Supporting Information

Prediction of the Imprinting Quality of Molecularly Imprinted Polymers via a Data-driven Similarity-based Clustering Approach

Bahar Dadfar[a,†], *Meriem Kassar*[a,†], *Berna Alemdag*[a,†], *Gözde Kabay*[a,*]

[a]Karlsruhe Institute of Technology (KIT), Institute of Functional Interfaces (IFG), Department for Bioengineering and Biosystems, 76344 Eggenstein-Leopoldshafen, Germany

E-mail: gozde.kabay@kit.edu

Table S1. Overview of commonly reported definitions of the IF in MIPs. The table lists the definition, experimental determination methods, and success criteria. In this study, IF values based on binding capacity and dissociation constant were used.

Equation	Experimental Methods	Success Criteria
Biding capacity-based: $IF_Q = \frac{Q_{MIP}}{Q_{NIP}}$ $Q = \frac{(C_i - C_f)V}{m}$	Batch rebinding: incubate MIP/NIP with known analyte initial concentration C_i, measure analyte free concentration C_f at equilibrium (using UV–Vis, HPLC, etc.).	$IF_Q > 1$: the higher the ratio, the stronger the contribution of specific recognition sites relative to nonspecific binding. Values significantly greater than 2–3 are often interpreted as strong evidence of imprinting efficiency.
Dissociation constant-based: $IF_{K_d} = \frac{K_{d,NIP}}{K_{d,MIP}}$ *Langmuir-type isotherm fitting:* $Q = \frac{Q_{max} \cdot C_f}{K_d + C_f}$	Adsorption isotherm experiments are fitted to models (mostly Langmuir,) to extract K_d, describing the affinity of the binding sites. Lower K_d means stronger binding.	$IF_{K_d} > 1$: A lower K_d for the MIP compared to the NIP indicates stronger binding (higher affinity). Ratios > 2 are typically considered indicative of efficient imprinting.

*C_i initial analyte concentration (mg/L), C_f free analyte concentration at equilibrium (mg/L), V volume of adsorption solution (L), m polymer mass (g), Q equilibrium adsorption capacity (mg/g), Q_{max} maximum binding capacity (mg/g), K_d dissociation constant (M), k' capacity factor, t_R retention time, t_0 column void time.

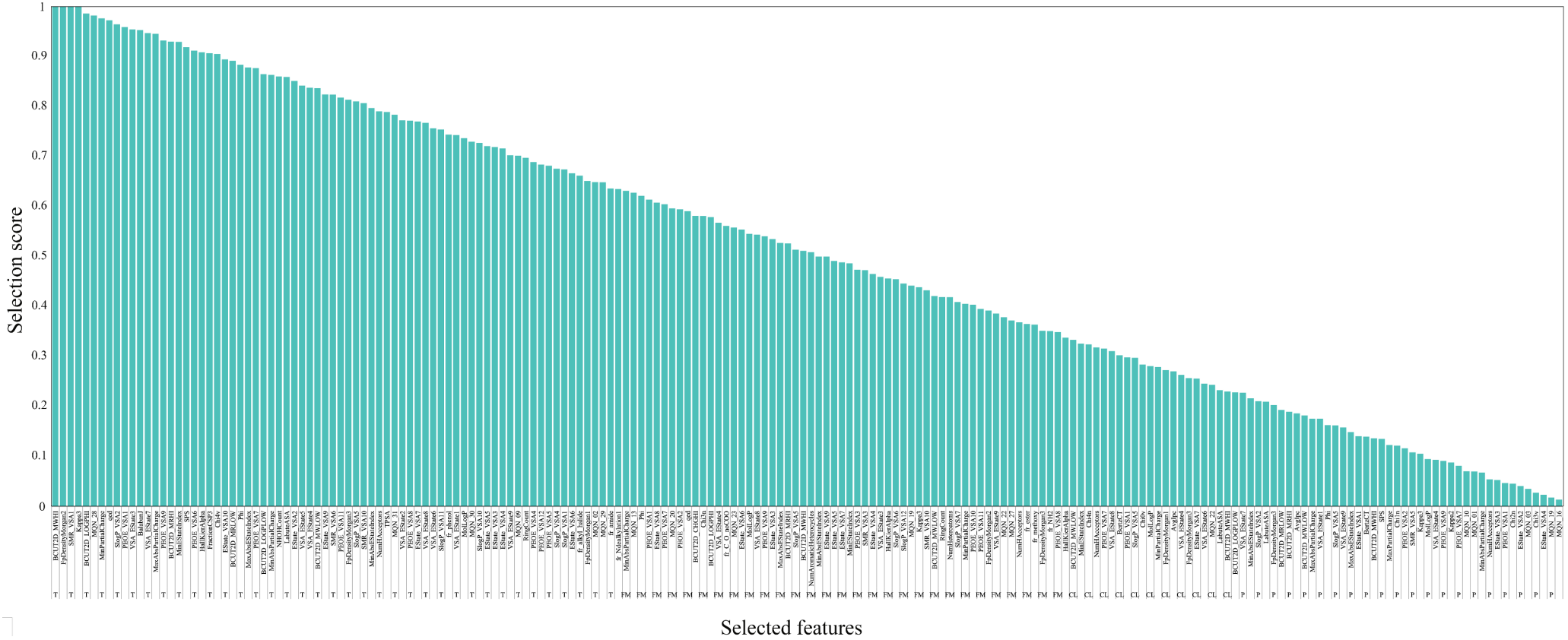

Figure S1. The stability-enhanced MRMR assigned scores for the selected features used in the similarity-based prediction model.