Exploring the degree of long-range order/disorder in indaceno-based photovoltaic small molecules by data-driven machine learning analysis

Hussein A.K. Kyhoiesh[1*], Karrar H. Salem[2], Azal S. Waheeb[3], Riyam A. Hasan[4], Hayder R. Salman[5], Ahmed A. Al-Kubaisi[6], Ashraf Y. Elnaggar[7], Islam H. El Azab[7], Mohamed H.H. Mahmoud[8]

[1]*National University of Science and Technology, Nasiriyah, Dhi Qar, 64001, Iraq*

[2]*College of Medical and Health Technologies, Al-Zahraa University for Women, Karbala, Iraq*

[3]*Department of Chemistry, College of Science, Al-Muthanna University, AL-Muthanna, Iraq*

[4]*Mazaya University College, Al-Zaytoun Street, Nasiriyah, Dhi-Qar, 64001, Iraq*

[5]*College of Pharmacy, Al-Mustaqbal University, 51001, Babylon, Iraq*

[6]*Department of Medical Laboratories Techniques, College of Health and Medical Technology, University of Al Maarif, Al Anbar, 31001, Iraq*

[7]*Department of Food Sciences and Nutrition, College of Science, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia*

[8]*Department of Chemistry, College of Science, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia*

[*]*Corresponding author: H.A.K. Kyhoiesh; Email: hussein.k.sultan@nust.edu.iq; phone; (+964)7807229491;*
https://orcid.org/0000-0001-9705-9196

Feature Extractions

The calculated descriptors were correlated with target properties to demonstrate their correlations. Among their most common descriptors, their molar weights (MW) were calculated by following equation (Eq.1).

$$MW = \sum_{i=1}^{n} m_i \qquad \text{(Eq. 1)}$$

here m_i is their atomic masses for an at i in the molecule.

The partition coefficient (LogP) is a measure for their hydrophobicity and is defined as a logarithm of their ratio of equilibrium concentrations in octanol and water (Eq.2). It indicates that how well a substance can partition for hydrophobic (like octanol) and a hydrophilic (like water) environment.

$$LogP = Log_{10}\left(\frac{C_{octanol}}{C_{water}}\right) \qquad (Eq.\ 2)$$

The electronegativity (EN) of an atom is a measure for their tendency to attract a bonding pair of electrons. It is a key concept to understand their chemical bonding reactivity to influence that how atoms can interact with each other for their molecules (Eq.3).

$$EN = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (Eq.\ 3)$$

where χ_i is the electronegativity of atom i_i.To calculate the zero-order molecular valence connectivity index (χ_o^v) [24], the hydrogen-suppressed skeleton of a molecule was used (Eq.4). The non-hydrogen atom with a distinct δ^v value, was reflected by their atomic valence delta (δ^v) [25].

$$\chi_o^v = \sum_{i=10}^{n} (\delta^v)^{-0.6} \qquad \text{(Eq. 4)}$$

Similarly, for each non-hydrogen atom in the molecule, the atomic number (Z), valence electrons (Z^v), and attached hydrogen atoms (h) were used to calculate the δ^v. By combining these factors, the formula generated a unique δ^v value for each atom (Eq. 4).

$$\delta = \frac{Z^h - h}{Z - Z^h - 1} \qquad \text{(Eq. 4)}$$

Correlations and Feature Scores

The Pearson correlation coefficient (r) [27] is used to quantify linear relationship between two variables like XX and Y*Y (*Eq. 5).

$$r = \frac{n\left(\sum XY\right) - \left(\sum X\right)\left(\sum Y\right)}{\sqrt{\left[n\sum X^2 - \left(\sum X\right)^2\right]}\sqrt{\left[n\sum Y^2 - \left(\sum Y\right)^2\right]}} \qquad (Eq.\ 5)$$

Where n denotes the number of data points, ΣXY is the sum of the product of paired scores, ΣX sum of X scores, ΣY represent the sum of Y scores, ΣX2is the sum of squared XX scores and ΣY2 shows sum of squared YY scores. While their feature importance scores are calculated by depending upon their applied models. For tree-based models, it can be calculated by their decrease in impurity such as Gini impurity or entropy to be brought by each feature of a trees in the model (Eq. 6).

$$FI_j = \frac{1}{T}\sum_{t=1}^{T} \Delta I_{jt} \qquad (Eq.\ 6)$$

Where *FIj* represent their feature importance score for feature *j*, *T* is its total number of trees in the model and Δ*Ijt* shows a decrease in its impurity for feature *j* in tree *t*

Reproducibility Protocol

To ensure the reproducibility of experiments, specific protocols regarding hyperparameters and computational settings were implemented. Random seeds were fixed at 42 for all stochastic operations in Python's random module and NumPy to maintain consistency in results. Data was partitioned using stratified splitting based on crystallinity labels, utilizing StratifiedShuffleSplit from Scikit-Learn, with the following ratios: 60% for training, 15% for validation, and 25% for testing. Additionally, all descriptors were standardized using StandardScaler to ensure each feature has a mean of 0 and a variance of 1, which enhances the performance of many machine learning algorithms.

Table S1. A sample of collected dataset with its SMILES notations for small molecules

No.	SMILES
1	CCCCC(CC)CC1(C2=CC3=C(C=C2C4=C1C=C(S4)C5=CC=C(C6=NSN=C56)/C=C\7/C(=O)N(C(=S)S7)CC)C(C8=C3SC(=C8)C9=CC=C(C1=NSN=C91)/C=C\1/C(=O)N(C(=S)S1)CC)(CC(CC)CCCC)CC(CC)CCCC)CC(CC)CCCC
2	CCCCC(CC)CC1(C2=CC3=C(C=C2C4=C1C=C(S4)C5=CC=C(C6=NSN=C56)C=C7C(=O)N(C(=S)S7)CC)C(C8=C3SC(=C8)C9=CC=C(C1=NSN=C91)C=C1C(=O)N(C(=S)S1)CC)(CC(CC)CCCC)CC(CC)CCCC)CC(CC)CCCC
3	CCCCC(CC)CC1(C2=CC3=C(C=C2C4=C1C=C(S4)C5=CC=C(C6=NSN=C56)/C=C/7\C(=O)N(C(=S)S7)CC)C(C8=C3SC(=C8)C9=CC=C(C1=NSN=C91)/C=C\1/C(=O)N(C(=S)S1)CC)(CC(CC)CCCC)CC(CC)CCCC)CC(CC)CCCC
4	CCCCCCCCC1(C2=CC3=C(C=C2C4=C1C=C(S4)C5=CC=C(C6=NSN=C56)/C=C\7/C(=O)N(C(=S)S7)CC)C(C8=C3SC(=C8)C9=CC=C(C1=NSN=C91)/C=C\1/C(=O)N(C(=S)S1)CC)(CCCCCCCC)CCCCCCCC)CCCCCCCC
5	CCCCCCCCC1(C2=CC3=C(C=C2C4=C1C=C(S4)C5=CC=C(C6=NSN=C56)/C=C\7/C(=O)N(C(=S)S7)CC)C(C8=C3SC(=C8)C9=CC=C(C1=NSN=C91)/C=C\1/C(=O)N(C(=S)S1)CC)(CCCCCCCC)CCCCCCCC)CCCCCCCC
6	CCCCC(CC)CC1(C2=CC3=CC4=C(C=C3C=C2C5=C1C=C(S5)C6=CC=C(C7=NSN=C67)/C=C\8/C(=O)N(C(=S)S8)CC)C(C9=C4SC(=C9)C1=CC=C(C2=NSN=C12)/C=C\1/C(=O)N(C(=S)S1)CC)(CCCCCCCC)CCCCCCCC)CCCCCCCC
7	CCCCC(CC)CC1(C2=CC3=CC4=C(C=C3C=C2C5=C1C=C(S5)C6=CC=C(C7=NSN=C67)/C=C\8/C(=O)N(C(=S)S8)CC)C(C9=C4SC(=C9)C1=CC=C(C2=NSN=C12)/C=C\1/C(=O)N(C(=S)S1)CC)(CC(CC)CCCC)CC(CC)CCCC)CC(CC)CCCC
8	CCCCC(CC)CC1(C2=CC3=CC4=C(C=C3C=C2C5=C1C=C(S5)C6=CC=C(C7=NSN=C67)/C=C\8/C(=O)N(C(=S)S8)CC)C(C9=C4SC(=C9)C1=CC=C(C2=NSN=C12)/C=C\1/C(=O)N(C(=S)S1)CC)(CC(CC)CCCC)CC(CC)CCCC)CC(C)CC
9	CCCCC(CC)CC1(C2=CC3=C(C=C2C4=C1C=C(S4)C5=CC=C(C6=NSN=C56)/C=C\7/C(=O)N(C(=C(C#N)C#N)S7)CC)C(C8=C3SC(=C8)C9=CC=C(C1=NSN=C91)/C=C\1/C(=O)N(C(=C(C#N)C#N)S1)CC)(CC(CC)CCCC)CC(CC)CCCC)CC(CC)CCCC
10	CCCCC(CC1(C2=C(SC(=C2)C3=CC=C(C4=NSN=C34)/C=C/5\SC(=S)N(C5=O)CC)C6=C1C=C(S6)C7=CC=C(C8=NSN=C78)/C=C/9\SC(=S)N(C9=O)CC)CC(CCCC)CC)CC
11	CCCCCCCCC1=CC=C(C=C1)C2(C3=C(C4=C2C5=C(S4)C6=C(S5)C7=C(C6(C8=CC=C(C=C8)C)C9=CC=C(C=C9)C)C=C(S7)C1=CC=C(/C(=N/S)/C1=N)/C=C\1/C(=O)N(C(=S)S1)CC)SC(=C3)C1=CC=C(C2=NSN=C12)/C=C\1/C(=O)N(C(=S)S1)CC)C1=CC=C(C=C1)C
12	CCN1C(=S)S/C(=C\C2=CC=C(C3=NSN=C23)C4=CC5=C(S4)C6=C(C5(C7=CC=C(C=C7)C)C8=CC=C(C=C8)C)C9=C(S6)C2=C(S9)C3=C(C2(C2=CC=C(C=C2)C)C2=CC=C(C=C2)C)C=C(S3)C2=CC=C(C3=NSN=C23)/C=C/2\SC(=S)N(C2=O)CC)/C1=O
13	CCCCCCCCC1=CC=C(C=C1)C2(C3=C(SC4=C3SC5=C4C(C6=C5SC(=C6)C7=CC=C(C8=NSN=C78)/C=C/9\SC(=S)N(C9=O)CC)(C1=CC=C(C=C1)CCCCCCCC)C1=CC=C(C=C1)CCCCCCCC)C1=C2C=C(S1)C1=CC=C(C2=NSN=C12)/C=C/1\SC(=S)N(C1=O)CC)C1=CC=C(C=C1)CCCCCCCC

	CCC1=CC=C(C=C1)C2(C3=C(C4=C2C5=C(S4)C6=C(S5)C7=C(C6(C8=CC=C(C=C8)C)C9=CC=C(C=C9)C)C=C(S7)C1=CC=C(C2=NSN=C12)/C=C\1/C(=O)N(C(=S)S1)CC)SC(=C3)C1=CC=C(C2=NSN=C12)/C=C\1/C(=O)N(C(=S)S1)CC)C1=CC=C(C=C1)C
14	CCN1C(=O)/C(=C/C2=CC=C(C3=NSN=C23)C4=CC5=C(S4)C6=CC7=C(C=C6C5(C)C)C8=C(C7(C)C)C=C(S8)C9=CC=C(C2=NSN=C92)/C=C\2/C(=O)N(C(=S)S2)CC)/SC1=S
15	CCN1C(=O)/C(=C/C2=CC=C(C3=NSN=C23)C4=CC5=C(S4)C6=CC7=C(C=C6C5(CC(C)C)CC(C)C)C8=C(C7(CC(C)C)CC(C)C)C=C(S8)C9=CC=C(C2=NSN=C92)/C=C\2/C(=O)N(C(=S)S2)CC)/SC1=S
16	CCN1C(=S)S/C(=C\C2=CC=C(C3=NSN=C23)C4=CC5=C(S4)C6=C(C5(CC(C)C)CC(C)C)C=C(S6)C7=CC=C(C8=NSN=C78)/C=C/9\SC(=S)N(C9=O)CC)/C1=O
17	CCN1C(=O)/C(=C/C2=CC=C(C3=NSN=C23)C4=CC5=C(S4)C6=C(C5(CC(C)C)CC(C)C)C=C7C=C8C(=CC7=C6)C(C9=C8SC(=C9)C2=CC=C(C3=NSN=C23)/C=C(/C(=O)N(C)CC)\SC=S)(CC(C)C)CC(C)C)/SC1=S
18	CCN1C(=O)/C(=C/C2=CC=C(C3=NSN=C23)C4=CC5=C(S4)C6=CC7=C(C=C6[Ge]5(C(C)C)C(C)C)C8=C([Ge]7(C(C)C)C(C)C)C=C(S8)C9=CC=C(C2=NSN=C92)/C=C\2/C(=O)N(C(=S)S2)CC)/SC1=S
19	CCN1C(=O)/C(=C/C2=CC=C(C3=NSN=C23)C4=CC5=C(S4)C6=C([Ge]5(CC(C)C)CC(C)C)C=C7C=C8C(=CC7=C6)[Ge](C9=C8SC(=C9)C2=CC=C(C3=NSN=C23)/C=C(/C(=O)N(C)CC)\SC=S)(CC(C)C)CC(C)C)/SC1=S
20	CCN1C(=O)/C(=C/C2=CC=C(C3=NSN=C23)C4=CC5=C(S4)C6=CC7=C(C=C6C5(CC(C)C)CC(C)C)C8=C(C7(CC(C)C)CC(C)C)C=C(S8)C9=CC=C(C2=NSN=C92)/C=C\2/C(=O)N(C(=C(C#N)C#N)S2)CC)/SC1=C(C#N)C#N
21	CCN1C(=O)/C(=C/C2=CC=C(C3=NSN=C23)C4=CC5=C(S4)C6=CC7=C(C=C6C5(C)C)C8=C(C7(C)C)C=C(S8)C9=CC=C(C2=NSN=C92)/C=C\2/C(=O)N(C(=C(C#N)C#N)S2)CC)/SC1=C(C#N)C#N
22	CCCCC(CC)CC1(C2=C(C3=C1C=C(S3)C4=CC=C(C5=NSN=C45)/C=C\6/C(=O)N(C(=S)S6)CC)SC(=C2)C7=CC=C(C8=NSN=C78)/C=C\9/C(=O)N(C(S9)S)CC)CC(CC)CCC
23	CCCCCCCCCC1=CC=C(C=C1)C2(C3=C(C4=C2C5=C(S4)C6=C(S5)C7=C(C6(C8=CC=C(C=C8)CCCCCCCC)C9=CC=C(C=C9)CCCCCCCC)C=C(S7)C1=CC=C(C2=NSN=C12)/C=C\1/C(=O)N(C(=S)S1)CC)SC(=C3)C1=CC=C(C2=NSN=C12)/C=C\1/C(=O)N(C(=S)S1)CC)C1=CC=C(C=C1)C
24	CCCCCCCCC1(C2=CC3=C(C=C2C4=C1C=C(S4)C5=CC=C(C6=NSN=C56)/C=C\7/C(=S)N(C(=O)S7)CC)C(C8=C3SC(=C8)C9=CC=C(C1C9NSN1)/C=C\1/C(=O)N(C(=S)S1)CC)(CCCCCCCC)CCCCCCCC)CCCCCCCC
25	CCCCC(CC)CC1(C2=CC3=C(C=C2C4=C1C=C(S4)C5=CC=C(C6=NSN=C56)/C=C\7/C(=O)N(C(S7)S)CC)C(C8=C3SC(=C8)C9=CC=C(C1=NSN=C91)/C=C\1/C(=O)N(C(=S)S1)CC)(CC(CC)CCCC)CC(CC)CCCC)CC(CC)CCCC
26	CCCCCCC1=S=C(C=C1)C2=CC=C(S2)C3=CC=C(C4=NSN=C34)C5=CC6=C(S5)C7=CC8=C(C=C7C6(CC(CC)CCCC)CC(CC)CCCC)C9=C(C8(CC(CC)CCCC)CC(CC)CCCC)C=C(S9)C1=CC=C(S1)C1=C2C(=C(N(C2=O)CCCCCC)C2=CC=C(S2)C2=CC3=C(S2)C2=C(C3(CC(CC)CCCC)CC(CC)CCCC)C=C3C(=C2)C(C2=C3SC(=C2)C2=CC=C(C3=NSN=C23)C2=CC=C(S2)C2=CC=C(S2)CCCCCC)(CC(CC)CCCC)CC(CC)CCCC)C(=O)N1CCCCCC
27	

	CCCCC(CC)CC1=CC=C(C=C1)C2(C3=CC4=C(C=C3C5=C2C=C(S5)C6=CC=C(C7=NSN=C67)/C=C\8/C(=O)N(C(=S)S8)CC)C(C9=C4SC(=C9)C1=C2C(=C(C=C1)/C=C\1/C(=O)N(C(=S)S1)CC)N=S=N2)(C1=CC=C(C=C1)CC(CC)CCCC)C1=CC=C(C=C1)CC(CC)CCCC)C1=CC=C(C=C1)CC(CC)CCCC
28	
	CCCCCCCCCC1=CC=C(C=C1)C2(C3=CC4=C(C=C3C5=C2C=C(S5)C6=CC=C(C7=NSN=C67)/C=C\8/C(=O)N(C(=S)S8)CC)C(C9=C4SC(=C9)C1=C2C(=C(C=C1)/C=C\1/C(=O)N(C(=S)S1)CC)N=S=N2)(C1=CC=C(C=C1)CCCCCCCC)C1=CC=C(C=C1)CC
29	CCCCCC)C1=CC=C(C=C1)CCCCCCCC
	CCCCCCCCC1(C2=CC3=C(C=C2C4=C1C=C(S4)C5=CC=C(C6=NSN=C56)/C=C/7\C(=O)N(C(=S)S7)CC)C(C8=C3SC(=C8)C9=CC=C(C1=NSN=C91)/C=C/1\C(=O)N(C
30	(=S)S1)CC)(CCCCCCCC)CCCCCCCC)CCCCCCCC
	CCCCC(CC)CC1(C2=CC3=C(C=C2C4=C1C=C(S4)C5=CC=C(C6=NSN=C56)/C=C/7\C(=O)N(C(=S)S7)CC)C(C8=C3SC(=C8)C9=C1C(=C(C=C9)/C=C/2\C(=O)N(C(=S)S
31	2)CC)N=S=N1)(CC(CC)CCCC)CC(CC)CCCC)CC(CC)CCCC
	CCCCCCC1=CC=C(C=C1)C2(C3=CC4=C(C=C3C5=C2C=C(S5)C6=CC=C(C7=NSN=C67)/C=C\8/C(=O)N(C(=S)S8)CC)C(C9=C4SC(=C9)C1=C2C(=C(C=C1)/C=C\1/C(=O)N(C(=S)S1)CC)N=S=N2)(C1=CC=C(C=C1)CCCCCC)C1=CC=C(C=C1)CCCCC
32	C)C1=CC=C(C=C1)CCCCCC
	CCCCCCCCC1(C2=CC3=C(C=C2C4=C1C=C(S4)C5=CC=C(C6=NSN=C56)C=C7C(=O)N(C(=S)S7)CC)C(C8=C3SC(=C8)C9=CC=C(C1=NSN=C91)C=C1C(=O)N(C(=S)
33	S1)CC)(CCCCCCCC)CCCCCCCC)CCCCCCCC
	CCCCCCC1=CC=C(C=C1)C2(C3=CC4=C(C=C3C5=C2C=C(S5)C6=CC=C(C7=NSN=C67)/C=C/8\C(=O)N(C(=S)S8)CC)C(C9=C4SC(=C9)C1=CC=C(C2=NSN=C12)/C=C\1/C(=O)N(C(=S)S1)CC)(C1=CC=C(C=C1)CCCCCC)C1=CC=C(C=C1)CCCCCC)C
34	1=CC=C(C=C1)CCCCCC
	CCCCC(CC)CC1(C2=CC3=C(C=C2C4=C1C=C(S4)C5=CC=C(C6=NSN=C56)/C=C\7/C(=O)N(C(=O)S7)CC)C(C8=C3SC(=C8)C9=CC=C(C1=NSN=C91)/C=C\1/C(=O)N(
35	C(=S)S1)CC)(CC(CC)CCCC)CC(CC)CCCC)CC(CC)CCCC
	CCCCCCC1=CC=C(C=C1)C2(C3=CC4=C(C=C3C5=C2C6=C(S5)C=C(S6)/C=C\7/C(=C(C#N)C#N)C8=CC(=C(C=C8C7=O)F)F)C(C9=C4SC1=C9SC(=C1)/C=C\1/C(=C(C#N)C#N)C2=CC(=C(C=C2C1=O)F)F)(C1=CC=C(C=C1)CCCCCC)C1=CC=C(C=C1)
36	CCCCCC)C1=CC=C(C=C1)CCCCCC
	CCCCCCC1=CC=C(C=C1)C2(C3=CC4=C(C=C3C5=C2C6=C(S5)C=C(S6)C=C7C(=C(C#N)C#N)C8=CC(=C(C=C8C7=O)F)F)C(C9=C4SC1=C9SC(=C1)C=C1C(=C(C#N)C#N)C2=CC(=C(C=C2C1=O)F)F)(C1=CC=C(C=C1)CCCCCC)C1=CC=C(C=C1)CC
37	CCCC)C1=CC=C(C=C1)CCCCCC
	CCCCCCC1=CC=C(C=C1)C2(C3=CC4=C(C=C3C5=C2C6=C(S5)C=C(S6)/C=C/7\C(=C(C#N)C#N)C8=CC(=C(C=C8C7=O)F)F)C(C9=C4SC1=C9SC(=C1)/C=C\1/C(=C(C#N)C#N)C2=CC(=C(C=C2C1=O)F)F)(C1=CC=C(C=C1)CCCCCC)C1=CC=C(C=C1)
38	CCCCCC)C1=CC=C(C=C1)CCCCCC
	CCCCCCC1=CC=C(C=C1)C2(C3=CC4=C(C=C3C5=C2C6=C(S5)C=C(S6)/C=C/7\C(=C(C#N)C#N)C8=CC(=C(C=C8C7=O)F)F)C(C9=C4SC1=C9SC(=C1)/C=C/1\C(=C(C#N)C#N)C2=CC(=C(C=C2C1=O)F)F)(C1=CC=C(C=C1)CCCCCC)C1=CC=C(C=C1)
39	CCCCCC)C1=CC=C(C=C1)CCCCCC
	CCCCCCC1=CC=C(C=C1)C2(C3=CC4=C(C=C3C5=C2C6=C(S5)C=C(S6)/C=C\7/C(=C(C#N)C#N)C8=CC=CC=C8C7=O)C(C9=C4SC1=C9SC(=C1)/C=C\1/C(=C(C#N)C
40	

	#N)C2=CC=CC=C2C1=O)(C1=CC=C(C=C1)CCCCCC)C1=CC=C(C=C1)CCCCCC)C1=CC=C(C=C1)CCCCCC
41	CCCCCCC1(C2=CC3=C(C=C2C4=C1C=C(S4)/C=C\5/C(=C(C#N)C#N)C6=CC=CC=C6C5=O)C(C7=C3SC(=C7)/C=C\8/C(=C(C#N)C#N)C9=CC=CC=C9C8=O)(CCCCCC)CCCCCC)CCCCCC
42	CCCCCCC1=CC=C(C=C1)CS/C(=C\2/C3=C(C4=C2C=C5C(=C4)C(=C(SC6=CC=C(C=C6)CCCCCC)SC7=CC=C(C=C7)CCCCCC)C8=C5SC(=C8)/C=C\9/C(=C(C#N)C#N)C1=CC(=C(C=C1C9=O)F)F)SC(=C3)/C=C\1/C(=C(C#N)C#N)C2=CC(=C(C=C2C1=O)F)F)/SCC1=CCC(C=C1)(C)CCCCCC
43	CCCCCCC1=CC=C(C=C1)CSC(=C2C3=CC4=C(C=C3C5=C2C=C(S5)/C=C\6/C(=C(C#N)C#N)C7=CC(=C(C=C7C6=O)F)F)C(=C(SCC8=CC=C(C=C8)CCCCCC)SCC9=CC=C(C=C9)CCCCCC)C1=C4SC(=C1)/C=C\1/C(=C(C#N)C#N)C2=CC(=C(C=C2C1=O)F)F)SCC1=CC=C(C=C1)CCCCCC
44	CC1=CC=C(C=C1)C2(C3=CC4=C(C=C3C5=C2C6=C(S5)C=C(S6)/C=C\7/C(=C(C#N)C#N)C8=CC(=C(C=C8C7=O)F)F)C(C9=C4SC1=C9SC(=C1)/C=C\1/C(=C(C#N)C#N)C2=CC(=C(C=C2C1=O)F)F)(C1=CC=C(C=C1)C)C1=CC=C(C=C1)C)C1=CC=C(C=C1)C
45	CC1=CC=C(C=C1)C2(C3=C(C4=C2C(=C5C(=C4F)C(C6=C5SC(=C6)/C=C\7/C(=C(C#N)C#N)C8=CC=CC=C8C7=S)(C9=CC=C(C=C9)C)C1=CC=C(C=C1)C)F)SC(=C3)/C=C\1/C(=C(C#N)C#N)C2=CC=CC=C2C1=O)C1=CC=C(C=C1)C
46	CC1=CC=C(C=C1)C2(C3=C(C(=C4C(=C3F)C5=C(C4(C6=CC=C(C=C6)C)C7=CC=C(C=C7)C)C8=C(S5)C=C(S8)/C=C\9/C(=C(C#N)C#N)C1=CC=CC=C1C9=O)F)C1=C2C2=C(S1)C=C(S2)/C=C\1/C(=C(C#N)C#N)C2=CC=CC=C2C1=O)C1=CC=C(C=C1)C
47	CC1=CC=C(C=C1)C2(C3=C(C4=C2C(=C5C(=C4F)C(C6=C5SC(=C6)C7=CC=C(S7)/C=C\8/C(=C(C#N)C#N)C9=CC=CC=C9C8=O)(C1=CC=C(C=C1)C)C1=CC=C(C=C1)C)F)SC(=C3)C1=CC=C(S1)/C=C\1/C(=C(C#N)C#N)C2=CC=CC=C2C1=O)C1=CC=C(C=C1)C
48	CC1=C(C2=C(S1)C3=C(C2(C)C)C(=C4C(=C3F)C(C5=C4SC(=C5F)CCCC6(C(=C(C#N)C#N)C7=CC=CC=C7C6=O)C)(C)C)F)F
49	CC1=C(C2=C(S1)C3=CC4=C(C=C3C2(C)C)C5=C(C4(C)C)C(=C(S5)CCCC6(C(=C(C#N)C#N)C7=CC=CC=C7C6=O)C)F)F
50	CC1=CC2=C(S1)C3=C(C2(C)C)C(=C4C(=C3F)C(C5=C4SC(=C5)CCCC6(C(=C(C#N)C#N)C7=CC=CC=C7C6=O)C)(C)C)F
51	CC1=CC=C(C=C1)C2(C3=C(C(=C4C(=C3F)C5=C(C4(C6=CC=C(C=C6)C)C7=CC=C(C=C7)C)C8=C(S5)C=C(S8)C9=CC=C(S9)/C=C\1/C(=C(C#N)C#N)C3=CC=CC=C3C1=O)F)C1=C2C2=C(S1)C=C(S2)C1=CC=C(S1)/C=C\1/C(=C(C#N)C#N)C2=CC=CC=C2C1=O)C1=CC=C(C=C1)C
52	CC1=CC2=C(S1)C3=C(S2)C4=C(C3(C)C)C(=C5C(=C4F)C(C6=C5SC7=C6SC(=C7)C/C=C\8/C(=C(C#N)C#N)C9=CC=CC=C9C8=O)(C)C)F
53	CCCCCCCCC1=CC=C(C=C1)C2(C3=CC4=C(C=C3C5=C2C=C(S5)/C=C\6/C(=C(C#N)C#N)C7=CC(=C(C=C7C6=O)F)F)C(C8=C4SC(=C8)/C=C\9/C(=C(C#N)C#N)C1=CC(=C(C=C1C9=O)F)F)(C1=CC=C(C=C1)CCCCCC)C1=CC=C(C=C1)CCCCCCCCC)C1=CC=C(C=C1)CCCCCC
54	CCCCCCC1=CC=C(C=C1)C2(C3=CC4=C(C=C3C5=C2C6=C(S5)C=C(S6)/C=C\7/C(=C(C#N)C#N)C8=C(C7=O)C(=C(S8)F)F)C(C9=C4SC1=C9SC(=C1)/C=C\1/C(=C(C#

	N)C#N)C2=C(C1=O)C(=C(S2)F)F)(C1=CC=CC=C1)C1=CC=C(C=C1)CCCCCC)C1=CC=C(C=C1)CCCCCC
55	CC1=CC=C(C=C1)C2(C3=C(C4=C2C(=C5C(=C4F)C(C6=C5SC(=C6)/C=C\7/C(=C(C#N)C#N)C8=CC=CC=C8C7=S)(C9=CC=C(C=C9)C)C1=CC=C(C=C1)C)F)SC(=C3)/C=C\1/C(=C(\C#N)/[N+]#[C-])/C2=CC=CC=C2C1=O)C1=CC=C(C=C1)C
56	CC1(C2=CC3=C(C=C2C4=C1C5=C(S4)C(=C(S5)C=C6C(=C(C#N)C#N)C7=CC=CC=C7C6=O)F)C(C8=C3SC9=C8SC(=C9F)C=C1C(=C(C#N)C#N)C2=CC=CC=C2C1=O)(C)C)C
57	CC1(C2=CC3=C(C=C2C4=C1C(=C(S4)/C=C\5/C(=C(C#N)C#N)C6=CC=CC=C6C5=O)F)C(C7=C3SC(=C7F)/C=C\8/C(=C(C#N)C#N)C9=CC=CC=C9C8=O)(C)C)C
58	CC1(C2=C(C(=C3C(=C2F)C4=C(C3(C)C)C5=C(S4)C(=C(S5)/C=C\6/C(=C(C#N)C#N)C7=CC=CC=C7C6=O)F)F)C8=C1C9=C(S8)C(=C(S9)/C=C\1/C(=C(C#N)C#N)C2=CC=CC=C2C1=O)F)C
59	CC1(C2=CC3=C(C=C2C4=C1C5=C(S4)C(=C(S5)/C=C\6/C(=C(C#N)C#N)C7=CC=CC=C7C6=O)F)C(C8=C3SC9=C8SC(=C9F)/C=C\1/C(=C(C#N)C#N)C2=CC=CC=C2C1=O)(C)C)C
60	CC1(C2=C(C(=C3C(=C2F)C4=C(C3(C)C)C5=C(S4)C=C(S5)/C=C\6/C(=C(C#N)C#N)C7=CC=CC=C7C6=O)F)C8=C1C9=C(S8)C=C(S9)/C=C\1/C(=C(C#N)C#N)C2=CC=CC=C2C1=O)C
61	CC1(C2=C(C(=C3C(=C2F)C4=C(C3(C)C)C)C(=C(S4)/C=C\5/C(=C(C#N)C#N)C6=CC=CC=C6C5=O)F)F)C7=C1C(=C(S7)/C=C\8/C(=C(C#N)C#N)C9=CC=CC=C9C8=O)F)C
62	CC1(C2=C(C3=C1C(=C4C(=C3F)C(C5=C4SC(=C5)/C=C\6/C(=C(C#N)C#N)C7=CC=CC=C7C6=O)(C)C)F)SC(=C2)/C=C\8/C(=C(C#N)C#N)C9=CC=CC=C9C8=O)C
63	CSC(=C1C2=CC3=C(C=C2C4=C1C=C(S4)/C=C\5/C(=C(C#N)C#N)C6=CC(=C(C=C6C5=O)F)F)C(=C(SC)SC)C7=C3SC(=C7)/C=C\8/C(=C(C#N)C#N)C9=CC(=C(C=C9C8=O)F)F)SC
64	CCCCCCC1=CC(=CC(=C1)C2(C3=CC4=C(C=C3C5=C2SC(=C5)/C=C\6/C(=C(\C#N)/[N+]#[C-])/C7=CC(=C(C=C7C6=O)F)F)C(C8=C4C=C(S8)/C=C\9/C(=C(/C#N)\[N+]#[C-])/C1=CC(=C(C=C1C9=O)F)F)(C1=CC(=CC(=C1)CCCCCC)CCCCCC)CCCCCC)C1=CC(=CC(=C1)CCCCCC)CCCCCC)C1=CC(=CC(=C1)CCCCCC)CCCCCC)CCCCCC
65	CCCCCCC1=CC=C(C=C1)C2(C3=CC4=C(C=C3C5=C2C6=C(S5)C=C(S6)/C=C\7/C(=C(C#N)C#N)C8=CC(=C(C=C8C7=O)F)F)C(C9=C4SC1=C9SC(=C1)/C=C\1/C(=C(C#N)C#N)C2=CC(=C(C=C2C1=O)F)F)(C1=CC=C(C=C1)C)C1=CC=C(C=C1)CCCCCC)C1=CC=C(C=C1)CCCCCC
66	CC1(C2=CC3=C(C=C2C4=C1C(=C(S4)C=C5C(=C(C#N)C#N)C6=CC=CC=C6C5=O)F)C(C7=C3SC(=C7F)C=C8C(=C(C#N)C#N)C9=CC=CC=C9C8=O)(C)C)C
67	CC1(C2=C(C(=C3C(=C2F)C4=C(C3(C)C)C5=C(S4)C(=C(S5)C=C6C(=C(C#N)C#N)C7=CC=CC=C7C6=O)F)F)C8=C1C9=C(S8)C(=C(S9)C=C1C(=C(C#N)C#N)C2=CC=CC=C2C1=O)F)C
68	CC1(C2=C(C(=C3C(=C2F)C4=C(C3(C)C)C5=C(S4)C=C(S5)C=C6C(=C(C#N)C#N)C7=CC=CC=C7C6=O)F)C8=C1C9=C(S8)C=C(S9)C=C1C(=C(C#N)C#N)C2=CC=CC=C2C1=O)C
69	CC1(C2=C(C(=C3C(=C2F)C4=C(C3(C)C)C)C(=C(S4)C=C5C(=C(C#N)C#N)C6=CC=CC=C6C5=O)F)F)C7=C1C(=C(S7)C=C8C(=C(C#N)C#N)C9=CC=CC=C9C8=O)F)C

70	CC1(C2=C(C3=C1C(=C4C(=C3F)C(C5=C4SC(=C5)C=C6C(=C(C#N)C#N)C7=CC=CC=C7C6=O)(C)C)F)SC(=C2)C=C8C(=C(C#N)C#N)C9=CC=CC=C9C8=O)CC1CCC2(C3=C(SC(=C3)/C=C/4\C(=O)C5=CC(=C(C=C5C4=C(C#N)C#N)F)F)C6=C2C=C(S6)/C=C/7\C(=O)C8=CC(=C(C=C8C7=C(C#N)C#N)F)F)CC1
71	CCCCCCC1=CC(=CC(=C1)C2(C3=C(C(=C4C(=C3F)C5=C(C4(C6=CC(=CC(=C6)CCCCC)CCCCCC)CCCCCC)C7=CC(=CC(=C7)CCCCCC)CCCCCC)C8=C(S5)C=C(S8)/C=C\9/C(=C(C#N)C#N)C1=CC(=C(C=C1C9=O)F)F)F)C1=C2C2=C(S1)C=C(S2)/C=C\1/C(=C(C#N)C#N)C2=CC(=C(C=C2C1=O)F)F)C1=CC(=CC(=C1)CCCCCC)CCCCCC)CC
72	CCCC CC1=CC2=C(C=C1F)C(=O)/C(=C\C3=CC4=C(S3)C5=C(C46CCC(CC6)(C)C)C=C(S5)/C=C\7/C(=C(\C#N)/[N+]#[C-])/C8=CC(=C(C=C8C7=O)F)F)/C2=C(C#N)[N+]#[C-]
73	CC.CC(F)(F)F.C#CC(=C\1C2=CC=CC=C2C(=O)/C1=C\C3=CC4=C(S3)C5=C(C46CCCC6)C=C(S5)/C=C\7/C(=C(\C#N)/[N+]#[C-])C8=CC(=C(C=C8C7=O)F)F)/C2=C(C#N)[N+]#[C-]
74	CC(F)(F)F.CC(F)(F)F.C1CCC2(C3=C(SC(=C3)/C=C/4\C(=O)C5=CC=CC=C5C4=C(C#N)C#N)C6=C2C=C(S6)/C=C/7\C(=O)C8=CC=CC=C8C7=C(C#N)C#N)CC1
75	C[C@H]1CC2(C[C@@H]1C)C3=C(C4=C2C=C(S4)/C=C\5/C(=C(C#N)[N+]#[C-])C6=C(C5=O)C=C(C(=C6)C)F)SC(=C3)/C=C\7/C(=C(\C#N)/[N+]#[C-]
76	CC1=CC=C(C=C1)C2(C3=CC4=C(C=C3C5=C2C=C(S5)/C=C\6/C(=C(C#N)C#N)C7=CC(=C(C=C7C6=O)F)F)C(C8=C4SC(=C8)/C=C\9/C(=C(\C#N)/[N+]#[C-])/C1=CC(=C(C=C1C9=O)F)F)(C1=CC=C(C=C1)C)C1=CC=C(C=C1)C)C1=CC=C(C=C1)C
77	CC1(CCC2(C3=C(SC(=C3)/C=C/4\C(=O)C5=CC(=C(C=C5C4=C(C#N)C#N)F)F)C6=C2C=C(S6)/C=C/7\C(=O)C8=CC(=C(C=C8C7=C(C#N)C#N)F)F)CC1)C
78	CC1=CC=C(C=C1)C2(C3=CC4=C(C=C3C5=C2C=C(S5)/C=C\6/C(=C(C#N)C#N)C7=CC(=C(C=C7C6=O)F)F)C(C8=C4SC(=C8)/C=C\9/C(=C(C#N)C#N)C1=CC(=C(C=C1C9=O)F)F)(C1=CC=C(C=C1)C)C1=CC=C(C=C1)C)C1=CC=C(C=C1)C
79	CCCCCCC1=CC(=CC(=C1)C2(C3=CC4=C(C=C3C5=C2SC(=C5)/C=C\6/C(=C(C#N)C#N)C7=CC(=C(C=C7C6=O)F)F)C(C8=C4C=C(S8)/C=C\9/C(=C(C#N)C#N)C1=CC(=C(C=C1C9=O)F)F)(C1=CC(=CC(=C1)CCCCCC)CCCCCC)C1=CC(=CC(=C1)CCCCCC)CCCCCC)C1=CC(=CC(=C1)CCCCCC)CCCCCC)CCCCCC
80	CCCCC(CC)CC1(C2=C(C3=C1C=C(S3)C4=CC(=C(C=C4F)C5=CC6=C(S5)C7=C(C6(CC(CC)CCCC)CC(CC)CCCC)C=C(S7)/C=C\8/C(=C(C#N)C#N)C9=CC=CC=C9C8=O)F)SC(=C2)/C=C\1/C(=C(C#N)C#N)C2=CC=CC=C2C1=O)CC(CC)CCCC
81	CCCCCCC1=CC=C(C=C1)C2(C3=CC4=C(C=C3C5=C2C=C(S5)/C=C/6\C=C/C(=C\7/C=C/C(=C\8/C(=C(\C#N)/[N+]#[C-])/C9=CC(=C(C=C9C8=O)F)F)/S7)/S6)C(C1=C4SC(=C1)/C=C/1\C=C/C(=C\2/C=C/C(=C\3/C(=C(/C#N)\[N+]#[C-])/C4=CC(=C(C=C4C3=O)F)F)/S2)/S1)(C1=CC=C(C=C1)CCCCCC)C1=CC=C(C=C1)CCCCCC)C1=CC=C(C=C1)CCCCCC
82	CCCCCCC1=CC=C(C=C1)C2(C3=CC4=C(C=C3C5=C2C=C(S5)/C=C/6\C=C/C(=C\7/C(=C(\C#N)/[N+]#[C-])/C8=CC(=C(C=C8C7=O)F)F)/S6)C(C9=C4SC(=C9)/C=C/1\C=C/C(=C\2/C(=C(/C#N)\[N+]#[C-

	])/C3=CC(=C(C=C3C2=O)F)F)/S1)(C1=CC=C(C=C1)CCCCCC)C1=CC=C(C=C1)CCCCCC)C1=CC=C(C=C1)CCCCCC
	CCCCCCC1=CC=C(C=C1)C2(C3=CC4=C(C=C3C5=C2C=C(S5)/C=C/6\C=C7C(=C/C(=C\8/C(=C(\C#N)/[N+]#[C-])/C9=CC(=C(C=C9C8=O)F)F)/S7)S6)C(C1=C4SC(=C1)/C=C/1\C=C2C(=C/C(=C\3/C(=C(/C#N)\[N+]#[C-])/C4=CC(=C(C=C4C3=O)F)F)/S2)S1)(C1=CC=C(C=C1)CCCCCC)C1=CC=C(C=C1)
85	CCCCCC)C1=CC=C(C=C1)CCCCCC
	CCCCCCC1=CC=C(C=C1)C2(C3=CC4=C(C=C3C5=C2C6=C(S5)C=C(S6)/C=C\7/C(=C(C#N)C#N)C8=CC=CC=C8C7=O)C(C9=C4SC1=C9SC(=C1)C=C1C(=C(C#N)C#N)C2=CC=CC=C2C1=O)(C1=CC=C(C=C1)CCCCCC)C1=CC=C(C=C1)CCCCCC)C1
86	=CC=C(C=C1)CCCCCC
	CCCCCCC1=CC=C(S1)C2(C3=CC4=C(C=C3C5=C2C6=C(S5)C=C(S6)/C=C\7/C(=C(C#N)C#N)C8=CC=CC=C8C7=O)C(C9=C4SC1=C9SC(=C1)/C=C\1/C(=C(C#N)C#N)C2=CC=CC=C2C1=O)(C1=CC=C(S1)CCCCCC)C1=CC=C(S1)CCCCCC)C1=CC=C
87	(S1)CCCCCC
	CCCCCCC1=CC=C(C=C1)C2(C3=CC4=C(C=C3C5=C2C6=C(S5)C=C(S6)C=C7C(=C(C#N)C#N)C8=CC=CC=C8C7=O)C(C9=C4SC1=C9SC(=C1)C=C1C(=C(C#N)C#N)C2=CC=CC=C2C1=O)(C1=CC=C(C=C1)CCCCCC)C1=CC=C(C=C1)CCCCCC)C1=
88	CC=C(C=C1)CCCCCC
	CCCCCCC1=CC=C(C=C1)C2(C3=CC4=C(C=C3C5=C2C6=C(S5)C=C(S6)/C=C/7\C(=C(C#N)C#N)C8=CC=CC=C8C7=O)C(C9=C4SC1=C9SC(=C1)/C=C\1/C(=C(C#N)C#N)C2=CC=CC=C2C1=O)(C1=CC=C(C=C1)CCCCCC)C1=CC=C(C=C1)CCCCCC)C
89	1=CC=C(C=C1)CCCCCC
	CCCCCCCCCCCCCCCCCC1(C2=CC3=C(C=C2C4=C1C=C(S4)/C=C\5/C(=C(C#N)C#N)C6=CC=CC=C6C5=O)C(C7=C3SC(=C7)/C=C\8/C(=C(C#N)C#N)C9=CC=CC=C9C8=O)(CCCCCCCCCCCCCCCCCC)CCCCCCCCCCCCCCCCCC)CCCCCCCCCCCCCCCCC
90	C
	CCCCCCC1=CC=C(C=C1)C2(C3=CC4=C(C=C3C5=C2C6=C(S5)C=C(S6)C=C7C(=C(C#N)C#N)C8=C(C7=O)C=CC(=C8)C)C(C9=C4SC1=C9SC(=C1)C=C1C(=C(C#N)C#N)C2=C(C1=O)C=CC(=C2)C)(C1=CC=C(C=C1)CCCCCC)C1=CC=C(C=C1)CCC
91	CCC)C1=CC=C(C=C1)CCCCCC
	CCCCCCC1=CC=C(S1)C2(C3=CC4=C(C=C3C5=C2C6=C(S5)C=C(S6)C=C7C(=C(C#N)C#N)C8=CC=CC=C8C7=O)C(C9=C4SC1=C9SC(=C1)C=C1C(=C(C#N)C#N)C2=CC=CC=C2C1=O)(C1=CC=C(S1)CCCCCC)C1=CC=C(S1)CCCCCC)C1=CC=C(S1)
92	CCCCCC
	CCCCCCC1=CC=C(C=C1)C2(C3=CC4=C(C=C3C5=C2C6=C(S5)C=C(S6)/C=C/7\C(=C(C#N)C#N)C8=CC=CC=C8C7=O)C(C9=C4SC1=C9SC(=C1)/C=C/1\C(=C(C#N)C#N)C2=CC=CC=C2C1=O)(C1=CC=C(C=C1)CCCCCC)C1=CC=C(C=C1)CCCCCC)C
93	1=CC=C(C=C1)CCCCCC
	CCCCCCC1=CC=C(C=C1)C2(C3=CC4=C(C=C3C5=C2C6=C(S5)C=C(S6)/C=C\7/C(=C(C#N)C#N)C8=C(C7=O)C=CC(=C8)C)C(C9=C4SC1=C9SC(=C1)/C=C\1/C(=C(C#N)C#N)C2=C(C1=O)C=CC(=C2)C)(C1=CC=C(C=C1)CCCCCC)C1=CC=C(C=C1)C
94	CCCCC)C1=CC=C(C=C1)CCCCCC
	CCCCCCC1=CC=C(C=C1)C2(C3=CC4=C(C=C3C5=C2C6=C(S5)C7=C(S6)C=C(S7)/
95	C=C\8/C(=C(C#N)C#N)C9=C(C8=O)C(=CC=C9)F)C(C1=C4SC2=C1SC1=C2SC(=C1

	)/C=C\1/C(=C(C#N)C#N)C2=C(C1=O)C(=CC=C2)F)(C1=CC=C(C=C1)CCCCCC)C1 =CC=C(C=C1)CCCCCC)C1=CC=C(C=C1)CCCCCC
96	CCCCCCC1=CC=C(C=C1)C2(C3=CC4=C(C=C3C5=C2C6=C(S5)C7=C(S6)C=C(S7)/ C=C\8/C(=C(C#N)C#N)C9=C(C8=O)C=C(C=C9)F)C(C1=C4SC2=C1SC1=C2SC(=C1)/C=C\1/C(=C(C#N)C#N)C2=C(C1=O)C=C(C=C2)F)(C1=CC=C(C=C1)CCCCCC)C1 =CC=C(C=C1)CCCCCC)C1=CC=C(C=C1)CCCCCC
97	CCCCCCC1=CC=C(C=C1)C2(C3=CC4=C(C=C3C5=C2C6=C(S5)C7=C(S6)C=C(S7)/ C=C\8/C(=C(C#N)C#N)C9=CC(=C(C=C9C8=O)F)F)C(C1=C4SC2=C1SC1=C2SC(= C1)/C=C\1/C(=C(C#N)C#N)C2=CC(=C(C=C2C1=O)F)F)(C1=CC=C(C=C1)CCCCCC)C1=CC=C(C=C1)CCCCCC)C1=CC=C(C=C1)CCCCCC
98	CCCCCCC1(C2=CC3=C(C=C2C4=C1C=C(S4)/C=C\5/C(=C(C#N)C#N)C6=CC(=C(C =C6C5=O)F)F)C(C7=C3SC(=C7)/C=C\8/C(=C(C#N)C#N)C9=CC(=C(C=C9C8=O)F) F)(CCCCCC)CCCCCC)CCCCCC
99	CCCCCCC1=CC=C(C=C1)C2(C3=CC4=C(C=C3C5=C2C6=C(S5)C=C(S6)/C=C\7/C(=C(C#N)C#N)C8=C(C7=O)C=C(C=C8)F)C(C9=C4SC1=C9SC(=C1)/C=C\1/C(=C(C# N)C#N)C2=C(C1=O)C=C(C=C2)F)(C1=CC=C(C=C1)CCCCCC)C1=CC=C(C=C1)CC CCCC)C1=CC=C(C=C1)CCCCCC

Table S2. Training/validation MSE values

Model	Train AUC	Val AUC	Test AUC	Test MSE
SVM-RBF	1.000	0.997	0.999	0.01
Random Forest	1.000	0.995	0.998	0.00
SVM-Linear	0.994	0.981	0.986	0.64

Table S3. Model-Specific Hyperparameters (optimized via 5-fold cross-validation)

Model	**Hyperparameters**
SVM-RBF	C=10.0, gamma=0.01, epsilon=0.1
Random Forest	n_estimators=500, max_depth=15, min_samples_split=5, max_features='sqrt'
SVM-Linear	C=1.0, epsilon=0.1

Table S4. Listing all molecular descriptors used in the study

Descriptor	Definition	Significance
Chi0v	Valence molecular connectivity index (order 0)	Molecular size/branching
Chi1v	Valence molecular connectivity index (order 1)	Molecular connectivity
Chi2v	Valence molecular connectivity index (order 2)	Shape complexity
Chi3v	Valence molecular connectivity index (order 3)	3D structural complexity
Chi4v	Valence molecular connectivity index (order 4)	Saturation effects
kappa1	Kappa shape index (order 1)	Molecular flexibility
kappa2	Kappa shape index (order 2)	Molecular compactness
kappa3	Kappa shape index (order 3)	Complexity relative to linear chains
HallKierAlpha	Hall-Kier alpha value	Molecular cyclicity
Ipc	Information content index	Structural symmetry
EState_VSA1	E-state approximation to VDWSA (bin 1)	Hydrophobicity
EState_VSA2	E-state approximation to VDWSA (bin 2)	Polarity
SMR_VSA1	MR approximation to VDWSA (bin 1)	Polarizability
PEOE_VSA1	Partial charge approximation to VDWSA (bin 1)	Electrostatic interactions
NumRadicalElectrons	Number of radical electrons	Reactivity
MaxPartialCharge	Maximum atom partial charge	Charge distribution
MinPartialCharge	Minimum atom partial charge	Charge distribution
LabuteASA	Approximate surface area (Labute method)	Packing efficiency
TPSA	Topological polar surface area	Solubility/membrane permeability
RadiusOfGyration	Radius of gyration	Molecular compactness
InertialShapeFactor	Inertial shape factor	Molecular elongation
PrincipalMoments	Principal moments of inertia (x, y, z)	3D shape anisotropy
NPR1	Normalized principal moment ratio 1	Shape deviation from sphere
NPR2	Normalized principal moment ratio 2	Shape deviation from sphere
HeavyAtomCount	Number of non-hydrogen atoms	Molecular size
NHOHCount	Number of NH or OH groups	Hydrogen bonding

Descriptor	Definition	Significance
NumHeteroatoms	Number of heteroatoms (O, N, S, P)	Reactivity/solubility
NumRotatableBonds	Number of non-terminal single bonds (excl. amides)	Molecular flexibility
NumAromaticRings	Number of aromatic rings	Rigidity/π-stacking
NumAliphaticRings	Number of aliphatic rings	Flexibility
FractionCSP3	Fraction of SP^3-hybridized carbon atoms	Saturation
SlogP_VSA1	SlogP approximation to VDWSA (bin 1)	Lipophilicity
MolLogP	Wildman-Crippen logP	Partition coefficient
MolMR	Wildman-Crippen MR	Polarizability
BalabanJ	Balaban's J index	Molecular branching
BertzCT	Bertz complexity index	Structural complexity
VSA_EState1	E-state/VSA hybrid (bin 1)	Electronic accessibility