SUPPLEMENTARY INFORMATION

NMR structure of emfourin, a novel protein metalloprotease inhibitor: insights into the mechanism of action

Timur N. Bozin[1,2,3,*], Igor M. Berdyshev[1,*], Ksenia N. Chukhontseva[1], Maria A. Karaseva[1], Petr V. Konarev[4], Anna M. Varizhuk[5], Dmitry M. Lesovoy[3], Alexander S. Arseniev[3], Sergey V. Kostrov[1], Eduard V. Bocharov[3,5], and Ilya V. Demidyuk[1,#]

Affiliations

[1] Institute of Molecular Genetics of National Research Centre "Kurchatov Institute," Moscow, Russia;
[2] National Research Centre "Kurchatov Institute," Moscow, Russia;
[3] Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia;
[4] Shubnikov Institute of Crystallography of the Federal Scientific Research Centre "Crystallography and Photonics," Russian Academy of Sciences, Moscow, Russia;
[5] Moscow Institute of Physics and Technology (State University), Dolgoprudny, Russia.

* These authors contributed equally and share first authorship.

[#] Corresponding author. Tel: +7 499 196 18 53; E-mail: duk@img.ras.ru

SUPPLEMENTARY METHODS

Details of NMR structure calculations

The spatial structure of M4in was calculated using CYANA software. Upper inter-proton distance constraints were derived via standard CYANA calibration procedure from nuclear Overhauser effect (NOE) cross-peak intensities that was analyzed in three different 3D ^{15}N- and ^{13}C-edited NMR spectra: ^{1}H-^{15}N-NOESY-HSQC, ^{1}H-^{13}C-NOESY-HSQC for aliphatic and aromatic spectral regions (with the mixing time of 100 ms in all three cases). All pre-proline peptide bonds were clearly identified as being in trans-configuration based on characteristic NOE signals. Stereospecific assignments were obtained by the analysis of local conformations in CYANA as well as J-coupling constant analysis from the 3D ^{1}H-^{15}N-HNHB spectrum.

The constraints on the angles φ and ψ of the main chain were obtained from the analysis of C^{α}, C^{β}, C', N^{H}, H^{N}, and H^{α} chemical shifts in the TALOS-N program. The constraints on the χ^1 angles of the side chains were obtained from the joint analysis of the TALOS-N data (prediction of the distribution of conformers) and spectral data of two 2D spin-echo difference ^{1}H-^{13}C-CT-HSQC experiments (2D ^{13}C-{^{15}N} and 2D ^{13}C-{^{13}CO}) to measure $^{3}J_{NC}{}^{\gamma}$ and $^{3}J_{C'C}{}^{\gamma}$, respectively. The constraints on the χ^2 angle of Leu7 were established by the analysis of NOESY cross-peaks and 2D ^{13}C-{^{13}CA} spin-echo difference ^{1}H-^{13}C-CT-HSQC experiment to measure $^{3}J_{C}{}^{\alpha}{}_{C}{}^{\delta}$. For the formation of a staggered conformation with two adjacent sp^3-hybridized atoms, additional restrictions were imposed on the torsion angles: they were taken equal to 60º, -60º, or 180º (within the range of permissible deviations ±30º). For certain residues located in the flexible areas, where the angles φ and ψ were not unambiguously determined and, after preliminary calculations, fell into clearly energetically unfavorable areas on the Ramachandran map, the CYANA macro ramaaco was applied to correct for the introduction of additional restrictions on the angles φ and ψ.

The donors of hydrogen bonds in the main chain were identified according to the values of the amide proton chemical shift temperature coefficients ($\Delta\delta^{1}H^{N}/\Delta T$, ppb/K), which were determined from ^{1}H,^{15}N-HSQC spectra at various temperatures (298, 300, 303, 305, and 308 K). The ^{1}H-^{15}N-CLEANEX spectrum, in which the signal intensity is directly proportional to the proton chemical exchange of a given amide group with the solvent and therefore is inversely proportional to the incidence of a proton in a hydrogen bond, was used for the same purpose. When calculating the structure, the protons of amide groups whose absolute temperature coefficients did not exceed 5 ppb/K and/or produced no or weak signal in the CLEANEX spectrum were considered as donor candidates (see Fig S1A and B for details). Some side-chain donors were employed to explain the slowdown in the chemical exchange ($H^{\gamma 1}$ of Thr11, H^{γ} of Ser87, and $H^{\gamma 1}$ of Thr53). The corresponding hydrogen bond acceptors were determined based on the preliminary structure calculations. The corresponding hydrogen bond restraints were employed in subsequent calculations for d(O,N), and d(O,H^{N}) distances in accordance with the hydrogen bond distance criteria in CYANA.

Back calculation of NOESY spectra peak list was used to identify missing signals in the spectra, which made it possible to define the lower distance limits. At the final stages of structure calculation, the interproton lower distance restraint of 4 Å was used and then the standard CYANA simulated annealing protocol was applied to calculate 100 random structures using the angle and distance restraints, and 20 resulting NMR structures of M4in with the lowest values of target function were selected.

Expression plasmids for M4in mutant genes

To construct pD70A, pD70S, pD70N, and pD70E, fragments of the M4in gene were PCR amplified from pM4in-His/S plasmid using the D70A-mf, D70S-mf, D70N-mf, and D70E-mf primers, respectively, with the R-1 primer (sequences of all used oligonucleotides are shown in Table S3). In the case of pD70A, the PCR product was purified, digested with *Xma*I (Sibenzyme, Russia) and *Xho*I (New England Biolabs, USA), and ligated into pM4in-His/S digested with the same enzymes. To construct pD70N, pD70S, and pD70E, the PCR products were purified, digested with *Xho*I, and ligated into pM4in-His/S digested with *Sma*I (Sibenzyme, Russia) and *Xho*I.

A two-step PCR amplification was used to construct pF21S. At the first step, a fragment of the M4in gene was amplified from pM4in-His/S using the F21S-mf/R-1 primer pair. At the second step, the resulting PCR product was used as a template for amplification of the full-length gene with the F21S-f/R-1 primers. The PCR product was purified, digested with *Fau*NDI (Sibenzyme, Russia) and *Xho*I, and cloned into pM4in-His/S digested with the same enzymes.

Overlap extension PCR was used to construct pA22S, pF23S, pF21S-F23S, pF21S-A22S, pA22S-F23S, pF21S-A22S-F23S, and pD70A-F21S-A22S-F23S. pM4in-His/S was used as the amplification template and the cloning, except for pD70A-F21S-A22S-F23S, where pF21S-A22S-F23S served as the template and the vector. At the first step, two overlapping modified fragments of the M4in gene were PCR amplified using the A22S-mf/R-2 and F-1/A22S-mr, F23S-mf/R-2 and F-2/F23S-mr, F21S-F23S-mf/R-2 and F-2/F21S-F23S-mr, F21S-A22S-mf/R-3 and F-2/F21S-A22S-mr, A22S/F23S-mf/R-3 and F-2/A22S-F23S-mr, F21S-A22S-F23S-mf/R-3 and F-2/F21S-A22S-F23S-mr, and D70A-F21S-A22S-F23S-mf/R-2 and F-1/D70A-F21S-Ala22-F23S-mr primer pairs, respectively. In the second step, the fragments obtained were used as primers and templates in the overlap extension followed by the amplification of the full-length sequence with pairs of primers F-1/R-2, F-2/R-2, and F-2/R-3, respectively. The resulting products were purified, digested with *Fau*NDI and *Xho*I, and cloned into pM4in-His/S or pF21S-A22S-F23S digested with the same enzymes. All cloned fragments were confirmed by sequencing.

SUPPLEMENTARY RESULTS

Oligomeric state and overall parameters of protealysin (PLN) and emfourin (M4in) in solution

The information about the oligomeric state of PLN and M4in was obtained from the SAXS experiments. The processed scattering data and the computed distance distribution functions are displayed in Fig 2. The overall parameters extracted from the SAXS data are summarized in Table S1.

The excluded volume V_p of the PLN particle (53 ± 4) nm^3 suggests its monomeric state in solution at 0.63-10.0 mg/ml, in agreement with the empirical finding for globular proteins that the hydrated volume in nm^3 should numerically be about 1.5-1.7 times larger than the molecular mass in kDa (1) (the calculated molecular mass of the monomer is 32.7 kDa). The experimental R_g and D_{max} (1.96 ± 0.04 nm and 6.1 ± 0.2 nm, respectively) point to a rather compact structure. The bell-shaped distance distribution function $p(r)$ for PLN (Fig 2B) is also consistent with the compact shape of the protein.

The experimental V_p of M4in (19 ± 3) nm^3 indicates that this protein is also monomeric in solution at 1.0-16.0 mg/ml (the calculated molecular mass of the monomer is 12.7 kDa). In contrast to PLN, R_g and D_{max} values (1.68 ± 0.03 nm and 5.5 ± 0.2 nm, respectively) point to an elongated shape of the molecule and its $p(r)$ function (Fig 2B) displays an asymmetric tail typical for elongated particles.

SUPPLEMENTARY FIGURES

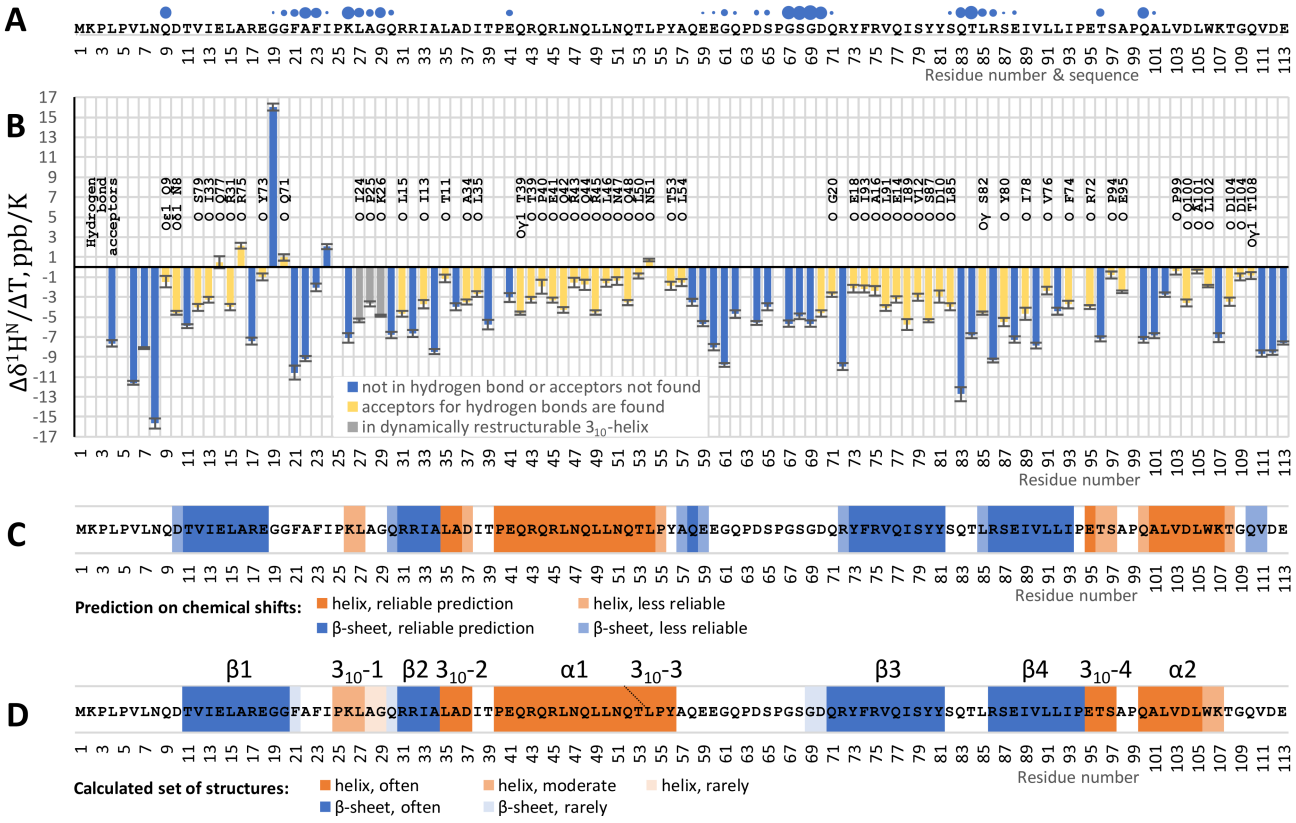

Figure S1. Hydrogen bonds and secondary structure of M4in.

A. CLEANEX data on chemical exchange of amide protons with water molecules. The area of circles is proportional to the volume of appropriate CLEANEX peaks, and those in turn increase with the chemical exchange rate.

B. Chemical shift temperature coefficients ($\Delta\delta^1H^N/\Delta T$, ppb/K) of amide protons. The amide protons where acceptors for hydrogen bonds were found (signed in the figure) are indicated yellow, the amide protons in dynamically restructurable 3_{10}-helix are indicated grey. Additionally, four hydrogen bonds with sidechain donors were introduced. These are T11-$OH^{\gamma1}$···O=V6, T53-$OH^{\gamma1}$···O=L49, S87- OH^{γ}···O=R86, W106-$NH^{\varepsilon1}$···O=G29.

C. Chemical shifts based secondary structure prediction made by three different programs (2) and averaged over probability (reliable prediction means that it was made by at least 2 out of 3 programs, less reliable means that it was made by a single program).

D. The frequency of occurrence of residues in a certain secondary structure in the calculated set. Often is 19-20 times; moderate, 10-13 times; rare, 1-2 times; no other frequencies were observed.

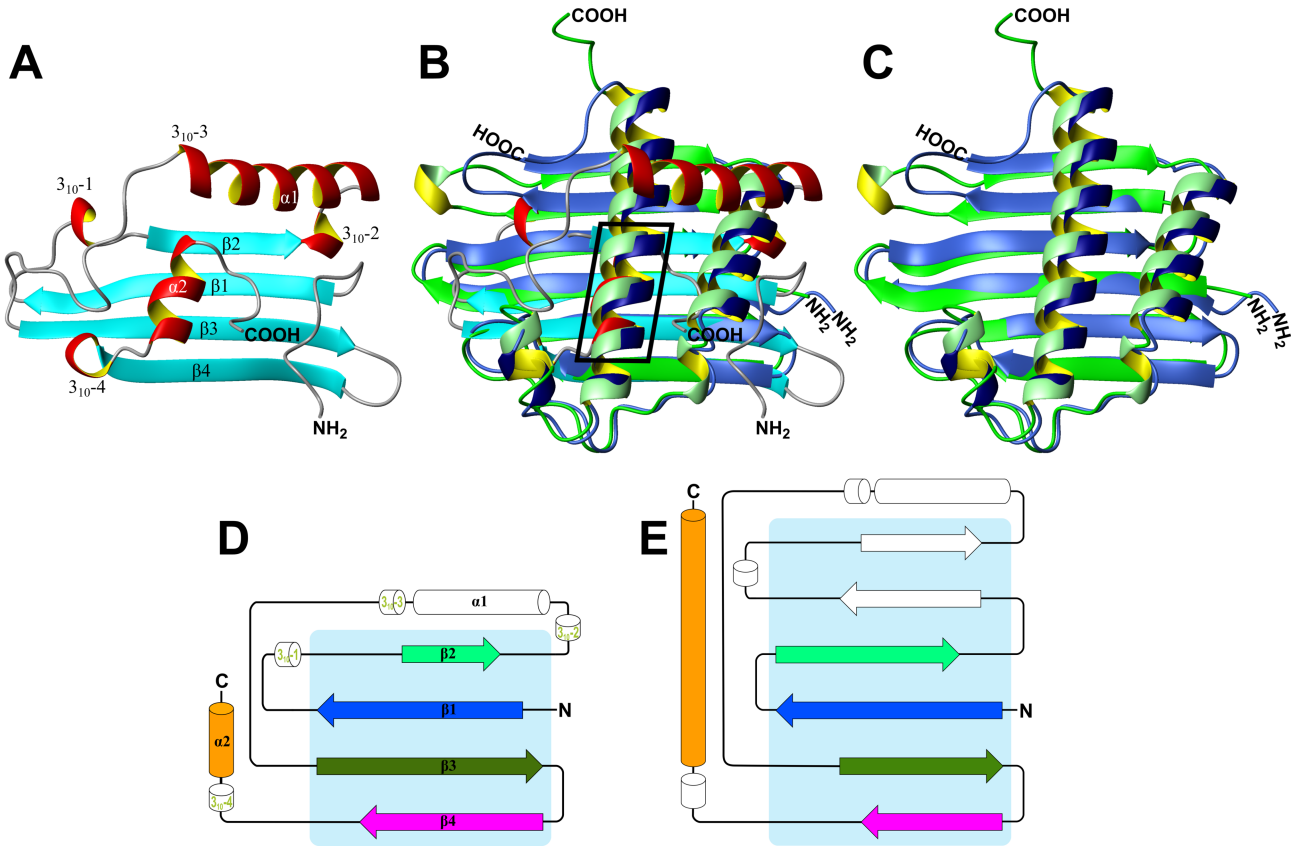

Figure S2. Comparison of 3D structures of emfourin and proteins of mago nashi family (http://pfam.xfam.org/family/PF02792).

A. Emfourin (PDB ID 6ZYG, model 1).

B. Superposition of emfourin (cyan & red), *Drosophila melanogaster* mago nashi protein (PDB ID 2X1G, chain B, blue and navy) and the human mago nashi protein homolog (PDB ID 2HYI, chain G, green and pale green). The emfourin α2-helix coinciding with one of mago nashi α-helices is enclosed in a quadrangle.

C. Superposition of *Drosophila melanogaster* mago nashi protein (blue and navy) and human mago nashi protein homolog (green and pale green).

Topology diagrams of emfourin (D) and mago nashi protein homolog (E). The β-sheet is marked with a blue rectangle.

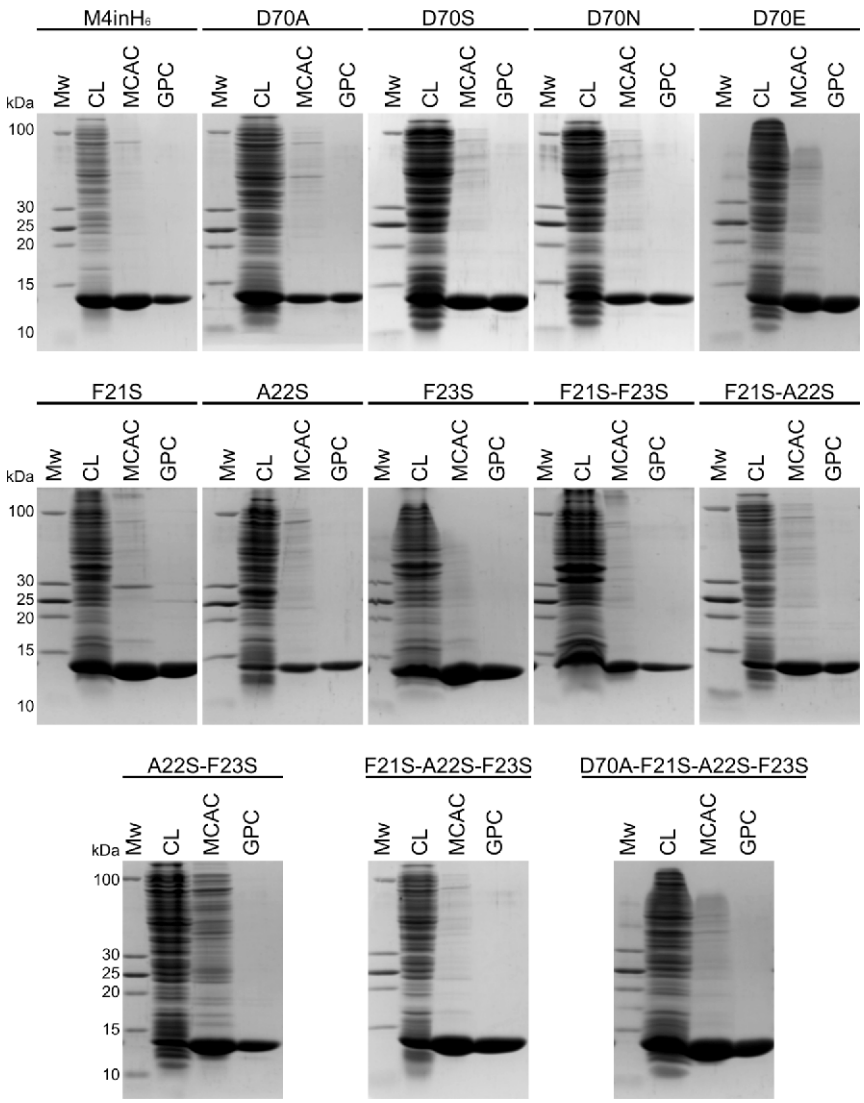

Figure S3. Purification of recombinant M4in variants: SDS-PAGE analysis.

Protein electrophoresis (SDS-PAGE) was performed in vertical 15% polyacrylamide gels with 0.1% sodium dodecylsulfate after Laemmli. Proteins were visualized by Coomassie brilliant blue staining. PageRuler Unstained Low Range Protein Ladder (ThermoFisher Scientific, USA) was used as molecular weight standards. Mw, molecular weight standards; CL, supernatant of cell lysate; MCAC, sample after metal-chelate affinity chromatography; GPC, sample after gel permeation chromatography. Protein designations correspond to those in Table 2.

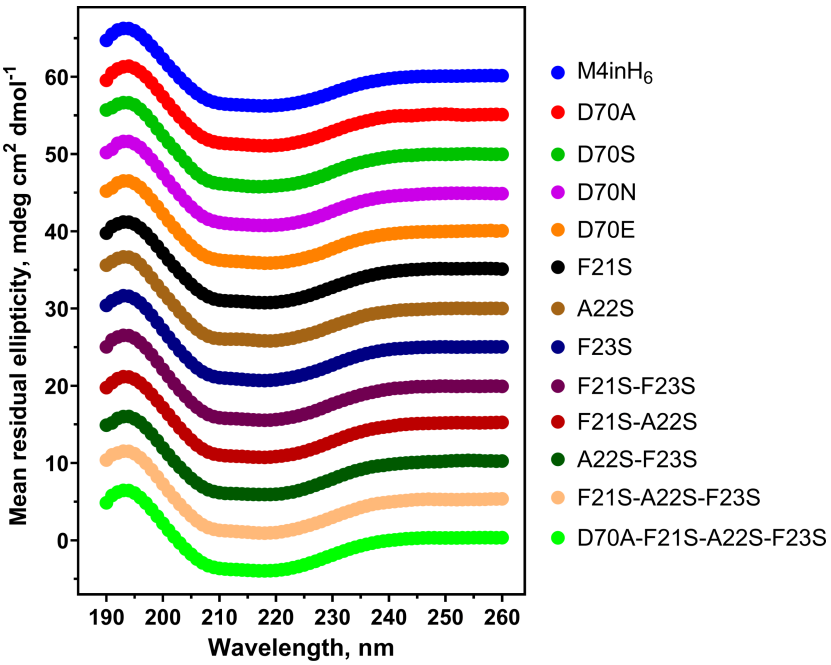

Figure S4. Comparison of far-UV CD spectra for M4in variants.

The curves are displaced up by 5 mdeg·cm^2·dmol^{-1} for clarity. The designations correspond to those in Table 2.

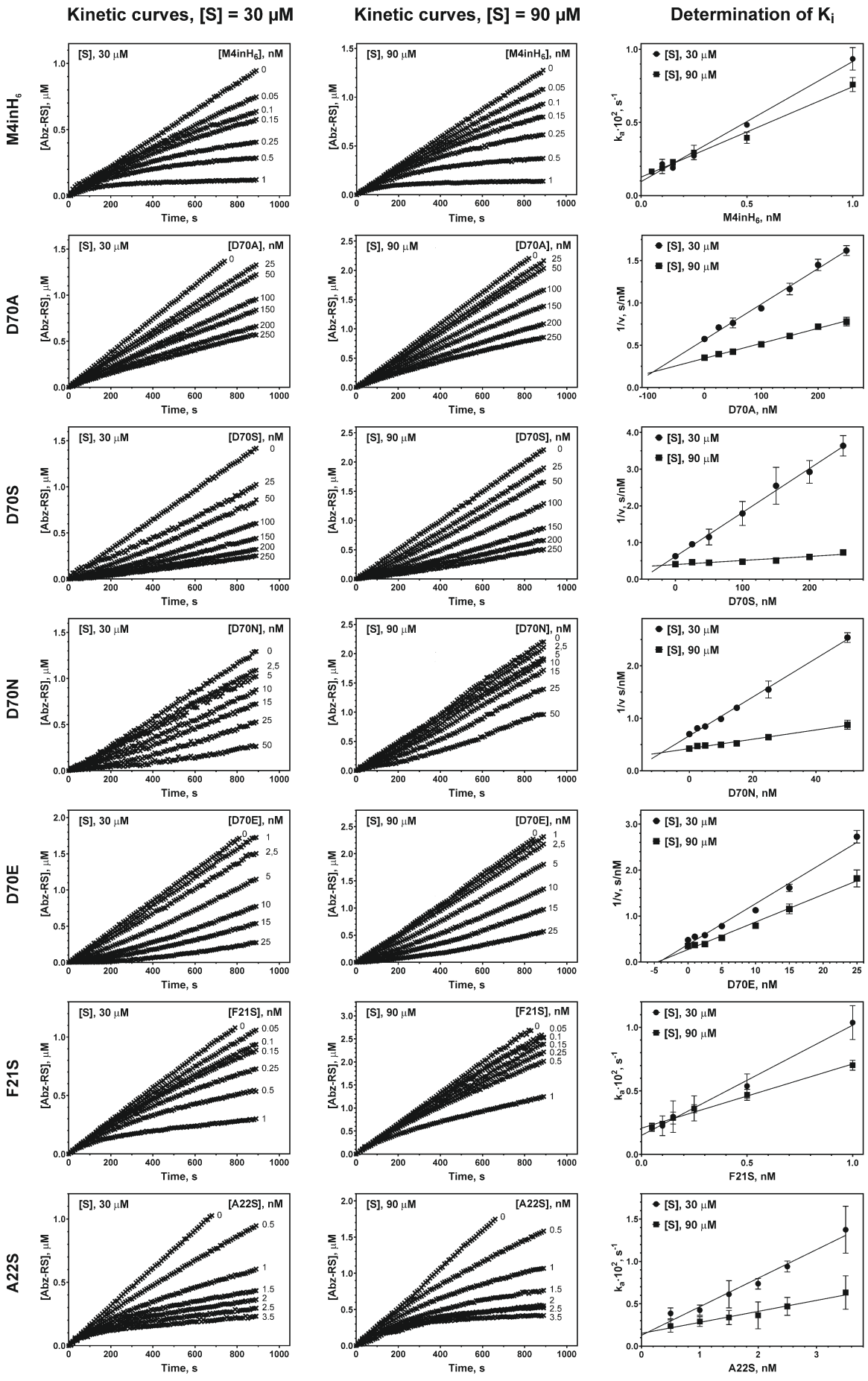

Figure S5. (Legend on next page)

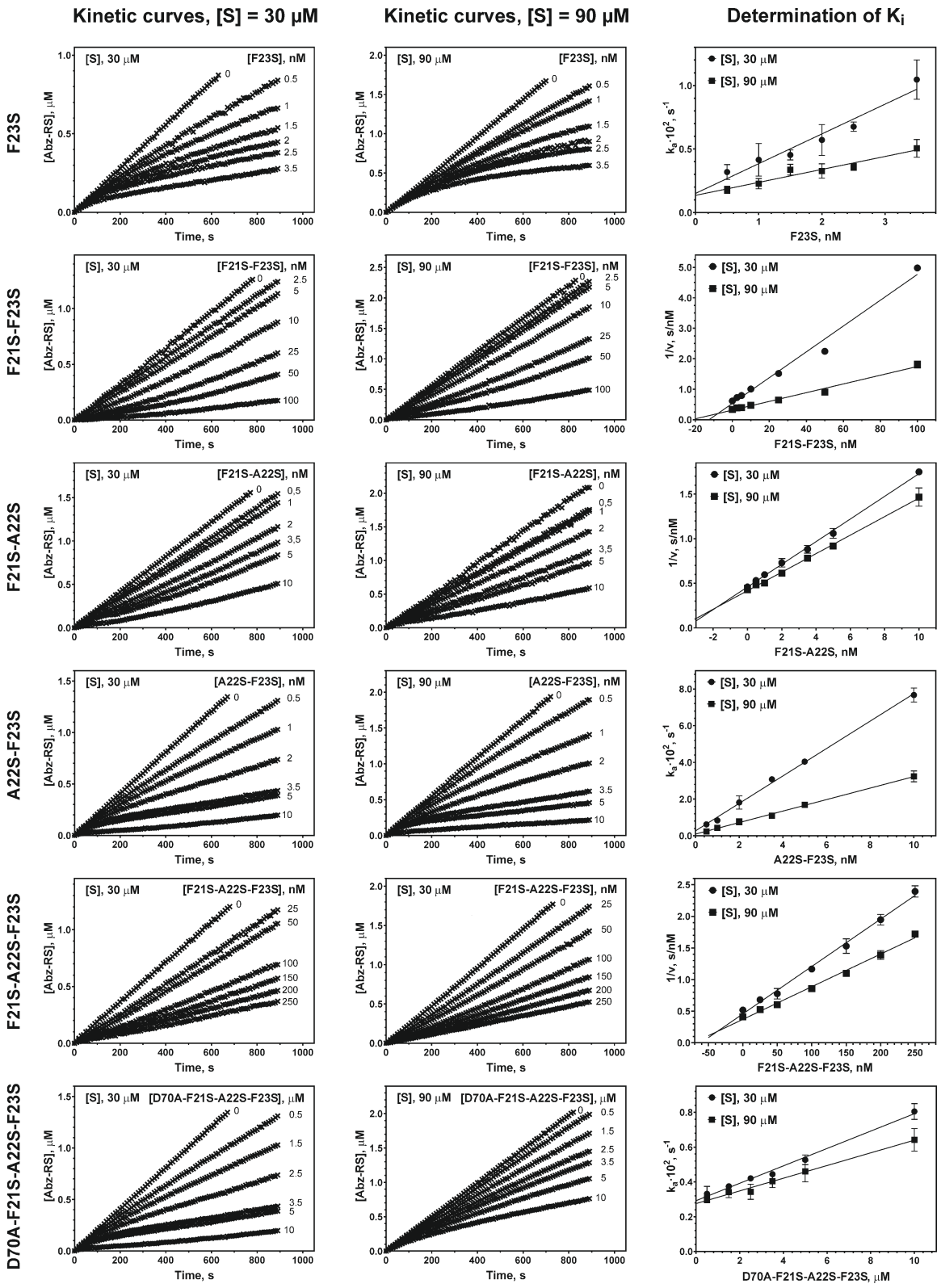

Figure S5. Hydrolysis of Abz-RSVIK(Dnp) by protealysin in presence of emfourin variants and determination of inhibition constant.

Representative sets of kinetic curves obtained for different inhibitor concentrations and two substrate concentrations 30 μM (left column) and 90 μM (middle column) are shown. The inhibition constants were calculated by the graphical method using the Morrison plot for slow-binding inhibition ($M4inH_6$, F21S, A22S, F23S, A22S-F23S, and D70A-F21S-A22S-F23S) and Dixon plot for classical inhibition (D70A, D70S, D70N, D70E, F21S-F23S, F21S-A22S, and F21S-A22S-F23S) (right column); the coefficient of determination (R^2) was above 0.98 in all cases. Values are represented as mean ± SD of three independent experiments. Protein designations correspond to those in Table 2.

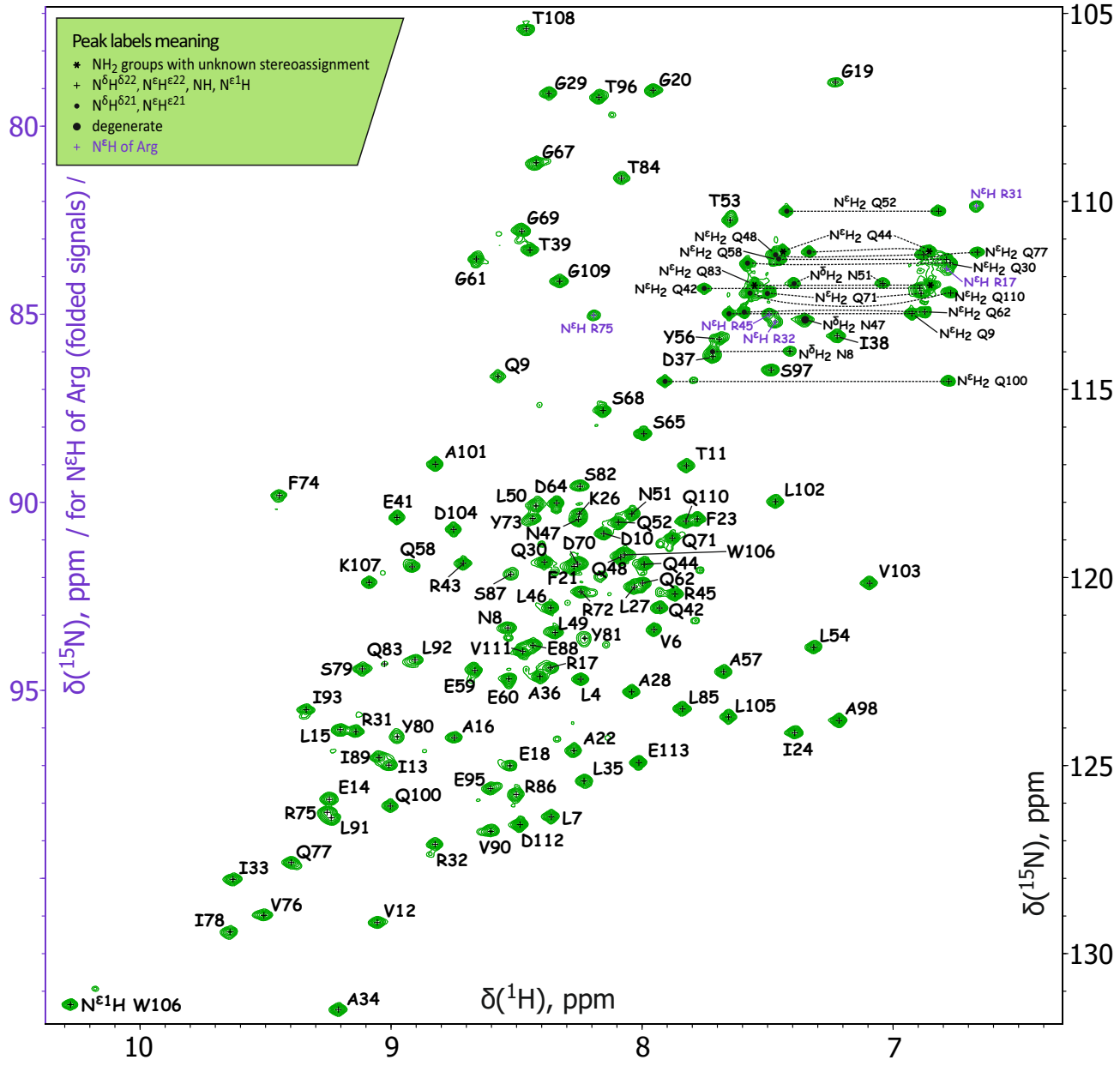

Figure S6. 2D ^{1}H-^{15}N-HSQC NMR spectrum of M4in.

The spectrum of the ^{13}C-^{15}N-labeled M4in sample (0.37 mM, with 0.05% NaN_3, 18 mM phosphate (Na_2HPO_4/NaH_2PO_4) in 10% D_2O/90% H_2O, pH 6.5, ionic strength ~35 mM) was acquired on a 600 MHz NMR spectrometer at 303 K. The cross-peak assignments are labeled. The conformational exchange led to the appearance of additional low-intensity peaks in the spectrum, which are not assigned.

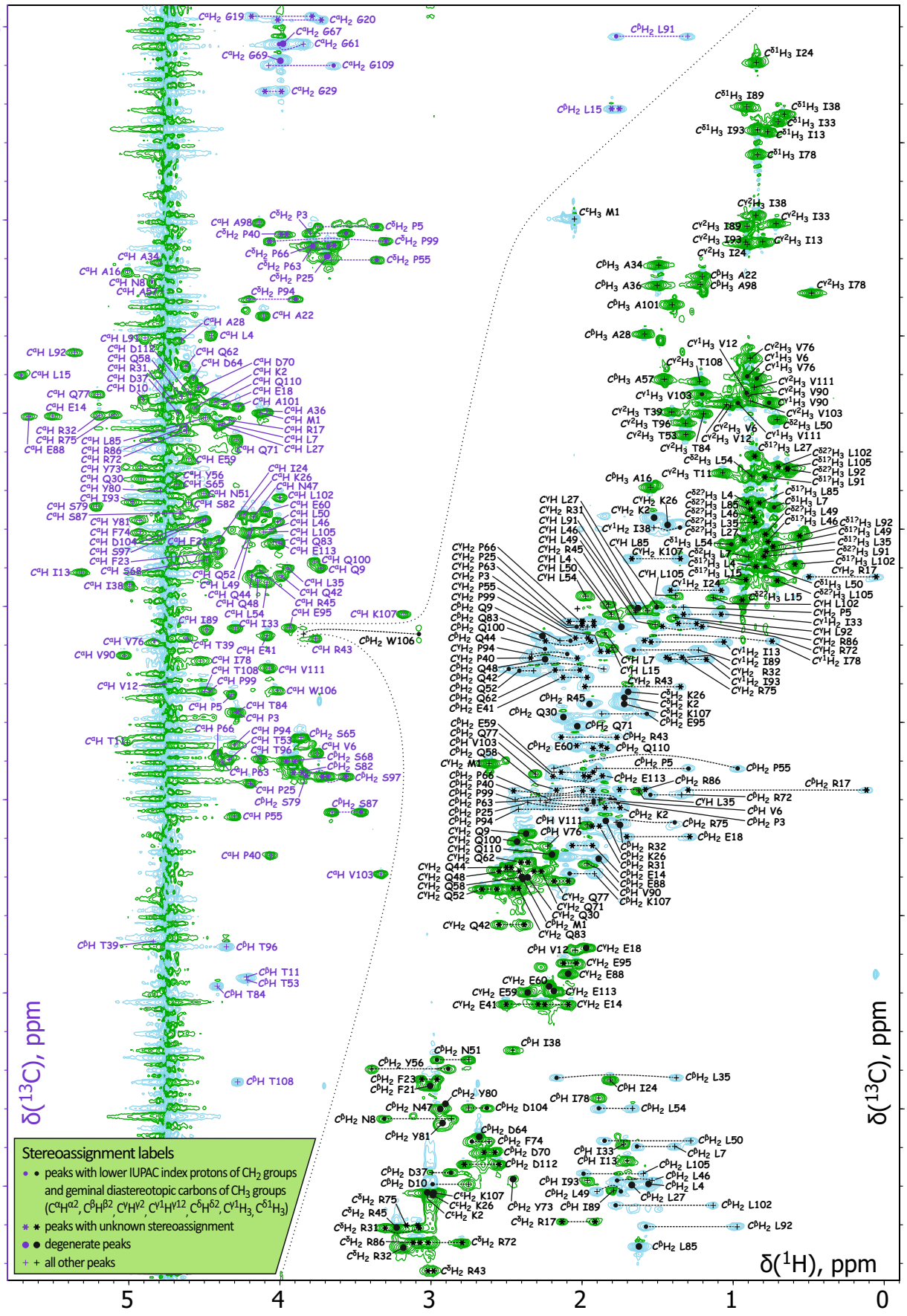

Figure S7. 2D ^{1}H-^{13}C-HSQC NMR spectrum of M4in.

The spectrum of the ^{13}C-^{15}N-labeled M4in sample (see description under Fig S6) was acquired on a 600 MHz NMR spectrometer at 303 K. Only major conformation peak assignments are labeled.

SUPPLEMENTARY TABLES

Table S1. Overall parameters calculated from SAXS data analysis.

Sample	Concentration (mg / mL)	R_g (nm)	D_{max} (nm)	V_p (nm^3)	MW (kDa)	χ^2_{ab}	χ^2_{pdb}
PLN-M4in complex (1:1 stoichiometry)	3.0, 6.0, 12.0	2.35±0.05	7.6±0.2	71±5	45.4	0.97	1.14
PLN protein	0.63, 1.25, 2.5, 5.0, 10.0	1.96±0.04	6.1±0.2	53±4	32.7	0.99	1.15
M4in protein	1.0, 2.0, 4.0, 8.0, 16.0	1.68±0.03	5.5±0.2	19±3	12.7	1.16	1.20

R_g, radius of gyration; D_{max}, maximum size of the particle; V_p, excluded volume of the hydrated particle; *MW*, molecular weight of the particle; χ^2_{ab} *and* χ^2_{pdb} discrepancies (chi-square values) for the fits from *ab initio* DAMMIN models and atomic PDB models calculated by CRYSOL (the crystallographic model of PLN (PDB ID: 2VQX), NMR model of M4in (PDB ID: 6ZYG), and the rigid body model of PLN-M4in complex obtained by ROSIE.

Table S2. Purification of emfourin variants.

Emfourin variant[a]	Purification stage[b]	Total protein (mg)	Amount of protein of interest[c] (%)	Yield[d] (%)	Purification rate
M4inH$_6$	CL	108	39	100	1.0
	MCAC	28	87	58	2.2
	GPC	4	99	9	2.5
D70A	CL	162	30	100	1.0
	MCAC	28	78	45	2.6
	GPC	4	99	8	3.3
D70S	CL	166	38	100	1.0
	MCAC	41	77	50	3.6
	GPC	4	99	6	4.5
D70N	CL	311	27	100	1.0
	MCAC	47	83	47	3.1
	GPC	14	99	17	3.7
D70E	CL	166	38	100	1.0
	MCAC	41	77	50	2.0
	GPC	7	99	11	2.6
F21S	CL	122	34	100	1.0
	MCAC	18	88	38	2.6
	GPC	3	99	7	2.9
A22S	CL	168	14	100	1.0
	MCAC	20	67	57	4.8
	GPC	2	99	8	7.1
F23S	CL	196	34	100	1.0
	MCAC	43	85	55	2.5
	GPC	4	99	6	2.9
F21S-F23S	CL	240	30	100	1.0
	MCAC	46	88	56	2.9
	GPC	8	99	11	3.3
F21S-A22S	CL	122	11	100	1.0
	MCAC	12	67	62	6.3
	GPC	2	99	15	9.3
A22S-F23S	CL	187	12	100	1.0
	MCAC	11	71	35	6.0
	GPC	2	99	9	8.4
F21S-A22S-F23S	CL	214	23	100	1.0
	MCAC	27	75	41	3.3
	GPC	4	99	8	4.3
D70A-F21S-A22S-F23S	CL	241	35	100	1.0
	MCAC	46	72	53	2.1
	GPC	4	99	7	2.8

[a] Protein designations correspond to Table 2.
[b] CL, cell lysate supernatant; MCAC, metal-chelate affinity chromatography; GPC, gel permeation chromatography.
[c] Target protein amount was evaluated densitometrically from SDS-PAGE gels stained with Coomassie brilliant blue R-250.
[d] Yield was calculated from the protein of interest.

Table S3. Primers used.

Primer	Sequence*
F-1	GTAATACGACTCACTATAGGG
F-2	TAATACGACTCACTATAGGG
R-1	TATGCTAGTTATTGCTCAG
R-2	ATGCTAGTTATTGCTCAGCGG
R-3	TATGCTAGTTATTGCTCAGCGG
D70A-mf	CTCA**CCCGGG**AGCGGTGCCCAACGCTA
D70S-mf	GGGAGCGGTAGCCAACGCTA
D70N-mf	GGGAGCGGTAACCAACGCTA
D70E-mf	GGGAGCGGTGAACAACGCTA
F21S-mf	*CAATCAGGATACGGTCAT*TGAGCTGGCGCG TGAAGGCGGTAGCGCCTT
F21S-f	TACA**CATATG**AAACCTTTGCCGGTGCT*CAA TCAGGATACGGTCAT*
A22S-mf	*AGGCGGTTTCTCCTTTATTCCC*A
A22S-mr	*GGAATAAAGGAGAAACCGCCT*TC
F23S-mf	*GGTTTCGCCTCTATTCC*CAAACTG
F23S-mr	*GGAATAGAGGCGAAACC*GCCTTCAC
F21S-A22S-mf	*GGTTCCTCCTTTATTCCCAAAC*TG
F21S-A22S-mr	*GTTTGGGAATAAAGGAGGAACC*GCCTTCAC
A22S-F23S-mf	*GGTTTCTCCTCTATTCC*CAAACTG
A22S-F23S-mr	*GGAATAGAGGAGAAACC*GCCTTCAC
F21S-F23S-mf	*GGTAGCGCCTCTATTCC*CAAACT
F21S-F23S-mr	*GGAATAGAGGCGCTACC*GCCTTCAC
F21S-A22S-F23S-mf	*GGTTCCTCCTCTATTCC*CAAACTG
F21S-A22S-F23S-mr	*GGAATAGAGGAGGAACC*GCCTTCAC
D70A-F21S-A22S-F23S-mf	*GGAGCGGTGCCCAACGCTA*
D70A-F21S-A22S-F23S-mr	*TAGCGTTGGGCACCGCTCC*

* Restriction sites are boldfaced, sequences corresponding to mutations are underlined, and overlapping sequences are italicized. All primers were from Evrogen (Russia).

Supplementary references

1. Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D., Konarev, P. V., and Svergun, D. I. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J Appl Crystallogr* **45**, 342-350.
2. Bozin, T. N., Chukhontseva, K. N., Lesovoy, D. M., Filatov, V. V., Kozlovskiy, V. I., Demidyuk, I. V., and Bocharov, E. V. (2021) NMR assignments and secondary structure distribution of emfourin, a novel proteinaceous protease inhibitor. *Biomol NMR Assign* **15**, 361-366.