Supporting Information: Fragment arrays for early druggability assessments

J. Aretz[a,b], Y. Kondoh[c,d], K. Honda[c,d], U. R. Anumala[e], M. Nazaré[e], N. Watanabe[c,d,f], H. Osada[c,d] and C. Rademacher[a,b]

[a] Department of Biomolecular Systems, Max Planck Institute of Colloids and Interfaces, Potsdam, Germany.
[b] Department of Biology, Chemistry, and Pharmacy, Freie Universität Berlin, Berlin, Germany.
[c] Antibiotics Laboratory, RIKEN, Wako, Japan.
[d] Chemical Biology Research Group, RIKEN Center for Sustainable Resource Science, Wako, Japan.
[e] Leibniz Institut für Molekulare Pharmakologie (FMP), Berlin, Germany.
[f] Industrial Biotechnology Research Laboratory, School of Biological Sciences, Universiti Sains Malaysia, Penang, Malaysia.

Table of content

Materials and Methods 2

Supporting figures 4

Figure S1: Quality controls of the proteins used in this study. 4

Figure S2: Scan of an array with the highest gain to test for autofluorescence of the fragments. 5

Figure S3: Correlation of signal intensities of monomeric DC-SIGN carbohydrate recognition domain (CRD) with tetrameric DC-SIGN extracellular domain (ECD). 6

Figure S4: Recovery rates for hits from ^{19}F NMR screenings in array experiments using a PEG linker for fragment immobilization. 6

Figure S5: Correlation of the hit rate of the array screening with the ^{19}F NMR screening using a PEG linker for immobilization. 7

Chemoinformatic analysis 8

Figure S6: Difference between “non-hitters” and “regular-hitters” using MACCS fingerprints (Student’s *t*-test). 8

Additional discussion of chemoinformatic analysis 12

Figure S7: Principle component analysis of MACCS fingerprints of the fragment library and the “regular hitters” data set. 13

Figure S8: Known CA2 inhibitors and related fragments present on the chemical fragment array. 14

Figure S9: MNK inhibitors from fragment-based ligand design. 15

Figure S10: Chemical arrays using 0.2 µM protein in the presence of HEK293T cell lysate. 15

Notes and References 16

Materials and Methods

Fragments used in this study

Fluorinated fragments and their selection were published earlier[1]. Hits from a fragment-based NMR screening against murine Langerin will be published independently of this report (Aretz *et al.*, unpublished data). Quality controls were carried out for each compound using ^{1}H NMR (data not shown).

Protein preparation

DC-SIGN, murine and human Langerin as well as MNK were expressed as described elsewhere[1, 2]. Bovine Carbonic Anhydrase II (CA2) was purchased from Sigma-Aldrich (C3934, St. Louis, MO).

Protein quality controls

Purity and size of the proteins was assessed using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) with subsequent Coumassie staining.

Fluorescent Labelling of C-type lectins

2 mL of protein solution was dialyzed overnight against 1 L of HBS buffer, pH 7.6, with 5 mM calcium chloride at 4°C. The protein solution was transferred into a 5 mL pear shaped flask and 30 mM mannose was added while stirring at room temperature. 1 mg Chromeo-642-NHS-ester dye (Active Motif, Carlsbad, CA) was dissolved in 10 µL DMF. From this stock solution, 1 µL was added to the protein solution in 0.2 µL steps. Afterwards, the reaction was carried out for 1 h at room temperature and then quenched with a final concentration of 50 mM ethanolamine, pH 8.5, for 20 min. Then, the reaction mixture was rebuffered against TBS, pH 7.6, using Zeba Spin Columns (MWCO 7,000, Pierce Biotechnology, Waltham, MA) after the addition of 10 mM EDTA. After rebuffering, a final concentration of 30 mM calcium chloride was added and the protein was purified via a mannan affinity chromatography as described earlier[3].

Fluorescent Labeling of MNK and CA2

MNK was diafiltered using Amicon Ultra-15 spin filters (MWCO 10,000, EMD Millipore, Billerica, MA) against HBS buffer, pH 7.6 and afterwards concentrated to 1 mL volume. 3 mg of CA2 were dissolved in HBS buffer, pH 7.6. The protein solution was transferred into 15 mL falcon tubes. 1 mg Chromeo-642-NHS-ester dye was dissolved in 10 µL DMF. From this stock solution, in total 2 µL were added to the protein solution in 0.2 µL steps. Afterwards, the reaction was carried out for 1 h at room temperature in the dark and then quenched by adding 4 mL TBS buffer, pH 7.8 for 1 h. Excessive dye was removed by diafiltration using Amicon Ultra-15 spin filters and TBS buffer, pH 7.8 until the flow-through was colorless.

Preparation of chemical arrays

Chemical arrays were prepared using photo affinity proline linker as described previously[4].

Optimization of the incubation conditions

To optimize the incubation conditions, we tested various methods to block the arrays (skim milk powder and BSA), to incubate the protein samples (4°C, 25°C, 1 h, overnight, with and without rotating, in presence and absence of BSA or skim milk), to detect protein binding (directly labelled, primary and secondary labelled detection) and to wash the arrays (quick and extensive). Extensive and multiple washing steps were not beneficial for signal quality, so we used directly fluorescently labelled proteins and a quick washing procedure using cold buffer (thrice for 5 s). On the other hand, this enhanced the background signal in samples with higher protein concentrations and lead to a background that was not evenly distributed in samples without mixing. A higher background signal was additionally observed when blocking the arrays with skim milk. This can probably be traced back to interactions from components of the skim milk with our lectins. In the end, using low concentrations of labelled protein after blocking the arrays with BSA while rotating during incubation

yielded the best results for Langerin and DC-SIGN (Fig. 1 A, D). The rigid proline linker outperformed the flexible PEG linker with respect to background signal and signal to noise.

Fragment immobilization

For the initial test of the array, compounds were printed in three concentrations (2.5 mM, 5 mM, and 10 mM dissolved in DMSO). For most of the compounds it was beneficial to use a higher concentration for immobilization. For Langerin 18 signals were significantly enhanced on the 10 mM spots compared to three that showed this behaviour on the 2.5 mM spots. The results for DC-SIGN were comparable (20 to three). The overall recovery rates for hits against Langerin and DC-SIGN for the 10 mM spots were 69% and 55%, respectively. Including the hits that were unique for the 2.5 mM spots, the recovery rates were 71% and 59%, respectively. Choosing higher concentrations is on average more efficient if it is not possible to optimize the immobilization concentration for each compound. This effect is probably caused by compounds with a high absorption at $\lambda = 365$ nm which is the wavelength used during the photoreaction.

Performing the micro array experiment

The arrays were blocked with 2% BSA in TBS-T (TBS with 0.05% Tween-20) for at least 1 h at room temperature while shaking. After washing thrice for 5 min with TBS-T while shaking, the protein sample was applied in a Microarray Hybridization Chamber (G2534A, Agilent Technologies, Santa Clara, CA) at 0.2 µM. Samples were incubated overnight at 25°C in a vertical rotator. Afterwards, the arrays were washed thrice with cold TBS-T with 2 mM calcium chloride for 5 s and dried by centrifugation. Then, the arrays were scanned using a GenePix 4300A microarray scanner (Molecular Devices, Sunnyvale, CA).

Preparation of HEK293T cell lysate

A confluent 60 mm petri dish of HEK293T cells was harvested by scraping and resuspended in and washed with cold PBS (800 x *g*, 4°C). After another centrifugation step, the cells were resuspended in 500 µL of sample buffer and lysed by ultrasonication (3 x 10 s).

NMR screening

The screening was performed as described previously for DC-SIGN and human Langerin[1]. The same method was applied for murine Langerin as well[1]. For CA2, the nanomolar inhibitor 6-Ethoxy-2-benzothiazolesulfonamide (333328, Sigma-Aldrich) was added for competition at a final concentration of 200 µM[5]. For MNK, ATP and ManNAc were added at a final concentration of 10 mM each. Compounds that changed their signal intensity either as competitor (3%) or potential allosteric binder (5%) were considered as hits (8% in total, Aretz, *et al.*, 2016, DOI: 10.1139/cjc-2015-0603).

Data analysis of the chemical fragment arrays

Signal intensities were calculated by subtracting the mean background signals from the mean values of each spot using GenePix Pro 7 (Molecular Devices). Then, the signal data for each compound and for the DMSO spots were grouped in KNIME 2.11.0[6] and analyzed using an ANOVA followed by Dunnett's test in GraphPad Prism 6.0 (GraphPad Software, La Jolla, CA, USA) or R 2.15.0[7]. MACCS (Molecular ACCess System) fingerprints were calculated in KNIME 2.11.0 using the MACCS Keys MOE node (Chemical Computing Group, Montreal, Canada)[8].

Supporting figures

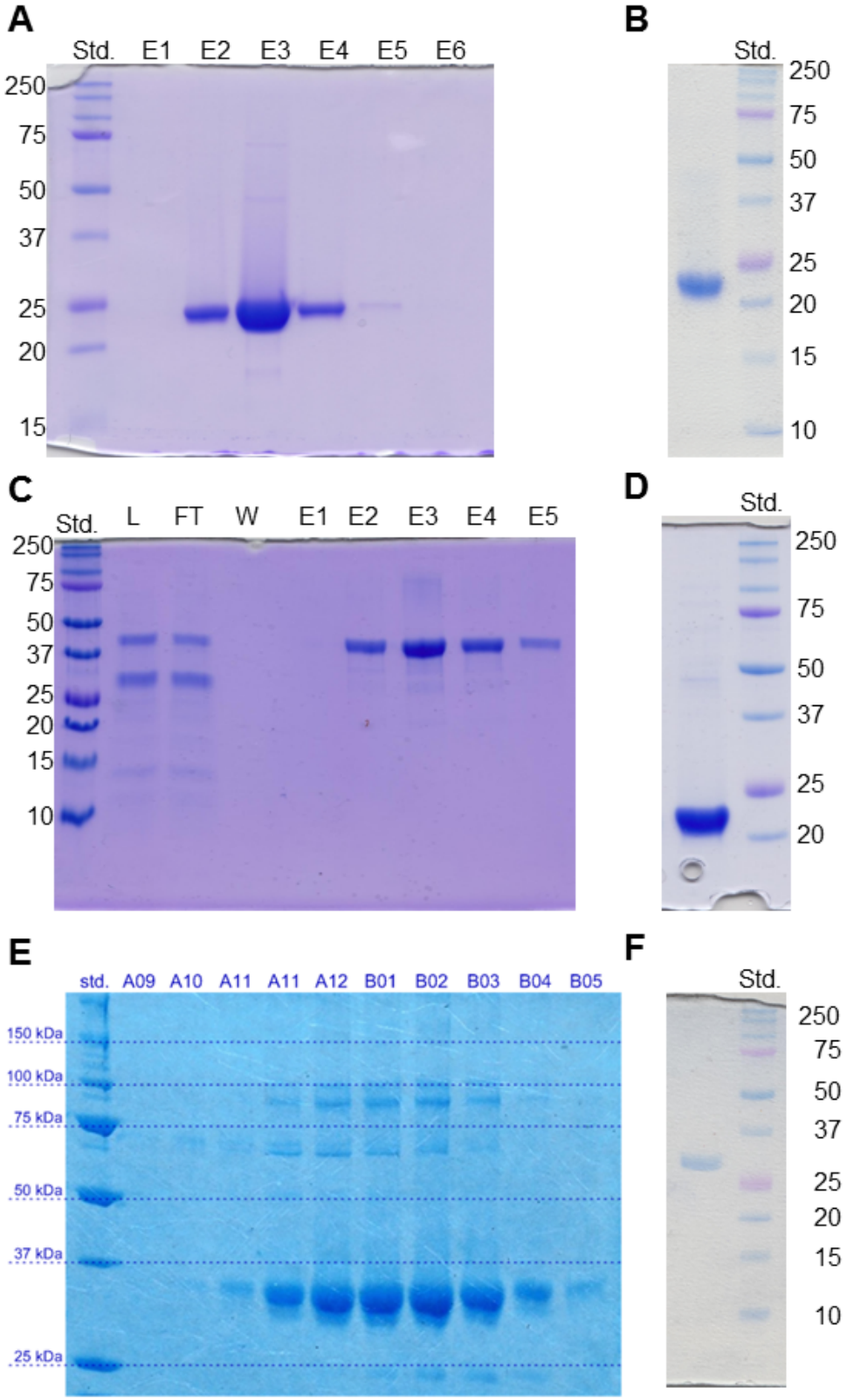

Fig. S1: Quality controls of the proteins used in this study. To monitor the purity and size of the protein preparations, SDS-PAGE was performed for murine Langerin (23 kDa, A), human Langerin (23 kDa, B), DC-SIGN extracellular domain (42 kDa, C), DC-SIGN carbohydrate recognition domain (20 kDa, D), MNK (35 kDa, E), and CA2 (30 kDa, F). Abbrevations used: Std.: standard, E: elution, L: load, FT: flow-through, W: wash.

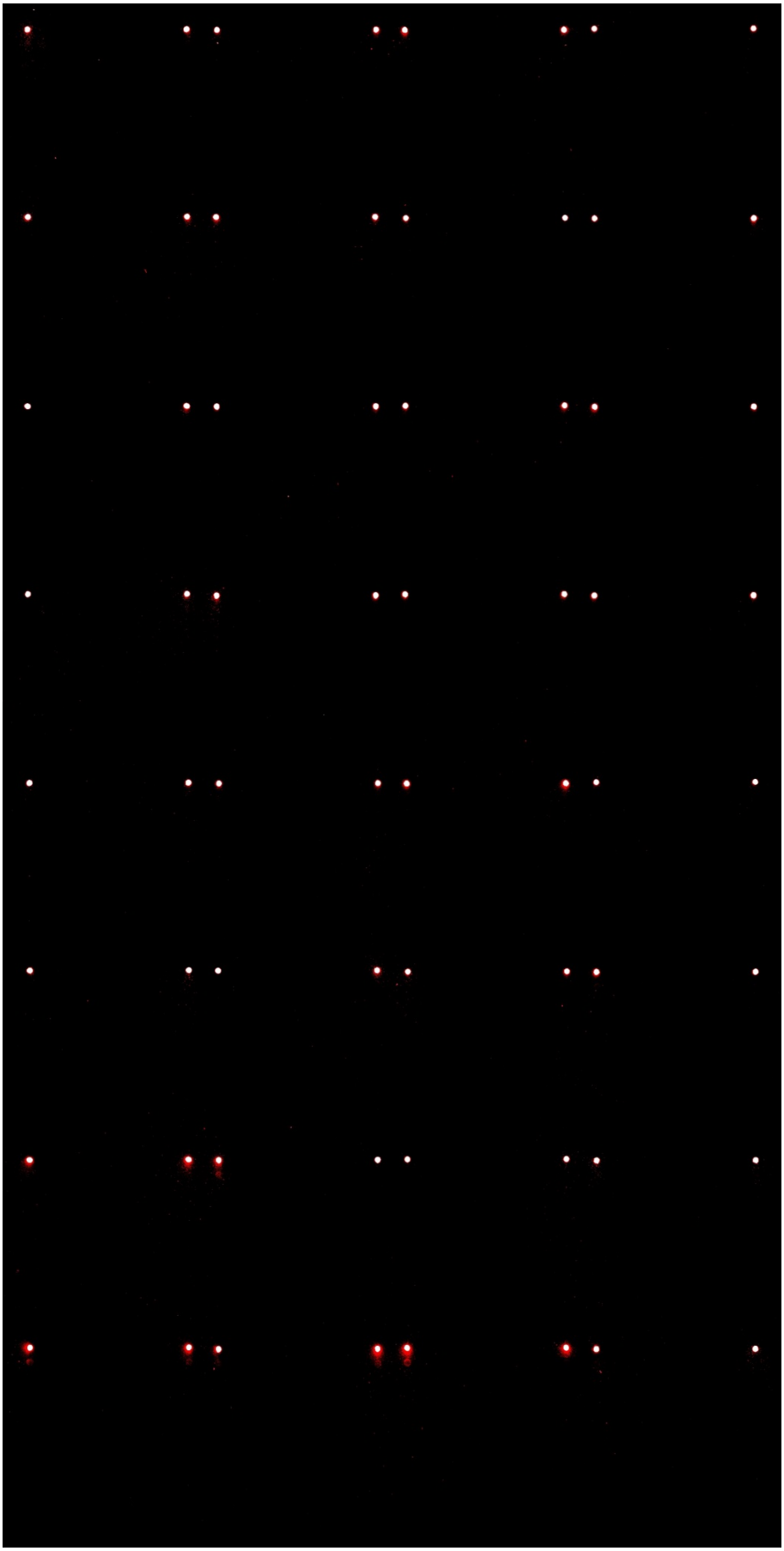

Fig. S2: Scan of an array with the highest gain to test for autofluorescence of the fragments. Only the position markers of each block gave rise to detectable signals.

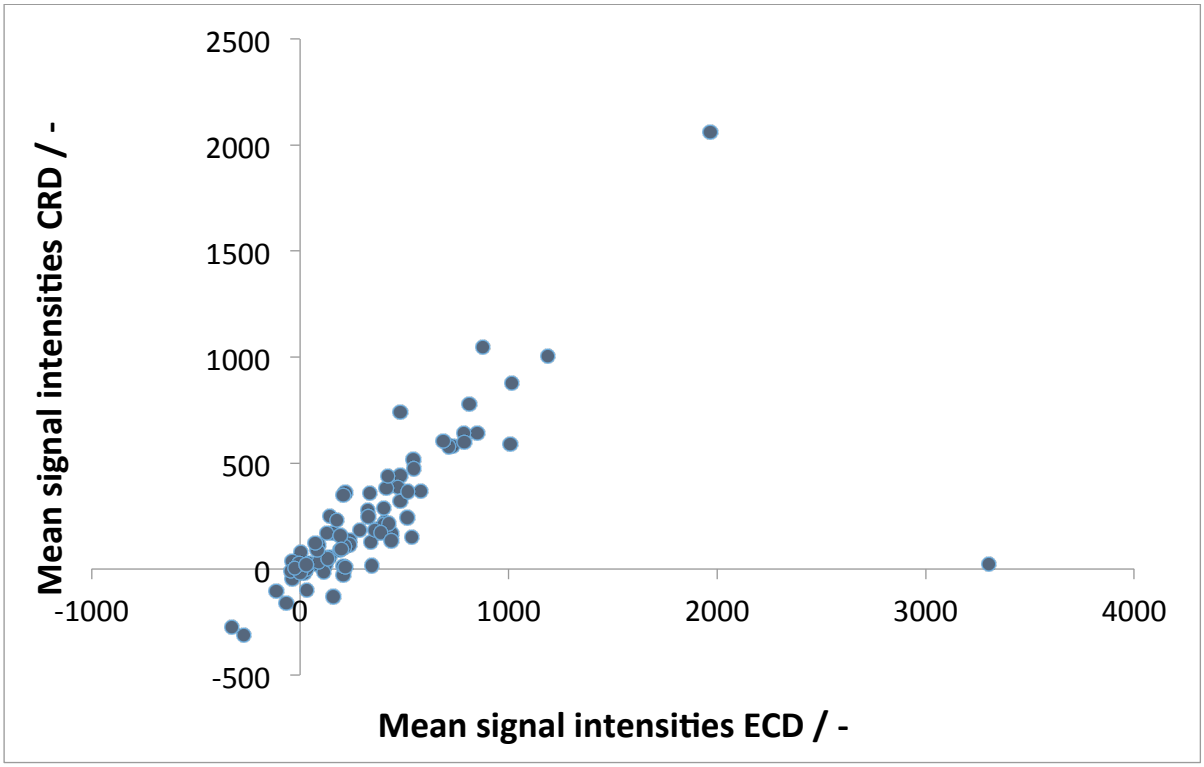

Fig. S3: Correlation of signal intensities of monomeric DC-SIGN carbohydrate recognition domain (CRD) with tetrameric DC-SIGN extracellular domain (ECD).

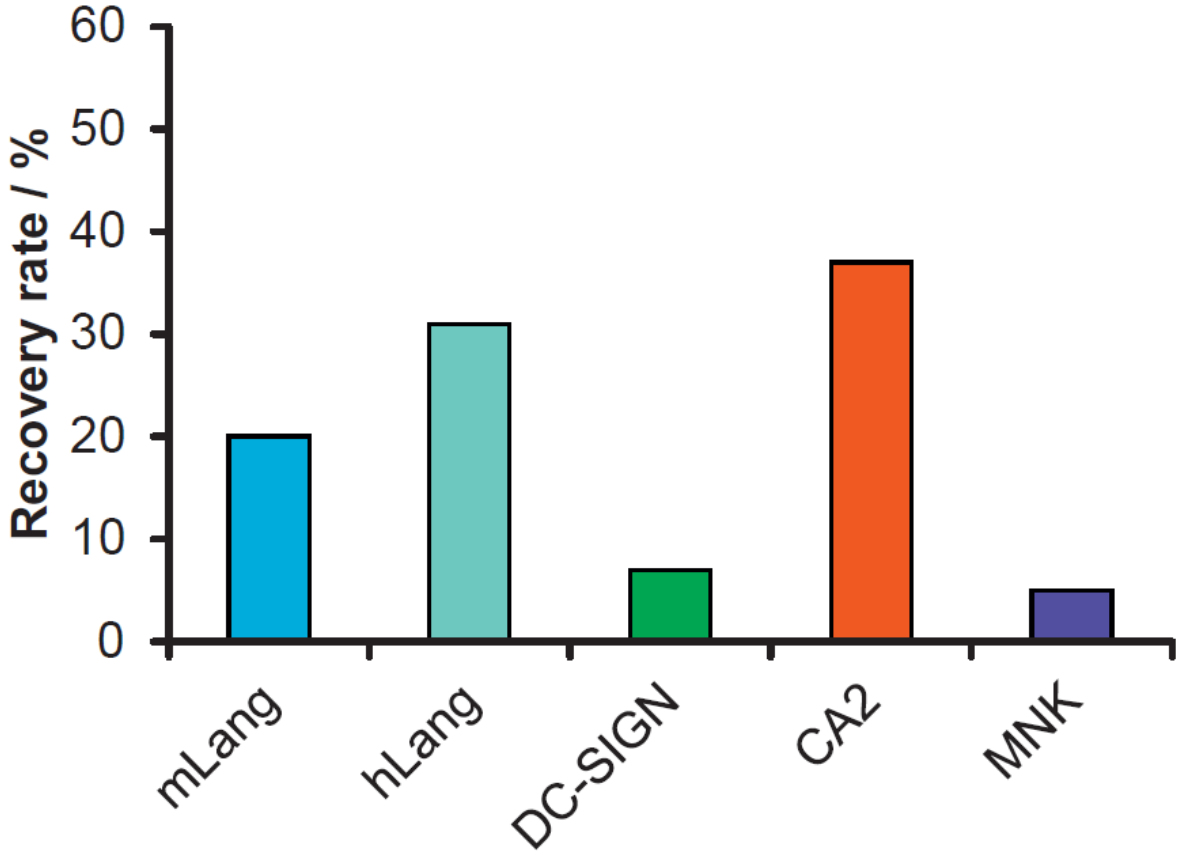

Fig. S4: Recovery rates for hits from ^{19}F NMR screenings in array experiments using a PEG linker for fragment immobilization.

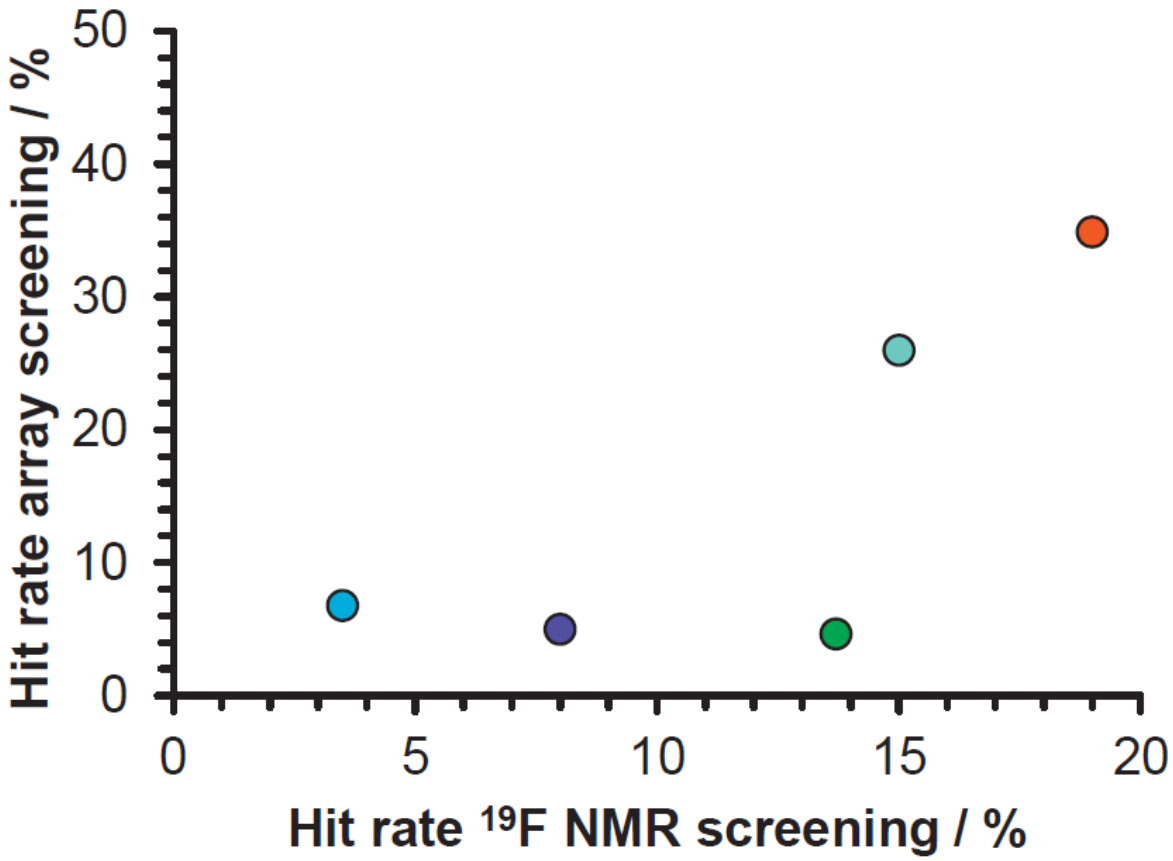

Fig. S5: Correlation of the hit rates of the array screening with the ^{19}F NMR screening using a PEG linker for immobilization.

Chemoinformatic analysis

MACCS fingerprint	p value
1: #isotopes	1,000
2: #atoms with atomic number > 103	1,000
3: #group IVA, VA and VIA periods 4-6	1,000
4: #Actinides	1,000
5: #group IIIB, IVB elements	1,000
6: #Lanthanides	1,000
7: #group VB, VIB, VIIB elements	1,000
8: #heteroatoms in 4-membered rings	0,696
9: #group VIIIB elements	1,000
10: #alkaline earth elements	1,000
11: #atoms in 4 ring	0,931
12: #group IB, IIB elements	1,000
13: #N connected to 1 O and 2 C	0,557
14: #S atoms in S-S groups	1,000
15: #C connected to 3 O	1,000
16: #heteroatoms in 3-membered rings	1,000
17: #C in CC triple bonds	0,747
18: #group IIIA elements	1,000
19: #atoms in 7 ring	0,970
20: #silicon atoms	1,000
21: #C = bonded to C and 3 heavy atoms	0,158
22: #atoms in 3 ring	0,984
23: #C bonded 1 N and 2 O	0,318
24: #O-N single bonds	0,820
25: #C bonded to at least 3 N atoms	0,627
26: #C in 3 ring bonds and a double bond	0,318
27: #iodine atoms	1,000
28: #XCH2X, where X<>C	0,568
29: #phosphorous atoms	1,000
30: #non-C Q4 bonded to >= 3 C	1,000
31: #halogens connected to non carbons	1,000
32: #S bonded to an N and a C	0,376
33: #S atoms bonded to N	0,297
34: #CH2= units	1,000
35: #alkali (group IA) elements	1,000
36: #S atoms in rings	0,852
37: #C bonded to >= 1 O & >=2 N	0,703
38: #C bonded >= 2 N and 1 C	0,848
39: #S atoms bonded to 3 O	0,318
40: #S single bonded to OQ2	0,318
41: #N in C#N	0,645
42: #fluorine atoms	0,030
43: #X-H heteroatoms 2 bonds from another	0,008

44: #other elements	1,000
45: #N atoms adjacent to -C=C	0,979
46: #bromine atoms	0,419
47: #S two bonds from an N	0,969
48: #non C bonded to >= 3 O	0,318
49: #charged atoms	0,195
50: #C in C=C bonded to >= 3 C	0,196
51: #S bonded to a C and an O	0,507
52: #N bonded to N	0,368
53: #QH 4 bonds from another QH	0,012
54: #QH 3 bonds from another QH	0,000
55: #S bonded to >=2 O	0,409
56: #N bonded to >= 2O and >= 1 C	1,000
57: #O in rings	0,622
58: #S bonded to >=2 non-carbon atoms	0,409
59: #non-aromatic S-[a]	0,984
60: #[S+]-[O-]	0,409
61: #SQ3	0,409
62: #non-ring bonds that connect rings	0,279
63: #N atoms in double bonds with O	1,000
64: #non-ring S attached to a ring	0,620
65: #N in aromatic bonds with C	0,960
66: #CX4 bonded to >=3 carbons	0,986
67: #S attached to heteroatoms	0,454
68: #QH bonded to another QH	0,318
69: #QH bonded to another Q	0,814
70: #N bonded to two non-C heavy atoms	0,920
71: #N bonded to O	0,480
72: #O separated by 3 bonds	0,304
73: #S in double/charge separated bonds	0,555
74: #dimethyl substituted atoms	0,405
75: #N non-ring bonded to a ring	0,310
76: #C in C=C bonded to >= 3 heavy atoms	0,654
77: #N separated by 2 bonds	0,954
78: #N double bonded to C	0,535
79: #N separated by 3 bonds	0,091
80: #N separated by 4 bonds	0,786
81: #S attached to Q >= 3 atoms	0,854
82: #heteroatoms attached to a CH2	0,002
83: #heteroatoms in 5 ring	0,616
84: #NH2 groups	0,729
85: #N bonded to >= 3 C	0,756
86: #CH2 or CH3 separated by non-C	0,542
87: #halogens bonded to any ring	0,159
88: #sulfurs	0,869
89: #O separated by 4 bonds	0,845
90: #het. 3 bonds from a CH2	0,083

91: #het. 4 bonds from a CH2	0,038
92: #C bonded to >=1 N, >=1 C & >= 1 O	0,212
93: #methylated heteroatoms	0,459
94: #N bonded to non C	0,712
95: #O 3 bonds from an N	0,129
96: #atoms in 5-rings	0,445
97: #O 4 bonds from an N	0,242
98: #het. in 6-ring	0,891
99: #C in C=C	0,698
100: #N attached to CH2	0,088
101: #atoms in 8-ring or higher	0,911
102: #O bonded to non C heavy atoms	0,304
103: #chlorine atoms	0,939
104: #hets. 2 bonds from a CH2	0,001
105: #hets. ring bonded to a 3-ring bond X	0,939
106: #X bonded to >= 3 non-C	0,068
107: #XQ>3 bonded to at least 1 halogen	0,477
108: #CH3 4 bonds from a CH2	0,979
109: #O attached to CH2	0,802
110: #O 1 C from an N	0,157
111: #N 2 bonds from a CH2	0,138
112: #atoms with coordination number >= 4	0,020
113: #O in non-aromatic bonds to an [a]	0,514
114: #CH3 attached to CH2	0,355
115: #CH3 2 bonds from a CH2	0,752
116: #CH3 3 bonds from a CH2	0,586
117: #N 2 bonds from an O	0,587
118: (key(147)-1 if key(147)>1; else 0)	0,605
119: #N in double bonds	0,535
120: (key(137)-1 if key(137)>1; else 0)	0,339
121: #N in rings	0,413
122: #N with coordination number >=3	0,858
123: #O separated by 1 C	0,052
124: #het-het bonds	0,706
125: Is # AROMATIC RING > 1?	0,075
126: #non-ring O bonded to 2 heavy atoms	0,704
127: (key(143)-1 if key(143)>1; else 0)	0,622
128: #CH2s separated by 4 bonds	0,680
129: #CH2s separated by 3 bonds	0,212
130: (key(124)-1 if key(124)>1; else 0)	0,690
131: (# het atoms with H)	0,000
132: #O 2 bonds from CH2	0,161
133: #N non-ring bonded to a ring	0,610
134: #halogens	0,039
135: #N in a non-aromatic bond with [a]	0,835
136: Bit: is there more than 1 O=	0,567
137: Total # ring HETEROCYCLE atoms	0,580

138: (key(153)-1 if key(153)>1; else 0)	0,191
139: #OH groups	0,000
140: (key(164)-3 if key(164)>3; else 0)	0,059
141: (key(160)-2 if key(160)>2; else 0)	1,000
142: (key(161)-2 if key(161)>1; else 0)	0,340
143: #non ring O connected to a ring	0,425
144: #atoms separated by (!:):(!:)	0,420
145: #6M RING > 1	0,041
146: Key(164)-2 if key(164)>2; else 0	0,044
147: #CH2 attached to CH2	0,444
148: #non-C with coordination number >=3	0,839
149: (key(160)-1 if key(160)>1; else 0)	0,713
150: #X separated by (!r)-r-(!r)	0,675
151: #NH	0,059
152: #C bonded to >=2 C and 1 O	0,681
153: #non-carbons attached to CH2	0,141
154: #O in C=O	0,094
155: #non-ring CH2	0,015
156: #XN where coord. # of X>=3	0,800
157: #O in C-O single bonds	0,012
158: #N in C-N single bonds	0,125
159: Key(164)-1 if key(164)>1; else 0	0,043
160: #CH3 groups	0,245
161: #N	0,420
162: #aromatics	0,598
163: #atoms in 6 rings	0,031
164: #oxygens	0,053
165: #ring atoms	0,002
166: Is there more than 1 fragment?	0,116

Fig. S6: Difference between “non-hitters” and “regular-hitters” using MACCS fingerprints (Student’s *t*-test).

Additional discussion of chemoinformatic analysis

Notably several compounds did not show binding in any of the performed experiments even though they were identified as binders by NMR while others did in most of the cases. To explore common features we analyzed these chemoinformatically (Fig. S6). For this purpose, the 281 fluorinated fragments were separated into three groups: compounds that never showed a significantly enhanced signal in at least one experiment were considered as "non-hitters" (Dunnett's test against DMSO controls, $p < 0.05$) while compounds that hit in at least 80% of the performed experiments were considered as "frequent hitters" (Dunnett's test against DMSO controls, $p < 0.001$). All other compounds were considered as "regular hitters".

"Non-hitters" were in average significantly smaller than "regular hitters" concerning MW and number of non-hydrogen atoms (21 Da and 2 HA, *t*-test, equal variances not assumed, $p < 0.001$). They also have less and smaller ring systems (#6M Ring > 1, #atoms in 6 membered rings and #ring atoms are significantly lower). Interestingly, the amount of heteroatoms in both groups was the same while the number of X-H heteroatoms was significantly lower in the "non-hitter" group (*t*-test, equal variances not assumed, $p < 0.001$). The important groups for this effect were hydroxyls but not amine groups (#OH groups significantly lower, #NH groups and #NH2 groups not affected significantly). In a study analyzing the favored reaction products of a diazirine containing photoaffinity linker with different small organic molecules, a tendency towards a favored reaction with hydroxyl groups is observed as well[9]. This leads to the assumption that some binding epitopes are more prone to be impaired by the linker conjugation. In a larger molecule this effect is less likely as the linker has more attachment sites. Hydroxyl groups react faster with the linker and thus a reaction is more probable[9]. Thus, this group may protect other necessary binding epitopes. A higher number of hydroxyl groups per molecule may decrease the likelihood for an essential hydroxyl group to react. Alternatively, the immobilization density could be increased if fragments contain more hydroxyl groups. A feature of compounds that is known to impair the immobilization density is absorption at $\lambda = 365$ nm, the wavelength at which the photoaffinity reaction is performed. Finally, new binding epitopes may arise during the reaction of the linker, because a trifluoromethyl group and a benzamide are attached to the fragment. This effect was already observed for some compounds from the Langerin SAR that did not bind in the SPR assay but on the array.

A limited number of "frequent hitters" was observed, with only twelve fragments binding to 80% of the targets during every screening round. This low number renders a statistical analysis difficult. The only feature that was significantly enhanced in this group was the number of aromatic features (#aromatics). The higher number of aromatic substructures suggests that larger aromatic substituents may be more susceptible to false positive behavior due to non-specific binding.

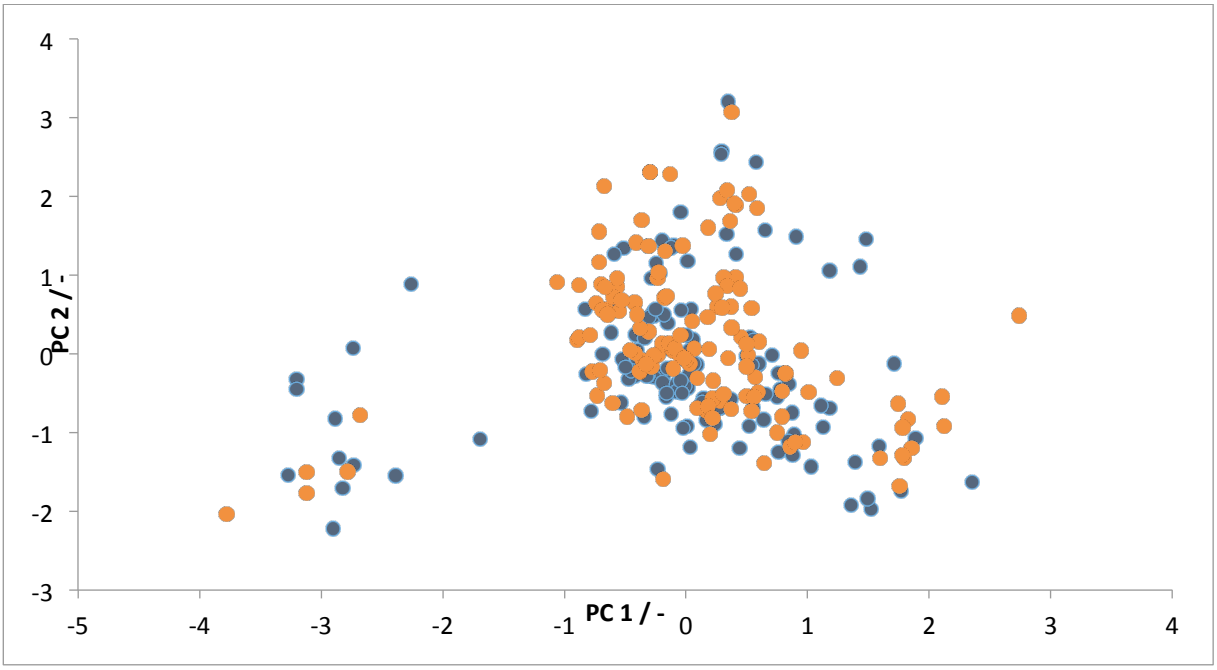

Fig. S7: Principle component analysis of MACCS fingerprints of the fragment library (blue) and the "regular hitters" data set (orange).

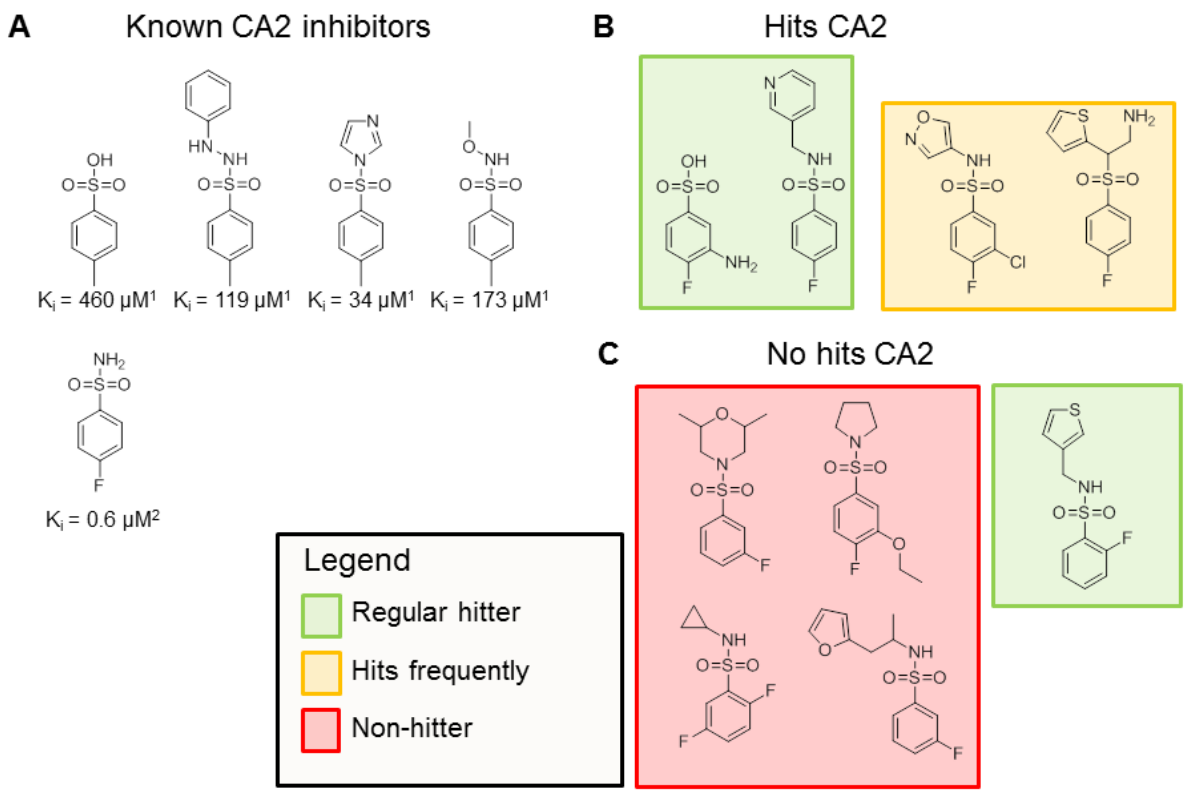

Fig. S8: Known CA2 inhibitors[5] and related fragments present on the chemical fragment array. To analyze whether fragment arrays are able to identify suitable starting points for drug design, we found nine compounds in the immobilized library that resemble known CA2 inhibitors. (A) Previously identified inhibitors for bovine CA2 [10, 11]. (B+C) Nine fragments present on the array with high substructure similarity to known CA2 inhibitors. (B) Four were identified as hits against CA2 on the array while (C) five compounds were not identified as hits. The compounds are classified as "regular hitters" (green), "frequent hitters" (yellow) and "non-hitters" (red, see "additional discussion of chemoinformatic analysis").

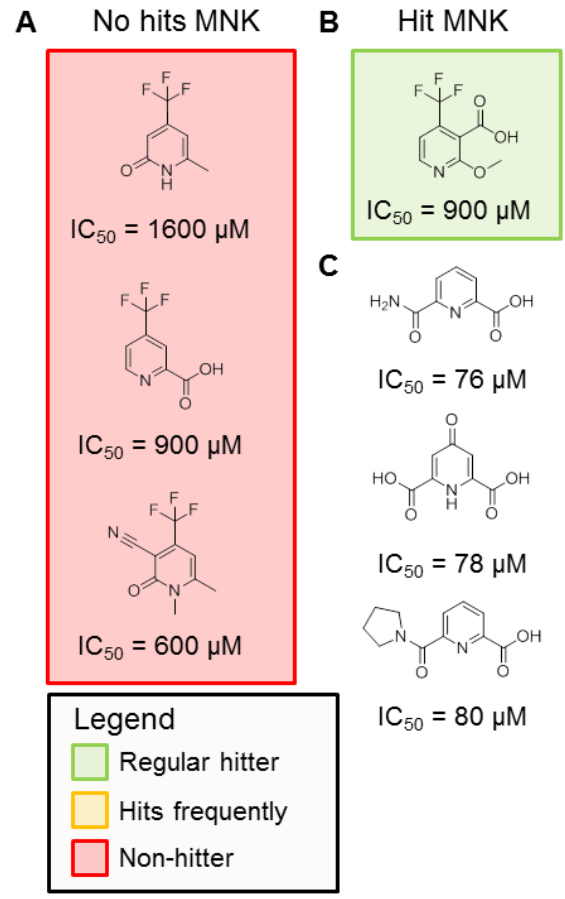

Fig. S9: MNK inhibitors from fragment-based ligand design. We previously identified picolinic acid derivatives as highly efficient fragment inhibitors against MNK (Aretz, *et al.*, 2016, DOI: 10.1139/cjc-2015-0603). Four of these inhibitors were immobilized on the fragment array. (A) Three of these four inhibitors were not identified, (B) while one fragment was a hit during the array screening. The compounds are classified as "regular hitters" (green), "frequent hitters" (yellow) and "non-hitters" (red, see "additional discussion of chemoinformatic analysis"). (C) Previously described MNK inhibitors with picolinic acid scaffold (Aretz, *et al.*, 2016, DOI: 10.1139/cjc-2015-0603).

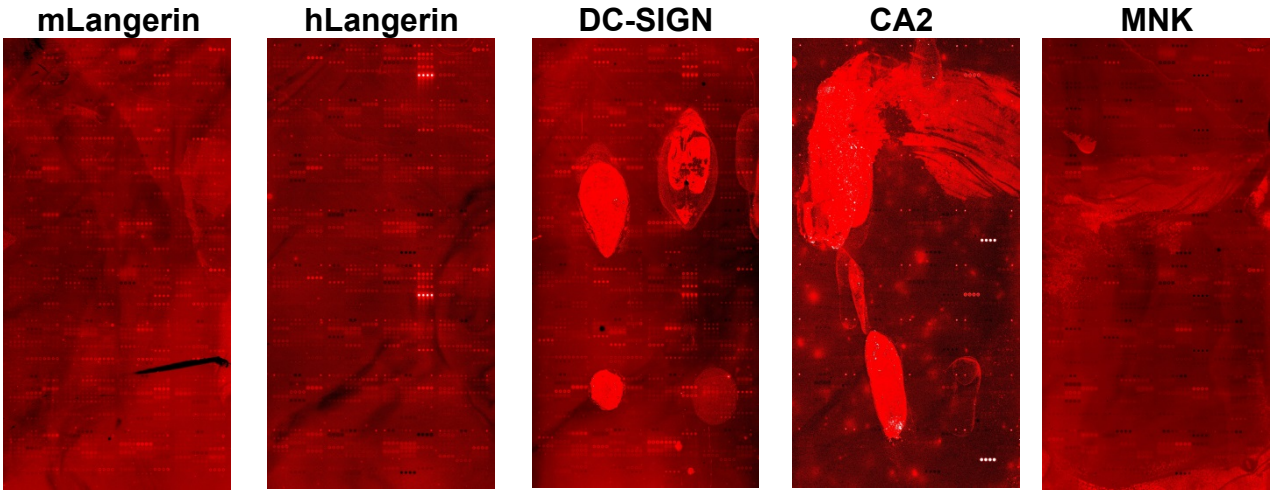

Fig. S10: Chemical arrays using 0.2 µM protein in the presence of HEK293T cell lysate.

Notes and References

1. J. Aretz, E. C. Wamhoff, J. Hanske, D. Heymann and C. Rademacher, *Frontiers in immunology*, 2014, **5**, 323.
2. J. Martinez, L. D. Nguyen, S. Hinderlich, R. Zimmer, E. Tauberger, W. Reutter, W. Saenger, H. Fan and S. Moniot, *J Biol Chem*, 2012, **287**, 13656-13665.
3. N. S. Stambach and M. E. Taylor, *Glycobiology*, 2003, **13**, 401-410.
4. Y. Kondoh, K. Honda and H. Osada, *Methods in molecular biology*, 2015, **1263**, 29-41.
5. V. M. Krishnamurthy, G. K. Kaufman, A. R. Urbach, I. Gitlin, K. L. Gudiksen, D. B. Weibel and G. M. Whitesides, *Chem Rev*, 2008, **108**, 946-1051.
6. N. C. Michael R. Berthold, Fabian Dill, Thomas R. Gabriel, Tobias Koetter, Thorsten Meinl, Peter Ohl, Christoph Sieb, Kilian Thiel, Bernd Wiswedel, in *Studies in Classification, Data Analysis, and Knowledge Organization*, Springer, 2007.
7. R Development Core Team, *Journal*, 2012, **2.15.0**.
8. I. Chemical Computing Group, *Journal*, 2013.
9. T. Suzuki, T. Okamura, T. Tomohiro, Y. Iwabuchi and N. Kanoh, *Bioconjug Chem*, 2015, **26**, 389-395.
10. F. Briganti, R. Pierattelli, A. Scozzafava and C. T. Supuran, *Eur J Med Chem*, 1996, **31**, 1001-1010.
11. V. M. Krishnamurthy, B. R. Bohall, C. Y. Kim, D. T. Moustakas, D. W. Christianson and G. M. Whitesides, *Chem Asian J*, 2007, **2**, 94-105.