WEBMODELLER	A practical introduction to WebModeller which includes the usage of Blast, Clustal w, Modeler and other Bioinformatics stuff for the new and the advanced user.

Table of Contents

Overview

Webmodeller an integration of biology and computing, a solution to your modelling problems. Webmodeller is a complete and stand alone program which allows you to obtain a best model for your novel protein sequence using local modeller. It includes ClustalW program to obtain a phylogeny relationships between your model and the best 5 hits which it obtains by running blast p locally on pdb amino acid database.
You can visualize your modeller reports, Blast reports and ClustalW files instantly when the job is finished and can carry on your further research.
It enable you to do modelling and obtaining best model for your query protein without having the knowledge of modeller and its input files viz, template pdb file, operating python file, alignment file.

Intended Audience

Webmodeller is very basic but very powerful. Its simple user interactive web interface welcomes novice and advanced user to gain benefits of complex programming running at back hand. Webmodeller is appropriate both for simple as well as advanced users of Bioinformatics.

Before you Begin

Before you Begin, you should know what is :

- Protein Sequence or amino acid sequence : When amino acids connected by peptide bonds lie in an organised order it forms a protein sequence and its codes for some big protein which may be an important constituent of cell cycle, metabolism etc. Generally, a protein sequence is reported from the N-terminal end containing free amino group to C-terminal end having free carboxyl group.

  Click here to read more.[1]

- NCBI BLAST : BLAST basically abbreviates to Basic Local Alignment Search Tool. This tool is provided freely by NCBI where the inputted sequence is searched for the similarities against the databases such as nr(non-redundant), pdbaa(protein data bank amino acid database), swiss prot database etc. There are various blast programs such as Blastn, Blastp, Tblastn, Tblastp, TblastX, SNP blast, Mega blast etc.

  Click here to read more.[2]

- ClustalW : It is a multiple sequence alignment program which produce biologically meaningful alignments of inputted protein or DNA sequences. It basically calculates the best match for inputted sequences and lines them according to the similarities, identities and differences between them. It produces alignment file which can be viewed in text form or in browser and other is the dnd file or the tree file which can be viewed in any tree viewing software such as NJPLOT etc.

  Click here to read more.[3]

- Modeller : Modeller is a modelling program which is used for homology or comparative modelling of proteins. Its is command line program and somewhat difficult to operate but once known its very useful. For using modeller, the user needs to provide 3 files, one of which is the alignment file which includes the sequence of user model and the sequence of most significant hit obtained in the blast report. The second file is the template pdb file which the user downloads from the www.rcsb.org for the most significant hit obtained from the blast output. Third file is the operations file which holds the instructions to run the modeller. It can be instruction file with the extension .top or a python file which has an extension of .py. Number of models to be outputted are specified in this file. When all the files are ready then the command is run from the CUI (command line interface) like gnome terminal in Linux and after some time the models are generated from which the best model is generated which is basically the model with lowest energy(molpdf).

  Click here to read more.[4]

Sample Data

```
PQITLYQRPLVTIKIGGQLKEALLKTGADDTVLEEMSLPGEWKPKLIGGIGGFIKVRQYDQILIEICGHKAIGTVLVGPT
PVNILGRNLLTQIGCTLNF
```

This amino acid sequence is a perfect example for sample data but webmodeller runs on a fool proof and very flexible system. If user inputs any kind of illegal characters, special characters, any digits or numerals, the data will be automatically formatted to pick up only the alphabets and will change into uppercase letters to increase the human readability and then sends it to blast and carries out the other things on that.

* Try to copy only the sequence not the protein name or any other extra alphabets as in all it will ruin the results to some extent.

Lets Start

When the webmodeller is accessed at http://bioinf3.bioc.le.ac.uk/~ss533/input.html , the following screen appears which is the homepage of webmodeller.

Step 1 : Get Familiarize with Webmodeller.

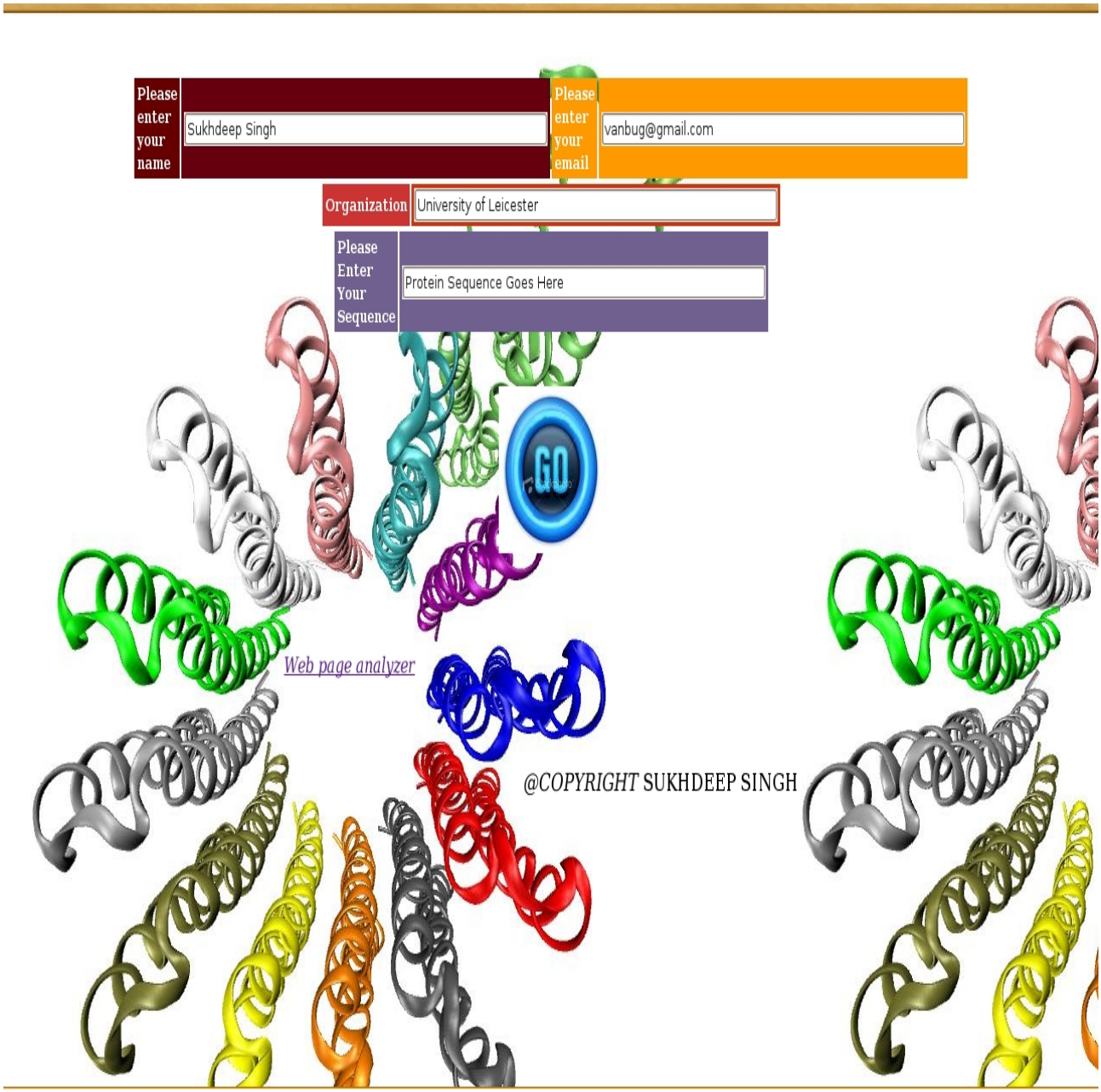

Fig.1 Webmodeller's Main Page

Step 2 : About various input fields

Webmodeller as already said is very user friendly and runs on an interactive environment.
These are 4 input fields viz,

- Name input field : Here user enters his your name. *Default is Sukhdeep Singh*
- Email id : Here user enters his email. *Default is vanbug@gmail.com*
- Organization : Here user enters his organization. *Default is University of Leicester*
- Sequence Field : Here user enters his query sequence.

As I have told that webmodeller runs on a fool proof system, so if a person doesn't enters his/her name or email id, then respective message is displayed as 2 persons name and email cant be same.

Step 3 : Web Page Analyzer

There is a link given to Web page analyzer which opens an on-line web analyzer where the different statistics are displayed for the inputted link.
*Webmodeller has passed every check in that.

Step 4: Things to do

User should have a clear focus in mind that what he/she is doing and why is he/she doing.
The inputted sequence can be a novel protein sequence which has been just found out in the lab and whose structure is not known yet, so its model structure needs to be generated on the basis of its most related protein in family or out of family.
So, webmodeller is a place where user can input its protein sequence which goes into a pipeline where its follows blast, then clustalw and in the end modeller.

Step 5 : Things required

After familiarizing yourself with certain things, this is the time to start now.
At this point, you should have 2 things:
1) A protein sequence to model.
2) An aim what to do and what to obtain.

Step 6 : GO

So after filling all input fields, press the big GO button to enjoy the ride.
After pressing the GO, it will go second page where you can see all the stuff such as reports and results.

What's happening

Now its the time to see what is happening at back side of Webmodeller.
A user should be able to understand that on what is he working and in actual what is going on.

Step 1 : Understanding what's happening there

When the user enters the sequence and fills in all the other details, and presses GO button, then webmodeller takes user to the next level.
The sequence goes through a number of steps viz.;

- Displaying the user sequence in browser.
- Running blast locally with pdb amino acid database on the query sequence.
- Finding Template for query sequence which is the best hit from blast.
- Downloading the pdb file of template from ftp of protein data bank.
- Generating alignment file for modeller.
- Generating user options as view reports, download reports and structures.
- Generating buttons such as Run modeller and Run clustalw.

Step 2 : Getting Results

After entering the sequence on the first page, the user sequence is displayed on the second page. Then, the BLAST-p (Basic Local Alignment Search Tool – Protein) starts on the query sequence choosing pdb database and produces a result file. From that result file the the most significant hit is taken out by pattern match and is displayed as " The template found out for your sequence is ...." Then , after few moments, the web page appears completely and user can see various options on the screen such as

- Run ClustalW
- Run Modeller
- Download Results
- Download Template pdb file
- View Alignment file
- View Blast report
- Home
- Contact Me
- Click for Resume

When the user get the message that his template is this, then its his wish, he can *run ClustalW* which takes him to a new page where he can view the Clustal Alignment file and Clustal Tree file.

Also, he can press *run modeller* and then go to modeller page where the modeller will run when he will press run modeller and the he can view the modeller report and also he can download the top 5 pdb files of the models generated by the modeller.

He can now either download the results by pressing *download results* in the form of Blast report, alignment file and modelller file or he can view them separately by pressing the respective buttons.

Home button is there to go to home page.

Contact Me link is there, when pressed it takes the user to a new page where he can send the query to webmaster and will receive the same email to him also.

Click for resume button when pressed gives the resume of the creator of Webmodeller.

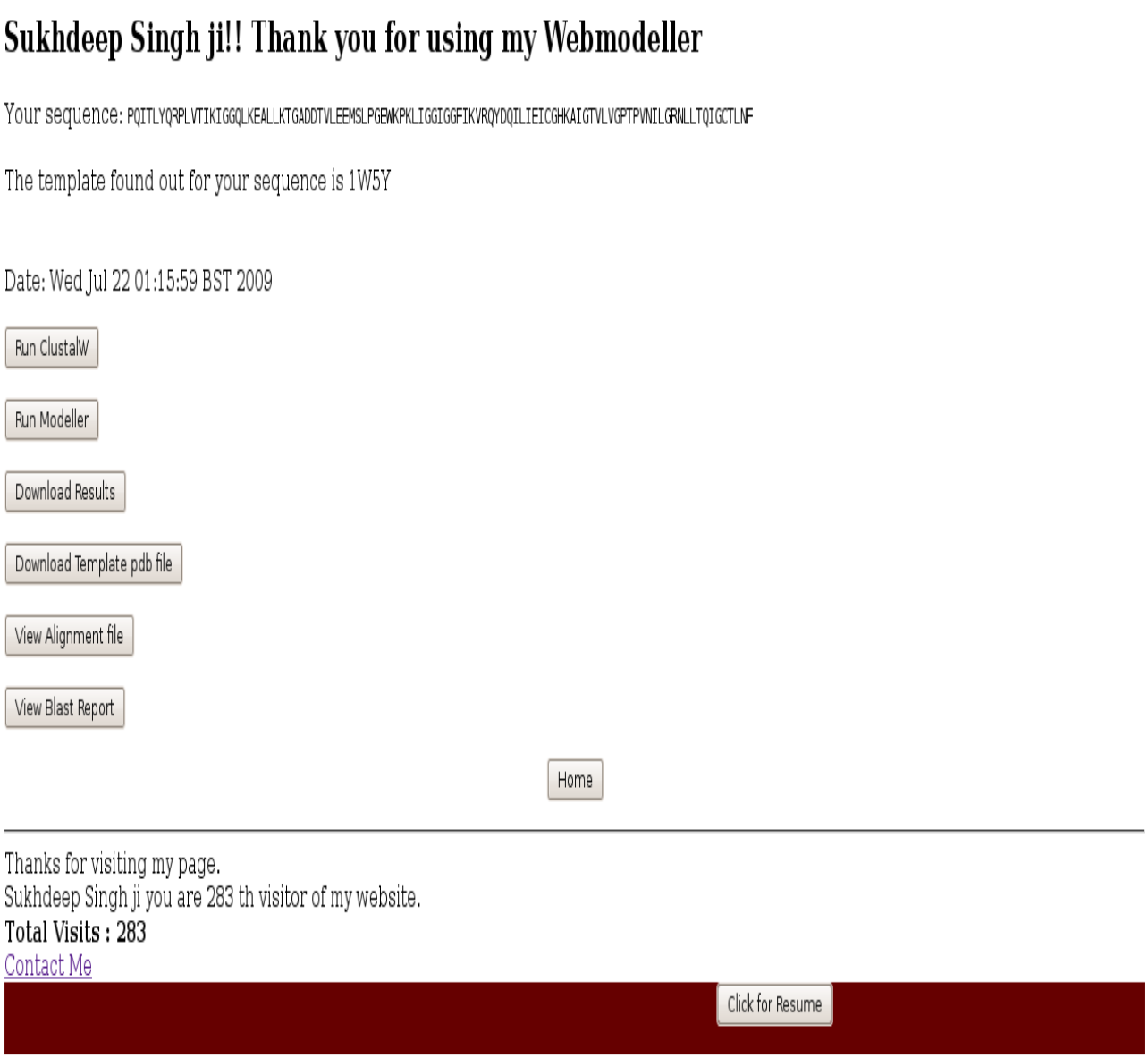

Fig.2 Webmodeller Results page

Step 3 : Am I Satisfied

Last thing to consider for the user is am I satisfied, so if the hunger is satisfied them I am very grateful, if not then the user can take the pdb files of the models along with the other repost files and can study them and can go to other databases also such as NCBI, SWISSPDB, EXPASY etc. If anything user wants to say about webmodeller, he/she can email me by pressing contact me.

Summary

Webmodeller is a stand alone programme for the user's modelling problems solved by combining very powerful programs such as Modeller, Blast and Clustal W. Its a platform designed to support the group of people who are working in the field of research for finding some thing new and good for the humanity. When the novel protein sequence is identified in the lab or in vivo, then the first step is that is there any other similar protein like that which might be performing vital work in some organisms body. So, we do blast to search for the similar protein and after that the most significant hit is taken where the scientist says that its similar to this protein found in this organism and playing this role. So, if its protein playing role in some disease then the next step is to model the protein or binding the ligand so that it does not bind the target and does not cause the disease. For ligand binding and docking, scientist needs a protein structure and if its novel then it must go through NMR spectroscopy or X-Ray spectroscopy to deduce the structure.

So, scientist can use *swissmodel* which works on multi step process, takes more time and is less complicated or can use *modeller* which work on single step process, takes less time and is more complicated as involves the construction of 3 different files to run it.

So, Webmodeller comes into play and for the users which do not know anything about modeller, blast or clustalw can just enter their sequence and just can enjoy the results on one platform rather than going to NCBI's BLAST, EBI's ClustalW or need not to download and install and make files for Modeller as its integrated in Webmodeller.

Resources

1. http://en.wikipedia.org/wiki/Peptide_sequence
2. http://blast.ncbi.nlm.nih.gov/Blast.cgi
3. http://www.ebi.ac.uk/Tools/clustalw2/index.html
4. http://www.salilab.org/modeller/
5. http://www.rcsb.org/pdb/home/home.do