Introduction

In this practical we will try to predict the structure for one of the target sequences from CASP 10. We will build homology models using MODELLER and assess the quality of the models.

You need to choose a template for your query sequence (from CASP) using HMMer

The idea is to create two alignments for the template with your target sequence using
1) pairwise sequence alignment
2) alignment from HMMer

After this we will assess each of the models based on GDT scores, comparing it to the solution structure.

This practical will be done in groups of 3 students. I would suggest that you split the work above, for example one alignment type per person. You need to indicate who has done what.

Instructions for handing in your assignment

Please submit your assignment report in pdf format. You can submit one report per group.

Make sure to hand in at least all material so that your analysis is reproducible. For example: identifiers, files used, alignments, names and settings of programs used, scripts used. Make sure that each file you hand is also explained inside the report together with the filename.

1 - CASP Target

You will start with building a model for CASP10 (2012) T0651. Get the sequence for this target. Why do some tagets have a ‘D1/D2’ suffix, and why is this suffix not stated in the target list? What is the PDB code of the experimentally solved ‘solution’ structure? What kind of structure is this? Give a detailed description / analysis (secondary structure content, function, classification, picture, ligands, experimental method and date of deposit).

[10]

2 - Fold recognition server

Use the HHpred server to find a template for your target sequence. You are not allowed to choose any templates that were deposited after May 2012. Why not?.

Pick a template using HMMer (you may use the online server). Explain why you choose this template, based on which scores and what this indicates. How accurate do you think your model will be, based on the results of HHMer?

Give the alignment from the server between the target and template. How was this alignment

created?

[10]

3 - Description of template

Describe your template structure in detail (just as in Q1). Based on what you know from the solution structure, do you think you will be able to create a reasonable model?

[5]

4 - Sequence Alignment

You should create an alignment that is based on sequence alone (You can use a server on EBI - see BB). Give the pairwise sequence alignments between your target the template. Give the name of the program you used, and any options you chose. Give a short justification for using local or global alignment.

[3]

Using pairwise sequence alignment between template and target, you do not use all possible information. Please indicate which information you cannot include in your pairwise sequence alignment, that could be helpful for making an accurate alignment. Also make a suggestion how such information could be included.

[5]

Note that you need the a PIR alignment format as input for Modeller, you may use an online converter, see notes on BB.

5 - Modeller

Use Modeller to create models from the alignments and templates. You should should end up with 2 models, for each alignment technique. However, you should initially make multiple models for each alignment and pick the “best one” based on the DOPE score.

Discuss the options you used - you may try different setting, make sure to use the same options for each alignment so your results remain comparable. Discuss how you decided on your final 2 models. You should also describe any problems you had with Modeller, and your solution to the problems.

Give your models, and any script you used to create these. And the DOPE scores for all the models.

[15]

6 - GDT

Calculate GDT_TS scores between each of your models and the solution structure. You can use the LGA program. Discuss your results. What does your GDT score indicate? What was your best model, and why do you think that modelling route was most successful? Was the DOPE score predictive?

[7]

7 - Structural Comparison

Choose a suitable method to compare the template structure with the solution structure. You can use online servers. Explain the choice of method.

Use the same comparison method to compare your models with the solution structure.

Give the scores from your structural comparison - and explain what they mean. Is it fair to compare these scores? Discuss the 'added value' of Modeller: is the model created by Modeller closer to the solution structure than your template?

[10]

8 - Paper by Forrest et al.

Forrest et al. compare alignments for transmembrane proteins, using an approach similar to the method you have just used. Why may alignment programs for trans-membrane proteins be less accurate? Describe how they obtained the results given in Figure 1 & 2. What results do they find in figures 1 & 2? How could you explain these results?

[15]

9 - Free modelling

- Create a structural alignment between your template and your “solution” structure, and create a new model structure that uses this alignment. How does this perform compared to your sequence based alignments? Is this as you would expect, discuss.
- Try to improve your GDT_TS score (you may not succeed). You could try using FUGUE, multiple templates, alignment correction by hand, functional input. Explain what you have done in detail, and explain if this approach indeed has increased the GDT_TS score and why. Also explain if you could have made the same choices if you did not have the solution structure.
- Redo your analysis on a more difficult CASP target
- Try to model a larger part / additional parts of your query sequence, and redo your analysis.

[up to 20 points, approximately 10 points per attempt]

10 - Group

Please indicate which people were in your group, and how each person contributed.