Protein sub-family analysis

Geoff Barton
Division of Computational Biology
College of Life Sciences
University of Dundee, UK

twitter:@gjbarton
blog: geoffbarton.wordpress.com

www.compbio.dundee.ac.uk
www.jalview.org

Jalview Training Course – May 2015
Identification of functional sites

- Whole alignment methods
  - Simple visualisation
  - Calculation of “conservation values”

- Sub-family analysis
  - AMAS analysis
  - Tree determinant positions
  - “Evolutionary trace”
Example Multiple Sequence alignment of 27 SH2 domains
AMAS method of calculating conservation

Taylor Venn diagram of amino acid properties

Count maximum number of set boundaries that must be crossed to include all amino acids at an alignment position

This gives a measure of the physico-chemical property variability at the alignment position

For comprehensive review of methods see:

Example Multiple Sequence alignment of 27 SH2 domains

Eyeball the alignment...
Identification of functional sites

- Whole alignment methods
  - Simple visualisation
  - Calculation of “conservation values”

- Sub-family analysis
  - AMAS analysis
  - “Tree determinant” positions
  - “Evolutionary trace”
Sequence analysis of the Annexins: An example of sub-family analysis

- “Large” number of sequences (for 1990)
- Possess multiple domains
- Unknown tertiary structure at the time of analysis

Annexins

Calcium and phospholipid binding

Wide family - 22 known sequences
(Insect - Human)

Found in many cell types

Implicated in

membrane fusion
exocytosis
cell signalling
anti-inflammatory properties
Annexins

Homologous domains
calcium and phospholipid
binding

Variable
N-terminal
region

1 2 3 4

Longer link

Short linkers

Annexin VI has 8 repeats
Sequence Analysis of Annexin Domains

Dot-Plot comparison of Human Annexin I with itself.

Four repeats (domains ?) are visible.
Annexin Predictions

1. 5 Helices

2. Core residues (hydrophobic patterns)

3. Conserved Glu in repeats II
   and Arg in repeats IV
   form a salt bridge

4. Helix a in repeat III shorter

5. Not like uteroglobin

6. Helix a - helix b loop important in repeat III
Annexin V showing Glu-Arg salt bridge between helix 2 of domain II and helix 2 of domain IV
Analysis of similarities and differences between sub-families can reveal functionally important residues

Generalise lessons learned in Annexin study
Annexin Domains

Domain 3 & 7

Domain 4 & 8

Domain 1 & 5

Domain 2 & 6

S.D. Score

10 15 20 25 30 35
Principles of sub-family analysis

See what happens to conservation when you put two sub-families together.

Does it stay high?

Does it go from high to low?

Imply position is important to both sub-families, but the position is important for novel features of the two sub-families.

Imply position is important to both but the position is important for novel features of the two sub-families.
References on Sub-family analysis

- Methods Enzymol paper includes summary of first paper.
- Copies are available on: http://www.compbio.dundee.ac.uk/ftp/pdf/
Identification of functional sites

- Whole alignment methods
  - Simple visualisation
  - Calculation of “conservation values”

- Sub-family analysis
  - AMAS analysis
  - “Tree determinant” positions
  - “Evolutionary trace”
Tree determinants

- Positions in the alignment that are most responsible for the topology of the phylogenetic tree derived from the complete alignment

- These positions may be functionally important.
Generate Local Matrix

Generate Global Distance Matrix

Saved Global Matrix

Compare

Correlation Coefficient
Partial Multiple Sequence Alignment (Globins)
Observed Haeme Contacts (green)
Observed Protein Contacts (yellow)

Prediction (window size 3, interval 2)
Papers on tree determinants/predicting functional sites


Practicals on sub-family analysis