

Supplementary Information

Applying deep learning to iterative screening of medium-sized molecules for protein–protein interaction-targeted drug discovery

Yugo Shimizu^a, Tomoki Yonezawa^a, Yu Bao^a, Junichi Sakamoto^b, Mariko Yokogawa^a,
Toshio Furuya^c, Masanori Osawa^a and Kazuyoshi Ikeda^{*a,d}

^a Division of Physics for Life Functions, Keio University Faculty of Pharmacy, 1-5-30 Shibakoen, Minato-ku, Tokyo 105-8512, Japan. *E-mail: ikeda-kz@pha.keio.ac.jp

^b Axcelead Drug Discovery Partners, Inc., 26-1, Muraoka-Higashi 2-chome, Fujisawa, Kanagawa 251-0012, Japan

^c Drug Discovery Department, Research & Development Division, PharmaDesign, Inc., Hatchobori 2-19-8, Chuo-ku, Tokyo 104-0032, Japan

^d HPC- and AI-driven Drug Development Platform Division, Center for Computational Science, RIKEN, Yokohama230-0045, Japan

Supplementary Figures

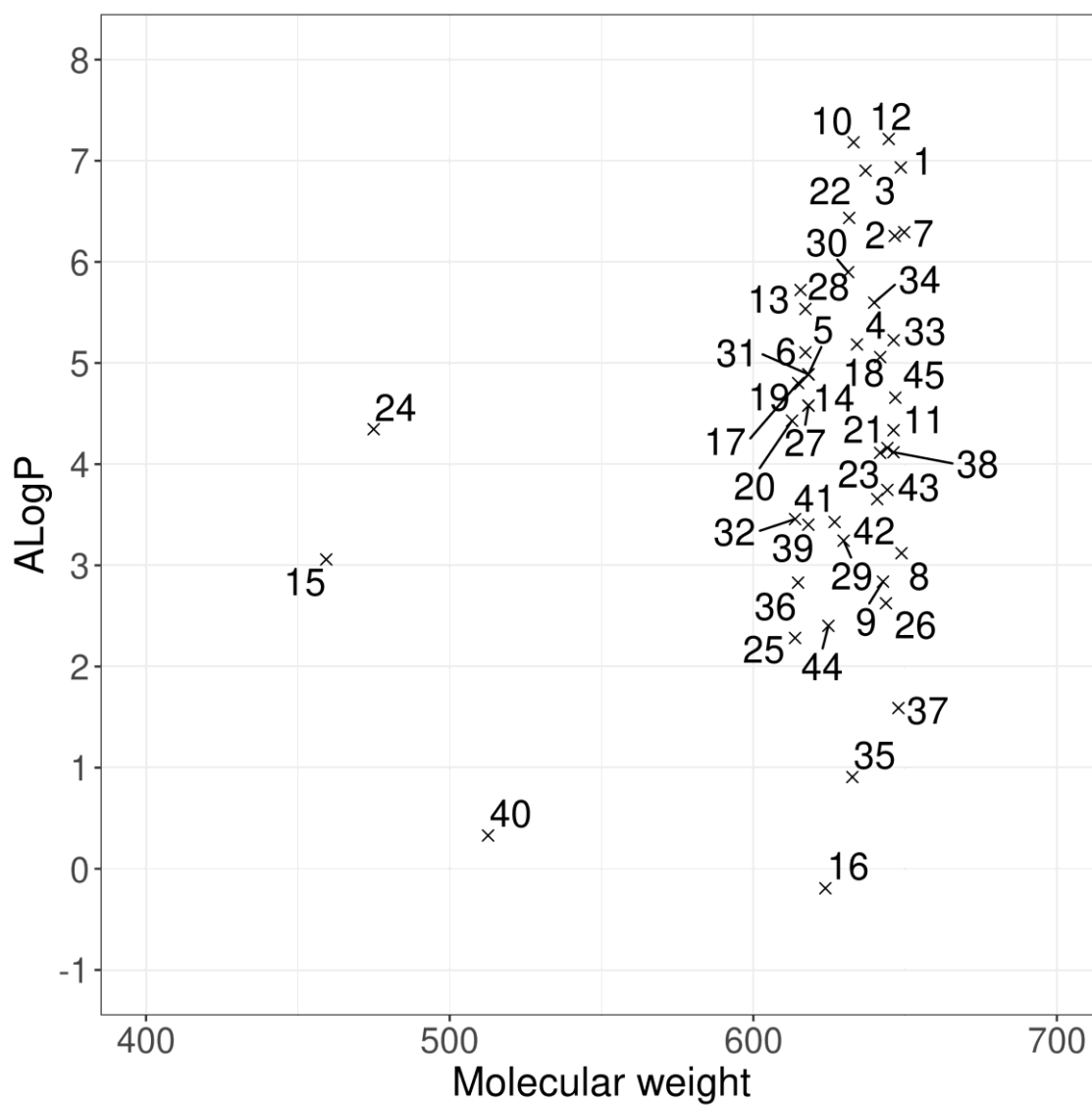
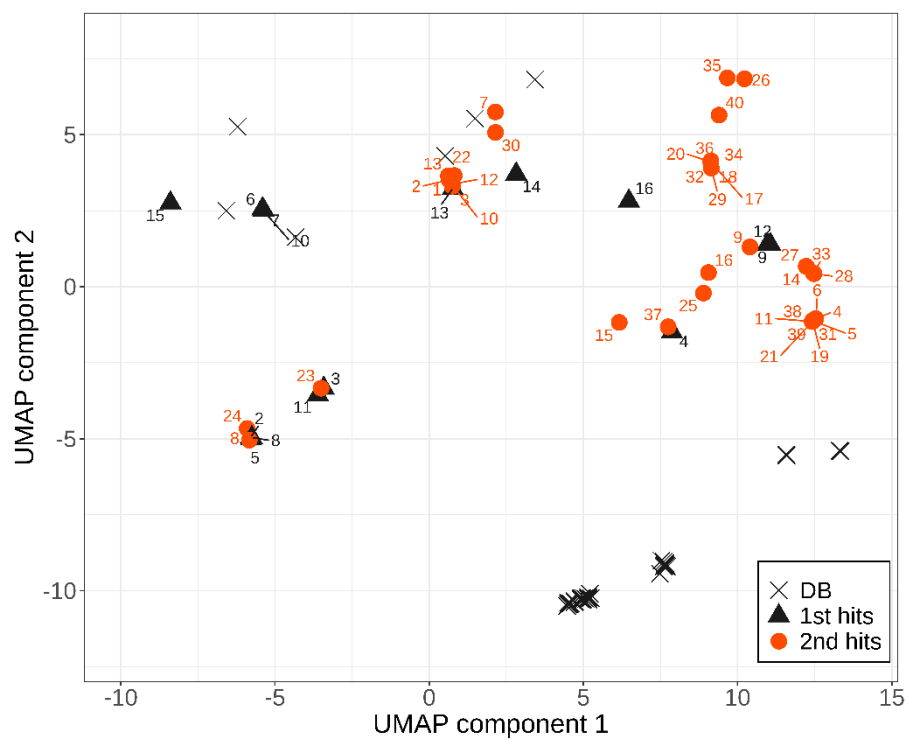


Fig. S1 Distribution of molecular properties (molecular weight and AlogP) of the hit compounds 1-45.

(A)



(B)

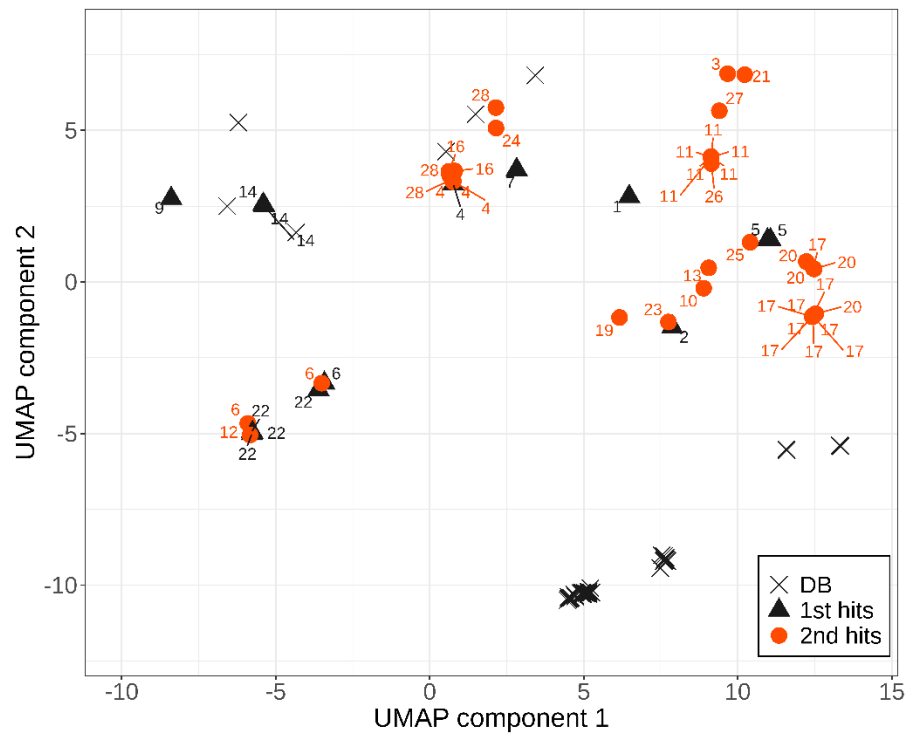


Fig. S2 Chemical space of the active compounds in the public databases and the specific hit compounds of the first and second assays with compound IDs (A) and cluster IDs (B).

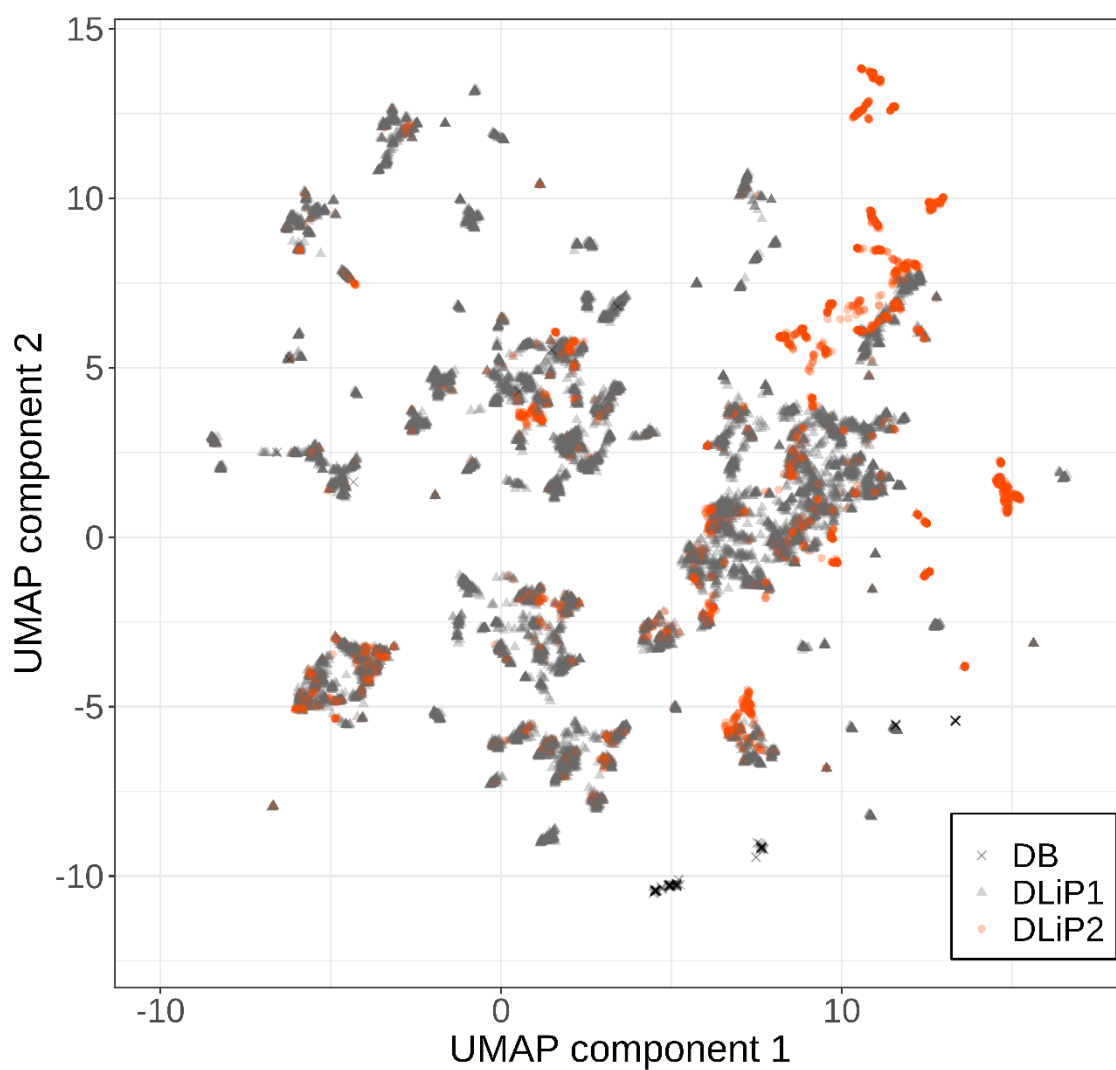


Fig. S3 Chemical space of the DLiP1 and DLiP2 library compounds and active compounds of the DB dataset.

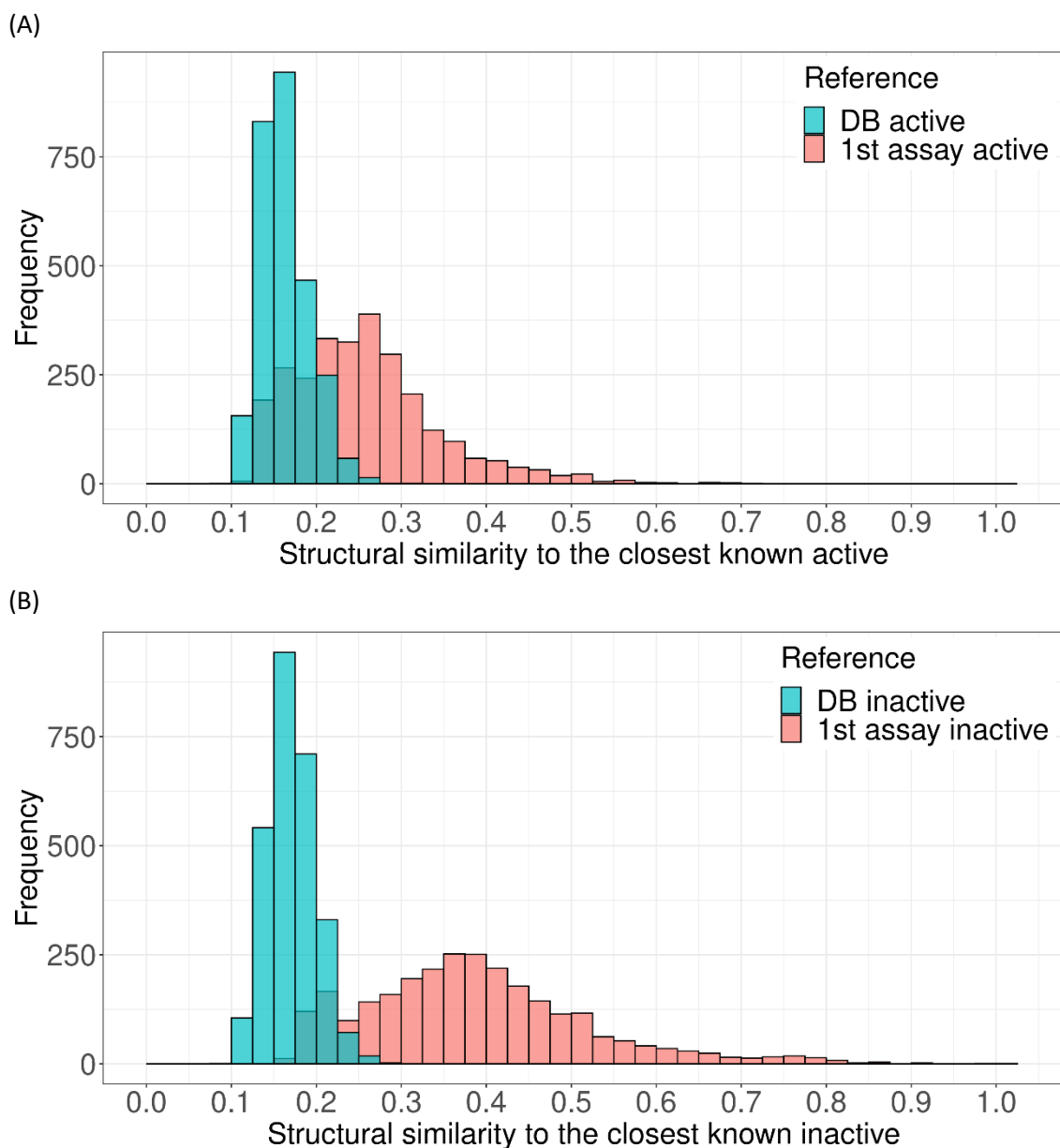


Fig. S4 Distribution of structural similarities (Tanimoto coefficient of FCFP₆) from the DLip2 library compounds to their closest known Keap1/Nrf2 PPI inhibitory (A) and non-inhibitory (B) compounds (from public databases: cyan; from our first assay: light red).

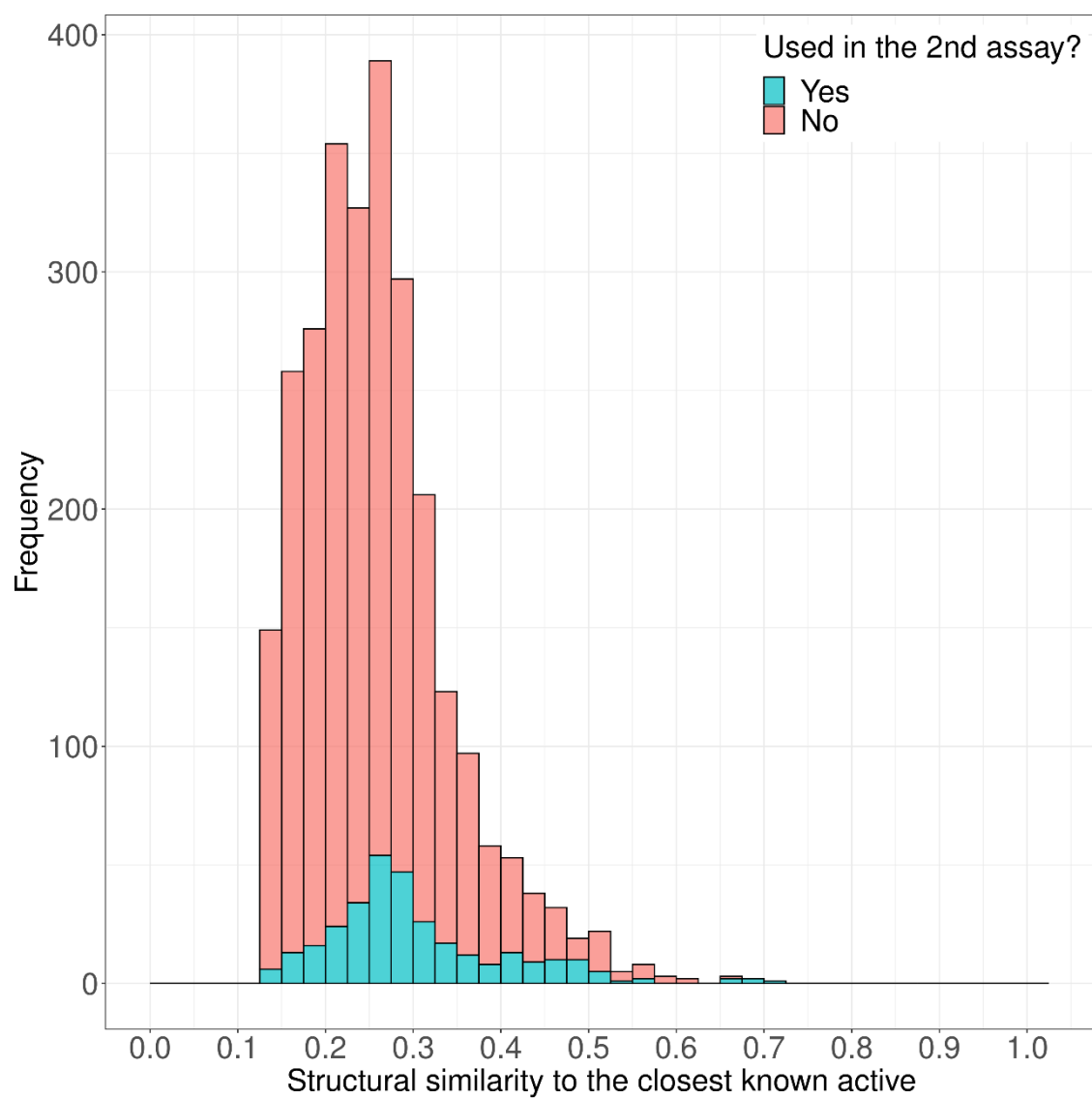


Fig. S5 Distribution of structural similarities (Tanimoto coefficient of FCFP₆) from the DLiP2 library compounds to their closest known Keap1/Nrf2 PPI inhibitory compounds. The assayed compounds and not assayed ones are shown in cyan and light red, respectively.

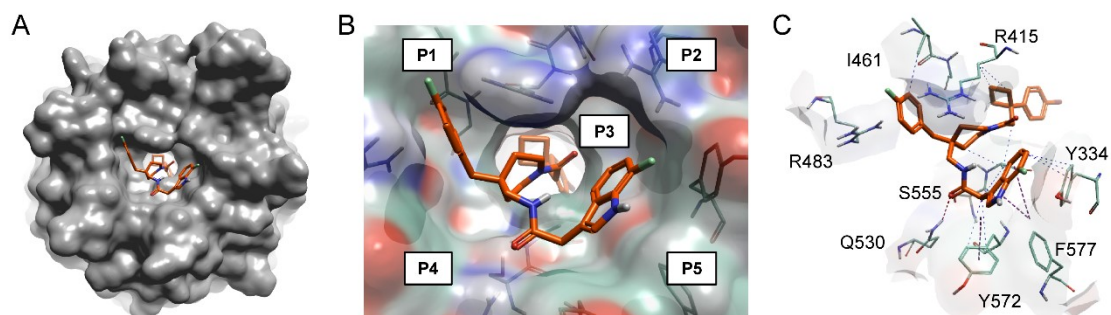


Fig. S6 Prediction of the binding pose of the hit compound **1** using docking simulation against 4XMB. (A) The whole structure of the Keap1 Kelch domain is displayed in opaque gray using a surface model. The docked structure of the compound is shown in a stick model. (B) The predicted binding pose of the compound in the ligand-binding pocket is represented using a transparent surface model. Amino acid residues in the ligand-binding site are highlighted using a thin stick model. The positions of subpockets (P1–P5) in the Kelch domain are shown. (C) The interactions of the compound with the Kelch domain are shown. Hydrogen bonds and aromatic–aromatic interactions are indicated by dashed lines.

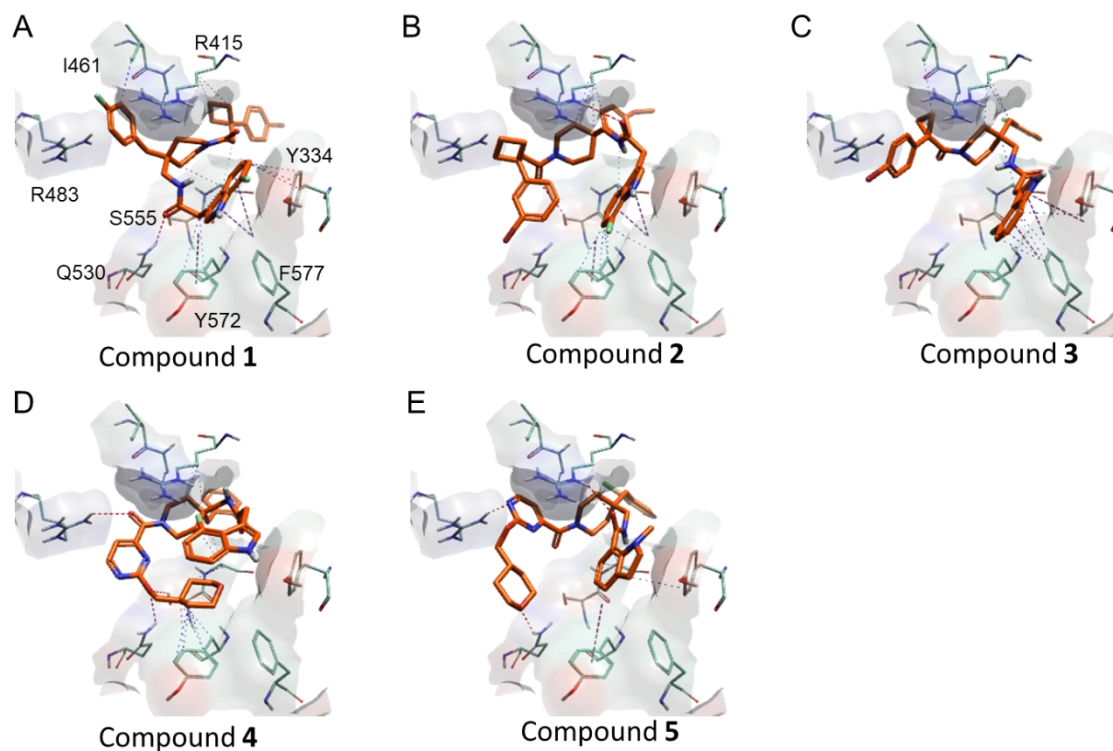


Fig. S7 Predicted binding poses of the hit compounds (compound 1–5) using docking calculations against 4XMB. Amino acid residues in the ligand-binding site are highlighted using thin stick models. The predicted binding modes of compounds are displayed by orange-colored stick models. Hydrogen bonds and aromatic–aromatic interactions are indicated by dashed lines. In these docking poses, it is commonly observed that each compound's bicyclic structure interacts with the aromatic residues (Tyr334, Tyr572, and Phe577) in the Keap1 Kelch domain.

Supplementary Tables

Table S1 (in a separate file) Structures of the hit compounds and their inhibitory rates against Keap1/Nrf2 and Bcl6/F1325. Similarities between the hit compounds and their closest known active compounds in the first assay (FA) and the DB dataset are shown. Cluster ID indicates clusters compounds belong to. DNN-hybrid, DNN-hybrid-PI, RF-FA, RF-DB-PI, and RF-DB-TI columns indicate ranks of predicted scores of the hit compounds calculated by the models. LF dG indicates estimated protein–ligand binding energy in the docking calculation. NA: not available. NI: no inhibition. NS: not selected.

Experimental

Data collection for machine learning (ML)

Three compound–activity datasets were used as training data sources for ML. The first was a database (DB) dataset used in our previous work¹, including 214 compounds with activity information against the Keap1/Nrf2 protein–protein interaction (PPI) extracted from three public databases: ChEMBL², TIMBAL³, and 2P2I.⁴ In accordance with our previous work¹, the compounds from ChEMBL and TIMBAL were defined as either “active” or “inactive” based on whether their minimum standardized activity values were $\leq 10 \mu\text{M}$ or not, respectively, while all compounds from the 2P2I database were defined as “active” regardless, resulting in 108 active and 106 inactive compounds. The second dataset was the first assay (FA) dataset that included 619 compounds with inhibition rates of the TR-FRET assay for the Keap1/Nrf2 PPI from our previous work.¹ The activity of each compound from the FA dataset was defined as “active,” “weak active,” or “inactive” based on inhibition rates of $\geq 15\%$, 5–15%, and $< 5\%$ at $100 \mu\text{M}$ concentration, respectively. The third dataset was a miscellaneous putative inactive (MISC-PI) dataset that included 12,973 compounds with minimum standardized activity values corresponding to $\leq 10 \mu\text{M}$ against at least one of 103 miscellaneous PPI targets, excluding Keap1/Nrf2. Compounds and their activity information of the MISC-PI dataset was extracted from Database of Chemical Library of Protein-Protein Interaction (DLiP).⁵

ML

1. Feature preparation

The structures of the largest fragments of compounds were used to calculate the features for ML. Two sets of features were calculated for each compound. One was a fingerprint (FP) set that included functional structural fingerprints (FCFP_6)⁶ and five simple molecular properties (molecular weight, AlogP, numbers of H acceptors, H donors, and rotatable bonds) calculated using Pipeline Pilot.⁷ The FP set was created in the same manner as that used in a previous study.¹ The other set of features was molecular descriptor set that included two-dimensional (2D) molecular descriptors calculated using CDK 1.1-SNAPSHOT,⁸ RDKit 2018.03.4.0,⁹ and Mordred 1.1.1,¹⁰ and did not include fingerprints.

2. Creating models

Classification models predicting the inhibitory activity of compounds against Keap1/Nrf2 PPI were created using random forest (RF)¹¹ and deep neural network (DNN) algorithms.¹²

RF models

Two RF models (RF-DB-TI and RF-DB-PI) were created in our previous work.¹ In this work, two RF models (RF-FA and RF-hybrid) were additionally created as well using the R package ranger 0.10.1^{13,14} as described in the previous work. The RF-DB-TI and RF-DB-PI models were created as two-class classification (active or inactive) models, while the RF-FA and RF-hybrid models were created as three-class classification (active, weak active, or inactive) models.

DNN models

To create balanced datasets (*i.e.*, the number of positive data equals the number of negative data) and increase the amount of data as appropriate for deep learning, a data rebalancing procedure was utilized. SMOTE¹⁵ was used as the oversampling technology, and random sampling with nearest neighbor clustering¹⁶ was used as the undersampling technology. The training dataset for the DNN-hybrid model was rebalanced into 3,000 positives and 3,000 negatives and that for the DNN-hybrid-PI model was rebalanced into 4,000 positives and 4,000 negatives. Normalization was performed using a scaler fitted to the training dataset. Features with all NaN values were removed. The correlation coefficients between all pairs of features were calculated, and when the feature pair had a correlation coefficient > 0.9, one of the pair was randomly selected and removed. In the DNN models, five layers, including an input layer, three hidden layers, and an output layer, were implemented to classify compounds into active or inactive. The size of the input layer was equal to the size of the input feature, the size of the second layer (*i.e.*, the first hidden layer) was equal to the size of the input layer, the size of the third and fourth layers was fixed at 200, and the size of the fifth layer (*i.e.*, the output layer) was fixed at 2. The neural network was implemented using Keras 2.3.1,¹⁷ Python 3.7, and TensorFlow 2.0.¹⁸ Early stopping¹⁹ was utilized in the neural network implementation with patience of 10, where 1/5 of the original FA datasets (*i.e.*, datasets before data rebalancing) were used as validation sets and were not included in the training sets. The advantage of using early stopping is that it can reduce overfitting during the learning process by detecting the appropriate time to stop the training process. L2 norm regularization was utilized in the implemented neural network with a parameter of 0.01. This regularization also helps reduce overfitting.

Molecular diversity

1. Diversity in molecular shapes

The principal moments of inertia (PMI) were employed to calculate the diversity of the molecular shapes in the library compounds. Three-dimensional (3D) structures (single conformations) were generated for the library compounds using Omega2 software.²⁰ The PMI values and normalized principal moment ratios of 1 and 2 (NPR1 and NPR2) were calculated using the RDKit.⁹ The two

normalized ratios were plotted on a 2D triangular graph and used to compare the diversity of the molecular shapes of a given group of compounds.

2. Diversity in structural fingerprints

Calculation of Similarity/Distance and clustering

Tanimoto similarity/Jaccard distance of fingerprints (FCFP₆) was used to calculate structural similarity/distance between compounds. Hit compounds of our assays were clustered using a Cluster Molecules component in Pipeline Pilot.⁷ The clustering was performed using FCFP₆ distance and a MaximumDissimilarity center selection method with MaximumDistance parameter 0.5.

Visualization of chemical space

Chemical space in terms of structural diversity was visualized using uniform manifold approximation and projection (UMAP). UMAP is a dimension reduction technique that can project close points in a high dimensional space onto close points in a low dimensional space.²¹ We used UMAP-learn 0.5.3²² with the Jaccard metric to transform FCFP₆ fingerprint bit arrays of compounds into 2D components (UMAP component 1 and 2).

Docking

Docking calculations were performed using Flare v6.0.²³ We used a 3D structure of the Keap1 protein (4XMB)²⁴ from PDB as the target protein for the docking calculations. For the ligand structure, the SD files of compounds were used as input, and the 3D structure generation function implemented in Flare was used to generate the ligand conformations. The calculation mode of docking in Flare was selected as "very accurate but slow," and the maximum number of poses was set to 50. The docking grid definition was set using 41P (2,2'-(naphthalene-1,4-diylbis(((4-methoxyphenyl)sulfonyl)azanediyl))diacetamide). The average docking score (LF dG) of five inactive compounds against Keap1/Nrf2 PPI randomly selected from the ChEMBL database was -9.919 (SD 2.085).

Bioassay materials

Library compounds were synthesized in the AMED project. The hKeap1 Kelch domain (residues Ala321–Thr609), hBcl6 domain (residues Ala5–Glu129), Nrf2 (TAMRA-LQLDEETGEFLPIQ-NH2) peptide, and F1325 (TAMRA-Abu(4)-VWYTDIRMRDWM) peptide were prepared in the same protocol to our previous work.¹

Synthetic procedures of the hit compounds

The tested compounds were synthesized by combinatorial chemistry. The building blocks (BBs) of the hit compounds (**1–45**) are shown in Table S2. Their synthetic procedures for connecting BBs to obtain the compounds (**1–45**) were classified into five reaction types: A–E (Table S2).

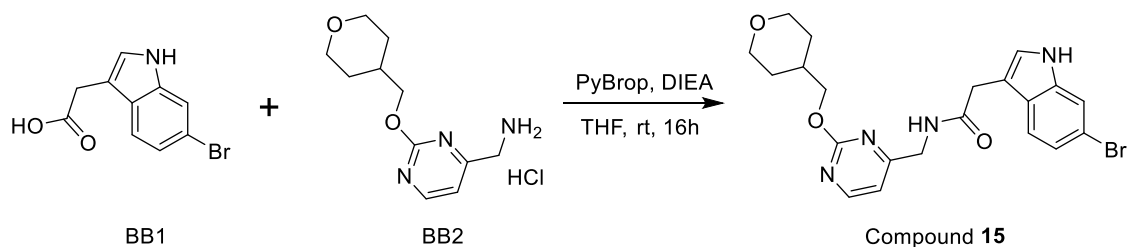
Table S2 The reaction types and BBs used in the synthesis of the hit compounds (**1–45**).

Compound	BBs			Reaction type
1	K-050-4148	K-032-382	K-050-3546	B
2	K-050-4148	K-045-706	K-050-3545	B
3	K-050-4145	K-032-384	K-050-048	B
4	K-050-4145	K-032-383	K-013-2655	B
5	K-013-2751	K-032-383	K-013-2655	B
6	K-013-2750	K-032-383	K-013-1971	B
7	K-050-4168	K-045-692	K-050-1472	B
8	K-013-2078	K-008-301	K-050-3286	B
9	K-013-2690	K-028-243	K-002-430	B
10	K-050-4159	K-032-384	K-050-048	B
11	K-013-1955	K-032-383	K-013-2655	B
12	K-050-4159	K-032-382	K-050-3546	B
13	K-050-4700	K-045-705	K-050-054	B
14	K-013-2750	K-032-384	K-013-2666	B
15	K-050-4155	K-049-625		A
16	K-002-324	K-002-098	K-013-2740	E
17	K-050-4637	K-032-394	K-013-1998	B
18	K-050-070	K-032-394	K-013-1998	B
19	K-013-2740	K-032-383	K-013-2655	B
20	K-050-4638	K-032-394	K-013-1998	B
21	K-013-9002	K-032-383	K-013-2655	B
22	K-050-4700	K-032-382	K-050-3546	B
23	K-049-701	K-008-340	K-050-3570	B
24	K-008-340	K-050-253		D
25	K-013-2646	K-045-612	K-002-427	B
26	K-013-2655	K-050-3552	K-054-0040	B
27	K-013-2749	K-032-384	K-013-2666	B
28	K-013-2750	K-032-385	K-013-1893	B
29	K-013-2691	K-032-393	K-013-2644	B
30	K-050-4613	K-032-392	K-014-138	C

31	K-013-2749	K-032-383	K-013-2655	B
32	K-044-047	K-032-394	K-013-1998	B
33	K-013-2446	K-032-385	K-013-1893	B
34	K-050-3547	K-032-394	K-013-1998	B
35	K-013-1218	K-013-2009	K-054-0046	B
36	K-044-050	K-032-394	K-013-1998	B
37	K-008-421	K-013-2643	K-002-148	C
38	K-013-1960	K-032-383	K-013-2655	B
39	K-013-9001	K-032-383	K-013-2655	B
40	K-013-2642	K-050-3557	K-054-0040	B
41	K-008-2586	K-002-322	K-013-514	B
42	K-050-3513	K-045-612	K-002-427	B
43	K-013-1419	K-032-383	K-013-2655	B
44	K-008-2570	K-002-319	K-013-1935	B
45	K-013-1897	K-050-2521	K-028-182	B

The synthetic procedures of the five reaction types are shown below.

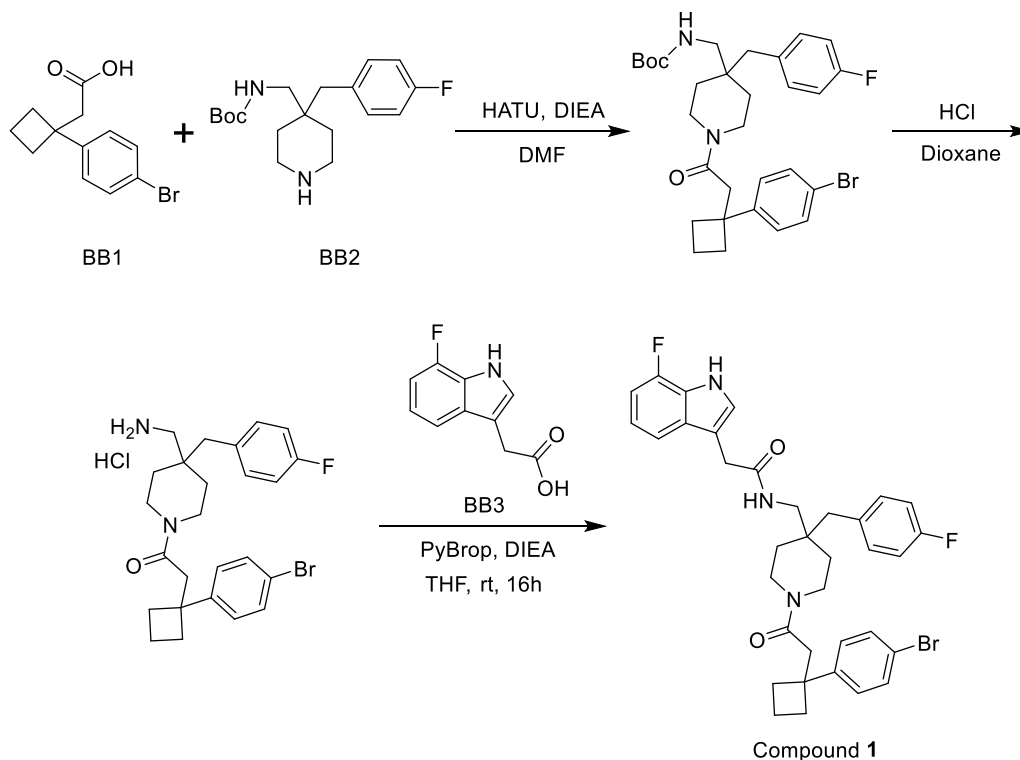
The synthetic procedures of reaction type A (Compound **15**): connecting two BBs.



Scheme S1 Synthesis of compound **15**.

To a stirred mixture of 2-(6-bromo-1*H*-indol-3-yl)acetic acid (25.4 mg, 0.100 mmol, 1.00 equiv) and 2-((tetrahydro-2*H*-pyran-4-yl)methoxy)pyrimidin-4-yl)methanamine hydrochloride (26.0 mg, 0.100 mmol, 1 equiv) in THF (2 mL) was added PyBrop (70mg, 0.15 mmol, 1.5 equiv) and DIEA (38 mg, 0.29 mmol, 2.91 equiv) at rt. The reaction was stirred at rt for 16 hours. The resulting mixture was concentrated under vacuum. The residue was purified by Prep-HPLC with the following conditions: Column, Xselect CSH OBD Column 30*150mm 5um, n; mobile phase, Water (0.1%FA) and ACN (30% Phase B up to 60% in 10 min), Detector, UV, 254 nm, to afford 2-(6-bromo-1*H*-indol-3-yl)-*N*-((2-((tetrahydro-2*H*-pyran-4-yl)methoxy)pyrimidin-4-yl)methyl)acetamide (5.1 mg, 11.1%) as a light brown solid.

The synthetic procedures of reaction type B (Compound 1): connecting three BBs.

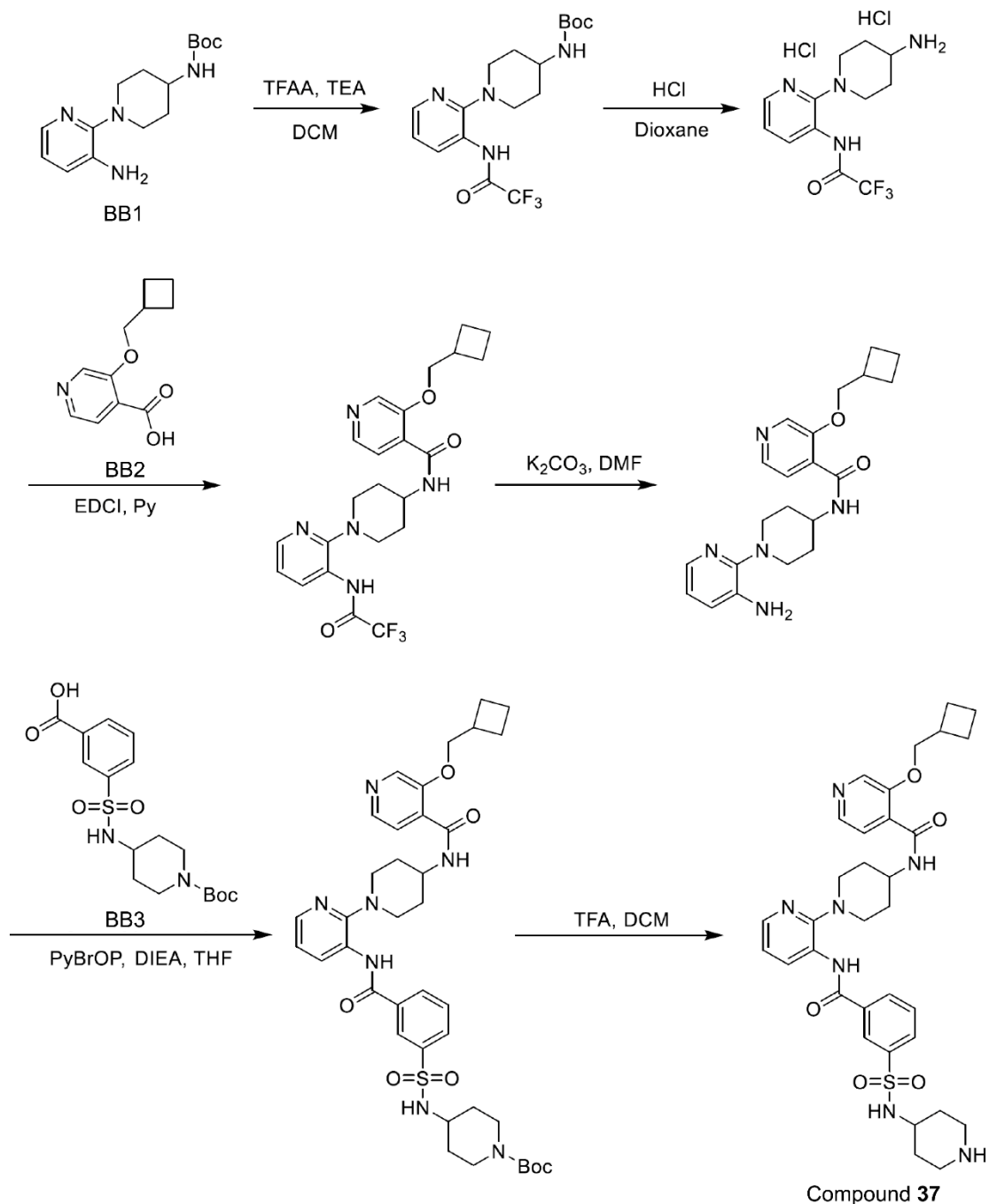


Scheme S2 Synthesis of compound 1.

To a stirred solution of 2-(1-(4-bromophenyl)cyclobutyl)acetic acid (26.9 mg, 0.100 mmol, 1.00 equiv) and *tert*-butyl ((4-(4-fluorobenzyl)piperidin-4-yl)methyl)carbamate (32.24 mg, 0.100 mmol, 1 equiv) in DMF (2 mL) was added DIEA (38.89 mg, 0.300 mmol, 3 equiv) and HATU (57.20 mg, 0.150 mmol, 1.5 equiv). The resulting mixture was stirred at rt for 16 hours. The resulting mixture was concentrated under reduced pressure. The residue was applied onto a silica gel column with dichloromethane/methanol (10:1~5:1). This resulted in *tert*-butyl ((1-(2-(1-(4-bromophenyl)cyclobutyl)acetyl)-4-(4-fluorobenzyl)piperidin-4-yl)methyl)carbamate (25mg, 43.59%). To a stirred solution of *tert*-butyl ((1-(2-(1-(4-bromophenyl)cyclobutyl)acetyl)-4-(4-fluorobenzyl)piperidin-4-yl)methyl)carbamate (25 mg, 0.044 mmol, 1.00 equiv) in 1,4-dioxane (1 mL) was added HCl (4M in 1,4-dioxane) (1 mL) at rt. The resulting mixture was stirred at rt for 2 hours. The resulting mixture was concentrated under reduced pressure. The residue in 1-(4-(aminomethyl)-4-(4-fluorobenzyl)piperidin-1-yl)-2-(1-(4-bromophenyl)cyclobutyl)ethan-1-one hydrochloride (24 mg). To a stirred mixture of 1-(4-(aminomethyl)-4-(4-fluorobenzyl)piperidin-1-yl)-2-(1-(4-bromophenyl)cyclobutyl)ethan-1-one hydrochloride (24 mg, 0.047 mmol, 1.00 equiv) and 2-(7-fluoro-1*H*-indol-3-yl)acetic acid (9.1 mg, 0.047 mmol, 1.0 equiv) in THF (2 mL) was added PyBrop (32.6 mg, 0.07 mmol, 1.5 equiv) and DIEA (19.4 mg, 0.15 mmol, 3.19 equiv) at rt. The reaction was stirred at rt for 16 hours. The resulting mixture was concentrated under vacuum. The residue was purified by Prep-HPLC with the following conditions: Column, Xselect CSH OBD

Column 30*150mm 5um; mobile phase, Water (0.1%FA) and ACN (30% Phase B up to 60% in 10 min), Detector, UV, 254 nm, to afford 1-(2-(4-(3-chlorobenzyl)tetrahydro-2H-pyran-4-carboxamido)-3-fluorophenyl)piperidine-4-carboxylic acid (6.3 mg, 20.67%) as a light brown solid.

The synthetic procedures of reaction type C (Compound **37**): deprotection of an amine after connecting three BBs.

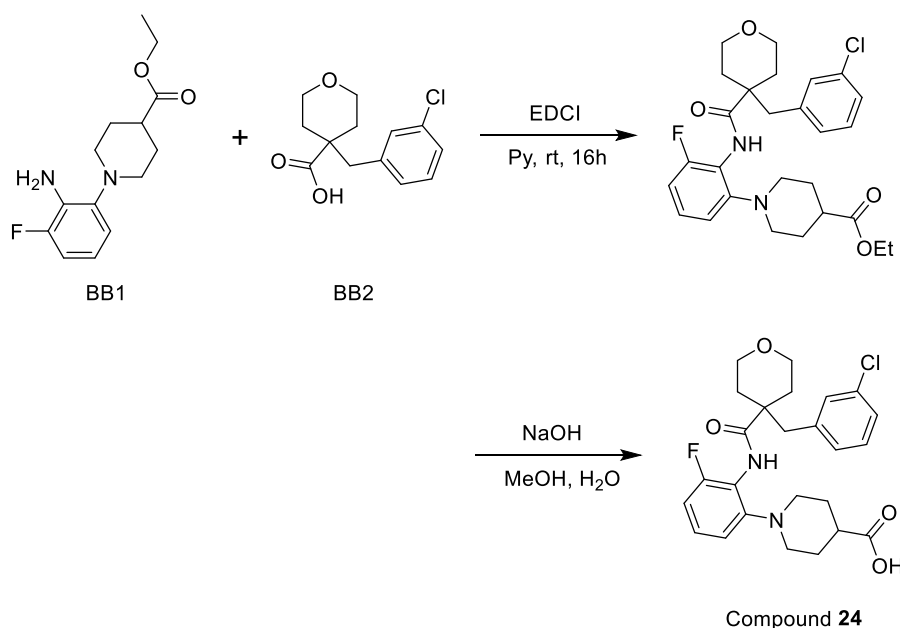


Scheme S3 Synthesis of compound **37**.

Into a 10-mL round bottom flask, a solution of *tert*-butyl (1-(3-aminopyridin-2-yl)piperidin-4-yl)carbamate (100 mg, 0.34 mmol, 1.00 equiv) in DCM (5 mL) was added dropwise TEA (68.8 mg, 0.68 mmol, 2.00 equiv) and TFAA (107 mg, 0.51 mmol, 1.50 equiv) at 0°C under N₂ atmosphere. The reaction mixture was stirred for 16 hours at room temperature. The reaction was washed with brine (3x5mL), dried over anhydrous Na₂SO₄, and concentrated under vacuum. The residue was applied onto a silica gel column with dichloromethane/methanol (10:1). This resulted in 105 mg of *tert*-butyl (1-(3-(2,2,2-trifluoroacetamido)pyridin-2-yl)piperidin-4-yl)carbamate. To a stirred solution of *tert*-butyl (1-(3-(2,2,2-trifluoroacetamido)pyridin-2-yl)piperidin-4-yl)carbamate (105 mg, 0.27 mmol, 1.00 equiv) in 1,4-dioxane (1 mL) was added HCl (4M in 1,4-dioxane) (1 mL) at rt. The resulting mixture was stirred at rt for 2 hours. The resulting mixture was concentrated under reduced pressure. The residue in *N*-(2-(4-aminopiperidin-1-yl)pyridin-3-yl)-2,2,2-trifluoroacetamide dihydrochloride (~98 mg). To a stirred solution of *N*-(2-(4-aminopiperidin-1-yl)pyridin-3-yl)-2,2,2-trifluoroacetamide dihydrochloride (98 mg, 0.27 mmol, 1 equiv) and 3-(cyclobutylmethoxy)isonicotinic acid (56 mg, 0.27 mmol, 1 equiv) in Py (2 mL) was added EDCI (104 mg, 0.54 mmol, 2.0 equiv) at rt. The reaction was stirred at rt for 16 hours. The resulting mixture was concentrated under vacuum. The residue was applied onto a silica gel column with dichloromethane/methanol (10:1). This resulted in 3-(cyclobutylmethoxy)-*N*-(1-(3-(2,2,2-trifluoroacetamido)pyridin-2-yl)piperidin-4-yl)isonicotinamide (~90 mg). To a stirred solution of 3-(cyclobutylmethoxy)-*N*-(1-(3-(2,2,2-trifluoroacetamido)pyridin-2-yl)piperidin-4-yl)isonicotinamide (90 mg, 0.19 mmol, 1 equiv) in DMF (2 mL) was added K₂CO₃ (53 mg, 0.38 mmol, 2.0 equiv) at rt. The reaction was stirred at rt for 16 hours. The resulting mixture was concentrated under vacuum. The residue was applied onto a silica gel column with dichloromethane/methanol (5:1). This resulted in *N*-(1-(3-aminopyridin-2-yl)piperidin-4-yl)-3-(cyclobutylmethoxy)isonicotinamide (~51 mg). To a stirred mixture of *N*-(1-(3-aminopyridin-2-yl)piperidin-4-yl)-3-(cyclobutylmethoxy)isonicotinamide (51 mg, 0.13 mmol, 1.00 equiv) and 3-(*N*-(1-(*tert*-butoxycarbonyl)piperidin-4-yl)sulfamoyl)benzoic acid (50 mg, 0.13 mmol, 1.0 equiv) in THF (2 mL) was added PyBrop (93 mg, 0.20 mmol, 1.5 equiv) and DIEA (50 mg, 0.39 mmol, 3 equiv) at rt. The reaction was stirred at rt for 16 hours. The resulting mixture was concentrated under vacuum. The residue was applied onto a silica gel column with dichloromethane/methanol (5:1), to afford *tert*-butyl 4-((3-((2-(4-(3-(cyclobutylmethoxy)isonicotinamido)piperidin-1-yl)pyridin-3-yl)carbamoyl)phenyl)sulfonamido)piperidine-1-carboxylate (~40 mg). To a stirred solution of *tert*-butyl 4-((3-((2-(4-(3-(cyclobutylmethoxy)isonicotinamido)piperidin-1-yl)pyridin-3-yl)carbamoyl)phenyl)sulfonamido)piperidine-1-carboxylate (40 mg, 0.053mmol, 1.00 equiv) in DCM (1 mL) was added TFA (1 mL) at rt. The resulting mixture was stirred at rt for 2 hours. The resulting mixture was concentrated under reduced pressure. The residue was purified by Prep-HPLC with the following conditions: Column: Shim-pack XR-ODS,3.0*50 mm,2.2 um; Mobile

Phase A: Water 0.05%TFA, Mobile Phase B: ACN 0.05%TFA; Flow rate: 1.2 mL/min; Gradient: 5%B to 100%B in 2.2 min, hold 1.0 min; Detector, UV, 254 nm, to afford 3-(cyclobutylmethoxy)-*N*-(1-(3-(3-(*N*-(piperidin-4-yl)sulfamoyl)benzamido)pyridin-2-yl)piperidin-4-yl)isonicotinamide (4.6 mg, 13.30%) as an off-white solid.

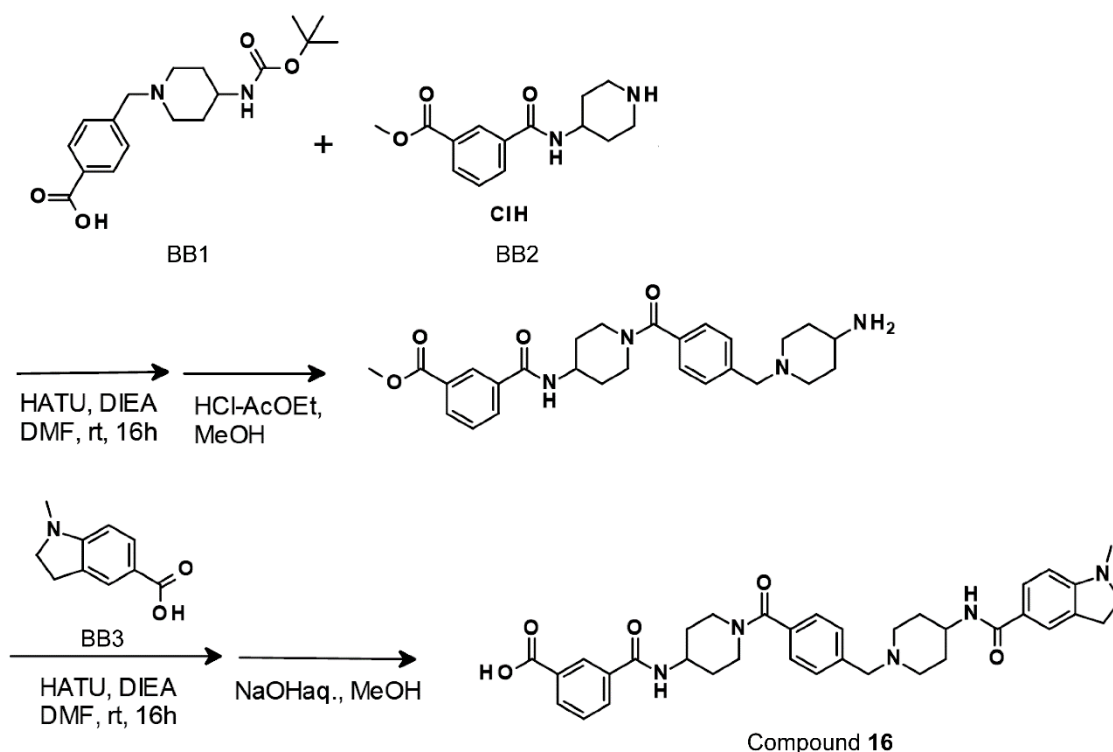
The synthetic procedures of reaction type D (Compound **24**): hydrolysis of an ester after connecting two BBs.



Scheme S4 Synthesis of compound **24**.

To a stirred mixture of ethyl 1-(2-amino-3-fluorophenyl)piperidine-4-carboxylate (26.63 mg, 0.100 mmol, 1.00 equiv) and 4-(3-chlorobenzyl)tetrahydro-2*H*-pyran-4-carboxylic acid (25.47 mg, 0.100 mmol, 1 equiv) in Pyridine (2 mL) was added EDCI (38.34 mg, 0.2 mmol, 2 equiv). The resulting mixture was stirred at rt for 16 hours. The resulting mixture was concentrated under vacuum. This resulted in 100mg crude ethyl 1-(2-(4-(3-chlorobenzyl)tetrahydro-2*H*-pyran-4-carboxamido)-3-fluorophenyl)piperidine-4-carboxylate. To a stirred mixture of ethyl 1-(2-(4-(3-chlorobenzyl)tetrahydro-2*H*-pyran-4-carboxamido)-3-fluorophenyl)piperidine-4-carboxylate (100 mg crude) in CH₃OH (1 mL) was added NaOH (20 mg, 0.5 mmol, 5 equiv) in 1ml H₂O at room temperature. The resulting mixture was stirred overnight at room temperature. The resulting mixture was concentrated under vacuum. The residue was purified by Prep-HPLC with the following conditions: Column, Xselect CSH OBD Column 30*150mm 5um; mobile phase, Water(0.1%FA) and ACN (30% Phase B up to 60% in 10 min), Detector, UV, 254 nm, to afford 1-(2-(4-(3-chlorobenzyl)tetrahydro-2*H*-pyran-4-carboxamido)-3-fluorophenyl)piperidine-4-carboxylic acid (5.3 mg, 11.12%) as an off-white solid.

The synthetic procedures of reaction type E (Compound **16**): hydrolysis of an ester after connecting three BBs.



Scheme S5 Synthesis of compound **16**.

To a stirred solution of 4-[[4-[[[1,1-dimethylethoxy]carbonyl]amino]-1-piperidinyl]methyl]benzoic acid (10.0 mg, 0.03 mM, 1.0 equiv) and methyl 3-[[4-piperidinylamino]carbonyl]benzoate hydrochloride (9.0 mg, 0.03 mM, 1.0equiv) in DMF (1 mL) was added HATU (17mg, 0.045 mM, 1.5 equiv) and DIEA (15.5 mg, 0.12 mM, 4 equiv) at rt. The reaction was stirred at rt for 16 hours. The resulting mixture was concentrated under reduced pressure. To a stirred solution of the residue was added methanol (0.5mL) and HCl (4M in ethyl acetate) (0.5mL) at rt. The reaction was stirred at rt for 1 hour. The resulting mixture was concentrated under reduced pressure. The residue was purified by Prep-HPLC with the following conditions: Column, SunFire Prep C18 OBD Column, 19*100mm, 5 μ m; mobile phase, Water (0.1% HCO₂H) and CH₃CN (0.1% HCO₂H) (5% Phase B up to 25% in 10 min); Detector, MS, to afford methyl 3-[[4-4-[[4-amino-1-piperidinyl]methyl]benzoyl] piperidinylamino]carbonyl]benzoate (3.6mg, 25.1%) as a colorless semi-solid. To a stirred solution of methyl 3-[[4-4-[[4-amino-1-piperidinyl]methyl]benzoyl] piperidinylamino]carbonyl]benzoate (3.6mg, 0.0075mM, 1.00 equiv) and 1-methylindoline-5-carboxylic acid (5.3 mg, 0.03 mM, 4.0equiv) in DMF (0.5 mL) was added HATU (17mg, 0.045 mM, 6.0 equiv) and DIEA (15.5 mg, 0.12 mM, 16 equiv) at rt. The reaction was stirred at rt for 16 hours. The resulting mixture was concentrated under reduced pressure. To a stirred solution of the residue was added methanol (0.5mL) and NaOH (1M in H₂O)

(0.5mL) at rt. The reaction was stirred at rt for 12 hours and then added HCl (1M in H₂O) (0.5mL) at rt. The resulting mixture was concentrated under reduced pressure. The residue was purified by Prep-HPLC with the following conditions: Column, SunFire Prep C18 OBD Column, 19*100mm, 5µm; mobile phase, Water (0.1% HCO₂H) and CH₃CN (0.1% HCO₂H) (5% Phase B up to 25% in 10 min); Detector, UV, MS, to afford 3-((4-(4-((4-(1-methylindoline-5-carboxamido)piperidin-1-yl)methyl)benzoyl)piperidin-1-yl)carbamoyl)benzoic acid (1.4mg, 29.9%) as a colorless semi-solid.

The identity of the compounds was confirmed, based on the MS data of the hit compounds (Table S3) and the MS/NMR data of the BBs used for their synthesis (Tables S4).

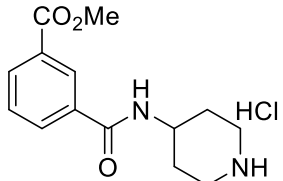
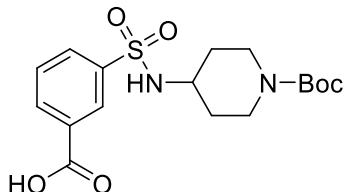
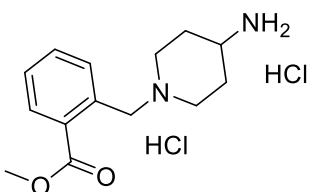
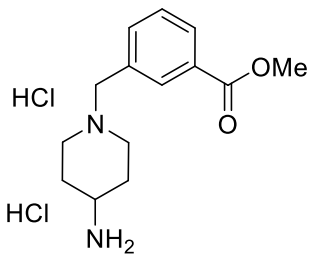
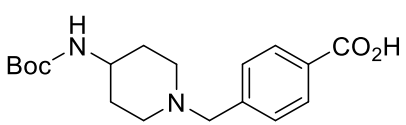
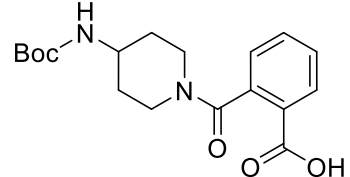
Table S3 MS data and HPLC purity of the hit compounds (**1–45**).

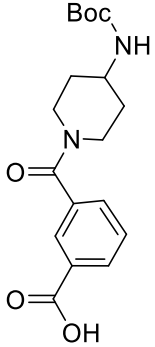
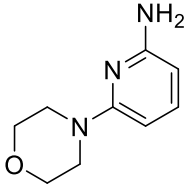
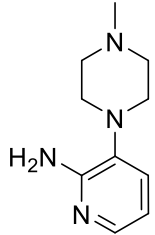
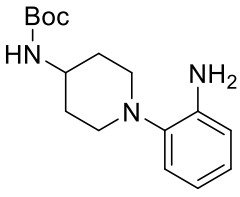
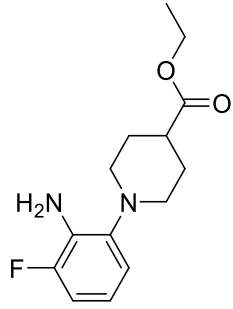
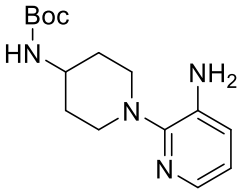
Compound	Molecular weight	Target MS [Exact MS+H]	Purity (%)
1	648.6	648&650	99.24
2	646.6	646&648	93.78
3	637.0	636&638	98.63
4	634.2	634	98.85
5	618.2	618	99.82
6	617.2	617	92.86
7	649.3	650	99.03
8	648.3	649	92.16
9	642.7	643	99.63
10	633.0	632&634	97.22
11	646.2	646	99.63
12	644.6	644&646	92.04
13	614.2	615&617	98.03
14	618.2	618	95.88
15	459.3	459&461	95.71
16	623.7	623	100.00
17	614.8	615	95.93
18	641.8	642	92.73
19	618.2	618	94.18
20	612.3	613	98.92
21	644.2	644	92.53
22	631.6	631&633	99.29

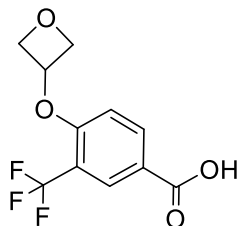
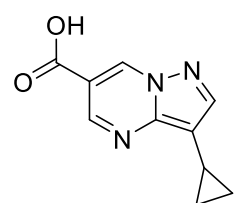
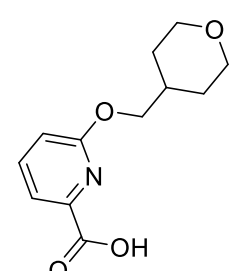
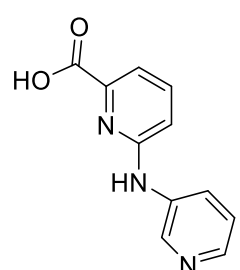
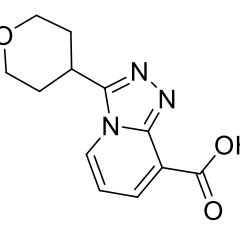
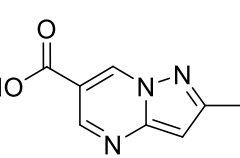
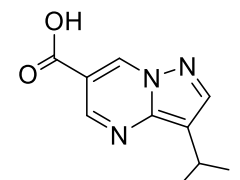
23	641.8	642	98.79
24	475.0	475	98.13
25	613.3	614	97.78
26	643.7	644	98.82
27	618.2	618	97.45
28	617.2	617	91.74
29	629.8	630	99.31
30	631.3	631	98.82
31	618.2	618	97.22
32	613.3	614	97.39
33	646.2	646	96.70
34	639.8	640&642	93.68
35	632.6	633	98.50
36	614.7	615	97.33
37	647.3	648	99.42
38	646.2	646	99.79
39	618.1	618	99.45
40	512.3	513	97.82
41	626.4	627	96.81
42	640.4	641	95.24
43	644.2	644	99.32
44	624.3	625	96.52
45	646.3	647	93.23

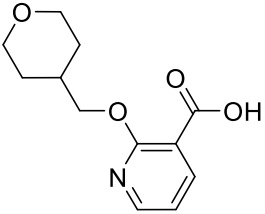
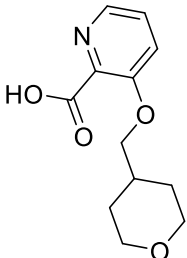
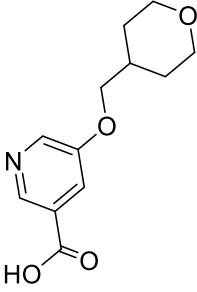
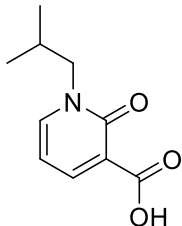
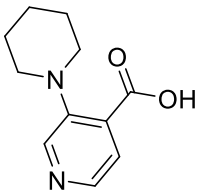
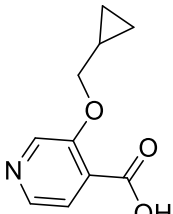
The purity and MS information of the hit compounds (**1–15**, **17–45**) were obtained via LC-MS on a Shimadzu LCMS-2020 [gradient from 5% MeCN/95% H₂O to 100% MeCN/0% H₂O (+0.05% trifluoroacetic acid) in 2 min with a Shim-pack XR-ODS column or gradient from 10% MeCN/90% H₂O to 95% MeCN/5% H₂O (+5mM NH₄HCO₃) in 2 min with a Kinetex EVO C18 column]. The purity and MS information of the hit compound **16** were obtained via LC-MS on a Waters ACQUITY UPLC I-Class [gradient from 5% MeCN/95% H₂O to 100% MeCN/0% H₂O (+0.1% Formic acid) in 5 min with a ACQUITY UPLC BEH C18 1.7µm 2.1mm×50mm]. The purity of the samples was assessed using a UV detector at 254 nm.

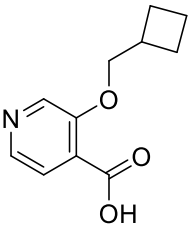
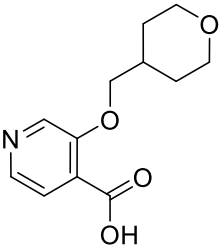
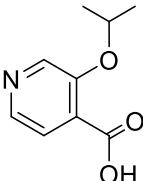
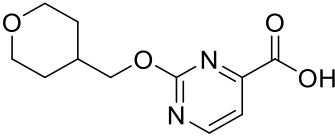
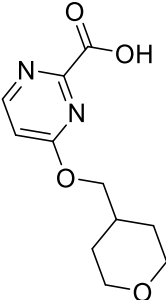
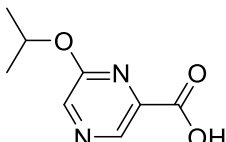
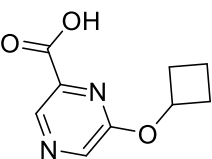
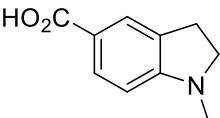
Table S4 The structures, mp, and MS data of BBs used in the synthesis.

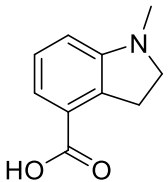
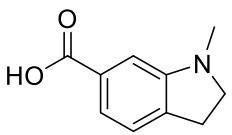
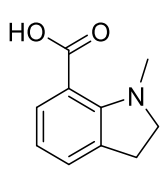
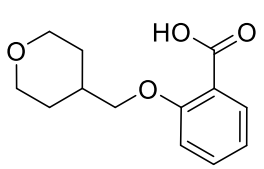
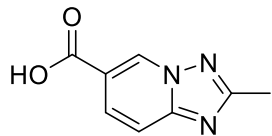
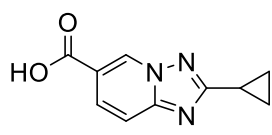
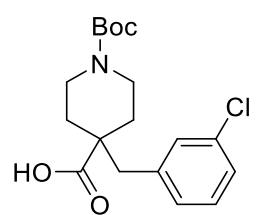
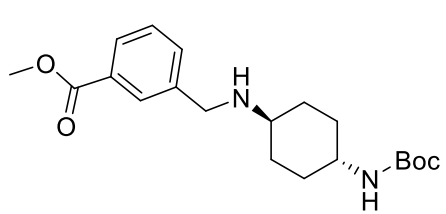
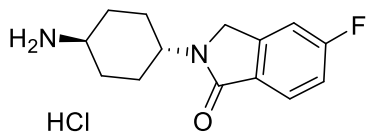
BB	Structure	mp (°C)	Molecular weight	MS (<i>m/z</i>)
K-002-098		125–128	298.77	[M+1-HCl]: 263
K-002-148		161–163	384.45	[M+1-Boc]: 285
K-002-319		150–153	321.24	[M+1-2HCl]: 249
K-002-322		256–258	321.24	[M+1-2HCl]: 249
K-002-324		218–220	334.42	[M+1]: 335
K-002-427		238–240	348.40	[M+1]: 349

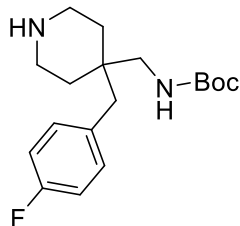
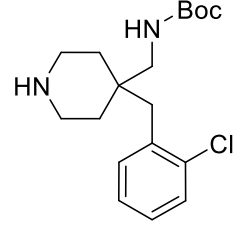
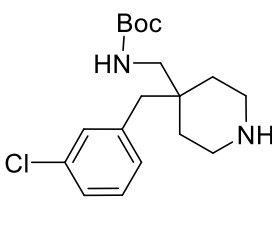
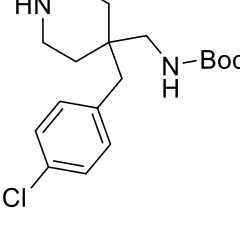
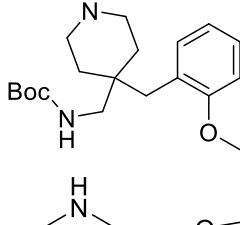
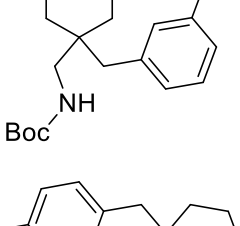
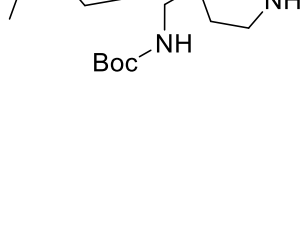
K-002-430		155–160	348.40	[M+1-t-Bu]: 293
K-008-2570		83–87	179.22	[M+1]: 180
K-008-2586		168–172	192.27	[M+1]: 193
K-008-301		80–85	291.40	[M+1]: 292
K-008-340		38–41	266.32	[M+1]: 267
K-008-421		140–145	292.38	[M+1]: 293

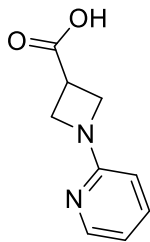
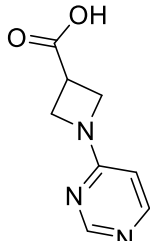
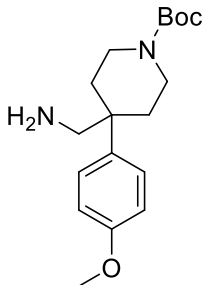
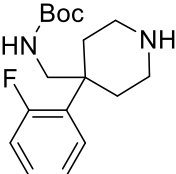
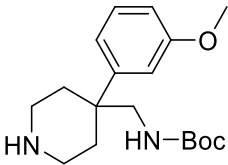
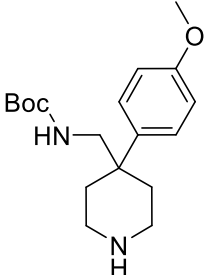
K-013-1218		138–140	262.18	[M-1]: 261
K-013-1419		226–228	203.20	[M+1]: 204
K-013-1893		83–85	237.26	[M+1]: 238
K-013-1897		245–247	215.21	[M+1]: 216
K-013-1935		193–196	247.25	[M+1]: 248
K-013-1955		223–225	205.22	[M+1]: 206
K-013-1960		183–185	205.22	[M+1]: 206

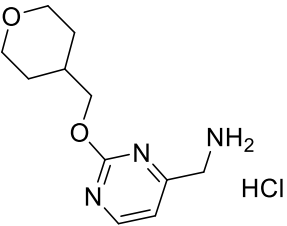
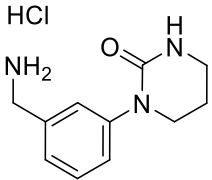
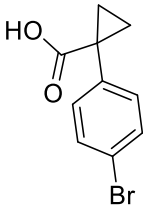
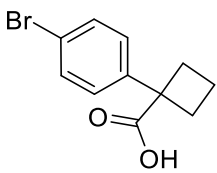
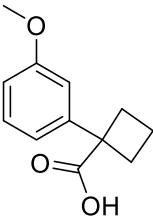
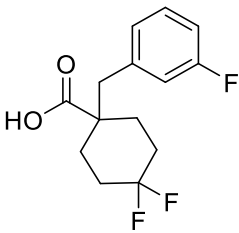
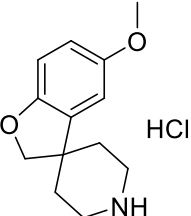
K-013-1971		112–114	237.26	[M+1]: 238
K-013-1998		110–112	237.26	[M+1]: 238
K-013-2009		145–147	237.26	[M+1]: 238
K-013-2078		128–130	195.22	[M+1]: 196
K-013-2446		145–147	206.25	[M+1]: 207
K-013-2642		>290	193.20	[M+1]: 194

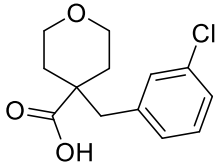
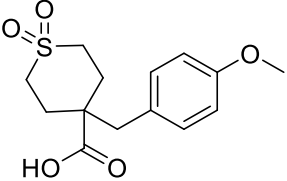
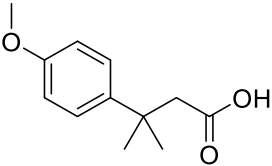
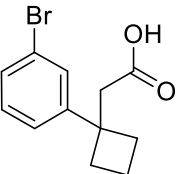
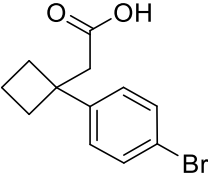
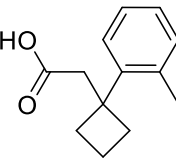
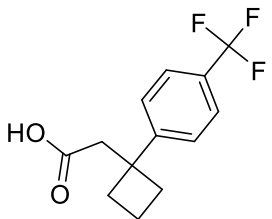
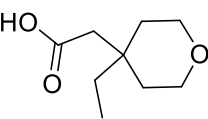
K-013-2643		175–177	207.23	[M+1]: 208
K-013-2644		166–168	237.26	[M+1]: 238
K-013-2646		173–175	181.19	[M+1]: 182
K-013-2655		170–172	238.24	[M+1]: 239
K-013-2666		108–110	238.24	[M+1]: 239
K-013-2690		169–171	182.18	[M+1]: 183
K-013-2691		182–184	194.19	[M+1]: 195
K-013-2740		239–241	177.20	[M+1]: 178

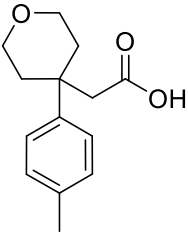
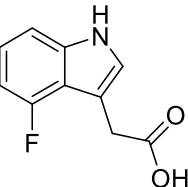
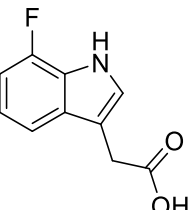
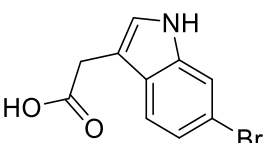
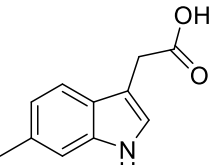
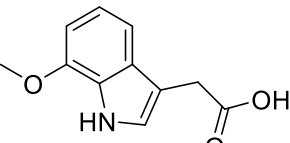
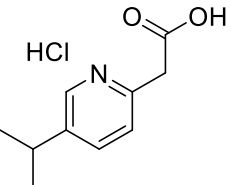
K-013-2749		176–178	177.20	[M+1]: 178
K-013-2750		160–162	177.20	[M+1]: 178
K-013-2751		99–101	177.20	[M+1]: 178
K-013-514		58–62	236.27	[M+1]: 237
K-013-9001		>250	177.16	[M+1]: 178
K-013-9002		>250	203.20	[M+1]: 204
K-014-138		139–141	353.84	[M-1]: 352
K-028-182		268–270	362.47	[M+1]: 363
K-028-243		325–327	284.76	[M+1-HCl]: 249

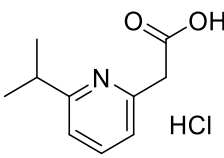
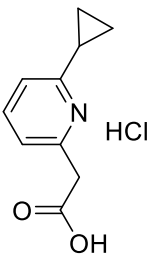
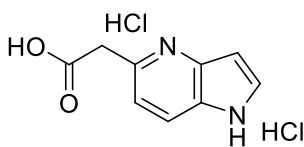
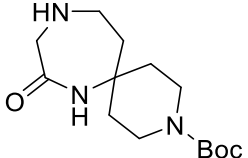
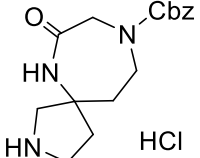
K-032-382		170–172	322.42	[M+1]: 323
K-032-383		72–76	338.88	[M+1]: 339
K-032-384		51–53	338.88	[M+1]: 339
K-032-385		97–101	338.88	[M+1]: 339
K-032-392		54–56	334.46	[M+1]: 335
K-032-393		121–123	334.46	[M+1]: 335
K-032-394		83–85	334.46	[M+1]: 335

K-044-047		201–203	178.19	[M-1]: 179
K-044-050		160–168	179.18	[M-1]: 180
K-045-612		67–69	320.43	[M+1-t-Bu]: 265
K-045-692		145–147	308.40	[M+1]: 309
K-045-705		63–65	320.43	[M+1]: 321
K-045-706		120–122	320.43	[M+1]: 321

K-049-625		158–160	259.73	[M+1-HCl]: 224
K-049-701		208–210	241.72	[M+1-HCl]: 206
K-050-048		156–158	241.08	[M-1]: 239/241
K-050-054		120–122	255.11	[M-1]: 253/255
K-050-070		64–66	206.24	[M-1]: 205
K-050-1472		116–118	272.27	[M]: 272
K-050-2521		262–265	255.74	[M+1-HCl]: 220

K-050-253		98–100	254.71	[M]: 254
K-050-3286		250–254	298.35	[M-1]: 297
K-050-3513		73–75	208.26	[M-1]: 207
K-050-3545		76–78	269.14	[M-1]: 267/269
K-050-3546		75–77	269.14	[M+1]: 269/271
K-050-3547		67–69	204.27	[M+1+CH ₃ CN]: 246
K-050-3552		(oil)	258.24	[M]: 258
K-050-3557		30–32	172.22	[M+1]: 173

K-050-3570		114–116	234.30	[M-1]: 233
K-050-4145		160–162	193.18	[M+1]: 194
K-050-4148		164–166	193.18	[M-1]: 192
K-050-4155		192–194	254.08	[M+1]: 254/256
K-050-4159		174–176	189.21	[M+1]: 190
K-050-4168		99–101	205.21	[M+1]: 206
K-050-4613		120–125	215.68	[M+1-HCl]: 180

K-050-4637		160–162	215.68	[M+1-HCl]: 180
K-050-4638		155–157	213.66	[M+1-HCl]: 178
K-050-4700		178–180	249.09	[M+1-2HCl]: 177
K-054-0040		145–149	283.37	[M+1]: 284
K-054-0046		>200	339.82	[M+1-HCl]: 304

NMR data of BBs.

K-002-098:

^1H NMR (300 MHz, D_2O): δ 8.16 (s, 1H), 8.05–8.02 (d, 1H), 7.86–7.83 (d, 1H), 7.52–7.46 (t, 1H), 4.11–4.01 (m, 1H), 3.84 (s, 3H), 3.47–3.38 (m, 2H), 3.14–3.01 (m, 2H), 2.18–2.03 (m, 2H), 1.83–1.69 (m, 2H).

K-002-148:

^1H NMR (300 MHz, DMSO): δ 13.54 (s, 1H), 8.36 (s, 1H), 8.18–8.15 (m, 1H), 8.07–8.03 (m, 1H), 7.98–7.95 (m, 1H), 7.75–7.70 (m, 1H), 3.72–3.67 (m, 2H), 3.21–3.16 (m, 1H), 2.76–2.72 (m, 2H), 1.54–1.20 (m, 13H).

K-002-319:

^1H NMR (300 MHz, $\text{DMSO}-d_6$): δ 10.21 (s, 1H), 8.68 (d, $J = 42.6$ Hz, 3H), 8.15–7.92 (m, 1H), 7.92–7.46 (m, 3H), 4.61 (dd, $J = 25.1, 4.8$ Hz, 2H), 3.89 (d, $J = 3.4$ Hz, 3H), 3.42 (s, 2H), 3.19 (d, $J = 11.6$ Hz, 3H), 2.04 (dt, $J = 37.6, 13.5$ Hz, 4H).

K-002-322:

^1H NMR (300 MHz, D_2O): δ 8.10–8.05 (m, 2H), 7.72–7.69 (d, 1H), 7.61–7.56 (t, 1H), 4.37 (s, 2H), 3.89 (s, 3H), 3.64–3.46 (m, 3H), 3.17–3.09 (m, 2H), 2.29–2.24 (m, 2H), 1.88–1.84 (m, 2H).

K-002-324:

^1H NMR (300 MHz, DMSO): δ 7.97–7.94 (d, 2H), 7.57–7.54 (d, 2H), 4.23 (s, 2H), 3.47 (m, 1H), 3.26–3.22 (m, 2H), 3.01–2.94 (m, 2H), 1.92–1.89 (m, 2H), 1.64–1.53 (m, 2H), 1.32 (s, 9H).

K-002-427:

^1H NMR (300 MHz, $\text{DMSO}-d_6$): δ 13.03 (s, 1H), 7.88 (d, $J = 7.7$ Hz, 1H), 7.60 (t, $J = 7.4$ Hz, 1H), 7.47 (t, $J = 7.6$ Hz, 1H), 7.22 (d, $J = 7.5$ Hz, 1H), 6.85 (s, 1H), 4.26 (d, $J = 13.5$ Hz, 1H), 3.17 (d, $J = 13.3$ Hz, 1H), 2.88 (s, 2H), 1.35 (s, 13H).

K-002-430:

^1H NMR (300 MHz, DMSO): δ 13.13 (s, 1H), 8.03–7.95 (m, 1H), 7.86 (s, 1H), 7.62–7.54 (m, 2H), 6.89–6.86 (d, 7.5 Hz, 1H), 4.40–4.20 (m, 1H), 3.60–3.41 (m, 2H), 3.20–2.91 (m, 2H), 1.91–1.61 (m, 2H), 1.44–1.18 (m, 11H).

K-008-2570:

^1H NMR (300 MHz, Chloroform-*d*): δ 7.29 (d, $J = 7.9$ Hz, 1H), 5.99 (d, $J = 8.1$ Hz, 1H), 5.92 (d, $J = 7.8$ Hz, 1H), 4.23 (s, 2H), 3.89–3.73 (m, 4H), 3.56–3.35 (m, 4H).

K-008-2586:

^1H NMR (300 MHz, CD_3OD): δ 7.67–7.65 (d, 1H), 7.31–7.28 (d, 1H), 6.68–6.64 (m, 1H), 2.95 (m, 4H), 2.65 (m, 4H), 2.37 (s, 3H).

K-008-301:

^1H NMR (300 MHz, $\text{DMSO-}d_6$): δ 6.97–6.71 (m, 3H), 6.66 (dd, $J = 7.9, 1.6$ Hz, 1H), 6.52 (td, $J = 7.5, 1.6$ Hz, 1H), 3.33 (s, 1H), 2.99 (dt, $J = 12.3, 3.6$ Hz, 2H), 2.64–2.49 (m, 2H), 1.92–1.69 (m, 2H), 1.66–1.46 (m, 2H), 1.40 (s, 9H).

K-008-340:

^1H NMR (300 MHz, DMSO): δ 6.90–6.69 (m, 2H), 6.55 (td, $J = 8.1, 6.2$ Hz, 1H), 4.61 (s, 2H), 4.11 (q, $J = 7.1$ Hz, 2H), 3.07 (s, 2H), 2.59 (t, $J = 10.5$ Hz, 2H), 2.44 (d, $J = 11.0$ Hz, 1H), 2.04–1.66 (m, 4H), 1.21 (t, $J = 7.1$ Hz, 3H).

K-008-421:

^1H NMR (300 MHz, $\text{DMSO-}d_6$): δ 7.54 (dd, $J = 4.7, 1.7$ Hz, 1H), 7.00–6.81 (m, 2H), 6.77 (dd, $J = 7.7, 4.7$ Hz, 1H), 4.70 (s, 2H), 3.32 (s, 3H), 2.62 (td, $J = 12.2, 2.5$ Hz, 2H), 1.81 (d, $J = 12.4$ Hz, 2H), 1.55 (qd, $J = 11.8, 3.8$ Hz, 2H), 1.40 (s, 9H).

K-013-1218:

^1H NMR (300 MHz, $\text{DMSO-}d_6$): δ 13.12 (s, 1H), 8.10 (dq, $J = 4.8, 2.2$ Hz, 2H), 6.96 (d, $J = 9.2$ Hz, 1H), 5.71–5.35 (m, 1H), 5.11–4.79 (m, 2H), 4.70–4.30 (m, 2H).

K-013-1419:

^1H NMR (300 MHz, DMSO): δ 13.56 (s, 1H), 9.36 (s, 1H), 8.78 (s, 1H), 8.19 (s, 1H), 2.09–2.00 (m, 1H), 0.98–0.80 (m, 4H).

K-013-1893:

^1H NMR (300 MHz, CDCl_3): δ 7.86–7.79 (m, 2H), 7.06–7.00 (m, 1H), 4.20–4.18 (d, 2H), 4.07–4.02 (m, 2H), 3.50–3.41 (m, 2H), 2.18–2.03 (m, 1H), 1.79–1.73 (m, 2H), 1.57–1.51 (m, 2H).

K-013-1897:

^1H NMR (300 MHz, DMSO): δ 13.01 (s, 1H), 9.51 (s, 1H), 8.98 (s, 1H), 8.42–8.38 (m, 1H), 8.13–8.11 (m, 1H), 7.79–7.73 (m, 1H), 7.50–7.47 (m, 1H), 7.33–7.29 (m, 1H), 7.08–7.05 (m, 1H).

K-013-1935:

^1H NMR (300 MHz, CD_3OD): δ 8.99 (dd, $J = 6.9, 1.0$ Hz, 1H), 8.54 (dd, $J = 7.1, 1.0$ Hz, 1H), 7.52 (t, $J = 7.0$ Hz, 1H), 4.11 (dt, $J = 11.6, 3.5$ Hz, 2H), 3.72 (ddd, $J = 13.6, 10.5, 6.1$ Hz, 3H), 2.09 (td, $J = 10.2, 8.9, 3.7$ Hz, 4H).

K-013-1955:

^1H NMR (400 MHz, DMSO): δ 13.53 (s, 1H), 9.40 (s, 1H), 8.83 (s, 1H), 6.74 (s, 1H), 3.19–3.12 (m, 1H), 1.35–1.33 (d, 6H).

K-013-1960:

^1H NMR (300 MHz, DMSO): δ 13.56 (s, 1H), 9.41 (s, 1H), 8.82 (s, 1H), 8.33 (s, 1H), 3.32–3.22 (m, 1H), 1.35–1.33 (d, 6H).

K-013-1971:

^1H NMR (300 MHz, DMSO): δ 12.89 (s, 1H), 8.32–8.30 (d, 1H), 8.10–8.07 (d, 1H), 7.08–7.04 (m, 1H), 4.20–4.18 (d, 2H), 3.95–3.85 (m, 2H), 3.36–3.28 (m, 2H), 2.08–1.93 (m, 1H), 1.69–1.65 (m, 2H), 1.40–1.30 (m, 2H).

K-013-1998:

^1H NMR (300 MHz, DMSO): δ 13.05 (s, 1H), 8.14–8.12 (d, 1H), 7.61–7.58 (d, 1H), 7.49–7.44 (m, 1H), 3.94–3.85 (m, 4H), 3.36–3.28 (m, 2H), 2.02–1.95 (m, 1H), 1.67–1.64 (m, 2H), 1.41–1.28 (m, 2H).

K-013-2009:

^1H NMR (400 MHz, DMSO): δ 13.45 (s, 1H), 8.67 (s, 1H), 8.50 (s, 1H), 7.74 (s, 1H), 3.99–3.90 (m, 2H), 3.89–3.79 (m, 2H), 3.37–3.31 (m, 2H), 2.08–1.99 (m, 1H), 1.75–1.65 (m, 2H), 1.36–1.26 (m, 2H).

K-013-2078:

^1H NMR (400 MHz, DMSO): δ 14.71 (s, 1H), 8.41–8.39 (d, 1H), 8.24–8.22 (d, 1H), 6.76–6.73 (m, 1H), 3.96–3.94 (d, 2H), 2.10–2.00 (m, 1H), 0.89–0.88 (d, 6H).

K-013-2446:

^1H NMR (300 MHz, DMSO): δ 8.81 (s, 1H), 8.51–8.49 (d, 1H), 7.70–7.69 (d, 1H), 3.16–3.12 (m, 4H), 1.74–1.59 (m, 6H).

K-013-2642:

^1H NMR (400 MHz, DMSO- d_6): δ 13.34 (s, 1H), 8.50 (s, 1H), 8.27 (d, J = 4.8 Hz, 1H), 7.49 (d, J = 4.7 Hz, 1H), 4.04 (d, J = 6.8 Hz, 2H), 1.42–1.07 (m, 1H), 0.75–0.47 (m, 2H), 0.47–0.20 (m, 2H).

K-013-2643:

^1H NMR (400 MHz, DMSO- d_6): δ 13.30 (s, 1H), 8.52 (s, 1H), 8.28 (d, J = 4.7 Hz, 1H), 7.49 (d, J = 4.7 Hz, 1H), 4.14 (d, J = 6.2 Hz, 2H), 2.85–2.62 (m, 1H), 2.17–1.94 (m, 2H), 1.94–1.74 (m, 4H).

K-013-2644:

^1H NMR (300 MHz, Methanol- d_4): δ 8.32 (s, 1H), 8.16 (d, J = 4.8 Hz, 1H), 7.44 (d, J = 4.8 Hz, 1H), 4.07–3.82 (m, 4H), 3.43 (td, J = 11.8, 2.1 Hz, 2H), 2.08 (ttd, J = 10.2, 6.4, 3.3 Hz, 1H), 1.77 (ddd, J = 13.1, 4.1, 2.0 Hz, 2H), 1.45 (dtd, J = 13.4, 11.9, 4.5 Hz, 2H).

K-013-2646:

^1H NMR (300 MHz, DMSO- d_6): δ 13.23 (s, 1H), 8.48 (s, 1H), 8.21 (d, J = 4.7 Hz, 1H), 7.43 (dd, J = 4.8, 0.6 Hz, 1H), 4.74 (hept, J = 6.1 Hz, 1H), 1.25 (d, J = 6.0 Hz, 6H).

K-013-2655:

^1H NMR (400 MHz, DMSO): δ 8.84–8.83 (d, 1H), 7.60–7.59 (d, 1H), 4.23–4.21 (d, 2H), 3.90–3.86 (m, 2H), 3.37–3.33 (m, 2H), 2.10–1.99 (m, 1H), 1.69–1.65 (m, 2H), 1.35–1.25 (m, 2H).

K-013-2666:

^1H NMR (400 MHz, DMSO- d_6): δ 8.63 (s, 1H), 7.10 (d, J = 5.5 Hz, 1H), 4.23 (d, J = 6.4 Hz, 2H), 3.86 (dd, J = 11.3, 4.2 Hz, 2H), 3.31 (t, J = 11.6 Hz, 2H), 2.12–1.91 (m, 1H), 1.77–1.53 (m, 2H), 1.32 (qd, J = 12.3, 4.5 Hz, 2H).

K-013-2690:

^1H NMR (400 MHz, DMSO): δ 8.72 (s, 1H), 8.43 (s, 1H), 5.36–5.30 (m, 1H), 1.35–1.33 (d, 6H).

K-013-2691:

^1H NMR (300 MHz, DMSO- d_6): δ 13.53 (s, 1H), 8.70 (d, J = 0.5 Hz, 1H), 8.43 (d, J = 0.6 Hz, 1H), 5.32–4.95 (m, 1H), 2.43–2.27 (m, 2H), 2.19–1.91 (m, 2H), 1.91–1.43 (m, 2H).

K-013-2740:

^1H NMR (300 MHz, DMSO- d_6): δ 7.65 (dd, $J = 8.2, 1.8$ Hz, 1H), 7.53 (d, $J = 1.8$ Hz, 1H), 6.45 (d, $J = 8.3$ Hz, 1H), 3.40 (t, $J = 8.5$ Hz, 2H), 2.92 (t, $J = 8.4$ Hz, 2H), 2.77 (s, 3H).

K-013-2749:

^1H NMR (400 MHz, DMSO- d_6): δ 7.22–6.97 (m, 2H), 6.67 (dd, $J = 7.1, 1.7$ Hz, 1H), 3.30 (d, $J = 8.0$ Hz, 2H), 3.18 (t, $J = 8.5$ Hz, 2H).

K-013-2750:

^1H NMR (400 MHz, DMSO- d_6): δ 12.60 (s, 1H), 7.23 (dd, $J = 7.5, 1.5$ Hz, 1H), 7.11 (dd, $J = 7.5, 1.3$ Hz, 1H), 6.96 (d, $J = 1.5$ Hz, 1H), 3.29 (t, $J = 8.3$ Hz, 2H), 3.00–2.81 (m, 2H), 2.71 (s, 3H).

K-013-2751:

^1H NMR (400 MHz, DMSO): δ 12.48 (s, 1H), 7.33–7.31 (d, 1H), 7.15–7.13 (d, 1H), 6.61–6.57 (m, 1H), 3.42–3.38 (m, 2H), 2.93–2.88 (m, 2H), 2.78 (s, 3H).

K-013-514:

^1H NMR (300 MHz, CDCl_3): δ 8.21–8.18 (d, 1H), 7.59–7.53 (t, 1H), 7.18–7.13 (t, 1H), 7.06–7.03 (d, 1H), 4.11–4.03 (m, 4H), 3.51–3.42 (t, 2H), 2.29–2.21 (m, 1H), 1.80–1.75 (m, 2H), 1.57–1.51 (m, 2H).

K-013-9001:

^1H NMR (300 MHz, Methanol- d_4): δ 9.53 (s, 1H), 8.56 (d, $J = 9.0$ Hz, 1H), 8.04 (d, $J = 9.1$ Hz, 1H), 2.73 (s, 3H).

K-013-9002:

^1H NMR (300 MHz, DMSO- d_6): δ 13.51 (s, 1H), 9.24 (s, 1H), 7.98 (dd, $J = 9.3, 1.6$ Hz, 1H), 7.72 (d, $J = 9.2$ Hz, 1H), 2.20 (tt, $J = 8.1, 4.8$ Hz, 1H), 1.27–0.75 (m, 4H).

K-014-138:

^1H NMR (300 MHz, DMSO): δ 12.69 (s, 1H), 7.33–7.26 (m, 2H), 7.16 (s, 1H), 7.12–7.07 (m, 1H), 3.80–3.75 (d, 2H), 2.80 (m, 4H), 1.86–1.82 (m, 2H), 1.38–1.25 (m, 11H).

K-028-182:

^1H NMR (300 MHz, DMSO): δ 8.16 (s, 1H), 8.01–7.99 (d, 1H), 7.88–7.81 (m, 1H), 7.63–7.58 (m, 1H), 4.23 (s, 2H), 3.88 (s, 3H), 3.23–3.16 (m, 1H), 3.04–2.97 (m, 1H), 2.16–2.12 (m, 2H), 1.92–1.84 (m, 2H), 1.51–1.38 (m, 11H), 1.25–1.13 (m, 2H).

K-028-243:

^1H NMR (300 MHz, CD_3OD): δ 7.82–7.77 (m, 1H), 7.52–7.46 (m, 1H), 7.37–7.23 (m, 1H), 4.51 (s, 2H), 4.21–4.11 (m, 1H), 3.24–3.15 (m, 1H), 2.21–2.11 (m, 2H), 2.02–1.98 (m, 2H), 1.90–1.84 (m, 2H), 1.81–1.76 (m, 2H).

K-032-382:

^1H NMR (400 MHz, CD_3OD): δ 7.32–7.24 (m, 2H), 7.17–7.07 (m, 2H), 3.15–3.05 (m, 5H), 2.75 (s, 2H), 1.63–1.47 (m, 13H).

K-032-383:

^1H NMR (300 MHz, CD_3OD): δ 7.48–7.35 (m, 1H), 7.35–7.14 (m, 3H), 3.32–3.05 (m, 6H), 2.87 (d, $J = 18.0$ Hz, 2H), 1.78–1.55 (m, 4H), 1.48 (d, $J = 3.8$ Hz, 9H).

K-032-384:

^1H NMR (400 MHz, CD_3OD): δ 7.38–7.19 (m, 3H), 7.13 (dt, $J = 7.4, 1.6$ Hz, 1H), 3.06 (s, 2H), 2.87 (t, $J = 5.7$ Hz, 4H), 2.67 (s, 2H), 1.48 (s, 9H), 1.44–1.33 (m, 4H).

K-032-385:

^1H NMR (300 MHz, DMSO): δ 7.40–7.26 (m, 2H), 7.18 (d, $J = 8.2$ Hz, 2H), 6.88 (t, $J = 6.2$ Hz, 1H), 2.87 (d, $J = 6.1$ Hz, 2H), 2.73 (t, $J = 5.6$ Hz, 4H), 2.56 (s, 2H), 1.39 (s, 9H), 1.19 (ddt, $J = 25.4, 13.1, 6.8$ Hz, 4H).

K-032-392:

^1H NMR (300 MHz, CD_3OD): δ 7.29–7.19 (m, 1H), 7.11 (dd, $J = 7.5, 1.8$ Hz, 1H), 6.98 (d, $J = 8.2$ Hz, 1H), 6.96–6.84 (m, 1H), 3.84 (s, 3H), 3.01 (d, $J = 7.3$ Hz, 6H), 2.69 (s, 2H), 1.46 (s, 13H).

K-032-393:

^1H NMR (300 MHz, CD_3OD): δ 7.24–7.18 (m, 1H), 6.82–6.75 (m, 3H), 3.78 (s, 3H), 3.12–3.08 (m, 6H), 2.65 (s, 2H), 1.58–1.54 (m, 4H), 1.46 (s, 9H).

K-032-394:

^1H NMR (300 MHz, Methanol- d_4): δ 7.14–7.05 (m, 2H), 6.93–6.75 (m, 2H), 3.77 (s, 3H), 3.10 (s, 6H), 2.62 (s, 2H), 1.54 (t, $J = 5.9$ Hz, 4H), 1.46 (s, 9H).

K-044-047:

^1H NMR (300 MHz, DMSO- d_6): δ 12.62 (s, 1H), 8.05 (ddd, $J = 5.0, 2.0, 0.9$ Hz, 1H), 7.49 (ddd, $J = 8.7, 7.1, 1.9$ Hz, 1H), 6.62 (ddd, $J = 7.2, 4.9, 1.0$ Hz, 1H), 6.37 (dd, $J = 8.4, 1.0$ Hz, 1H), 4.08 (t, $J = 8.4$ Hz, 2H), 3.95 (dd, $J = 8.0, 5.9$ Hz, 2H), 3.50 (tt, $J = 8.7, 5.9$ Hz, 1H).

K-044-050:

^1H NMR (400 MHz, D_2O): δ 8.49 (s, 1H), 7.96–7.94 (m, 1H), 6.62–6.54 (m, 1H), 4.49–4.28 (m, 4H), 3.68–3.55 (m, 1H).

K-045-612:

^1H NMR (400 MHz, DMSO): δ 7.25–7.23 (d, 2H), 6.92–6.90 (d, 2H), 3.74 (s, 3H), 3.59–3.56 (m, 2H), 2.89 (m, 2H), 2.55 (s, 2H), 2.02–1.98 (m, 2H), 1.66–1.59 (m, 2H), 1.38 (s, 9H), 1.15 (m, 2H).

K-045-692:

^1H NMR (400 MHz, CDCl_3): δ 7.32–7.04 (m, 4H), 4.32–4.30 (m, 1H), 4.22 (m, 1H), 3.51–3.49 (m, 2H), 3.23 (m, 2H), 2.93 (m, 2H), 2.39–2.37 (m, 2H), 2.01 (m, 2H), 1.37 (s, 9H).

K-045-705:

^1H NMR (400 MHz, DMSO): δ 7.29–7.25 (m, 1H), 6.91–6.89 (m, 1H), 6.83–6.79 (m, 2H), 6.66–6.63 (m, 1H), 3.74 (s, 3H), 3.02–2.95 (m, 4H), 2.59–2.50 (m, 2H), 2.08–2.05 (m, 2H), 1.79–1.74 (m, 2H), 1.31 (s, 9H).

K-045-706:

^1H NMR (400 MHz, CDCl_3): δ 7.25–7.23 (d, 2H), 6.94–6.92 (d, 2H), 3.84 (s, 3H), 3.29–3.27 (m, 2H), 3.02–2.98 (m, 2H), 2.82–2.76 (m, 2H), 2.40 (m, 1H), 2.11–2.08 (m, 2H), 1.82–1.77 (m, 2H), 1.42 (s, 9H).

K-049-625:

^1H NMR (300 MHz, CD_3OD): δ 8.76–8.60 (m, 1H), 7.41–7.22 (m, 1H), 4.32–4.26 (m, 2H), 4.00–3.96 (m, 2H), 3.86 (s, 2H), 3.50–3.43 (m, 2H), 2.19 (ttt, $J = 10.5, 6.6, 3.8$ Hz, 1H), 1.80 (ddd, $J = 13.1, 4.1, 2.0$ Hz, 2H), 1.67–1.35 (m, 2H).

K-049-701:

^1H NMR (300 MHz, D_2O): δ 7.43–7.38 (m, 1H), 7.28–7.23 (m, 3H), 4.09 (s, 2H), 3.60–3.56 (m, 2H), 3.32–3.28 (m, 2H), 2.02–1.94 (m, 2H).

K-050-048:

^1H NMR (400 MHz, CDCl_3): δ 7.46–7.44 (m, 2H), 7.25–7.18 (m, 2H), 1.75–1.69 (d, 2H), 1.26–1.21 (d, 2H).

K-050-054:

^1H NMR (300 MHz, CDCl_3): δ 7.55–7.36 (m, 2H), 7.26–7.11 (m, 2H), 3.00–2.75 (m, 2H), 2.64–2.38 (m, 2H), 2.12 (dp, $J = 11.4, 8.6$ Hz, 1H), 1.90 (dtt, $J = 11.1, 9.3, 4.9$ Hz, 1H).

K-050-070:

^1H NMR (400 MHz, DMSO): δ 12.34 (s, 1H), 7.29–7.23 (m, 1H), 6.89–6.78 (m, 3H), 3.75 (s, 3H), 2.73–2.66 (m, 2H), 2.43–2.36 (m, 2H), 1.96–1.93 (m, 1H), 1.91–1.83 (m, 1H).

K-050-1472:

^1H NMR (400 MHz, DMSO): δ 12.77 (s, 1H), 7.34–7.29 (m, 1H), 7.08–7.04 (m, 1H), 6.99–6.85 (m, 2H), 2.85 (s, 2H), 1.99–1.96 (m, 4H), 1.81–1.67 (m, 2H), 1.54–1.35 (m, 2H).

K-050-2521:

^1H NMR (300 MHz, Methanol- d_4): δ 6.85–6.62 (m, 3H), 4.46 (s, 2H), 3.76 (s, 3H), 3.54–3.39 (m, 2H), 3.24–3.02 (m, 2H), 2.13 (td, $J = 13.8, 12.8, 4.4$ Hz, 2H), 1.97 (d, $J = 14.5$ Hz, 2H).

K-050-253:

^1H NMR (300 MHz, DMSO): δ 12.63 (s, 1H), 7.30–7.28 (m, 2H), 7.16 (s, 1H), 7.10–7.07 (m, 1H), 3.78–3.72 (m, 2H), 3.28–3.24 (m, 2H), 2.81 (s, 2H), 1.84–1.80 (m, 2H), 1.54–1.46 (m, 2H).

K-050-3286:

^1H NMR (300 MHz, DMSO- d_6): δ 13.01 (s, 1H), 7.21–6.95 (m, 2H), 6.95–6.67 (m, 2H), 3.88–3.60 (m, 3H), 3.15 (d, $J = 14.4$ Hz, 2H), 2.94 (t, $J = 13.2$ Hz, 2H), 2.83 (s, 2H), 2.24 (d, $J = 14.4$ Hz, 2H), 1.92 (t, $J = 13.2$ Hz, 2H).

K-050-3513:

^1H NMR (400 MHz, CD_3OD): δ 7.34–7.30 (d, 2H), 6.87–6.83 (d, 2H), 3.78 (s, 3H), 2.60 (s, 2H), 1.45 (s, 6H).

K-050-3545:

^1H NMR (300 MHz, CDCl_3): δ 7.33 (dd, $J = 7.1, 1.5$ Hz, 2H), 7.22–7.08 (m, 2H), 2.83 (s, 2H), 2.54–2.28 (m, 4H), 2.10 (dp, $J = 11.6, 8.6$ Hz, 1H), 1.89 (dtt, $J = 11.5, 9.1, 4.3$ Hz, 1H).

K-050-3546:

^1H NMR (300 MHz, CDCl_3): δ 7.45–7.29 (m, 2H), 7.13–6.91 (m, 2H), 2.78 (s, 2H), 2.48–2.23 (m, 4H), 2.07 (dq, $J = 11.6, 8.5$ Hz, 1H), 1.84 (dtt, $J = 11.5, 9.2, 4.5$ Hz, 1H).

K-050-3547:

^1H NMR (300 MHz, $\text{DMSO}-d_6$): δ 11.80 (s, 1H), 7.20–6.79 (m, 4H), 2.73 (s, 2H), 2.46–2.28 (m, 4H), 2.18 (s, 3H), 2.13–1.94 (m, 1H), 1.73 (ddt, $J = 10.6, 7.4, 4.0$ Hz, 1H).

K-050-3552:

^1H NMR (300 MHz, CDCl_3): δ 7.55 (d, $J = 8.1$ Hz, 2H), 7.38–7.26 (m, 2H), 2.86 (s, 2H), 2.43 (dddd, $J = 14.7, 12.4, 10.7, 6.6$ Hz, 4H), 2.23–2.05 (m, 1H), 1.90 (dtt, $J = 11.4, 9.2, 4.4$ Hz, 1H).

K-050-3557:

^1H NMR (400 MHz, $\text{DMSO}-d_6$): δ 12.02 (s, 1H), 3.69–3.44 (m, 4H), 2.27 (s, 2H), 1.66–1.19 (m, 6H), 0.82 (t, $J = 7.5$ Hz, 3H).

K-050-3570:

^1H NMR (400 MHz, DMSO): δ 11.77 (s, 1H), 7.24–7.15 (m, 3H), 7.03–7.01 (m, 1H), 3.72–3.66 (m, 2H), 3.48–3.43 (m, 2H), 2.57 (s, 2H), 2.34 (s, 3H), 2.13–2.09 (m, 2H), 2.01–1.96 (m, 2H).

K-050-4145:

^1H NMR (300 MHz, CD_3OD): δ 7.45 (dd, $J = 8.7, 5.3$ Hz, 1H), 7.12 (t, $J = 1.0$ Hz, 1H), 7.00 (dd, $J = 10.0, 2.3$ Hz, 1H), 6.77 (ddd, $J = 9.8, 8.7, 2.3$ Hz, 1H), 3.68 (d, $J = 0.9$ Hz, 2H).

K-050-4148:

^1H NMR (300 MHz, CD_3OD): δ 10.84 (s, 1H), 7.34 (d, $J = 7.9$ Hz, 1H), 7.29–7.15 (m, 1H), 6.96 (td, $J = 7.9, 4.7$ Hz, 1H), 6.83 (dd, $J = 11.5, 7.8$ Hz, 1H), 3.74 (d, $J = 0.8$ Hz, 2H).

K-050-4155:

^1H NMR (300 MHz, $\text{DMSO}-d_6$): δ 12.15 (s, 1H), 11.02 (s, 1H), 7.49 (d, $J = 1.7$ Hz, 1H), 7.41 (d, $J = 8.4$ Hz, 1H), 7.23 (d, $J = 2.4$ Hz, 1H), 7.07 (dd, $J = 8.4, 1.8$ Hz, 1H), 3.59 (d, $J = 0.8$ Hz, 2H).

K-050-4159:

^1H NMR (300 MHz, DMSO): δ 12.11 (s, 1H), 10.73 (s, 1H), 7.37–7.35 (d, J = 7.8 Hz, 1H), 7.12 (s, 2H), 6.82–6.79 (d, J = 7.8 Hz, 1H), 3.59 (s, 2H), 2.37 (s, 3H).

K-050-4168:

^1H NMR (300 MHz, DMSO- d_6): δ 12.08 (s, 1H), 10.95 (s, 1H), 7.23–6.96 (m, 2H), 6.86 (t, J = 7.8 Hz, 1H), 6.60 (d, J = 7.6 Hz, 1H), 3.86 (s, 3H), 3.68–3.46 (m, 2H).

K-050-4613:

^1H NMR (300 MHz, DMSO- d_6): δ 8.78 (d, J = 6.1 Hz, 1H), 8.07–7.68 (m, 2H), 3.16 (hept, J = 6.9 Hz, 1H), 1.26 (d, J = 6.9 Hz, 6H).

K-050-4637:

^1H NMR (400 MHz, D_2O): δ 8.35 (t, J = 8.0 Hz, 1H), 7.78 (dd, J = 8.2, 1.2 Hz, 1H), 7.67 (dd, J = 7.9, 1.1 Hz, 1H), 3.27 (hept, J = 7.0 Hz, 1H), 1.31 (d, J = 7.0 Hz, 6H).

K-050-4638:

^1H NMR (400 MHz, DMSO- d_6): δ 8.38–8.15 (m, 1H), 7.63 (dd, J = 13.5, 7.8 Hz, 1H), 7.41 (dd, J = 29.2, 8.2 Hz, 1H), 4.27–4.07 (m, 2H), 2.69–2.52 (m, 1H), 1.34 (dd, J = 7.7, 4.9 Hz, 2H), 1.15 (tq, J = 4.8, 3.0, 1.9 Hz, 2H).

K-050-4700:

^1H NMR (400 MHz, DMSO- d_6): δ 12.58 (s, 1H), 8.48 (d, J = 8.4 Hz, 1H), 8.18 (t, J = 3.0 Hz, 1H), 7.56 (d, J = 8.3 Hz, 1H), 6.99–6.67 (m, 1H).

K-054-0040:

^1H NMR (400 MHz, D_2O): δ 3.53 (s, 4H), 3.37–3.18 (m, 2H), 3.12–2.93 (m, 2H), 1.92–1.74 (m, 4H), 1.65 (ddd, J = 13.8, 9.4, 4.1 Hz, 2H), 1.36 (s, 9H).

K-054-0046:

^1H NMR (400 MHz, D_2O): δ 7.33 (dd, J = 18.5, 9.1 Hz, 5H), 5.20–4.98 (m, 2H), 4.27–4.01 (m, 2H), 3.73–3.45 (m, 2H), 3.36 (dt, J = 12.6, 7.6 Hz, 2H), 3.20 (d, J = 10.7 Hz, 2H), 2.36–1.80 (m, 4H).

Bioassay

The PPI inhibitory activities of compounds against Keap1/Nrf2 and Bcl6/F1325 (as a counter) were measured using a TR-FRET assay,^{25,26} as described in our previous work.¹

References

- 1 Y. Shimizu, T. Yonezawa, J. Sakamoto, T. Furuya, M. Osawa and K. Ikeda, *Sci. Rep.*, 2021, **11**, 7420.
- 2 D. Mendez, A. Gaulton, A. P. Bento, J. Chambers, M. De Veij, E. Félix, M. P. Magariños, J. F. Mosquera, P. Mutowo, M. Nowotka, M. Gordillo-Marañón, F. Hunter, L. Junco, G. Mugumbate, M. Rodríguez-Lopez, F. Atkinson, N. Bosc, C. J. Radoux, A. Segura-Cabrera, A. Hersey and A. R. Leach, *Nucleic Acids Res.*, 2019, **47**, D930–D940.
- 3 A. P. Higuieruelo, H. Jubb and T. L. Blundell, *Database*, 2013, **2013**, 1–5.
- 4 S. Milhas, B. Raux, S. Betzi, C. Derviaux, P. Roche, A. Restouin, M.-J. Basse, E. Rebuffet, A. Lugari, M. Badol, R. Kashyap, J.-C. Lissitzky, C. Eydoux, V. Hamon, M.-E. Gourdel, S. Combes, P. Zimmermann, M. Aurrand-Lions, T. Roux, C. Rogers, S. Müller, S. Knapp, E. Trinquet, Y. Collette, J.-C. Guillemot and X. Morelli, *ACS Chem. Biol.*, 2016, **11**, 2140–2148.
- 5 K. Ikeda, Y. Maezawa, T. Yonezawa, Y. Shimizu, T. Tashiro, S. Kanai, N. Sugaya, Y. Masuda, N. Inoue, T. Niimi, K. Masuya, K. Mizuguchi, T. Furuya and M. Osawa, *Front. Chem.*, 2023, **10**, 1090643.
- 6 D. Rogers and M. Hahn, *J. Chem. Inf. Model.*, 2010, **50**, 742–754.
- 7 Dassault Systèmes BIOVIA, BIOVIA Pipeline Pilot (release 2018), Dassault Systèmes, San Diego, CA, 2018.
- 8 E. L. Willighagen, J. W. Mayfield, J. Alvarsson, A. Berg, L. Carlsson, N. Jeliaskova, S. Kuhn, T. Pluskal, M. Rojas-Chertó, O. Spjuth, G. Torrance, C. T. Evelo, R. Guha and C. Steinbeck, *J. Cheminform.*, 2017, **9**, 33.
- 9 RDKit: Open-Source Cheminformatics Software, <https://www.rdkit.org>, (accessed December 2022)
- 10 H. Moriwaki, Y.-S. Tian, N. Kawashita and T. Takagi, *J. Cheminform.*, 2018, **10**, 4.
- 11 L. Breiman, *Mach. Learn.*, 2001, **45**, 5–32.
- 12 I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, MIT Press, Cambridge, MA, 2016.
- 13 M. N. Wright and A. Ziegler, *J. Stat. Softw.*, 2017, **77**, 1–17.
- 14 R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2020.
- 15 N. V. Chawla, K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer, *J. Artif. Intell. Res.*, 2002, **16**, 321–357.
- 16 P. Cunningham and S. J. Delany, *ACM Comput. Surv.*, 2022, **54**, 1–25.
- 17 F. Chollet and others, Keras, <https://keras.io>, 2015
- 18 M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G.

- Irving, M. Isard and others, in *OSDI*, 2016, vol. 16, pp. 265–283.
- 19 R. Caruana, S. Lawrence and C. Giles, in *Advances in Neural Information Processing Systems*, eds. T. Leen, T. Dietterich and V. Tresp, MIT Press, Cambridge, MA, 2000, vol. 13.
- 20 P. C. D. Hawkins, A. G. Skillman, G. L. Warren, B. A. Ellingson and M. T. Stahl, *J. Chem. Inf. Model.*, 2010, **50**, 572–584.
- 21 L. McInnes, J. Healy and J. Melville, *arXiv*, 2018.
- 22 L. McInnes, J. Healy, N. Saul and L. Großberger, *J. Open Source Softw.*, 2018, **3**, 861.
- 23 Flare version 6.0, Cresset, Litlington, Cambridgeshire, UK, <https://www.cresset-group.com/flare/>, 2022
- 24 A. D. Jain, H. Potteti, B. G. Richardson, L. Kingsley, J. P. Luciano, A. F. Ryuzoji, H. Lee, A. Kronic, A. D. Mesecar, S. P. Reddy and T. W. Moore, *Eur. J. Med. Chem.*, 2015, **103**, 252–268.
- 25 S. Sogabe, K. Sakamoto, Y. Kamada, A. Kadotani, Y. Fukuda and J. Sakamoto, *Biochem. Biophys. Res. Commun.*, 2017, **486**, 620–625.
- 26 T. Sameshima, T. Yamamoto, O. Sano, S. Sogabe, S. Igaki, K. Sakamoto, K. Ida, M. Gotou, Y. Imaeda, J. Sakamoto and I. Miyahisa, *Biochemistry*, 2018, **57**, 1369–1379.