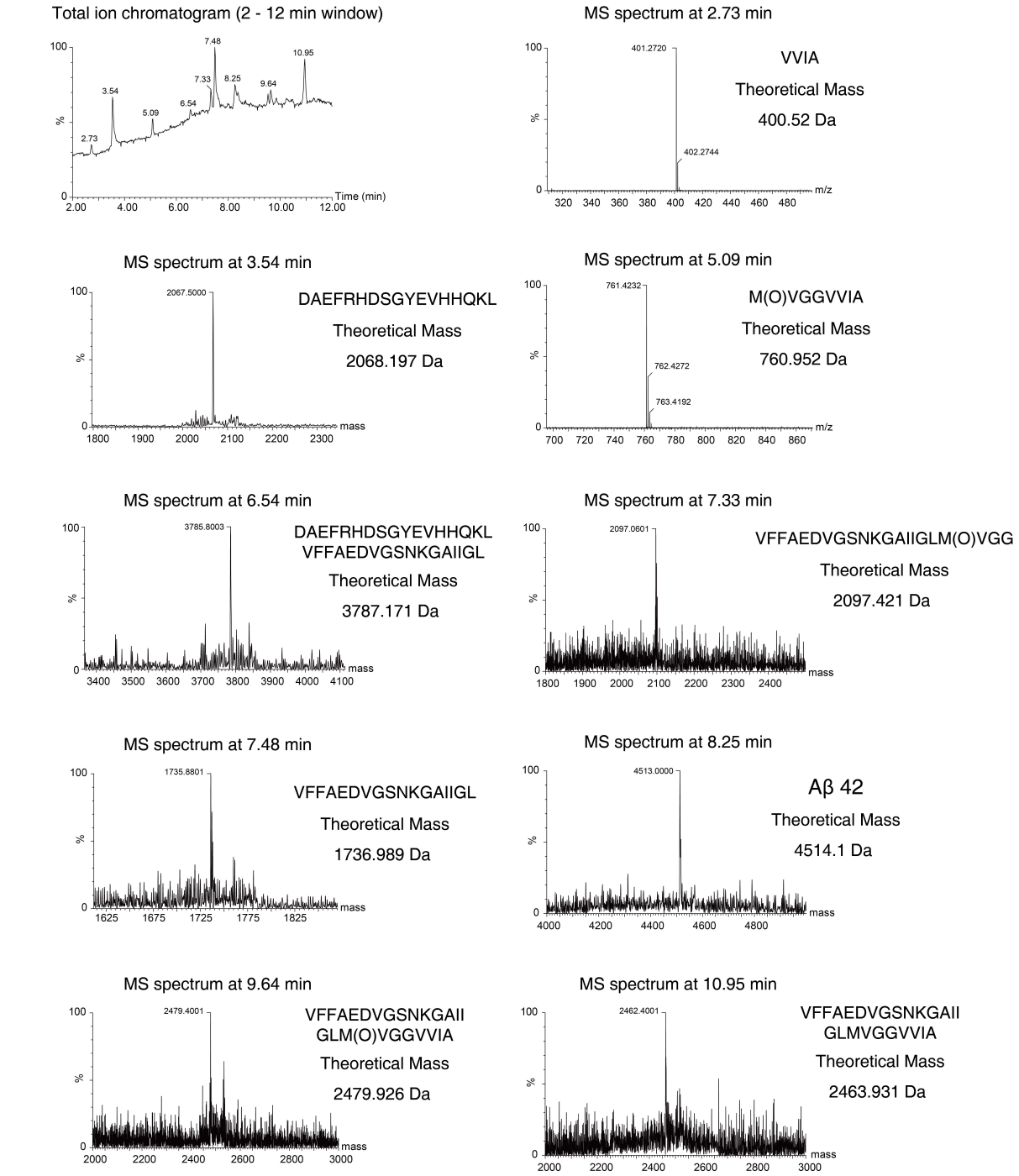

**Supplementary Fig. S13 | Mass spectrometric validation of Aβ42 peptide degradation.** LC-MS analysis of Aβ42 proteolysis by DP221-S3. TICs and mass spectra reveal fragments corresponding to cleavage at multiple positions across the full length Aβ42 sequence. M(O) denotes methionine oxidation.




**Supplementary Fig. S14 | Mass spectrometric validation of Aβ42 peptide degradation.** LC-MS analysis of Aβ42 proteolysis by DP622-S2 Y91F. TICs and mass spectra of annotated peaks confirm cleavage of the full length Aβ42 peptide at the intended site. M(O) denotes methionine oxidation.

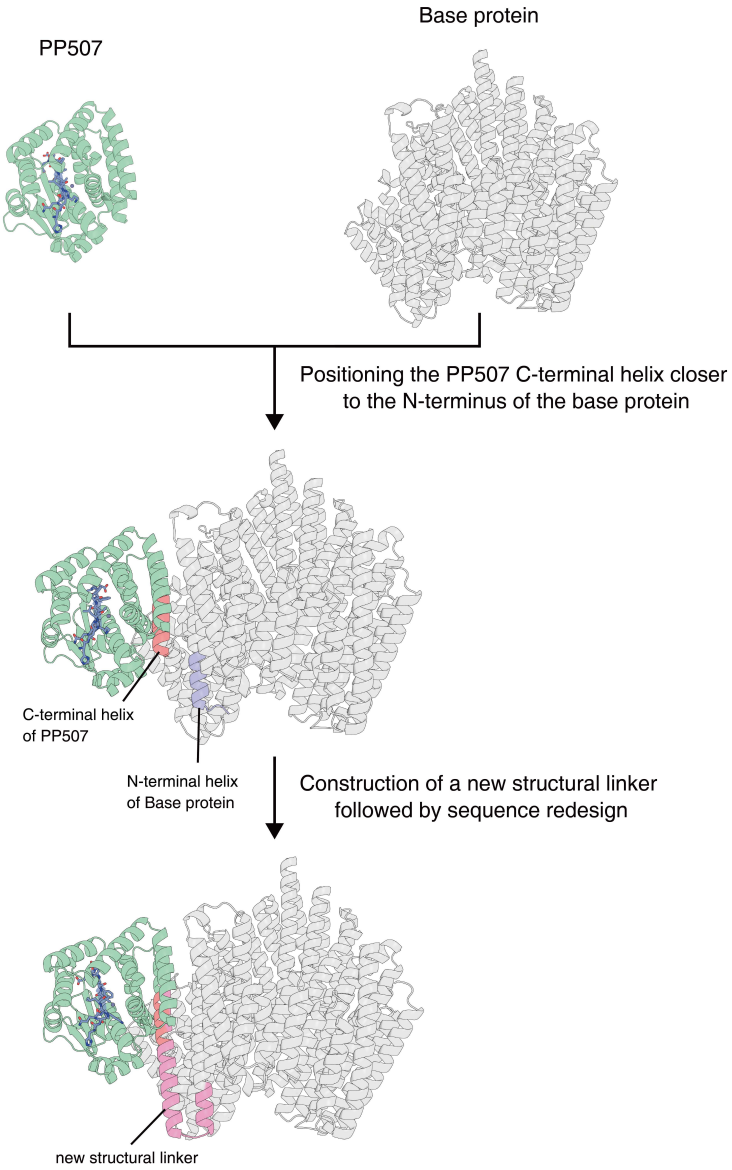



**Supplementary Fig. S15 | Mass spectrometric validation of Aβ42 peptide degradation.** LC-MS analysis of Aβ42 proteolysis by DP221-S3 H191A. TICs and mass spectra of annotated peaks confirm cleavage of the full length Aβ42 peptide at the intended site. M(O) denotes methionine oxidation.

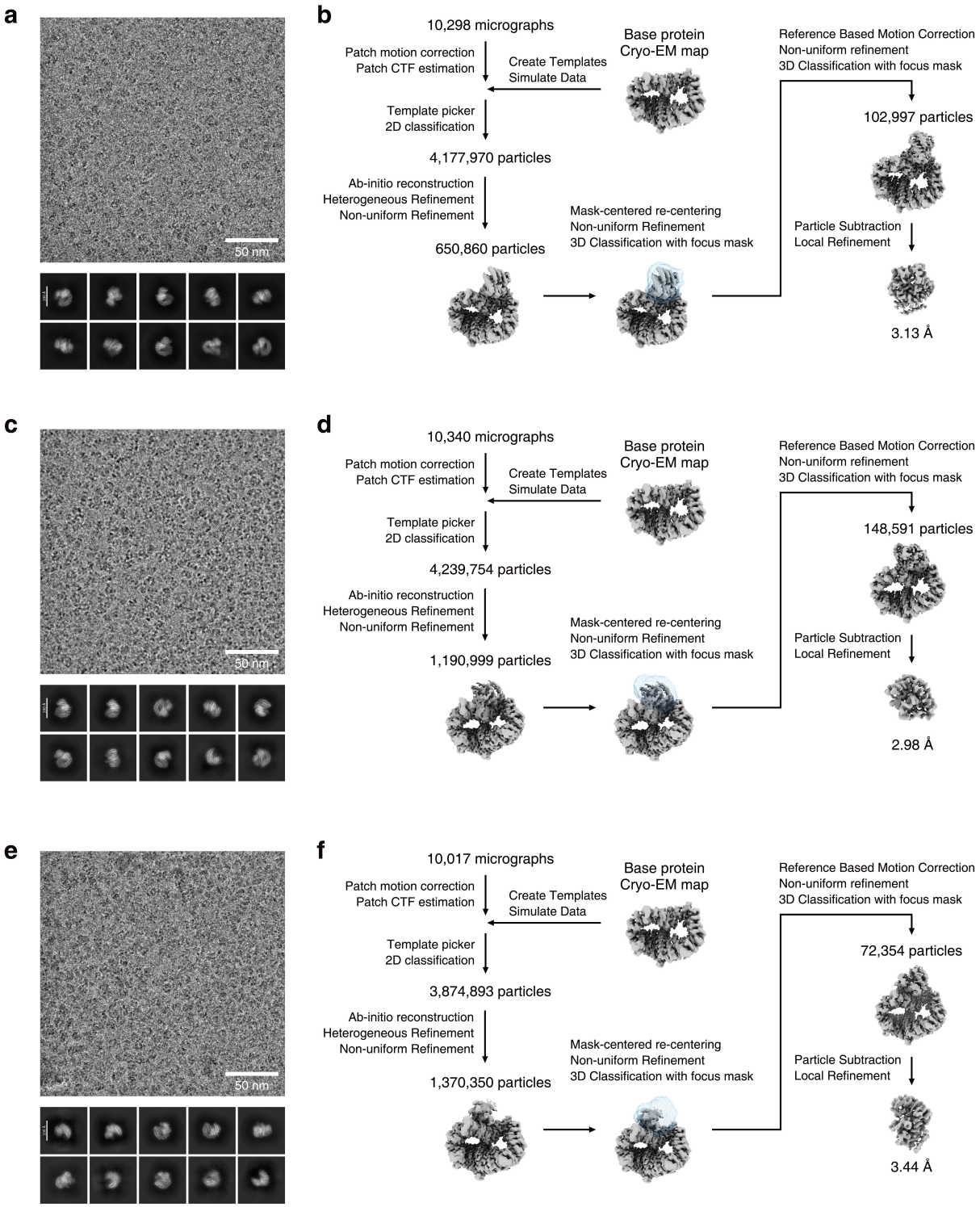



**Supplementary Fig. S16 | Mass spectrometric validation of Aβ42 peptide degradation.** LC-MS analysis of Aβ42 proteolysis by a cocktail of PP507-S1, DP622-S2 Y91F, and DP221-S3 H191A. The total ion chromatogram (TIC) from 2 to 12 minutes displays the elution profile of the cleavage products, accompanied by the corresponding mass spectra (m/z or deconvoluted

mass) for the annotated peaks. These spectra confirm the efficient degradation of the full-length synthetic Aβ42 peptide into distinct fragments. M(O) denotes methionine oxidation.

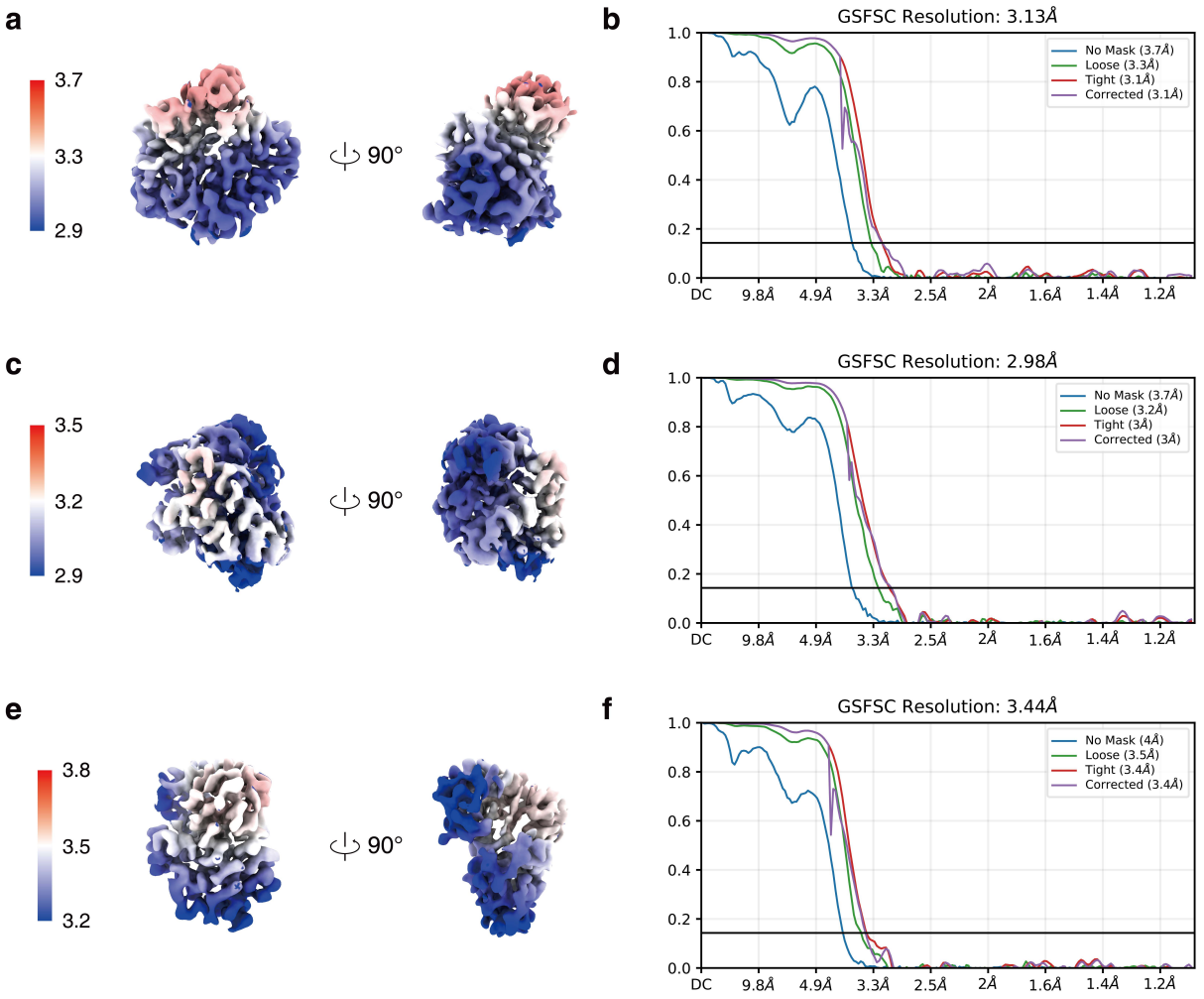



**Supplementary Fig. S17 | Construction of the fusion protein for cryo-EM analysis.** Initial spatial arrangement was performed in PyMOL by positioning the C-terminal helix of the designed protease (PP507 as a representative example) adjacent to the N-terminus of the scaffolding base protein (1,200 amino acids in length, *de novo* generated by ProteusFlow). ProteusFlow was then employed to *de novo* generate a structured linker connecting the two domains. To stabilize the overall fusion architecture and increase structural rigidity, the inter-domain amino acid interface near the junction was computationally redesigned.

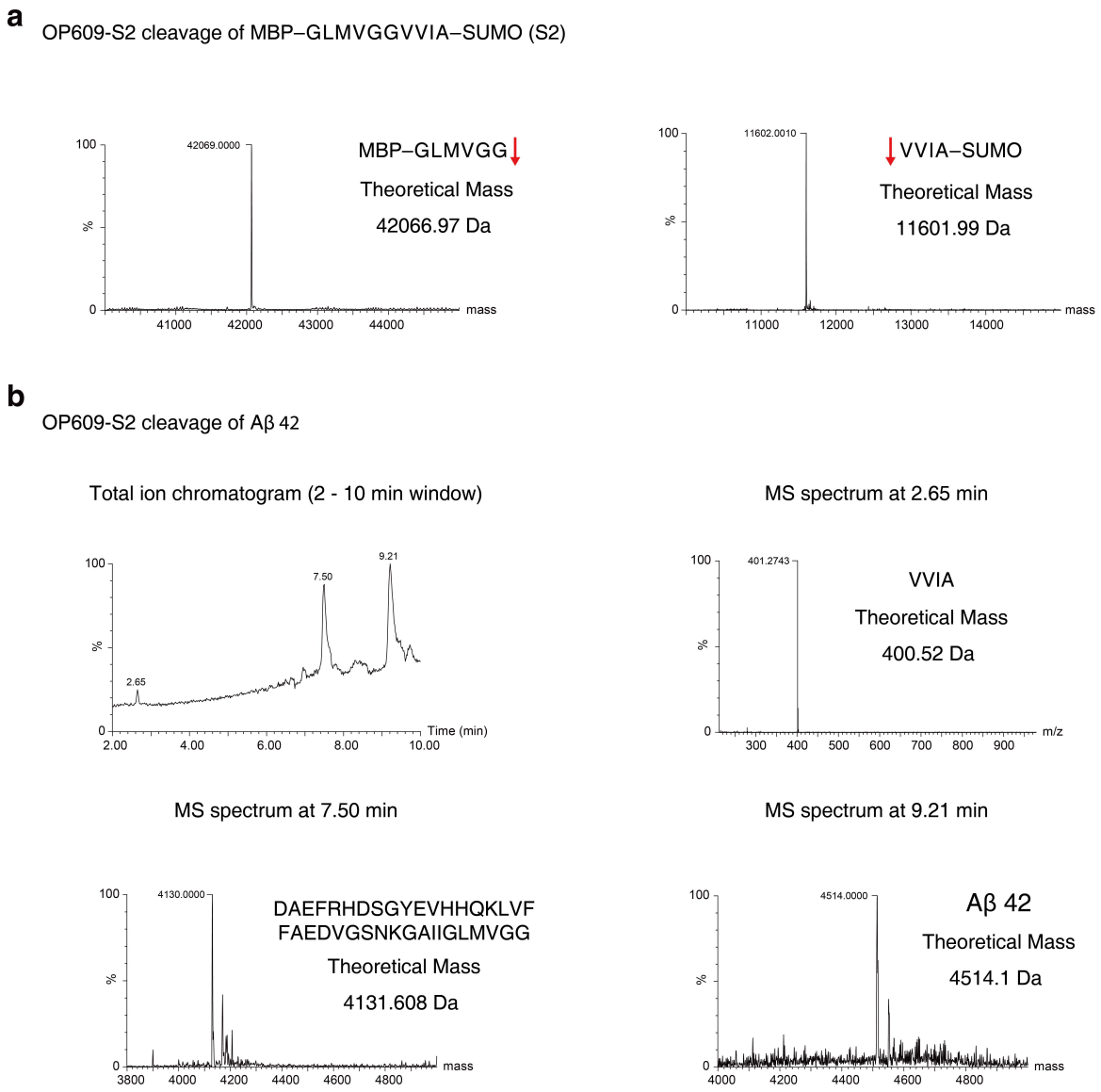



**Supplementary Fig. S18 | Cryo-EM data processing and validation of PP507 E110Q, DP622 E96Q and DP221 E85Q.** Representative cryo-EM micrographs and two-dimensional (2D) class averages for PP507 E110Q (**a**), DP622 E96Q (**c**), and DP221 E85Q (**e**). Flowcharts

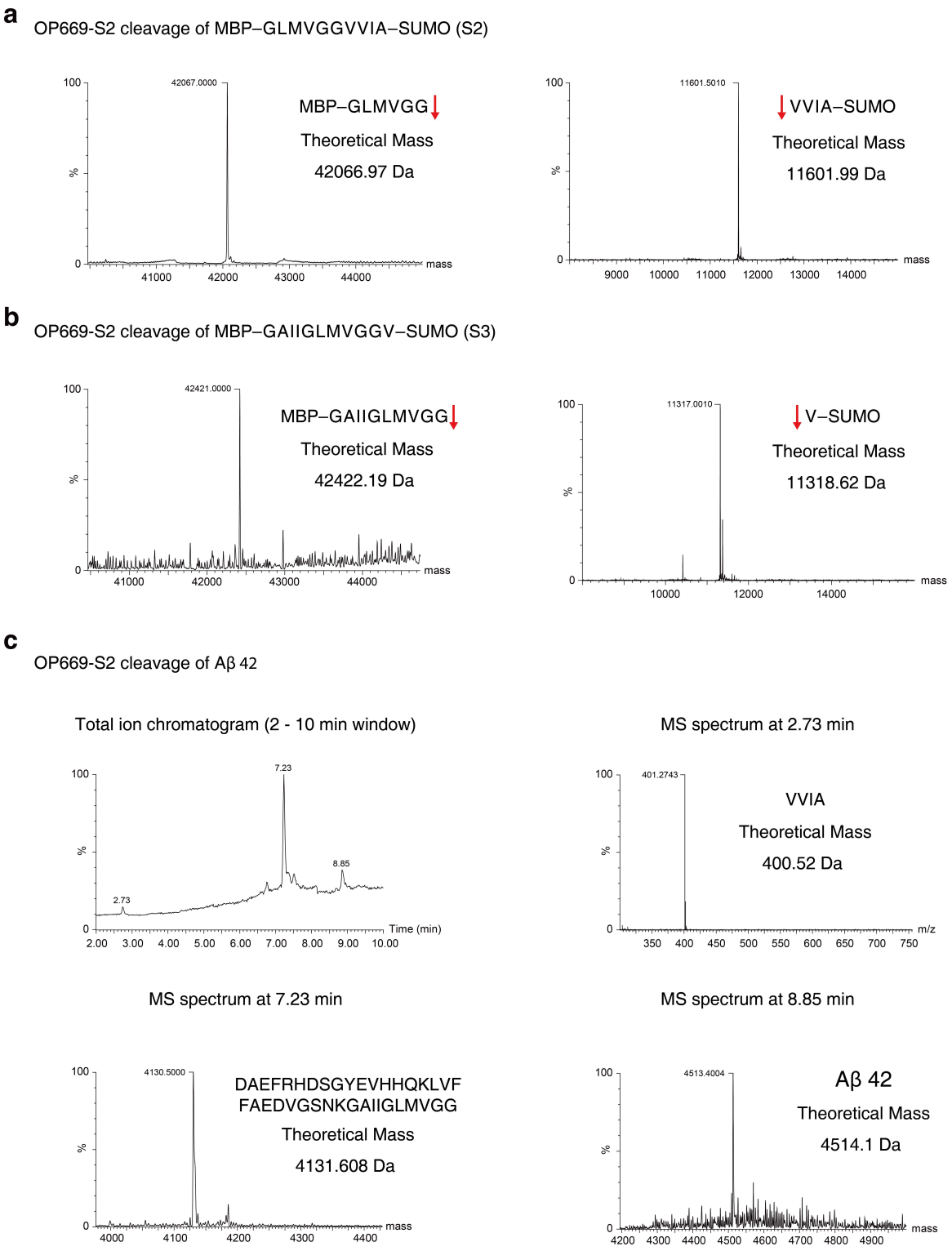
detailing the cryo-EM single-particle data processing workflows for PP507 E110Q (**b**), DP622 E96Q (**d**), and DP221 E85Q (**f**).




**Supplementary Fig. S19 | Local resolution estimation and FSC curves for PP507 E110Q, DP622 E96Q, and DP221 E85Q.** Cryo-EM density maps colored by local resolution (Å) for PP507 E110Q (**a**), DP622 E96Q (**c**), and DP221 E85Q (**e**). Gold-standard Fourier shell correlation (FSC) curves for the final 3D reconstructions of PP507 E110Q (**b**), DP622 E96Q (**d**), and DP221 E85Q (**f**), indicating the overall estimated resolutions at the 0.143 threshold.



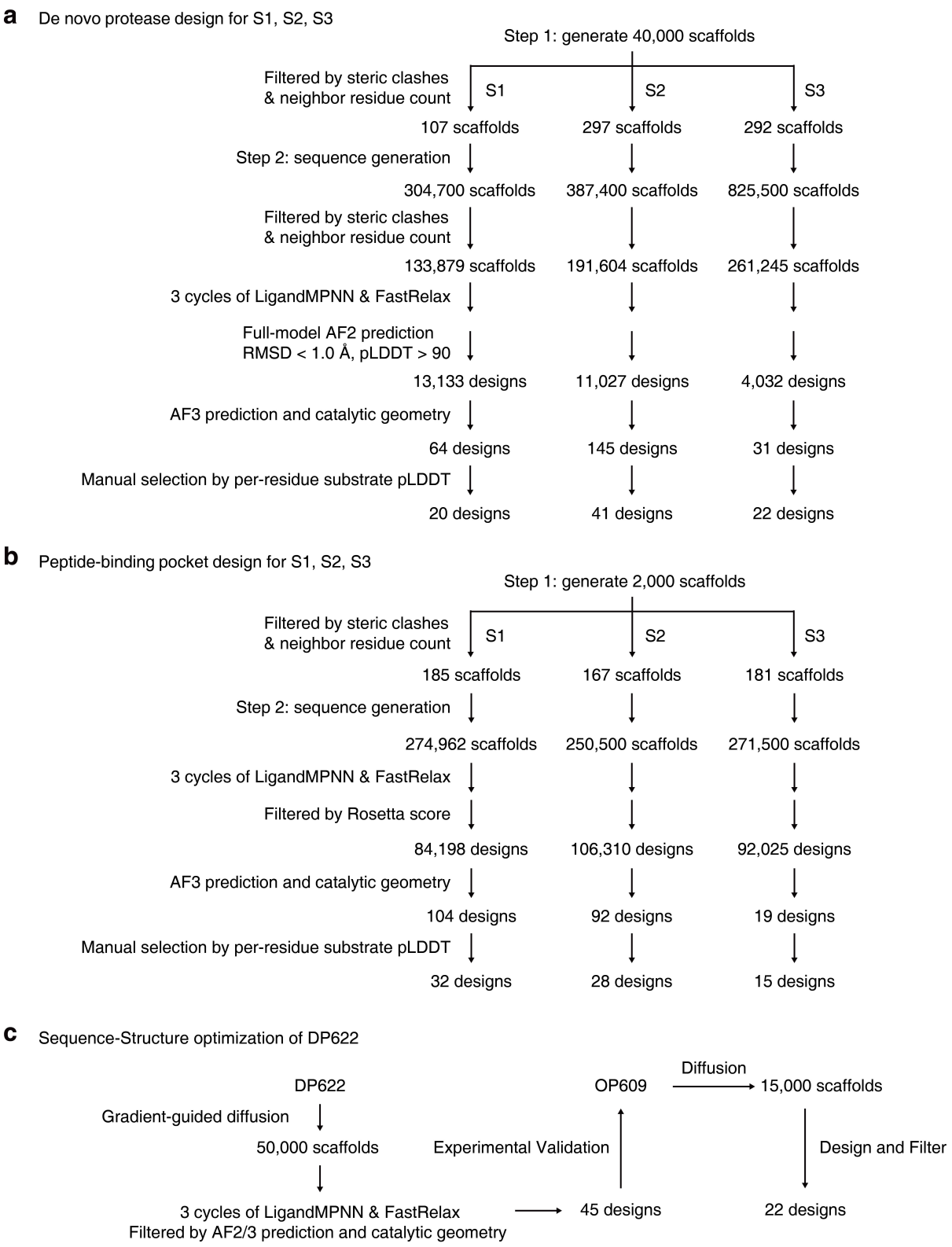

**Supplementary Fig. S20 | Mass spectrometric identification of the OP609-S2 cleavage site.** **a** OP609-S2 cleaving S2 precisely at the intended site. **b** TICs and mass spectra of Aβ42 cleavage products. Elution profiles of annotated peaks confirm precise cleavage of the full length Aβ42 peptide at the intended site by OP609-S2.




**Supplementary Fig. S21 | Mass spectrometric identification of the OP669-S2 cleavage site.**

**a, b** OP669-S2 cleaving both S2 and S3 precisely at the intended position. **c** TICs and mass

spectra of Aβ42 cleavage products. Elution profiles of annotated peaks confirm precise cleavage of the full length Aβ42 peptide at the intended site by OP669-S2.




**Supplementary Fig. S22 | Computational workflow for the *de novo* design of proteases. a** Computational pipeline for the *de novo* protease design (DP) strategy. **b** Computational

pipeline for the peptide binding pocket design (PP) strategy. **c** Flow-based structural optimization of DP622-S2.

**Supplementary Table S1 | Amino acid sequences of fusion substrates (MBP-Aβ_segment-SUMO).**

| Name | Amino Acid Sequence |
|---|---|
| MBP-Aβ-SUMO Substrate 1 | MAKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIF WAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNP PKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAG AKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTF KGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAK DPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTRITKGGGSGG GSGGGSHHQKLVFFAEDGGGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLR RLMEAFAKRQGKEMDSLTFLYDGIEIQADQTPEDLDMEDNDIIEAHREQIGG |
| MBP-Aβ-SUMO Substrate 2 | MAKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIF WAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNP PKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAG AKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTF KGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAK DPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTRITKGGGSGG GSGGGSGLMVGGVVIAGGGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLR RLMEAFAKRQGKEMDSLTFLYDGIEIQADQTPEDLDMEDNDIIEAHREQIGG |
| MBP-Aβ-SUMO Substrate 3 | MAKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIF WAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNP PKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAG AKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTF KGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAK DPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTRITKGGGSGG GSGGGSGAIIGLMVGGVGGGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLR RLMEAFAKRQGKEMDSLTFLYDGIEIQADQTPEDLDMEDNDIIEAHREQIGG |

Note: Color coding denotes protein domains: Green = MBP tag; Orange = Aβ peptide segment; Red = SUMO tag.

**Supplementary Table S2 | Success rates of metalloprotease designs by strategy and target cleavage site.**

| Design strategy | Cleavage site | Ordered designs | Soluble expressed designs | High-activity design (>80% overnight cleavage) | Success rate |
|---|---|---|---|---|---|
| PP | S1 | 32 | 30 | 2 | 6% |
| PP | S2 | 28 | 27 | 0 | 0% |
| PP | S3 | 15 | 14 | 0 | 0% |
| DP | S1 | 20 | 19 | 1 | 5% |
| DP | S2 | 41 | 35 | 1 | 2% |
| DP | S3 | 22 | 16 | 1 | 5% |

**Supplementary Table S3 | Amino acid sequences of FRET assay substrates (mTurquoise2-Aβ_segment-mVenus).**

| Name | Amino Acid Sequence |
|---|---|
| mTurquoise2-Aβ-mVenus Substrate 1 | MGHHHHHHHHHSSGLEVLFQGPGGTVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDA TYGKLTLKFICTTGKLPVPWPTLVTTLSWGVQCFARYPDHMKQHDFFKSAMPEGYVQERTIF FKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYFSDNVYITADKQKNGI KANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLE FVTAAGITLGMDELYKGGGSGGGSHHQKLVFFAEDGGGSGGGSVSKGEELFTGVVPILVELD GDVNGHKFSVSGEGEGDATYGKLTLKLICTTGKLPVPWPTLVTTLGYGLQCFARYPDHMKQ HDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLE YNYNSHNVYITADKQKNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSY QSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYK |
| mTurquoise2-Aβ-mVenus Substrate 2 | MGHHHHHHHHHSSGLEVLFQGPGGTVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDA TYGKLTLKFICTTGKLPVPWPTLVTTLSWGVQCFARYPDHMKQHDFFKSAMPEGYVQERTIF FKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYFSDNVYITADKQKNGI KANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLE FVTAAGITLGMDELYKGGGSGGGSGLMVGGVVIAGGGSGGGSVSKGEELFTGVVPILVELD GDVNGHKFSVSGEGEGDATYGKLTLKLICTTGKLPVPWPTLVTTLGYGLQCFARYPDHMKQ HDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLE YNYNSHNVYITADKQKNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSY QSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYK |
| mTurquoise2-Aβ-mVenus Substrate 3 | MGHHHHHHHHHSSGLEVLFQGPGGTVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDA TYGKLTLKFICTTGKLPVPWPTLVTTLSWGVQCFARYPDHMKQHDFFKSAMPEGYVQERTIF FKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYFSDNVYITADKQKNGI KANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLE FVTAAGITLGMDELYKGGGSGGGSGAIIGLMVGGVGGGSGGGSVSKGEELFTGVVPILVELD GDVNGHKFSVSGEGEGDATYGKLTLKLICTTGKLPVPWPTLVTTLGYGLQCFARYPDHMKQ HDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLE YNYNSHNVYITADKQKNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSY QSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYK |

| | |
|---|---|
| mTurquoise2-Aβ-mVenus extended Substrate 2 | MGHHHHHHHHSSGLEVLFQGPGGTVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDA<br>TYGKLTLKFICTTGKLPVPWPTLVTTLSWGVQCFARYPDHMKQHDFFKSAMPEGYVQERTIF<br>FKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYFSDNVYITADKQKNGI<br>KANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLE<br>FVTAAGITLGMDELYKGGGSGGGSNKGAIIGLMVGGVVIAGGGSGGGSVSKGEELFTGVVPI<br>LVELDGDVNGHKFSVSGEGEGDATYGKLTLKLICTTGKLPVPWPTLVTTLGYGLQCFARYPD<br>HMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILG<br>HKLEYNYNSHNVYITADKQKNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNH<br>YLSYQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYK |

Note: Color coding denotes protein domains: Blue = mTurquoise2; Orange = Aβ peptide; Yellow = mVenus.

**Supplementary Table S4 | Cryo-EM data collection, refinement and validation statistics.**

| | PP507 E110Q (EMD-69321; PDB 23WM) | DP622 E96Q (EMD-69322; PDB 23WN) | DP221 E85Q (EMD-69323; PDB 23WO) |
|---|---|---|---|
| **Data collection and processing** | | | |
| Magnification | 215000× | 215000× | 215000× |
| Voltage (kV) | 300 | 300 | 300 |
| Electron exposure (e–/Å$^2$) | 50 | 50 | 50 |
| Defocus range (μm) | -1.0 ~ -2.0 | -1.0 ~ -2.0 | -1.0 ~ -2.0 |
| Pixel size (Å) | 0.57 | 0.57 | 0.57 |
| Symmetry imposed | C1 | C1 | C1 |
| Initial particle images (no.) | 4,177,970 | 4,239,754 | 3,874,893 |
| Final particle images (no.) | 102,997 | 148,591 | 72,354 |
| Map resolution (Å) | 3.13 | 2.98 | 3.44 |
| FSC threshold | 0.143 | 0.143 | 0.143 |
| **Refinement** | | | |
| Initial model used | In silico | In silico | In silico |
| Model resolution (Å) | 3.3 | 3.3 | 3.8 |
| FSC threshold | 0.5 | 0.5 | 0.5 |
| Map sharpening *B* factor (Å$^2$) | -90 | -90 | -90 |
| Model composition | | | |
| Non-hydrogen atoms | 2116 | 1783 | 1666 |
| Protein residues | 265 | 233 | 221 |
| Ligands | ZN: 1 | ZN: 1 | ZN: 1 |
| *B* factors (Å$^2$) | | | |
| Protein | 71.82 | 46.89 | 23.12 |
| Ligand | 115.39 | 65.99 | 43.83 |
| R.m.s. deviations | | | |
| Bond lengths (Å) | 0.003 | 0.006 | 0.005 |
| Bond angles (°) | 0.583 | 1.118 | 0.931 |
| Validation | | | |
| MolProbity score | 1.82 | 2.46 | 2.36 |
| Clashscore | 5.59 | 12.52 | 14.54 |
| Poor rotamers (%) | 0.45 | 2.27 | 1.21 |
| Ramachandran plot | | | |
| Favored (%) | 91.19 | 88.65 | 86.64 |
| Allowed (%) | 8.05 | 9.61 | 9.68 |
| Disallowed (%) | 0.77 | 1.75 | 3.69 |