PADRE: Pairwise Atomic Density Reverse Engineering

What is PADRE?

Pairwise Atomic Density Reverse Engineering (PADRE) implements an algorithm to compute local molecular shape by "reverse engineering" the Lennard-Jones potential. Specifically, PADRE computes an estimate of the atomic density exponent at a collection of arbitrary points about a molecule. The atomic density at a point is defined as the slope of log(N) vs. log(r), where N is the number of atomic centers within r Angstroms from the point. In crevices, the density tends to be high (N accumulates rapidly), while on protrusions, the density tends to be low.

Since PADRE can compute the density at arbitrary points, it can be used to evaluate the shape on a surface about the molecule computed by some other means. In this regard, PADRE is specifically designed to work with Michel Sanner's MSMS program.

Why use PADRE?

PADRE is designed to aid in the molecular modeling of proteins. In particular, the program is able to rapidly elucidate features of interest, such as crevices, grooves and protrusions. The “topographical” information produced by PADRE can help researchers easily pinpoint the most prominent features of a protein, regions which are likely to participate in interactions with other molecules.

Who and Where

PADRE was developed at the San Diego Supercomputer by Julie C. Mitchell and Rex Kerr. PADRE is distributed by the Computational Center for Macromolecular Structures. Please visit our website http://www.sdsc.edu/CCMS

References

Please reference the following manuscript in any scientific publications that use PADRE’s atomic density methods for shape characterization:

[1] Mitchell, J.C., Kerr, R. and Ten Eyck, L.F. (2001) “Rapid Atomic Density Methods for Molecular Shape Characterization,” J. Mol. Graph. Model, 19(3), pp. 324-329.

Installing PADRE

Supported Platforms

PADRE has been compiled and successfully run on all of the following platforms:

Platform	OS Version
SunOS (solaris)	5.5
IRIX	6.5
IRIX64	6.2
OSF1 (alpha)	4.0
Linux	2.2.16
Darwin (Mac OS X)	10beta

To compile PADRE on Unix-based systems, a C compiler is required. Compiled executables are available for common Unix platforms and Macintosh OS 8/9/10. The Mac OS 8/9 version is available by special request to mitchell@sdsc.edu

Installing PADRE

Installing PADRE should be very simple.

1. Download the file "fadepadre.tar.gz"
2. Type "gunzip fadepadre.tar.gz"
3. Type "tar xf fadepadre.tar"
4. Type "cd FADEPADRE"
5. Type "make" (Note: you may skip this if you use a precompiled binary)

Hopefully, this will work.

Some caveats

If you do not have "gmake" installed on your system, change the value of "MAKE" in the root "Makefile" to "make" For some systems, you may also need to change the C compiler. You can do this permanently by editing the "configure" file in the "utils" directory. You can do this temporarily (after configuration) by editing the "Makefile" in the "src" directory. If you are still having problems, please use the compiled executable for your system.

Bug reports

Please report any bugs to mitchell@sdsc.edu. Include the terminal output for the example you are having problems with, and indicate the operating system you are using.

Running PADRE

Getting Help

Running PADRE without arguments or with -?, -h or -help will print out a help screen.

```
PADRE -examples
```

gives several examples of useful arguments to give PADRE.

```
PADRE -advanced
```

prints out a help screen on the more advanced arguments.

Input and Output Files

PADRE input types:

.xyz	Location of atomic centers specified one per line.
.vert	Location of evaluation points specified one per line.
.face	For VRML output, specifies triplets of vertices for each face
.color	For VRML output specifies a mapping between color and density
.PDB	Protein Databank files.

PADRE output types:

.pad	Atomic density at each evaluation point
vrml	VRML file showing the result in 3D coded by color.

XYZ and Vert file formats

.xyz and .vert files are expected to consist of the following:

Each line must contain at least three numbers specifying the X, Y, and Z coordinates of an atomic center (in Angstroms). Numbers after the first three are ignored.

Face file format

Each line must contain at least three integers interpreted as the first, second, and third vertices of a triangular simplex. Numbers after the first three are ignored. The integers refer to the vertices given in the .vert file. Numbering starts at 1.

PDB file format

The PDB file format current as of this writing is described at

http://www.rcsb.org/pdb/docs/format/pdbguide2.2/guide2.2_frame.html

PADRE relies on the ATOM and HETATOM labels to determine which lines contain positions of atomic centers, on column 22 to determine the chain, and on columns 13-14 and 77-78 to determine if a given atom is a hydrogen.

Color file format

Each line must contain at least four numbers interpreted as RGB intensity (0.0 to 1.0) followed by a relative spacing between the current color and the next color. These colors are mapped to the range of densities found with linear interpolation. *This should be considered a "beta" feature, and both the feature and file format could potentially change in the future (or simply be ignored).*

PAD file format

The output (.pad) file format is

```
x  y  z  distance  exponent  lj_potential
```

for list output. When only is grid is being output and the -grid_out option is given, the output file will contain a matrix of density exponents only.

VRML file format

VRML is too complex of a markup language to describe in detail here. The program attempts to place the camera position such that the entire molecule is in view and set the scale such that manipulation occurs at a reasonable rate. The camera is placed initially to look down along the Z axis from above the center of the molecule.

Molscript file format

A description of this format can be found in the documentation for Molscript.

Command-line options

Overview

Running PADRE with no arguments will cause it to print out a help screen.

```
PADRE <filename>
```

will read <filename> (or <filename>.xyz if the .xyz isn't specified), perform an analysis on the default grid, and write to <filename>.pad. If the filename ends with .pdb or .PDB, PADRE will assume it is a Protein Databank file. If no extension is given, and a.xyz file is not found, it will try to read a file with the extension .pdb.

The general syntax is

```
PADRE [options] <filename> [<output filename>]
```

The output filename, if specified, will be used instead of <filename>.pad.

Grid evaluation options

```
-g=X[xY[xD]]                    Set grid dimensions and spacing
-g                              Use planar grid
-orient=theta/phi               Set orientation of plane normal vector
-units=x1/y1/z1/x2/y2/z2
-center           centers grid at the molecule center.
-center=X/Y/Z.    centers grid at coordinates X,Y,Z.
```

The default grid is 50x50 with spacing of 1.0 Angstroms. To change the grid size, use **-g=X** for a square grid or **-g=XxY** for a non-square grid. For instance,

```
-g=50x100
```

would specify a grid of 50 points in the X direction and 100 in the Y with spacing of 1.0 Angstroms. To change the grid spacing, use **-g=XxYxD**, where D is the number of Angstroms between successive grid points. For instance,

```
g=50x100x0.5
```

would specify a 50x100 grid with 0.5 Angstrom spacing (covering a total area of 25x50 Angstroms).

-orient=theta/phi adjusts the orientation of the normal vector to the plane of the grid. Theta measures the angle from vertical; phi measures the angle from the +x axis (counterclockwise). Units are degrees. For instance, to specify a grid with a unit normal in the X direction (a Y-Z plane), use

```
-orient=90/0
```

-units=x1/y1/z1/x2/y2/z2 specifies unit vectors for the edge of each cell in the grid. If the vectors are not normalized to the spacing (default=1, or specified with -g), they will be. If they are not at right angles, the orthogonal component of the second vector will be used. For instance,

```
-units=0/1/0/0/0/1
```

specifies a grid in the Y-Z plane.

-center translates the center of the grid to the center of the molecule. You can translate to an arbitrary coordinate by using -center=X/Y/Z. For instance,

```
-center=3.7/2.9/-1.4
```

uses (3.7, 2.9, -1.4) as the center of the grid. If you want to specify a location relative to the center of the molecule, use both -center and -center=X/Y/Z. That is, if you want the grid center to be 2 Angstroms in the x direction from the molecule center, use

```
-center   -center=2.0/0.0/0.0
```

Note that -center only translates the atomic centers, not the evaluation points.

Vertex and surface evaluation options

-p	Compute over a set of points (<filename>.vert)
-p=<filename>	Override default naming for input .vert file
-s=<filename>	Override default naming for input .face file

Note: See notes on the “bloat” utility before running a surface computation.

Note: If -p and -g are both given, both will be computed and output.

Output options

-quiet	No screen output (not even errors)
-nop	No output files (computations only)
-g	Compute over a grid
-g=X[xY[xD]]	Compute over grid with specified sizes
-grid_out	Write grid in matrix format
-v	Produce VRML output
-v=<filename>	Produce VRML and write to specified filename
-m	Produce Molscript output
-m=<filename>	Produce Molscript and write to specified filename
-o	Output to list (.pad) format
-o=<filename>	Override default naming for output .pad file

PDB file options

Note that PDB files must end in .pdb or .PDB to be properly read by PADRE. If both .xyz and .PDB files exist, and the extension is not specified, PADRE will use the .xyz file.

-usehet	use HETATM records (ignored by default)
-useh	use hydrogens (ignored by default)
-chain=<letters>	use specified chains only (case sensitive)

Miscellaneous options

-c=<filename>	Reads a .color file
-min=m	Set minimum density exponent for color interpolation
-max=M	Set maximum density exponent for color interpolation
-buffer=N	Set maximum number of simplices and points (default is 262144)
-r0=R	Sets the Lennard-Jones radius (default is 3.0)

Note: the algorithm does not accurately compute the density at distances closer to the molecule than r0.

Software Utilities

bloat

A "bloat.pl" Perl script is provided in the "bin/utils/scripts" directory. There is a wrapper "bloat" in the "/bin/work" directory. This script can expand or contract the size of atoms given in a XYZR file. The syntax for the "bloat" wrapper is

```
bloat   input_file   dr  > output_file
```

where dr is the number of Angstroms to add to the radius. (Negative numbers may be used.)

Bloat can be useful when constructing a surface that is a certain distance outside (or inside) the traditional molecular surface. For instance, one could use the output from bloat as input to MSMS. PADRE does not use the radius information itself.

Note: PADRE's **density exponents are essentially meaningless at points on or inside the molecular surface.** To view shape characteristics **near the surface**, we advise using bloat with a value of dr that is not significantly smaller than the value of r0 used in the Lennard-Jones computations. The default value for r0 is 3.0 Angstroms.

msmspadre

A “msmspadre.awk” AWK script is provided in the “bin/utils/scripts” There is a wrapper “msmspadre” in the work directory. This script sets up a PADRE computation over a surface, starting with a PDB file and using the molecular surface software MSMS. For more information on installing and running MSMS, see

http://www.scripps.edu/pub/olson-web/people/sanner/html/msms_home.html

PADRE understands the .face and .vert files output by MSMS. MSMS needs to be invoked with the -no_header option for PADRE to correctly interpret the files. The syntax for the “msmspadre” wrapper is

```
msmspadre  basename
```

where “basename.pdb” is the PDB file for which a surface is desired. We have set useful default values for the computation. The user can change these and/or write a more elaborate script.

Note: There is a “msms” wrapper in the “bin/work” which directory **must be edited** to point to the MSMS binary executable.

Troubleshooting and Known Bugs

PADRE and multiple structures

Q: "Why does my output look funny? PADRE thinks everything is a crevice."

A: You probably have multiple structures in your PDB file, which is frequently the case with structures obtained by NMR. Because PADRE's methods rely on an atom counting scheme, this could skew the results.

Shell script wrappers

Q: "I am having problems running the scripts in the /bin/work directory."

A: The files in this directory are wrappers to call executables in the /bin/exec and /bin/utils directories. The scripts automatically figure out which system you are running and call the appropriate executable. Your problem likely comes from the fact that you do not have the Korn shell (ksh) installed on your Unix system. This shell is used to allow more than 9 arguments to be called in the script. If you do not have the Korn shell, you can try executing the script using other shells (for example, try "sh PADRE …" or "bash PADRE …" to see if this fixes the problem. Alternately, you can just use the actual executables located in the /bin/exec and /bin/utils directories.

Mac OS X

Q: "I double-click on the PADRE executable, but it doesn't open."

A: Executables for Macintosh OS X are compiled as native Unix tools. This means that you need to run them from a Unix terminal. There is a program called "Terminal" included with Mac OS X for this purpose. If you are not familiar with navigating Unix-based filesystems, the book "Unix in a Nutshell" by Arnold Robbins can help you get started.