Looking for the unusual: Hunting protein interactions with help of structural and evolutionary thinking... and a bit of common sense

Dietlind L. Gerloff
BME Department

gerloff@soe.ucsc.edu

www.soe.ucsc.edu/~gerloff
The Gerloff group is active in many areas...
www.soe.ucsc.edu/~gerloff/group

Facilitating Software

• **Data Visualisation/User Interface Design**
  www.malariagenomeexplore.org
  (UCSC: Herbert Lee)

Data Sharing between Laboratories

• **Peer-to-Peer Infrastructure “OpenKnowledge”**
  www.openk.org
  (UCSC: Jonathan Magasin)
Today: Protein-Protein Interaction Challenges + Structural Bioinformatics

Applied Bioinformatics
- CDK/cyclin complex prediction:
  Partner prediction between paralogous sets

Method Development
- 1-D Electrostatic Surface Profiles:
  Site prediction on paralogous domains
1-D Surface Profiles:

A Simplified Representation of Electrostatics on Model Surfaces for Protein Interaction Prediction

New potential PPI sites on Complement Receptor 1 (CR1)

University of Edinburgh:
Shakir Ali
Dinesh Soares
Rupert König
Paul Barlow

UC Santa Cruz:
Marcos Woehrmann
Eric Scott, John Archie

University of Wyoming:
David Liberles
Background Motivation

- Surface properties are informative for predicting protein-protein interactions (binding site and/or partner)
- Use modelled structures (not high-resolution crystal structures)

3-D context though approximate
Background Motivation

- Surface properties are informative for predicting protein-protein interactions (binding site and/or partner).
- Use modelled structures (not high-resolution crystal structures).

True CDK2  
Model CDK2  
Template (CDK6) 

(Grasp Electrostatic Surfaces)
Background Motivation

- Surface properties are informative for predicting protein-protein interactions (binding site and/or partner)
- Use modelled structures (not high-resolution crystal structures)
- 3-D context though approximate
  - Method for “screening GRASP cartoons”
  - Tackle some of currently neglected targets/problems
Two different applications for screening/comparison tools

- compare surfaces of homologous proteins:

  Related work: PIPSA (Blomberg et al., Proteins 1999, 37:379-87);
  SURF’S UP (Sasin et al., J Biosci 2007, 32:97-100).

- evaluate hypothetical interfaces in hetero-dimeric (non-homologous) proteins:

  Related work: MOLSURFER (Gabadoulline et al., NAR 2003, 31:3349-51).
Two different applications for screening/comparison tools

- compare surfaces of homologous proteins:


  binding site ?
  (partner unknown)

- evaluate hypothetical interfaces in hetero-dimeric (non-homologous) proteins:

  Related work: MOLSURFER (Gabadoulline et al., *NAR* 2003, 31:3349-51).

  binding partner ?
How would electrostatic surface comparison be helpful?

Conserved function:

• Specific interactions (molecular recognition):
  - conservation pressure better detected by sequence
  - partner proteins differ between homologs
How would electrostatic surface comparison be helpful?

**Conserved function:**

- Specific interactions (molecular recognition):
  - conservation pressure better detected by sequence
  - partner proteins differ between homologs

- Unspecific interactions:
  - recurrent observation of a patch within a family (?)

**Varying/New function:**

- In some cases, new/different interaction sites may exhibit “excessive” change in surface charge distribution, due to adaptation to the binding partner -

  look for unusually different sites
Complement Receptor 1 (CR1)

http://www.bionmr.chem.ed.ac.uk/bionmr/public_html/ccp-db.html
Soares et al. Peds (2005), 18:379-88
1-D Electrostatic Profiles

= (context-dependent) charge of solvent-accessible atoms, apportioned per amino acid

Example: CR1~16 (order sequential)

PDB2PQR\(^1\):  
- add hydrogens  
- compute charge

NACCESS\(^2\):  
- > 0.15 accessible

1-D Profile: 
summed per aa

\(^1\): Dolinsky et al. (2004), *NAR* 32:W665-7 \(^2\): Hubbard & Thornton (1993)
Order: 3, 61, 1, 6, 9, 10, 18

Structural Neighbour-Enriched Orders:
Generate structural neighbour-enriched orders (no knowledge of location of binding site):

• Tabulate matrix of spatially adjacent residues (Cβ-Cβ < 8Å)
• Path-walking algorithm favouring 3-D neighbors nearby
• Compute 1000 different combinations of residue orders

Order1:
1 27 29 32 31 66 67 68 73 74 4 7 8 18 21 20 33 63 65 75 78 79 82 11 10 9 6 28 23 37 35 19 36 34 59 58 80 81 84 85 49 12 40 41 38 53 52 51 50 86 42

Order2:
8 9 7 6 4 28 23 35 18 37 12 10 11 82 79 78 75 74 73 67 68 1 31 20 19 38 40 49 81 80 63 65 66 27 29 32 33 21 36 53 51 50 42 41 86 85 84 52 58 59 34

Order3:
78 75 65 63 34 59 58 79 80 82 81 11 9 7 4 7 3 6 67 66 31 32 33 35 20 19 18 38 53 52 51 50 84 85 86 42 41 40 49 10 8 37 6 28 4 27 1 29 23 21 36 12

Random orders: 85% of windows with 0 or 1 (of 6) pairs adjacent
Structure-based random orders: 73% with 2, 3 or 4 (of 6) pairs adjacent
1-D Electrostatic Profiles

= (context-dependent) charge of solvent-accessible atoms, apportioned per amino acid

Example: CR1~16, CR1~02 (in one neighbor-enriched order)

For each order:
- compare pair-wise
- sliding window
- “distance” value

Distance matrix (by module)

1000 orders, each 1 tree/clustering

consensus tree
Sequence Tree:

“Surface Tree”:

(Ali et al., manuscript in prep.)
Sequence Tree:

“Surface Tree“:

(Ali et al., manuscript in prep.)
Major Problem: Harly any information to validate

Able to “screen GRASP cartoons” - but are these binding sites?

Corroboration / overlay with other clues:

• Local nature of the signal (surface patch)
• Adaptive Evolution?

$K_a/K_s$ or $(K_a - K_s)$ analysis of coding sequences:

$K_a$: non-synonymous mutation rate
$K_s$: synonymous mutation rate

$\rightarrow K_a > K_s$ gene is under positive selection
Tertiary Windowing for $K_a/K_s$
(David Liberles, University of Wyoming)

- Local (3-D structural) $K_a$ vs $K_s$ calculations
- Attribute to individual lineages in the evolutionary tree

For example:

CR1 orthologous genes from: human, chimp, rhesus, baboon

Tertiary Windowing for $K_a/K_s$
(David Liberles, University of Wyoming)

$K_a > K_s$ ~ neutral
too few contacts
Tertiary Windowing for $K_a/K_s$
(David Liberles, University of Wyoming)

**Protein sequence variation**
- green: module 02 special
- blue: module-specific
- cyan: highly variable
Tertiary Windowing for $K_a/K_s$  
(David Liberles, University of Wyoming)

$K_a > K_s$  
~ neutral  
too few contacts
Take Home Bioinformatics

- a simplistic new representation for surface properties
- various applications - but difficult to test!
  not discussed: potential use in evaluating complementarity in hypothetical complexes with known binding sites (partner prediction)

- Needs more examples - what is the application range?

We need to produce methods that can be used on modelled structures …
Take Home
Complement Receptor 1

• could the potential interaction site be the 24-25 junction?
• intriguingly, there are three polymorphisms on module 25 that are attributed a protective effect against malaria…
  if this were due to a PPI with a human partner protein our prediction could be very important!

• “Go fishing” experiments next?

A (cheap) experiment seems the way forward…
  wish us luck 😊