Project PyPlot

First, Solve the problem. Then, write the code.

-John Johnson

ABSTRACT

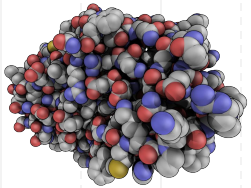
Project PyPlot is a simple implementation of multiple libraries made using the language Python. It connects to the mySQL database and using the link in the database, downloads the pdb file and renders it in PyMol software. It further displays several information related to the protein.

TARGET A

Part A of the project aims at successfully unifying the PDB localDatabase and the PyMol Software using a common Python Script

TARGET B

The second part aims at optimizing the usability of the protein data for various outputs obtained from the individual PDB file of the protein

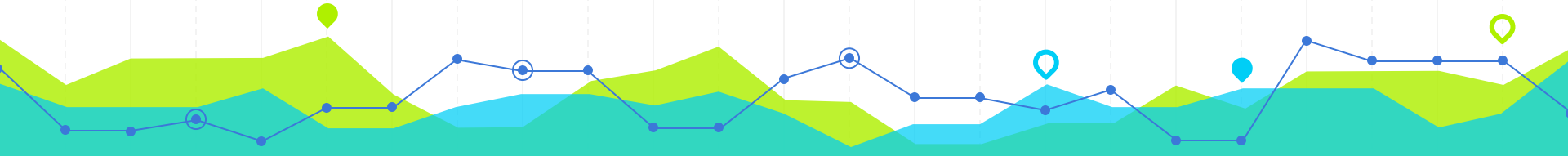


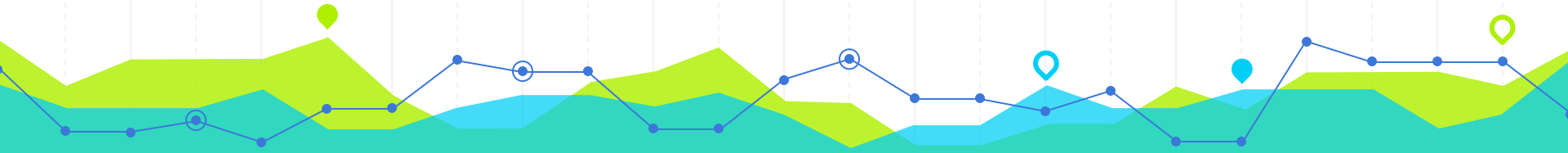
+

MySQL™

PDB

HELLO!





KNOW OUR PROJECT

Let’s start with the first set of slides

1

LET'S REVIEW SOME TERMINOLOGIES

mySQLdb
mySQLdb is an thread-compatible interface to the popular MySQL database server that provides the Python database API
Usage in Project: Create Database

URLib2
The urllib2 module provides an updated API for using internet resources identified by URLs.
Usage in project: HTTP requests

PyMol
PyMOL is an open-source, user-sponsored, molecular visualization system now owned by Schrodinger Inc.

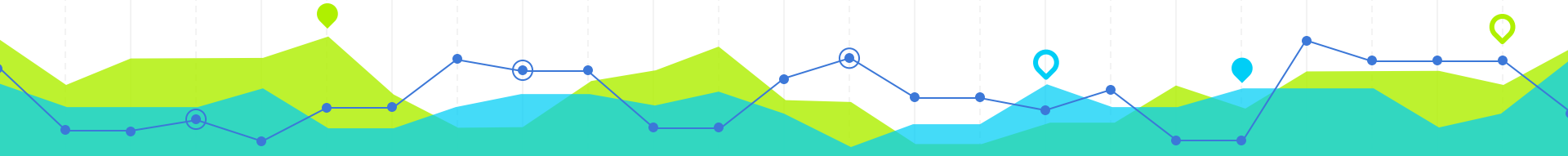
Protein
Proteins are large biomolecules, consisting of one or more long chains of amino acid residues. They perform a vast variety of functions.

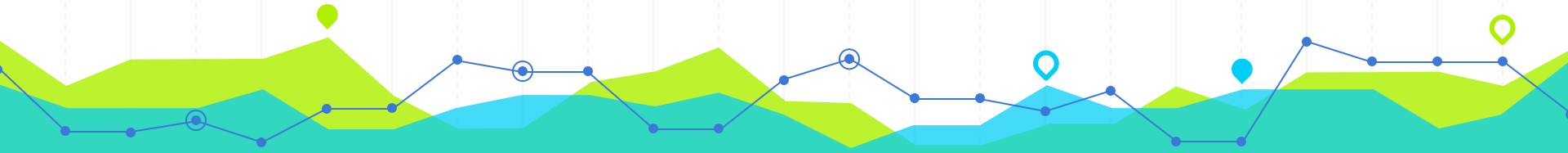
PDB
It is the textual file format describing the 3D structures of molecules held in the Protein Data Bank.

PDBtools
Set of scripts/framework/softwares which help to manipulate and optimize the data provided in the PDB files.

Protein Data Bank

- The Protein Data Bank (PDB) is a crystallographic database for the three-dimensional structural data of large biological molecules, such as proteins and nucleic acids. The data, typically obtained by X-ray crystallography, NMR spectrocopy, are freely accessible on the Internet via the websites of its member organisations (PDBe, PDBj, and RCSB). The PDB is overseen by an organization called the Worldwide Protein Data Bank, wwPDB.

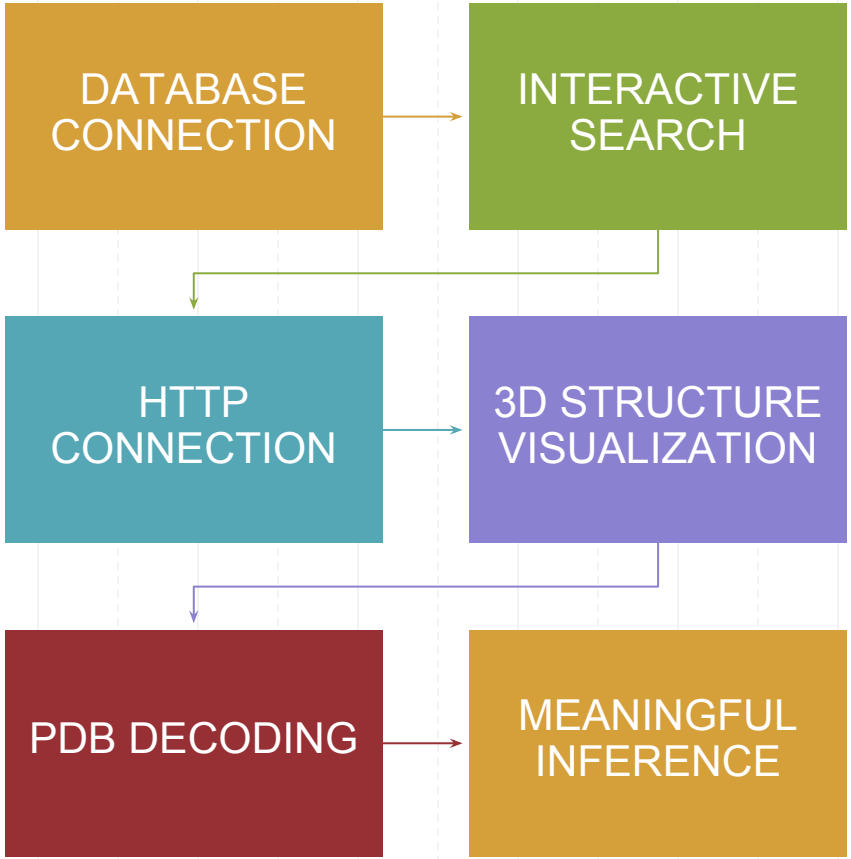




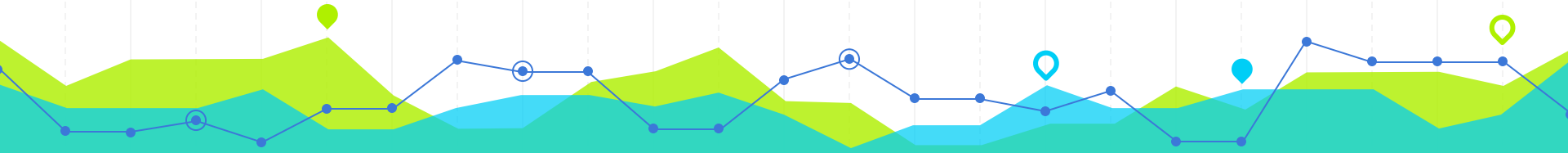
WORKFLOW

2

Let’s look at the flow of our program



OUR PROCESS IS EASY, BUT IT HAS A HUGE POTENTIAL FOR AGGRESSIVE EXPANSION

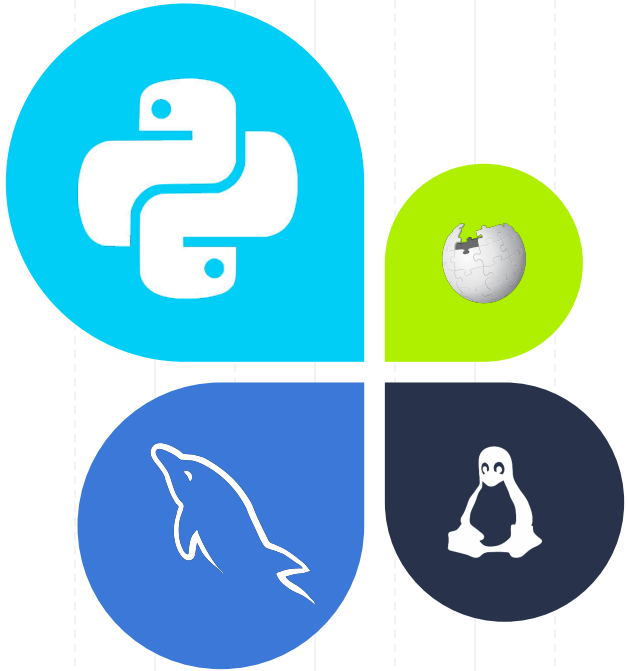


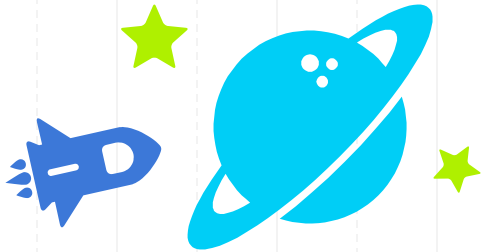
SOFTWARE TOUR

3

Let’s Understand the need of OpenSource in today’s emerging technological world.

USED SOFTWARES IN OUR PROJECT



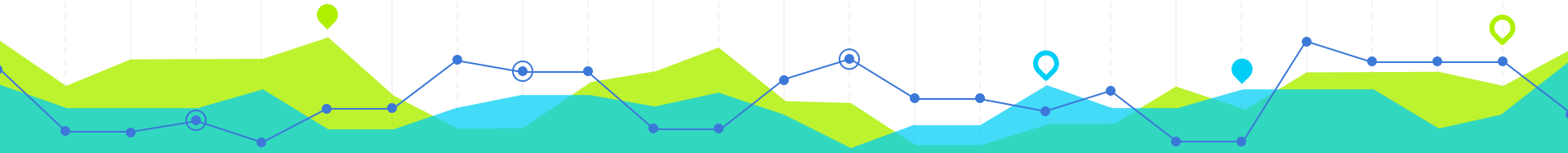


PYMOL

OpenSource . UserSponsored . Molecular Visualization System.

POWER OF PYMOL SOFTWARE

- Perform Rendering of complex protein structures
- Powerful atom-atom selection from the entire Amino Acid sequence.
- Identification of various atoms present in the protein

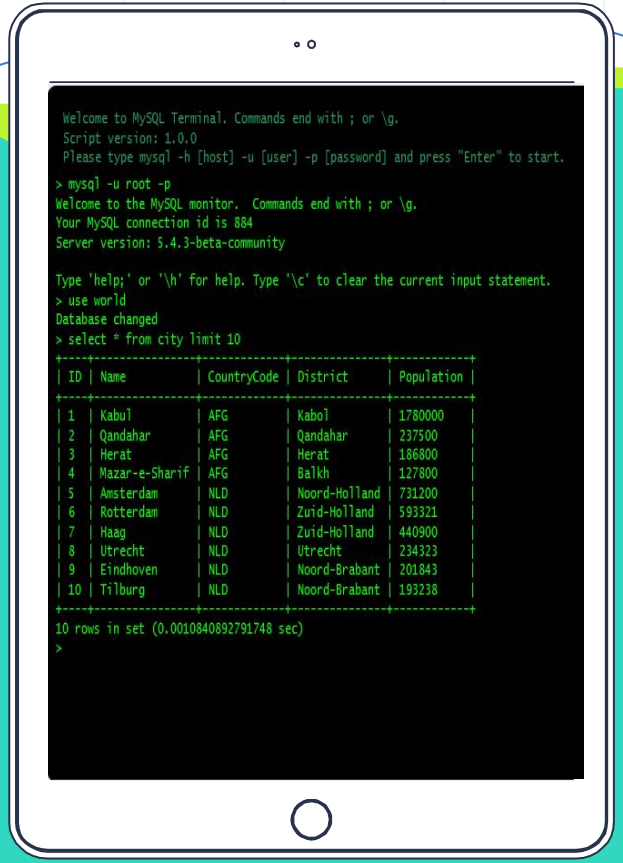


SNEAK PEAK 4

Let’s take a quick look at the project subparts.

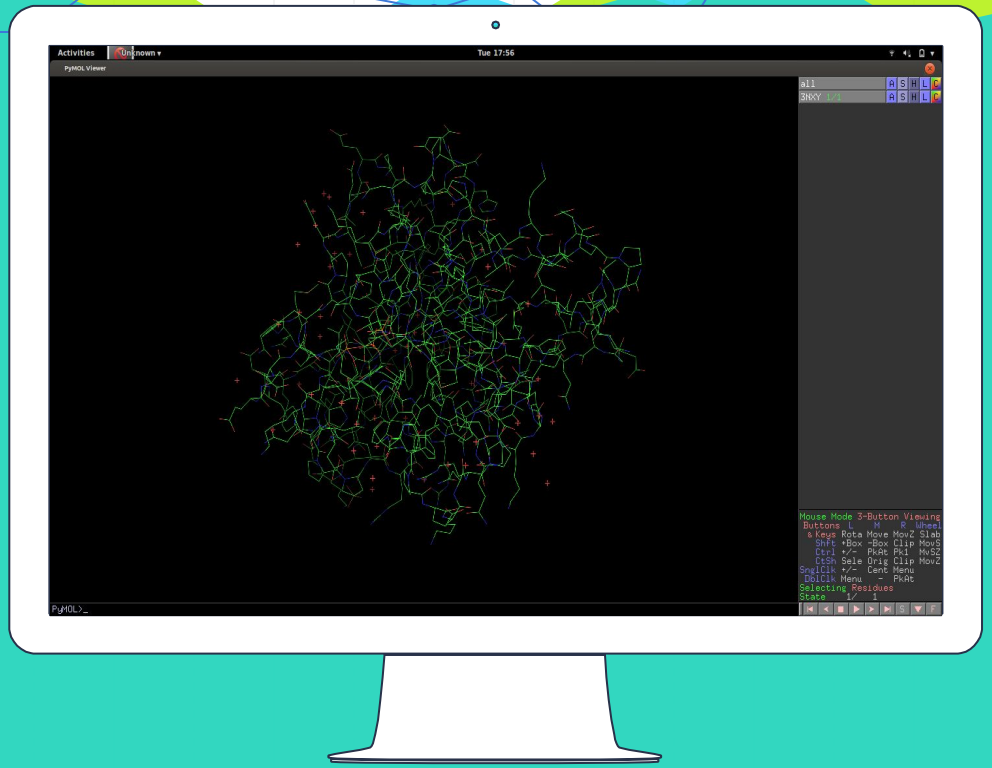
MYSQL Database

This demonstrates the PDB database of our project.



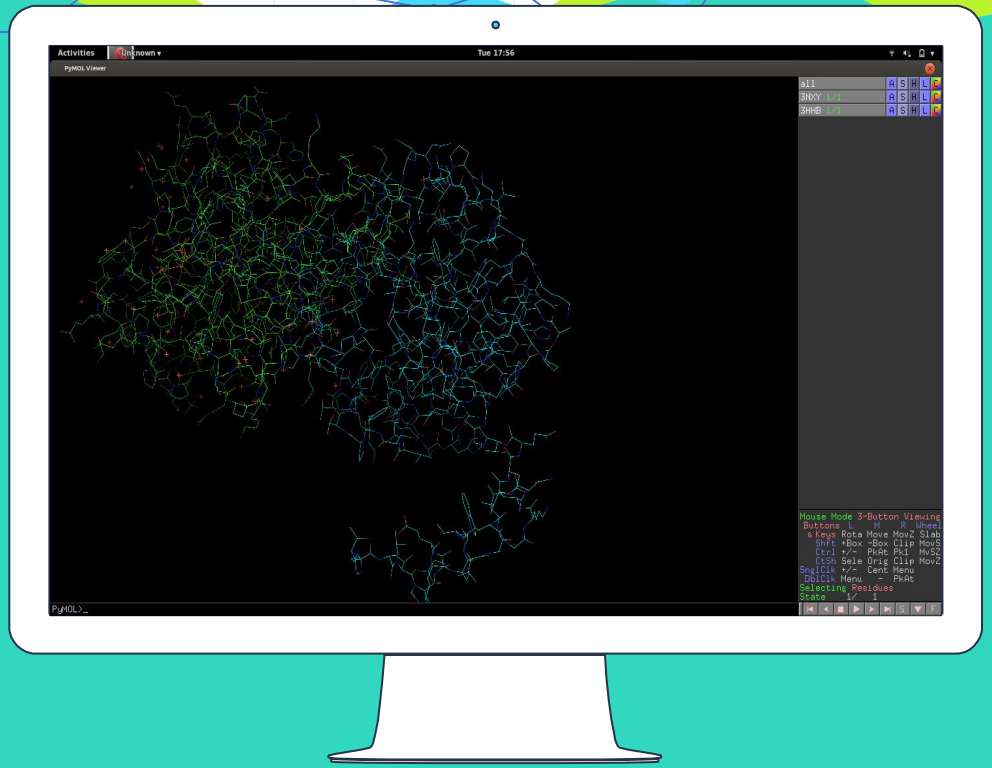
3NXY.pdb

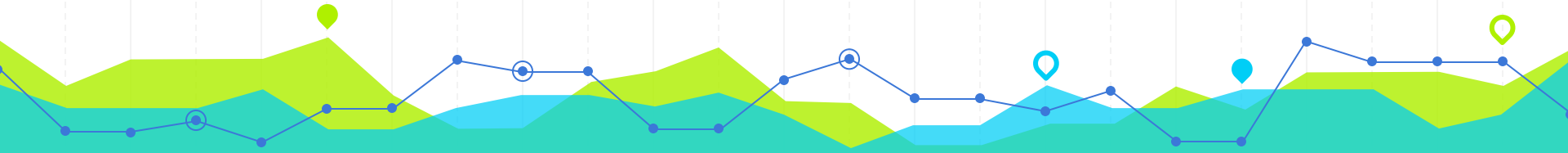
File sample as viewed on PyMol Software



3HHB.pdb

Another file sample from PyMol Software.





PROJECT LOGICS

5

Let’s look at the Data Structure and Algorithm concepts used in this project.

DATA STRUCTURES AND ALGORITHMS

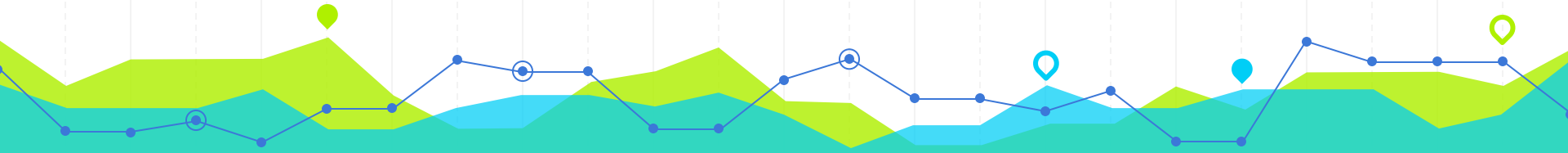
DATA STRUCTURES

- Structured Query Language software implemented, which uses a Binary Search Tree with highly complex algorithms.

-Static Linear String Arrays are used since no dynamicity was essential in the project.

ALGORITHMS

- String Search algorithms for finding exact data in the PDB files.
- Nested Loop Join algorithm, and other search algorithms are used.



IMPLICATIONS

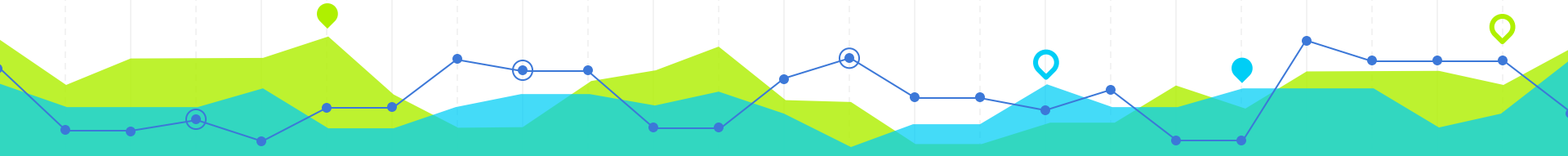
6

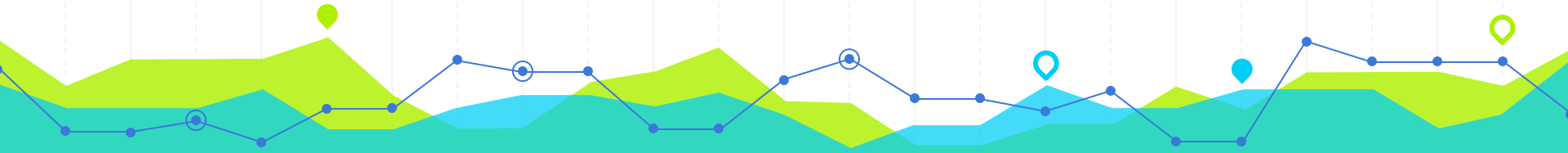
Let’s view the implication of the project to the next level.

PROJECT OUTREACH

Binding Site of any ligand (such as drug) can be inferred using these tools from our project.

- Identify regions of tight atomic packing. (Atom to Atom distance)
- Filter out sites that are "too exposed" to solvent. (Solvent Accessibility Percentage)
- Use hydrophobic/hydrophilic classifications. (Directly from PDB)





STATISTICS

7

Let’s look at the project reach statisticaly.

3,234,830,000 base-pairs

Human genome Database

110,790 files

RCSB PDB Database (as of August 2014)

100%

Reach of our Project!

THANKS!

Willing to contribute?

https://github.com/rishiraj824/PymolDSA

RISHI RAJ

ABHINAV DAS

CREDITS

Thanks to RCSB, PyProt, Biostar, mySQLdb, Stack Overflow, PyMol, Python.org and of course Google.
Thanks to Arnold Emersion Sir, SBST, for his constant guidance.

SPECIAL CREDITS