NETWORK VISUALIZATIONS (APPLICATION TO PROTEIN STRUCTURES)

Ron Michael V. Acda

Activity 10

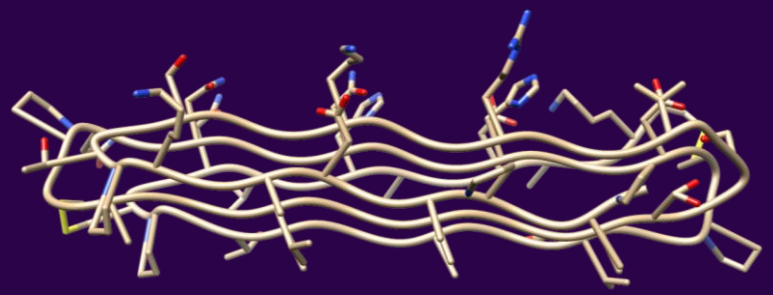
MOTIVATION

In structural bioinformatics, networks can be used as a powerful visualization tool. Since we were given a little bit of leeway to choose what we can do, I chose to apply network visualization in protein structures using the Chimera 1.16 software in conjunction with Python. These networks are weighted graphs, with the weights represented as colors instead of numerical values. These weighted graphs can be used to deduce hydrophilic and hydrophobic domains in a protein.

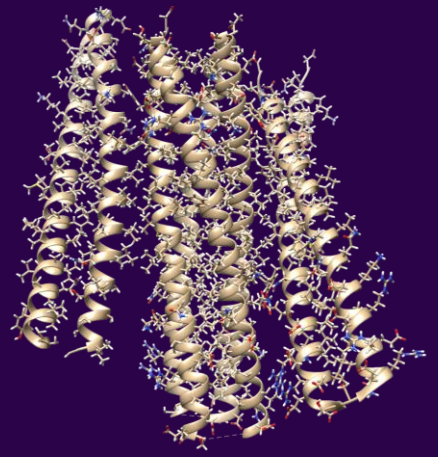
I chose to implement this since biophysics is my field of interest, and I think it would be really cool to demonstrate a little bit of science behind protein structures.

INTRODUCTION

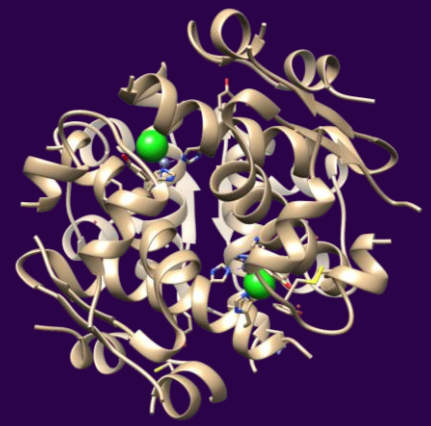
UCSF Chimera allows for the 3D visualization of protein structures as stored in the Protein Data Bank (PDB). Here are some of the cartoon models of some proteins.



Antifreeze protein, used by snow fleas to protect against the cold (PDB ID: 2PNE)



A transmembrane protein (a protein that is embedded in the cell membrane) of *E. coli* (PDB ID: 6B87)



Human insulin (PDB ID: 1ZNJ)

OBJECTIVE

My primary objective here is to construct a 3D network of hydrophobic and hydrophilic residues by connecting the centroids of the residues with a “weighted” edge.
Instead of a numerical value, the edge weights are denoted by an RGB color denoting hydrophobicity/hydrophilicity

SOME DEFINITIONS

Hydrophobic residues: molecules (particularly, amino acids) in the protein that has a side chain that “repels” water. These residues tend to face away from an aqueous environment.

Hydrophilic residues: molecules in the protein that “love” water molecules, trying to face an aqueous environment as much as possible.

METHODOLOGY

The Python scripts are executed in the Chimera 1.16 IDLE interface. The pertinent modules are:

```
import numpy as np
from StructBio.Scenographics.solids import Solids
from StructBio.Utilities.miscFunctions import residueKDhColor
from StructBio.Utilities.shellNeighbors import Shell
```

1. The Solids module allows us to place 3D objects in the 3D visualization.
2. The residueKDhColor allows us to color nodes and edges according to the hydrophobicity value. Those that have bluish colors are hydrophilic, and those that are reddish are hydrophobic.
3. The Shell module allows us to use an algorithm to compute the “nearest contacts” of a residue.

METHODOLOGY

```
class ResidueCenter:
    def __init__(self, resObj):
        self.residue = resObj

    def coord(self):
        res = self.residue
        cbExists = res.findAtom('CB')
        if not cbExists:
            coords = np.array(res.findAtom('CA').coord())
        else:
            coords = np.array([0.0,0.0,0.0])
            counter = 0
            for atom in res.atoms:
                if atom.name not in ['CA','C','N','O']:
                    coords += np.array(atom.coord())
                    counter += 1
            coords = coords/counter
        return coords
```

The nodes are the centroids or the “geometric center” of the residues. Chimera makes extensive use of object-oriented programming, so we will use OOP.

The input for the ResidueCenter class is a Chimera Residue object, and the output of the coord function is a numpy array of the coordinates.

METHODOLOGY

```
phobic_spheres = Solids() #Hydrophilic residues
philic_spheres = Solids() #Hydrophobic residues
HH = Solids() #hydrophobic-hydrophobic
HP = Solids() #hydrophobic - hydrophilic
PP = Solids() #hydrophilic-hydrophilic
```

Now we instantiate the Solid objects to prepare the graph network. The phobic and philic spheres are the hydrophobic and hydrophilic nodes, while the HH, HP, and PP will represent the tubes or the colored edges connecting nodes.

METHODOLOGY

```
for res in res_centroids:
   for neighbor in rCshell.getAtomsInShell(res):
      if neighbor.residue.id.chainId < res.residue.id.chainId:
         continue
      if (neighbor.residue.id.chainId == res.residue.id.chainId and
         neighbor.residue.id.position < res.residue.id.position):
         continue
      kdColor = tuple(0.5*( np.array(residueKDhColor(res.residue)) +
np.array(residueKDhColor(neighbor.residue)) ) )

      if res.residue.kdHydrophobicity < 0 and neighbor.residue.kdHydrophobicity < 0:
         PP.addTube(res.coord(), neighbor.coord(), 0.3, kdColor)
      if res.residue.kdHydrophobicity >= 0 and neighbor.residue.kdHydrophobicity >= 0:
         HH.addTube(res.coord(), neighbor.coord(), 0.3, kdColor)
      if neighbor.residue.kdHydrophobicity*res.residue.kdHydrophobicity < 0:
         HP.addTube(res.coord(), neighbor.coord(), 0.3, kdColor)
```

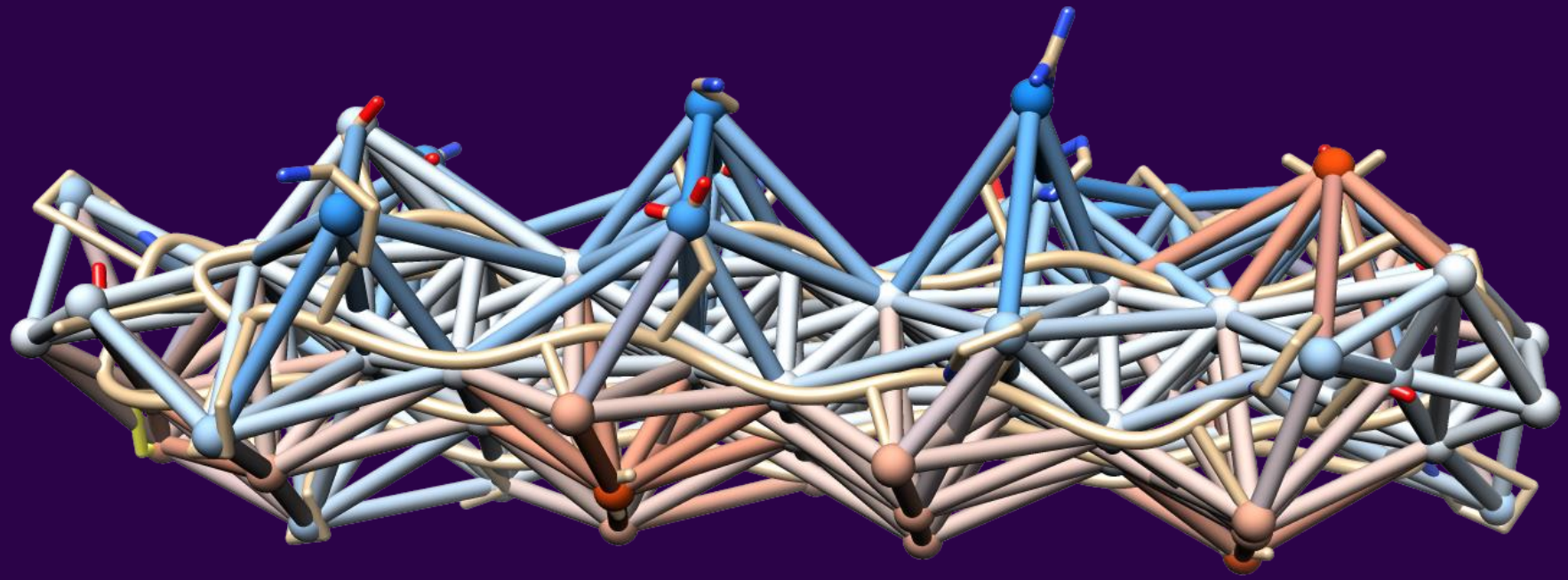
Now we connect the nearest neighbor of each node by a colored edge. The RGB value of the edge represents its “weight”, with a value equal to the average of the RGB values of the two connecting nodes.

METHODOLOGY

```
phobic_spheres.display()
philic_spheres.display()
HH.display()
HP.display()
PP.display()
```

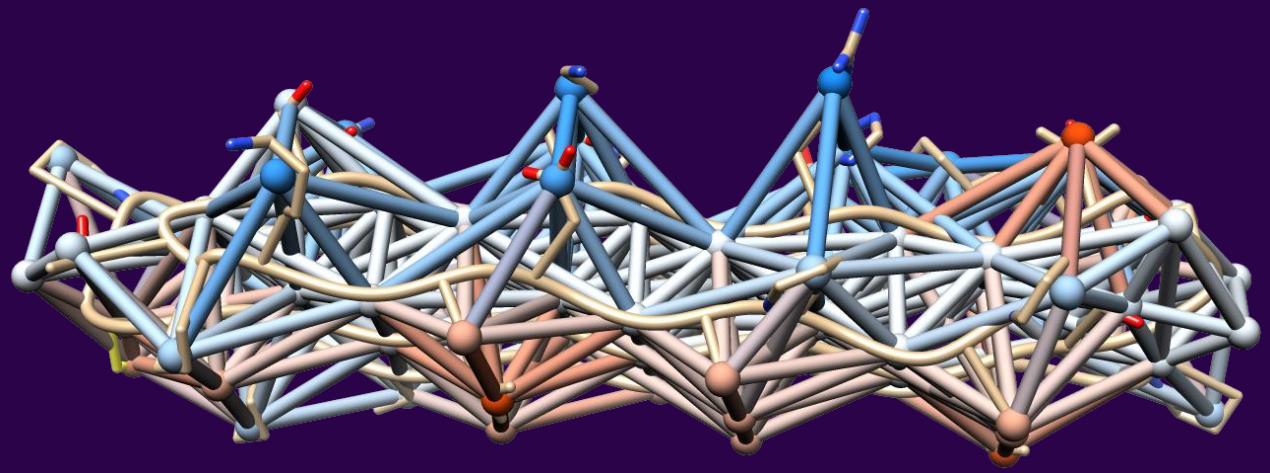
We then display the graph

RESULTS AND DISCUSSION



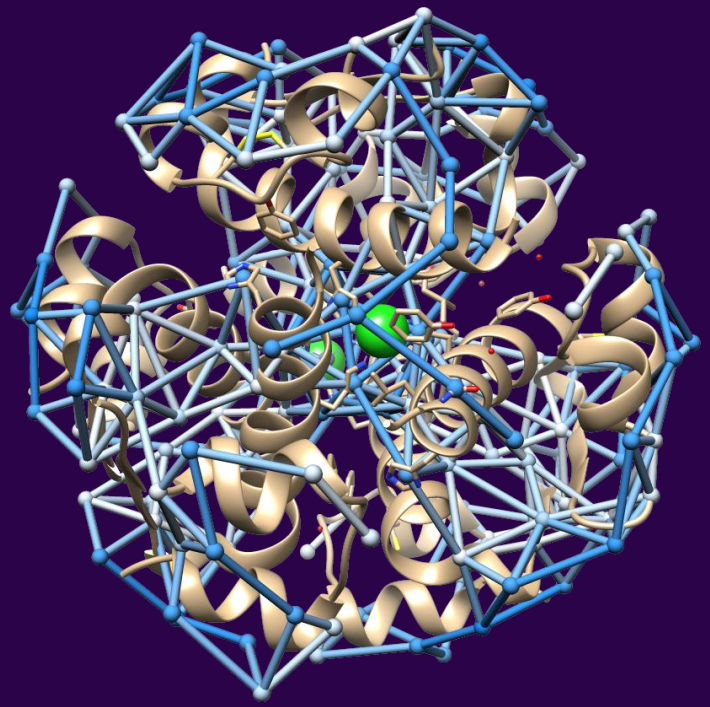
Antifreeze protein residue network

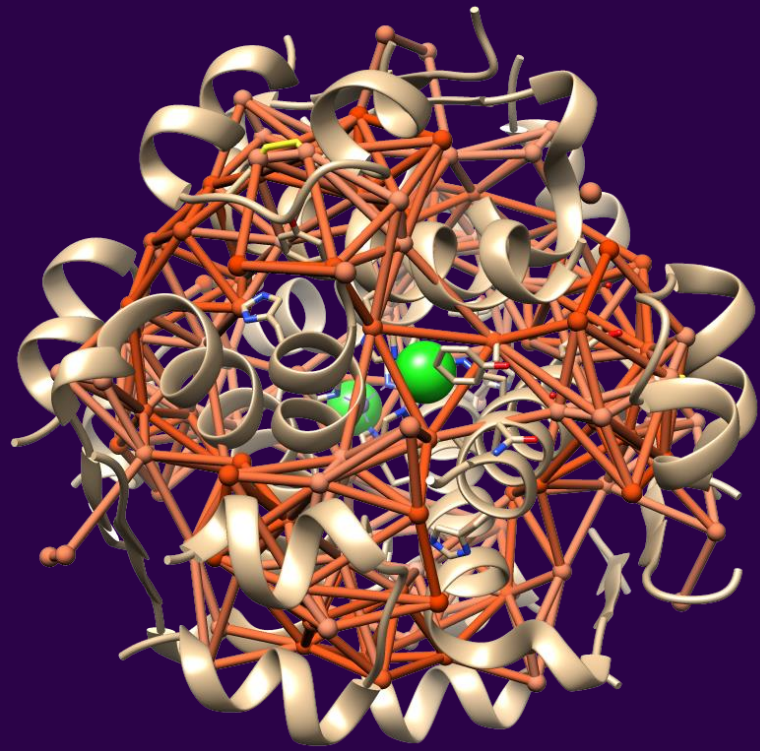
RESULTS AND DISCUSSION



Notice that the protein essentially is two-faced, with one face being “water-loving” and the other “water-fearing”. From here, we can deduce which side faces an aqueous environment.

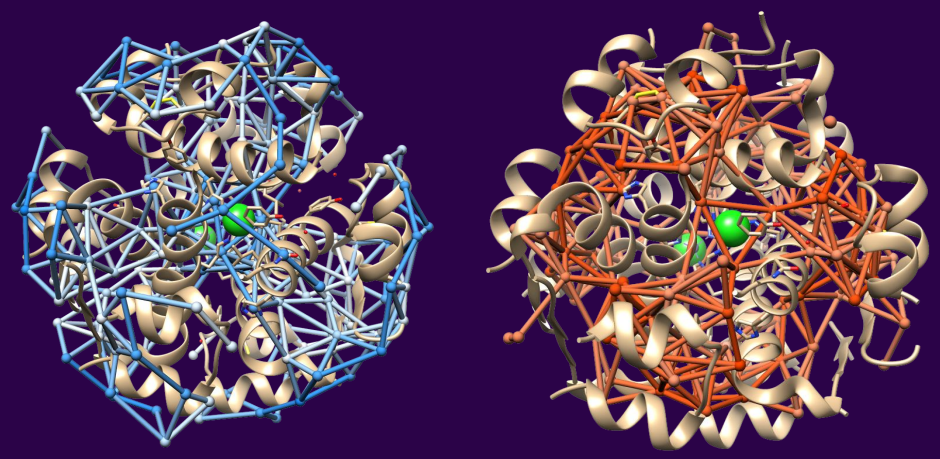
RESULTS AND DISCUSSION





Human insulin residue network

RESULTS AND DISCUSSION

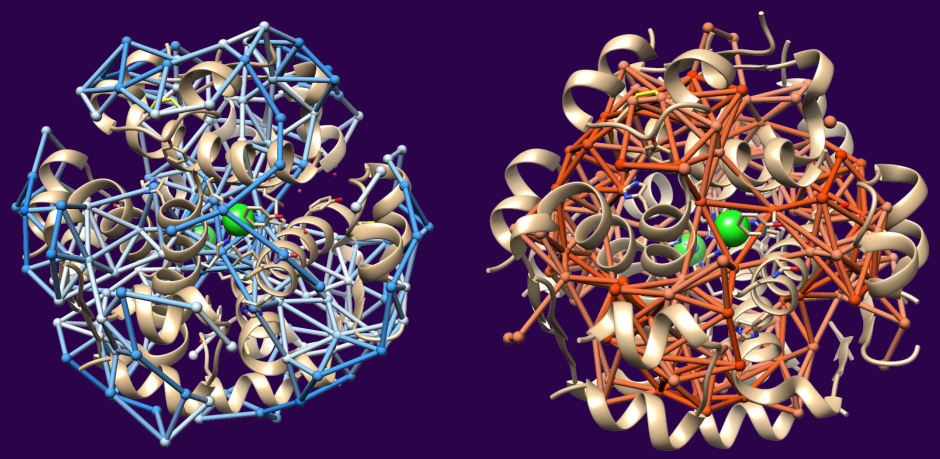


Human insulin residue network

Insulin is an example of a globular protein due to its spherical or globular shape.

These proteins are found in aqueous environments. Thus, it is thermodynamically favorable for them to have a structure whose hydrophilic residues are on the surface, and the hydrophobic residues on the inside and “shielded”

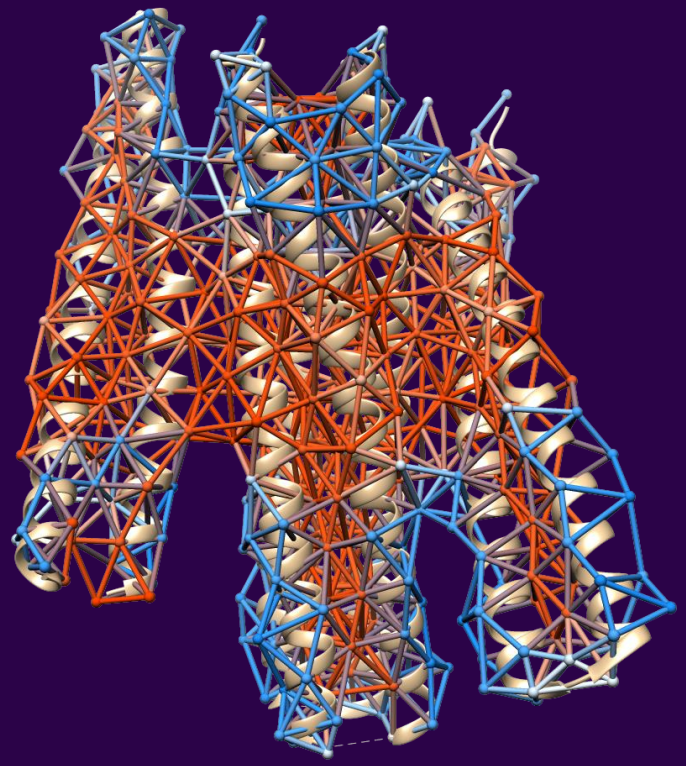
RESULTS AND DISCUSSION



Human insulin residue network

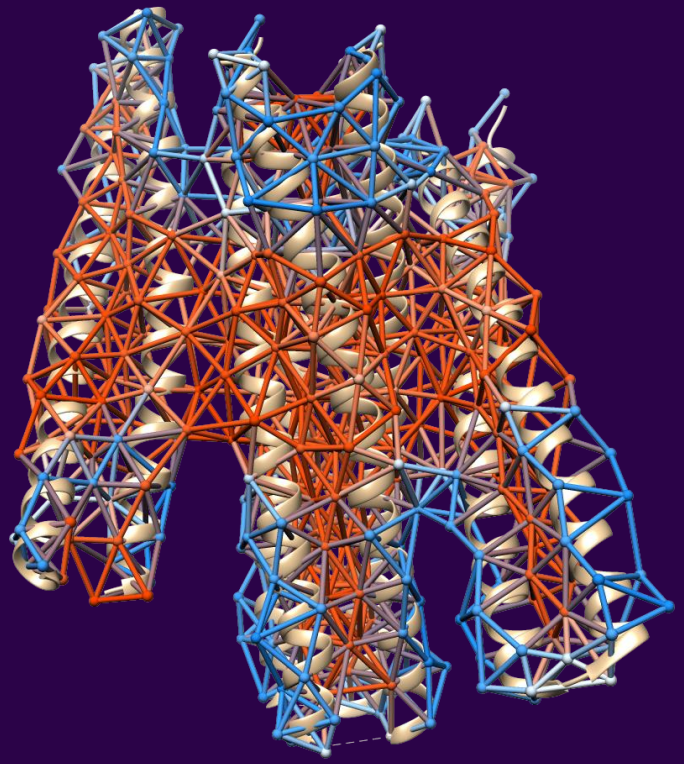
Indeed, we can see here that the external surface is surrounded by hydrophilic residues, and the hydrophobic residues are located inside.

RESULTS AND DISCUSSION



E. coli transmembrane protein

RESULTS AND DISCUSSION



E. coli transmembrane protein

Transmembrane proteins are embedded across the cell membrane. Since the interior of the cell membrane is devoid of water, we indeed expect the middle sections of transmembrane proteins to be hydrophobic.

The sections of the protein that “jut out” of the cell membrane face aqueous environments (extracellular and cytosolic environments), so we expect these regions to be hydrophilic.

CONCLUSION

Residue networks can be used to easily visualize the hydrophobic and hydrophilic domains of a protein. Knowledge of these domains is key to the energetics and thermodynamics of protein folding and structure, and the environments in which they can be found.

REFERENCES

E. Pettersen, T. Goddard, C. Huang, G. Couch, D. Greenblatt, M. E.C., and T. Ferrin, UCSF
Chimera–a visualization system for exploratory research and analysis, J. Comput Chem. 25 (2004).

F. Burkowski, Computational and Visualization Techniques for Structural Bioinformatics using
Chimera (Chapman Hall/CRC, Abingdon, Oxfordshire OX14 4RN, UK, 2014).

SELF-REFLECTION

Self-score: 110/100

I was a little bit glad that we were given leeway to what we can do in this activity as this gives me an excuse to brush up on protein structures and learn a little bit more about biophysics/bioinformatics. This is an exciting example of how graph theory can be used in computational biology and biophysics.