# SUPPLEMENTARY INFORMATION

# *De Novo* Design and Hit-to-Lead Optimization of Aurora A Kinase Inhibitors

Tiago Rodrigues,[†] Filip Roudnicky,[†] Christian P. Koch,[†] Takayuki Kudoh,[†,‡] Daniel Reker,[†] Michael Detmar,[†] Gisbert Schneider[†]*

[†]Eidgenössische Technische Hochschule (ETH), Department of Chemistry and Applied Biosciences, Wolfgang-Pauli-Str. 10, 8093 Zürich, Switzerland
[‡]Department of Medical and Bioengineering Science, Graduate School of Natural Science and Technology, Okayama University, Okayama, Japan

## Table of Contents

1. General information 2
2. *De novo* design and post-run analysis 2
3. Chemistry 3
   3.1 General procedure for amide bond formation 3
   3.2 General procedure for reductive amination 4
4. *In vitro* tests 6
   4.1 Aurora A kinase $IC_{50}$ determination 6
   4.2 Phenotypic assays 6
5. Kinase panel screen data 6
6. Optimal number of clusters for *k*-means clustering (Fig. S1) 8
7. Scaffold diversity and count (Table S1) 9
References 11

## 1. General information

All starting materials and solvents were obtained from ABCR Chemicals, Aldrich, Fluka, or Acros, and were used without further purification.

Analytical HPLC-MS was carried out in a Shimadzu LC-MS2020 system, equipped with a Nucleodur $C_{18}$ HTec column, under an appropriate gradient of acetonitrile : $H_2O$ (+ 0.1% TFA in each phase), and a total flow rate of 0.5 mL/min. The mass spectrometer was operated in positive-ion mode with ESI. Preparative HPLC was carried out on a Shimadzu LC-8A system, coupled to a Nucleodur 100-5 $C_{18}$ HTec column, and a SPD-20A UV/Vis detector.

Proton and carbon nuclear magnetic resonance spectra ($^1$H and $^{13}$C NMR, respectively) were recorded on Bruker Avance 400 spectrometer. Chemical shifts (δ) are reported in units of parts per million (ppm) downfield from $SiMe_4$ (δ 0.0) and relative to the respective solvent's peak. Multiplicities are given as: s (singlet), d (doublet), br. s (broad singlet) and m (multiplet). $^1$H-$^1$H Coupling constants (*J*) are reported in Hertz (Hz).

Melting points were measured on a Büchi M-560 apparatus and are uncorrected.

## 2. *De novo* design and post-run analysis

VX680 was standardized using the "wash", "protonate", and "add hydrogens" options from MOE 2011.10 (Chemical Computing Group, Montreal, Canada). The software DOGS was run, on an eight-core MacPro (OS X 10.6.8) computer with the options α = 0.4 and 0.6 (favoring a balance of explorative and conservative designs), and 100 start fragments (broad sampling of chemical space). Similarity between the designs and the template was computed using the ISOAK graph kernel method on reduced graph representations.[1,2] Molecular properties were computed with MOE 2011.10. The compound list constructed by DOGS was analyzed in terms of underlying atomic scaffolds.[3] Scaffold analysis was executed on the KNIME[4] platform employing RDkit[5] nodes for scaffold extraction and KNIME-native nodes for frequency determination. Frequencies of unique scaffolds were determined by comparing their canonical SMILES.[6] Nearest neighbor analysis, revealing kinase reference inhibitors and their inhibited kinases, was conducted by utilizing the topological auto-

correlation descriptor CATS[7] with default parameters. A projection of CATS descriptor space to a 2D map was conducted with Stochastic Neighbor Embedding (*t*-SNE[8]). Finally *k*-means[9] clustering was employed to the designed compounds' descriptors, identifying an optimal *k* of 9. A neighbor-joining[10] tree was built and labeled with highly inhibited kinases[11] (<20% remaining kinase activity) by the nearest reference inhibitor to the respective cluster means.

Automated ligand docking was performed with the software GOLD 5.1 (Genetic Optimization for Ligand Docking), which searches ligand poses using a genetic algorithm.[12] The co-crystalized structure of VX-680 in the ATP binding pocket of Aurora A kinase (PDB 3E5A)[13] was used. Crystalized ligands were extracted, hydrogen atoms were added to residues and the query ligands docked at the ATP binding site (defined by VX-680 coordinates). Five hundred runs per ligand, under standard settings, were performed and the docking poses scored with ASP (Astex Statistical Potential)[12,14] scoring function. The best poses were energy-minimized with MOE (MMFF94x force field).

## 3. Chemistry

### 3.1 General procedure for amide bond formation

Carboxylic acid (1.0 molar eq.), EDCI (1.25 molar eq.), HOBt (1.65 molar eq.) and amine (0.85 molar eq.) were dissolved in dry $CH_2Cl_2$ (20 mL / mmol carboxylic acid) and DMF (0.5 mL / mmol carboxylic acid). The reaction mixture was stirred at room temperature for 48 hours. The solvent was evaporated under reduced pressure and the crude product purified via preparative HPLC (Eluent A: ACN + 0.1% TFA; Eluent B: $H_2O$ + 0.1% TFA) with a 30-95% ACN gradient, run over 16 minutes. Evaporation of solvent afforded white solids.

### *N*-(4-(*N*-(5-methyl-1,3,4-thiadiazol-2-yl)sulfamoyl)phenyl)cyclopropane carboxamide, 2a

73%; m.p. > 250 ºC; $^1$H NMR (DMSO-$d_6$, 400.13 MHz): δ 0.83 (4H, d, $J$ = 6 Hz, $CH_2$), 1.80 (1H, m, CH), 2.51 (3H, s, $CH_3$), 7.71 (2H, d, $J$ = 8.8 Hz, Ar-H), 7.75 (2H, d, $J$ = 8.8 Hz, Ar-H), 10.55 (1H, s, NH). $^{13}$C-NMR (100.61 MHz, DMSO-$d_6$): δ 7.59,

14.64, 16.02, 118.54, 126.88, 135.52, 142.70, 154.37, 167.68, 172.23. HPLC-MS ($H_2O$ + 0.1% TFA / ACN + 0.1% TFA, RP18, 254 nm, $ESI^+$) purity: 96% (area normalization), *m/z* = 339 $[M+1]^+$; HRMS-ESI calc. ($C_{13}H_{14}N_4O_3S_2+H^+$): 339.0580, found: 339.0568.

**3-methyl-*N*-(4-(*N*-(5-methyl-1,3,4-thiadiazol-2-yl)sulfamoyl)phenyl)-1*H*-pyrazole-5-carboxamide, 2b**

58%; m.p. > 250 ºC; $^1$H NMR ($CD_3OD$, 400.13 MHz): δ 2.25 (3H, s, $CH_3$), 2.39 (3H, s, $CH_3$), 6.51 (1H, s, Ar-H), 7.73 (2H, d, *J* = 9.0 Hz, Ar-H), 7.79 (2H, d, *J* = 9.0 Hz, Ar-H). $^{13}$C-NMR (100.61 MHz, DMSO-$d_6$): δ 10.55, 16.03, 104.86, 119.67, 126.60, 130.62, 135.87, 142.33, 142.48, 154.41, 160.78, 167.77. HPLC-MS ($H_2O$ + 0.1% TFA / ACN + 0.1% TFA, RP18, 254 nm, $ESI^+$) purity: 99.9% (area normalization), *m/z* = 379 $[M+1]^+$; HRMS-ESI calc. ($C_{14}H_{14}N_6O_3S_2+H^+$): 379.0642, found: 379.0644.

***N*-(4-(*N*-(5-methyl-1,3,4-thiadiazol-2-yl)sulfamoyl)phenyl)-4-nitro-1*H*-pyrazole-5-carboxamide, 3a**

28%; m.p. > 250 ºC; $^1$H NMR (DMSO-$d_6$, 400.13 MHz): δ 2.50 (3H, s, $CH_3$), 7.80 (2H, d, *J* = 9.2 Hz, Ar-H), 7.86 (2H, d, *J* = 9.2 Hz, Ar-H), 8.99 (1H, br. s, NH), 11.05 (1H, s, Ar-H), 13.95 (1H, br. s, NH), 14.33 (1H, br. s, NH); HPLC-MS ($H_2O$ + 0.1% TFA / ACN + 0.1% TFA, RP18, 254 nm, $ESI^+$) purity: 99.9% (area normalization), *m/z* = 410 $[M+1]^+$; HRMS-ESI calc. ($C_{13}H_{11}N_7O_5S_2+H^+$): 410.0336, found: 410.0337.

***N*-(4-(*N*-(5-methyl-1,3,4-thiadiazol-2-yl)sulfamoyl)phenyl)-3-nitro-1*H*-pyrazole-5-carboxamide, 3b**

30%, m.p. > 250 ºC; $^1$H NMR (DMSO-$d_6$, 400.13 MHz): δ 2.48 (3H, s, $CH_3$), 7.82 (2H, d, *J* = 8.8 Hz, Ar-H), 7.92 (2H, d, *J* = 8.8 Hz, Ar-H), 10.77 (1H, s, Ar-H); HPLC-MS ($H_2O$ + 0.1% TFA / ACN + 0.1% TFA, RP18, 254 nm, $ESI^+$) purity: 100% (area normalization), *m/z* = 410 $[M+1]^+$; HRMS-ESI calc. ($C_{13}H_{11}N_7O_5S_2+H^+$): 410.0336, found: 410.0331.

**3.2 General procedure for reductive amination**[15]

Compound **3a** or **3b** (1 molar eq.) was dissolved in MeOH (20 mL / mmol) and 2 drops of AcOH, decaborane (0.3 molar eq.) and 10% Pd/C (55 mg / mmol) were

added to the solution. The suspension was refluxed for 30 minutes under $N_2$ and cooled down to room temperature. The required aldehyde (1.1 molar eq.) and decaborane (0.2 molar eq.) were added to the reaction mixture. The mixture was stirred at room temperature for 3 hours. The catalyst was filtered off and the solvent evaporated under reduced pressure. The crude product was purified from preparative HPLC. (Eluent A: ACN + 0.1% TFA; Eluent B: $H_2O$ + 0.1% TFA) with a 5-50% ACN gradient, run over 16 minutes.

**3-((4-cyanobenzyl)amino)-*N*-(4-(*N*-(5-methyl-1,3,4-thiadiazol-2-yl)sulfamoyl) phenyl)-1*H*-pyrazole-5-carboxamide, 4a**

White solid; 16%; m.p. > 250 ºC; $^1$H NMR (DMSO-$d_6$, 400.13 MHz): δ 2.47 (3H, s, $CH_3$), 4.37 (2H, s, $CH_2$), 5.97 (1H, br. s, NH), 7.55 (2H, d, $J$ = 9.2 Hz, Ar-H), 7.73 (2H, d, $J$ = 9.2 Hz, Ar-H), 7.80 (2H, d, $J$ = 9.2 Hz, Ar-H), 7.92 (2H, d, $J$ = 9.2 Hz, Ar-H), 10.21 (1H, s, Ar-H). $^{13}$C-NMR (100.61 MHz, DMSO-$d_6$): δ 16.02, 47.27, 88.70, 109.36, 118.92, 119.67, 126.64, 126.77, 127.99, 128.04, 132.14, 135.97, 142.33, 146.47, 154.43, 159.87, 167.76. HPLC-MS ($H_2O$ + 0.1% TFA / ACN + 0.1% TFA, RP18, 254 nm, ESI$^+$) purity: 99.9% (area normalization), $m/z$ = 495 $[M+1]^+$; HRMS-ESI calc. ($C_{21}H_{18}N_8O_3S_2$+$H^+$): 495.1016, found: 495.1017.

***N*-(4-(*N*-(5-methyl-1,3,4-thiadiazol-2-yl)sulfamoyl)phenyl)-4-((pyridin-4-ylmethyl)amino)-1*H*-pyrazole-5-carboxamide, 4b**

Yellow oil; 30%; $^1$H NMR ($CD_3OD$, 400.13 MHz): δ 2.51 (3H, s, $CH_3$), 4.68 (2H, s, $CH_2$), 7.08 (1H, s, Ar-H), 7.85 (2H, d, $J$ = 7.0 Hz, Ar-H), 7.93 (2H, d, $J$ = 9.2 Hz, Ar-H), 8.11 (2H, d, $J$ = 9.2 Hz, Ar-H), 8.77 (2H, d, $J$ = 7.0 Hz, Ar-H); $^{13}$C-NMR (100.61 MHz, $CD_3OD$): δ 16.26, 50.62, 115.17, 120.60, 120.71, 126.64, 128.28, 128.35, 137.01, 137.57, 142.40, 143.75, 156.13, 164.26, 164.84. HPLC-MS ($H_2O$ + 0.1% TFA / ACN + 0.1% TFA, RP18, 254 nm, ESI$^+$) purity: 99.9% (area normalization), $m/z$ = 471 $[M+1]^+$; HRMS-ESI calc. ($C_{19}H_{18}N_8O_3S_2$+$H^+$): 471.1016, found: 471.1029

## 4. *In vitro* tests

### 4.1 Aurora A kinase $IC_{50}$ determination

Tests were carried out at Reaction Biology Corp. as follows: Substrate was dissolved in freshly prepared buffer (20 mM Hepes pH 7.5, 10 mM $MgCl_2$, 1 mM EGTA, 0.02% Brij35, 0.02 mg/ml BSA, 0.1 mM $Na_3VO_4$, 2 mM DTT, 1% DMSO) and the required co-factors were delivered to the solution. Aurora A kinase was added to the substrate solution and mixed gently. Compounds in DMSO were added to the kinase solution followed by $^{33}$P-ATP (specific activity 500 µCi/µl), to initiate the reaction. The kinase reaction was incubated for 120 minutes at room temperature. Staurosporine was used as positive control ($IC_{50}$ = 2.94 nM). Compounds were tested in 10 different concentrations, starting at 100 µM subsequent 3-fold serial dilution. All reactions were performed at 1 µM ATP.[11]

### 4.2 Phenotypic assays

Human T47D ductal breast cancer cells ($1x10^4$) were seeded onto non-coated 96-well plates. Quinduplicate wells were treated or not with several concentrations of **2b**, **4a** or **4b** in Dulbecco's modified Eagle medium containing 1% fetal bovine serum (FBS). After 72 hours, cells were incubated with 5-methylumbelliferylheptanoate (MUH) as described.[16] The fluorescence intensity, proportional to the number of viable cells, was measured using a Spectra Max GEMINI EM fluorescence reader (Molecular Devices). All studies were performed in triplicate and representative data are shown. Statistical analyses were performed using one-way ANOVA with two-sided post-hoc Dunnett test.

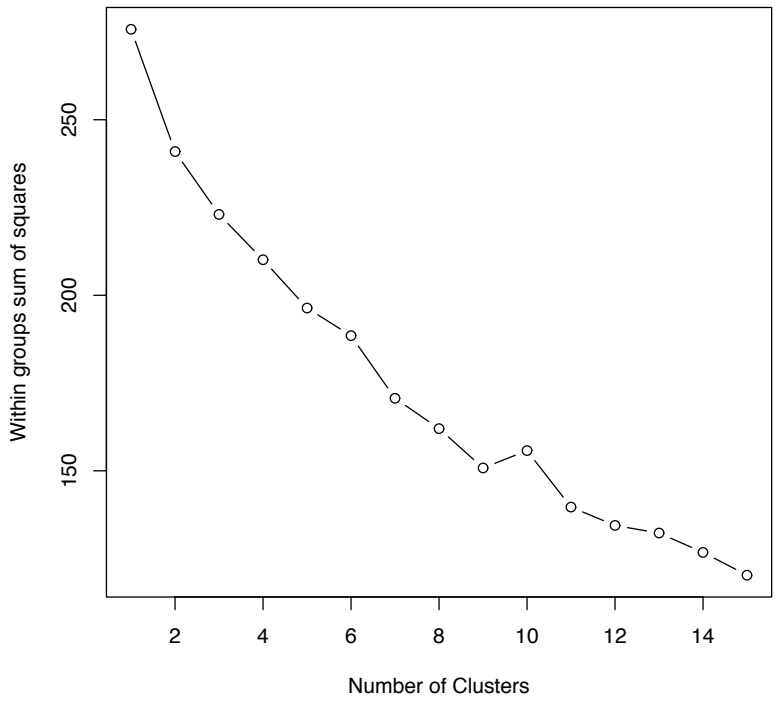
## 5. Kinase panel screen data

Assays were performed by Cerep (Le Bois l'Evêque, B.P. 30001, 86600 Celle l'Evescault, France) in the "ExpresS Diversity Kinase Profile" on a fee-for-service basis (URL: www.cerep.fr).

Compound screening results (**2b**, **4a** and **4b**) against a 48-kinase panel at 10 μM.

| Kinase | % Inhibition | | |
|---|---|---|---|
| | **2b** | **4a** | **4b** |
| Abl kinase | 0 | **87** | **50** |
| Akt1/PKBa | 3 | 0 | 0 |
| AurA kinase | **56** | **52** | 10 |
| CaMK2a | **41** | 11 | **49** |
| CDC2/CDK1 | 0 | 25 | 0 |
| CDK2 | 23 | **58** | 39 |
| CHK1 | 0 | 5 | 0 |
| CHK2 | **65** | 31 | 11 |
| CK1a | 1 | 10 | 12 |
| c-Met kinase | 6 | 20 | 20 |
| EGFR kinase | 2 | 30 | **45** |
| EphA2 kinase | 9 | **40** | **57** |
| EphA3 kinase | 17 | 23 | 33 |
| EphB4 | 0 | 16 | 3 |
| $ERK_2$ | 17 | 3 | 31 |
| FGFR1 kinase | 9 | 5 | 13 |
| FGFR2 kinase | 30 | **62** | **78** |
| FGFR3 kinase | 19 | 9 | 16 |
| GSK3b | 8 | **58** | 33 |
| MAP4K4 | **69** | **95** | **43** |
| IKKa | 0 | 6 | **49** |
| IRAK4 | 0 | 10 | 18 |
| IRK | 6 | 0 | 0 |
| JAK3 | 24 | **45** | 12 |
| JNK1 | 11 | 6 | **40** |
| KDR kinase | 28 | **75** | 2 |
| Lck kinase | 9 | 0 | 5 |
| MAPKAPK2 | 0 | 11 | 0 |
| MARK1 | 4 | 1 | 0 |
| MKK6 | 0 | 1 | 0 |
| MNK2 | 14 | 0 | 3 |
| MST4 kinase | 26 | 10 | 0 |
| NEK2 | 3 | 5 | 5 |
| p38a kinase | 1 | 10 | 20 |
| PAK2 | 4 | 0 | 0 |
| PAK4 | 21 | 1 | 0 |
| PDK1 | 18 | 0 | 10 |
| Pim2 kinase | 3 | 5 | 12 |
| PKA | 4 | 0 | 6 |
| PKCb2 | 0 | 2 | 0 |
| PLK1 | 15 | 24 | **76** |
| RAF-1 kinase | 5 | 20 | 0 |
| ROCK1 | 3 | 5 | 17 |
| SGK1 | 14 | 15 | **49** |
| SIK | 6 | 6 | 0 |
| Src kinase | 0 | **42** | 7 |
| TAOK2 (TAO1) | 21 | **52** | 27 |
| TRKA | 15 | **77** | 0 |

## 6. Optimal number of clusters for *k*-means clustering (Fig S1)




## 7. Scaffold diversity and count (Table S1)

| No. | Scaffold CanonicalSmiles | Scaffold Frequency |
|---|---|---|
| 1 | c1ccc(-n2c(Cc3ncsc3)nc3c2cccc3)cc1 | 21 |
| 2 | c1ccc(Cc2onc(-c3cnccc3)n2)cc1 | 12 |
| 3 | c1ccc(-c2ncn(-c3cnccc3)c2)cc1 | 8 |
| 4 | O=C(NCCc1ccccc1)Cn1cnc2c1ncnc2 | 7 |
| 5 | O=C(Cc1ncsc1)Nc1ccccc1 | 6 |
| 6 | c1ccc(-c2[nH]nc(-c3ncncn3)n2)cc1 | 6 |
| 7 | O=C(C1CC1)Nc1c2c(Nc3ccccc3)nn(-c3cccnc3)c2ncn1 | 4 |
| 8 | O=C(CC#Cc1ccccc1)Nc1scc(Cc2[nH]c3c(cccc3)n2)n1 | 4 |
| 9 | c1ccc(-c2noc(Cc3ncsc3)n2)cc1 | 4 |
| 10 | O=C(NCCc1ccccc1)CCc1onc(-c2cn[nH]c2)n1 | 3 |
| 11 | O=C(Nc1ccccc1)CCc1onc(-c2[nH]cnc2)n1 | 3 |
| 12 | O=C(Nc1ccccc1)CCc1onc(-c2cn[nH]c2)n1 | 3 |
| 13 | c1ccc(Cc2onc(-c3ccc(-c4nc[nH]c4)cc3)n2)cc1 | 3 |
| 14 | c1ccc(NC(Cc2csc(-n3cnc4c(c3=O)cccc4)n2)=O)cc1 | 3 |
| 15 | c1nc2c(n1Cc1onc(-c3ccccc3)n1)ncnc2 | 3 |
| 16 | O=C(Cc1ncsc1)Nc1c(Sc2ccccc2)cccc1 | 2 |
| 17 | O=C(Cc1ncsc1)Nc1ccnn1-c1ccccc1 | 2 |
| 18 | O=C(Cn1cccc1)Nc1ccccc1 | 2 |
| 19 | O=C(NCCc1ccccc1)Cn1cnc2c(CCC[NH+]3CC[NH2+]CC3)ncnc12 | 2 |
| 20 | O=C(Nc1cccc(-c2ccnc(N3CC[NH+](CC4CC4)CC3)c2)c1)Cc1cscn1 | 2 |

| | | |
|---|---|---|
| 21 | O=C(Nc1cccc(-c2cnc(N3CC[NH+](CC4CC4)CC3)cc2)c1)Cc1cscn1 | 2 |
| 22 | O=C(Nc1ccccc1)C1C(c2ccc(-c3noc(Cc4ncsc4)n3)cc2)CC=CC1 | 2 |
| 23 | c1ccc(/C=C/c2ncncn2)cc1 | 2 |
| 24 | c1ccc2c(CCNC(=O)Cc3nc(-n4cnc5c(c4=O)cccc5)sc3)c[nH]c2c1 | 2 |
| 25 | O=C(C1CC1)Nc1ncnc2n(-c3ccncc3)nc(Nc3ccccc3)c12 | 1 |
| 26 | O=C(COc1ccc(-c2nc[nH]c2)cc1)Nc1c(Sc2ccccc2)cccc1 | 1 |
| 27 | O=C(Cc1ccccc1Nc1ccccc1)Nc1c(C2=CC[NH2+]CC2)cccc1 | 1 |
| 28 | O=C(Cc1cscn1)Nc1n(-c2ccccc2)ncc1-c1ccc(N2CC[NH2+]CC2)nc1 | 1 |
| 29 | O=C(Cc1cscn1)Nc1n(-c2ccccc2)ncc1-c1ccnc(N2CC[NH2+]CC2)c1 | 1 |
| 30 | O=C(Cc1ncsc1)Nc1ccc(Oc2ccccc2)cc1 | 1 |
| 31 | O=C(Cc1ncsc1)Nc1ccc(Sc2ccccc2)cc1 | 1 |
| 32 | O=C(Cc1ncsc1)Nc1ccccc1C1=CC[NH2+]CC1 | 1 |
| 33 | O=C(Cc1ncsc1)Nc1ccccc1Sc1ccc(-c2noc(C3=CCC[NH2+]C3)n2)cc1 | 1 |
| 34 | O=C(Cc1ncsc1)Nc1ccnn1-c1ccc(-c2noc(C3=CCC[NH2+]C3)n2)cc1 | 1 |
| 35 | O=C(Cn1cnc2c1ncnc2)NC(=S)Nc1ccccc1 | 1 |
| 36 | O=C(NC(C[NH2+]CC[NH+]1CC[NH2+]CC1)Cc1ccccc1)Cn1cc(Cc2ccccc2)nn1 | 1 |
| 37 | O=C(NC(Cc1ccccc1)Cn1cc(Cc2ccccc2)nn1)OCc1ccccc1 | 1 |
| 38 | O=C(NCCc1c[nH]c2ccc(OCc3ncccc3)cc12)C1CC1 | 1 |
| 39 | O=C(NCCc1ccccc1)COc1ccc(-c2noc(Cc3ncsc3)n2)cc1 | 1 |
| 40 | O=C(Nc1c(-c2nc(-c3conc3-c3ccccc3)on2)cccc1)c1ccccc1 | 1 |
| 41 | O=C(Nc1c(-c2noc(Cc3ncsc3)n2)cccc1)c1ccccc1 | 1 |
| 42 | O=C(Nc1c(Cc2ccccc2)cccc1)COc1ccc(-c2nc[nH]c2)cc1 | 1 |
| 43 | O=C(Nc1c(Cc2ccccc2)cncn1)C1CC1 | 1 |
| 44 | O=C(Nc1c(Cc2ccccc2)cncn1)c1ccn[nH]1 | 1 |
| 45 | O=C(Nc1c(Cc2ccccc2)cncn1)c1cn[nH]c1 | 1 |
| 46 | O=C(Nc1cc(Cc2ccccc2)ccc1)COc1ccc(-c2nc[nH]c2)cc1 | 1 |
| 47 | O=C(Nc1ccc(Cc2ccc([NH2+]Cc3cnc[nH]3)cc2)cc1)C1CC1 | 1 |
| 48 | O=C(Nc1ccc(Cc2ccccc2)cc1)COc1ccc(-c2nc[nH]c2)cc1 | 1 |
| 49 | O=C(Nc1cccc(CC[NH+]2CC[NH2+]CC2)c1)OCC1=C(Cn2cc(Cc3ccccc3)nn2)CC=CC1 | 1 |
| 50 | O=C(Nc1ccccc1-c1ccnc(N2CC[NH+](CC3CC3)CC2)c1)Cc1cscn1 | 1 |
| 51 | O=C(Nc1ccccc1-c1cnc(N2CC[NH+](CC3CC3)CC2)cc1)Cc1cscn1 | 1 |
| 52 | O=C(Nc1ccccc1Sc1ccc(CCC[NH+]2CCN(c3ccccc3)CC2)cc1)C1CC1 | 1 |
| 53 | O=C(c1ccn[nH]1)Nc1cc(CCc2ncccc2)ccc1Nc1ccccc1 | 1 |
| 54 | O=C(c1cn[nH]c1)Nc1cc(CCc2ncccc2)ccc1Nc1ccccc1 | 1 |
| 55 | c1c(Cc2ccccc2)nnn1C1C[NH2+]C(Cc2ccccc2)C1 | 1 |
| 56 | c1c(Cc2ccccc2)nnn1CC([NH2+]c1ncccc1)c1ccccc1 | 1 |
| 57 | c1c(Cc2ccccc2)nnn1CC([NH2+]c1ncncc1)c1ccccc1 | 1 |
| 58 | c1cc(-c2c[nH]cn2)ccc1CC[NH+]1CC[NH2+]CC1 | 1 |
| 59 | c1cc(-c2noc(Cc3ncsc3)n2)cc(C2CC=CCC2)c1 | 1 |
| 60 | c1cc(C(=O)Nc2ccc(Cc3ccc(NC(=O)c4cccnc4)cc3)cc2)[nH]n1 | 1 |
| 61 | c1cc(C(=O)Nc2ccc(S(=O)(=O)Nc3ncccn3)cc2)[nH]n1 | 1 |
| 62 | c1cc(C(=O)Nc2ccc(S(=O)(=O)Nc3scnn3)cc2)[nH]n1 | 1 |
| 63 | c1cc(C(=O)Nc2ccc(S(Nc3ccc(CC[NH+]4CC[NH2+]CC4)nn3)(=O)=O)cc2)[nH]n1 | 1 |
| 64 | c1cc(CC[NH+]2CC[NH2+]CC2)c(-c2[nH]nc(-c3ncncn3)n2)cc1 | 1 |
| 65 | c1cc(CC[NH+]2CC[NH2+]CC2)cc(-c2[nH]nc(-c3ncncn3)n2)c1 | 1 |

| | | |
|---|---|---|
| 66 | c1cc(CCc2scnc2)cc(-c2c[nH]cn2)c1 | 1 |
| 67 | c1cc(C[NH2+]CC[NH+]2CC[NH2+]CC2)cc(-c2noc(Cc3ncsc3)n2)c1 | 1 |
| 68 | c1cc(Cc2nc(-c3cc(NC(=O)NC4=NC(=O)CN4)ccc3)no2)on1 | 1 |
| 69 | c1cc(S(Nc2ccc(CC[NH+]3CC[NH2+]CC3)nn2)(=O)=O)ccc1NC(=O)c1cn[nH]c1 | 1 |
| 70 | c1cc2ccc(NC(C(Cc3cnc[nH]3)NC(C3CC3)=O)=O)cc2cc1 | 1 |
| 71 | c1cc2ncn(-c3ncc(CCNC(=O)C4CC4)s3)c(=O)c2cc1 | 1 |
| 72 | c1ccc(-c2c(NC(=O)C3CC3)nc(-n3cnc4ccccc4c3=O)nn2)cc1 | 1 |
| 73 | c1ccc(-c2ncn(-c3ccncc3)c2)cc1 | 1 |
| 74 | c1ccc(C(=O)Nc2c(CCc3ncccc3)cc(NC(C3CC3)=O)cc2)cc1 | 1 |
| 75 | c1ccc(C(N(CC2CC2)C(C2[NH2+]CC[NH2+]C2)=O)c2ccccc2NC(c2ncn[nH]2)=O)cc1 | 1 |
| 76 | c1ccc(C2(C(NCCc3cnc(-n4cnc5c(c4=O)cccc5)s3)=O)CC2)cc1 | 1 |
| 77 | c1ccc(C2CC2C(NCCc2cnc(-n3cnc4c(c3=O)cccc4)s2)=O)cc1 | 1 |
| 78 | c1ccc(Cc2cn(CC(Nc3scc(CC(Nc4ccccc4C4=CC[NH2+]CC4)=O)n3)=O)nn2)cc1 | 1 |
| 79 | c1ccc(NC(Cn2cc(COC(=O)Nc3ccccc3)nn2)=O)cc1 | 1 |
| 80 | c1cn(CC(Nc2scc(CC(Nc3ccccc3C3=CC[NH2+]CC3)=O)n2)=O)nn1 | 1 |
| 81 | c1cnc(NS(=O)(=O)c2ccc(NC(=O)c3cn[nH]c3)cc2)nc1 | 1 |
| 82 | c1cnc(NS(c2ccc(NC(=O)C3CC3)cc2)(=O)=O)nc1 | 1 |
| 83 | c1cncc(C(=O)Nc2ccc(Cc3ccc(NC(c4cn[nH]c4)=O)cc3)cc2)c1 | 1 |
| 84 | c1coc(NS(c2ccc(NC(=O)C3CC3)cc2)(=O)=O)n1 | 1 |
| 85 | n1cnc(-c2n[nH]c(-c3ccc(CC[NH+]4CC[NH2+]CC4)cc3)n2)nc1 | 1 |
| 86 | n1cnc(NS(=O)(=O)c2ccc(NC(=O)C3CC3)cc2)cc1 | 1 |
| 87 | n1csc(NS(c2ccc(NC(=O)C3CC3)cc2)(=O)=O)n1 | 1 |
| 88 | n1csc(NS(c2ccc(NC(=O)c3cn[nH]c3)cc2)(=O)=O)n1 | 1 |

## References

(1) M. Rupp, E. Proschak, G. Schneider, *J. Chem. Inf. Model.* **2007**, *47*, 2280-2286.

(2) M. Rupp, G. Schneider, *Mol. Inf.* **2010**, *29*, 266-273.

(3) G. W. Bemis, M. A. Murcko, *J. Med. Chem.* **1996**, *39*, 2887-2893.

(4) M. R. Berthold, N. Cebron, F. Dill, T. R. Gabriel, T. Kötter, T. Meinl, P. Ohl, K. Thiel, B. Wiswedel, *ACM SIGKDD Explorations Newsletter* **2009**, *11*, 26-31.

(5) RDKit: Open-source cheminformatics; http://www.rdkit.org.

(6) D. Weininger, A. Weininger, J. L. Weininger, *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 97-101.

(7) G. Schneider, W. Neidhart, T. Giller, G. Schmid, *Angew. Chem. Int. Ed.* **1999**, *38*, 2894-2896.

(8) G. Hinton, S. T. Roweis, *Adv. Neural Inf. Process. Syst.* **2002**, *15*, 833-840.

(9) H. Steinhaus, *Bull. Acad. Polon. Sci* **1956**, *1*, 801-804.

(10) N. Saitou, M. Nei, *Mol. Biol. Evol.* **1987**, *4*, 406-425.

(11) T. Anastassiadis, S. W. Deacon, K. Devarajan, H. Ma, J. R. Peterson, *Nat. Biotechnol.* **2011**, *29*, 1039-1045.

(12) G. Jones, P. Willett, R. C. Glen, *J. Mol. Biol.* **1995**, *245*, 43-53.

(13) B. Zhao, A. Smallwood, J. Yang, K. Koretke, K. Nurse, A. Calamari, R. B. Kirkpatrick, Z. Lai, *Protein Sci.* **2008**, *17*, 1791-1797.

(14) G. Jones, P. Willett, R. C. Glen, A. R. Leach, R. Taylor, *J. Mol. Biol.* **1997**, *267*, 727–748.

(15) J. W. Bae, Y. J. Cho, S. H. Lee, C.-O. M. Yoon, C. M. Yoon, *Chem. Commun.* **2000**, 1857

(16) M. Detmar, S. Tenorio, U. Hettmannsperger, Z. Ruszczak, C. E. Orfanos, *J. Invest. Dermatol.* **1992**, *98*, 147-153.