Project Requirements and Suggestions

Overview

Here are some suggestions for your final project. This will involve the write up and presentation of a topic in structural bioinformatics. This may be based on one or more research papers, for example, a comparison of research results or a verification of some research result involving computations associated with protein structure.

General guidelines for content

As stated earlier, the topic must deal with 3D structure of biomolecules. For example:

- Proteins
- Proteins interacting with proteins
- Proteins interacting with a ligand (docking), or
- RNA.

It is expected that the topic will likely be related to some algorithm or procedure that serves some practical purpose related to the 3D structure or provides some insight or analysis about such a structure.

Note that such a study need not involve the entire protein. Scientists studying proteins will often concentrate most of their efforts on a protein domain or the binding pocket of the protein.

Project options

- Option A (literature survey):
    - Pick a problem that you consider to be interesting.
    - Search the literature for algorithms dealing with this problem.
    - Discuss the relative strengths of each approach.

- Option B (empirical evaluation):
    - Pick a problem that you consider to be interesting.
    - Implement and compare two or more algorithms to solve this problem.

You are not limited to these options, but if you choose another type of study then it has to be assessed by the instructor and it should involve as much effort as either of these listed options.

Time lines

All projects are to be handed in on Monday 9 AM April 16, 2012 in RCH 209. This will be followed by the project presentations.

Your project write up should be about 20 pages, double spaced, say 5000 words in length. A presentations should be, at most, 20 minutes in length followed by 5 minutes for questions.

By Mar. 13 at the latest, you should send me a one page description of the project so that I can offer suggestions and so that I can tell you whether or not the topic is acceptable. Students will be allowed to pick their presentation slot with a priority that depends on the submission of their proposal (so if you are first to submit a proposal then you get first pick of the presentation slots, second student get second pick, etc.).

Your project proposal should include the following:

- State the option that you have chosen.
- Clearly state the problem being studied with an emphasis on the structural aspects.
- What are the algorithms involved in this study?
- Be sure to include a set of references to papers that you will be using.

References and leads for project ideas

The following list is an assortment of web pages that may provide some starting points for you when you search for a topic:

1. 3Dsig:
   A fine selection of current research topics can be found in the archived proceedings of previous 3Dsig meetings. Start at: http://bcb.med.usherbrooke.ca/3dsig12/previous_speakers.php and review some of the abstracts in the proceedings. If you find a topic of interest you can usually do a Google search to get more information, for example, a complete paper on the research topic.

2. Molecule of the month:
   http://www.pdb.org/pdb/static.do?p=education_discussion/molecule_of_the_month/index.html.
   This is an excellent starting point for various studies in proteins. Goodsell usually relates function to structure and you can read the research papers associated with the protein to get more information. Recall that every protein in the PDB has at least one reference to a research paper dealing with that protein.

3. For an excellent starting point when you need information about a specific protein, try: www.sbkb.org/.

Sample leads for a project

At times a project can begin with a simple question that leads to ideas that may improve an algorithm. Consider the following leads:

Lead 1: Caver

Look at: http://www.caver.cz/index.php?sid=120. This algorithm is based on Dijkstra's algorithm (remember CS341?). The algorithm estimates the diameter of the largest sphere that can pass into the tunnel. It assumes the protein tunnel is lined with side chains that do not move. Question: Can we improve the algorithm by computing the placement of sidechains so that the narrowest part of the tunnel is made wider? In assignment 2 you did dead end elimination to reduce the number of rotamers for a side chain. There are other algorithms that are more effective in the elimination of rotamers. See for example, the Goldstein algorithm: http://en.wikipedia.org/wiki/Dead-end_elimination. After rotamers are eliminated, the conformation of side chains can be done in several ways, for example: http://dunbrack.fccc.edu/scwrl4/SCWRL4Paper.pdf. Getting back to the original question: Can we find a conformation with a wider tunnel?

Lead 2: Pockets

Currently, researchers understand the structure and function of many proteins, but there are thousands of proteins for which we know structure while having no knowledge about their functions. As a first step in trying to understand function, researchers often employ software that investigates the surface of a protein in an attempt to discover surface "pockets" that might function as binding sites. See:

Pérot, S., Sperandio, O., Miteva, M.A., Camproux, A.-C., and Villoutreix, B.O., Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery, *Drug Discovery Today*, Vol. 15, No. 15/16, 2010, 656-667.

There are several algorithms for this purpose. A few of these programs are described in[1]:

Levitt D, Banaszak L. POCKET: a computer graphics method for identifying and displaying protein cavities and their surrounding amino acids. *J. Mol. Graph* 1992, **10**:229-234.

Hendlich M, et al.: LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *J. Mol. Graph. Mod.* 1997, **15**:359-363

Morita, M., Nakamura, S., and Shimizu, K., Highly accurate method for ligand-binding site prediction in unbound state (apo) protein structures. *Proteins*, 73, 2008, 468-479.

[1] Another rather different approach can be found in: Bock, M.E., Garutti, C., and Guerra, C., Cavity detection and matching for binding site recognition. *Theoretical Computer Science*, **408**, 2008, 151-162.

Zhang, Z., Li, Y., Lin, B., Schroeder, M., Huang, B., Identification of cavities on protein surfaces using multiple computational approaches for drug binding site prediction. *Bioinformatics*, Vol. 27, No. 15, 2011, 2083-2088.

A good overview of the topic can be found in: rostlab.org/cms/uploads/media/20100511_Schafferhans_PP_BindingSiteAnalysis.pdf. In this case, the question is: How do these algorithms compare in terms of efficiency, success, and application goals?

Lead 3: Icosahedral surfaces

Launch Chimera and do the following menu invocation:
Tools… **Higher-Order Structure**… **Icosahedron Surface**.
It might seem somewhat strange that Chimera would have a special menu entry for such a structure but it ties in with multiscale modeling of virus capsids: http://www.cgl.ucsf.edu/chimera/experimental/flatten_icosahedron/flaticos.html.
Icosahedrons have also been studied by researchers dealing with fullerenes: http://en.wikipedia.org/wiki/Fullerene. See material related to C_{540}.
The vertices of an icosahedron are easily computed. See the Cartesian coordinates section of: http://en.wikipedia.org/wiki/Icosahedron.
It has also been shown that the icosahedron can be the first step in a sphere tessellation that uses triangles to represent the surface of a sphere. See: http://www.geometrictools.com/Documentation/TessellateSphere.pdf. A Google search will give other references for sphere tessellation.
What would be project goals for a sphere tessellation? There are several possibilities. Here is one: The algorithms can be modified to allow the generation of partial spheres. This is the type of structure used to explain the donor and acceptor regions for hydrogen bonding. See the discussion at: http://www.simbiosys.ca/sprout/visualisation/index.html.
The goal of such a project would be to generate the surfaces that define the donor and acceptor regions in a molecular scene after some particular hydrogen bond is specified.

Lead 4: Hydrophobicity

Hydrophobic interactions play a significant role in the formation of a protein's tertiary structure. In addition, hydrophobicity of a binding site is a very important consideration in the design of drugs. To visualize the hydrophobic attributes on a protein surface one can use the menu invocation: **Presets**… **Interactive 3 (hydrophobicity surface)** to get a surface that is coloured according to the Kyte-Doolittle hydrophobicity of the amino acid just below the surface. The colours range from dodger blue (hydrophilic) to white to orange red (hydrophobic). Here is the question: How does this compare with the assessment of hydrophobicity discussed in:
Brylinski, M., Kochanczyk, M., Broniatowska, E., and Roterman, I., Localization of ligand binding site in proteins identified *in silico*. *J. Mol. Model.* 2007, **13**, 665-675?
A worthwhile part of this project would be the development of a script to do the visualizations seen in this paper.

The Contents of the Project Report

The following points are suggestions for the content of your report:

Option A: Literature survey

1. Introduction and motivation
    a. Clearly state the problem or topic.
    b. How is the problem or topic important?
2. Literature survey
   What are the techniques or algorithms used by various researchers? It will be important for you to organize this material in a way that makes it understandable and clear.
3. Critical evaluation
    a. How well do the techniques or algorithms achieve their objectives?
    b. What are the strengths and weaknesses of the various approaches?
    c. What problems or issues remain to be studied?
4. Conclusion
    a. What can we learn from these papers?
    b. Can you suggest any new avenues of investigation?
5. References

Option B: Empirical evaluation

1. Introduction and motivation
    a. Clearly state the problem or topic.
    b. How is the problem or topic important?
2. Algorithms to be studied
   Clearly describe the algorithms that you are comparing. Be sure to justify the choice of these algorithms. In other words: Why do they hold any promise of being successful?
3. Critical evaluation and comparison
    a. How do the algorithms compare?
    b. What are their strengths and weaknesses? The comparison should involve a performance evaluation, assessment of complexity, ease of use, and any other measures that you consider to be relevant.
4. Conclusion
    a. Considering your evaluation, what algorithm is the best?
    b. What can we learn from these algorithms?
    c. Can you suggest any new avenues of investigation?
5. References

Project Marks

Here is a project marking template. **Note**: I would highly recommend that students do their project using Microsoft Word. The grammar checker is very useful. Marks will be deducted for extensive bad grammar that is so bad that it could easily have been detected by Microsoft Word.

Technical Style of the report:

Grammar and Spelling:	__________	(/8)
Organization:	__________	(/6)
Clarity:	__________	(/6)

Content:

Introduction and motivation:	__________	(/10)
Background material:	__________	(/10)
Critical evaluation:	__________	(/25)
Discussion/Conclusion:	__________	(/10)
References	__________	(/5)

Presentation style:

Grammar and Spelling:	__________	(/4)
Organization:	__________	(/4)
Clarity (speaking & pacing):	__________	(/4)
Design of slides (font size, pictures, examples):	__________	(/4)
Interaction with audience: (enthusiasm, handling questions)	__________	(/4)

Total: __________ (/100)