EMBO Practical Course

Computational analysis of protein-protein interactions in cell function and disease

01 – 06 December 2019 | Bangalore, India

EMBO PPI Bioinformatics Course
Bangalore, 3-12-19

Modular Protein Architecture and the Construction of Cell Regulatory Systems: From Health to Disease

Toby J. Gibson
Structural & Computational Biology Unit
EMBL, Heidelberg

Biology is an information science: Bioinformatics is now very often a part of interdisciplinary biomedical studies

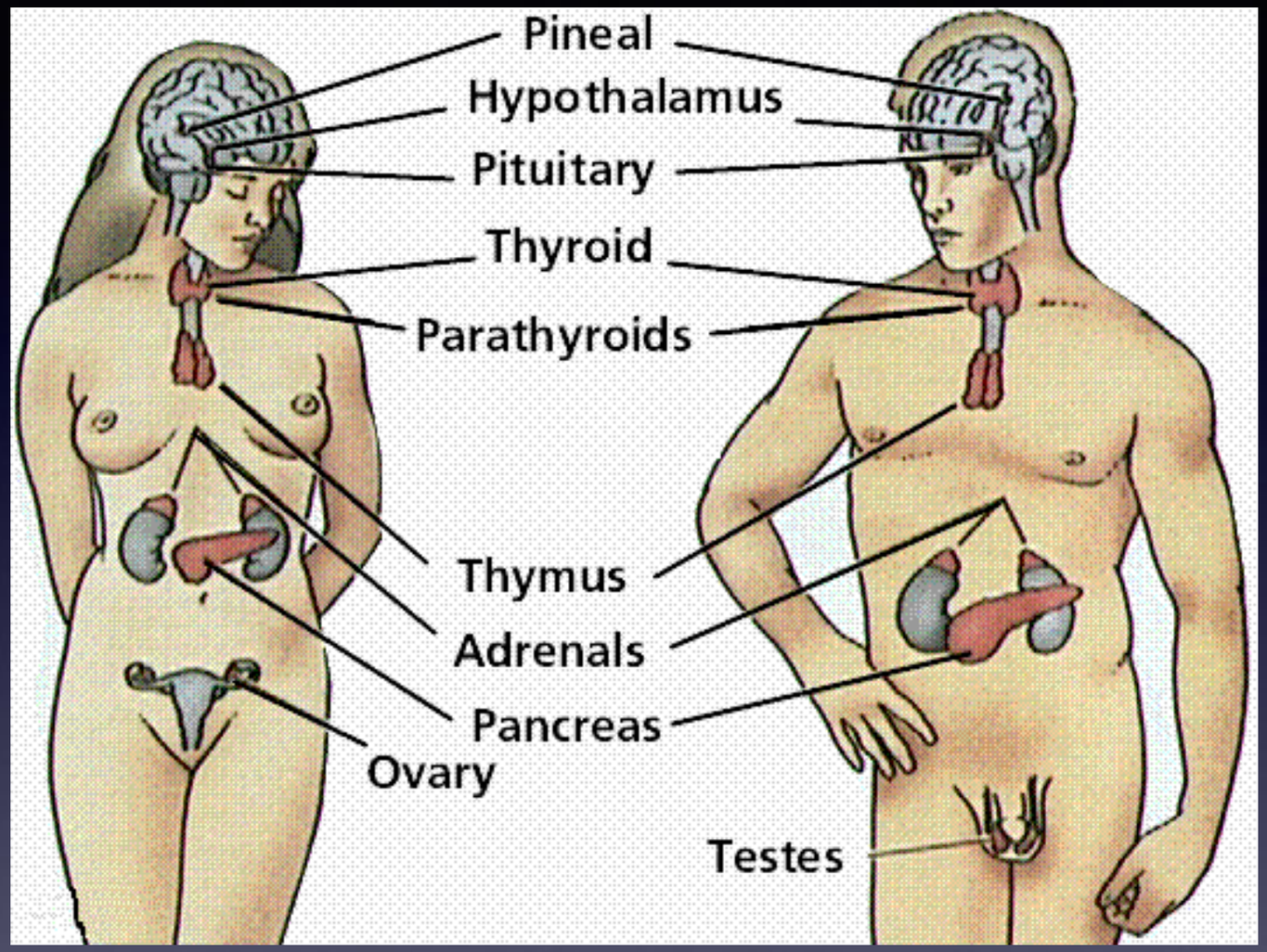
This talk visits some relevant topics to try and set background context about why PPI bioinformatics is useful

In particular, spatio-temporal signaling using modular proteins and aspects of cancer, viral and bacterial diseases

Interdisciplinary Signalling Research requires addressing communication at multiple levels between cells and within cells

Let's drill down from the top level

Organismal Level Signalling



Insulin, endocrine system etc. could be researched long before in-cell signalling

Circadium Rhythm has organismal and cellular level components. So why would anyone study circadium rhythm in Hela cells?

PubMed | hela circadium rhythm

Create RSS Create alert Advanced

Format: Summary **Sort by**: Most Recent **Per page**: 20

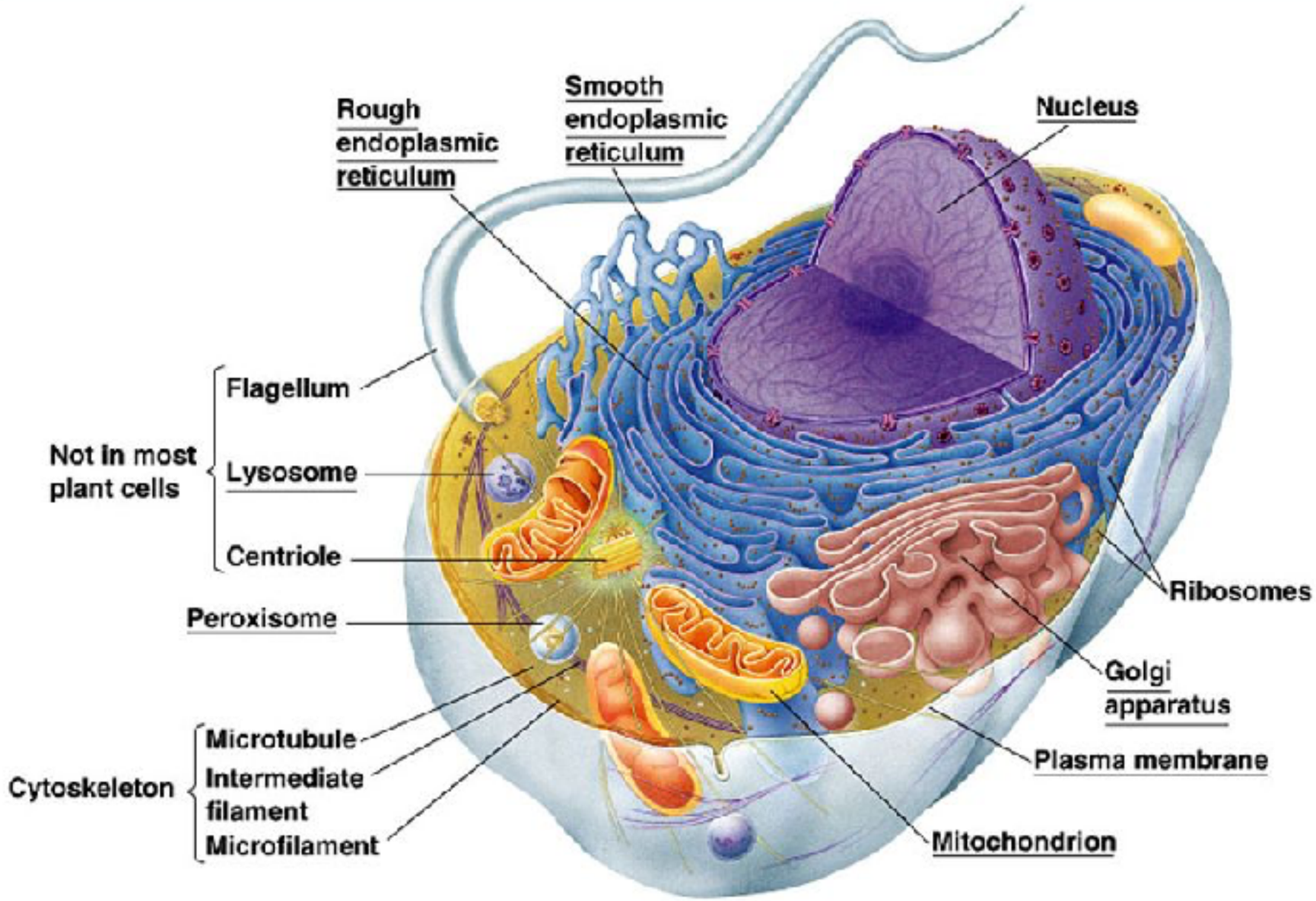
Send to

Search results

Items: 1 to 20 of 40 << First < Prev Page 1 of 2 Next > Last >>

In-cell signalling is primarily studied in cancer cells adapted to tissue culture: (1) Biased (2) Broken

A Generic Animal Cell

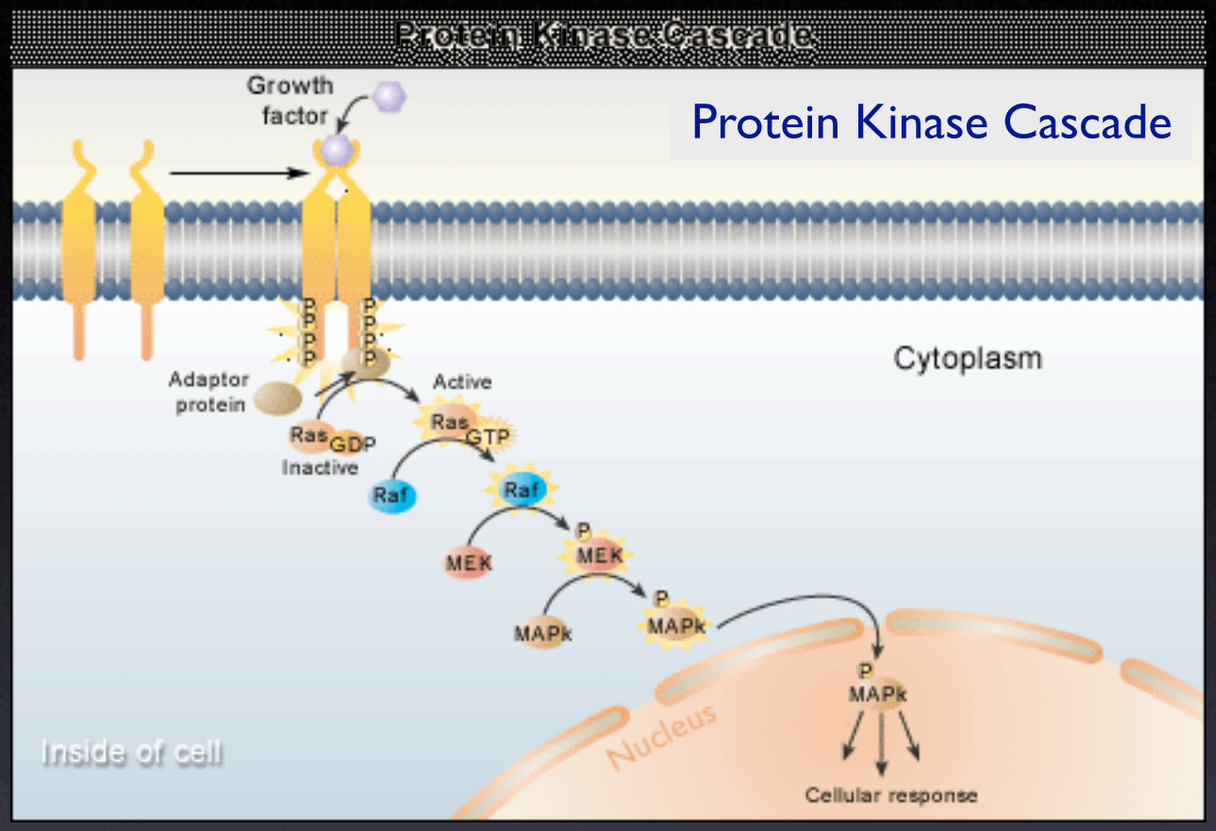


Copyright © 2003 Pearson Education, Inc., publishing as Benjamin Cummings.

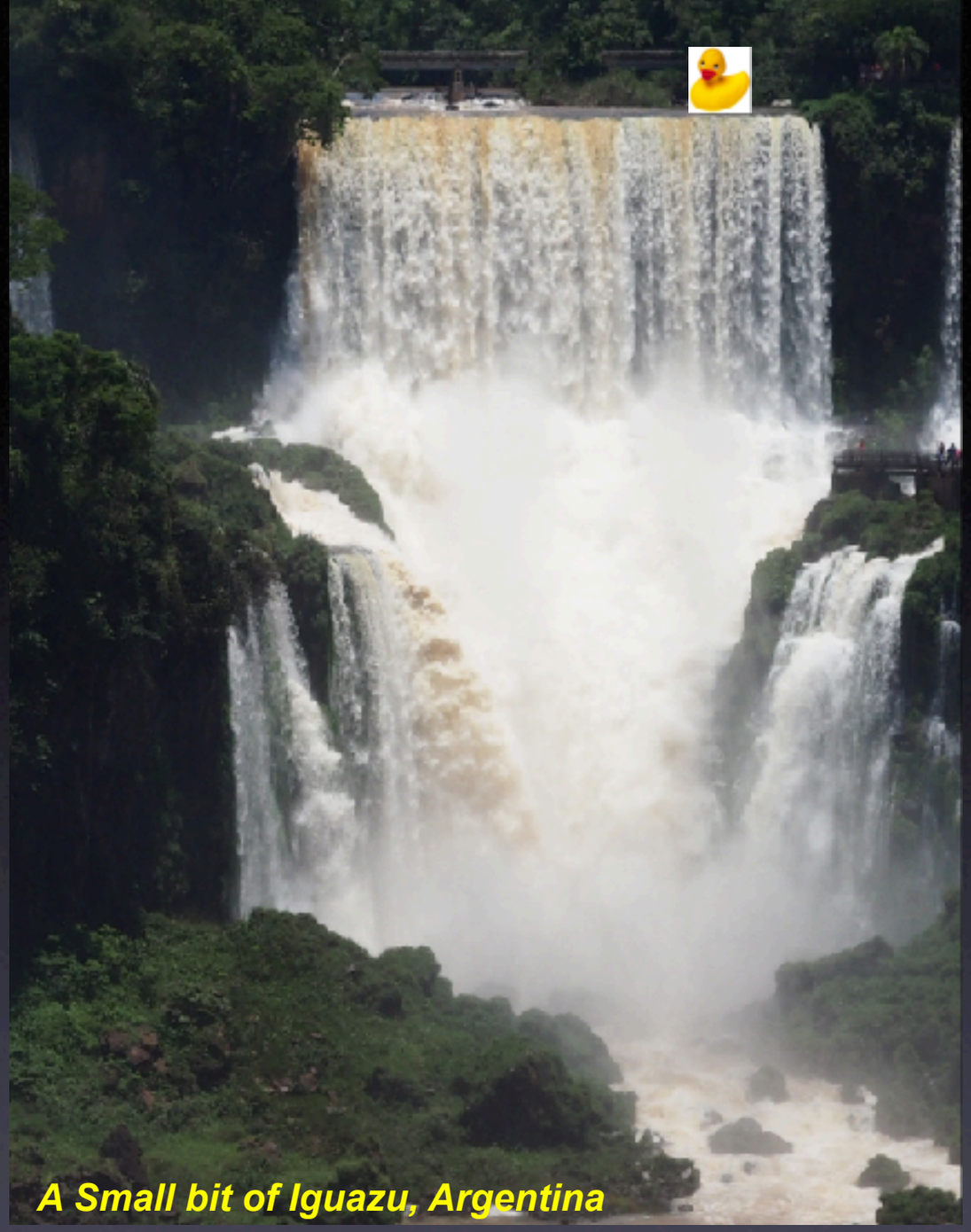
Copyright © 2004 Pearson Education, Inc. publishing as Benjamin Cummings

Molecules have to get into cell compartments but also be able to get around them

Cascades



source http://www.biology.arizona.edu/cell_BIO/



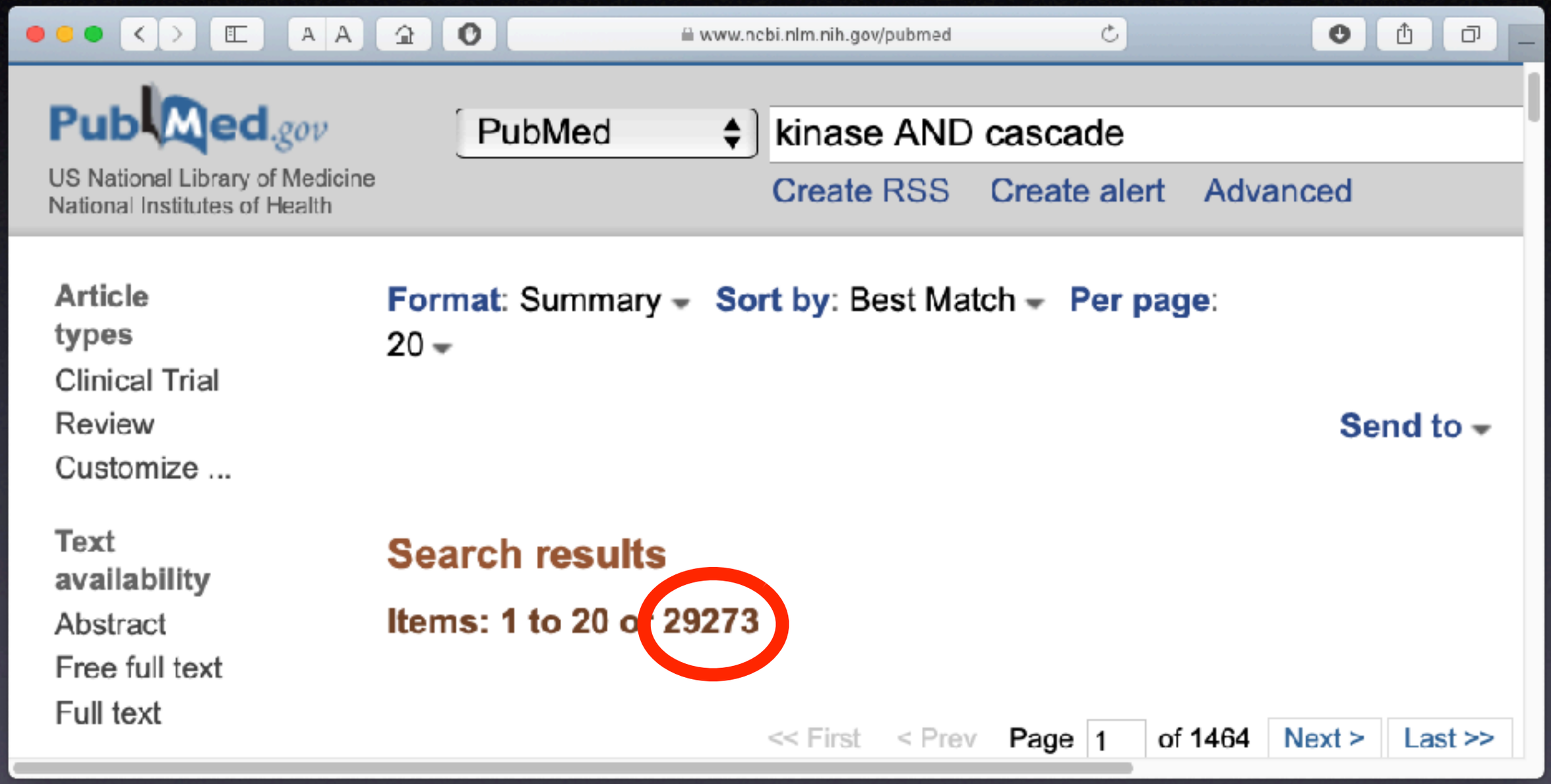
Properties

Linearity
Uneven
Accellerating
Unregulated
Uncertain end point?

Cascading mechanisms are neither accurate nor precise

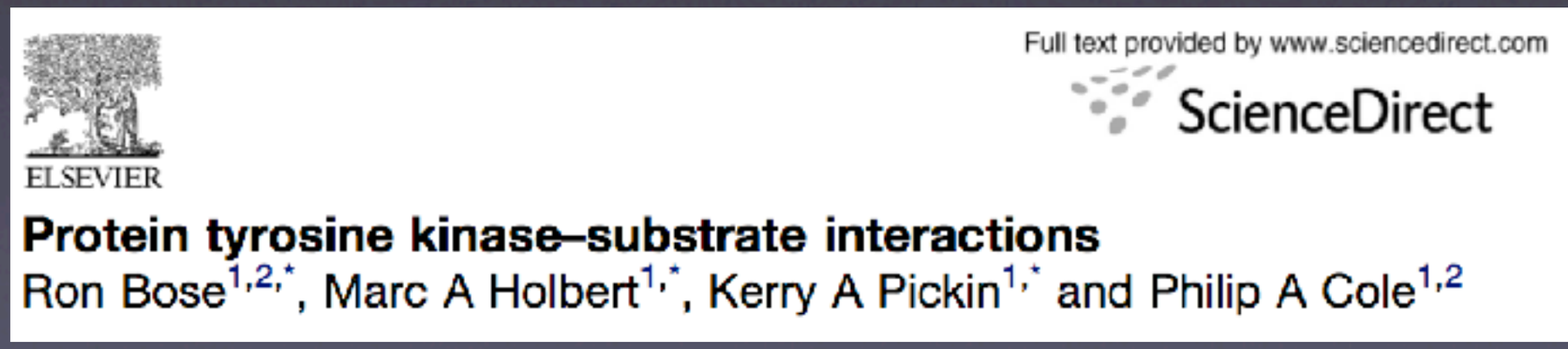
Vast literature on kinase cascades

They must be well understood by now



Most Tyrosine Kinases have very limited sequence specificity

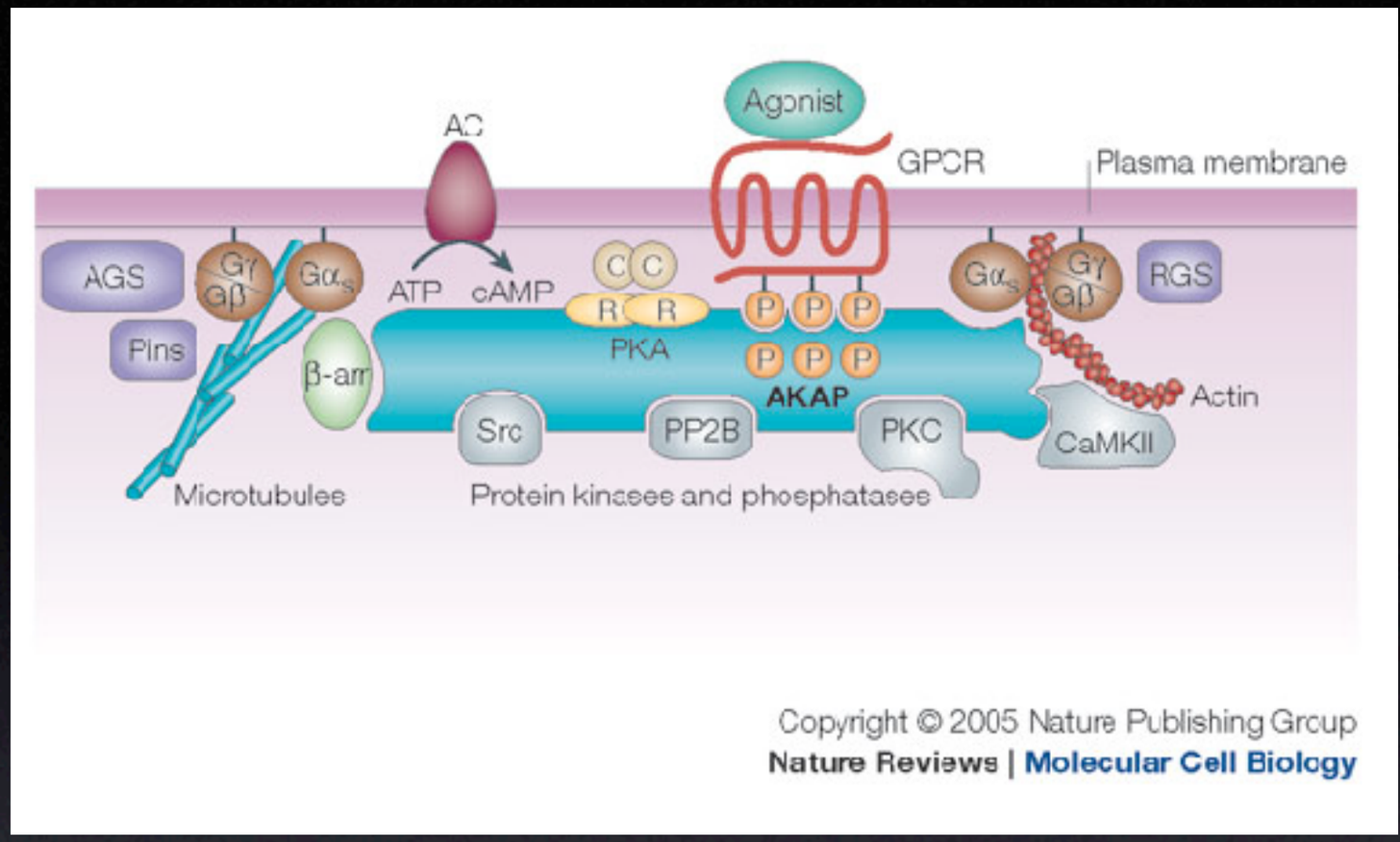
- *in vivo* TK substrate detection remains difficult
- *in vivo* substrates ≠ good in *vitro* peptides
- Cannot define a simple sequence pattern at phosphosite
- Problem: how do they avoid each other's substrates?



Curr. Op. Str. Bio. (2006) 16, 668

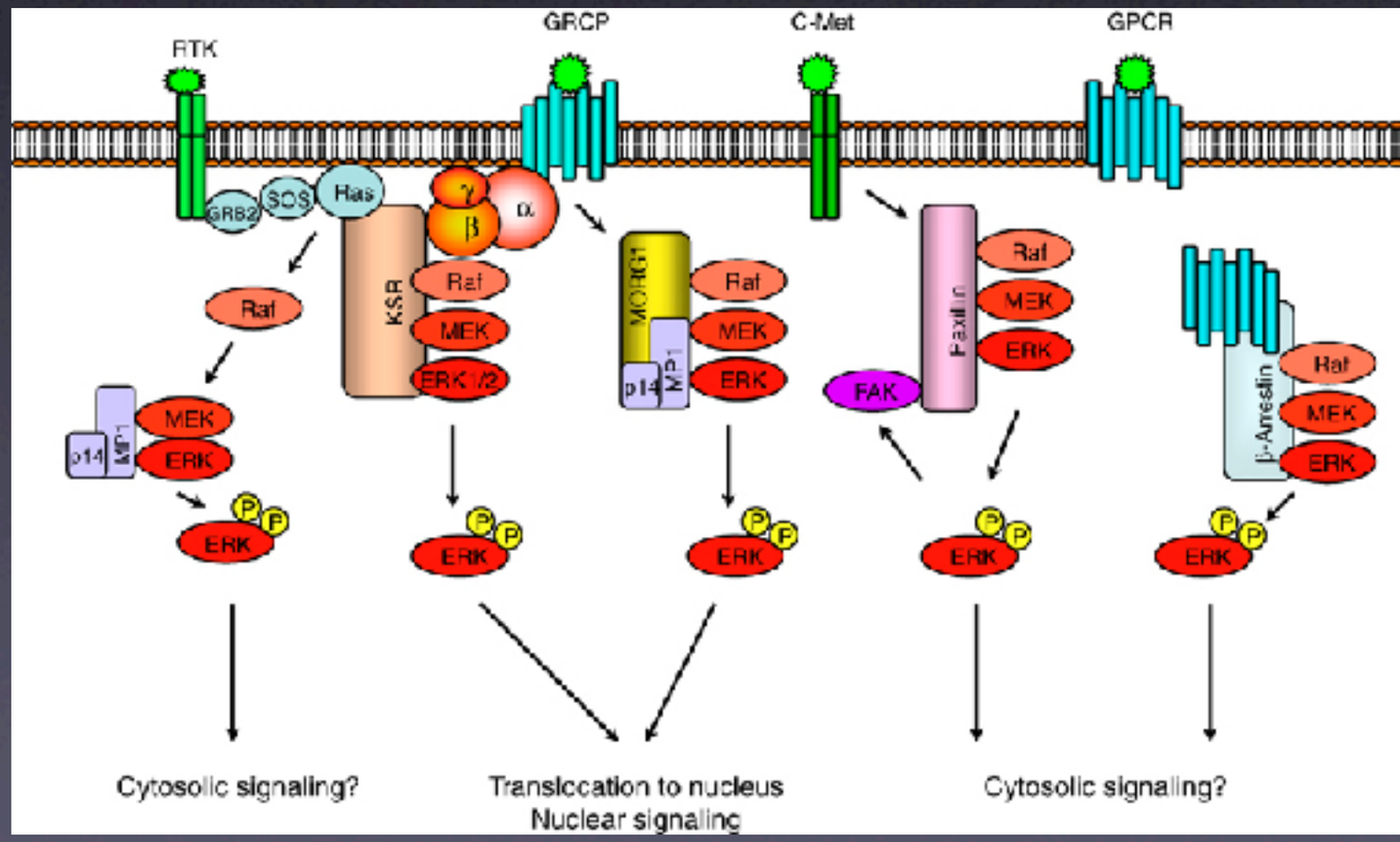
Solution to kinase substrate specificity problem: Scaffolding

PKA/Src/PKC scaffold



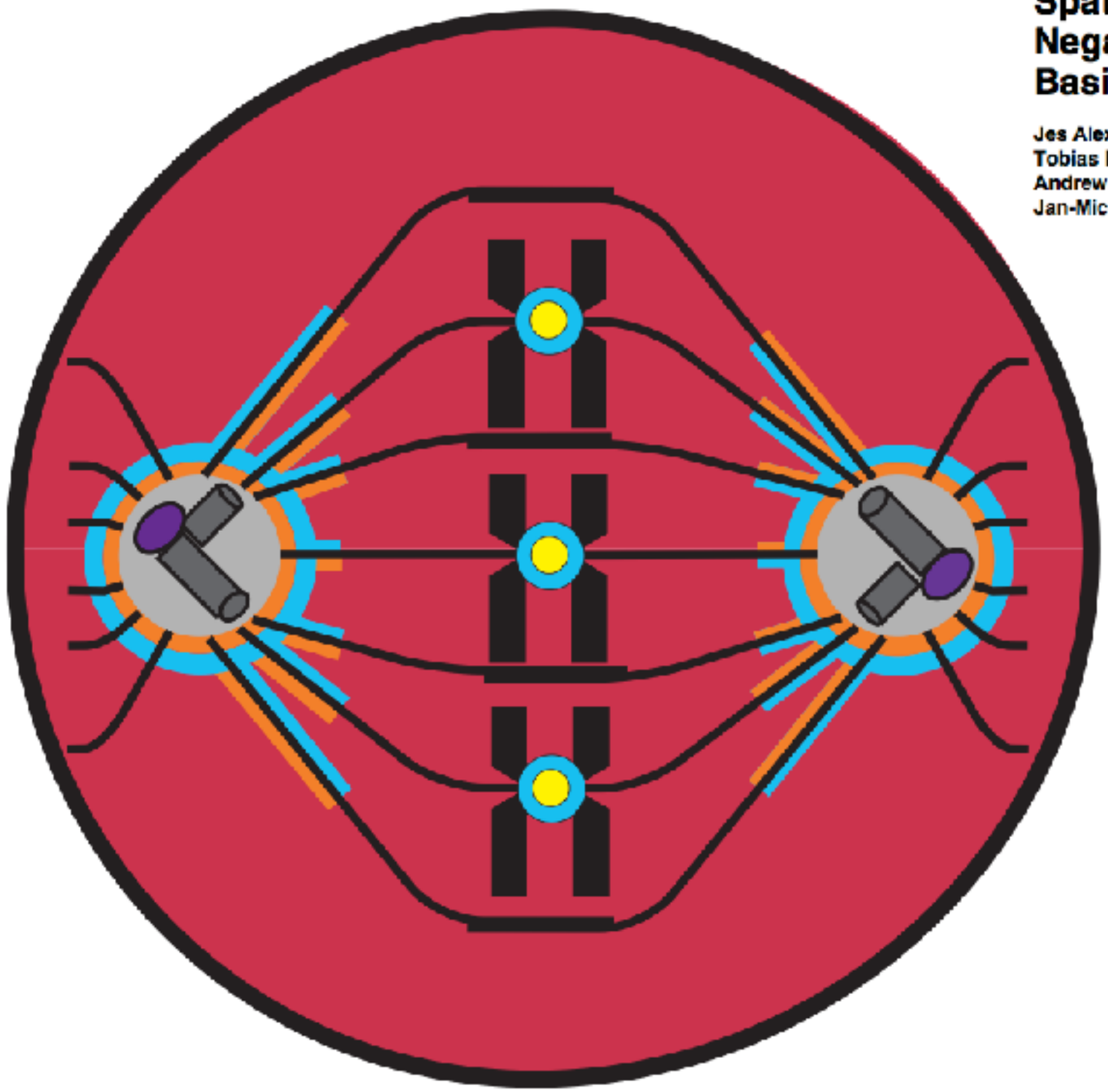
Malbon (2005) NRMCB, 6, 689

Map kinase scaffolds



Dhanasekaran (2007) Oncogene, 26, 3185

Spatial Exclusivity of Mitotic Kinases



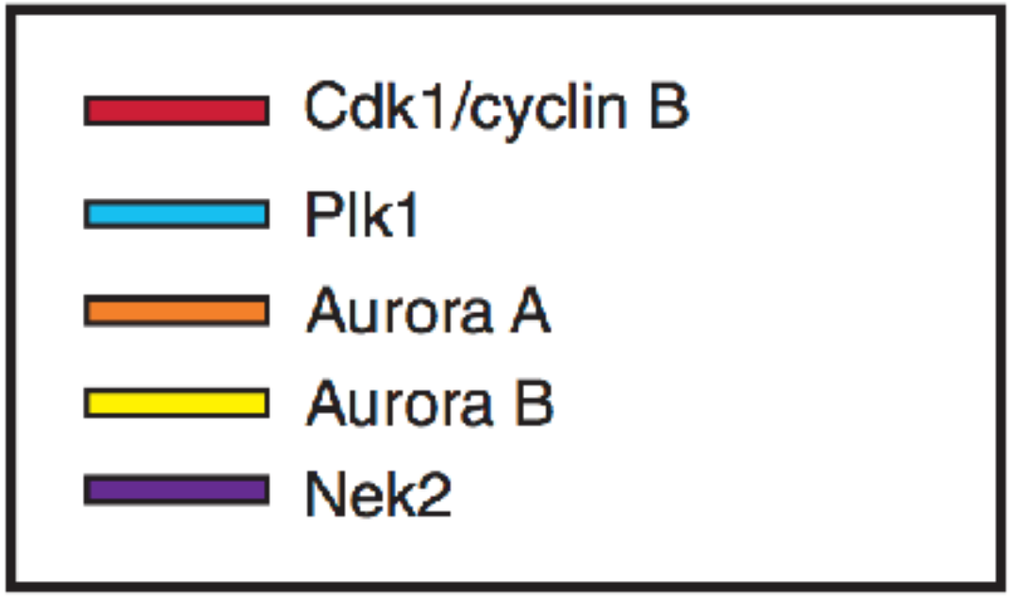
RESEARCH ARTICLE

MITOTIC KINASES

Sci. Sig., 6-2011

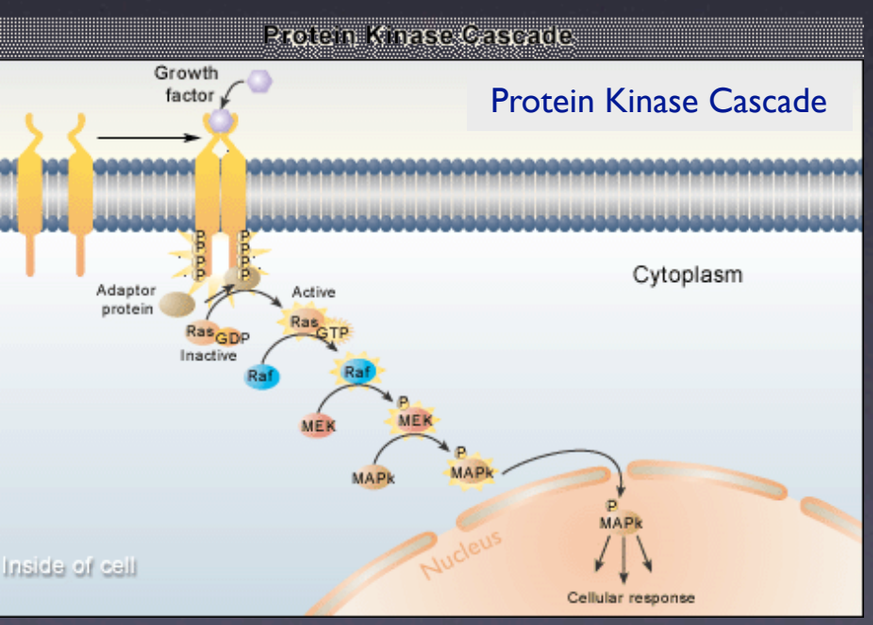
Spatial Exclusivity Combined with Positive and Negative Selection of Phosphorylation Motifs Is the Basis for Context-Dependent Mitotic Signaling

Jes Alexander,[1*] Daniel Lim,[1] Brian A. Joughin,[1] Björn Hegemann,[2†] James R. A. Hutchins,[2] Tobias Ehrenberger,[1] Frank Ivins,[3] Fabio Sessa,[4] Otto Hudecz,[2] Erich A. Nigg,[5] Andrew M. Fry,[6] Andrea Musacchio,[4] P. Todd Stukenberg,[7] Karl Mechtler,[2] Jan-Michael Peters,[2] Stephen J. Smerdon,[3] Michael B. Yaffe[1,8‡]



Implications of kinase localisation, complexes, scaffolds and weak substrate specificity

- Kinase-substrate interactions occur in scaffolded complexes including bound positive and negative regulators
- Kinases do not find their substrates by simple free diffusion



source http://www.biology.arizona.edu/

- *"Kinase Cascade" is one of the worst analogies in Biology and its meme needs to become extinct*
- Widely used Reaction-Diffusion equations are inappropriate and/or insufficient for modelling kinase signalling

Cell Regulation: Cooperative and Spatially arranged

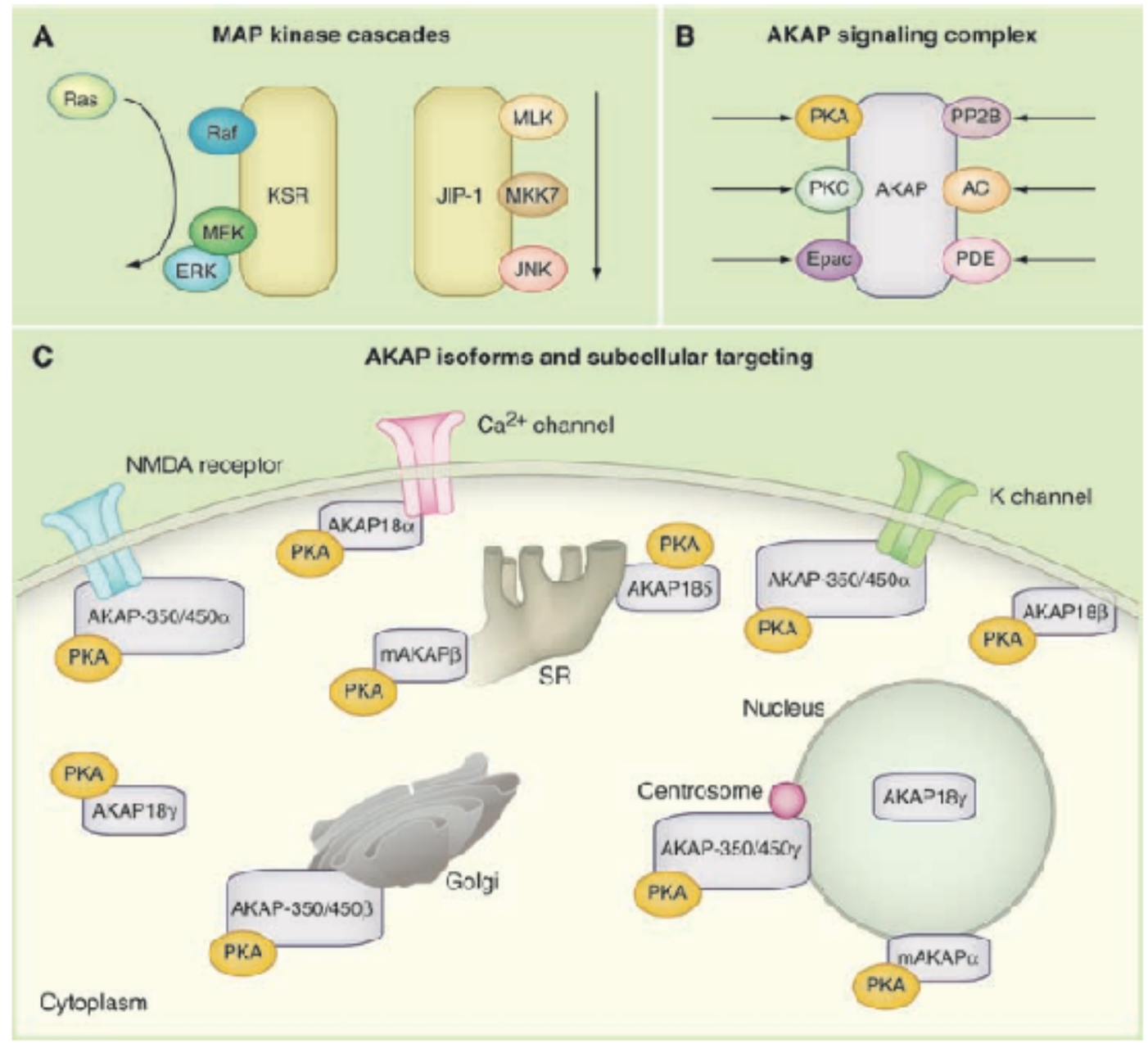
Spatial Cell Biology

REVIEW

Cell Signaling in Space and Time: Where Proteins Come Together and When They're Apart

John D. Scott[1,*] and Tony Pawson[2,3,*]

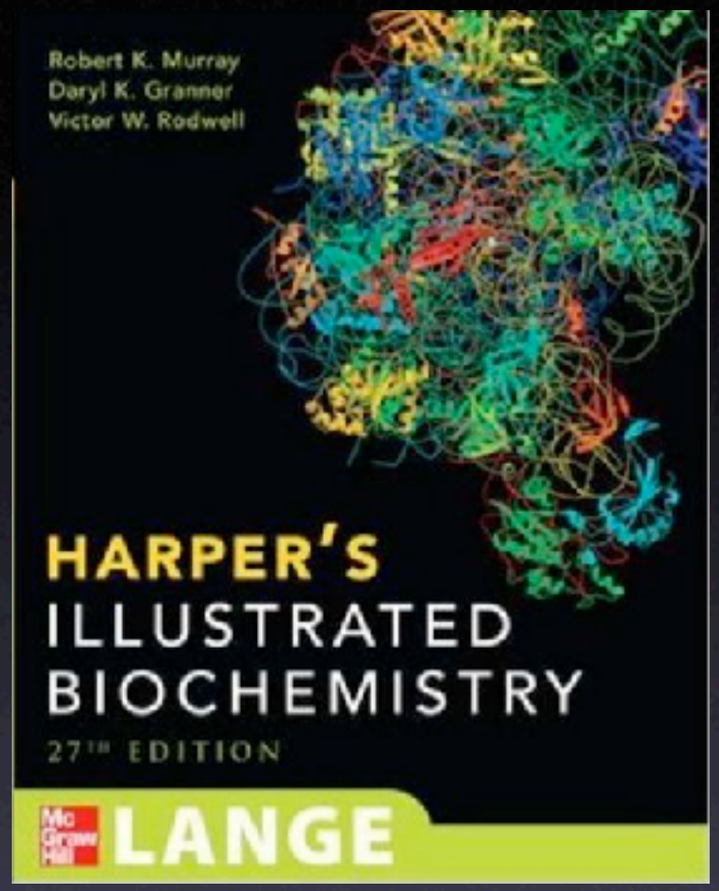
Science, 326, 2009

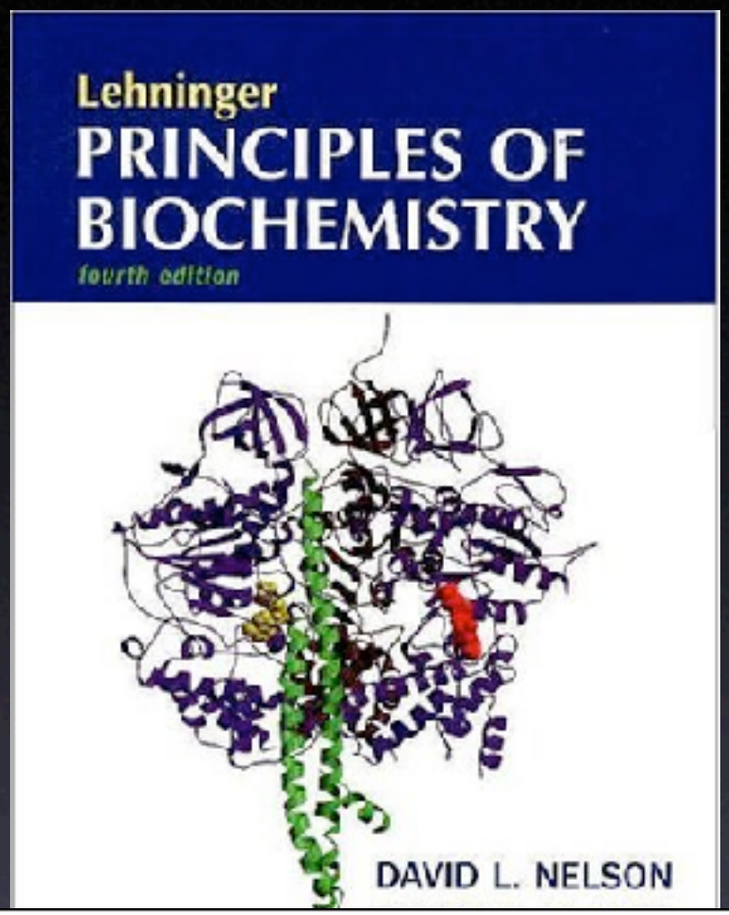


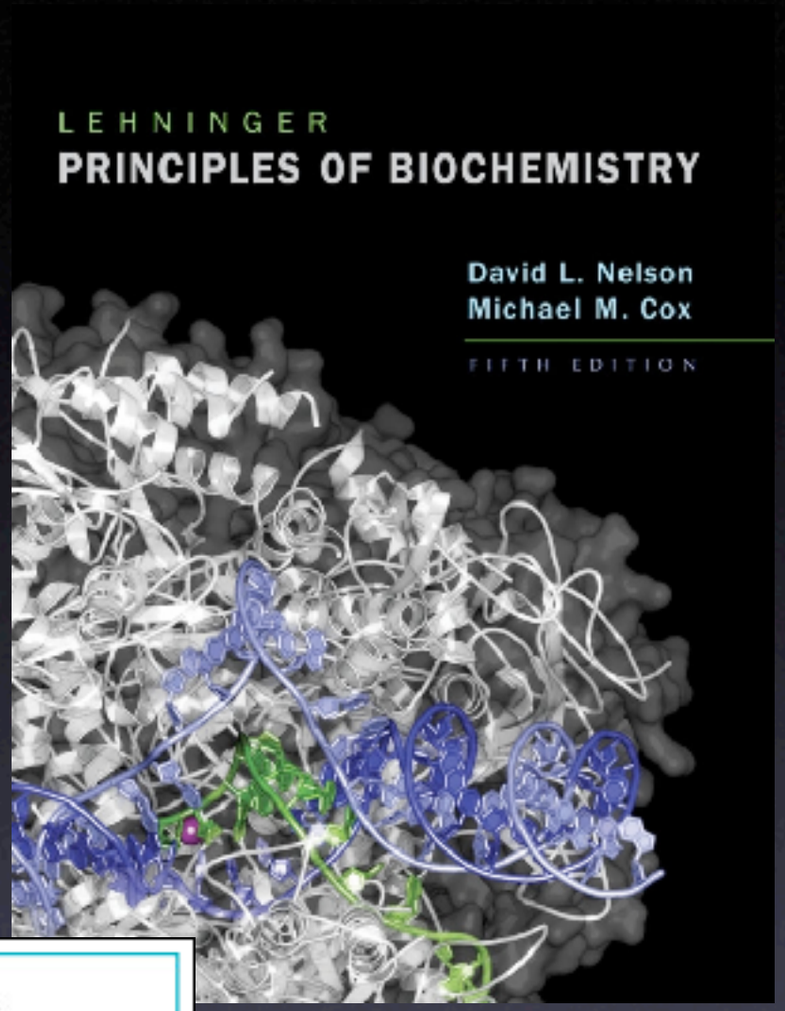
Tony Pawson, Cell (2004)

While there is still much debate about these ideas, the spatial segregation of signaling pathways is likely to be an important topic for the future.

Biochemistry Text Book Figures: Engines of the Cell







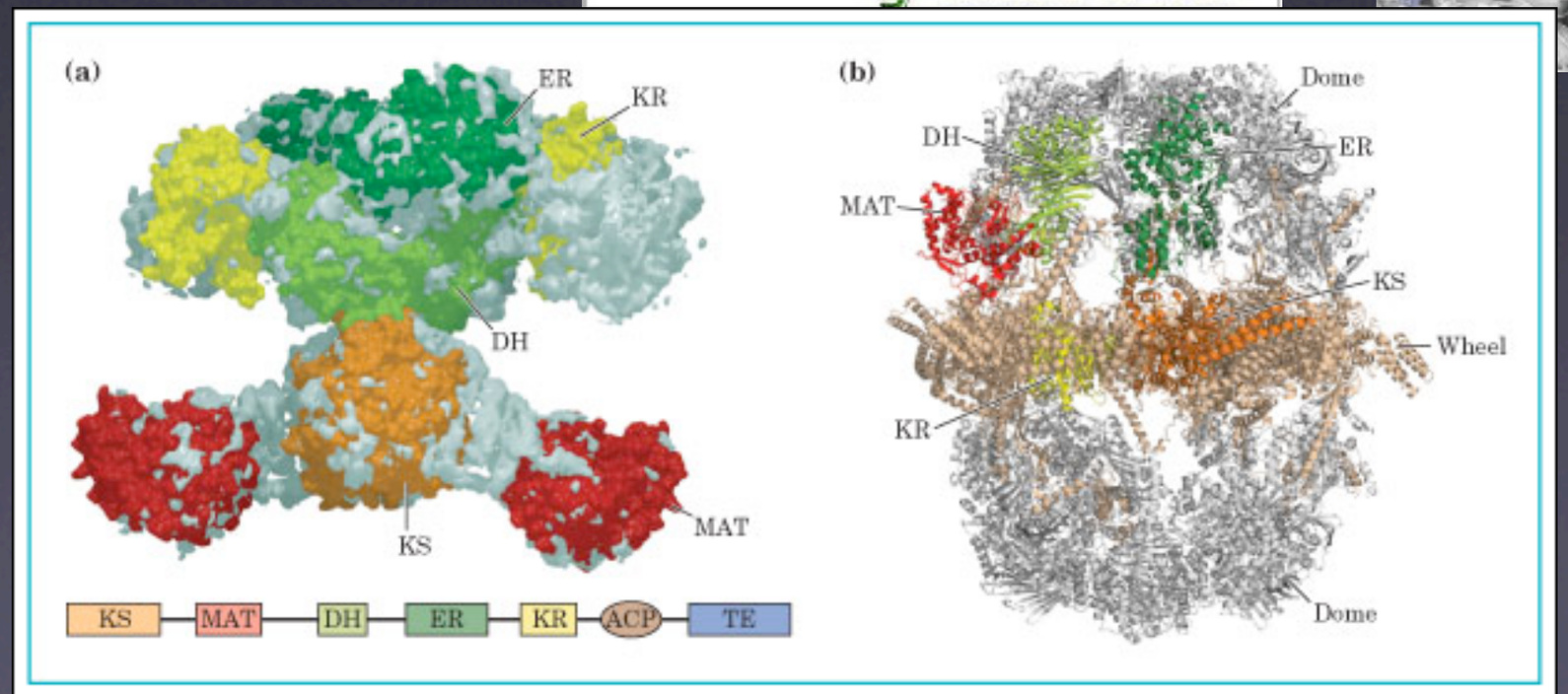


FIGURE 21-3 The structure of fatty acid synthase type I systems.

Complexes
Complexes
Complexes

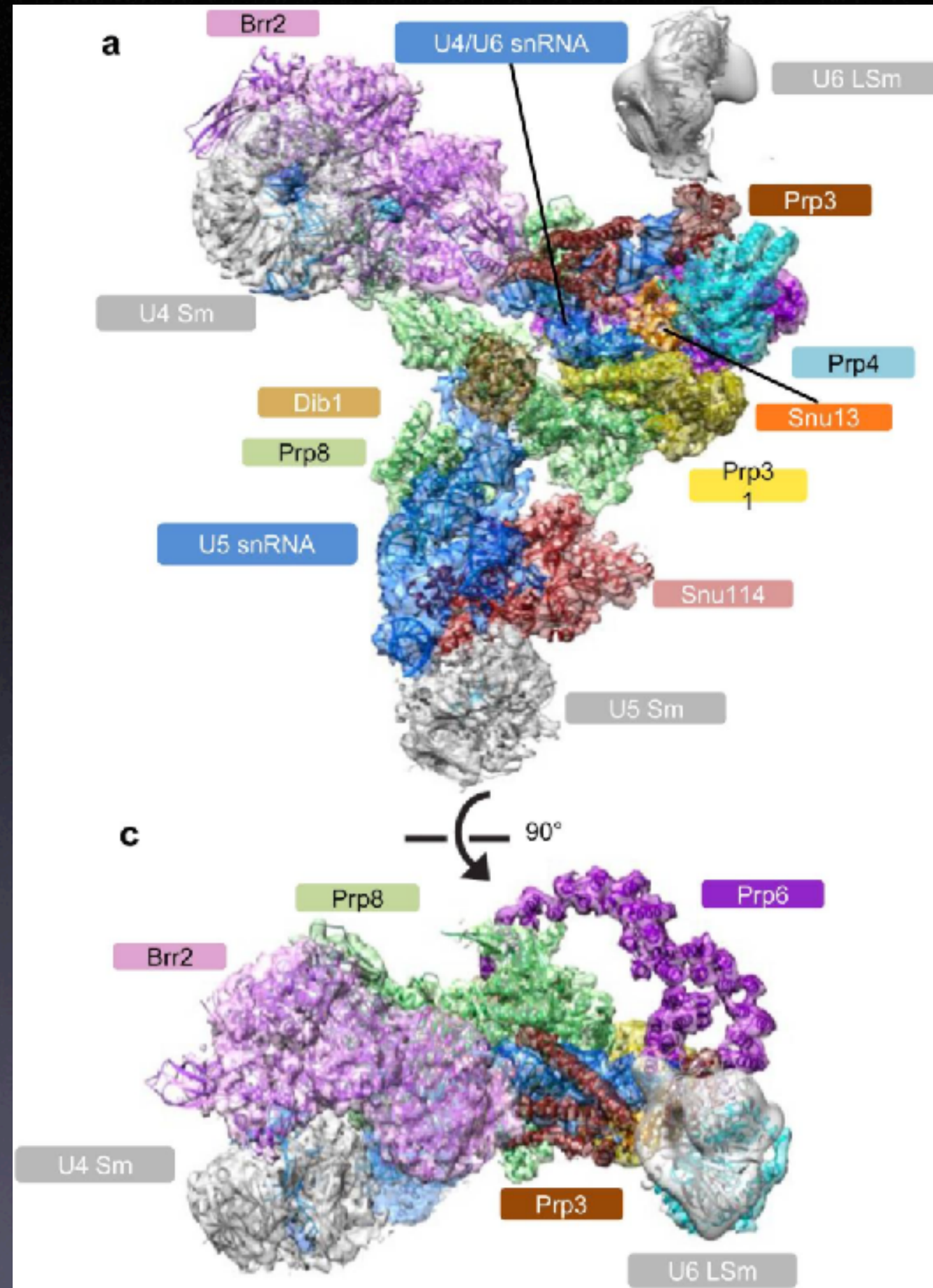
Truth and clarity are complementary

Niels Bohr

U4/U6.U5 tri-snRNP by cryo-EM

A Cryo-EM complex that is:
Large
Flexible
Dynamic
Has big holes
Very hard to crystallise

There are plenty of complexes that are:
Much larger
More flexible
More dynamic
Have even bigger holes
Will never be crystallised



Nguyen et al. (2015) Nature,

Some types of complex tolerate stoichiometry violations

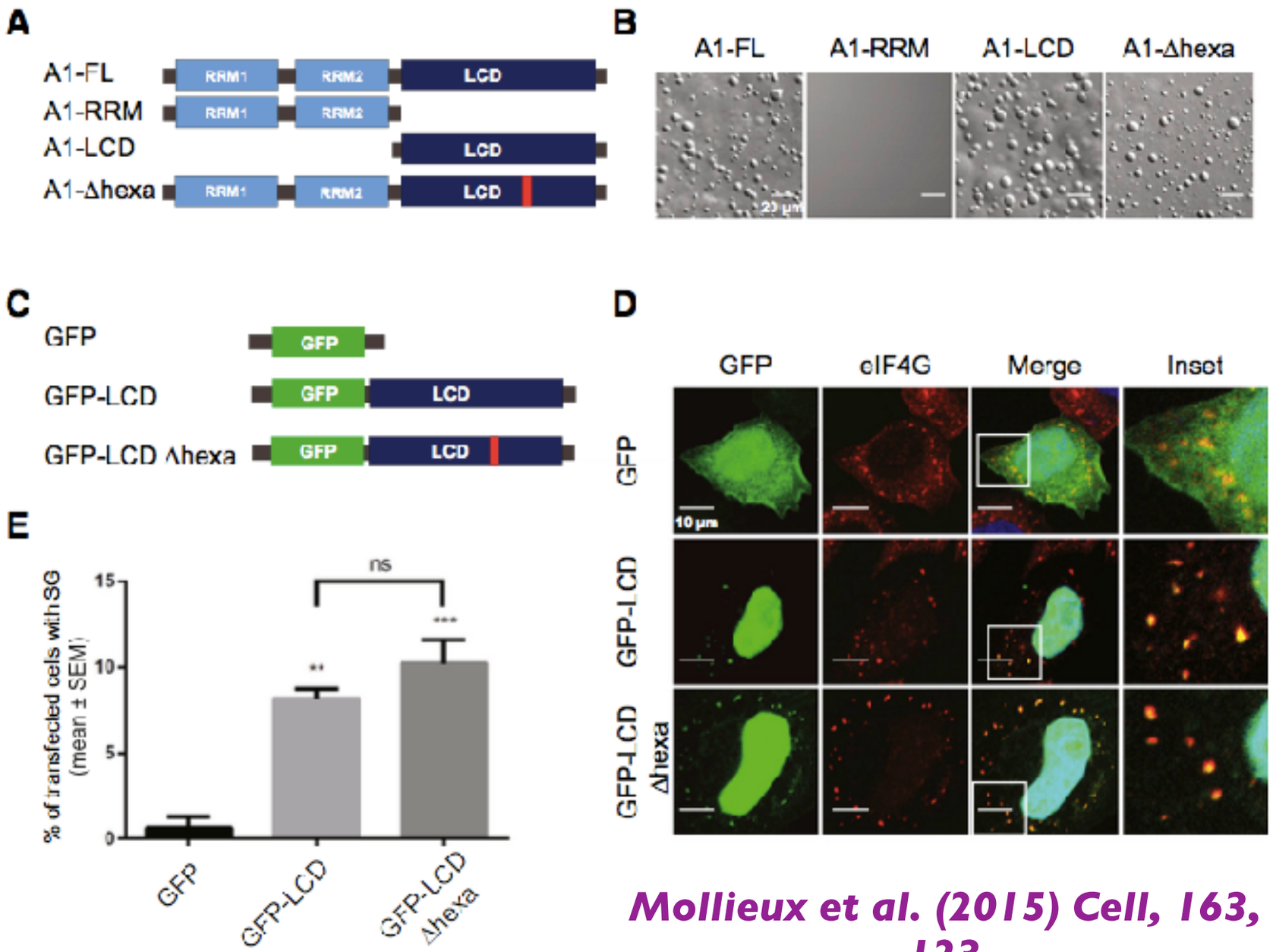


Figure 2. Liquid-Liquid Phase Separation by hnRNPA1 Is Mediated by the C-Terminal Low Complexity Sequence Domain and Is Distinct from Fibrillization

(A) Schematic of the structure of hnRNPA1 full length (A1-FL), the N terminus comprising the two folded RNA recognition motifs (A1-RRM), the low complexity sequence domain (A1-LCD), and the mutant with a deletion of residues 259–264 (Kim et al., 2013) (A1-Δhexa).

(B) DIC images of A1-FL, A1-RRM, A1-LCD, and A1-Δhexa at 140 μM protein, 150 mg/ml Ficoll in 50 mM HEPES, 300 mM NaCl, and 5 mM DTT.

(C) Schematic of the constructs transiently expressed in HeLa cells.

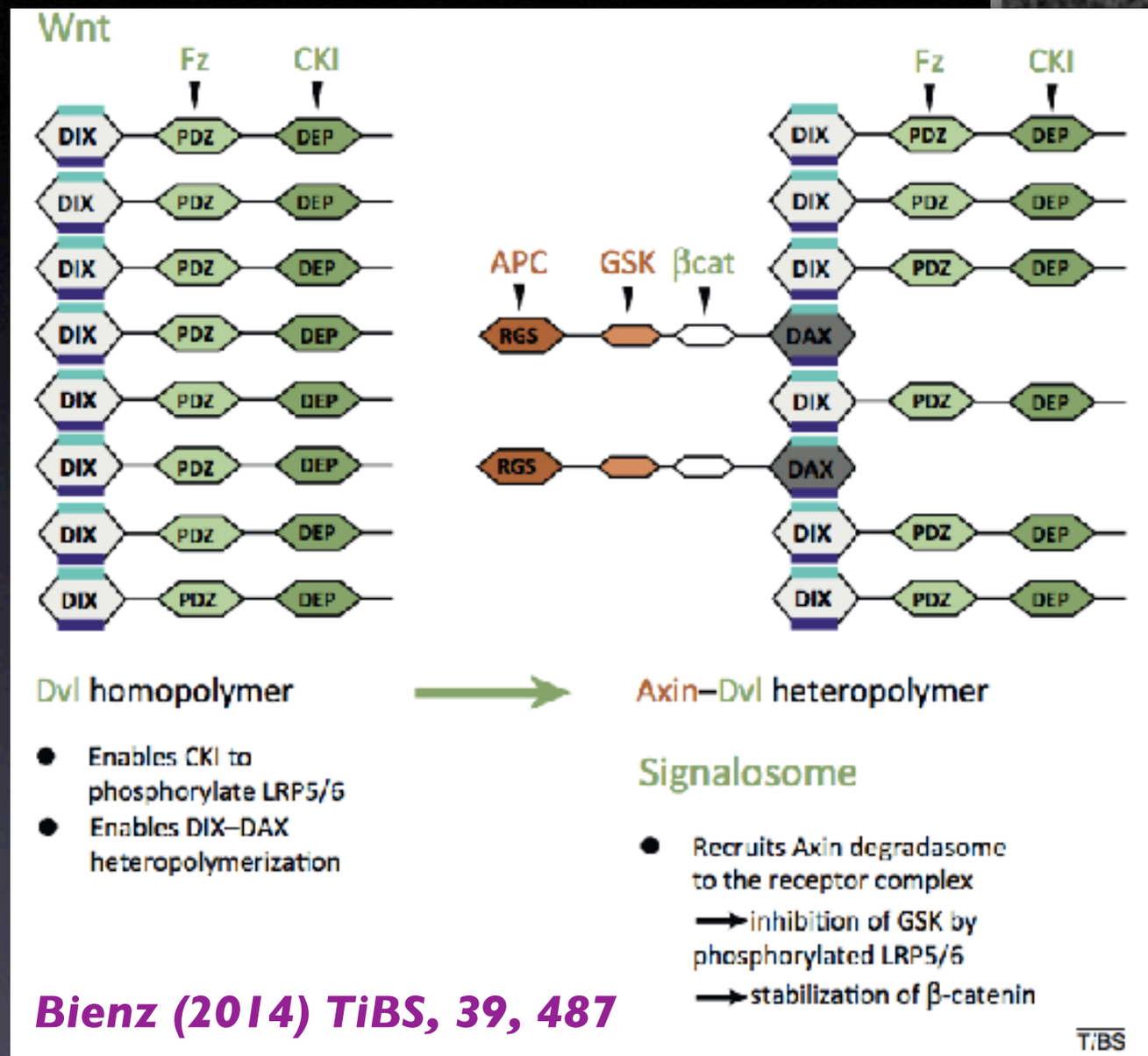
(D) Representative confocal microscopy images of HeLa cells transfected with constructs presented in (C), treated with 0.5 mM sodium arsenite for 15 min, and immunostained with anti-eIF4G (red) and DAPI (blue).

(E) Quantification for data in (D). The percentage of transfected cells displaying GFP signal in SGs ([number of cells with GFP-positive SGs/number of GFP-expressing cells] × 100) was plotted as mean ± SEM; n = 100 cells; $^{**}p < 0.005$, $^{***}p < 0.001$ by one-way ANOVA, Tukey's post hoc test.

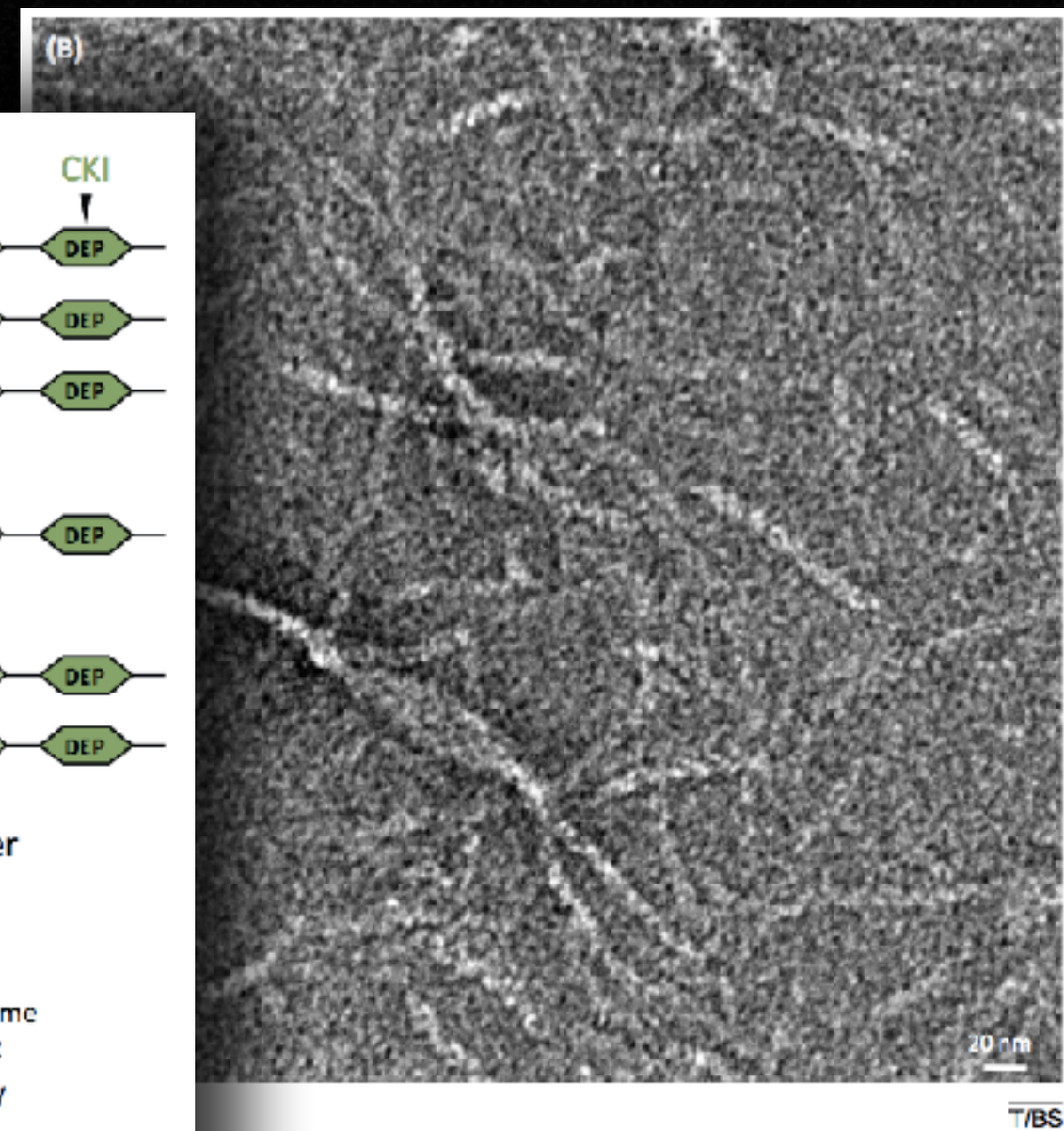
Mollieux et al. (2015) Cell, 163, 123

Liquid phase separation "complexes" scale to any size (e.g. stress granules, nucleoli)

Some types of complex tolerate stoichiometry violations



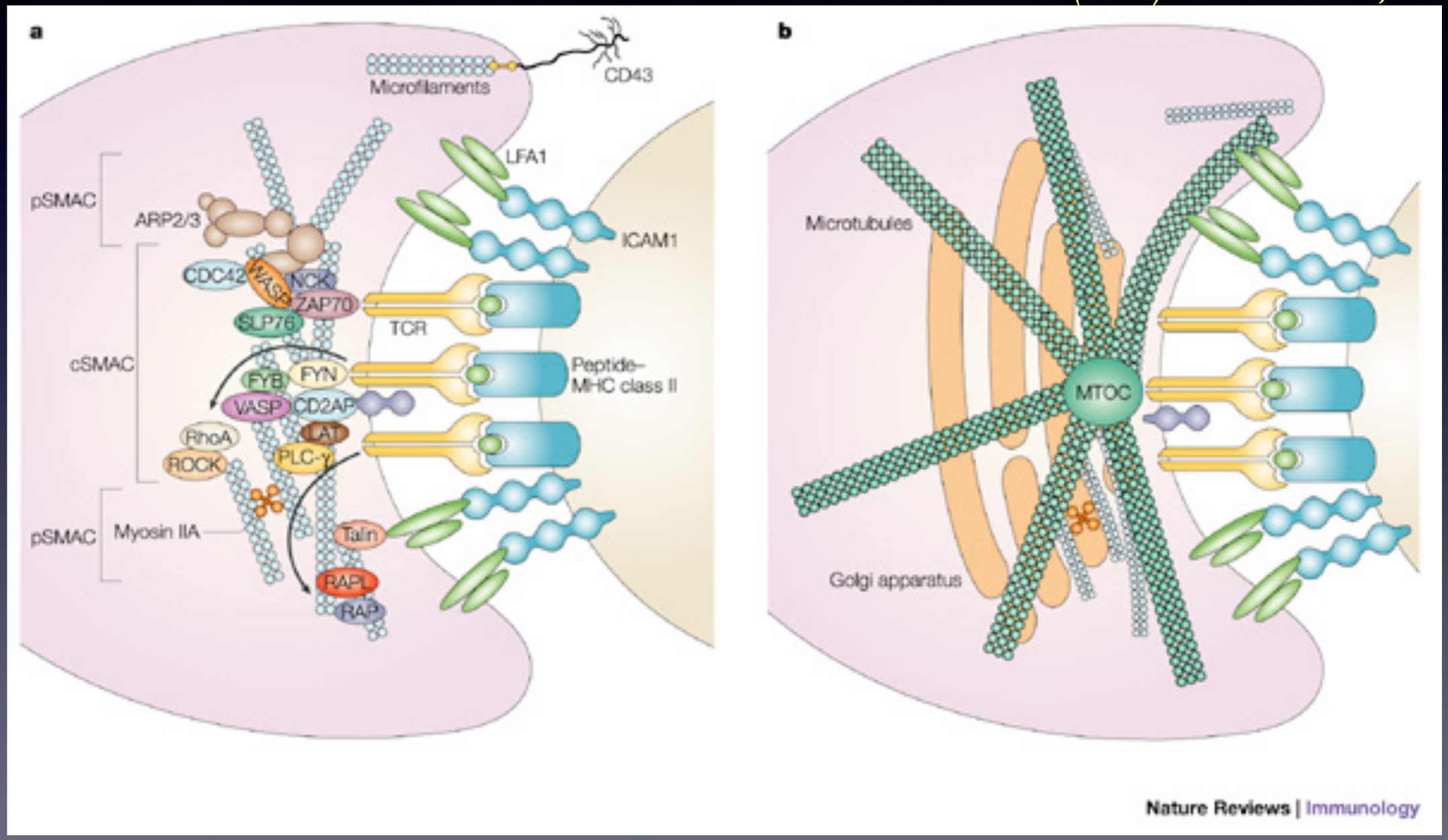
Bienz (2014) TiBS, 39, 487



Wnt-signalling Polymeric helical signalosomes have different proportions of component proteins: They also act as scaffolds to assemble other regulatory proteins/complexes

The Immunological Synapse - A platform for multisignal input and output in T Cell activation

Vicente-Manzanares et al. (2004) Nat. Rev. Imm. 4, 110

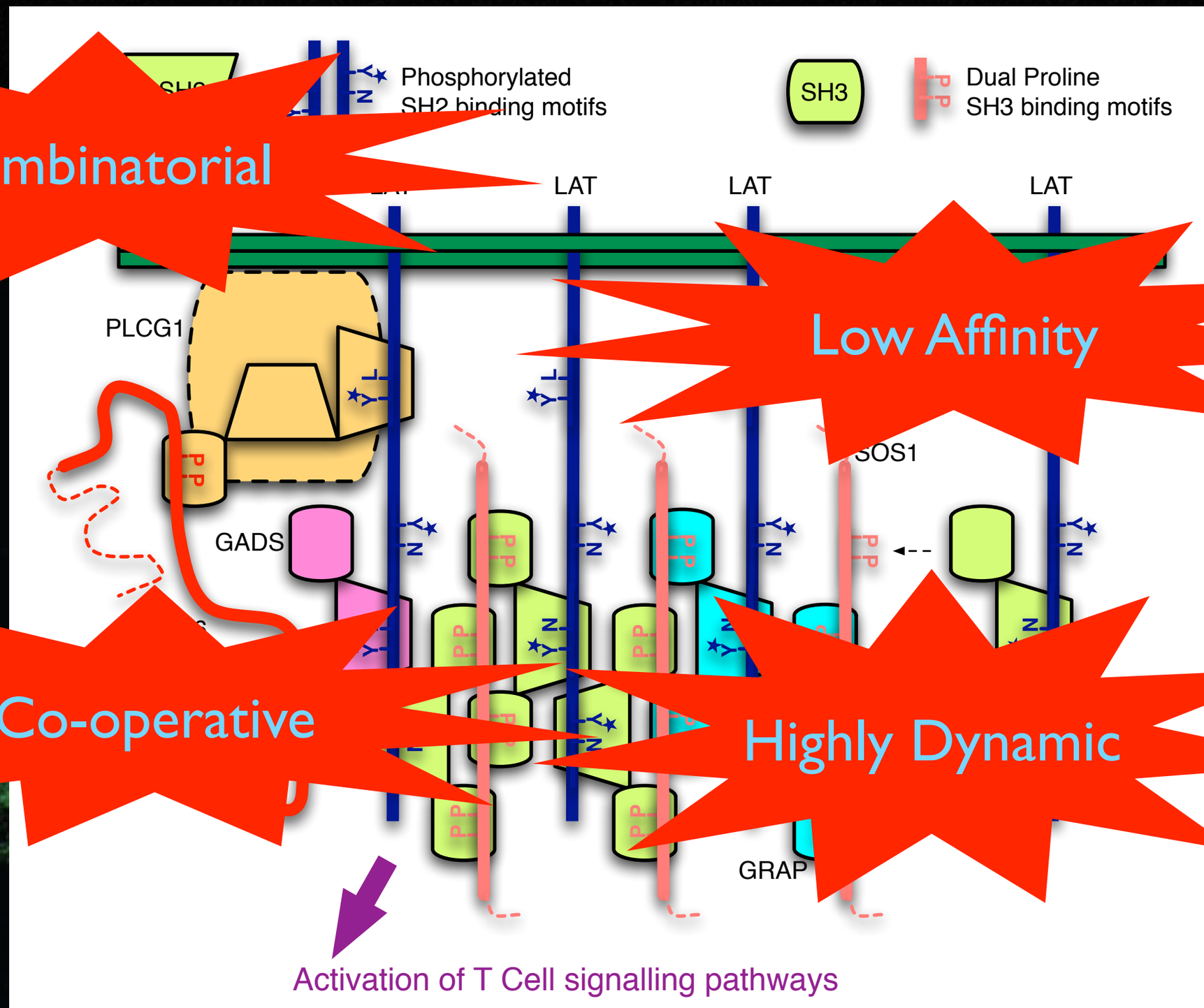


Effectively the whole cell becomes a giant macromolecular complex

Initiation and propagation of T Cell signalling

Multivalent assembly of the LAT signalling complex by short linear motifs

LAT is phosphorylated by TKs ZAP-70 and/or SYK



Combinatorial

Low Affinity

Co-operative

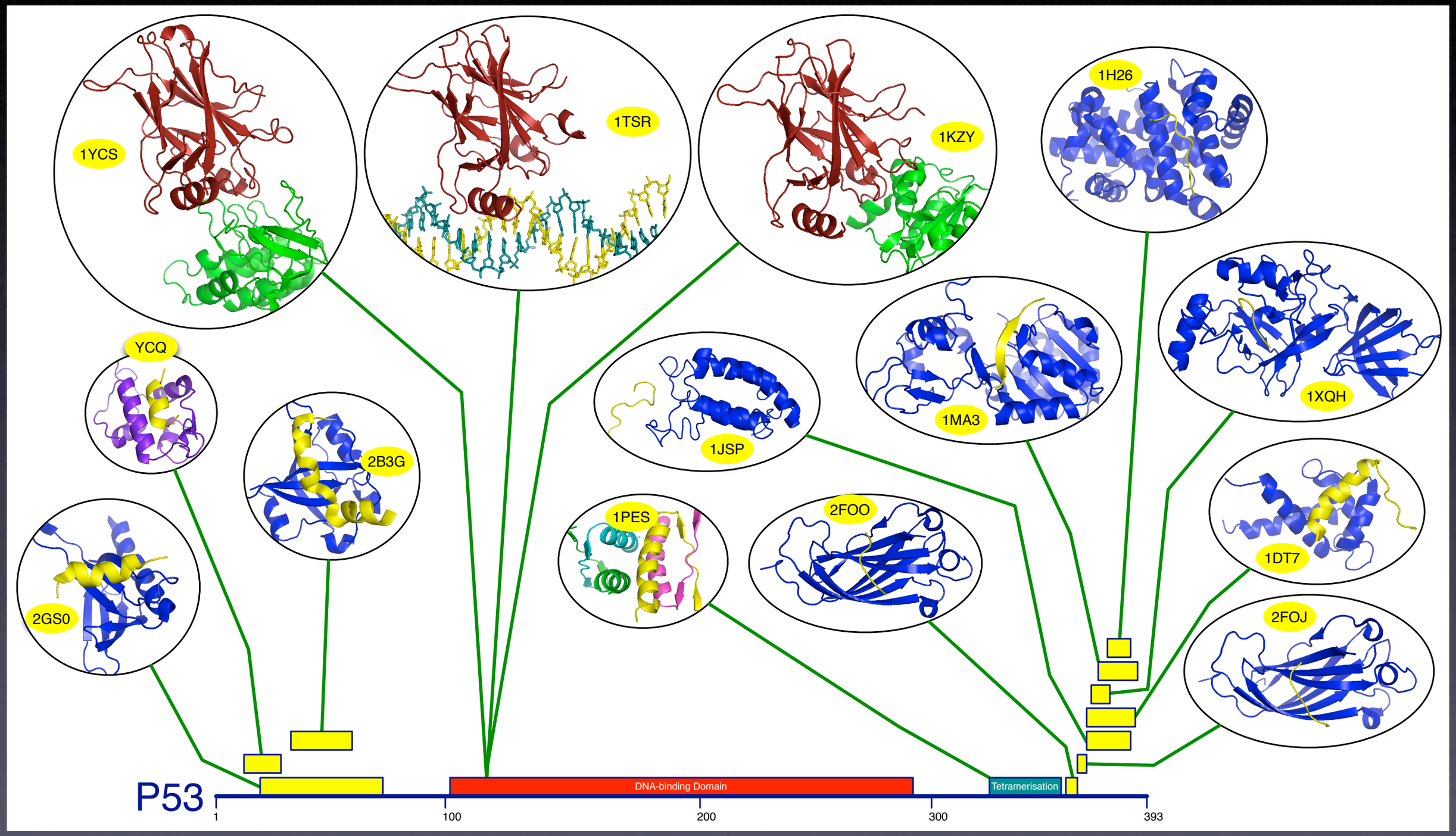
Highly Dynamic

After Houtman et al. (2006) NS&MB, 13, 798

Balagopalan et al. (2010) CSHPB, 13, a005512

Solved structures of p53 fragments

p53 exhibits switch-like properties



Gibson, TIBS 10/09

p53 is a famous example of a protein with a modular structural architecture

Thousands of cell regulatory proteins are composed of three main protein module classes:

- Natively folded domains
- Intrinsically Disorded Domains (IDDs)
- Short Linear Motifs (SLiMs)

You have been and will continue to learn more about these modules during the course

Degrons are a type of SLiM

SCIENCE SIGNALING | REVIEW

Amy Wheeler
@DJAhmei

I misread this paper's title as "Dragons in cancer"
https://t.co/cmdQlFEUhH

14 Mar 2017

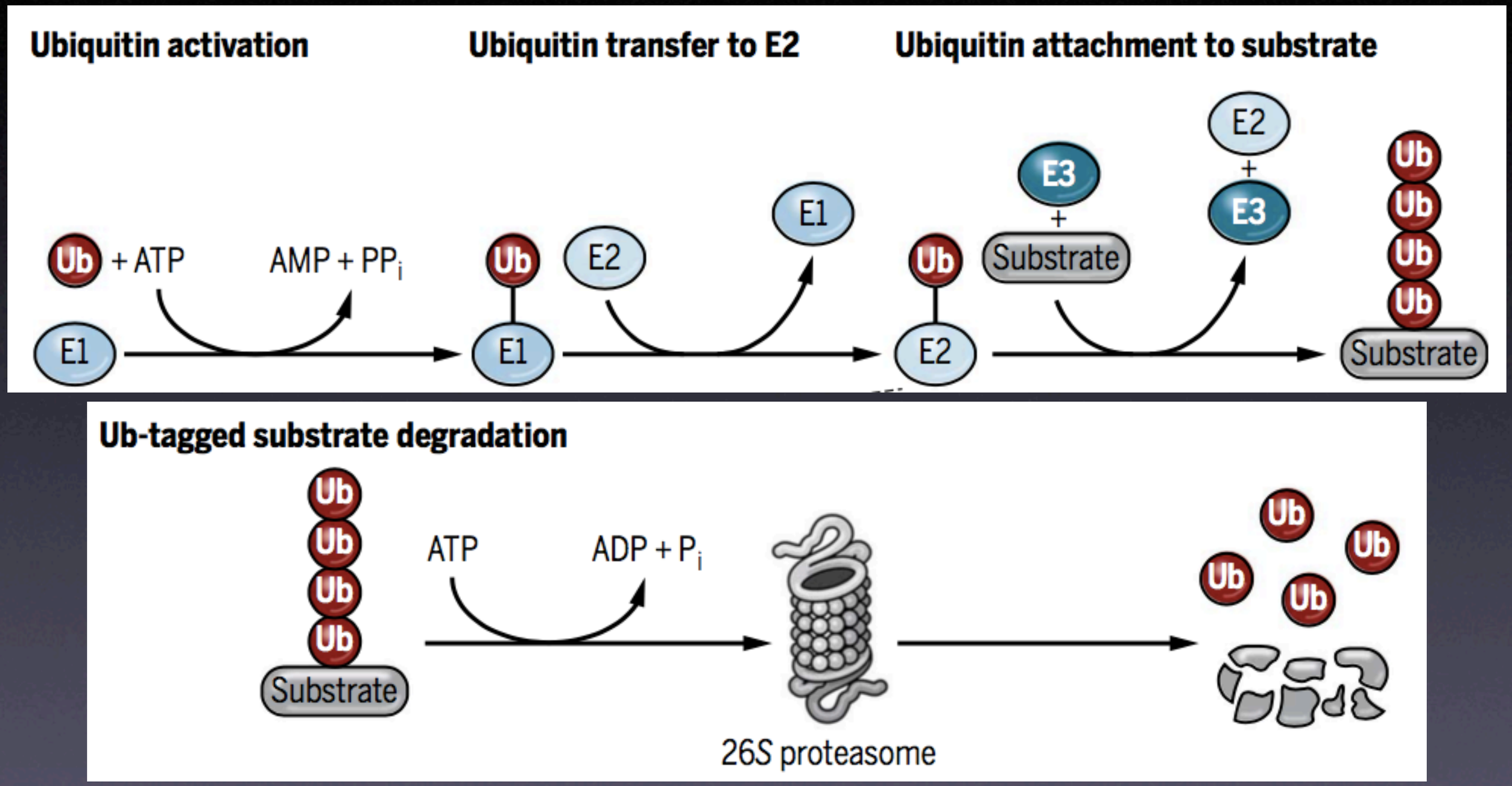
CANCER

Degrons in cancer

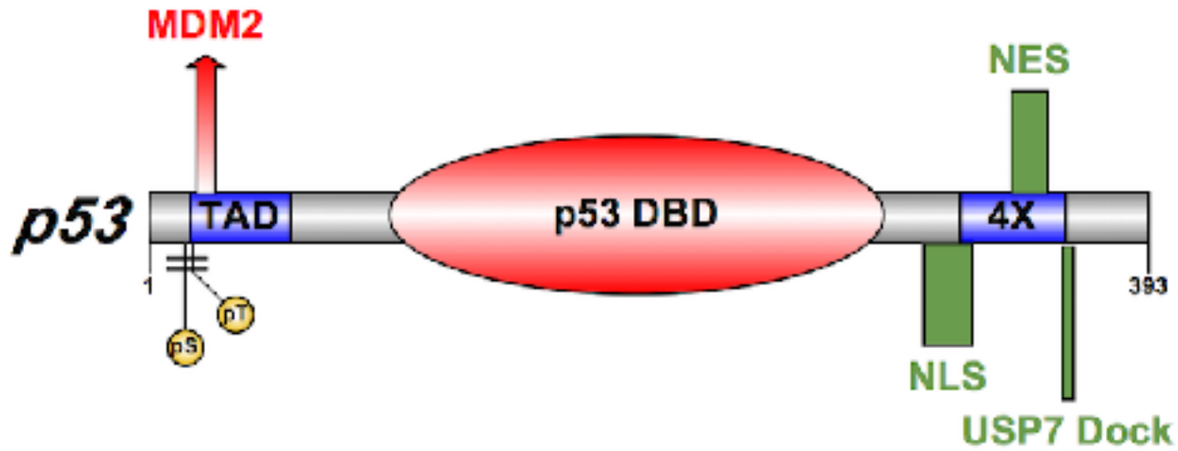
Bálint Mészáros,[1]* Manjeet Kumar,[2]* Toby J. Gibson,[2†] Bo

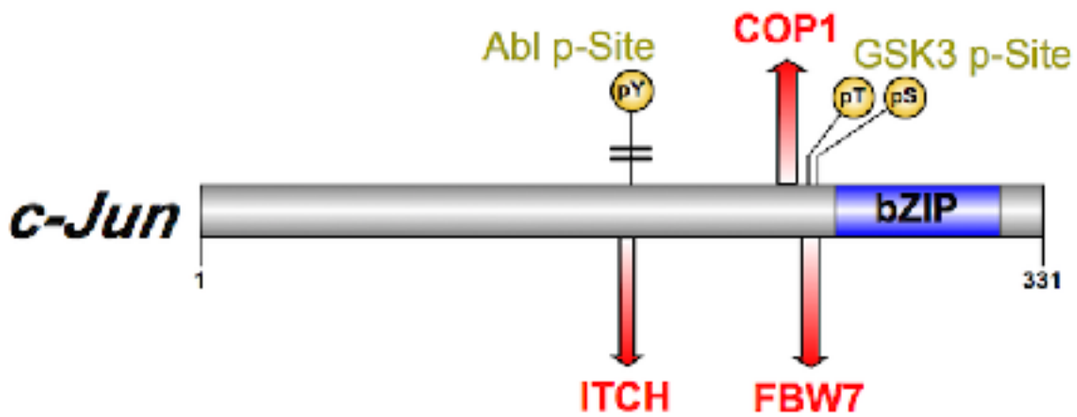
Degrons are the elements that are used by E3 ubiquitin ligases to targ
short linear motifs embedded within the sequences of modular prote
they are important for many different cellular processes, such as prog
cellular hypoxia. Degrons enable the elimination of proteins that are no longer required, preventing their possible dysfunction. Although the human genome encodes ~600 E3 ubiquitin ligases, only a fraction of these enzymes have well-defined target degrons. Thus, for most cellular proteins, the destruction mechanisms are poorly understood. This is important for many diseases, especially for cancer, a disease that involves the enhanced expression of oncogenes and the persistence of encoded oncoproteins coupled with reduced abundance of tumor suppressors. Loss-of-function mutations occur in the degrons of several oncoproteins, such as the transcription factors MYC and NRF2, and in various mitogenic receptors, such as NOTCH1 and several receptor tyrosine kinases. Mutations eliminating the function of the β-catenin degron are found in many cancers and are considered one of the most abundant mutations driving carcinogenesis. In this Review, we describe the current knowledge of degrons in cancer and suggest that increased research on the "dark degrome" (unknown degron-E3 relationships) would enhance progress in cancer research.

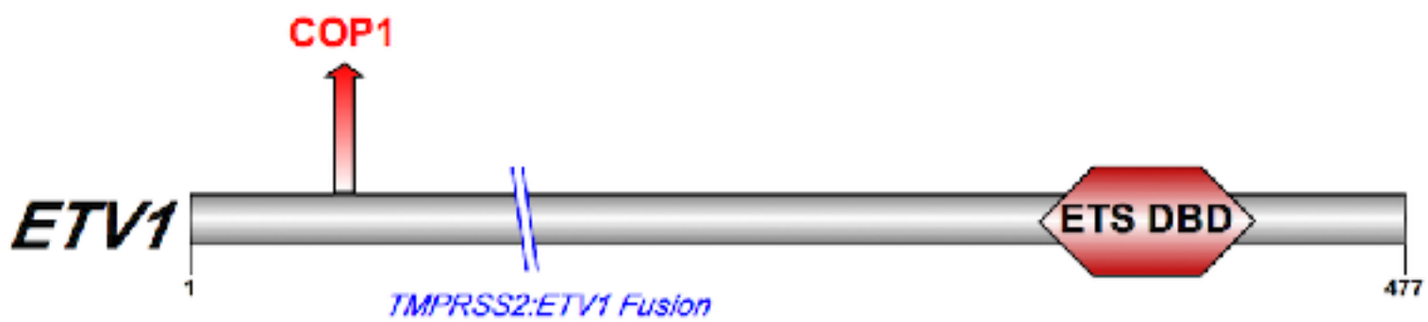
Proteins are recycled by the ubiqutin-proteasome system (UPS)

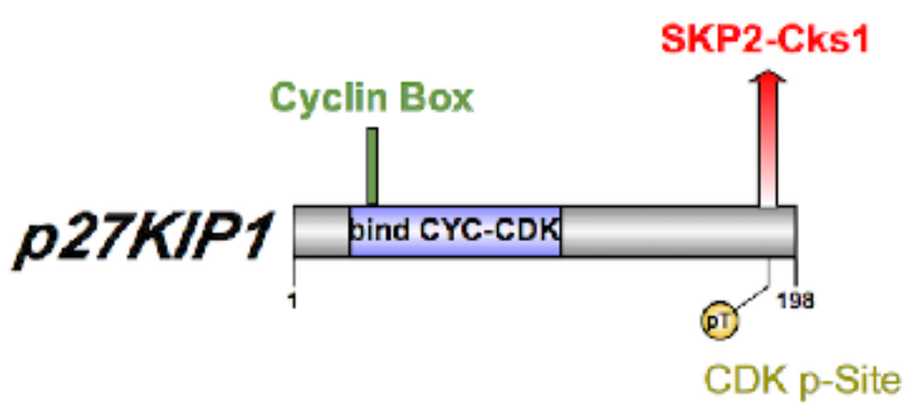


E3 ligases recognise "degrons" on the target proteins


MDM2
NES
p53
TAD
p53 DBD
4X
1
393
pT
pS
NLS
USP7 Dock


Abl p-Site
COP1
GSK3 p-Site
pY
pT
pS
c-Jun
bZIP
1
331
ITCH
FBW7


COP1
ETV1
ETS DBD
1
477
TMPRSS2:ETV1 Fusion


SKP2-Cks1
Cyclin Box
p27KIP1
bind CYC-CDK
1
198
pT
CDK p-Site

A. p53 - MDM2

B. p27 - SKP2-CKS1

C. Beta-Catenin - TrCP

D.TRB1 - COP1

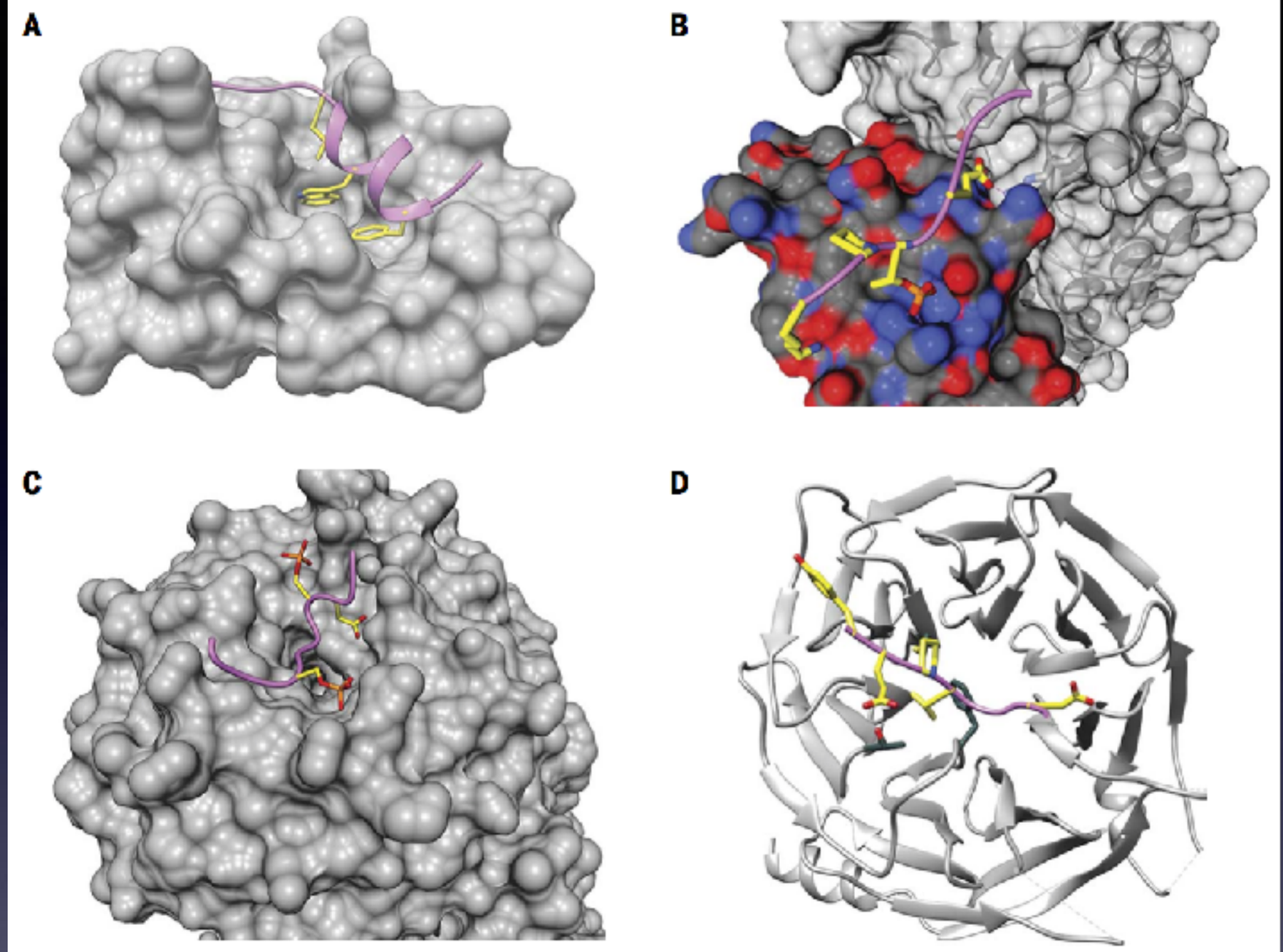


Fig. 3. E3 ligase–degron complexes. (**A**) p53-MDM2, the p53 degron peptide, enters a deep hydrophobic pocket on MDM2. The three key hydrophobic residues from p53 peptide are shown [Protein Data Bank (PDB): 1YCR]. (**B**) Structure of SKP1-SKP2-CKS1 in complex with the $p27^{KIP1}$ phosphodegron peptide. SKP2 is rendered in gray surface, whereas CKS1 is represented in dark gray surface, with red (oxygen) and blue (nitrogen) polar functional groups. The small peptide from the $p27^{KIP1}$ is rendered in ribbon representation with purple color: Its phosphothreonine interacts with positively charged CKS1 surface (blue region), which provides the phosphospecificity. SKP1 is not shown for clarity (PDB: 2AST). (**C**) Doubly phosphorylated β-catenin degron motif (see Table 1) in complex with β-TrCP1 and SKP1. Both of these molecules are shown in gray surface representation, whereas the β-catenin fragment is rendered as purple ribbon (PDB: 1P22). (**D**) COP1 E3 ligase with TRB1 degron. TRB1 binds on the conserved surface of COP1 in an extended manner. COP1 with WD40 repeats (β-propellers) is rendered as ribbon, whereas TRB1 peptide is rendered in stick representation and colored purple. (PDB: 5IGQ). Interchain H bonds are represented in magenta. Motif-defining positions have been rendered as yellow sticks for all the cases.

Many degrons get mutated in cancer

-

But only in tumour promoters, never in tumour suppressors

-

Degrons will need to be scored for targeted therapies

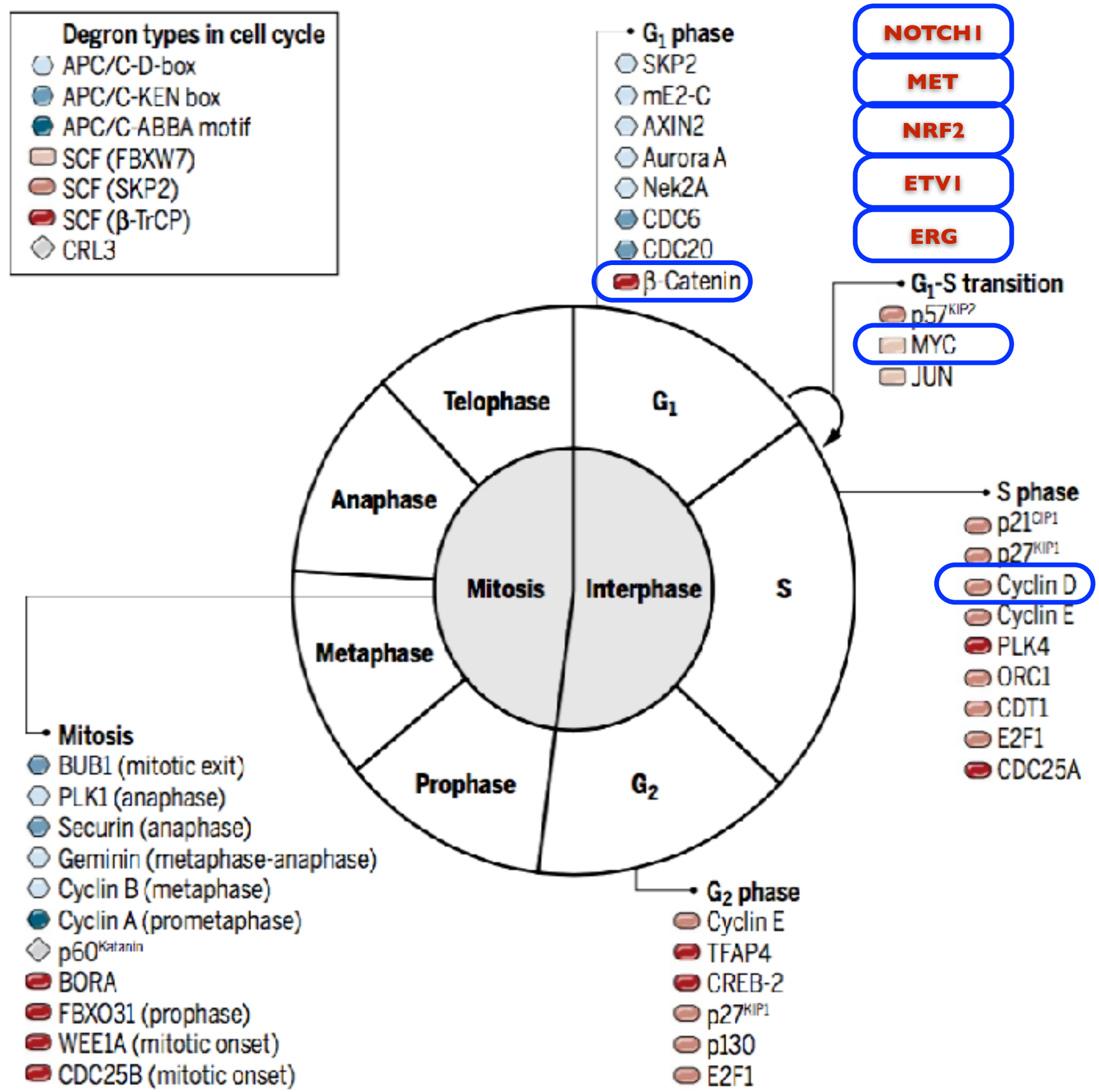
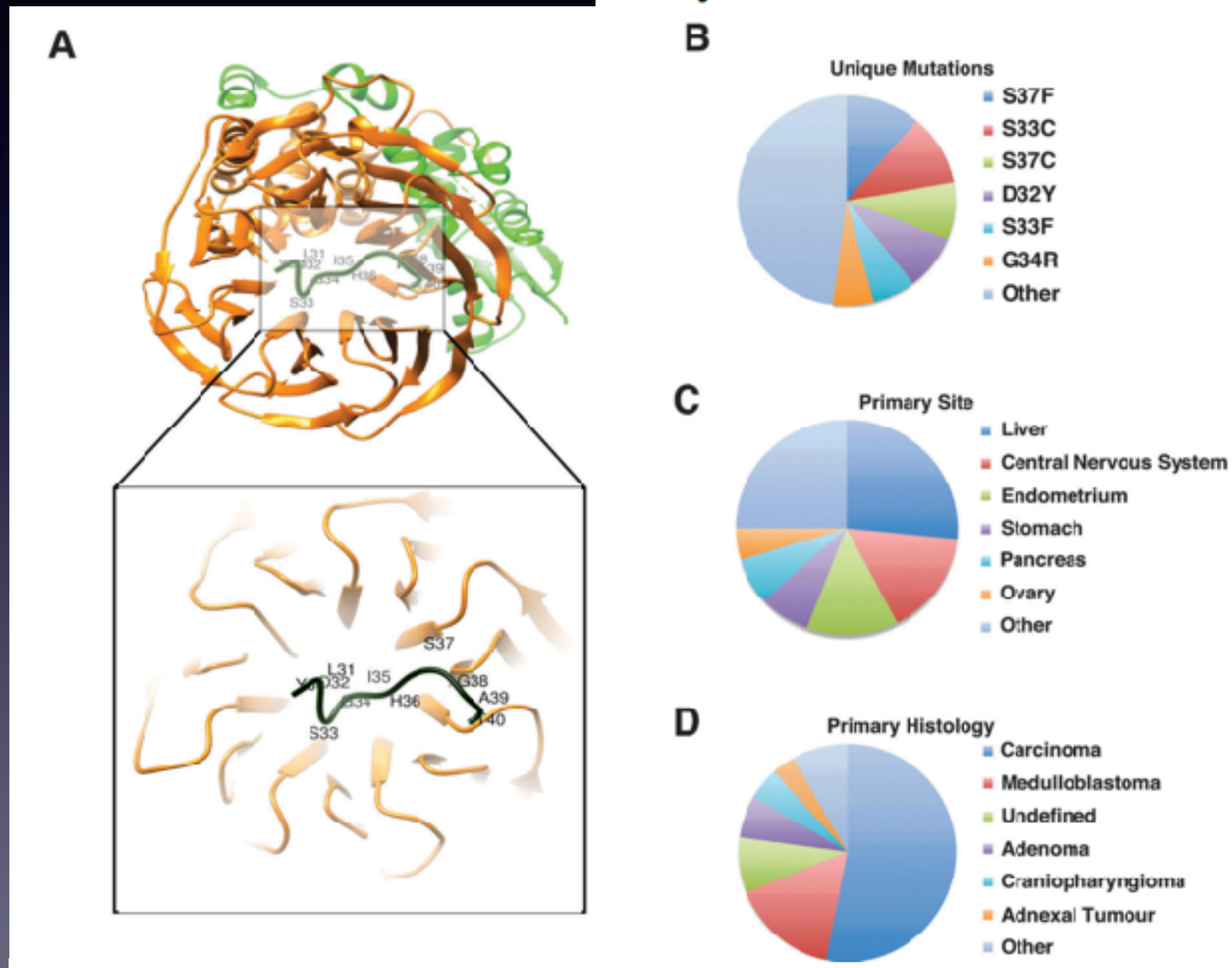


Fig. 6. Cell cycle phase diagram showing where degrons function in the cycle. Degron-containing protein names are categorized adjacent to the cell cycle phase during which the degradation of the protein enables progression of the cell cycle.

Proteome-wide analysis of human disease mutations in short linear motifs: neglected players in cancer?†

Bora Uyar,*[a] Robert J. Weatheritt,[bc] Holger Dinkel,[a] Norman E. Davey[ad] and Toby J. Gibson*[a]

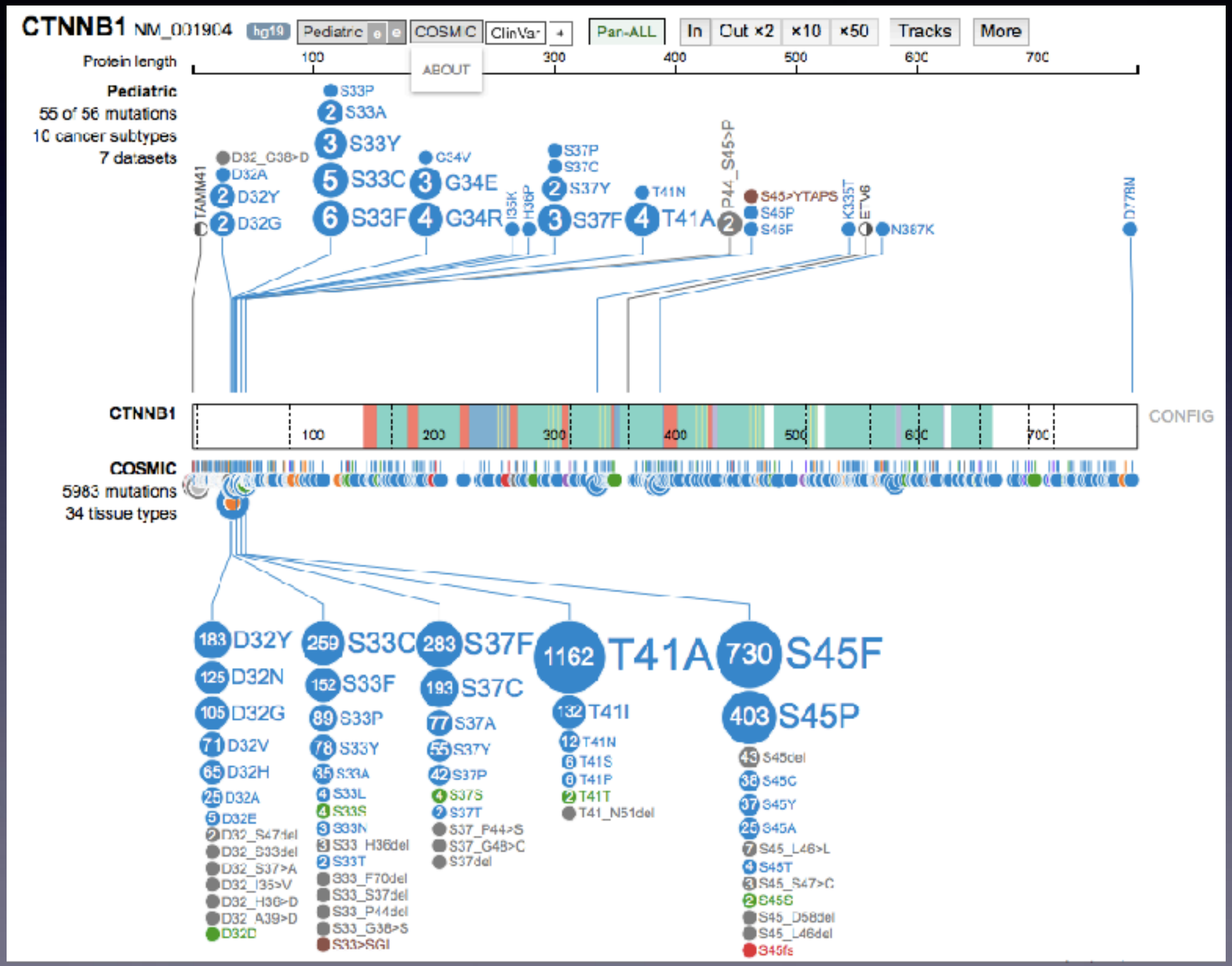
Mol. BioSyst., 2014, **10**, 2626–2642



The Beta-Catenin DSGxxS degron is the most mutated SLiM in cancer

According to the available disease-related missense mutation datasets, the most recurrently mutated experimentally validated SLiM is the conserved proteasomal degradation motif ("degron") in the highly disordered N-terminal region of β-catenin (Fig. 3A). This motif (DEG_SCF_TRCP1_1, $_{32}D_PSGIH_PS_{37}$)

Cancer mutations in Beta-Catenin target its degron (displayed by PeCan, St Judes)



Degrons are at the **nexus** between mutation and gene expression changes in cancer

Their roles in cancer are a manifestation of the importance of correct gene dosage in cell regulation

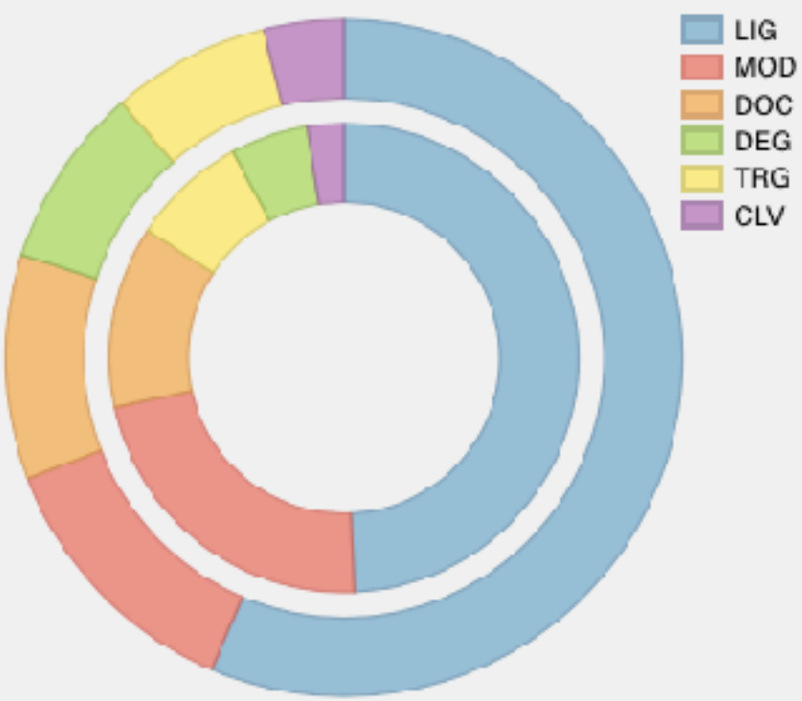
Nucleic Acids Research, 2019 1
doi: 10.1093/nar/gkz1030

ELM—the eukaryotic linear motif resource in 2020

Manjeet Kumar[1,*,†], Marc Gouw[1,†], Sushama Michael[1,†], Hugo Sámano-Sánchez[1,2], Rita Pancsa[3], Juliana Glavina[4], Athina Diakogianni[1], Jesús Alvarado Valverde[1], Dayana Bukirova[1,5], Jelena Čalyševa[1,2], Nicolas Palopoli[6], Norman E. Davey[7], Lucía B. Chemes[4,*] and Toby J. Gibson[1,*]

ELM database statistics

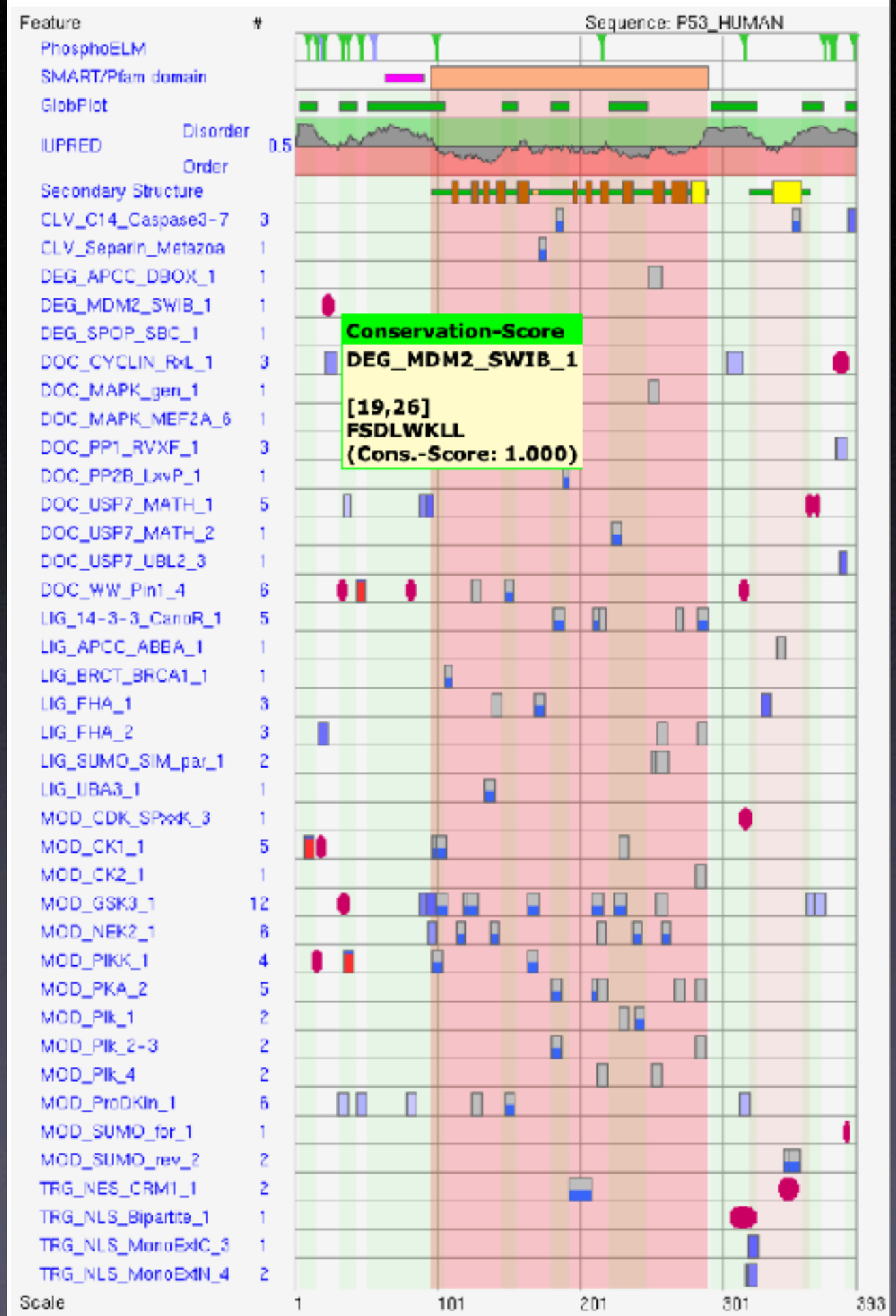
Classes & Instances



Name:	# Elm classes:	# Elm instances:
LIG:	163	1,738
DOC:	31	442
TRG:	22	281
CLV:	11	89
DEG:	25	188
MOD:	37	785
Total:	289	3,523

ELM is at http://elm.eu.org

p53 output of the ELM server: Red indicates the annotated motifs



Robustness of biological systems

Complexity and robustness

J. M. Carlson*† and John Doyle†

*Department of Physics, University of California, Santa Barbara, CA 93106; and †Control and Dynamical Systems, California Institute of Technology, Pasadena, CA 91125

Carlson and Doyle (2002) PNAS, 66, 2538

...By robustness, we mean the maintenance of some desired system characteristics despite fluctuations in the behavior of its component parts or its environment....

BIOLOGICAL ROBUSTNESS

Hiroaki Kitano

Abstract | Robustness is a ubiquitously observed property of biological systems. It is considered to be a fundamental feature of complex evolvable systems. It is attained by several underlying principles that are universal to both biological organisms and sophisticated engineering systems.

Kitano (2004) Nat Rev Genet, 5, 826

Increases in system complexity, due to selection for robustness, introduce a new issue: **system fragility**

A good example is the Internet which is:

"robust yet fragile" (RYF)

that is, unaffected by random component failures but vulnerable to targeted attacks on its key components (e.g. DDoS attacks).

Sify.com / Finance / Crime / Indian education institutions hit hard by hackers: Report

Indian education institutions hit hard by hackers: Report

Viral hijack of cell regulation:

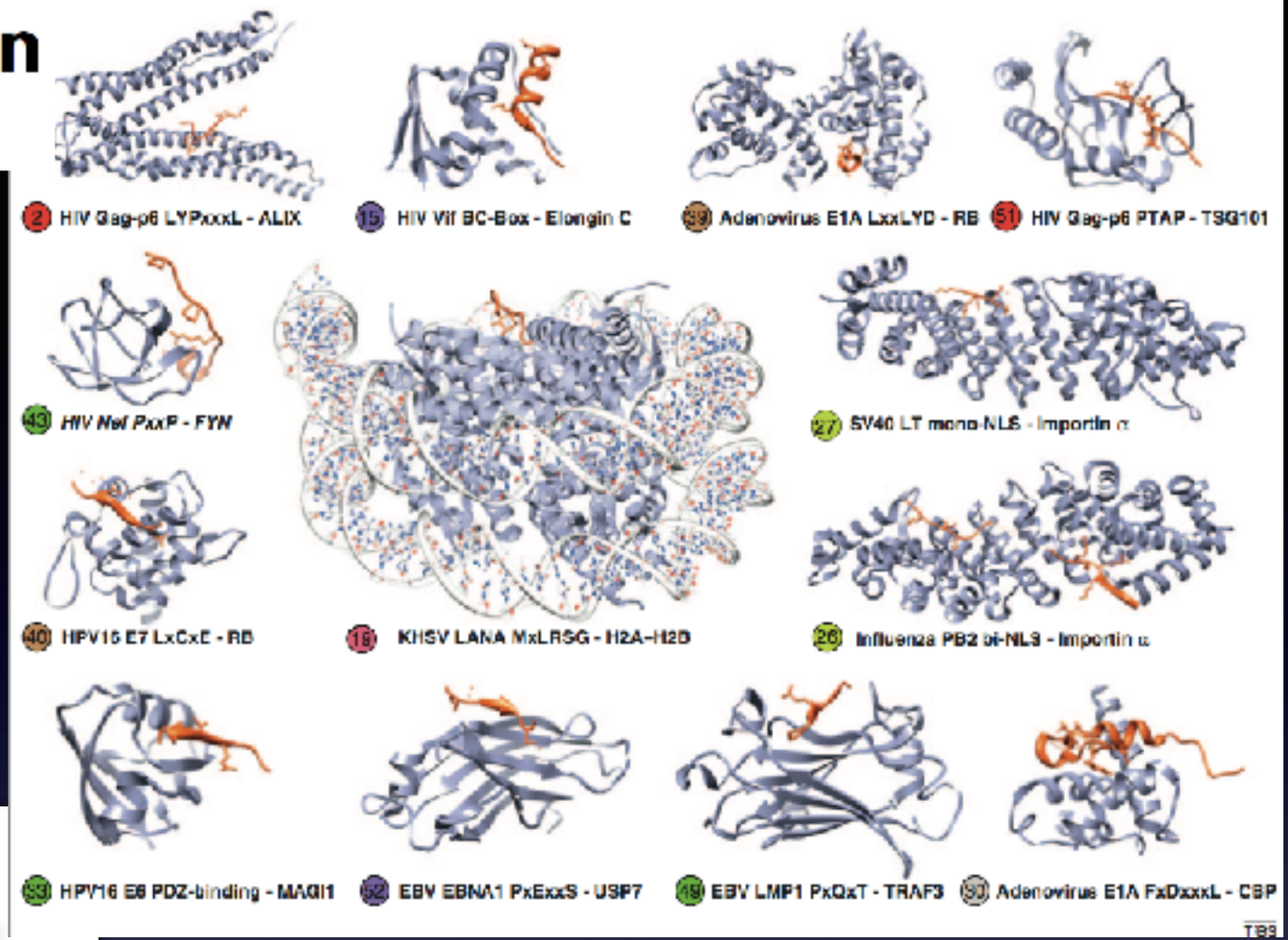
Fragility in a complex system

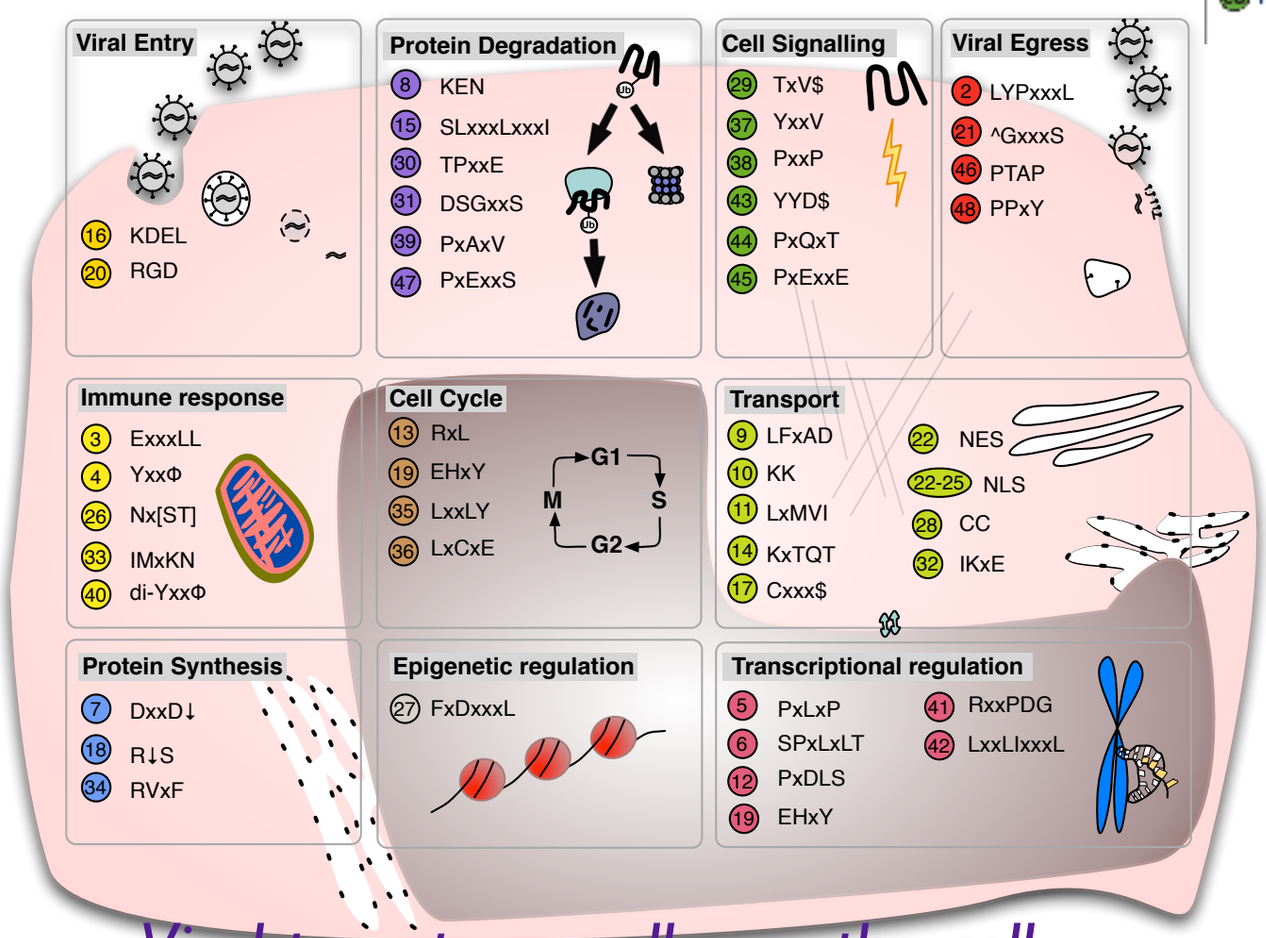
How viruses hijack cell regulation

Norman E. Davey[1], Gilles Travé[2] and Toby J. Gibson[1]

TiBS (2011) 36, 159

More than a third of the motif classes annotated in our ELM Resource (http:elm.eu.org) are already known to be used by viruses





Viral targets are all over the cell

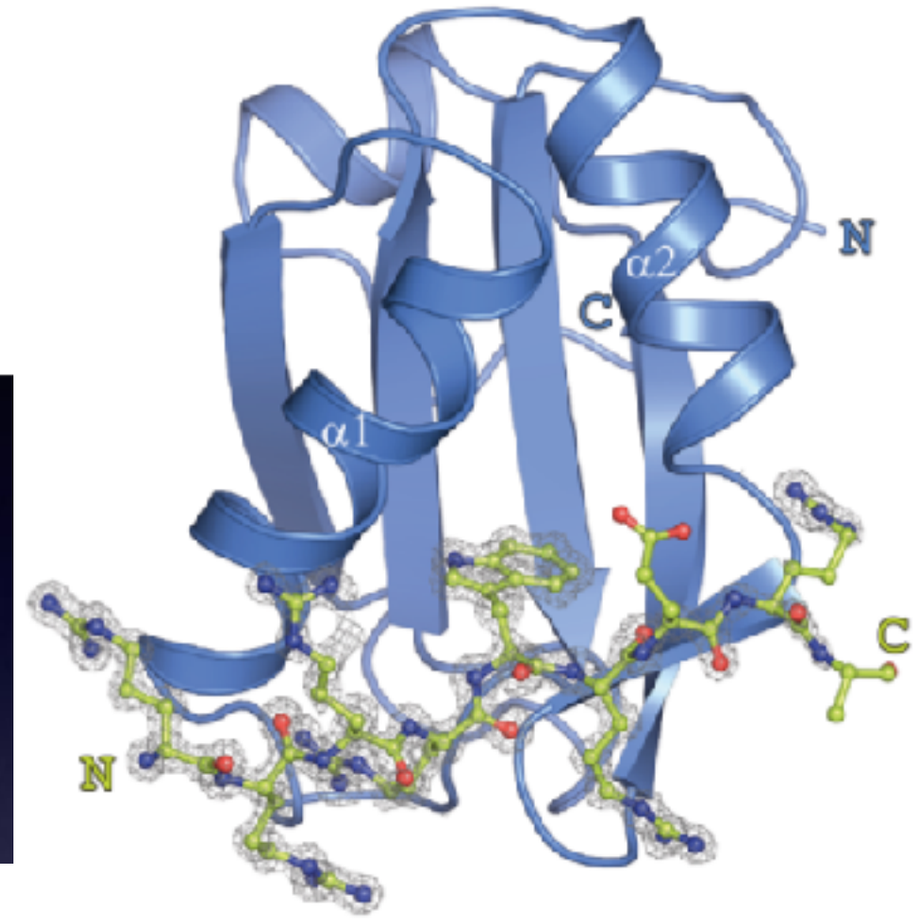
- Why is there "always" a cellular protein motif interaction for a virus to subvert?
- What does this tell us about the nature of the cell?

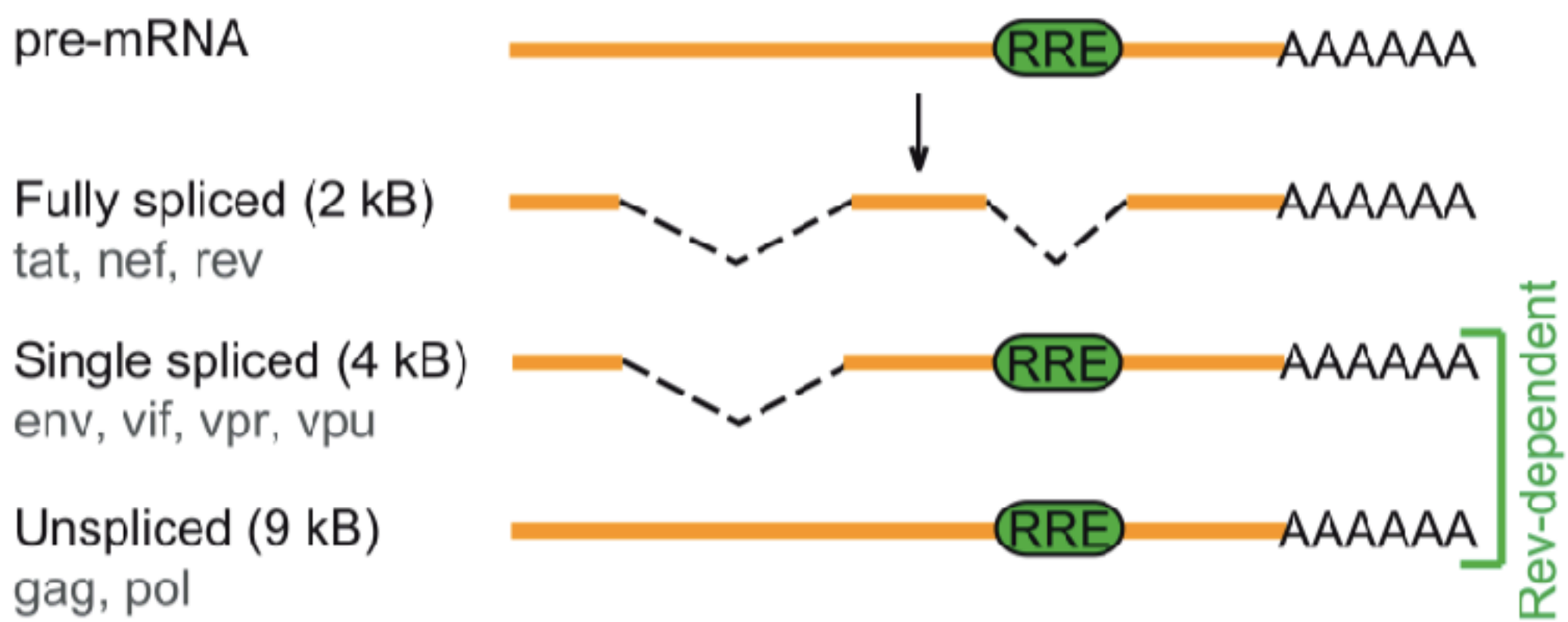
Published online 20 March 2019 Nucleic Acids Research, 2019, Vol. 47, No. 9 4859–4871 A
doi: 10.1093/nar/gkz185

Modulation of HIV-1 gene expression by binding of a ULM motif in the Rev protein to UHM-containing splicing factors

Marta Pabis[1,2,†], Lorenzo Corsini[1,2,†], Michelle Vincendeau[3,4], Konstantinos Tripsianes[5], Toby J. Gibson[6], Ruth Brack-Werner[3] and Michael Sattler[1,2,*]

```
REV_HV1H2   QARRNRRRRWREERQRQIHSI
REV_HV1OY   QARRNRRRRWREERQRQIRKI
REV_HV1Y2   QARRNRRRRWREERQRQIRSI
REV_HV1Z6   QARRNRRRRWRARQRQIHSI
REV_HV2D2   SQRRNRRRRWKRRGLQILAL
REV_HV2SB   RQRRNRRRRWERRWKQILAL
REV_HV2RO   SQRRNRRRRWKQRWRQILAL
REV_SIVCZ   KARRNRRRRWRARQKQISEI
REV_SIVMK   NQRRQRRRRWRRRWQQLLAL
REV_SIVVG   RQRRRARRRWRQAQEQLRAL
REV_SIVSP   NQRRRRRRRWRQRWQQILAL
```

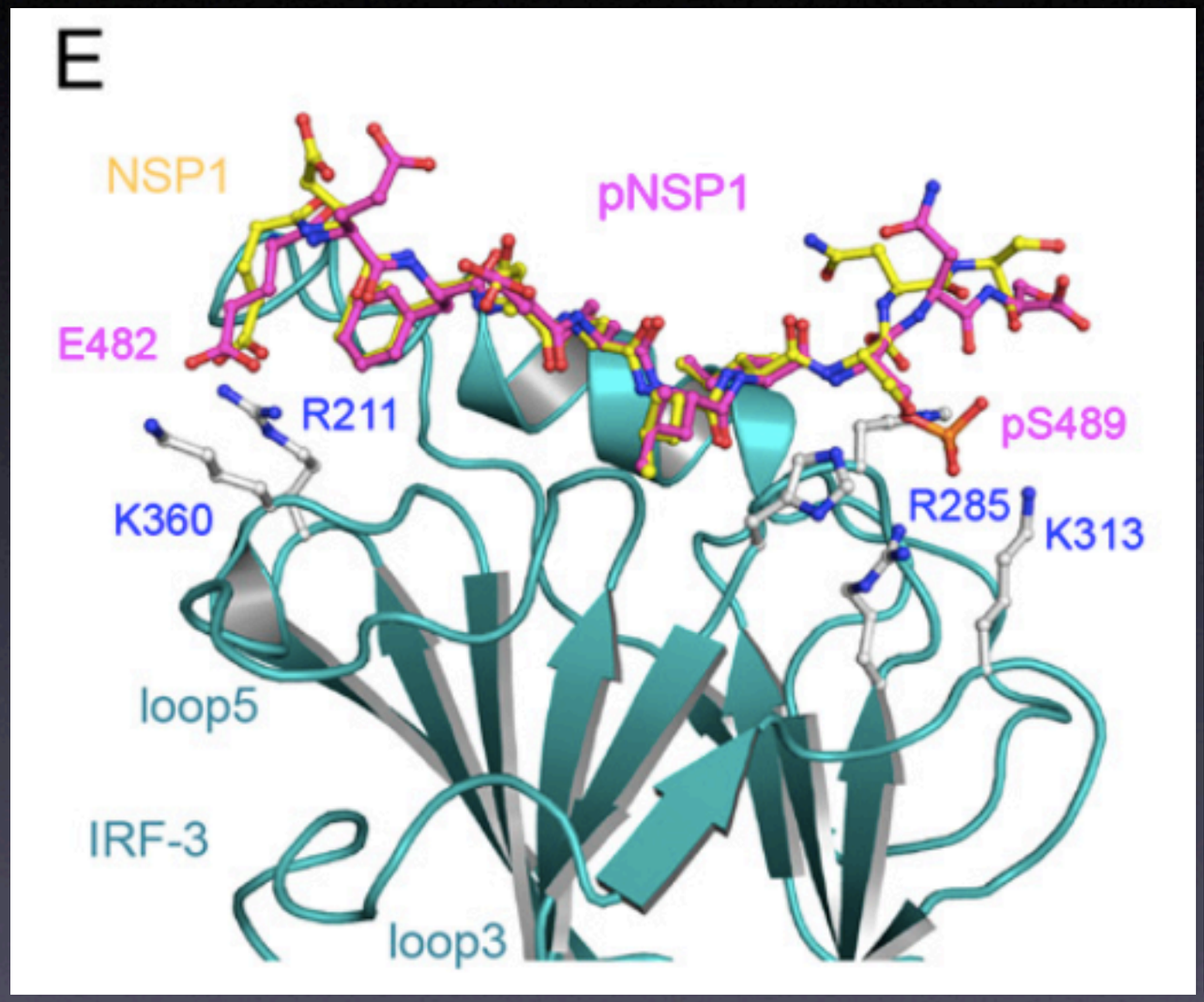




Structural basis for concerted recruitment and activation of IRF-3 by innate immune adaptor proteins

Baoyu Zhao[a,1], Chang Shu[a,1,2], Xinsheng Gao[b], Banumathi Sankaran[c], Fenglei Du[a], Catherine L. Shelton[d,e], Andrew B. Herr[d,e], Jun-Yuan Ji[b], and Pingwei Li[a,2]

approved April 27, 2016



Significance

Type I IFNs are key cytokines involved in antiviral immunity. A number of innate sensing pathways regulate the induction of type I IFNs. These pathways converge at the activation of the transcription factor IRF-3 (IFN regulatory factor 3). Three different adaptors mediate the recruitment of IRF-3 using a conserved structural motif. In this study, we determined the molecular mechanisms by which these adaptors recruit IRF-3 upon phosphorylation, the mechanism of IRF-3 activation, and how rotavirus subverts these signaling mechanisms to evade innate immune surveillance. These results provide critical insights into the molecular basis of innate immunity against microbial and viral infections.

The *p*L*x*IS motif of rotavirus NSP1 targets IRF-3 for degradation

See ELM Entry LIG_IRF3_LxIS_1

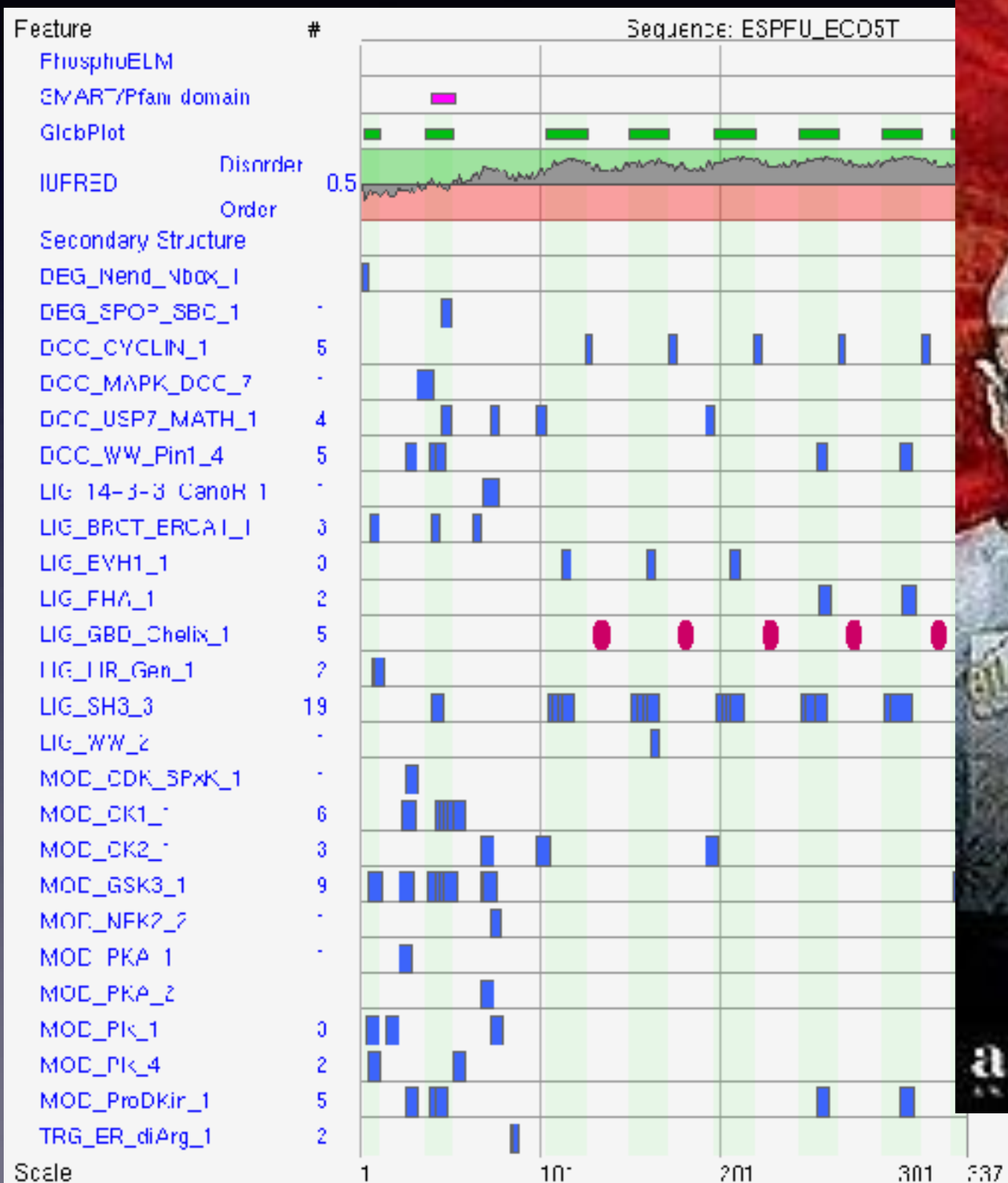
Historical Fact:
Virologists first identified many of the short linear motifs used by the cell

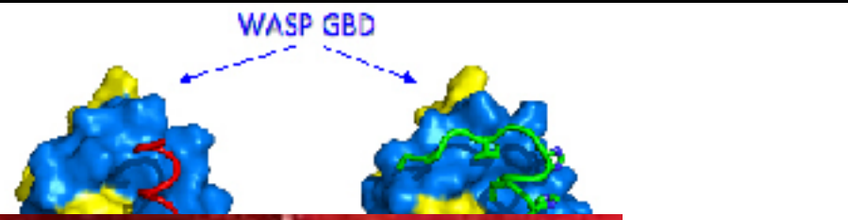
There will be many more to find: Even so, viral motif biology is well ahead of the invasive cellular pathogen field

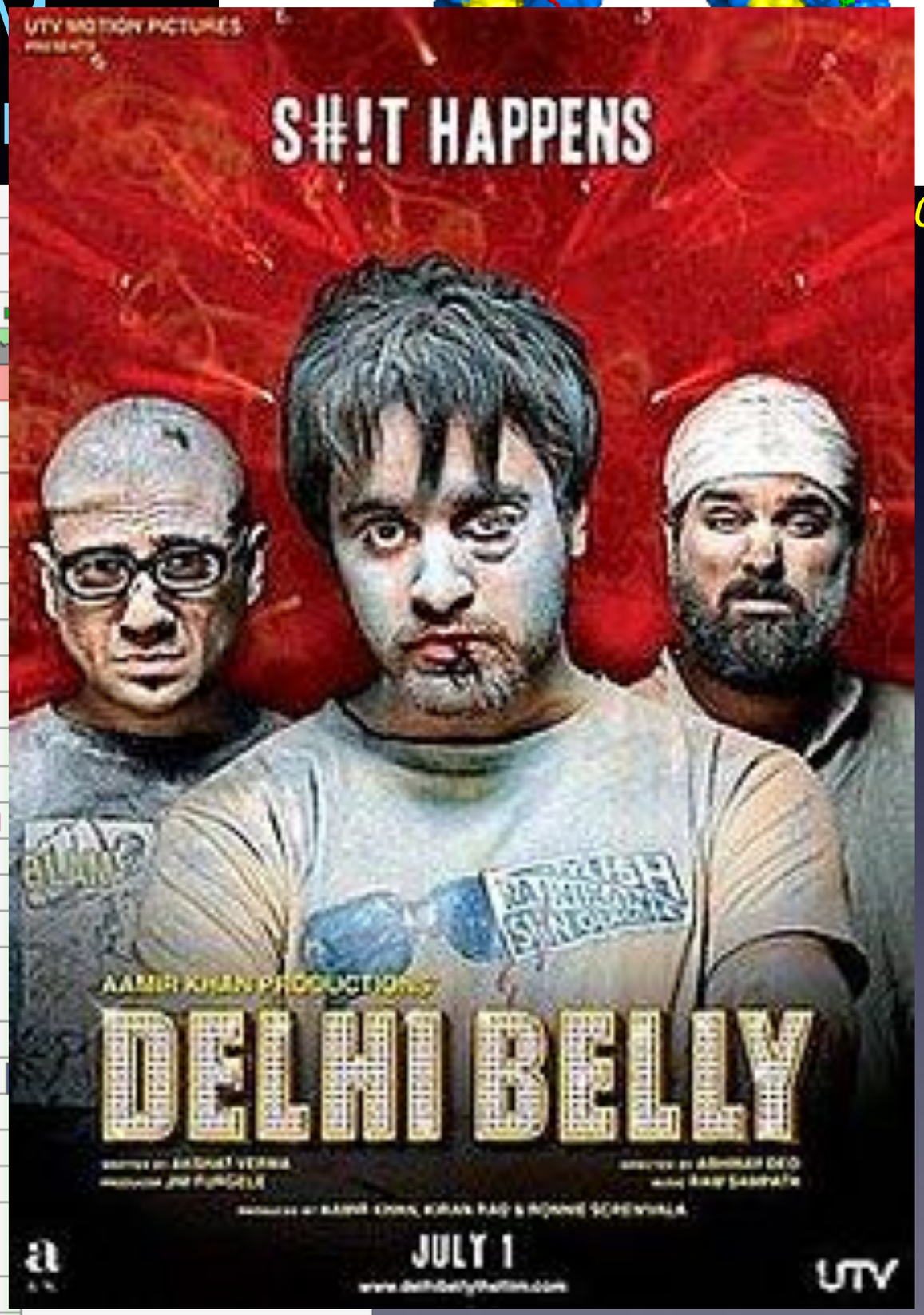
Living pathogens use three main approaches to perturb host cells

- Toxins
  - *Staphylococcus*: Pore-forming toxin
    - *The cell will die*
- Enzymatic irreversible inactivation
  - *Yersinia*: Acetylates MapK activation loop Ser and Thr
    - *The cell will eventually die*
- Linear motif hijack
  - *Toxoplasma*: MapK docking motifs activate MapK
    - *Cell can be re-tuned and kept alive indefinitely*

E. coli EHEC EspFu effector protein output from ELM
Repeated Actin regulatory

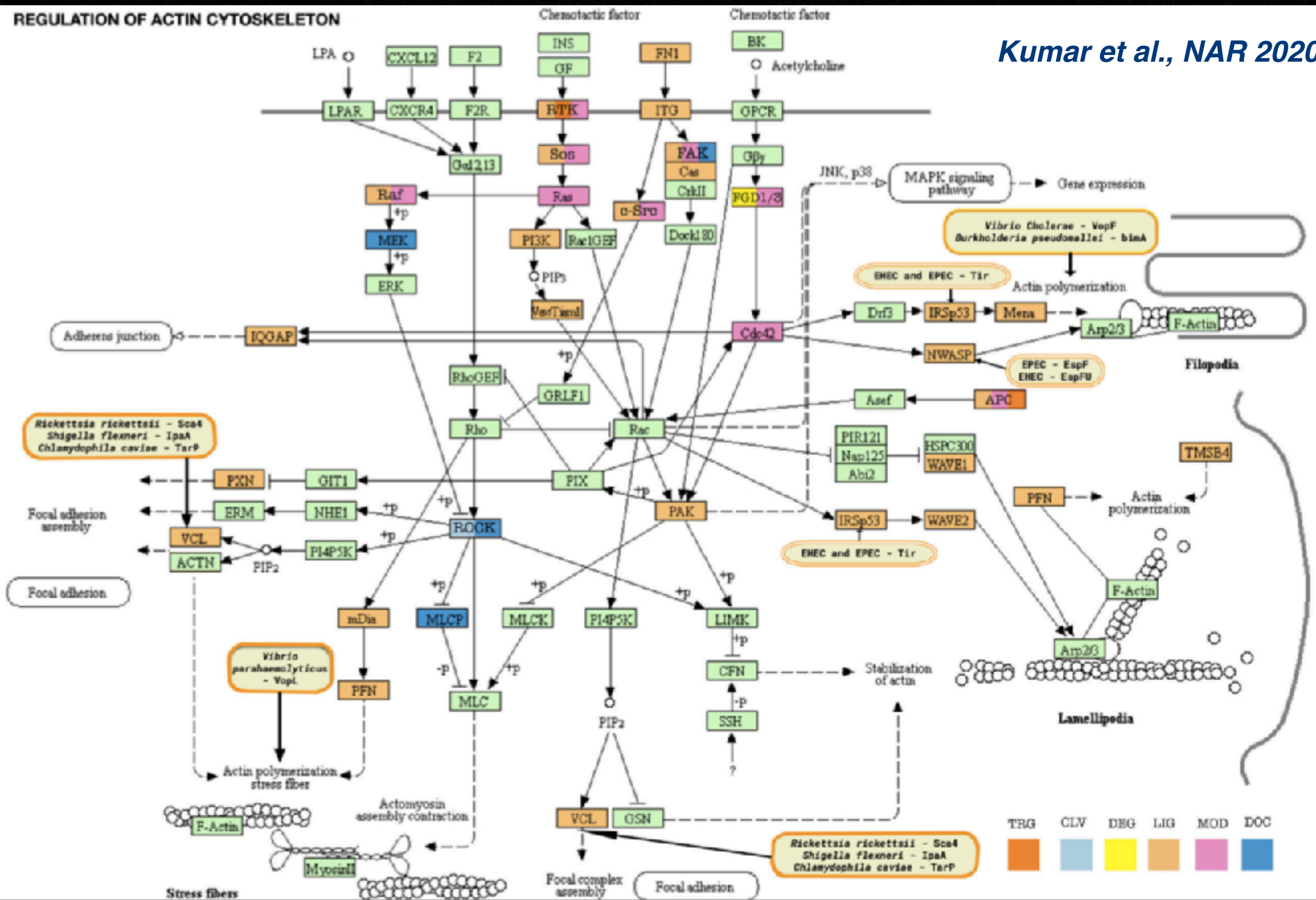




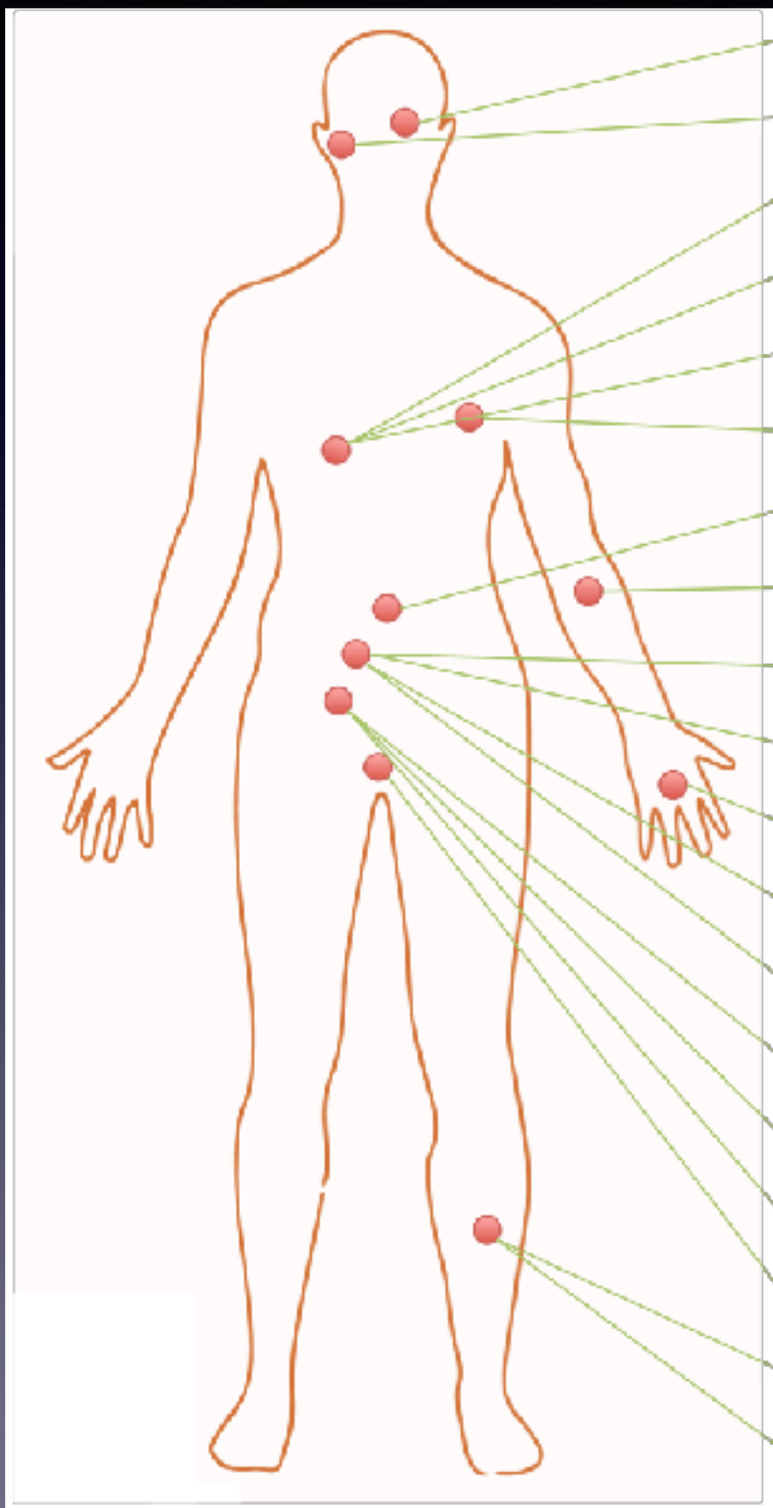


09 (2008)

Bacterial pathogens modulate actin filaments from many points in the actin regulatory network



We are collecting reported examples of effector SLiMs for benchmarking e.g. Human Bacterial Pathogens Using Motif Mimics



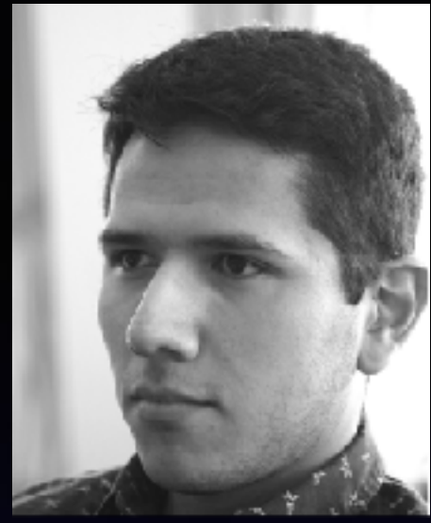
Chlamydia trachomatis: IncG	14-3-3 binding motif
Pseudomonas aeruginosa: PilY1, ExoA	RGD binding motif, KDEL
Mycobacterium tuberculosis: GarA	FHA phosphopeptide ligand
Burkholderia mallei: BimA	WH2 domain-binding motif
Legionella pneumophila: Lpl2806	FHA phosphopeptide ligand
Corynebacterium glutamicum: OdhI	Prenyl
Helicobacter pylori: CagA, CagL	EPIYA,
Listeria monocytogenes: ActA	EVH1 b
***E. coli* ETEC:** LTIIA	KDEL
***E. coli* EPEC:** Tir, EspI , Map, EspF	SH2, P
Bartonella henselae: BepD, BepE, BepF	EPIYA
***Salmonella* Typhimurium:** SpvC	MAPK
***E. coli* EHEC:** EspF(U), NleH, Tir	SH3, P
Vibrio cholerae: ctxA, VopF	KDEL, WH2 domain-binding motif
Vibrio parahaemolyticus: VopL	WH2 domain-binding motif
Shigella flexneri: OspF, IpaA	MAPK docking motif, Vh1 binding motif
Haemophilus ducreyi: LspA1, LspA2	EPIYA
Anaplasma phagocytophilum: AnkA, AmpA	EPIYA, sumoylation site
Rickettsia rickettsii: Sca4	Vh1 binding motif

Hugo Carlos Tue 4:45 p.m.
Corynebacterium axilliary
Bartonella henselae produces cat scratch
C. trachomatis: genitourinary, pulmonary, ocular
P. aeruginosa: ear infections
Anaplasmosis, by a tick also, detected in blood.
H. ducreyi: Chancro in genitals
Listeria: blood, spinal fluid, placenta
Vibrio: diagnosed by a rectal swab
Shigella: detected in stools
Salmonella: Blood and intestinal cell
Ricketsia: by a tick
Burkholderia mallei: lung, bloodstream

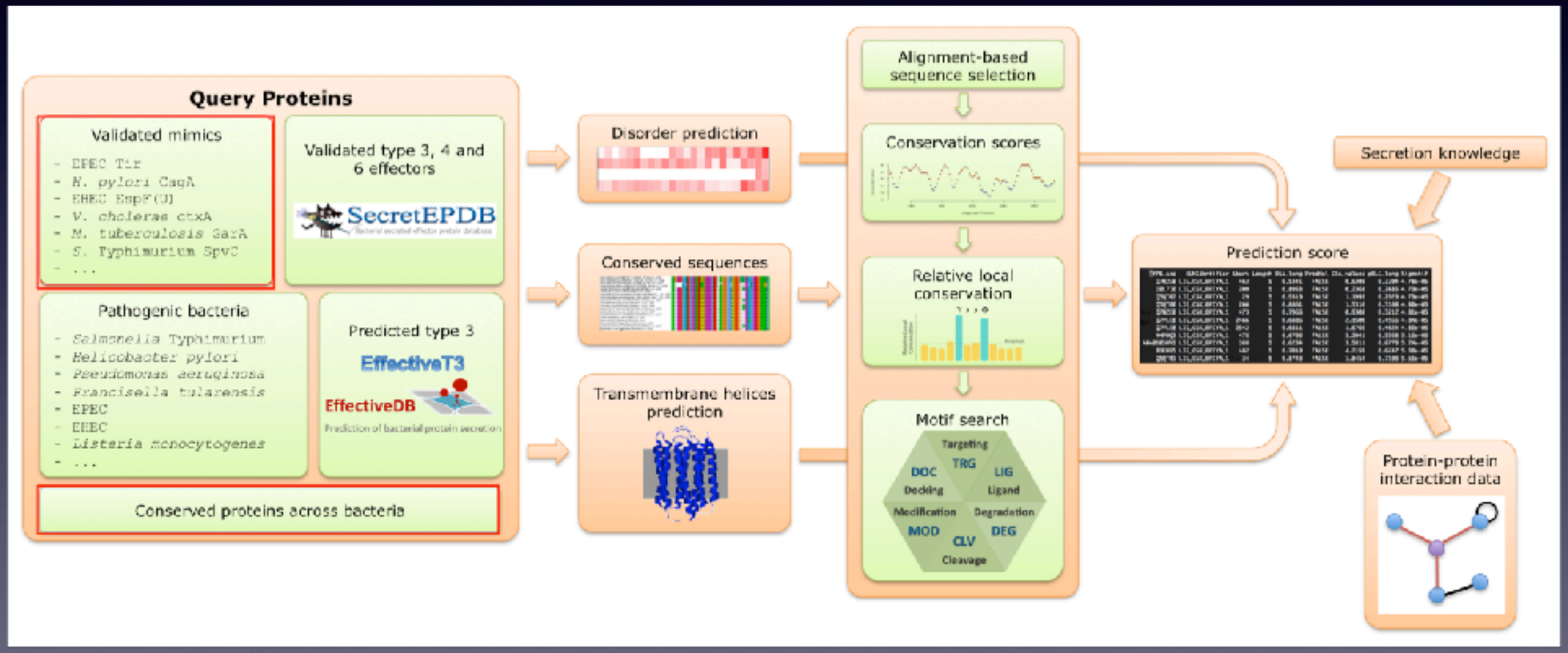
Slide from Hugo Samano

Pipeline to predict SLiM mimics in bacterial proteins

Motif / Disorder / Conservation / Secreted(?)

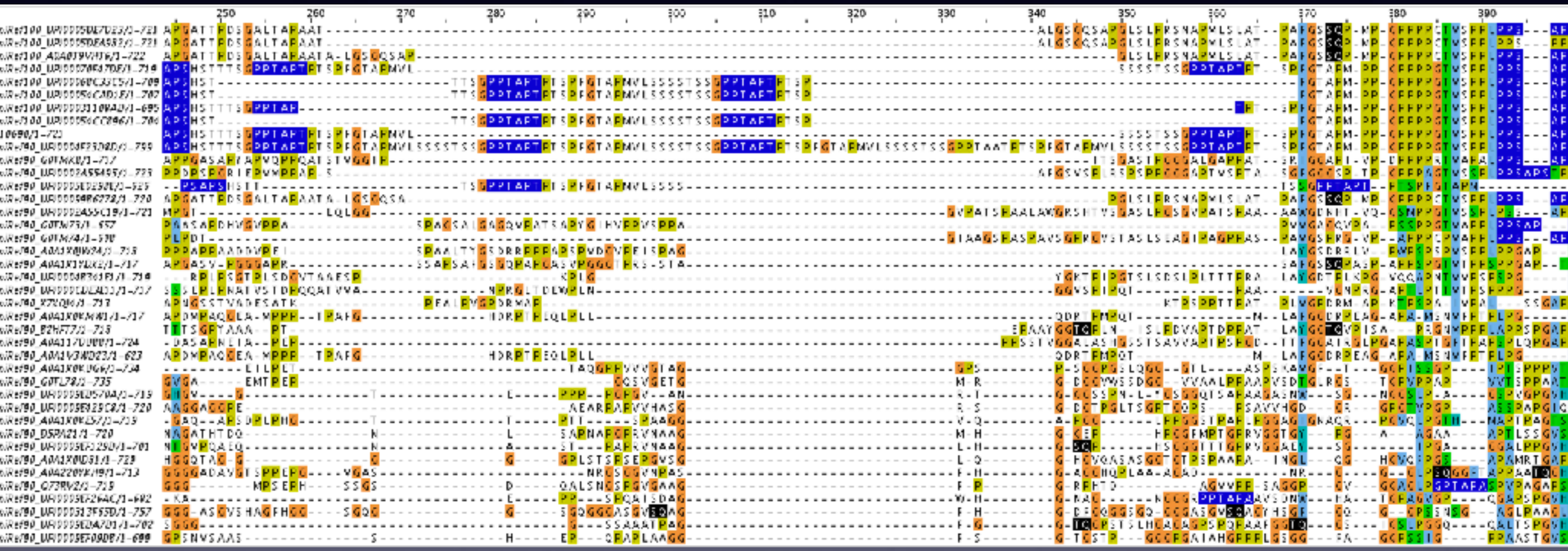


Hugo Samano



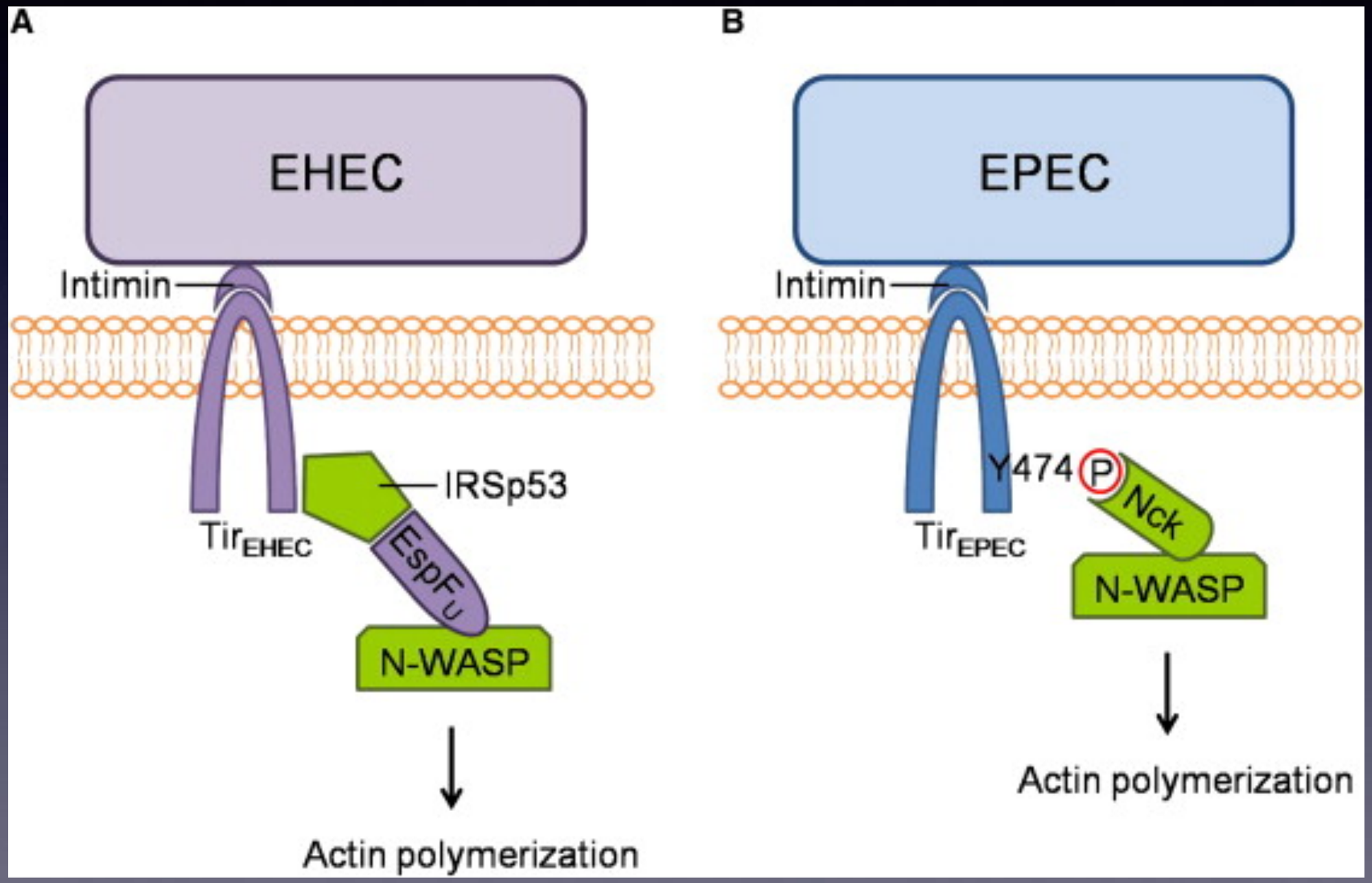
Prediction Score (Logistic Regression) = f(disorder prediction, conservation, RLC, evolutionary distance)

Mycobacterium tuberculosis protein Rv2082 contains multiple candidate TSG101 interacting PTAP motifs which could potentially play a role in ESCRT-mediated vesicle regulation



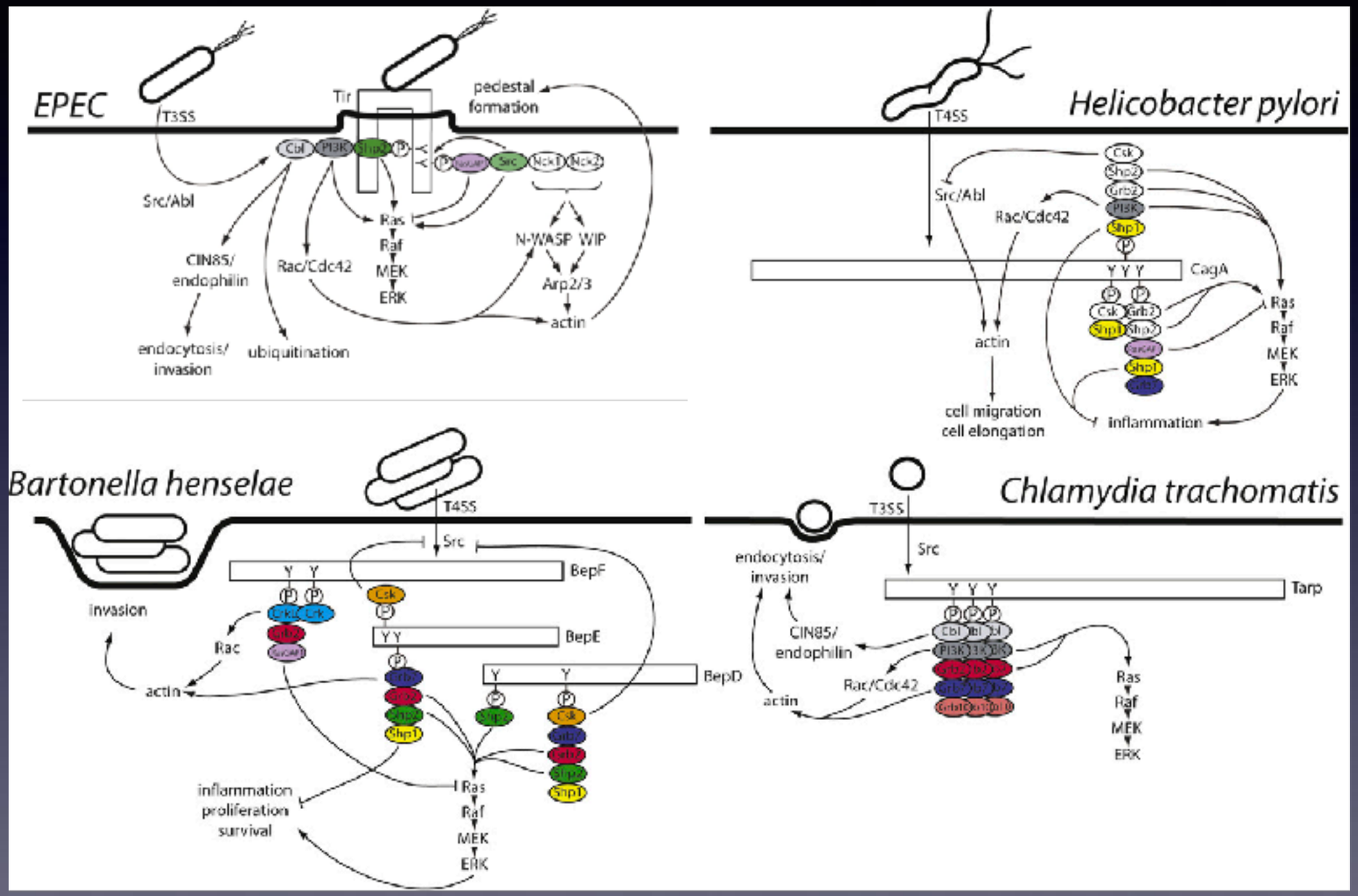
Other Mycoplasma do not have PTAP motifs in their versions of Rv2082

Pathogenic *E. coli* EHEC and EPEC use Abl and Src kinases to polymerise actin and make pedestals



Yi et al. Cell Host Micro. (2009)

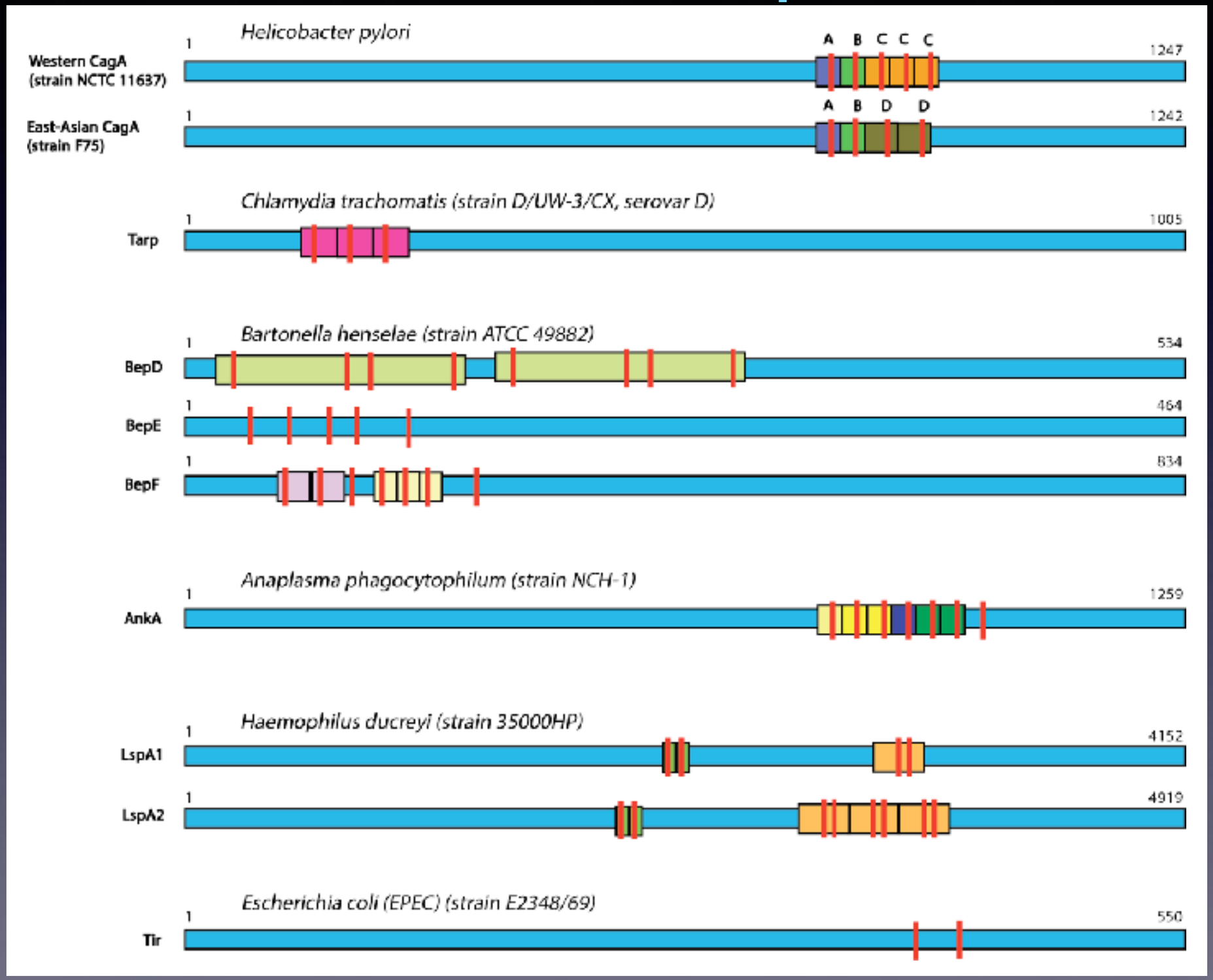
Cellular proteins that interact with tyrosine-phosphorylated bacterial proteins



SILAC experiment recovered interactions with tyrosine phosphorylated bacterial proteins

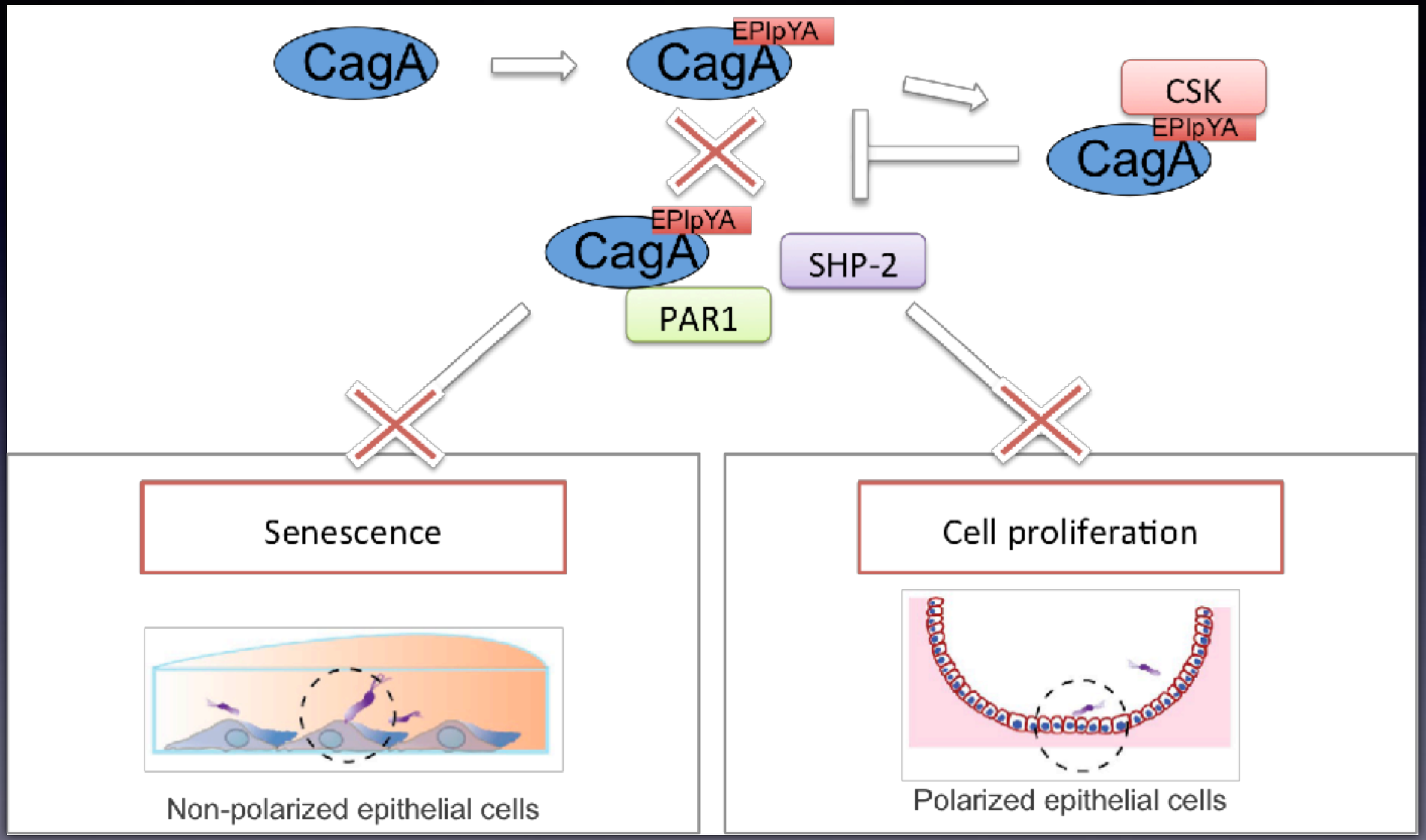
Selbach, M., et al. 2009

EPIYA-like motifs in *H. pylori* CagA and other bacterial effector proteins



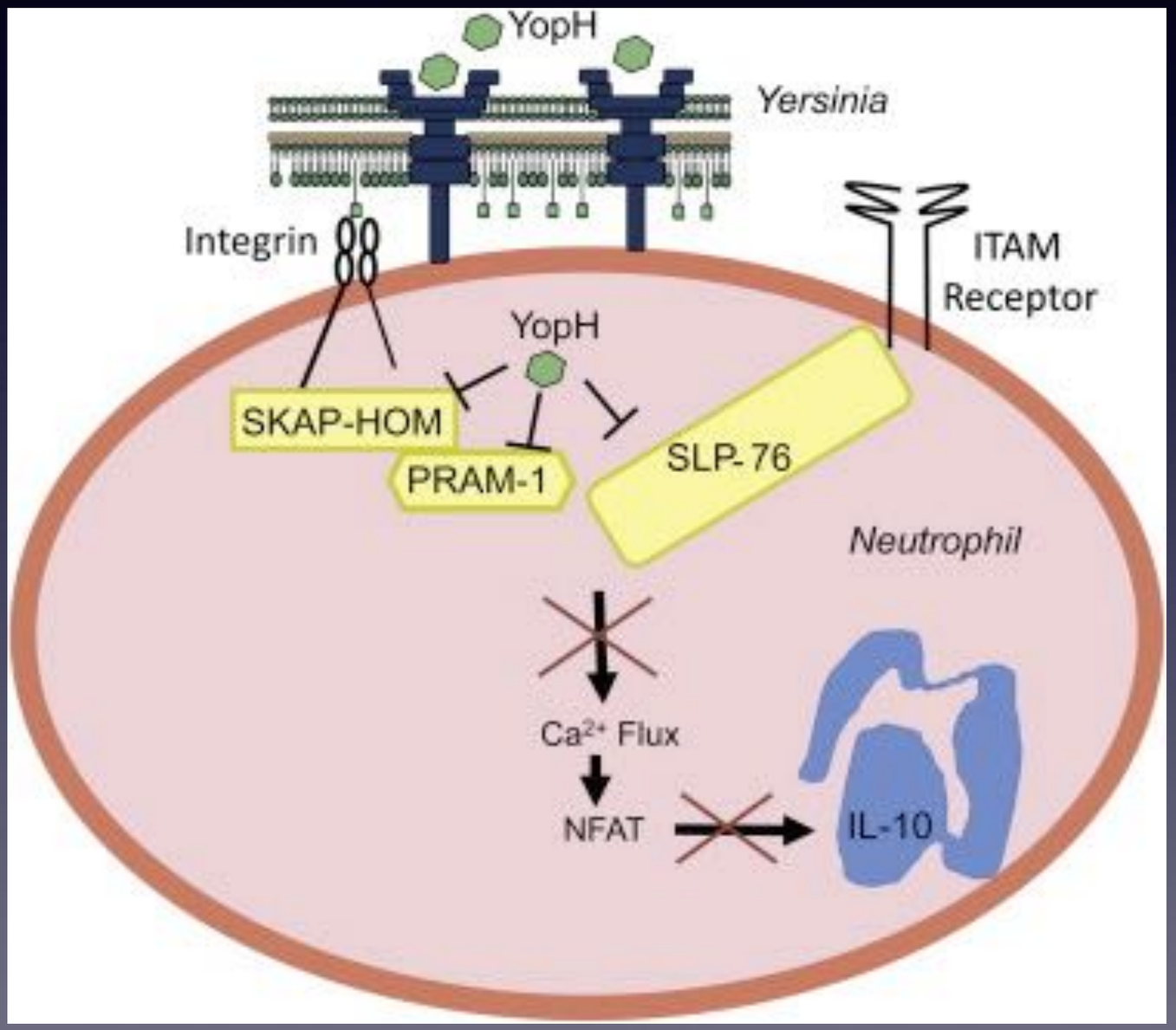
Hugo Samano, based on Hayashi, T., et al., 2013

Phosphorylated EPIYA Motifs in *H. pylori* enable CagA to activate CSK, inhibiting SRC-like kinases and phagocytosis as well as contributing to host cell fate



Tsutsumi, R., et al., 2003
Based on Saito, T., et al., 2010

Yersinia pestis introduces a tyrosine phosphatase to impair neutrophil responses



YopH selectively dephosphorylates host proteins

Calcium flux is blocked and cytokine production altered

Rolan, H. G., et al., 2013

Pseudomonas aeruginosa internalization is Abl kinase dependent

OPEN ACCESS Freely available online PLoS PATHOGENS

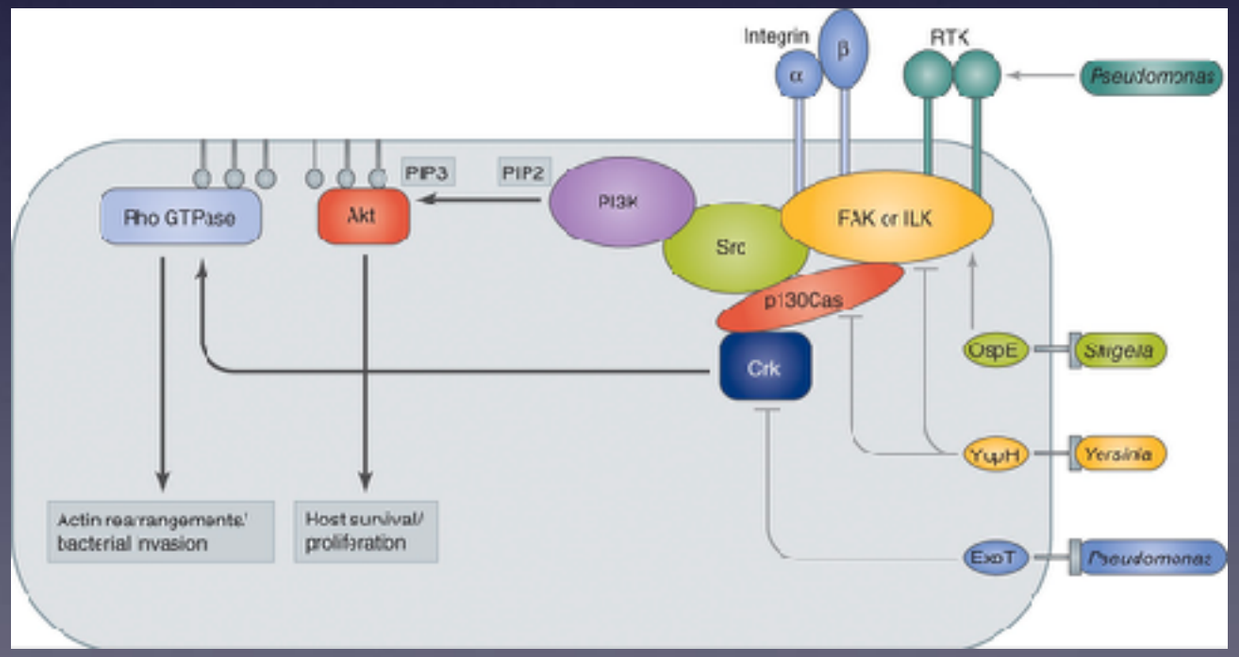
RNAi Screen Reveals an Abl Kinase-Dependent Host Cell Pathway Involved in *Pseudomonas aeruginosa* Internalization

Julia F. Pielage[1,2,3], Kimberly R. Powell[4], Daniel Kalman[4], Joanne N. Engel[1,2,3]*

~80 genes were inactivated using RNAi

Abl is a component of the host signaling pathway that leads to internalisation of *P. aeruginosa*

ExoT abrogates *P. aeruginosa*–induced Abl-dependent Crk phosphorylation



Pielage, J. F., et al., 2008
Krachler A. M., et al., 2011

"Drug the Cell to Cure the Pathogen?"

*The Concept of **Host-Directed Therapy** is not new: Originally **HDT** had a strong immunity focus*

Ann Intern Med. 1995 Feb 1;122(3):218-22.

Host-directed and immune-based therapies for human immunodeficiency virus infection.

Lederman MM[1].

Author information

1 Department of Medicine, University Hospitals of Cleveland, OH 44106.

Abstract

This essay reviews the rationale underlying host-directed or immune-based therapeutic strategies for human immunodeficiency virus (HIV) infection and its complications. These approaches have seen only limited trial in the past 10 years, but as more is learned about the immunopathogenesis of HIV disease and as the limitations of virus-directed therapies become more apparent, the need to pursue other therapeutic avenues has become increasingly important. Moreover, properly designed trials of host-directed and immune-based therapies can provide key insights into the pathogenesis of HIV disease that may be otherwise unattainable through in vitro studies.

Can we use Tyrosine Kinase inhibitors to treat bacterial infections?

Pathogenic perturbation of signalling networks might enable a general therapeutic strategy:

Can inhibitors of signalling set the cell into a condition that is hostile to the pathogen?

Would pathogens find it difficult to evolve resistence to such therapies?

Starting 2020: Phase II Trial for Imatinib in TB

https://clinicaltrials.gov/ct2/show/NCT03891901

NIH U.S. National Library of Medicine

ClinicalTrials.gov

Find Studies ▾ About Studies ▾ Submit Studies ▾ Resources ▾ About Site ▾

Home > Search Results > Study Record Detail

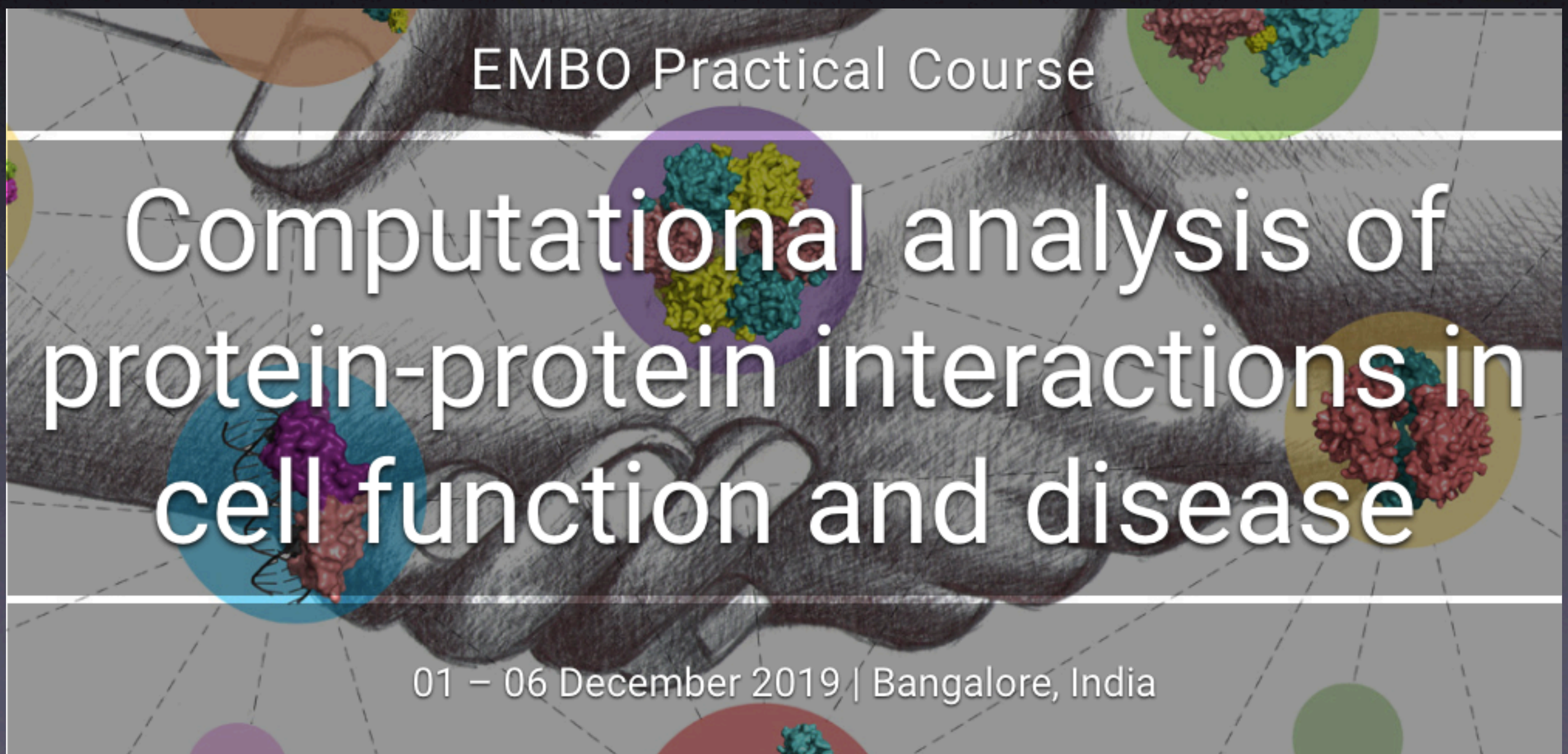
Save this study

Trial record **1 of 1** for: imatinib | Tuberculosis

Previous Study | Return to List | Next Study

A Clinical Trial of the Safety, Pharmacokinetics and Hematologic Effects of Imatinib on Myelopoiesis in Adults When Given With and Without Isoniazid and Rifabutin (IMPACT-TB)

The safety and scientific validity

ClinicalTrials.gov Identifier:
NCT03891901

Bioinformatics of PPIs has an essential role to play in current and future biomedical research