SUPPORTING INFORMATION FOR:

Accelerating the Discovery of High-Mobility Molecular Semiconductors: A Machine Learning Approach

Tahereh Nematiaram,*[a] Zenon Lamprou,[b] and Yashar Moshfeghi[b]

[a] Department of Pure and Applied Chemistry, University of Strathclyde, 295 Cathedral Street, Glasgow G1 1XL, U.K.
[b] Department of Computer and Information Sciences, University of Strathclyde, 26 Richmond Street, Glasgow G1 1XH, U.K.

*tahereh.nematiaram@strath.ac.uk

The distributions of charge transfer integrals of $|J_1|$, $|J_2|$ and associated 2D values are shown using three violin plots in Figure S1. Each plot depicts the probability density of the respective data points, offering insights into their distribution, spread, and concentration. Left panel, representing $|J_1|$, shows a distribution with a median of 0.036 eV, indicating a high density of lower values. Medium Panel, for $|J_2|$, exhibits a similar pattern with slight variations in density and spread and a median of 0.014 eV, suggesting a consistent concentration of lower values. Right panel, representing 2D, displays a distinct shape with a wider distribution, reflecting a broader range of values and a higher density spread across different values, with a median of 0.605. This dataset is notably imbalanced and can be categorised into three performance groups: (i) high-performance, which includes structures with $|J_1|$ values of at least 0.1 eV and *2D* values equal to or greater than 0.05. Figure S2 displays diagrams of randomly selected molecules in this category, labelled with their CSD identifiers; (ii) low-performance, which includes structures where $|J_1|$ is less than 0.1 eV and 2D is less than 0.05; and (iii) moderate-performance, containing all other structures that do not fall into the high or low-performance categories.

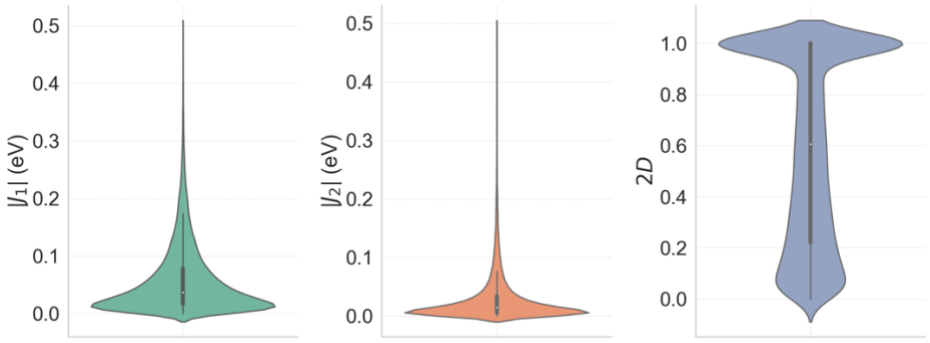

Figure S1. From left to right: Violin plots show the distribution of the largest $|J_1|$, the second largest transfer integral $|J_2|$, and the *2D* parameter across the dataset.

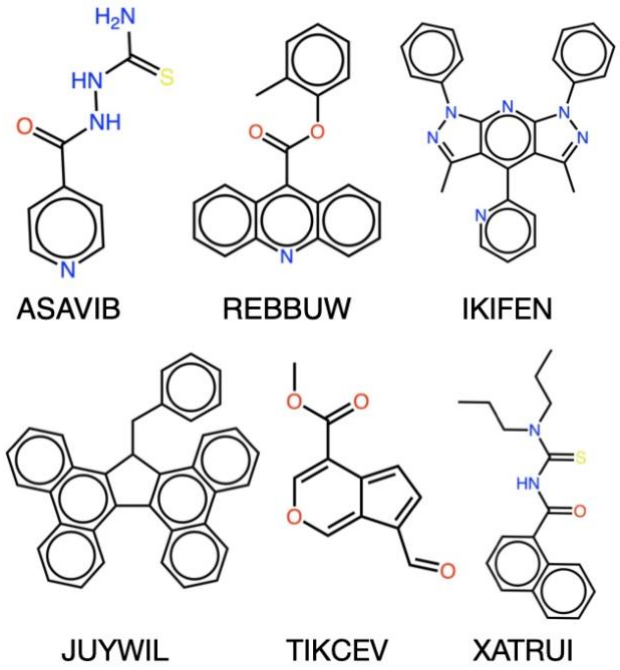

Figure S2. A random selection of CSD molecules with highly isotropic bands.