Exploring Chemical Space and Commercial Compounds for Kinase Inhibitors and Fragments

Mark. L. Nelson [1], and Hugo O. Villar[2] and Mark R. Hansen[2]
Frontier Scientific, Inc., Logan, UT and Newark, DE[1], and Altoris, Inc., San Diego, CA[2]

Background

Kinases account for 25-30% of all targets screened for in the Pharmaceutical industry, and represent over 500 distinct proteins in diverse families implicated in cellular growth, dysregulation and in cancer. General screening libraries of structurally diverse compounds have seen smaller, focused libraries being studied, from either corporate collections, synthesized by directed combinatorial synthesis, or from commercial suppliers as targeted chemotypes. While there has been determination of "privileged" structures, the rules for kinase inhibitors have evolved for enriching collections of kinase active compounds.

Our chemoinformatics program, **CHEM Exchange™**, is based upon Altoris' SARvision technology, which allows the user to input commercially available chemical collections as structure-data files. The program then searches for kinase fragments and building blocks throughout the commercially available chemicals, allowing for the efficient design, procurement and ultimately, synthesis of kinase inhibitors through either de novo synthesis, or through synthesis using fragments, privileged building blocks and heterocyclic scaffolds.

Currently, chemical aggregation of commercial compounds is available online, and searching for chemotypes is limited to structure-based searches. Now custom databases of commercially available compounds can be compiled by the user, and compounds **that actually exist** can be found readily using input of fragments, chemotypes and building blocks found useful as kinase inhibitors.

Chemical Diversity and Commercial Compounds

While the Chemical Abstracts Service CAS REGISTRY[SM] claims to have over 100M compounds in its database, and the ZINC15 chemical database states that 100M compounds are commercially available, the reality is database aggregations of all planetary chemical compounds actually available is far less. While eMolecules purports to have available 7.5M and Ambinter claims 7M, there is no actual count of commercially available compounds to rely on. Additionally, some databases harbor redundant, non-existent or hypothetical compounds, limiting their usefulness and impeding chemical synthesis and research.

Comparing scaffolds content of different libraries

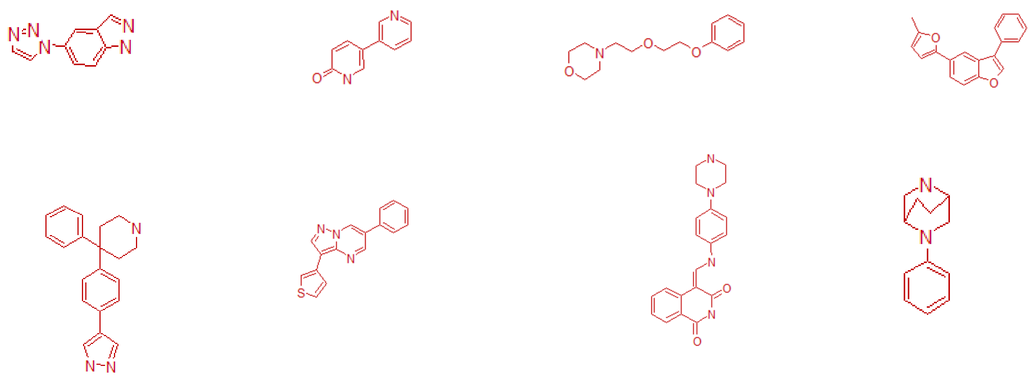
Using SARvision|SM the number of unique scaffolds present in ChEMBL Sarfari Kinase dataset, as well as 4 other vendor catalogs was determined. The percentage of kinase scaffolds populated in each vendor catalog with at least 3 molecules, was determined and it is shown in table 1 :

Library	Unique Scaffolds	Molecules	%Kinase
ChemBL- Kinase Sarfari	9320	53966	100
Cayman	2510	7797	8.9
Frontier	7182	81228	21.8
Maybridge	7332	58558	19.0
Sigma-Aldrich	14411	234400	23.2
ALL 4 VENDORS	25383	387321	33.0

The results show that neither number of total molecules or the number of scaffolds they contain is a good predictor of suitability for tailoring to a particular target class.
While the examples shown are carried out with relatively small chemical collections, the analysis can be done for much larger datasets, and it is largely limited only by computational resources.

What scaffolds are not found with selected vendors

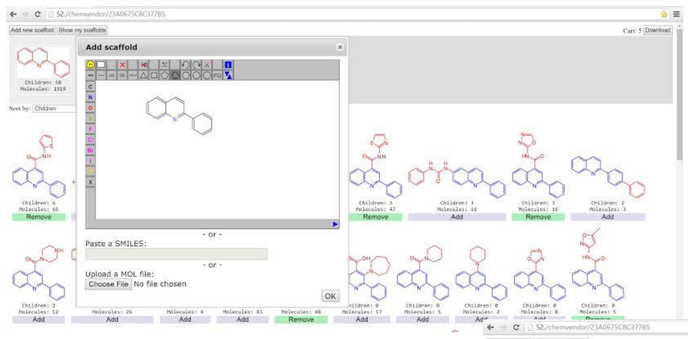
With this approach the user can identify scaffolds that are present in the ChEMBL Kinase Sarfari library, but are not found in any of the chosen vendor catalogs. Some of the frequently reported scaffolds in Kinase Sarfari that are not found in any of the 4 catalogs selected include:



In these situations, users may want to expand the search using a much larger and diverse chemical collection. For those scaffolds that are not present or are underrepresented in the users' existing collection **CHEM Exchange™** offers a way to browse chemicals by chemotype.

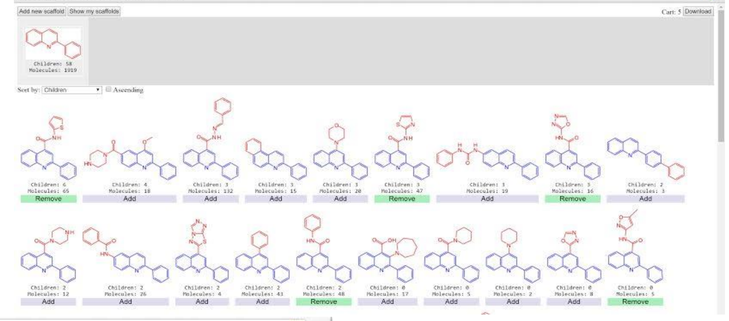
Browsing for scaffolds

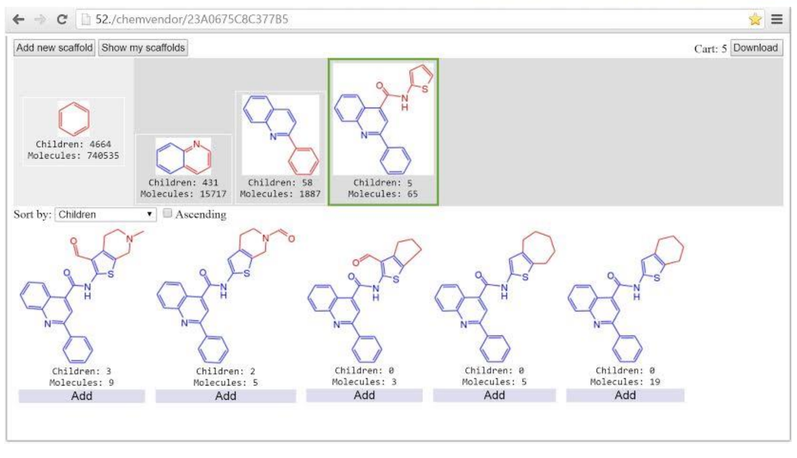
Once a set of scaffolds of interest have been identified, through **CHEM Exchange™** users can quickly browse through vendor catalogs, looking at scaffolds or substructures of increasing complexity:



1. Users can draw a scaffold or substructure of interest using the interface provided in the program.
Using the substructure drawn, the program will seek other more complex substructures to populate the table.

2. For each child of the scaffold drawn, the Interface will display the number of children of increasing complexity as well as the number of molecules that contain that substructure.





3. Users repeat the operation for as many scaffolds as desired, and can then choose which to be added or removed from consideration. Finally, the molecules associated can be downloaded for further processing and ordering.

CHEM Exchange™ A Program for Compound Procurement

Our new web service, named **CHEM Exchange™** will allow chemists to create their own chemical database or subscribe to those vetted and provided by Frontier Scientific, Inc., using commercially available compound collections that are provided by the vendor, and allows for input of chemical fragments, scaffolds and chemotypes, for generating useful and commercially available starting materials for drug design and organic synthesis.

The user inputs a collection of chemical structures, and SARvision uses its scaffold perception algorithms to identify chemotypes present in the dataset and displayed in a hierarchy from simpler to complex, via **CHEM Exchange™ .** The ability to browse for scaffolds rather than searching for specific substructures, provides a new powerful way to enhance chemical libraries.

In this presentation, we will show how analysis of chemical collections at the scaffold level can can be powerful tool when designing chemical libraries for screening. To that end, SARvision is used to determine the overlap between different chemical libraries, and select those that can be more relevant and better tailored for a specific target class.

Conclusions

- Comparing and analyzing libraries at the level of scaffolds rather than individual molecules can be a powerful way to select compounds
- The similarity of chemical collections at the scaffold level can quickly be gaged, and provide a rapid way to assess the suitability of different libraries towards the selection of compounds for a given compound of interest.
- Analysis at the scaffold level as done in SARvision provides a complementary tool to the selection of compounds based on cheminformatics approaches based on molecular descriptors
- While explorations of the Kinase Inhibitor class of compounds is described here, **CHEM Exchange™** is suitable for all chemotypes, substructures, functional groups and their biological targets.
- Even large chemical collections may lack some chemotypes commonly present in tailored libraries.

References

1. SARvision|SM, Altoris Inc., San Diego, CA www.chemapps.com
2. https://www.ebi.ac.uk/chembl/sarfari/kinasesarfari
3. http://zinc15.docking.org