Correlating spectra to structure for water in, and on, crystals -Predictions and/or insight?

Jolla Kullgren[1], Andreas Röckert[1] and Kersti Hermansson
[1] Chemistry - Ångström, Uppsala University, Uppsala, Sweden

jolla@kemi.uu.se

Robust correlation curves are crucial for extracting structural insights from spectra. In this presentation, we evaluate several key structural descriptors (features) based on their ability to capture OH vibrational frequencies of water molecules at surfaces and in crystalline bulk hydrates (including a rich plethora of ice polymorphs). The results include data from Refs. [1-3] as well as new and previously unpublished data.
Our descriptors were selected to represent a progression from a geometric machine-learning (ML) descriptor towards those with gradually more physics (chemistry) contents, and with varying short- and long range effects (See Figure 1).
To evaluate the performance of the descriptions, we use a data-set specifically designed for the purpose with structures optimized at the Density Functional Theory level. Our analysis reveals that the simple geometric hydrogen-bond descriptor, R(H···O), is robust, straightforward, and remarkably accurate for correlating vibrational frequency and structure. When we use a more elaborate machine learning (ML) descriptor encoding more contributions from atoms in the external environment of the water we achieve a better correlation than with R(H···O). This comes at the cost of reduced interpretability and natural leads to the question: Can we reach a good balance between prediction and insight?

References

[1] Andreas Röckert, Jolla Kullgren, Kersti Hermansson, J. Chem. Theory Comput., 12 (2022), 7683–769.
[2] Andreas Röckert, Jolla Kullgren, Daniel Sethio, Lorenzo Agosta Kersti Hermansson, J. Chem. Phys., 159 (2023), 044705.
[3] Jolla Kullgren, Andreas Röckert, Kersti Hermansson, Phys. Chem. C, 127 (2023), 13740–13750.

Figures

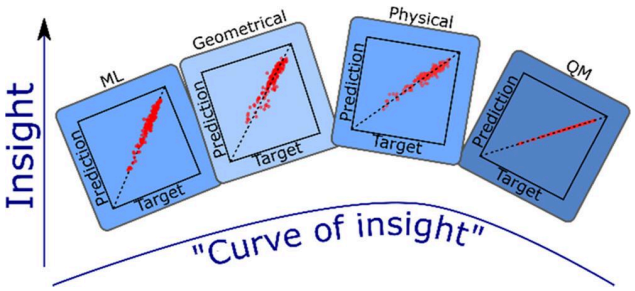

Figure 1. A schematic overview of descriptors with varying degrees of complexity, physical insight and level of predictive power. The figure is taken from Ref. [3].