Appendix C: Cleaning pdb Files

Many pdb files have extraneous information and often do not conform to file standards. You may have to "clean" your pdb file before loading it into PyRosetta. You can do this through a variety of methods.

- From within a UNIX shell:

```
grep "^ATOM" 1ABC.pdb > 1ABC.clean.pdb
```

- From within a DOS shell:

```
findstr /b "ATOM" 1ABC.pdb > 1ABC.clean.pdb
```

- Using a PyRosetta `toolbox` method `cleanATOM`

```
from toolbox import cleanATOM
cleanATOM("1YY8.pdb")
```

  In this example, the `toolbox` method `cleanATOM` will create a file called `1YY8.clean.pdb`, with all non-`ATOM` lines removed. Warning: this method will overwrite any other files of the same name in its directory.

All of these methods will remove all lines that do not begin with `ATOM` in the pdb file and create a new "clean" pdb file named `1ABC.clean.pdb`.

One could also easily write a script to clean multiple pdb files at once. Here is an example Bash (UNIX shell) script that will clean all pdb files in a single directory:

```
#!/bin/sh
for pdbfile in *.pdb
do
        echo "cleaning $pdbfile..."
        grep ^ATOM $pdbfile > ${pdbfile%.pdb}.clean.pdb
done
```

If after cleaning a pdb file, PyRosetta still gives errors, you might use your text editor to open the pdb file and edit or remove the offending data lines manually.