

# ProDy

Protein Dynamics & Sequence Analysis

## **ProDy Tutorial**

*Release*

**Ahmet Bakan, Cihan Kaya**

June 07, 2024



<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Structural Ensemble Analysis . . . . .	1
1.2	Elastic Network Models . . . . .	1
1.3	Trajectory Analysis . . . . .	1
1.4	Visualization . . . . .	1
<b>2</b>	<b>How to Start</b>	<b>3</b>
2.1	Using ProDy . . . . .	3
2.2	Interactive Usage . . . . .	3
2.3	Using Documentation . . . . .	4
<b>3</b>	<b>ProDy Basics</b>	<b>7</b>
3.1	File Parsers . . . . .	7
3.2	Analysis Functions . . . . .	8
3.3	Plotting Functions . . . . .	8
3.4	Protein Structures . . . . .	9
3.5	Atom Groups . . . . .	10
3.6	ProDy Verbosity . . . . .	11
<b>4</b>	<b>Atom Groups</b>	<b>13</b>
4.1	Building an Atom Group . . . . .	13
4.2	Storing data in AtomGroup . . . . .	17
<b>5</b>	<b>Atom Selections</b>	<b>19</b>
5.1	Atom Selections . . . . .	19
5.2	Operations on Selections . . . . .	21
<b>6</b>	<b>Hierarchical Views</b>	<b>25</b>
6.1	Hierarchical Views . . . . .	25
6.2	Chains . . . . .	26
6.3	Residues . . . . .	28
6.4	Atoms . . . . .	29
6.5	State Changes . . . . .	30
<b>7</b>	<b>Structure Analysis</b>	<b>31</b>
7.1	Measure geometric properties . . . . .	31
7.2	Compare and align structures . . . . .	31
7.3	Writing PDB files . . . . .	32
<b>8</b>	<b>Dynamics Analysis</b>	<b>35</b>
8.1	PCA Calculations . . . . .	35

8.2	ANM Calculations . . . . .	36
8.3	Comparative Analysis . . . . .	37
8.4	Output Data Files . . . . .	37
8.5	External Data . . . . .	38
8.6	Plotting Data . . . . .	38
8.7	More Examples . . . . .	39
<b>9</b>	<b>Sequence Analysis</b>	<b>41</b>
9.1	Access Pfam . . . . .	41
9.2	Parse MSA . . . . .	41
9.3	Sequences . . . . .	41
9.4	Analysis . . . . .	42
<b>10</b>	<b>Applications Tutorial</b>	<b>45</b>
10.1	Align PDB files . . . . .	45
10.2	ANM calculations . . . . .	45
10.3	PCA calculations . . . . .	46

---

## INTRODUCTION

*ProDy* is an application programming interface (API) designed for structure-based analysis of protein dynamics, in particular for inferring protein dynamics from large heterogeneous structural ensembles. It comes with several command line applications ([ProDy Applications](#)<sup>1</sup>) and graphical user interface for visualization ([Normal Mode Wizard](#)<sup>2</sup>). This tutorial shows core features of *ProDy* and some basic analysis tasks. You can find links to more detailed and advanced tutorials below.

### 1.1 Structural Ensemble Analysis

*ProDy* is primarily designed for analysis of *large* heterogeneous structural datasets for a protein composed of sequence homologs, mutants, or ligand bound structures that have with missing loops or terminal residues. Dominant patterns in structural variability are extracted by principal component analysis (PCA) of the ensemble. Helper functions allow for comparison of dynamics inferred from experiments with theoretical models and simulation data. For detailed usage examples see [Ensemble Analysis](#)<sup>3</sup>.

### 1.2 Elastic Network Models

*ProDy* can be used for normal mode analysis (NMA) of protein dynamics based on elastic network models (ENMs). Flexible classes allow for developing and using customized gamma functions in ENMs and numerous helper functions allow for comparative analysis of experimental and theoretical datasets. See [Elastic Network Models](#)<sup>4</sup> for detailed usage examples.

### 1.3 Trajectory Analysis

In addition to analysis of experimental data and theoretical models, *ProDy* can be used to analyze trajectories from molecular dynamics simulations, such as for performing essential dynamics analysis (EDA). *ProDy* supports DCD file format, but trajectories in other formats can be parsed using other Python packages and analyzed using *ProDy*. See [Trajectory Analysis](#)<sup>5</sup> for detailed usage examples.

### 1.4 Visualization

Finally, results from *ProDy* calculations can be visualized using NMWiz, which is a **VMD** plugin GUI. NMWiz can also be used for submitting *ProDy* calculations for molecules in VMD. See [NMWiz Tutorial](#)<sup>6</sup> for analysis of various types of datasets and visualization of protein dynamics.

---

<sup>1</sup><http://prody.csb.pitt.edu/manual/apps/prody/index.html#prody-apps>

<sup>2</sup>[http://prody.csb.pitt.edu/tutorials/nmwiz\\_tutorial/intro.html#nmwiz](http://prody.csb.pitt.edu/tutorials/nmwiz_tutorial/intro.html#nmwiz)

<sup>3</sup>[http://prody.csb.pitt.edu/tutorials/ensemble\\_analysis/index.html#ensemble-analysis](http://prody.csb.pitt.edu/tutorials/ensemble_analysis/index.html#ensemble-analysis)

<sup>4</sup>[http://prody.csb.pitt.edu/tutorials/enm\\_analysis/index.html#enm-analysis](http://prody.csb.pitt.edu/tutorials/enm_analysis/index.html#enm-analysis)

<sup>5</sup>[http://prody.csb.pitt.edu/tutorials/trajectory\\_analysis/index.html#trajectory-analysis](http://prody.csb.pitt.edu/tutorials/trajectory_analysis/index.html#trajectory-analysis)

<sup>6</sup>[http://prody.csb.pitt.edu/tutorials/nmwiz\\_tutorial/index.html#nmwiz-tutorial](http://prody.csb.pitt.edu/tutorials/nmwiz_tutorial/index.html#nmwiz-tutorial)



## HOW TO START

### 2.1 Using ProDy

ProDy can be used in a number of ways:

1. interactively in a Python shell,
2. as a command line program via [ProDy Applications](#)<sup>7</sup>,
3. from within VMD via [Normal Mode Wizard](#)<sup>8</sup>,
4. or as a toolkit for developing new software.

#### 2.1.1 Python for beginners

Familiarity with Python programming language will help when using *ProDy*. If you are new to Python, or to programming, you may start with one of the following tutorials:

- [The Python Tutorial](#)<sup>9</sup>
- [Python Scientific Lecture Notes](#)<sup>10</sup>
- [A Primer on Python for Life Science Researchers](#)<sup>11</sup>

### 2.2 Interactive Usage

In the rest of this tutorial, we assume that you will be typing commands in a Python shell. *ProDy* will automatically download PDB files and save them to current working directory, so you may want start Python from inside of a directory that you make for this tutorial:

```
$ mkdir prody_tutorial  
$ cd prody_tutorial
```

#### 2.2.1 Start Python shell

For best interactive usage experience, we strongly recommend that you use **IPython** instead of the standard Python shell. IPython shell provides many user-friendly features, such as dynamic introspection and help, and also convenient integration of **NumPy** and **Matplotlib**.

If you have installed IPython, type in:

---

<sup>7</sup><http://prody.csb.pitt.edu/manual/apps/prody/index.html#prody-apps>

<sup>8</sup>[http://prody.csb.pitt.edu/tutorials/nmwiz\\_tutorial/intro.html#nmwiz](http://prody.csb.pitt.edu/tutorials/nmwiz_tutorial/intro.html#nmwiz)

<sup>9</sup><http://docs.python.org/tutorial/>

<sup>10</sup><http://scipy-lectures.github.com/>

<sup>11</sup><https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2098836/>

```
$ ipython
```

If you also installed Matplotlib, use:

```
$ ipython --pylab
```

--pylab option will import Matplotlib and NumPy automatically, and is equivalent to the following:

```
In [1]: from pylab import *
In [2]: ion() # turn interactive mode on
```

If you don't have IPython yet, use:

```
$ python
```

On Windows, after you make the directory, make a Shift+right click in it in Windows Explorer and then select *Open command window here* option. Then start C:\Python27\python.exe. Alternatively, you may run **IDLE (Python GUI)** or **Python (command line)** from the start menu.

### 2.2.2 Import from ProDy

We import all *ProDy* functions and classes into the current namespace as follows:

```
In [3]: from prody import *
```

There are other ways to import *ProDy* contents. You may use `import prody as pd` and prefix all functions calls with `pd.`, if you prefer not to overcrowd the target namespace. Alternatively, if you want to use contents of a specific module, such as `prody.proteins`, you can use `from prody.proteins import *`. You should, however, avoid using `from prody.proteins.pdbfile import *`, because location of methods in submodules may change without notice.

## 2.3 Using Documentation

ProDy documentation is quite comprehensive and you can access it in a number of different ways. In interactive sessions, API reference can be accessed using the built-in Python function `help()`<sup>12</sup>:

```
help(select) # help on select module
help(fetchPDB) # help on parsePDB function
```

This function prints documentation on screen, and you will need to type `q` to exit from help view. If you are using the interactive Python shell (iPython), you can also get help using `?`:

### 2.3.1 Searching documentation

You can search entire documentation, including manual and tutorial pages, by typing in a keyword, function, or class name. Try searching for *selections* to get to [Atom Selections](#)<sup>13</sup>, for example.

<sup>12</sup><http://docs.python.org/library/functions.html#help>

<sup>13</sup><http://prody.csb.pitt.edu/manual/reference/atomic/select.html#selections>

### Toolbox

Show code snippets for copying:

Show code and output

Enter search terms or a module, class or function name:

selections Go

### 2.3.2 Copying code snippets

When reading online documentation, you can use *Show code* button on the right hand side panel to display only code snippets. From this view, you can copy code directly into a file, i.e. click *Select* and then `Ctrl+C` to have the text in your clipboard. To return to the documentation click the *Close* button.





## PRODY BASICS

We start with importing everything from ProDy package:

```
In [1]: from prody import *
In [2]: from pylab import *
In [3]: ion()
```

Functions and classes are named such that they should not create a conflict with any other package. In this part we will familiarize with different categories of functions and methods.

### 3.1 File Parsers

Let's start with parsing a protein structure and then keep working on that in this part. File parser function names are prefixed with `parse`. You can get a list of parser functions by pressing `TAB` after typing in `parse`:

```
In [4]: parse<TAB>
parseArray      parseCIFStream  parseEMD        parseHiC        parseMSA
parsePDBHeader  parsePQR        parseSTAR       parseChainsList parseDCD
parseEMDStream  parseHiCStream  parseNMD        parsePDBStream  parsePSF
parseSTRIDE     parseCIF        parseDSSP       parseHeatmap    parseModes
parsePDB        parsePfamPDBs  parseSparseMatrix
```

When using `parsePDB()`, usually an identifier will be sufficient, If corresponding file is found in the current working directory, it will be used, otherwise it will be downloaded from PDB servers.

Let's parse structure `:pdb:5uoj` of p38 MAP kinase (MAPK):

```
In [5]: p38 = parsePDB('5uoj') # returns an AtomGroup object
In [6]: p38 # typing in variable name will give some information
Out [6]: <AtomGroup: 5uoj (3138 atoms)>
```

We see that this structure contains 2962 atoms.

Now, similar to listing parser function names, we can use tab completion to inspect the `p38` object:

```
In [7]: p38.num<TAB>
p38.numAtoms    p38.numChains    p38.numFragments  p38.numSegments
p38.numBonds    p38.numCoordsets p38.numResidues
```

This action printed a list of methods with `num` prefix. Let's use some of them to get information on the structure:

```
In [8]: p38.numAtoms()
Out [8]: 3138
```

```
In [9]: p38.numCoordsets() # returns number of models
```

```
Out [9]: 1
```

```
In [10]: p38.numResidues() # water molecules also count as residues
```

```
Out [10]: 718
```

## 3.2 Analysis Functions

Similar to parsers, analysis function names start with `calc`:

```
In [11]: calc<TAB>
```

<code>calcADPAxes</code>	<code>calcChainsNormDistFluct</code>	<code>calcCrossProjection</code>
<code>calcDistFlucts</code>	<code>calcFractVariance</code>	<code>calcADPs</code>
<code>calcCollectivity</code>	<code>calcCumulOverlap</code>	<code>calcENM</code>
<code>calcGNM</code>	<code>calcAngle</code>	<code>calcCovariance</code>
<code>calcDeformVector</code>	<code>calcEnsembleENMs</code>	<code>calcGyradius</code>
<code>calcANM</code>	<code>calcCovOverlap</code>	<code>calcDihedral</code>
<code>calcEnsembleSpectralOverlaps</code>	<code>calcMeff</code>	<code>calcCenter</code>
<code>calcCrossCorr</code>	<code>calcDistance</code>	<code>calcEntropyTransfer</code>
<code>calcMSAOccupancy</code>	<code>calcMSF</code>	<code>calcPairDeformationDist</code>
<code>calcPsi</code>	<code>calcSignatureCollectivity</code>	<code>calcSpecDimension</code>
<code>calcOccupancies</code>	<code>calcPercentIdentities</code>	<code>calcRankorder</code>
<code>calcSignatureCrossCorr</code>	<code>calcSpectralOverlap</code>	<code>calcOmega</code>
<code>calcPerturbResponse</code>	<code>calcRMSD</code>	<code>calcSignatureFractVariance</code>
<code>calcSqFlucts</code>	<code>calcOverallNetEntropyTransfer</code>	<code>calcPhi</code>
<code>calcRMSF</code>	<code>calcSignatureOverlaps</code>	<code>calcSubspaceOverlap</code>
<code>calcOverlap</code>	<code>calcProjection</code>	<code>calcShannonEntropy</code>
<code>calcSignatureSqFlucts</code>	<code>calcTempFactors</code>	<code>calcTransformation</code>
<code>calcTree</code>		

Let's read documentation of `calcGyradius()` function and use it to calculate the radius of gyration of p38 MAPK structure:

## 3.3 Plotting Functions

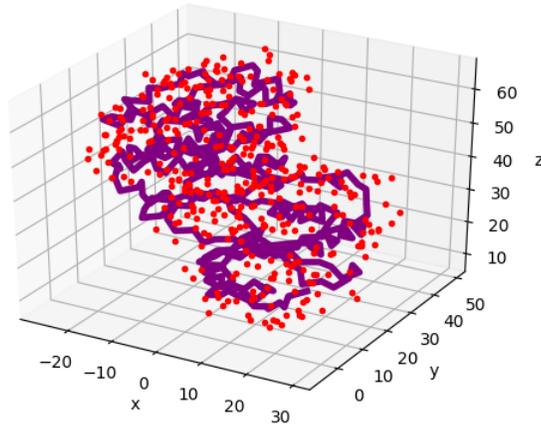
Likewise, plotting function names have `show` prefix and here is a list of them:

```
In [12]: show<TAB>
```

<code>showAlignment</code>	<code>showCrossProjection</code>	<code>showDomainBar</code>
<code>showHeatmap</code>	<code>showMeanMechStiff</code>	<code>showNormDistFunct</code>
<code>showAtomicLines</code>	<code>showCumulFractVars</code>	<code>showDomains</code>
<code>showLines</code>	<code>showMechStiff</code>	<code>showNormedSqFlucts</code>
<code>showAtomicMatrix</code>	<code>showCumulOverlap</code>	<code>showEllipsoid</code>
<code>showLinkage</code>	<code>showMode</code>	<code>showOccupancies</code>
<code>showContactMap</code>	<code>showDiffMatrix</code>	<code>showEmbedding</code>
<code>showMap</code>	<code>showMSAOccupancy</code>	<code>showOverlap</code>
<code>showCrossCorr</code>	<code>showDirectInfoMatrix</code>	<code>showFractVars</code>
<code>showMatrix</code>	<code>showMutinfoMatrix</code>	<code>showOverlaps</code>
<code>showOverlapTable</code>	<code>showScaledSqFlucts</code>	<code>showSignatureCollectivity</code>
<code>showSignatureSqFlucts</code>	<code>showVarianceBar</code>	<code>showPairDeformationDist</code>
<code>showSCAMatrix</code>	<code>showSignatureCrossCorr</code>	<code>showSignatureVariances</code>
<code>showPerturbResponse</code>	<code>showShannonEntropy</code>	<code>showSignatureDistribution</code>
<code>showSqFlucts</code>	<code>showProjection</code>	<code>showSignatureID</code>
<code>showSignatureMode</code>	<code>showTree</code>	<code>showProtein</code>
<code>showSignatureAtomicLines</code>	<code>showSignatureOverlaps</code>	<code>showTree_networkx</code>

We can use `showProtein()` function to make a quick plot of p38 structure:

```
In [13]: showProtein(p38);
```



This of course does not compare to any visualization software that you might be familiar with, but it comes handy to see what you are dealing with.

## 3.4 Protein Structures

Protein structures (`.pdb` or `.cif` files) will be the standard input for most *ProDy* calculations, so it is good to familiarize with ways to access and manage PDB file resources.

### 3.4.1 Fetching PDB files

First of all, *ProDy* downloads PDB files when needed (these are compressed on the PDB webservice). If you prefer saving decompressed files, you can use `fetchPDB()` function as follows:

```
In [14]: fetchPDB('5uoj', compressed=False)
Out [14]: '5uoj.pdb'
```

Note that *ProDy* functions that fetch files or output files return filename upon successful completion of the task. You can use this behavior to shorten the code you need to write, e.g.:

```
In [15]: parsePDB(fetchPDB('5uoj', compressed=False)) # same as p38 parsed above
Out [15]: <AtomGroup: 5uoj (3138 atoms)>
```

We downloaded and save an uncompressed PDB file, and parsed it immediately.

### 3.4.2 PDB file resources

Secondly, *ProDy* can manage local mirrors of the PDB server or a local PDB folder, as well as using a server close to your physical location for downloads:

- One of the [wwPDB](http://www.wwpdb.org/)<sup>14</sup> FTP servers in US, Europe or Japan can be picked for downloads using `wwPDBServer()`.

<sup>14</sup><http://www.wwpdb.org/>

- A local PDB mirror can be set for faster access to files using `pathPDBMirror()`.
- A local folder can be set for storing downloaded files for future access using `pathPDBFolder()`.

If you are in the Americas now, you can choose the PDB server in the US as follows:

```
In [16]: wwPDBServer('us')
```

If you would like to have a central folder, such as `~/Downloads/pdb`, for storing downloaded PDB files (you will need to make it), do as follows:

```
In [17]: mkdir ~/Downloads/pdb;
```

```
In [18]: pathPDBFolder('~/Downloads/pdb')
```

Note that when these functions are used, ProDy will save your settings in `.prodyrc` file stored in your home folder.

## 3.5 Atom Groups

As you might have noticed, `parsePDB()` function returns structure data as an `AtomGroup` object. Let's see for `p38` variable from above:

```
In [19]: p38
```

```
Out [19]: <AtomGroup: 5uoj (3138 atoms)>
```

You can also parse a list of `.pdb` files into a list of `AtomGroup` objects:

```
In [20]: ags = parsePDB('5uoj', '3h5v')
```

```
In [21]: ags
```

```
Out [21]: [<AtomGroup: 5uoj (3138 atoms)>, <AtomGroup: 3h5v (9392 atoms)>]
```

If you want to provide a list object you need to provide an asterisk (\*) to let Python know this is a set of input arguments:

```
In [22]: pdb_ids = ['5uoj', '3h5v']
```

```
In [23]: ags = parsePDB(pdb_ids)
```

```
In [24]: ags
```

```
Out [24]: [<AtomGroup: 5uoj (3138 atoms)>, <AtomGroup: 3h5v (9392 atoms)>]
```

Data from this object can be retrieved using `get` methods. For example:

```
In [25]: p38.getResnames()
```

```
Out [25]: array(['ARG', 'ARG', 'ARG', ..., 'HOH', 'HOH', 'HOH'], dtype='<S6')>
```

```
In [26]: p38.getCoords()
```

```
Out [26]: array([[ 25.325,   3.794,  22.831],
 [ 24.258,   4.528,  22.091],
 [ 23.399,   3.547,  21.279],
 ...,
 [ -1.774,  28.702,  32.891],
 [ -1.648,  31.544,  31.756],
 [-25.711,  35.476,  55.232]])
```

To get a list of all methods use tab completion, i.e. `p38.<TAB>`. We will learn more about atom groups in the following chapters.

### 3.5.1 Indexing

An individual Atom can be accessed by indexing AtomGroup objects:

```
In [27]: atom = p38[0]
In [28]: atom
Out[28]: <Atom: N from 5uoj (index 0)>
```

Note that all get/set functions defined for AtomGroup instances are also defined for Atom instances, using singular form of the function name.

```
In [29]: atom.getResname()
Out[29]: 'ARG'
```

### 3.5.2 Slicing

It is also possible to get a slice of an AtomGroup. For example, we can get every other atom as follows:

```
In [30]: p38[::2]
Out[30]: <Selection: 'index 0:3138:2' from 5uoj (1569 atoms)>
```

Or, we can get the first 10 atoms, as follows:

```
In [31]: p38[:10]
Out[31]: <Selection: 'index 0:10:1' from 5uoj (10 atoms)>
```

### 3.5.3 Hierarchical view

You can also access specific chains or residues in an atom group. Indexing by a single letter identifier will return a Chain instance:

```
In [32]: p38['A']
Out[32]: <Chain: A from 5uoj (718 residues, 3138 atoms)>
```

Indexing atom group with a chain identifier and a residue number will return Residue instance:

```
In [33]: p38['A', 100]
Out[33]: <Residue: ASN 100 from Chain A from 5uoj (8 atoms)>
```

See [Atomic classes](#)<sup>15</sup> for details of indexing atom groups and *Hierarchical Views* for more on hierarchical views.

## 3.6 ProDy Verbosity

Finally, you might have noticed that ProDy prints some information to the console after parsing a file or doing some calculations. For example, PDB parser will print what was parsed and how long it took to the screen:

```
@> 5uoj (./5uoj.pdb.gz) is found in the target directory.
@> 2962 atoms and 1 coordinate sets were parsed in 0.08s.
```

This behavior is useful in interactive sessions, but may be problematic for automated tasks as the messages are printed to stderr. The level of verbosity can be controlled using `confProDy()` function, and calling it as `confProDy(verbosity='none')` will stop all information messages permanently.

<sup>15</sup><http://prody.csb.pitt.edu/manual/reference/atomic/index.html#atomic>



## ATOM GROUPS

The example below shows how to build an `AtomGroup` from scratch. We start by importing everything from the ProDy package and the NumPy package:

```
In [1]: from prody import *
In [2]: from numpy import *
In [3]: from pylab import *
In [4]: ion()
```

### 4.1 Building an Atom Group

The best way to start constructing an `AtomGroup` is by setting the coordinates first. Number of atoms will be automatically set according to the size of the coordinate data array:

```
In [5]: wtr1 = AtomGroup('Water')
In [6]: wtr1
Out[6]: <AtomGroup: Water (0 atoms; no coordinates)>
In [7]: coords = array([[1, 0, 0], [0, 0, 0], [0, 0, 1]], dtype=float)
In [8]: coords
Out[8]:
array([[1., 0., 0.],
       [0., 0., 0.],
       [0., 0., 1.]])
In [9]: wtr1.setCoords(coords)
In [10]: wtr1
Out[10]: <AtomGroup: Water (3 atoms)>
```

#### 4.1.1 Attributes

Attributes must be passed in a list or an array whose size is the same as the number of atoms.

```
In [11]: wtr1.setNames(['H', 'O', 'H'])
In [12]: wtr1.setResnums([1, 1, 1])
In [13]: wtr1.setResnames(['WAT', 'WAT', 'WAT'])
```

Accessing data will return a copy of the data. This means you can manipulate the return values safely without affecting the original values. For example:

```
In [14]: names = wtr1.getNames()
In [15]: names[0] = names[0] + '1'
In [16]: names
Out[16]: array(['H1', 'O', 'H'], dtype='<S6')

```

```
In [17]: wtr1.getNames()
Out[17]: array(['H', 'O', 'H'], dtype='<S6')

```

### 4.1.2 Atoms

Atoms are represented by instance of `Atom`.

#### Iteration

Atoms in an `AtomGroup` can be iterated over

```
In [18]: for a in wtr1:
.....:     print(a)
.....:
Atom H (index 0)
Atom O (index 1)
Atom H (index 2)

```

#### Indexing

Atoms in an atom group can be accessed via indexing:

```
In [19]: a = wtr1[0]
In [20]: a
Out[20]: <Atom: H from Water (index 0)>

```

```
In [21]: a.getCoords()
Out[21]: array([1., 0., 0.])

```

### 4.1.3 Coordinate sets

Let's add another coordinate set to the atom group:

```
In [22]: wtr1.addCoordset(array([[0, 1, 0], [0, 0, 0], [0, 0, 1.1]], dtype=float))
In [23]: wtr1
Out[23]: <AtomGroup: Water (3 atoms; active #0 of 2 coordsets)>

```

Note that the number of coordinate sets is now 2, but the active coordinate set index is still 0. Active coordinate set index can be changed for `AtomGroup` as follows:

```
In [24]: a.setACSIndex(1)
In [25]: a
Out[25]: <Atom: H from Water (index 0; active #1 of 2 coordsets)>

```

Changing active coordinate set for an atom group, does not affect the active coordinate set of the atom group:

```
In [26]: wtr1
Out [26]: <AtomGroup: Water (3 atoms; active #0 of 2 coordsets)>
```

Coordinates for the atom group will be returned from the active coordinate set

```
In [27]: a.getCoords()
Out [27]: array([0., 1., 0.])
```

```
In [28]: a.setACSIndex(0)

In [29]: a.getCoords()
Out [29]: array([1., 0., 0.])
```

## Iterations

Coordinate sets can also be iterated over for Atom and AtomGroup instances:

```
In [30]: for xyz in a.iterCoordsets():
        ....:     print(xyz)
        ....:
[1. 0. 0.]
[0. 1. 0.]
```

### 4.1.4 Copying and Merging

Now let's make another copy of this water:

```
In [31]: wtr2 = wtr1.copy()

In [32]: wtr2
Out [32]: <AtomGroup: Water (3 atoms; active #0 of 2 coordsets)>
```

#### Translate copy

Let's translate the coordinates of wtr2 so that it does not overlap with wtr1

```
In [33]: wtr2.setCoords(wtr2.getCoords() + 2)

In [34]: wtr2.getCoords()
Out [34]:
array([[3., 2., 2.],
       [2., 2., 2.],
       [2., 2., 3.]])
```

Above operation only translated the coordinate set at index 0

```
In [35]: wtr2.setACSIndex(1)

In [36]: wtr2.getCoords()
Out [36]:
array([[0. , 1. , 0. ],
       [0. , 0. , 0. ],
       [0. , 0. , 1.1]])

In [37]: wtr2.setCoords(wtr2.getCoords() + 2) # translate the 2nd coordset as well
```

#### Change attributes

Before we merge wtr1 and wtr2, let's change resid's of wtr2:

```
In [38]: wtr2.setResnums( [2, 2, 2] )
```

```
In [39]: wtr2.getResnums()
```

```
Out [39]: array([2, 2, 2])
```

We can do this in an alternate way too:

```
In [40]: wtr2.select('all').setResnums(2)
```

```
In [41]: wtr2.getResnums()
```

```
Out [41]: array([2, 2, 2])
```

## Merge two copies

Let's merge the two water atom groups:

```
In [42]: wtrs = wtr1 + wtr2
```

```
In [43]: wtrs
```

```
Out [43]: <AtomGroup: Water + Water (6 atoms; active #0 of 2 coordsets)>
```

```
In [44]: wtrs.getCoords()
```

```
Out [44]:
```

```
array([[1., 0., 0.],
       [0., 0., 0.],
       [0., 0., 1.],
       [3., 2., 2.],
       [2., 2., 2.],
       [2., 2., 3.]])
```

```
In [45]: wtrs.getNames()
```

```
Out [45]: array(['H', 'O', 'H', 'H', 'O', 'H'], dtype='|S6')
```

```
In [46]: wtrs.getResnums()
```

```
Out [46]: array([1, 1, 1, 2, 2, 2])
```

## 4.1.5 Hierarchical views

Hierarchical views of atom groups are represented by `HierView`.

Residues (and also chains) in an atom group can also be iterated over

```
In [47]: for res in wtrs.getHierView().iterResidues():
.....:     print(res)
.....:
```

```
WAT 1
WAT 2
```

## 4.1.6 Renaming an atom group

Finally, it's possible to change the name of `wtrs` from "Water + Water" to something shorter:

```
In [48]: wtrs.setTitle('2Waters')
```

```
In [49]: wtrs
```

```
Out [49]: <AtomGroup: 2Waters (6 atoms; active #0 of 2 coordsets)>
```

## 4.2 Storing data in AtomGroup

Now let's get an atom group from a PDB file:

```
In [50]: structure = parsePDB('5uoj')
```

In addition to what's in a PDB file, you can store arbitrary atomic attributes in `AtomGroup` objects.

### 4.2.1 Set a new attribute

For the purposes of this example, we will manufacture atomic data by dividing the residue number of each atom by 10:

```
In [51]: myresnum = structure.getResnums() / 10.0
```

We will add this to the atom group using `AtomGroup.setData()` method by passing a name for the attribute and the data:

```
In [52]: structure.setData('myresnum', myresnum)
```

We can check if a custom atomic attribute exists using `AtomGroup.isDataLabel()` method:

```
In [53]: structure.isDataLabel('myresnum')
Out[53]: True
```

### 4.2.2 Access subset of data

Custom attributes can be accessed from selections:

```
In [54]: calpha = structure.calpha
```

```
In [55]: calpha.getData('myresnum')
```

```
Out[55]:
```

```
array([ 0.5,  0.6,  0.7,  0.8,  0.9,  1. ,  1.1,  1.2,  1.3,  1.4,  1.5,
        1.6,  1.7,  1.8,  1.9,  2. ,  2.1,  2.2,  2.3,  2.4,  2.5,  2.6,
        2.7,  2.8,  2.9,  3. ,  3.1,  3.2,  3.3,  3.4,  3.5,  3.6,  3.7,
        3.8,  3.9,  4. ,  4.1,  4.2,  4.3,  4.4,  4.5,  4.6,  4.7,  4.8,
        4.9,  5. ,  5.1,  5.2,  5.3,  5.4,  5.5,  5.6,  5.7,  5.8,  5.9,
        6. ,  6.1,  6.2,  6.3,  6.4,  6.5,  6.6,  6.7,  6.8,  6.9,  7. ,
        7.1,  7.2,  7.3,  7.4,  7.5,  7.6,  7.7,  7.8,  7.9,  8. ,  8.1,
        8.2,  8.3,  8.4,  8.5,  8.6,  8.7,  8.8,  8.9,  9. ,  9.1,  9.2,
        9.3,  9.4,  9.5,  9.6,  9.7,  9.8,  9.9, 10. , 10.1, 10.2, 10.3,
       10.4, 10.5, 10.6, 10.7, 10.8, 10.9, 11. , 11.1, 11.2, 11.3, 11.4,
       11.5, 11.6, 11.7, 11.8, 11.9, 12. , 12.1, 12.2, 12.3, 12.4, 12.5,
       12.6, 12.7, 12.8, 12.9, 13. , 13.1, 13.2, 13.3, 13.4, 13.5, 13.6,
       13.7, 13.8, 13.9, 14. , 14.1, 14.2, 14.3, 14.4, 14.5, 14.6, 14.7,
       14.8, 14.9, 15. , 15.1, 15.2, 15.3, 15.4, 15.5, 15.6, 15.7, 15.8,
       15.9, 16. , 16.1, 16.2, 16.3, 16.4, 16.5, 16.6, 16.7, 16.8, 16.9,
       17. , 17.1, 17.2, 17.9, 18. , 18.1, 18.2, 18.3, 18.4, 18.5, 18.6,
       18.7, 18.8, 18.9, 19. , 19.1, 19.2, 19.3, 19.4, 19.5, 19.6, 19.7,
       19.8, 19.9, 20. , 20.1, 20.2, 20.3, 20.4, 20.5, 20.6, 20.7, 20.8,
       20.9, 21. , 21.1, 21.2, 21.3, 21.4, 21.5, 21.6, 21.7, 21.8, 21.9,
       22. , 22.1, 22.2, 22.3, 22.4, 22.5, 22.6, 22.7, 22.8, 22.9, 23. ,
       23.1, 23.2, 23.3, 23.4, 23.5, 23.6, 23.7, 23.8, 23.9, 24. , 24.1,
       24.2, 24.3, 24.4, 24.5, 24.6, 24.7, 24.8, 24.9, 25. , 25.1, 25.2,
       25.3, 25.4, 25.5, 25.6, 25.7, 25.8, 25.9, 26. , 26.1, 26.2, 26.3,
       26.4, 26.5, 26.6, 26.7, 26.8, 26.9, 27. , 27.1, 27.2, 27.3, 27.4,
       27.5, 27.6, 27.7, 27.8, 27.9, 28. , 28.1, 28.2, 28.3, 28.4, 28.5,
       28.6, 28.7, 28.8, 28.9, 29. , 29.1, 29.2, 29.3, 29.4, 29.5, 29.6,
```

```
29.7, 29.8, 29.9, 30. , 30.1, 30.2, 30.3, 30.4, 30.5, 30.6, 30.7,  
30.8, 30.9, 31. , 31.1, 31.2, 31.3, 31.4, 31.5, 31.6, 31.7, 31.8,  
31.9, 32. , 32.1, 32.2, 32.3, 32.4, 32.5, 32.6, 32.7, 32.8, 32.9,  
33. , 33.1, 33.2, 33.3, 33.4, 33.5, 33.6, 33.7, 33.8, 33.9, 34. ,  
34.1, 34.2, 34.3, 34.4, 34.5, 34.6, 34.7, 34.8, 34.9, 35. , 35.1,  
35.2, 35.3])
```

### 4.2.3 Make selections

Custom atomic attributes can be used in selections:

```
In [56]: mysel = structure.select('0 < myresnum and myresnum < 10')  
In [57]: mysel  
Out [57]: <Selection: '0 < myresnum and myresnum < 10' from 5uoj (779 atoms)>
```

This gives the same result as the following selection:

```
In [58]: structure.select('0 < resnum and resnum < 100') == mysel  
Out [58]: True
```

### 4.2.4 Save attributes

It is not possible to save custom attributes in PDB files, but `saveAtoms()` function can be used them to save in disk for later use:

```
In [59]: saveAtoms(structure)  
Out [59]: '5uoj.ag.npz'
```

Let's load it using `loadAtoms()` function:

```
In [60]: structure = loadAtoms('5uoj.ag.npz')  
In [61]: structure.getData('myresnum')  
Out [61]: array([ 0.5,  0.5,  0.5, ..., 77.3, 77.4, 77.5])
```

### 4.2.5 Delete an attribute

Finally, when done with an attribute, it can be deleted using `AtomGroup.delData()` method:

```
In [62]: structure.delData('myresnum')  
Out [62]: array([ 0.5,  0.5,  0.5, ..., 77.3, 77.4, 77.5])
```

## ATOM SELECTIONS

This part gives more information on properties of `AtomGroup` objects. We start with making necessary imports. Note that every documentation page contains them so that the code within the can be executed independently. You can skip them if you have already done them in a Python session.

```
In [1]: from prody import *
In [2]: from pylab import *
In [3]: ion()
```

### 5.1 Atom Selections

`AtomGroup` instances have a plain view of atoms for efficiency, but they are coupled with a powerful atom selection engine. You can get well defined atom subsets by passing simple keywords or make rather sophisticated selections using composite statements. Selection keywords and grammar are very much similar to those found in **VMD\_**. Some examples are shown here:

#### 5.1.1 Keyword selections

Now, we parse a structure. This could be any structure, one that you know well from your research, for example.

```
In [4]: structure = parsePDB('5uoj')
In [5]: protein = structure.select('protein')
In [6]: protein
Out [6]: <Selection: 'protein' from 5uoj (2763 atoms)>
```

Using the "protein" keyword we selected 2833 atoms out of 2962 atoms. `Atomic.select()` method returned a `Selection` instance. Note that all get and set methods defined for the `AtomGroup` objects are also defined for `Selection` objects. For example:

```
In [7]: protein.getResnames()
Out [7]: array(['ARG', 'ARG', 'ARG', ..., 'LEU', 'LEU', 'LEU'], dtype='<S6')>
```

#### 5.1.2 Select by name/type

We can select backbone atoms by passing atom names following "name" keyword:

```
In [8]: backbone = structure.select('protein and name N CA C O')
In [9]: backbone
Out [9]: <Selection: 'protein and name N CA C O' from 5uoj (1372 atoms)>
```

Alternatively, we can use "backbone" to make the same selection:

```
In [10]: backbone = structure.select('backbone')
```

We select acidic and basic residues by using residue names with "resname" keyword:

```
In [11]: charged = structure.select('resname ARG LYS HIS ASP GLU')
In [12]: charged
Out [12]: <Selection: 'resname ARG LYS HIS ASP GLU' from 5uoj (843 atoms)>
```

Alternatively, we can use predefined keywords "acidic" and "basic".

```
In [13]: charged = structure.select('acidic or basic')
In [14]: charged
Out [14]: <Selection: 'acidic or basic' from 5uoj (843 atoms)>
In [15]: set(charged.getResnames())
Out [15]: {'ARG', 'ASP', 'GLU', 'HIS', 'LYS'}
```

### 5.1.3 Composite selections

Let's try a more sophisticated selection. We first calculate the geometric center of the protein atoms using `calcCenter()` function. Then, we select the  $C\alpha$  and  $C\beta$  atoms of residues that have at least one atom within 10 Å away from the geometric center.

```
In [16]: center = calcCenter(protein).round(3)
In [17]: center
Out [17]: array([ 0.918, 17.423, 40.248])
In [18]: sel = structure.select('protein and name CA CB and same residue as '
    ....:                        '((x-1)**2 + (y-17.5)**2 + (z-40.0)**2)**0.5 < 10')
    ....:
In [19]: sel
Out [19]: <Selection: 'protein and nam...)**2)**0.5 < 10' from 5uoj (70 atoms)>
```

Alternatively, this selection could be done as follows:

```
In [20]: sel = structure.select('protein and name CA CB and same residue as '
    ....:                        'within 10 of center', center=center)
    ....:
In [21]: sel
Out [21]: <Selection: 'index 567 570 5... 1631 1645 1648' from 5uoj (70 atoms)>
```

### 5.1.4 Selections simplified

In interactive sessions, an alternative to typing in `.select('protein')` or `.select('backbone')` is using dot operator:

```
In [22]: protein = structure.protein
In [23]: protein
Out [23]: <Selection: 'protein' from 5uoj (2763 atoms)>
```

You can use dot operator multiple times:

```
In [24]: bb = structure.protein.backbone
In [25]: bb
Out [25]: <Selection: '(backbone) and (protein)' from 5uoj (1372 atoms)>
```

This may go on and on:

```
In [26]: ala_ca = structure.protein.backbone.resname_ALA.alpha
In [27]: ala_ca
Out [27]: <Selection: '(alpha) and ((...and (protein)))' from 5uoj (26 atoms)>
```

### 5.1.5 More examples

There is much more to what you can do with this flexible and fast atom selection engine, without the need for writing nested loops with comparisons or changing the source code. See the following pages:

- [Atom Selections](#)<sup>16</sup> for description of all selection keywords
- [Intermolecular Contacts](#)<sup>17</sup> for selecting interacting atoms

## 5.2 Operations on Selections

Selection objects can be used with bitwise operators:

### 5.2.1 Union

Let's select  $\beta$ -carbon atoms for non-GLY amino acid residues, and  $\alpha$ -carbons for GLYs in two steps:

```
In [28]: betas = structure.select('name CB and protein')
In [29]: len(betas)
Out [29]: 328
In [30]: gly_alphas = structure.select('name CA and resname GLY')
In [31]: len(gly_alphas)
Out [31]: 15
```

The above shows that the p38 structure contains 15 GLY residues.

These two selections can be combined as follows:

```
In [32]: betas_gly_alphas = betas | gly_alphas
In [33]: betas_gly_alphas
Out [33]: <Selection: '(name CB and pr...nd resname GLY)' from 5uoj (343 atoms)>
In [34]: len(betas_gly_alphas)
Out [34]: 343
```

The selection string for the union of selections becomes:

```
In [35]: betas_gly_alphas.getSelstr()
Out [35]: '(name CB and protein) or (name CA and resname GLY)'
```

<sup>16</sup><http://prody.csb.pitt.edu/manual/reference/atomic/select.html#selections>

<sup>17</sup>[http://prody.csb.pitt.edu/tutorials/structure\\_analysis/contacts.html#contacts](http://prody.csb.pitt.edu/tutorials/structure_analysis/contacts.html#contacts)

Note that it is also possible to yield the same selection using selection string (name CB and protein) or (name CA and resname GLY).

### 5.2.2 Intersection

It is as easy to get the intersection of two selections. Let's find charged and medium size residues in a protein:

```
In [36]: charged = structure.select('charged')
In [37]: charged
Out[37]: <Selection: 'charged' from 5uoj (843 atoms)>
In [38]: medium = structure.select('medium')
In [39]: medium
Out[39]: <Selection: 'medium' from 5uoj (720 atoms)>
```

```
In [40]: medium_charged = medium & charged
In [41]: medium_charged
Out[41]: <Selection: '(medium) and (charged)' from 5uoj (192 atoms)>
In [42]: medium_charged.getSelstr()
Out[42]: '(medium) and (charged)'
```

Let's see which amino acids are considered charged and medium:

```
In [43]: set(medium_charged.getResnames())
Out[43]: {'ASP'}
```

What about amino acids that are medium or charged:

```
In [44]: set((medium | charged).getResnames())
Out[44]: {'ARG', 'ASN', 'ASP', 'CYS', 'GLU', 'HIS', 'LYS', 'PRO', 'THR', 'VAL'}
```

### 5.2.3 Inversion

It is also possible to invert a selection:

```
In [45]: only_protein = structure.select('protein')
In [46]: only_protein
Out[46]: <Selection: 'protein' from 5uoj (2763 atoms)>
In [47]: only_non_protein = ~only_protein
In [48]: only_non_protein
Out[48]: <Selection: 'not (protein)' from 5uoj (375 atoms)>
In [49]: water = structure.select('water')
In [50]: water
Out[50]: <Selection: 'water' from 5uoj (375 atoms)>
```

The above shows that 5uoj does not contain any non-water hetero atoms.

### 5.2.4 Addition

Another operation defined on the `Select` object is addition (also on other `AtomPointer` derived classes).

This may be useful if you want to yield atoms in an AtomGroup in a specific order. Let's think of a simple case, where we want to output atoms in 5uoj in a specific order:

```
In [51]: protein = structure.select('protein')
In [52]: water = structure.select('water')
In [53]: water_protein = water + protein
In [54]: writePDB('5uoj_water_protein.pdb', water_protein)
Out[54]: '5uoj_water_protein.pdb'
```

In the resulting file, the water atoms will precedes the protein atoms.

## 5.2.5 Membership

Selections also allows membership test operations:

```
In [55]: backbone = structure.select('protein')
In [56]: calpha = structure.select('calpha')
```

Is calpha<sup>18</sup> a subset of backbone<sup>19</sup>?

```
In [57]: calpha in backbone
Out[57]: True
```

Or, is water in protein selection?

```
In [58]: water in protein
Out[58]: False
```

Other tests include:

```
In [59]: protein in structure
Out[59]: True

In [60]: backbone in structure
Out[60]: True

In [61]: structure in structure
Out[61]: True

In [62]: calpha in calpha
Out[62]: True
```

## 5.2.6 Equality

You can also check the equality of selections. Comparison will return True if both selections refer to the same atoms.

```
In [63]: calpha = structure.select('protein and name CA')
In [64]: calpha2 = structure.select('calpha')
In [65]: calpha == calpha2
Out[65]: True
```

<sup>18</sup><http://prody.csb.pitt.edu/manual/reference/atomic/flags.html#term-calpha>

<sup>19</sup><http://prody.csb.pitt.edu/manual/reference/atomic/flags.html#term-backbone>



## HIERARCHICAL VIEWS

This part describes how to use hierarchical views. We start by importing everything from the ProDy package:

```
In [1]: from prody import *
```

### 6.1 Hierarchical Views

Then we parse a structure to get an `AtomGroup` instance which has a plain view of atoms:

```
In [2]: structure = parsePDB('3mkb')
In [3]: structure
Out [3]: <AtomGroup: 3mkb (4776 atoms)>
```

A hierarchical view of the structure can be simply get by calling the `AtomGroup.getHierView()` method:

```
In [4]: hv = structure.getHierView()
In [5]: hv
Out [5]: <HierView: AtomGroup 3mkb (4 chains, 946 residues)>
```

#### 6.1.1 Indexing

Indexing `HierView` instances return `Chain`:

```
In [6]: hv['A']
Out [6]: <Chain: A from 3mkb (254 residues, 1198 atoms)>
In [7]: hv['B']
Out [7]: <Chain: B from 3mkb (216 residues, 1193 atoms)>
In [8]: hv['Z'] # This will return None, which means chain Z does not exist
```

The length of the `hv` variable gives the number of chains in the structure:

```
In [9]: len(hv)
Out [9]: 4
In [10]: hv.numChains()
Out [10]: 4
```

It is also possible to get a `Residue` by directly indexing the `HierView` instance:

```
In [11]: hv['A', 100]
Out [11]: <Residue: MET 100 from Chain A from 3mkb (8 atoms)>
```

Insertion codes can also be passed:

```
In [12]: hv['A', 100, 'B']
```

This does not return anything, since residue 100B does not exist.

### 6.1.2 Iterations

One can iterate over HierView instances to get chains:

```
In [13]: for chain in hv:
.....:     print(chain)
.....:
Chain A
Chain B
Chain C
Chain D
```

It is also possible to get a list () of chains simply as follows:

```
In [14]: chains = list(hv)

In [15]: chains
Out[15]:
[<Chain: A from 3mkb (254 residues, 1198 atoms)>,
 <Chain: B from 3mkb (216 residues, 1193 atoms)>,
 <Chain: C from 3mkb (245 residues, 1189 atoms)>,
 <Chain: D from 3mkb (231 residues, 1196 atoms)>]
```

### 6.1.3 Residues

In addition, one can also iterate over all residues:

```
In [16]: for i, residue in enumerate(hv.iterResidues()):
.....:     if i == 4: break
.....:     print(residue)
.....:
ALA 1
PHE 2
THR 3
GLY 4
```

## 6.2 Chains

We obtain a Chain object for chain A by indexing the HierView object.

```
In [17]: chA = hv['A']

In [18]: chA
Out[18]: <Chain: A from 3mkb (254 residues, 1198 atoms)>
```

The length of the chain is equal to the number of residues in it:

```
In [19]: len(chA)
Out[19]: 254

In [20]: chA.numResidues()
Out[20]: 254
```

## 6.2.1 Indexing

Indexing a Chain instance with a scalar returns an Atom like indexing an Atomgroup:

```
In [21]: chA[1]
Out [21]: <Atom: CA from 3mkb (index 1)>

In [22]: chA[1000]
Out [22]: <Atom: C from 3mkb (index 1000)>
```

Indexing a Chain instance with a tuple returns an Residue. This can also be achieved with the `prody.atomic.chain.Chain.getResidue()`<sup>20</sup> method.

```
In [23]: chA[(1,)]
Out [23]: <Residue: ALA 1 from Chain A from 3mkb (5 atoms)>

In [24]: chA.getResidue(1)
Out [24]: <Residue: ALA 1 from Chain A from 3mkb (5 atoms)>
```

The use of a tuple is necessary to allow inclusion of insertion codes. In this case there aren't any so None is returned.

```
In [25]: chA[1, 'A'] # Residue 1 with insertion code A also does not exist
```

## 6.2.2 Iterations

Iterating over a chain also yields residues:

```
In [26]: for i, residue in enumerate(chA):
.....:     if i == 4: break
.....:     print(residue)
.....:
ALA 1
PHE 2
THR 3
GLY 4
```

Note that water atoms, each constituting a residue, are also part of a chain if they are labeled with that chain's identifier.

This enables getting a list () of residues simply as follows:

```
In [27]: chA_residues = list(chA)

In [28]: chA_residues[:4]
Out [28]:
[<Residue: ALA 1 from Chain A from 3mkb (5 atoms)>,
<Residue: PHE 2 from Chain A from 3mkb (11 atoms)>,
<Residue: THR 3 from Chain A from 3mkb (7 atoms)>,
<Residue: GLY 4 from Chain A from 3mkb (4 atoms)>]

In [29]: chA_residues[-4:]
Out [29]:
[<Residue: HOH 471 from Chain A from 3mkb (1 atoms)>,
<Residue: HOH 485 from Chain A from 3mkb (1 atoms)>,
<Residue: HOH 490 from Chain A from 3mkb (1 atoms)>,
<Residue: HOH 493 from Chain A from 3mkb (1 atoms)>]
```

## 6.2.3 Get data

All methods defined for AtomGroup class are also defined for Chain and Residue classes:

<sup>20</sup><http://prody.csb.pitt.edu/manual/reference/atomic/chain.html#prody.atomic.chain.Chain.getResidue>

```
In [30]: chA.getCoords()
Out [30]:
array([[ -2.139,  17.026, -13.287],
       [ -1.769,  15.572, -13.111],
       [ -0.296,  15.257, -13.467],
       ...,
       [ -5.843,  17.181, -16.86 ],
       [-13.199,  -9.21 , -49.692],
       [ -0.459,   0.378, -46.156]])

In [31]: chA.getBetas()
Out [31]: array([59.35, 59.14, 58.5 , ..., 57.79, 47.77, 40.77])
```

## 6.2.4 Selections

Finally, you can select atoms from a Chain instance:

```
In [32]: chA_backbone = chA.select('backbone')

In [33]: chA_backbone
Out [33]: <Selection: '(backbone) and (chain A)' from 3mkb (560 atoms)>

In [34]: chA_backbone.getSelstr()
Out [34]: '(backbone) and (chain A)'
```

As you see, the selection string passed by the user is augmented with “chain” keyword and identifier automatically to provide internal consistency:

```
In [35]: structure.select(chA_backbone.getSelstr())
Out [35]: <Selection: '(backbone) and (chain A)' from 3mkb (560 atoms)>
```

## 6.3 Residues

```
In [36]: chA_res1 = chA.getResidue(1)

In [37]: chA_res1
Out [37]: <Residue: ALA 1 from Chain A from 3mkb (5 atoms)>
```

### 6.3.1 Indexing

Residue instances can be indexed to get individual atoms:

```
In [38]: chA_res1['CA']
Out [38]: <Atom: CA from 3mkb (index 1)>

In [39]: chA_res1['CB']
Out [39]: <Atom: CB from 3mkb (index 4)>

In [40]: chA_res1['X'] # if atom does not exist, None is returned
```

### 6.3.2 Iterations

Iterating over a residue instance yields Atom instances:

```
In [41]: for i, atom in enumerate(chA_res1):
.....:     if i == 4: break
```

```

.....:     print(atom)
.....:
Atom N (index 0)
Atom CA (index 1)
Atom C (index 2)
Atom O (index 3)

```

This makes it easy to get a `list()` of atoms:

```

In [42]: list(chA_res1)
Out [42]:
[<Atom: N from 3mkb (index 0)>,
 <Atom: CA from 3mkb (index 1)>,
 <Atom: C from 3mkb (index 2)>,
 <Atom: O from 3mkb (index 3)>,
 <Atom: CB from 3mkb (index 4)>]

```

### 6.3.3 Get data

All methods defined for `AtomGroup` class are also defined for `Residue` class:

```

In [43]: chA_res1.getCoords()
Out [43]:
array([[ -2.139,  17.026, -13.287],
       [ -1.769,  15.572, -13.111],
       [ -0.296,  15.257, -13.467],
       [  0.199,  14.155, -13.155],
       [ -2.752,  14.639, -13.898]])

In [44]: chA_res1.getBetas()
Out [44]: array([59.35, 59.14, 58.5 , 59.13, 59.02])

```

### 6.3.4 Selections

Finally, you can select atoms from a `Residue` instance:

```

In [45]: chA_res1_bb = chA_res1.select('backbone')

In [46]: chA_res1_bb
Out [46]: <Selection: '(backbone) and ... and (chain A)' from 3mkb (4 atoms)>

In [47]: chA_res1_bb.getSelstr()
Out [47]: '(backbone) and (resnum 1 and (chain A))'

```

Again, the selection string is augmented with the chain identifier and residue number ([resnum<sup>21</sup>](http://prody.csb.pitt.edu/manual/reference/atomic/fields.html#term-resnum)).

## 6.4 Atoms

The lowest level of the hierarchical view contains `Atom` instances.

```

In [48]: chA_res1_CA = chA_res1['CA']

In [49]: chA_res1_CA
Out [49]: <Atom: CA from 3mkb (index 1)>

```

<sup>21</sup><http://prody.csb.pitt.edu/manual/reference/atomic/fields.html#term-resnum>

### Get atomic data

All methods defined for `AtomGroup` class are also defined for `Atom` class with the difference that method names are singular (except for coordinates):

```
In [50]: chA_res1_CA.getCoords()
Out[50]: array([-1.769,  15.572, -13.111])

In [51]: chA_res1_CA.getBeta()
Out[51]: 59.14
```

## 6.5 State Changes

A `HierView` instance represents the state of an `AtomGroup` instance at the time it is built. When chain identifiers or residue numbers change, the state that hierarchical view represents may not match the current state of the atom group:

```
In [52]: chA.setChid('X')

In [53]: chA
Out[53]: <Chain: X from 3mkb (254 residues, 1198 atoms)>

In [54]: hv['X'] # returns None, since hierarchical view is not updated

In [55]: hv.update() # this updates hierarchical view

In [56]: hv['X']
Out[56]: <Chain: X from 3mkb (254 residues, 1198 atoms)>
```

When this is the case, `HierView.update()` method can be used to update hierarchical view.

## STRUCTURE ANALYSIS

ProDy comes with many functions that can be used to calculate structural properties and compare structures. We demonstrate only some of these functions. For more detailed examples, see [Structure Analysis](#)<sup>22</sup> tutorial.

```
In [1]: from prody import *
In [2]: from pylab import *
In [3]: ion()
```

### 7.1 Measure geometric properties

Let's parse a structure:

```
In [4]: p38 = parsePDB('5uoj')
```

Functions for analyzing structures can be found in `measure`<sup>23</sup> module. For example, you can calculate phi ( $\phi$ ) and psi ( $\psi$ ) for the 10th residue, or the radius of gyration of the protein as follows:

```
In [5]: calcPhi(p38[10,])
Out[5]: -124.86176953710802

In [6]: calcPsi(p38[10,])
Out[6]: 132.5985032413337

In [7]: calcGyradius(p38)
Out[7]: 22.31823191315965
```

### 7.2 Compare and align structures

You can also compare different structures using some of the methods in `proteins` module. Let's parse another p38 MAP kinase structure

```
In [8]: bound = parsePDB('1zz2')
```

You can find similar chains in structure 5uoj and 1zz2 using `matchChains()` function:

```
In [9]: apo_chA, bnd_chA, seqid, overlap = matchChains(p38.protein, bound.protein)[0]
In [10]: apo_chA
Out[10]: <AtomMap: Chain A from 5uoj -> Chain A from 1zz2 from 5uoj (337 atoms)>
```

---

<sup>22</sup>[http://prody.csb.pitt.edu/tutorials/structure\\_analysis/index.html#structure-analysis](http://prody.csb.pitt.edu/tutorials/structure_analysis/index.html#structure-analysis)

<sup>23</sup><http://prody.csb.pitt.edu/manual/reference/measure/index.html#module-prody.measure>

```
In [11]: bnd_chA
Out[11]: <AtomMap: Chain A from 1zz2 -> Chain A from 5uoj from 1zz2 (337 atoms)>

In [12]: seqid
Out[12]: 99.40652818991099

In [13]: overlap
Out[13]: 98
```

Matching  $C\alpha$  atoms are selected and returned as `AtomMap` instances. We can use them to calculate RMSD and superpose structures.

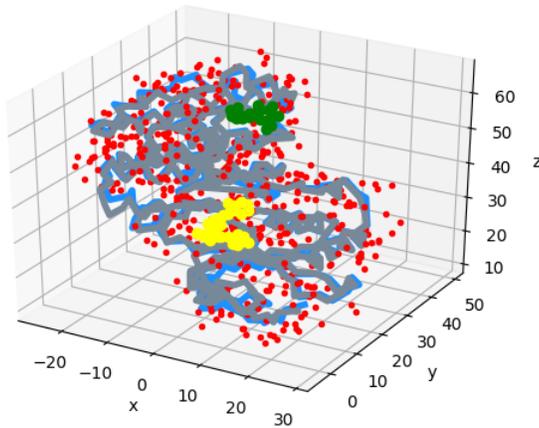
```
In [14]: calcRMSD(bnd_chA, apo_chA)
Out[14]: 72.93561517918914

In [15]: bnd_chA, transformation = superpose(bnd_chA, apo_chA)

In [16]: calcRMSD(bnd_chA, apo_chA)
Out[16]: 1.8248053285019694
```

```
In [17]: showProtein(p38);

In [18]: showProtein(bound);
```



## 7.3 Writing PDB files

PDB files can be written using the `writePDB()` function. The function accepts objects containing or referring to atomic data.

Output selected atoms:

```
In [19]: writePDB('5uoj_calphas.pdb', p38.select('calpha'))
Out[19]: '5uoj_calphas.pdb'
```

Output a chain:

```
In [20]: chain_A = p38['A']  
In [21]: writePDB('5uoj_chain_A.pdb', chain_A)  
Out [21]: '5uoj_chain_A.pdb'
```

As you may have noticed, this function returns the file name after it is successfully written. This is a general behavior for ProDy output functions. For more PDB writing examples see [Write PDB file](#)<sup>24</sup>.

---

<sup>24</sup>[http://prody.csb.pitt.edu/tutorials/structure\\_analysis/pdbfiles.html#writepdb](http://prody.csb.pitt.edu/tutorials/structure_analysis/pdbfiles.html#writepdb)



## DYNAMICS ANALYSIS

In this section, we will show how to perform quick PCA and ANM analysis using a solution structure of [:wiki:'Ubiquitin'](#). If you started a new Python session, import ProDy contents:

```
In [1]: from prody import *
In [2]: from pylab import *
In [3]: ion()
```

### 8.1 PCA Calculations

Let's perform principal component analysis (PCA) of an ensemble of NMR models, such as [:pdb:'2k39'](#). The same example is presented in the [Ensemble Analysis](#)<sup>25</sup> where more details are provided.

First, we prepare the ensemble:

```
In [4]: ubi = parsePDB('2k39', subset='calpha')
In [5]: ubi_selection = ubi.select('resnum < 71')
In [6]: ubi_ensemble = Ensemble(ubi_selection)
In [7]: ubi_ensemble.iterpose()
```

Then, we perform the PCA calculations:

```
In [8]: pca = PCA('Ubiquitin')
In [9]: pca.buildCovariance(ubi_ensemble)
In [10]: pca.calcModes()
```

This analysis provides us with a description of the dominant changes in the structural ensemble. Let's see the fraction of variance for top ranking 4 PCs:

```
In [11]: for mode in pca[:4]:
.....:     print(calcFractVariance(mode).round(2))
.....:
0.13
0.09
0.08
0.07
```

---

<sup>25</sup>[http://prody.csb.pitt.edu/tutorials/ensemble\\_analysis/index.html#ensemble-analysis](http://prody.csb.pitt.edu/tutorials/ensemble_analysis/index.html#ensemble-analysis)

PCA data can be saved on disk using `saveModel()` function:

```
In [12]: saveModel(pca)
Out [12]: 'Ubiquitin.pca.npz'
```

This function writes data in binary format, so is an efficient way of storing data permanently. In a later session, this data can be loaded using `loadModel()` function.

## 8.2 ANM Calculations

Anisotropic network model (ANM) analysis can be performed in two ways:

The shorter way, which may be suitable for interactive sessions:

```
In [13]: anm, atoms = calcANM(ubi_selection, selstr='calpha')
```

The longer and more controlled way:

```
In [14]: anm = ANM('ubi') # instantiate ANM object
In [15]: anm.buildHessian(ubi_selection) # build Hessian matrix for selected atoms
In [16]: anm.calcModes() # calculate normal modes
In [17]: saveModel(anm)
Out [17]: 'ubi.anm.npz'
```

Anisotropic Network Model (ANM)<sup>26</sup> provides a more detailed discussion of ANM calculations. The above longer way gives more control to the user. For example, instead of building the Hessian matrix using uniform force constant and cutoff distance, customized force constant functions (see [Custom Gamma Functions](#)<sup>27</sup>) or a pre-calculated matrix (see `ANM.setHessian()`) may be used.

Individual Mode instances can be accessed by indexing the ANM instance:

```
In [18]: slowest_mode = anm[0]
In [19]: print( slowest_mode )
Mode 1 from ANM ubi
In [20]: print( slowest_mode.getEigval().round(3) )
1.714
```

Note that indices in Python start from zero (0). 0th mode is the 1st non-zero mode in this case. Let's confirm that normal modes are orthogonal to each other:

```
In [21]: (anm[0] * anm[1]).round(10)
Out [21]: -0.0
In [22]: (anm[0] * anm[2]).round(10)
Out [22]: 0.0
```

As you might have noticed, multiplication of two modes is nothing but the `dot()` product of mode vectors/arrays. See [Normal Mode Algebra](#)<sup>28</sup> for more examples.

<sup>26</sup>[http://prody.csb.pitt.edu/tutorials/enm\\_analysis/anm.html#anm](http://prody.csb.pitt.edu/tutorials/enm_analysis/anm.html#anm)

<sup>27</sup>[http://prody.csb.pitt.edu/tutorials/enm\\_analysis/gamma.html#gamma](http://prody.csb.pitt.edu/tutorials/enm_analysis/gamma.html#gamma)

<sup>28</sup>[http://prody.csb.pitt.edu/tutorials/enm\\_analysis/normalmodes.html#normalmode-operations](http://prody.csb.pitt.edu/tutorials/enm_analysis/normalmodes.html#normalmode-operations)

## 8.3 Comparative Analysis

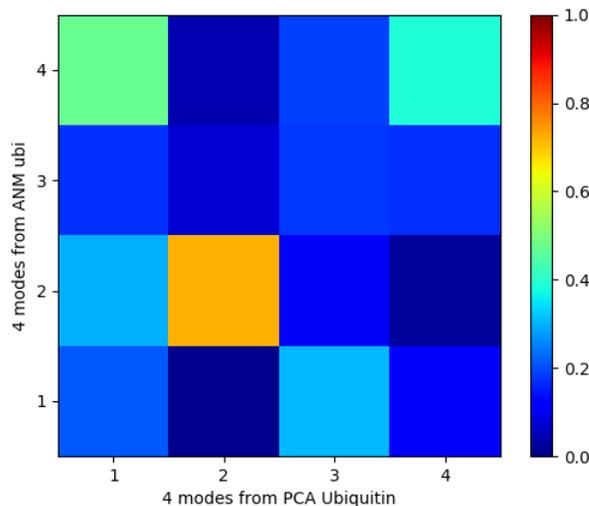
*ProDy* comes with many built-in functions to facilitate a comparative analysis of experimental and theoretical data. For example, using `printOverlapTable()` function you can see the agreement between experimental (PCA) modes and theoretical (ANM) modes calculated above:

```
In [23]: printOverlapTable(pca[:4], anm[:4])
```

```
Overlap Table
          ANM ubi
          #1     #2     #3     #4
PCA Ubiquitin #1 -0.21 +0.30 -0.17 -0.47
PCA Ubiquitin #2 +0.01 +0.72 +0.08 +0.05
PCA Ubiquitin #3 +0.31 +0.11 +0.18 +0.19
PCA Ubiquitin #4 +0.11 -0.02 -0.17 -0.39
```

Output above shows that PCA mode 2 and ANM mode 2 for ubiquitin show the highest overlap (cosine-correlation).

```
In [24]: showOverlapTable(pca[:4], anm[:4]);
```



This was a short example for a simple case. [Ensemble Analysis](#)<sup>29</sup> section contains more comprehensive examples for heterogeneous datasets. [Analysis](#)<sup>30</sup> shows more analysis function usage examples and [Dynamics Analysis](#)<sup>31</sup> module documentation lists all of the analysis functions.

## 8.4 Output Data Files

The `writeNMD()` function writes PCA results in NMD format. NMD files can be viewed using the [Normal Mode Wizard](#)<sup>32</sup> VMD plugin.

```
In [25]: writeNMD('ubi_pca.nmd', pca[:3], ubi_selection)
```

```
Out [25]: 'ubi_pca.nmd'
```

Additionally, results can be written in plain text files for analysis with other programs using the `writeArray()` function:

<sup>29</sup>[http://prody.csb.pitt.edu/tutorials/ensemble\\_analysis/index.html#pca](http://prody.csb.pitt.edu/tutorials/ensemble_analysis/index.html#pca)

<sup>30</sup>[http://prody.csb.pitt.edu/tutorials/ensemble\\_analysis/xray\\_analysis.html#pca-xray-analysis](http://prody.csb.pitt.edu/tutorials/ensemble_analysis/xray_analysis.html#pca-xray-analysis)

<sup>31</sup><http://prody.csb.pitt.edu/manual/reference/dynamics/index.html#dynamics>

<sup>32</sup>[http://prody.csb.pitt.edu/tutorials/nmwiz\\_tutorial/intro.html#nmwiz](http://prody.csb.pitt.edu/tutorials/nmwiz_tutorial/intro.html#nmwiz)

```
In [26]: writeArray('ubi_pca_modes.txt', pca.getArray(), format='%8.3f')
Out[26]: 'ubi_pca_modes.txt'
```

## 8.5 External Data

Normal mode data from other NMA, EDA, or PCA programs can be parsed using `parseModes()` function for analysis.

In this case, we will parse ANM modes for p38 MAP Kinase calculated using [ANM server](#)<sup>33</sup> as the external software. We use `oanm_eigvals.txt` and `oanm_slwevs.txt` files from the ANM server.

You can either download these files to your current working directory from [here](#) or obtain them for another protein from the ANM server.

```
In [27]: nma = parseModes(normalmodes='oanm_slwevs.txt',
....: eigenvalues='oanm_eigvals.txt',
....: nm_usecols=range(1,21), ev_usecols=[1], ev_usevalues=range(6,26))
....:
```

```
In [28]: nma
```

```
Out[28]: <NMA: oanm_slwevs (20 modes; 351 atoms)>
```

```
In [29]: nma.setTitle('5uoj ANM')
```

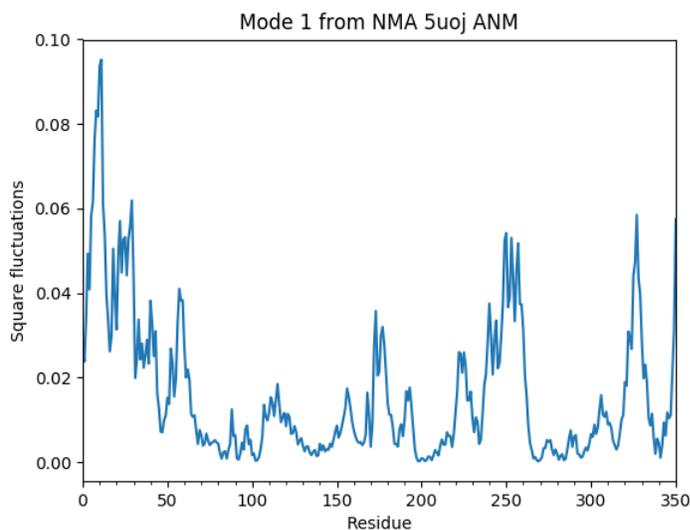
```
In [30]: slowmode = nma[0]
```

```
In [31]: print(slowmode.getEigval().round(2))
0.18
```

## 8.6 Plotting Data

If you have [Matplotlib](#), you can use functions whose name start with `show` to plot data:

```
In [32]: showSqFlucts(slowmode);
```



<sup>33</sup><http://ignmtest.cccb.pitt.edu/cgi-bin/anm/anm1.cgi>

Plotting<sup>34</sup> shows more plotting examples and Dynamics Analysis<sup>35</sup> module documentation lists all of the plotting functions.

## 8.7 More Examples

For more examples and details see Elastic Network Models<sup>36</sup> and Ensemble Analysis<sup>37</sup> tutorials.

---

<sup>34</sup>[http://prody.csb.pitt.edu/tutorials/ensemble\\_analysis/xray\\_plotting.html#pca-xray-plotting](http://prody.csb.pitt.edu/tutorials/ensemble_analysis/xray_plotting.html#pca-xray-plotting)

<sup>35</sup><http://prody.csb.pitt.edu/manual/reference/dynamics/index.html#dynamics>

<sup>36</sup>[http://prody.csb.pitt.edu/tutorials/enm\\_analysis/index.html#enm-analysis](http://prody.csb.pitt.edu/tutorials/enm_analysis/index.html#enm-analysis)

<sup>37</sup>[http://prody.csb.pitt.edu/tutorials/ensemble\\_analysis/index.html#ensemble-analysis](http://prody.csb.pitt.edu/tutorials/ensemble_analysis/index.html#ensemble-analysis)



## SEQUENCE ANALYSIS

Evol component of ProDy package brought new fast, flexible, and efficient features for handling multiple sequence alignments and analyzing sequence evolution. Here, we just give a brief introduction to these features. For more detailed examples, see [Evol Tutorial](#)<sup>38</sup>.

```
In [1]: from prody import *
In [2]: from pylab import *
In [3]: ion()
```

### 9.1 Access Pfam

First, let's fetch an MSA file from Pfam<sup>39</sup> database:

```
In [4]: filename = fetchPfamMSA('pkinase', alignment='seed')
In [5]: filename
```

We downloaded the seed alignment for Protein Kinase (**:pfam:'Pkinase'**) family.

### 9.2 Parse MSA

As you might guess, we will parse this file using `parseMSA()` function:

```
In [6]: msa = parseMSA('pkinase_seed.sth')
In [7]: msa
Out[7]: <MSA: pkinase_seed (38 sequences, 419 residues)>
```

### 9.3 Sequences

You can access Sequence objects by indexing MSA:

```
In [8]: seq = msa[0]
In [9]: seq
Out[9]: <Sequence: TTK_HUMAN (pkinase_seed[0]; length 419; 267 residues and 152 gaps)>
```

---

<sup>38</sup>[http://prody.csb.pitt.edu/tutorials/evol\\_tutorial/index.html#evol-tutorial](http://prody.csb.pitt.edu/tutorials/evol_tutorial/index.html#evol-tutorial)

<sup>39</sup><http://pfam.xfam.org>

```
In [10]: print(seq)
YSILKQIGSGGSSKV..FQVLN.EKKQIYAIKYVNLEEADNQTL.....DSYRNEIAYLNKLQQ.HSDKIIRLYDYEIT.....DQYIY..
```

You can also slice MSA objects and iterate over sequences:

```
In [11]: for seq in msa[:4]:
.....:     print(repr(seq))
.....:
<Sequence: TTK_HUMAN (pkinase_seed[0]; length 419; 267 residues and 152 gaps)>
<Sequence: MKK1_YEAST (pkinase_seed[1]; length 419; 268 residues and 151 gaps)>
<Sequence: STE7_YEAST (pkinase_seed[2]; length 419; 276 residues and 143 gaps)>
<Sequence: BYR1_SCHPO (pkinase_seed[3]; length 419; 255 residues and 164 gaps)>
```

## 9.4 Analysis

Evol component includes several functions for calculating conservation and coevolution properties of amino acids, which are shown in [Evol Tutorial](#)<sup>40</sup>. Here, let's take a look at `calcMSAOccupancy()` and `showMSAOccupancy()` functions:

```
In [12]: occ = calcMSAOccupancy(msa, count=True)
```

```
In [13]: occ.min()
```

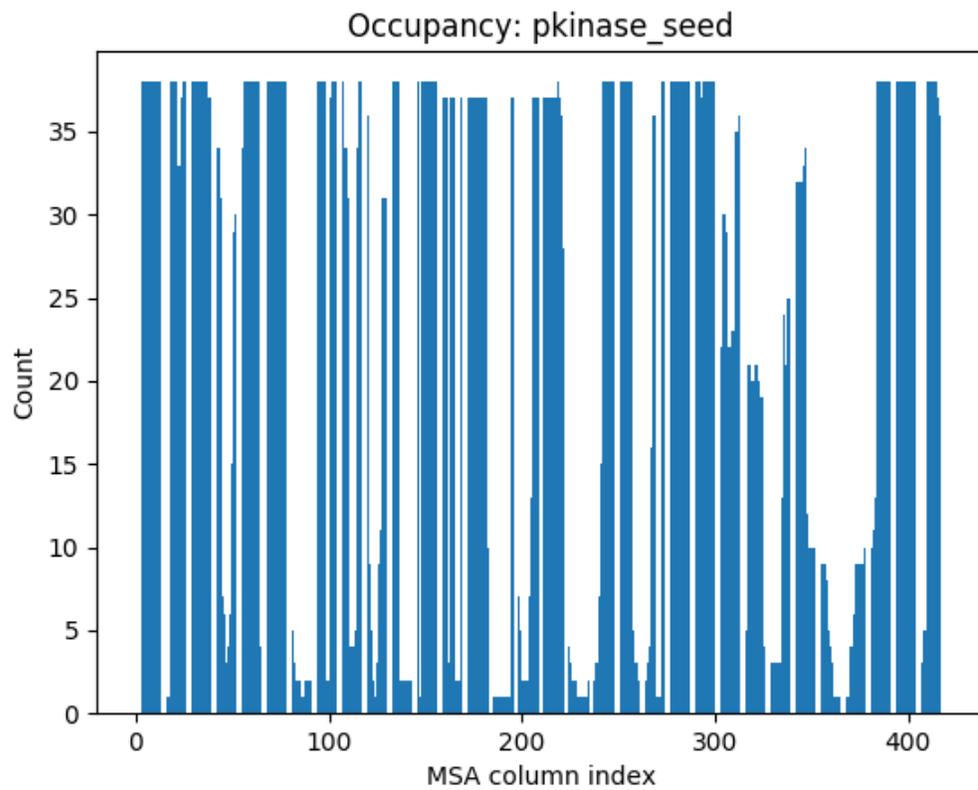
```
Out [13]: 1.0
```

This shows that an amino acid is present only in one of the sequences in the MSA.

```
In [14]: showMSAOccupancy(msa, count=True);
```

---

<sup>40</sup>[http://prody.csb.pitt.edu/tutorials/evol\\_tutorial/index.html#evol-tutorial](http://prody.csb.pitt.edu/tutorials/evol_tutorial/index.html#evol-tutorial)



You see that many residues are not present in all sequences. You will see how to refine such MSA instances in [Evol Tutorial](#)<sup>41</sup>.

<sup>41</sup>[http://prody.csb.pitt.edu/tutorials/evol\\_tutorial/index.html#evol-tutorial](http://prody.csb.pitt.edu/tutorials/evol_tutorial/index.html#evol-tutorial)



## APPLICATIONS TUTORIAL

You can use [ProDy Applications](#)<sup>42</sup> to perform some automated tasks, such as ANM/GNM/PCA calculations, fetching and aligning PDB files, making atom selections, or identify contacts. ProDy applications are handled by a script that comes with all installation packages. You can run the script from a central location such as `/usr/local/bin`:

```
$ prody -h
```

or from the current working directory:

```
$ ./prody -h
```

or:

```
$ prody -h
```

or on Windows:

```
$ C:\Python27\Scripts\prody -h
```

These lines will print available ProDy applications. You can get more help on a specific commands as follows:

```
$ prody anm -h
```

### 10.1 Align PDB files

`prody align`<sup>43</sup> command can be used to download and align structures for given PDB identifiers:

```
$ prody align 5uoj 1r39 1zz2
```

Structures will be automatically downloaded from wwPDB FTP servers and saved in the current working directory. Additionally, you can configure ProDy to use a local mirror of PDB or to store downloaded files in a local folder. See *ProDy Basics* part of the tutorial.

### 10.2 ANM calculations

`prody anm`<sup>44</sup> can be used to perform ANM calculations:

```
$ prody anm 5uoj -a -A
```

`-a` and `-A` options will make ProDy output all data and figure files.

---

<sup>42</sup><http://prody.csb.pitt.edu/manual/apps/prody/index.html#prody-apps>

<sup>43</sup><http://prody.csb.pitt.edu/manual/apps/prody/align.html#prody-align>

<sup>44</sup><http://prody.csb.pitt.edu/manual/apps/prody/anm.html#prody-anm>

## 10.3 PCA calculations

`prody pca`<sup>45</sup> can be used to perform PCA calculations. The following example will perform PCA calculations for  $C\alpha$  atoms of the p38 MAP kinase using files:

- ProDy Tutorial files (ZIP)
- ProDy Tutorial files (TGZ)

```
$ tar -xzf p38_trajectory.tar.gz
$ prody pca -a -A --select calpha --pdb p38.pdb p38_100frames.dcd
```

### Acknowledgments

Continued development of Protein Dynamics Software *ProDy* and associated programs is partially supported by the NIH<sup>46</sup>-funded Biomedical Technology and Research Center (BTRC) on *High Performance Computing for Multiscale Modeling of Biological Systems* (MMBios<sup>47</sup>) (P41 GM103712).

---

<sup>45</sup><http://prody.csb.pitt.edu/manual/apps/prody/pca.html#prody-pca>

<sup>46</sup><http://www.nih.gov/>

<sup>47</sup><http://mmbios.org/>