Chemical fragment analysis of halogen bonds in protein binding sites

Edith Chan

University College London, UK

13 March 2016
A Medicinal Chemist’s Problem

- A common question from medicinal chemists - what compounds should be made for a given protein target?
A Medicinal Chemist’s Problem

- A common question from medicinal chemists - what compounds should be made for a given protein target?
- Chemists very good at generating new ideas when seeing other structures.
Chemical Strategy of Medicinal Chemists

- Screening to find leads
- Natural ligand
- Virtual screening (hit rate improvement of 10x)
- SAR studies
- Fragment based approaches

- Med chem need
  - A simple way to select likely binding molecules for a protein binding site
Fragment-Based Drug Discovery

- FBDD has become an established and successful paradigm for the past 15 years.
- Small chemical structures are screened to probe the binding site and then to identify larger molecules to bind.
- Most platforms are laboratory based – like X-ray, NMR.
Design of Fragments: Current Status

- Physicochemical filtering Ro3 (Ro2\(\frac{1}{2}\)), small, soluble
- Screen out reactive groups
- Chemical handles for further manipulation
- Most groups use 500-2000 compounds, some use 10,000+
- A good binding affinity - ligand efficiency often used to express binding affinity of fragments
  
  - Typical good LE is > 0.2

\[
LE = \frac{-\Delta G}{HAC} = \frac{-RT \ln(K_d)}{HAC}
\]

LE = ligand efficiency, HAC = heavy atom count
Common Computational Approaches

- Fragment /property centered
  - Analyzing drug-related databases for molecular frameworks, property, diversity, and privileged scaffolds, etc.

DrugBank, WDI

- HD
- MW
- HA
- PSA
- CLOGP

DIVERSITY

Rule of 3
Common Computational Approaches

- 3D fragments experiment-centred
  - Most current fragments are based around aromatic or heteroaromatic structures
  - Use Diversity orientated synthesis to design new fragment based libraries

Route to three-dimensional fragments using diversity-oriented synthesis

Alvin W. Hung\textsuperscript{a}, Alex Ramek\textsuperscript{b}, Yikai Wang\textsuperscript{c}, Taner Kaya\textsuperscript{d}, J. Anthony Wilson\textsuperscript{e}, Paul A. Clemons\textsuperscript{f}, and Damian W. Young\textsuperscript{a,1}

\textsuperscript{a}Chemical Biology Program, Broad Institute of Harvard and Massachusetts Institute of Technology, 7 Cambridge Center, Cambridge, MA 02142; and
\textsuperscript{b}Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138

Edited by Stuart L. Schreiber, Broad Institute, Cambridge, MA, and approved March 3, 2011 (received for review December 22, 2010)

PNAS | April 26, 2011 | vol. 108 | no. 17 | 6799–6804
Target Centred

• Given a new or unknown target, what kind of scaffolds or fragments should chemists try?

• Using X-ray structures to define fragments

• A cheminformatics database identifying the chemical motifs or fragments that preferentially interact with particular protein side chains in the binding site.
Our Approach

- Use the PDB as a source of ligand-protein information – deriving the interacting “fragment”
  - High quality information
  - Previous study on Chemical Fragments that Hydrogen Bond to Asp, Glu, Arg, His Side Chains in Protein Binding Sites

Definition of fragment
- The largest ring assembly containing the atoms involved in halogen bond(s) to main or side chains

Definition of halogen bonds
- F, Cl, Br, I

![Diagram of halogen bonds](image)
To date: The Protein Data Bank (PDB) has about 110K structures, with ~88K that have co-crystallized ligand(s).

PDB 3D co-crystallized structures provide info on protein-ligand interactions.

Analyses of specific ligand/side chain interactions tended to focus on single ligand atoms rather than chemical fragments.
Protein-Ligand Interactions – 3D

Interactions in 3D: Hbonds, hydrophobic, van der Waals, etc

Halogen Bond
Protein-ligand interactions – 2D
• Interacting motif is the largest ring assembly containing the atoms involved in the halogen bond(s).

• Other substituents on the ring assembly are removed if they are not involved in the halogen bond to the relevant protein main or side chain.
Data set from PDB

- The data set for this study was compiled in October 2015.
- Approximately 2600 PDB structures.
- Halbond between ligands and specific protein side chains were extracted from the data files in PDBsum.
- PDBsum uses the HBPLUS program to calculate potential hydrogen bonds and non-bonded contacts.
- The 3D coordinates of the ligand and interacting side chain of interest were then extracted from the parent PDB file and translated into MDL SD format for further processing.
Halogen containing ligands in the PDB

- 2009 unique ligand in 2574 ligands/pdb complexes
- Halogens interact with protein sidechains via
  - halogen bond and
  - cation and pi
- In this study, we will concentrate on halogen bond only
Ligands – Halogen bond

- Data set is rather small: 641 ligands, represented by 266 unique ligands that have specific halogen bond with protein’s residues.

- We used a cut-off resolution of 2Å while most other studies have used 3Å.

- There are many examples in the latter cases where the bond length between halogen and sidechains were not accurately determined.

<table>
<thead>
<tr>
<th></th>
<th>Ligands</th>
<th>Unique ligands</th>
<th>Unique fragments</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>516</td>
<td>183</td>
<td>44</td>
</tr>
<tr>
<td>Cl</td>
<td>89</td>
<td>60</td>
<td>13</td>
</tr>
<tr>
<td>Br</td>
<td>30</td>
<td>18</td>
<td>7</td>
</tr>
<tr>
<td>I</td>
<td>6</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>total</td>
<td>641</td>
<td>266</td>
<td>65</td>
</tr>
</tbody>
</table>
Fragments

• In the PDB, there are far more data that show F halogen bonding with protein sidechain

• There are many more unique fragments in F containing fragments than other halogens

<table>
<thead>
<tr>
<th></th>
<th>Ligands</th>
<th>Unique ligands</th>
<th>Unique fragments</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>516</td>
<td>183</td>
<td>44</td>
</tr>
<tr>
<td>Cl</td>
<td>89</td>
<td>60</td>
<td>13</td>
</tr>
<tr>
<td>Br</td>
<td>30</td>
<td>18</td>
<td>7</td>
</tr>
<tr>
<td>I</td>
<td>6</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>total</td>
<td>641</td>
<td>266</td>
<td>65</td>
</tr>
</tbody>
</table>
Fluorine

- Quite a diverse set of fragments (188 unique ligands, 44 unique fragments)
- However dominated by phenyl-F or phenyl-CF$_3$

Unique ligands 44 18
Unique PDB 72 20

Table S3: Ligand fragments containing F

The table shows a list of ligand fragments containing the halogen F, bound to proteins in the PDB. The counts show the number of unique ligand molecules containing the fragment and number of PDB files in which an interaction between the halogen atom and the protein is observed. Further information can be found by following the links.
Fluorine

- Non aromatic fragments.
F and amino acids

- 517 F halogen bonds, 2/3 bond with “nitrogen”; 1/3 with “oxygen”
  - 64% N (22% main and 42% side) and 36% O
- Not many cases bond with S (Met)
- Do not bond with negatively charged residues such as Asp or Glu – F is exclusively H acceptor!

- Sidechain O: S, T, and Y
- Sidechain N:
  - N, Q, W – HB donor
- Main chain N:
  - 16 of the 20 amino acids – HB donor, no specificity
Chlorine

- Not many unique fragments (13 cases)
- However dominated by phenyl-Cl
Cl and amino acids

- 90 Cl halogen bonds, 64% N (22% main and 42% side) and 36% O
- Do not bond with negatively charged residues such as Asp or Glu. Very similar to F!

- Sidechain O: S, T, and Y
  - Sidechain N:
    - H, R, K – positively charged
    - N, Q – HB donor
  - Main chain N:
    - 11 of the 20 amino acids – HB donor, no specificity
Halogen Bond

- A halogen atom can both donate and accept in a halogen bond.
- No F or Cl is detected to form hydrogen bonds with Asp or Glu
  - confirming that a binding site’s Asp and Glu are always hydrogen bond acceptors as well as negatively charged in protein structures.
- F and Cl attached to ring systems are detected with low frequency for Arg and His side chains, another example of HA.
Bromide

- Not many fragments found.
- Like other halogens, phenyl-Br is most frequently used.
- -Br in this dataset interacts with
  - Nside (R, N, Q) and
  - Oside (S, T)
  - Nmain (R, L, F)
Iodine

- Not many cases.
- Only 1 fragment from 5 ligands is found
- Interact with O-side:
  - T, Y
- No interaction with N detected.
Results in webpages


**LigFrag - a database of protein side chain/ligand-fragment interactions**

*J. Med. Chem. Supporting information*

**An Analysis of the Chemical Fragments that Hydrogen Bond to Asp, Glu, Arg, His Side Chains in Protein Binding Sites**

A.W. Edith Chan, Roman A. Laskowski, and David L. Selwood

*J Med Chem, 2010, 53, 3088-3094*

Tables S1-S4 in this Supporting Information show the chemical fragments that interact via hydrogen bonds to the side chains of Asp, Glu, Arg and His in protein binding sites.

Table S5 lists the excluded ligands.

<table>
<thead>
<tr>
<th>Table</th>
<th>Ligand</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>Asp - Aspartate</td>
</tr>
<tr>
<td>S2</td>
<td>Glu - Glutamate</td>
</tr>
<tr>
<td>S3</td>
<td>Arg - Arginine</td>
</tr>
<tr>
<td>S4</td>
<td>His - Histidine</td>
</tr>
<tr>
<td>S5</td>
<td>Excluded ligands</td>
</tr>
</tbody>
</table>

**Fragments that have halogen bonds with amino acid side chains in protein binding sites**

The links show a list of ligand fragments that have halogen bond with protein side chains.

- Fluorine
- Chlorine
- Bromine
- Iodine
Bond length

- Typical HBond values between 2 heteroatoms
  - 2.5 – 3.5 Å (15 - 20 kcal/mol)
- For F: bond to N is shorter than to O, stronger halogen bond.
- Also 2 times more to N than to O
- For Cl, bond to N and O are very similar, larger value than F, exclusively like a Hbond donor
- -F is a Hbond acceptor.
Protein families

- Our data set contains structures from:
  - all enzyme classes
  - various receptor families.
- 3 protein families dominate:
  - Oxidoreductase
  - Hydrolase
  - Transferase

The domination of enzymatic classes over receptor families reflect the nature that almost in all the cases, halogen bonding interaction could be useful in enzymatic active sites. For example, replacing Hydrogen bonds.
Chemistry consideration

• Availability of starting material
  – Quick reactant searches for –halogen + free amine
  – -F, ~6000 hits; -Cl ~4000 hits; -Br ~300 and -I ~100

• Reactivity of halogen:
  – Cross reactivity, need to have a halogen incorporated in the fragment
  – Final product with halogen cannot be reactive, therefore, reflecting the high frequency of phenyl or aromatic halogens in the final “ligands” as well as less frequent Br or I in any final molecules.

• Fluorine in final product
  – less reactive than Cl
  – Direct replacement of a labile hydrogen atom
  – Better PK and metabolism, etc

Conclusion

• More data are found from –F and –Cl; favourite chemistry choice
• -F has the most diversity of fragments;
  – most frequently used; exclusively a HA
• -Cl is represented by many fragments;
• -Br or –I has low representation
Acknowledgments

- Thanks go to Dr Roman Laskowski (EBI) for allowing us to use the backend PDBsum data

- Thanks to Dr David Steadman (UCL) for insightful chemistry discussion.