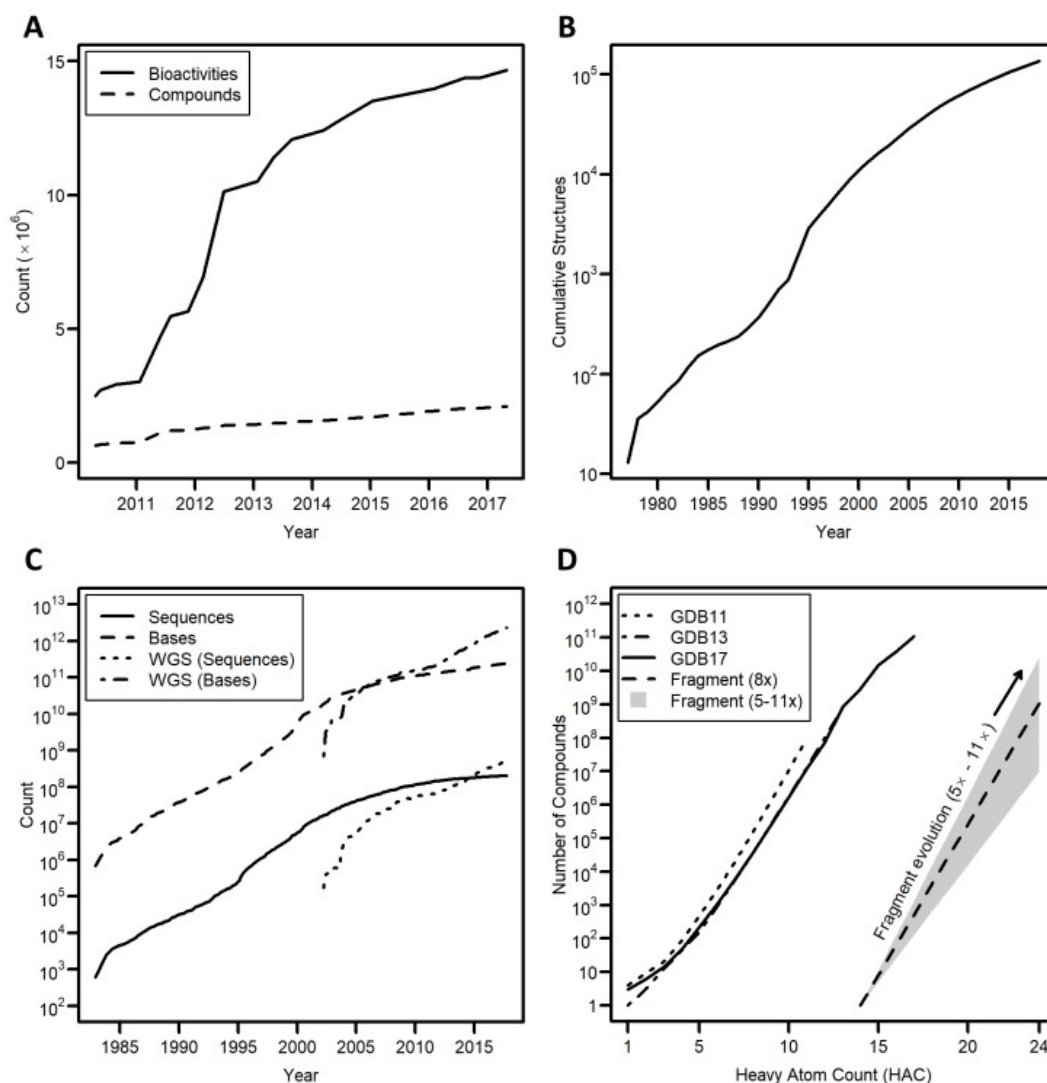# Supplementary Materials

Five Years of the KNIME Vernalis Cheminformatics Community Contribution

Stephen D. Roughley[1,*]

[1]*Department of Chemistry & Cheminformatics, Vernalis Research Ltd, Granta Park, Great Abington, Cambridge CB21 6GB, UK*

Figure S1 – Growth of publicly available data (Larger version of Figure 1)

Data collated as follows:

A – ChEMBL – The data were compiled from the individual release files
'chembl_XX_release_notes.txt' for each release at
ftp://ftp.ebi.ac.uk/pub/databases/chembl/ChEMBLdb/releases/, where 'XX' is the release number

B – RCSB PDB Structures – The data were downloaded from
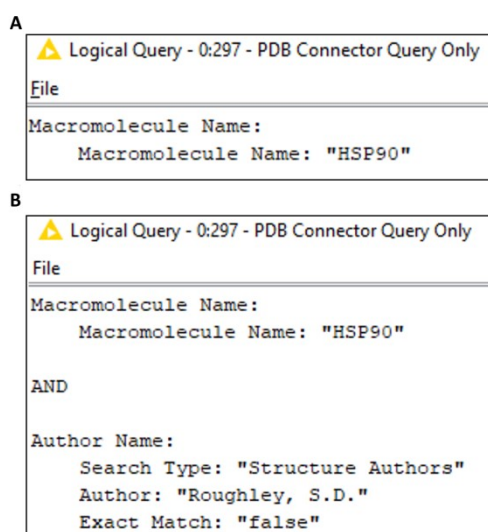https://www.rcsb.org/pdb/statistics/contentGrowthChart.do?content=total&seqid=100

C – NCBI GenBank – The data were downloaded from
https://www.ncbi.nlm.nih.gov/genbank/statistics/

D – GCB 11 / GDB 13 / GDB 17 – The data were compiled from the references cited in the main text

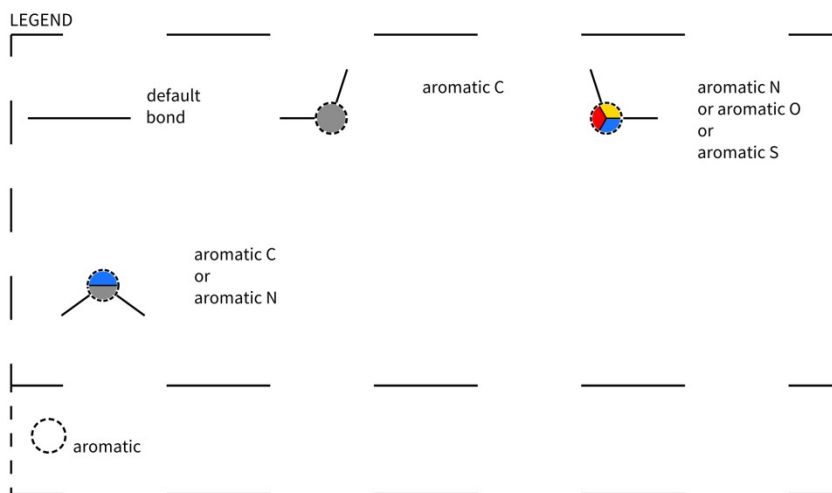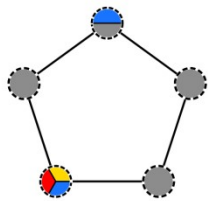Figure S2. XML Queries for the simplified query views shown in Figure 6:

A
```
⚠ Logical Query - 0:297 - PDB Connector Query Only
File
Macromolecule Name:
    Macromolecule Name: "HSP90"
```

B
```
⚠ Logical Query - 0:297 - PDB Connector Query Only
File
Macromolecule Name:
    Macromolecule Name: "HSP90"

AND

Author Name:
    Search Type: "Structure Authors"
    Author: "Roughley, S.D."
    Exact Match: "false"
```

6A:

```xml
<orgPdbQuery>
    <queryType>org.pdb.query.simple.MoleculeNameQuery</queryType>
    <macromoleculeName>HSP90</macromoleculeName>
</orgPdbQuery>
```

6B:

```xml
<orgPdbCompositeQuery version="1.0">
        <queryRefinement>
                <queryRefinementLevel>0</queryRefinementLevel>
                <orgPdbQuery>
                        <queryType>org.pdb.query.simple.MoleculeNameQuery</queryType>
                        <macromoleculeName>HSP90</macromoleculeName>
                </orgPdbQuery>
        </queryRefinement>
        <queryRefinement>
                <queryRefinementLevel>1</queryRefinementLevel>
                <conjunctionType>and</conjunctionType>
                <orgPdbQuery>
                        <queryType>org.pdb.query.simple.AdvancedAuthorQuery</queryType>
                        <searchType>1</searchType>
                        <audit_author.name>Roughley, S.D.</audit_author.name>
                        <exactMatch>false</exactMatch>
                </orgPdbQuery>
        </queryRefinement>
</orgPdbCompositeQuery>
```
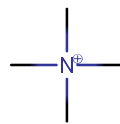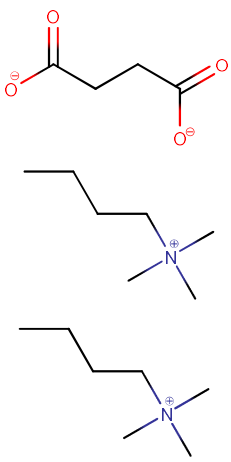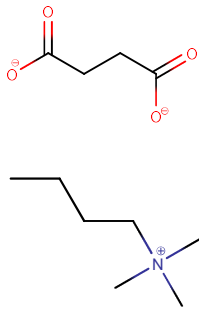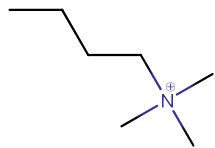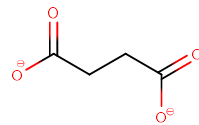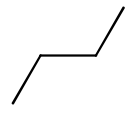
Figure S3. SMARTSviewer representation of `c1[n,o,s]c[c,n]c1`

c1[n,o,s]c[c,n]c1

LEGEND

—— default bond	aromatic C	aromatic N or aromatic O or aromatic S
aromatic C or aromatic N		
aromatic		

Table S1 Structural Representations of SMILES in Table 3
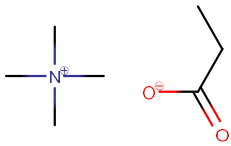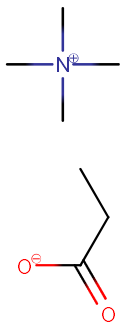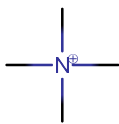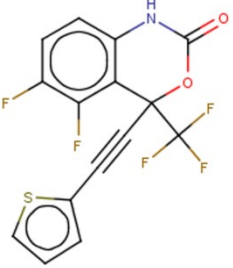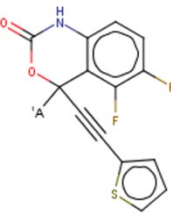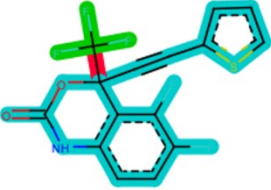
ID	SMILES	Default	1st Unique	Longest SMILES
1	C[N+](C)(C)C.CCC(=O)[O-]	C[N+](C)(C)C, CCC(=O)[O-]	C[N+](C)(C)C	C[N+](C)(C)C
2	C[N+](C)(C)CCCC.C[N+](C)(C)CCCC.[O-]C(=O)CCC(=O)[O-]	C[N+](C)(C)CCCC, [O-]C(=O)CCC(=O)[O-]	C[N+](C)(C)CCCC	[O-]C(=O)CCC(=O)[O-]
3	C1C.CC1	C1C[a], CC1[a]	C1C[a]	C1C[a]

Figure S4 – Rendering of MMP Fragmentation

Rendering Key – Green – 'Value', Cyan – 'Key', Red – Cut bond(s). The colour scheme can be changed in the node settings.