



IIB-UNSAM, Buenos Aires

21-11-2017

*Guidelines for Detecting, Validating
and Cataloguing
Short Linear Motifs*

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



Sturgeon's Law

90% of everything is crap

There are many sources of error
Here is the “matrix of blame”

Your fault	Somebody else's fault
Unintentional Error	Intentional Error

Not everything in
biological science is
what it seems to be

Mass Spec - The most sensitive tool in Biological Research

But what are you NOT getting?

- Proteins/peptides that don't fly
- Everything that you don't ask for:
 - Complex PTMs
 - Phospho- Methyl- Acetyl-Peptides
 - Unanticipated PTMs
 - e.g. Arginylation of the N-terminus
 - Anything else...

Attributing functions to genes and gene products

Neil S. Greenspan

Wolstein Research Building, Room 5130, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106-7288, USA

Trends Biochem Sci. 3-2011

What does “function” actually mean?

The units of function in regulatory proteins are the modules (domains, motifs, PTM sites) and the macromolecular complexes. The protein itself is **NOT** a unit of function !!!

REVIEW

Open Access



Experimental detection of short regulatory motifs in eukaryotic proteins: tips for good practice as well as for bad

Toby J. Gibson^{1*}, Holger Dinkel¹, Kim Van Roey^{1,2} and Francesca Diella¹

Abstract

It has become clear in outline though not yet in detail how cellular regulatory and signalling systems are constructed. The essential machines are protein complexes that effect regulatory decisions by undergoing internal changes of state. Subcomponents of these cellular complexes are assembled into molecular switches. Many of these switches employ one or more short peptide motifs as toggles that can move between one or more sites within the switch system, the simplest being on-off switches. Paradoxically, these motif modules (termed short linear motifs or SLiMs) are both hugely abundant but difficult to research. So despite the many successes in identifying short regulatory protein motifs, it is thought that only the “tip of the iceberg” has been exposed. Experimental and bioinformatic motif discovery remain challenging and error prone. The advice presented in this article is aimed at helping researchers to uncover genuine protein motifs, whilst avoiding the pitfalls that lead to reports of false discovery.

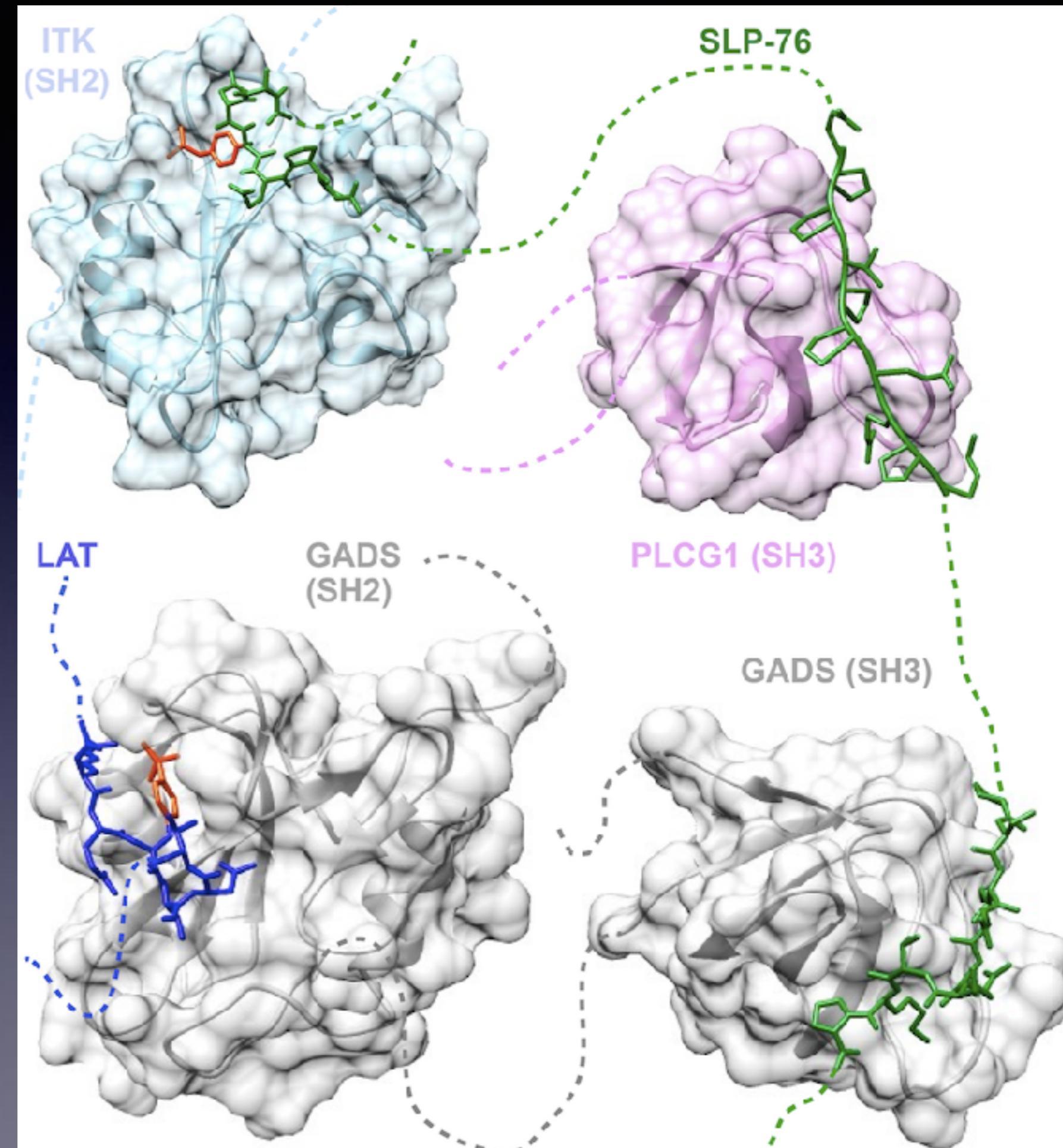
Keywords: Linear motifs, Bioinformatics, Molecular switches, Protein complexes, Cell regulation, Experimental design

The transience of transient overexpression

Toby J Gibson, Markus Seiler & Reiner A Veitia

Much of what is known about mammalian cell regulation has been achieved with the aid of transiently transfected cells. However, overexpression can violate balanced gene dosage, affecting protein folding, complex assembly and downstream regulation. To avoid these problems, genome engineering technologies now enable the generation of stable cell lines expressing modified proteins at (almost) native levels.

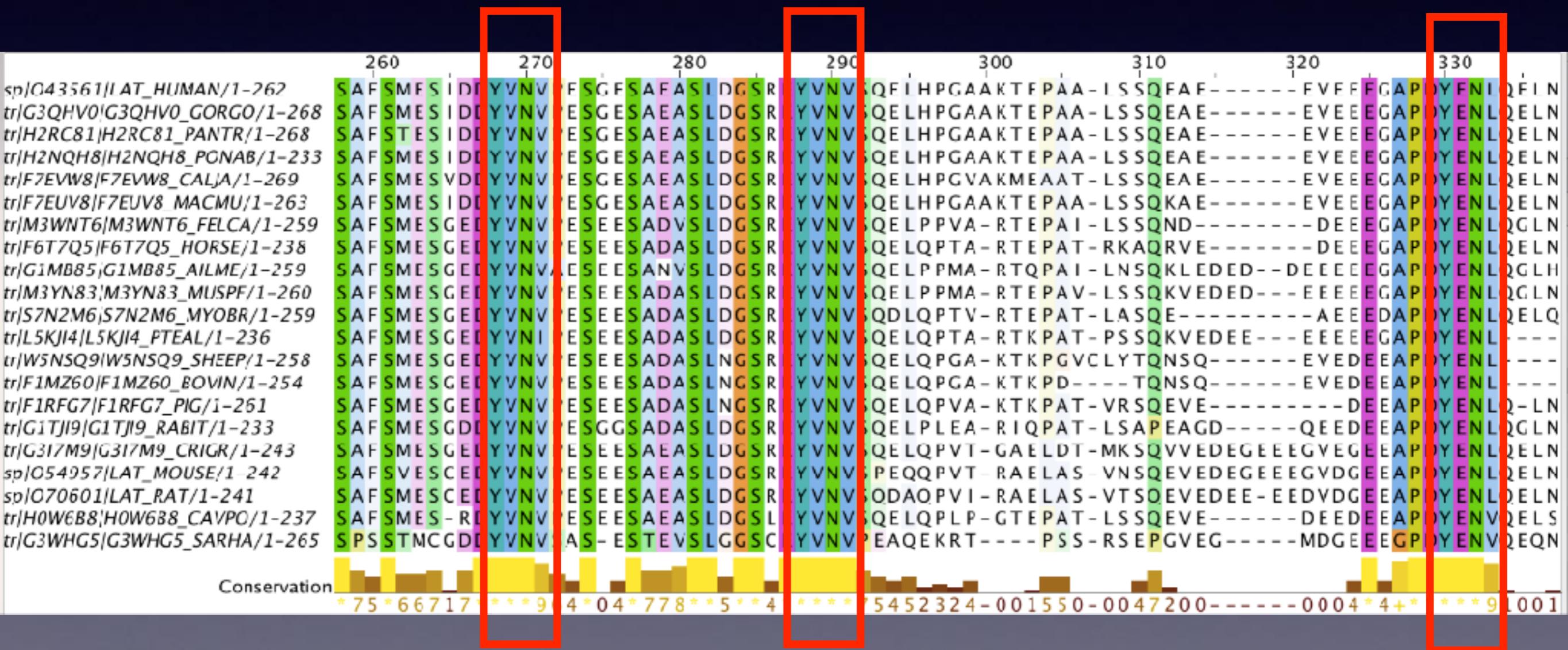
SLiMs are short peptide motifs,
usually in IDP
and always
working cooperatively



Sequence conservation is expected for SLiMs

Phospho-tyrosine motifs that bind SH2 domains are present in the LAT T-Cell activation protein

YxN and Yxx[VL] SH2 motifs are superposed



Caution!

Most sequence conservation is for folded
protein domains

So how do we know if our motif is
accessible?

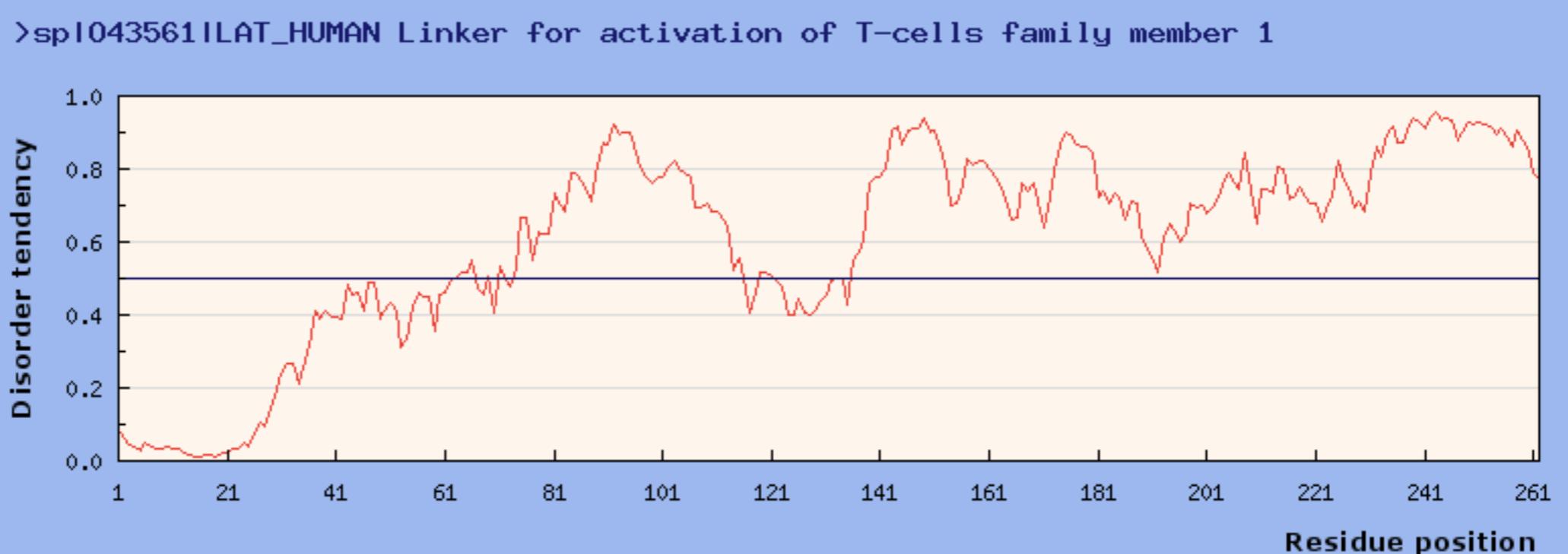
Sometimes we really don't :-(But often
there are useful clues

LAT is predicted to be intrinsically unstructured: All it's amino acids are accessible to make interactions



Prediction of Intrinsically Unstructured Proteins

- IUPred
- Theory
- How to use
- ANCHOR**
- Related links
- Downloads



The transience of transient overexpression

Toby J Gibson, Markus Seiler & Reiner A Veitia

Nature Methods (2013) NCB 10, 715

Aurora Kinase B
has a KEN box
degron recognised
by the APC/C E3
ligase.

An RxxL D-Box
motif was
incorrectly assigned
and is inaccessible.

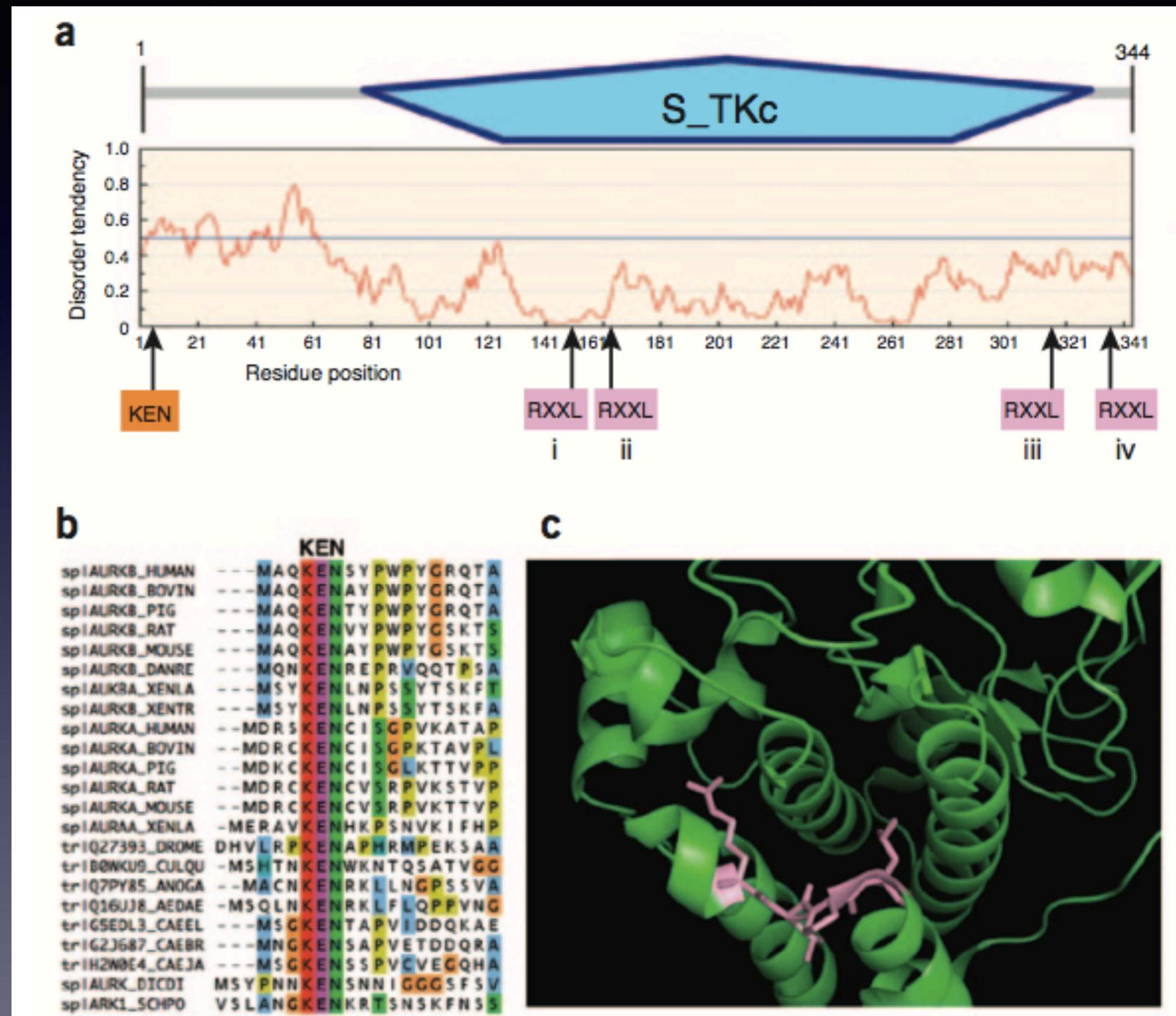
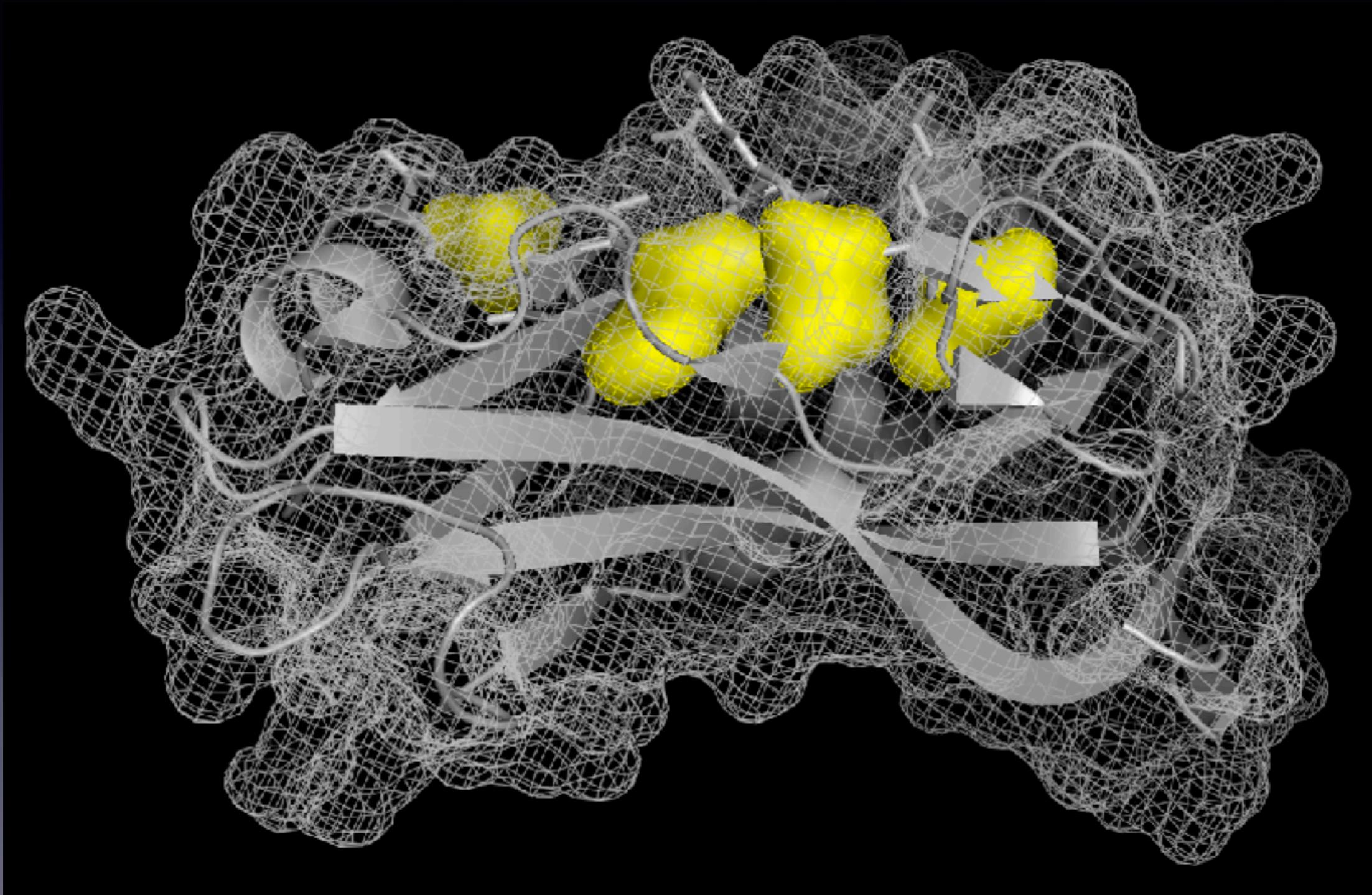


Figure 2 | Experimentally investigated candidates for the anaphase destruction motif in aurora B.

A typical false NES reported in TBX5

The Lx2,3Lx2,3LxL signature are core packing residues buried in the T-Box domain



Claimed by Kulisz and Simon (2008) MCB, 28, 1553

Rejected by Stirnimann et al. (2010 JMB 400, 71)

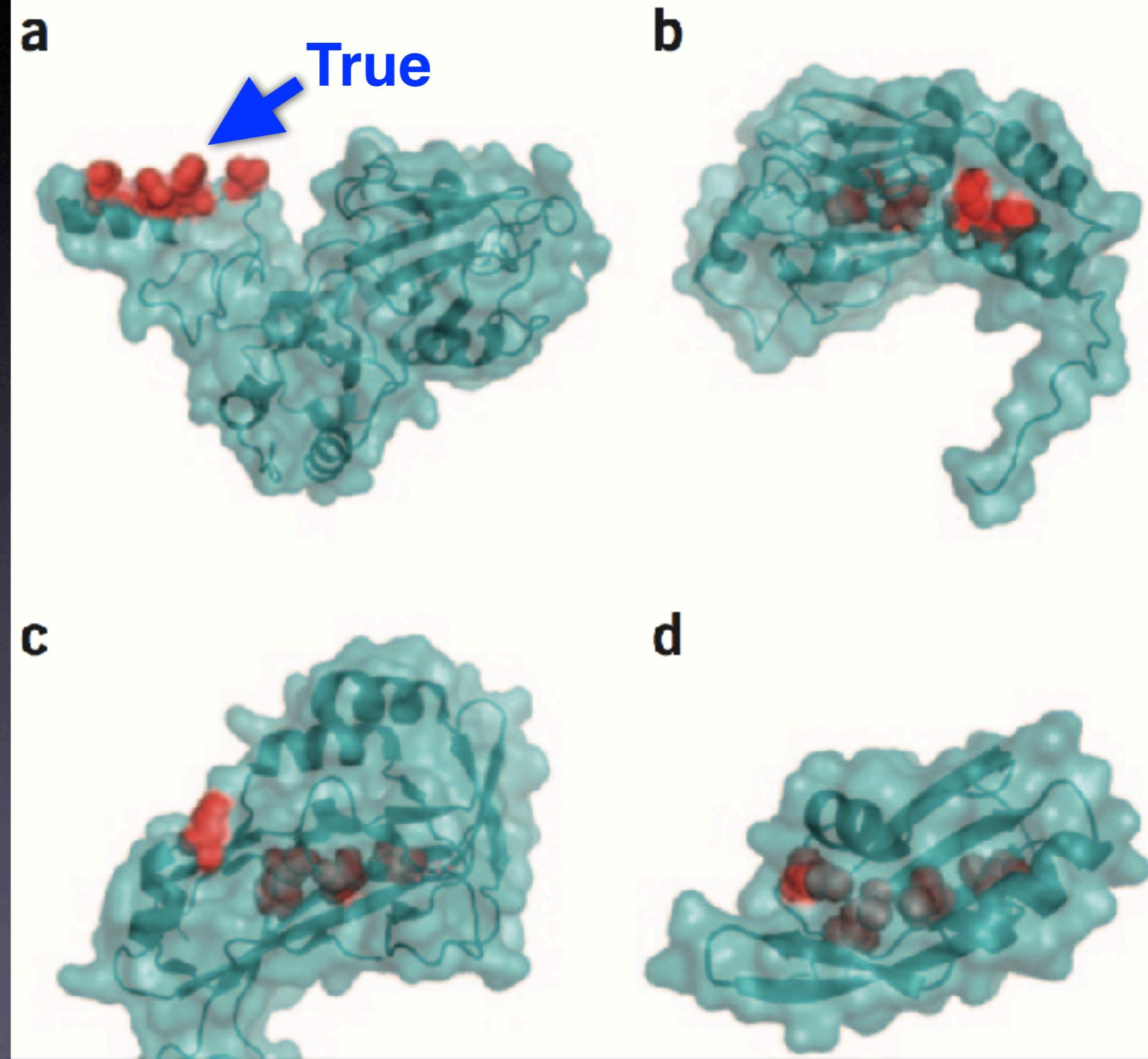
The transience of transient overexpression

Toby J Gibson, Markus Seiler & Reiner A Veitia

Nature Methods (2013) NCB 10, 715

Three examples
of buried
Nuclear Export
Sequence motifs

NESes have 4
hydrophobic
residues so the
motif matches
buried core
residues



The transience of transient overexpression

Toby J Gibson, Markus Seiler & Reiner A Veitia

Nature Methods (2013) NCB 10, 715

A table listing buried and therefore false NES motifs
We stopped counting at 10...

Table 1 | Postulated nuclear export signals (NESs) in wrong structural contexts

UniProt protein name	UniProt accession	NES amino acid positions	NES amino acid sequence ^a	Correct hydrophobic spacing? ^b	Protein domain harboring this sequence	Ref.	Protein Data Bank structure	Status
ABL1	P00520	1083–1093	LESNLRELQIC ^c	Yes	F-actin binding domain (FABD)	46	1ZZP	Rejected by Hantschel et al. ³⁰
AN32B	Q92688	109–120	LKKLECLKSLDL	Yes	Histone chaperone domain	47	2ELL	This article
CTNA2	P26232	146–155	VMRLLSHLKI ^c	Yes	Inside four-helix bundle	48	1DOW	This article
FAK1	Q00944	90–102	LQSEEVHWLHLDM	No	FERM	49	2J0K	This article
LIMK1	P53667	231–242	IRNVPLDEIDLL ^c	No	PDZ	50	2YUB	Retracted article ^d
OAZ1	P54370	114–134	RVLSIQSTLTEAKQVTWRAVVW	No	Inside β-fold	51	1Z00	This article
SMAD1	Q15797	406–414	LTKMCTIRM	Yes	MH2	52	1KHU	This article
STAT1	P42224	302–314	WDRTFSLFQQLIQC ^c	Yes	Inside four-helix bundle	53	1YVL	This article
TBX5	Q99593	152–160	LVSFQKLKL	Yes	T-box	54	2X6V	Rejected by Stirnimann et al. ³²
UPF3	P48412	88–97	LVIRLLPPNL ^c	No	RNP	55	1UW4	Rejected by Kadlec et al. ³¹

To check structural integrity, make and purify the mutant protein in an expression system

Spectroscopic methods like Circular Dichroism (CD) can quickly assess whether a protein is folded

It is important to remember that regulatory proteins are usually bound in large macromolecular complexes

Spatial Exclusivity of Mitotic Kinases

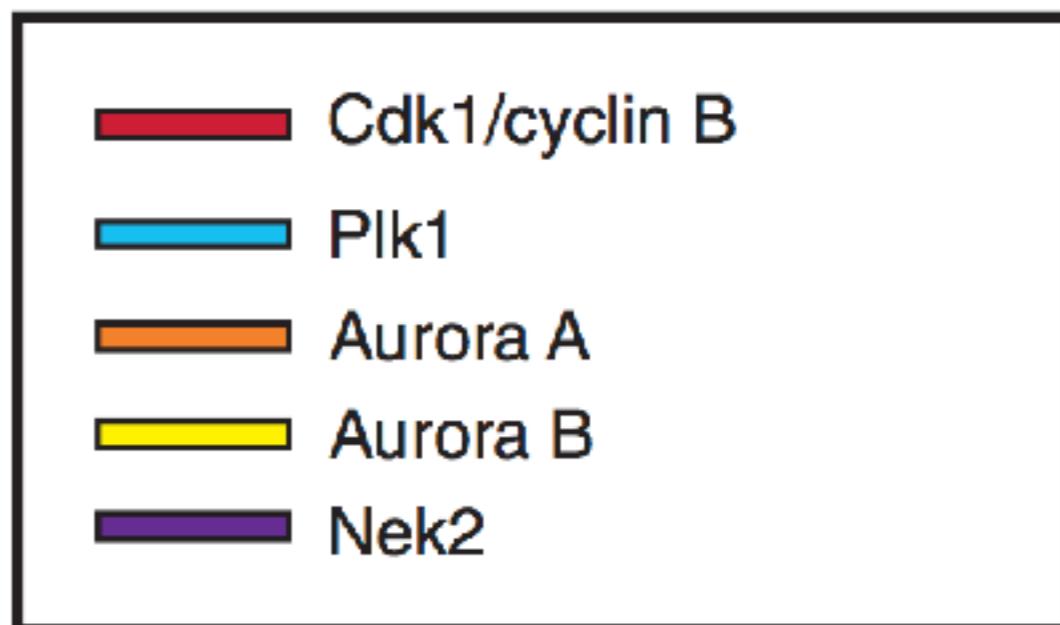
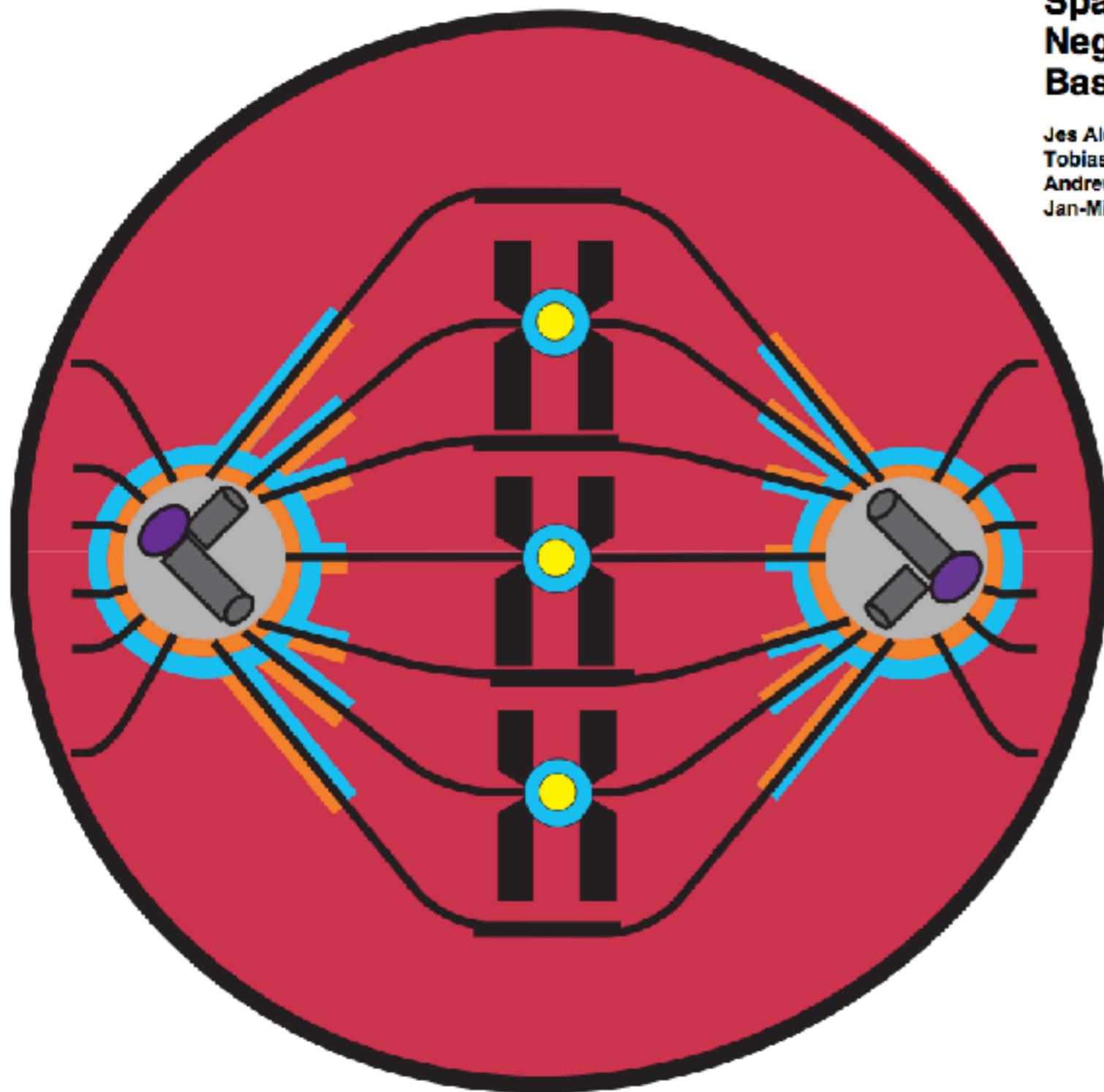
RESEARCH ARTICLE

MITOTIC KINASES

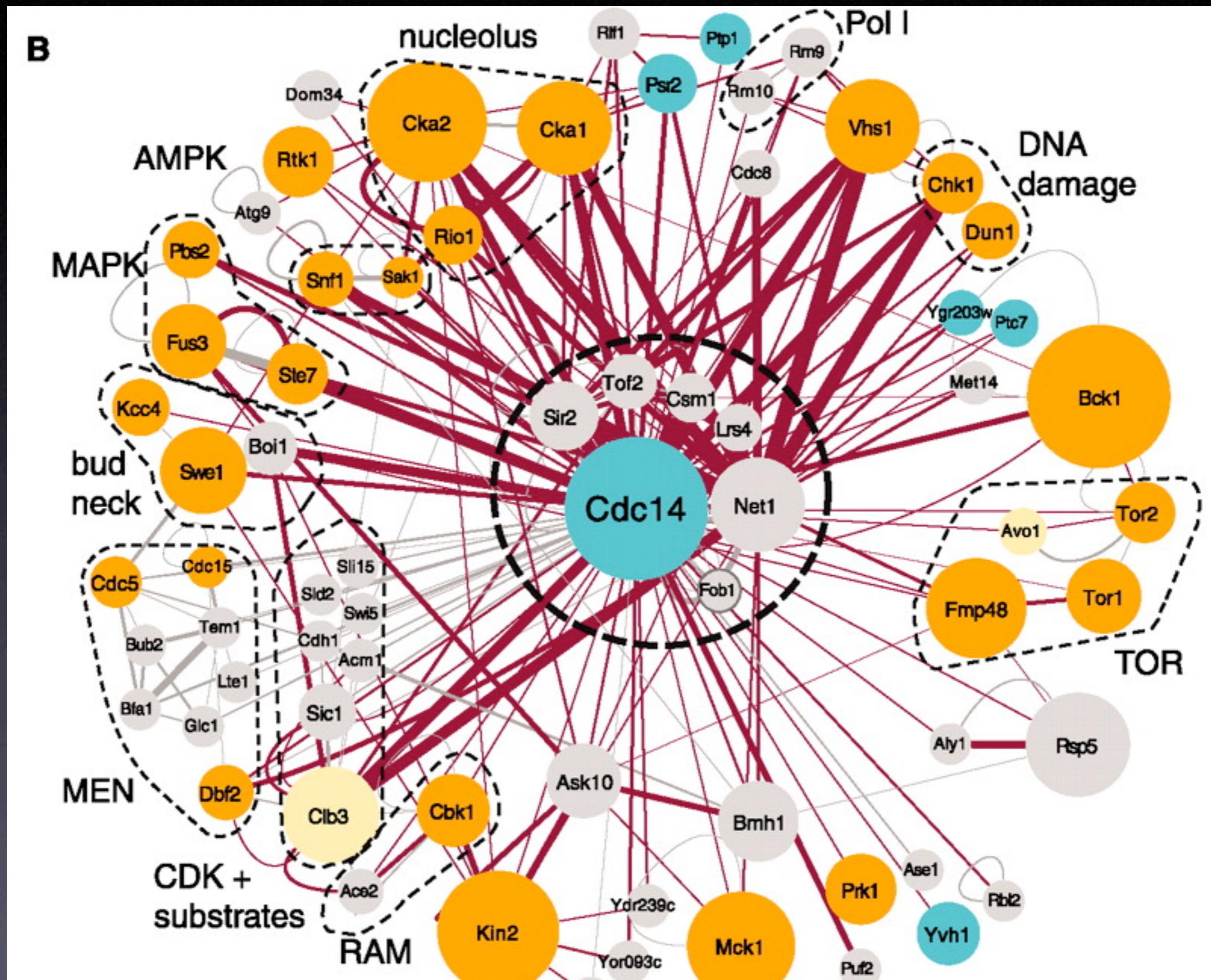
Sci. Sig., 6-2011

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‡}



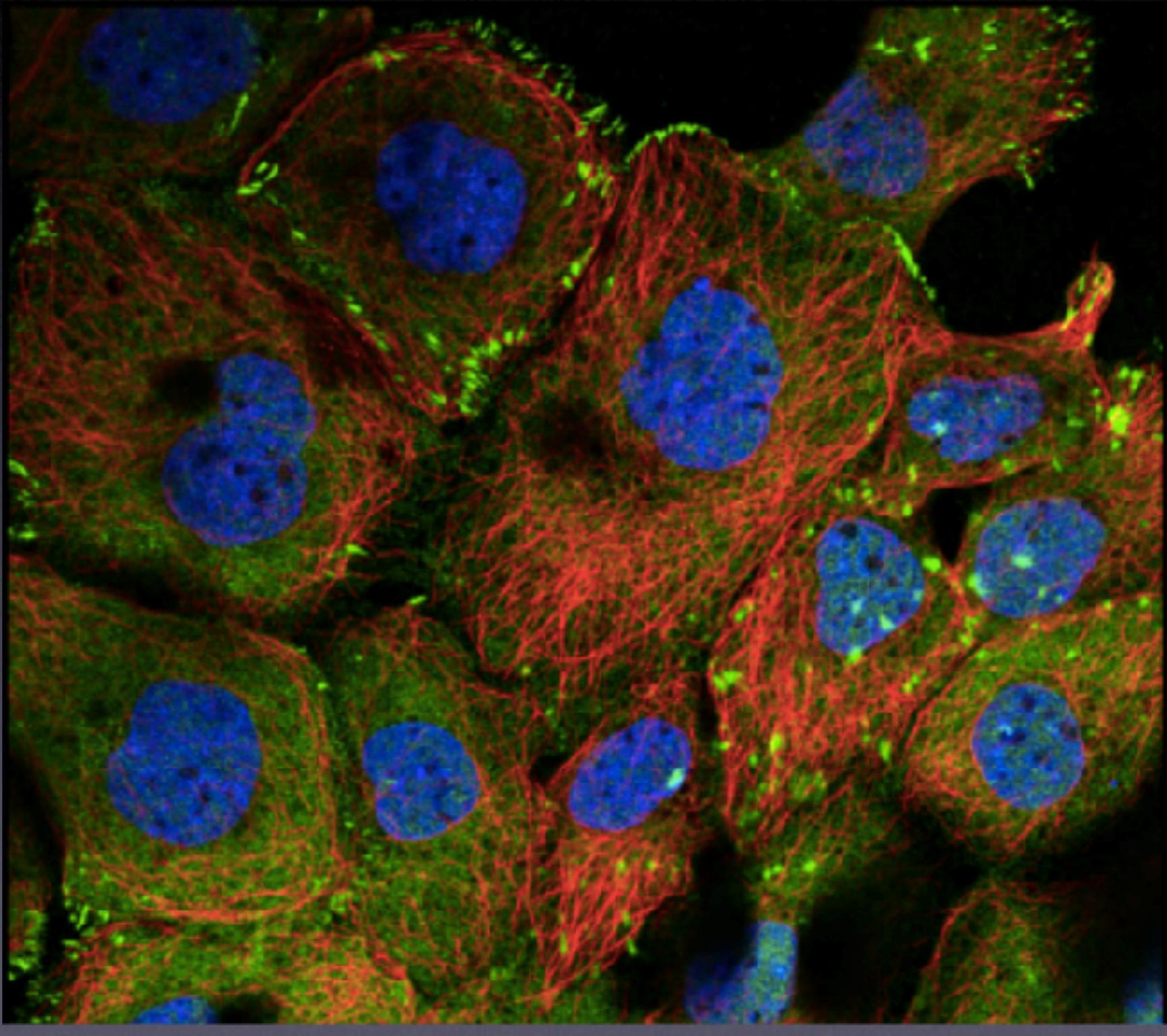
Yeast Cdc14 phosphatase interaction network



The Human Protein Atlas

can be
informative for
cell location

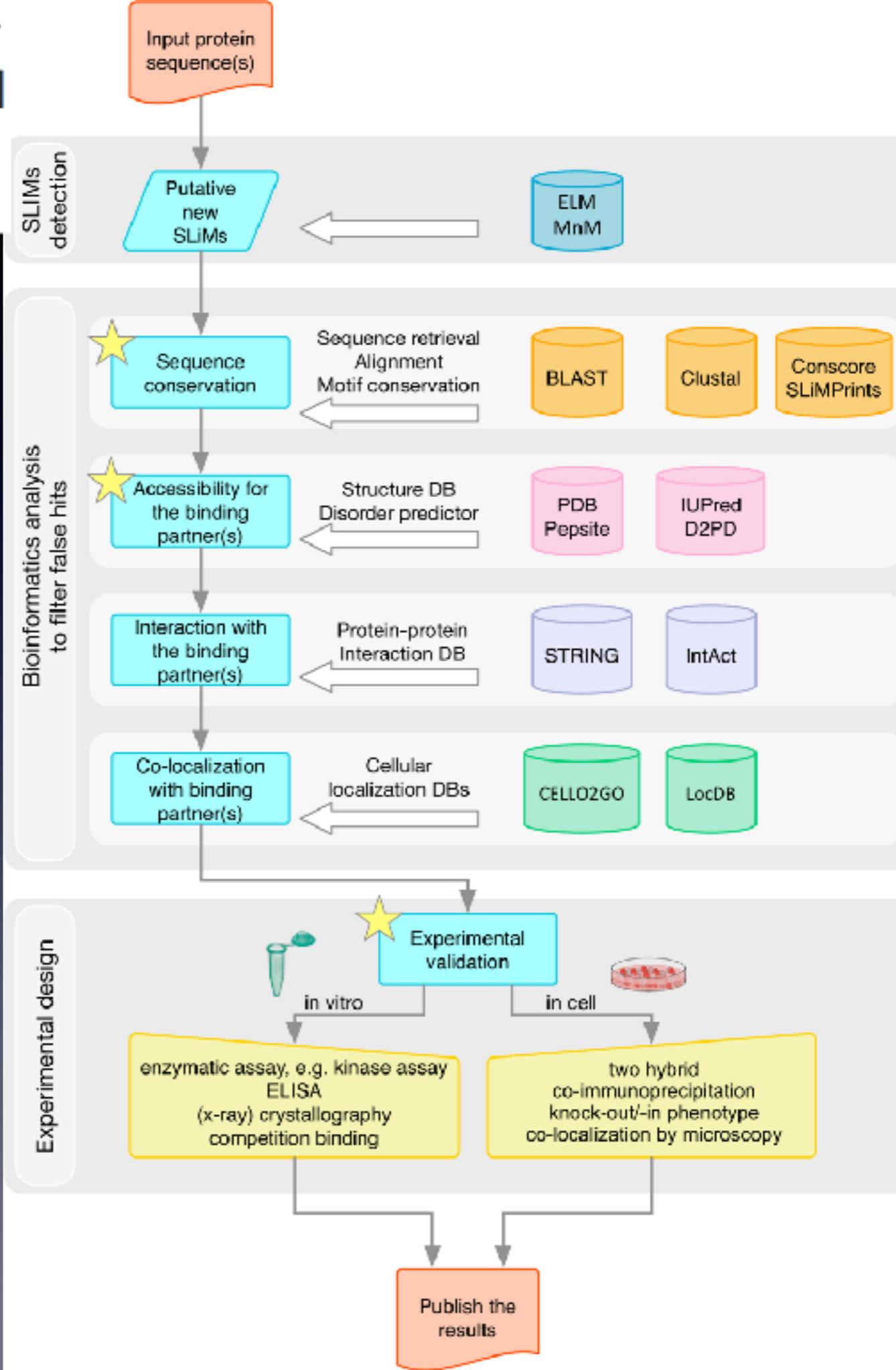
E.g Paxillin is
restricted in cell
location. It
mainly stains
Adhesion Foci



Experimental detection of short regulatory motifs in eukaryotic proteins: tips for good practice as well as for bad

Toby J. Gibson^{1*}, Holger Dinkel¹, Kim Van Roey^{1,2} and Francesca Diella¹

Guidelines: Pipeline for bioinformatics to support (or not) the motif and inform experimental design

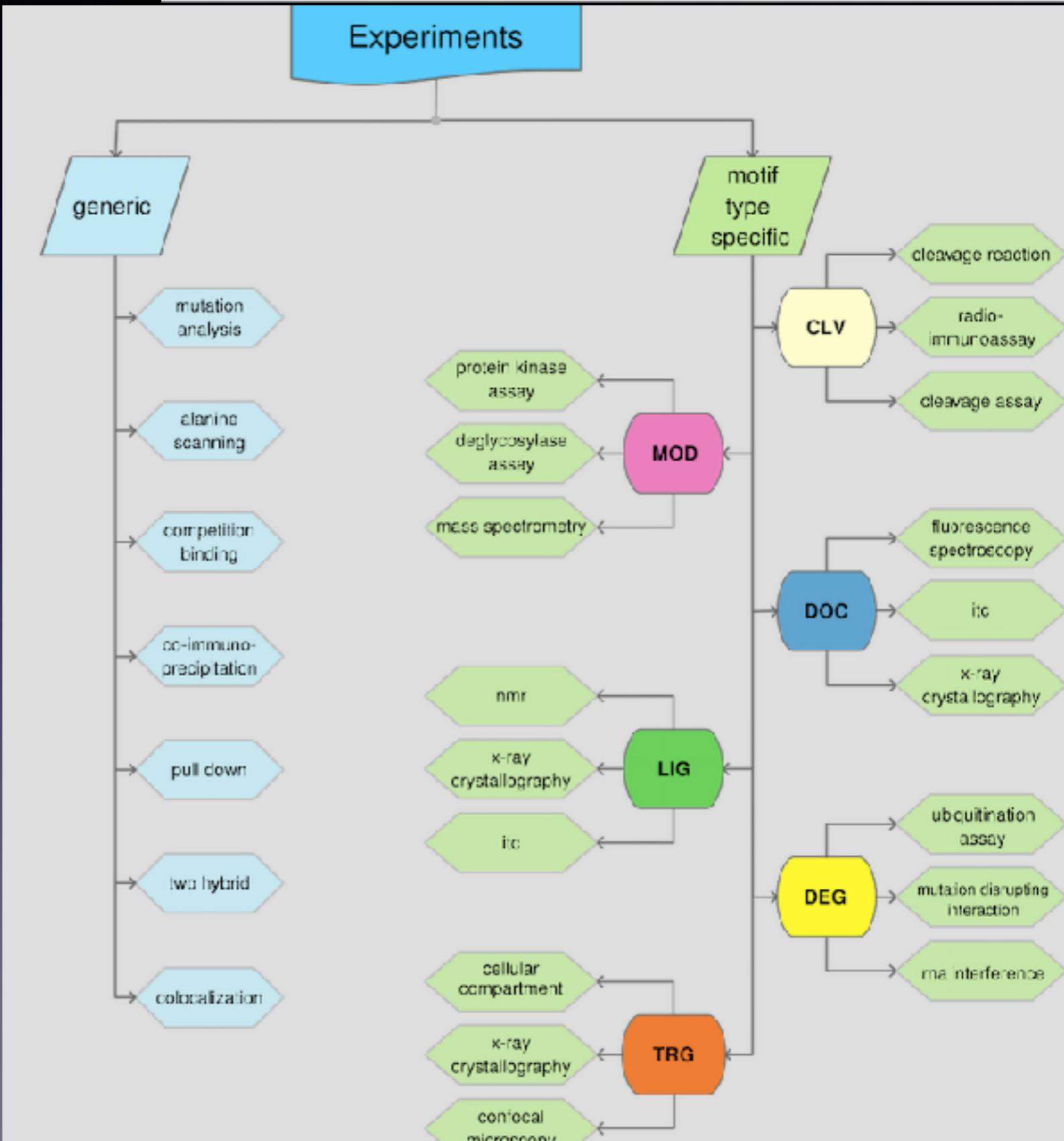


Guidelines:

Classify key experimental methods that are used in linear motif discovery
(Based on ELM Curation. More comprehensive methods table also included)

Experimental detection of short regulatory motifs in eukaryotic proteins: tips for good practice as well as for bad

Toby J. Gibson^{1*}, Holger Dinkel¹, Kim Van Roey^{1,2} and Francesca Diella¹



Experimental detection of short regulatory motifs in eukaryotic proteins: tips for good practice as well as for bad

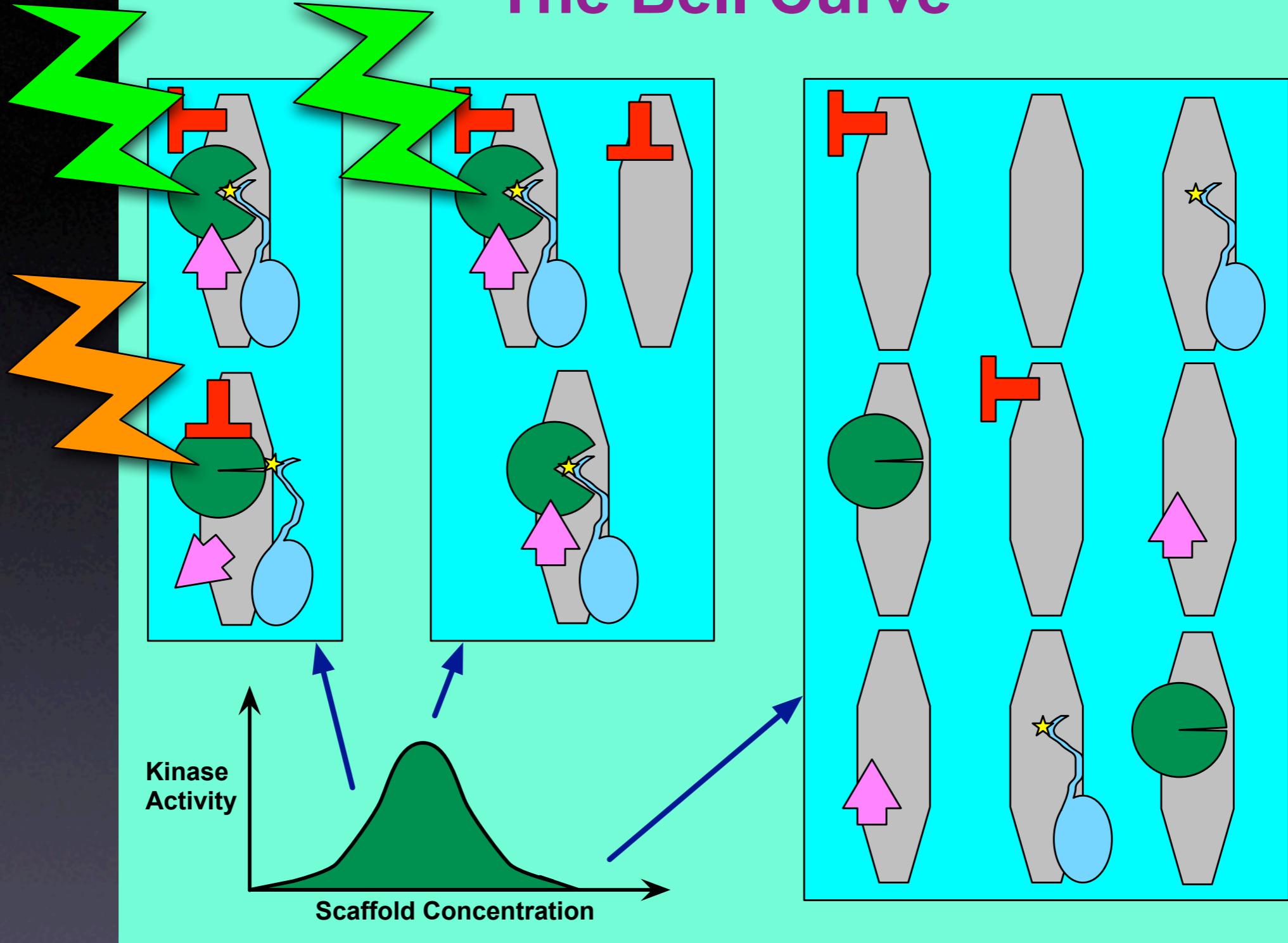
Toby J. Gibson^{1*}, Holger Dinkel¹, Kim Van Roey^{1,2} and Francesca Diella¹

How reliable is a reported motif?

Table 2: Rule of thumb quality scoring scheme.

Score	Evidence
-1	contradictory evidence
0	no evidence
1	indirect supporting evidence
2	direct supporting evidence for binding but not for in-cell function
2	Evidence in-cell that proteins associate, but direct supporting evidence for motif binding <i>in vitro</i> is lacking
3	direct supporting evidence for both binding and in-cell function

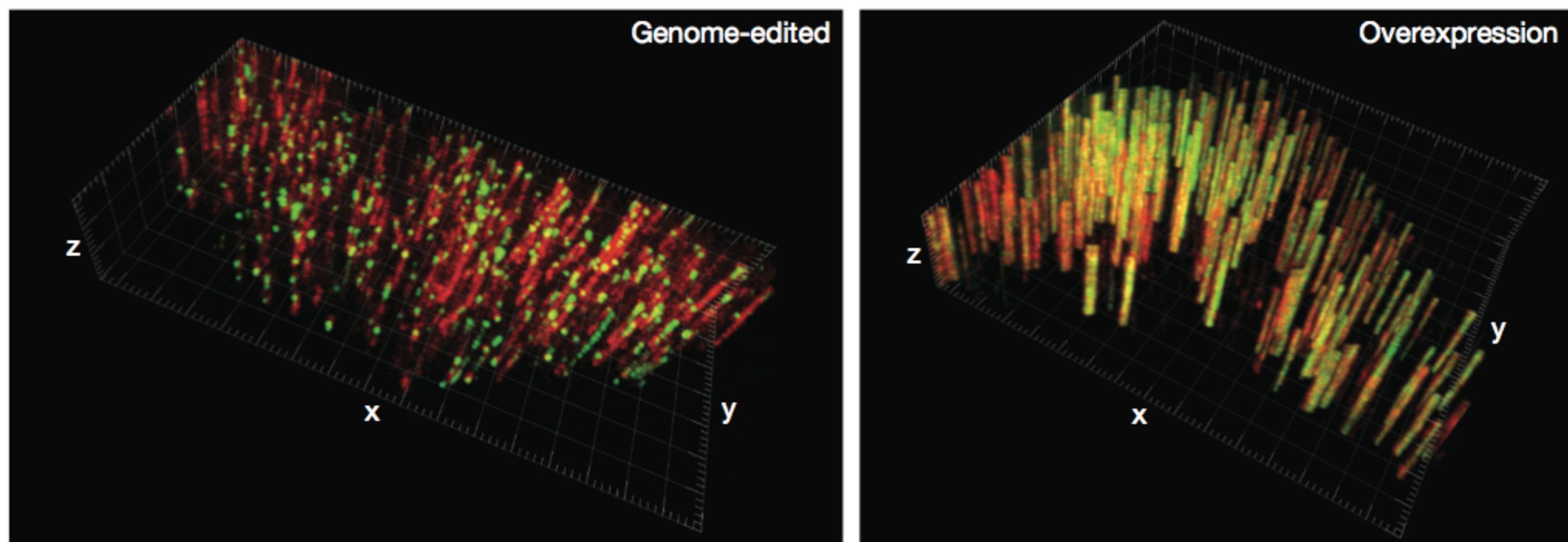
The Bell Curve



effect of KSR varied dramatically with the level of KSR protein expressed. In *Xenopus* oocytes, KSR functioned as a positive regulator of Ras signaling when expressed at low levels, whereas at high levels of expression, KSR blocked Ras-dependent signal transduction. Likewise, overexpression of *Drosophila* KSR blocked R7 photore-

Transient overexpression experiments may give misleading results

Kymographs with red *rfp-clathrin* (vesicles) and bound green *gfp-dynamin* motor proteins



DG Drubin Lab: Doyon et al. (2011) NCB 13, 331

The transience of transient overexpression

Toby J Gibson, Markus Seiler & Reiner A Veitia

Much of what is known about mammalian cell regulation has been achieved with the aid of transiently transfected cells. However, overexpression can violate balanced gene dosage, affecting protein folding, complex assembly and downstream regulation. To avoid these problems, genome engineering technologies now enable the generation of stable cell lines expressing modified proteins at (almost) native levels.

Nature Methods (2013) NCB 10, 715

Table 2. Contrasting issues with transient overexpression experiments relative to native expression

Features of Cell Regulation / Effect on Experiment	Over Expression	Native Expression
Low molecule number (e.g. <1000 per cell)	X	✓
Spatially arranged protein	X	✓
Coupled mRNA transport / Spatial translation	Overload system	✓
Mutants that are (unknowingly) unfolded	Amyloid/aggregation	?
Balanced gene dosage of regulators	X	✓
Kinases and their substrates are scaffolded	X	✓
Laser bleaching to study diffusion (or other motion) of a signalling protein	Meaningless	✓
Protein complex by Co-IP	???	✓
Proteomics	X	✓
Reproducibility	??	✓
Synchronised cell population	X	✓
Differentiate from stem cell	X	✓

Mutagenesis / Overexpression / Western Blot

are widely used to study cell regulation

However, on their own, they cannot validate a SLiM and its interaction partner:

Use transfections for discovery but not for validation

Instead of measuring concentration,
[the cell] counts molecules

Sydney Brenner, 2007

Experimental detection of short regulatory motifs in eukaryotic proteins: tips for good practice as well as for bad

Toby J. Gibson^{1*}, Holger Dinkel¹, Kim Van Roey^{1,2} and Francesca Diella¹

Guidelines:

What are the worst mistakes made by experimentalists?

Experimentalists start to go wrong when they overestimate the (normally low) likelihood that a given candidate motif might be real. A lack of understanding of protein sequence/structure relationships and of how sequence evolution and residue conservation can help assessing candidates will mean that the chance to

- Not accessible: domains, structure, IUP
- Wrong cell compartments
- Interpreting negative results as positive evidence
- Believing *in vitro* results prove a cellular function
- Believing in-cell experiments prove an interaction
 - Overinterpreting transient over-expression

The Bioinformatics tools we show in the course can help you plan your experiments.

But the results are not tablets of stone

You still have to think about your experiments critically and worry about potential artefacts in your results