**CReM: chemically reasonable mutations framework for structure generation**

Pavel Polishchuk

Table S1. PAINS patterns found in stochastically generated compounds using the PAINS-less ChEMBL fragment database and context radius 1.

|  |  |  |
| --- | --- | --- |
|  | Number of compounds | PAINS pattern |
|  | 10 | anil\_alk\_C(1) |
|  | 1 | anil\_alk\_D(1) |
|  | 626001 | anil\_di\_alk\_A(478) |
|  | 31554 | anil\_di\_alk\_B(251) |
|  | 34928 | anil\_di\_alk\_C(246) |
|  | 332531 | anil\_di\_alk\_D(198) |
|  | 743604 | anil\_di\_alk\_E(186) |
|  | 155 | anil\_di\_alk\_F(14) |
|  | 4544 | anil\_di\_alk\_G(9) |
|  | 2 | anil\_di\_alk\_I(4) |
|  | 4 | anil\_di\_alk\_J(3) |
|  | 20 | anil\_di\_alk\_K(2) |
|  | 1309 | anil\_NH\_alk\_B(3) |
|  | 43 | anil\_NH\_alk\_C(2) |
|  | 4 | anil\_NH\_alk\_D(2) |
|  | 19 | anil\_no\_alk(40) |
|  | 1 | anil\_OC\_alk\_D(2) |
|  | 2 | anil\_OC\_alk\_E(1) |
|  | 2 | anil\_OC\_alk\_F(1) |
|  | 3 | anil\_OC\_no\_alk\_A(8) |
|  | 4 | anil\_OC\_no\_alk\_C(3) |
|  | 23 | anisol\_A(5) |
|  | 22 | anisol\_B(2) |
|  | 3 | anthranil\_acid\_A(19) |
|  | 1 | anthranil\_acid\_G(1) |
|  | 350 | anthranil\_one\_A(38) |
|  | 17 | catechol\_A(92) |
|  | 1 | cyanamide\_A(1) |
|  | 5 | cyano\_cyano\_B(3) |
|  | 2 | cyano\_ene\_amine\_C(3) |
|  | 1 | cyano\_imine\_B(17) |
|  | 263 | cyano\_keto\_A(2) |
|  | 10 | dhp\_keto\_A(9) |
|  | 2 | diazox\_A(3) |
|  | 1170 | dyes5A(27) |
|  | 66 | ene\_cyano\_A(19) |
|  | 6 | ene\_cyano\_C(6) |
|  | 259 | ene\_cyano\_E(1) |
|  | 1 | ene\_cyano\_G(1) |
|  | 2 | ene\_one\_B(2) |
|  | 60 | ene\_one\_D(1) |
|  | 45 | ene\_one\_ester(24) |
|  | 662 | ene\_one\_hal(17) |
|  | 72 | ene\_one\_yne\_A(1) |
|  | 2 | ene\_rhod\_C(13) |
|  | 4 | furan\_acid\_A(4) |
|  | 2 | het\_5\_B(4) |
|  | 99 | het\_thio\_N\_5B(2) |
|  | 2 | het\_thio\_N\_5C(1) |
|  | 87 | hzone\_anil\_di\_alk(35) |
|  | 163 | hzone\_enamin(30) |
|  | 1 | hzone\_furan\_A(6) |
|  | 261 | hzone\_phenol\_A(479) |
|  | 21 | hzone\_phenol\_B(215) |
|  | 3 | hzone\_pyrrol(64) |
|  | 4 | hzone\_thiophene\_B(4) |
|  | 4 | imidazole\_B(2) |
|  | 6 | imine\_ene\_A(5) |
|  | 13771 | imine\_one\_A(321) |
|  | 5 | imine\_one\_B(4) |
|  | 184 | imine\_phenol\_A(3) |
|  | 22 | indol\_3yl\_alk(461) |
|  | 2 | keto\_keto\_beta\_D(5) |
|  | 5509 | mannich\_A(296) |
|  | 262 | pyrrole\_A(118) |
|  | 5 | pyrrole\_B(29) |
|  | 15 | pyrrole\_C(8) |
|  | 1 | pyrrole\_D(5) |
|  | 27 | pyrrole\_E(5) |
|  | 1 | pyrrole\_G(4) |
|  | 16 | pyrrole\_L(1) |
|  | 3 | pyrrole\_M(1) |
|  | 23 | pyrrole\_N(1) |
|  | 17 | pyrrole\_O(1) |
|  | 3 | rhod\_sat\_C(3) |
|  | 48 | sulfonamide\_B(41) |
|  | 3 | sulfonamide\_E(2) |
|  | 3 | sulfonamide\_F(1) |
|  | 7 | thiaz\_ene\_B(17) |
|  | 92 | thiazole\_amine\_A(4) |
|  | 1 | thiazole\_amine\_G(2) |
|  | 247 | thiazol\_SC\_A(3) |
|  | 5 | thio\_amide\_C(2) |
|  | 39 | thio\_amide\_D(2) |
|  | 1 | thio\_amide\_E(1) |
|  | 7 | thio\_amide\_F(1) |
|  | 84 | thio\_carbam\_A(1) |
|  | 40 | thio\_ester\_B(4) |
|  | 1 | thio\_keto\_het(2) |
|  | 5889 | thio\_ketone(43) |
|  | 4 | thiophene\_amino\_Ab(40) |
|  | 13 | thiophene\_amino\_B(12) |
|  | 1 | thiophene\_amino\_D(3) |
|  | 2 | thiophene\_amino\_G(2) |
|  | 15 | thiophene\_amino\_H(2) |
|  | 1 | thiophene\_D(2) |
|  | 1 | thiophene\_F(1) |
|  | 6 | thiophene\_hydroxy(28) |
|  | 2 | thio\_urea\_C(9) |
|  | 2 | thio\_urea\_D(8) |
|  | 1 | thio\_urea\_G(5) |
|  | 2 | thio\_urea\_L(1) |

Table S2. PAINS patterns found in stochastically generated compounds using the PAINS-less ChEMBL fragment database and context radius 2.

|  |  |  |
| --- | --- | --- |
|  | Number of compounds | PAINS pattern |
|  | 7 | anil\_alk\_D(1) |
|  | 316 | anil\_di\_alk\_A(478) |
|  | 639 | anil\_di\_alk\_B(251) |
|  | 27928 | anil\_di\_alk\_C(246) |
|  | 863 | anil\_di\_alk\_D(198) |
|  | 343 | anil\_di\_alk\_E(186) |
|  | 3 | anil\_di\_alk\_F(14) |
|  | 2 | anil\_di\_alk\_I(4) |
|  | 21 | anil\_NH\_alk\_B(3) |
|  | 11 | anil\_no\_alk(40) |
|  | 2 | anil\_OC\_alk\_E(1) |
|  | 2 | anil\_OH\_alk\_A(8) |
|  | 1 | anisol\_A(5) |
|  | 32 | anisol\_B(2) |
|  | 13 | anthranil\_one\_A(38) |
|  | 11 | catechol\_A(92) |
|  | 1 | cyano\_ene\_amine\_C(3) |
|  | 5 | cyano\_keto\_A(2) |
|  | 1 | diazox\_A(3) |
|  | 245 | dyes5A(27) |
|  | 10 | ene\_cyano\_A(19) |
|  | 3 | ene\_cyano\_E(1) |
|  | 2 | ene\_one\_D(1) |
|  | 25 | ene\_one\_hal(17) |
|  | 10 | ene\_rhod\_B(16) |
|  | 2 | het\_55\_A(2) |
|  | 1 | hzone\_enamin(30) |
|  | 26 | hzone\_phenol\_A(479) |
|  | 84 | hzone\_phenol\_B(215) |
|  | 7 | hzone\_pipzn(79) |
|  | 4 | hzone\_thiophene\_B(4) |
|  | 5 | imine\_one\_A(321) |
|  | 10 | imine\_phenol\_A(3) |
|  | 135 | mannich\_A(296) |
|  | 100 | pyrrole\_A(118) |
|  | 3 | pyrrole\_D(5) |
|  | 2 | pyrrole\_E(5) |
|  | 3 | pyrrole\_N(1) |
|  | 6 | sulfonamide\_B(41) |
|  | 17 | tetrazole\_A(1) |
|  | 4 | thiaz\_ene\_B(17) |
|  | 19 | thiazole\_amine\_B(3) |
|  | 26 | thiazole\_amine\_L(1) |
|  | 66 | thio\_carbam\_A(1) |
|  | 2 | thio\_ester\_B(4) |
|  | 2 | thio\_ketone(43) |
|  | 13 | thiophene\_amino\_Ab(40) |
|  | 13 | thiophene\_amino\_G(2) |
|  | 4 | thiophene\_D(2) |
|  | 1 | thiophene\_E(2) |
|  | 25 | thiophene\_hydroxy(28) |
|  | 1 | thio\_thiomorph\_Z(1) |

Table S3. PAINS patterns found in stochastically generated compounds using the PAINS-less ChEMBL fragment database and context radius 3.

|  |  |  |
| --- | --- | --- |
|  | Number of compounds | PAINS pattern |
|  | 767 | anil\_di\_alk\_A(478) |
|  | 181 | anil\_di\_alk\_B(251) |
|  | 214474 | anil\_di\_alk\_C(246) |
|  | 656 | anil\_di\_alk\_D(198) |
|  | 1305 | anil\_di\_alk\_E(186) |
|  | 5 | anil\_di\_alk\_F(14) |
|  | 32 | anil\_di\_alk\_G(9) |
|  | 38 | anil\_di\_alk\_I(4) |
|  | 171 | anil\_NH\_alk\_B(3) |
|  | 36 | anil\_no\_alk(40) |
|  | 2 | anil\_OC\_no\_alk\_A(8) |
|  | 3 | anil\_OH\_alk\_A(8) |
|  | 529 | dyes5A(27) |
|  | 73 | ene\_cyano\_A(19) |
|  | 1 | ene\_rhod\_C(13) |
|  | 8 | het\_thio\_5\_A(8) |
|  | 51 | het\_thio\_5\_C(2) |
|  | 2 | hzone\_phenol\_A(479) |
|  | 1 | hzone\_phenol\_B(215) |
|  | 4 | hzone\_phenone(7) |
|  | 11 | imine\_one\_A(321) |
|  | 185 | mannich\_A(296) |
|  | 96 | pyrrole\_A(118) |
|  | 2 | pyrrole\_D(5) |
|  | 4 | pyrrole\_G(4) |
|  | 3 | thiazole\_amine\_B(3) |
|  | 1 | thio\_urea\_D(8) |
|  | 5 | thio\_urea\_E(7) |

Table S4. PAINS patterns found in stochastically generated compounds using the PAINS-less ChEMBL fragment database and context radius 4.

|  |  |  |
| --- | --- | --- |
|  | Number of compounds | PAINS pattern |
|  | 1 | anil\_alk\_B(1) |
|  | 3 | anil\_alk\_bim(9) |
|  | 13 | anil\_alk\_D(1) |
|  | 1249 | anil\_di\_alk\_A(478) |
|  | 615 | anil\_di\_alk\_B(251) |
|  | 4887 | anil\_di\_alk\_C(246) |
|  | 10556 | anil\_di\_alk\_D(198) |
|  | 12476 | anil\_di\_alk\_E(186) |
|  | 2 | anil\_di\_alk\_F(14) |
|  | 66 | anil\_di\_alk\_G(9) |
|  | 59 | anil\_NH\_alk\_B(3) |
|  | 5 | anil\_NH\_alk\_D(2) |
|  | 111 | anil\_no\_alk(40) |
|  | 9 | anisol\_B(2) |
|  | 1 | ene\_rhod\_B(16) |
|  | 1 | het\_thio\_5\_C(2) |
|  | 1 | hzone\_anil(14) |
|  | 2 | hzone\_phenol\_A(479) |
|  | 1 | hzone\_phenol\_B(215) |
|  | 3 | hzone\_phenone(7) |
|  | 1 | mannich\_A(296) |
|  | 4 | misc\_aminoacid\_A(1) |
|  | 1 | pyrrole\_G(4) |
|  | 120 | thiazole\_amine\_B(3) |
|  | 1 | thiazol\_SC\_A(3) |
|  | 1 | thio\_urea\_B(9) |

Table S5. PAINS patterns found in stochastically generated compounds using the PAINS-less ChEMBL fragment database and context radius 5.

|  |  |  |
| --- | --- | --- |
|  | Number of compounds | PAINS pattern |
| 1. | 29 | anil\_di\_alk\_C(246) |