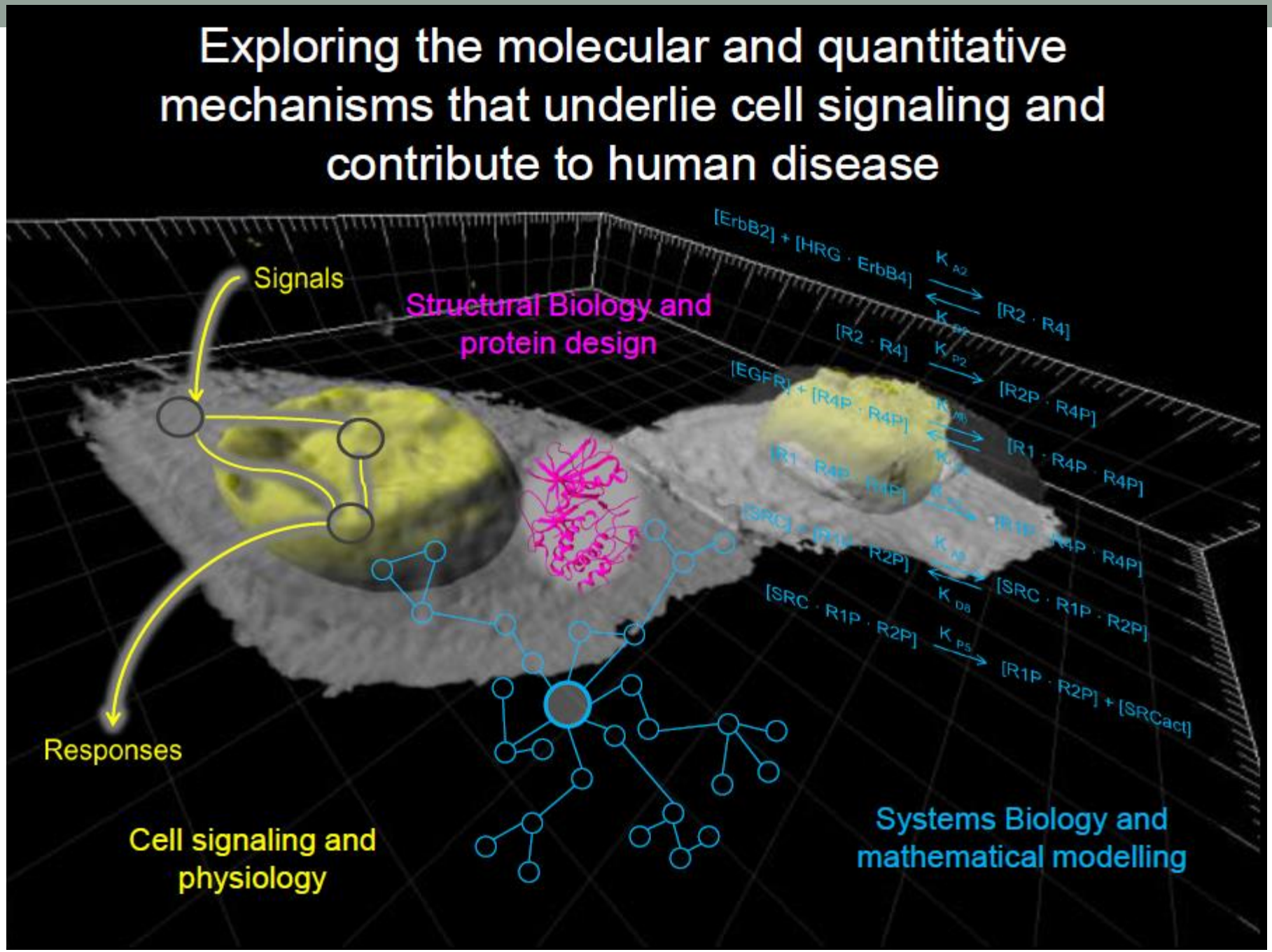
Integrative and quantitative analysis of disease mutations in protein interaction networks and implications for personalized medicine

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
Signals
Structural Biology and protein design
Responses
Cell signaling and physiology
Systems Biology and mathematical modelling
[ErbB2] + [HRG · ErbB4]
[R2 · R4]
[R2 · R4]
[R2P · R4P]
[EGFR] + [R4P · R4P]
[R1 · R4P · R4P]
[SRC · R1P · R2P]
[SRC · R1P · R2P]
[R1P · R2P] + [SRCact]

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

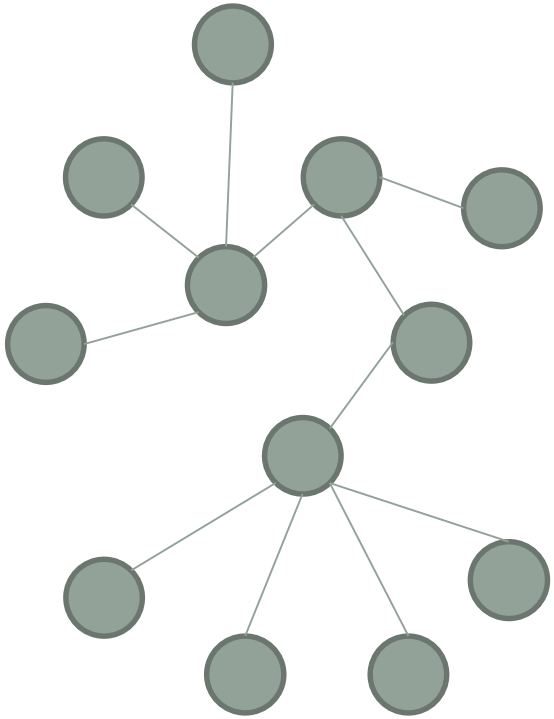
1. RASopathy vs cancer mutations: a matter of quantity
2. Rhodopsin stability and disease onset
3. BRAF mutation frequency: prediction of oncogenic drivers

VI. Summary tools & websites

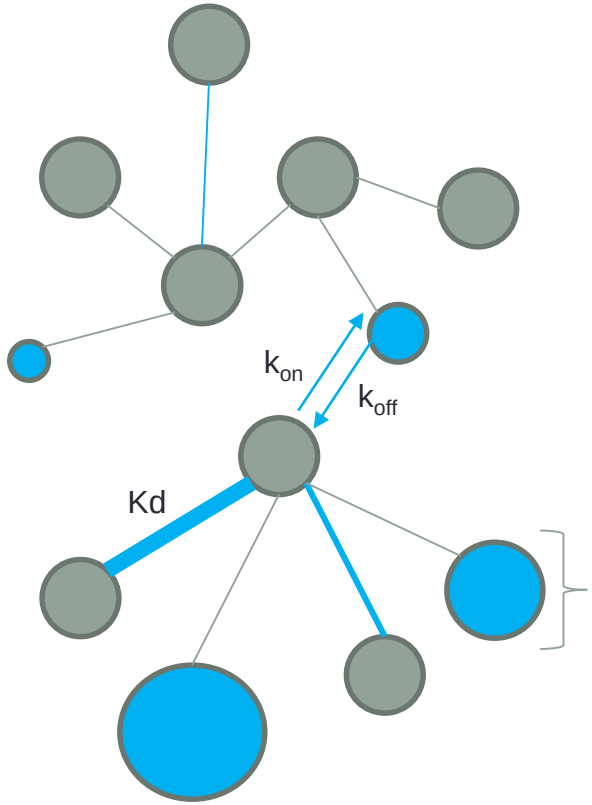
VII. Wrap up/ discussion/ conclusions

Quantitative information in protein-protein interaction (PPI) networks

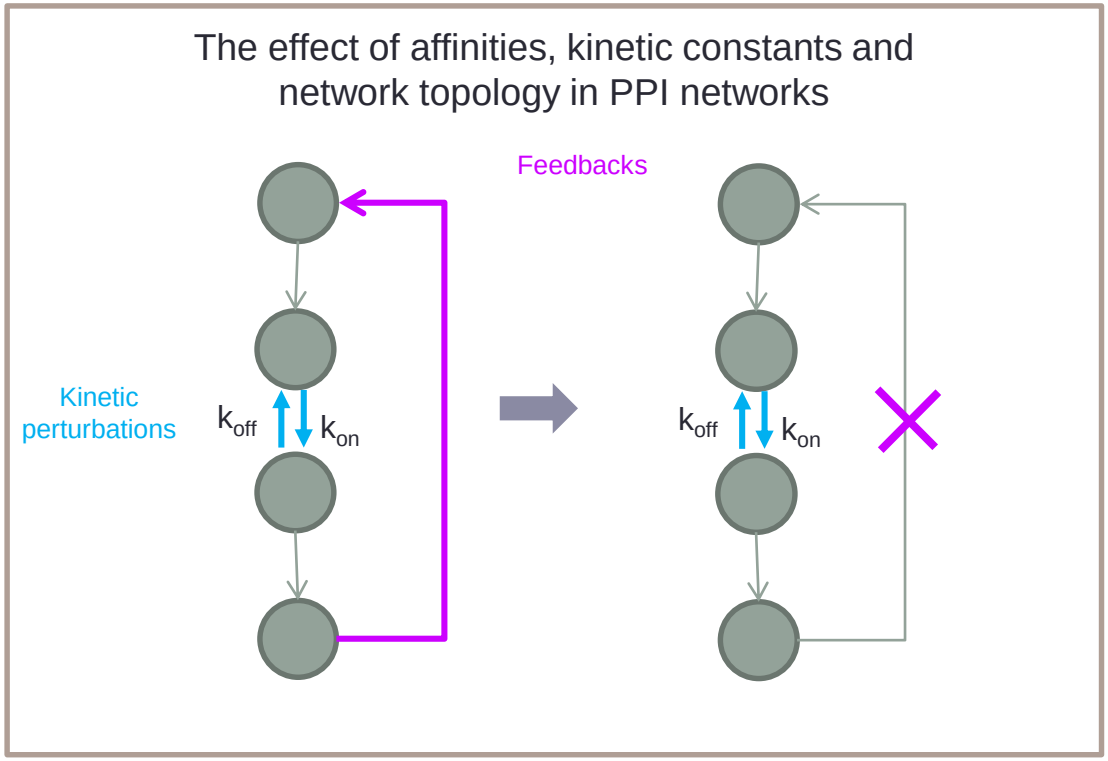
Qualitative PPI networks



Quantitative PPI networks

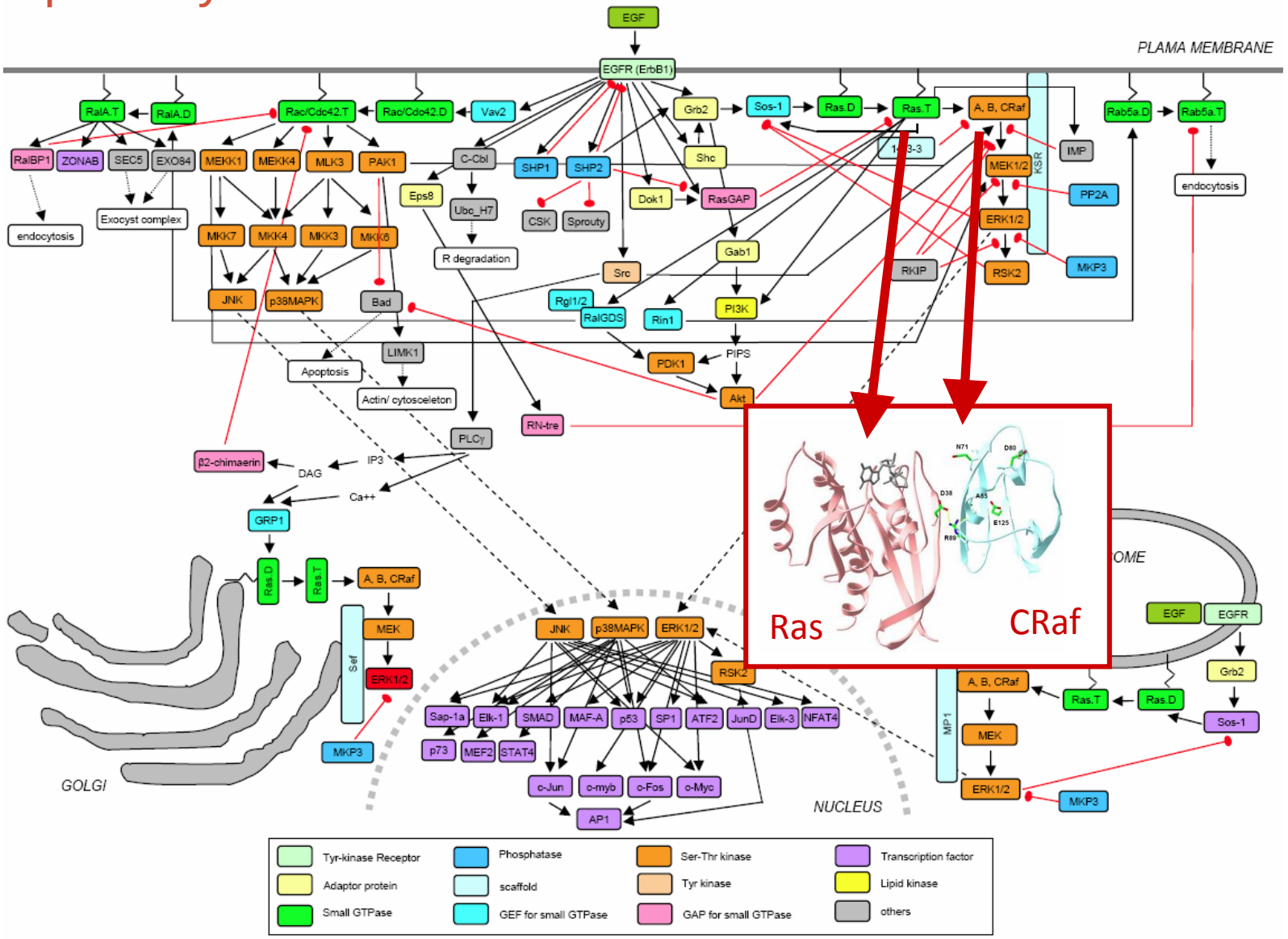


Considering protein abundances and affinities/ kinetic constants



Kiel & Serrano, Science Signal, 2009

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



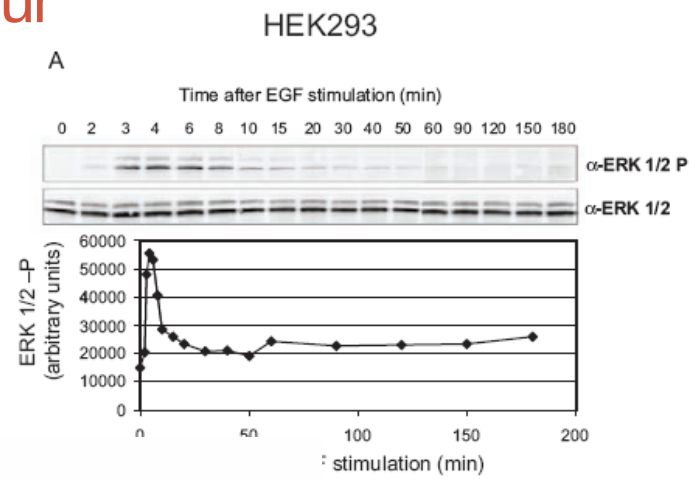
Different network 'wiring' /feedbacks causes the different behaviour

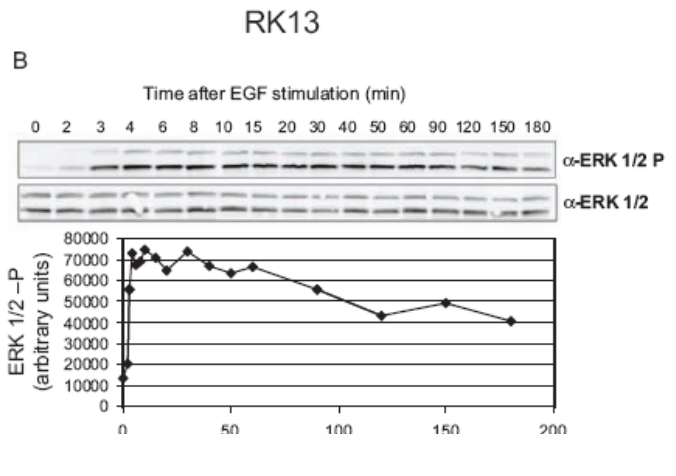
HEK293 cells

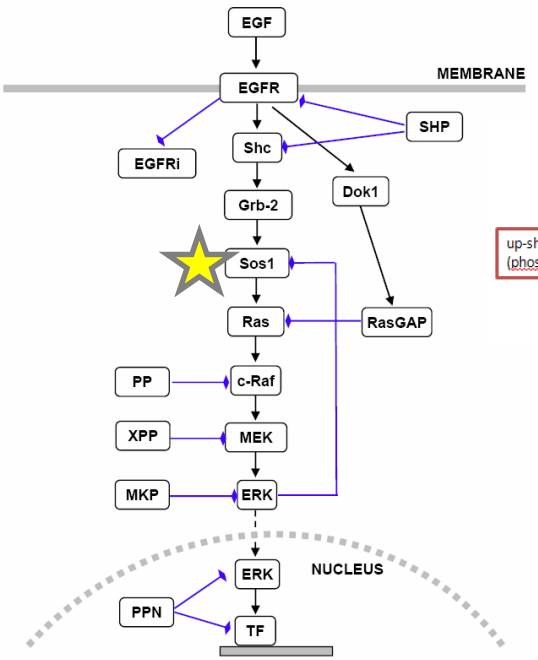
Transient response

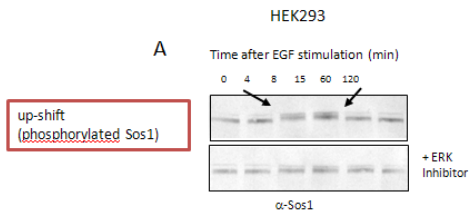
RK13 cells

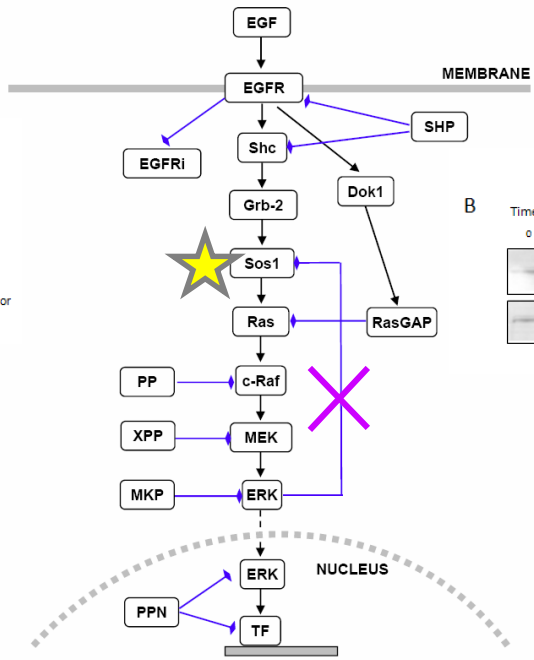
Sustained response

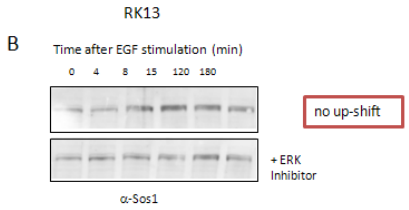






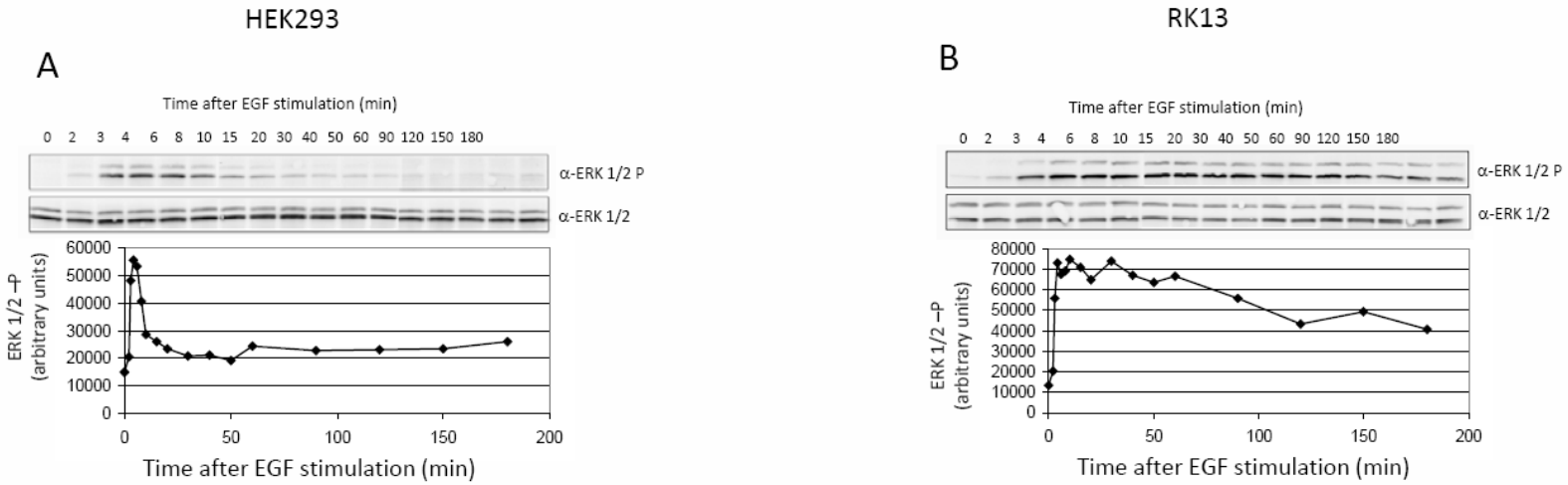


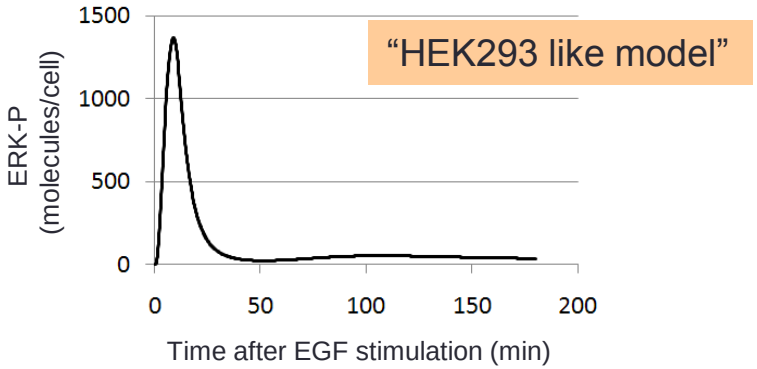




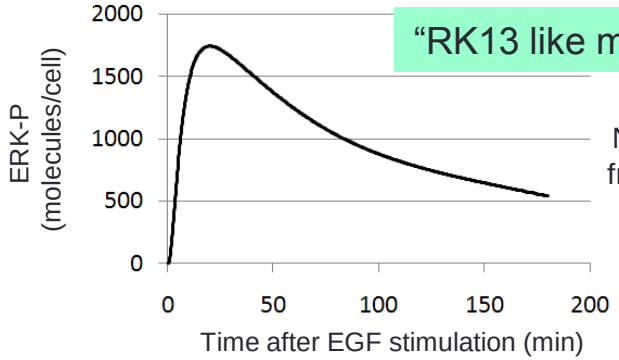
Kiel & Serrano, Sci Signal, 2009

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





"HEK293 like model"



"RK13 like model"

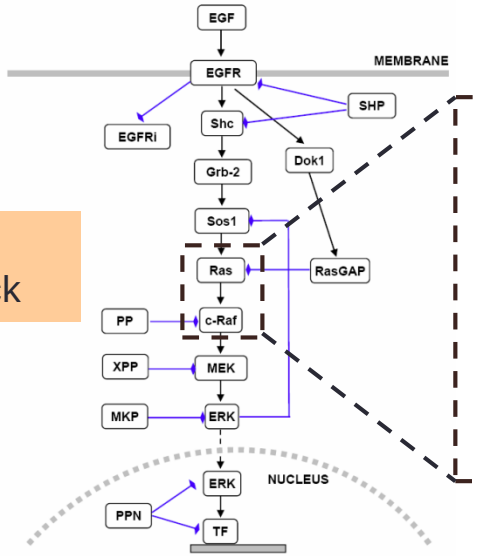
No negative feedback from ERK-P to Sos1 in the RK13-like model

- Good agreement of experiment and model predictions

Kiel & Serrano, Sci Signal, 2009

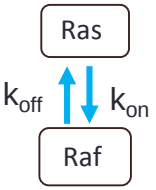
Model predictions: different cell type-specific wiring results in different responses to mutations with affinity perturbations

Strong feedback

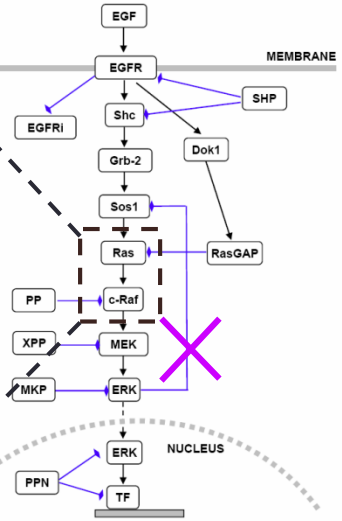


Subtle affinity changes

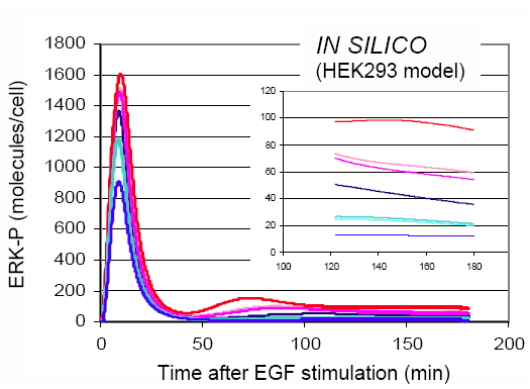
Kinetic perturbations



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

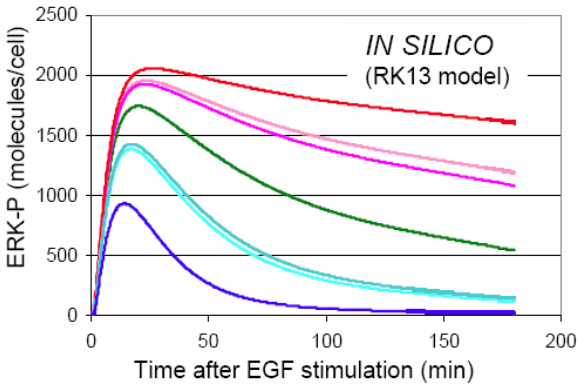


Weak feedback



No significant changes

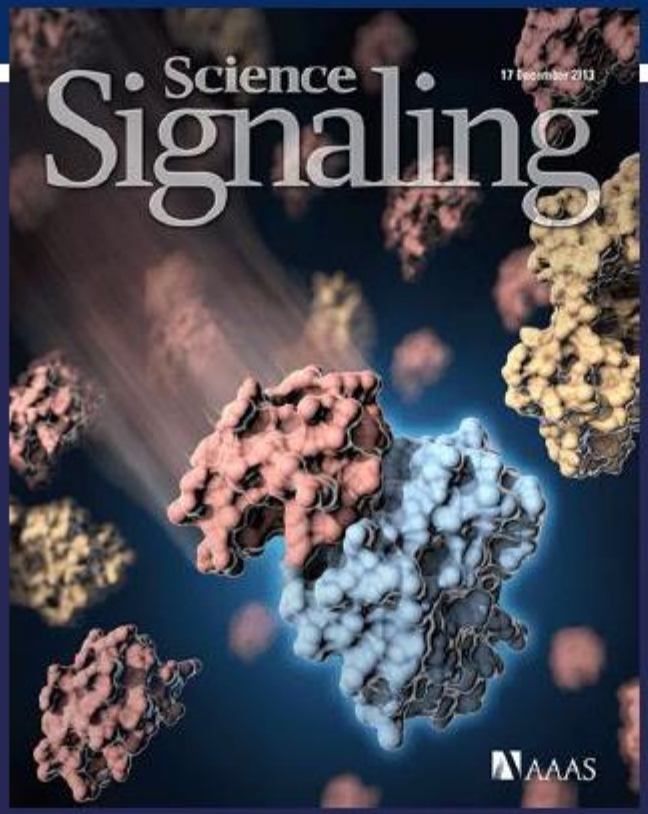
- Mutations can have different cell type (**patient!**)-specific effects



Significant differences

Kiel & Serrano, Sci Signal, 2009

Kiel et al, 2013



Interaction competition

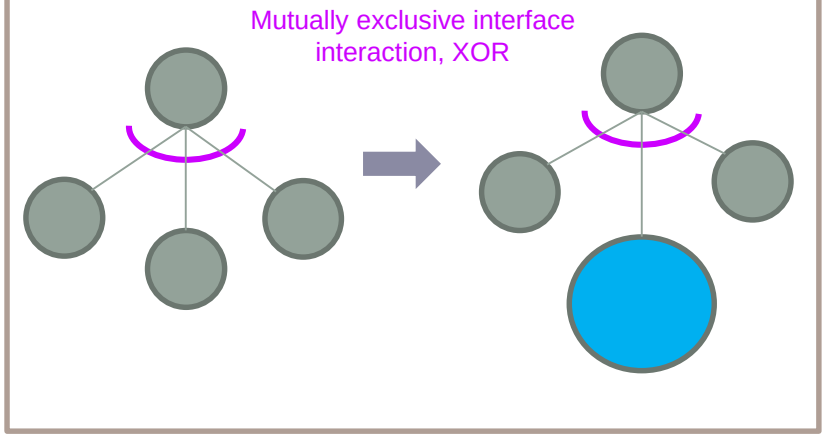
A means to redirect signaling flows?

CELL BIOLOGY

2013: Signaling Breakthroughs of the Year

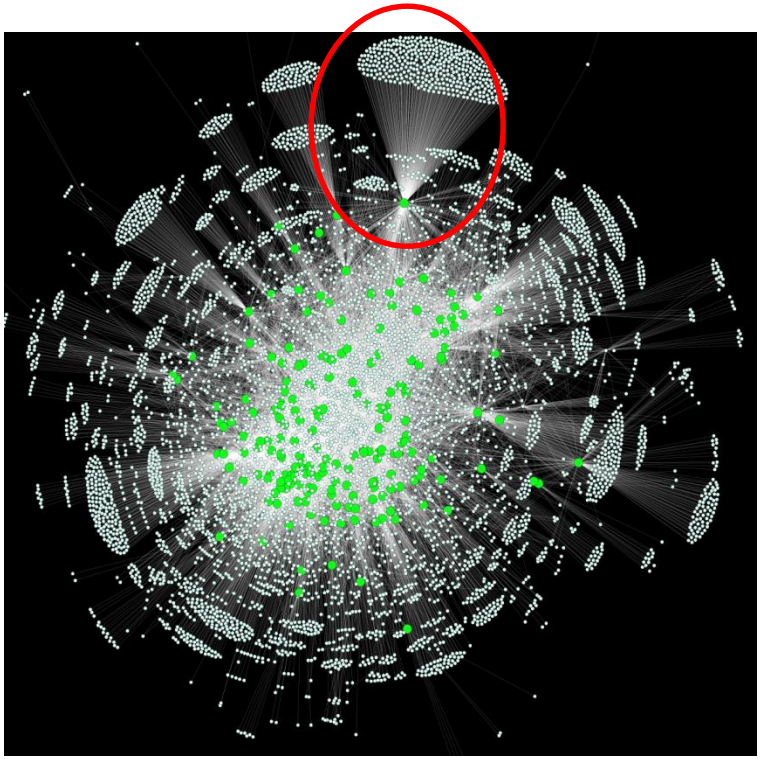
Another mechanism by which hubs decode different inputs is exemplified in a study by Kiel *et al.* (*7*). The authors showed that competition among proteins binding to a hub can govern how cells produce different responses to the activation of the same receptor. For proteins with a similar affinity for a common hub that bind in a mutually exclusive fashion, the relative abundance of these

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

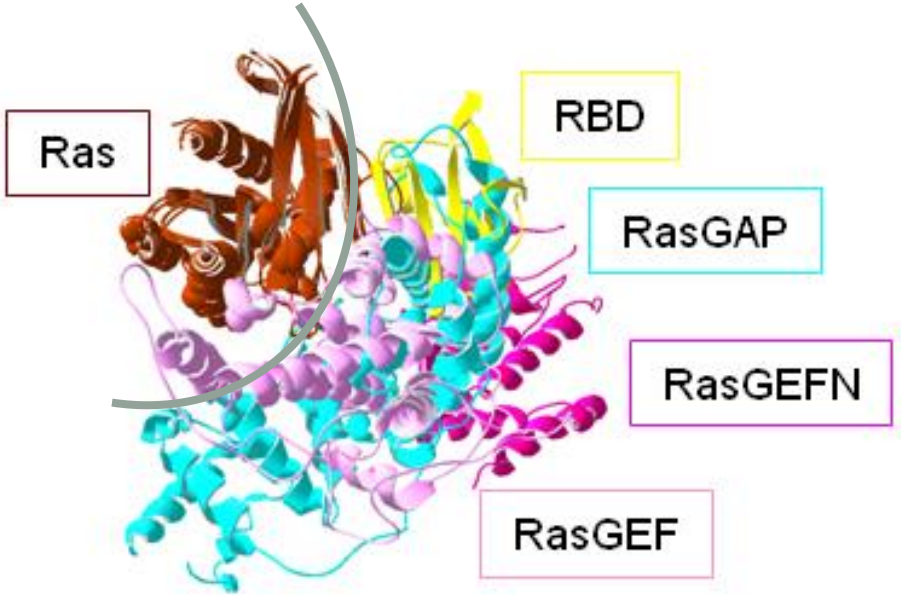


How could interaction competition and protein concentration affect downstream signaling?

Signaling complexes: > 300 partners for one protein??

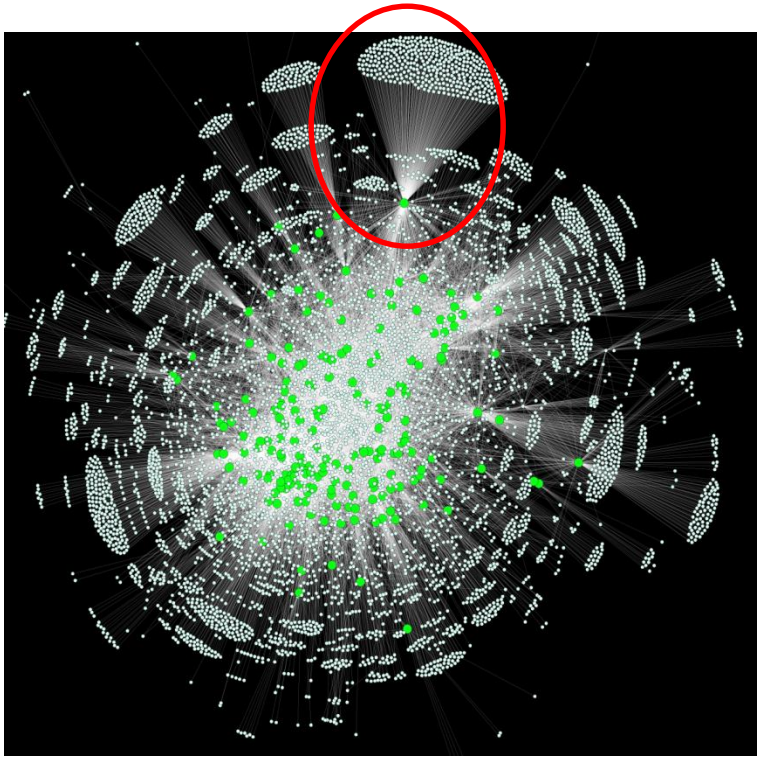


Some proteins will use similar binding surfaces for interaction with other molecules: 'mutually exclusive interactions'/ 'XOR'

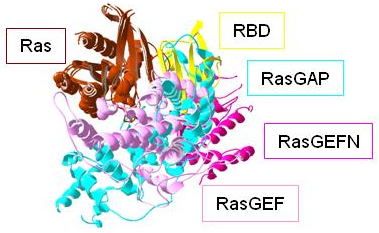


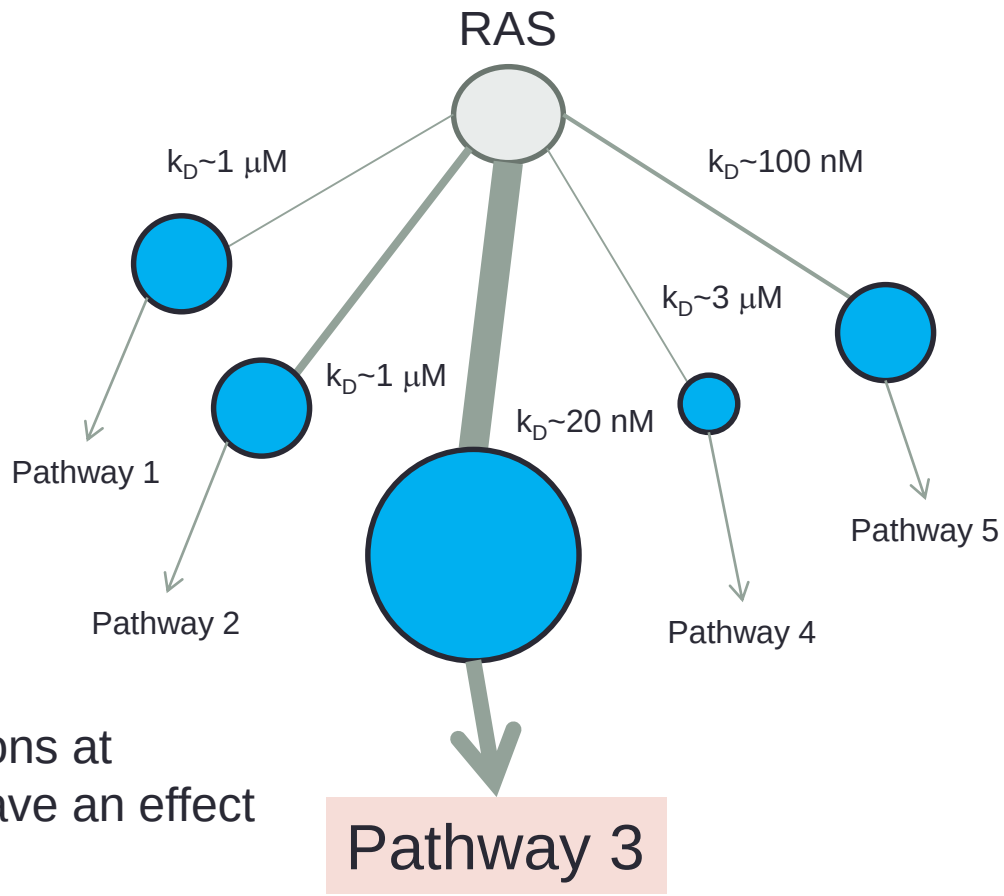
How could interaction competition and protein concentration affect downstream signaling?

Signaling complexes: > 300 partners for one protein??



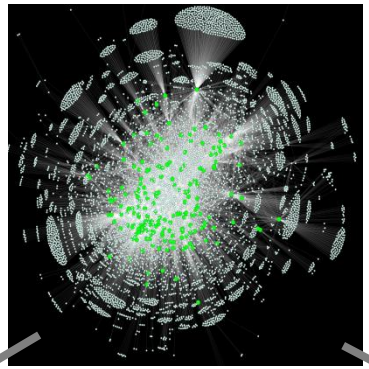
In a simple world: concentration and k_D will determine the signaling output

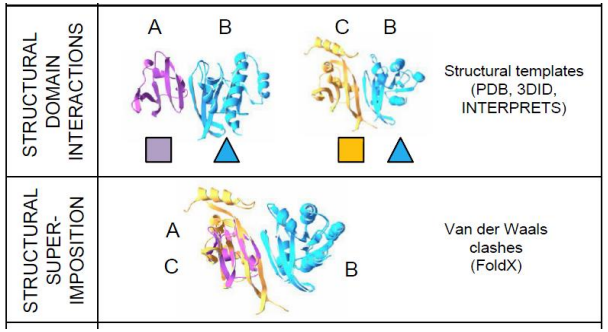




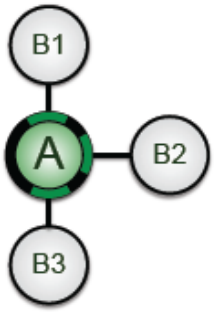
Changes in concentration (ie mutations at promoters, enhancers etc..) could have an effect in signalling

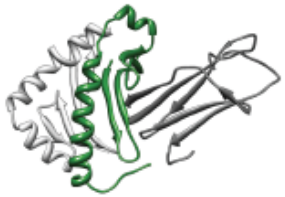
A bioinformatics tool to distinguish mutually exclusive from compatible interactions in large-scale PPI





Compatible ('AND')

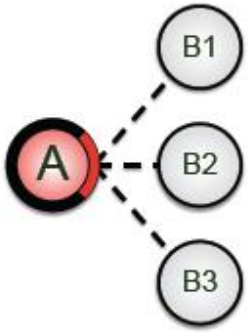


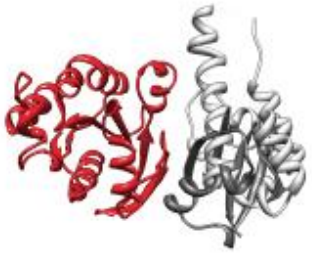


SAPIN (structural analysis of protein interaction networks) webserver

http://sapin.crg.es/

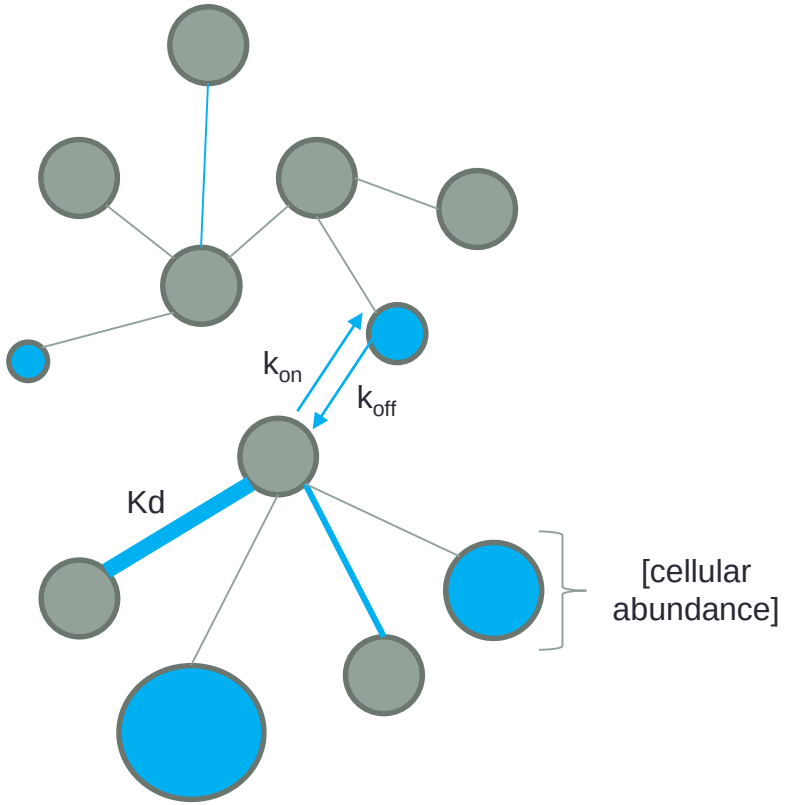
Exclusive ('XOR')





Yang et al, Bioinformatics, 2012

Experimental methods to quantify protein abundances, affinities, and kinetic constants



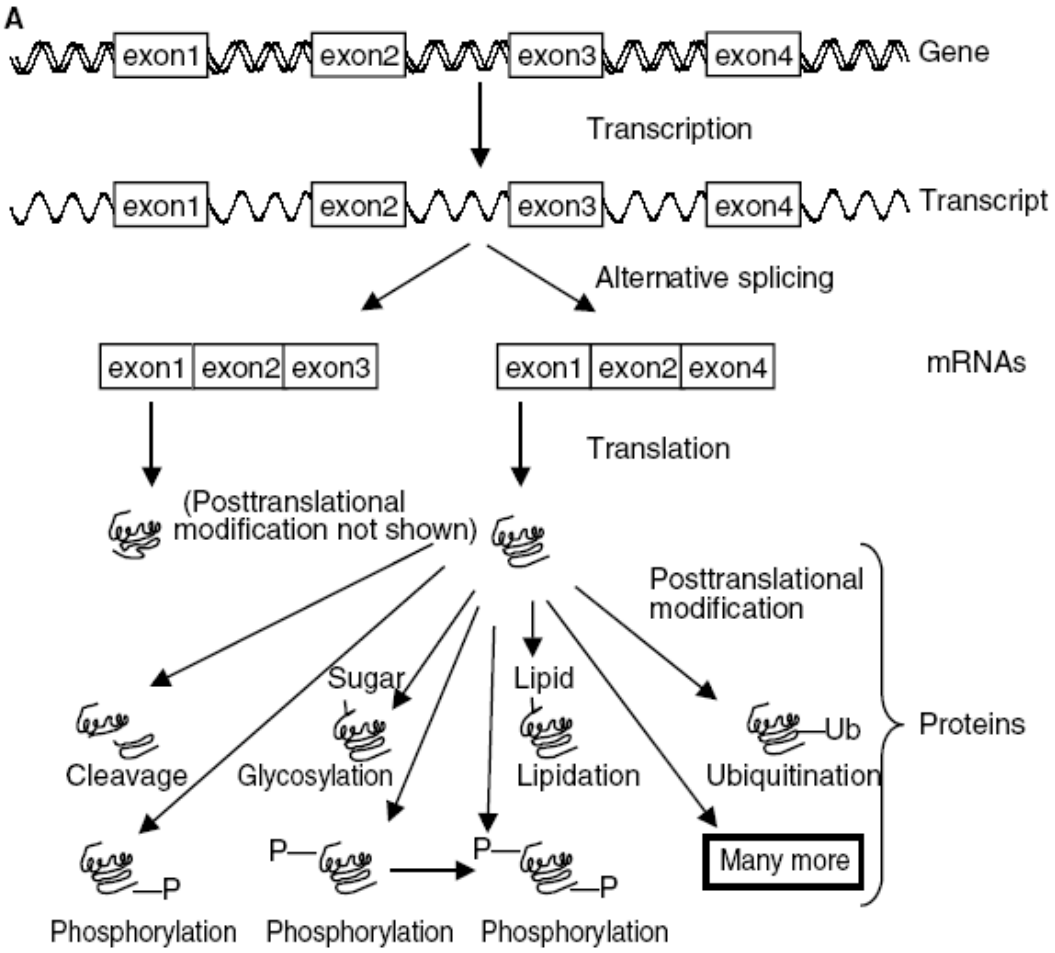
Why proteomics in times of deep RNA sequencing?

- mRNA does not translate1:1 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



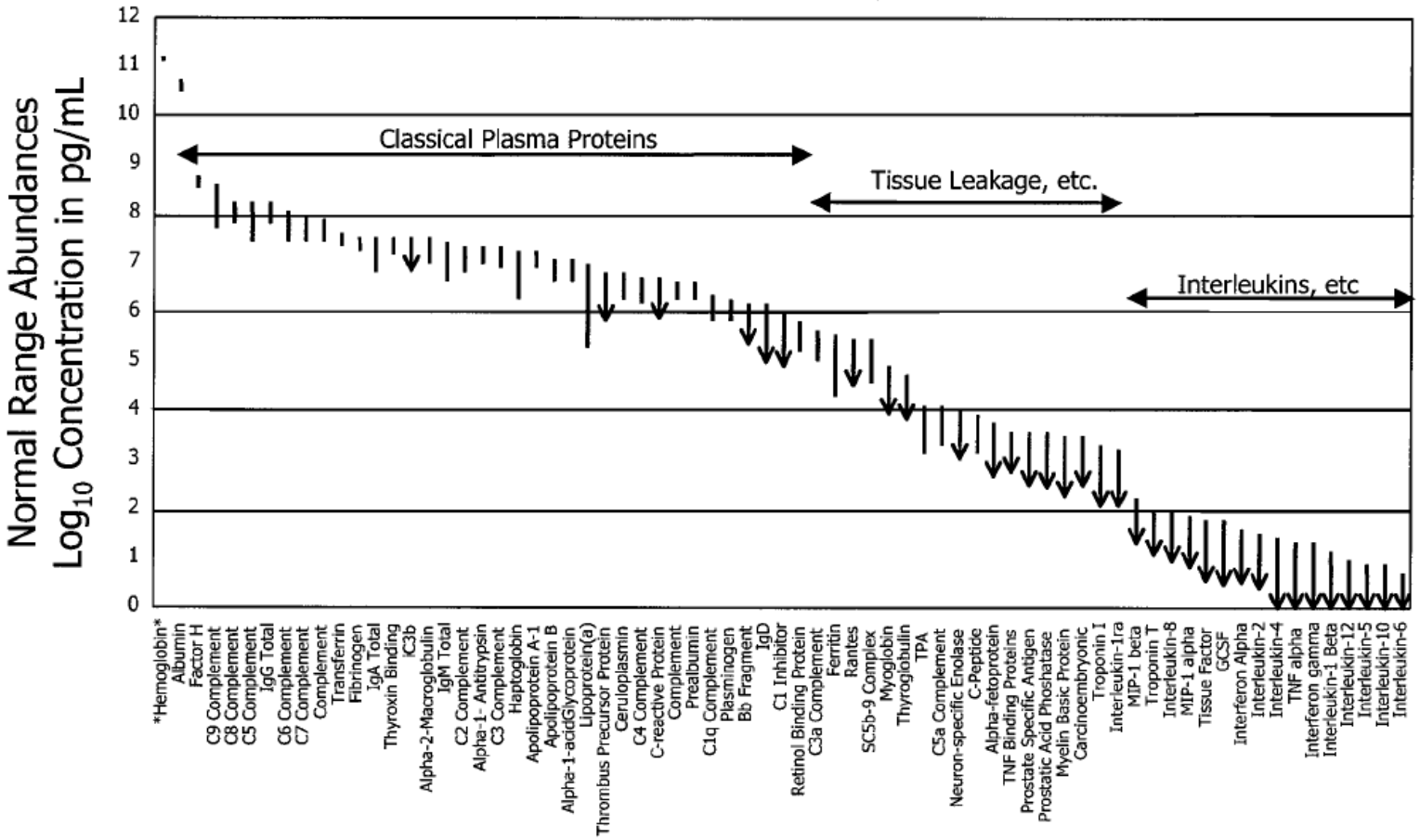
30,000 coding genes per cell

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

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

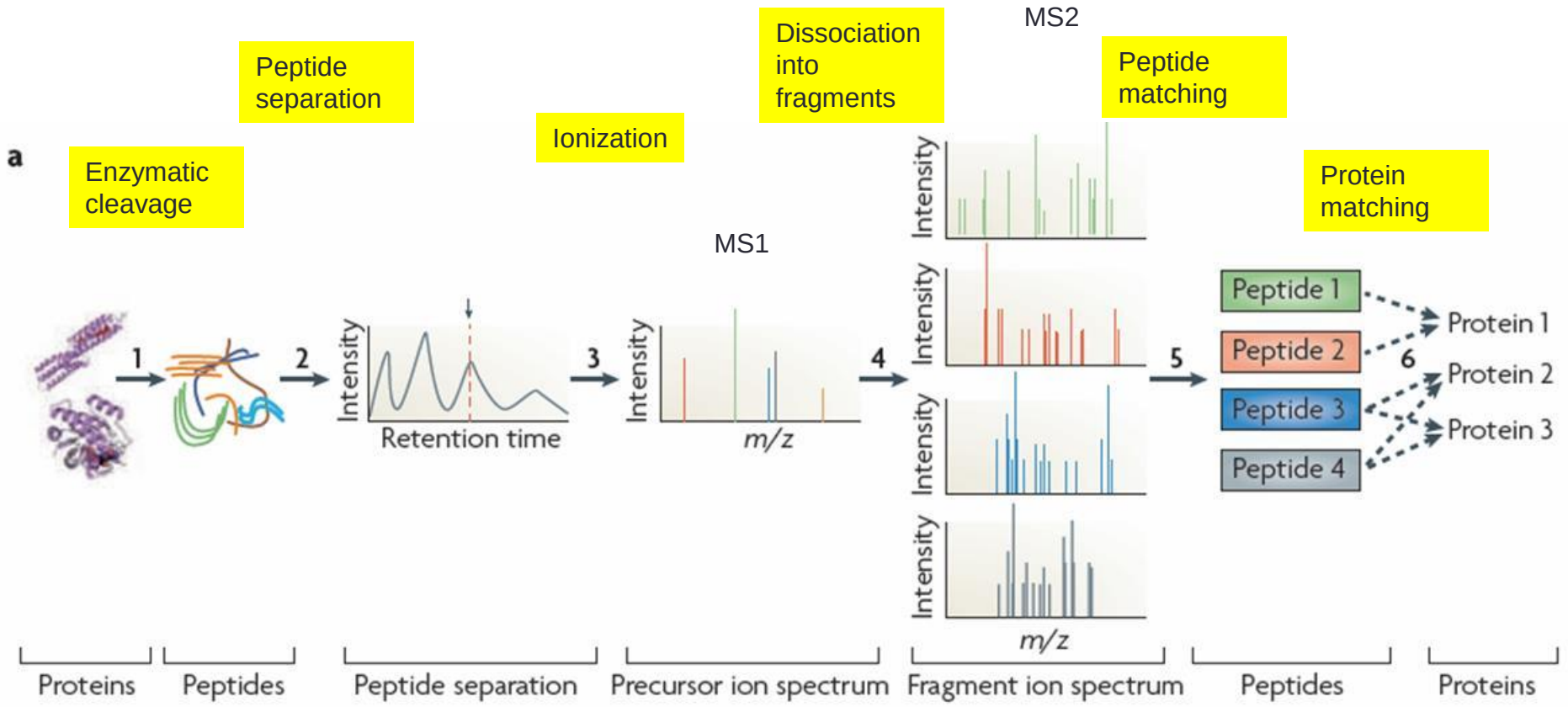
Peng and Gygi, JMS, 2001

High dynamic range of the proteome



Anderson and Anderson, MCP, 2002

Protein identification by mass spectrometry



- ❑ Address problem of cellular <u>complexity</u> by fractionation, e.g. liquid chtromatography
- ❑ Address problem of cellular <u>dynamic range</u> by better and better (and better…) mass spectrometers…

Ahrens et al, 2010

Human deep proteome mapping

Molecular Systems Biology 7; Article number 549; doi:10.1038/msb.2011.82
Citation: *Molecular Systems Biology* 7: 549

www.molecularsystemsbiology.com

REPORT

The quantitative proteome of a human cell line

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

[1] European Molecular Biology Laboratory, Heidelberg, Germany, [2] Biozentrum, University of Basel, Basel, Switzerland, [3] Department of Immunotechnology, BMC, Lund, Sweden, [4] Biognosys AG, Schlieren, Switzerland, [5] Department of Computer Science, ETH Zurich, Zurich, Switzerland, [6] Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland; [7] Competence Center for Systems Physiology and Metabolic Diseases, Zurich, Switzerland and [8] Department of Science, University of Zurich, Zurich, Switzerland
[9] 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

www.molecularsystemsbiology.com

REPORT

Deep proteome and transcriptome mapping of a human cancer cell line

Nagarjuna Nagaraj[1], Jacek R Wisniewski[1], Tamar Geiger[1], Juergen Cox[1], Martin Kircher[2], Janet Kelso[2], Svante Pääbo[2] and Matthias Mann[1,*]

[1] Department for Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany and [2] 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[1,2], Sneha M. Pinto[3], Derese Getnet[1,4], Raja Sekhar Nirujogi[3], Srikanth S. Manda[3], Raghothama Chaerkady[1,2], Anil K. Madugundu[3], Dhanashree S. Kelkar[3], Ruth Isserlin[5], Shobhit Jain[5], Joji K. Thomas[3], Babylakshmi Muthusamy[3], Pamela Leal-Rojas[1,6], Praveen Kumar[3], Nandini A. Sahasrabuddhe[3], Lavanya Balakrishnan[3], Jayshree Advani[3], Bijesh George[3], Santosh Renuse[3], Lakshmi Dhevi N. Selvan[3], Arun H. Patil[3], Vishalakshi Nanjappa[3], Aneesha Radhakrishnan[3], Samarjeet Prasad[1], Tejaswini Subbannayya[3], Rajesh Raju[3], Manish Kumar[3], Sreelakshmi K. Sreenivasamurthy[3], Arivusudar Marimuthu[3], Gajanan J. Sathe[3], Sandip Chavan[3], Keshava K. Datta[3], Yashwanth Subbannayya[3], Apeksha Sahu[3], Soujanya D. Yelamanchi[3], Savita Jayaram[3], Pavithra Rajagopalan[3], Jyoti Sharma[3], Krishna R. Murthy[3], Nazia Syed[3], Renu Goel[3], Aafaque A. Khan[3], Sartaj Ahmad[3], Gourav Dey[3], Keshav Mudgal[7], Aditi Chatterjee[3], Tai-Chung Huang[1], Jun Zhong[1], Xinyan Wu[1,2], Patrick G. Shaw[1], Donald Freed[1], Muhammad S. Zahari[2], Kanchan K. Mukherjee[8], Subramanian Shankar[9], Anita Mahadevan[10,11], Henry Lam[12], Christopher J. Mitchell[3], Susarla Krishna Shankar[10,11], Parthasarathy Satishchandra[13], John T. Schroeder[14], Ravi Sirdeshmukh[3], Anirban Maitra[15,16], Steven D. Leach[1,17], Charles G. Drake[16,18], Marc K. Halushka[15], T. S. Keshava Prasad[3], Ralph H. Hruban[15,16], Candace L. Kerr[19]†, Gary D. Bader[5], Christine A. Iacobuzio-Donahue[15,16,17], Harsha Gowda[3] & Akhilesh Pandey[1,2,3,4,15,16,20]

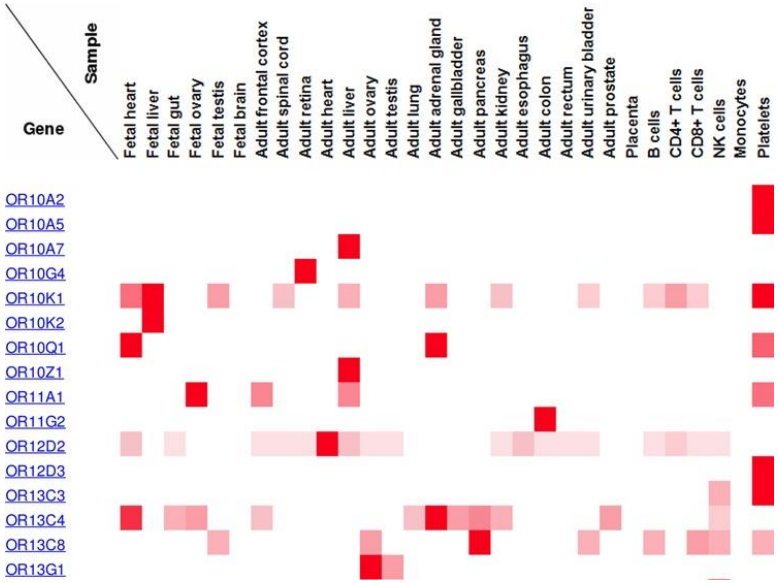
ARTICLE 2014 Kuster lab

doi:10.1038/nature13319

Mass-spectrometry-based draft of the human proteome

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

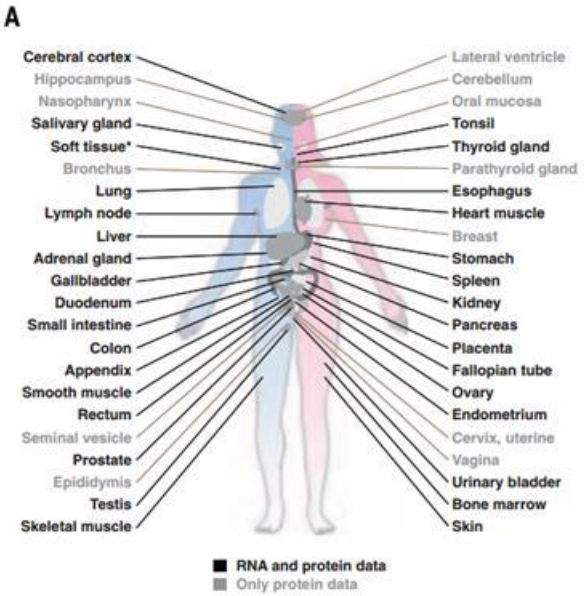
Many proteins are identified with peptides belonging to more than one protein (e.g. isoforms)

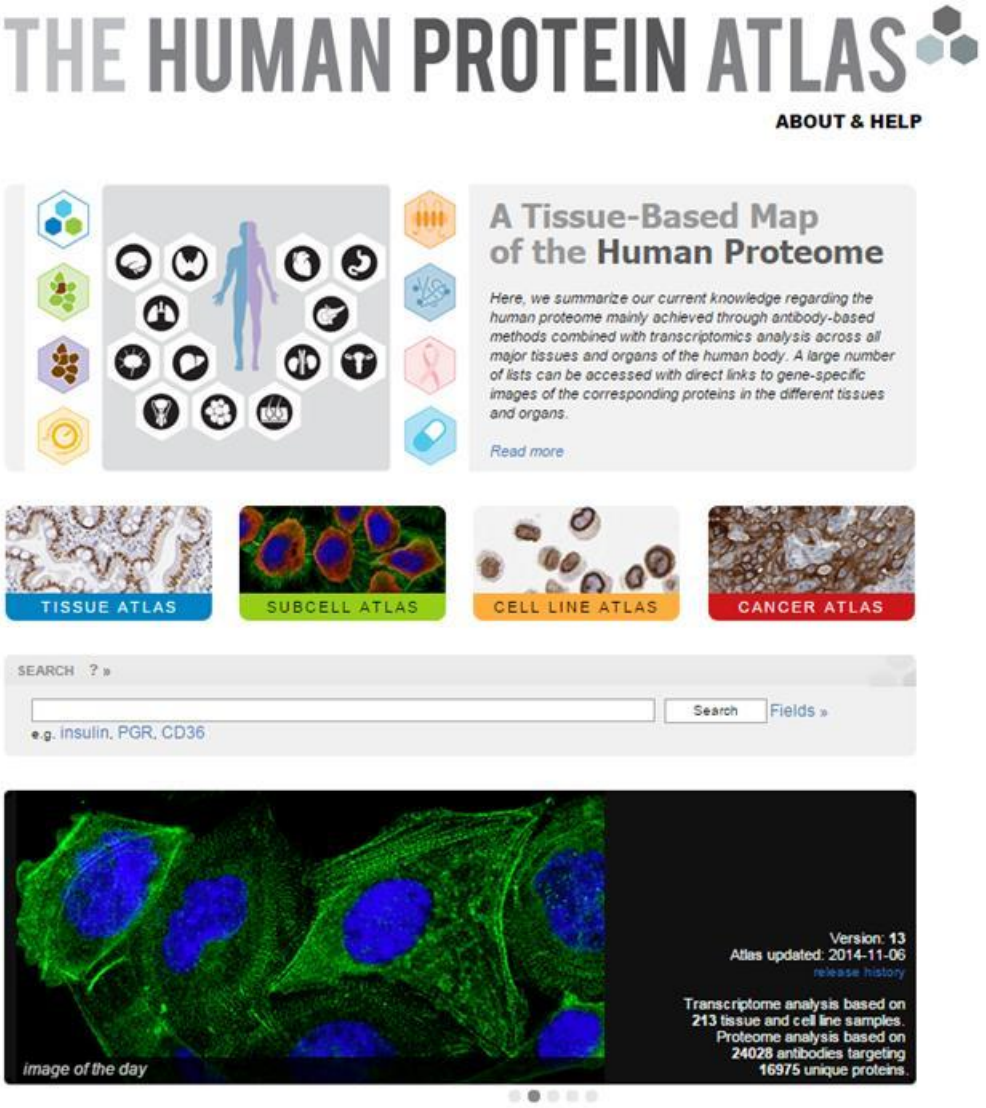


Ezkurdia et al, J Proteome Res, 2014

Antibody-based proteomics: only semi-quantitative abundances

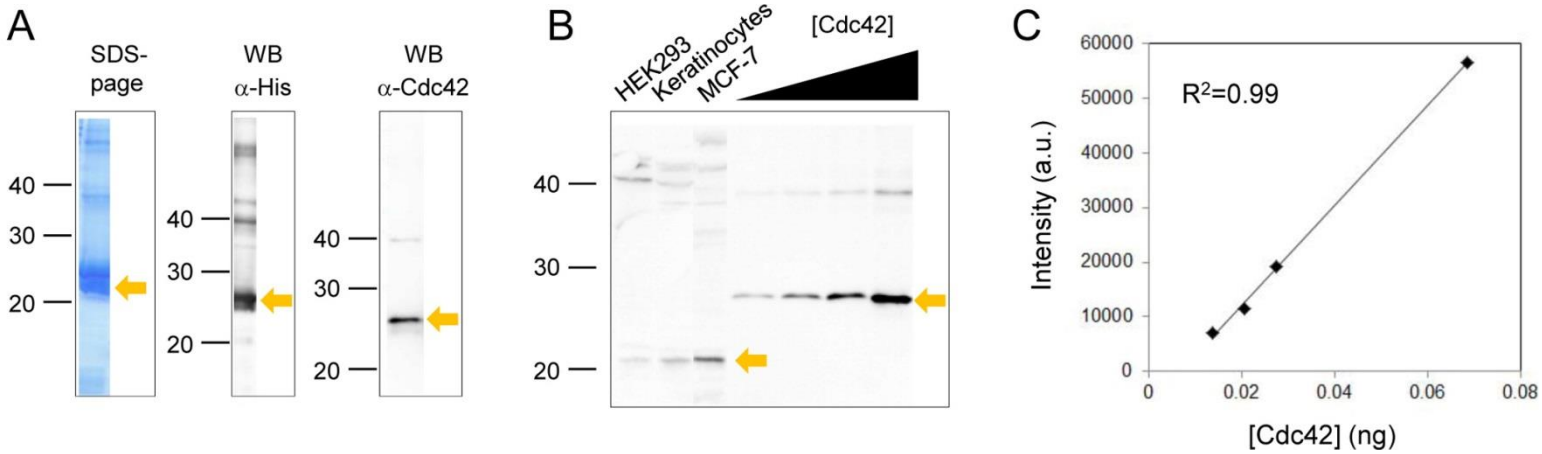
- Tissue-based map of the human proteome
- 44 major tissues and organs in the human body
- 24,028 antibodies corresponding to 16,975 protein-encoding genes



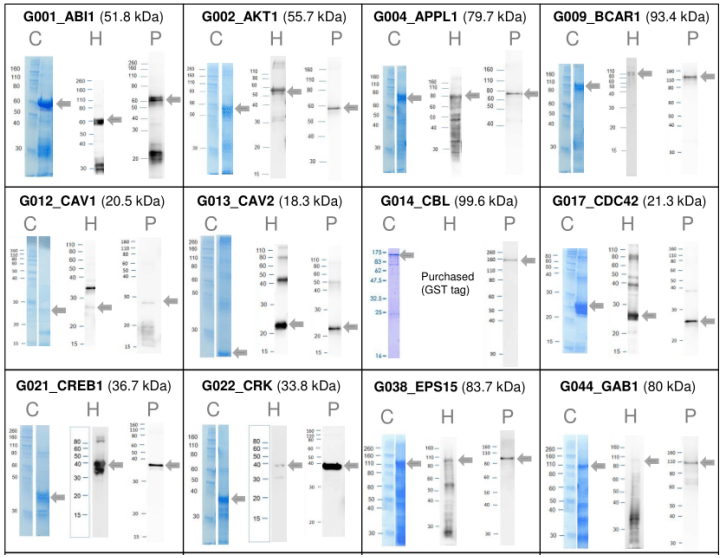


Uhlen et al, Science, 2015

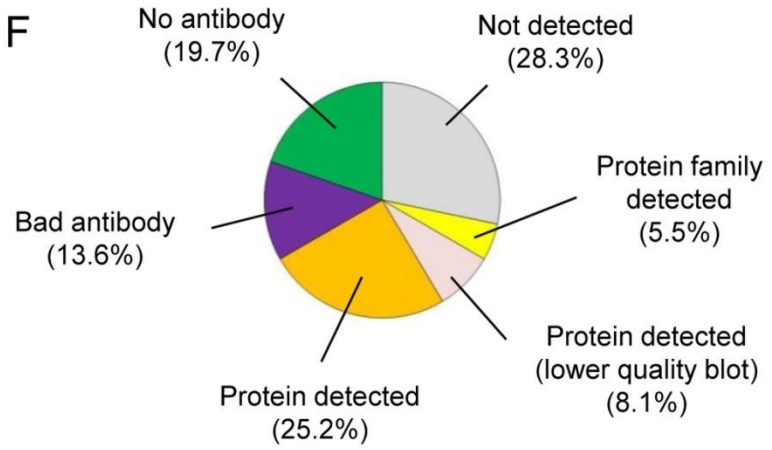
Quantitative Western blotting



Protein standards: expression, purification and quantification

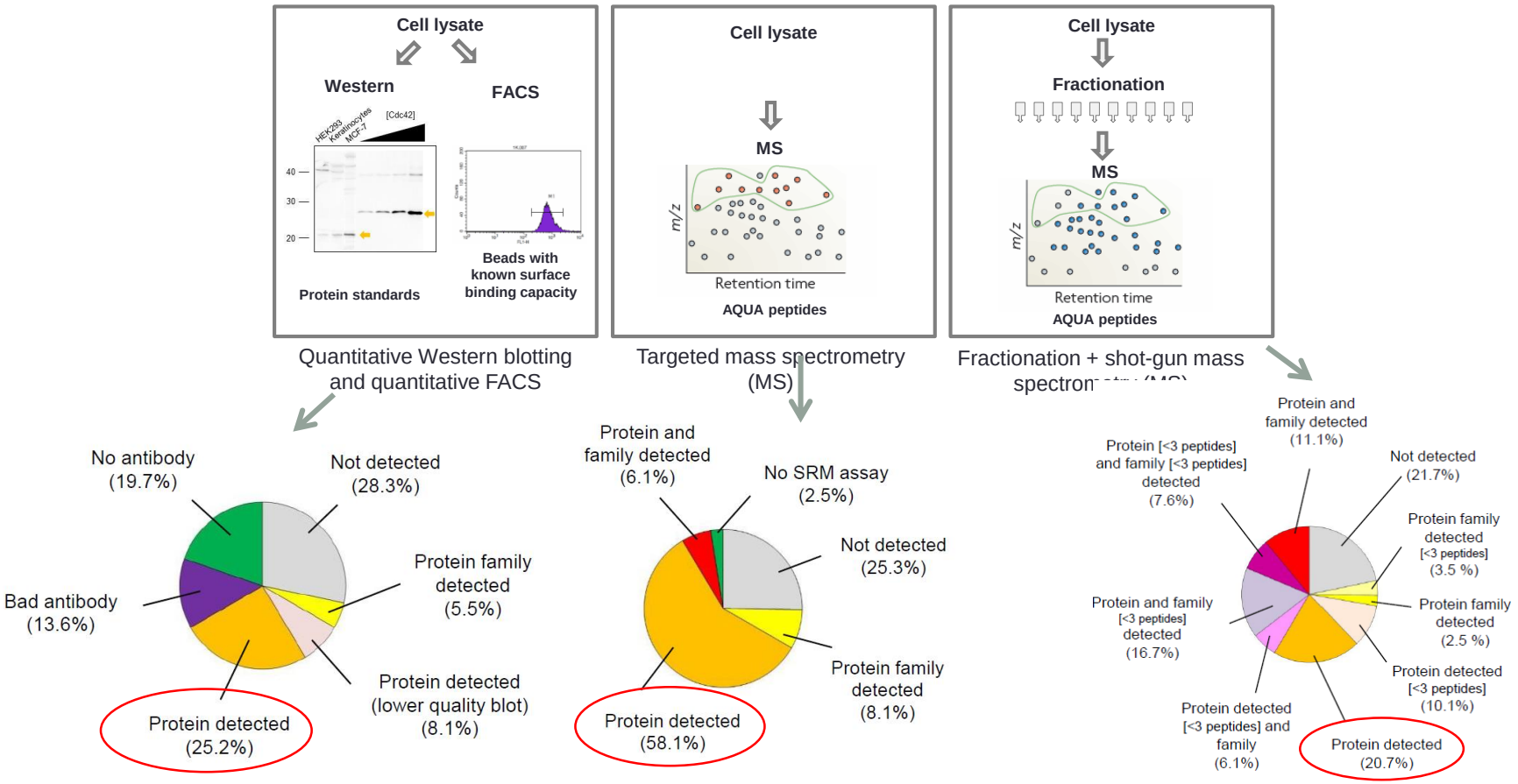


Summary statistic for quantitative Western blotting of 198 ErbB-related proteins



Kiel et al, J Prot Res, 2014

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



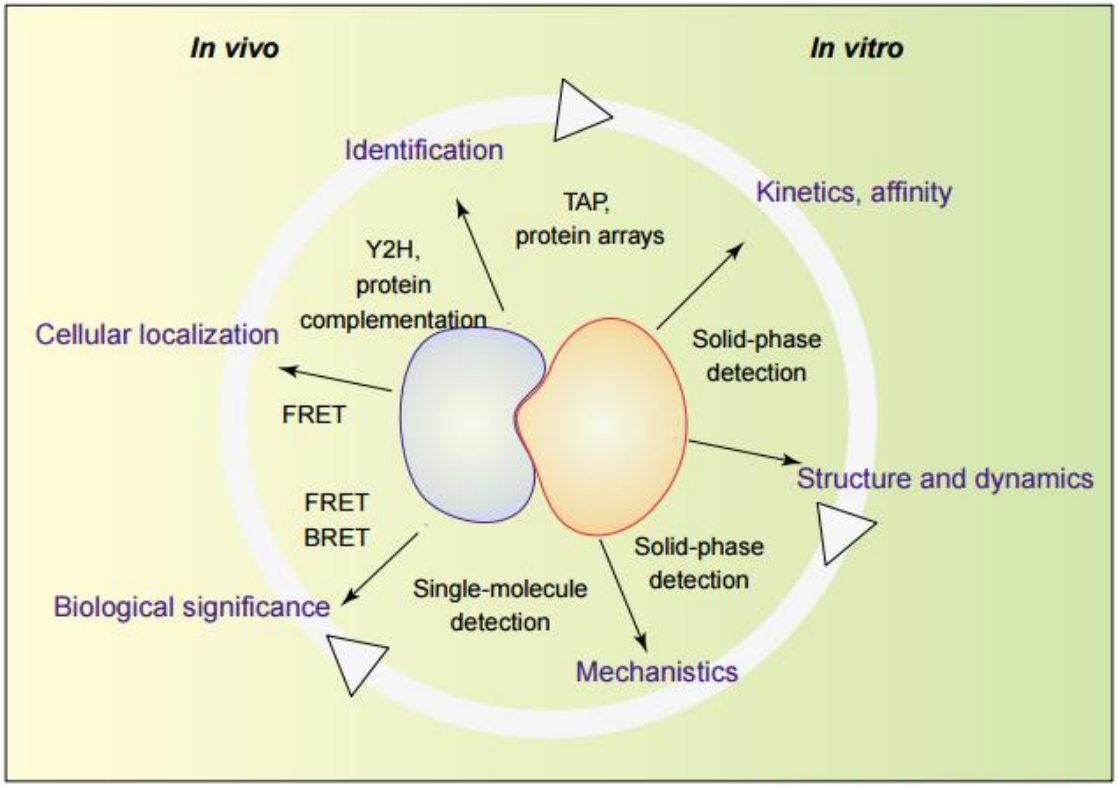
- SRM has a higher sensitivity compared to quantitative western blotting (but some proteins are only detected by Western blotting)
- Problem with isoforms and protein families: as a consequence of frequent gene duplication events in mammals, often similar proteins (e.g. AKT1 and AKT2) cannot be distinguished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

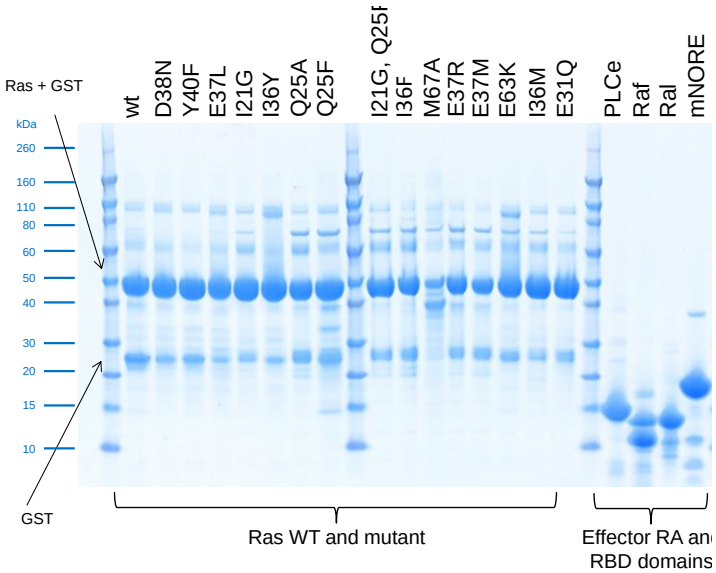
- most *in vivo* techniques are high-throughput, but do not provide affinities (only qualitative binding detection)
- *in vitro* techniques can provide affinities and kinetic constants, but are not high-throughput methods

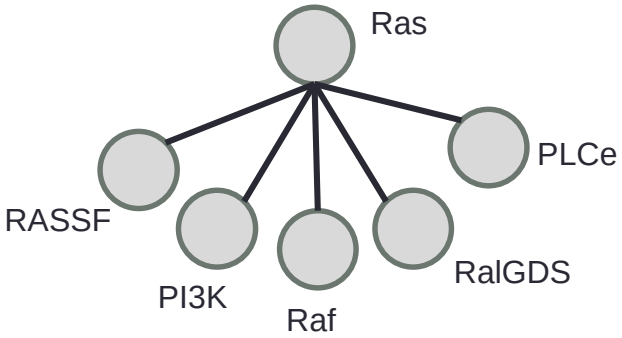


Piehler, Curr Opin Struct Biol, 2005

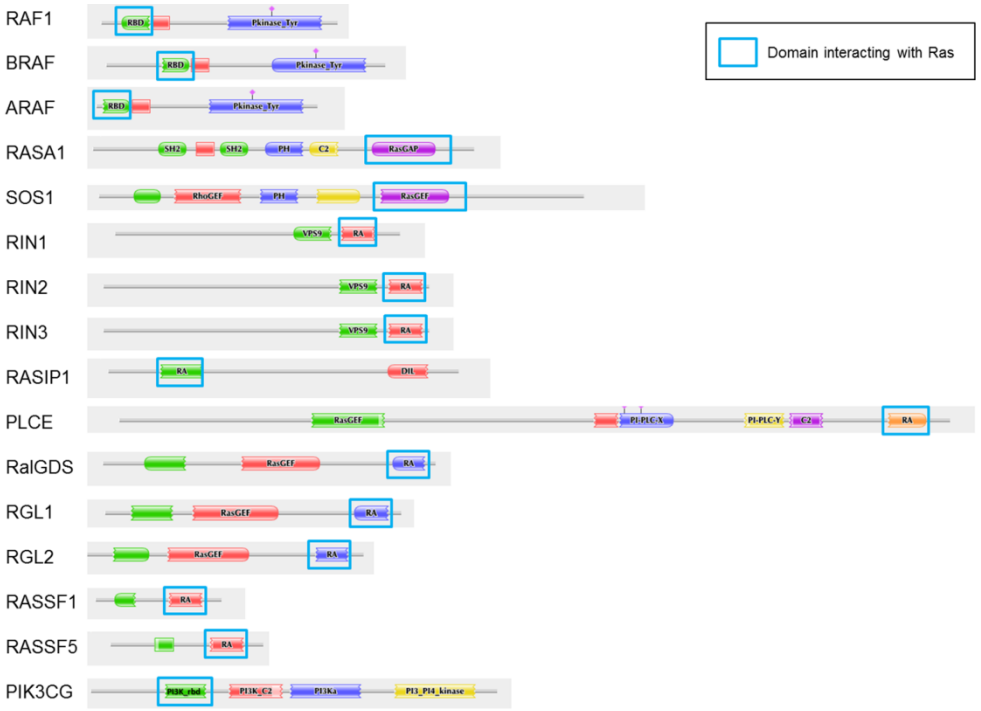
Measuring protein affinities in vitro requires the expression and purification of proteins (e.g. using bacteria)

Example: Bacterial expressed and purified Ras protein mutants and interactors



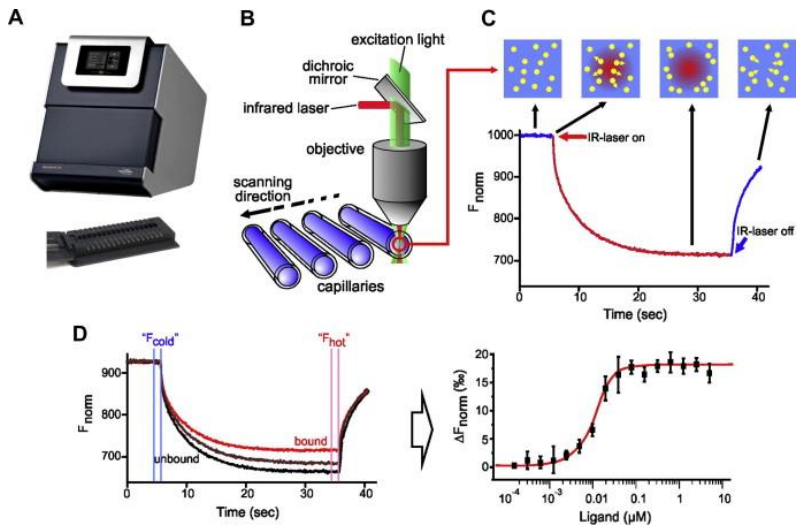


Large proteins are often not soluble: expression and purification of protein domains

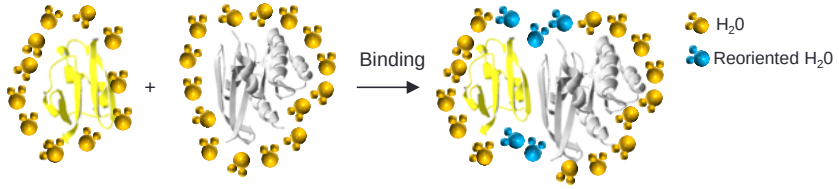


Two main methods to measure affinities and kinetic constants

Microscale thermophoresis



Jerabek-Willemsen et al, J Mol Struct, 2014



Amine-covalent labelled RBDs (fluorophore)

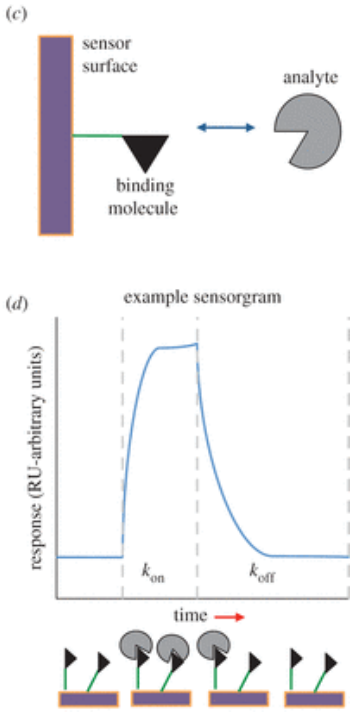
Ras WT and Mut (serial dilutions)

Fluorescence signal (depends on charge, size and hydration shell

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

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

Surface plasmon resonance



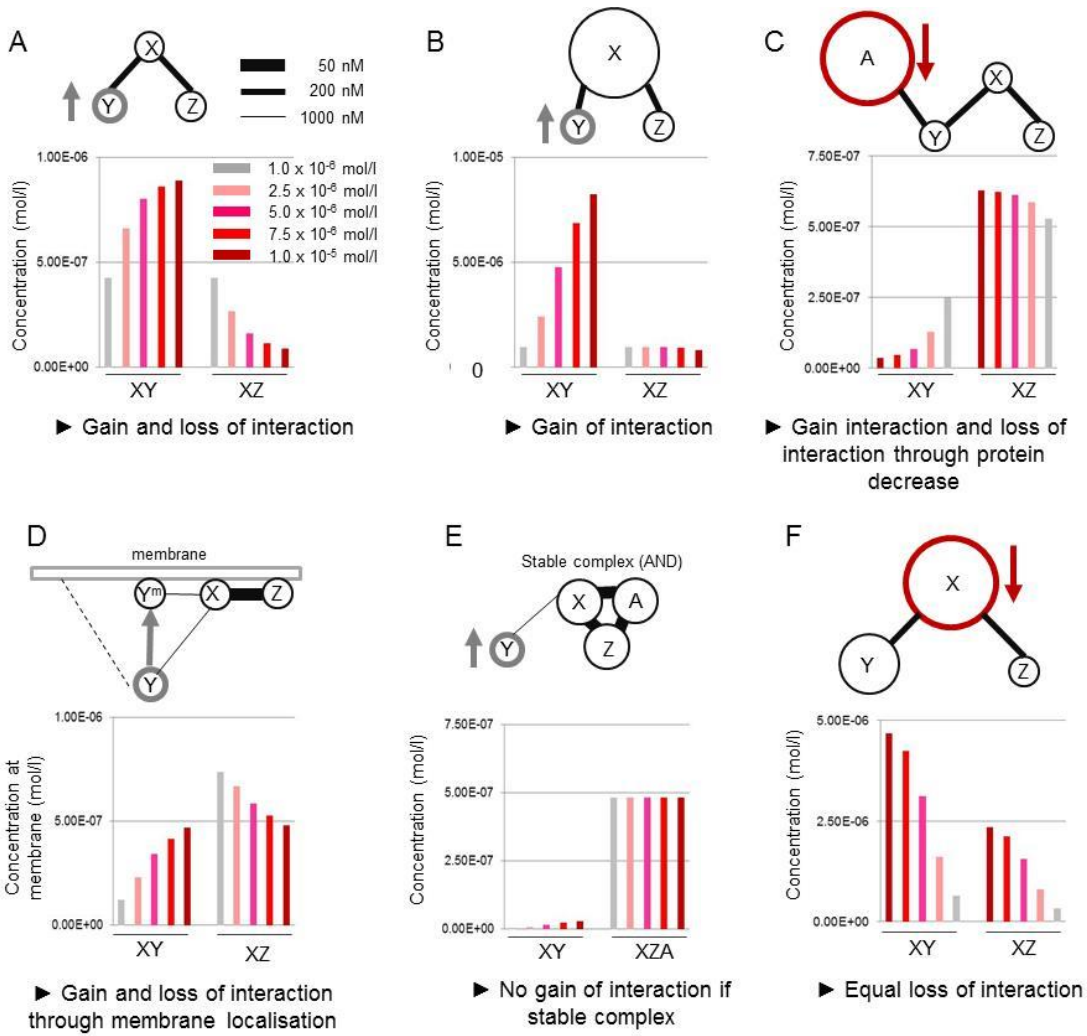
Kastritis et al, 2012

Optical method to measure the refractive index near a sensor surface

- Provides kinetic constants (k_{on} and k_{off})

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

The effect of abundance variation at XOR network motifs

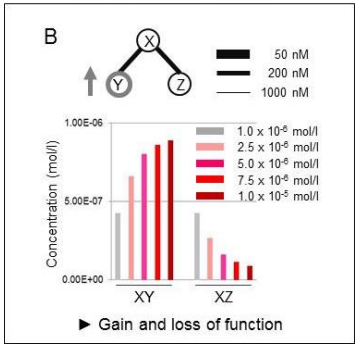


- The output/ function depends on both, network structure and abundance: we need to know the network very well to understand

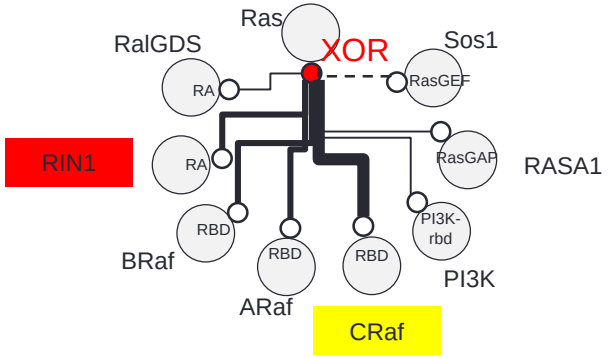
Kiel et al, Sci Signal, 2013

Competition at the Ras XOR node

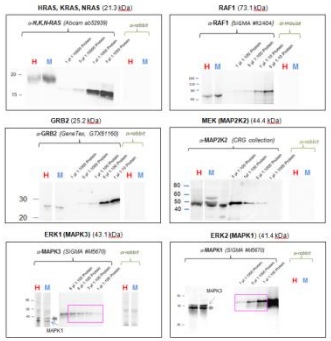
Network motif



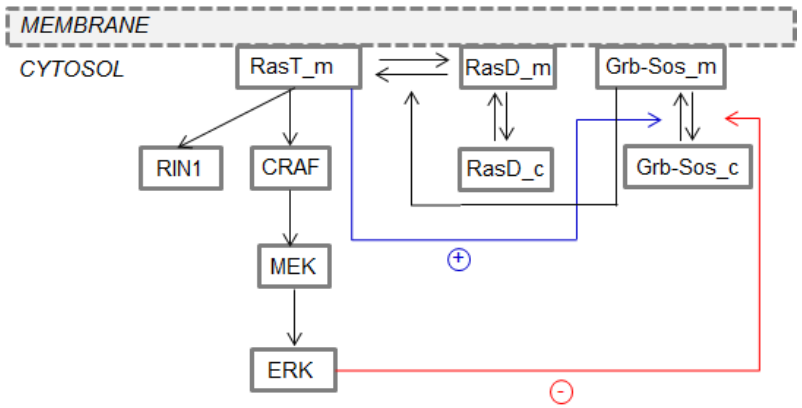
The Ras XOR node



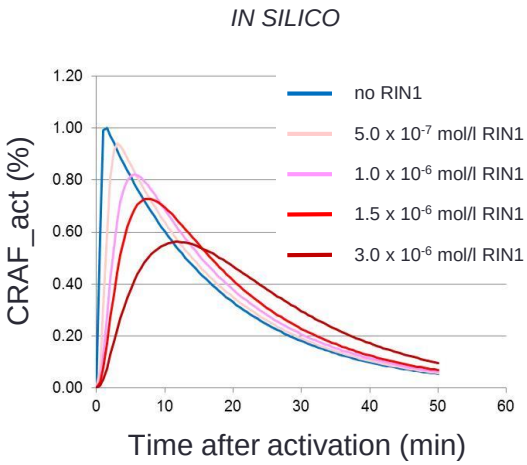
Experimental abundances



A



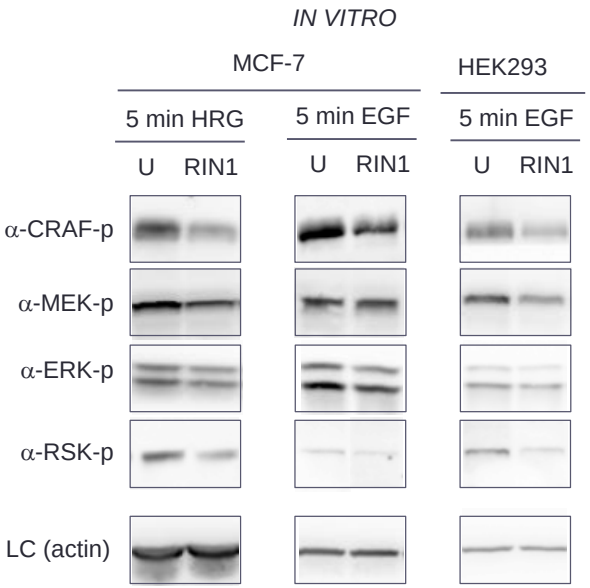
B

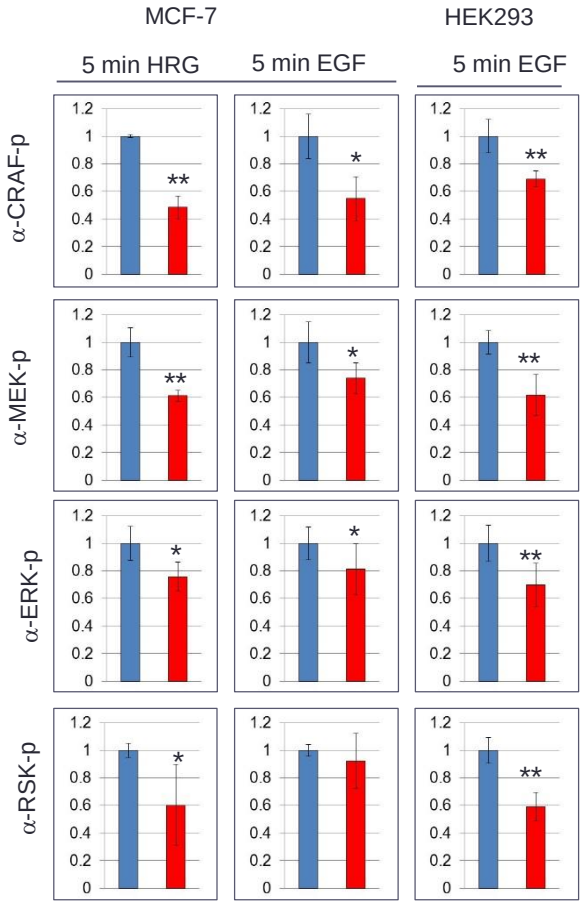


Mathematical network modeling: increasing RIN1 to 10-fold higher of CRAF expression should decrease CRAF activation

Kiel et al, Sci Signal, 2013

Experimental testing of competition at the Ras node

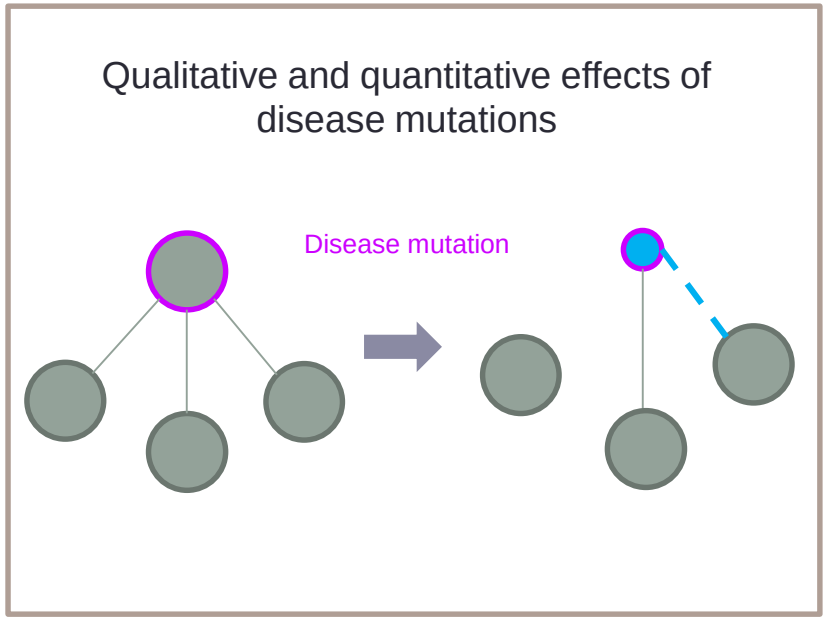




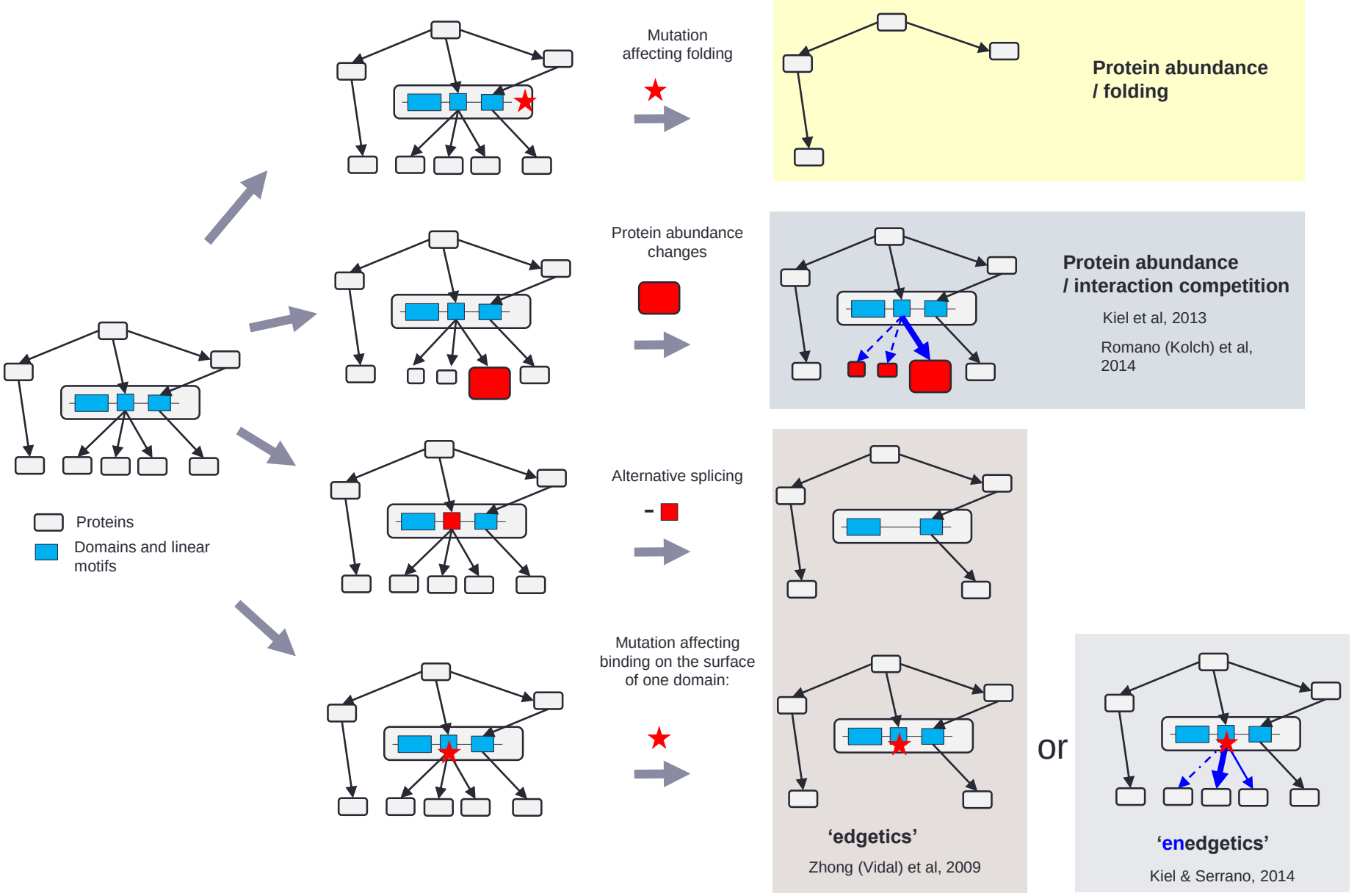
Expression of RIN1 in MCF-7 and HEK293 cells decreases CRAF, MEK, and ERK activation

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

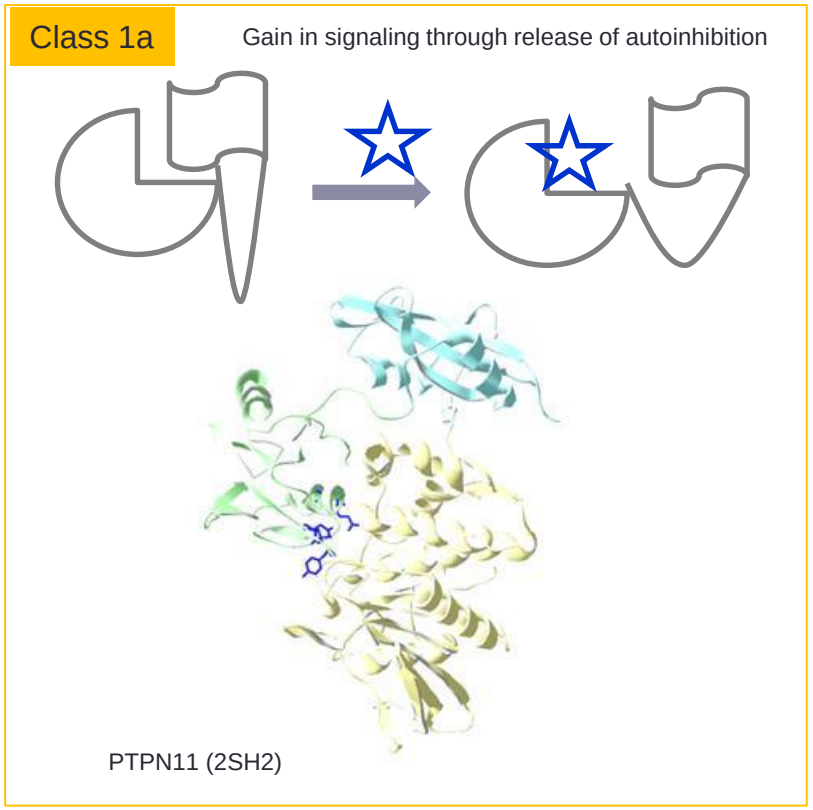
Kiel et al, Sci Signal, 2013


Qualitative and quantitative effects of disease mutations
Disease mutation

General concepts of interaction (‘edge’) rewiring

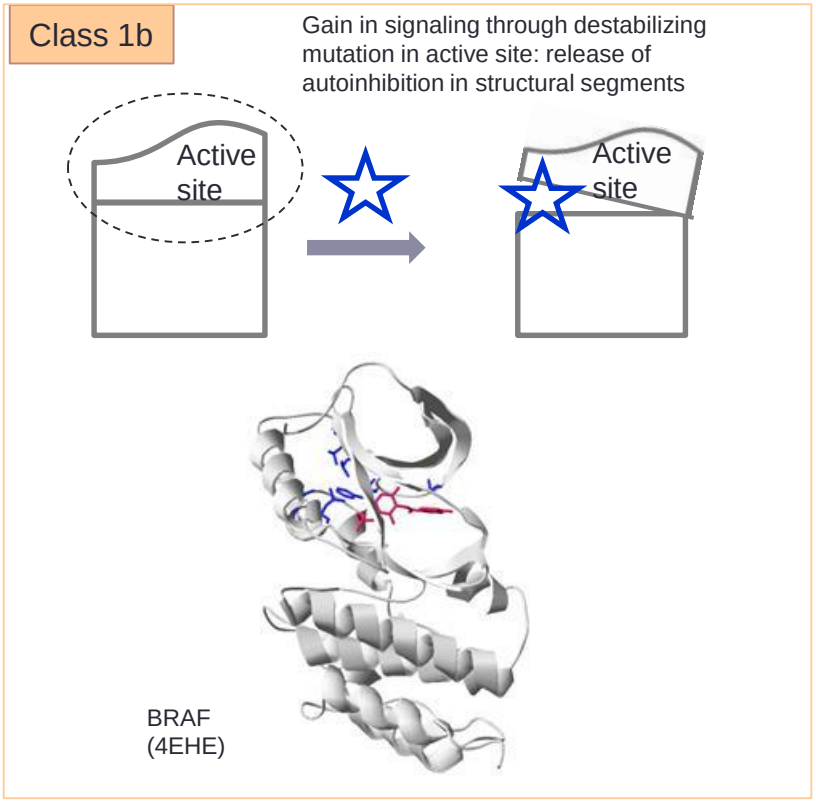


Examples how missense mutations can affect the network: a 3D structural perspective



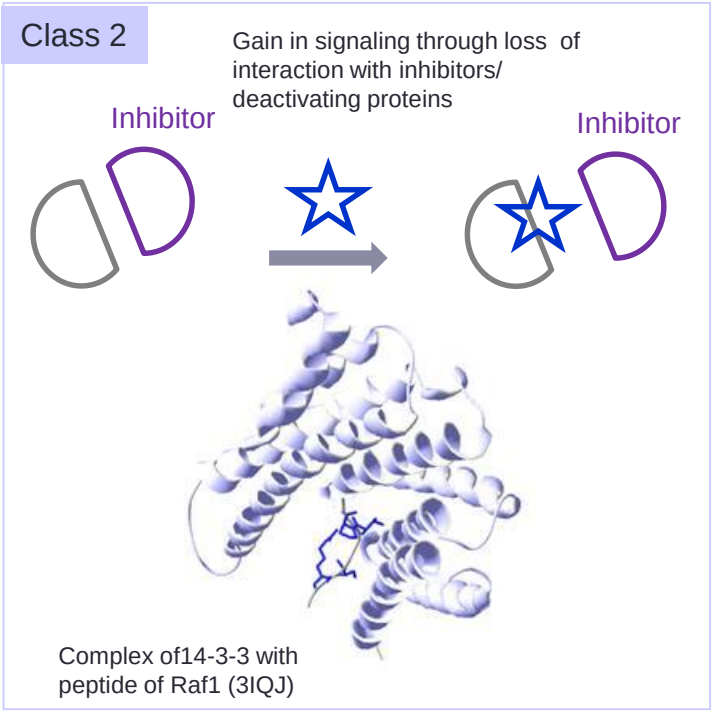
Kiel & Serrano, Mol Sys Biol, 2014

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



Kiel & Serrano, Mol Sys Biol, 2014

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

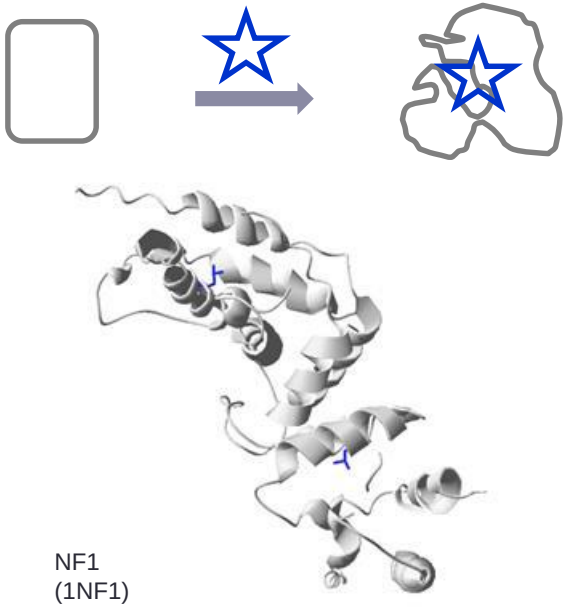


Kiel & Serrano, Mol Sys Biol, 2014

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



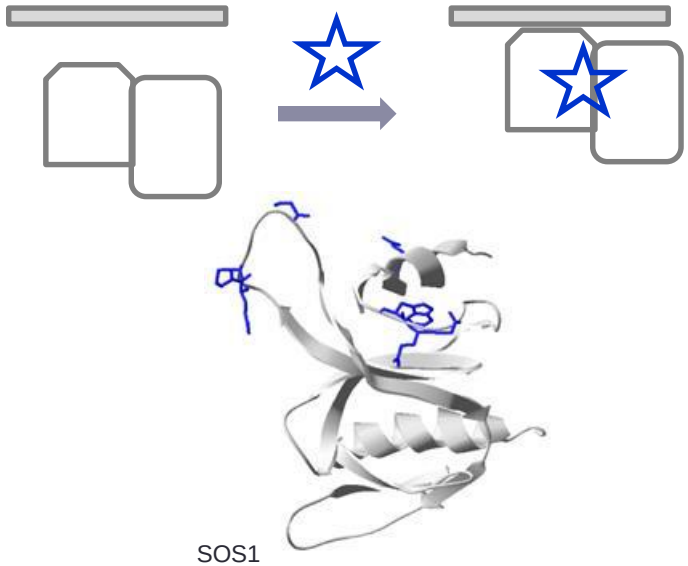
NF1
(1NF1)

Kiel & Serrano, Mol Sys Biol, 2014

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

Class 4

Gain in signaling through mutation of domains involved in membrane recruitment

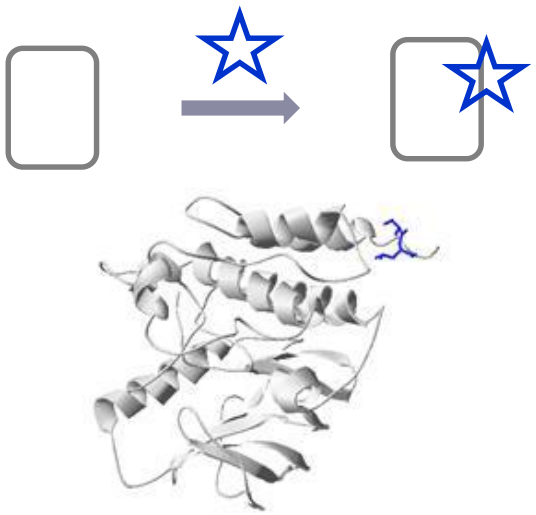


SOS1
(1DBH)

Kiel & Serrano, Mol Sys Biol, 2014

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

Class 5 No effect; location on surface



SOS1
(1DBH)

Kiel & Serrano, Mol Sys Biol, 2014

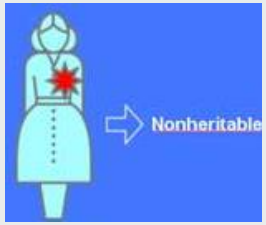
Example 1: RASopathy and cancer disease mutations

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)
- 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

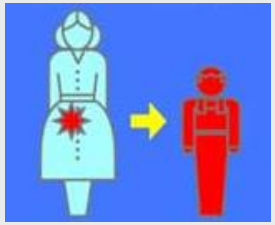
Somatic mutations

- 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)

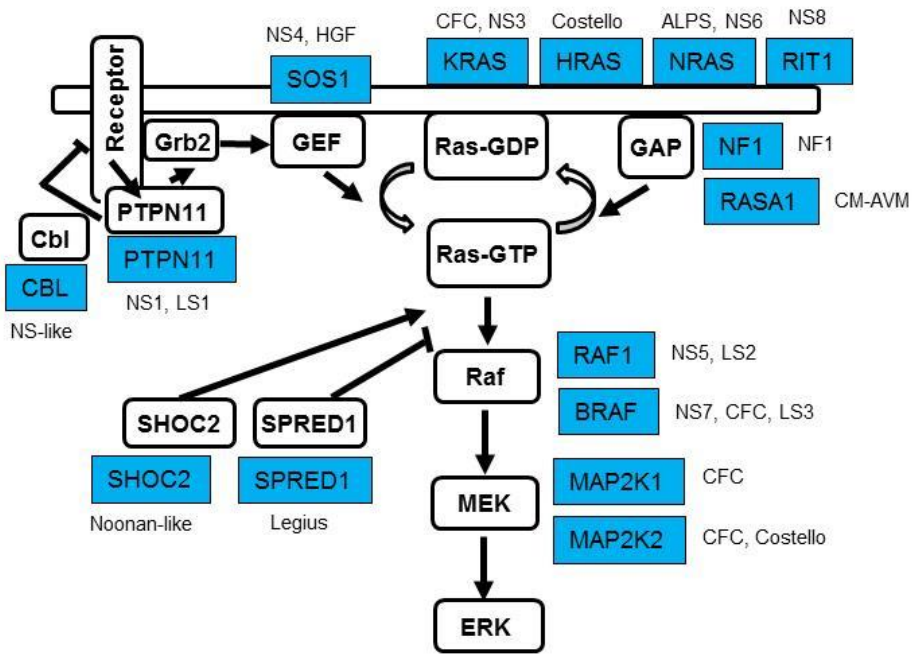


Christina Kiel

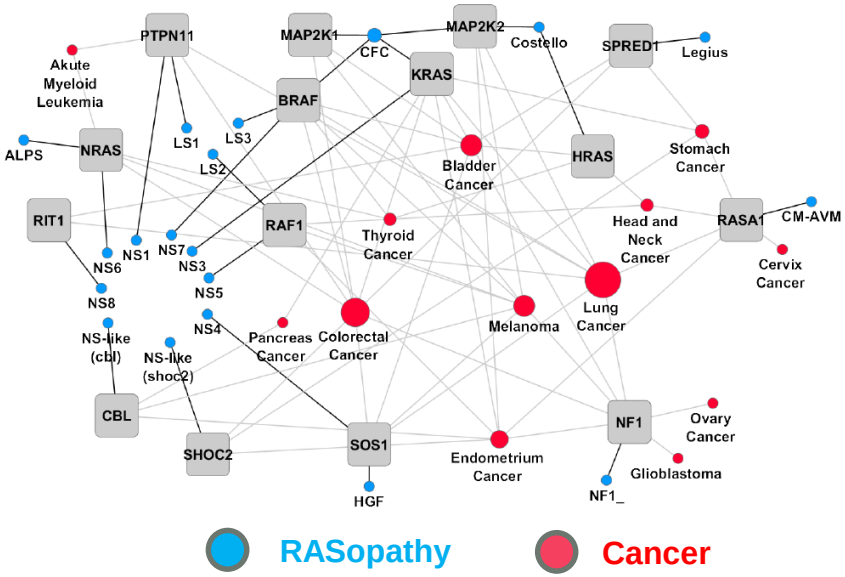
Hannah Benisty

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

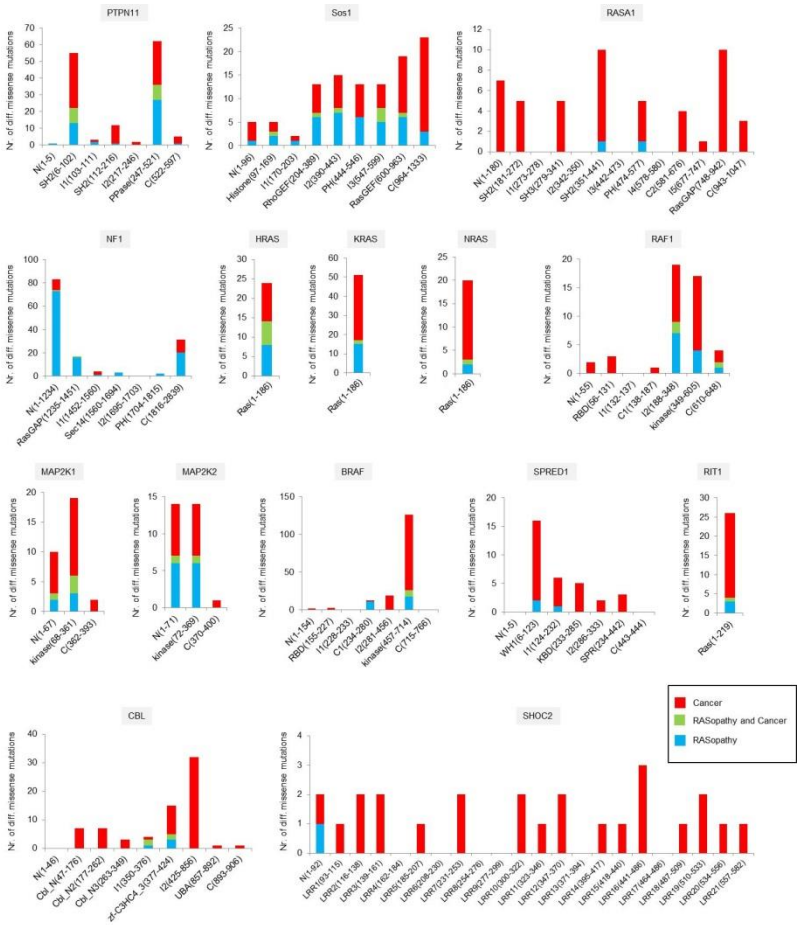


- Proteins in Ras/MAPK syndromes ('RASopathies') are also found in cancer



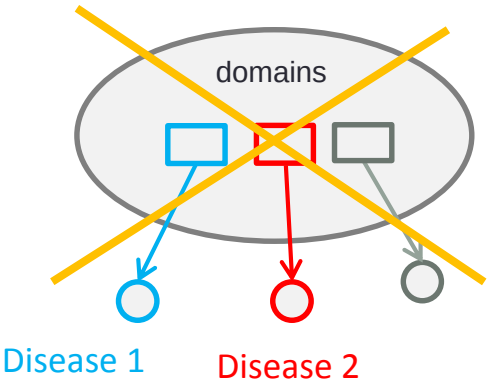
Location of mutations in different domains does not explain the difference between RASopathy and cancer mutations

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



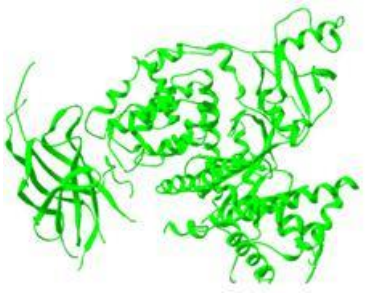
Kiel & Serrano, Mol Sys Biol, 2014

FoldX-based energy calculations of proteins

3D Structural information

A force field for energy calculations and protein design

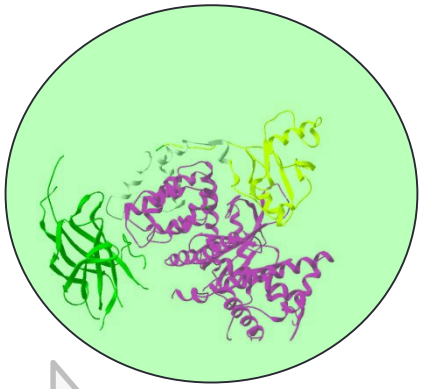
Schymkowitz et et al, *Nucleic Acids Res*, 2005



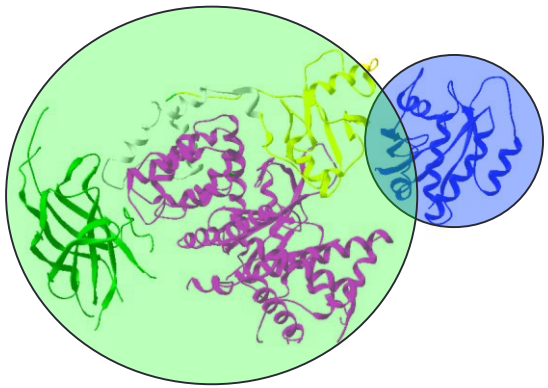
+ FoldX = ΔG

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

✓ Total free energy

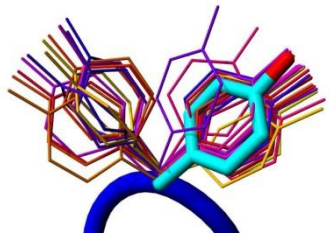


✓ Interaction energy



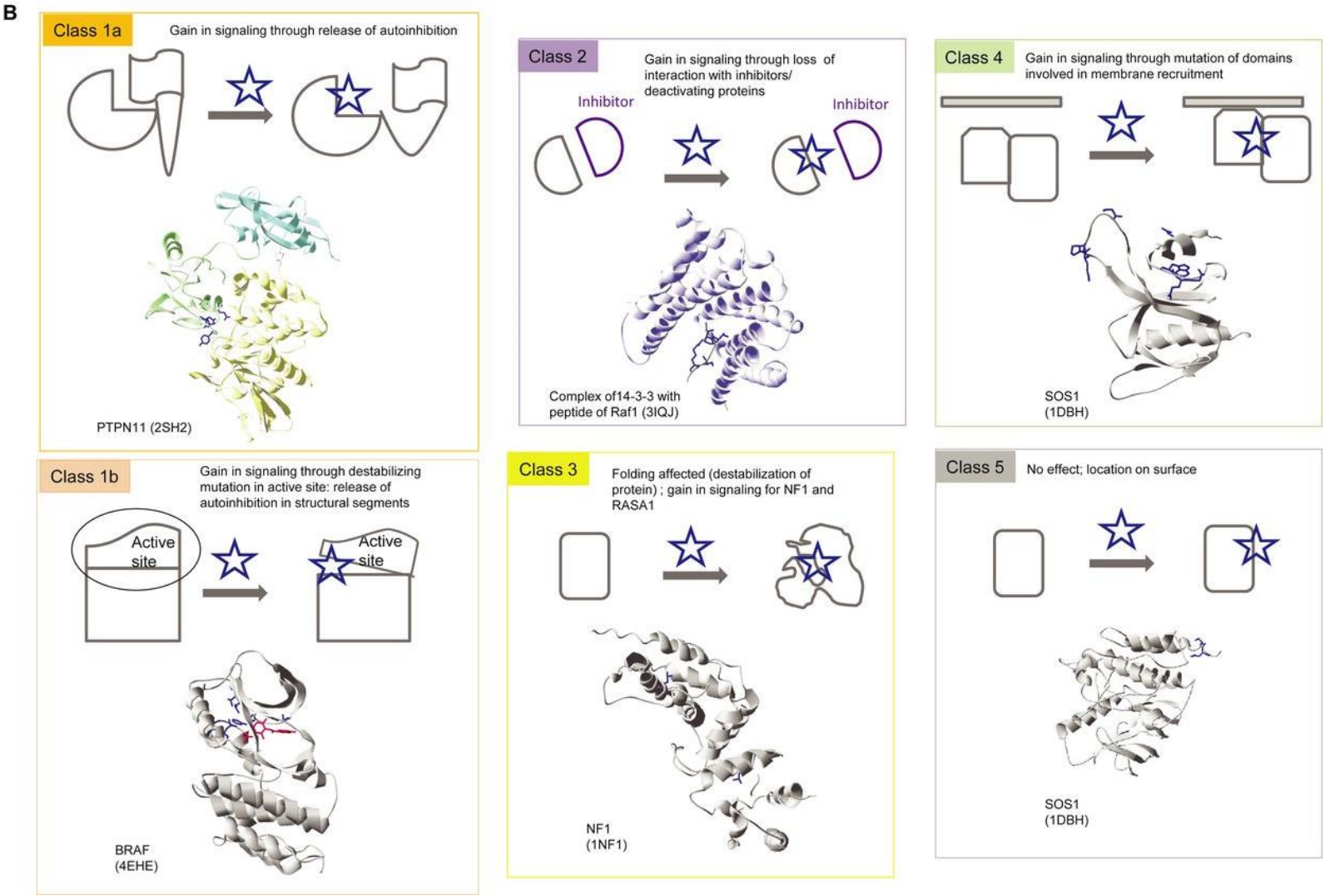
✓ Mutagenesis

A rotamer library to replace the 20 amino acids



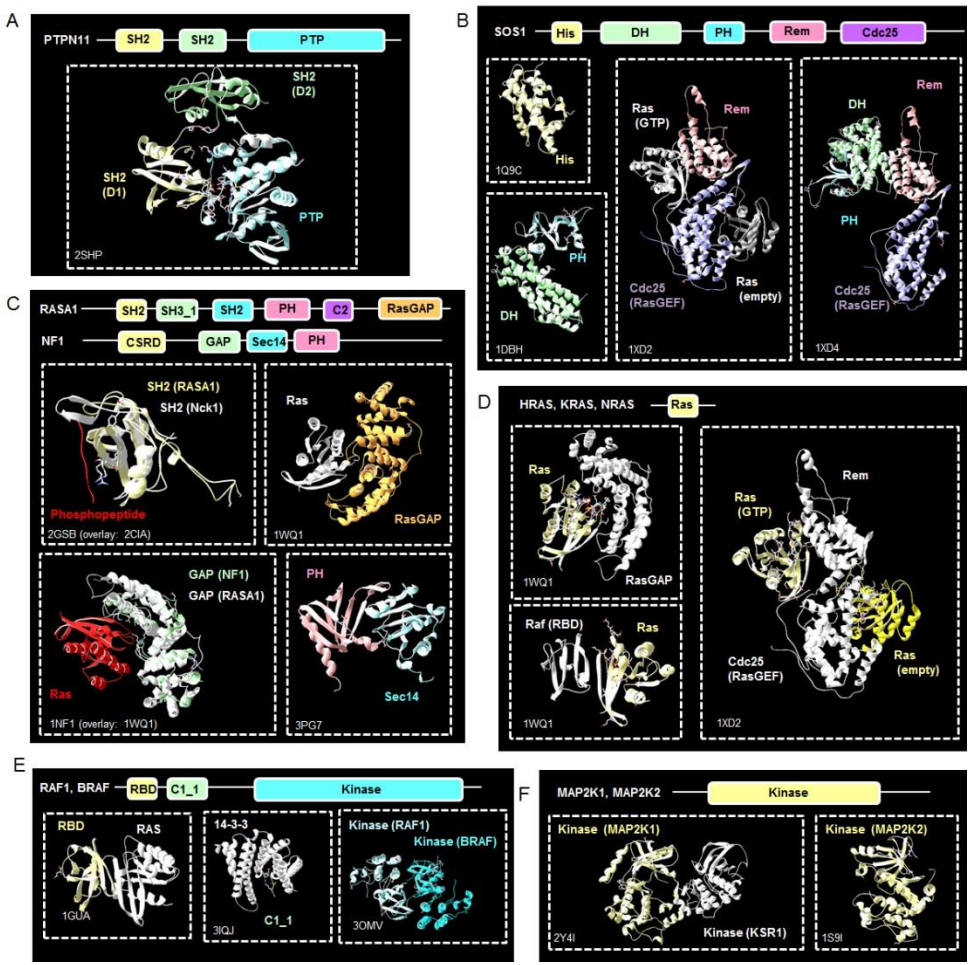
Protein design

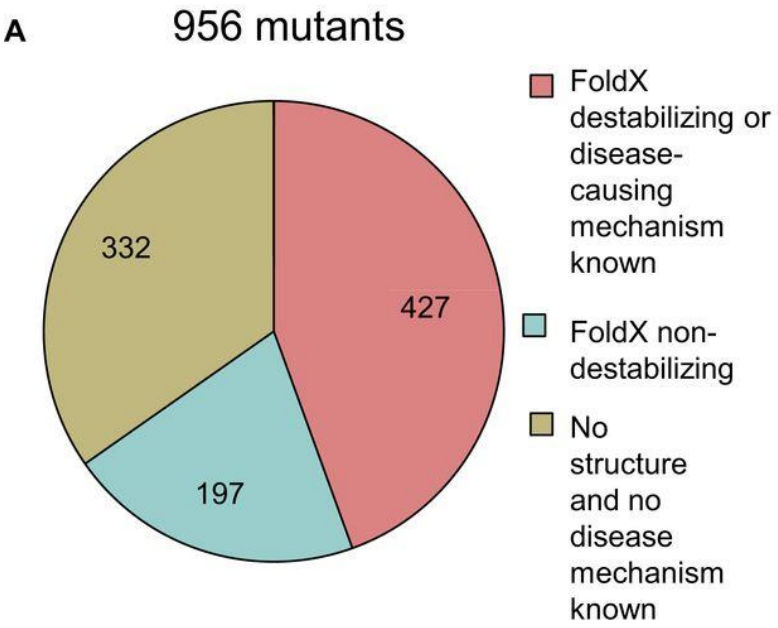
Analysis of 956 missense mutations in RASopathies and cancer based on structural information and FoldX energies



Kiel & Serrano, Mol Sys Biol, 2014

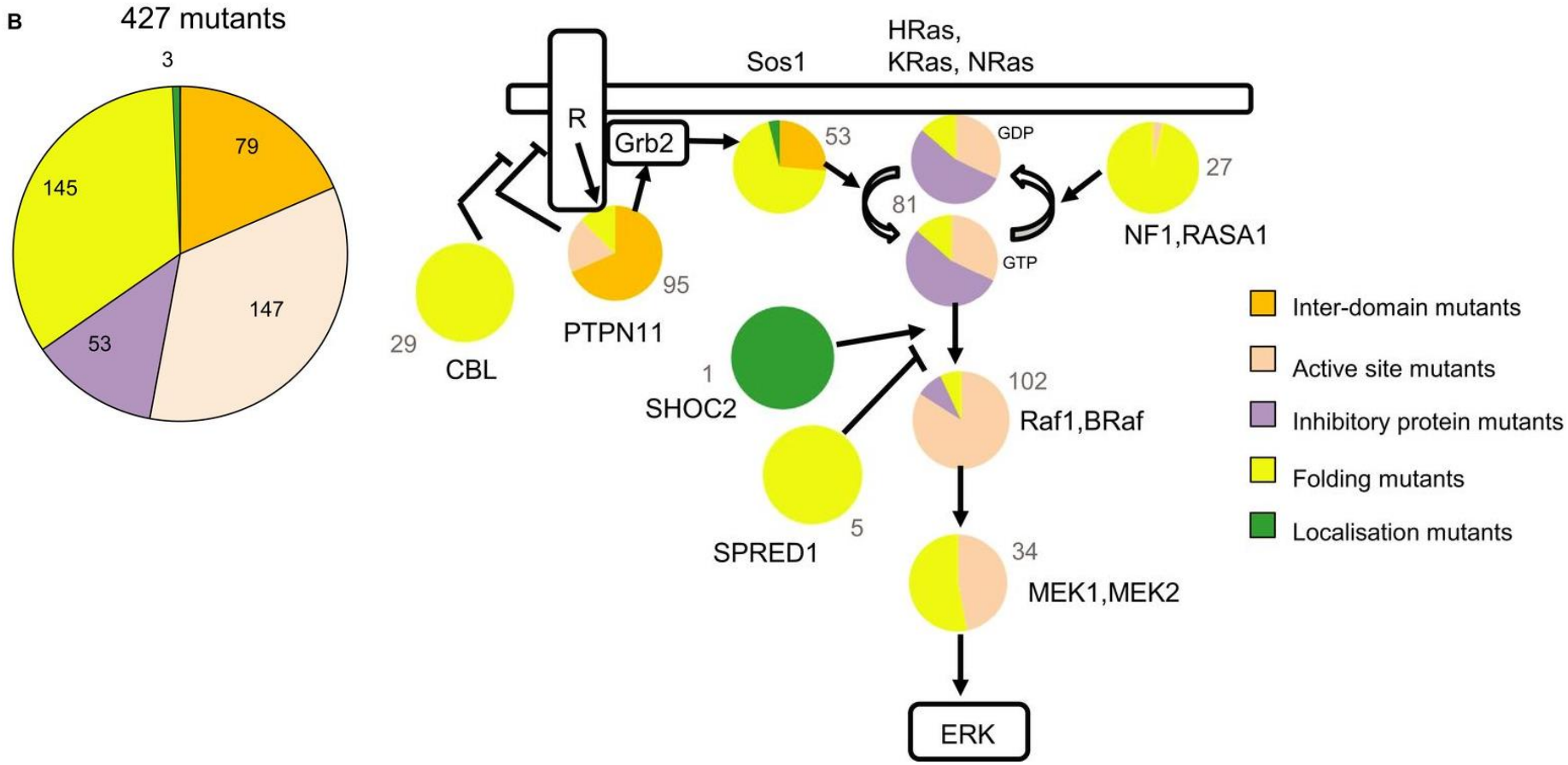
Analysis of 956 missense mutations in RASopathies and cancer: high structural coverage





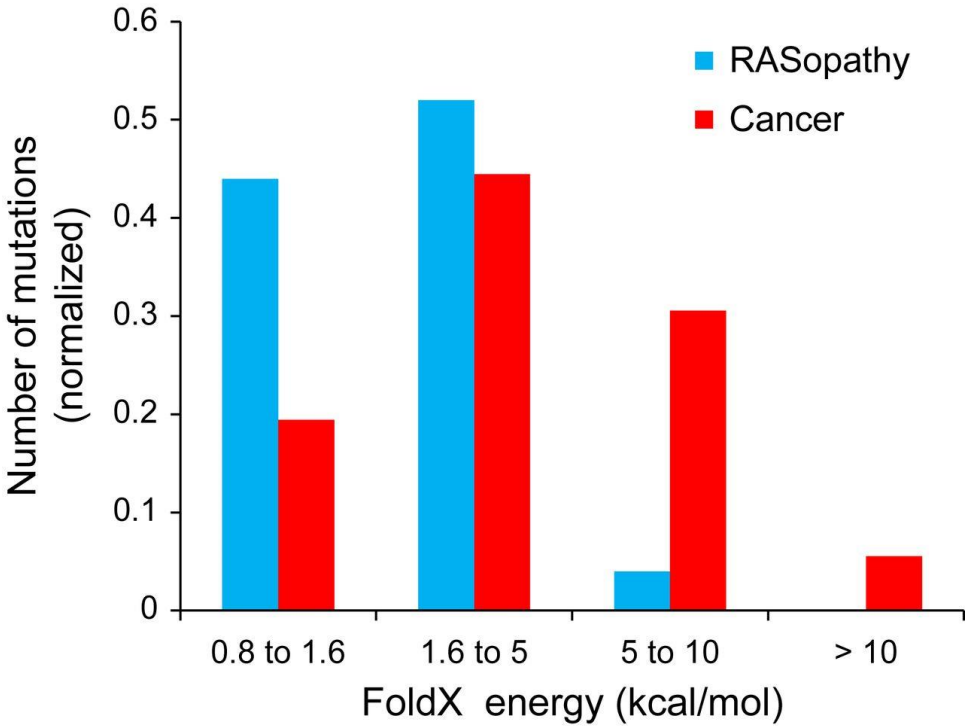
Kiel & Serrano, Mol Sys Biol, 2014

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



Kiel & Serrano, Mol Sys Biol, 2014

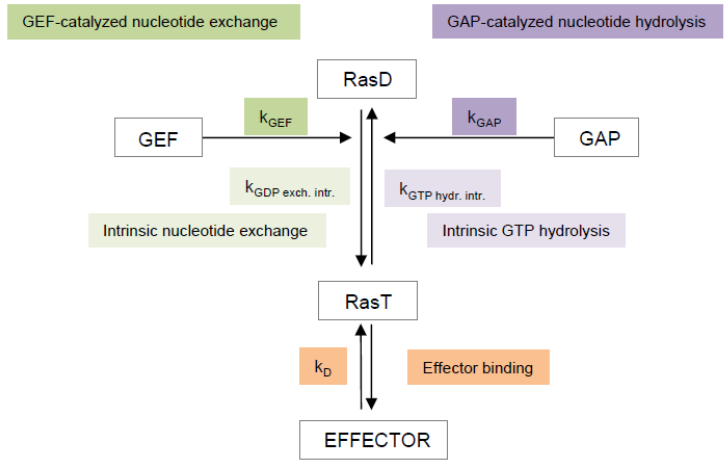
Cancer mutations tend to have higher destabilization values (on average)



Kiel & Serrano, Mol Sys Biol, 2014

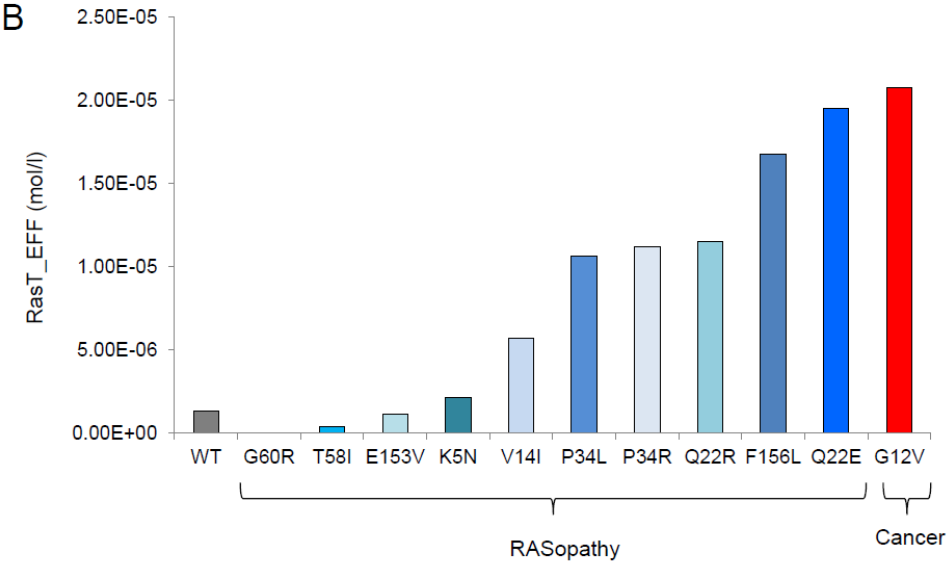
Quantitative effects on protein stability, or activity could explain in some cases the different phenotype: cancer or RASopathy

A



Simulation of Ras activation

B



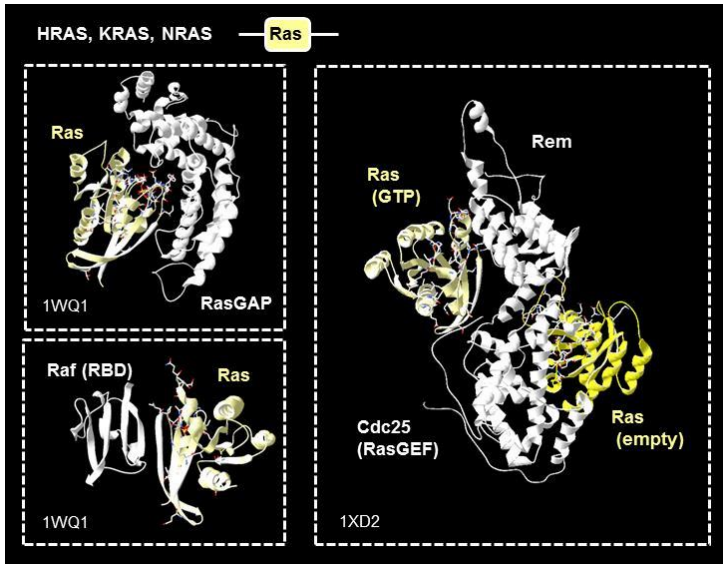
‘Enedgetics’: quantitative edge effects

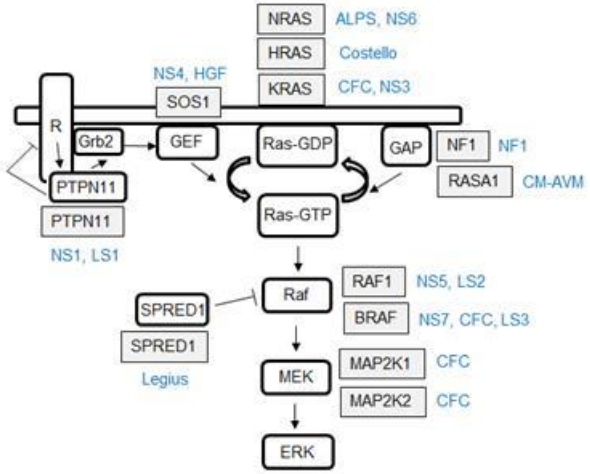
‘Edgetics’ + energies = ‘enedgetics’

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

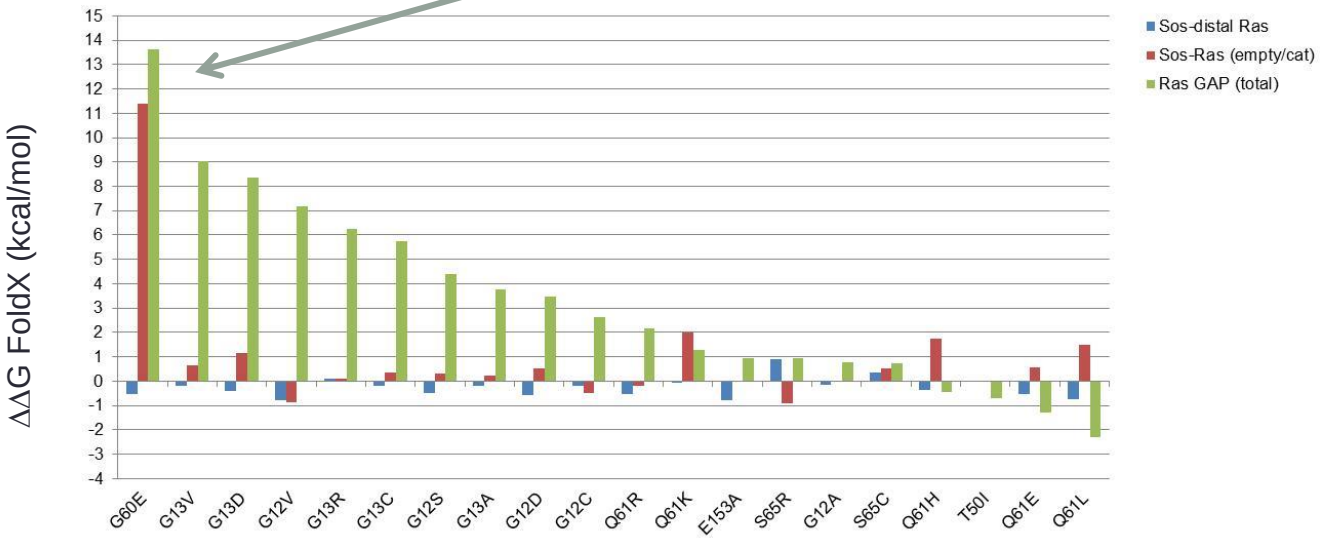
Kiel & Serrano, Mol Sys Biol, 2014

Compensatory effects of mutations on different interaction partners





NRAS G60E

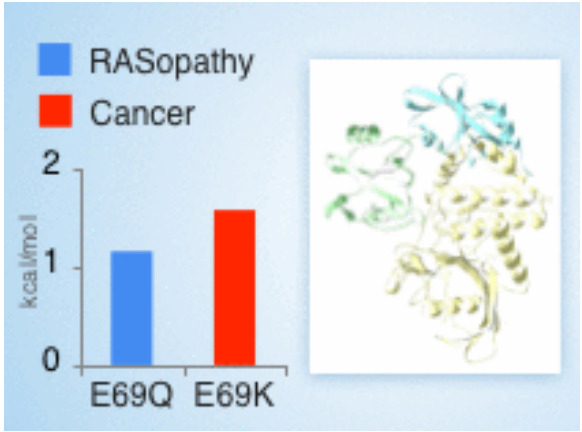


Kiel & Serrano, Mol Sys Biol, 2014

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.

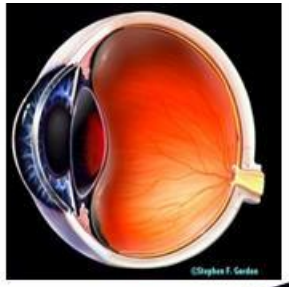
➢ Combined network-based and structural analyses show that quantitative changes rather than all-or-none rewiring underlie the difference between RASopathy and Cancer mutations.

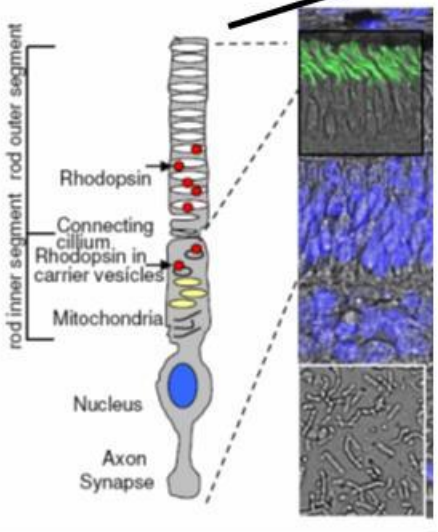


Example 2: Rhodopsin disease mutations

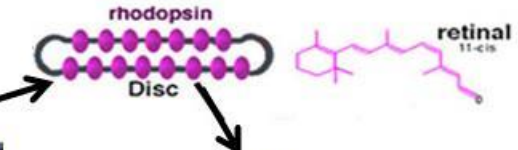
Rhodopsin: involved in light perception in rod outer segment

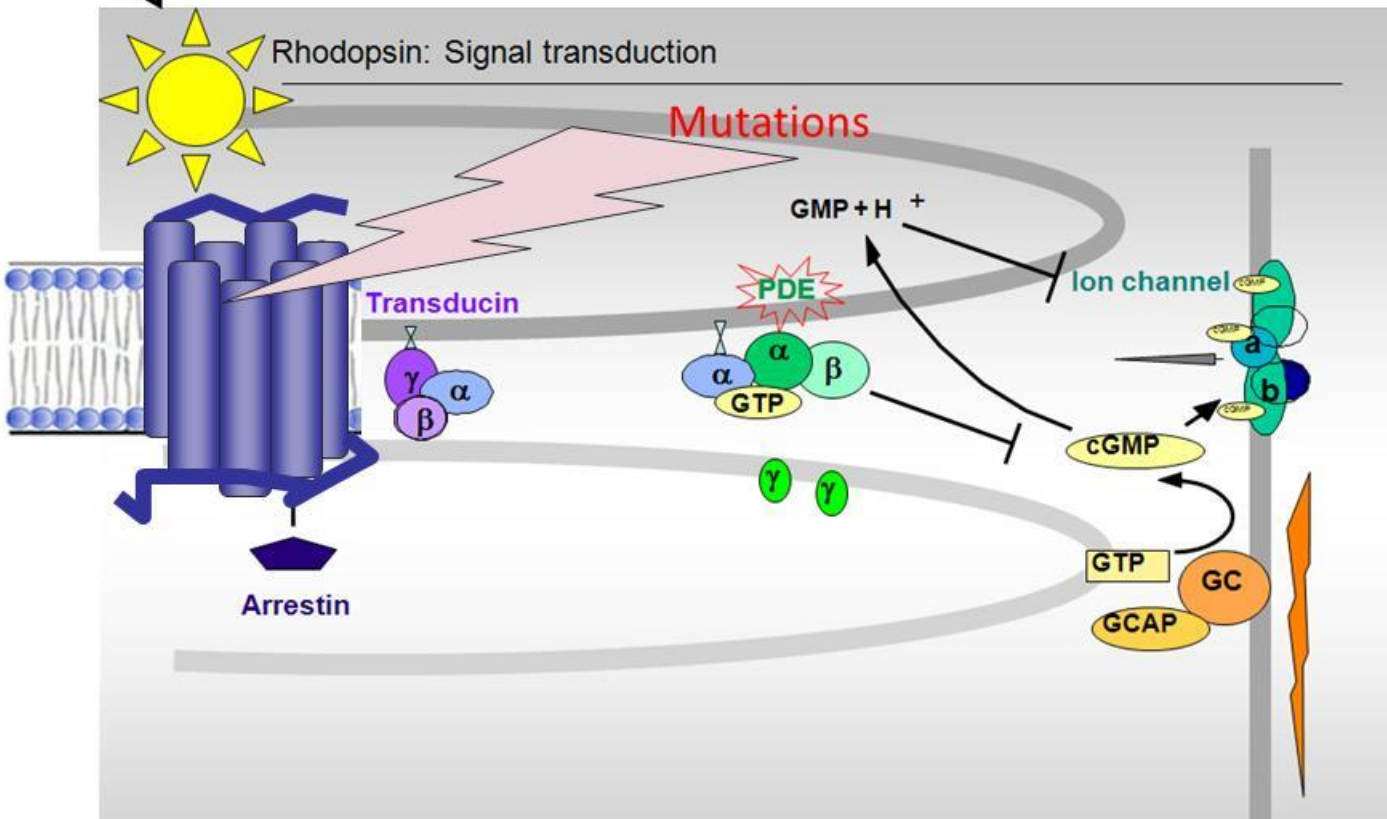
Understanding disease mutations in rhodopsin, a common cause of retinitis pigmentosa (RP)





outer segment



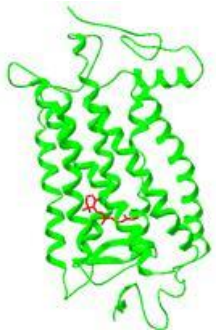


Analysis of 103 mutations in rhodopsin linked to RP

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

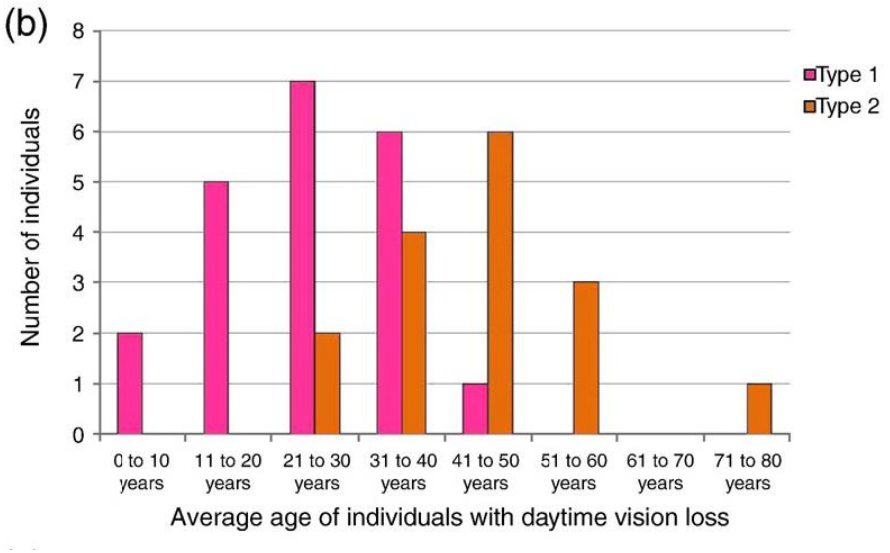
Energy changes

FoldX +



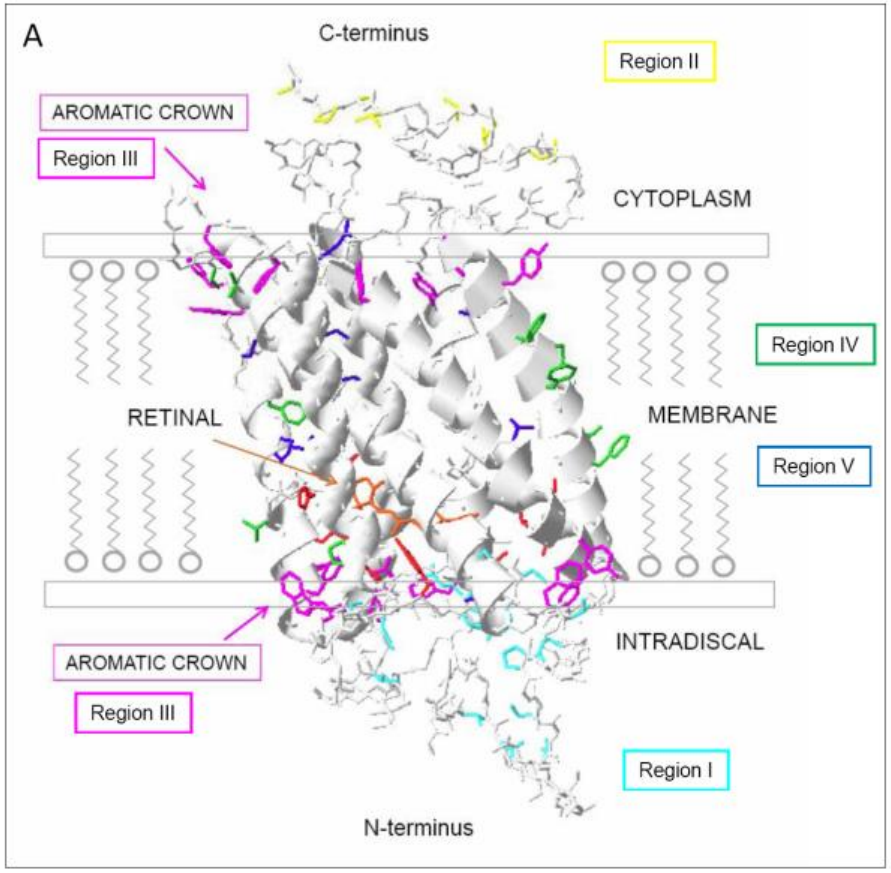
?

correlation



Rakoczy et al, J Mol Biol, 2011

Several consideration 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.

Rakoczy et al, J Mol Biol, 2011

For analyzing Region IV mutants (residues pointing outside and facing the lipid bilayer): use a different algorithm

Linking amino acid sequence to membrane insertion efficiency

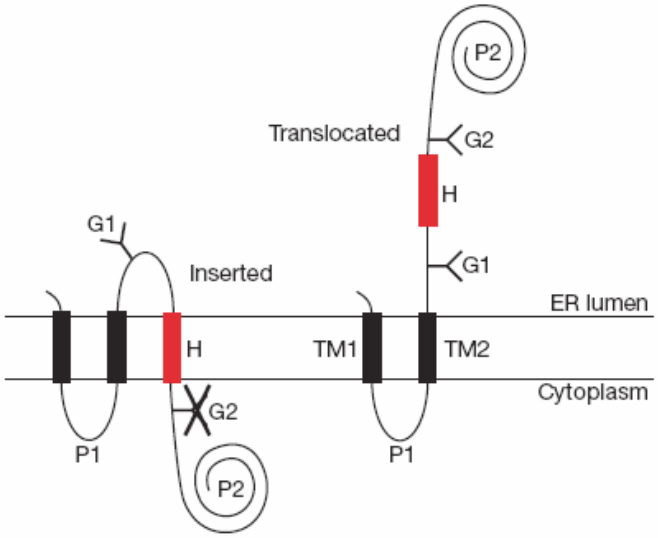
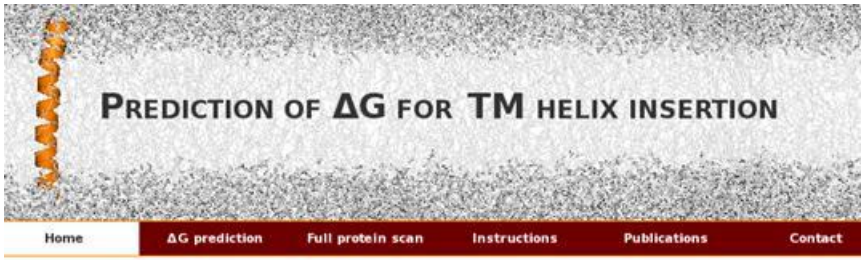


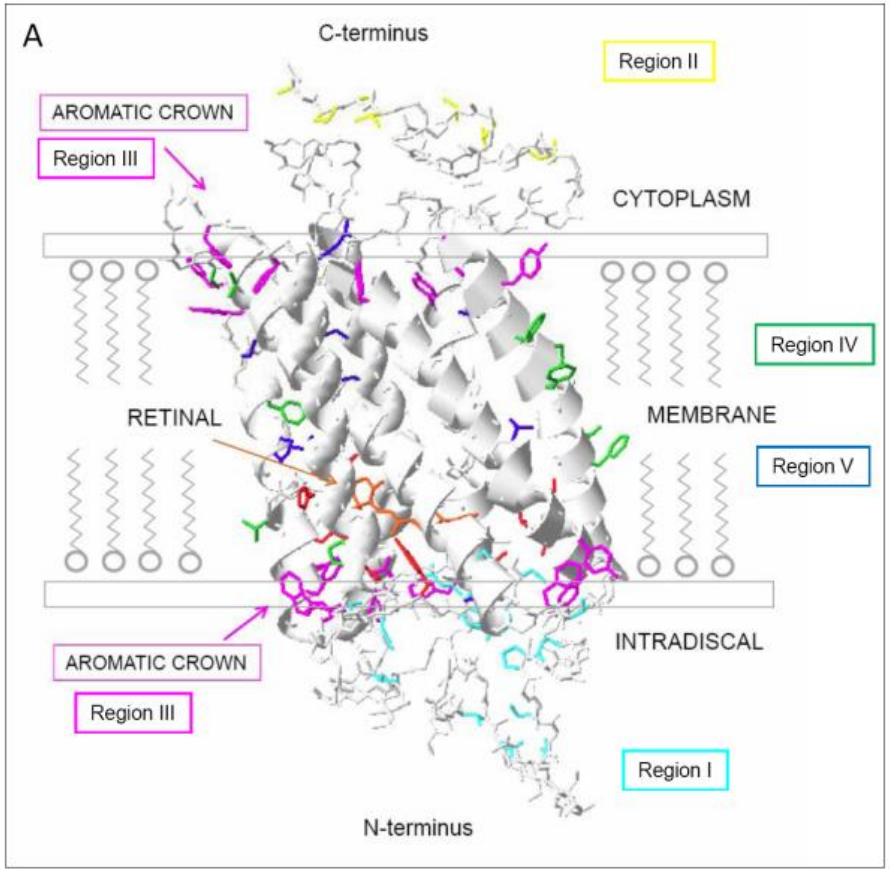
Figure 1 | The Lep model protein. *Escherichia coli* leader peptidase (Lep) has two TM helices (TM1 and TM2) and a large luminal domain (P2). It inserts into rough microsomes in an N_{lum}–C_{lum} orientation. H-segments (red) are engineered into the P2 domain with two flanking Asn-X-Thr glycosylation 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).

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.



Hessa/ von Heijne et al, Nature, 2007

Several consideration 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 consideration 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

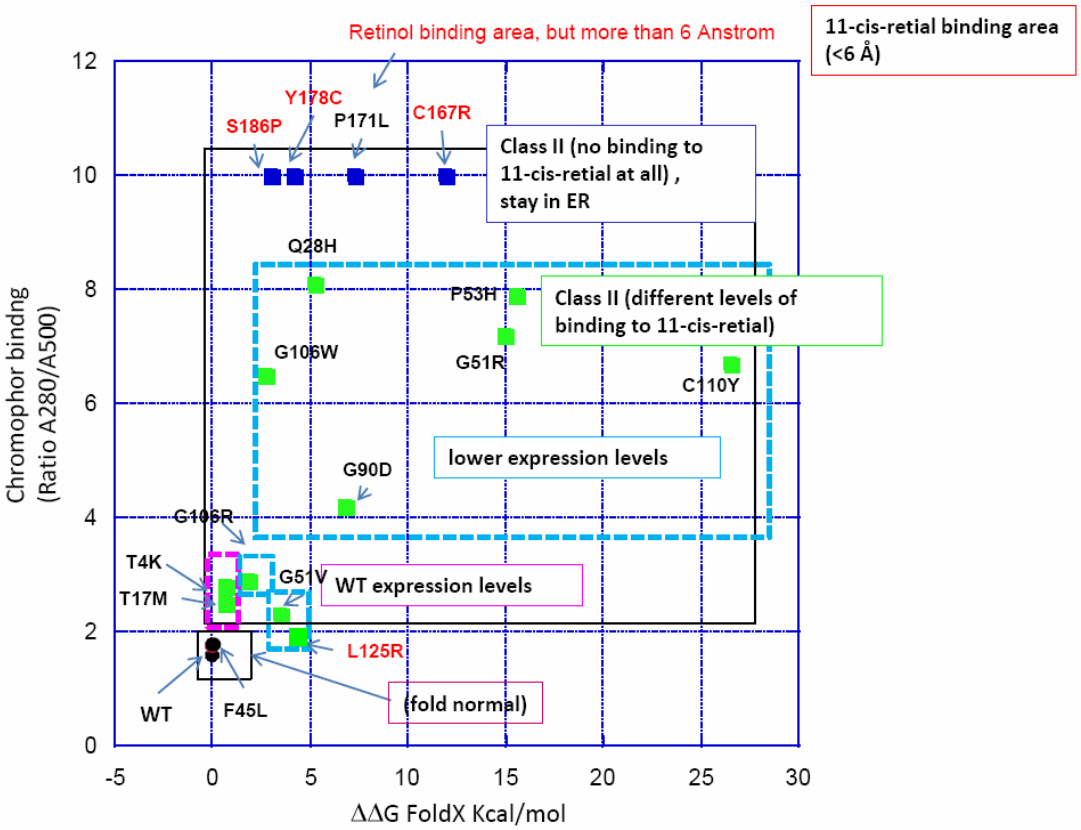
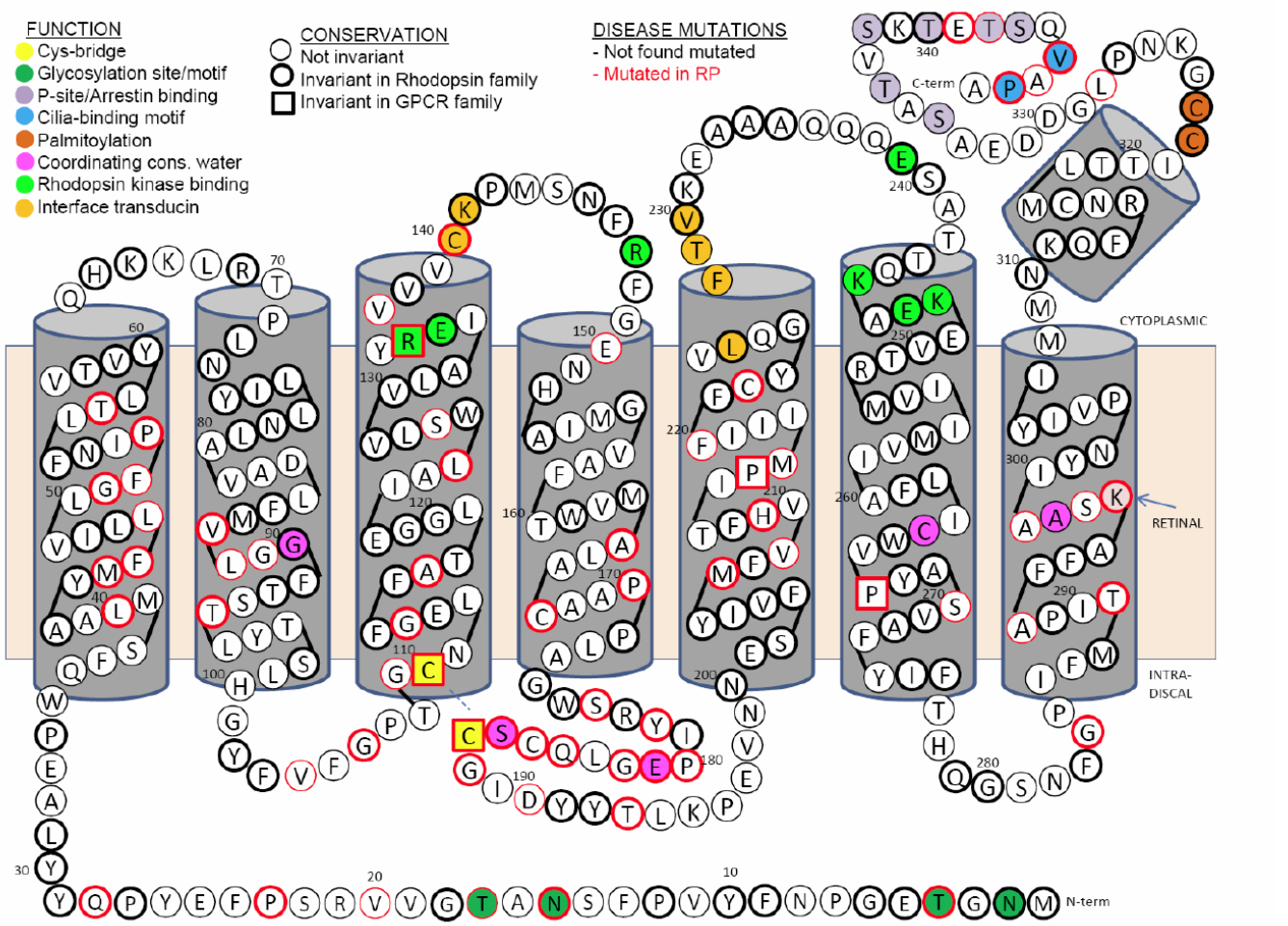


Figure S9 Correlation of FoldX ΔΔG values with in vitro expression levels and chromophore binding properties. In vitro data for protein expression levels and chromophore binding capabilities were taken from Kaushal and Khorana [19] and Briscoe et al. [49]. Involvement of mutants in different Classes (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).

Rakoczy et al, J Mol Biol, 2011

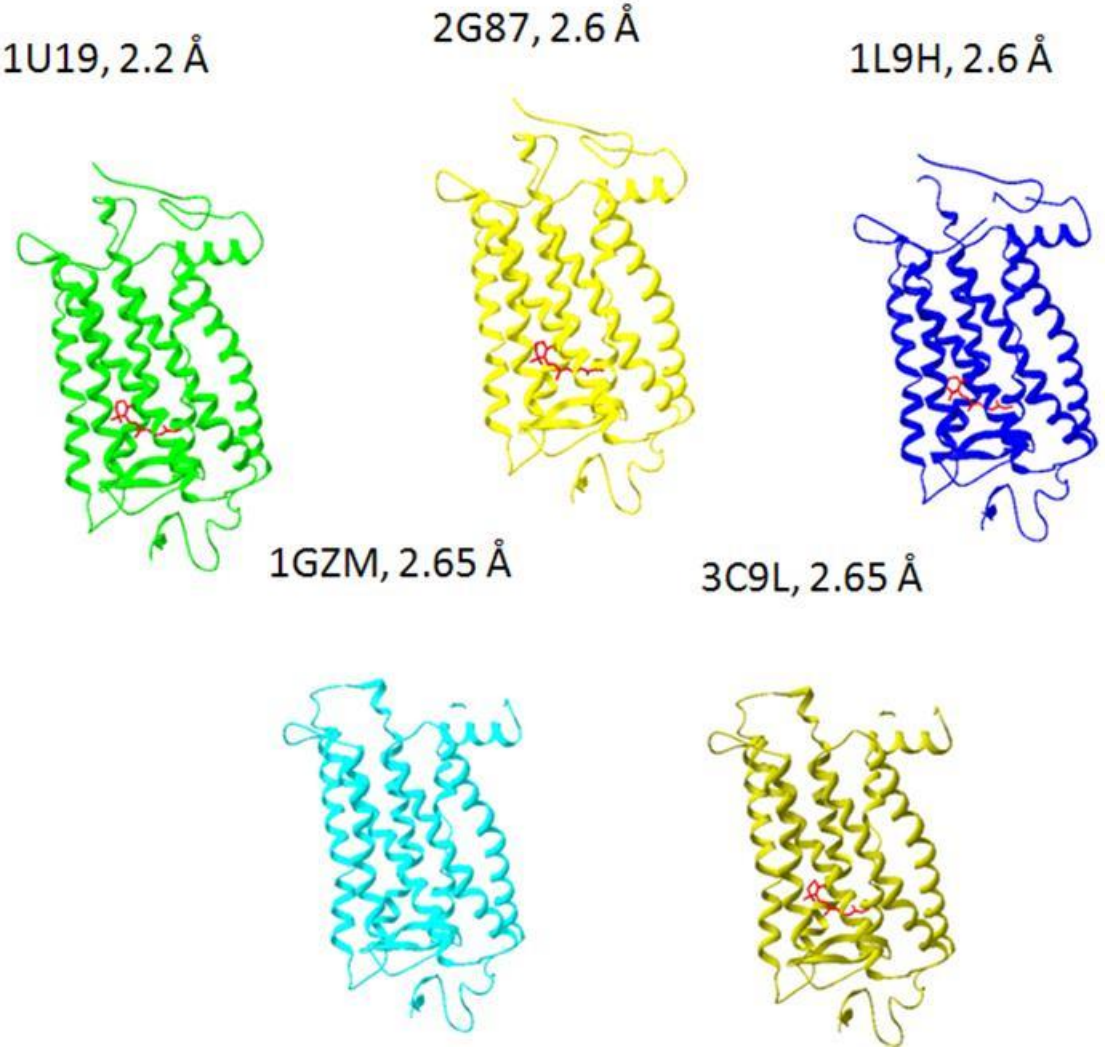
Several consideration for studying the effect of missense mutations in rhodopsin

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.



Rakoczy et al, J Mol Biol, 2011

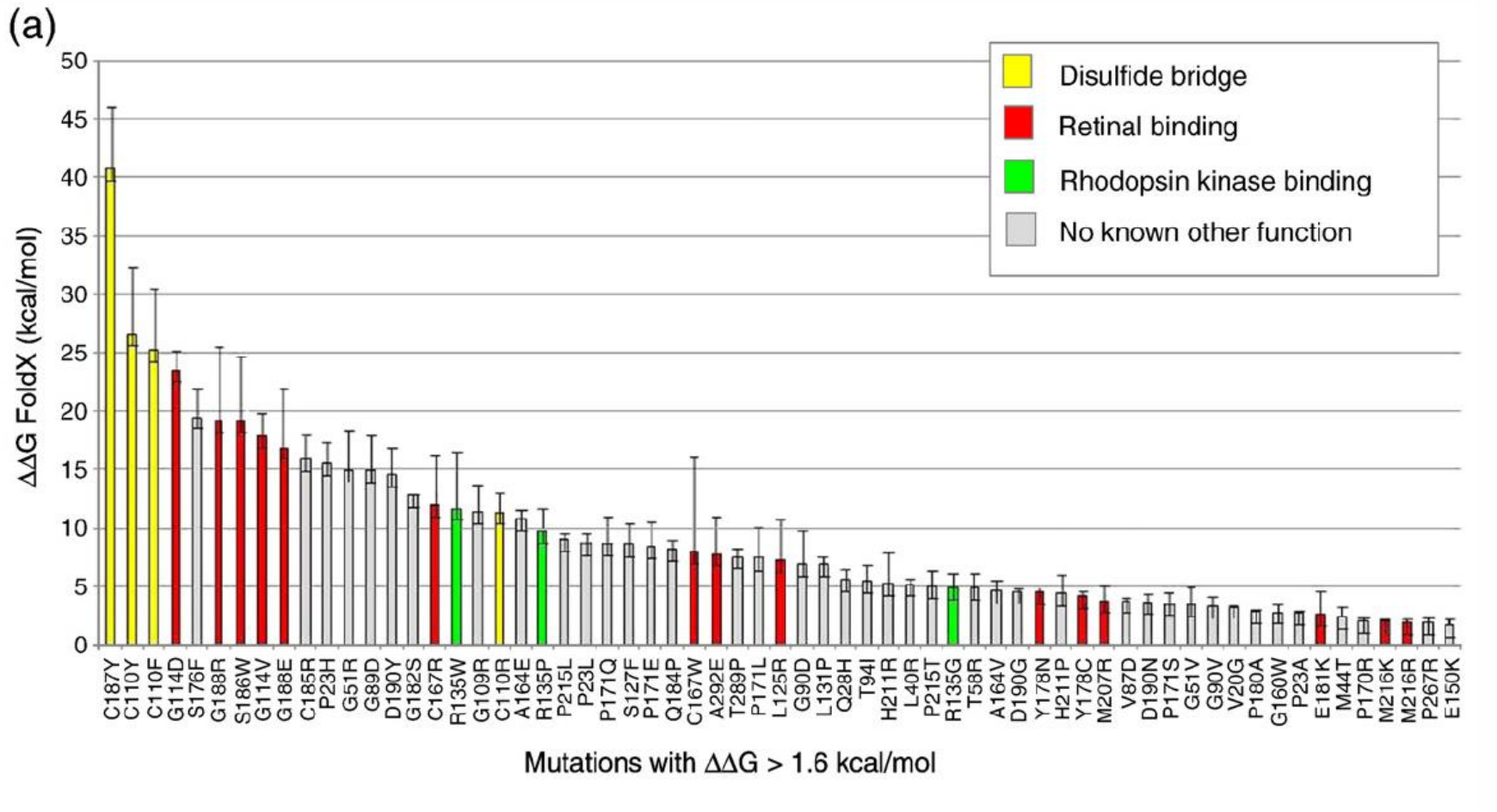
Five structures of bovine rhodopsin were selected (<2.6 Å) for mutagenesis and protein stability analysis using FoldX



Rakoczy et al, J Mol Biol, 2011

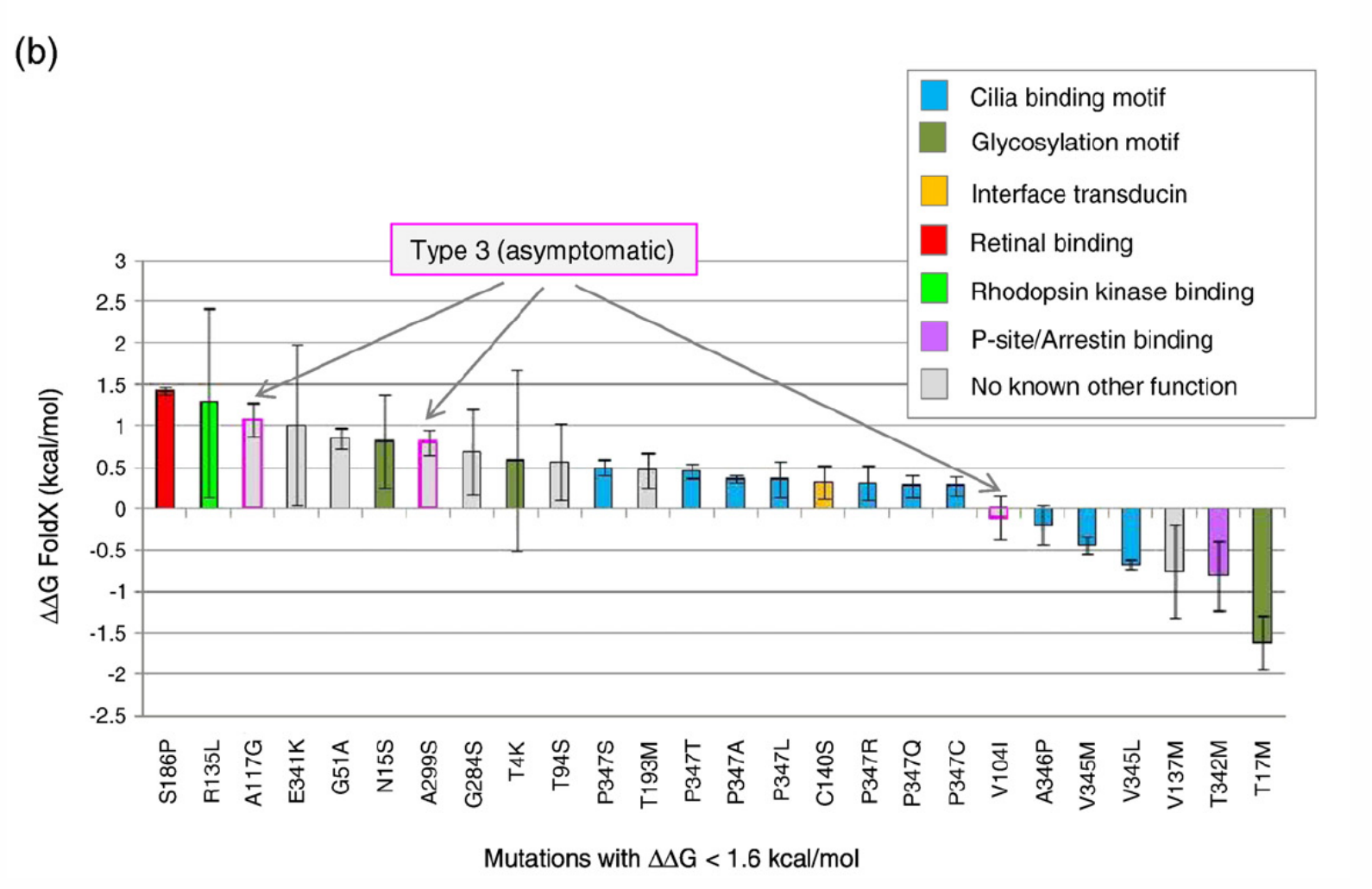
FoldX energy results and involvement in other function

Mutants that are destabilizing (ΔΔG > 1.6 kcal/ mol)



Rakoczy et al, J Mol Biol, 2011

FoldX energy results and involvement in other function



- 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

V20G	P23A	P23L	P23H	Q28H	L40R	M44T	G51R	G51V	T58R
V87D	G89D	G90D	G90V	T94I	G106R	G106W	G109R	C110Y	C110F
C110R	G114D	G114V	L125R	S127F	L131P	R135G	R135W	R135P	E150K
A164V	A164E	C167R	C167W	P170R	P171E	P171S	P171L	P171Q	S176F
Y178N	Y178C	P180A	E181K	G182S	Q184P	C185R	S186W	C187Y	G188R
G188E	D190N	D190G	D190Y	M207R	H211P	H211R	P215T	P215L	M216R (bovine: Leu)
M216K (bovine: Leu)	T289P	A292E							

Destabilization

FoldX >1.6kcal: 63

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

F45L	L46R	F52Y
P53R	F56Y	V209M
F220C	F220L	P267L
P267R	S270R (bovine: Gly)	S297R (bovine: Thr)

Total number of mutations: 103

Bad structure: 2

C222R	L328P

FoldX < 1.6kcal: 26

Functional residues:19

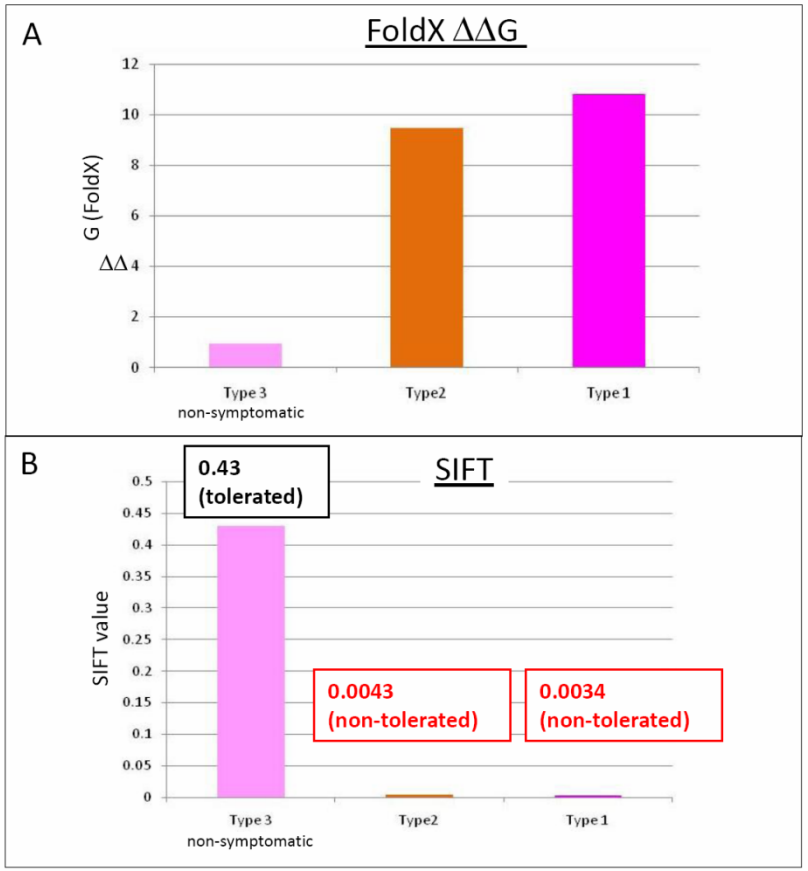
T4K	N15S	T17M
T94S	C140S	R185L
S186P	E341K	T342M
V345M	V345L	A346P
P347A	P347S	P347Q
P347L	P347R	P347T
P347C		

Unexplained/ misdiagnosed: 4

G51A	G284S
V137M	T193M

Asymptomatic No known function: 3

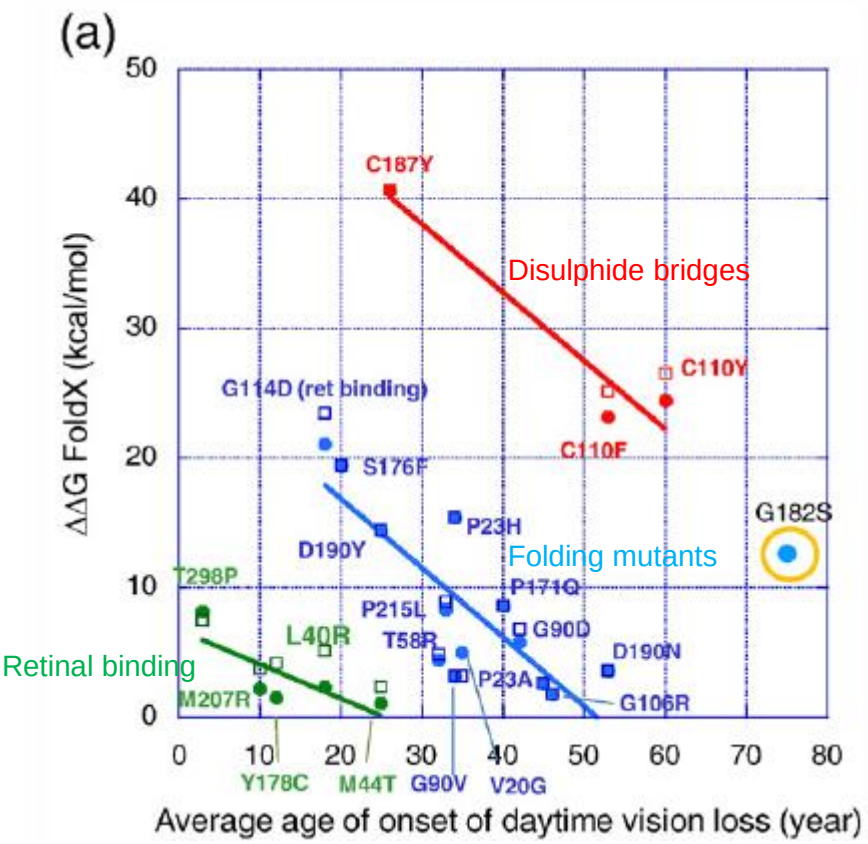
V104I	A117G	A299S

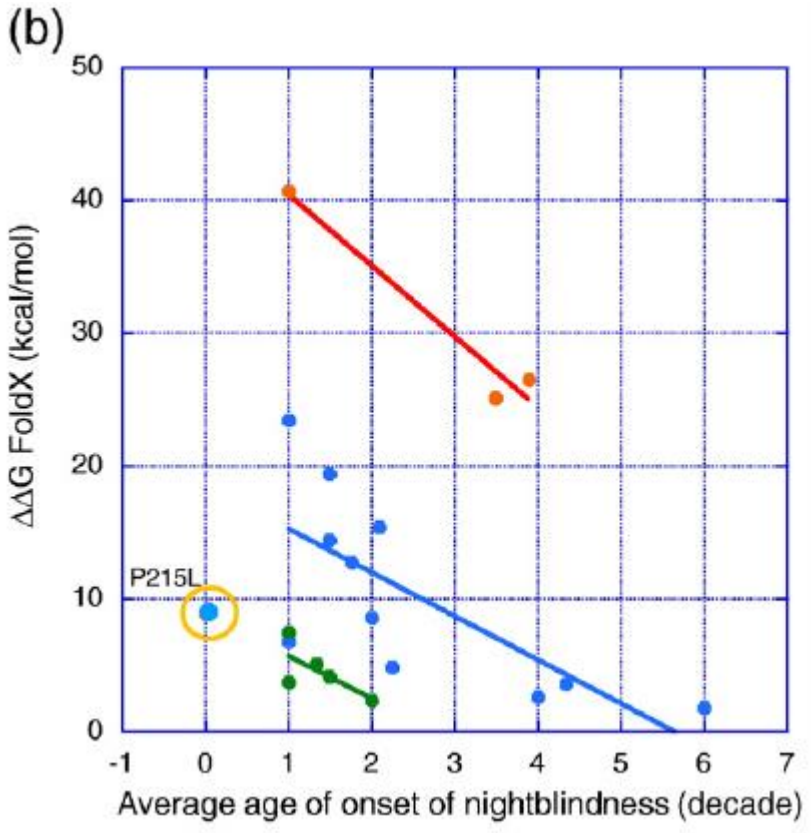


Rakoczy et al, J Mol Biol, 2011

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

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





Rakoczy et al, J Mol Biol, 2011

Conclusions example 2: Rhodopsin mutations

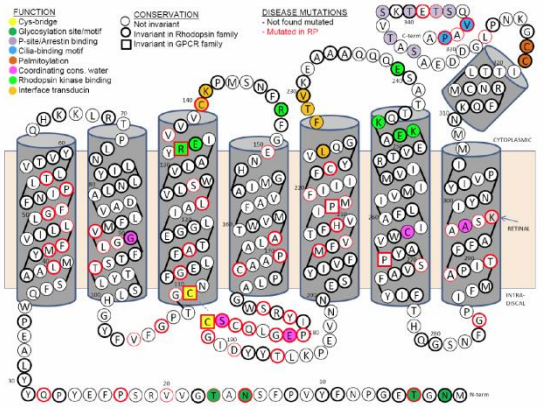
- The majority of the mutants is located within the hydrophobic core of the corresponding proteins and are therefore likely to cause misfolding.

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:

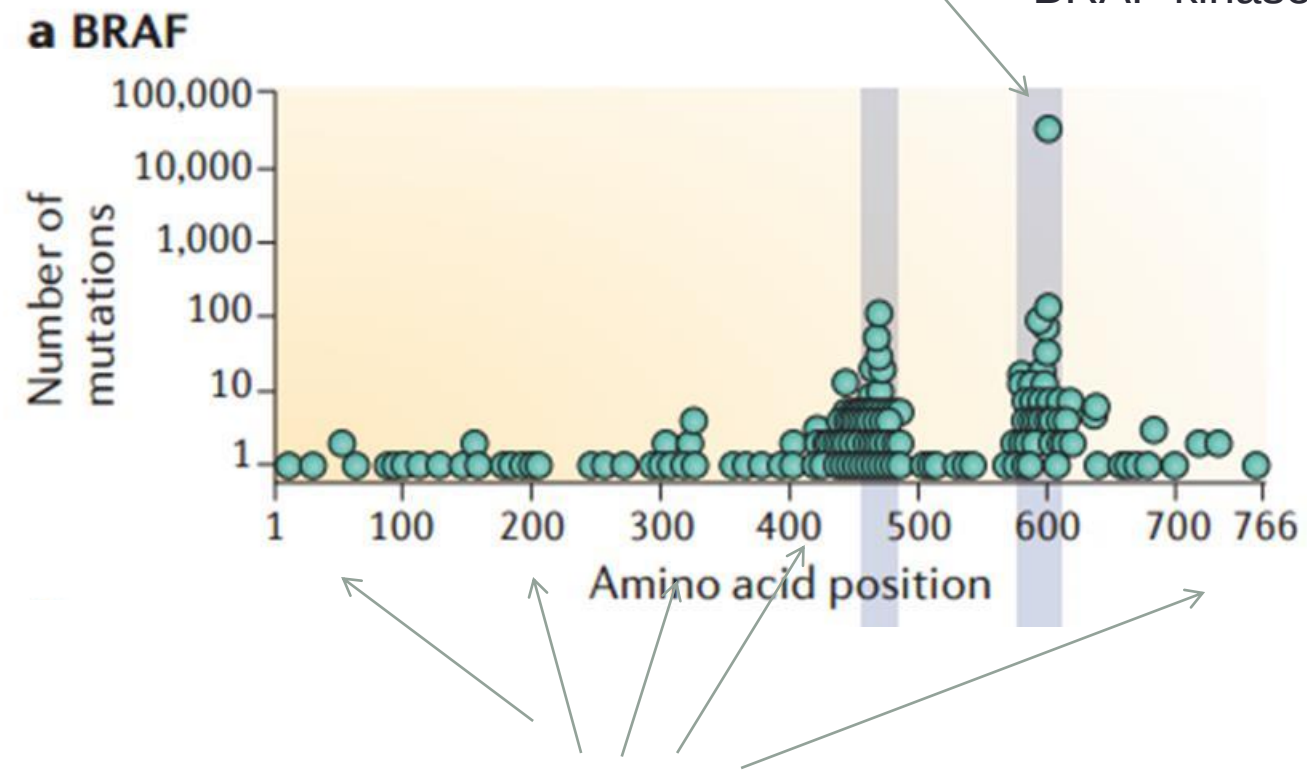
➢ a high level of functional understanding was necessary for our analysis and the observed energy-phenotype correlation.



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



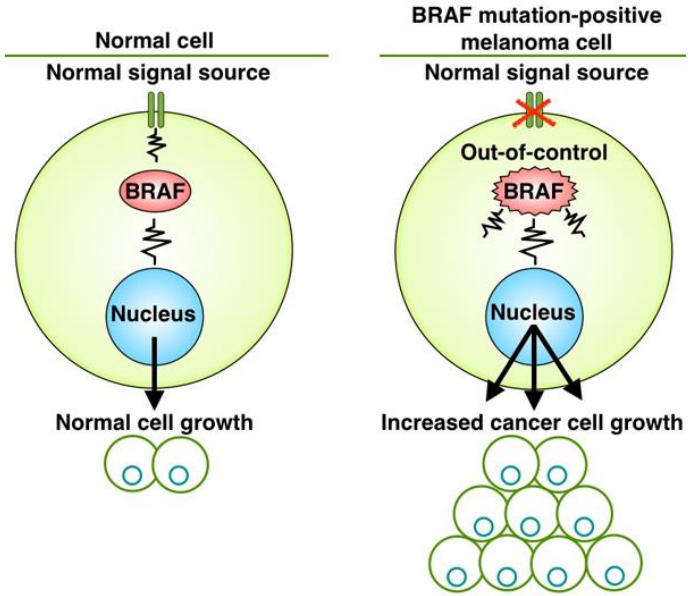
Shall we only treat patients which harbour V600E mutations or also patients with non-V600E mutations?

Catalytic activity of kinases is usually tightly controlled

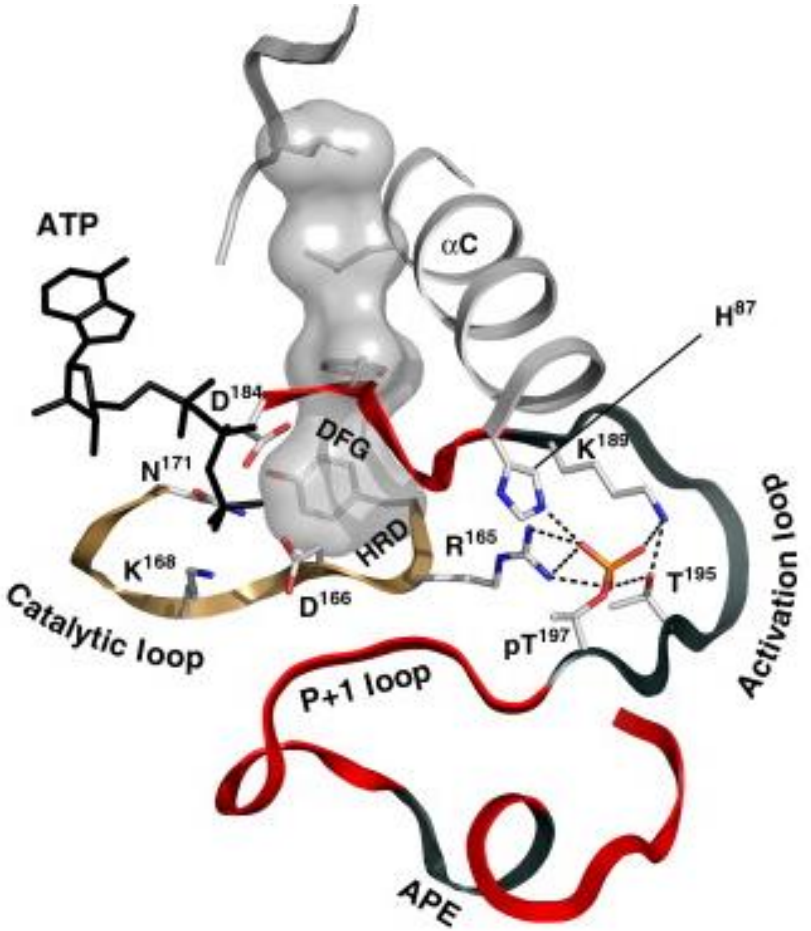
Mechanisms for kinase activation are:

- 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)

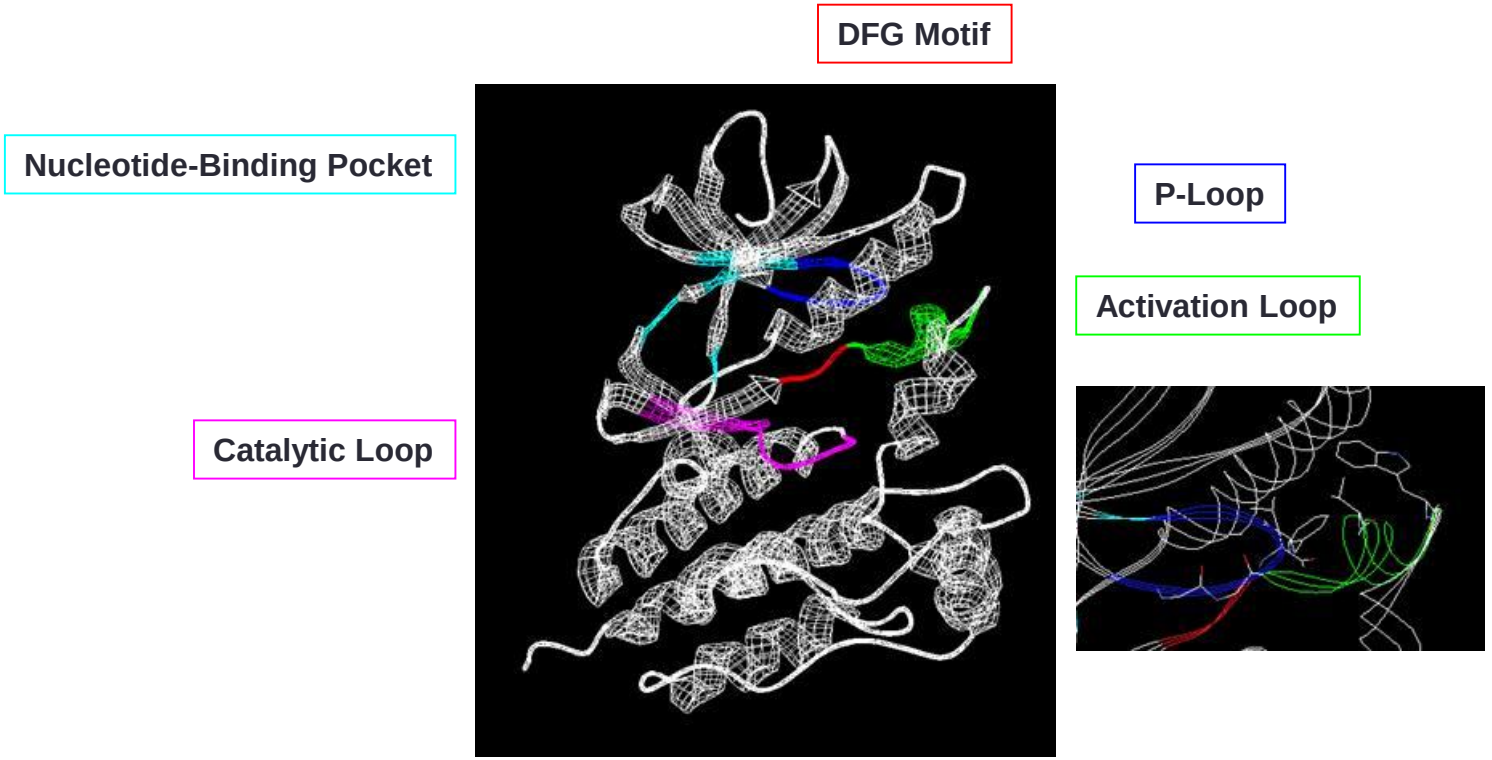


- phosphorylation in the activation segment causes structural **rearrangements of the activation segment and the aC helix**. This reorients the DFG loop resulting in activation of the kinase

Taylor & Kornev, TIBS, 2011

BRAF kinase activation though oncogenic mutations (e.g. V600E)

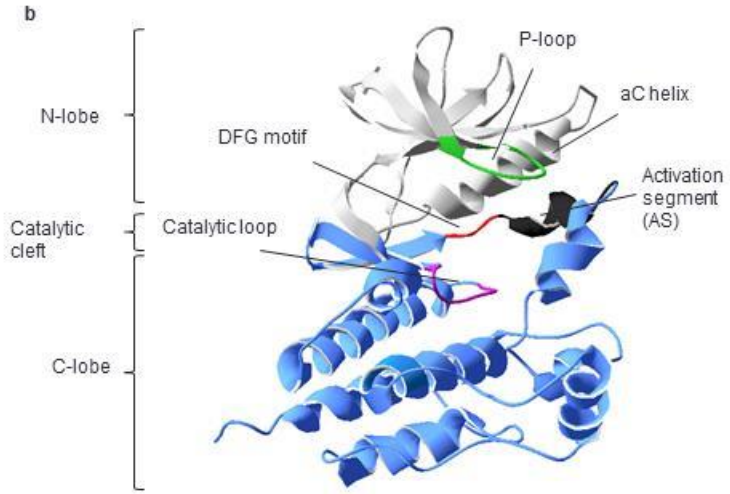
V600E mimics the negative charge of the neighbouring phosphorylated Thr599-P

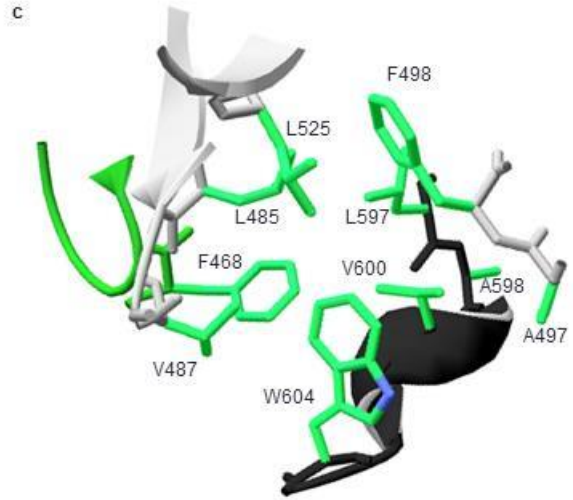


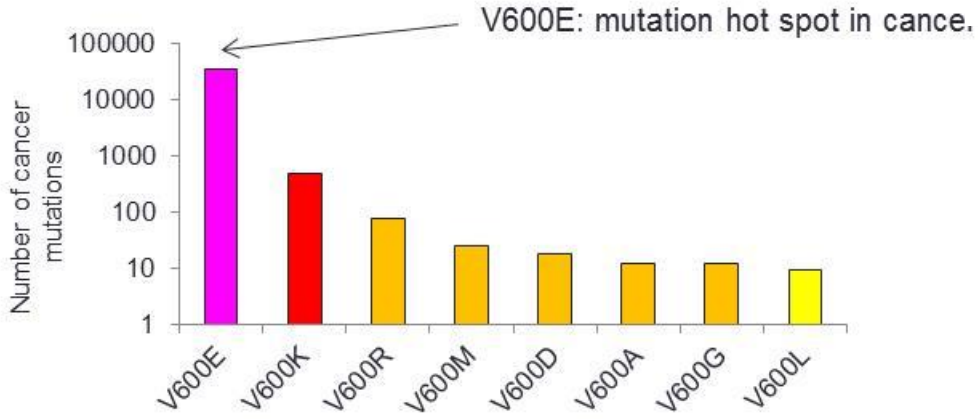
Activation loop residues: form **strong hydrophobic interactions with the P-loop in the inactive conformation of the kinase, locking the kinase in its inactive state** until the activation loop is phosphorylated, destabilizing these interactions with the presence of negative charge. This triggers the 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







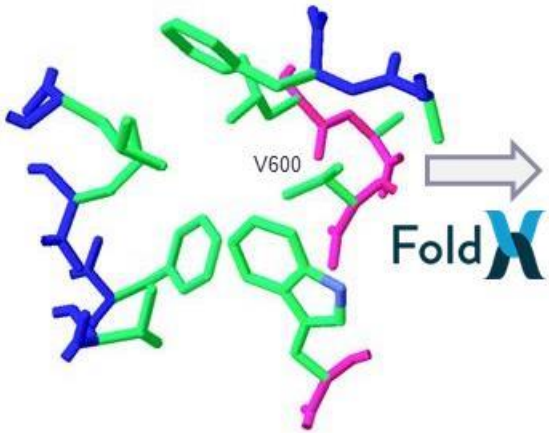
Differences in mutation frequencies: a quantitative effect?

Kiel et al, Elife, 2016

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

Total destabilization:

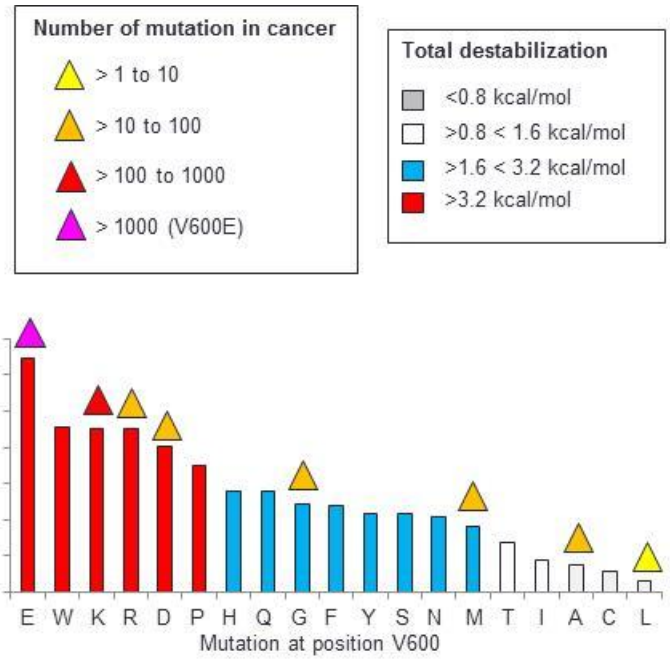
FoldX ΔΔG (BRAF_T)



A C D E F G H I K L M N P Q R S T V W Y

FoldX = ΔG

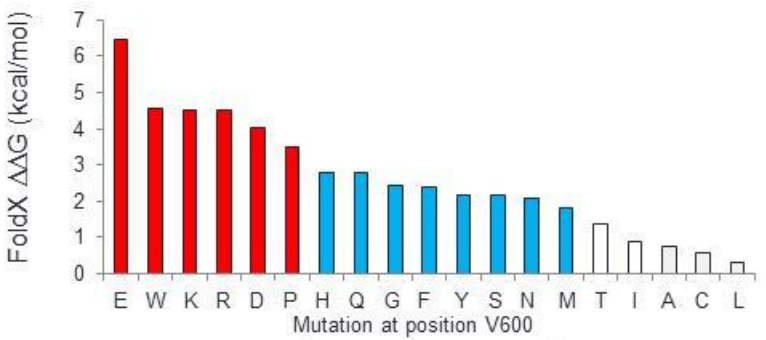
Destabilization of inactive state



No destabilization of active state (data not shown)

Kiel et al, Elife, 2016

Distinguishing driver from passenger mutations



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

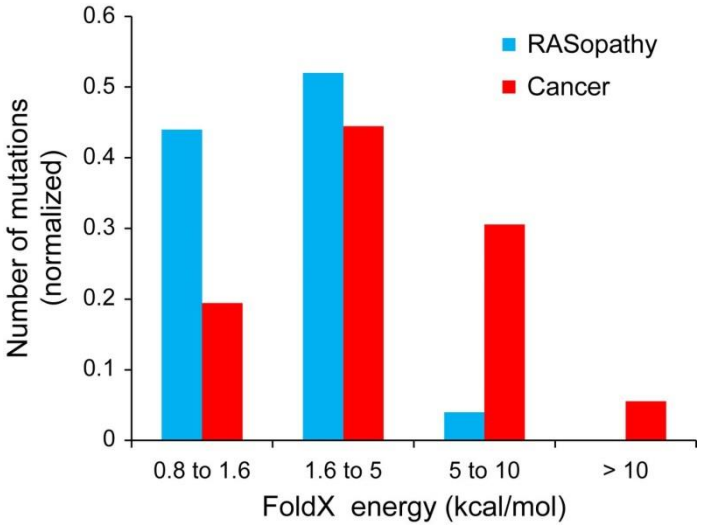
V600G behaves more like a RASopathy mutation

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, KJ[1]; Bunag, C[2]; Estep, AL[2]; Jones, JR[1]; Bolt, CH[1]; Rogers, RC[1]; Rauen, KA[3]; Everman, DB[1]

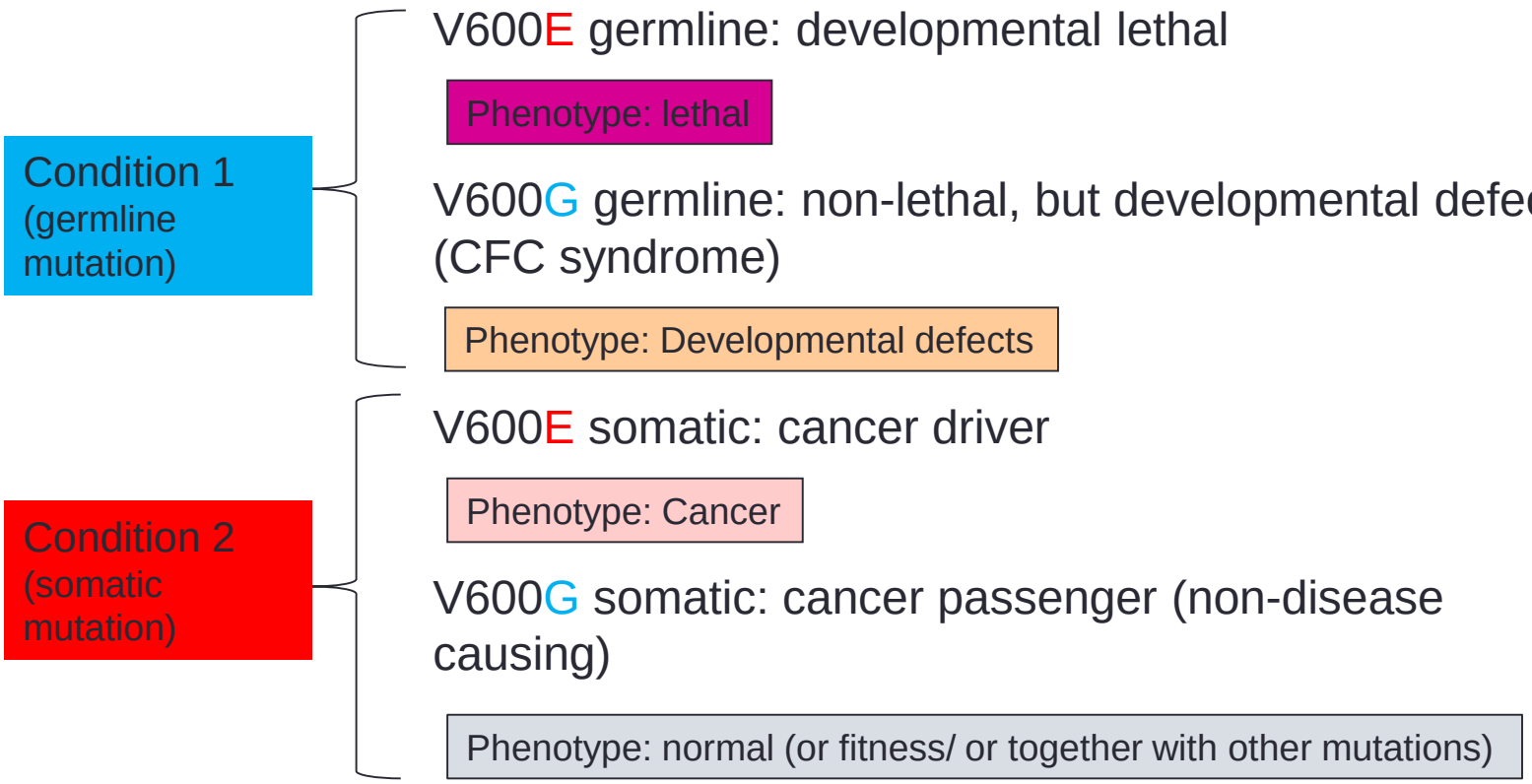
Clinical Genetics, Volume 79, issue 5 (May 2011), p. 468-474.
ISSN: 0009-9163 DOI: 10.1111/j.1399-0004.2010.01495.x
Blackwell Publishing Ltd



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

Kiel & Serrano, 2014

Different energy thresholds for germline and somatic mutations? 'Condition-dependent phenotypes'



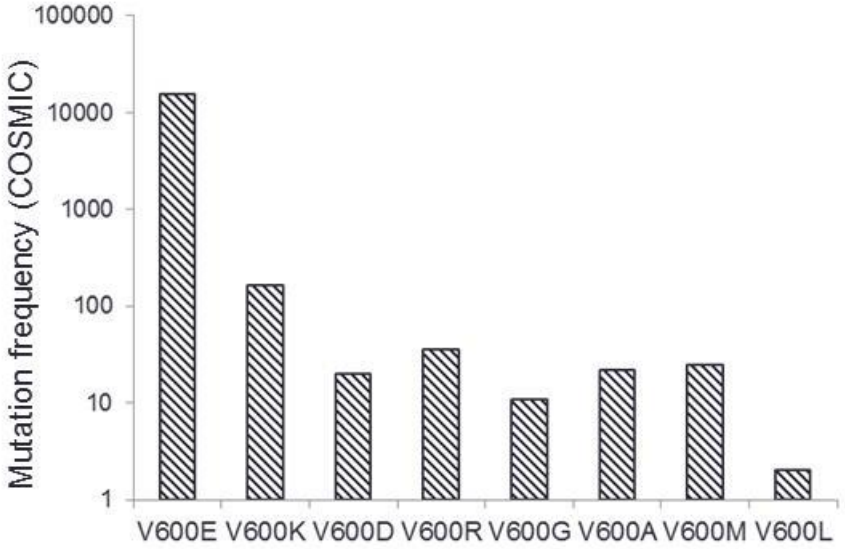
Why different cancer frequencies for V600E, V600D and V600K?

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

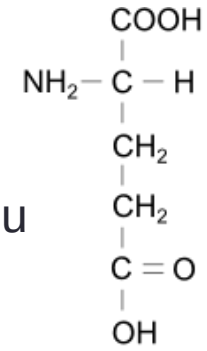
Why different cancer frequencies for V600E, V600D and V600K?

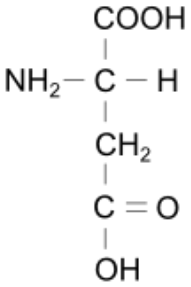


V600E: 15474 frequency
V600D: 20 frequency

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

Glu





Asp

Why different cancer frequencies for V600E, V600D and V600K?

1st letter	Second Letter: U	Second Letter: C	Second Letter: A	Second Letter: G	3rd letter
U	UUU Phe UUC UUA Leu UUG	UCU UCC Ser UCA UCG	UAU Tyr UAC UAA Stop UAG Stop	UGU Cys UGC UGA Stop UGG Trp	U C A G
C	CUU CUC Leu CUA CUG	CCU CCC Pro CCA CCG	CAU His CAC CAA Gln CAG	CGU CGC Arg CGA CGG	U C A G
A	AUU AUC Ile AUA AUG Met	ACU ACC Thr ACA ACG	AAU Asn AAC AAA Lys AAG	AGU Ser AGC AGA Arg AGG	U C A G
G	GUU GUC Val GUA GUG	GCU GCC Ala GCA GCG	GAU Asp GAC GAA Glu GAG	GGU GGC Gly GGA GGG	U C A G

V600E: GAG

V600K: AAG

V600R: AGG

V600D: GAC/T

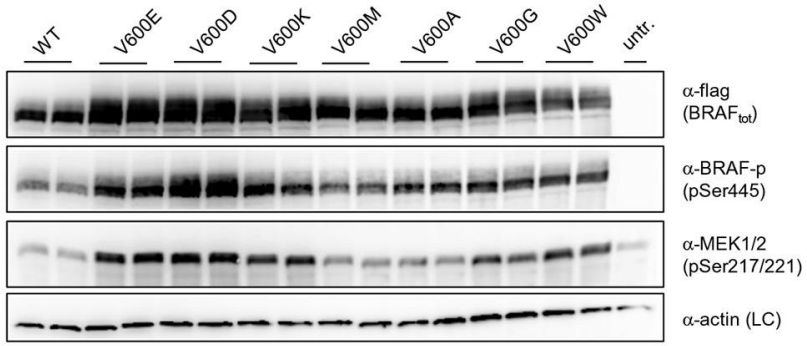
- The higher mutation frequency of V600E compared to V600D can be explained based on the number of nucleotide substitutions needed: V600D requires 2 nucleotide substitutions

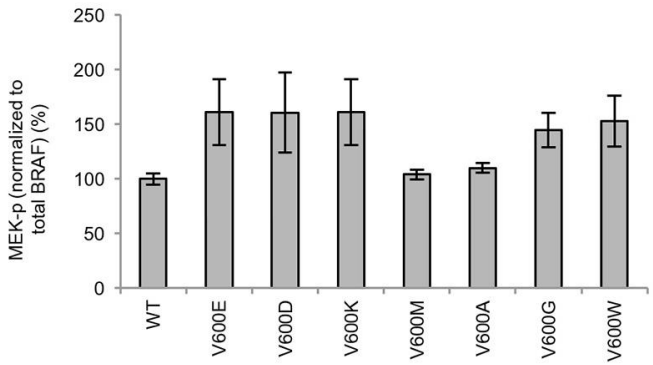
Experimentally validate the effect of BRAF mutations by monitoring downstream MEK activation (HEK293 cells)

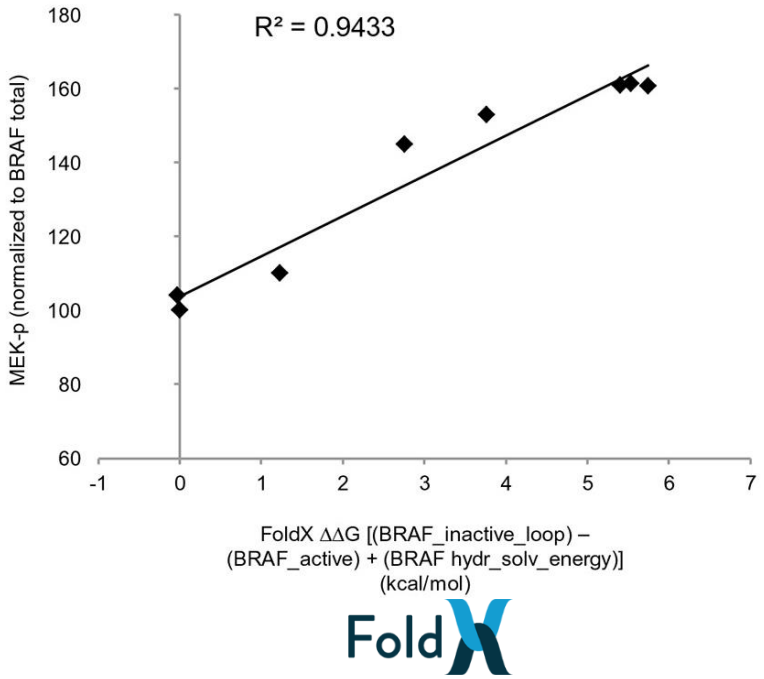
Day1:
Seed HEK293 cells

Day2:
Transfect flag-BRAF WT and mutants

Day3:
Lyse cells and Western blot

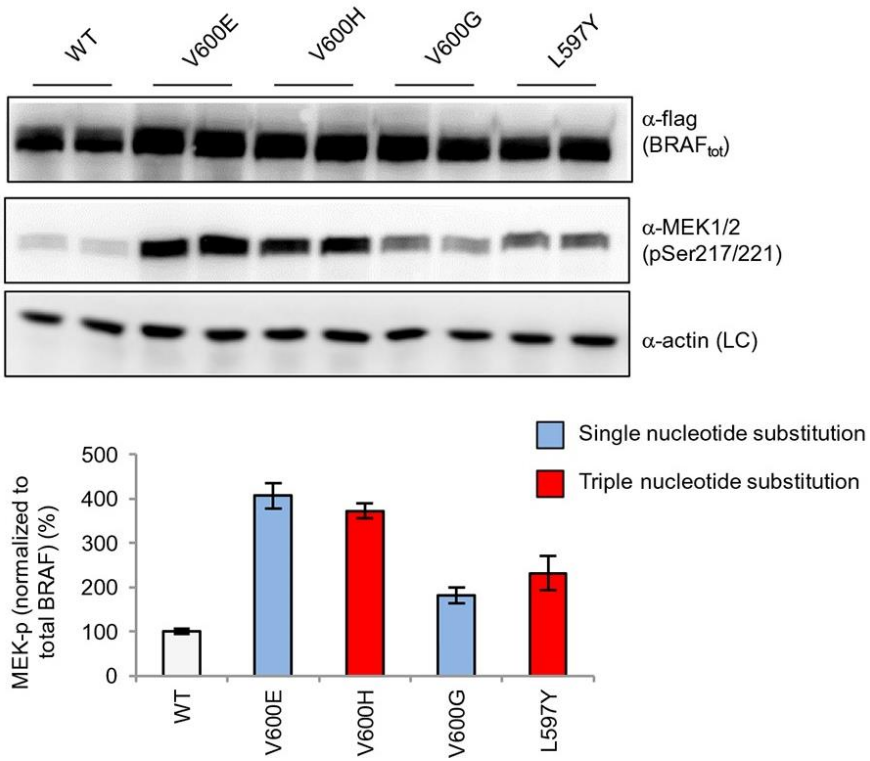






Kiel et al, Elife, 2016

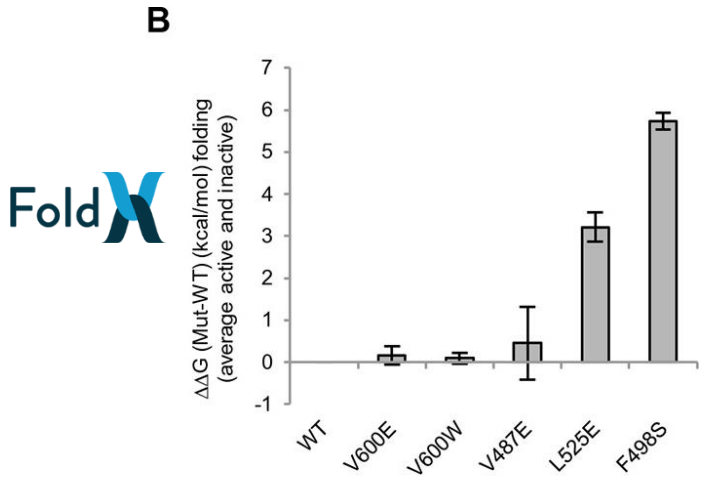
Experimentally validate the effect of BRAF mutations by monitoring downstream MEK activation (HEK293 cells)

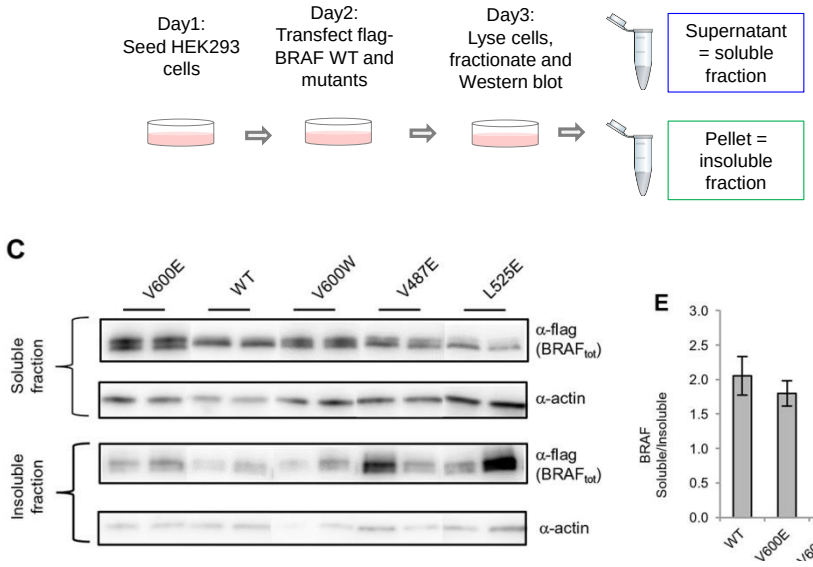


- V600H (requires 3 nucleotide substitutions) is as active as V600E, but NOT found in cancer. Similarly L597Y is not found in rasopathy patients.

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

EXCELENCIA SEVERO OCHOA

CRG Centre for Genomic Regulation

SynSignal

PRIMES Protein Interaction Machines in Oncogenic EGF Receptor Signaling

erc

GOBIERNO DE ESPAÑA MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

Serrano lab

Luis Serrano

Cell signaling team

Violeta Beltran
Hannah Benisty
Tony Ferrar
Jae-Seong Yang
Martin Schaefer
Javier Delgado
Kiana Toufighi
Besray Unal

CRG core facilities

Proteomics Core Facility

Advanced Light Microscopy Unit

Protein Screening

FACS Core Facility

Bioinformatics Core Facility

Collaborations cell signaling team:

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

Edda Klipp, Friedemann Uschner (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

erc 2009-2014

miniCELL

MYCOSYNVAC

erc

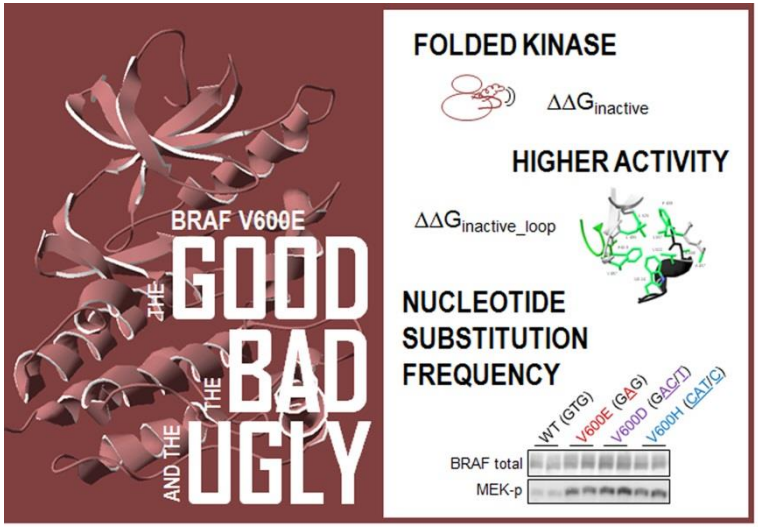
Conclusions example 3: Why BRAF V600E?

- 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.

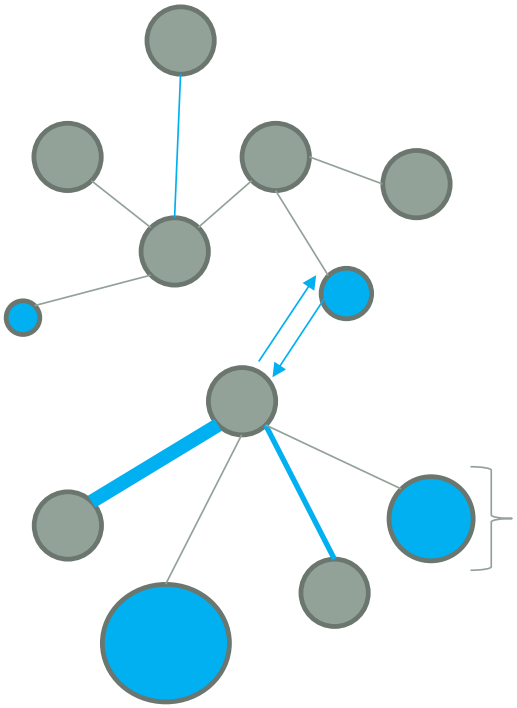
Why BRAF V600E?

- 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

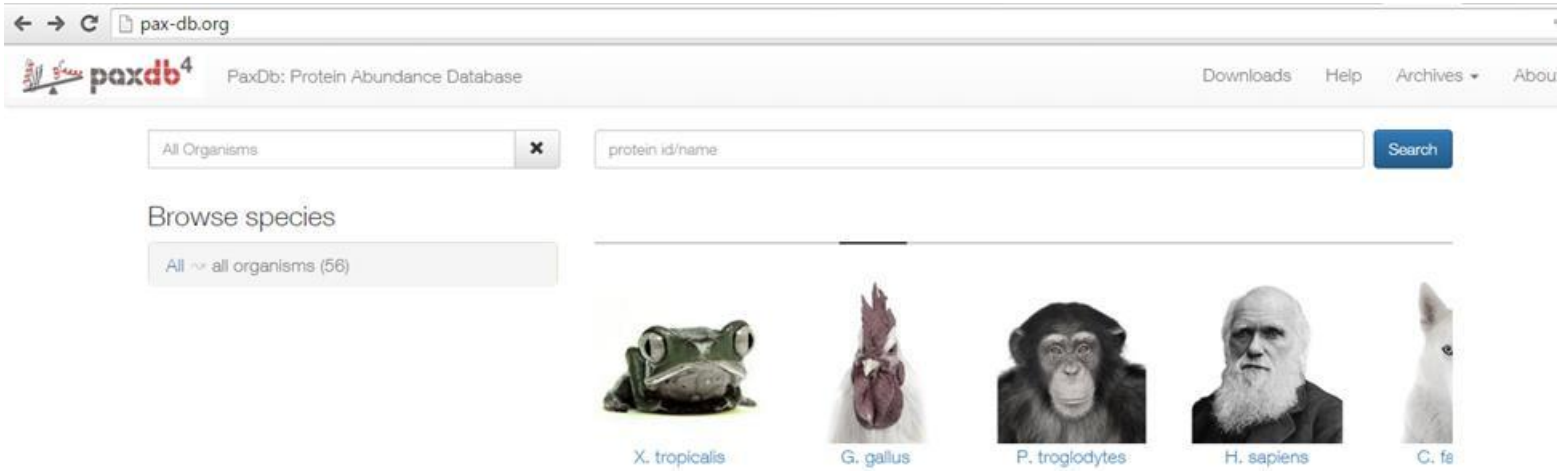
BindingDB

dsysmap

paxdb4

BIONUMB3RS THE DATABASE OF USEFUL BIOLOGICAL NUMBERS

Protein abundances



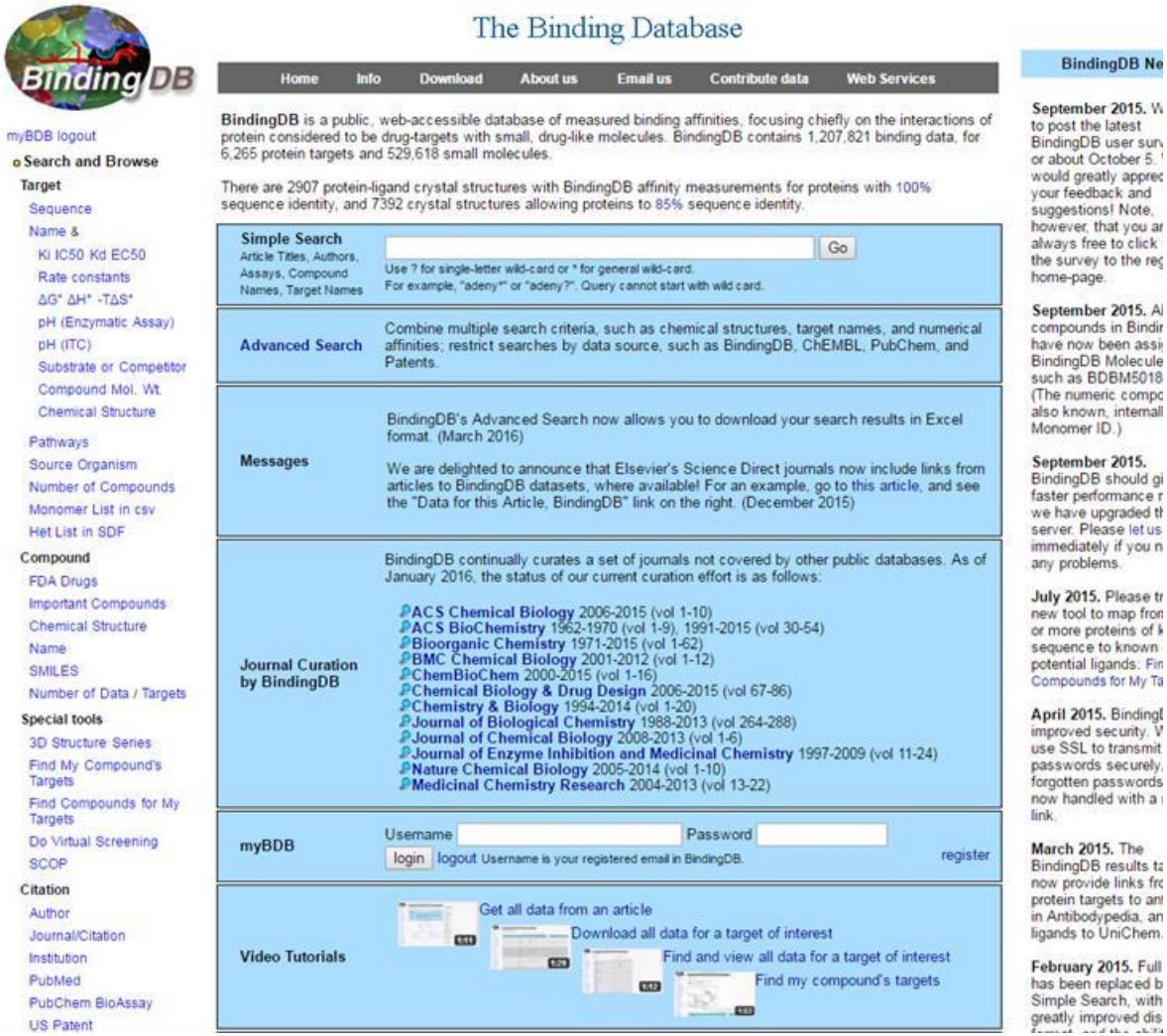
Data Overview

species name

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

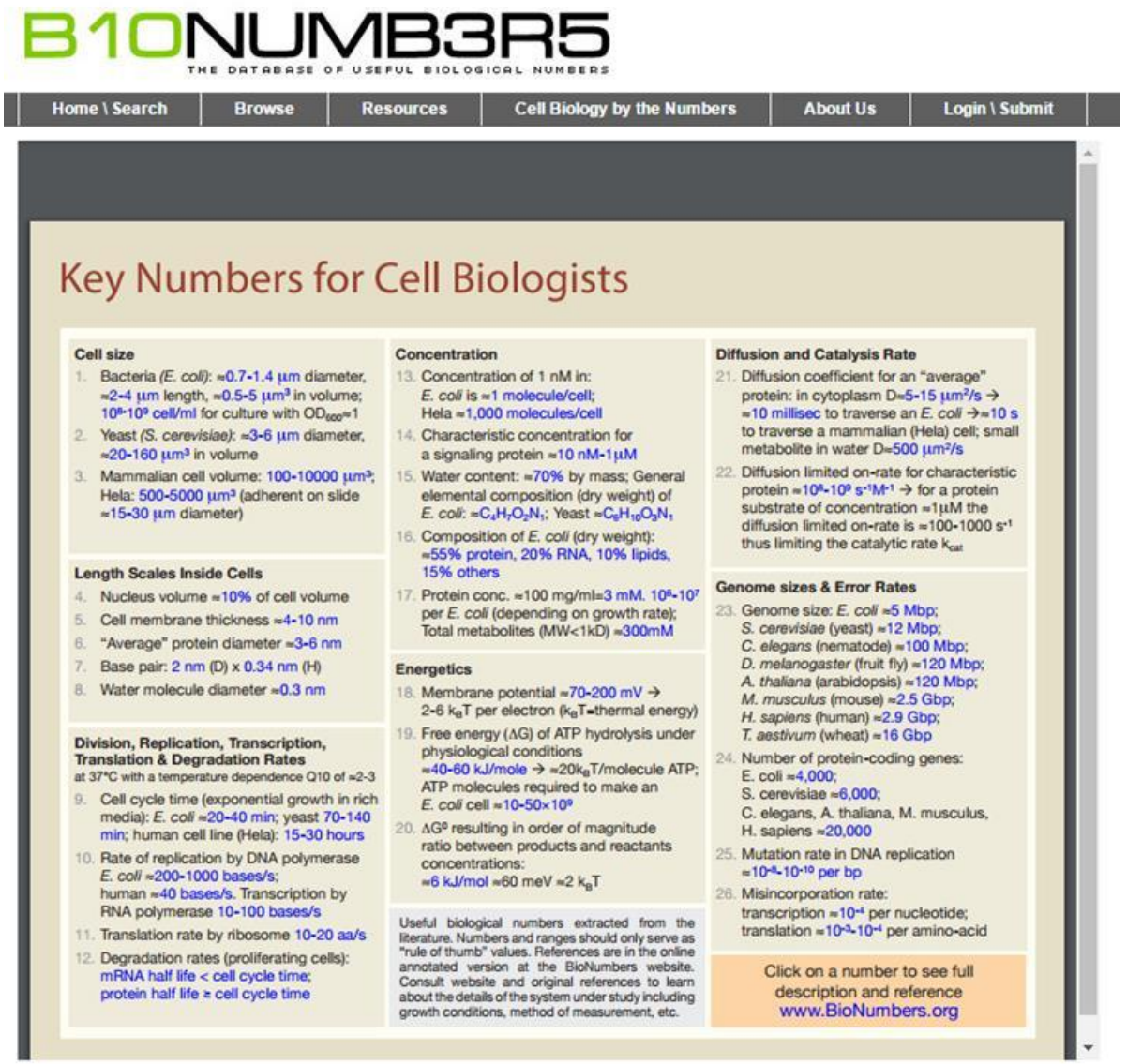
http://pax-db.org/

Affinities and kinetic constants



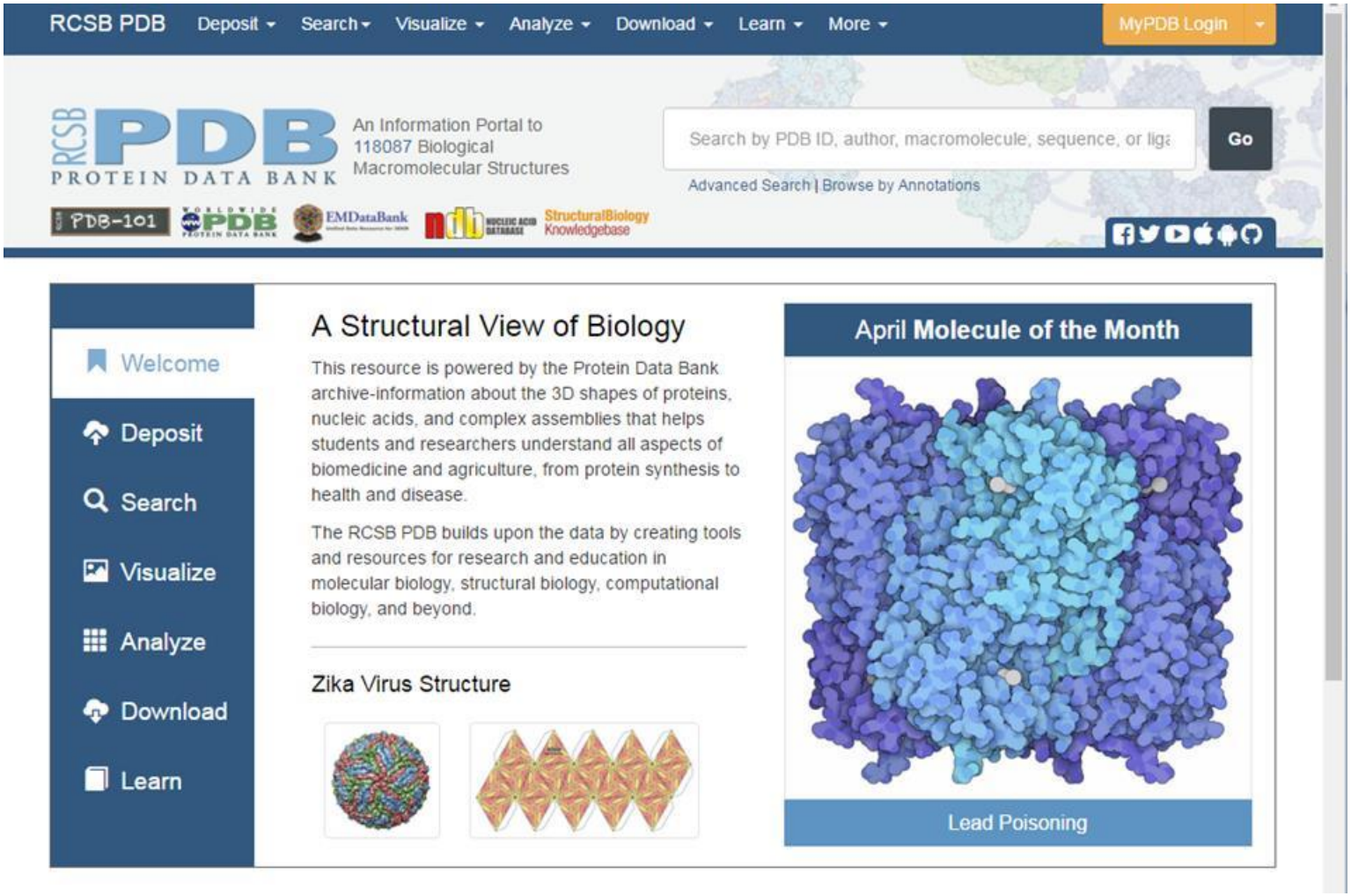
https://www.bindingdb.org/bind/index.jsp

General 'numbers' in biology



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

Protein structures

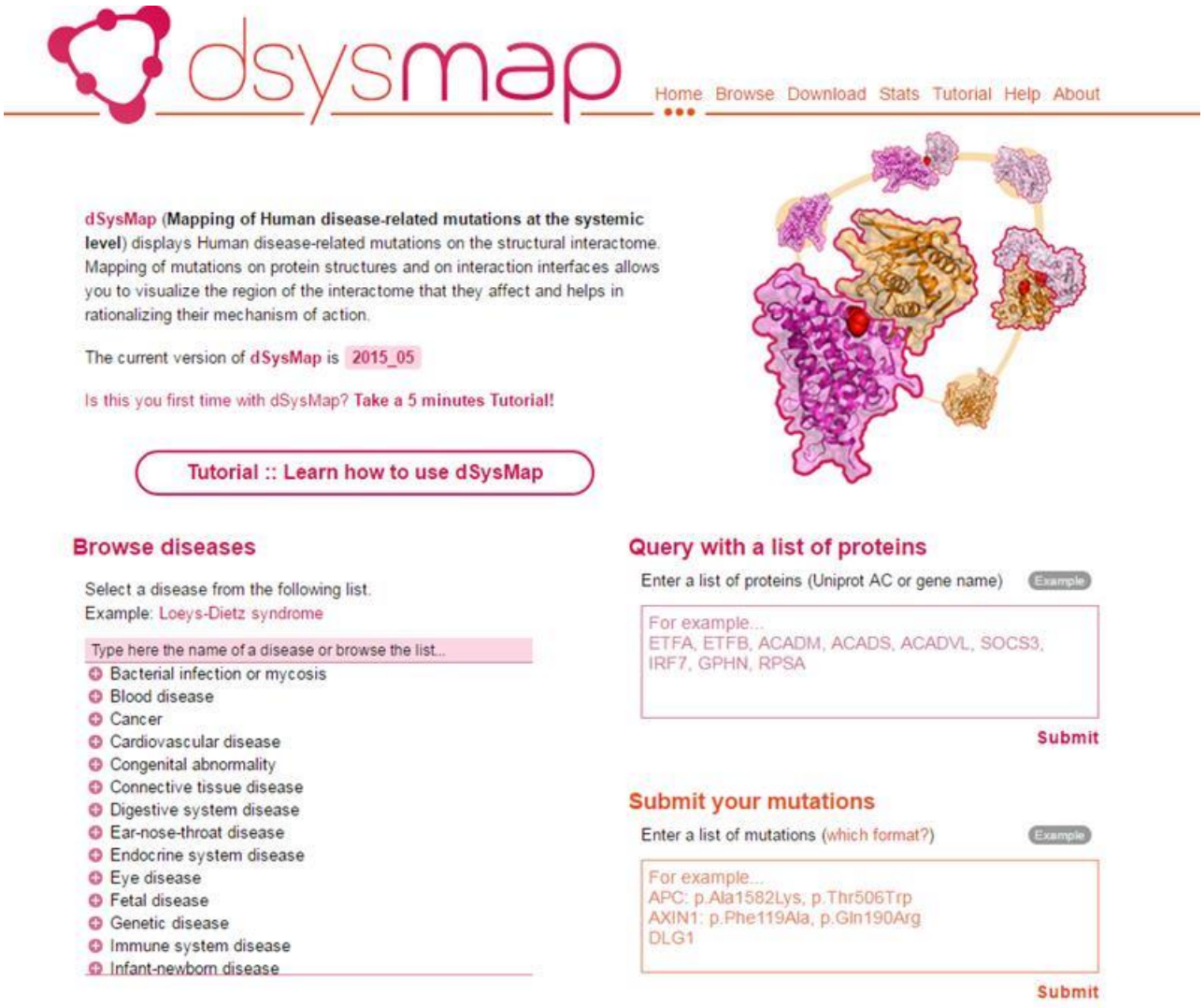


http://www.rcsb.org/pdb/home/home.do

3D structures of protein interactions

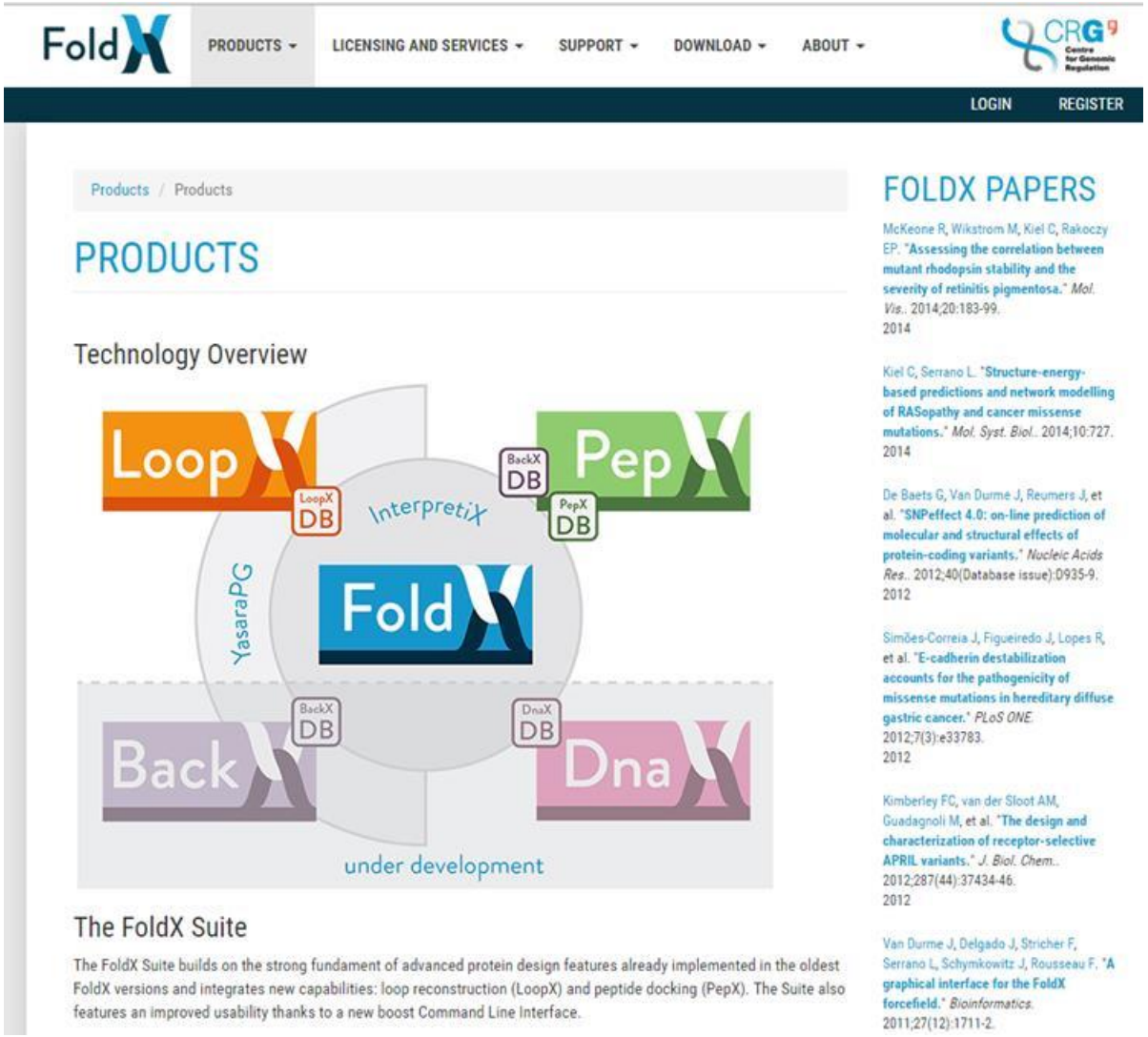
http://interactome3d.irbbarcelona.org/

3D structures of protein interactions/ mapping of disease mutations



http://dsysmap.irbbarcelona.org/

Protein design



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

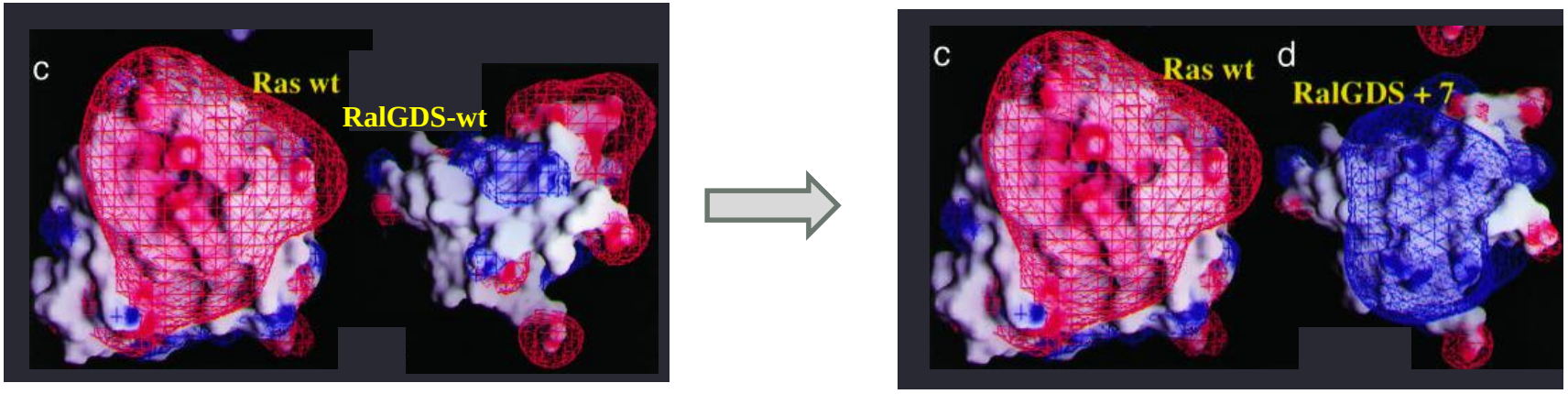
Affinity (Dissociation constant) → $K_d = \frac{k_{off}}{k_{on}}$

k_{off}: Dissociation rate constant

k_{on}: Association rate constant

E.g.:

↑ Increase k_{on}: improve electrostatic surface complementarity; *'electrostatic steering'*



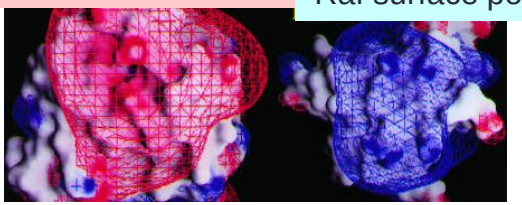
Kiel et al., PNAS, 2004

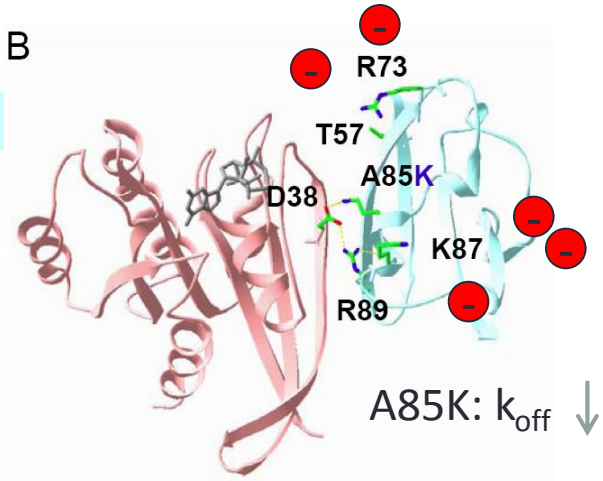
↑ Increase k_{off}: mutate hot-spot residues in the interface

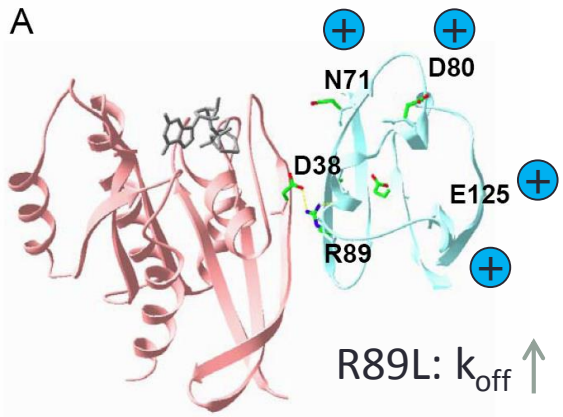
Summary of the protein mutant design

Ras surface negative

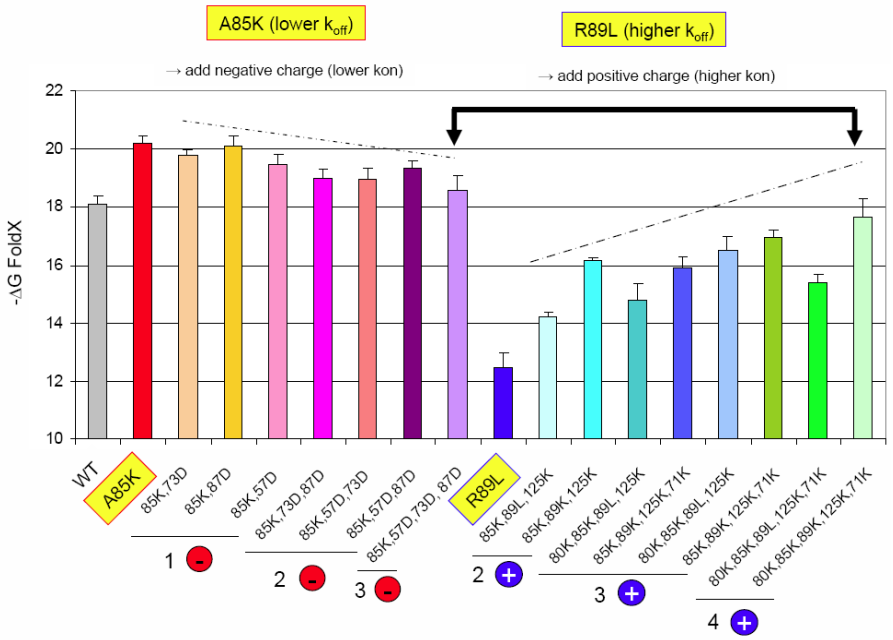
Raf surface positive





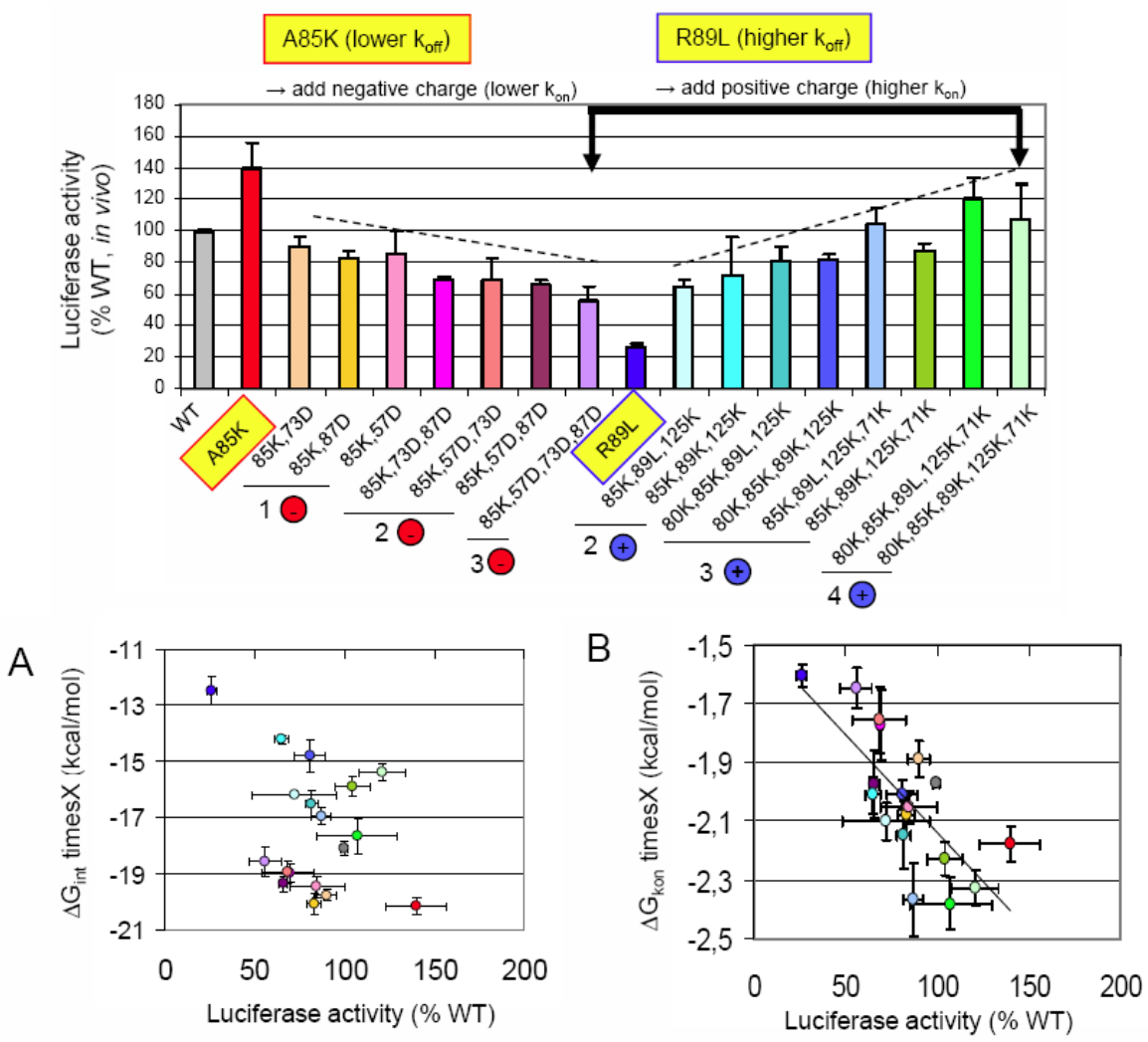


FoldX



Kiel & Serrano, Sci Signal, 2009

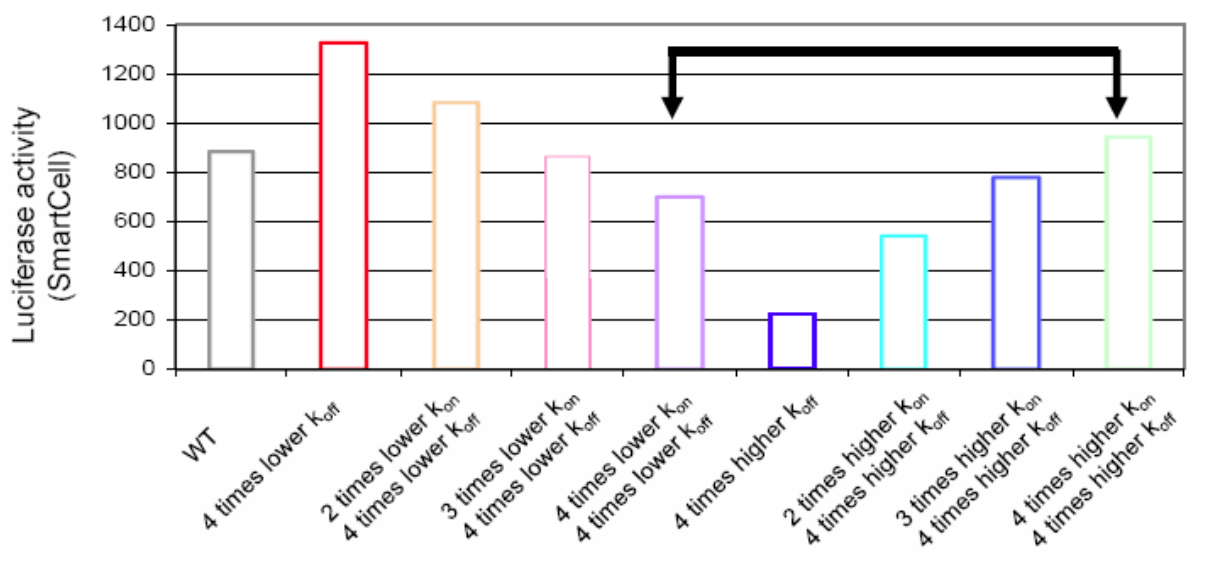
Analysis of all mutants in RK13 cells (luciferase activity assay)



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

Kiel & Serrano, Sci Signal, 2009

Results from 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

Kiel & Serrano, Sci Signal, 2009