

**BME 110L / BIOL 181L**

## **Computational Biology Tools**

**[www.soe.ucsc.edu/classes/bme110/Winter09](http://www.soe.ucsc.edu/classes/bme110/Winter09)**

**January 27:**

- Difficulties with JAVA applets/webstart on your laptops?
- A note re PSI-BLAST (something we noticed last year)
- Caring about proteins (rather than DNA): overview
  - single protein sequence analysis
  - protein databases
  - a mention of protein structure (more later in the course)  
(Slides by Prof. Carol Rohl).

Posted for self-study (and/or asking questions):

B4D-derived slides "Searching Sequence Databases" (Winter'08)

Accompanying Reading (B4D): Chp 4+6

## **A few things re In-Class "Jan22"**

- **JAVA tools must work ON YOUR LAPTOPS  
(though JDotter may be quite slow)**

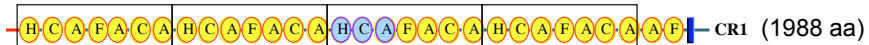
try in your own time - the outputs and options of the two programs should look roughly the same

a quick overview of CR1 right after this -

it is a repetitive protein (check out the Dotlet examples for another one)

does what you see in the dot plots make sense now?

# Complement Receptor 1 (CR1)



Automated large-scale protein structure modelling of individual human Complement Control Protein (CCP)-modules

The five human Regulators of complement activation (RCA) proteins - factor H, complement receptor type 1 (CR1), membrane cofactor protein (MCP), decay accelerating factor (DAF), and C4b-binding protein (C4BP) - are crucial for ensuring that a complement-mediated immune response is proportionate and directed against the infectious agent. They are built up from multiple examples of the complement control protein (CCP)-module (also known as sushi domains and SCR repeats). The five RCA proteins contain a total of 69 modules. Each module has between 50-71 amino acids and is separated from its neighbour by short linking sequences. The 3D structures of a diverse range of CCP Modules, including eleven RCA modules, have been experimentally determined over the last decade. Their compact all beta-domain fold is characterised by a consensus "scaffold" of conserved residues, involving two disulphide bridges and a buried tryptophan. Using this knowledge, a master alignment for this family and other CCP modules has been proposed. This was used to guide a strategy for automated modelling of the individual modules, based on most similar homologues with known structures. A total of 49 medium-resolution models were obtained using the program Modeller (Sali, Blundell, 1993). These models provide a basis for rationalising the extensive existing functional data on RCA proteins that have arisen from mutagenesis experiments. The large scale modelling process was extended to include a larger representative set of human CCP module containing proteins. As a result, a total of 136 modules were modelled by homology from 26 different CCP-containing proteins. The models will be made available publicly through this website and updated as more similar template structures become available.

List of **modelled CCP modules** with names and accession numbers

List of **experimentally solved CCP modules**

Table of residue lengths of CCP module domain boundaries and inter-CCP module linkers in RCA proteins [PDF]

Table of buried surface area for CCP bimodules [PDF]

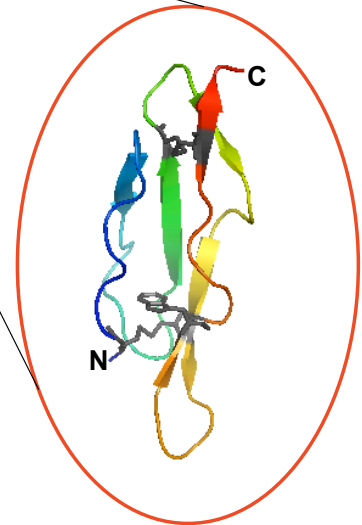
The molecular architectures of the set of human CCP module containing proteins are taken from **SMART** (Schultz et al., 1998; Letunic et al., 2002) and the CCP modules are numbered according to SMART annotations. In this work, module domain boundaries for each CCP module were considered from one residue before the first cysteine until three residues after the fourth cysteine (both inclusive).

Click directly on the CCP module to open or save the PDB file. You need to install **Clustal** to directly load up the molecule within your browser. All experimentally solved structure modules have an "EXP-" label preceding the module name (Viewed best using Internet Explorer). If a particular module was not modelled, a label "NO MODEL-" precedes the module name. In these cases, the link provided should take you directly to the SMART annotation for that bit of the protein sequence. Please use the labels provided as a guide, else directly check against the complete tabulated list provided above.

For collaborative enquiries on modelling your CCP module of choice (if not on the list, email Paul Barlow [Paul.Barlow@bru.ac.uk](mailto:Paul.Barlow@bru.ac.uk))

Human CCP module containing proteins involved in complement

DAF MCP VCP



<http://www.bru.ed.ac.uk/~dinesh/ccp-db.html>  
Soares et al. *PEDS* (2005), 18:379-88

## PSI-BLAST issues & notes

- **The single most important thing is that you know precisely how to “talk E-values”**
  - orders of magnitude are expressed as (e.g. for 0.0001) 1E-04 or  $10^{-4}$  or maybe 1e-04 (but not  $10^{-4}$ )
  - know what “better than” means w.r.t. the numerical value, etc.
- **It is true that the unchecking things manually was merely educational (this could have been done via inclusion value)**
- **The procedure you were shown was reasonably conservative (again, the unchecking, or inclusion cut-off) - this was to minimize the risk of “PSSM corruption”**

## PSI-BLAST issues & notes

- **We noticed some “inexplicable” behaviour which probably has to do with overly sophisticated www-page design causing parameter settings that sometimes are not as on the form...**

just to be absolutely sure that this doesn't happen we may recommend on the exam instructions that you restart your browser before you run PSI-BLAST (if we had a question asking you to do so)

we recommend NOT to use Internet Explorer since that browser seemed to even write funny characters into FASTA files when copy-pasting

- **In any case, if you describe what you did+ what parameters you changed, then of course we'd realise if a false result snuck in because of something like this - so no worries :-)**