Table S7. Key Descriptors Utilized in the Regression Model Constructed using the Set 4 descriptors (functional group counts, atom-centered fragments, atom-type E-state indices, and pharmacophore descriptors) (model no. 15 in Table 2). The model employed KKN as the regression algorithm with 'Boruta’ as a feature selection method.

| **Descriptor** | **Correlated Descriptors** | **Correlation coefficient(s)** | **Activity Relationship** |
| --- | --- | --- | --- |
| SaaaC (Sum of aaaC E-states, i.e. E-states for aromatic carbon atoms that have no hydrogen atoms attached and are connected to three other aromatic atoms) | NaaaC (Number of atoms of type aaaC) | r=0.98 | Lower values → higher activity |
| C-033 (R–CH..X) | None | N/A | Higher values → higher activity |
| nCconj (number of non-aromatic conjugated C(sp2)) | None | N/A | Higher values → higher activity |
| C-034 (R–CR..X) | nPyrroles (number of pyrrole rings), N-073 (Ar2NH / Ar3N / Ar2N-Al / R..N..R), SaasN (sum of aasN E-states), NaasN (number of atoms of type aasN) | r=0.89 – 0.90 | Higher values → very slightly higher activity |
| SaaCH (Sum of aaCH E-states) | nBM (number of multiple bonds)  nAB (number of aromatic bonds)  nCsp2 (number of sp2 hybridized Carbon atoms)  ARR (aromatic ratio)  piPC04 (molecular multiple path count of order 4)  piPC05 (molecular multiple path count of order 5)  P\_VSA\_e\_2 (P\_VSA-like on Sanderson electronegativity, bin 2)  P\_VSA\_i\_2 (P\_VSA-like on ionization potential, bin 2)  P\_VSA\_s\_4 (P\_VSA-like on I-state, bin 4)  VSA\_ppp\_con (P\_VSA-like on potential pharmacophore points, conjugated atoms)  P\_VSA\_ppp\_cyc (P\_VSA-like on potential pharmacophore points, atoms belonging to cycles)  Eta\_betaP\_A (eta pi and lone pair average VEM count)  Eta\_D\_beta\_A ( eta average measure of electronic features)  SpMAD\_EA(bo) (spectral mean absolute deviation from edge adjacency mat. weighted by bond order)  nCar (number of aromatic C(sp2))  C-025 (R–CR–R)  Uc (unsaturation count)  Ui (unsaturation index)  P\_VSA\_MR\_6 (P\_VSA-like on Molar Refractivity, bin 6)  P\_VSA\_ppp\_ar (P\_VSA-like on potential pharmacophore points, aromatic atoms)  nCbH (number of unsubstituted benzene C(sp2))  C-024 (R–CH–R)  NaaCH (Number of atoms of type aaCH) | r > 0.80  r > 0.90 | Higher values → slightly higher activity |
| C-001 (CH3R / CH4) | P\_VSA\_LogP\_1 (P\_VSA-like on LogP, bin 1)  NsCH3 (Number of atoms of type sCH3)  nCp (number of terminal primary C(sp3))  SsCH3 (Sum of sCH3 E-states) | r > 0.89  r > 0.90 | Higher values → very slightly higher activity |
| nPyrroles (number of pyrrole rings) | C-034 (R–CR..X)  SaasN (Sum of aasN E-states) | r > 0.81 | Higher values → very slightly higher activity |
| H-053 (H attached to C0(sp3) with 2X attached to next C) | None | NA | A flattened inverted U shape |
| SssCH2 (Sum of ssCH2 E-states) | X1Av (average valence connectivity index of order 1)  X3Av (average valence connectivity index of order 3)  X4Av (average valence connectivity index of order 4)  X5Av (average valence connectivity index of order 5)  ChiA\_B(s) (average Randic-like index from Burden matrix weighted by I-State)  Eta\_L\_A (eta average local composite index)  C-002 (CH2R2)  SpPosA\_B(e) (normalized spectral positive sum from Burden matrix weighted by Sanderson electronegativity)  SpPosA\_B(i) (normalized spectral positive sum from Burden matrix weighted by ionization potential)  AVS\_B(s) (average vertex sum from Burden matrix weighted by I-State | r > 0.80  r < - 0.80 | Higher values → slightly lower activity |

Table S8. Key Descriptors Utilized in the Regression Model Constructed using the Set 4 descriptors (functional group counts, atom-centered fragments, atom-type E-state indices, and pharmacophore descriptors) (model no. 16 in Table 2). The model employed BART as the regression algorithm and 'gaselect’ as a feature selection method.

| **Descriptor** | **Correlated Descriptors** | **Correlation coefficient(s)** | **Activity Relationship** |
| --- | --- | --- | --- |
| CATS2D\_07\_DL (CATS2D Donor-Lipophilic at lag 07) | CATS2D\_02\_DL (CATS2D Donor-Lipophilic at lag 02)  CATS2D\_09\_DL (CATS2D Donor-Lipophilic at lag 09) | r > 0.81 | Higher values → higher activity |
| C-034 (R–CR..X) | nPyrroles (number of pyrrole rings), N-073 (Ar2NH / Ar3N / Ar2N-Al / R..N..R), SaasN (sum of aasN E-states), NaasN (number of atoms of type aasN) | r= 0.89 – 0.90 | Higher values → higher activity |
| nPyrimidines (number of Pyrimidines) | nSO2N (number of sulfonamides (thio-/dithio-)  C-032 (X–CX–X)  N-071 (Ar-NAl2)  S-110 (R-SO2-R)  NddssS (number of atoms of type ddssS)  B01[N-S] (presence/absence of N – S at topological distance 1)  B01[O-S] (presence/absence of O – S at topological distance 1)  B03[N-S] (presence/absence of N – S at topological distance 3)  B09[O-S] (presence/absence of O – S at topological distance 9)  B09[S-F] (presence/absence of S – F at topological distance 9)  F01[N-S] (frequency of N – S at topological distance 1)  F01[O-S] (frequency of O – S at topological distance 1)  F03[N-S] (frequency of N – S at topological distance 3)  F09[O-S] (frequency of O – S at topological distance 9)  F09[S-F] (frequency of S – F at topological distance 9) | r= 0.80 – 0.87 | Higher values → higher activity |
| CATS2D\_07\_DA (CATS2D Donor-Acceptor at lag 07) | None | N/A | Higher values → lower activity |
| C-033 (R–CH..X) | None | N/A | Higher values → higher activity |
| CATS2D\_03\_DL (CATS2D Donor-Lipophilic at lag 03) | TIE (E-state topological parameter)  BAC (Balaban centric index)  J\_A (Balaban-like index from adjacency matrix)  J\_X (Balaban-like index from chi matrix)  H\_Dt (Harary-like index from detour matrix)  Wi\_D/Dt (Wiener-like index from distance/detour matrix)  AVS\_Dz(v) (average vertex sum from Barysz matrix weighted by van der Waals volume)  SpAbs\_Dz(v) (graph energy from Barysz matrix weighted by van der Waals volume)  SpPos\_Dz(v) (spectral positive sum from Barysz matrix weighted by van der Waals volume)  SpMax\_Dz(v) (leading eigenvalue from Barysz matrix weighted by van der Waals volume)  SpAD\_Dz(v) (spectral absolute deviation from Barysz matrix weighted by van der Waals volume)  J\_B(m) (Balaban-like index from Burden matrix weighted by mass)  J\_B(v) (Balaban-like index from Burden matrix weighted by van der Waals volume)  J\_B(p) Balaban-like index from Burden matrix weighted by polarizability  J\_B(i) (Balaban-like index from Burden matrix weighted by ionization potential)  ATSC5e (Centred Broto-Moreau autocorrelation of lag 5 weighted by Sanderson electronegativity)  ATSC6e (Centred Broto-Moreau autocorrelation of lag 6 weighted by Sanderson electronegativity)  ATSC5s (Centred Broto-Moreau autocorrelation of lag 5 weighted by I-state)  ATSC6s (Centred Broto-Moreau autocorrelation of lag 6 weighted by I-state)  ATSC7s (Centred Broto-Moreau autocorrelation of lag 7 weighted by I-state)  P\_VSA\_LogP\_4 (P\_VSA-like on LogP, bin 4)  P\_VSA\_MR\_2 (P\_VSA-like on Molar Refractivity, bin 2)  P\_VSA\_v\_2 (P\_VSA-like on van der Waals volume, bin 2)  P\_VSA\_p\_2 (P\_VSA-like on polarizability, bin 2)  nHDon (number of donor atoms for H-bonds (N and O))  H-050 (H attached to heteroatom)  SdO (Sum of dO E-states)  CATS2D\_00\_DD (CATS2D Donor-Donor at lag 00 (number of H bond donor atoms))  CATS2D\_03\_DD (CATS2D Donor-Donor at lag 03)  CATS2D\_08\_DD (CATS2D Donor-Donor at lag 08)  CATS2D\_09\_DD (CATS2D Donor-Donor at lag 09)  CATS2D\_03\_DA (CATS2D Donor-Acceptor at lag 03)  CATS2D\_05\_DA (CATS2D Donor-Acceptor at lag 05)  CATS2D\_08\_DA (CATS2D Donor-Acceptor at lag 08)  CATS2D\_09\_DA (CATS2D Donor-Acceptor at lag 09)  CATS2D\_04\_DL (CATS2D Donor-Lipophilic at lag 04)  F04[C-O] (Frequency of C – O at topological distance 4)  F08[O-O] (Frequency of O – O at topological distance 8)  Hy (hydrophilic factor)  TPSA(NO) (topological polar surface area using N,O polar contributions)  TPSA(Tot) (topological polar surface area using N,O,S,P polar contributions)  SAacc (surface area of acceptor atoms from P\_VSA-like descriptors) | r > 0.80 | Higher values → lower activity |
| SaaaC (Sum of aaaC E-states, i.e. E-states for aromatic carbon atoms that have no hydrogen atoms attached and are connected to three other aromatic atoms) | NaaaC (Number of atoms of type aaaC) | r=0.98 | Lower values → higher activity |
| O-056 (alcohol) | P\_VSA\_MR\_3 ( P\_VSA-like on Molar Refractivity, bin 3)  nOHs (number of secondary alcohols)  nROH (number of hydroxyl groups) | r > 0.96  r = 0.81 | Higher values → lower activity |
| C-002 (CH2R2) | X3Av (average valence connectivity index of order 3)  X4Av (average valence connectivity index of order 4)  X5Av (average valence connectivity index of order 5)  ChiA\_B(s) (average Randic-like index from Burden matrix weighted by I-State)  Eta\_L\_A (eta average local composite index)  SssCH2 (Sum of ssCH2 E-state)  nCs (number of total secondary C(sp3))  NssCH2 (Number of atoms of type ssCH2) | r > 0.80  r > 0.90 | Sawtooth-like curve (maximum activity for lowest value) |
| C-006 (CH2RX) | None | N/A | Higher values → lower activity |

Table S9. Key Descriptors Utilized in the Regression Model Constructed using the Set 4 descriptors (functional group counts, atom-centered fragments, atom-type E-state indices, and pharmacophore descriptors) (model no. 20 in Table 2). The model employed BART as the regression algorithm and 'Boruta’ as a feature selection method.

| **Descriptor** | **Correlated Descriptors** | **Correlation coefficient(s)** | **Activity Relationship** |
| --- | --- | --- | --- |
| nPyrroles (number of pyrrole rings) | C-034 (R–CR..X)  SaasN (Sum of aasN E-states) | r > 0.81 | Higher values → very slightly higher activity |
| H-053 (H attached to C0(sp3) with 2X attached to next C) | None | NA | An inverted U shape |
| nCconj (number of non-aromatic conjugated C(sp2)) | None | N/A | Higher values → higher activity (slight effect) |
| H-046 (H attached to C0(sp3) no X attached to next C) | H% (percentage of H atoms)  X0Av (average valence connectivity index of order 0)  Eta\_L\_A (eta average local composite index)  SpPosA\_B(e) (normalized spectral positive sum from Burden matrix weighted by Sanderson electronegativity)  SpPosA\_B(i) ( normalized spectral positive sum from Burden matrix weighted by ionization potential)  Eta\_beta\_A (eta average VEM count)  Eta\_FL\_A (eta average local functionality index) | r > 0.81  r < -0.82 | Sawtooth-like curve (maximum activity for lowest value) |
| SaaaC (Sum of aaaC E-states, i.e. E-states for aromatic carbon atoms that have no hydrogen atoms attached and are connected to three other aromatic atoms) | NaaaC (Number of atoms of type aaaC) | r=0.98 | Sawtooth-like curve (maximum activity at minimum values) |
| CATS2D\_06\_AL (CATS2D Acceptor-Lipophilic at lag 06) | CATS2D\_07\_AL (CATS2D Acceptor-Lipophilic at lag 07) | r = 0.87 | Higher values → lower activity |
| C-033 (R–CH..X) | None | N/A | Higher values → higher activity |
| CATS2D\_04\_AA (CATS2D Acceptor-Acceptor at lag 04) | F04[O-O] (Frequency of O – O at topological distance 4) | r=0.81 | ≥3 → Stronger activity |
| C-034 (R–CR..X) | nPyrroles (number of pyrrole rings), N-073 (Ar2NH / Ar3N / Ar2N-Al / R..N..R), SaasN (sum of aasN E-states), NaasN (number of atoms of type aasN) | R=0.89 – 0.90 | Inverted U-shaped curve |
| nCrt (number of ring tertiary C) | nCt, C-003, SpMin1\_Bh(s) (smallest eigenvalue n. 1 of Burden matrix weighted by I-state) | 0.80 – 0.88 | U-shaped curve |