Holly Foster 26/03/2020

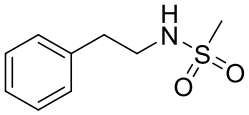
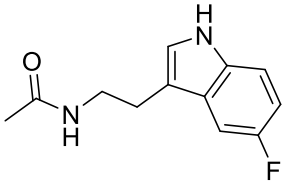
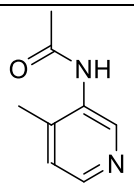
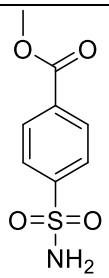
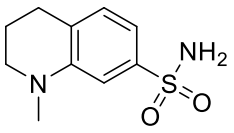
COVID-19 Enumeration of Active Site Fragment Hits

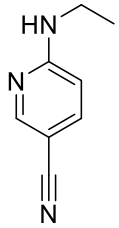
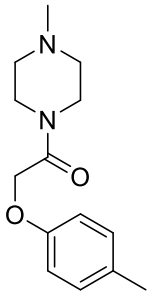
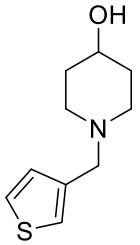
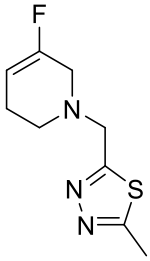
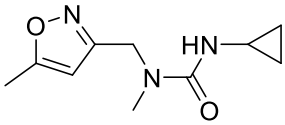
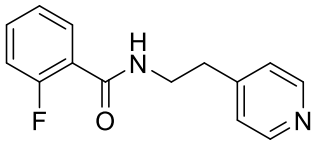
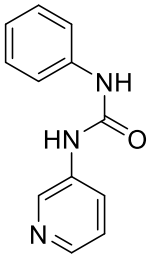
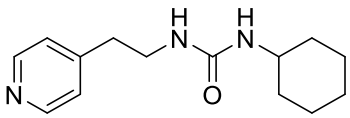
Aim: Enumerate the non-covalent active site fragment hits from the X-Chem screen using well documented one-step coupling techniques and compounds available from enamine.

Enumeration of Fragment Hits

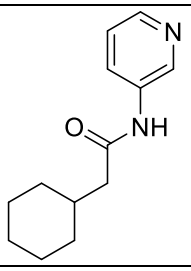
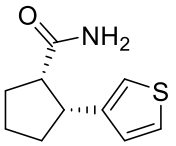
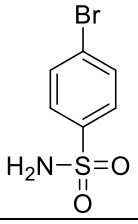
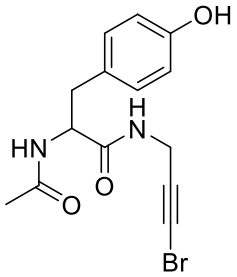
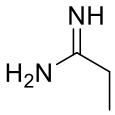
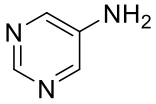
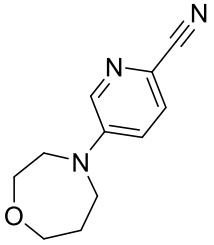
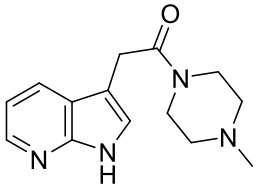
Key steps of the proposed protocol:

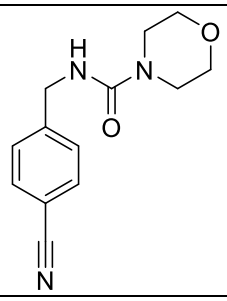
1. Identified existing functional groups in the fragments that may be readily enumerated using well documented coupling chemistries. The aim was to minimise modifications to the fragment hit and grow with existing linker vectors to maximise retention of binding of the original fragment.
2. Visually inspected the crystal structures with fragments to assess if growth via the enumeration method suggested would be tolerated spatially and would not alter the pose of the fragment.
3. It was important to investigate if the starting material derived from the fragment hit was commercially available to expediate the synthesis stage. POB0129 is not readily available and Z45617795 is not available from enamine, so these compounds could be filtered out.
4. Using a bespoke KNIME workflow the suitable fragments were enumerated, filtered (for unwanted functional groups and PAINS) and were exported to Excel.

ID	Structure	Enumeration Ideas	Spatially tolerated?	Requisite fragment purchasable from enamine/other suppliers*
Z45617795		Sulfonamide	Yes	Yes*
Z1220452176		Amide coupling	Small space for growth and extension in this region may be solvent exposed	Yes
Z1129283193		Amide coupling	Yes	Yes
Z18197050		Sulfonamide and transesterification. Convert ester to amide – amide couplings	More space for transesterification/amide coupling	Acid and sulfonyl chloride – both yes
Z1367324110		Sulfonamide and N-alkylation	More space for alkylation. Sulfonamide extension may	Amine and sulfonyl chloride – both yes

			be solvent exposed	
Z219104216		N-alkylation	Small space for growth and extension in this region may be solvent exposed	Yes
Z285634836		N-Alkylation	No. Methyl points directly at cavity.	N.A.
Z285634899		O-Alkylation/N-Alkylation with alternative cyclic amines	Yes. Portion to add should be reasonably flexible.	Yes for fragment and bromomethyl thiophene
Z1619978933		No easy methods.	N.A.	N.A.
Z369936976		N-Alkylation of methyl urea	Yes	Not available.
Z1310876699		No easy methods.	N.A.	N.A.
Z44592329		N-Alkylation of urea nearest to phenyl	Yes	Not available.
Z111507846		No easy methods.	N.A.	N.A.

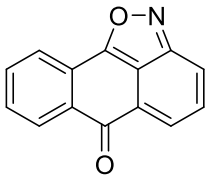
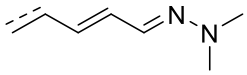
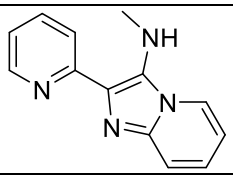
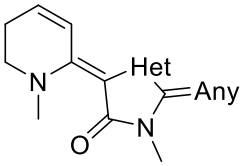
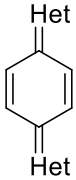
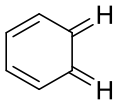
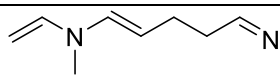
Holly Foster 26/03/2020

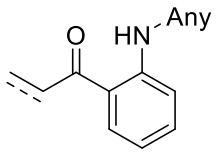
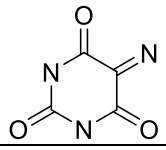
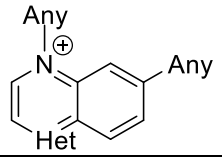
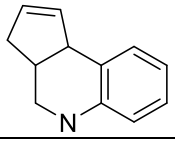
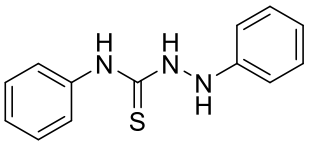
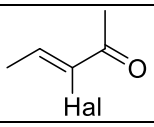
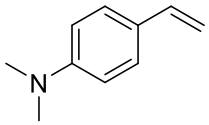
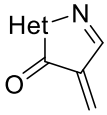
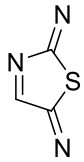
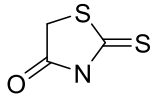
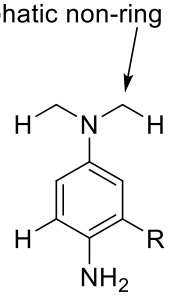
Z31792168		Amide coupling – change acid	Yes	Yes
POB0129		Amide coupling	Small fragments should be tolerated	No
NCL-00023830		Sulfonamide, Suzuki	Sulfonamide only as Br pointing at pocket	Yes
NCL-00024905		Amide coupling, phenol alkylation and Suzuki	Phenol and Br pointing towards pocket – no room for growth. Amide would just be solvent exposed.	N.A.
Z1741969146		No easy methods.	N.A.	N.A.
Z1741970824		No easy methods.	N.A.	N.A.
Z1348371854		No easy methods.	N.A.	N.A.
Z1400780201		N-Alkylation	No – methyl piperazine pockets directly at a pocket.	N.A.

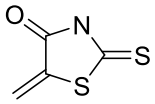
Z509756472		No easy methods.	N.A.	N.A.

Development of Custom PAINS Filter

It has been well documented that PAINS filters based on the seminal PAINS paper by Holloway and Baell in 2010 (https://pubs.acs.org/doi/10.1021/jm901137j) can falsely identify PAINS, thus removing potential active compounds from a screening collection. A recent study by Tropsha *et al*. (*J. Chem. Inf. Model*. 2017, **57**, 417−427) analysed data from 153,339 unique compounds that were tested in six different PPI assays (available in PubChem) to establish 'real' PAINS based on the available assay data. Compounds were described as frequent hitters if they were active in two out of six assays, and infrequent hitters if they were active in one or zero assays. The enrichment percentage was defined as the number of frequent over infrequent hitters. For this project structures which had an EV > 10% were included in the custom pains filter. See the below table for information on these structures, where: $N_{FH\text{-}PAINS}$ = number of frequent hitters, $N_{IH\text{-}PAINS}$ = number of infrequent hitters, N_{PAINS} = total number of PAINS, and EV% = enrichment percentage.

PAINS Alert	Substructure	$N_{FH\text{-}PAINS}$	$N_{IH\text{-}PAINS}$	N_{PAINS}	EV,%
Quinone_B(5)		3	1	4	300.0
Imine_ene_A(5)		3	2	5	150.0
Het_65_Db(5)		4	3	7	133.3
Ene_rhod_J(3)		1	3	4	33.3
Quinone_A(370)		47	160	207	29.4
Quinone_D(2)		9	36	45	25.0
Dyes5A(27)		1	5	6	20.0

Anthranil_one_A(38)		3	16	19	18.8
Imine_one_sixes(27)		3	16	19	18.8
Het-pyridiniums_A(39)		5	29	34	17.2
Anil_alk-ene(51)		2	12	14	16.7
Thio_urea_D(8)		1	6	7	16.7
Ene_one_hal(17)		14	89	103	15.7
Anil_di_alk_B(251)		16	116	132	13.8
Ene_five_het_A(201)		7	51	58	13.7
Het_thio_5_imine_A(1)		1	9	10	11.1
Rhod_sat_A(33)		4	39	43	10.3
Azo_A(34)	N=N	9	114	123	7.9
Anil_di_alk_A(478)	Aliphatic non-ring atoms  R = H, CH or OCH_2-CH	14	1083	1097	1.3

Ene_rhod_A(235)		4	593	597	0.67

Another interesting paper by Vidler *et al.* (https://pubs.acs.org/doi/10.1021/acsmedchemlett.8b00097#) based on a similar analysis of a smaller collection of compounds from industry suggested that the Azo_A(34), Anil-di_alk_A(478) and Ene-rhod_a(235) were particularly nuisance PAINS. Although these structures had a lower enrichment percentage than was defined, there was lots of data points for these compounds suggesting that they are some of the most common PAINS, so were also included in the filter.

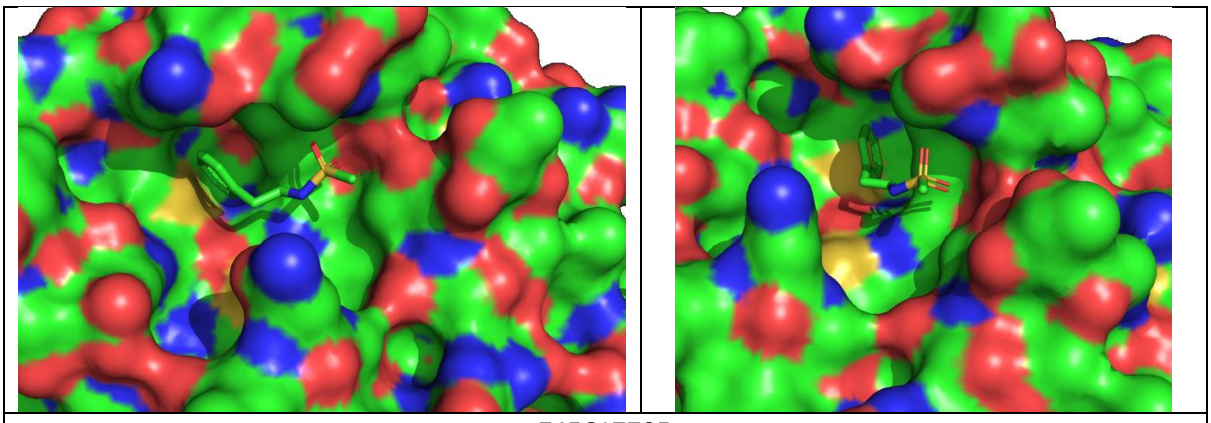
Property Filter using FAF

Can we use known data to guide this filtering process?

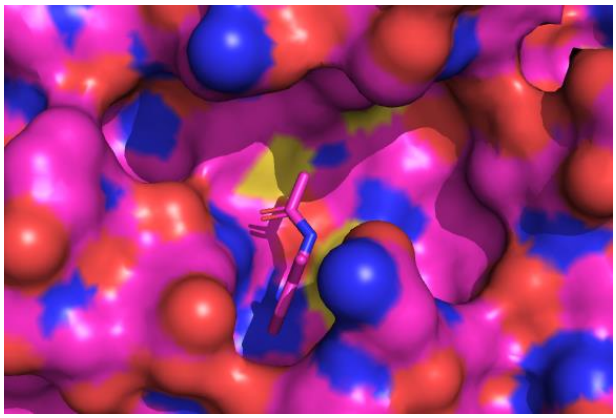
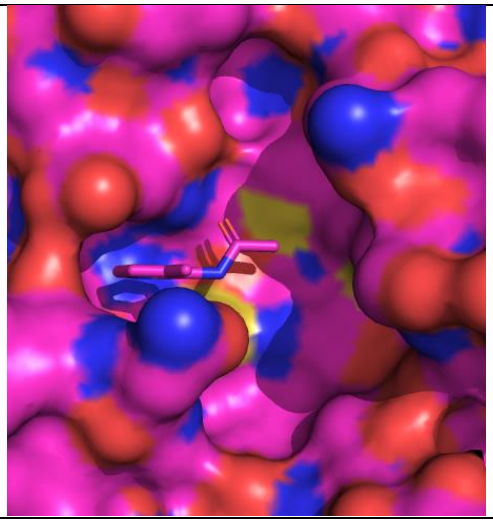
- 14 known protease inhibitors identified using ChEMBL
- Tested FAF4drugs with these 14 compounds and interesting predictions are made like bioavailability, but process is slow and need SMILES files
- Can write SMILES files for enumeration output using KNIME but with all the compounds it takes a VERY long time to do this
- Alternatively can get basic properties using KNIME, dock and then generate extra properties
- Filter properties in KNIME based on the 14 known compounds?
- Filtered based on Lipinski rules at this stage as still gives room to add mass to these potential hit compounds since the 14 known compounds are large
- FAF4 https://mobyle.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py#jobs::FAF-Drugs4.C21889833053112
- FAF4 help sheet https://fafdrugs4.rpbs.univ-paris-diderot.fr/mobyle.html

Top three fragment hits

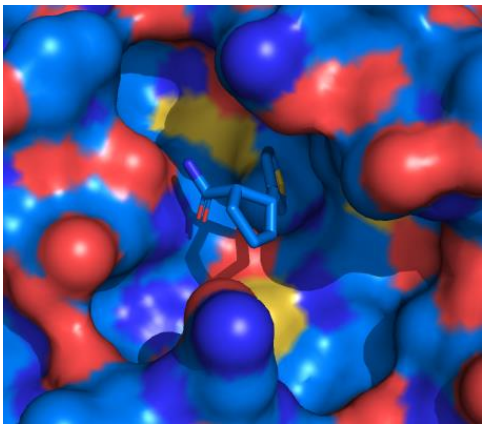
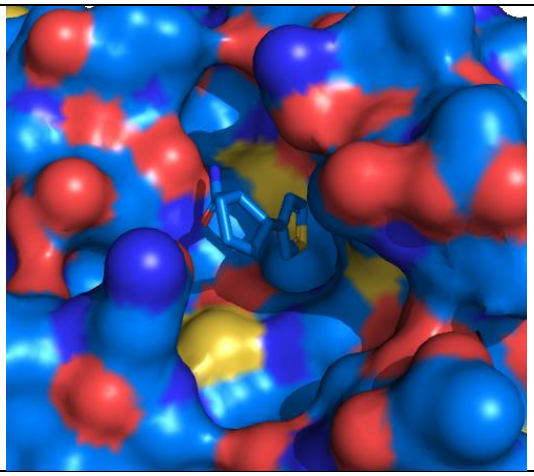
Based on a visual inspection of the fragment poses (see below) and the proposed enumeration method the following three fragments look to be the most promising: Z45617795, Z1129283193 and POB0129, as their enumeration may lead to growth into new pockets hopefully without disrupting the known binding pose, thereby maintaining this initial interaction. Z1129283193 enumeration could be the most fruitful as has space to expand into the remaining pockets. Modifications to the cyclopentyl of POB0129 may enable more binding interactions with the pocket, but this would require more synthesis (slowing the process) to faciltate these modifications.



Z45617795

Z1129283193

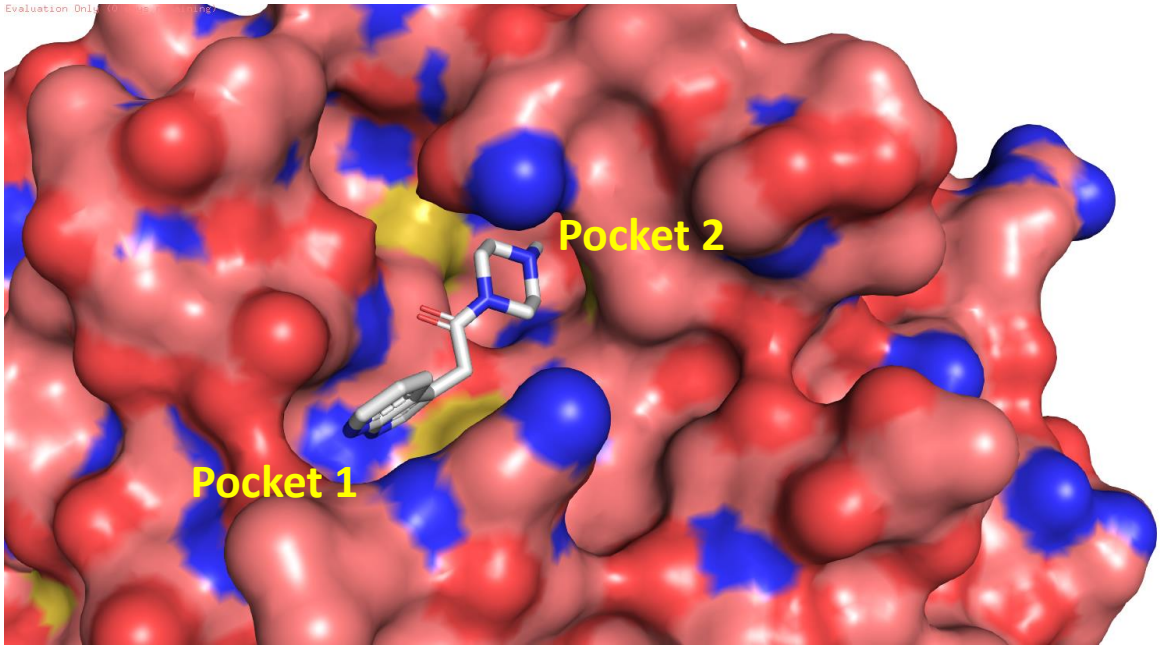



POB0129

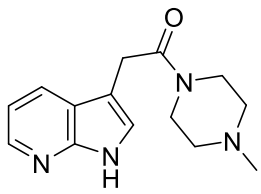
Other General Medicinal Chemistry Observations

Visually inspected the non-covalent fragment hits. Most tended to occupy two possible binding sites (indicated below with Z1400780201).

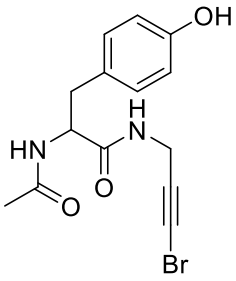
Holly Foster 26/03/2020



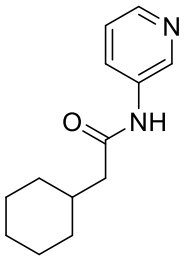
The following fragments may be superior in terms of potency as occupy both sites (see below for images): Z1400780201, NCL-00024905, Z31792168, Z44592329, Z1129283193.



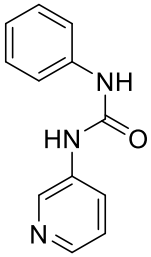
Z1400780201



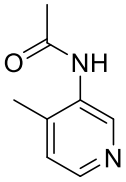
NCL-00024905



Z31792168



Z44592329

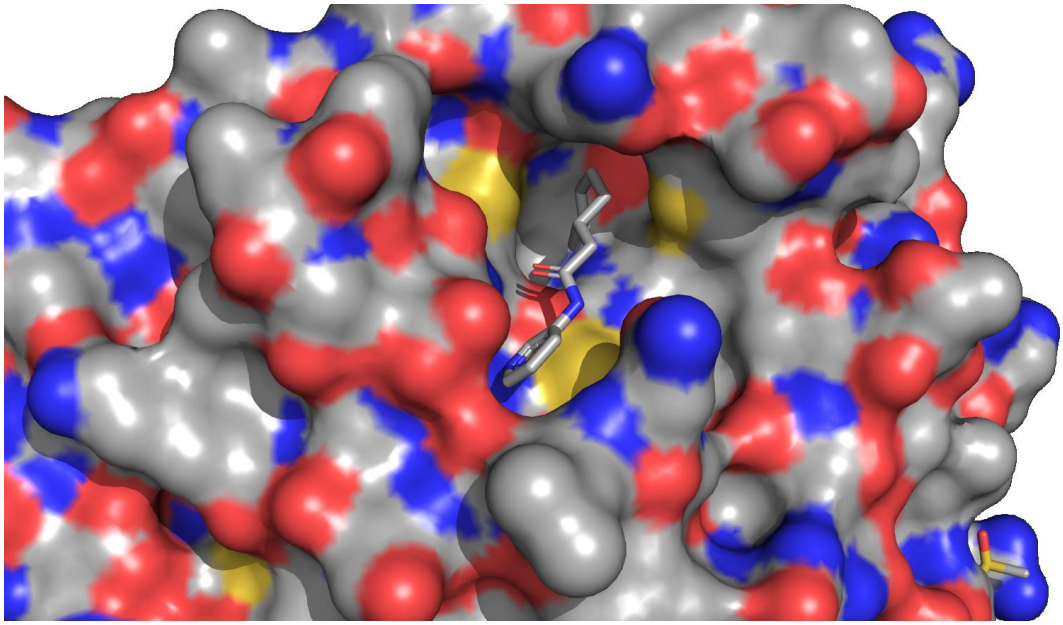


Z1129283193

All of the above fragments have an aromatic ring and amide linker which could be important motifs for binding. Modifications to the piperazine of Z1400780201 or cyclohexyl of Z31792168 or phenyl of Z31792168 may extend this fragment to unoccupied pockets and increase binding affinity.

Z31792168:

Holly Foster 26/03/2020



Z44592329: