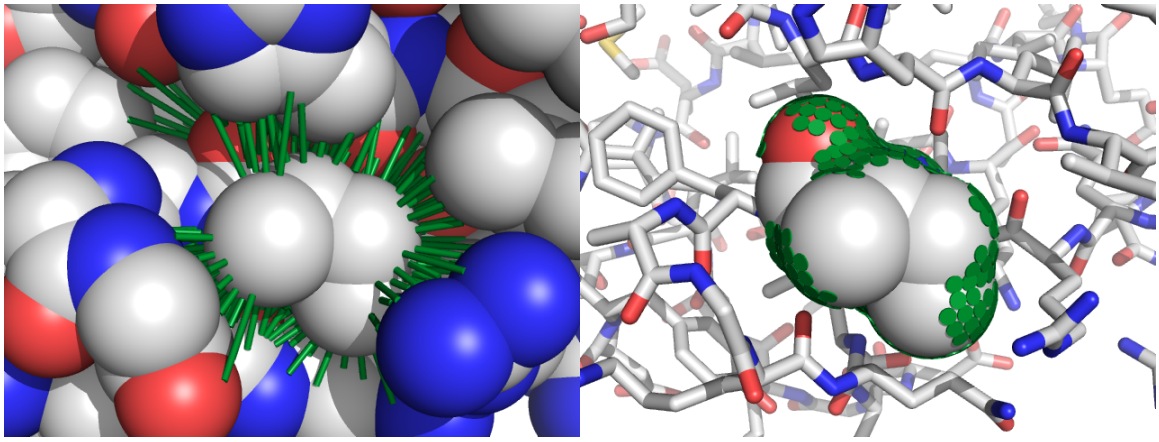
Occluded Surface (OS)



OS is a package of programs to calculate the **occluded surface** and **atomic packing** of protein model structures.

Occluded surface is defined as the molecular surface that is less than 2.8 Å from the surface of neighboring non-bonded atoms. That is, if a water molecule **cannot** fit between two atoms they occlude each other. Occluded surface is similar to buried surface but is more sensitive to packing geometry than buried surface identified using a rolling probe.

To calculate occluded surface, normals at the molecular surface are extended outward until they intersect neighboring van der Waals surface. The collection of extended normals, and their respective lengths, defines the **packing** of each atom in a structural model. The figure above on the left illustrates this concept. The occluded surface of individual atoms in a residue is illustrated by the green patches in the figure above on the right.

- Either the **normals** or the molecular surface **patches** (above right) that represent the occluded surface may be displayed in PyMOL using scripts provided.
- Analysis of the **occluded surface** enables one to identify amino acid residues in *unusual occluded surface environments* compared to a database of high resolution structures. Residues in an unusual environment are frequently **incorrectly modeled**, involved in **crystal contacts**, or involved in **ligand binding**.
- A weighted parameter, **Psr**, related to the **occluded surface** is within a narrow range for native protein structures; if a protein model has a **Psr** value outside this range the model is probably wrong in at least some areas.
- Analysis of the collection of extended normals (**raylengths**) and their associated occluded surface enables an estimation of the **packing** of the protein both at the **residue level** and for the protein as a **whole**. The *output list* identifies all atoms which interact with each atom in the structure, the surface area involved in the interaction and the average raylength for that interacting patch of surface area.
- A combination of **occluded surface area** and **average length of the normals** (raylengths) may be used to obtain the **Occluded Surface Packing (OSP)** value for each residue. These residue **OSP** values are useful for identifying regions of close packing versus areas of loose packing in a protein.
- Analysis of **inter-chain occluded surface** allows a detailed calculation of protein-protein interactions.

Input

Use of the program requires four files:

1. A **PDB file** of coordinates. This must be a '**clean**' file with no *alternate conformations*; no *hydrogens*; and only *one chain*. You should *renumber* your multi-chain files to be *sequentially* numbered, otherwise the output will be unintelligible because chain identifiers are not kept.

A shell script to clean up a PDB file called *extract* is included with the distribution. Type,

```
extract pdb[code].ent
```

and a new "clean" PDB file called, *pdb[code].cln* will be created.

OS expects a PDB file with a single chain, with the residues sequentially numbered, it ignores chain ID. If you have a multiple-chain protein, remove the TER records and renumber the PDB file sequentially. A simple FORTRAN program to renumber your clean PDB file is included in the OS distribution. Edit your pdb[c*ode*].*pdb* (or pdb[c*ode*].*cln*) file so that the first residue has the number you want to start with. Then launch the [*path*]/*os_v75*/*bin*/*renum* program from the command line and follow the prompt. Your renumbered file will be called *new.pdb*.

You may use **OS** to calculate the occluded surface and raylengths of non-amino acid molecules in a structural model. Simply change the **HETATM** to **ATOM** and make sure that the rest of each record corresponds to PDB format and the numbering is sequential. In addition, you should make sure that each element in your PDB file has an entry in the **radii** file. This file is temporarily deposited in your *cwd* for the **OS** run. If you want to make your own, copy it to your directory using the command in ***os.run***. Then edit the file.

Note: The analysis programs, *occsurf* and *respak* recognize only the 20 amino acid names and CYT (for OPLS dissulfide linked residues) and UNK (for all others including ligands). If you want analysis of ligand packing change the PDB file or the "prot.srf" file so that all three-letter residue names are one of the above.

2. A **command file**, *os.run*

A sample file is in the **OS** distribution. You must edit the path to the installation to define the OSDIR variable correctly. You should **uncomment the appropriate lines** in *os.run* to run the different programs in the package.

3. A **data file**, *os.fil*,

The data file, *os.fil*, tells **OS**

1. The name of the clean **PDB** file

2. The **range of residues** to calculate the occluded surface for (all atoms are considering as potential occluding atoms, but the occluded surface is only calculated for those given in the range)

3. Whether or not to **create a file for display of the extended rays** using **PyMOL** (This option should be used for only a few residues at most – the files are large)

The format of *os.fil* is as follows;

```
filename.pdb
2
167
n
```

An example to display the rays as in the at the top of this document is:

```
1stn.pdb
23
23
y
```

The "**y**" in the last line causes a file called ***rays*** to be generated.

4. An **atom radii** file, *radii*

A default radii file is supplied and will be read by the program. If you want to use your own radii set, edit the file, *radii*, in the *data* subdirectory of the OS distribution.

Execution

If the above files have been deposited in your working directory, and the *os.run* file has been edited to define the executable file location, to start simply type:

```
./os.run
```

Note that the various FORTRAN source files have been compiled for several computer architectures. You may have to re-compile if the executables are not compatible with your machine. See the INSTALL file in the *os_v75/bin/* directory.

Output

1. The *prot.srf* file.

The **prot.srf** file contains all the information from the calculation of occluded surface and extension of surface normals. The additional programs in this package analyze the data in this file.

A sample of a typical **prot.srf** file is explained below:

```
...
AVG for ATOM: CG2   28.374 es    2.792 os   31.166 ts    0.527 Rln THR   13
INF THR   13@CG2_>LEU  14@O___     7 pts   1.412 A2    0.175 Rlen  3.74 Dxx
INF THR   13@CG2_>ILE  15@CG1_     5 pts   0.858 A2    0.840 Rlen  5.94 Dxx
INF THR   13@CG2_>MET  26@CE__     1 pts   0.299 A2    0.991 Rlen  6.52 Dxx
INF THR   13@CG2_>ILE  15@CA__     1 pts   0.223 A2    0.964 Rlen  6.11 Dxx
    ES_Total      :   57.962                                        THR   13
    ES_Backbone   :   10.702                                        THR   13
    ES_Side Chain :   47.260                                        THR   13
    OS_Total      :   32.587                                        THR   13
    OS_Backbone   :   25.113                                        THR   13
    OS_Side Chain :    7.474                                        THR   13
    Ave_Raylength :    0.430                                        THR   13
    MC_RAYLENGTH  :    0.426                                        THR   13
    SC_RAYLENGTH  :    0.438                                        THR   13
...
```

where,

For the **AVG** line:

es is the exposed surface area in Angstroms for that atom
os is the occluded surface area in Angstroms for that atom
ts is the total surface area in Angstroms for that atom
Rln is the mean raylength for that atom

The first **INF** line means the following:

The **CG2** atom of **THR 13** is occluded by the **O** atom of **LEU 14**;
seven (**7**) surface points (**pts**) with a total surface area of **1.412** square Angstoms (**A2**) is occluded;
the mean fractional raylength (**Rlen**) for this occluded patch is **0.175**;
the distance between the **CG2** and **O** atom centers is **3.74** Angstroms (**Dxx**).

In the residue summary section:

ES_Total is the total exposed surface area for THR 13
ES_Backbone is the exposed surface area of the N, CA, C, O atoms
ES_Side Chain is the exposed surface area of the side chain atoms
OS_Total is the total exposed surface area
OS_Backbone is the exposed surface area of the N, CA, C, O atoms
OS_Side Chain is the exposed surface area of the side chain atoms
Ave_Raylength is the mean fractional raylength for the residue (Equals 1.0 if there is no occluding van der Waals surface within 2.8 Angstroms of the molecular surface of THR 13; equals 0.0 if 100% of the molecular surface of THR 13 were in contact with van der Waals surface of other atoms.)
MC_RAYLENGTH and SC_RAYLENGTH break down the raylength averages to main chain and side chain components.

2. Evaluation of occluded surface environment.

The **occluded surface** area for each residue is compared to the distribution of occluded surface areas for that type of amino acid in a data set of high resolution structures. The output is found in the file, *prot.eval* as shown below.

The two important parameters here are **Ri** and **Psr**. A plot of **Ri** tells you which residues are in unusual environments. And **Psr** should be between 0.60 and 0.74 or your structure is probably wrong.

```
...
Sw:           26.87    THR  316
St:           90.10    THR  316
Ri:            0.30    THR  316
Sw:           52.73    PRO  317
St:           92.00    PRO  317
Ri:            0.57    PRO  317
    Total_Res:      317
    Pw:        20144.56
    Pi:        30304.49
    Psr:           0.66
```

where,

Sw is a weighted surface area of the residue. This parameter is a function of both the occluded surface area of the residue and the distribution of occluded surface areas found in a data set of high resolution structures.

St is the total surface area for that residue type.

Ri is a weighted parameter describing how "usual" the occluded surface environment is:
Below a value of 0.2 indicates that the residue is in a highly unusual environment.
A value of 1.0 indicates that the residue is in a completely average environment for that residue type.

Total_Res is the total number of residues

Pw is the weighted surface area of the whole protein

Pi is the ideal weighted surface area of a protein with the same amino acid composition.

Psr is the normalized protein surface ratio. A value between 0.60 - 0.74 is found for *properly folded proteins*. If your structure has a **Psr** value outside of this range it is probably wrong.

3. Evaluation of the packing of the protein model.

The **Occluded Surface Packing Value (OSP)** is calculated for each residue. This value is 0.0 for completely exposed residues and is 1.0 for the impossible case where 100% of molecular surface is in contact with other van der Waals surface. The average for a protein is dependent on the molecular weight of the protein.

For each atom that has occluded surface, the value of **(os*[1-raylength])** is calculated.
Here **raylength** is the fractional length of the extended normal. (A fractional length of 1.0 is equivalent to the diameter of a water molecule, 2.8 Ang.) These values are summed for each residue to give the parameter listed in column four of the output file below. This parameter is divided by the ***total surface area*** for that amino acid type to give a **normalized occluded surface packing value** listed under "**OSP**" in the output file.

Thus, if the complete residue surface is occluded, and all rays have a length of zero, the packing value is the same as the total residue surface area and **OSP** = 1.0.

Output is in the file, *prot.pak*

```
Resnum  Resname       OS     os*[1-raylen]   OSP
   74     THR       36.93       27.61       0.263
   75      ?         0.0         0.0        0.0
   76     ILE       41.35       20.30       0.178
   77     ILE       41.15       24.81       0.219
   78     PHE       56.07       38.81       0.269
   79     GLY       46.99       35.25       0.659
   80     VAL       55.52       38.29       0.397
    ...
```

Note: If a residue has zero occluded surface the residue name will not be listed as for residue 75, above.

The output file, *pak.plt*, contains columns 1 and 5 from the above file **without** the column labels for plotting **residue numbe**r *versus* **OSP**.

Note for ligands in OS: To calculate the ligand/protein **`OS`** and **`os*[1-raylen]`** change the name of the ligand to **`UNK`** in the **prot.srf** file. The program **respak** will not calculate parameters for unrecognized residue names.

4. Evaluation of interchain occluded surface.

Only **intermolecular occlusion** is considered here. To run **intchos** concatenate the two chains (A and B, or segid A and segid B) or two interacting molecues into one PDB file and renumber the residues using **renum**. Then enter the numbers of the first and last residues of one chain or one molecule in the indicated place in **os.run** and run **os.run**.

The program **int**[er]**ch**[ain]**os** is in the **OS** installation and can be run standalone if you have a **prot.srf** file in your directory. The program **intchos** will ask you for the residue numbers of the first and last residues in one chain or one molecule. The output will be in **intch.os**.

For each atom that has occluded surface, the value of **os** and **(os*[1-raylength])** is calculated. These are summed for each residue to give the parameters listed in columns three and four of the output file, **intch.os**, as shown below.

Output is in the file, ***intch.os***

```
Resnum  Resname       OS     os*[1-raylen]
  ...
  74     THR        0.45        0.00
  75      ?         0.0         0.0
  76     ILE       17.29        4.34
  77      ?         0.0         0.0
  78     PHE        4.05        1.95
  79     GLY       23.35       17.54
  80     VAL       23.16       14.20
  81      ?         0.0         0.0
  82     ALA        8.05        4.02
  83     GLY       25.96       18.73
  84     VAL       21.07       12.72
  ...
```

Note: If a residue has zero interchain occluded surface the residue name will not be listed as for residues 75, 77, and 81above.

Note for ligand docking OS: To calculate the ligand/protein **OS** and **os*[1-raylen]** change the name of the ligand to **UNK** in the **prot.srf** file. The program **intchos** will not calculate parameters for unrecognized residue names.

5. Display of the occluded surface and/or normal using PyMOL.

The file ***raydist.lst*** (output by **OS**) contains the coordinates, lengths, and normal vectors for the rays of the residue(s) you specified in the *os.fil* file (if you entered a "**y**" on the last line).

Copy the *os_v75/bin/Disp_dots.py* and *Disp_rays.py* files to your working directory (where the ***raydist.lst*** file is located).

Launch **PyMOL** in your working directory, load your protein model, and in the command window type,

`select os_res, [resid]` [whatever_residues_listed_in_os.fil]

Display the **os_res** object as spheres.

Then type the following in the command window,

`run Display_dots.py` and follow the prompt.

or

`run Display_rays.py` and follow the prompt.

References

If results derived from the calculation of occluded surface are published please cite the following:
Pattabiraman, N., Ward, K.B. and Fleming, P.J. (1995) Occluded Molecular Surface: Analysis of Protein Packing, *Journal of Molecular Recognition*, **8**:334-344. (This is the original description of the OS method).

Fleming,P.J.and F.M.Richards (2000) Protein Packing:Dependence on Protein Size, Secondary Structure and Amino Acid Composition. *J.Mol.Biol.* 299, 487-498. (This is the most complete description of occluded surface packing and includes packing results for a dataset of 152 proteins).

Vorobjev, Y.N. and Hermans, J. (1997) SIMS: Computation of a Smooth Invariant Molecular Surface. *Biophysical Journal*, **73**:722-732. (SIMS is used to calculate the dot surface for OS).

An example of increased packing in the stabilization of a hyperthermophilic protein:

Brian S. DeDecker, Ronan O'Brien, Patrick J. Fleming, James H. Geiger, Stephen P. Jackson, Paul B. Sigler (1996) The Crystal Structure of a Hyperthermophilic Archaeal TATA-box Binding Protein, *J. Mol. Biol.*, **264**, 1072 - 1084.

A correlation of OS with the energetics of protein-protein interaction is described in the following paper:
Fleming, KG, Ackerman, AL and DM Engelman (1997) "The Effect of Point Mutations on the Free Energy of Transmembrane alpha-Helix Dimerization" *J. Mol. Biol.* **272:** 266-275.

OS was used to show that most NMR structures have different packing than X-ray crystal structures in the following report:
Ratnaparkhi et al. (1998) Discrepancies between the NMR and X-ray Structures of Uncomplexed Barstar: Analysis Suggests That Packing Densities of Protein Structures Determined by NMR Are Unreliable, *Biochemistry*, **37**, 6958-6966.

A correlation of OS with the energetics of cavity formation in proteins is described in the following paper:
Ratnapharkhi, G.S. and Varadarajan, R. (2000) "Thermodynamic and Structural Studies of Cavity Formation in Proteins Suggest That Loss of Packing Interactions Rather Than the Hydrophobic Effect Dominates the Observed Energetics. *Biochemistry* **39:** 12365-12374.