Review of: "Fragment libraries designed to be functionally diverse recover protein binding information more efficiently than standard structurally diverse libraries"

Yun Shi

Potential competing interests: One reference cited in this review (https://doi.org/10.3390/molecules24152838) is the work of the reviewing author. The author declares no other competing interests.

This is a great study that examines the functional diversity of fragment libraries, providing valuable insights into the design of fragment libraries. The principles explored in this study may also be applied towards the design of high-throughput screening libraries with “Ro5” compounds in addition to “Ro3” compounds. I'd like to raise three points that could further improve this manuscript.

First, the authors used a limited dataset of 524 protein-fragment complexes from 11 targets and 347 (XChem). Is it possible to expand this dataset to include more protein-fragment complexes that are available publicly in Protein Databank? I think many more (protein) targets would have fragment-bound crystal structures published.

Second, the authors looked at interaction fingerprints (IFP) at the atomic level and the residual level. Why not include analyses for the molecular (target) level, looking at interactions between the whole fragment (not individual atoms) and the protein? From a drug discovery perspective, “function” could also mean the ability of a fragment to bind certain targets but not others.

Third, the authors used (normalized) richness, i.e. number of (unique) IFPs, as a metric for diversity, such as in Figure 2. Could true diversity (**https://doi.org/10.2307/1934352)** also be used? It has been used to analyze fragment libraries (e.g. **https://doi.org/10.3390/molecules24152838)** before as it not only counts the number of features, i.e. the proportion of IFPs relative to the total IFPs of the entire dataset, but also takes into account the evenness of IFP distribution. A library with better true diversity may ultimately improve the target selectivity of fragment hits when molecular (target) level IFPs were used.

Other minor points: for the Materials and Methods section, could the authors come up with a flow chart to better inform the audience about their methodology? In the Results section, when the authors state “90 pairs … identical interactions”, do they mean a similarity value of 1, i.e. 100% identical? On Figure 1b, the pale blue-coloured fragment seems to have additional interactions with the protein than the other two fragments.