**Laboratory of Bioinformatics 2018-12-20**

Notes on project writing

Every statement claiming something that has been demonstrated must be followed by the respective reference (like in every normal article)

Introduce the sequence you’re modelling, then the family the protein belongs to and key information on the active site. Describe it well! → derive info from the original article

If there are ions (or other cofactors) it should come up in the introduction.

English: don’t put too many aims in one sentence.

functional annotation: done after the evaluation of the structure on the basis of the given template of the same family. the procedure was carried out by adopting homology comparison.

Any introduction should contain: the aim of the report (functional annotation); information derived from the paper describing the template.

introduction: name of the effort(functional annotation), information derived from the paper describing the template

Try to focus on this concept: the protein on TREMBL is poorly annotated, what we are doing is trying to improve this by functional annotations by means of building by homology.

* Information derived from a paper describing the paper
* models
* struct. comparison

In a very formal paper, in the introduction we put at the end another sentence: “we were able to demonstrate that the protein is a laccase, and we were able to transfer a set of GO terms etc. etc.”

methods make you able to transfer a rich set of go terms from the template to the target protein l lm

you may do whatever you want to do, the important is that you make some considerations

Bring URLs of used tools or sources in references

Could you please point me to where in the heavens and hells I can find the current release from RCSB PDB? I had some difficulties with it too

the ec number of the protein was derived from ec pdb so you can *e*lso have the reference and the release for ec pdb

PDBSUM: spatial restraints that were derived from pdb sum .

Highlight the relationship between the Uniprot and PDB entries for the Tve laccase (we use the PDB file for the structure and sequence, but eventually the pl

Each project, as soon as they’re ready, should be addressed to Casadio, Aggazio and Babbi altogether.

[rita.casadio@unibo.it](mailto:rita.casadio@unibo.it) [giulia.babbi3@unibo.it](mailto:giulia.babbi3@unibo.it) [francesco.aggazio2@unibo.it](mailto:francesco.aggazio2@unibo.it)

the template protein was found browsing (or aligning) the target with blastp

“browsing the target against the uniprotkbblablabla with blastp version blabla”

* retrieving the template (using the uniprot kb database in the section related to structure)

nb:what is blosum 62? It’s the substitution matrix used for scoring the pairwise matches (similar to PAM).

* in order to compare the template sequence to the target sequence, we run lalign

(global or local alignment)

* modeller was downloaded from…. and was used to generate 3 models of the target structure.

Avoid passive forms

jce and modeller have been installed in house, describe the technical operation.

Present tense is suggested instead of was was was was

( think she said sequence but meant residues)

coverage=number of sequence in the template/number of sequence in the target

if you are using the sequence of the pdb were there is no signal peptide, so computing the coverage you have to delete the signal peptide also from the target.

template selection:

role of the protein family in ensuring the structure conservation in relation to the function conservation.

* the procedure of building by homology needs the identification of the proteins in a protein family.

FAMILy:laccase.

functional annotation requires to attribute the target within a family with a known function

in order to do it you try to browse the target against the sequences of uniprot endowed with a structure

browsing the uniprot-pdb with blast we ended up with a list of template .

MUst compute the pairwise alignment of target/template with right sequence

why the e value is 0.0?this is the highest score?

I don’t think the E-value can ever be exactly zero, it’s probably 0.0000000000 something

or can it be zero? help [Not real answers, but something to be aware of: 1) E value is a function of the size of the database of sequences against of which the query is blasted, and could be estimated empirically (e.g. randomizing the databases sequences, keeping the symbols’ frequence) so it could actually be == 0. 2) Extremely small numbers cannot/it’s not useful to/it’s not convenient to be represented in bits. In many instances values < 10^-16 are rounded to zero]

identity 62%.

Results → Should be focused on what you want to do

functional annotation is very much depending on the protein family on which the protein is associated to.

function is conserved through evolution , when the structure is conserver independently on the sequence conservation.

problem for the template selection:

the template is the structure that corresponds to a given function within a protein family.

As a rule-of-thumb, a protein that at some point has to pass a membrane to reach its definitive location is in 90% of the cases endowed with a signal peptide.

When in doubt can produce more models starting from different templates → choose the template that produces the best model

Highlight the conservation of crucial amino acids in the lalign output image.

It’s not that we remove the signal peptide to increase the quality of the alignment, but we do it instead to adapt target to template.

Figure references: in text body and image caption the references should have the same format (if I write “figure 9” in text body I should also write “figure 9” in the caption and viceversa if I use Figure 9)

input preparation: for modeller

Evaluation of the various models: be consistent in the description! Explain in a linear way how you evaluate the various parameters and then also comment the Ramachandran plot of the chosen protein. Don’t just splash it there.

“this is the ramachandran plot elò elò” cocco bello

in order to verify if the active site is conserve (and so the function of the protein):

“my suggestion is set the background white: because otherwise PRRRRRRRRRRT”

make accurate comparison of distances of copper ions.

what’s going on in the enviroment of the different coppers:Determining the neighborhood of the his residues. This is done by measuring conservation of all the relative distances in the active site. This will make us be able to conclude that our protein is/is not an active protein.

you can do a table commenting the difference between the template and the target you observed in rasmol

(relative distances ec…)

Observations to add

* active site conformation, copper distances
* disulfide bridges
* conservation of Asn (glycosylation sites). How many are conserved? (Davide says 4 over 8 so let’s treasure this little piece of information)

functional annotation:we produced a model of the sequence that faso poor annotate as laccase for the computation of the structure , we proved that…

we can transfer the GO terms to our target.

Measure all

We are not doing experiments => we partecipate in a research and say what to characterize and what not

Check that 2 sequences start at the same point!

Check the reference article results for inspiration

electron densities are those convincing us of the structure => fitting them with the skeleton of the putative model

when you have structure of template and have domains you may even say when superimposing that they are fairly well conserved

try to mimic the description of the domains from the article =>

topological organization at the level of the ??

The model produced has a similar domain architecture overall. Highlight the disulphide bonds

Multicopper

ramachandran plot: projection of the protein in the plane. this plane is a representation of all the possible pairwise values of the couples of dihedral angles that the main chain can assume in the space or in the solvent.)

alpha elixes mean: specific psi and phi values at the level of the backbone

i have a distribution of values reflected in the red region (alpha elix).

now i have to compare proteins with the same resolution.

THEY INVENTED THIS WAY OF REPRESENTING THE SPACE OF THE PROTEIN.

you have constrains in the flexibility of the protein that corresponds to the peptide bonds of the protein.

these are fixed point. and whatever remains flexible are the c-alpha atoms (covalently bond with the lateral side residues).

when im saying alpha helix i mean a series of phi psi values at the level of the backbone that strictly have a certain range of values => from 4 to 35-50 residues depending on the protein

alpha helix => sistematically in the range of -45 - -50

**a ramachandran plot is like squeezing a protein in a plane => output:reference region in the background with all the lateral side chains represented in this plane**

no distinction between parallel and antiparallel betasheets

we are now describing the protein in the plane at 360 degrees

by convention the alpha helixes are in the -- region

beta sheets in the -+ (or +-, didnt understand) region

50% of residues in the pdb are in coil regions, 35% in alpha helixes and 25% in beta sheets

she put a table with the exact values for alpha helixes

i should compare proteins with the same resolution

even in 1gyc there are residues in disallowed regions => nothing is perfect in life not even a crystal

-prolins are not treated the same => they are the biggest residues

-glycins are too easy

<http://servicesn.mbi.ucla.edu/PROCHECK/>

FIND A TETRAMER FOR p53

BIG TRICK (\*thrick) IN THE LAST EXERCISE => FUNCTIONAL ANNOTATION

!!!TRANSGLUTAMINASE!!!

culo :-)

her googling skills are astounding

In che città è il Martelli in questa foto?? venezia? Possibile

Evolution stopped when achieved Pier Luigi’s smile

**Mattteo Bolner you are a hero**

**You made her happy**

**And our afternoon not miserable (AGREE)thank you**

**orco ladropi**

unothing is perfect in life

except pierluigi’s smile

This hammer has 100% chance to steal your man.

più lo sguardo più mi immanoro