

# 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

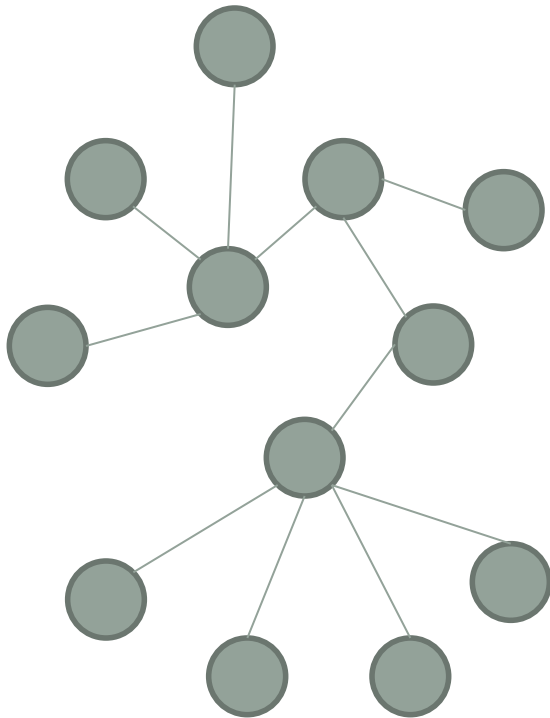
CRG Barcelona



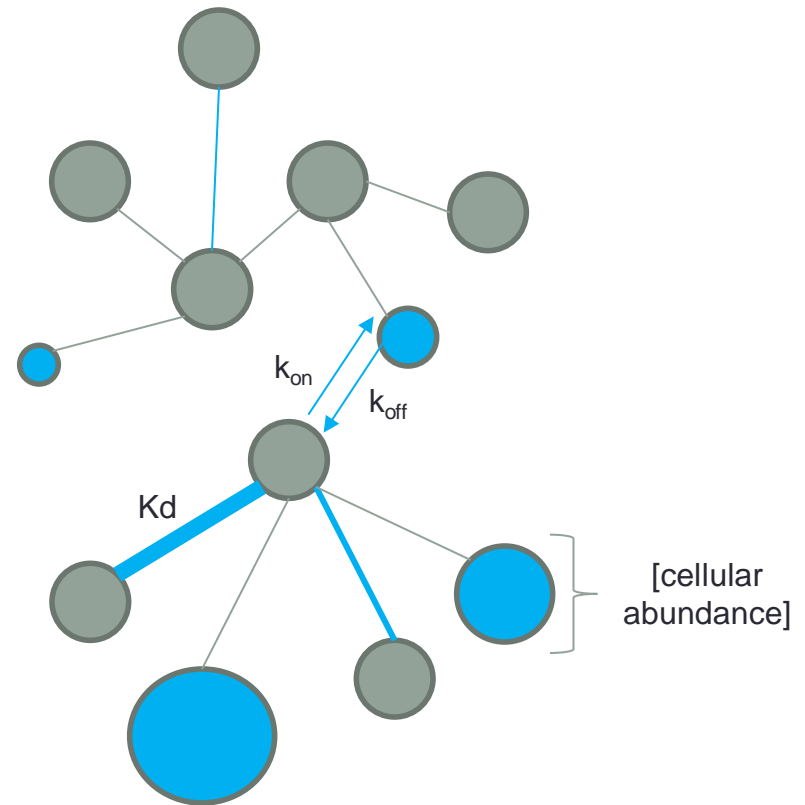
CRG Barcelona: <http://crg.eu>

# Quantitative information in protein-protein interaction (PPI) networks

Qualitative PPI networks



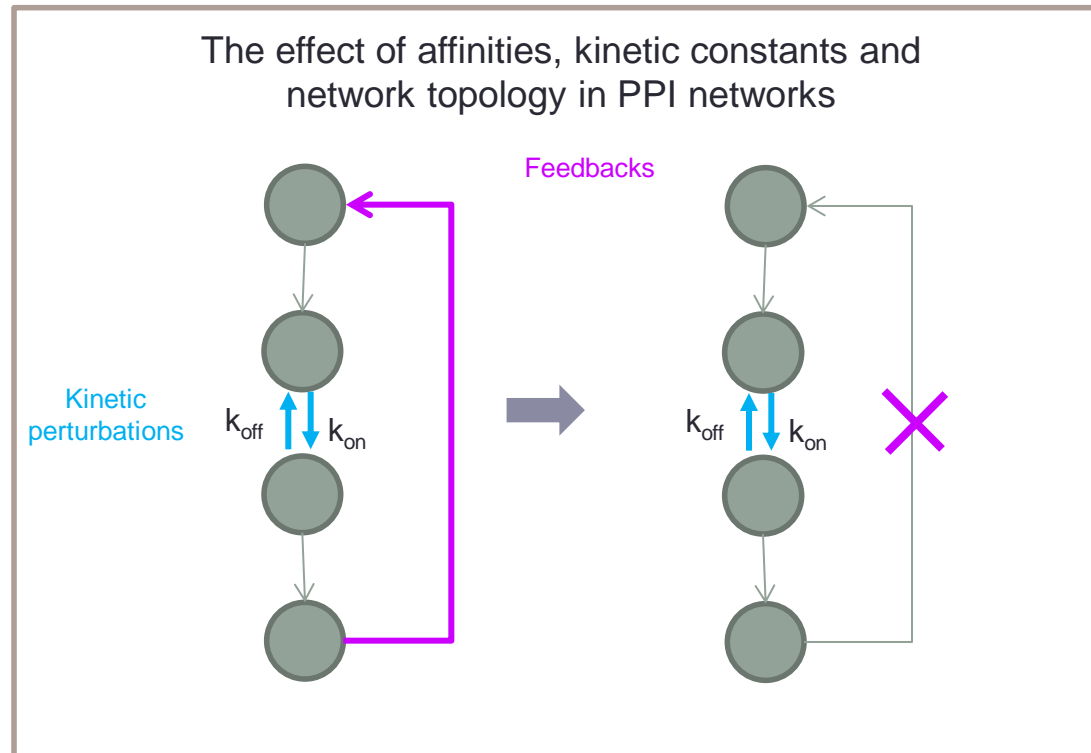
Quantitative PPI networks



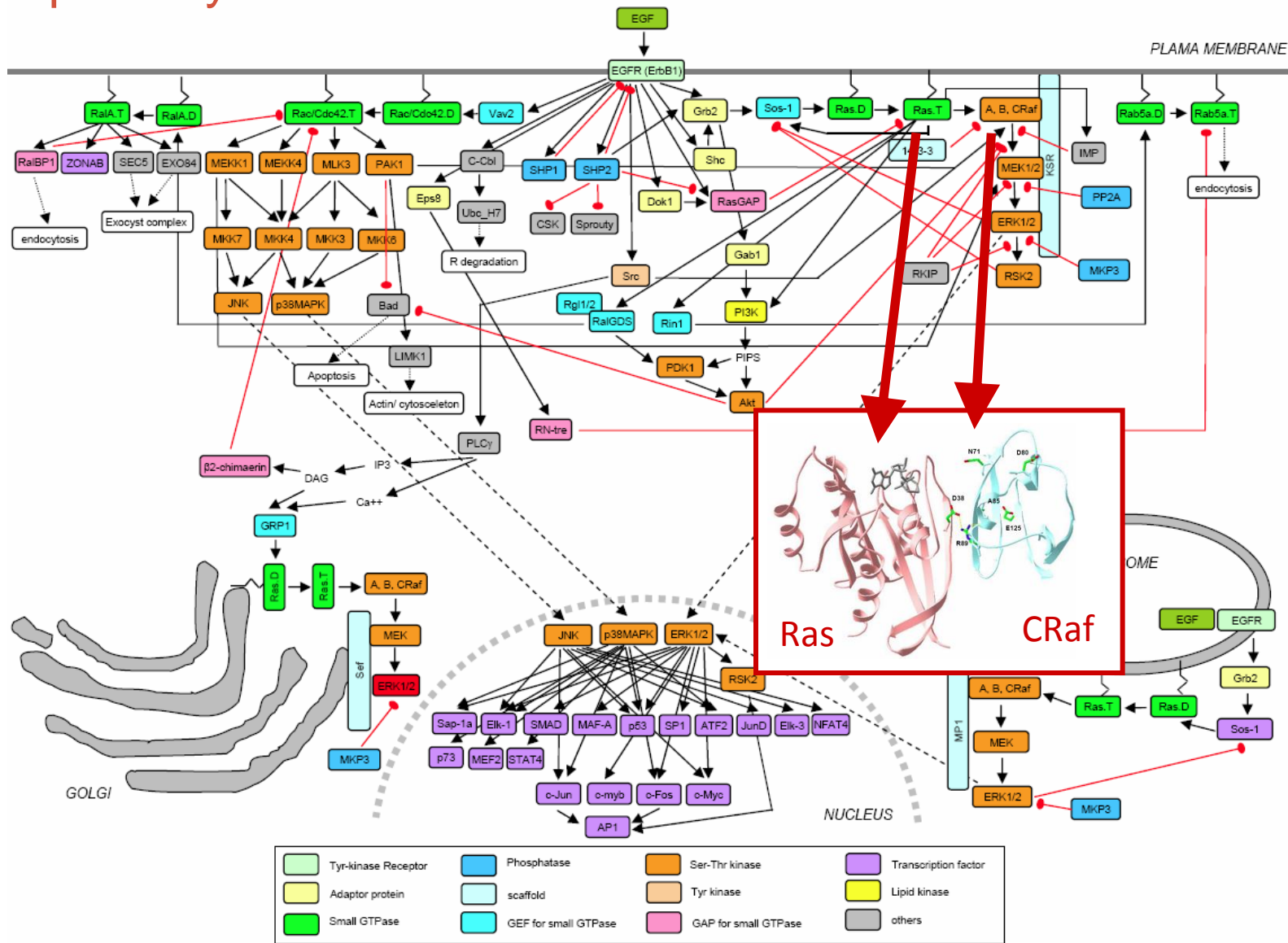
Considering protein abundances and affinities/  
kinetic constants

# Outline

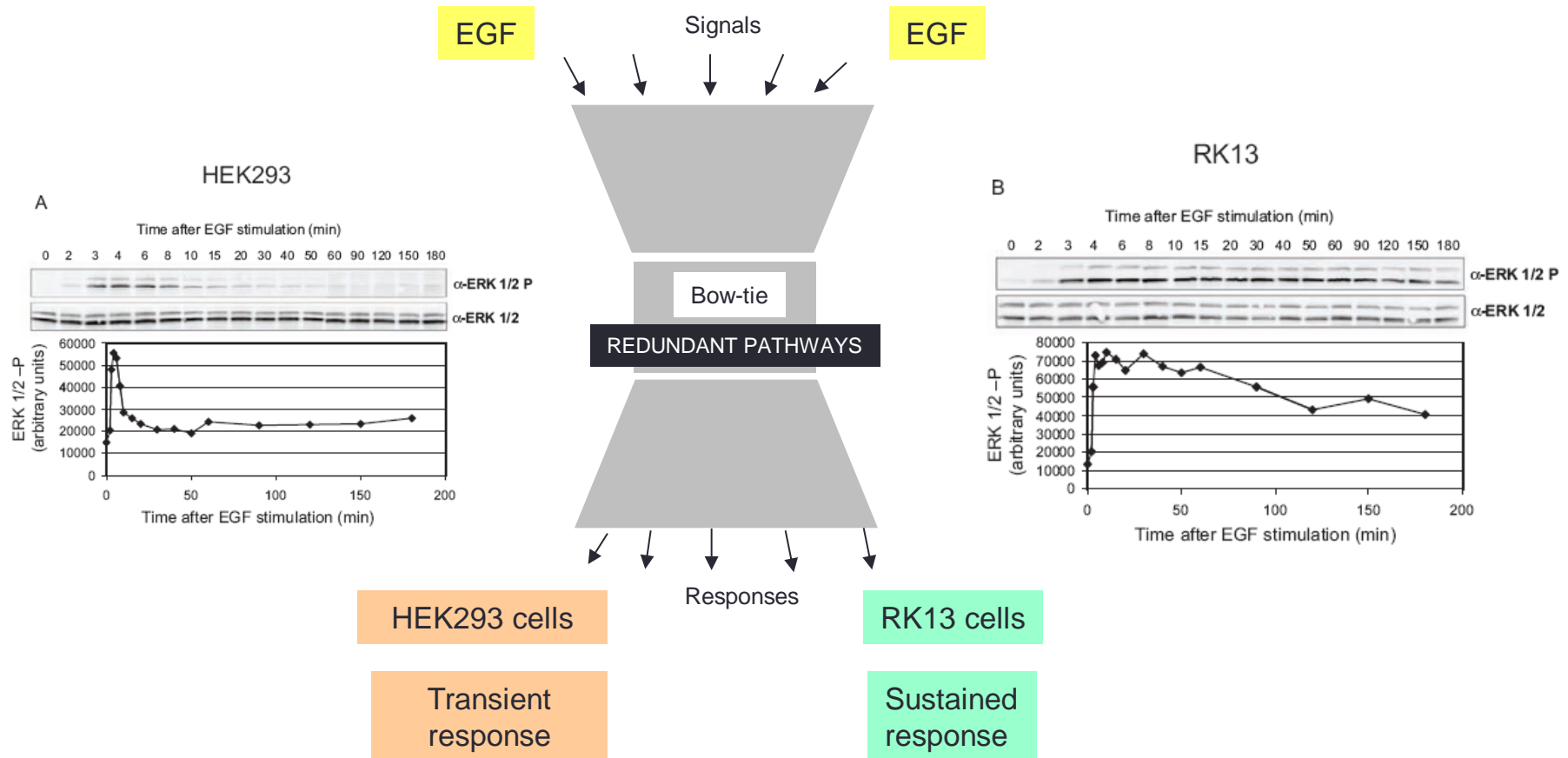
- I.** The effect of affinities, kinetic constants and network topology in PPI networks
- II.** The effect of protein abundance perturbations and interaction competition in PPI networks
- III.** Methods to quantify protein abundances, affinities, and kinetic constants
- IV.** Disease mutations and their principle effect on PPI networks
- V.** Examples for quantitative effects in disease networks
  1. RASopathy vs cancer mutations: a matter of quantity
  2. BRAF mutation frequency: prediction of oncogenic drivers
- VI.** Summary tools & websites
- VII.** Wrap up/ discussion/ conclusions



# Epidermal growth factor (EGF) activates the RAS-RAF-MEK-ERK pathway



# Cell type-specific ERK activation in HEK293 and RK13 cells



Different signaling response (ERK-p) with similar ligand (EGF)

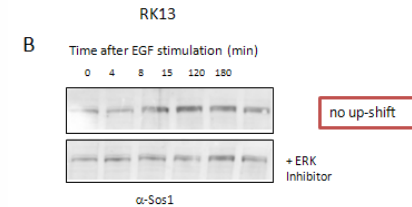
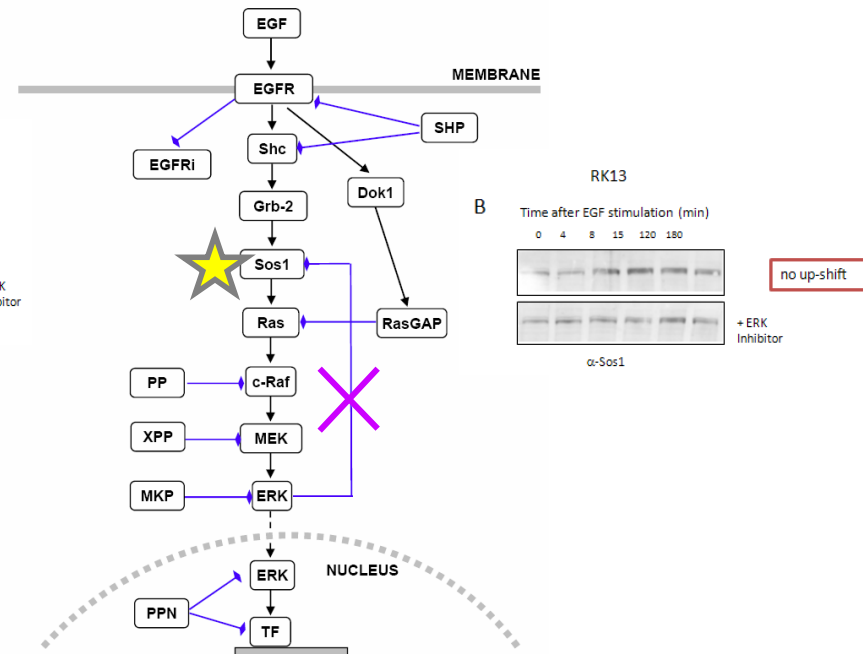
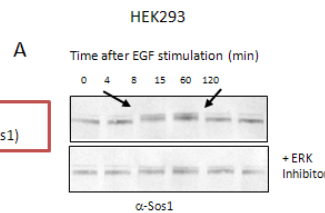
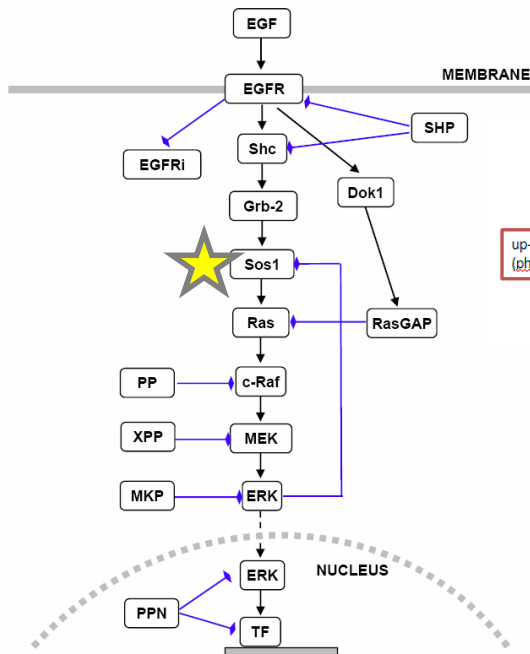
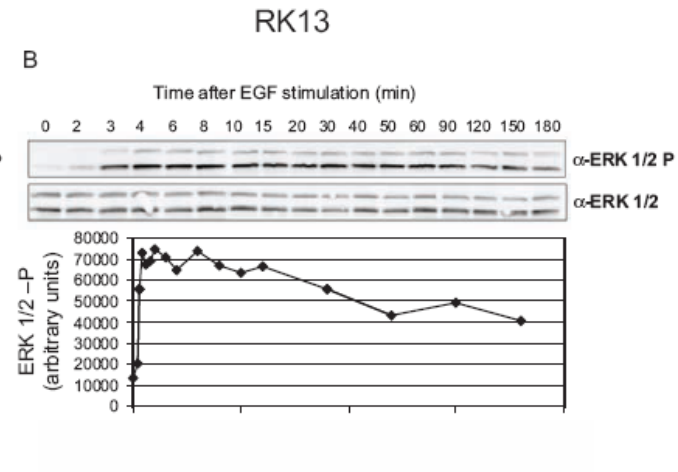
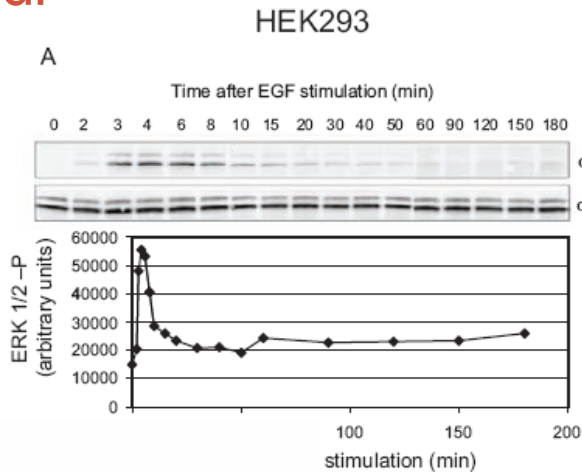
# Different network 'wiring' /feedbacks causes the different behaviour

HEK293 cells

Transient response

RK13 cells

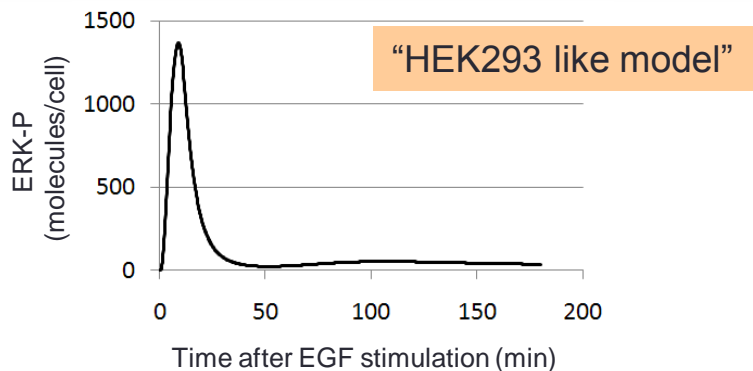
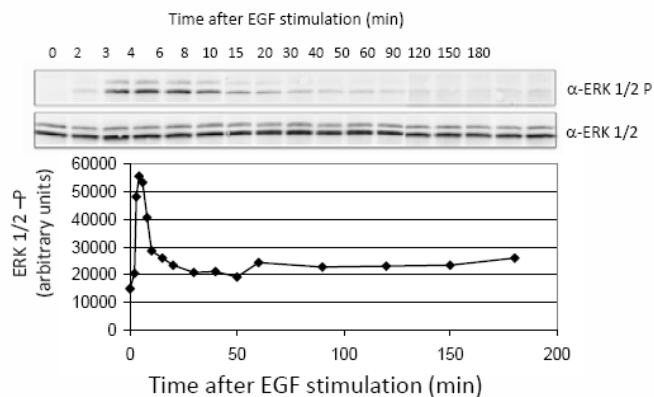
Sustained response



# A simple computer model of ERK activation in HEK293 and RK13 cells

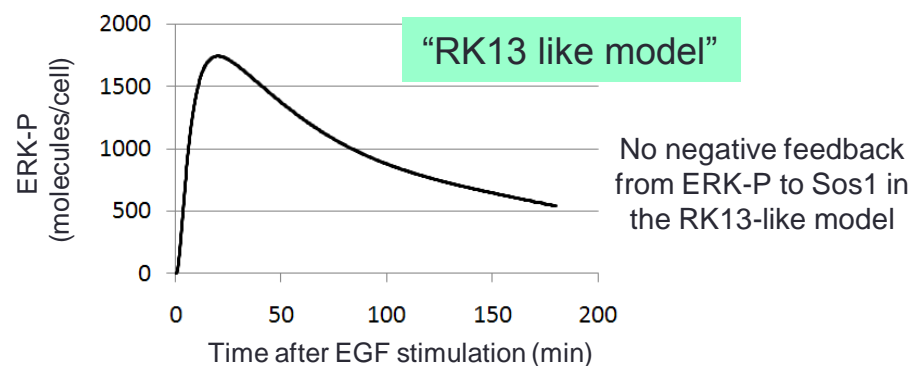
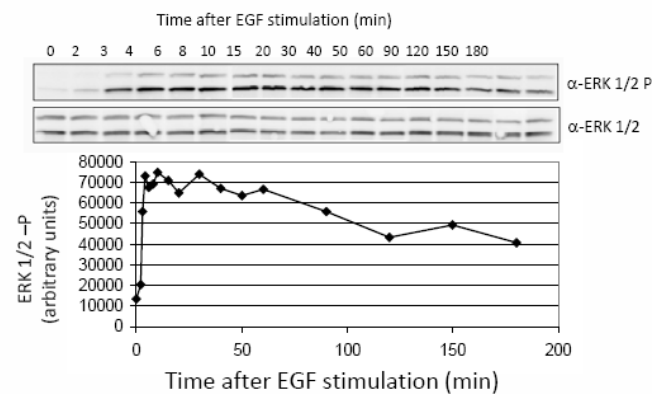
HEK293

A



RK13

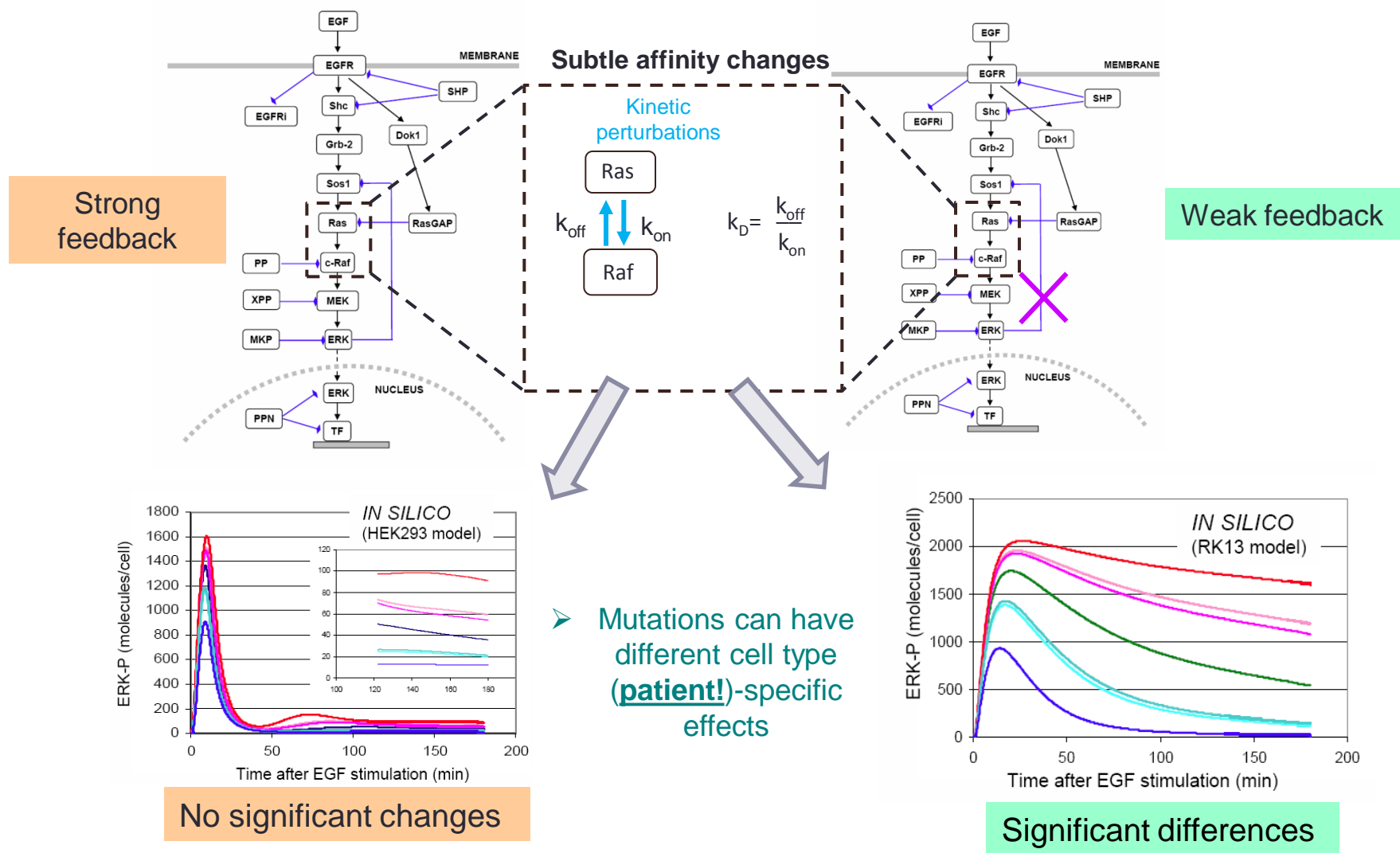
B



➤ Good agreement of experiment and model predictions



# Model predictions: different cell type-specific wiring results in different responses to affinity perturbations



## Experimental validation of the role of kinetic parameters in MCF7 cells (weak feedback)

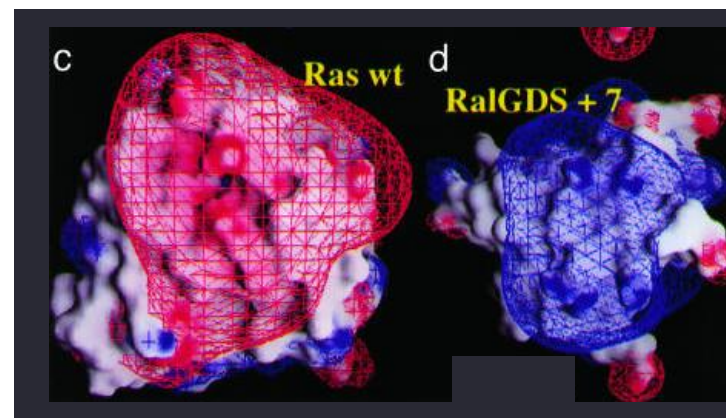
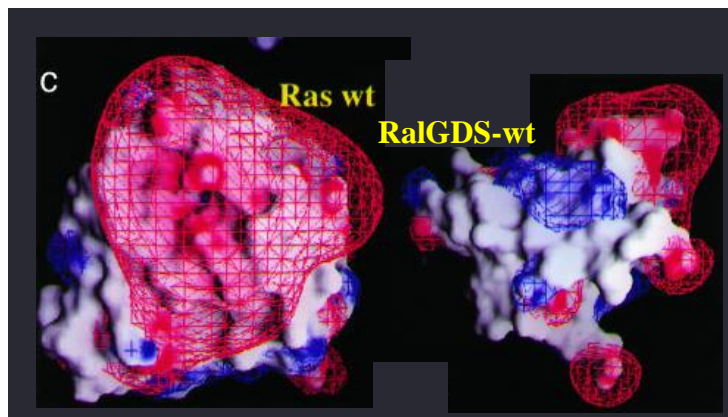
Experimental design of mutants that introduce kinetic perturbations

Affinity  
(Dissociation constant)  $\longrightarrow$   $K_d = \frac{k_{\text{off}}}{k_{\text{on}}}$

$k_{\text{off}}$   $\longleftarrow$  Dissociation rate constant  
 $k_{\text{on}}$   $\longleftarrow$  Association rate constant

E.g.:

$\uparrow$  Increase  $k_{\text{on}}$ : improve electrostatic surface complementarity; '*electrostatic steering*'



Kiel et al., PNAS, 2004

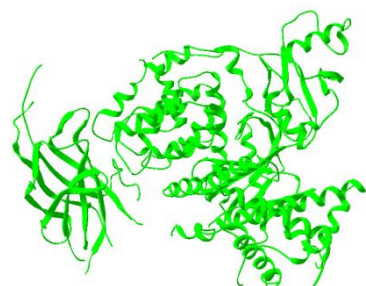
$\uparrow$  Increase  $k_{\text{off}}$ : mutate hot-spot residues in the interface

# FoldX-based energy calculations of proteins

3D Structural information

A force field for energy calculations and protein design

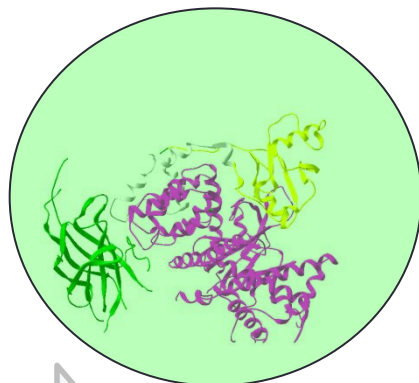
Schymkowitz et al, *Nucleic Acids Res*, 2005



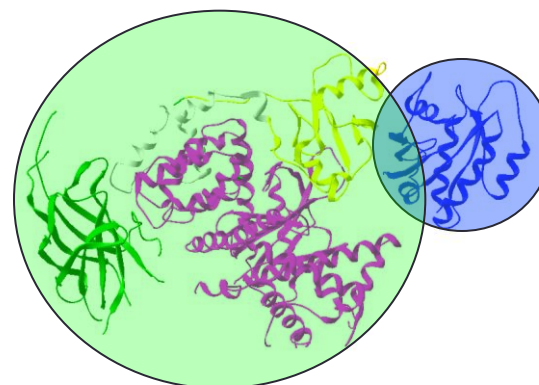
$$+ \text{FoldX} = \Delta G$$

Relation to affinity:  $\Delta G = RT \ln K_d$

✓ Total free energy

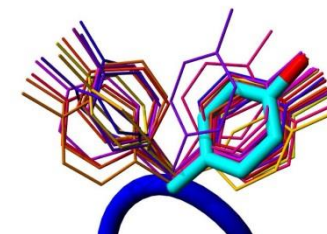


✓ Interaction energy



✓ Mutagenesis

A rotamer library to replace the 20 amino acids

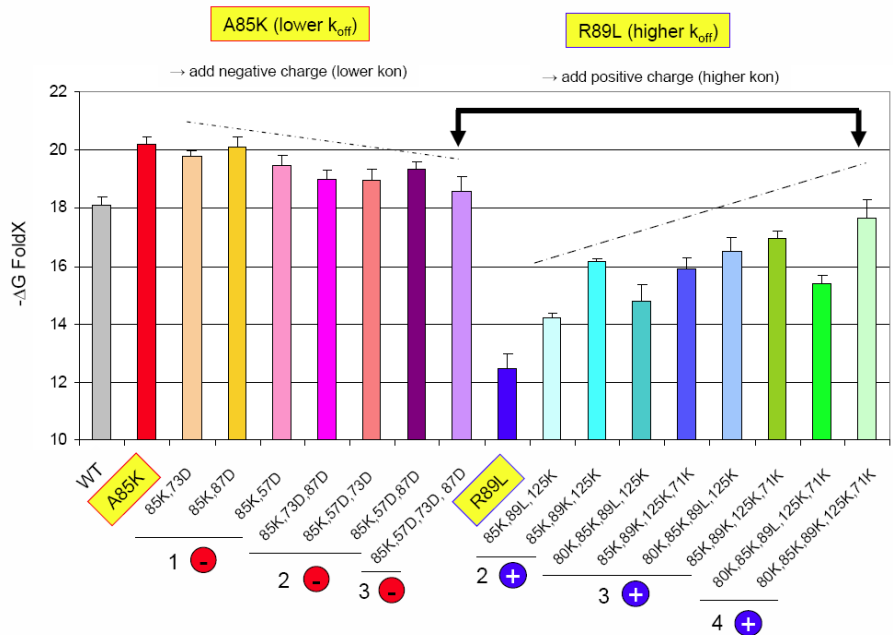
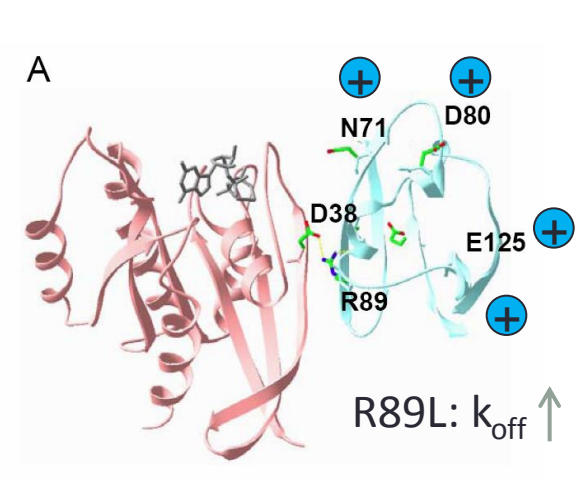
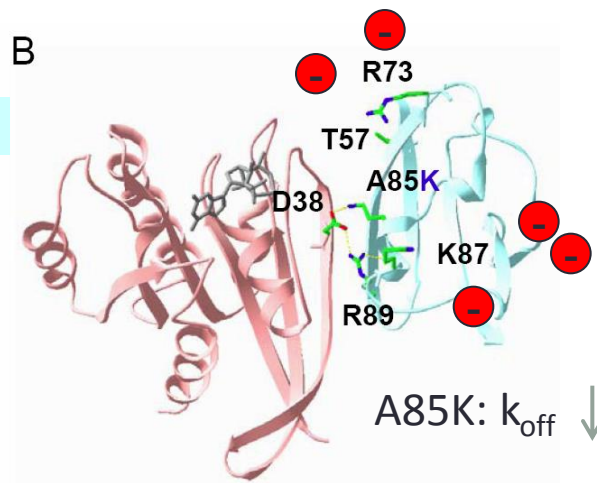
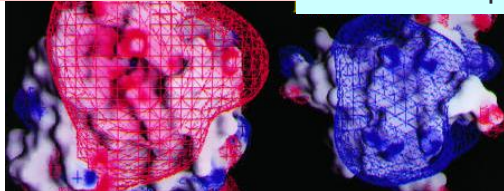


Protein design

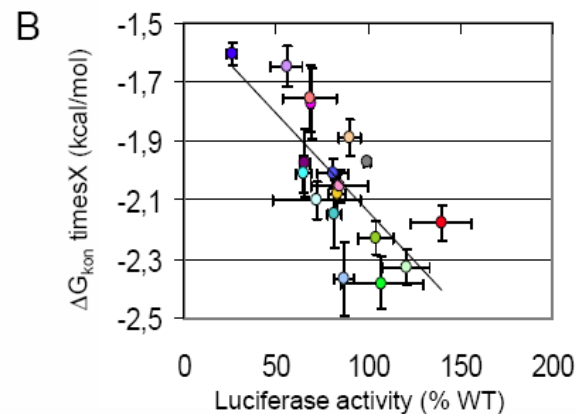
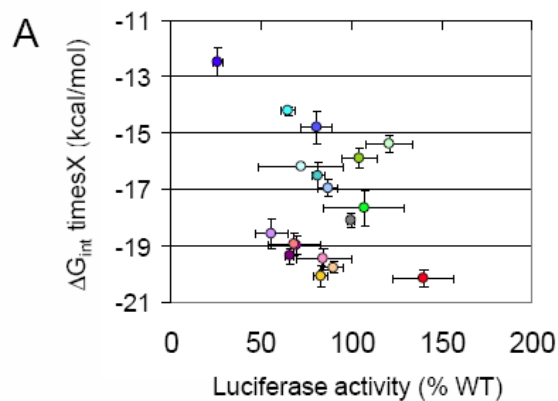
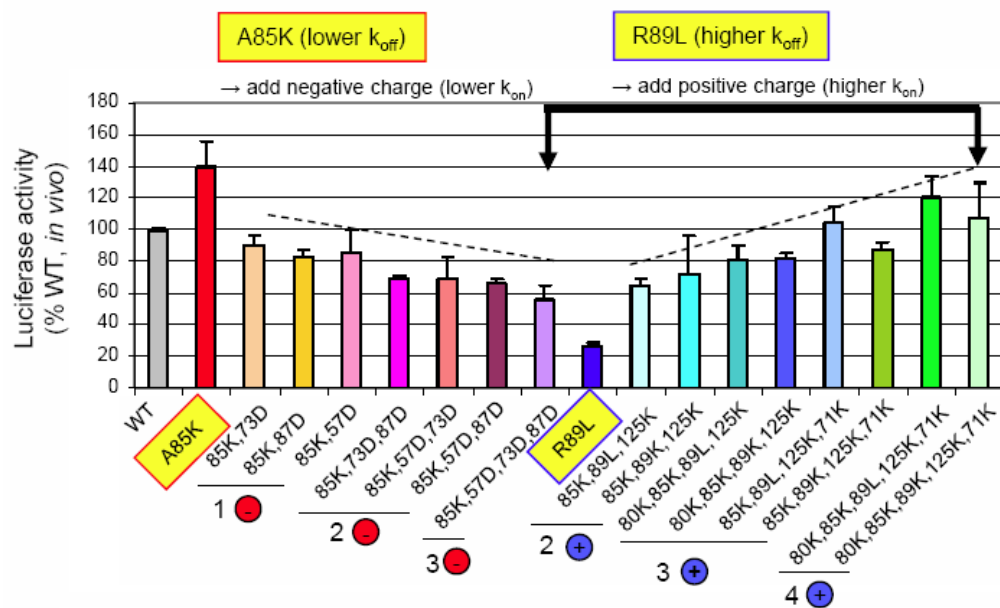
# Summary of the protein mutant design

Ras surface negative

Raf surface positive

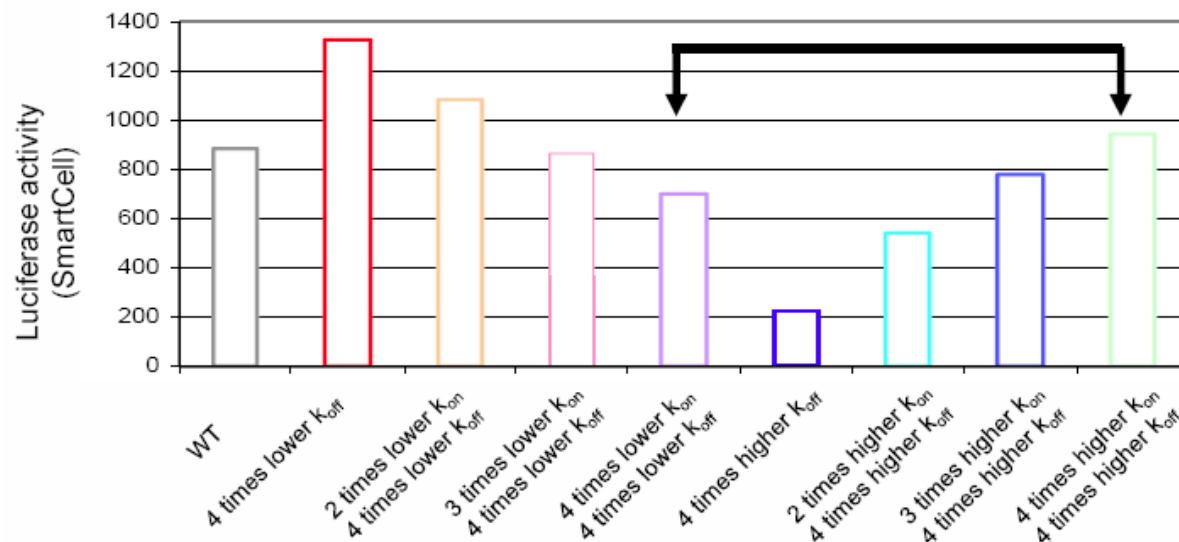


# Analysis of all mutants in RK13 cells (luciferase activity assay)



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

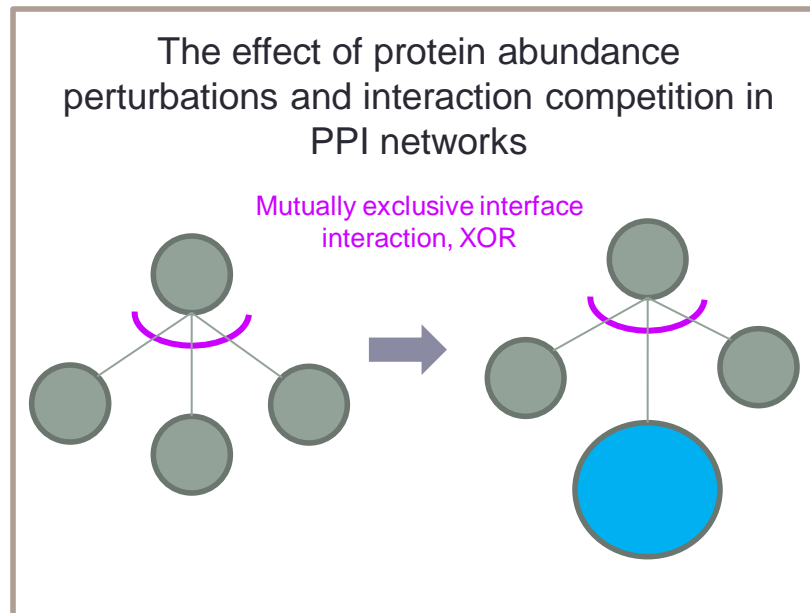
## Results from the network model for designed mutants



Confirms experimental findings:

Mutant with 4 time lower  $k_{on}$  and 4 times lower  $k_{off}$  (same  $K_D$ ) has less predicted luciferase activity (and opposite for mutant with 4 times higher  $k_{on}/k_{off}$ )

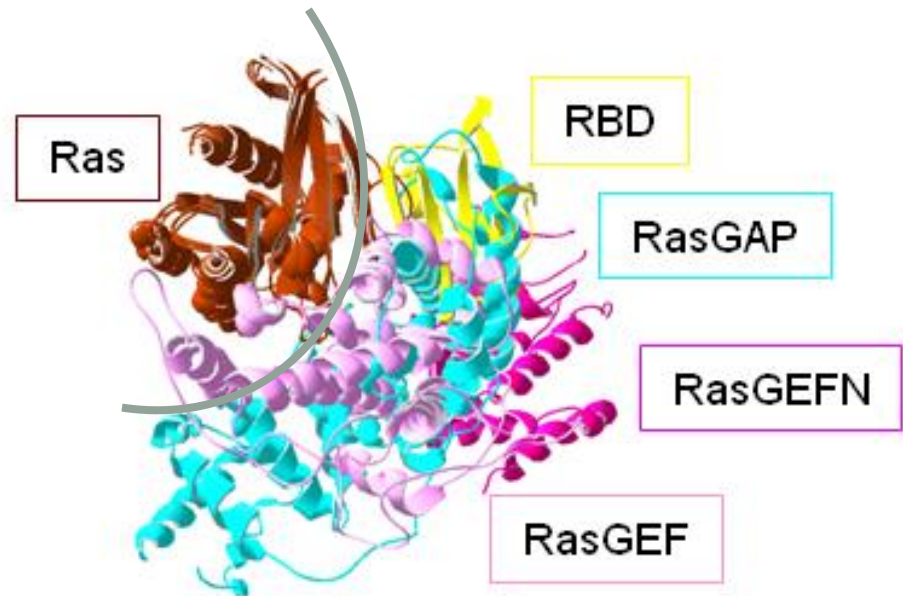
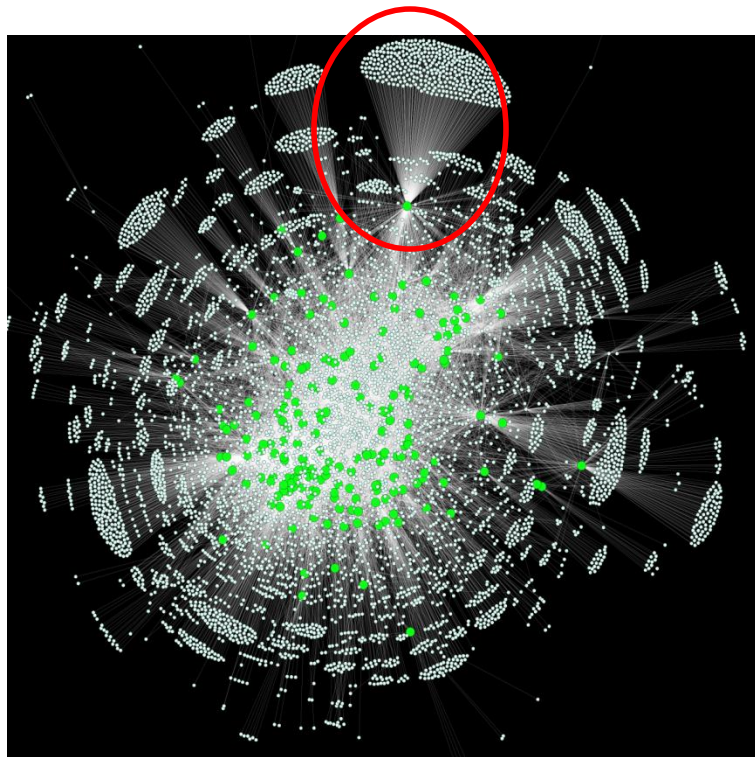
- Experiments and simulations suggest that association rate constants of Ras-Raf complex formation are important for signaling



## 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'

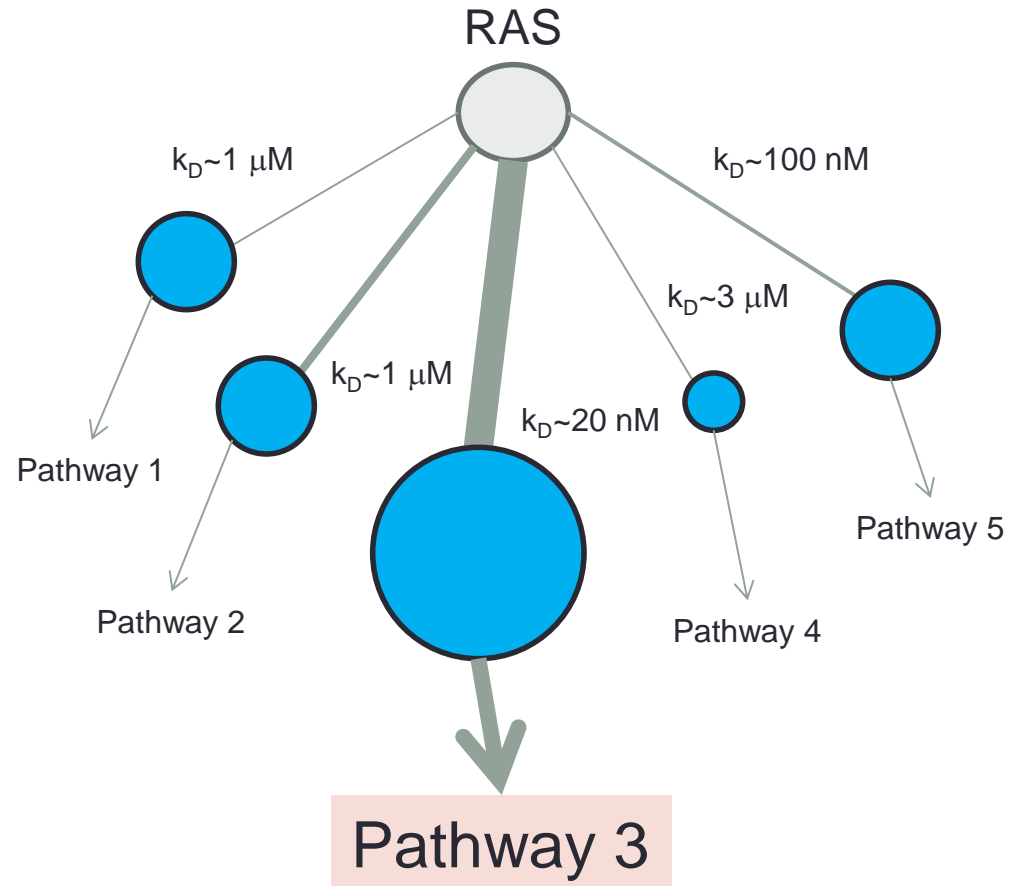
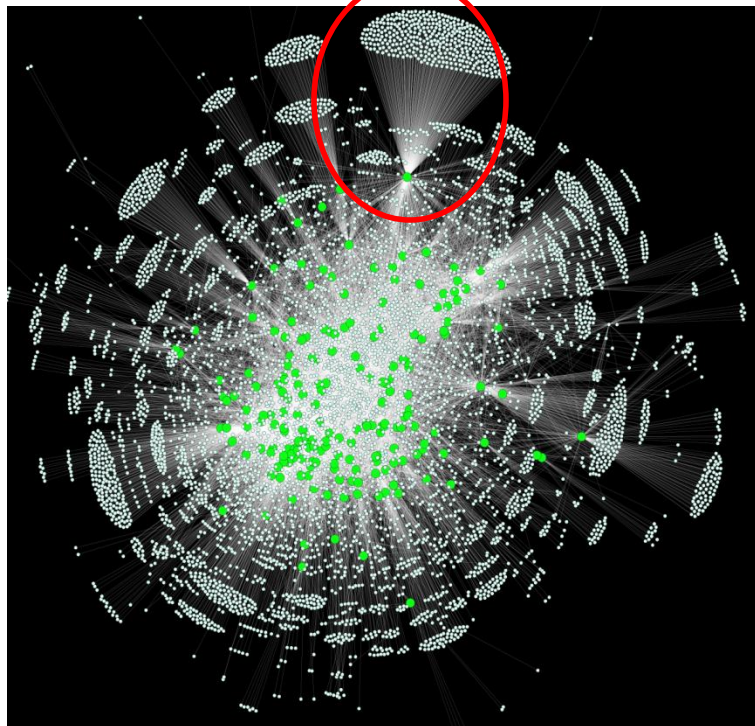
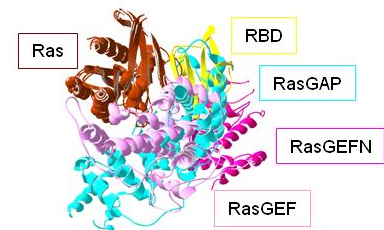




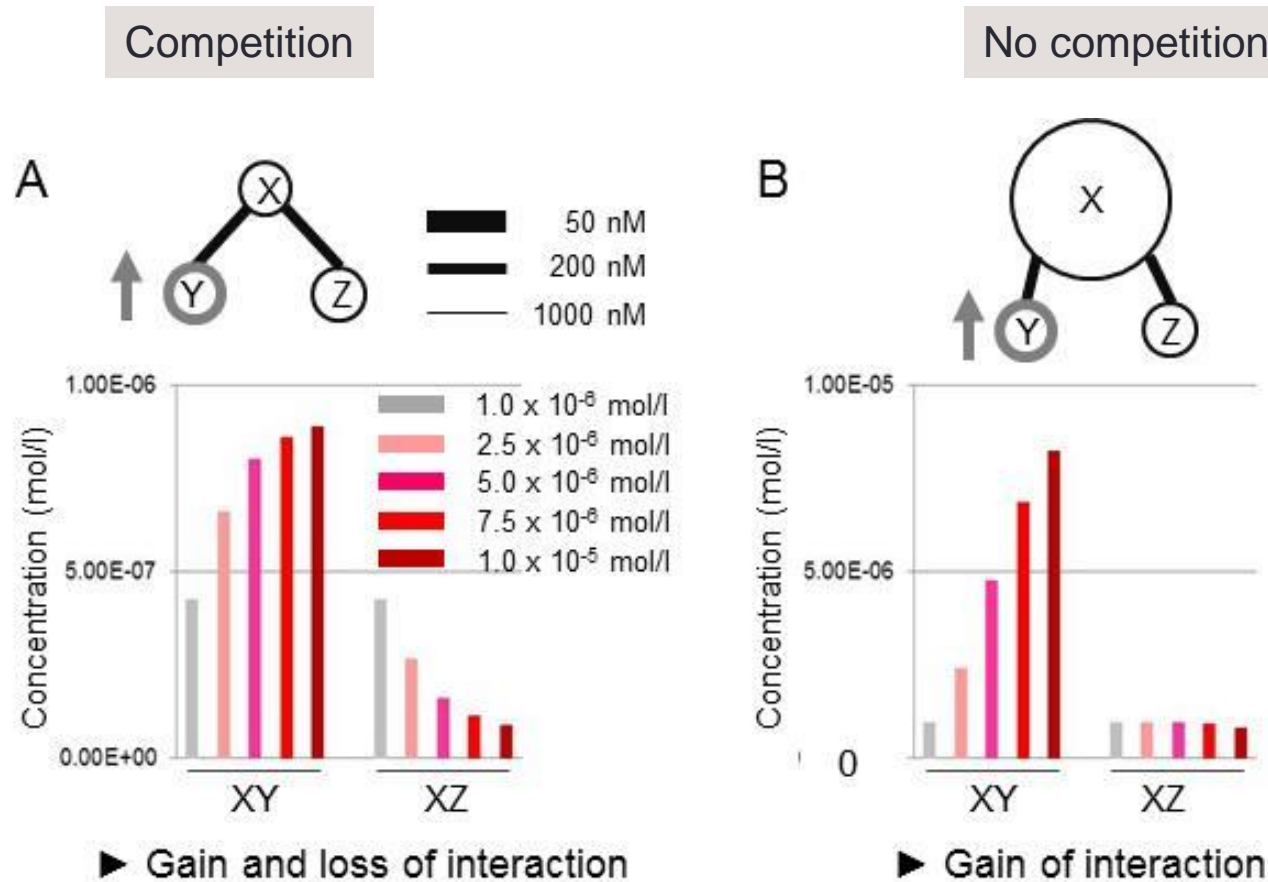
# How could interaction competition and protein concentration affect downstream signaling?

Signaling complexes: > 300 partners for one protein??

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



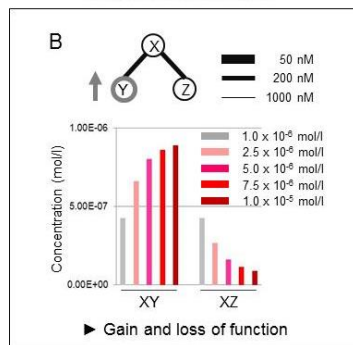
# The effect of abundance variation at XOR network motifs



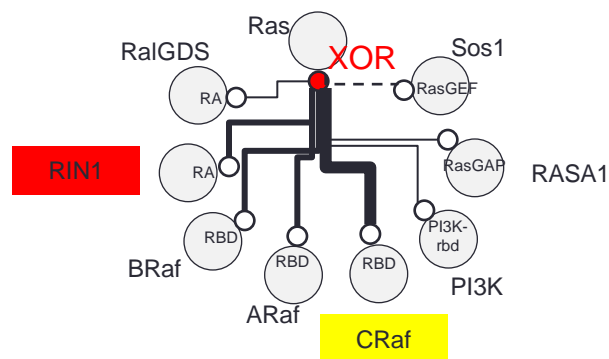
➤ The output/ function depends on both, network structure and abundance

# Competition at the Ras XOR node

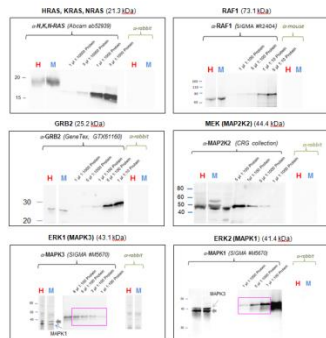
## Network motif



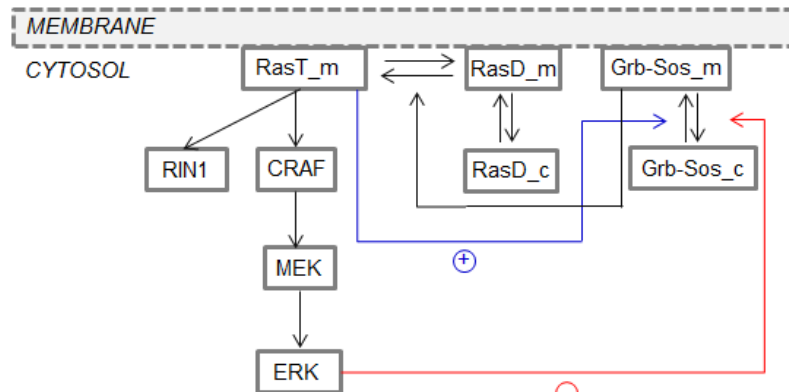
## The Ras XOR node



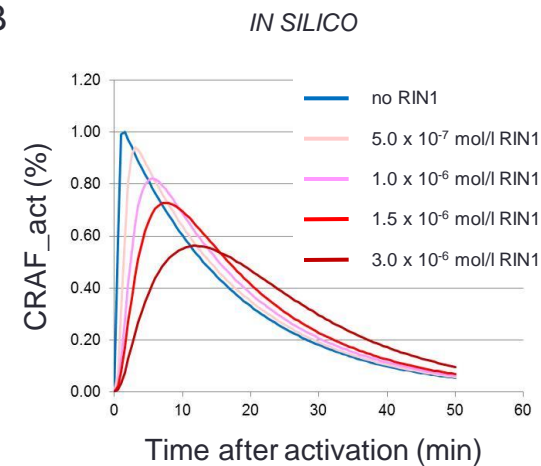
## Experimental abundances



### A

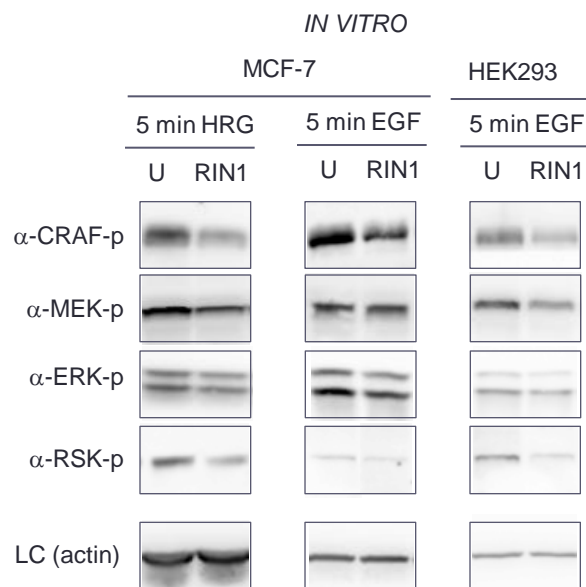


### B

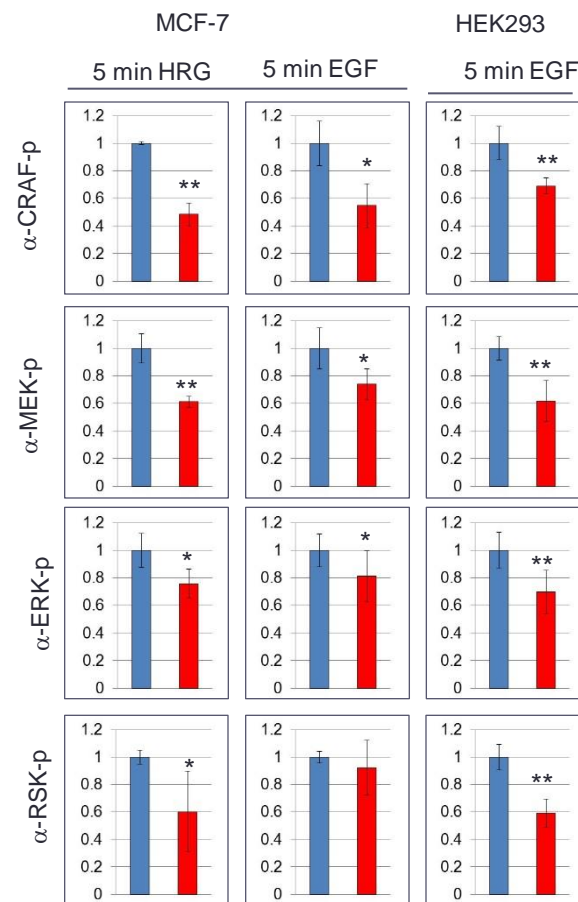


Mathematical network modeling: increasing RIN1 to 10-fold higher of CRAF expression should decrease CRAF activation

# Experimental testing of competition at the Ras node

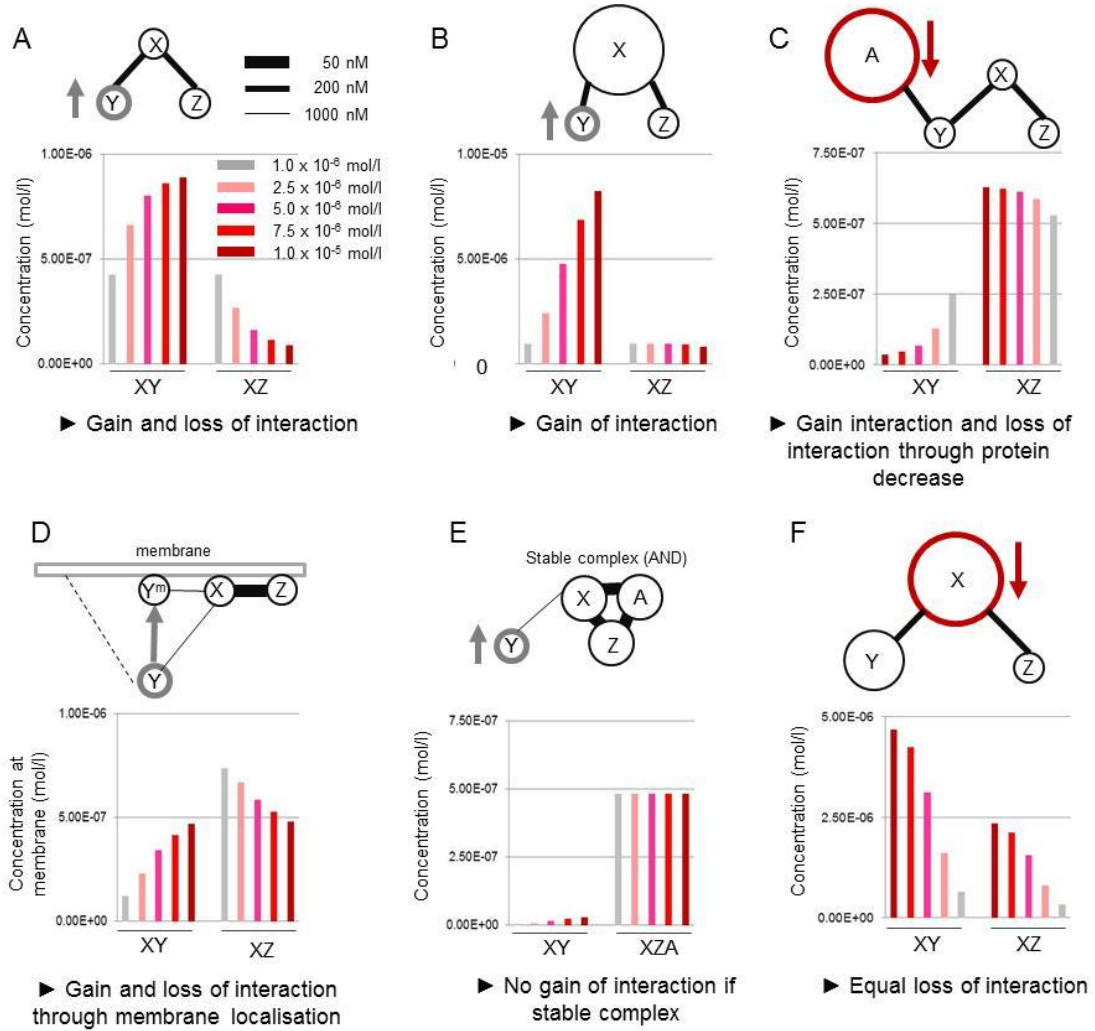


Expression of RIN1 in MCF-7 and HEK293 cells decreases CRAF, MEK, and ERK activation



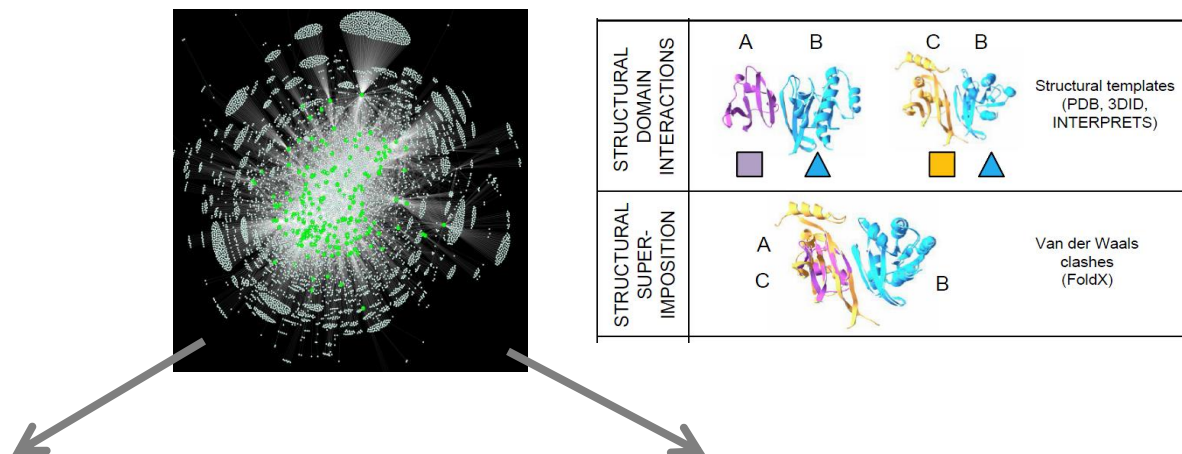
- Alterations in the abundance of one of two hub-binding partners affected downstream signaling

# The effect of abundance variation at XOR network motifs

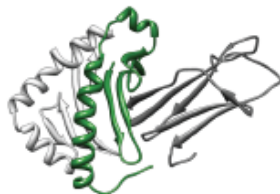
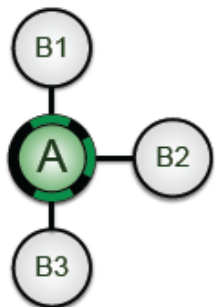


➤ The output/ function depends on both, network structure and abundance: we need to know the network very well to understand

# A bioinformatics tool to distinguish mutually exclusive from compatible interactions in large-scale PPI



Compatible ('AND')

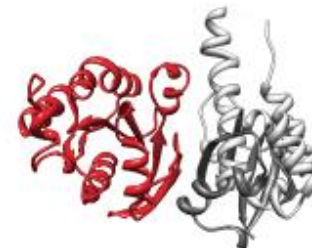
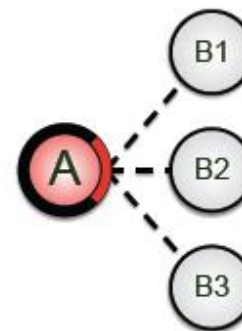


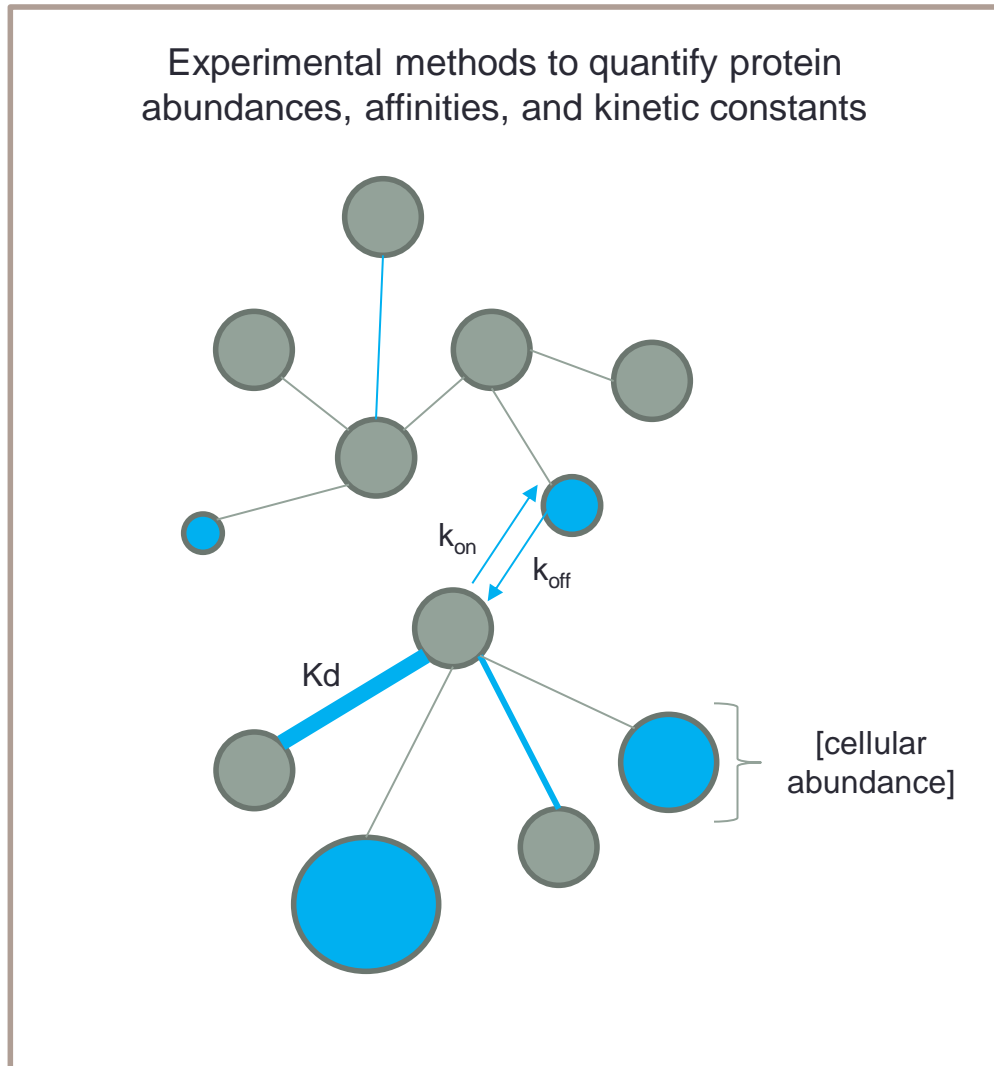
**SAPIN** (structural analysis of protein interaction networks)

webservice

<http://sapin.crg.es/>

Exclusive ('XOR')





## Why proteomics in times of deep RNA sequencing?

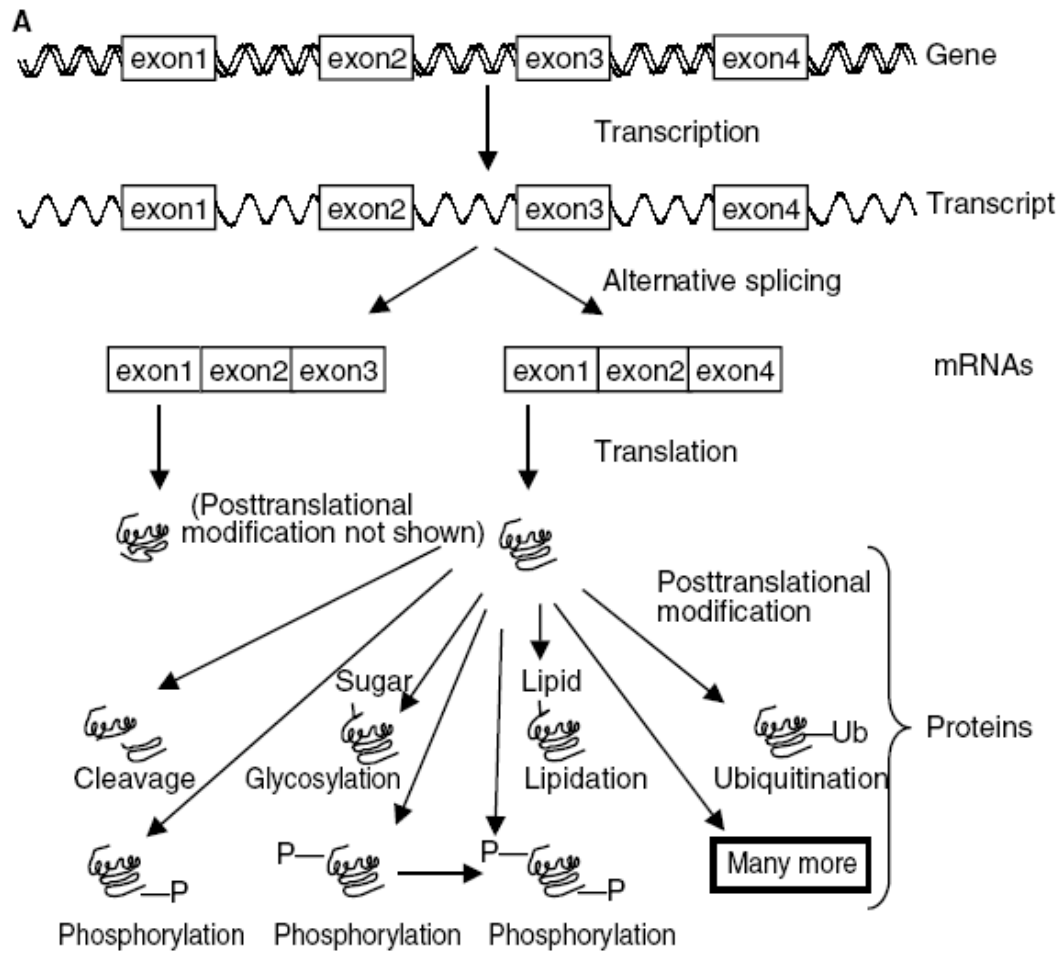
- ❑ mRNA does not translate 1:1 into protein; keywords:
  - (i) translation efficiency,
  - (ii) mRNA stability,
  - (iii) protein stability,
- ❑ Posttranslational modification (PTMs) of proteins, e.g. phosphorylation

Two main aims: IDENTIFICATION and QUANTIFICATION

Two main techniques: MASS SPECTROMETRY and ANTIBODY-BASED



# High complexity of the proteome

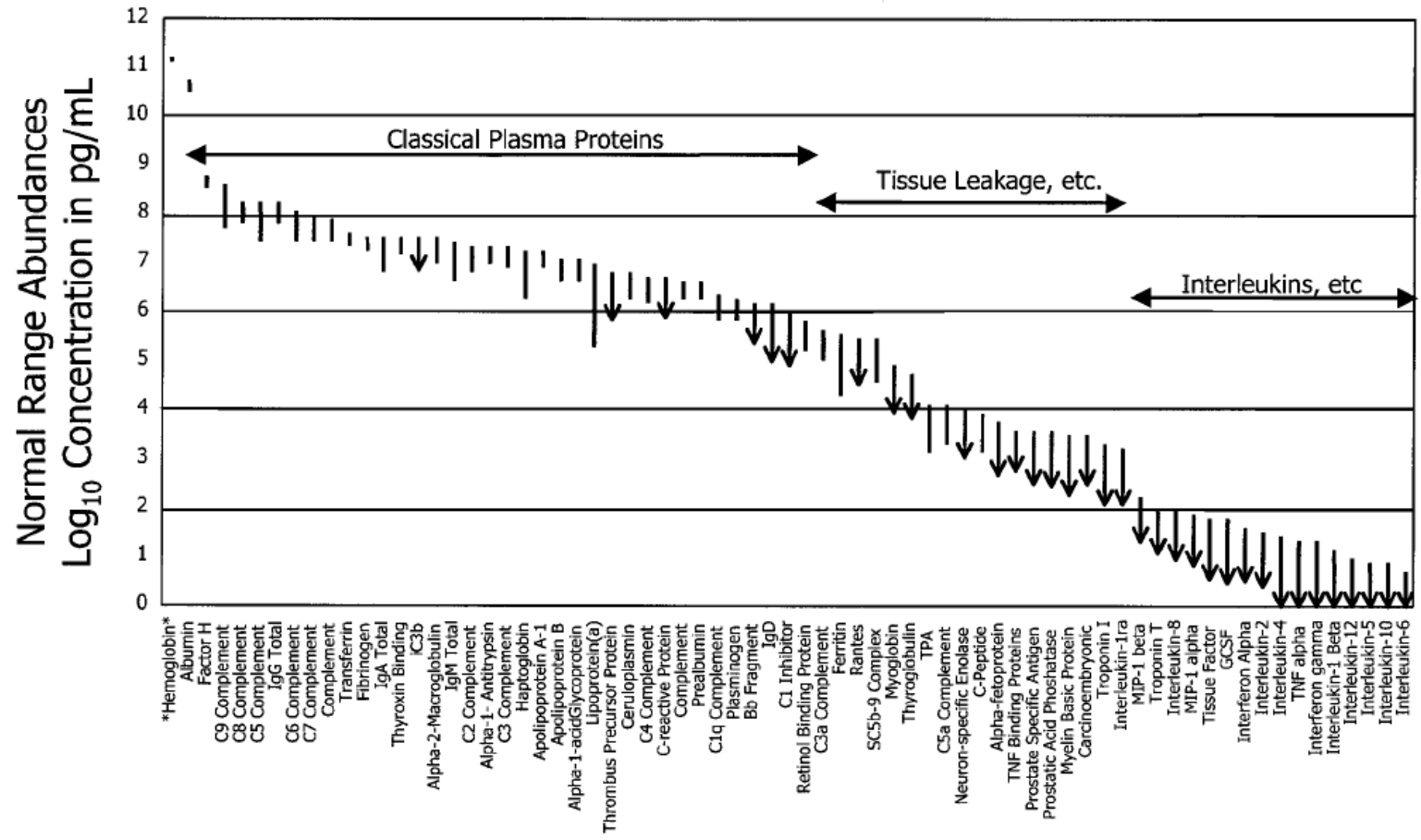


30,000 coding genes per cell

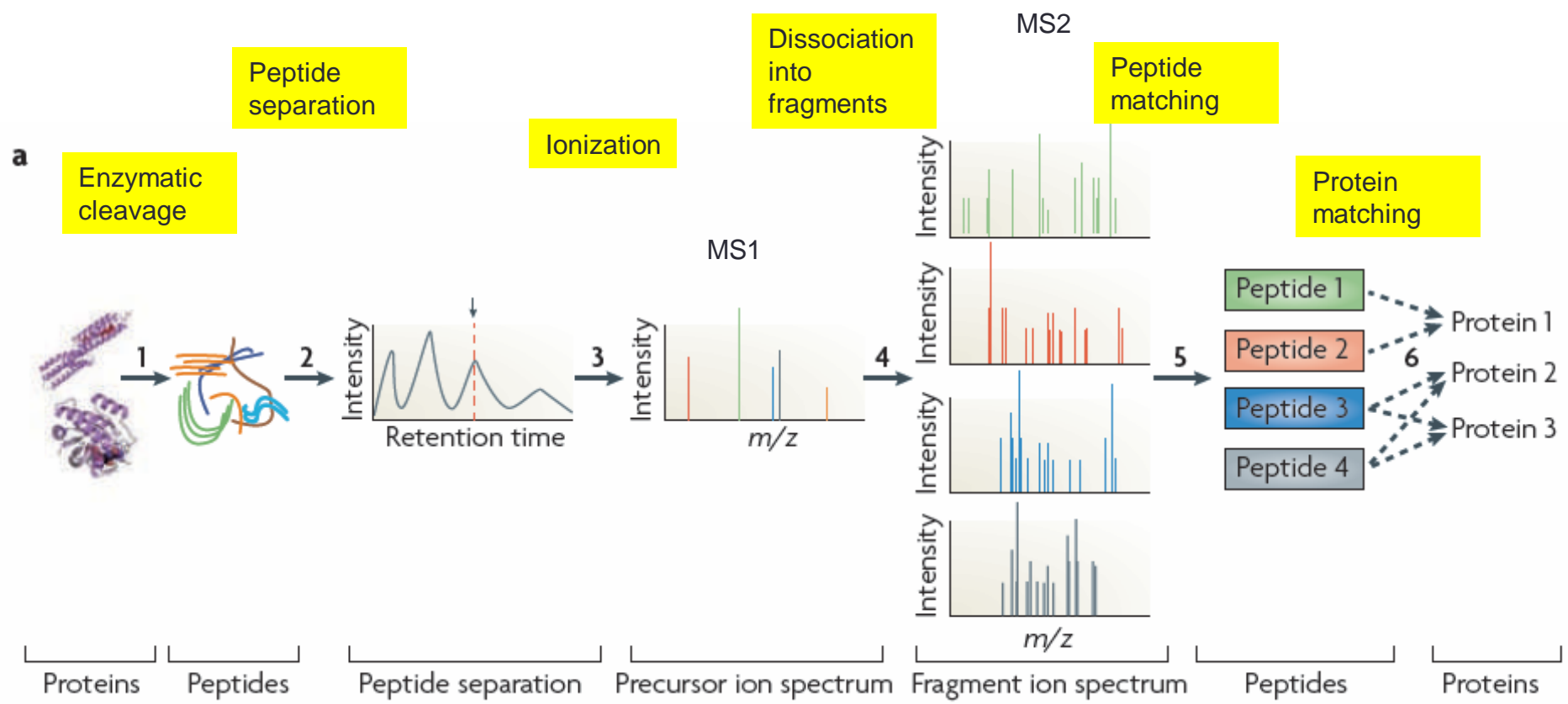
Alt.splicing: 2-3 x 30,000  
= 90,000 proteins

Post-translational modifications  
> 10 x 90,000  
= 900,000 proteins

# High dynamic range of the proteome



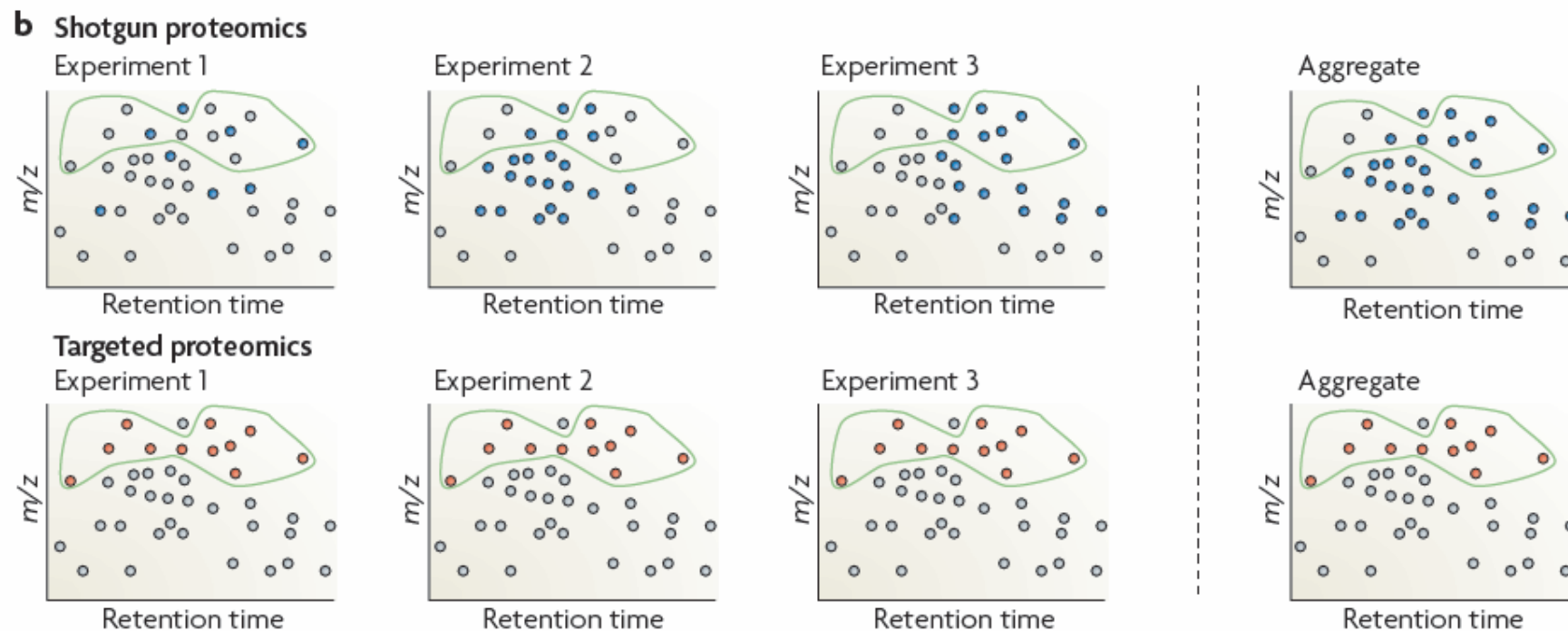
# Protein identification by mass spectrometry



- ❑ 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...

## 'Shotgun' compared to 'targeted' approach

Targeted proteomics is the method of choice for studying (a limited number of) signaling proteins



# Human deep proteome mapping

Molecular Systems Biology 7; Article number 549; doi:10.1038/msb.2011.82  
 Citation: *Molecular Systems Biology* 7: 549  
 © 2011 EMBO and Macmillan Publishers Limited All rights reserved 1744-4292/11  
 www.molecularsystemsbiology.com



## REPORT

### The quantitative proteome of a human cell line

Martin Beck<sup>1,9</sup>, Alexander Schmidt<sup>2,9</sup>, Johan Malmstroem<sup>3,4</sup>, Manfred Claassen<sup>5</sup>, Alessandro Ori<sup>1</sup>, Anna Szymborska<sup>1</sup>, Franz Herzog<sup>6</sup>, Oliver Rinner<sup>4</sup>, Jan Ellenberg<sup>1</sup> and Ruedi Aebersold<sup>6,7,8,\*</sup>

<sup>1</sup> European Molecular Biology Laboratory, Heidelberg, Germany, <sup>2</sup> Biozentrum, University of Basel, Basel, Switzerland, <sup>3</sup> Department of Immunotechnology, BMC, Lund, Sweden, <sup>4</sup> Biognosys AG, Schlieren, Switzerland, <sup>5</sup> Department of Computer Science, ETH Zurich, Zurich, Switzerland, <sup>6</sup> Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland; <sup>7</sup> Competence Center for Systems Physiology and Metabolic Diseases, Zurich, Switzerland and <sup>8</sup> Department of Science, University of Zurich, Zurich, Switzerland

<sup>9</sup> These authors contributed equally to this work

\* Corresponding author. Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Wolfgang-Pauli-Strasse 16, Zurich 8093, Switzerland. Tel.: +41 44 633 1071; Fax: +41 44 633 1051; E-mail: aebersold@imsb.biol.ethz.ch

Received 15.8.11; accepted 29.9.11

- R. Aebersold lab

~10,000 proteins quantified

Beck et al, MSB, 2011

Molecular Systems Biology 7; Article number 548; doi:10.1038/msb.2011.81  
 Citation: *Molecular Systems Biology* 7: 548  
 © 2011 EMBO and Macmillan Publishers Limited All rights reserved 1744-4292/11  
 www.molecularsystemsbiology.com



## REPORT

### Deep proteome and transcriptome mapping of a human cancer cell line

Nagarjuna Nagaraj<sup>1</sup>, Jacek R Wisniewski<sup>1</sup>, Tamar Geiger<sup>1</sup>, Juergen Cox<sup>1</sup>, Martin Kircher<sup>2</sup>, Janet Kelso<sup>2</sup>, Svante Pääbo<sup>2</sup> and Matthias Mann<sup>1,\*</sup>

<sup>1</sup> Department for Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany and <sup>2</sup> Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

\* Corresponding author. Department for Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Am Klopferspitz 18, D-82152 Martinsried, Germany. Tel.: +49 89 8578 2557; Fax: +49 89 8578 2219; E-mail: mmann@biochem.mpg.de

Received 15.7.11; accepted 29.10.11

- M Mann lab

10,255 proteins quantified

Nagaraj et al, MSB, 2011

# Human deep proteome mapping: where are we now? Complete?

ARTICLE

2014 Pandey lab

doi:10.1038/nature13302

## A draft map of the human proteome

Min-Sik Kim<sup>1,2</sup>, Sneha M. Pinto<sup>3</sup>, Derese Getnet<sup>1,4</sup>, Raja Sekhar Nirujogi<sup>3</sup>, Srikanth S. Manda<sup>3</sup>, Raghothama Chaerkady<sup>1,2</sup>, Anil K. Madugundu<sup>3</sup>, Dhanashree S. Kelkar<sup>3</sup>, Ruth Isserlin<sup>5</sup>, Shobhit Jain<sup>5</sup>, Joji K. Thomas<sup>3</sup>, Babylakshmi Muthusamy<sup>1</sup>, Pamela Leal-Rojas<sup>1,6</sup>, Praveen Kumar<sup>3</sup>, Nandini A. Sahasrabudhe<sup>3</sup>, Lavanya Balakrishnan<sup>3</sup>, Jayshree Advani<sup>3</sup>, Bijesh George<sup>3</sup>, Santosh Renuse<sup>3</sup>, Lakshmi Dhevi N. Selvan<sup>3</sup>, Arun H. Patil<sup>3</sup>, Vishalakshi Nanjappa<sup>3</sup>, Aneeha Radhakrishnan<sup>3</sup>, Samarjeet Prasad<sup>1</sup>, Tejaswini Subbannayya<sup>3</sup>, Rajesh Raju<sup>3</sup>, Manish Kumar<sup>3</sup>, Sreelakshmi K. Sreenivasamurthy<sup>3</sup>, Arivusudar Marimuthu<sup>3</sup>, Gajanan J. Sathe<sup>3</sup>, Sandip Chavan<sup>3</sup>, Keshava K. Datta<sup>3</sup>, Yashwanth Subbannayya<sup>3</sup>, Apeksha Sahu<sup>3</sup>, Soujanya D. Yelamanchi<sup>3</sup>, Savita Jayaram<sup>3</sup>, Pavithra Rajagopalan<sup>3</sup>, Jyoti Sharma<sup>3</sup>, Krishna R. Murthy<sup>3</sup>, Nazia Syed<sup>3</sup>, Renu Goel<sup>3</sup>, Aafaque A. Khan<sup>3</sup>, Sartaj Ahmad<sup>3</sup>, Gourav Dey<sup>3</sup>, Keshav Mudgal<sup>3</sup>, Aditi Chatterjee<sup>3</sup>, Tai-Chung Huang<sup>3</sup>, Jun Zhong<sup>3</sup>, Xinyan Wu<sup>1,2</sup>, Patrick G. Shaw<sup>1</sup>, Donald Freed<sup>1</sup>, Muhammad S. Zahari<sup>2</sup>, Kanchan K. Mukherjee<sup>8</sup>, Subramanian Shankar<sup>9</sup>, Anita Mahadevan<sup>10,11</sup>, Henry Lam<sup>12</sup>, Christopher J. Mitchell<sup>1</sup>, Susarla Krishna Shankar<sup>10,11</sup>, Parthasarathy Satishchandra<sup>13</sup>, John T. Schroeder<sup>14</sup>, Ravi Sirdeshmukh<sup>3</sup>, Anirban Maitra<sup>15,16</sup>, Steven D. Leach<sup>1,17</sup>, Charles G. Drake<sup>16,18</sup>, Marc K. Halushka<sup>15</sup>, T. S. Keshava Prasad<sup>3</sup>, Ralph H. Hruban<sup>15,16</sup>, Candace L. Kerr<sup>19†</sup>, Gary D. Bader<sup>5</sup>, Christine A. Iacobuzio-Donahue<sup>15,16,17</sup>, Harsha Gowda<sup>3</sup> & Akhilesh Pandey<sup>1,2,3,4,15,16,20</sup>

ARTICLE

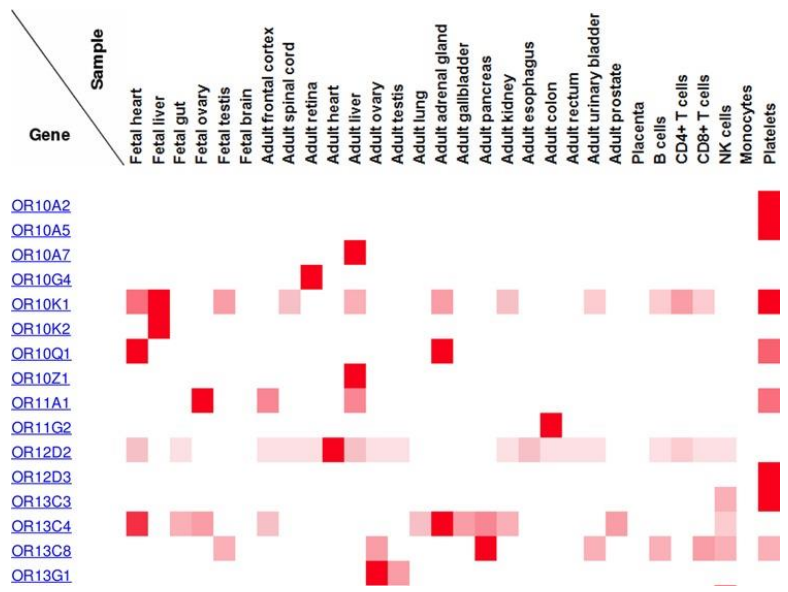
2014 Kuster lab

doi:10.1038/nature13319

## Mass-spectrometry-based draft of the human proteome

Mathias Wilhelm<sup>1,2\*</sup>, Judith Schlegl<sup>2\*</sup>, Hannes Hahne<sup>1\*</sup>, Amin Moghaddas Gholami<sup>1\*</sup>, Marcus Lieberenz<sup>2</sup>, Mikhail M. Savitski<sup>3</sup>, Emanuel Ziegler<sup>2</sup>, Lars Butzmann<sup>2</sup>, Siegfried Gessulat<sup>2</sup>, Harald Marx<sup>1</sup>, Toby Mathieson<sup>3</sup>, Simone Lemeer<sup>1</sup>, Karsten Schnatbaum<sup>4</sup>, Ulf Reimer<sup>2</sup>, Holger Wenschuh<sup>4</sup>, Martin Mollenhauer<sup>5</sup>, Julia Slotta-Huspenina<sup>5</sup>, Joos-Hendrik Boese<sup>2</sup>, Marcus Bantscheff<sup>3</sup>, Anja Gerstmair<sup>2</sup>, Franz Faerber<sup>2</sup> & Bernhard Kuster<sup>1,6</sup>

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

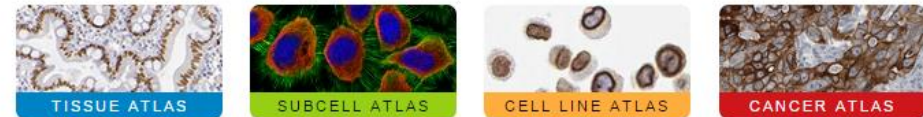
## THE HUMAN PROTEIN ATLAS

ABOUT & HELP

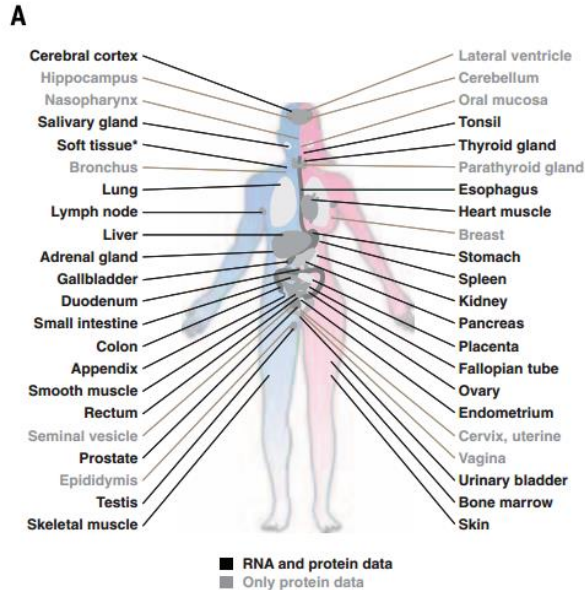
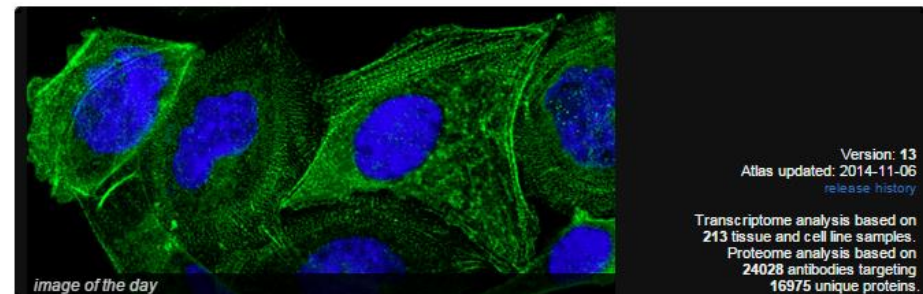
**A Tissue-Based Map of the Human Proteome**

Here, we summarize our current knowledge regarding the human proteome mainly achieved through antibody-based methods combined with transcriptomics analysis across all major tissues and organs of the human body. A large number of lists can be accessed with direct links to gene-specific images of the corresponding proteins in the different tissues and organs.

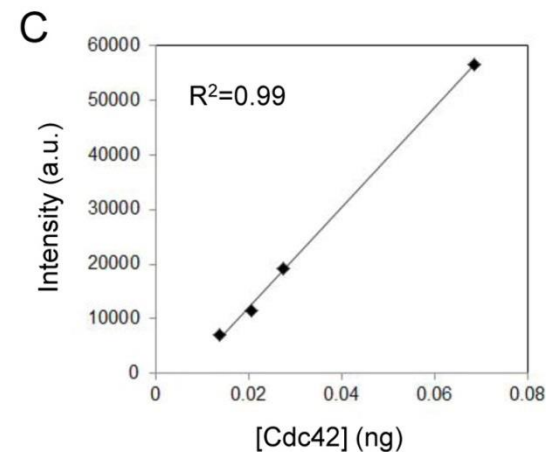
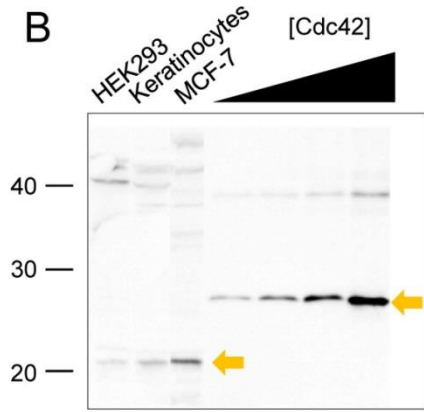
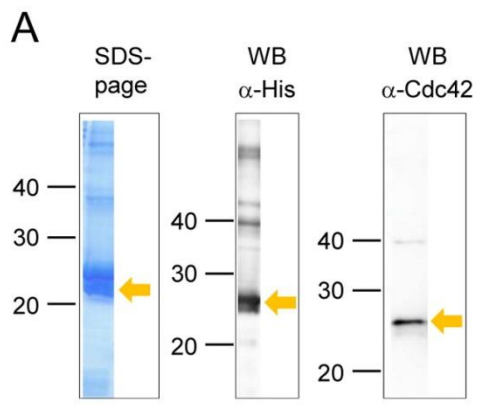
[Read more](#)



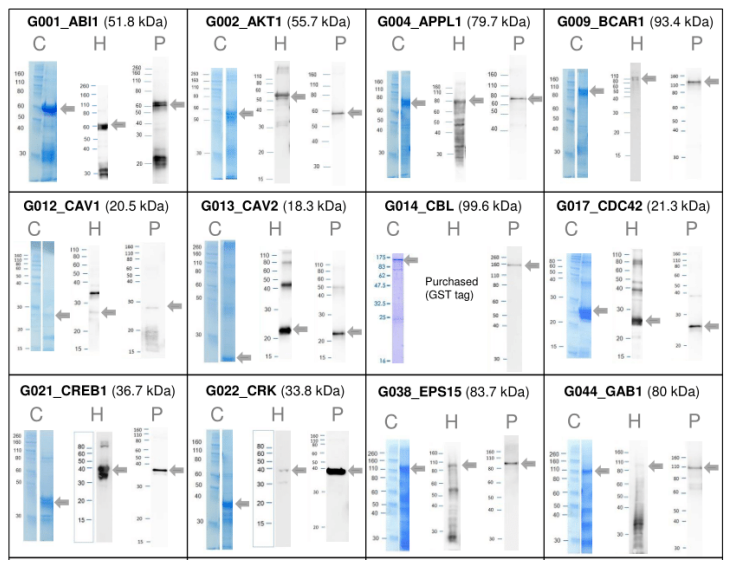
SEARCH ? »


[Fields »](#)


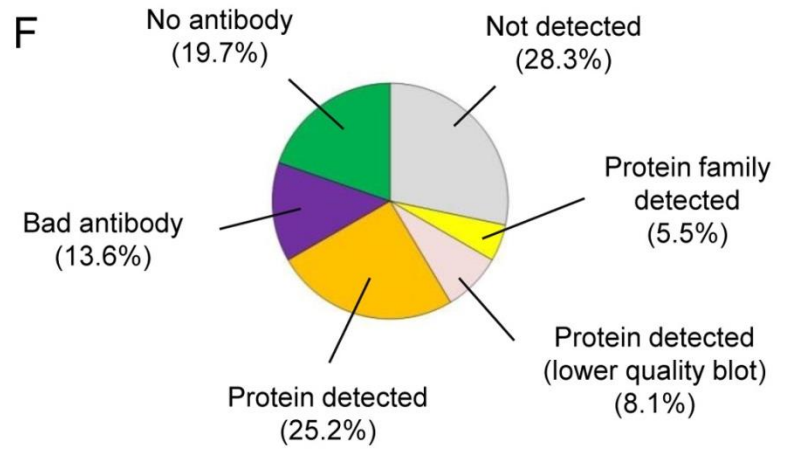
# Quantitative Western blotting



Protein standards: expression, purification and quantification

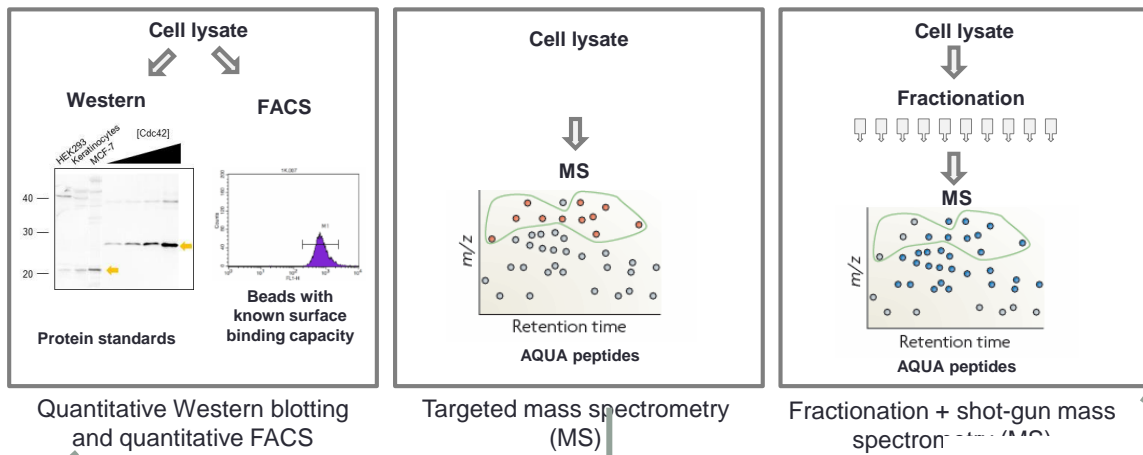


Summary statistic for quantitative Western blotting of 198 ErbB-related proteins





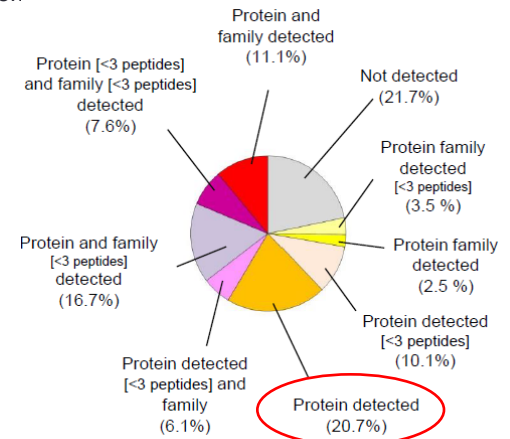
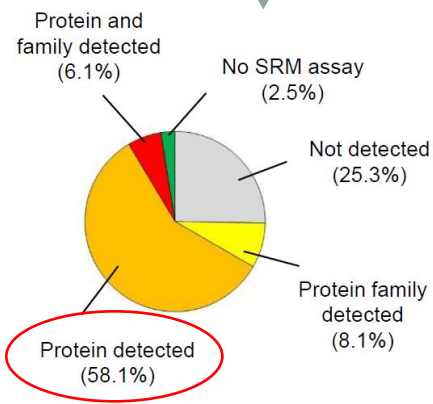
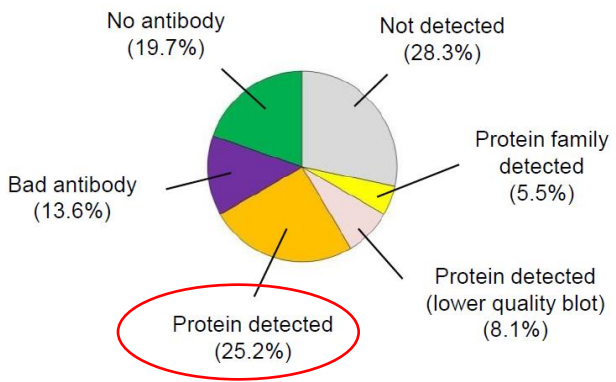
# Combining different quantitative approaches to quantify 198 proteins in the ErbB signaling pathway



Quantitative Western blotting and quantitative FACS

Targeted mass spectrometry (MS)

Fractionation + shot-gun mass spectrometry (MS)

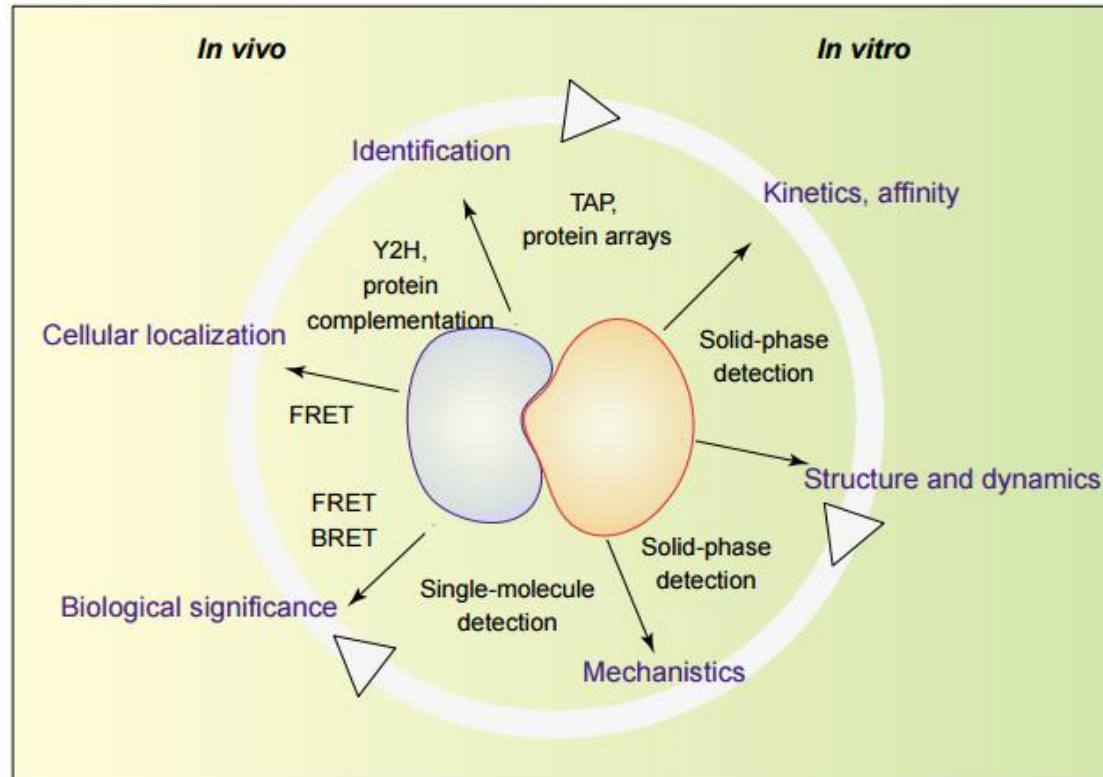


- SRM has a higher sensitivity compared to quantitative western blotting (but some proteins are only detected by Western blotting)
- Problem with isoforms and protein families: as a consequence of frequent gene duplication events in mammals, often similar proteins (e.g. AKT1 and AKT2) cannot be distinguished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

## Measuring protein interactions *in vivo* and *in vitro*

The challenge:

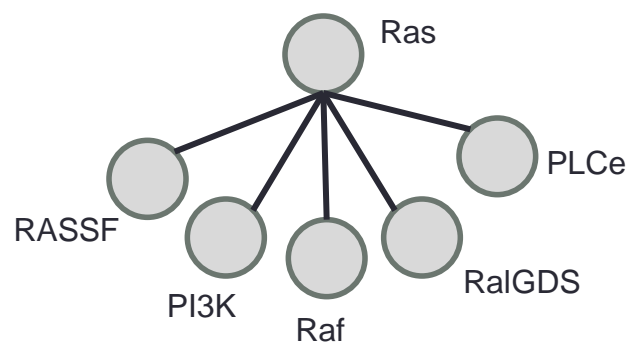
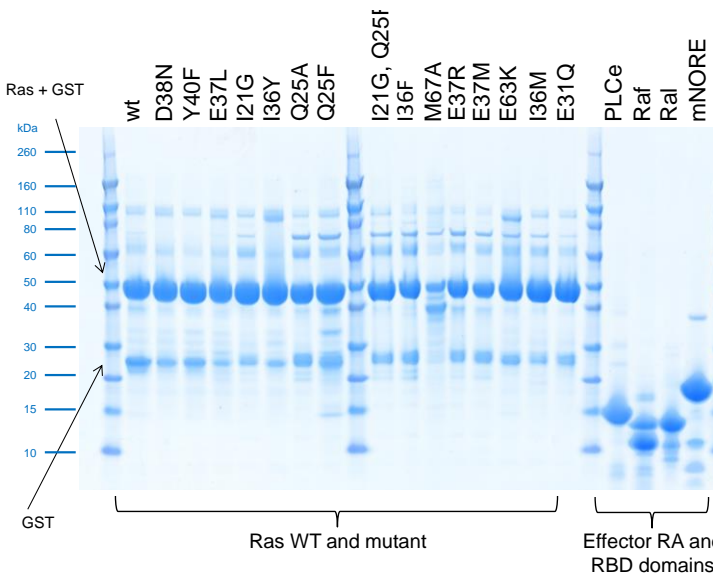
- most *in vivo* techniques are high-throughput, but do not provide affinities (only qualitative binding detection)
- *in vitro* techniques can provide affinities and kinetic constants, but are not high-throughput methods



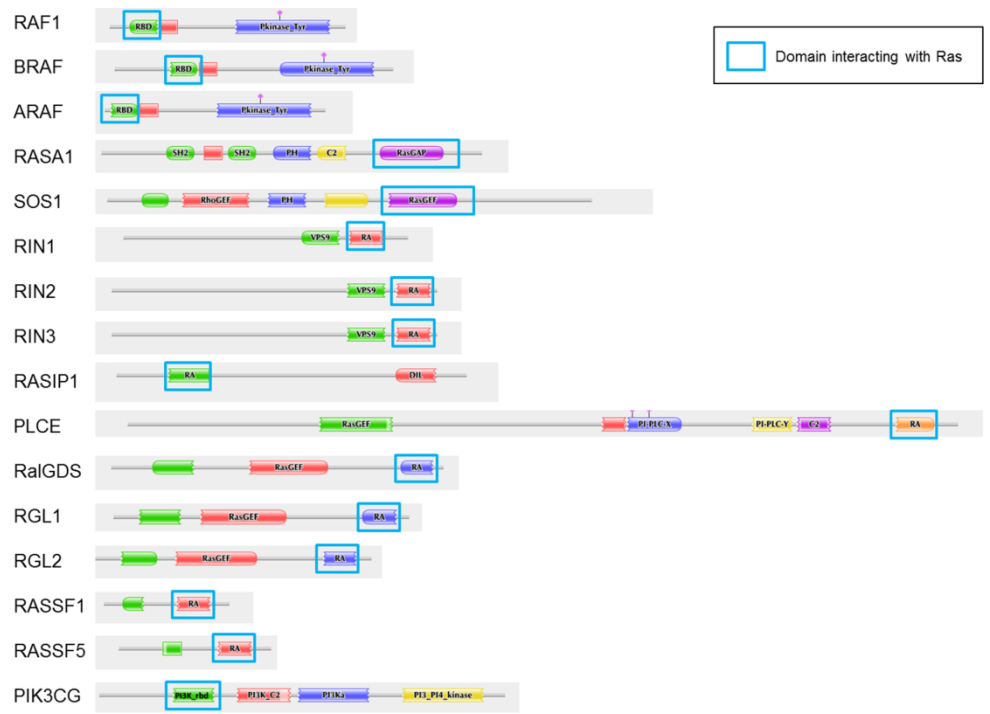
Current Opinion in Structural Biology

# 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

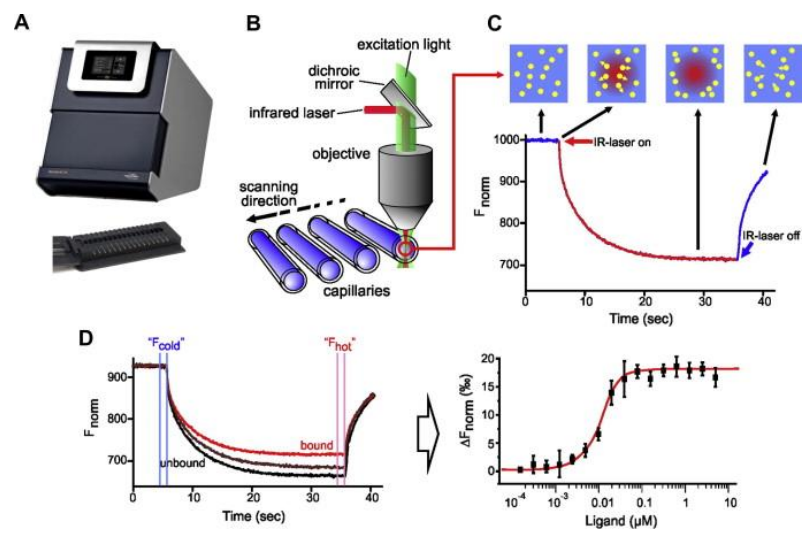


Large proteins are often not soluble: expression and purification of protein domains

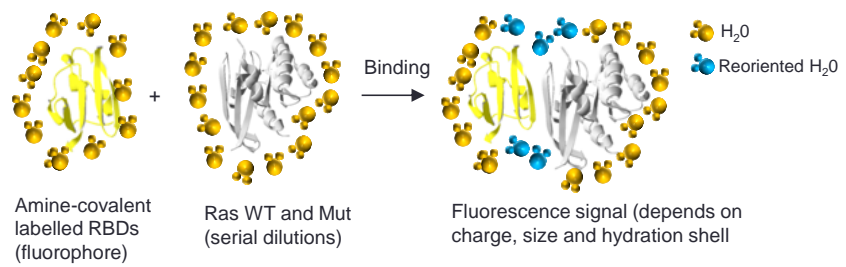


# Two main methods to measure affinities and kinetic constants

## Microscale thermophoresis



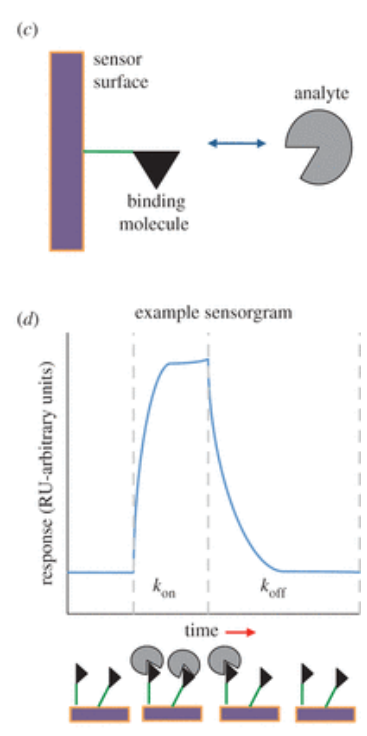
Jerabek-Willemsen et al, J Mol Struct, 2014



➤ Provides only the affinity in equilibrium ( $K_d$  value), but not kinetic constants

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

## Surface plasmon resonance

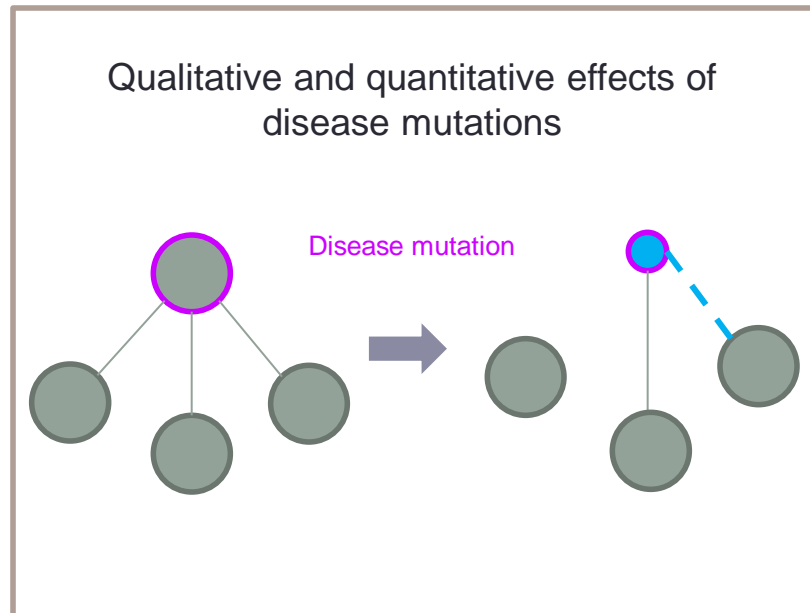


Optical method to measure the refractive index near a sensor surface

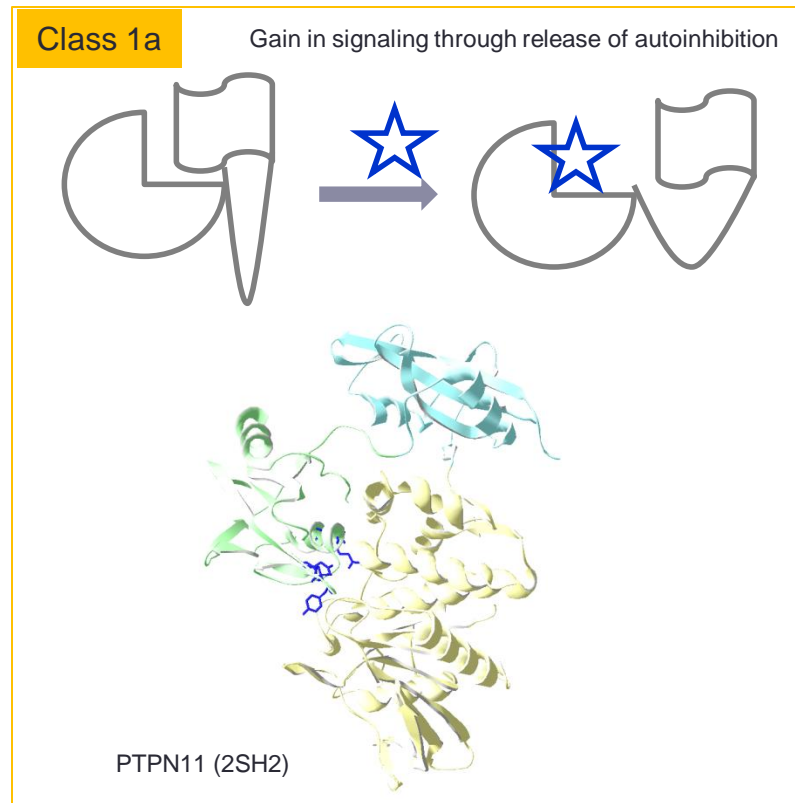
Kastritis et al, 2012

➤ Provides kinetic constants ( $k_{on}$  and  $k_{off}$ )

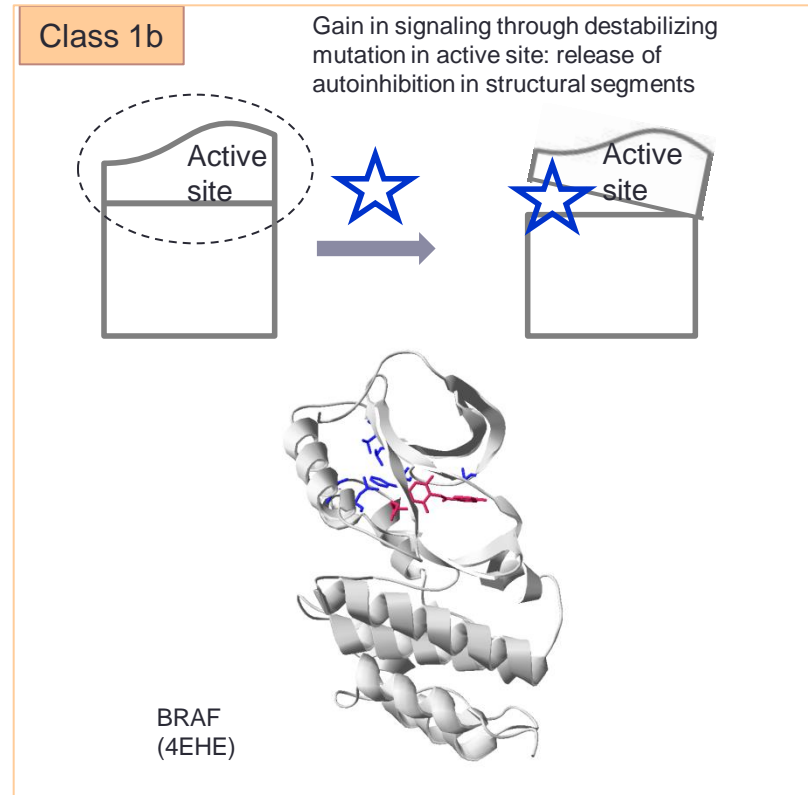
$$K_d = \frac{k_{off}}{k_{on}}$$



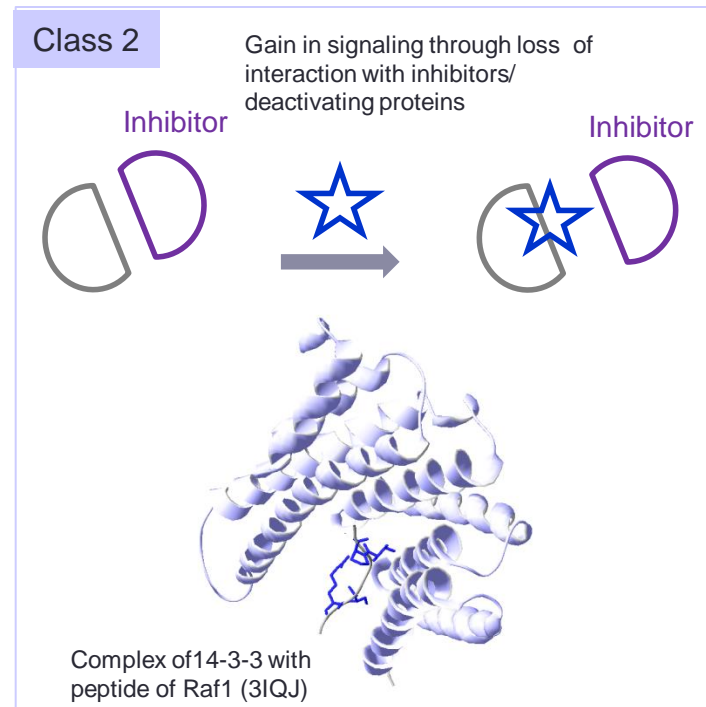
# Examples how missense mutations can affect the network: a 3D structural perspective



# Examples how missense mutations can affect the network: a 3D structural perspective

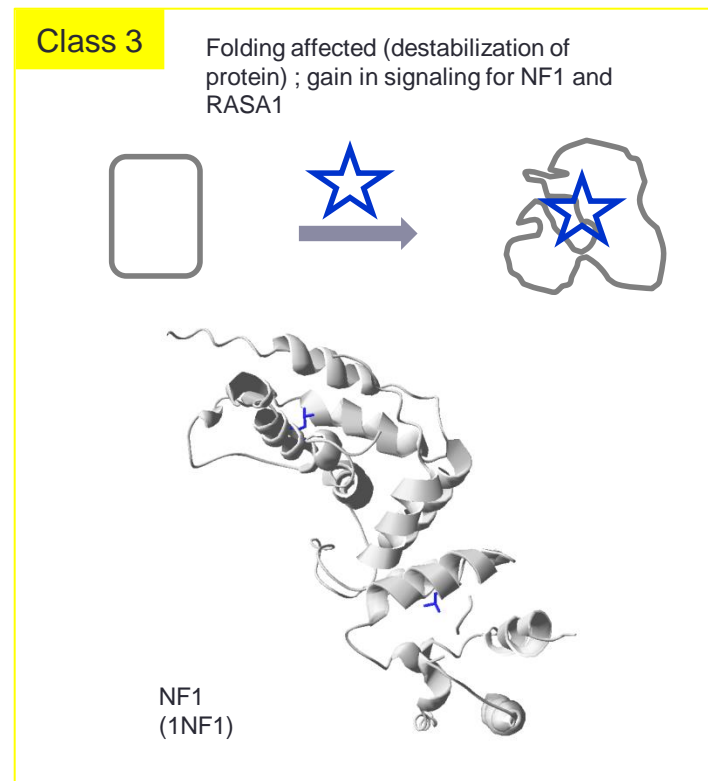


# Examples how missense mutations can affect the network: a 3D structural perspective

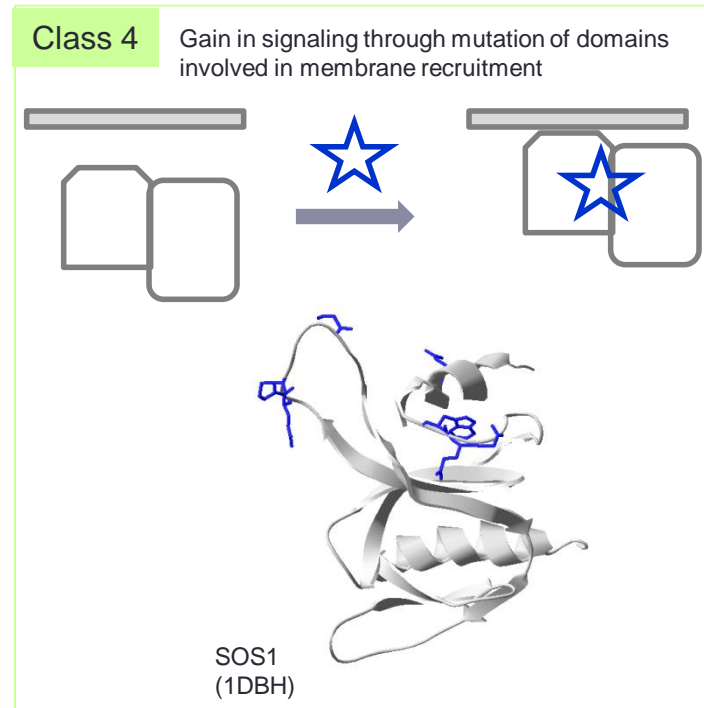




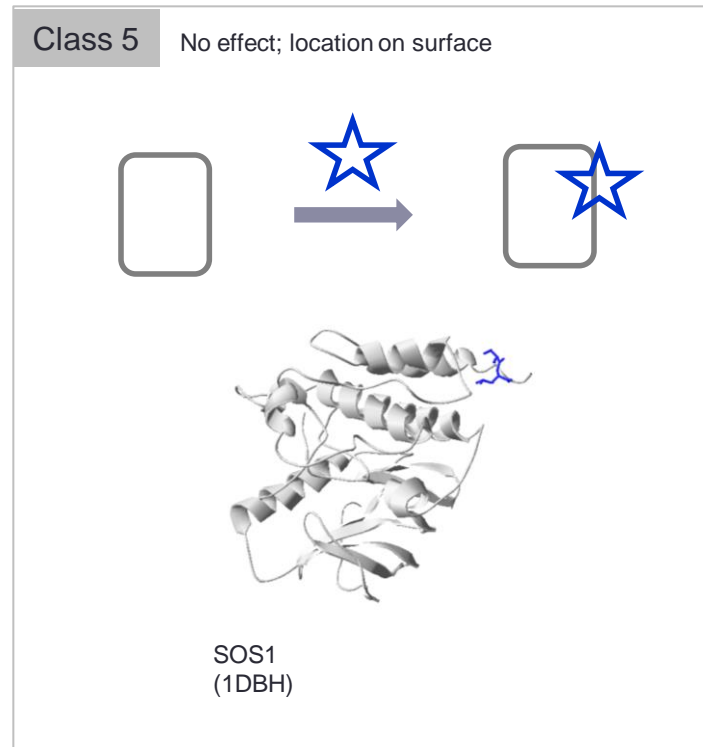
# Examples how missense mutations can affect the network: a 3D structural perspective



# Examples how missense mutations can affect the network: a 3D structural perspective



# Examples how missense mutations can affect the network: a 3D structural perspective

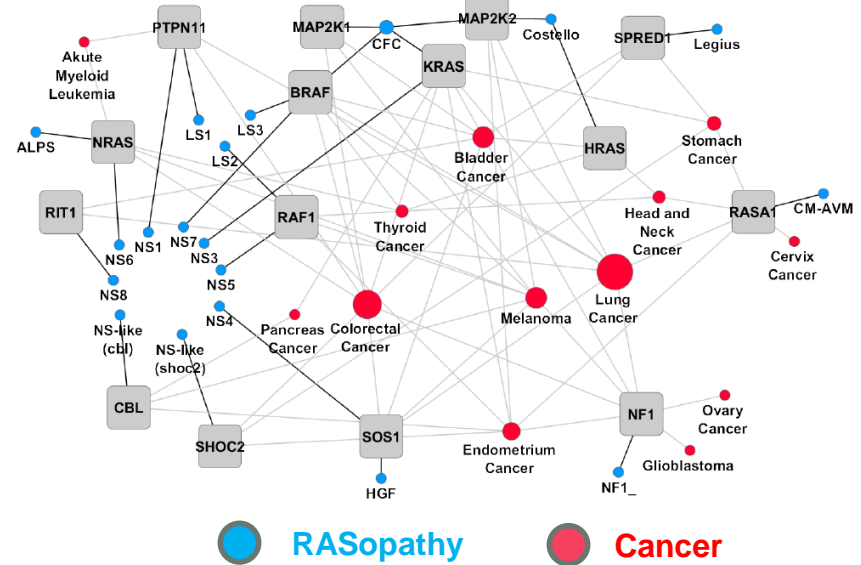
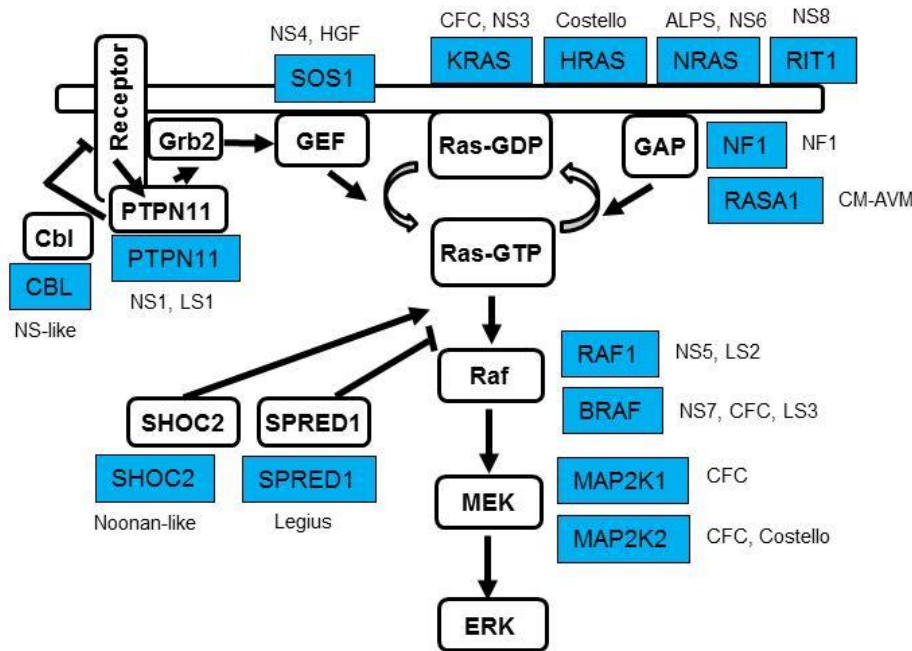


## Example 1: RASopathy and cancer disease mutations

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

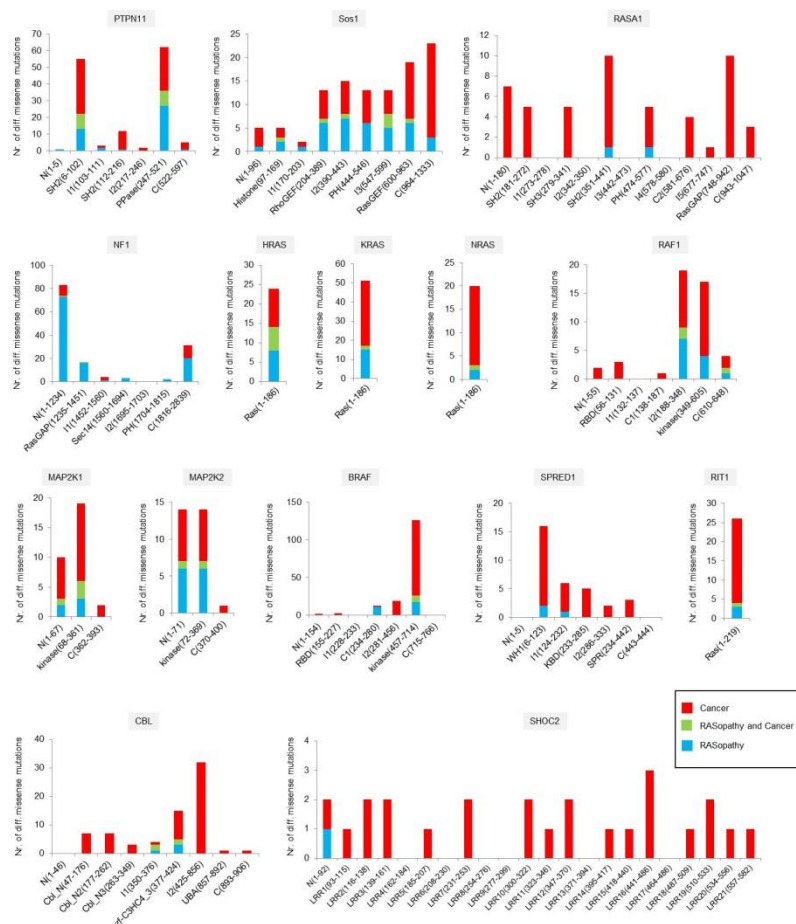
- Ras/MAPK syndromes ('RASopathies') are a class of developmental disorders caused by germline mutations

- Proteins in Ras/MAPK syndromes ('RASopathies') are also found in cancer

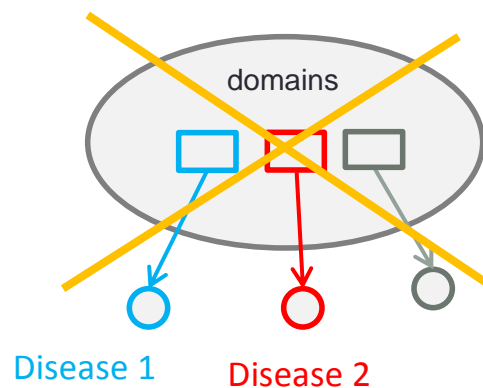
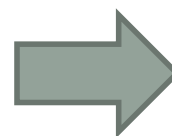


# Location of mutations in different domains does not explain the difference between RASopathy and cancer mutations

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

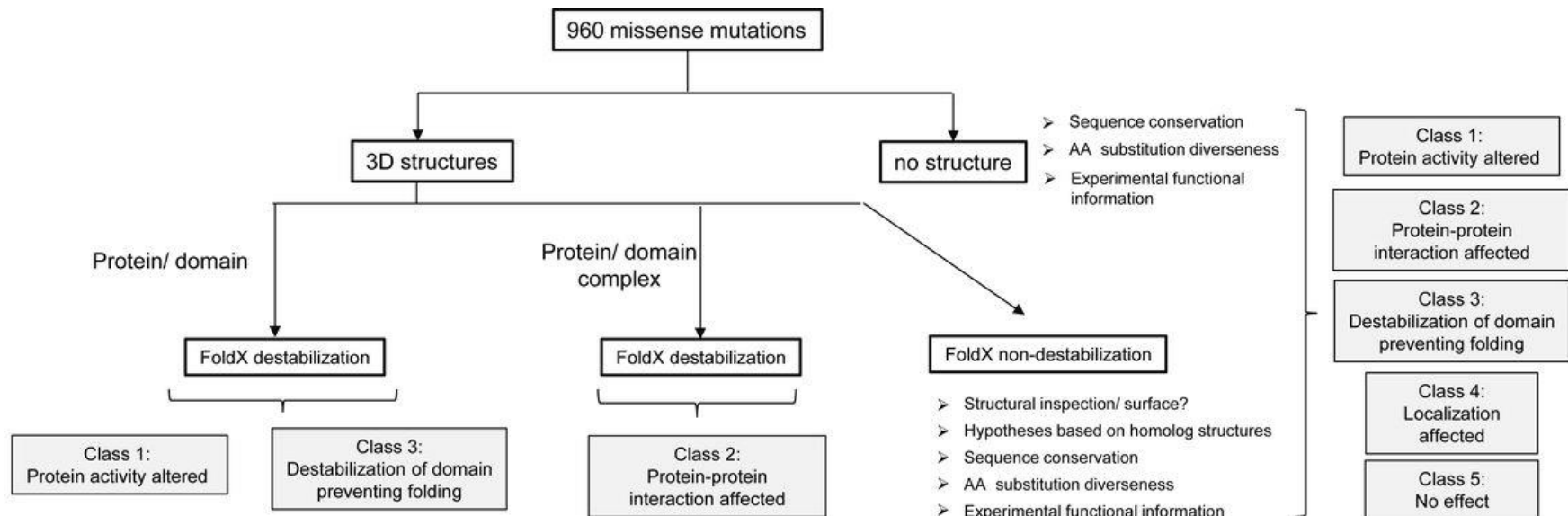


**'Edgetics' does not explain it**  
 Domain localization of mutation does not explain why a particular mutation will cause RASopathy or cancer



# Analysis of 956 missense mutations in RASopathies and cancer based on structural information and FoldX energies

## Pipeline:

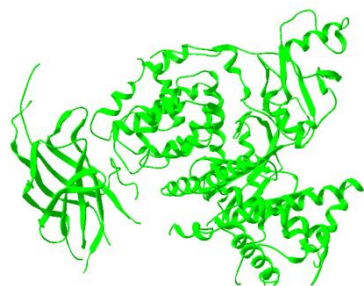


# FoldX-based energy calculations of proteins

3D Structural information

A force field for energy calculations and protein design

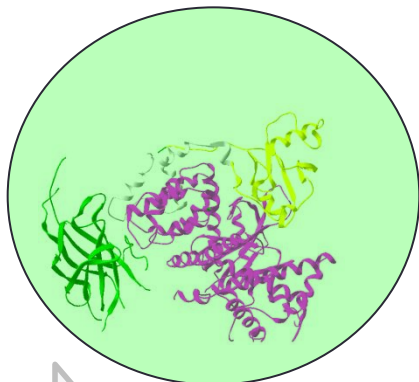
Schymkowitz et al, *Nucleic Acids Res*, 2005



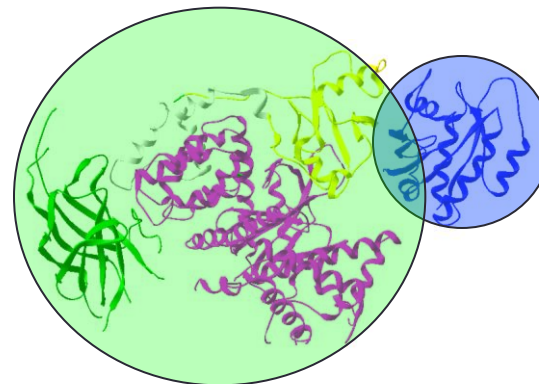
$$+ \text{FoldX} = \Delta G$$

Relation to affinity:  $\Delta G = RT \ln K_d$

✓ Total free energy

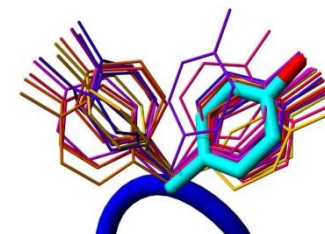


✓ Interaction energy

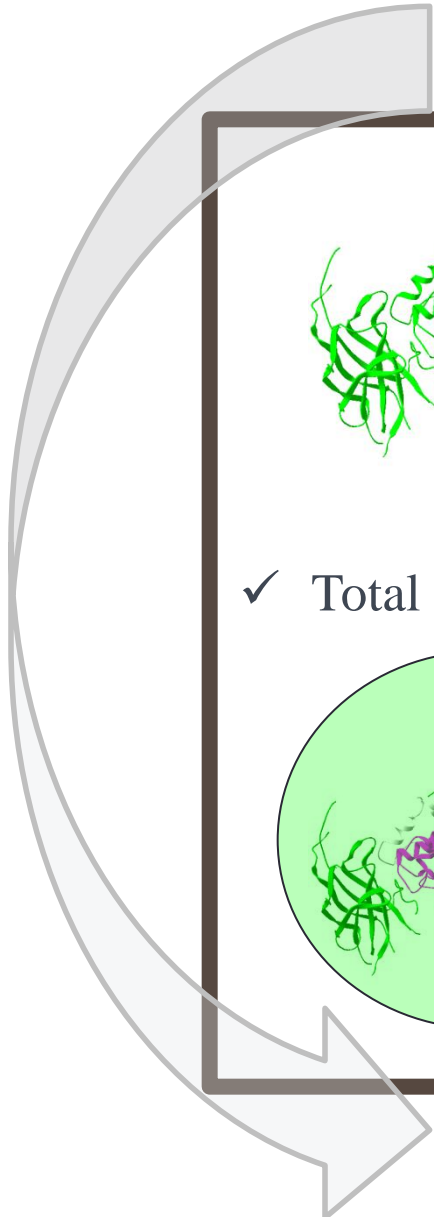


✓ Mutagenesis

A rotamer library to replace the 20 amino acids

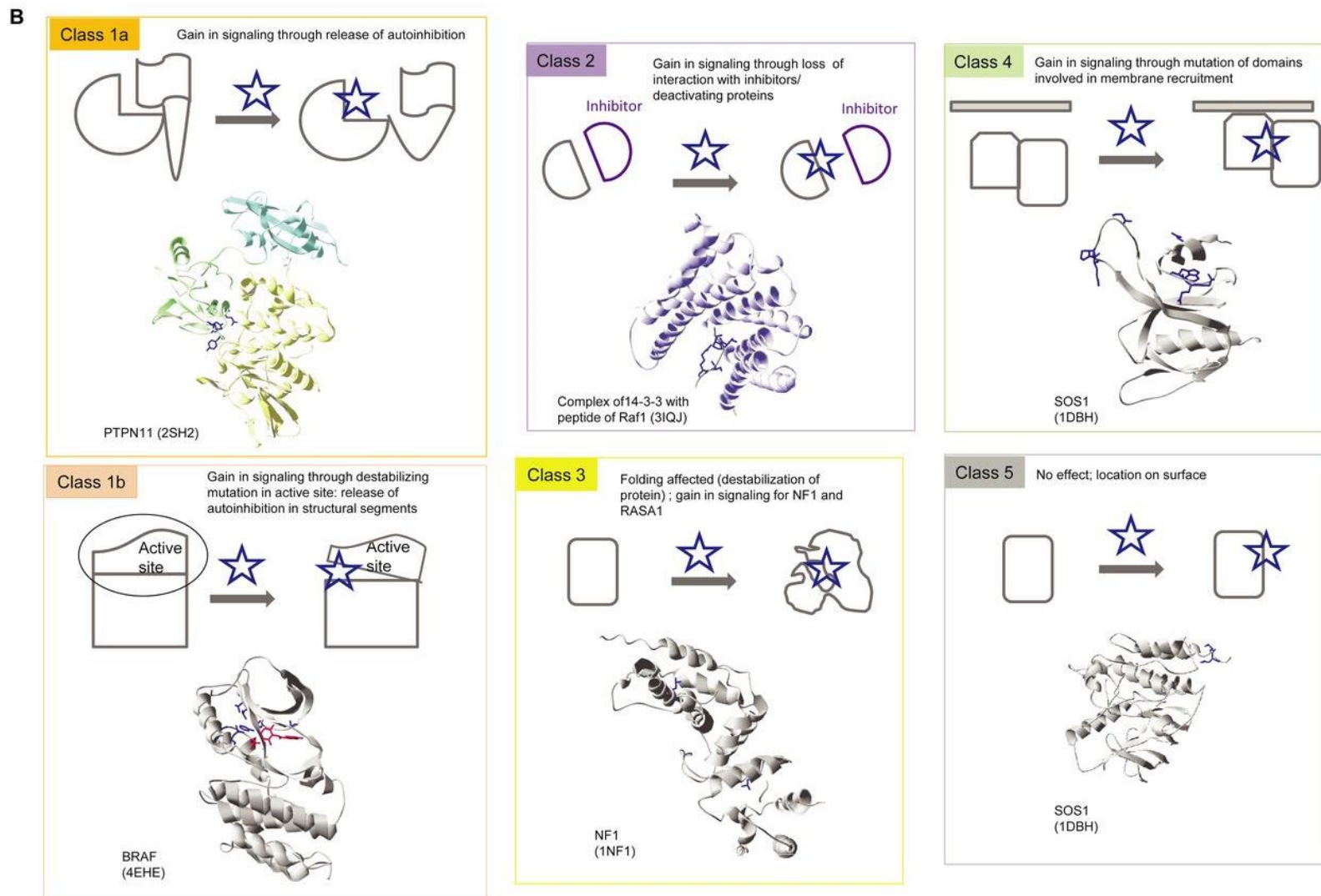


Protein design

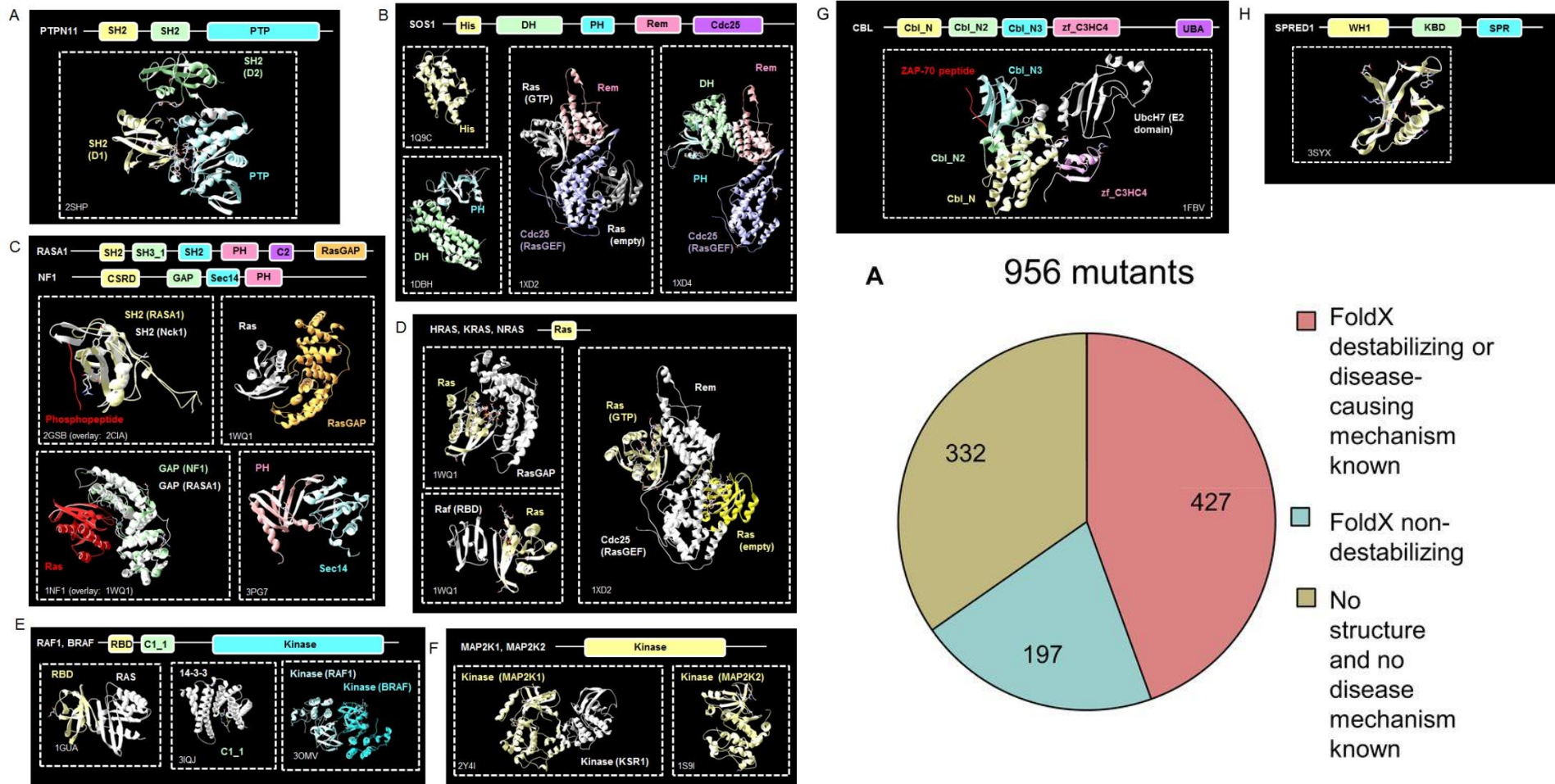




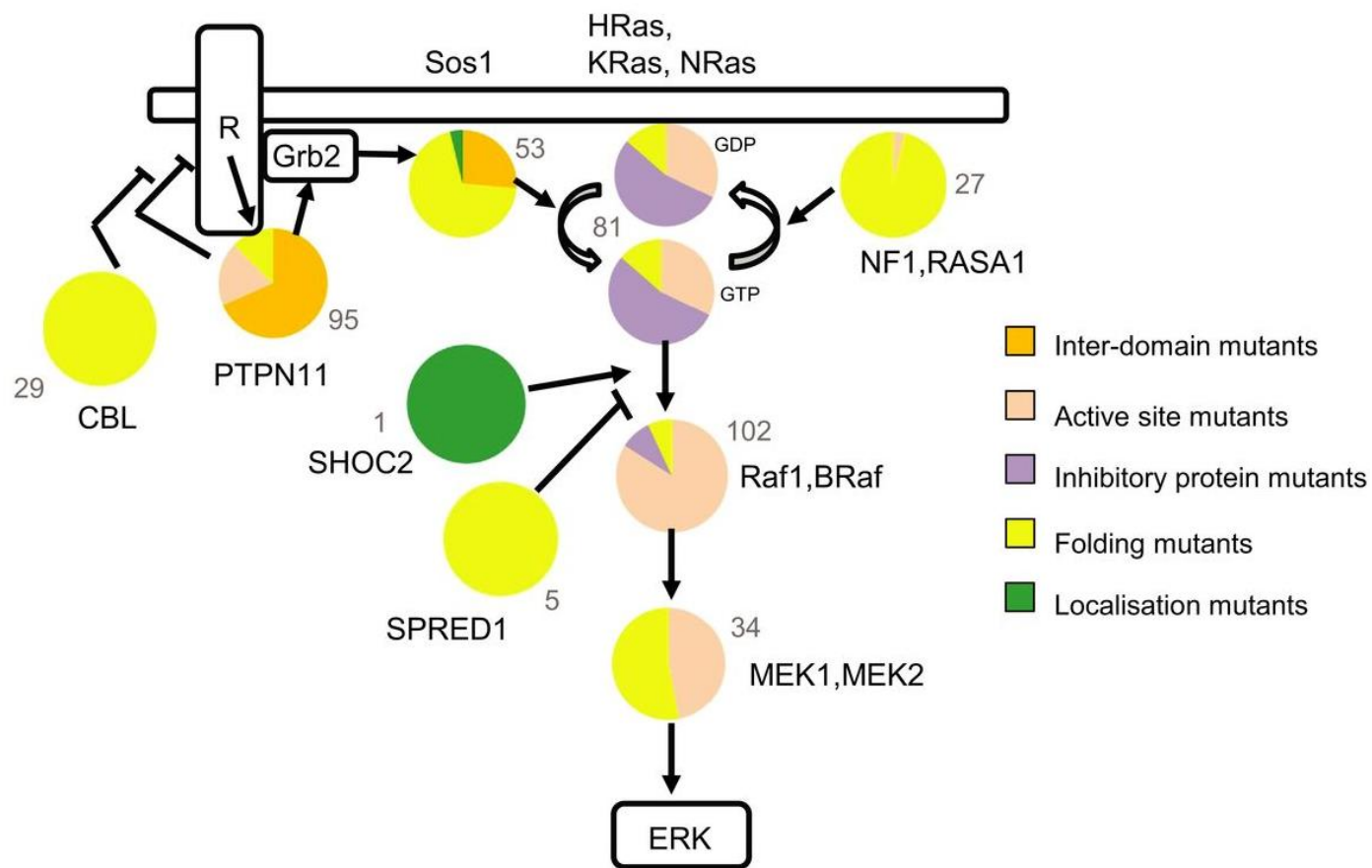
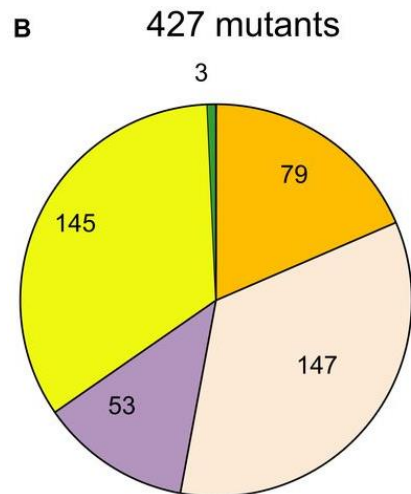
# Analysis of 956 missense mutations in RASopathies and cancer based on structural information and FoldX energies



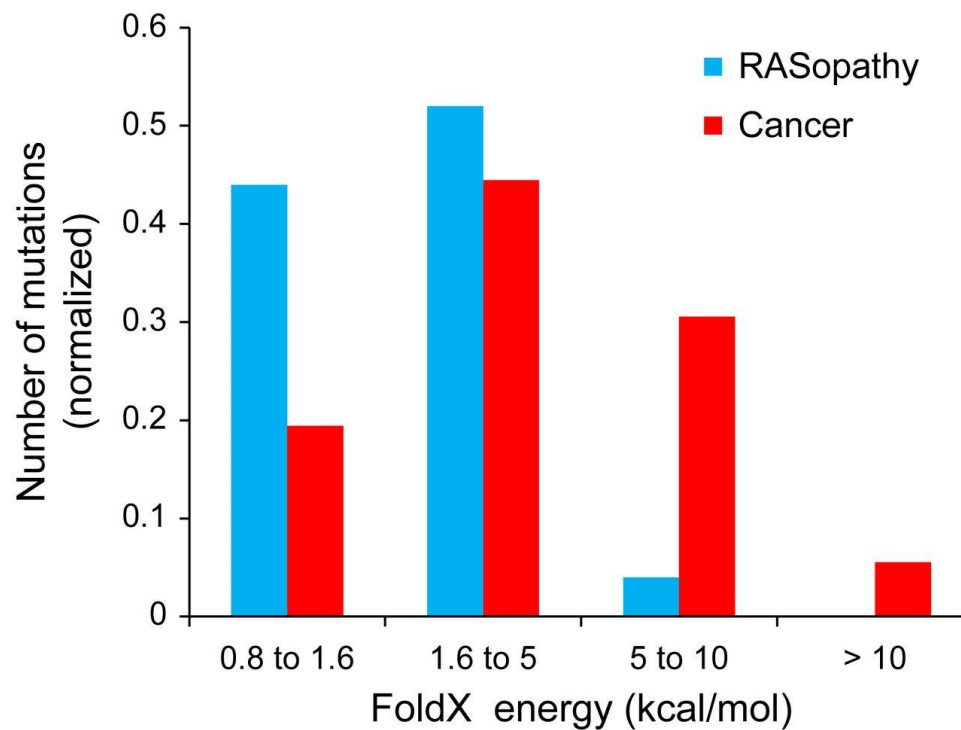
# Analysis of 956 missense mutations in RASopathies and cancer: high structural coverage



# Multiple effects of a mutation even for the same protein/ protein class

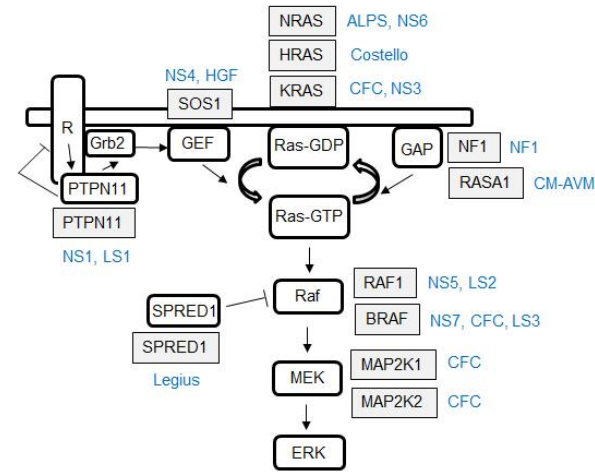
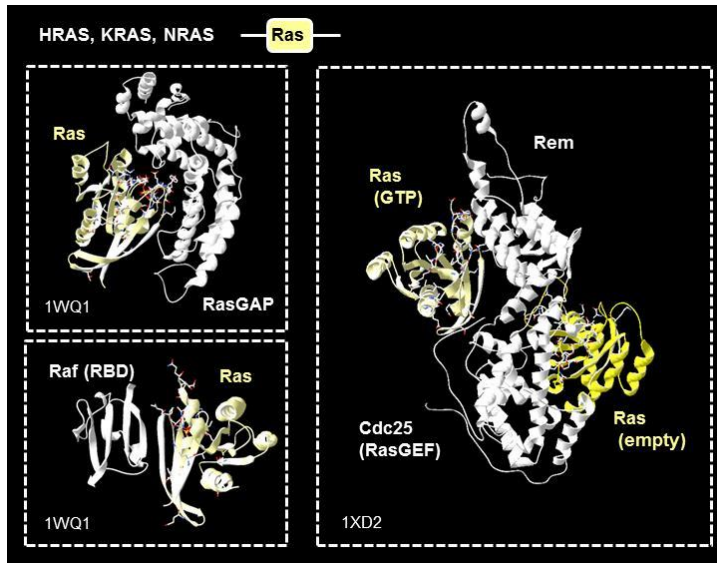


## Cancer mutations tend to have higher destabilization values (on average)

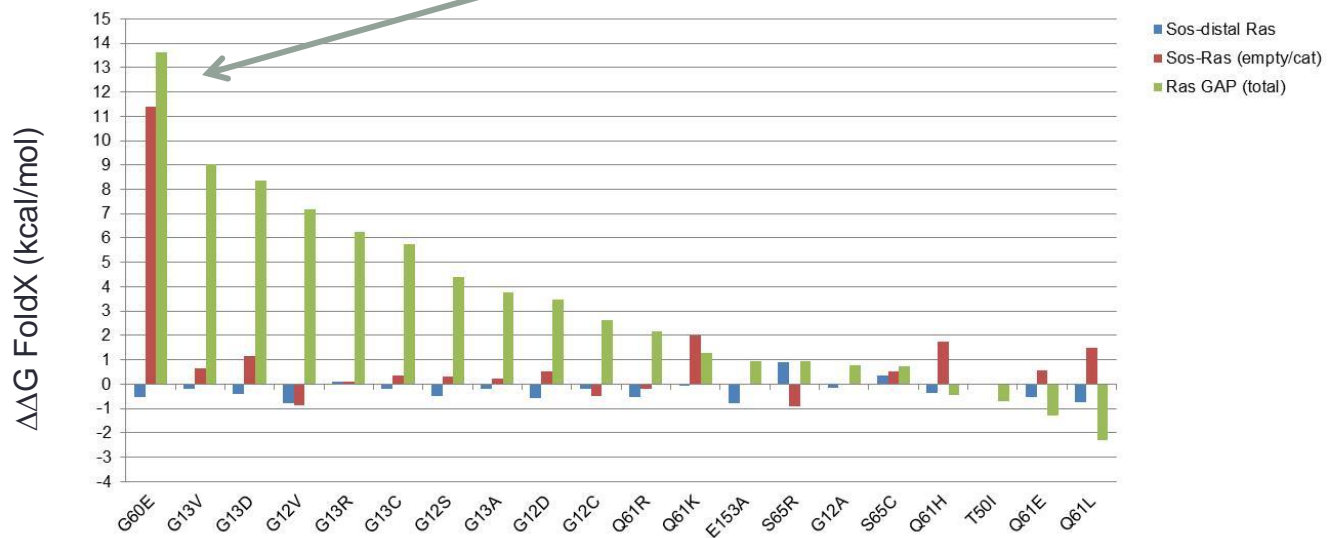


FoldX

# Compensatory effects of mutations on different interaction partners

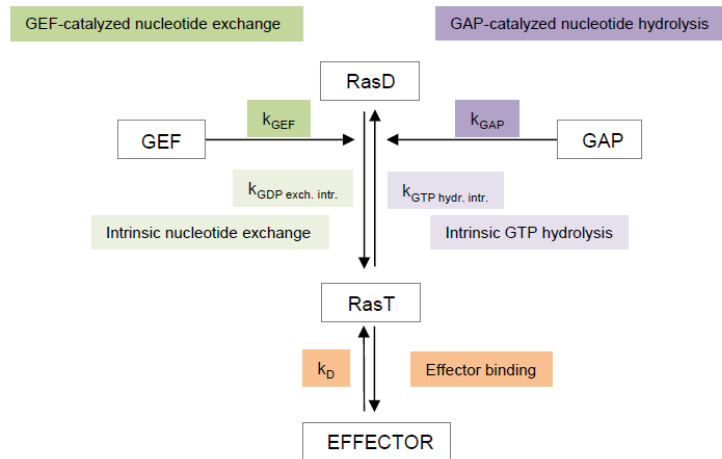


NRAS G60E



# Quantitative effects on protein stability, or activity could explain in some cases the different phenotype: cancer or RASopathy

A



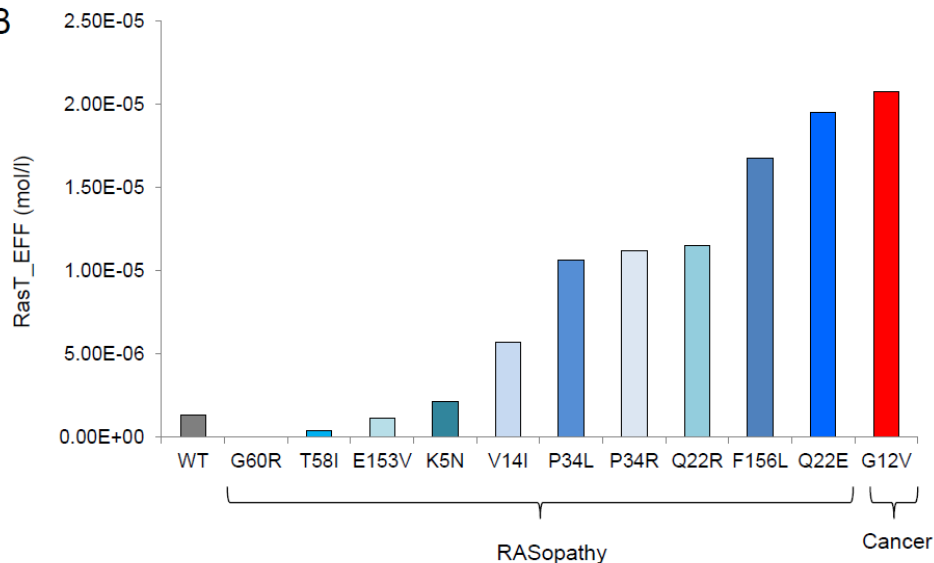
Simulation of Ras activation

‘Enegetics’: quantitative edge effects

‘Edgetics’ + energies = ‘enegetics’

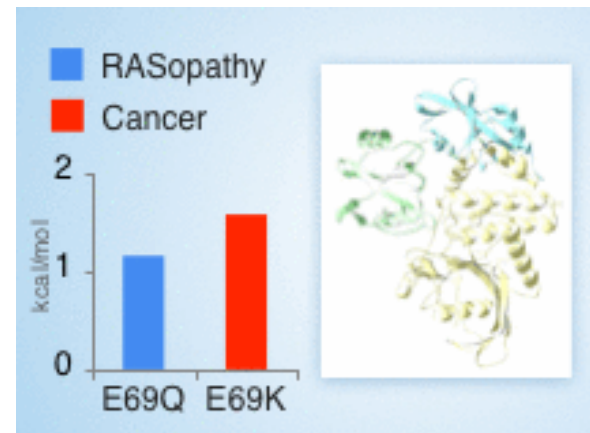
Quantitative effects on protein stability, activity, or folding explains in some cases the different phenotype

B



## 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.

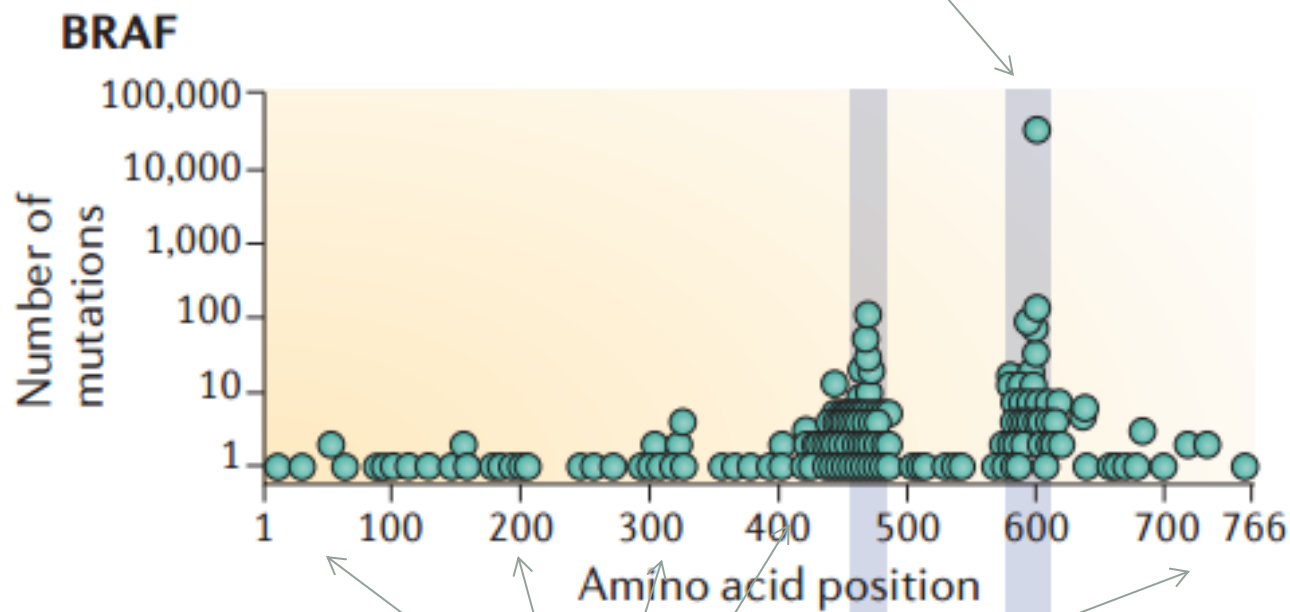


## Example 2: BRAF mutations in cancer. Why V600E?



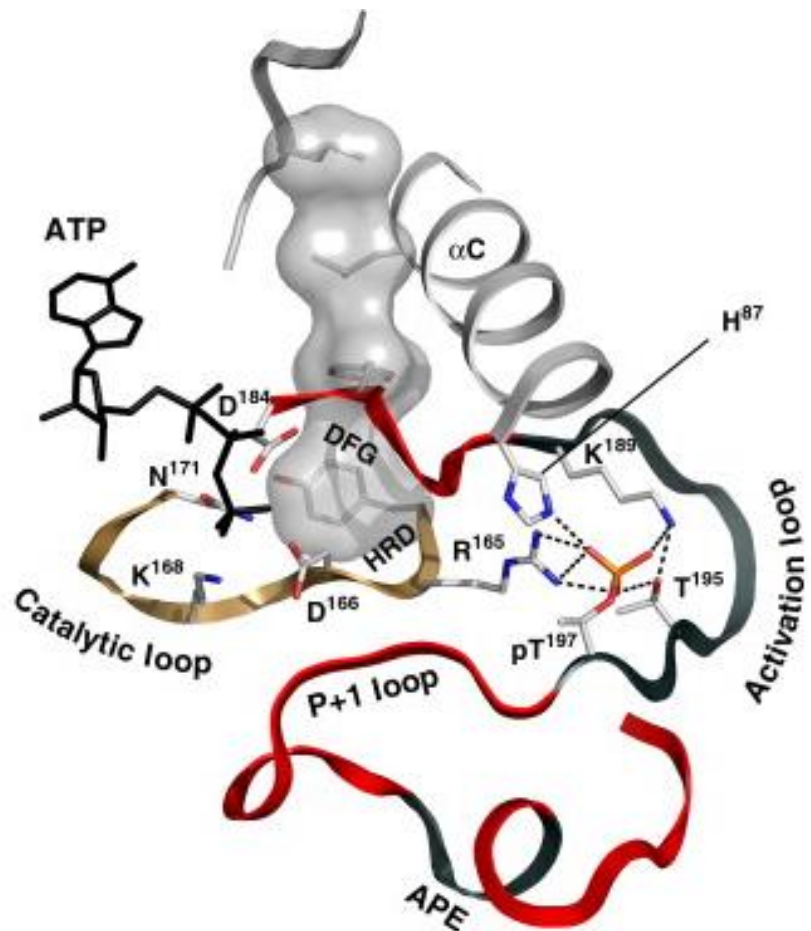
The most common BRAF mutation is V600E and induces constitutive kinase activation

Patients are treated with a BRAF kinase inhibitor



Shall we only treat patients which harbour V600E mutations or also patients with non-V600E mutations?

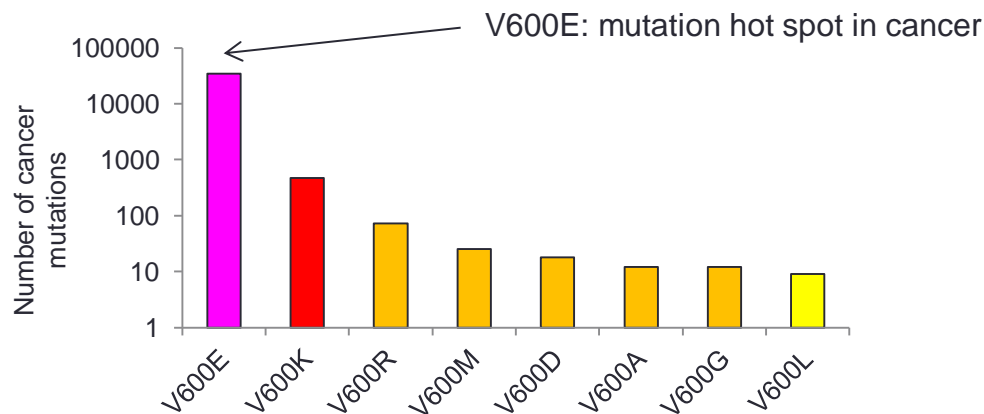
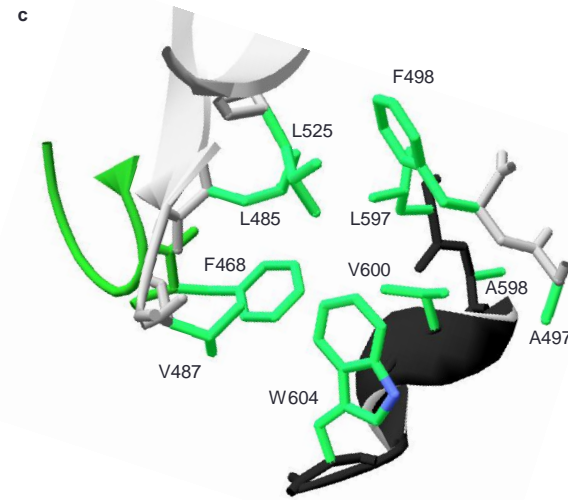
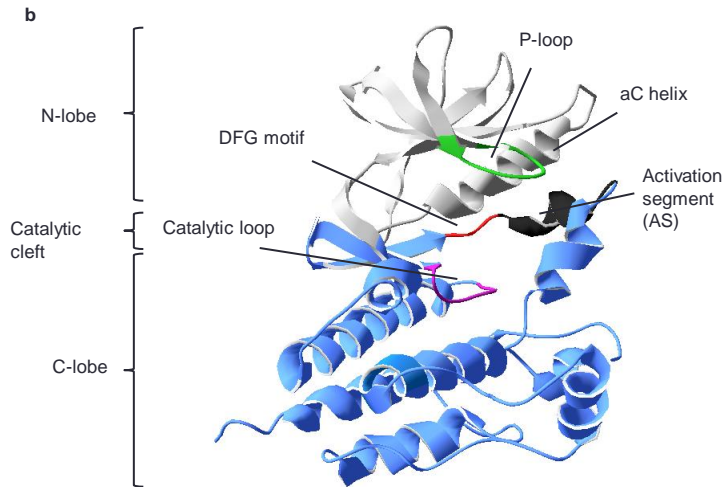
## Kinases are activated through mutations in the activation loop (activation segment)



- phosphorylation in the activation segment causes structural **rearrangements of the activation segment and the  $\alpha$ C helix**. This reorients the DFG loop resulting in activation of the kinase

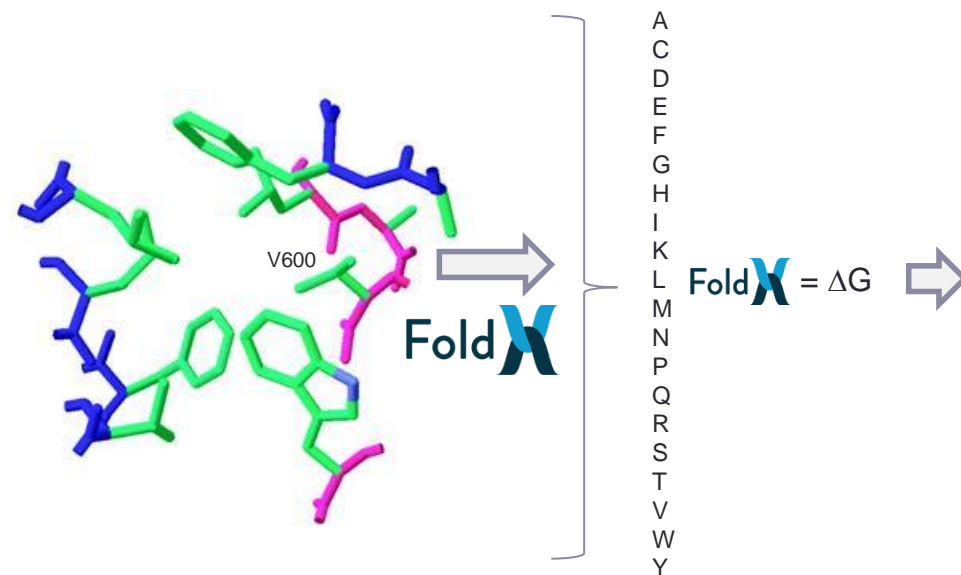
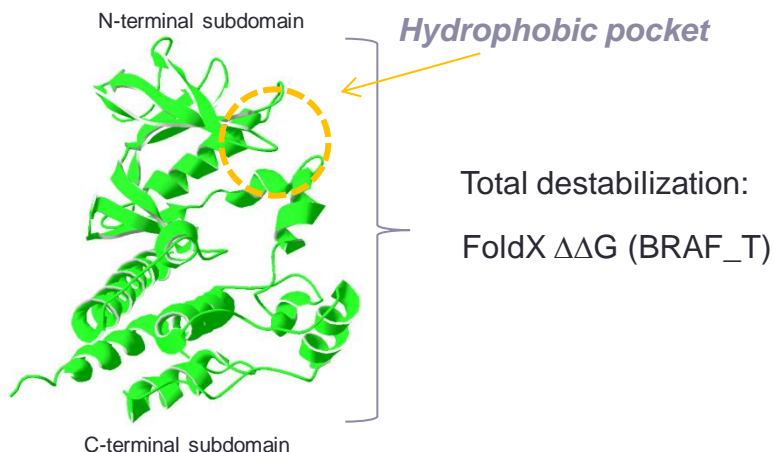
# Focus on the position Val600 in the kinase BRAF

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

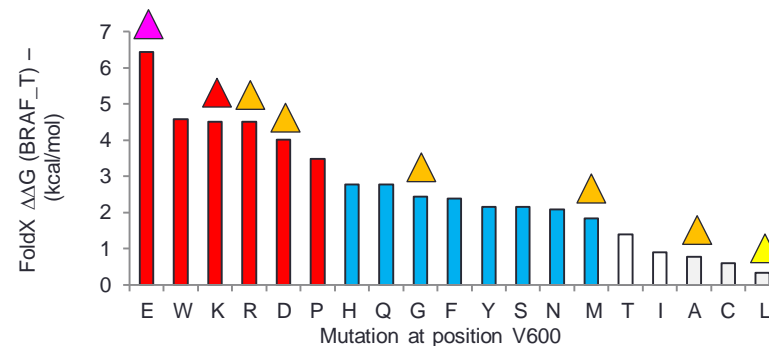
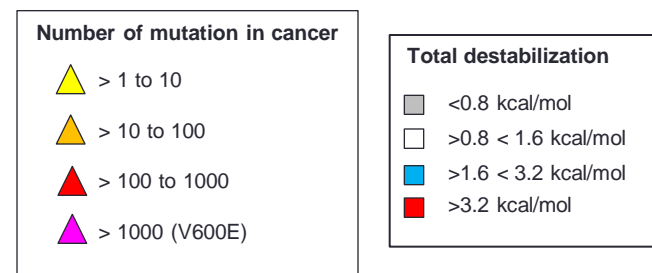


Differences in mutation frequencies: a quantitative effect?

# The V600E mutation causes a high destabilization of the inactive state (aC helix/AS hydrophobic pocket)

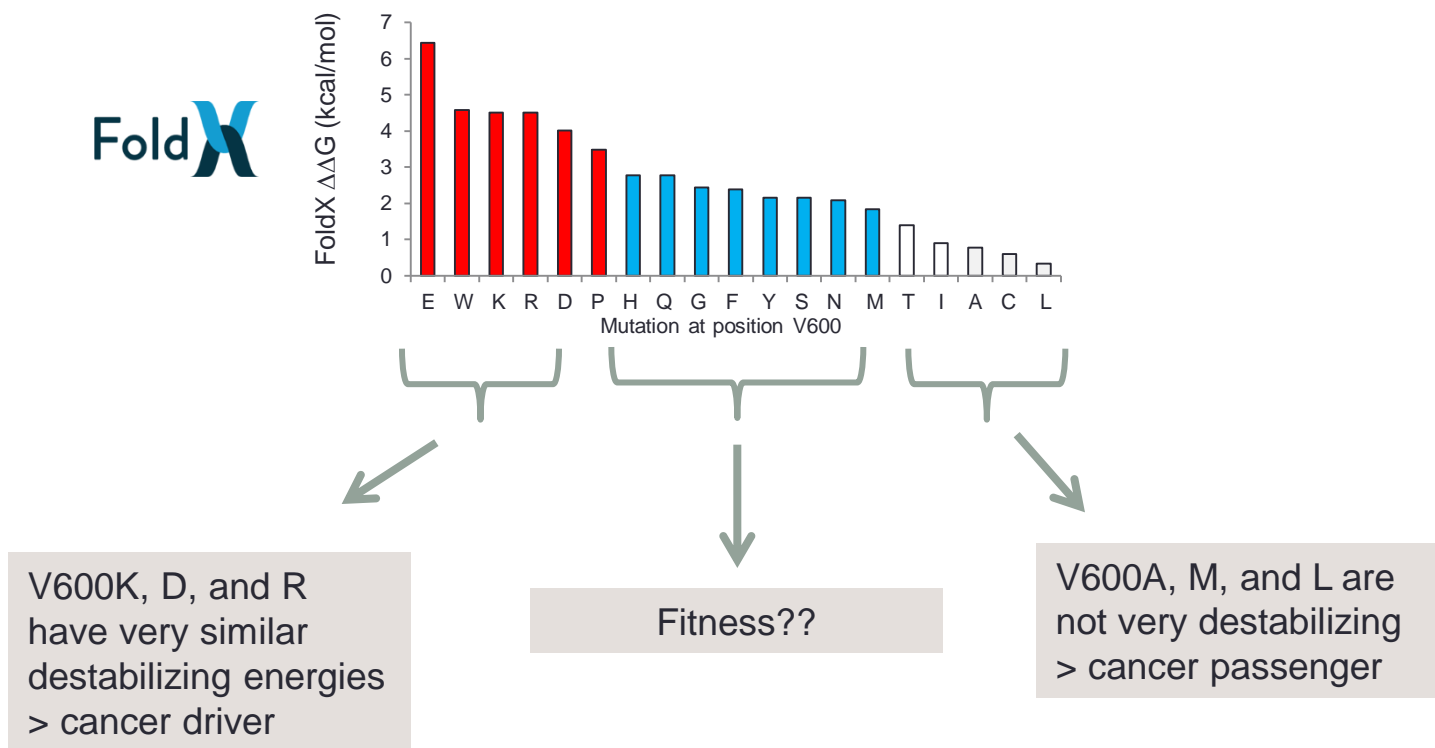


## Destabilization of inactive state



No destabilization of active state (data not shown)

## Distinguishing driver from passenger mutations



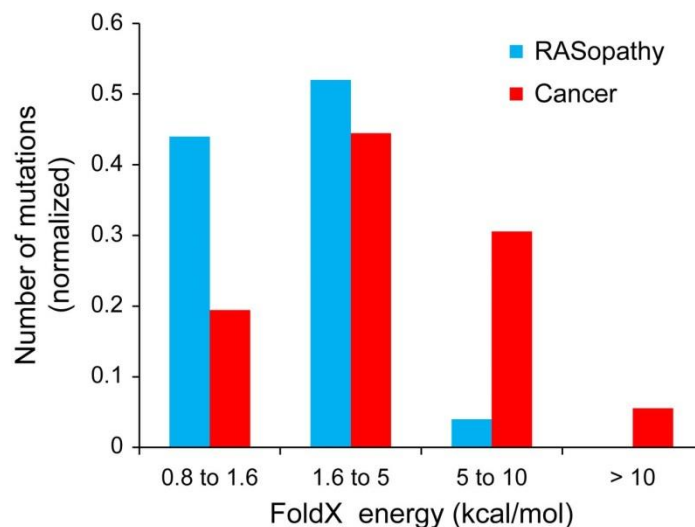
## V600G behaves more like a RASopathy mutation

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, KJ<sup>1</sup>](#); [Bunag, C<sup>2</sup>](#); [Estep, AL<sup>2</sup>](#); [Jones, JR<sup>1</sup>](#); [Bolt, CH<sup>1</sup>](#); [Rogers, RC<sup>1</sup>](#); [Rauen, KA<sup>3</sup>](#); [Everman, DB<sup>1</sup>](#)

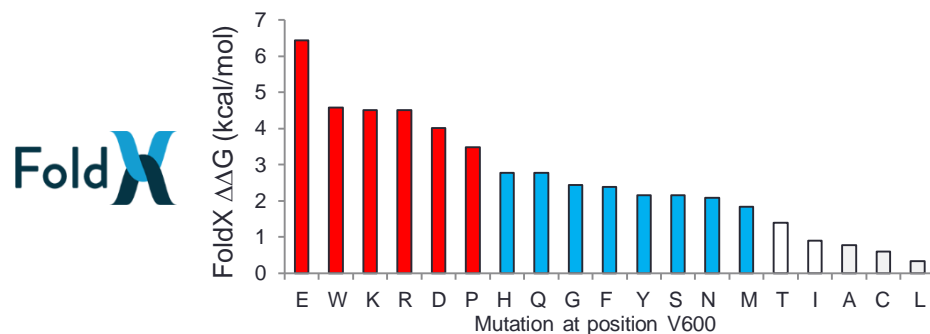
*Clinical Genetics*, Volume 79, issue 5 (May 2011), p. 468-474.  
ISSN: 0009-9163 DOI: 10.1111/j.1399-0004.2010.01495.x  
Blackwell Publishing Ltd



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

Kiel & Serrano, 2014

## Why different cancer frequencies for V600E, V600D and V600K?

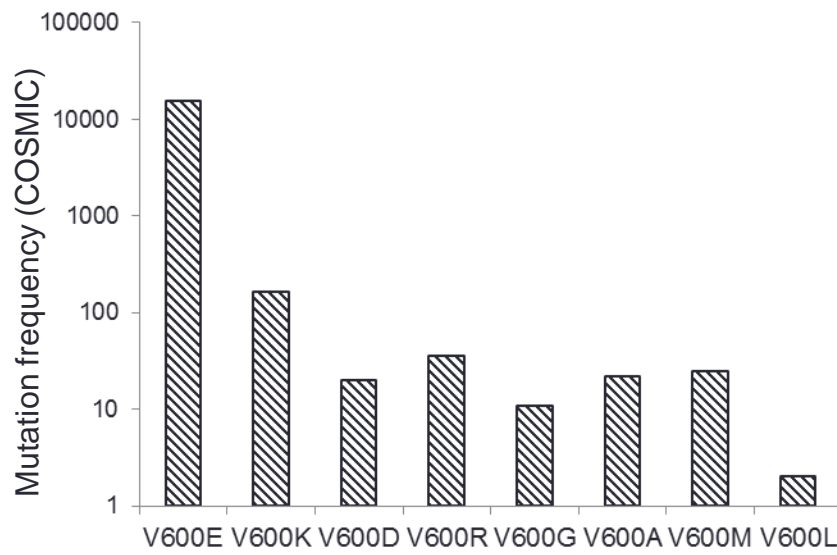


V600K, D, and R have very similar destabilizing energies

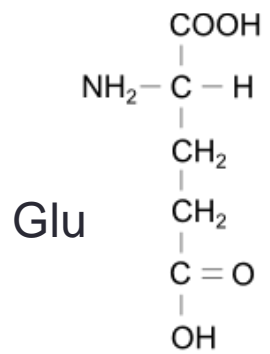
Why is V600E the by far most frequent mutation?

aa	frequency
Glu	15474
Lys	164
Arg	36
Met	25
Ala	22
Asp	20
Gly	11
Leu	2

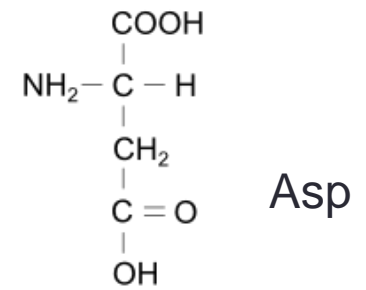
# Why different cancer frequencies for V600E, V600D and V600K?



V600E: 15474 frequency  
 V600D: 20 frequency



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?

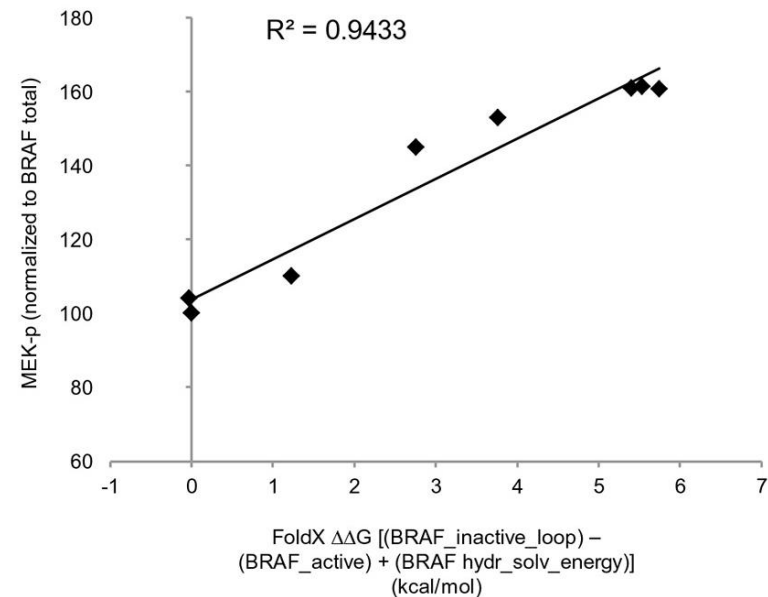
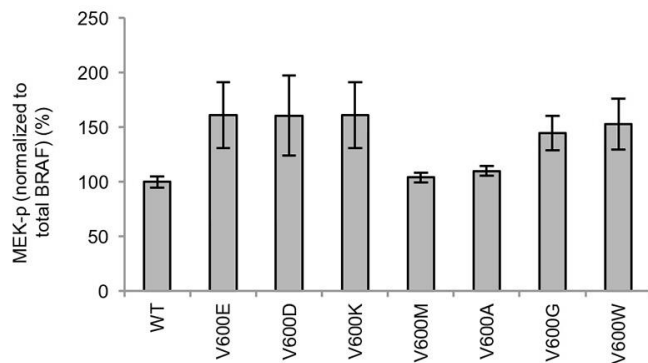
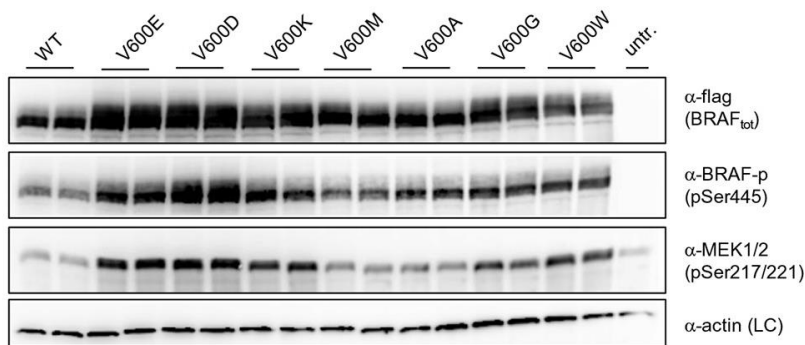
		Second Letter					
		U	C	A	G		
1st letter	U	UUU   Phe UUC   UUA   Leu UUG	UCU   Ser UCC   UCA   UCG	UAU   Tyr UAC   UAA   Stop UAG   Stop	UGU   Cys UGC   UGA   Stop UGG   Trp	U C A G	
	C	CUU   Leu CUC   CUA   CUG	CCU   Pro CCC   CCA   CCG	CAU   His CAC   CAA   Gln CAG	CGU   Arg CGC   CGA   CGG	U C A G	
	A	AUU   Ile AUC   AUA   AUG   Met	ACU   Thr ACC   ACA   ACG	AAU   Asn AAC   AAA   Lys AAG	AGU   Ser AGC   AGA   Arg AGG	U C A G	
	G	GUU   Val GUC   GUA   GUG	GCU   Ala GCC   GCA   GCG	GAU   Asp GAC   GAA   Glu GAG	GGU   Gly GGC   GGA   GGG	U C A G	

- 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

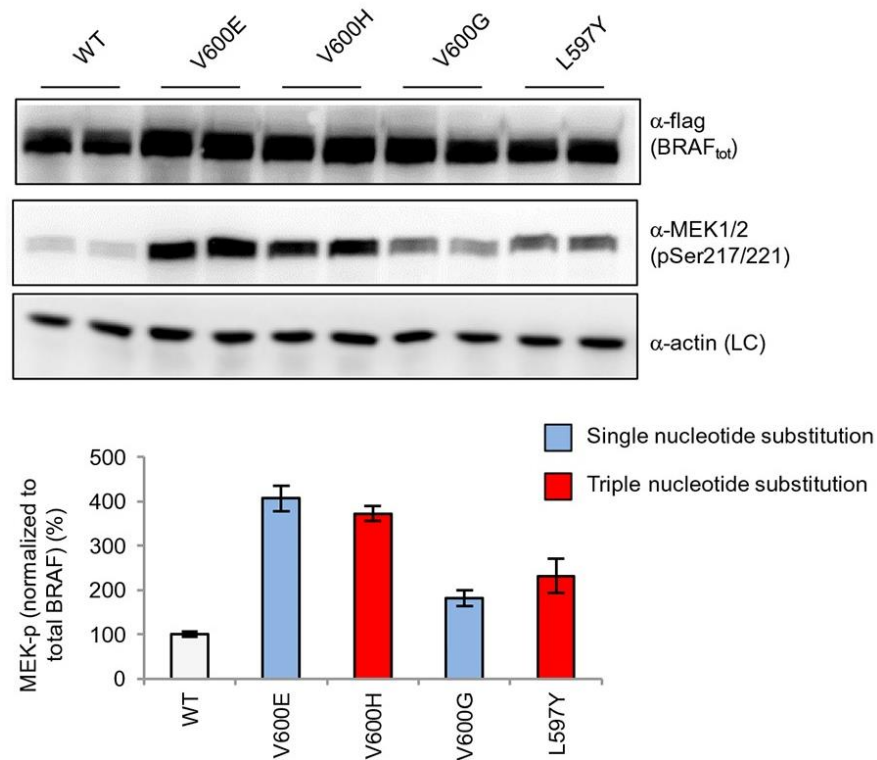
V600E: GAG

V600K: AAG  
V600R: AGG  
V600D: GAC/T

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



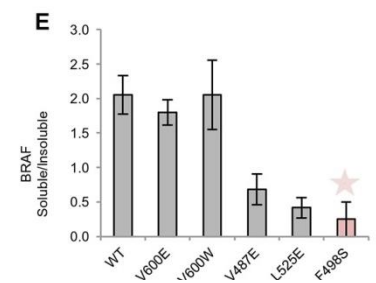
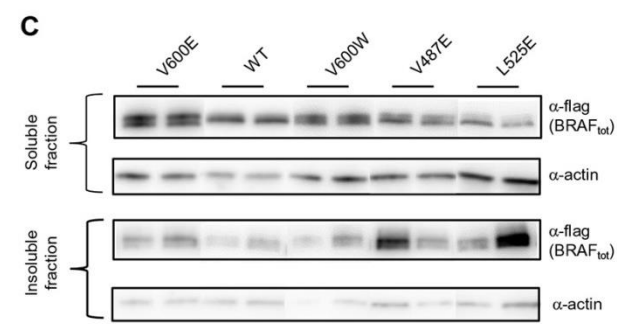
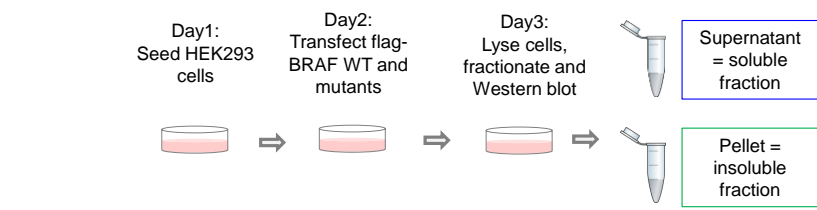
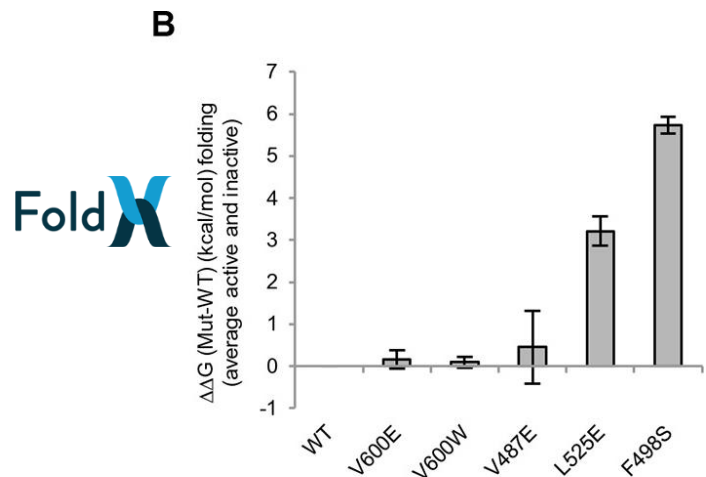
## Experimentally validate the effect of BRAF mutations by monitoring downstream MEK activation (HEK293 cells)



- V600E (requires 3 nucleotide substitutions) is as active as V600E, but NOT found in cancer

# 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)

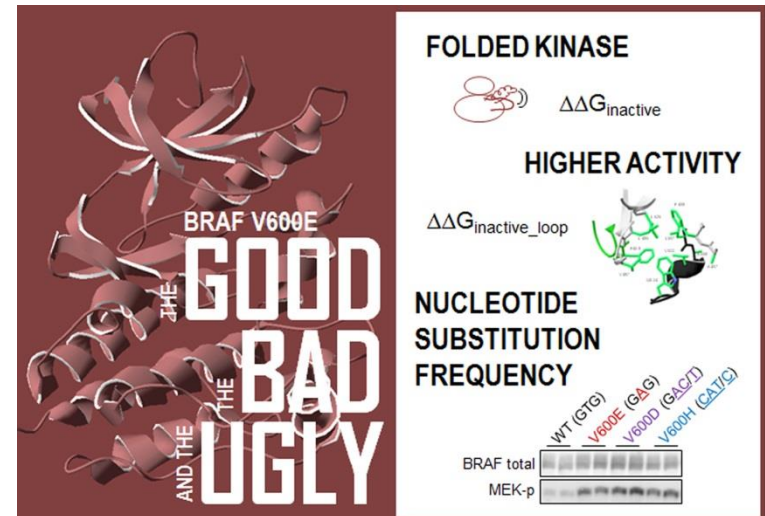
## Conclusions example 3: Why BRAF V600E?

- 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?

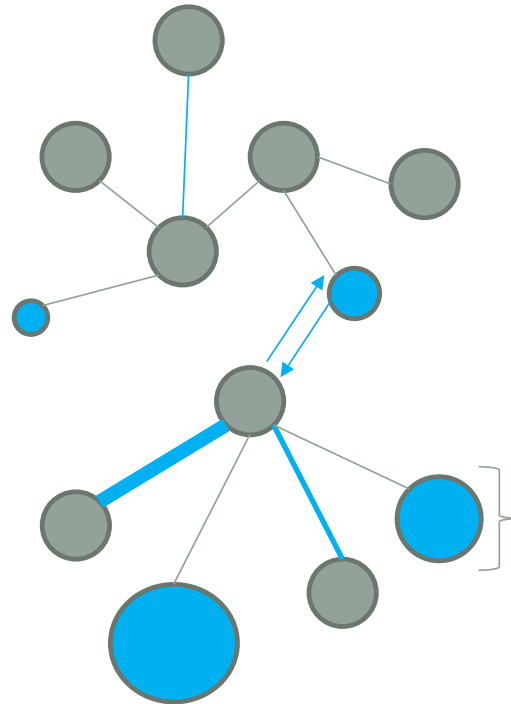
- V600E is the only single nucleotide substitution (Asp, Lys, and Arg, require two bases substitutions) that opens the AS through destabilization of autoinhibitory interactions, without significantly impairing the folding of the inactive or active kinase domain.

- 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.



Quantitative PPI networks

FoldX



RCSB PDB  
PROTEIN DATA BANK

Interactome3D

dsysmap

pubmed  
paxdb<sup>4</sup>

B10NUMB3R5  
THE DATABASE OF USEFUL BIOLOGICAL NUMBERS

# Protein abundances


← → ↻ pax-db.org

**paxdb<sup>4</sup>** PaxDb: Protein Abundance Database Downloads Help Archives ▾ About


All Organisms

Browse species


All ↗ all organisms (56)



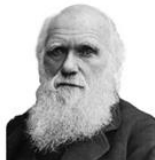
X. tropicalis




G. gallus



P. troglodytes



H. sapiens



C. fe

Data Overview

Species ▾	Predicted proteome size ▾	Datasets ▲	Proteins Covered ▾
<a href="#">Homo sapiens</a>	20457	170	98%
<a href="#">Mus musculus</a>	22668	75	90%
<a href="#">Arabidopsis thaliana</a>	27416	46	76%
<a href="#">Danio rerio</a>	26163	20	59%
<a href="#">Escherichia coli str. K-12 substr. MG1655</a>	4146	18	98%
<a href="#">Saccharomyces cerevisiae</a>	6692	17	96%
<a href="#">Caenorhabditis elegans</a>	20517	10	60%
<a href="#">Drosophila melanogaster</a>	13937	10	95%
<a href="#">Schizosaccharomyces pombe</a>	5144	8	90%
<a href="#">Zea mays</a>	92413	7	18%

# Affinities and kinetic constants



myBDB logout

## Search and Browse

### Target

Sequence

Name &

Ki IC50 Kd EC50

Rate constants

$\Delta G^*$   $\Delta H^*$   $-\Delta S^*$

pH (Enzymatic Assay)

pH (ITC)

Substrate or Competitor

Compound Mol. Wt.

Chemical Structure

Pathways

Source Organism

Number of Compounds

Monomer List in csv

Het List in SDF

### Compound

FDA Drugs

Important Compounds

Chemical Structure

Name

SMILES

Number of Data / Targets

### Special tools

3D Structure Series

Find My Compound's Targets

Find Compounds for My Targets

Do Virtual Screening

SCOP

### Citation

Author

Journal/Citation

Institution

PubMed

PubChem BioAssay

US Patent

## The Binding Database

Home Info Download About us Email us Contribute data Web Services

BindingDB is a public, web-accessible database of measured binding affinities, focusing chiefly on the interactions of protein considered to be drug-targets with small, drug-like molecules. BindingDB contains 1,207,821 binding data, for 6,265 protein targets and 529,618 small molecules.

There are 2907 protein-ligand crystal structures with BindingDB affinity measurements for proteins with 100% sequence identity, and 7392 crystal structures allowing proteins to 85% sequence identity.

### Simple Search

Article Titles, Authors, Assays, Compound Names, Target Names

Use ? for single-letter wild-card or \* for general wild-card.  
For example, "adeny\*" or "adeny?". 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 an example, go to [this article](#), and see the "Data for this Article, BindingDB" link on the right. (December 2015)

### Journal Curation by BindingDB

BindingDB continually curates a set of journals not covered by other public databases. As of January 2016, the status of our current curation effort is as follows:

- [ACS Chemical Biology](#) 2006-2015 (vol 1-10)
- [ACS BioChemistry](#) 1962-1970 (vol 1-9), 1991-2015 (vol 30-54)
- [Bioorganic Chemistry](#) 1971-2015 (vol 1-62)
- [BMC Chemical Biology](#) 2001-2012 (vol 1-12)
- [ChemBioChem](#) 2000-2015 (vol 1-16)
- [Chemical Biology & Drug Design](#) 2006-2015 (vol 67-86)
- [Chemistry & Biology](#) 1994-2014 (vol 1-20)
- [Journal of Biological Chemistry](#) 1988-2013 (vol 264-288)
- [Journal of Chemical Biology](#) 2008-2013 (vol 1-6)
- [Journal of Enzyme Inhibition and Medicinal Chemistry](#) 1997-2009 (vol 11-24)
- [Nature Chemical Biology](#) 2005-2014 (vol 1-10)
- [Medicinal Chemistry Research](#) 2004-2013 (vol 13-22)

### myBDB

Username  Password   
  Username is your registered email in BindingDB.

### Video Tutorials



### BindingDB News

**September 2015.** We to post the latest BindingDB user sun, or about October 5. would greatly appreciate your feedback and suggestions! Note, however, that you are always free to click the survey to the reg home-page.

**September 2015.** A compounds in Bindir have now been assi BindingDB Molecule such as BDBM5018 (The numeric comp also known, internal Monomer ID.)

**September 2015.** BindingDB should gi faster performance r we have upgraded tl server. Please let us immediately if you n any problems.

**July 2015.** Please tr new tool to map fror or more proteins of k sequence to known potential ligands: Fir Compounds for My Ta

**April 2015.** Bindin gl improved security. V use SSL to transmit passwords securely, forgotten passwords now handled with a link.

**March 2015.** The BindingDB results tz now provide links frc protein targets to an in Antibodypedia, an ligands to UniChem.

**February 2015.** Full has been replaced b Simple Search, with greatly improved dis



# General 'numbers' in biology

## B10NUMB3R5

THE DATABASE OF USEFUL BIOLOGICAL NUMBERS

[Home \ Search](#)
[Browse](#)
[Resources](#)
[Cell Biology by the Numbers](#)
[About Us](#)
[Login \ Submit](#)

### Key Numbers for Cell Biologists

#### Cell size

1. Bacteria (*E. coli*):  $\approx 0.7\text{-}1.4\ \mu\text{m}$  diameter,  $\approx 2\text{-}4\ \mu\text{m}$  length,  $\approx 0.5\text{-}5\ \mu\text{m}^3$  in volume;  $10^8\text{-}10^9$  cell/ml for culture with  $\text{OD}_{600} \approx 1$
2. Yeast (*S. cerevisiae*):  $\approx 3\text{-}6\ \mu\text{m}$  diameter,  $\approx 20\text{-}160\ \mu\text{m}^3$  in volume
3. Mammalian cell volume:  $100\text{-}10000\ \mu\text{m}^3$ ; HeLa:  $500\text{-}5000\ \mu\text{m}^3$  (adherent on slide  $\approx 15\text{-}30\ \mu\text{m}$  diameter)

#### Length Scales Inside Cells

4. Nucleus volume  $\approx 10\%$  of cell volume
5. Cell membrane thickness  $\approx 4\text{-}10\ \text{nm}$
6. "Average" protein diameter  $\approx 3\text{-}6\ \text{nm}$
7. Base pair:  $2\ \text{nm}$  (D) x  $0.34\ \text{nm}$  (H)
8. Water molecule diameter  $\approx 0.3\ \text{nm}$

#### Division, Replication, Transcription, Translation & Degradation Rates

- at  $37^\circ\text{C}$  with a temperature dependence  $Q_{10}$  of  $\approx 2\text{-}3$
9. Cell cycle time (exponential growth in rich media): *E. coli*  $\approx 20\text{-}40\ \text{min}$ ; yeast  $70\text{-}140\ \text{min}$ ; human cell line (HeLa):  $15\text{-}30\ \text{hours}$
  10. Rate of replication by DNA polymerase *E. coli*  $\approx 200\text{-}1000\ \text{bases/s}$ ; human  $\approx 40\ \text{bases/s}$ . Transcription by RNA polymerase  $10\text{-}100\ \text{bases/s}$
  11. Translation rate by ribosome  $10\text{-}20\ \text{aa/s}$
  12. Degradation rates (proliferating cells): mRNA half life  $<$  cell cycle time; protein half life  $\approx$  cell cycle time

#### Concentration

13. Concentration of  $1\ \text{nM}$  in: *E. coli* is  $\approx 1\ \text{molecule/cell}$ ; HeLa  $\approx 1,000\ \text{molecules/cell}$
14. Characteristic concentration for a signaling protein  $\approx 10\ \text{nM}\text{-}1\ \mu\text{M}$
15. Water content:  $\approx 70\%$  by mass; General elemental composition (dry weight) of *E. coli*:  $\approx \text{C}_2\text{H}_7\text{O}_2\text{N}_1$ ; Yeast  $\approx \text{C}_6\text{H}_{10}\text{O}_3\text{N}_1$
16. Composition of *E. coli* (dry weight):  $\approx 55\%$  protein,  $20\%$  RNA,  $10\%$  lipids,  $15\%$  others
17. Protein conc.  $\approx 100\ \text{mg/ml} = 3\ \text{mM}$ .  $10^6\text{-}10^7$  per *E. coli* (depending on growth rate); Total metabolites (MW  $< 1\ \text{kD}$ )  $\approx 300\ \text{mM}$

#### Energetics

18. Membrane potential  $\approx 70\text{-}200\ \text{mV} \rightarrow 2\text{-}6\ \text{k}_B\text{T}$  per electron ( $\text{k}_B\text{T}$  = thermal energy)
19. Free energy ( $\Delta G$ ) of ATP hydrolysis under physiological conditions  $\approx 40\text{-}60\ \text{kJ/mole} \rightarrow \approx 20\ \text{k}_B\text{T/molecule ATP}$ ; ATP molecules required to make an *E. coli* cell  $\approx 10\text{-}50 \times 10^9$
20.  $\Delta G^\circ$  resulting in order of magnitude ratio between products and reactants concentrations:  $\approx 6\ \text{kJ/mol} = 60\ \text{meV} \approx 2\ \text{k}_B\text{T}$

Useful biological numbers extracted from the literature. Numbers and ranges should only serve as "rule of thumb" values. References are in the online annotated version at the BioNumbers website. Consult website and original references to learn about the details of the system under study including growth conditions, method of measurement, etc.

#### Diffusion and Catalysis Rate

21. Diffusion coefficient for an "average" protein: in cytoplasm  $D = 5\text{-}15\ \mu\text{m}^2/\text{s} \rightarrow \approx 10\ \text{millisec}$  to traverse an *E. coli*  $\rightarrow \approx 10\ \text{s}$  to traverse a mammalian (HeLa) cell; small metabolite in water  $D = 500\ \mu\text{m}^2/\text{s}$
22. Diffusion limited on-rate for characteristic protein  $\approx 10^6\text{-}10^9\ \text{s}^{-1}\text{M}^{-1} \rightarrow$  for a protein substrate of concentration  $\approx 1\ \mu\text{M}$  the diffusion limited on-rate is  $\approx 100\text{-}1000\ \text{s}^{-1}$  thus limiting the catalytic rate  $k_{\text{cat}}$

#### Genome sizes & Error Rates

23. Genome size: *E. coli*  $\approx 5\ \text{Mbp}$ ; *S. cerevisiae* (yeast)  $\approx 12\ \text{Mbp}$ ; *C. elegans* (nematode)  $\approx 100\ \text{Mbp}$ ; *D. melanogaster* (fruit fly)  $\approx 120\ \text{Mbp}$ ; *A. thaliana* (arabidopsis)  $\approx 120\ \text{Mbp}$ ; *M. musculus* (mouse)  $\approx 2.5\ \text{Gbp}$ ; *H. sapiens* (human)  $\approx 2.9\ \text{Gbp}$ ; *T. aestivum* (wheat)  $\approx 16\ \text{Gbp}$
24. Number of protein-coding genes: *E. coli*  $\approx 4,000$ ; *S. cerevisiae*  $\approx 6,000$ ; *C. elegans*, *A. thaliana*, *M. musculus*, *H. sapiens*  $\approx 20,000$
25. Mutation rate in DNA replication  $\approx 10^{-8}\text{-}10^{-10}$  per bp
26. Misincorporation rate: transcription  $\approx 10^{-4}$  per nucleotide; translation  $\approx 10^{-3}\text{-}10^{-4}$  per amino-acid

Click on a number to see full description and reference [www.BioNumbers.org](http://www.BioNumbers.org)

# Protein structures

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB Login

**RCSB PDB**  
PROTEIN DATA BANK

An Information Portal to  
118087 Biological  
Macromolecular Structures

Search by PDB ID, author, macromolecule, sequence, or ligand **Go**

Advanced Search | Browse by Annotations

PDB-101 WORLDWIDE PDB PROTEIN DATA BANK EMDatabank NUCLEIC ACID DATABASE StructuralBiology Knowledgebase

f t v a

- Welcome
- Deposit
- Search
- Visualize
- Analyze
- Download
- Learn

### A Structural View of Biology

This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

#### Zika Virus Structure

### April Molecule of the Month

Lead Poisoning

# 3D structures of protein interactions



Home Download References Statistics News Tutorials Help About

Interactome3D is a web service for the structural annotation of protein-protein interaction networks. Submit your interactions and the server will find all the available structural data for both the single interactors and the interactions themselves. Additionally you can also visualize and download structural information for interactions involving a set of proteins or interactomes for one of the precalculated organisms.

If you have any doubts read our section of [Frequently Asked Questions](#).

The current version of Interactome3D is **2015\_12** [Release notes](#)

### Submit your interactions

Enter a name for your dataset:  [Example](#)

Enter a list of interactions (max. 10000). Every interaction has to be entered in a separate line, as a pair of space-separated Uniprot ACs (\*): [?](#)

For example...  
A0A5B9 A0A5B9  
A0A5B9 P01848  
A0AQH0 O61443

...or upload your interactions from a file: [?](#)

Email (\*\*):

 **Tutorial**  
Learn how to use Interactome3D

### Query interactions with proteins

Enter a list of Uniprot ACs (\*) or gene names:  [Example](#)

Only show the proteins in the list [?](#)

### Browse for organism

Select one of the pre-calculated organisms:

- > Arabidopsis thaliana
- > Bacillus subtilis
- > Bos taurus
- > Caenorhabditis elegans
- > Campylobacter jejuni
- > Mus musculus
- > Mycobacterium tuberculosis
- > Mycoplasma pneumoniae
- > Plasmodium falciparum
- > Rattus norvegicus
- > Saccharomyces cerevisiae
- > Schizosaccharomyces pombe

# 3D structures of protein interactions/ mapping of disease mutations



dsysmap

[Home](#) [Browse](#) [Download](#) [Stats](#) [Tutorial](#) [Help](#) [About](#)

---

**dSysMap** (Mapping of Human disease-related mutations at the systemic level) displays Human disease-related mutations on the structural interactome. Mapping of mutations on protein structures and on interaction interfaces allows you to visualize the region of the interactome that they affect and helps in rationalizing their mechanism of action.

The current version of **dSysMap** is **2015\_05**

Is this your first time with dSysMap? [Take a 5 minutes Tutorial!](#)

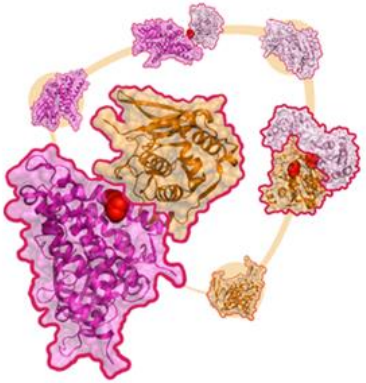
**Tutorial :: Learn how to use dSysMap**

**Browse diseases**

Select a disease from the following list.  
Example: [Loeys-Dietz syndrome](#)

Type here the name of a disease or browse the list...

- + Bacterial infection or mycosis
- + Blood disease
- + Cancer
- + Cardiovascular disease
- + Congenital abnormality
- + Connective tissue disease
- + Digestive system disease
- + Ear-nose-throat disease
- + Endocrine system disease
- + Eye disease
- + Fetal disease
- + Genetic disease
- + Immune system disease
- + Infant-newborn disease



**Query with a list of proteins**

Enter a list of proteins (Uniprot AC or gene name) Example

For example...

ETFA, ETFB, ACADM, ACADS, ACADVL, SOCS3, IRF7, GPHN, RPSA

**Submit**

**Submit your mutations**


Enter a list of mutations (which format?) Example

For example...


APC: p.Ala1582Lys, p.Thr506Trp  
AXIN1: p.Phe119Ala, p.Gln190Arg  
DLG1

**Submit**

# Protein design



[PRODUCTS ▾](#)
[LICENSING AND SERVICES ▾](#)
[SUPPORT ▾](#)
[DOWNLOAD ▾](#)
[ABOUT ▾](#)



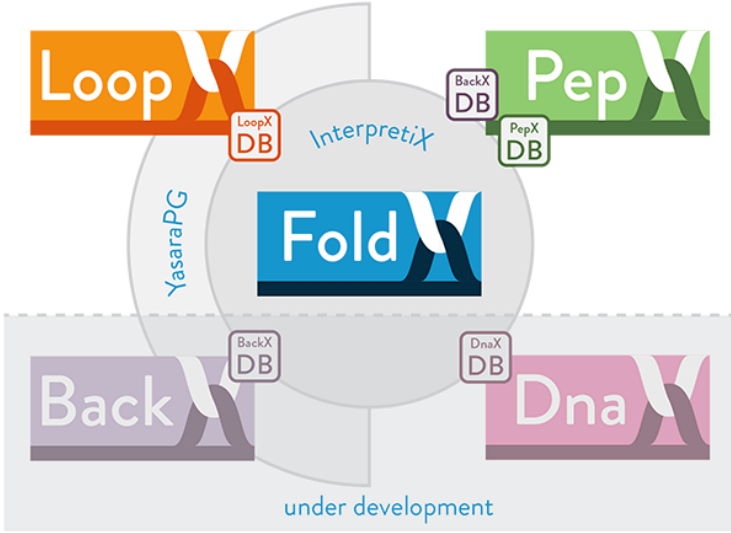
[LOGIN](#)   [REGISTER](#)

---

[Products](#) / [Products](#)

## PRODUCTS

### Technology Overview



under development

### The FoldX Suite

The FoldX Suite builds on the strong fundament 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.

### FOLDX PAPERS

McKeone R, Wikstrom M, Kiel C, Rakoczy EP. "Assessing the correlation between mutant rhodopsin stability and the severity of retinitis pigmentosa." *Mol. Vis.* 2014;20:183-99. 2014

Kiel C, Serrano L. "Structure-energy-based predictions and network modelling of RASopathy and cancer missense mutations." *Mol. Syst. Biol.* 2014;10:727. 2014

De Baets G, Van Durme J, Reumers J, et al. "SNPEffect 4.0: on-line prediction of molecular and structural effects of protein-coding variants." *Nucleic Acids Res.* 2012;40(Database issue):D935-9. 2012

Simões-Correia J, Figueiredo J, Lopes R, et al. "E-cadherin destabilization accounts for the pathogenicity of missense mutations in hereditary diffuse gastric cancer." *PLoS ONE*. 2012;7(3):e33783. 2012

Kimberley FC, van der Sloot AM, Guadagnoli M, et al. "The design and characterization of receptor-selective APRIL variants." *J. Biol. Chem.* 2012;287(44):37434-46. 2012

Van Durme J, Delgado J, Stricher F, Serrano L, Schymkowitz J, Rousseau F. "A graphical interface for the FoldX forcefield." *Bioinformatics*. 2011;27(12):1711-2.

## 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