Electronic Supplementary Material (ESI) for Faraday Discussions.

Materials Discovery by Chemical Analogy: Role of Oxidation States in Structure Prediction

Daniel W. Davies,[1] Keith T. Butler,[1] Olexandr Isayev,[2] and Aron Walsh[3,4,a)]

[1)] *Centre for Sustainable Chemical Technologies and Department of Chemistry, University of Bath, Claverton Down, Bath BA2 7AY, UK*

[2)] *Laboratory of Molecular Modeling, Division of Chemical Biological and Medicinal Chemistry, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, North Carolina 27599, USA.*

[3)] *Department of Materials Science and Engineering, Yonsei University, Seoul 03722, Korea*

[4)] *Department of Materials, Imperial College London, Exhibition Road, London SW7 2AZ, UK*

(Dated: 7 February 2018)

[a)]Electronic mail: a.walsh@imperial.ac.uk

I. SUPPLEMENTAL DATA ITEMS

A. All species fractions

Distribution of all metal species included in the dataset, normalised by the total number of compounds containing a given species. Anions on the x-axes are in order of decreasing electronegativity. Numbers to the left of each species on the y-axes show the raw number occurrences of each oxidation state for each metal. Continued on the next page.

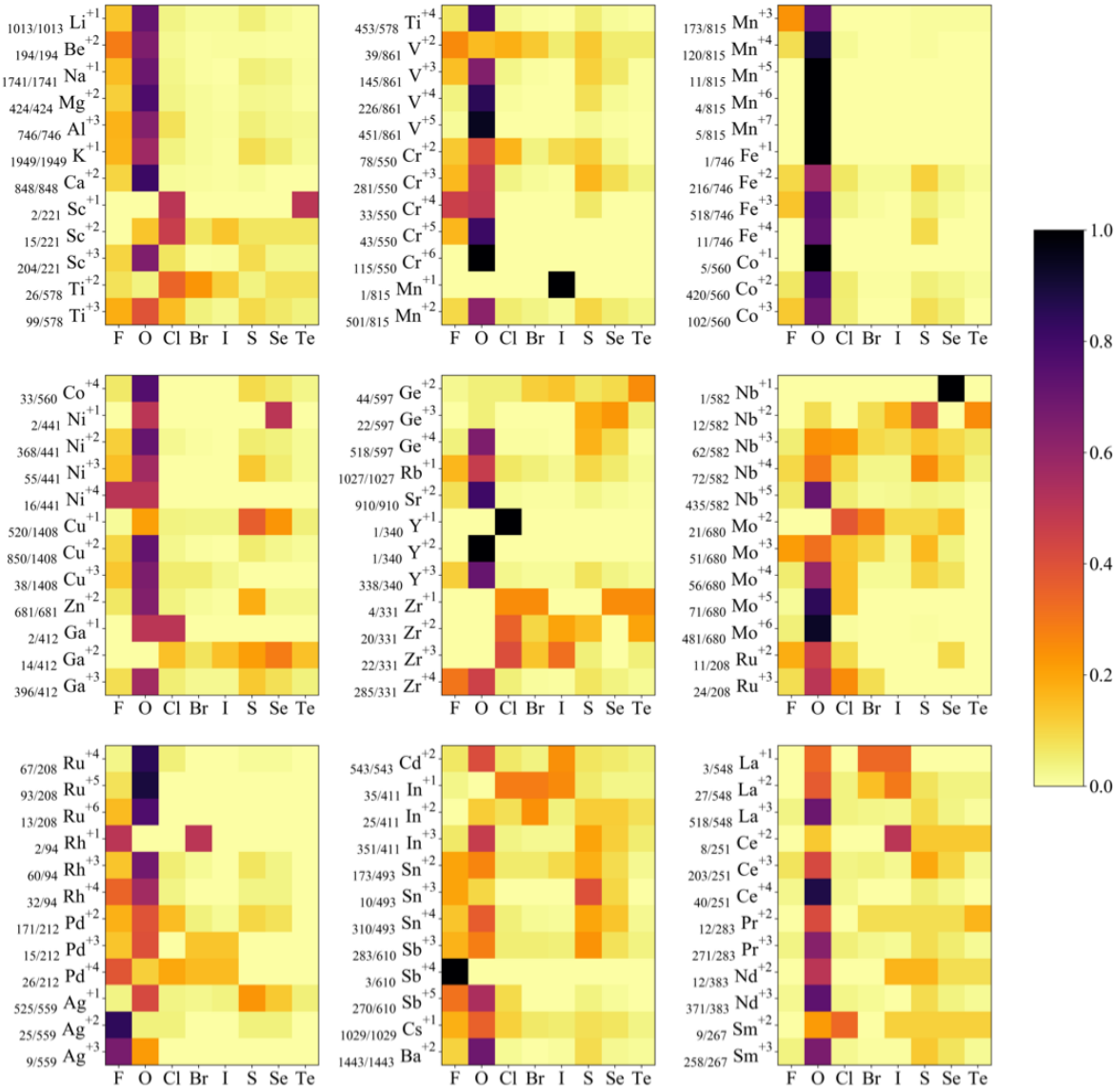

FIG. S1: Species Fractions part I.

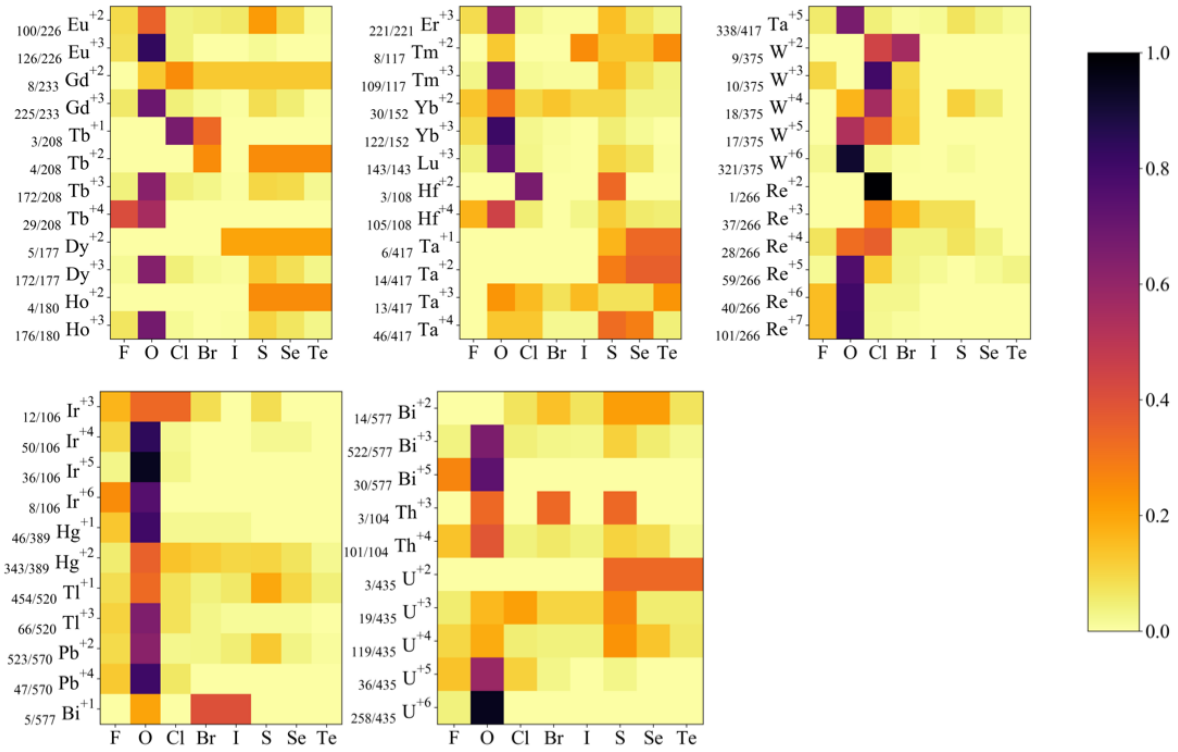

FIG. S2: Species Fractions part II.

B. Further species distributions

Additional plots of the distribution of some metal species in the dataset. The trends discussed in the main manuscript are also seen in the third row d-block metals (Figure S3) and the Lanthanides (Figure S4).

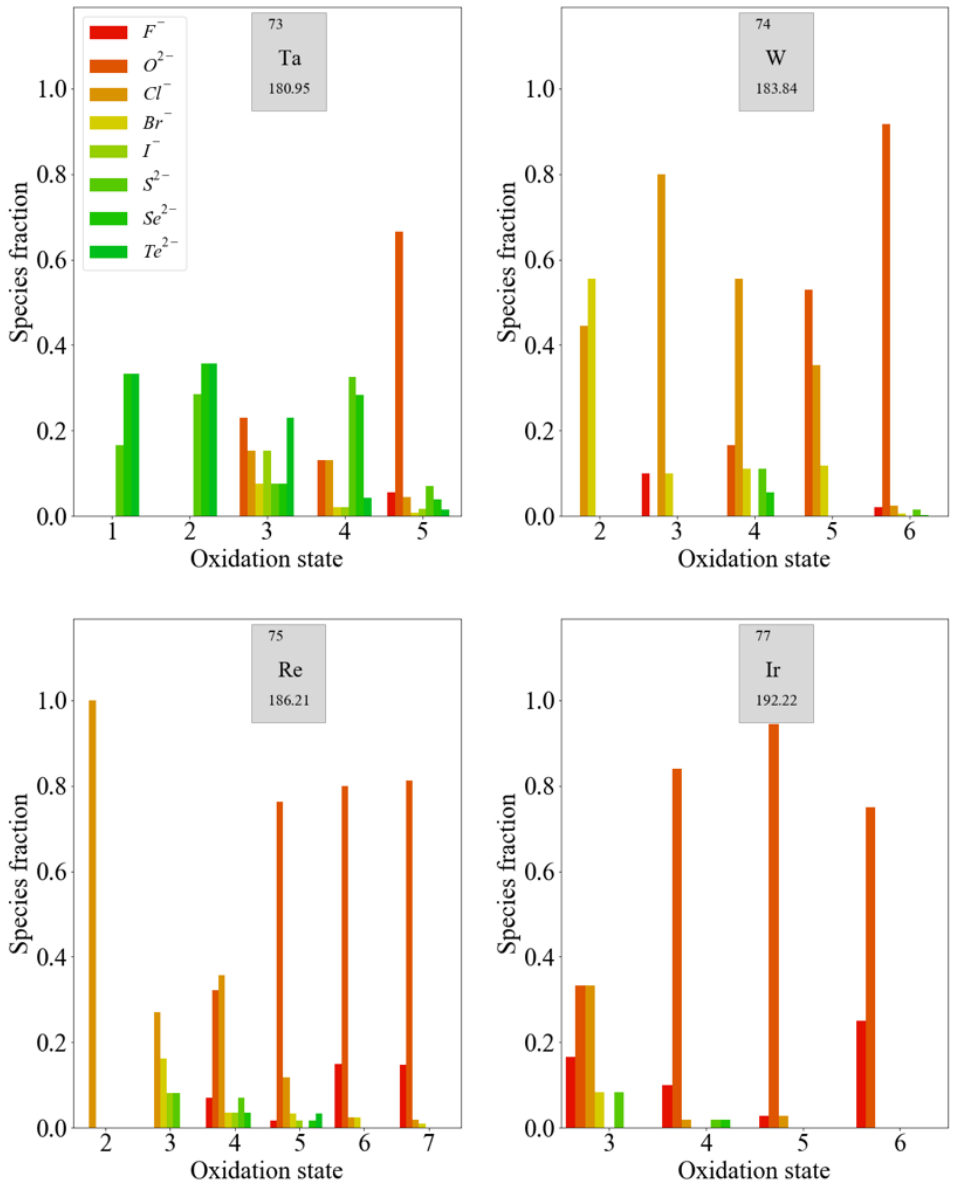

FIG. S3: Distribution of some third row transition metal species. The color scale represents the electronegativity of the most electronegative anion present in the compound from dark red (F, most electronegative) to dark green (Te, least electronegative).

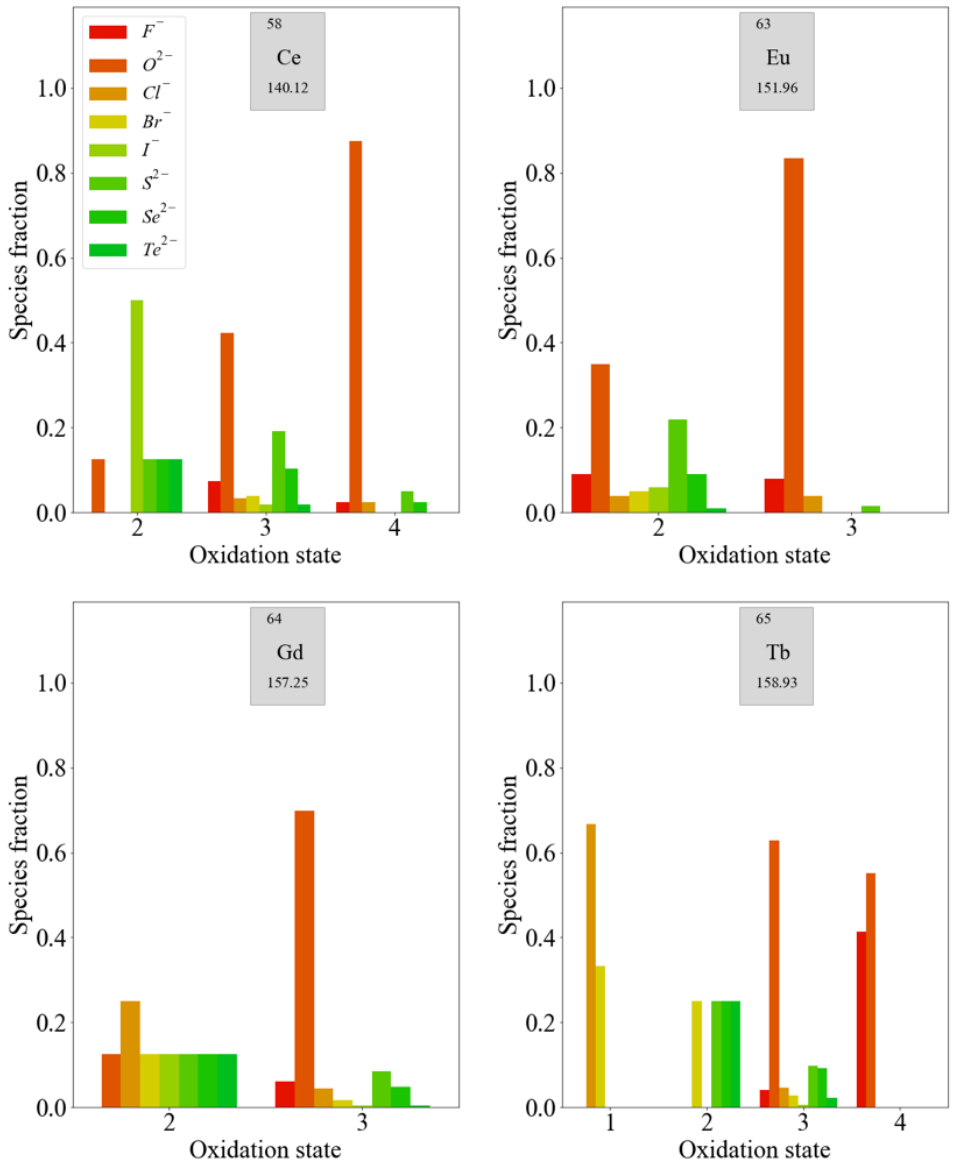

FIG. S4: Distribution of some Lanthanide metal species. The color scale represents the electronegativity of the most electronegative anion present in the compound from dark red (F, most electronegative) to dark green (Te, least electronegative).

C. All species–anion probabilities

Graphical representation of the lookup table used by the probabilistic model. The number of compounds containing a given species with the most electronegative anion is normalised by the total number of compounds containing the metal with the most electronegative anion. Anions on the x-axes are in order of decreasing electronegativity. Numbers to the left of each species on the y-axes show the raw number occurrences of each oxidation state for each metal. Continued on the next page.

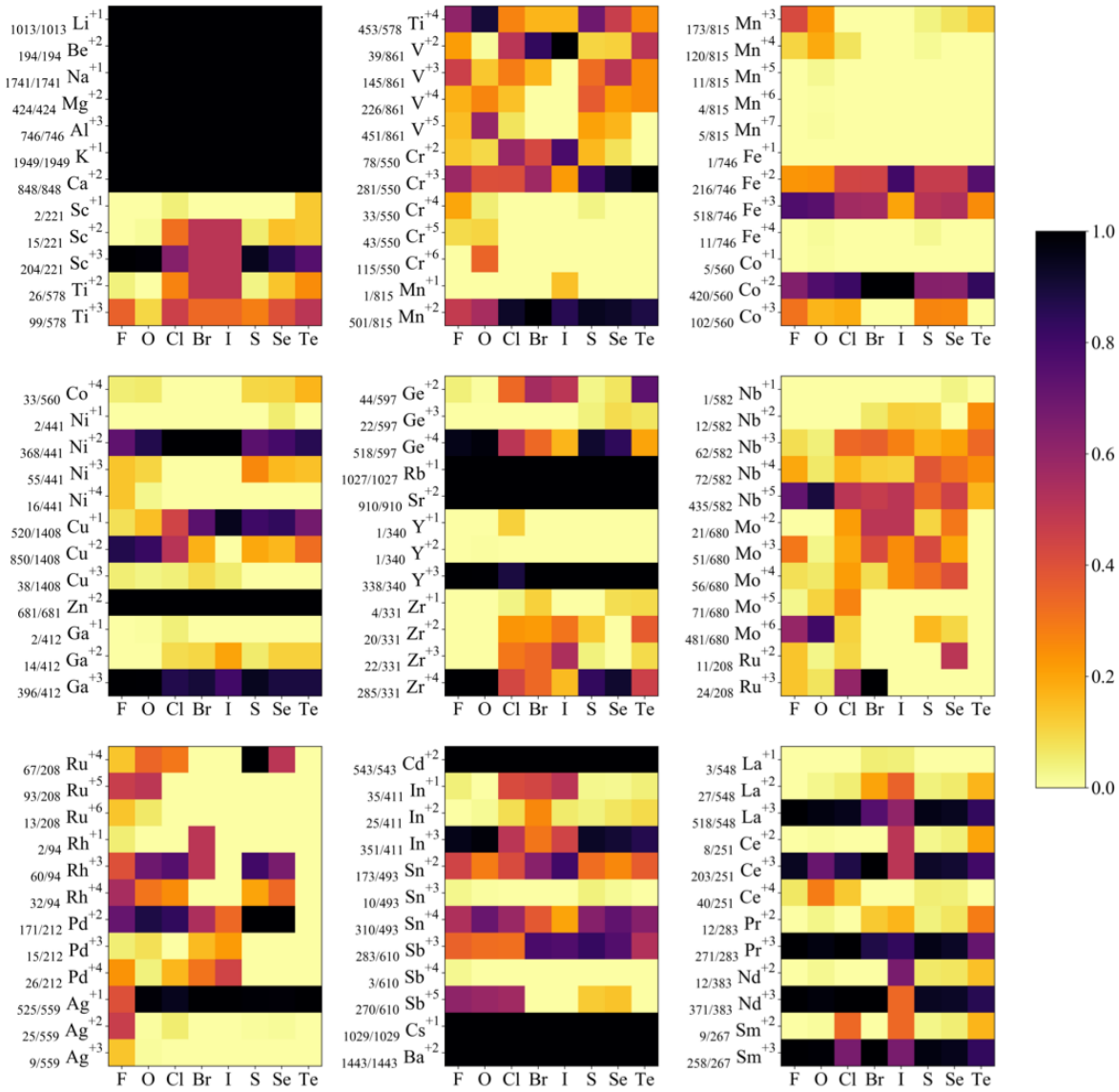

FIG. S5: Species–anion probabilities lookup table part I.

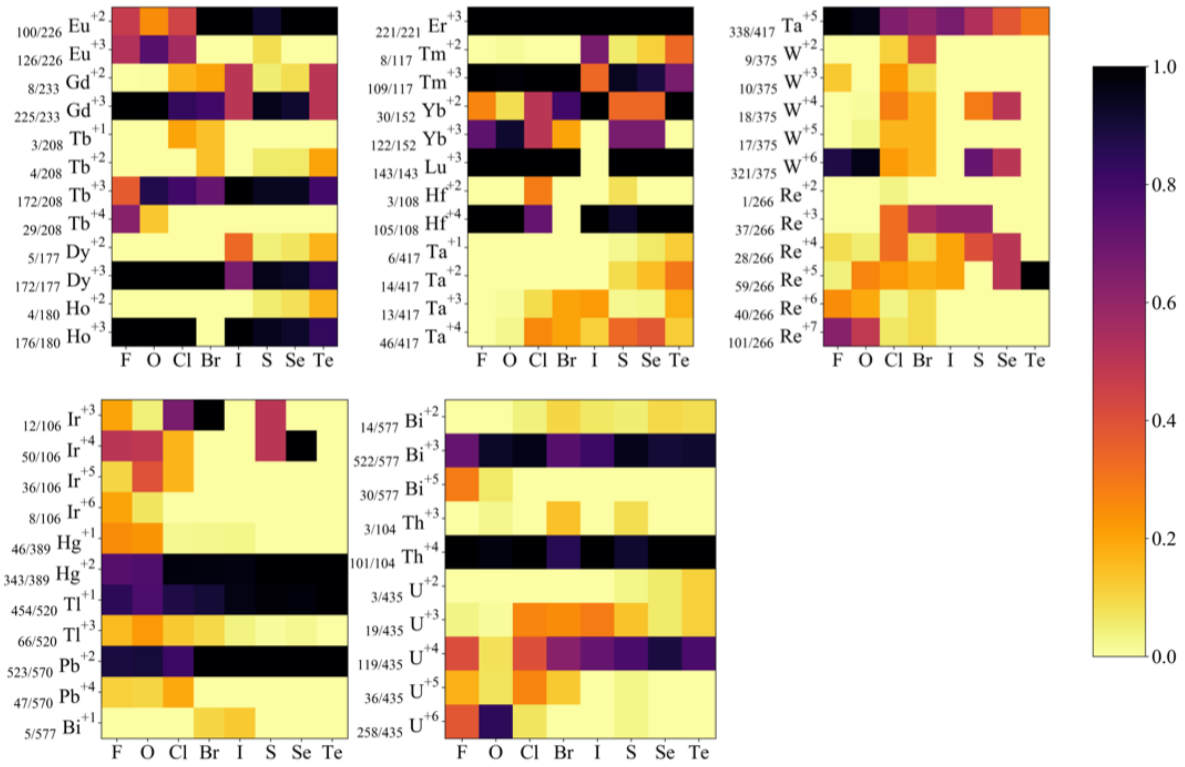

FIG. S6: Species–anion probabilities lookup table part II.