

## **Comparison of automated and manual patent chemistry extraction methods**

Luca Bartek, Stefan Senger, George Papadatos, Anna Gaulton

Patents are an important source of chemical information for researchers, as they are usually the first place where new chemical entities are published. With an increasing number of patents each year, it has become challenging to keep databases up-to-date by using traditional methods (manual curation) and engines have been developed to automatically extract compounds from new patents. But how reliable are these computer-assisted methods? We have compared two automatically generated sources, IBM SIIP and SureChEMBL, using two manually curated sources (SciFinder and Reaxys) as a “gold standard”.