Integrative and quantitative analysis of disease mutations in protein interaction networks and implications for personalized medicine

Christina Kiel, Staff scientist
Department of Systems Biology, Luis Serrano group
CRG Barcelona

CRG Barcelona: http://crg.eu
Exploring the molecular and quantitative mechanisms that underlie cell signaling and contribute to human disease.
Factors that could affect signaling

Outline

I. The effect of affinities, kinetic constants and network topology in PPI networks

II. The effect of protein abundance perturbations and interaction competition in PPI networks

III. Methods to quantify protein abundances, affinities, and kinetic constants

IV. Disease mutations and their principle effect on PPI networks

V. Examples for quantitative effects in disease networks
   1. RASopathy vs cancer mutations: a matter of quantity
   2. Rhodopsin stability and disease onset
   3. BRAF mutation frequency: prediction of oncogenic drivers

VI. Summary tools & websites

VII. Wrap up/ discussion/ conclusions
Quantitative information in protein-protein interaction (PPI) networks

Qualitative PPI networks

Quantitative PPI networks

Considering protein abundances and affinities/kinetic constants

\[ \text{cellular abundance} \]

\[ K_d \]

\[ k_{on} \]

\[ k_{off} \]
The effect of affinities, kinetic constants and network topology in PPI networks

Kiel & Serrano, Science Signal, 2009
Epidermal growth factor (EGF) activates the RAS-RAF-MEK-ERK pathway

I. Kinetic perturbations and network topology
Different network ‘wiring’ /feedbacks causes the different behaviour

I. Kinetic perturbations and network topology

HEK293 cells

Transient response

RK13 cells

Sustained response

Kiel & Serrano, Sci Signal, 2009
A simple computer model of ERK activation in HEK293 and RK13 cells

I. Kinetic perturbations and network topology

- No negative feedback from ERK to Sos1 in the RK13-like model

- Good agreement of experiment and model predictions

Kiel & Serrano, Sci Signal, 2009
Model predictions: different cell type-specific wiring results in different responses to mutations with affinity perturbations.

- Strong feedback
- Weak feedback

Kinetic perturbations

- Mutations can have different cell type (patient!)-specific effects

IN SILICO (HEK283 model)

No significant changes

IN SILICO (RK13 model)

Significant differences

Kiel & Serrano, Sci Signal, 2009
II. Protein abundances and competition

The effect of protein abundance perturbations and interaction competition in PPI networks

Mutually exclusive interface interaction, XOR
II. Protein abundances and competition

How could interaction competition and protein concentration affect downstream signaling?

Signaling complexes: > 300 partners for one protein??

Some proteins will use similar binding surfaces for interaction with other molecules: ‘mutually exclusive interactions’/ ‘XOR’
II. Protein abundances and competition

How could interaction competition and protein concentration affect downstream signaling?

Signaling complexes: > 300 partners for one protein??

In a simple world: concentration and $k_D$ will determine the signaling output

Changes in concentration (ie mutations at promoters, enhancers etc..) could have an effect in signalling
A bioinformatics tool to distinguish mutually exclusive from compatible interactions in large-scale PPI

**SAPIN** (structural analysis of protein interaction networks)

[Webserver](http://sapin.crg.es/)

Yang et al, Bioinformatics, 2012
Experimental methods to quantify protein abundances, affinities, and kinetic constants

III. Quantitative experimental methods: protein abundances and interactions
Why proteomics in times of deep RNA sequencing?

- mRNA does not translate 1:1 into protein; keywords:
  (i) translation efficiency,
  (ii) mRNA stability,
  (iii) protein stability,

- Posttranslational modification (PTMs) of proteins, e.g. phosphorylation

Two main aims: IDENTIFICATION and QUANTIFICATION

Two main techniques: MASS SPECTROMETRY and ANTIBODY-BASED
30,000 coding genes per cell

Alt. splicing: 2-3 x 30,000 = 90,000 proteins

Post-translational modifications
> 10 x 90,000
= 900,000 proteins
High dynamic range of the proteome

III. Quantitative experimental methods: protein abundances and interactions
Address problem of cellular complexity by fractionation, e.g. liquid chromatography

Address problem of cellular dynamic range by better and better (and better…) mass spectrometers…

Ahrens et al, 2010
Human deep proteome mapping

Beck et al, MSB, 2011
~10,000 proteins quantified

Mann lab
10,255 proteins quantified
Nagaraj et al, MSB, 2011
Many proteins are identified with peptides belonging to more than one protein (e.g. isoforms)
Antibody-based proteomics: only semi-quantitative abundances

- Tissue-based map of the human proteome
- 44 major tissues and organs in the human body
- 24,028 antibodies corresponding to 16,975 protein-encoding genes
Quantitative Western blotting

III. Quantitative experimental methods: protein abundances and interactions

A. SDS-page, WB α-His, WB α-Cdc42

B. Hek293, Keratinocytes, MCF-7

C. Intensity (a.u.) vs [Cdc42] (ng)

Protein standards: expression, purification and quantification

Summary statistic for quantitative Western blotting of 198 ErbB-related proteins

- No antibody (19.7%)
- Not detected (28.3%)
- Protein family detected (5.5%)
- Protein detected (25.2%)
- Protein detected (lower quality blot) (8.1%)
- Bad antibody (13.6%)

Kiel et al, J Prot Res, 2014
Combining different quantitative approaches to quantify 198 proteins in the ErbB signaling pathway

- SRM has a higher sensitivity compared to quantitative western blotting (but some proteins are only detected by Western blotting)
- Problem with isoforms and protein families: as a consequence of frequent gene duplication events in mammals, often similar proteins (e.g. AKT1 and AKT2) cannot be distinguished using the peptides detected by MS. > they can only be assigned to a protein group/ family

Kiel et al, J Prot Res, 2014
Measuring protein interactions *in vivo* and *in vitro*

The challenge:
- Most *in vivo* techniques are high-throughput, but do not provide affinities (only qualitative binding detection).
- *In vitro* techniques can provide affinities and kinetic constants, but are not high-throughput methods.

*Piehler, Curr Opin Struct Biol, 2005*
Measuring protein affinities in vitro requires the expression and purification of proteins (e.g. using bacteria)

Example: Bacterial expressed and purified Ras protein mutants and interactors

Large proteins are often not soluble: expression and purification of protein domains

III. Quantitative experimental methods: protein abundances and interactions
III. Quantitative experimental methods: protein abundances and interactions

Two main methods to measure affinities and kinetic constants

Microscale thermophoresis

Surface plasmon resonance

- Provides only the affinity in equilibrium ($K_d$ value), but not kinetic constants
- Provides kinetic constants ($k_{on}$ and $k_{off}$)

$K_d = \frac{[A] \times [B]}{[AB]}$

$K_d = \frac{k_{off}}{k_{on}}$
The effect of abundance variation at XOR network motifs

- The output/ function depends on both, network structure and abundance: we need to know the network very well to understand.

Competition at the Ras XOR node

Experimental abundances

Network motif

The Ras XOR node

Mathematical network modeling: increasing RIN1 to 10-fold higher of CRAF expression should decrease CRAF activation
### Experimental testing of competition at the Ras node

**IN VITRO**

<table>
<thead>
<tr>
<th></th>
<th>MCF-7</th>
<th>HEK293</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5 min HRG</td>
<td>5 min EGF</td>
</tr>
<tr>
<td>U</td>
<td>RIN1</td>
<td>U</td>
</tr>
</tbody>
</table>

- **α-CRAF-p**
- **α-MEK-p**
- **α-ERK-p**
- **α-RSK-p**
- LC (actin)

**Expression of RIN1 in MCF-7 and HEK293 cells decreases CRAF, MEK, and ERK activation**

- Alterations in the abundance of one of two hub-binding partners affected downstream signaling

---

IV. Rewiring through disease mutations

Qualitative and quantitative effects of disease mutations

Disease mutation
IV. Rewiring through disease mutations

General concepts of interaction (‘edge’) rewiring

- Mutation affecting folding
  - Protein abundance / folding
  
- Protein abundance changes
  - Protein abundance / interaction competition
    - Kiel et al, 2013
    - Romano (Kolch) et al, 2014

- Alternative splicing

- Mutation affecting binding on the surface of one domain:

- ‘edgetics’
  - Zhong (Vidal) et al, 2009

- ‘enedgetics’
  - Kiel & Serrano, 2014
Examples how missense mutations can affect the network: a 3D structural perspective

Class 1a  Gain in signaling through release of autoinhibition

PTPN11 (2SH2)
Examples how missense mutations can affect the network: a 3D structural perspective
Examples how missense mutations can affect the network: a 3D structural perspective

Class 2

Gain in signaling through loss of interaction with inhibitors/deactivating proteins

Complex of 14-3-3 with peptide of Raf1 (3I0J)

Kiel & Serrano, Mol Sys Biol, 2014
Examples how missense mutations can affect the network: a 3D structural perspective

Class 3

Folding affected (destabilization of protein) ; gain in signaling for NF1 and RASA1

NF1
(1NF1)

Kiel & Serrano, Mol Sys Biol, 2014
Examples how missense mutations can affect the network: a 3D structural perspective
Examples how missense mutations can affect the network: a 3D structural perspective
Example 1: RASopathy and cancer disease mutations

RASopathies: Developmental syndromes of Ras/ MAPK pathway dysregulation

- RASopathies are a group of developmental disorders characterized by postnatal reduced growth, facial dysmorphism, cardiac defects, mental retardation, skin defects, musculo-skeletal defects, short stature, cryptorchidism.
- RASopathies are caused by **germline mutations** in genes that encode protein components of the Ras/12 proteins involved (HRAS, NF1, MAP2K1, MAP2K2, RASA1, SPRED1, SOS1, PTPN11, RAF1, KRAS, NRAS, BRAF).
- The majority of mutations result in increased signal transduction down the Ras/MAPK pathway, but usually to a smaller extent than somatic mutations associated with cancer.

### Somatic Mutations
- Occur in non-germline tissues.
- Are non-heritable (do not affect offspring).

### Germline Mutations
- Present in egg or sperm.
- Are heritable (all cells affected in offspring).
V. Examples: 1. RASopathy vs cancer

What are the differences in mutations of the same protein causing different disease (e.g. RASopathies or cancer)?

- Ras/MAPK syndromes (‘RASopathies’) are a class of developmental disorders caused by germline mutations.
- Proteins in Ras/MAPK syndromes (‘RASopathies’) are also found in cancer.

Kiel & Serrano, Mol Sys Biol, 2014
Location of mutations in different domains does not explain the difference between RASopathy and cancer mutations

Distribution of somatic and germline mutations in 98 different structural domains and inter-structural regions

‘Edgetics’ does not explain it
Domain localization of mutation does not explain why a particular mutation will cause RASopathy or cancer

Kiel & Serrano, Mol Sys Biol, 2014
FoldX-based energy calculations of proteins

3D Structural information

A force field for energy calculations and protein design

\[ + \text{FoldX} = \Delta G \]

Relation to affinity: \( \Delta G = RT \ln K_d \)

- Total free energy
- Interaction energy
- Mutagenesis

V. Examples: 1. RASopathy vs cancer
Analysis of 956 missense mutations in RASopathies and cancer based on structural information and FoldX energies

V. Examples: 1. RASopathy vs cancer

B

Class 1a
Gain in signaling through release of autoinhibition

PTPN11 (2SH2)

Class 1b
Gain in signaling through destabilizing mutation in active site; release of autoinhibition in structural segments

BRAF (4EHE)

Class 2
Gain in signaling through loss of interaction with inhibitors/deactivating proteins

Complex of 14-3-3 with peptide of Raf1 (3IQJ)

Class 3
Folding affected (destabilization of protein); gain in signaling for NF1 and RASA1

NF1 (1NF1)

Class 4
Gain in signaling through mutation of domains involved in membrane recruitment

SOS1 (1DBH)

Class 5
No effect; location on surface

SOS1 (1DBH)
Analysis of 956 missense mutations in RASopathies and cancer: high structural coverage

V. Examples: 1. RASopathy vs cancer

Kiel & Serrano, Mol Sys Biol, 2014
Multiple effects of a mutation even for the same protein/protein class

V. Examples: 1. RASopathy vs cancer

Kiel & Serrano, Mol Sys Biol, 2014
Cancer mutations tend to have higher destabilization values (on average)

V. Examples: 1. RASopathy vs cancer

Kiel & Serrano, Mol Sys Biol, 2014
Quantitative effects on protein stability, or activity could explain in some cases the different phenotype: cancer or RASopathy

Simulation of Ras activation

‘Enedgetics’: quantitative edge effects

‘Edgetics’ + energies = ‘enedgetics’
Quantitative effects on protein stability, activity, or folding explains in some cases the different phenotype

Kiel & Serrano, Mol Sys Biol, 2014
V. Examples: 1. RASopathy vs cancer

Compensatory effects of mutations on different interaction partners

Kiel & Serrano, Mol Sys Biol, 2014
Conclusions example 1: RASopathy vs cancer

- A systematic analysis of 956 RASopathy and cancer mutations based on structures and energy predictions is presented.
- Even for the same gene, different disease-causing mechanisms exist depending on the type of mutation.
- Energy changes are higher for cancer compared to RASopathy mutations.
- In some cases, RASopathy mutations show compensatory changes that, as predicted by network modelling, result only in minor pathway deregulation.

- Combined network-based and structural analyses show that quantitative changes rather than all-or-none rewiring underlie the difference between RASopathy and Cancer mutations.
Example 2: Rhodopsin disease mutations
Rhodopsin: involved in light perception in rod outer segment

Understanding disease mutations in rhodopsin, a common cause of retinitis pigmentosa (RP)
V. Examples: 2. Rhodopsin mutations

Analysis of 103 mutations in rhodopsin linked to RP

Is there a correlation between energy changes of rhodopsin missense-mutations and their potential effect on clinical severity of Retinitis Pigmentosa (RP)?

Energy changes + Fold

Rakoczy et al, J Mol Biol, 2011
Several considerations for studying the effect of missense mutations in rhodopsin

1) **Rhodopsin is a membrane protein**: can we use FoldX, a design algorithm developed for *soluble proteins*, for predicting the effect of mutants for a membrane protein?

Region I mutants (intradiscal):
- ✓ YES, not in membrane

Region II mutants (cytoplasm):
- ✓ YES, not in membrane

Region IV mutants (residues pointing outside and facing the lipid bilayer):
- NO, a mutation from hydrophobic to polar residue could be predicted favorable by FoldX, but would prevent proper integration of rhodopsin into the membrane.
For analyzing Region IV mutants (residues pointing outside and facing the lipid bilayer): use a different algorithm

This algorithm is based on experimental results, in which systematically designed 19-residue long amino acid sequences have been expressed and tested in-vitro for TM insertion.

**Figure 1 | The Lep model protein.** *Escherichia coli* leader peptidase (Lep) has two TM helices (TM1 and TM2) and a large luminal domain (P2). It inserts into rough microsomes in an N\textsubscript{lum}-C\textsubscript{lum} orientation. H-segments (red) are engineered into the P2 domain with two flanking Asn-X-Thr glycosylation acceptor sites (G1, G2). Constructs for which the H-segment is integrated into the endoplasmic reticulum membrane as a TM helix are glycosylated only on the G1 site (left), whereas those for which the H-segment is translocated across the membrane are glycosylated on both the G1 and G2 sites (right).

V. Examples: 2. Rhodopsin mutations

Several considerations for studying the effect of missense mutations in rhodopsin

1) **Rhodopsin is a membrane protein:** can we use FoldX, a design algorithm developed for *soluble proteins*, for predicting the effect of mutants for a membrane protein?

Region I mutants (intradiscal):
✓ **YES**, not in membrane

Region II mutants (cytoplasm):
✓ **YES**, not in membrane

Region IV mutants (residues pointing outside and facing the lipid bilayer):
**NO**, a mutation from hydrophobic to polar residue could be predicted favorable by FoldX, but would prevent proper integration of rhodopsin into the membrane.

Region V mutants (residues facing inside the helices):
**NO**, FoldX desolvation effect is possibly not appropriate since the reference state in soluble proteins is water and in membranes, lipids.
**BUT:** **VanderWaal’s clashes** of course will be the same for a soluble or membrane protein. To avoid issues related to the proper calibration of the desolvation effect for buried residues in membrane proteins for residues in Region V we determined both the *overall change in energy* and the **Vander Waals’ clashes**.

---

Rakoczy et al, J Mol Biol, 2011
Several considerations for studying the effect of missense mutations in rhodopsin

2) Retinal-free Rhodopsin is unstable: If an amino acid residue contributes to binding a mutation might not necessarily lead to destabilization (energies of retinal not calibrated) → We need to identify all residues in the retinal binding area, and treat the results of mutations involving these residues, separately.

Figure S9

**Figure S9** Correlation of FoldX ΔΔG values with in vitro expression levels and chromophore binding properties. In vitro data for protein expression levels and chromophore binding capabilities were taken from Kausial and Khorana [19] and Brescic et al. [49]. Involvement of mutants in different Classes (according to the Mendes [20] classification) are indicated. For mutants that show no binding of 11-cis-retinal at all (no peak at 500 nm), we assumed that it corresponds to a ratio A280/A500 of 10. (The reason to assume this was that when A500 is very low, the ratio is round 8 to 9).
V. Examples: 2. Rhodopsin mutations

Several considerations for studying the effect of missense mutations in rhodopsin

3) Rhodopsin is involved in other functions (e.g. binding to partner proteins): A mutation might cause disease but not be predicted destabilizing with FoldX → We need to know as much as possible about rhodopsin function.
Five structures of bovine rhodopsin were selected (<2.6 Å) for mutagenesis and protein stability analysis using FoldX.
FoldX energy results and involvement in other function

Mutants that are destabilizing (\(\Delta \Delta G > 1.6\) kcal/mol)

(a)

![Graph showing FoldX energy results with mutations and their effects on \(\Delta \Delta G\) values, highlighting mutants that are destabilizing with \(\Delta \Delta G > 1.6\) kcal/mol.](image)

Rakoczy et al, J Mol Biol, 2011
FoldX energy results and involvement in other function

V. Examples: 2. Rhodopsin mutations

Mutants that are not destabilizing, are usually involved in other functions, which can explain their disease-causing effect.

Rakoczy et al, J Mol Biol, 2011
FoldX calculations and comparing with phenotypic data

V. Examples: 2. Rhodopsin mutations

Rakoczy et al, J Mol Biol, 2011

[Diagram showing FoldX calculations and comparisons]

Residues cannot be analysed with FoldX: 12 (Region IV, Van Heijne)

- F45L
- L46R
- F52Y
- F53R
- F56Y
- V209M
- F220C
- F220L
- P267L
- P267R
- S270R
- S297R

Bad structure: 2
- C222R
- L328P

Unexplained/misdiagnosed: 4
- G51A
- G284S
- V137M
- T193M

FoldX > 1.6 kcal: 63

Destabilization

Total number of mutations: 103

FoldX < 1.6 kcal: 26

Functional residues: 19
- T4K
- N15S
- T17M
- T94S
- C140S
- R185L
- S186P
- E341K
- T342M
- V345M
- V345L
- A346P
- P347A
- P347S
- P347Q
- P347L
- P347R
- P347T
- P347C

Asymptomatic
- No known function: 3
- V104I
- A117G
- A299S

SIFT

0.43 (tolerated)

0.0043 (non-tolerated)

0.0034 (non-tolerated)
V. Examples: 2. Rhodopsin mutations

**Correlation of daytime vision loss and night blindness with FoldX energy calculations**

Different therapies should be used for the three different types of mutations:

- **Disulphide bridges**
- **Folding mutants**
- **Retinal binding**

Rakoczy et al, J Mol Biol, 2011
Conclusions example 2: Rhodopsin mutations

- The majority of the mutants is located within the hydrophobic core of the corresponding proteins and are therefore likely to cause misfolding.

Quantitative predictive assessment for the severity and onset of the disease:
- For folding mutations where sub-typing was available we found a significant correlation between FoldX energy changes and both the average onset age of night-blindness, daytime vision loss and visual acuity.

Most important conclusion:
- a high level of functional understanding was necessary for our analysis and the observed energy-phenotype correlation.
Example 3: BRAF mutations in cancer. Why V600E?
The most common BRAF mutation is V600E and induces constitutive kinase activation.

Patients are treated with a BRAF kinase inhibitor.

Shall we only treat patients which harbour V600E mutations or also patients with non-V600E mutations?
Catalytic activity of kinases is usually tightly controlled

Mechanisms for kinase activation are:

- phosphorylation
- additional domains or subunits of the kinase
- scaffolding proteins
- kinase dimerization

Mutations in kinases (e.g. BRAF) can cause constitutive kinase activation and over activation of downstream signaling, which can cause cancer
Kinases are activated through mutations in the activation loop (activation segment)

- phosphorylation in the activation segment causes structural rearrangements of the activation segment and the aC helix. This reorients the DFG loop resulting in activation of the kinase
BRAF kinase activation through oncogenic mutations (e.g. V600E)

V600E mimics the negative charge of the neighbouring phosphorylated Thr599-P

Activation loop residues: form strong hydrophobic interactions with the P-loop in the inactive conformation of the kinase, locking the kinase in its inactive state until the activation loop is phosphorylated, destabilizing these interactions with the presence of negative charge. This triggers the shift to the active state of the kinase. Specifically, L597 and V600 of the activation loop interact with G466, F468, and V471 of the P-loop to keep the kinase domain inactive until it is phosphorylated.
Focus on the position Val600 in the kinase BRAF

V600 is buried in a hydrophobic pocket formed by the activation segment (AS) and the aC helix

V600E: mutation hot spot in cancer.

Differences in mutation frequencies: a quantitative effect?
The V600E mutation causes a high destabilization of the inactive state (aC helix/AS hydrophobic pocket).

Kiel et al, Elife, 2016

No destabilization of active state (data not shown)
V. Examples: 3. Why BRAF V600E?

Distinguishing driver from passenger mutations

FoldX

V600K, D, and R have very similar destabilizing energies > cancer driver

Fitness??

V600A, M, and L are not very destabilizing > cancer passenger

Kiel et al, Elife, 2016
V600G behaves more like a RASopathy mutation

Google search for “V600G BRAF CFC syndrome”: V600G found as a RASopathy mutation 😊

Germline mutation in BRAF codon 600 is compatible with human development: de novo p.V600G mutation identified in a patient with CFC syndrome

Champion, K1; Buning, C2; Estep, AL2; Jones, JR1; Bolt, CH1; Rogers, RC1; Rauen, KA3; Everman, DB1

ISSN: 0009-9163 DOI: 10.1111/j.1399-0004.2010.01495.x
Blackwell Publishing Ltd

“enedgetics”
Cancer mutations tend to have higher destabilization values (on average)

Kiel & Serrano, 2014
Different energy thresholds for germline and somatic mutations? ‘Condition-dependent phenotypes’

Condition 1 (germline mutation)
- V600E germline: developmental lethal
  - Phenotype: lethal
- V600G germline: non-lethal, but developmental defects (CFC syndrome)
  - Phenotype: Developmental defects

Condition 2 (somatic mutation)
- V600E somatic: cancer driver
  - Phenotype: Cancer
- V600G somatic: cancer passenger (non-disease causing)
  - Phenotype: normal (or fitness/ or together with other mutations)

V. Examples: 3. Why BRAF V600E?
Why different cancer frequencies for V600E, V600D and V600K?

V600K, D, and R have very similar destabilizing energies

Why is V600E the by far most frequent mutation?

<table>
<thead>
<tr>
<th>aa</th>
<th>frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glu</td>
<td>15474</td>
</tr>
<tr>
<td>Lys</td>
<td>164</td>
</tr>
<tr>
<td>Arg</td>
<td>36</td>
</tr>
<tr>
<td>Met</td>
<td>25</td>
</tr>
<tr>
<td>Ala</td>
<td>22</td>
</tr>
<tr>
<td>Asp</td>
<td>20</td>
</tr>
<tr>
<td>Gly</td>
<td>11</td>
</tr>
<tr>
<td>Leu</td>
<td>2</td>
</tr>
</tbody>
</table>
V. Examples: 3. Why BRAF V600E?

Why different cancer frequencies for V600E, V600D and V600K?

Distinguishing cancer driver from passenger mutations:
Is V600E a driver mutation and V600D a passenger mutation?

On the molecular level: Glu and Asp have similar biochemical properties.
Why different cancer frequencies for V600E, V600D and V600K?

- The higher mutation frequency of V600E compared to V600D can be explained based on the number of nucleotide substitutions needed: V600D requires 2 nucleotide substitutions.
Experimentally validate the effect of BRAF mutations by monitoring downstream MEK activation (HEK293 cells)

Day1: Seed HEK293 cells

Day2: Transfect flag-BRAF WT and mutants

Day3: Lyse cells and Western blot

Kiel et al, Elife, 2016
V. Examples: 3. Why BRAF V600E?

Experimentally validate the effect of BRAF mutations by monitoring downstream MEK activation (HEK293 cells)

- **V600H** (requires 3 nucleotide substitutions) is as active as V600E, but NOT found in cancer. Similarly **L597Y** is not found in rasopathy patients.
Why are no mutations at other positions in the hydrophobic pocket - in a different position to Val600 - found frequently mutated in cancer?

FoldX prediction: other mutations in the hydrophobic pocket destabilize the pocket and may thereby release the AS, would also affect the folding of the inactive and/or active kinase

- Experimentally: lower BRAF expression levels (and MEK phosphorylation)

Kiel et al, Elife, 2016
Conclusions/ Wrap up

- Quantitative information is important to consider in PPI networks; however, it is often difficult to address these quantities experimentally.
- Protein quantification is not a solved problem; especially in mammalian cells, because of the problem of shared peptides for isoforms and splice variants.
- It is impossible to measure binding affinities and kinetic constants in a high-throughput manner (protein expression and purification needed).
- The effect of mutations can be assessed in a quantitative manner using protein design tools, provided 3D structural information is available.
- Structural analysis of mutations could suggest for different therapies for mutations happening at different regions of the protein.
- In GWAS analysis the number of base changes required for a mutation should be considered in the analysis. Two mutations with the same frequency, one could be neutral and the other deleterious if the first one requires on base change and the second one, two.
Acknowledgements

Collaborations cell signaling team:

Walter Kolch, Boris Kholodenko, Katja Rybakova, Lan Nguyen (SBI Dublin)

Edda Klipp, Friedemann Usschner (Humboldt-Universität zu Berlin)

David Lynn, Kenneth Bryan (EMBL Australia)

Owen Sansom, Fatih Ceteci (Beatson Cancer Research, Glasgow)

Jörg Stelling (ETH Zürich)

Marius Ueffing, Karsten Boldt, (University of Tübingen)

Francesc Posas, Klement Stojanovski, Gerhard Seisenbacher (UPF Barcelona)

Ola Söderberg (Uppsala University)

Philippe Bastiaens, Aneta Koseska (MPI Dortmund)

Imre Berger (EMBL Grenoble)
Acknowledgements

Christina Kiel
Hanna Benisty
Violeta Beltran
Martin Schiefer

Systems Biology Group
Conclusions example 3: Why BRAF V600E?

- BRAF mutation frequencies depend on the equilibrium between the destabilization of the hydrophobic pocket, the overall folding energy, the activation of the kinase and the number of bases required to change the corresponding amino acid.

Why BRAF V600E?

- V600E is the only single nucleotide substitution (Asp, Lys, and Arg, require two bases substitutions) that opens the AS through destabilization of autoinhibitory interactions, without significantly impairing the folding of the inactive or active kinase domain.

- The results underscore the importance of considering changes at both the DNA and protein level when attempting to understand why certain cancer-causing mutations are more common than others.
VI. Summary tools & websites

Quantitative PPI networks

- FoldX
- Binding DB
- PDB
- RCB
- 3D Interactome
- dsysmap
- paxdb
- NUMB3R5
Protein abundances

http://pax-db.org/

<table>
<thead>
<tr>
<th>Species</th>
<th>Predicted proteome size</th>
<th>Datasets</th>
<th>Proteins Covered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homo sapiens</td>
<td>20457</td>
<td>170</td>
<td>98%</td>
</tr>
<tr>
<td>Mus musculus</td>
<td>22668</td>
<td>75</td>
<td>90%</td>
</tr>
<tr>
<td>Arabidopsis thaliana</td>
<td>27418</td>
<td>46</td>
<td>76%</td>
</tr>
<tr>
<td>Danio rerio</td>
<td>26103</td>
<td>20</td>
<td>59%</td>
</tr>
<tr>
<td>Escherichia coli. K-12 substr. MG1655</td>
<td>4146</td>
<td>18</td>
<td>98%</td>
</tr>
<tr>
<td>Saccharomyces cerevisiae</td>
<td>8892</td>
<td>17</td>
<td>98%</td>
</tr>
<tr>
<td>Caenorhabditis elegans</td>
<td>20517</td>
<td>10</td>
<td>80%</td>
</tr>
<tr>
<td>Drosophila melanogaster</td>
<td>13397</td>
<td>10</td>
<td>95%</td>
</tr>
<tr>
<td>Schizosaccharomyces pombe</td>
<td>3144</td>
<td>8</td>
<td>90%</td>
</tr>
<tr>
<td>Zea mays</td>
<td>92413</td>
<td>7</td>
<td>18%</td>
</tr>
</tbody>
</table>
Affinities and kinetic constants

The Binding Database

BindingDB is a public, web-accessible database of measured binding affinities, focusing chiefly on the interactions of proteins considered to be drug targets with small, drug-like molecules. BindingDB contains 1,207,891 binding data for 6,265 protein targets and 528,618 small molecules.

There are 2507 protein-ligand crystal structures with BindingDB affinity measurements for proteins with 100% sequence identity, and 7392 crystal structures allowing proteins to 60% sequence identity.

Simple Search
- Article Title: Authors
- Use ? for single-letter wild card or * for general wild card
- For example: "alcohol" or "alcohol*": query cannot start with wild card.

Advanced Search
- Combine multiple search criteria, such as chemical structures, target names, and numerical affinities; restrict searches by data source, such as BindingDB, CHEMBL, PubChem, and Patents.

Messages
- BindingDB's Advanced Search now allows you to download your search results in Excel format. (March 2016)
- We are delighted to announce that Elsevier's Science Direct journals now include links from articles to BindingDB datasets, where available! For example, go to this article, and see the "Data for this Article, BindingDB" link on the right. (December 2015)

Journal Citation by BindingDB
- BindingDB continuously curates a set of journals not covered by other public databases. As of January 2010, the status of our current curation effort is as follows:
  - ACS Chemical Biology 2009-2015 (vol 1-10)
  - Bioorganic & Medicinal Chemistry 1991-2015 (vol 1-9)
  - EMBO Journal 2001-2015 (vol 1-7)
  - J Med Chem 2005-2015 (vol 1-16)
  - J Med Chem 2006-2015 (vol 1-10)
  - J Med Chem 2008-2015 (vol 1-20)
  - J Med Chem 2009-2015 (vol 1-22)
  - J Med Chem 2010-2015 (vol 1-26)
  - J Med Chem 2011-2015 (vol 1-30)
  - J Med Chem 2012-2015 (vol 1-35)
  - J Med Chem 2013-2015 (vol 1-40)
  - J Med Chem 2015-2015 (vol 1-50)

myBDB
- Username: Password
- Login / Logout
- Username is your registered email in BindingDB.
- register

Video Tutorials
- Get all data from an article
- Download all data for a target of interest
- Find and view all data for a target of interest
- Find my compound's targets

https://www.bindingdb.org/bind/index.jsp
VI. Summary tools & websites

http://bionumbers.hms.harvard.edu/
Protein structures

http://www.rcsb.org/pdb/home/home.do
3D structures of protein interactions

http://interactome3d.irbbarcelona.org/
VI. Summary tools & websites

3D structures of protein interactions/ mapping of disease mutations

http://dsysmap.irbbarcelona.org/
**Protein design**

The FoldX Suite builds on the strong foundation of advanced protein design features already implemented in the oldest FoldX versions and integrates new capabilities: loop reconstruction (LoopX) and peptide docking (PepX). The Suite also features an improved usability thanks to a new boost Command Line Interface.

http://foldxsuite.crg.eu/products#foldx
Experimental validation of the role of kinetic parameters in MCF7 cells (weak feedback)

Experimental design of mutants that introduce kinetic perturbations

I. Kinetic perturbations and network topology

\[ K_d = \frac{k_{off}}{k_{on}} \]

Affinity (Dissociation constant) \( \rightarrow \) Dissociation rate constant \( \rightarrow \) Association rate constant

E.g.:

↑ Increase \( k_{on} \): improve electrostatic surface complementarity; *‘electrostatic steering’*

↑ Increase \( k_{off} \): mutate hot-spot residues in the interface

Kiel et al., PNAS, 2004

↑ Increase \( k_{off} \): mutate hot-spot residues in the interface
Summary of the protein mutant design

Ras surface negative
Raf surface positive

I. Kinetic perturbations and network topology

Kiel & Serrano, Sci Signal, 2009
Analysis of all mutants in RK13 cells (luciferase activity assay)

Correlation between predicted changes in $k_{on}$ is very high, while correlation with affinity ($\Delta G$) is poorer

Kiel & Serrano, Sci Signal, 2009
Results from network model for designed mutants

Experiments and simulations suggest that association rate constants of Ras-Raf complex formation are important for signaling

Kiel & Serrano, Sci Signal, 2009