Supplementary Information

SAR Matrices: Automated Extraction of Information-Rich SAR Tables from Large Compound Data Sets

Anne Mai Wassermann, Peter Haebel, Nils Weskamp, and Jürgen Bajorath

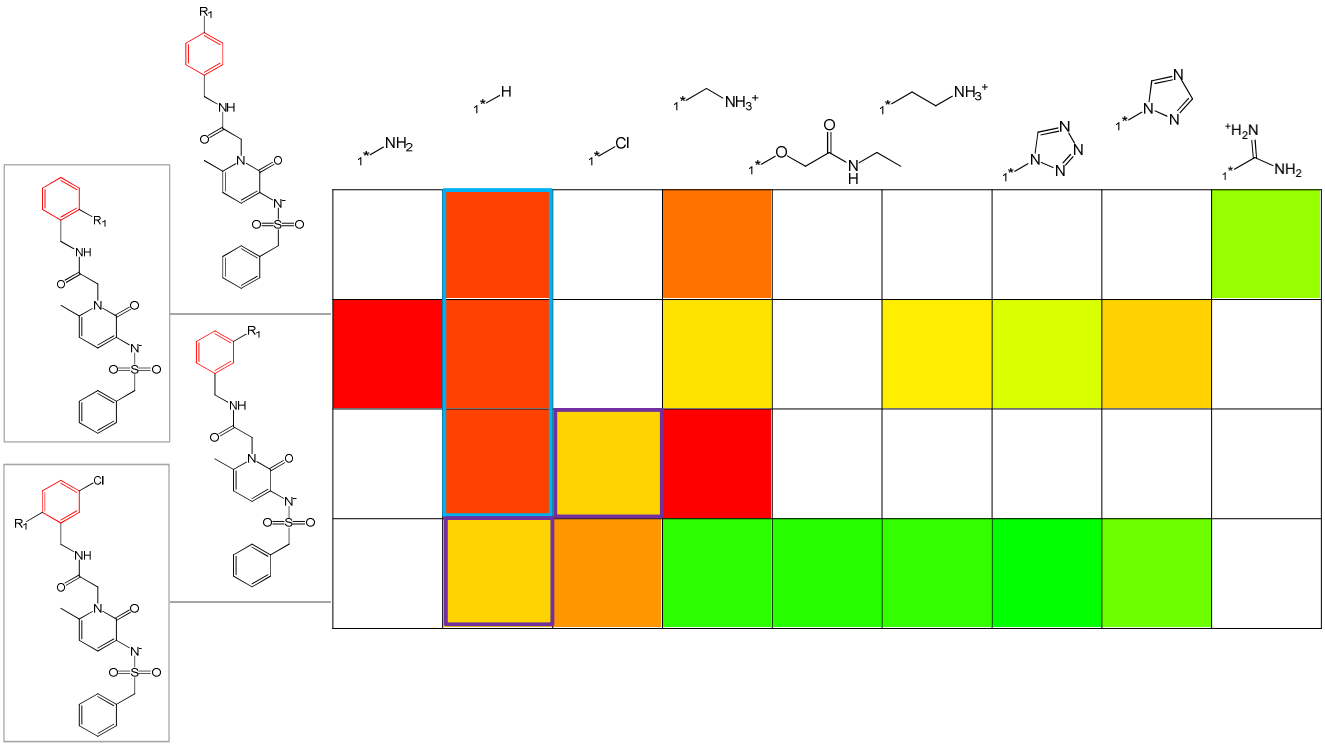
Supplementary Figures S1-S2

Supplementary Figure Legends

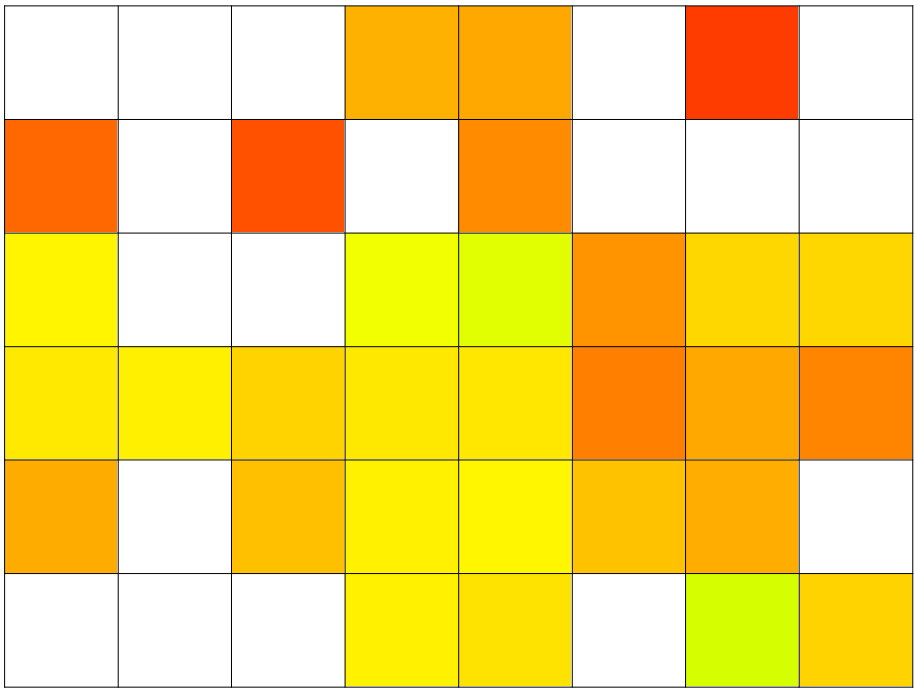
Supplementary Figure S1. Key-value pairing redundancies. Key-value pairings (matrix cells) that correspond to identical molecules are highlighted using color-coded box frames.

Supplementary Figure S2. Top-ranked matrices. For all four scoring schemes, top-ranked matrices are displayed. For simplicity, compound structures are omitted. (**a**) Top-ranked triple-cut matrix for the GlaxoSmithKline (GSK) data set using the KS-based scoring scheme. (**b**) Top-ranked single-cut matrix for the GSK data set according to the discontinuity-oriented scoring scheme. (**c**) Top-ranked single-cut matrix for the carbonic anhydrase I (CAI) data set using the substituent-oriented scoring scheme. (**d**) Top-ranked triple-cut matrix for the CAI data set according to SAR transfer scoring.

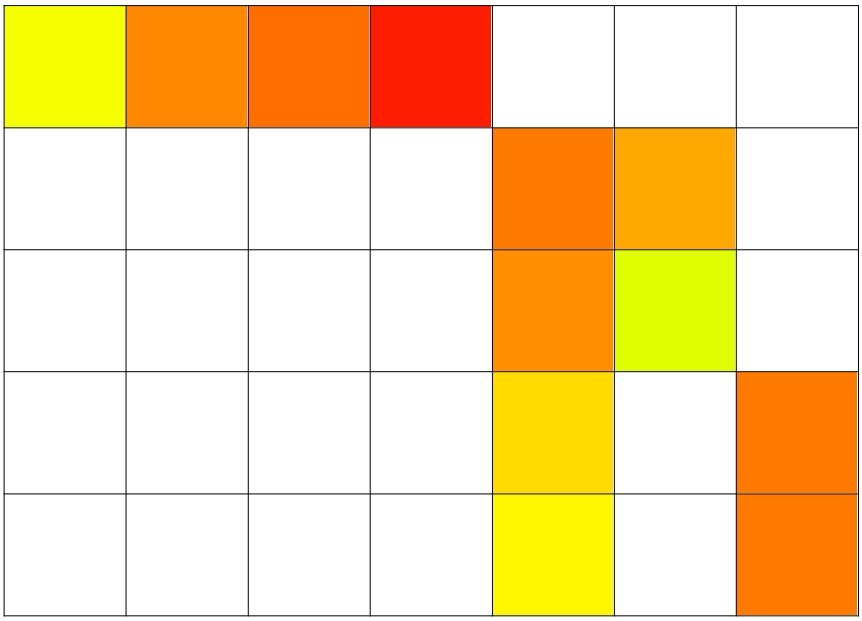
Supplementary Figure S1



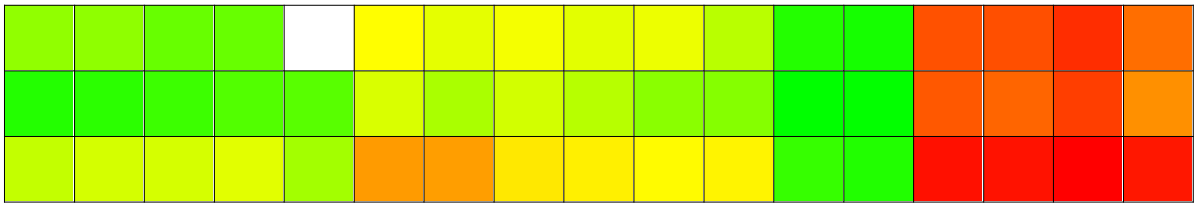
Supplementary Figure S2a



Supplementary Figure S2b



Supplementary Figure S2c



Supplementary Figure S2d