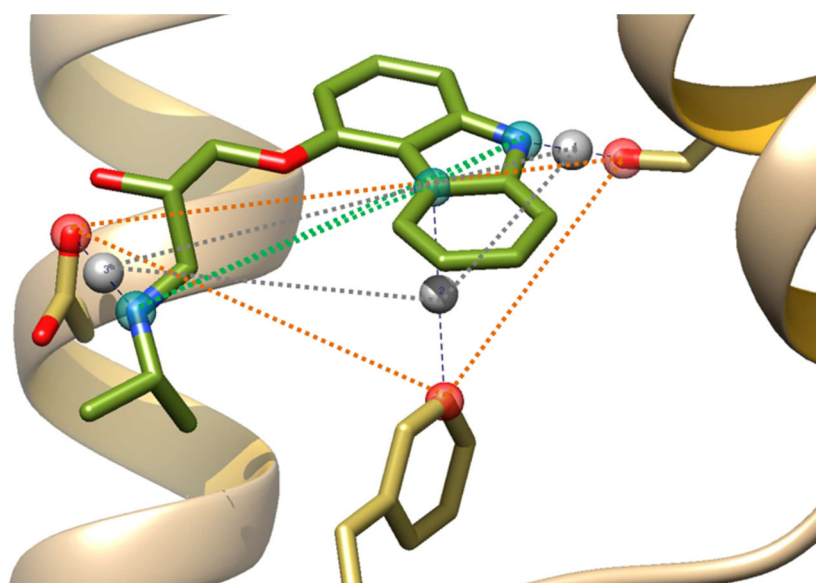# IChem: A Toolkit for detecting, comparing and predicting protein-ligand interactions

**Jérémy DESAPHY, Franck DA SILVA, Guillaume BRET and Didier ROGNAN**

*Laboratoire d'Innovation Thérapeutique, UMR 7200 CNRS-Université de Strasbourg, F-67400 ILLKIRCH*

Email: rognan@unistra.fr

# Literature Corner

Please have a look at these articles for detailed information on the basic principles and concepts underlying IChem usage.

| | |
|---|---|
| Marcou, G. and Rognan, D. (2007) Optimizing fragment and scaffold docking by use of molecular interaction fingerprints.*J. Chem. Inf. Model., * **47**, 195-207 | |
| Desaphy, J., Azdimousa, K., and Rognan, D. (2012) Comparison and druggability prediction of protein-ligand binding pockets from pharmacophore-annotated shapes., *J. Chem. Inf. Model.,* **52**, 2287-2299 | |
| Desaphy, J., Ducrot, P., Raimbaud, E. and Rognan, D. (2013) Encoding protein-ligand interaction patterns in fingerprints and graphs. *J. Chem. Inf. Model*, **53**, 623-637 | |
| Desaphy, J. and Rognan, D. (2014) scPDBFrag: a database of protein-ligand interaction patterns for bioisosteric replacements. *J.Chem. Inf. Model*., **54**, 1908-1918 | |
| Gabel, J., Desaphy, J. and Rognan, D. (2014) Beware of machine learning-based scoring functions - On the danger of developing black boxes. *J. Chem. Inf. Model*, **54**, 2807–2815 | |
| Da Silva, F., Desaphy, J., Bret, G. and Rognan, D. (2015) IChemPIC: A Random Forest Classifier of Biological and Crystallographic Protein-Protein Interfaces. *J. Chem. Inf. Model*, **55**, 2005–2014 | |
| Slynko, I., Da Silva, F., Bret, G. and Rognan, D. (2016) Docking pose selection by interaction pattern graph similarity: application to the D3R grand challenge 2015. *J. Comput.-Aided Mol. Des.,* **30**, 669-683. | |
| da Silva Figueiredo Celestino Gomes, P., Da Silva, F., Bret, G. and Rognan, D. (2018) Ranking docking poses by graph matching of protein-ligand interactions: lessons learned from the D3R Grand Challenge 2. *J. Comput.-Aided Mol. Des.*, **32**, 75-87. | |
| Da Silva, F., Desaphy, J. and Rognan D. (2018) IChem: A Versatile Toolkit for Detecting, Comparing, and Predicting Protein-Ligand Interactions. *ChemMedChem*, **13**, 507-510. | |
| Tran-Nguyen, V.K., Da Silva, F., Bret, G. and Rognan, D. (2019) All in One: Cavity Detection, Druggability Estimate, Cavity-Based Pharmacophore Perception and Virtual Screening. *J. Chem. Inf. Model.,* 59, 1, 573-585 | |
| Koensgen, F., Da Silva, F., Rognan, D. and Kellenberger, E. (2019) Unsupervised Classification of G-Protein Coupled Receptors and Their Conformational States Using IChem Intramolecular Interaction Patterns, J. Chem. Inf. Model., 59, 3611-3618 | |
| Da Silva, F., Bret, G., Teixeira, L., Gonzalez, C.F. and Rognan, D. (2019) Exhaustive repertoire of druggable cavities at protein-protein interfaces of known three-dimensional structure. J. Med. Chem, 2019, 62, 9732-9742 | |
| Tran-Nguyen, V.-K., Jacquemard, C. and Rognan D. (2020) LIT-PCBA: An unbiased dataset for machine learning and virtual screening. J. Chem. Inf. Model, 60, 4263-4273 | |
| Tran-Nguyen, V.-K., Bret, G. and Rognan, D. (2021) Accuracy of fast scoring functions to predict high-throughput screening data from docking poses: The simpler the better. J. Chem. Inf. Model., 67, 2788-2797 | |
| Volkov, M., Turk, A.J., Drizard, N., Martin, M., Hoffmann, B. Gaston-Mathé, Y. and Rognan, D. (2022) On the frustration to predict binding affinities from protein-ligand structures with deep neural networks. J. Med. Chem., 65, 7946–7958 | |

# What's new ?

IChem v 5.2.9 offers the possibility to automatically generate cavity-based pharmacophores from the sole 3D structure of a target of interest. This new functionality is embedded in the Volsite module and enables the user to define structure-based pharmacophores in three possible formats:

.chm for use in Biovia DiscoveryStudio or PipelinePilot

.pml for use in LigandScout

.mol2 for use in Shaper2 (in-house ligand alignment tool)

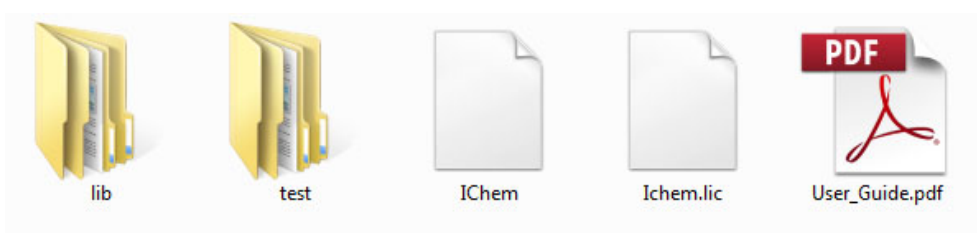For more details about the procedure, please have a look at:

Tran-Nguyen VK, Da Silva F, Bret G, Rognan D.
J Chem Inf Model. 2019, 59, 1, 573-585
All in One: Cavity Detection, Druggability Estimate, Cavity-Based Pharmacophore Perception and Virtual Screening.

# Installation

IChem is provided as a zipped archive file **(IChem.tgz)** containing the following material:



- **lib:** a directory containing necessary template files
- **test:** a directory containing some test input/output files
- **Ichem.lic** a license file
- **IChem:** A 64-bit Linux executable (CentOS 7.1)
- **User_Guide.pdf:** this manual

Source some environment variables and untar the IChem distribution in the IChem root directory

For csh users:   *setenv ICHEM_DIR /your_IChem_root_directory*
                 *setenv ICHEM_LIC  $ICHEM_DIR/IChem.lic*
                 *setenv ICHEM_LIB $ICHEM_DIR/lib*
                 *cp IChem.tgz $ICHEM_DIR*
                 *cd $ICHEMDIR; tar –cvfz IChem.tgz*

for bash users:  *ICHEM_DIR="/your_IChem_root_directory"; export ICHEM_DIR*
                 *ICHEM_LIC= ${ICHEM_DIR}/IChem.lic; export ICHEM_LIC*
                 *ICHEM_LIB= ${ICHEM_DIR}/lib; export ICHEM_LIB*
                 *cp IChem.tgz ${ICHEM_DIR}*
                 *cd ${ICHEM_DIR}; tar –cvfz IChem.tgz*

**Note to users:**
IChem commands with options/tool/arguments will be displayed in *italic* characters after the *">"* prompt
Input/output filenames will be displayed in **bold violet** characters
Terminal output will be displayed with a gray background

A copy of all test input/output files is given in the test directory of the IChem distribution. Before using IChem, please take the time to read the description of the technology. Articles to read will be mentioned by the following icon:

# A short introduction to IChem

IChem is a multi-task program for detecting, analyzing and comparing protein-ligand interactions. It is composed of several tools:

| Tool | Purpose |
| --- | --- |
| | |
| | **Binding mode tools** |
| **IFP** | Detect and save interactions as fingerprints (binding-site dependent) |
| **ints** | Detect and save interaction patterns( binding-site independent) |
| **grim** | Detect and save interactions as graphs |
| | |
| | **Generic tools** |
| **genkey** | License file generator (privileged usage) |
| **license** | Output license details |
| **pdbconv** | Convert PDB into MOL2 files |
| **realign** | Rotation/translation of atomic coordinates |
| **sims** | Fingerprint comparison |
| **utils** | Miscellaneous (buried surface area, ligand fragmentation) |
| | |
| | **Cavity detection** |
| **Volsite** | Cavity detection and druggability/ligandability prediction |
| | |
| | **Protein-Protein interfaces** |
| **DetectPPI** | Cavity detection and druggability/ligandability prediction |

# IChem Usage

The correct syntax for using IChem is:

> **> IChem** *options* **tool** *arguments*

Alternatively, options, tool and arguments can be stored in a parameter file (any name) that can be called from IChem with the –F option:

> **> IChem –F parameter_file**

Example of a parameter_file content
*options_1* *tool_1* *arguments_1*
*options_2* *tool_2* *arguments_2*
*…*
*options_n* *tool_n* *arguments_n*

Multiple IChem commands can be run from a single parameter file, just by adding for every line a novel list of options/tool/arguments

## Verbose and debugging mode:

IChem has been written in order to simplify standard output. I you want more details that that saved in the regular output, please use the  ***--verb*** option:

> **> IChem --verb** *options* **tool** *arguments*

Moreover, a debugging mode is available in IChem by calling the  ***--debug*** option with one or more of the following values, separated by a comma:

|  |  |
|---|---|
| **READ** | MOL2 file parser |
| **DINT** | Binding mode detection |
| **GRID** | Three-dimensional grid |
| **MOLD** | Molecular analysis |

> **> IChem –debug READ,DINT,MOLD** *options* **tool** *arguments*

# IChem menu

Just typing **$ICHEM_DIR/IChem** (or just **IChem** if $iCHEM_DIR is sourced in your path) gives you access to the full IChem menu

> ***> IChem***

## KEY Generator

**genKey** *year month day allowed_tools*

| | |
|---|---|
| year : | Expiring year  (4 digits : 2014) |
| month: | Expiring month (2 digits : 09) |
| day  : | Expiring day   (2 digits : 23) |
| allowed_tools: | each value is separated by space and must have a value of either 1 or 0. Must follow this order |
| 1st: | Licence Key generator |
| 2nd: | Molecular realignment |
| 3rd: | BSA Calculation |
| 4th: | IFP generator |
| 5th: | Interaction detection |
| 6th: | Graph Interaction Matching |
| 7th: | VolSite |
| 8th: | PDB to MOL2 conversion |
| 9th: | Patching MOL2 conversion |
| 10th: | Utils |
| 11th: | Sims |
| 12th: | Scoring |
| 13th: | Detection and analysis of PPi |

Example :
```
genKey 2014 1 28 0 0 1 0 0 0 0 0 1 0 1 1 1
        |<-DATE->|<--------TOOLS---------->
```

## realign - Molecular alignment
*realign rigidM mobilM applied1 applied2*


||-> rigidM : reference molecule to apply alignment to
||-> mobilM : comparison molecule to apply alignment from
||-> applied: molecule to apply rotation/translation to


[General options]
   -gmatch N (NAME) Use graph matching to align
     NAME     Atom Name matching
     ATMN     Atomic Name matching
     MOL2     MOL2 Type matching
     CALP     CAlpha Atom matching (protein only)
   --wMob     Also outputs the aligned mobilM
   -rule  R

By default, the program will perform an atom by atom match, without taking care of what kind of atom it match. If you want to perform a match by regarding only some atoms, this index_string is here to do so
    ex : -i '2-3|1-6|23-160'     Will match the second atom from the reference with the third from the comparison, the first with the sixth ...


## IFP - Interaction FingerPrint
  **IFP protein ligand**
  **IFP protein ligand ligand_ref**


[General options]
  -name  N (LIG)    Name of the fingerprint -Default: Name of the ligand
  --polar          Detect and output only polar interactions
  --metal          Detect and output only metal interactions
  --extended      Include within the fingerprint:
               |--> Metal/Acceptor interaction
               |--> Weak Hydrogen bonds
               |--> PI-Cation interactions
[testing options]
  -D_Hb  N (3.5)  Hbond length                        (Angstroem)
  -D_Hyd  N (4.5)  Hydrophobic length           (Angstroem)
  -D_Io  N (4.0)  Ionic length                      (Angstroem)
  -D_Me  N (2.8)  Metal/Acceptor length       (Angstroem)
  -D_Ar  N (4.0)  Aromatic interaction length    (Angstroem)
  -D_Pic  N (4.0)  Pi cation interaction length   (Angstroem)
  -a_H    N (Pi)   HBond angle                    (rad)
  -at_H   N (Pi/3) HBond tolerance angle         (rad)
  -a_ArFF  N (Pi)   Aromatic Face to Face interaction angle   (rad)
  -at_ArFF N (Pi/6) Aromatic Face to Face tolerance angle   (rad)
  -a_ArEF  N (Pi/2) Aromatic Edge to Face interaction angle (rad)
  -at_ArEF N (Pi/3) Aromatic Edge to Face tolerance angle   (rad)
  -a_Pic   N (Pi)   Pi cation interaction angle        (rad)
  -at_Pic  N (Pi/6) Pi cation tolerance angle        (rad)
  --ligD        Print all possible ligand interactions


Please note that ligand file can be a multi-mol2 file

## INTERACTION GENERATOR

*ints prot lig out*

  -type (CENT)      Alter positionning output, multiple values are allowed, separated by space

    PROT            InterPROT positionning

    LIG              InterLIG positionning

    CENT            Centered positionning

    MERG            Merged all 3 above

  -fgps              Fingerprint format

    STD              Standard   (1 0 21 0 0 3)

    SVM              SVM format (1:1 3:21 6:3)

    CMP              Compressed (1 [1 21 [2 3)

  --small           Compressed fingerprint

[General options]

  -name      (prot) Name of molecule in out file

  -logf          Name of log file

| | | | |
|---|---|---|---|
| -D_Hb | N (3.5) | Hbond length | (Angstroem) |
| -D_Hyd | N (4.5) | Hydrophobic length | (Angstroem) |
| -D_Io | N (4.0) | Ionic length | (Angstroem) |
| -D_Me | N (2.8) | Metal/Acceptor length | (Angstroem) |
| -D_Ar | N (4.0) | Aromatic interaction length | (Angstroem) |
| -D_Pic | N (4.0) | Pi cation interaction length | (rad) |
| -a_H | N (Pi) | HBond angle | (rad) |
| -at_H | N (Pi/3) | HBond tolerance angle | (rad) |
| -a_ArFF | N (Pi) | Aromatic Face to Face interaction angle | (rad) |
| -at_ArFF | N (Pi/6) | Aromatic Face to Face tolerance angle | (rad) |
| -a_ArEF | N (Pi/2) | Aromatic Edge to Face interaction angle | (rad) |
| -at_ArEF | N (Pi/3) | Aromatic Edge to Face tolerance angle | (rad) |
| -a_Pic | N (Pi) | Pi cation interaction angle | (rad) |
| -at_Pic | N (Pi/6) | Pi cation tolerance angle | (rad) |
| --noMerge | Don't merge hydrophobic interactions | | |
| --newH | Less permissive definition of hydrophobic interaction | | |

## GRIM - GRaph Interaction Matching :

*grim refProt refLig CompProt CompLig*                    *(1)*

*grim refInts compInts*                              *(2)*

*grim refProt refFile dockFile*                        *(3)*

[Note]

    (1)  use --multim2 to use multimol2

    (3)  refFile & dockFile can be multimol2 files

  [General options]

NOTE : INTERACTION GENERATION General Options also accessible

  -rn            N (Ref)   Reference name

  -cn            N (Comp)  Comparison name

  --values        Only output score and not alignment

  -sim            N (0)     Boolean telling whether the pair is similair or not

  -outInt         (MERG)  Output only one kind of interaction positionning

    MERG          All aligned interactions are outputed

    LIG            InterLIG positionning

    CENT          Centered positionning

    PROT          InterPROT positionning

NOTE : outInt useless when used with --values
-match          N (MERG)  Align only with a specific position
   MERG          Align with ALL interaction points
   LIG           Align only with ligand interaction points
   PROT          Align only with protein interaction points
   CENT          Align only with centered interaction points
-score          N (FCT)   Scoring method function
   STD           Scored by decreasing SumCl and increasing RMSD
   FCT           Scored with scoring function
--newH          Less permissive definition of hydrophobe interaction
[Alignment options]
-max    N (1)    Maximal number of outputed cliques.
-size   N (3)    Minimal size of a clique.
--all_cliques    Detect all cliques and not only maximal one
-dsame  N (0.5)   Maximal difference distance between to similar point(LIG-LIG, ...)
-dclose N (0.75)  Maximal difference distance between to close point (LIG-CENT or PROT-CEN
-dfar   N (1)    Maximal difference distance between to far point (LIG-CENT)

## VOLSITE - Cavity detection in a mol2 file
*volsite prot lig*                         *(1)*
*volsite prot*                             *(2)*

[General options]
-step  N (1.5)    Edge length of each box              (Angstroem)
-boxS  N (20)     Edge length of the main box          (Angstroem)
-b     N (55)     Minimal threshold for buriedness
-name  N          PDB Name for output cavity names
-n     N  (5)     Minimal neighbours for buried cavity boxes
-nPTS  N (70)     Minimal number of cubes to consider it a cavity
-NPTS  N (400)    Maximal number of cubes to consider it a cavity
--dna             Consider DNA as part of the protein
--cofactor        Consider cofactor as part of the protein
--solvent         Consider solvent as part of the protein
--hydrogen        Consider hydrogens
--desc            Write a descriptor file name descriptor.txt
--svm             Build a svm property file
-drog  N          Observed druggability
--pharm           Generate a pharmacophore (.chm) from cavity
--outExclu        Output exclusion sphere in pharmacophore file

## PDB Process
*pdbconv protein[.pdb|.mol2] output_dir pdb_id*

--wMOL2          Use MOL2 File as Input. PDB Options are not available
--wUnDrug        Output undruggable cavities
--noLig          PDB with no Ligand

By default all the following options are included. All chains will be kept
[PDB Options]
--HARMSIZE       Harmonize size line to 80 characters
--MSEMET         Change MSE to MET
--CSECYS         Change CSE to CYS
--MOVHET         move HETATM to the end of file
--ALTATM         select alternative atoms
--NUMATM         renumerotate atoms
--UPDMAS         update the MASTER line

--TOMOL2            convert to a molecular representation (instead of flat file)

if you use one of the option below, you MUST use also --TOMOL2 option or use --wMOL2 option

[MOL2 Options]
--RESTYP            apply Residue Class (cofactor/STD_AA/MOD_AA/Ligand ...)
--BONDSE            create bonds
--CLNUNW            clean unwanted residues
--MOL2TY            apply MOL2 types according to templates
--SPLITM            split molecule into protein/ligand/solvent
-SelChain N         List of chains to keep, separated by underscore
--SELWAT            select water molecules
--SELLIG            select ligand

## Buried Surface Area calculation
*utils bsa protein ligand*

## Ligand Fragmentation
*utils frag protein ligand*

## Fingerprint Similarity
*sims ref comp*                          *(1)*
*sims file*                              *(2)*
*sims RefInt CompInt*                    *(3)*

[General options]
--wInts             To use interactions instead of fingerprints (option 3 only)
--small             Use small fingerprint (option 3 only)
--binary            To add when the fingerprint is binary
-metric N (TC)      Select the metric
  TC                Tanimoto metric
  HM                Hamming distance
  RT                Ref Tversky
  FT                Fit Tversky
  DI                Dice
  SO                Soergel

## DetectPPI - Interface detection & characterization
*detectppi PDB_name name_file*                          *(1)*

*[General options]*
  *-c  C   : The name of all chains that might be in interaction*
  *-n (1)  : Maximum number of treated interfaces (ranked by decreasing size)*

*May have some bugs on structures with more than two chains*

# IChem Tutorial

## 1. License check and generation

The *IChem license* command enables you to see details of your license

> *IChem license*

The terminal output should look like this …

```
            Licence key generator : YES
            Molecular realignment : YES
               BSA Calculation : YES
                 IFP generator : YES
             Interaction detection : YES
     Graph Interaction Matching : YES
                      VolSite : YES
          PDB to MOL2 conversion : YES
       Patching MOL2 conversion : YES
                        Utils : YES
                         Sims : YES
                      Scoring : YES
         Detection, analysis of PPi : YES
Your licence expires on (Y/M/D): 2019/12/31
```

For each module, a YES/NO flag indicate whether you have rights to use the corresponding utility. The last line gives the expiration date of your current license.

If your license scheme allows you to generate license keys, the ***IChem*** *genkey* command permits you to define any possible license file

***genKey year month day allowed_tools***

  year : Expiring year  (4 digits : 2014)
  month: Expiring month (2 digits : 09)
  day  : Expiring day   (2 digits : 23)

 allowed_tools: each value is separated by space and must have a value of either 1 or 0. Must follow this order
  First  : Licence Key generator
  Second : Realigning molecules
  Third  : BSA Calculation
  Fourth : IFP generator
  Fifth  : Interaction detection
  Sixth  : Graph Interaction Matching
  Seventh: VolSite
  Eighth : PDB to MOL2 conversion
  Ninth  : Patching MOL2 conversion
  Tenth  : Utils
  Eleventh: sims
  Twelfth: scoring
  Thirteenth: Detection, analysis of PPi

---

**> *IChem genKey 2016 1 28 0 0 1 0 0 0 0 0 1 0 1 1***

---

grants you access until Jan.28th 2016 to 5 modules: BSA calculation,  Patching MOL2 conversion, sims and scoring.

## 2. Protein-Ligand Interaction Fingerprints (IFPs)

The *IChem IFP* command registers in a bit string the interactions between a protein (active site) and a ligand.

**IFP protein ligand**
**IFP protein ligand ligand_ref**

[General options]
   -name  N (LIG)    Name of the fingerprint -Default: Name of the ligand
   --polar            Detect and output only polar interactions
   --metal           Detect and output only metal interactions
   --extended       Include within the fingerprint:
                 |--> Metal/Acceptor interaction
                 |--> Weak Hydrogen bonds
                 |--> PI-Cation interactions
[testing options]
  -D_Hb   N (3.5)  Hbond length                       (Angstroem)
  -D_Hyd  N (4.5)  Hydrophobic length          (Angstroem)
  -D_Io    N (4.0)  Ionic length                     (Angstroem)
  -D_Me   N (2.8)  Metal/Acceptor length      (Angstroem)
  -D_Ar   N (4.0)  Aromatic interaction length   (Angstroem)
  -D_Pic  N (4.0)  Pi cation interaction length    (rad)
  -a_H    N (Pi)   HBond angle                     (rad)
  -at_H   N (Pi/3) HBond tolerance angle       (rad)
  -a_ArFF N (Pi)   Aromatic Face to Face interaction angle   (rad)
  -at_ArFF N (Pi/6) Aromatic Face to Face tolerance angle   (rad)
  -a_ArEF  N (Pi/2) Aromatic Edge to Face interaction angle (rad)
  -at_ArEF N (Pi/3) Aromatic Edge to Face tolerance angle   (rad)
  -a_Pic   N (Pi)   Pi cation interaction angle       (rad)
  -at_Pic  N (Pi/6) Pi cation tolerance angle        (rad)
  --ligD        Print all possible ligand interactions

Please note that ligand file can be a multi-mol2 file

For more information, see:

Marcou G, Rognan D.
J Chem Inf Model. 2007 Jan-Feb;47(1):195-207.
Optimizing fragment and scaffold docking by use of molecular interaction fingerprints.

In a first mode, interactions between an active site (mol2 file format) and one/several ligands (single or multimol2 file) are outputted in a table along with the IFP bit string. In a second mode, similarity of the IFP(s) to that of one (several) references is outputted in addition.

**-1st mode: computes an IFP between a protein active site and a ligand**

**Input directory:** $ICHEM_DIR/test/IFP
**Input files:** site.mol2, ligand.mol2
**Output file:** ligand.ifp

### > IChem IFP site.mol2 ligand.mol2 >ligand.ifp

The following information is contained in the output file (e.g. ligand.ifp) or at the terminal if no file redirection *(… >…)* is given:

```
HBond_LIG      |OD1   |78   |ASP 113-A  |O17  |1   |CAU   0-XX |0        |2.60795      |134.667
HBond_PROT     |ND2   |421  |ASN 312-A  |O17  |1   |CAU   0-XX |1        |2.76961      |159.236
Hydrophobic    |CZ    |349  |PHE 289-A  |C16  |2   |CAU   0-XX |3        |3.74597      | /
Ionic_LIG      |OD2   |79   |ASP 113-A  |N19  |4   |CAU   0-XX |4        |2.94042      | /
Ionic_LIG      |OD1   |78   |ASP 113-A  |N19  |4   |CAU   0-XX |5        |3.60381      | /
Hydrophobic    |CZ3   |46   |TRP 109-A  |C21  |6   |CAU   0-XX |6        |4.43435      | /
Hydrophobic    |CG2   |64   |THR 110-A  |C21  |6   |CAU   0-XX |7        |4.42961      | /
Hydrophobic    |CB    |163  |PHE 193-A  |C21  |6   |CAU   0-XX |8        |4.07199      | /
Hydrophobic    |CH2   |47   |TRP 109-A  |C22  |7   |CAU   0-XX |9        |3.72321      | /
Hydrophobic    |CZ    |438  |TYR 316-A  |C22  |7   |CAU   0-XX |10       |4.46257      | /
Hydrophobic    |CG2   |125  |VAL 117-A  |C15  |8   |CAU   0-XX |11       | 4.0788      | /
Hydrophobic    |CE2   |348  |PHE 289-A  |C15  |8   |CAU   0-XX |13       | 4.0318      | /
HBond_LIG      |OG    |248  |SER 203-A  |N7   |16  |CAU   0-XX |23       |3.31671      |127.808
Hydrophobic    |CG2   |185  |THR 195-A  |C1   |20  |CAU   0-XX |24       |4.32061      | /
Hydrophobic    |CZ3   |327  |TRP 286-A  |C16  |2   |CAU   0-XX |25       |4.05558      | /
Hydrophobic    |CG1   |124  |VAL 117-A  |C12  |11  |CAU   0-XX |28       |3.98025      | /
Hydrophobic    |CG1   |89   |VAL 114-A  |C11  |12  |CAU   0-XX |29       |3.83646      | /
Hydrophobic    |CB    |288  |SER 207-A  |C10  |13  |CAU   0-XX |30       |3.68825      | /
Hydrophobic    ll
|CG2   |90   |VAL 114-A  |C13  |10   |CAU   0-XX |31       |3.81147          | /
```

```
|A   M82|A   V86|A  W109|A  T110|A  D113|A  V114|A  L115|A  V117|A  T118|A  C191|A  F193|A  T195
|A  Y199|A  A200|A  I201|A  S203|A  S204|A  I205|A  S207|A  F208|A  W286|A  F289|A  F290|A  N293
|A  Y308|A  N312|A  Y316
000000000000000100000010000000000010110000000000000010000000000000000000001000000100000 0
000000000000000000000000010000000000000010000000000000100000010000000000000000000
0000000000100010000000
```

In the first section, all intermomecular interactions between active site and ligand are tabulated
1st column: type of interaction
2nd column: interacting protein atom name
3rd column: interacting protein atom number
4th column: interacting protein residue name & number
5th column: interacting ligand atom name
6th column: interacting ligand atom number
7th column: interacting ligand residue name & number
8th column: interaction distance
9th column: interaction angle (for H-bonds only)

In the second section, the interaction fingerprint is displayed with 7 bits/residue in normal mode.
The first line describes the chain number and residue number
The second line is the bit string (1: interaction, 0: no interaction). The order is the following:
1st position: hydrophobic
2nd position: aromatic (face-to-face)
3rd position: aromatic (edge-to-face)
4th position: h-bond (protein donor)
5th position: h-bond (ligand donor)
6th position: ionic (protein charged +)
7th position: ionic (ligand charged +)

**-2nd mode: compute the IFP similarity (TC coefficient) between docked ligand and reference**

**Input directory: $ICHEM_DIR/test/IFP**
**Input files: site.mol2, docked.mol2, ligand.mol2**
**Output file: docked_Tc.ifp**

---

**> IChem IFP site.mol2 docked.mol2 ligand.mol2 >docked_Tc.ifp**

---

The following information is contained in the output file (docked_Tc.ifp) or at the terminal if no file redirection *(… >…)* is given:

```
REF    |A  M82|A  V86|A W109|A T110|A D113|A V114|A L115|A V117|A T118|A C191|A F193|A T195|A
Y199|A A200|A I201|A S203|A S204|A I205|A S207|A F208|A W286|A F289|A F290|A N293|A Y308|A
N312|A Y316
REF
0000000000000001000000100000000001011000000000000010000000000000000000010000001000000000000000000
0000000000000001000000000000000001000000000000001000001000000000000000000000000000010001001000000
0
DOCKED |A  M82|A  V86|A W109|A T110|A D113|A V114|A L115|A V117|A T118|A C191|A F193|A T195|A
Y199|A A200|A I201|A S203|A S204|A I205|A S207|A F208|A W286|A F289|A F290|A N293|A Y308|A
N312|A Y316
DOCKED
0000000000000001000000100000010000110000000000000100000000000000000000010000001000000000000000000
0000000000000000000000000000001000000000000010000001000000000001000001000000000010001001000000
0
REF    REF    1
REF    DOCKED 0.705882
```
In the first section, IFPs are just given for the reference (REF) and the docked ligand (DOCKED)
In the second section, the Tc coefficients are given for both the REF and the DOCKED ligand

**-Non standard modes**

Polar mode: only 5 bits/residues are outputted (positions 4-7 of the standard IFP + metal coordination)

---

**> IChem *--polar* IFP *arguments***

---

Extended mode: 2 additional bits/residues with respect to the normal mode (position 8: pi-cation interactions; position 9: metal coordination)

---

**> IChem *--extended* IFP *arguments***

---

**-Interaction detection rules**

| | | | |
|---|---|---|---|
| -D_Hb | N (3.5) | Hbond length | (Angstroem) |
| -D_Hyd | N (4.5) | Hydrophobic length | (Angstroem) |
| -D_Io | N (4.0) | Ionic length | (Angstroem) |
| -D_Me | N (2.8) | Metal/Acceptor length | (Angstroem) |
| -D_Ar | N (4.0) | Aromatic interaction length | (Angstroem) |
| -D_Pic | N (4.0) | Pi cation interaction length | (Angstroem) |
| -a_H | N (Pi) | HBond angle | (rad) |
| -at_H | N (Pi/3) | HBond tolerance angle | (rad) |
| -a_ArFF | N (Pi) | Aromatic Face to Face interaction angle | (rad) |
| -at_ArFF | N (Pi/6) | Aromatic Face to Face tolerance angle | (rad) |

-a_ArEF         N (Pi/2) Aromatic Edge to Face interaction angle      (rad)
-at_ArEF         N (Pi/3) Aromatic Edge to Face tolerance angle      (rad)
-a_Pic         N (Pi)   Pi cation interaction angle            (rad)
-at_Pic         N (Pi/6) Pi cation tolerance angle            (rad)
--ligD         Print all possible ligand interactions

Interactions are detected based on default topological rules that can be modified whenever necessary:

> ***IChem** –D_Hb DHB –D_Hyd DHYD –D_Io DIO –D_Me DME –D_Ar DAR –D_PIC DIC –a_H AH –at_H ATH –a ArFF AARFF –at_ArFF ATARFF –a_ArEF –AAREF –at_AREF ATAREF –a_Pic APIC –at_PIC ATPIC **IFP** arguments*

*Default values are described in the following table:*

| Parameter | Description | Default Value |
|---|---|---|
| DHB | Donor-acceptor distance  threshold for H-bond interaction | 3.5 Å |
| DHYD | Donor-acceptor distance  threshold for hydrophobic  interaction | 4.5 Å |
| DIO | Anion-cation distance  threshold for ionic interaction | 4.0 Å |
| DME | Metal-acceptor distance  threshold for metal chelation | 2.8 Å |
| DAR | Distance  threshold between two aromatic ring centers | 4.0 Å |
| DPIC | Distance  threshold for pi cation interaction | 4.0 Å |
| AH | Angle threshold between donor-H-acceptor for H-bond | $\pi$ |
| ATH | Tolerance angle for H-bond | $\pi/3$ |
| AARFF | Angle threshold between the plane of two aromatic cyles for face to face aromatic interactions | $\pi$ |
| ATARFF | Tolerance angle for face to face aromatic interactions | $\pi/6$ |
| AAREF | Angle threshold between the plane of two aromatic cyles for edge to face aromatic interactions | $\pi/2$ |
| ATAREF | Tolerance angle for edge to face aromatic interactions | $\pi/6$ |
| APIC | Angle threshold for pic cation interactions | $\pi$ |
| ATPIC | Tolerance angle for pi cation interactions | $\pi/6$ |
| nomerge | After detecting all hydrophobic contacts, a filtering process is engaged to avoid too many hydrophobic pseudoatoms. Using this flag avoids the filtering process | no |

The interaction rules options can be used for the following IChem tools: **IFP**, **ints**, **grim**

**Some notes:**
- Beware: the mol2 file should comply with the standard mol2 file format from TRIPOS (http://www.tripos.com/data/support/mol2.pdf)
- If you do not have SYBYL, you can generate correct mol2 files with UCSF chimera (http://www.cgl.ucsf.edu/chimera)
- IFPs are active-site dependent (nbits/residue) and cannot be compared across active sites of different compositions (e.g. different number of residues).
- Please rather use the active site and not the protein as input, to restrict the size of the IFP and prevent generating bit strings with mostly "0" values.
- Compare your output with that given in $ICHEM_DIR/test/IFP/output

### 3. Interaction Fingerprint Triplet (TIFPs)

TIFPs are coordinate-frame invariant interaction fingerprints presenting the advantage to compare ligand binding to completely different active sites, whatever their composition.
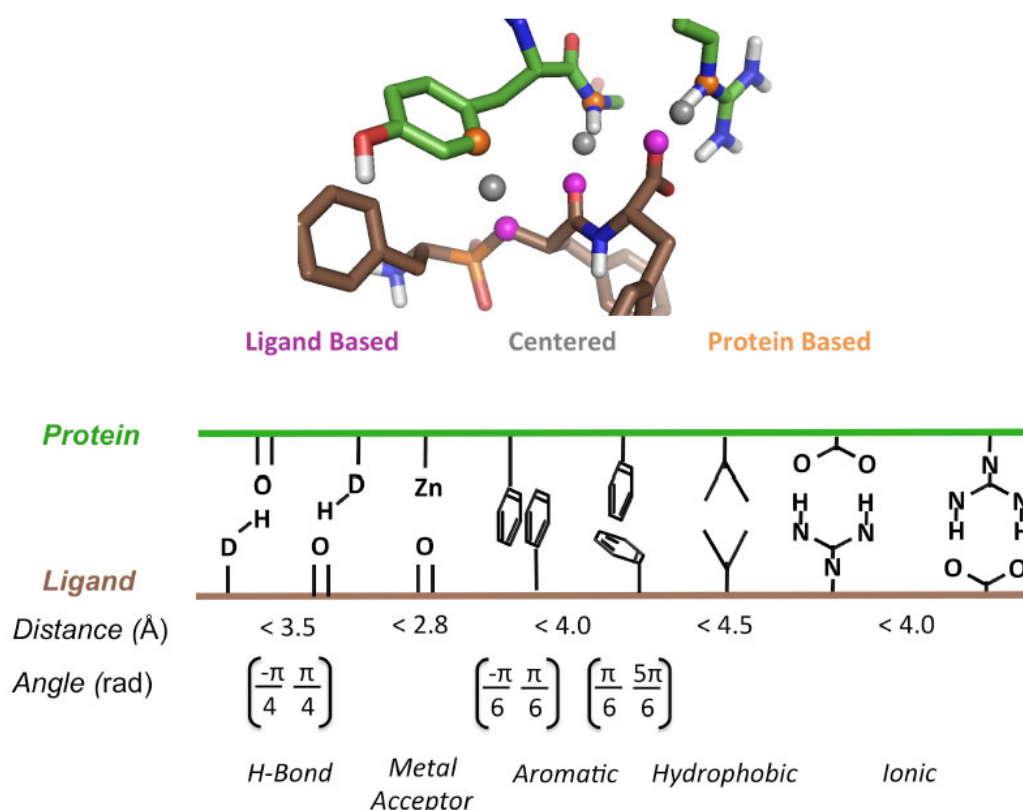
For more information, see:

Desaphy J, Raimbaud E, Ducrot P, Rognan D.
J Chem Inf Model. 2013 Mar 25;53(3):623-37
Encoding protein-ligand interaction patterns in fingerprints and graphs.

Interactions are encoded by pseudoatoms in a MOL2 file format with the following properties: interaction type, atomic coordinates (**Figure 1**).



*Figure 1. Encoding protein-ligand interaction by pseudoatoms. Each interaction is detected on the fly from standard topological rules and is represented by a pseudoatom with 3 possible atomic coordinates: the interacting ligand atom (LIG mode, violet balls), the interacting protein atom (PROT mode, orange balls), the barycenter of interacting protein and ligand atoms (CENT mode), the three aboved-described atoms (MERG mode).*

```
ints prot lig out
   ints prot lig outfile
      -type      (CENT) Alter positionning output
                        Multiple values allowed, separated by space
         PROT            InterPROT positionning
         LIG             InterLIG positionning
         CENT            Centered positionning
         MERG            Merged all 3 above

      -fgps     (STD)  Fingerprint format
         STD             Standard   (1 0 21 0 0 3)
         SVM             SVM format (1:1 3:21 6:3)
         CMP             Compressed (1 [1 21 [2 3)
      --small           Compressed fingerprint

   [General options]
      -name      (prot) Name of molecule in out file
      -logf            Name of log file
      -D_Hb    N (3.5)  Hbond length                      (Angstroem)
      -D_Hyd   N (4.5)  Hydrophobic length                (Angstroem)
      -D_Io    N (4.0)  Ionic length                      (Angstroem)
      -D_Me    N (2.8)  Metal/Acceptor length             (Angstroem)
      -D_Ar    N (4.0)  Aromatic interaction length       (Angstroem)
      -D_Pic   N (4.0)  Pi cation interaction length      (Angstroem)
      -a_H     N (Pi)   HBond angle                       (rad)
      -at_H    N (Pi/3) HBond tolerance angle             (rad)
      -a_ArFF  N (Pi)   Aromatic Face to Face interaction angle (rad)
      -at_ArFF N (Pi/6) Aromatic Face to Face tolerance angle   (rad)
      -a_ArEF  N (Pi/2) Aromatic Edge to Face interaction angle (rad)
      -at_ArEF N (Pi/3) Aromatic Edge to Face tolerance angle   (rad)
      -a_Pic   N (Pi)   Pi cation interaction angle        (rad)
      -at_Pic  N (Pi/6) Pi cation tolerance angle          (rad)
      --noMerge         Don't merge hydrophobic interactions
      --newH            Less permissive definition of hydrophobic interaction
```

**-Listing interactions and outputting interaction pseudoatoms**

**Input directory: $ICHEM_DIR/test/TIFP**
**Input files: site.mol2, ligand.mol2**
**Output files: 2rh1_ints.txt, 2rh1_site_INTS_C.mol2**

> **IChem** *–logf 2rh1_ints.txt –type CENT* **ints** *site.mol2  ligand.mol2*

The **2rh1_ints.txt** file contains the same information as the table outputted by the **IChem IFP** command (recall pages 13, 14)

The **2rh1_site_INTS_X.mol2** file describes the properties and atomic coordinates of the interaction pseudoatoms. Depending on the *–type* option (*CENT, PROT, LIG, MERG*), pseudoatoms are mapped onto ligand-interacting atoms (*-type LIG,*), protein-interacting atoms (*-type PROT*), barycenter of ligand and protein-interacting atoms (*-type CENT*, default option), or on all above-described atoms (-*type MERG*)

| Filename | -type mode |
|---|---|
| Header_INTS_C.mol2 | CENT mode |
| Header_INTS_L.mol2 | LIG mode |
| Header_INTS_P.mol2 | PROT mode |
| Header_INTS_M.mol2 | MERG mode |

**Atomic description of the interaction pseudoatoms:**

| Atom Name | Described interaction |
|-----------|----------------------|
| CA | hydrophobic |
| CZ | aromatic |
| O | Hydrogen-bond (interacting protein atom is acceptor) |
| OG | Hydrogen-bond (interacting protein atom is both acceptor and donor) |
| N | Hydrogen-bond (interacting protein atom is donor) |
| OD1 | Ionic (interacting protein atom is negatively charged) |
| NZ | Ionic (Interacting protein atom is positively charged) |
| ZN | Metal coordination |

**-Generating interaction fingerprints**

With respect to *IChem IFP* , *IChem ints* outputs generic and binding site-independent interaction fingerprints consisting in all possible combinations of triplets of interaction pseudoatoms

**Input directory: $ICHEM_DIR/test/TIFP**
**Input files: site.mol2, ligand.mol2**
**Output files: 2rh1.fgp**

**> IChem   -fgps STD ints site.mol2  ligand.mol2 2rh1_full.fgp**

The fgp output file can appended after every novel *IChem ints* command if the same outputfile name is given. By default, a standard fingerprint format is outputted, but the *–fgps* option enables you to control the fingerprint format:

| Format mode | *-fgps option* | Example | Description |
|-------------|----------------|---------|-------------|
| Standard | STD | 0 1 0 0 132 0 1 0 0 0 | Each value is separated by a space<br>Inefficient for long and sparse fingerprints (many null values) |
| SVM | SVM | 2:1: 5:132 7:1 | Only non-null values are outputted and defined by *POSITION:VALUE* separated by a space<br>Very efficient for long sparse fingerprints<br>Not efficient for long and dense fingerprints (few null values) |
| Compressed | CMP | [1 1 [2 132 [1 1 [3 | Non-null values are explicitly outputted. Null values are encoded by a "[" sign followed by the number of consecutive null values<br>Efficient for long and sparse fingerprints |

By default, the standard fingerprint is made of 12510 integers, most of which encoding triplets of pseudoatoms describing hydrophobic contacts. The full fingerprint can be pruned to remove rare triplets (based on the analysis of ca. 10 000 protein-ligand PDB Structures) and lead to a simpler fingerprint of 210 integers (*--small* option)

> ***IChem*  *--small –fgps STD* *ints* *site.mol2  ligand.mol2 2rh1_small.fgp***

### 4. Graph Matching (GRIM)

GRIM is a tool to match protein-ligand complexes using a graph matching algorithm focusing on protein-ligand interaction pseudoatoms. It thereby enables an interaction-based alignment of different protein-ligand complexes that can be quantified by an empirical scoring function (GrScore) and used to post-process docking poses by similarity to known protein-ligand interaction patterns.

In a recent international contest (D3R Docking Challenge 2015), **Grim was ranked 2nd out of 44** scoring functions to predict the binding mode of 36 inhibitors prior to the release of protein-bound X-ray coordinates (Gathiaka et al. J Comput.-Aided Mol. Des, 2016)

For more information, see:

Desaphy J, Raimbaud E, Ducrot P, Rognan D., J Chem Inf Model. 2013; 53: 623-637:
Encoding protein-ligand interaction patterns in fingerprints and graphs.

Slynko I, Da Silva F, Bret G, Rognan D. J Comput Aided Mol Des. 2016, 30: 669-683.
Docking pose selection by interaction pattern graph similarity: application to the D3R grand challenge 2015.

```
grim refProt refLig CompProt CompLig           (1)
grim refInts compInts                          (2)
grim refProt refFile dockFile                  (3)
```

```
[Note]
     (1)  use --multim2 to use multimol2 ligand files
     (3)  refFile & dockFile can be multimol2 files
[General options]
NOTE : INTERACTION GENERATION General Options also accessible
     -rn     N (Ref)   Reference name
     -cn     N (Comp)  Comparison name
     --values          Only output score and not alignment
     -sim    N (0)     Boolean telling whether the pair is similair or not
     -outInt   (MERG)  Output only one kind of interaction positionning
        MERG           All aligned interactions are outputed
        LIG            InterLIG positionning
        CENT           Centered positionning
        PROT           InterPROT positionning
      NOTE : outInt useless when used with --values
     -match  N (MERG)  Align only with a specific position
        MERG           Align with ALL interaction points
        LIG            Align only with ligand interaction points
        PROT           Align only with protein interaction points
        CENT           Align only with centered interaction points
     -score  N (FCT)   Scoring method function
        STD            Scored by decreasing SumCl and increasing RMSD
        FCT            Scored with scoring function
     --newH            Less permissive definition of hydrophobe interaction
[Alignment options]
     -max    N (1)     Maximal number of outputed cliques.
     -size   N (3)     Minimal size of a clique.
     --all_cliques     Detect all cliques and not only maximal one
     -dsame  N (0.5)   Maximal difference distance between to similar point(LIG-
LIG, ...)
     -dclose N (0.75)  Maximal difference distance between to close point (LIG-
CENT or PROT-CENT)
     -dfar   N (1)     Maximal difference distance between to far point (LIG-CENT)
-1st mode: Graph matching from structures
```

**Input directory: $ICHEM_DIR/test/GRIM**
**Input files: 2rh1_prot.mol2, 2rh1_lig.mol2, 4amj_prot.mol2, 4amj_lig.mol2**
**Output files: CompInts.mol2, Grifp_res.csv, GRIM_ints.mol2, GRIM_lig.mol2, GRIM.log, GRIM_prot.mol2, RefInts.mol2**

>**IChem** *-sim 1 -rn 2rh1 -cn 4amj* **grim** *2rh1_prot.mol2 2rh1_lig.mol2 4amj_prot.mol2 4amj_lig.mol2*

**-2ⁿᵈ mode: Graph matching from interaction pseudoatoms**
**Input directory: $ICHEM_DIR/test/GRIM**
**Input files: 2rh1_INTS_M.mol2, 4amj_INTS_M.mol2**
**Output files: CompInts.mol2, Grifp_res.csv, GRIM_ints.mol2, GRIM_lig.mol2, GRIM.log, GRIM_prot.mol2, RefInts.mol2**

> **IChem** *-sim 1 -rn 2rh1 -cn 4amj* **grim** *2rh1_INTS_M.mol2 4amj_INTS_M.mol2*

Whatever the mode, 7 files are outputted:
**CompInts.mol2:** interaction pseudoatoms (merged mode) of the complex to fit
**Grifp_res.csv:** summary of the GRIM alignment
**GRIM_ints.mol2:** aligned interaction pseudoatoms of the complex to fit
**GRIM_lig.mol2:** coordinates of the fitted ligand, aligned to the reference
**GRIM.log:** output of the GRIM alignment
**GRIM_prot.mol2:** coordinates of the fitted protein, aligned to the reference
**RefInts.mol2:** interaction pseudoatoms (merged mode) of the reference complex

The **Grifp_res.csv file** (appended at every novel IChem grim command) is a table that looks like this:

| NCli | Ref | Comp | Simil | LIG | CENTER | PROT | SumCl | RMSD | RLig | RCent | RProt | CLig | CCent | CProt | GrSc | RMSDAl | NPol |
|------|------|------|-------|-----|--------|------|-------|------|------|-------|-------|------|-------|-------|------|--------|------|
| 1 589 | 2rh1 10 | 4amj | 1 | 9 | 6 | 7 | 0.0685 | 0.2763 | 13 | 19 | 19 | 15 | 21 | 21 | 0.75450 | 0.5ll | |

NCli:       Id of the clique (starts at 0)
Ref:        Name of the reference (protein name if *–rn* is not used)
Comp:       Name of the comparison (protein name if *–cn* is not used)
Simil:      Similarity flag (empty is *–sim* is not used)
LIG:        Number of matched LIG pseudoatoms
CENTER:   Number of mgrimatched CENTERED pseudoatoms
PROT:       Number of matched PROT peudoatoms
SumCl:      SumCl score (clique-based)
RMSD:       root-mean square deviation (in Å) of the clique
RLig:       Number of reference LIG pseudoatoms
RCent:      Number of reference CENTERED pseudoatoms
RProt:      Number of reference PROT pseudoatoms
CLig:       Number of comparison LIG pseudoatoms
CCent:      Number of comparison CENTERED pseudoatoms
CProt:      Number of comparison PROT pseudoatoms
GrSC:       Grim score (empirical score)
RMSDAl:  root-mean square deviation (in Å) of the clique according to GrScore
NPol:       Number of matched polar interaction pseudoatoms

**-3ʳᵈ mode: Post-processing docking results by interaction graph matching**
**Input directory: $ICHEM_DIR/test/GRIM/screen**
**Input files: docked.mol2, ligand.mol2, site.mol2**
**Output files: Grimscreen_res.tsv, GrScreen.csv**

> **IChem**  *-sim 1* **grim** *site.mol2 ligand.mol2 docked.mol2*

The file docked.txt summarized the docking results obtained with Surflex-Dock: pose nr, docking score (pkd), rmsd to the true X-ray pose

```
Pose          pkd rmsd
ligand_000    8.63 1.14
ligand_001    8.35 1.08
ligand_002    8.12 1.00
ligand_003    8.07 1.14
ligand_004    7.54 1.19
ligand_005    7.10 1.28
ligand_006    7.04 1.41
ligand_007    6.90 1.08
ligand_008    6.72 1.17
ligand_009    6.67 1.34
ligand_010    6.46 1.23
ligand_011    6.34 1.47
ligand_012    6.31 2.02
ligand_013    6.30 3.67
ligand_014    6.28 1.50
ligand_015    6.28 1.27
ligand_016    6.20 2.09
ligand_017    6.16 1.18
ligand_018    6.07 3.07
ligand_019    6.04 3.54
```

All docking poses (docked.mol2) are matched to the X-ray pose (ligand.mol2) for rescoring based on the interaction pattern graph similarity score (GrScore). The full output is stored in the Grimscreen_res.tsv output file.

| NCli | Ref | Comp | Simil | LIG | CENTER | PROT | SumCl | RMSD | RLig | RCent | RProt | CLig | CCent | CProt | GrSc | RMSDA1 | NPol |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2rh1_site | REF | 1 | 9 | 13 | 13 | 0.1054 | 0.2403 | 13 | 19 | 19 | 13 | 19 | 18 | 0.8779 | 0.0049 | 11 |
| 1 | 2rh1_site | REF | 1 | 10 | 11 | 11 | 0.0924 | 0.1405 | 13 | 19 | 19 | 13 | 18 | 17 | 0.8611 | 0.0023 | 9 |
| 1 | 2rh1_site | REF | 1 | 11 | 14 | 16 | 0.1151 | 0.2133 | 13 | 19 | 19 | 14 | 21 | 21 | 0.9589 | 0.0026 | 14 |
| 1 | 2rh1_site | REF | 1 | 10 | 12 | 13 | 0.1064 | 0.2206 | 13 | 19 | 19 | 14 | 20 | 20 | 0.8907 | 0.0024 | 13 |
| 1 | 2rh1_site | REF | 1 | 8 | 9 | 5 | 0.0703 | 0.2583 | 13 | 19 | 19 | 13 | 17 | 16 | 0.7242 | 0.0075 | 7 |
| 1 | 2rh1_site | REF | 1 | 9 | 6 | 4 | 0.0538 | 0.2686 | 13 | 19 | 19 | 12 | 18 | 16 | 0.7091 | 0.0151 | 5 |
| 1 | 2rh1_site | REF | 1 | 9 | 5 | 5 | 0.0565 | 0.2744 | 13 | 19 | 19 | 13 | 20 | 17 | 0.7196 | 0.0186 | 6 |
| 1 | 2rh1_site | REF | 1 | 7 | 12 | 16 | 0.1199 | 0.2240 | 13 | 19 | 19 | 12 | 17 | 16 | 0.8909 | 0.0039 | 10 |
| 1 | 2rh1_site | REF | 1 | 11 | 11 | 13 | 0.1047 | 0.3423 | 13 | 19 | 19 | 14 | 23 | 22 | 0.8928 | 0.0030 | 14 |
| 1 | 2rh1_site | REF | 1 | 8 | 4 | 7 | 0.0573 | 0.4584 | 13 | 19 | 19 | 13 | 19 | 17 | 0.7161 | 0.0072 | 6 |
| 1 | 2rh1_site | REF | 1 | 5 | 6 | 6 | 0.0527 | 0.3720 | 13 | 19 | 19 | 14 | 21 | 18 | 0.6696 | 0.0075 | 7 |
| 1 | 2rh1_site | REF | 1 | 4 | 2 | 6 | 0.0313 | 0.4146 | 13 | 19 | 19 | 12 | 17 | 16 | 0.6313 | 0.0150 | 2 |
| 1 | 2rh1_site | REF | 1 | 3 | 3 | 3 | 0.0191 | 0.5820 | 13 | 19 | 19 | 10 | 16 | 15 | 0.5626 | 0.2846 | 0 |
| 1 | 2rh1_site | REF | 1 | 4 | 3 | 6 | 0.0303 | 0.5961 | 13 | 19 | 19 | 13 | 19 | 17 | 0.6219 | 0.0020 | 2 |
| 1 | 2rh1_site | REF | 1 | 4 | 3 | 6 | 0.0396 | 0.4205 | 13 | 19 | 19 | 12 | 17 | 15 | 0.6365 | 0.0020 | 3 |
| 1 | 2rh1_site | REF | 1 | 7 | 5 | 9 | 0.0665 | 0.3336 | 13 | 19 | 19 | 13 | 16 | 15 | 0.7444 | 0.0031 | 5 |
| 1 | 2rh1_site | REF | 1 | 5 | 8 | 6 | 0.0427 | 0.4845 | 13 | 19 | 19 | 12 | 19 | 18 | 0.6672 | 0.0300 | 2 |
| 1 | 2rh1_site | REF | 1 | 6 | 11 | 15 | 0.0853 | 0.2214 | 13 | 19 | 19 | 14 | 19 | 17 | 0.8506 | 0.0033 | 7 |
| 1 | 2rh1_site | REF | 1 | 4 | 1 | 2 | 0.0128 | 0.3104 | 13 | 19 | 19 | 12 | 19 | 15 | 0.5738 | 0.1752 | 0 |
| 1 | 2rh1_site | REF | 1 | 4 | 7 | 7 | 0.0341 | 0.4100 | 13 | 19 | 19 | 13 | 20 | 19 | 0.6660 | 0.0043 | 1 |

The pose with the highest Grimscore is not the top-ranked one (ligand_000) but pose 002 (highlighted in yellow). Interestingly, it is the pose with the smallest rmsd to the true X-ray pose.

## - GRIM-based alignment of protein-ligand complexes

*realign rigidM mobilM applied1 applied2*

||-> rigidM : reference molecule to apply alignment to
||-> mobilM : comparison molecule to apply alignment from
||-> applied: molecule to apply rotation/translation to

[General options]
   -gmatch N (NAME) Use graph matching to align
      NAME          Atom Name matching
      ATMN          Atomic Name matching
      MOL2          MOL2 Type matching
      CALP          CAlpha Atom matching (protein only)
   --wMob           Also outputs the aligned mobilM
   -rule   R

 By default, the program will perform an atom by atom match, without taking care of what kind of atom it match. If you want to perform a match by regarding only some atoms, this index_string is here to do so
    ex : -i '2-3|1-6|23-160'       Will match the second atom from the reference with the third from the comparison, the first with the sixth ...

**Input directory: $ICHEM_DIR/test/REALIGN**
**Input files: 4amj_INTS_M.mol2, GRIM_ints.mol2, 4amj_prot.mol2 4amj_lig.mol2**
**Output files: rot_4amj_INTS_M.mol2, rot_4amj_lig.mol2, rot_4amj_prot.mol2**

> *IChem --wMob realign GRIM_ints.mol2 4amj_INTS_M.mol2 4amj_lig.mol2 4amj_prot.mol2*

*IChem realign* will apply to the native ligand and/or protein input files (4amj_lig.mol2, 4amj_prot.mol2) a set of rotation/translations (deduced from the previous step) to align them to the reference GRIM input files (2rh1_lig.mol2, 2rh1_prot.mol2). The move mimics the transformation of interaction pseudoatoms previously saved: **4amj_INTS_M.mol2 → GRIM_ints.mol2**

The aligned coordinates of the complex to fit are in the rot_xxx.mol2 files.

## 5. Cavity detection and druggability prediction (VolSite)

*IChem VolSite* is a structure-based tool to automatically detect cavities at the surface of a target protein, and predict their ligandability (structural druggability)
For more information, see:

> Desaphy J, Azdimousa K, Kellenberger E, Rognan D.
> J Chem Inf Model. 2012 Aug 27;52(8):2287-99.
> Comparison and druggability prediction of protein-ligand binding sites from pharmacophore-annotated cavity shapes.

A new feature, introduced from the current version is the generation of cavity-based pharmacophores, as described in

> Tran-Nguyen VK, Da Silva F, Bret G, Rognan D.
> J Chem Inf Model. 2019, 59, 1, 573-585
> All in One: Cavity Detection, Druggability Estimate, Cavity-Based Pharmacophore Perception and Virtual Screening.

*IChem Volsite* can be run in two modes depending on whether coordinates of a bound ligand are given (ligand-restricted mode) or not (unrestricted mode).

```
volsite prot lig              (1)
volsite prot                  (2)

 [General options]
   -step  N (1.5)    Edge length of each box              (Angstroem)
   -boxS  N (20)     Edge length of the main box          (Angstroem)
   -b     N (55)     Minimal threshold for buriedness
   -name  N          PDB Name for output cavity names
   -n     N  (5)     Minimal neighbours for buried cavity boxes
   -nPTS  N (70)     Minimal number of cubes to consider it a cavity
   -NPTS  N (400)    Maximal number of cubes to consider it a cavity
   --dna             Consider DNA as part of the protein
   --cofactor        Consider cofactor as part of the protein
   --solvent         Consider solvent as part of the protein
   --hydrogen        Consider hydrogens
   --desc            Write a descriptor file name descriptor.txt
   --svm             Build a svm property file
   -drog  N          Observed druggability
   --pharm           Generate a pharmacophore (.chm) from cavity
   --outExclu        Output exclusion sphere in pharmacophore file
```

**-Detection of all possible cavities (unrestricted mode)**
**Input directory: $ICHEM_DIR/test/Volsite/unrestricted**
**Input files: protein.mol2**
**Output files: CAVITY_Nx_ALL.mol2 (x=1-9), VolSite_stat.csv**

---

> **> IChem Volsite protein.mol2**

---

VolSite produces a summary (**VolSite_stat.csv**) and a mol2 file (**CAVITY_Nx_ALL.mol2; x=1-6**) for each detected cavity. For the test example (**protein.mol2**), 6 cavities are detected at the surface of the protein and numbered (1 to 6) from the largest to the smallest.

The **VolSite_stat.csv** summary recapitulates the properties of all detected cavities as follows:

| Name | NCav | Size | Buriedness | NPts | Volume | CA | CZ | O | OG | OD1 | N | NZ | DU | Drugg |
|------|------|------|-----------|------|--------|----|----|---|----|-----|---|----|----|----|
| / | 1 | ALL | 84.9875 | 321 | 1083.38 | 87 | 76 | 11 | 27 | 8 | 61 | 34 | 17 | 1.12574 |
| / | 2 | ALL | 78.1878 | 149 | 502.875 | 63 | 15 | 4 | 15 | 8 | 12 | 13 | 19 | 0.71907 |
| / | 3 | ALL | 70.0841 | 133 | 448.875 | 30 | 9 | 6 | 5 | 0 | 32 | 28 | 23 | -0.0102194 |
| / | 4 | ALL | 71.5261 | 80 | 270 | 36 | 11 | 3 | 5 | 2 | 7 | 0 | 16 | 0.543419 |
| / | 5 | ALL | 71.0452 | 78 | 263.25 | 53 | 9 | 0 | 0 | 0 | 2 | 0 | 14 | 1.05675 |
| / | 6 | ALL | 71.3473 | 75 | 253.125 | 18 | 5 | 5 | 0 | 18 | 1 | 18 | 10 | -1.98361 |

1st column: name of the protein (by default, mol2 file protein header)
2nd column: Cavity number
3rd column: Size (ALL: no distance cut-off to the cavity center is applied to define cavity points)
4th column: Average buriedness of the cavity, in %
6th column: Cavity volume in $Å^3$
7th column: number of hydrophobic cavity points (CA)
8th column: number of aromatic cavity points (CZ)
9th column: number of hydrophogen-bond acceptor cavity points (O)
10th column: number of hydrogen-bond acceptor/donor cavity points (OG)
11th column: number of negatively ionizable cavity points (OD1)
12th column: number of hydrogen bond donor cavity points (N)
13th column: number of positively ionized cavity points (NZ)
14th column: number of dummy cavity points (DU): no protein atom < 4.5 Å
15th column: estimated druggability (Drugg): druggable if Drugg > 0; undruggable if Drugg < 0

**-Detection of a cavity around a particular ligand (ligand-restricted mode)**
**Input directory: $ICHEM_DIR/test/Volsite/ligand**
**Input files: protein.mol2, ligand.mol2**
**Output files: CAVITY_Nx_y.mol2 (x = 1-3; y = 4, 6, 8, 12, ALL), VolSite_stat.csv**

**> IChem *volsite* *protein.mol2 ligand.mol2***

The terminal output of the command indicates that a druggable cavity has been found (Druggability: 1.13).
VolSite produces a summary (**VolSite_stat.csv**) and a mol2 file (**CAVITY_Nx_y.mol2; x = 1; y = 4, 6, 8, 12, ALL**) for each detected cavity. For the test example (**protein.mol2**, **ligand.mol2**), 2 cavities are detected at the surface of the protein and numbered (x= 1 to 2) from the largest to the smallest. y indicates the maximal distance (in Å) between cavity points and any input ligand heavy atom. For each cavity, 5 truncation modes are then applied to define binding sites of increasing size (4, 6, 8, 10, 12 Å) around the input ligand. If y = ALL, no truncation is applied.

The **VolSite_stat.csv** summary recapitulates the properties of all detected cavities as follows:

| Name | NCav | Size | NPts | Volume | CA | CZ | O | OG | OD1 | N | NZ | DU | Recovery | Drugg |
|------|------|------|------|--------|----|----|---|----|-----|---|----|----|----------|-------|
| CAU | 1 | 4 | 115 | 388.125 | 31 | 37 | 3 | 12 | 0 | 22 | 6 | 4 | 53.913 | / |
| CAU | 1 | 6 | 160 | 540 | 44 | 44 | 6 | 20 | 0 | 30 | 10 | 6 | 38.75 | / |
| CAU | 1 | 8 | 205 | 691.875 | 58 | 56 | 6 | 22 | 1 | 37 | 15 | 10 | 30.2439 | / |
| CAU | 1 | 12 | 303 | 1022.62 | 83 | 74 | 9 | 26 | 8 | 56 | 30 | 17 | 20.462 | / |
| CAU | 1 | ALL | 321 | 1083.38 | 87 | 76 | 11 | 27 | 8 | 61 | 34 | 17 | 19.3146 | 1.12574 |

2

1st column: name of the ligand (by default, mol2 file ligand header)
2nd column: Cavity number
3rd column: Size (truncation mode)
4th column: Number of cavity points

5$^{th}$ column: Cavity volume in Å$^3$
6$^{th}$ column: number of hydrophobic cavity points (CA)
7$^{th}$ column: number of aromatic cavity points (CZ)
8$^{th}$ column: number of hydrophogen-bond acceptor cavity points (O)
9$^{th}$ column: number of hydrogen-bond acceptor/donor cavity points (OG)
10$^{th}$ column: number of negatively ionizable cavity points (OD1)
11$^{th}$ column: number of hydrogen bond donor cavity points (N)
12$^{th}$ column: number of positively ionized cavity points (NZ)
13$^{th}$ column: number of dummy cavity points (DU): no protein atom < 4.5 Å
14$^{th}$ column: % of the binding site enclosing the ligand
15$^{th}$ column: estimated druggability (Drugg): druggable if Drugg > 0; undruggable if Drugg < 0

**- Outputting cavity properties**

The *IChem --desc* command outputs cavity properties used by our SVM druggability predictor.

**Input directory: $ICHEM_DIR/test/Volsite/unrestricted**
**Input files: protein.mol2**
**Output files: CAVITY_Nx_ALL.mol2 (x=1-6), VolSite_stat.csv, 2rh1_protein_descriptor.txt**

**> IChem  --desc volsite protein.mol2**

The 2rh1_protein_descriptor.txt file outputs for every cavity a vector of 89 reals as follows:

| #  | Descriptor |
|----|------------|
| 1  | Volume |
| 2  | % of aromatic cavity points (CZ) |
| 3  | % of hydrophobic points (CA) |
| 4  | % of h-bond acceptor points (O) |
| 5  | % of negatively ionizable points (OD1) |
| 6  | % of h-bond acceptor & donor points (OG) |
| 7  | % of h-bond donor points (N) |
| 8  | % of positively ionizable points (NZ) |
| 9  | % of dummy points (DU) |
| 10 | Percent of CZ points with a projection value below 40 |
| 11 | Percent of CZ points with a projection value between 40 and 50 |
| 12 | Percent of CZ points with a projection value between 50 and 60 |
| 13 | Percent of CZ points with a projection value between 60 and 70 |
| 14 | Percent of CZ points with a projection value between 70 and 80 |
| 15 | Percent of CZ points with a projection value between 80 and 90 |
| 16 | Percent of CZ points with a projection value between 90 and 100 |
| 17 | Percent of CZ points with a projection value between 100 and 110 |
| 18 | Percent of CZ points with a projection value between 110 and 120 |
| 19 | Percent of CZ points with a projection value of 120 |
| 20 | Percent of CA points with a projection value below 40 |
| 21 | Percent of CA points with a projection value between 40 and 50 |
| 22 | Percent of CA points with a projection value between 50 and 60 |
| 23 | Percent of CA points with a projection value between 60 and 70 |
| 24 | Percent of CA points with a projection value between 70 and 80 |
| 25 | Percent of CA points with a projection value between 80 and 90 |

| 26 | Percent of CA points with a projection value between 90 and 100 |
| 27 | Percent of CA points with a projection value between 100 and 110 |
| 28 | Percent of CA points with a projection value between 110 and 120 |
| 29 | Percent of CA points with a projection value of 120 |
| 30 | Percent of O points with a projection value below 40 |
| 31 | Percent of O points with a projection value between 40 and 50 |
| 32 | Percent of O points with a projection value between 50 and 60 |
| 33 | Percent of O points with a projection value between 60 and 70 |
| 34 | Percent of O points with a projection value between 70 and 80 |
| 35 | Percent of O points with a projection value between 80 and 90 |
| 36 | Percent of O points with a projection value between 90 and 100 |
| 37 | Percent of O points with a projection value between 100 and 110 |
| 38 | Percent of O points with a projection value between 110 and 120 |
| 39 | Percent of O points with a projection value of 120 |
| 40 | Percent of OD1 points with a projection value below 40 |
| 41 | Percent of OD1 points with a projection value between 40 and 50 |
| 42 | Percent of OD1 points with a projection value between 50 and 60 |
| 43 | Percent of OD1 points with a projection value between 60 and 70 |
| 44 | Percent of OD1 points with a projection value between 70 and 80 |
| 45 | Percent of OD1 points with a projection value between 80 and 90 |
| 46 | Percent of OD1 points with a projection value between 90 and 100 |
| 47 | Percent of OD1 points with a projection value between 100 and 110 |
| 48 | Percent of OD1 points with a projection value between 110 and 120 |
| 49 | Percent of OD1 points with a projection value of 120 |
| 50 | Percent of OG points with a projection value below 40 |
| 51 | Percent of OG points with a projection value between 40 and 50 |
| 52 | Percent of OG points with a projection value between 50 and 60 |
| 53 | Percent of OG points with a projection value between 60 and 70 |
| 54 | Percent of OG points with a projection value between 70 and 80 |
| 55 | Percent of OG points with a projection value between 80 and 90 |
| 56 | Percent of OG points with a projection value between 90 and 100 |
| 57 | Percent of OG points with a projection value between 100 and 110 |
| 58 | Percent of OG points with a projection value between 110 and 120 |
| 59 | Percent of OG points with a projection value of 120 |
| 60 | Percent of N points with a projection value below 40 |
| 61 | Percent of N points with a projection value between 40 and 50 |
| 62 | Percent of N points with a projection value between 50 and 60 |
| 63 | Percent of N points with a projection value between 60 and 70 |
| 64 | Percent of N points with a projection value between 70 and 80 |
| 65 | Percent of N points with a projection value between 80 and 90 |
| 66 | Percent of N points with a projection value between 90 and 100 |
| 67 | Percent of N points with a projection value between 100 and 110 |
| 68 | Percent of N points with a projection value between 110 and 120 |
| 69 | Percent of N points with a projection value of 120 |
| 70 | Percent of NZ points with a projection value below 40 |
| 71 | Percent of NZ points with a projection value between 40 and 50 |
| 72 | Percent of NZ points with a projection value between 50 and 60 |
| 73 | Percent of NZ points with a projection value between 60 and 70 |
| 74 | Percent of NZ points with a projection value between 70 and 80 |
| 75 | Percent of NZ points with a projection value between 80 and 90 |
| 76 | Percent of NZ points with a projection value between 90 and 100 |

| 77 | Percent of NZ points with a projection value between 100 and 110 |
| 78 | Percent of NZ points with a projection value between 110 and 120 |
| 79 | Percent of NZ points with a projection value of 120 |
| 80 | Percent of DU points with a projection value below 40 |
| 81 | Percent of DU points with a projection value between 40 and 50 |
| 82 | Percent of DU points with a projection value between 50 and 60 |
| 83 | Percent of DU points with a projection value between 60 and 70 |
| 84 | Percent of DU points with a projection value between 70 and 80 |
| 85 | Percent of DU points with a projection value between 80 and 90 |
| 86 | Percent of DU points with a projection value between 90 and 100 |
| 87 | Percent of DU points with a projection value between 100 and 110 |
| 88 | Percent of DU points with a projection value between 110 and 120 |
| 89 | Percent of DU points with a projection value of 120 |

The projection value is the number of regularly-spaced 8 Å-long vectors emitted from each cavity points intercepting the protein surface (maximum of 120)

**Correspondence between projection value and buriedness**

| Projection value | Buriedness, % |
| --- | --- |
| < 40 | <33.3 |
| 40-50 | 33.3 - 41.6 |
| 50-60 | 41.6 – 50.0 |
| 60-70 | 50.0 - 58.3 |
| 70-80 | 58.3 - 66.6 |
| 80-90 | 66.6 – 75.0 |
| 90-100 | 75 .0 - 83.3 |
| 100-110 | 83.3 - 91.6 |
| 110-120 | 91.6 - 99.9 |
| 120 | 100 |

**- Generating cavity-based pharmacophores**

The *IChem --pharm* generates cavity-based pharmacophores ready to be used in BIOVIA, LigandScout or Shaper2.

**Input directory: $ICHEM_DIR/test/Volsite/pharm**
**Input files: protein.mol2, ligand.mol2**
**Output files: Pharmacophore.chm, Pharmacophore.mol2, Pharmacophore.plp,Pharmacophore.pml**

> *IChem  --pharm -pCA 0 -pCZ 1.56 -pO O -pN 0 -pOD1 0 -pNZ 0 -pZn 3.49* **volsite** *protein. mol2 ligand.mol2*

Usage of –p options (-p CA, -pCZ, -pO, -pN, -pOD1, -pNZ, -pZn) enables to filter out pharmacophoric features by PLP interaction energies (Gehlhaar et al. Chem. Biol. 1995, 2, 317-324) according to a

given threshold. We recommend using the above-described thresholds to optimize the accuracy of cavity-based pharmacophores.

4 files are generated: three in various pharmacophoric formats (BIOVIA .chm, LigandScout .pml, Shaper2 .mol2). The .plp file summarizes the PLP interaction energy (in kcal/mol) of each feature.

**Some notes:**

- The ligand may be a simple atom with user-defined coordinates!
- In ligand-restricted mode, druggability is only estimated for non-truncated full cavities (3rd column = ALL)
- The duggability model is only valid for standard VolSite parameters (step=1.5, boxes=20, b = 55, n =5)! If you change these parameters, keep in mind that the predicted druggability value has no meaning
- For large and accessible cavities (e.g. protein-protein interfaces), standard settings will output small-sized cavities. The most reliable approach to treat such flat cavities is to reduce the value of the *b* parameter (e.g. b= 45) until you get a set of cavity points that fills the expected interface area.
- If you want to consider accessory molecules (co-factor, nucleic acids, water) as part of your protein, use the corresponding options (--dna, --cofactor, --solvent, --hydrogen; see all arguments page 5)
- The additional –OutExclu parameter adds exclusion spheres on cavity-based pharmacophores created with the --pharm option.

## 6. PDB processing (pdbconv)

*Ichem **pdbconv*** is a tool to parse and process PDB files, automatically detect bound ligands (HET code) and their cavity, and estimates their druggability. It is the protocol that we currently use to set-up the sc-PDB database of druggable protein-ligand complexes ([http://bioinfo-pharma.u-strasbg.fr/scPDB](http://bioinfo-pharma.u-strasbg.fr/scPDB)). Fore more information on the parsing process, please see: [http://bioinfo-pharma.u-strasbg.fr/scPDB/ABOUT](http://bioinfo-pharma.u-strasbg.fr/scPDB/ABOUT)

   **pdbconv** *protein[.pdb|.mol2] output_dir pdb_id*

| | |
|---|---|
| --wMOL2 | Use MOL2 File as Input. PDB Options are not available |
| --wUnDrug | Output undruggable cavities |
| --noLig | PDB with no Ligand |

By default all the following options are included. All chains will be kept
[PDB Options]

| | |
|---|---|
| --HARMSIZE | Harmonize size line to 80 characters |
| --MSEMET | Change MSE to MET |
| --CSECYS | Change CSE to CYS |
| --MOVHET | move HETATM to the end of file |
| --ALTATM | select alternative atoms |
| --NUMATM | renumerotate atoms |
| --UPDMAS | update the MASTER line |
| --TOMOL2 | convert to a molecular representation (instead of flat file) |

if you use one of the option below, you MUST use also --TOMOL2 option or use --wMOL2 option

[MOL2 Options]

| | |
|---|---|
| --RESTYP | apply Residue Class (cofactor/STD_AA/MOD_AA/Ligand ...) |
| --BONDSE | create bonds |
| --CLNUNW | clean unwanted residues |
| --MOL2TY | apply MOL2 types according to templates |
| --SPLITM | split molecule into protein/ligand/solvent |
| -SelChain N | List of chains to keep, separated by underscore |
| --SELWAT | select water molecules |
| --SELLIG | select ligand |

**Input directory**: **$ICHEM_DIR/test/PDBCONV**
**Input file: 2RH1.pdb**
**Output directory: ./output**

---

**> IChem *pdbconv* 2RH1.pdb output 2rh1**

---

For each detected ligand, a mol2 file is given along with the corresponding ligand-free protein (mol2 file), and Volsite cavities (mol2 file) in the **output** directory (to be created before executing the command). In addition, cavity descriptors are stored in the **_descriptor.txt**.

## 7. Buried Surface Area calculation

The *IChem utils bsa* command computes the buried surface area of the ligand

   *utils bsa protein ligand*

**Input directory:** $ICHEM_DIR/test/BSA
**Input files:** protein.mol2, ligand.mol2
**Output file:** ligand.bsa

*> IChem utils bsa protein.mol2 ligand.mol2  > ligand.bsa*

The ouput is a 3 columns table:
```
REF     70.1934 347.125
```

1st column: protein-ligand header
2nd column: buried surface area (%)
3rd column: ligand volume ($Å^3$)

### 8. Protein-bound ligand fragmentation

The ***IChem utils frag*** command fragments in 3-D space a protein-bound ligand structure using a method aimed at detecting substituted ring cores. First, a ring perception algorithm is used to automatically detect aromatic and aliphatic rings. Acyclic atoms are then parsed to assign either a linker or substituent label, as whether to the corresponding bonds are connecting two rings or not. Linker atoms are left unchanged. In case of substituent atoms, single bonds involving the closest apolar carbon (in terms of bond distance) to any ring are later cleaved at the condition that the cleaved bond is at least 3 bonds away from the cyclic root atom. An anchoring atom (Z label) is then added to each of the remaining fragments to indicate the cleavage site

For more information, see:

Desaphy J, Rognan D.
J Chem Inf Model. 2014 Jul 28;54(7):1908-18
sc-PDB-Frag: a database of protein-ligand interaction patterns for Bioisosteric replacements.

***utils frag*** *protein ligand*

**Input directory: $ICHEM_DIR/test/Frag**
**Input file: protein.mol2, ligand.mol2**
**Output files: 2rh1_CAU_1_protein_FRAG_1_MOLE.mol2**, **2rh1_CAU_1_protein_FRAG_1_INTS.mol2**, **xx.mol2**

---

***> IChem utils frag*** *protein.mol2 ligand.mol2*

4 fragments are detected and their interaction with the target recorded in the terminal output. To be saved, a fragment should exhibit at least 5 interactions with the target.diff

Three mol2 files are outputted:
**2rh1_CAU_1_protein_FRAG_1_MOLE.mol2**: fragment from the original ligand
**2rh1_CAU_1_protein_FRAG_1_INTS.mol2:** interaction pseudoatoms (merged mode) for the fragment
**xxx.mol2:** interaction pseudoatoms (merged mode) for the entire ligand

**Some notes:**
- Protein and ligand 3D coordinate should be in the same coordinates frame in mol2 file format.
- Several fragments can be generated. Each fragment has an index (header_FRAG_index.mol2) along with the corresponding interaction pseudoatoms (header_FRAG_index_INTS.mol2)

## 9. Fingerprint similarity measures

The *IChem sims* command enables you to compute similarities with various metrics between two fingerprints.

## Fingerprint Similarity

   **sims** *ref comp*                                    *(1)*
   **sims** *file*                                    *(2)*
   **sims** *RefInt CompInt*                                    *(3)*

  [General options]
    --wInts        To use interactions instead of fingerprints (option 3 only)
    --small         Use small fingerprint (option 3 only)
    --binary       To add when the fingerprint is binary
    -metric N (TC)   Select the metric
      TC          Tanimoto metric (default)
      HM          Hamming distance
      RT          Ref Tversky
      FT          Fit Tversky
      DI          Dice
      SO          Soergel

**Input directory: $ICHEM_DIR/test/SIMILARITY**
**Input file: FP1.txt, FP2.txt, FP.txt**
**Output files: sim.txt, matrix.txt**

---

> *IChem  --binary **sims** FP1.txt FP2.txt >sim.txt*

---

Fingerprints of diverse nature are accepted
    STD          Standard   (1 0 21 0 0 3)
    SVM         SVM format (1:1 3:21 6:3)
    CMP         Compressed (1 [1 21 [2 3)

The output is a table listing the reference fingerprint, the target fingerprint and the similarity score:
FP1    FP2    0.809524

Alternatively, a full similarity matrix (all against all) can be generated by concatenating all fingerprints to compare (see FP.txt input file) and running IChem in matrix mode

---

> *IChem  --binary **sims** FP.txt >matrix.txt*

---

The output is a table listing the reference fingerprint, the target fingerprint and the similarity score as follows:
FP1    FP1    1
FP1    FP2    0.809524
FP1    FP3    0.652174
FP2    FP1    0.809524
FP2    FP2    1
FP2    FP3    0.714286
FP3    FP1    0.652174
FP3    FP2    0.714286
FP3    FP3    1

### 9. Detection and analysis of protein-protein interfaces (DetectPPI)

DetectPPI is a tool to detect and analyze protein-protein interfaces, and outputs cavities remote to every detected PPI

For more information, see:

> Da Silva, F., Desaphy, J., Bret, G. and Rognan, D.  (2015) IChemPIC: A Random Forest Classifier of Biological and Crystallographic Protein-Protein Interfaces. J. Chem. Inf. Model, 55, 2005–2014.

**Input directory: $ICHEM_DIR/test/PPI**
**Input file: 4NN6.pdb**
**Output files:**

| | | | |
|---|---|---|---|
| 4NN6_interaction_A_B.ints | 4NN6_prot_B_C.mol2 | CAVITY_all_N2_ALL.mol2 | CAVITY_B_N1_ALL.mol2 |
| 4NN6_interaction_A_C.ints | 4NN6_site_A_B.mol2 | CAVITY_all_N3_ALL.mol2 | CAVITY_B_N2_ALL.mol2 |
| 4NN6_interaction_B_C.ints | 4NN6_site_A_C.mol2 | CAVITY_all_N4_ALL.mol2 | CAVITY_B_N3_ALL.mol2 |
| 4NN6_ints_A_B.mol2 | 4NN6_site_B_C.mol2 | CAVITY_all_N5_ALL.mol2 | CAVITY_C_N1_ALL.mol2 |
| 4NN6_ints_A_C.mol2 | CAVITY_all_N10_ALL.mol2 | CAVITY_all_N6_ALL.mol2 | descriptor.sre |
| 4NN6_ints_B_C.mol2 | CAVITY_all_N11_ALL.mol2 | CAVITY_all_N7_ALL.mol2 | VolSite_Stat.csv |

> ***IChem  detectppi** 4NN6 4NN6.pdb*

IChem requires a PDB file with orientated polar hydrogen atoms. You can use any program for that specific task although we strongly recommend the usage of ProToss (http://protoss.zbh.uni-hamburg.de/) for that purpose.

PDB_interaction_X_Y.ints: interactions between protein chains X and Y
PDB_ints_X_Y.mol2: interaction pseudoatoms (centered mode) between protein chains X and Y
PDB_prot_X_Y.mol2: mol2file of protein chains X and Y
PDB_site_X_Y.mol2: mol2file of interacting residues from chains X and Y
CAVITY_all_X_ALL.mol2: Volsite cavities for the entire protein assembly
CAVITY_X_NY_ALL.mol2: Volsite cavities in chain X only
descriptor.sre: set of 45 descriptors of each PPI
VolSite_Stat.csv: Volsite output

Here is the meaning of the 45 PPI descriptors:

| Name | Description |
|---|---|
| **nPTS** | Total number of interaction points |
| **Hydro** | % of hydrophobic interaction points |
| **Aro** | % of aromatic interaction points |
| **Hbond** | % of hydrogen-bond interaction points |
| **Ionic** | % of of ionic bond interaction points |
| **Hydro1** | % of hydrophobic points (25 %<Burial <33.3%) |
| **Hydro2** | % of hydrophobic points (33.3 %<Burial <41.6%) |
| **Hydro3** | % of hydrophobic points (41.6 %<Burial <50%) |
| **Hydro4** | % of hydrophobic points (50 %<Burial <58.3%) |
| **Hydro5** | % of hydrophobic points (58.3 %<Burial <66.6%) |
| **Hydro6** | % of hydrophobic points (66.6 %<Burial <75%) |
| **Hydro7** | % of hydrophobic points 75 %<Burial <83.3%) |

| | |
|---|---|
| **Hydro8** | % of hydrophobic points (83.3 %<Burial <91.6%) |
| **Hydro9** | % of hydrophobic points (91.6%<Burial <100%) |
| **Hydro10** | % of hydrophobic points (Burial =100%) |
| **Aro1** | % of aromatic points (25 %<Burial <33.3%) |
| **Aro2** | % of aromatic points (33.3 %<Burial <41.6%) |
| **Aro3** | % of aromatic points (41.6 %<Burial <50%) |
| **Aro4** | % of aromatic points (50 %<Burial <58.3%) |
| **Aro5** | % of aromatic points (58.3 %<Burial <66.6%) |
| **Aro6** | % of aromatic points (66.6 %<Burial <75%) |
| **Aro7** | % of aromatic points 75 %<Burial <83.3%) |
| **Aro8** | % of aromatic points (83.3 %<Burial <91.6%) |
| **Aro9** | % of aromatic points (91.6%<Burial <100%) |
| **Aro10** | % of aromatic points (Burial =100%) |
| **Hbond1** | % of hydrogen bond points (25 %<Burial <33.3%) |
| **Hbond2** | % of hydrogen bond points (33.3 %<Burial <41.6%) |
| **Hbond3** | % of hydrogen bond points (41.6 %<Burial <50%) |
| **Hbond4** | % of hydrogen bond points (50 %<Burial <58.3%) |
| **Hbond5** | % of hydrogen bond points (58.3 %<Burial <66.6%) |
| **Hbond6** | % of hydrogen bond points (66.6 %<Burial <75%) |
| **Hbond7** | % of hydrogen bond points 75 %<Burial <83.3%) |
| **Hbond8** | % of hydrogen bond points (83.3 %<Burial <91.6%) |
| **Hbond9** | % of hydrogen bond points (91.6%<Burial <100%) |
| **Hbond10** | % of hydrogen bond points (Burial =100%) |
| **Ionic1** | % of ionic bond points (25 %<Burial <33.3%) |
| **Ionic2** | % of ionic bond points (33.3 %<Burial <41.6%) |
| **Ionic3** | % of ionic bond points (41.6 %<Burial <50%) |
| **Ionic4** | % of ionic bond points (50 %<Burial <58.3%) |
| **Ionic5** | % of ionic bond points (58.3 %<Burial <66.6%) |
| **pwIonic6** | % of ionic bond points (66.6 %<Burial <75%) |
| **Ionic7** | % of ionic bond points 75 %<Burial <83.3%) |
| **Ionic8** | % of ionic bond points (83.3 %<Burial <91.6%) |
| **Ionic9** | % of ionic bond points (91.6%<Burial <100%) |
| **Ionic10** | % of ionic bond points (Burial = 100%) |

Please note that we recommend the usage of our web interface (http://bioinfo-pharma.u-strasbg.fr/IChemPIC) to predict the relevance (biological, crystallographic) of any possible interface from a PDB file.