*Bioinformatics*, 2016, December

doi: 10.1093/bioinformatics/xxxxx

Advance Access Publication Date: DD Month YYYY

Manuscript Category

|  |
| --- |
| Subject Section  Comparative Protein Modelling  Sai Manikanta S Godavarthi, Deepika Joseph  Computer Science – EECS Department, Wichita State University, Wichita, Kansas  Abstract  **Motivation:**.  **Results:** With the given 10 proteins; 5 from CASP 11 and 5 from CASP 10 we have obtained better accuracy for 3 proteins comparing with the CASP results. Results for each of the protein are explained separately in the results section. We have also implemented a sample pipeline for automation of the process of modelling using modeller. Implemented few refinement steps on some proteins where the accuracy has been increased compared to the previous results before refinement.  **Contact:** sxgdavarthi@shockers.wichita.edu  **Supplementary information:** Supplementary data are available at *Bioinformatics* online. |

# Introduction

The term comparative protein modelling or homology modelling or template based modelling are referred as the same where our own main goal is to model a protein 3D structure using the know templates. Here the templates are those with similar sequence to that of our unknown query protein. Based on the known protein i.e. our template we will model the query protein using the properties and alignment of the known template. The template contains sufficient information of spatial arrangement of residues and internal structure which helps in predicting our model. Comparative modelling of protein sequence is more reliable than compared to that of ab initio methods as in the later, the model is entirely built using only the sequence rather depending on the template. Protein are one of most important functional units of our body, they do most of the work in cells and they are required for structure, function, and regulation of body’s tissues and organs. The three-dimensional structure of the protein determines the functionality of the protein. The four levels of proteins i.e. the primary structure which is a sequence of amino acid residues determine the peptide chain. In the secondary structure, hydrogen bonds between the amino acids creates alpha helix, which is a spiral or coiled molecule and pleated sheet, which looks like ribbon with regular peaks and valleys as a part of the fabric. The tertiary structure is for overall shape of the protein which are either globular or fibrous. Quaternary structures describe the proteins appearance.

Our method of protein modelling starts with taking a query sequence. Query sequence is the one which we want to model a three-dimensional structure for the protein then later identifying template and build model using the modeller tool and we validate our results using varies validation techniques available online and do structural analysis of the proteins using visualization tools.

# Methods

The whole process if protein modelling is described as follows:

1. Select the required protein and get the sequence.
2. Search for template in BLAST, PDB and other protein databases accordingly as required.
3. Find the better matching sequence for the query sequence, this is our template. We may have single or multiple templates. Align the template with our query.
4. Prepare our files in PIR format. Download the PDB formats for the templates. Save PIR formatted query file as .ali extension.
5. Start the modeller by giving input the query .ali file and other PDB files, according to the log file generated give inputs to the modeller.
6. Validate the chosen best model and determine the accuracy.

To align the given sequences, we have used multiple sequence alignment techniques for multiple templates selected, some of them include T-Coffee and Clustal-Omega. At some cases, we have chosen only one template, where we have used Needleman-Wunch algorithm to align the sequences and converted all of them into PIR format. The pipeline that we have implemented asks for sequence and automatically converts them into PIR format and saves query in .ali file and rest template sequences as .pir extension. The PDBs are automatically download once given the template IDs after script 1 execution. The program that we wrote has BioPython packages and uses NCBI pdb API call to download the PDB files given the template IDs. Alternatively, we can use REST API to download the pdb files. The REST API is of XML format.

For validation, we are using different tools available online and for visualization purpose we have used Chimera as our tool. The various online techniques that we have used to measure the accuracy of our model inclue TM-Score, Molprobity scores, and RMSD score.

All the steps are cleared explained clearly below where we have mentioned each of options that we have chosen for protein modelling.

The software distribution that we used for this project is described as below:

1. Test cases i.e. Protein queries are taken from CASP website. CASP10 and CASP 11 are chosen to select 10 proteins which include:
   * CASP11 targets T0856, T0843, T0806, T0837, T0792 and
   * CASP10 targets T0757, T0666, T0678, T0651, T0694
2. For template identification: BLAST, PDB, and SWISS-Model.
3. Sequence alignment: Needleman-Wunch algorithm, and T-Coffee
4. Software for protein modelling used is Modeller 9.17.
5. Various languages used are python 2.7 and JAVA (Needleman-Wunch algorithm)
6. Protein visualization software’s: Chimera, and Rasmol
7. Project Management and version control: Github

**Process automation:**

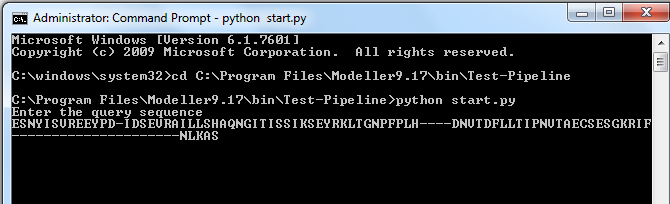
Implemented a sample pipeline where the process of using modeller is made lot easier compared to the standard approach. Instead changing inputs in the script files everytime, user needs to just enter the sequences initially and template IDs later for each of the script to run. The script automatically converts the given aligned query sequence into PIR format as saves it as .ali extension file. At each script execution, based on the output generated by the script and after evaluating the results, we given the template ID as input to the next script and the process is the same till the end. Sample screenshot for the pipeline is as below, where the program ask for user input query sequence.

Figure 1:start.py script; User entering query sequence

# Results

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

## Data Structure This is Heading 2 style this is heading 2 style

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

### 3.1.1 This is heading 3 style

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

1. The quick brown fox jumps over the lazy dog.
2. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.
3. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

* The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.
* The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.
* The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

## Unnumbered list style

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

 (1)

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

**Fig. 1. Relation between τ and *t*.** This example has only two continuous Steppers, S1 and S2.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

**Table 1.**Benchmark results of the cascade oscillators model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| |S| | Predicted cost | Timing | Predicted speed | Speed |
| 1 | S219.20(100%) | 68m43s | 1.00 | 1.00 |
| 2 | 29.10+219.10(~50%) | 35m13s | 2.00 | 1.95 |
| 4 | 219.20(100%) | 68m43s | 1.00 | 1.00 |
| 10 | 29.10+219.10(~50%) | 35m13s | 2.00 | 1.95 |
| 20 | 219.20(100%) | 68m43s | 1.00 | 9.5 |

This is table foot note sample text This is table foot note sample text This is table foot note sample text

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.The quick brown fox jumps over the lazy dog.

Acknowledgements

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

Funding

This work has been supported by the …..

*Conflict of Interest:* none declared.

References

Alexandrescu,A. (2001) Modern C++ Design: Generic Programming and Design Patterens Applied. Addision Wesley Professional, Boston.

Dormand,J.R. and Prince,P.J. (1980) A family of embedded Runge–Kutta formulae. *J. Comp. Appl. Math.*, **6**, 19–26.

Alexandrescu,A. (2001) *Modern C++ Design: Generic Programming and Design Patterens Applied.* Addision Wesley Professional, Boston.

Dormand,J.R. and Prince,P.J. (1980) A family of embedded Runge–Kutta formulae. *J. Comp. Appl. Math.*, **6**, 19–26.

Alexandrescu,A. (2001) *Modern C++ Design: Generic Programming and Design Patterens Applied.* Addision Wesley Professional, Boston.

Dormand,J.R. and Prince,P.J. (1980) A family of embedded Runge–Kutta formulae. *J. Comp. Appl. Math.*, **6**, 19–26.

Yoo,M.S. *et al*. (2003) Oxidative stress regulated genes in nigral dopaminergic neurnol cell: correlation with the known pathology in Parkinson’s disease. *Brain Res. Mol. Brain Res.*, **110**(Suppl. 1), 76–84.

Crenshaw, B.,III, and Jones, W.B.,Jr (2003) The future of clinical cancer management: one tumor, one chip. *Bioinformatics*, doi:10.1093/bioinformatics/btn000.

Auhtor,A.B. *et al.* (2000) Chapter title. In Smith, A.C. (ed.), *Book Title*, 2nd edn. Publisher, Location, Vol. 1, pp. ???–???.

Bardet, G. (1920) Sur un syndrome d’obesite infantile avec polydactylie et retinite pigmentaire (contribution a l’etude des formes cliniques de l’obesite hypophysaire). PhD Thesis, name of institution, Paris, France.