Supporting Information

for the article

Reemerging aspartic protease targets: Examining *C. neoformans* Major aspartyl peptidase 1 as a target for antifungal drug discovery

Robin Kryštůfek[1], Pavel Šácha[1], Jana Starková[1], Jiří Brynda[1,2], Martin Hradilek[1], Eva Tloušťová[1], Justyna Grzymska[3], Wioletta Rut[3], Michael J. Boucher[4], Marcin Drąg[3], Pavel Majer[1], Miroslav Hájek[1], Pavlína Řezáčová[1,2], Hiten D. Madhani[4,5], Charles S. Craik[6], Jan Konvalinka[1,]*

[1] Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Flemingovo n. 2, 16610, Prague 6, Czech Republic

[2] Institute of Molecular Genetics of the Czech Academy of Sciences, Vídeňská 1083, Prague 4, 14220, Czech Republic

[3]Department of Chemical Biology and Bioimaging, Wroclaw University of Science and Technology, Wybrzeze Wyspianskiego 27, 50-370, Wroclaw, Poland

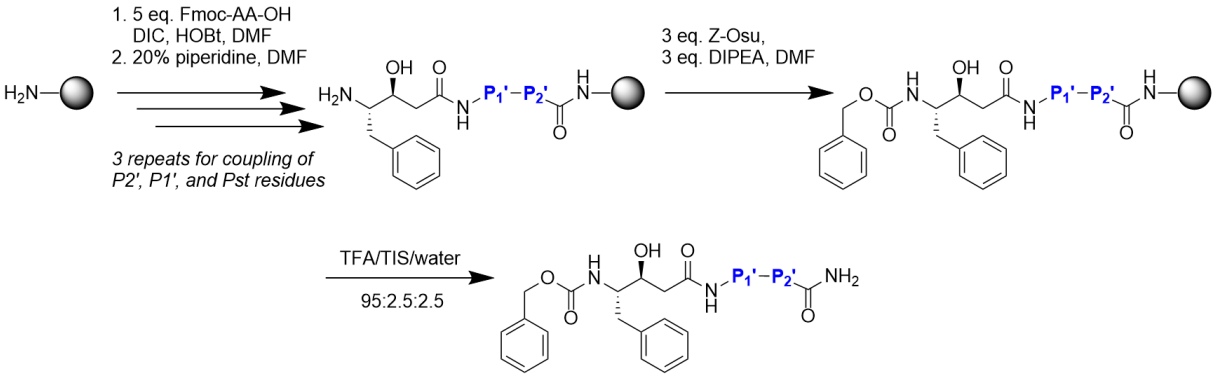
[4] Department of Biochemistry & Biophysics, University of California, San Francisco, San Francisco, CA 94158, USA

[5] Chan-Zuckerberg Biohub, San Francisco, CA 94158, USA

[6] Department of Pharmaceutical Chemistry, University of California San Francisco, UCSF Genentech Hall Box 2280, 600 16th St Rm S512, 94158, San Francisco, California, United States of America

Table of contents

Chemical syntheses



	cap		TSA		P1'		P2'
1a	Z	-	Pst	-	L-Glu	-	L-Abu - NH_2
2a	Z	-	Pst	-	L-Glu	-	L-Nva - NH_2
3a	Z	-	Pst	-	L-Glu	-	L-Val - NH_2
4a	Z	-	Pst	-	L-Glu	-	L-Nle - NH_2
5a	Z	-	Pst	-	L-Glu	-	L-Leu - NH_2
6a	Z	-	Pst	-	L-Glu	-	L-Ile - NH_2
7a	Z	-	Pst	-	L-Glu	-	L-Thr - NH_2
8a	Z	-	Pst	-	L-Glu	-	L-Met - NH_2
9a	Z	-	Pst	-	L-Glu	-	L-Trp - NH_2
10a	Z	-	Pst	-	L-Glu	-	L-Cha - NH_2
11a	Z	-	Pst	-	L-Glu	-	L-Phg - NH_2
12a	Z	-	Pst	-	L-Glu	-	L-Phe - NH_2
13a	Z	-	Pst	-	L-Glu	-	L-Hph - NH_2

	cap		TSA		P1'		P2'
1b	Z	-	Pst	-	L-Gln	-	L-Abu - NH_2
2b	Z	-	Pst	-	L-Gln	-	L-Nva - NH_2
3b	Z	-	Pst	-	L-Gln	-	L-Val - NH_2
4b	Z	-	Pst	-	L-Gln	-	L-Nle - NH_2
6b	Z	-	Pst	-	L-Gln	-	L-Ile - NH_2
7b	Z	-	Pst	-	L-Gln	-	L-Thr - NH_2
8b	Z	-	Pst	-	L-Gln	-	L-Met - NH_2
10b	Z	-	Pst	-	L-Gln	-	L-Cha - NH_2
11b	Z	-	Pst	-	L-Gln	-	L-Phg - NH_2
12b	Z	-	Pst	-	L-Gln	-	L-Phe - NH_2
13b	Z	-	Pst	-	L-Gln	-	L-Hph - NH_2

Fmoc-protected amino acids, Fmoc-(3S,4S)-AHPPA-OH (phenylstatine, Pst) transition state analogue (TSA) building block, and Rink amide MBHA resin were purchased from Iris Biotech (Marktredwitz, Germany) and Bachem (Bubendorf, Switzerland). N-(benzyloxycarbonyl)succinimide (Z-Osu) from Merck (Kenilworth, USA). Compounds were synthesized by standard Fmoc-chemistry solid phase peptide synthetic protocols on Rink amide MBHA resin support (subst. 0.69 mmol/g) in N,N-dimethylformamide (DMF) as a solvent. Syntheses were performed in 0.1 mmol scale, couplings with 5 equivalent amino acid excess to resin and 1-hydroxybenzotriazol (HOBt, 1.5 eq. to amino acid) and N,N'-diisopropylcarbodiimide (DIC, 1.5 eq. to amino acid) activation. Fmoc groups were removed with 20% piperidine. Coupling and deprotection as repeated to incorporate 3 residues for each compound. The N-(benzyloxycarbonyl) capping group was introduced with N-(benzyloxycarbonyl) succinimide (Z-Osu, 3 eq. to resin) in presence of N,N-diisopropylethylamine (DIPEA, 3 eq. to resin) in DMF as a solvent. Resin was then washed with dichloromethane and dried in vacuo. Compounds were deprotected and cleaved off the resin by incubation with a mixture of TFA/triisopropylsilane/water (95:2.5:2.5) for an hour.

Table S1. Data collection and refinement statistics.

	Apo-MayI	MayI-pepstatin A
Crystal data		
Space group	$C222_1$	$C222_1$
a, b, c (Å)	97.42, 112.06, 91.21	97.36 112.64 91.03
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Molecules per asymmetric unit	1	1
Matthews coefficient ($Å^3$ Da^{-1})	3.38	3.33
Solvent content (%)	63.62	63.06
Max. resolution (Å)	1.75	1.80
Data collection and processing		
Wavelength (Å)	0.918	1.542
Resolution limits (Å)	48.71-1.75 (1.86-1.75)	50.0-1.8 (1.91-1.80)
No. of observed reflections	264,390 (23,308)	138,736 (2,791)
No. of unique reflections	48515 (6,330)	39317 (2,229)
Multiplicity	5.4 (3.7)	3.5 (1.3)
R_{merge}[a]	0.135 (1.356)	0.087 (0.565)
$CC_{1/2}$[b]	0.997 (0.482)	0.997 (0.590)
Completeness (%)	95.8 (78.2)	84.3 (29.8)
$\langle I/\sigma I \rangle$	10.3 (1.0)	11.7 (0.99)
Refinement statistics		
Resolution (Å)	48.7-1.75 (1.80-1.75)	73.7-1.8 (1.85-1.80)
No. of reflections in working set	46,413 (2,589)	39,317 (2,229)
No. of reflections in test set	2101 (118)	1710 (30)
R_{work}[c] (%)	18.0 (39.5)	17.7 (44.5)
R_{free}[d] (%)	22.3 (39.3)	20.7 (42.9)
Average B-factor ($Å^2$)	29.6	24.6
RMSD bond length (Å)	0.013	0.012
RMSD angle (°)	1.6	1.6
Number of atoms in AU (protein/inhibitor/water molecules)	2642/0/346	2617/60/314
Ramachandran plot		
Most favored regions[e] (%)	97.99	97.65
Additional allowed regions[e] (%)	2.01	2.08
Disallowed regions[e] (%)	0.0	0.27
PDB code	6R5H	6R6A

Values in parentheses report the values in the highest resolution shell.

[a] $R_{merge} = \Sigma_{hkl} \Sigma_i |I_i(hkl)_i - \langle I(hkl) \rangle| / \Sigma_{hkl} \Sigma_i I_i (hkl)_i$.

[b] $CC_{(1/2)}$ is the correlation coefficient between random half data sets and from its value the Pearson correlation coefficient of the true level of signal can be calculated: $CC^* = \sqrt{2\frac{CC_{\frac{1}{2}}}{1} + CC_{\frac{1}{2}}}$ [S1].

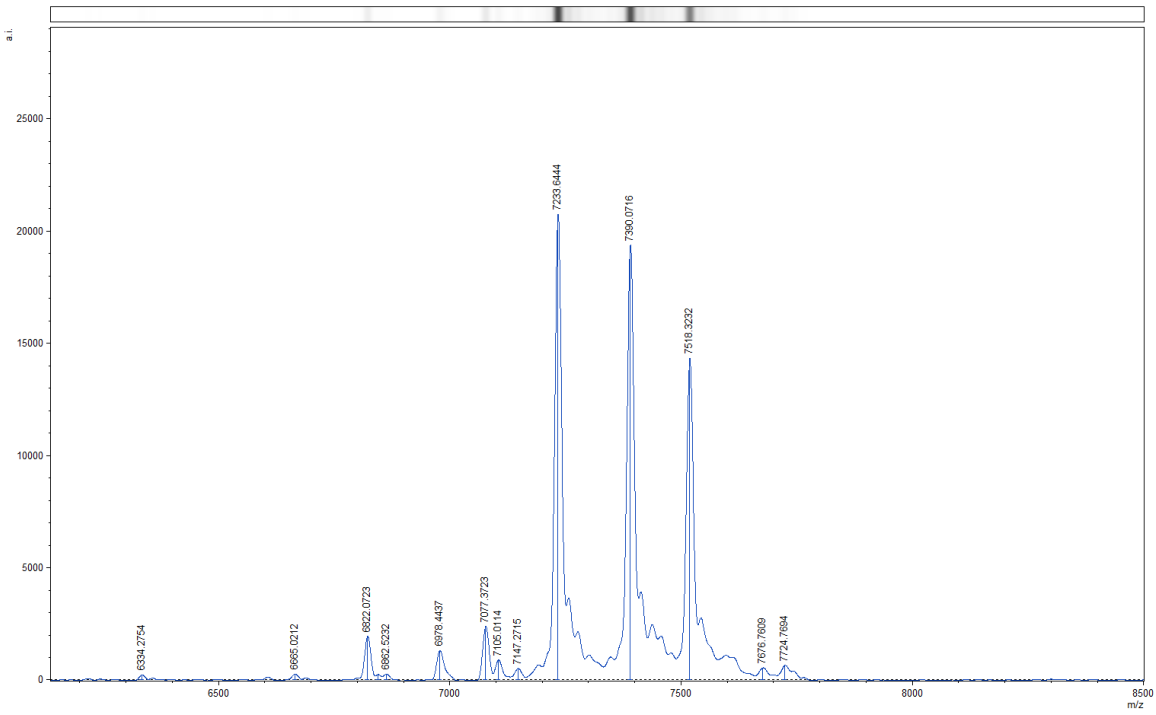
[c] R-value = $||F_o| - |F_c|| / |F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively.

[d] R_{free} is equivalent to the R-value but is calculated for 5% of the reflections chosen at random and omitted from the refinement process [S2].

[e] As determined by MolProbity [S3].

Table S2. Cytotoxicity of Z-Pst-*L*-Glu-Hph-NH_2 toward various human cell lines. Data were obtained through measurements of cell viability by luminometric CellTiter-Glo® 2.0 Cell Viability Assay (Promega).

	HeLa	CCRF-CEM	HL-60	MCF-7	HepG2
CC_{50}	> 80 μM	> 80 μM	> 80 μM	> 80 μM	> 80 μM



	m/z	
Δ	calcd.	found
-0,5	7077,87	7077,37
-0,41	7234,05	7233,64
-0,17	7390,24	7390,07
-0,09	7518,41	7518,32

RSAPSNSVPLTPLNSRQYSEDLAERQSWLLDQAKGLRSKYAPHLGERGQELRRRDIIDEGIT
RSAPSNSVPLTPLNSRQYSEDLAERQSWLLDQAKGLRSKYAPHLGERGQELRRRDIIDEGITR
RSAPSNSVPLTPLNSRQYSEDLAERQSWLLDQAKGLRSKYAPHLGERGQELRRRDIIDEGITRR
RSAPSNSVPLTPLNSRQYSEDLAERQSWLLDQAKGLRSKYAPHLGERGQELRRRDIIDEGITRRK

1 10 20 30 40 50 60 70 80

CNAG_05872 MHYLAVALPLLTLALAAPSNSVPLTPLNSRQYSEDLAERQSWLLDQAKGLRSKYAPHLGERGQELRRRDIIDEGITRRKRAN

Figure S1a. MALDI mass spectrometry analysis of May1 prodomain fragments. The mismatching N-terminal Arg-Ser dipeptide in the fragments is a remnant of the *Bgl*II restriction site product.

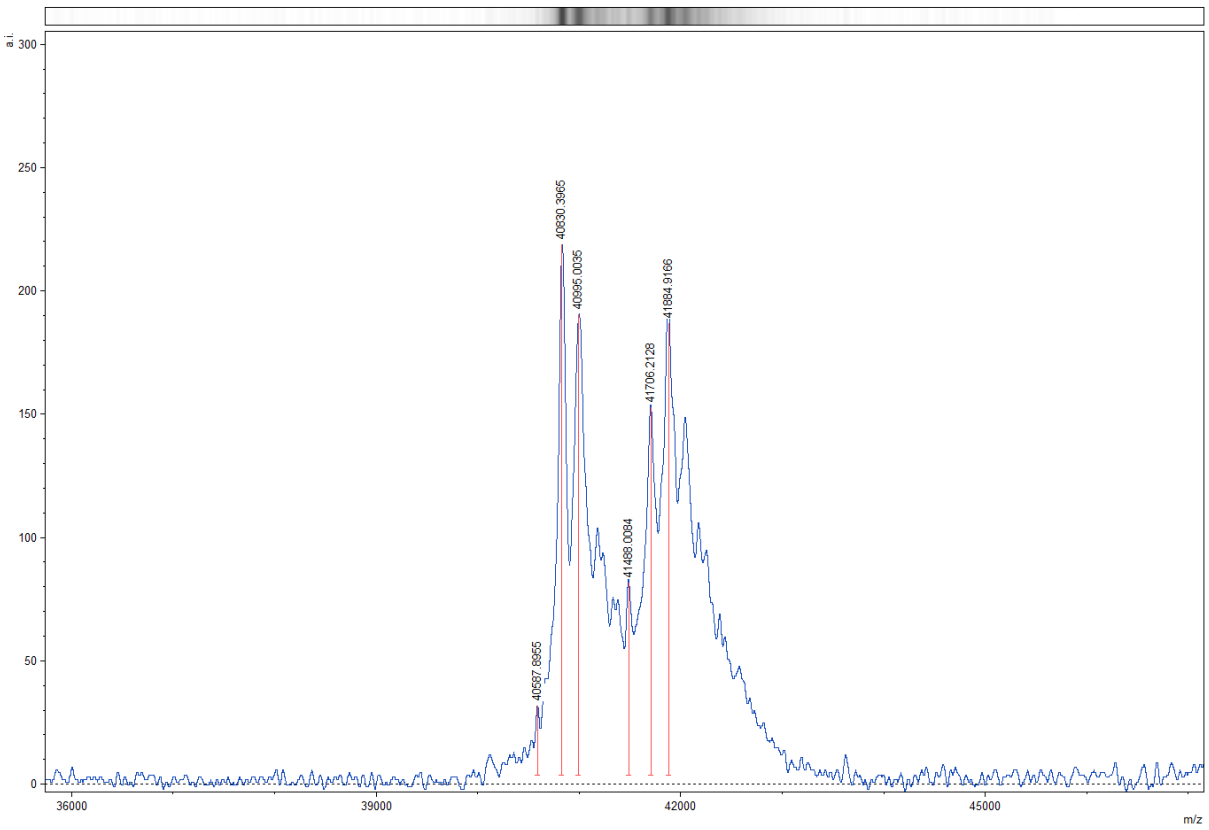


Figure S1b. (continued) MALDI mass spectrometry analysis of active May1 with its prodomain cleaved off. Largest possible average molecular weight of MayI fragment 77-434, including C-terminal tag and its biotinylation, is 40145.35 Da – this corresponds to the loss of the shortest detected prodomain fragment 15-76 (first entry in the table above).

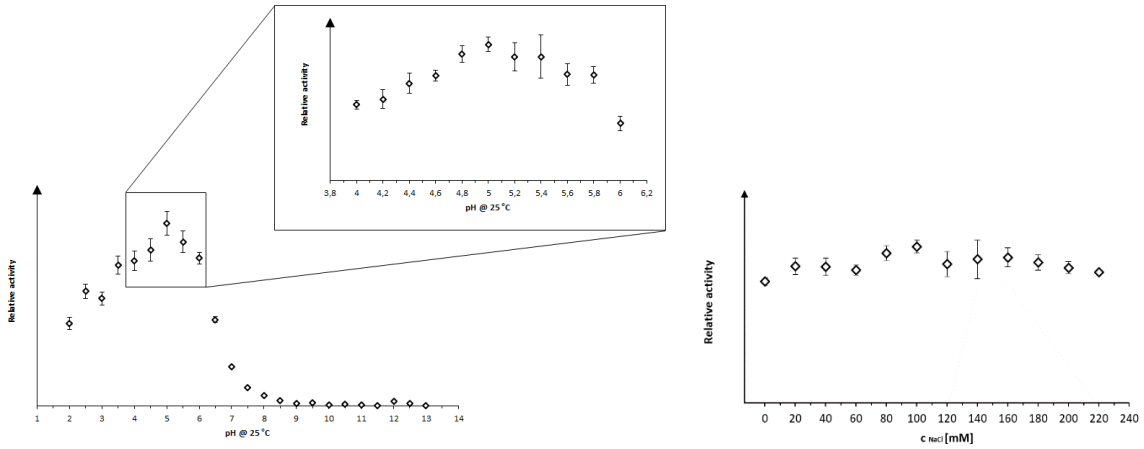


Figure S2. Activity profiles of May1 detected by cleavage of IQ-2 fluorescent substrate. Left side: Activity profile across pH 2.0 to 13.0 with steps of 0.5 in a Britton-Robinson buffer system. The activity maximum was further resolved with a series of buffers of pH 4.0 to 6.0 with steps of 0.2. Right side: Activity profile with varying salt concentration at pH 5.0.

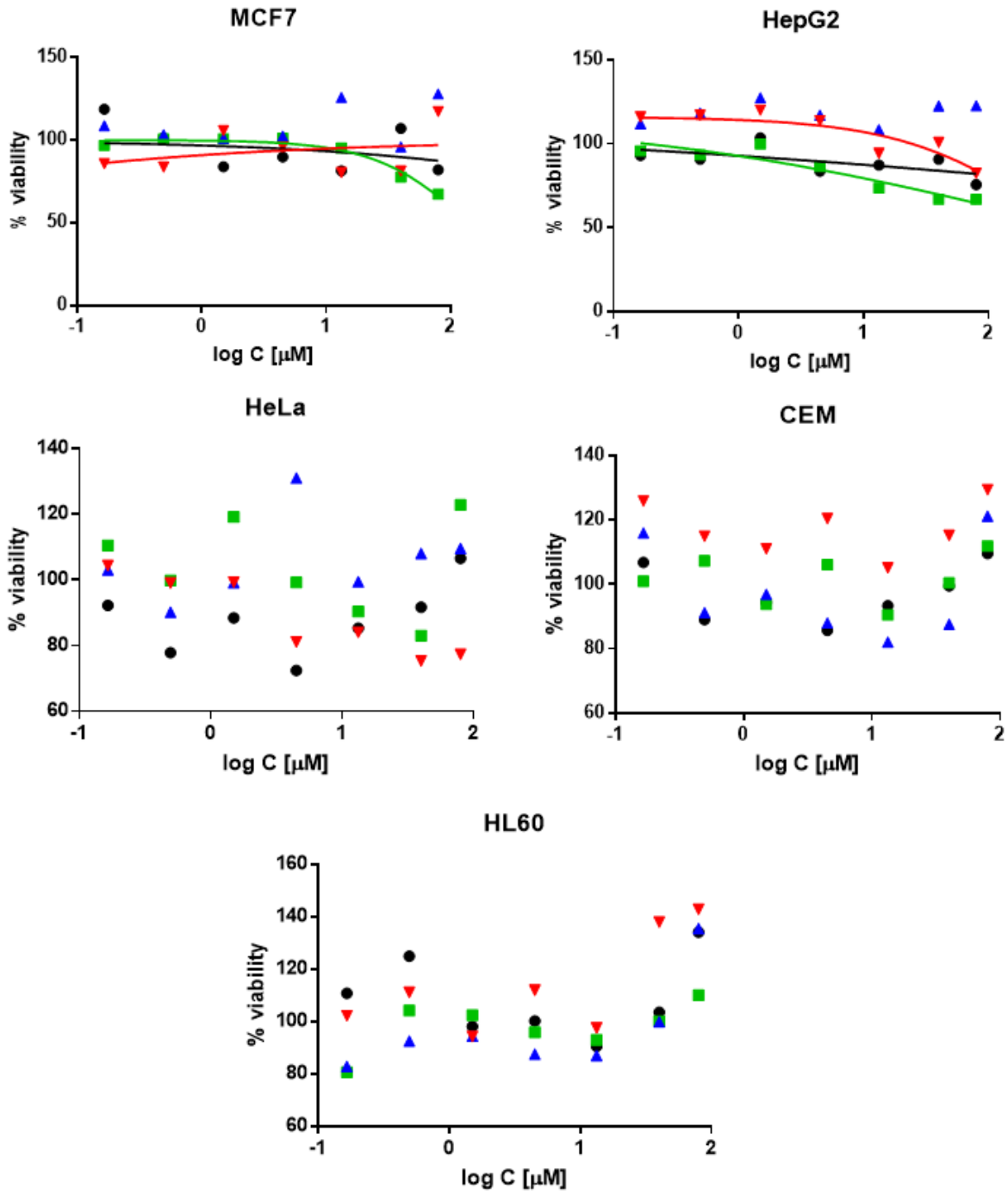


Figure S3. Cytotoxicity curves for five cell lines. Data were obtained through measurements of cell viability by luminometric CellTiter-Glo® 2.0 Cell Viability Assay (Promega).

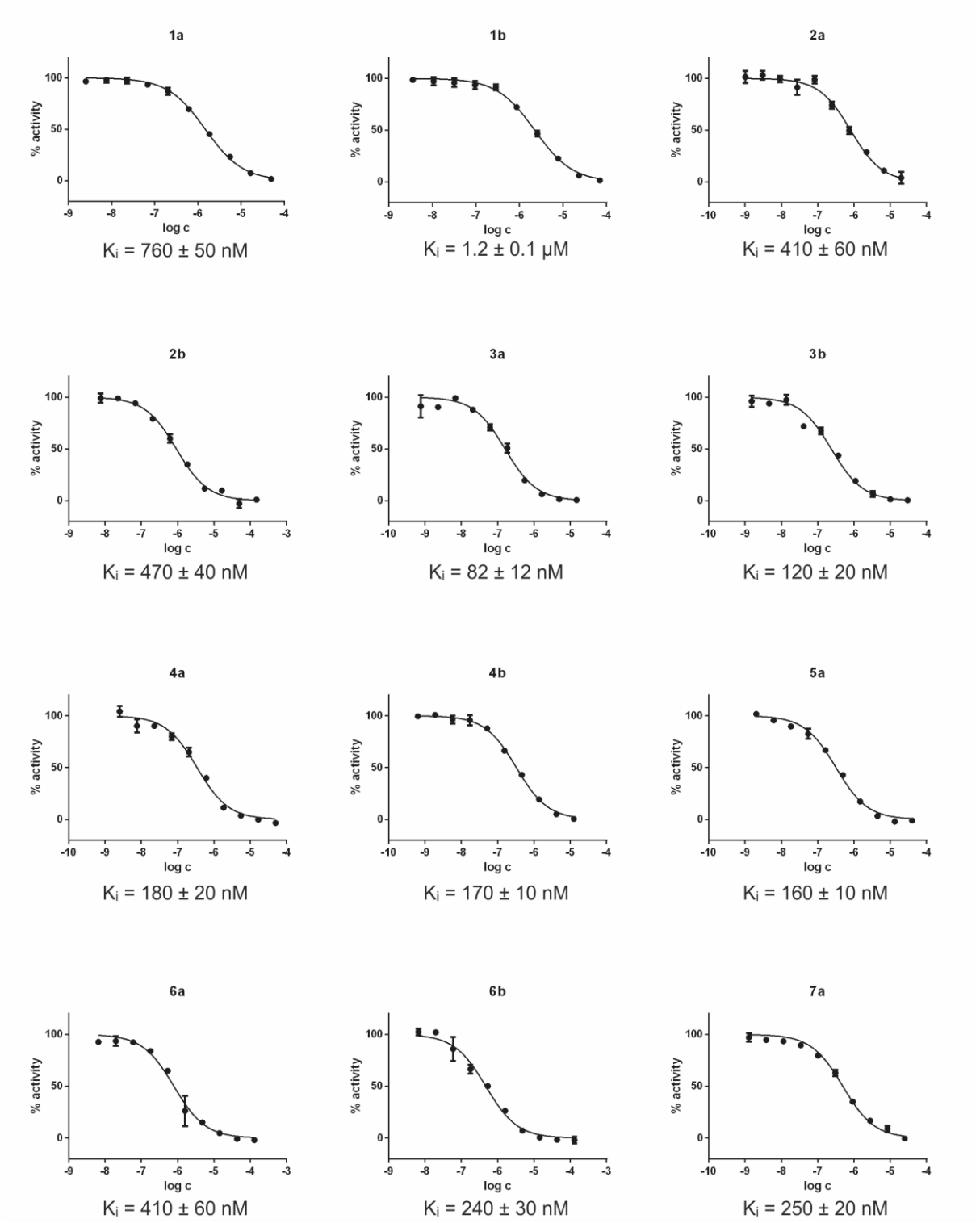


Figure S4a. May1 inhibition curves for compounds 1 – 13a and 1 – 13b with fitted K_i values and 95% CI. Figure continues on the next page.

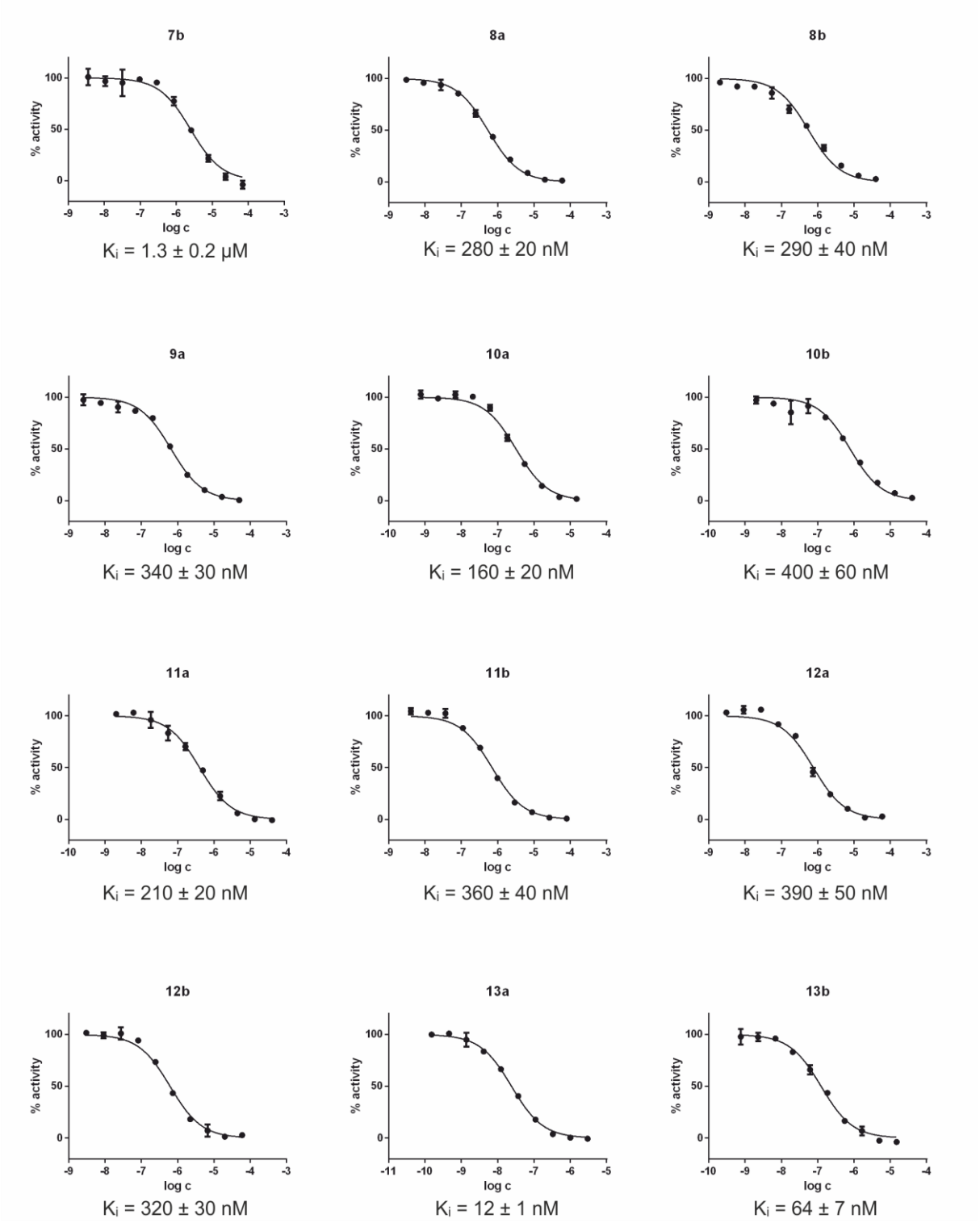
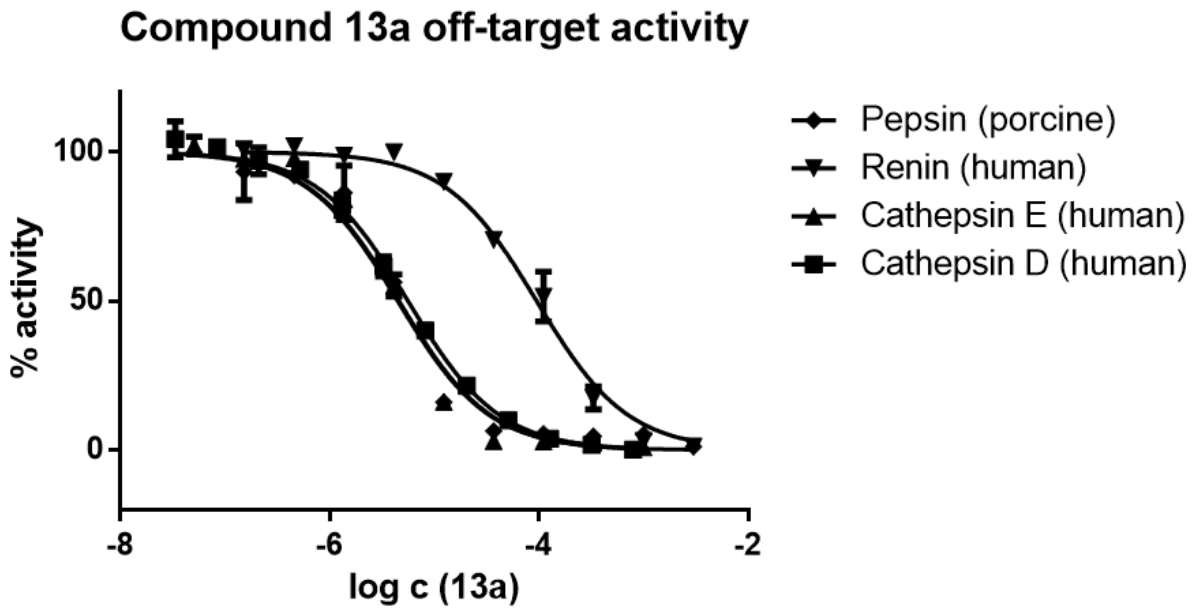


Figure S4b. (continued) May1 inhibition curves for compounds 1 – 13a and 1 – 13b with fitted K_i values and 95% CI.



enzyme	substrate K_m	substrate concentration	compound 13a K_i ± 95% CI
pepsin (porcine)	2.5 µM	10 µM	1.6 ± 0.3 µM
renin (human)	80 µM	150 µM	19 ± 2 µM
cathepsin E (human)	3.3 µM	30 µM	440 ± 70 nM
cathepsin D (human)	3.7 µM	30 µM	620 ± 50 nM

Figure S5. Inhibition curves for compound 13a against a panel of aspartic protease off-targets. Bottom table details the assay conditions used to calculate the fitted K_i values and 95% CI.

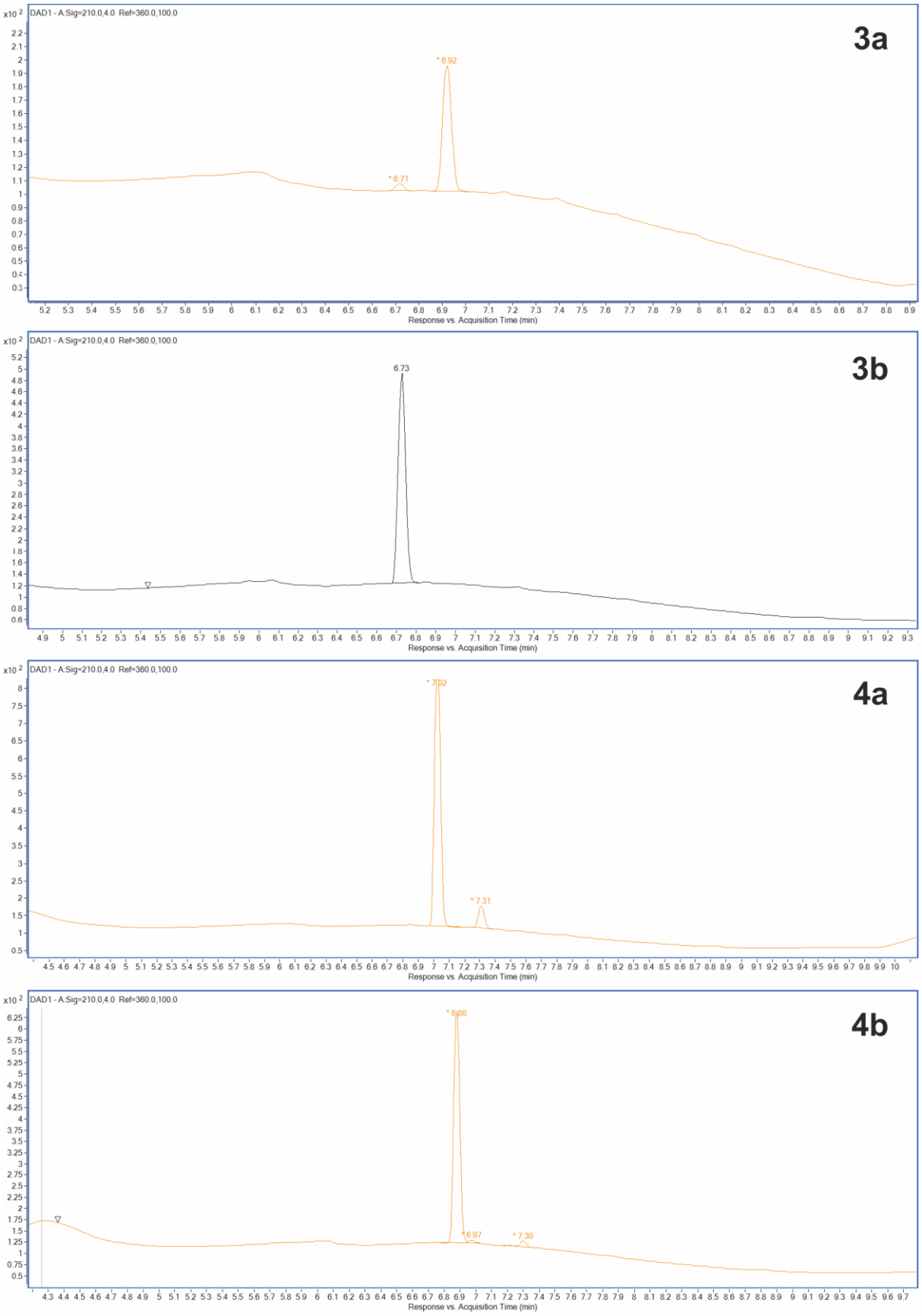


Figure S6a. LC traces of compounds 3a, 3b, 4a and 4b. Figure continues on the next page.

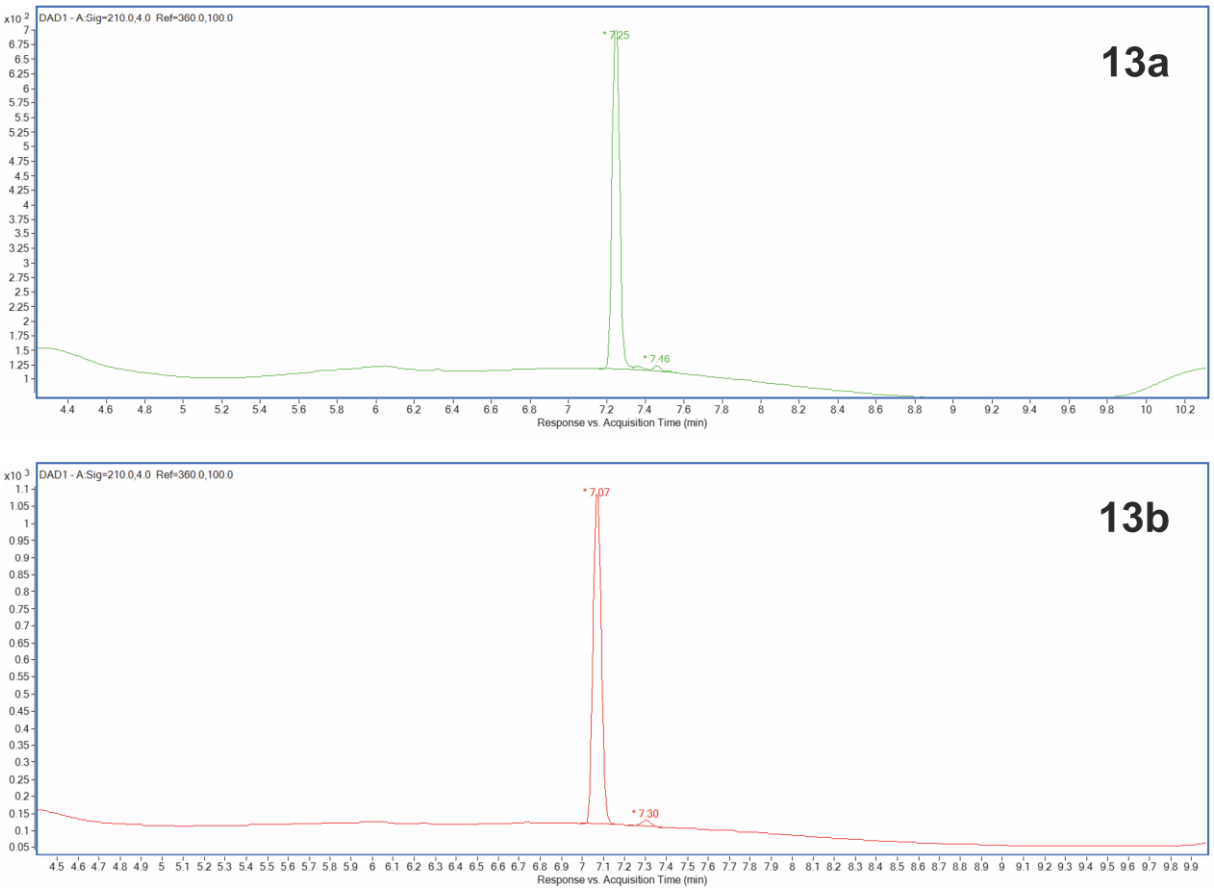


Figure S6b. (continued) LC traces of compounds 13a and 13b.

Supplementary references:

S1 Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030-1033 (2012).

S2 Brunger, A. T. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472-475 (1992).

S3 Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst D* **66**, 12-21, (2010).