



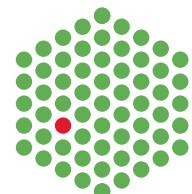
# **Short Linear Motifs in Pathogens**

**Hugo Samano-Sanchez**

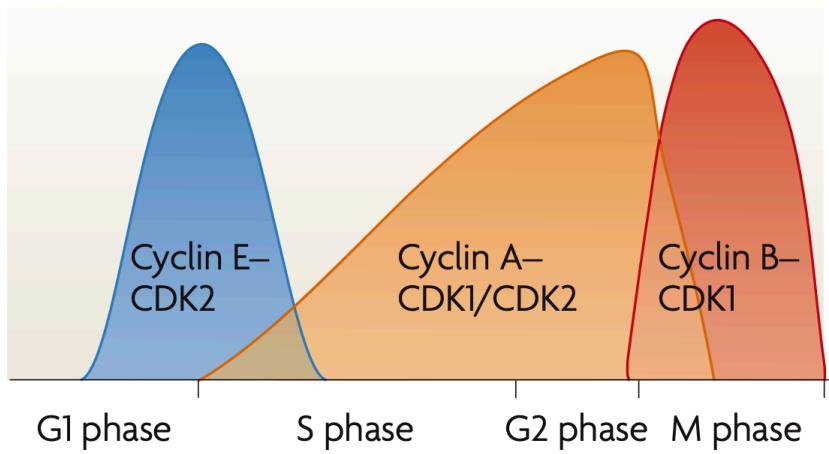
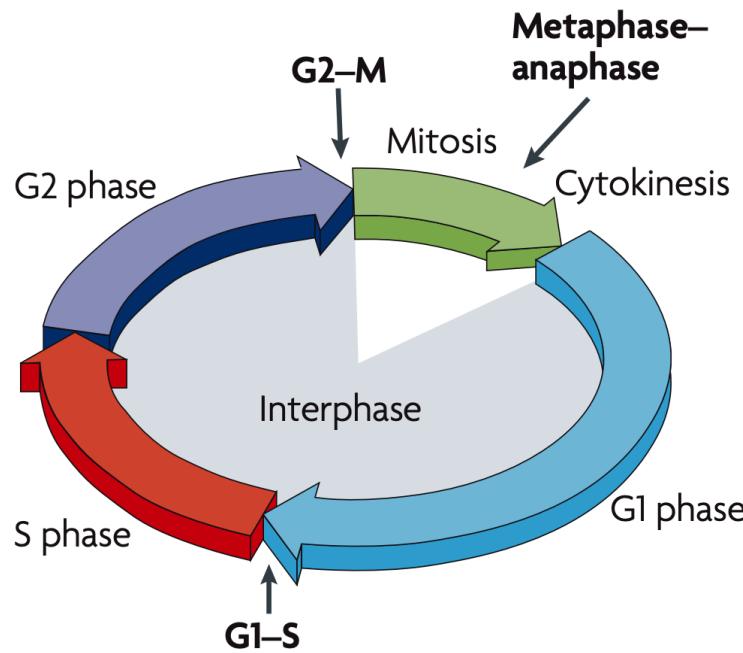
**Toby Gibson's group**

Predoc course 2019  
November 13<sup>th</sup>

**EMBL**



# Cell Cycle: A Highly Regulated Process



Hochegger, H., et al., 2008

Short Linear Motifs are involved



# Linear Motifs in Cell Cycle: <http://elm.eu.org>

The Eukaryotic Linear Motif resource for Functional Sites in Proteins

search ELM Database

ELM Home ELM Prediction ELM DB ELM Candidates ELM Information ELM downloads admin

## Welcome to the Eukaryotic Linear Motif (ELM) resource

This computational biology resource mainly focuses on annotation and detection of eukaryotic linear motifs (ELMs) by providing both a repository of annotated motif data and an exploratory tool for motif prediction. ELMs, or short linear motifs (SLIMs), are compact protein interaction sites composed of proteins ([Davey,2011](#),[Van Roey,2014](#)). They play crucial roles in intrinsically disordered regions of the proteins, which are enriched in aberrant SLIM function has been associated with several pathogens to manipulate their hosts' cellular machinery ([Davey,2011](#),[Van Roey,2014](#)).

### ELM Prediction

The ELM prediction tool scans user-submitted protein sequences for ELMs. Distinction is made between matches that correspond to putative curated in the ELM database and matches that correspond to putative short and degenerate, overprediction is likely and many putative predictive power is improved by using additional filters based on co-cellular compartment, evolutionary conservation and structural features.

#### Protein sequence

Enter Uniprot identifier or accession number: (auto-completion) e.g. EPN1\_HUMAN, P04637, TAU\_HUMAN, [RANDOM]

Or paste the sequence (Single letter code sequence only or FASTA form)

• Cell compartment (one or several): • Taxonom Type in spe

not specified extracellular nucleus cytosol peroxisome glycosome Golgi apparatus

• Motif Pro

#### ELM Prediction

The ELM prediction tool scans user-submitted protein sequences for ELMs. Distinction is made between matches that correspond to putative curated in the ELM database and matches that correspond to putative short and degenerate, overprediction is likely and many putative predictive power is improved by using additional filters based on co-cellular compartment, evolutionary conservation and structural features.

#### Protein sequence

Enter Uniprot identifier or accession number: (auto-completion) e.g. EPN1\_HUMAN, P04637, TAU\_HUMAN, [RANDOM]

Or paste the sequence (Single letter code sequence only or FASTA form)

• Cell compartment (one or several): • Taxonom Type in spe

not specified extracellular nucleus cytosol peroxisome glycosome Golgi apparatus

• Motif Pro

#### ELM DB

The Eukaryotic Linear Motif resource for Functional Sites in Proteins

search ELM Database

ELM Home ELM Prediction ELM DB ELM Candidates ELM Information ELM downloads admin

## Welcome to the Eukaryotic Linear Motif (ELM) resource

This computational biology resource mainly focuses on annotation and detection of eukaryotic linear motifs (ELMs) by providing both a repository of annotated motif data and an exploratory tool for motif prediction. ELMs, or short linear motifs (SLIMs), are compact protein interaction sites composed of proteins ([Davey,2011](#),[Van Roey,2014](#)). They play crucial roles in intrinsically disordered regions of the proteins, which are enriched in aberrant SLIM function has been associated with several pathogens to manipulate their hosts' cellular machinery ([Davey,2011](#),[Van Roey,2014](#)).

### ELM Pathways

The Eukaryotic Linear Motif resource for Functional Sites in Proteins

search ELM Database

ELM Home ELM Prediction ELM DB ELM Candidates ELM Information ELM downloads admin

## Pathways linked from ELM instances

Please select a taxon or enter any search term:

The following list contains taxons for which sequences at ELM have been mapped to pathways annotated at the "Kyoto Encyclopedia of Genes and Genomes" (KEGG) database.

- Arabidopsis thaliana (4)
- Ashbya gossypii ATCC 10895 (1)
- Bos taurus (65)
- Caenorhabditis elegans (5)
- Canis lupus familiaris (1)
- Canis lupus familiaris (3)
- Danio rerio (6)
- Drosophila melanogaster (12)
- Echinodermata (1)
- Venustus galloisi (1)
- Homo sapiens (231)
- Mus musculus (10)
- Oryzias latipes (29)
- Plasmodium falciparum 3D7 (1)
- Rattus norvegicus (139)
- Saccharomyces cerevisiae (26)
- Saccharomyces cerevisiae S288c (10)
- Schizosaccharomyces pombe (11)
- Schizosaccharomyces pombe 972h- (6)
- Schizosaccharomyces pombe (1)
- Strongylocentrotus purpuratus (1)
- Sus scrofa (30)
- Vibrio cholerae (2)
- Xenopus laevis (23)
- All (792)

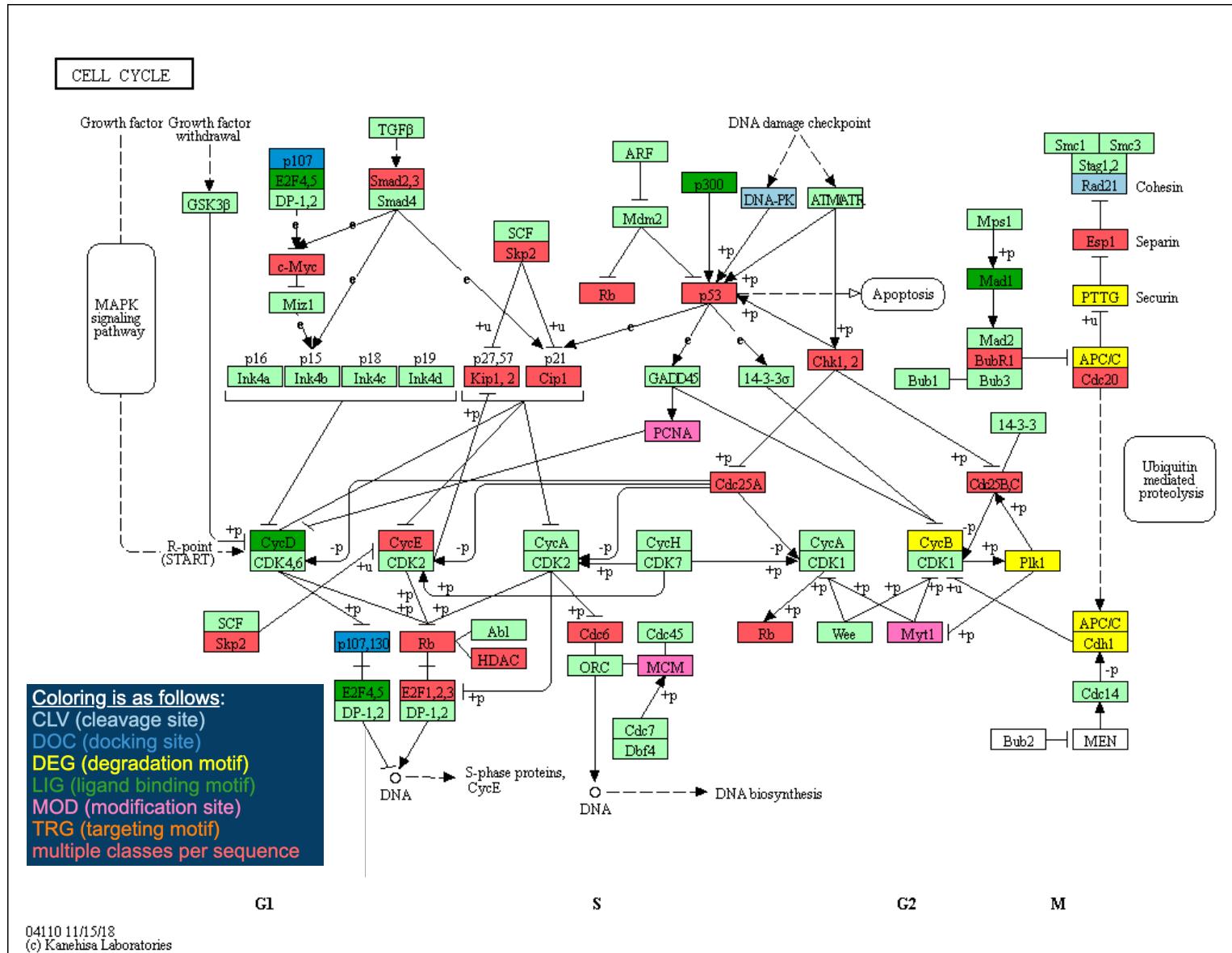
Please cite: the eukaryotic linear motif resource - 2018 update. (PMID: [2913](#))

ELM data can be downloaded & distributed for non-commercial use according to the ELM license.

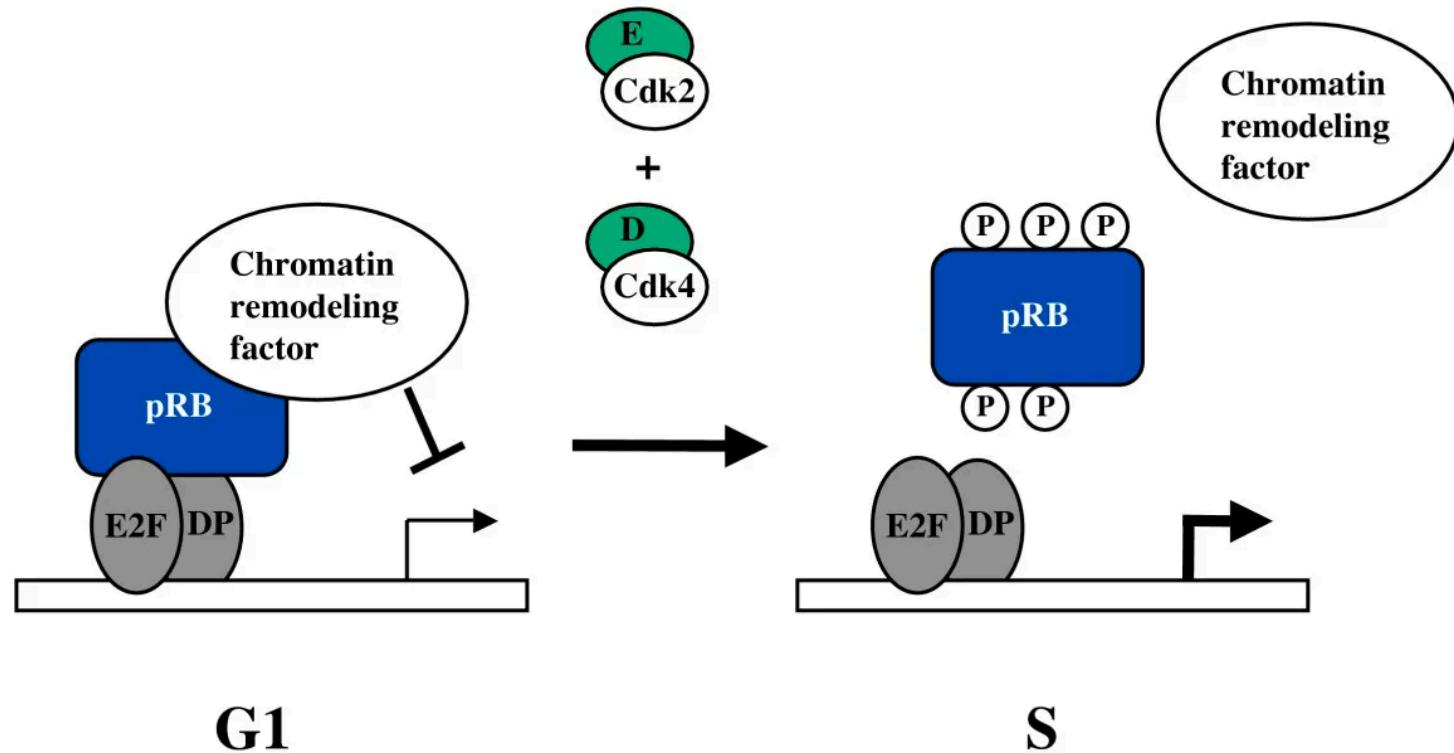
Pathway entry Pathway name # Instances # Sequences

Taxon	Pathway entry	Pathway name	# Instances	# Sequences
hsa02010	ABC transporters	5	3	
hsa0221	Acute myeloid leukemia	48	21	
hsa04520	Adherens junction	44	19	
hsa04920	Adipocytokine signaling pathway	24	13	
hsa04261	Adrenergic signaling in cardiomyocytes	19	14	
hsa05034	African trypanosomiasis	2	2	
hsa05960	Alcoholism	33	20	
hsa05330	Aldosterone regulated sodium reabsorption	14	8	
hsa05010	Allograft rejection	5	4	
hsa05010	Alzheimer's disease	25	14	
hsa05020	Amino sugar and nucleotide sugar metabolism	1	1	
hsa05148	Amiodarone metabolism	24	13	
hsa05011	Amphetamine addiction	22	10	
hsa05152	AMPK signaling pathway	33	19	
hsa05014	Amyotrophic lateral sclerosis (ALS)	27	8	
hsa04612	Antigen processing and presentation	13	10	
hsa04210	Apoptosis	44	18	
hsa05040	Arachidonate metabolism	1	1	
hsa05412	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	19	9	
hsa00953	Ascorbate and aldarate metabolism	1	1	
hsa05310	Asthma	3	2	
hsa05320	Autoimmune thyroid disease	6	5	
hsa04360	Axon guidance	31	19	
hsa05100	Bacterial invasion of epithelial cells	40	18	
hsa03017	Beta-1,4-galactosidase	31	7	
hsa03022	Basal transcription factors	1	1	
hsa03410	Base excision repair	16	9	
hsa04662	B cell receptor signaling pathway	58	28	
hsa04976	Bile secretion	7	4	
hsa05100	Biosynthesis of antibiotics	2	2	
hsa05219	Bladder cancer	59	17	
hsa04020	Calcium signaling pathway	39	22	
hsa04024	cAMP signaling pathway	63	34	
hsa04973	Carbohydrate digestion and absorption	5	2	
hsa01200	Carbon metabolism	1	1	
hsa04020	Cardiac muscle contraction	4	3	
hsa04514	Cation channel activity (C4)	17	14	
hsa04110	Cell cycle	102	37	
hsa05230	Central carbon metabolism in prokaryotes	55	15	

# Linear Motifs in Cell Cycle



# Retinoblastoma (pRb): A Cell Cycle Regulator

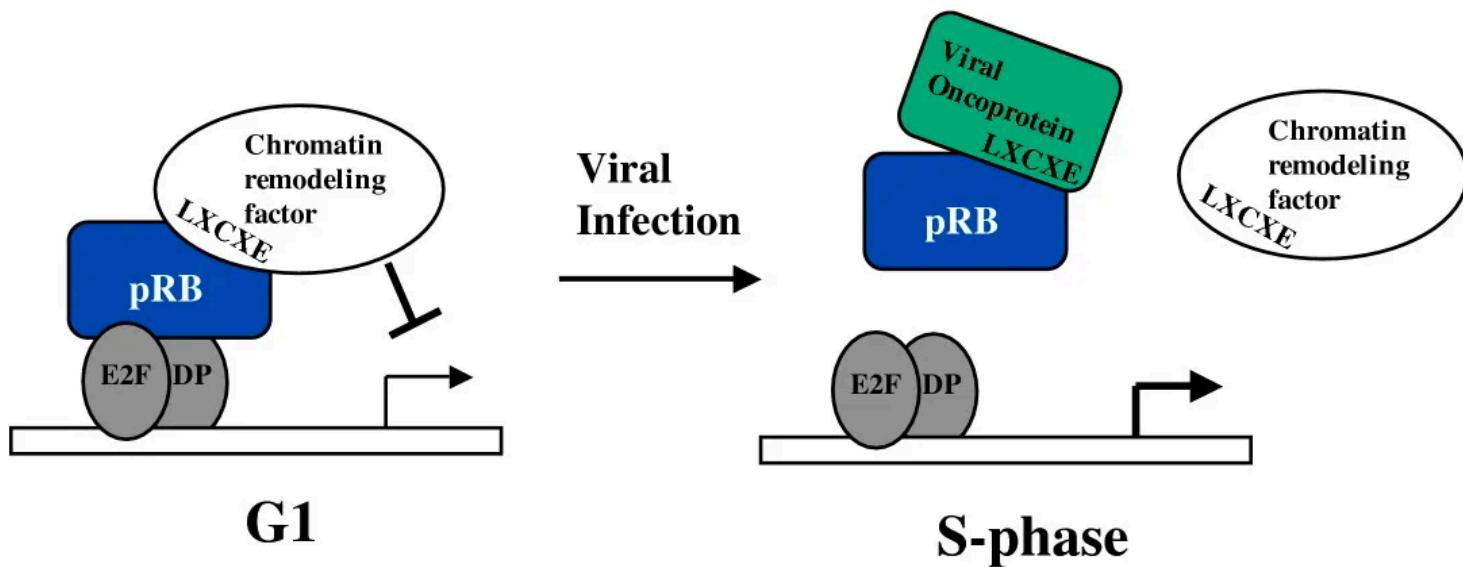


LxCxE (pRb) motif: <http://elm.eu.org>

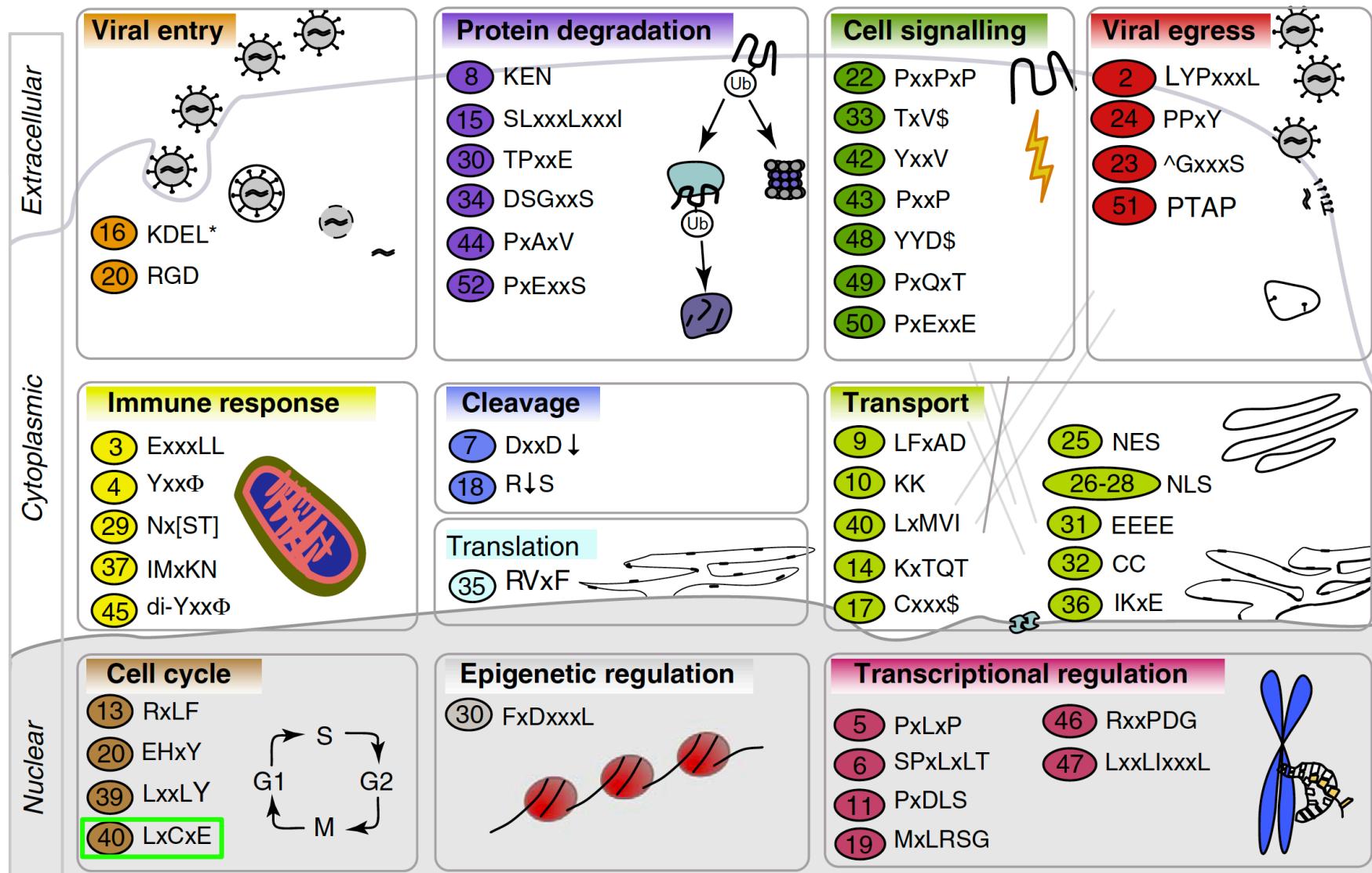
EMBL



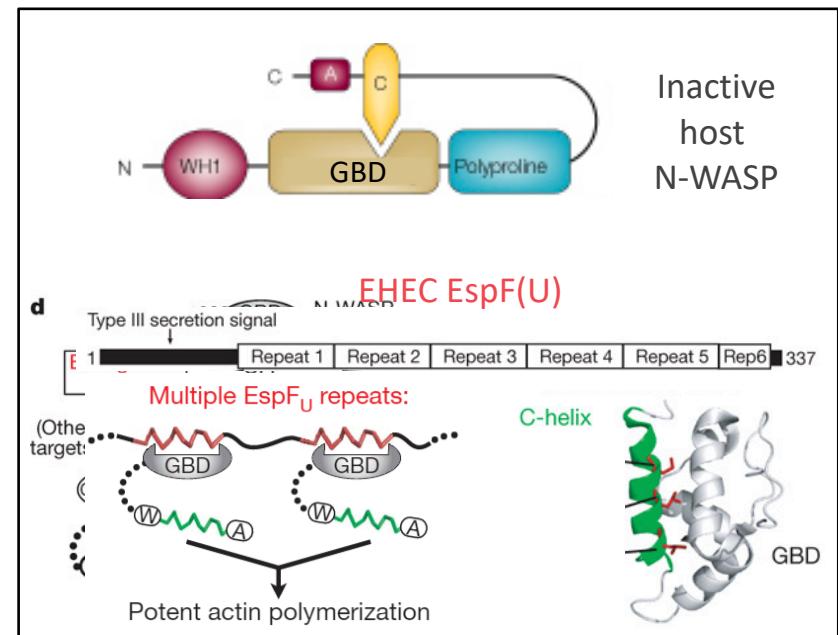
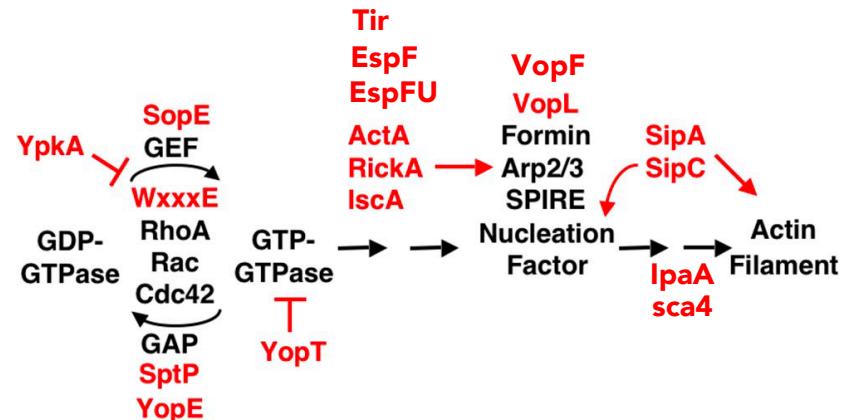
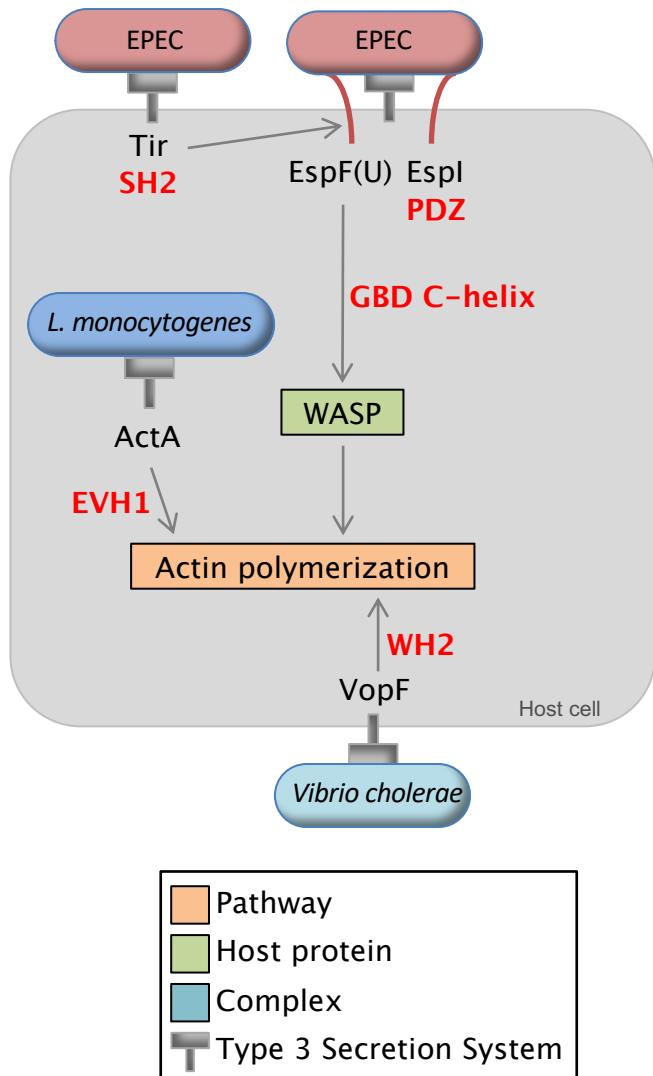
# Viral Hijack of Retinoblastoma (pRb)



# How Viruses Hijack Cell Regulation

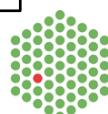


# Pathogenic Bacteria Exploit Host Actin Polymerization

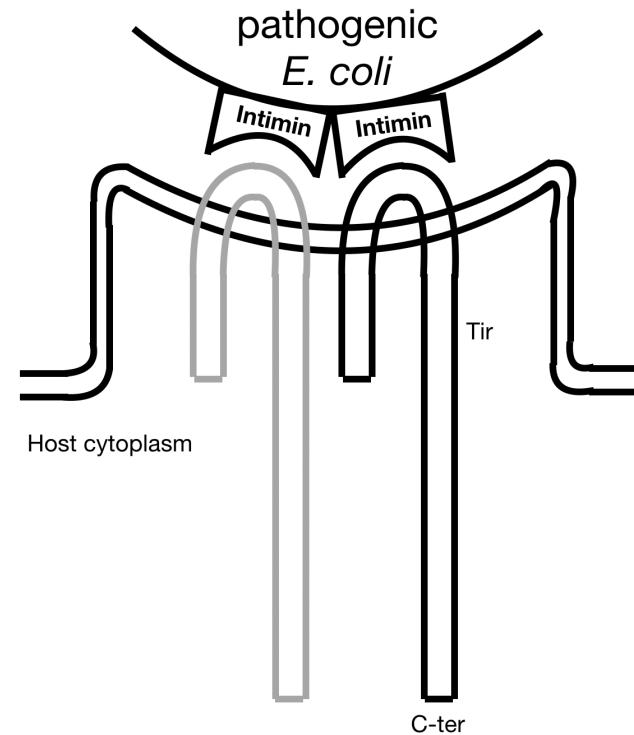
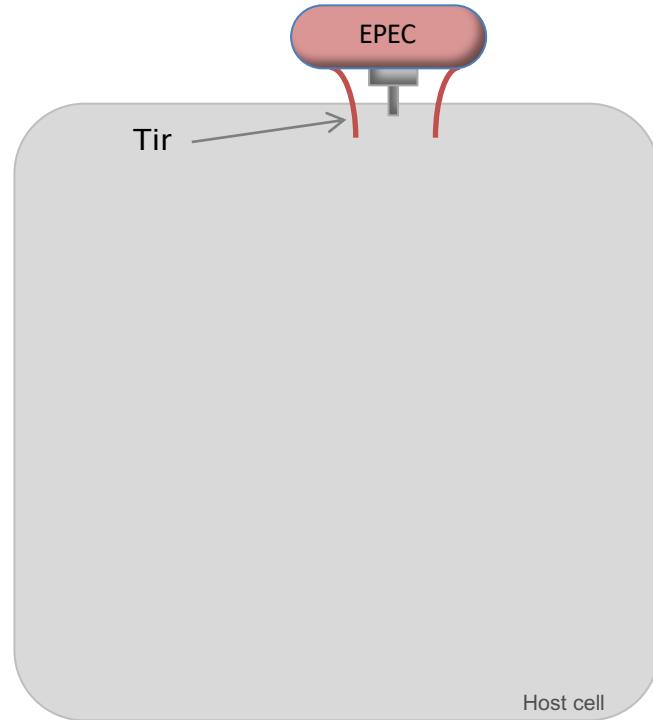


Based on Liverman, A. D., et al., 2007  
and Based on Salle N.A., et al., 2008

EMBL



# Pathogenic Bacteria Exploit Host Actin Polymerization



Tir: translocated intimin receptor

Based on Liverman, A. D., et al., 2007  
and Based on Salle N.A., et al., 2008

# Practical

---

1. Go to <http://www.uniprot.org>
2. Search for Tir protein from *Escherichia coli*
  - Use: name:tir organism: "escherichia coli"
3. Remove identical sequences
  - Use: UniRef 100%
4. Download all sequences
5. Open the sequences in JalView
6. Remove sequences too short or too long (Tir is ~540 aas)
  - Sort them by length
  - Remove sequences with < 535 aas
7. Align them (use ClustalO)

# Practical

---

## 8. Manually improve the alignment

- Remove potential low quality sequences
- Remove column gaps

## 9. Remove misaligned sequences

## 10. Sort the sequences by sequence similarity

- Use the Sort By Pairwise Identity tool
- Refine the sorting manually (Select + ↑↓ arrows)

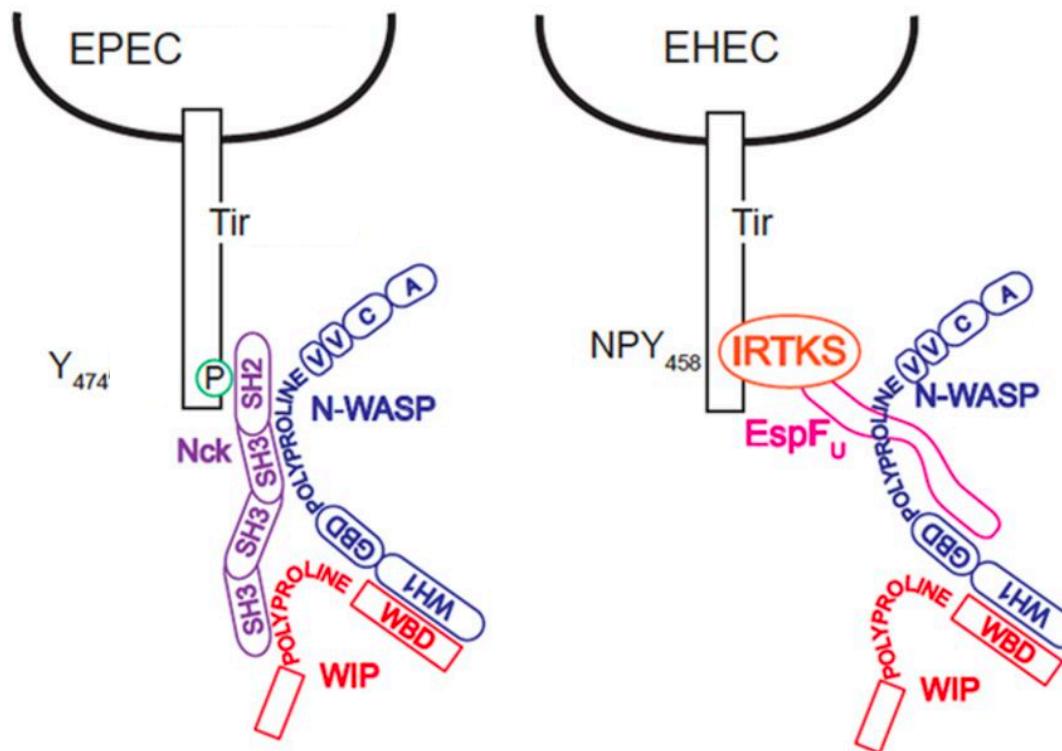
## 11. Identify conserved and partially conserved Tyrosine residues

- Ctrl+F Y

## 12. Search for SLiMs:

- [I-BAR-binding motif](#) (NPY)
- and [NCK-binding motif](#) (YX<sub>1</sub>X<sub>2</sub>X<sub>3</sub>X<sub>4</sub>)

# Pathogenic *E. coli* Strains Differentially Use Mimics of SLIMs



Modified from Garber J.J., et al., 2012

<b>EPEC strains</b>	<i>UniRef100_Q9KWH9/1-552</i>	NPYAEVGEPRNSLST - RQQEEHI - YDEV - - AADP
	<i>UniRef100_A0A3U8E574/1-552</i>	NPYAEVGEPRNSLST - RQQEEHI - YDEV - - AADP
	<i>UniRef100_A4PHQ3/1-550</i>	NPYADVGR <sub>T</sub> SR - - - QELPEEH <sub>I</sub> - YDOVAAD - EP
	<i>UniRef100_A0A070SWV7/1-552</i>	NPYADVGRASR - - - QELPEEH <sub>I</sub> - YDQVAGDLEP
	<i>UniRef100_A0A3T6XQJ5/1-552</i>	NPYAEVGEPRNSLST - RQQEEHI - YDEV - - AADP
	<i>UniRef100_A0A376PPI9/1-552</i>	NPYAEVGEPRNSLST - RQQEEHI - YDEV - - AADP
	<i>UniRef100_A0A3Y6B6T5/1-553</i>	NPYAQVGEPRNSLST - RQQEEHI - YDEV - - AADP
	<i>UniRef100_A0A3K9RFU0/1-552</i>	NPYADVKTSLHD <sub>S</sub> Q - VPTSNNSNTSVQNMGNTDSV
<b>EHEC strains</b>	<i>UniRef100_A0A3Y8BQK6/1-552</i>	NPYADVKTSLHD <sub>S</sub> Q - VPTSNNSNTSVQNMGNTDSV
	<i>UniRef100_A0A3T8SP97/1-554</i>	NPYDDVKTSLHD <sub>S</sub> Q - VPTSNNSNTSVQNMGNTDSV
	<i>UniRef100_A0A3J1EUH7/1-555</i>	NPYADVKTSLHD <sub>S</sub> Q - VPTSNNSNTSVQNMGNTDSV
	<i>UniRef100_B5ARQ7/1-556</i>	NPYADVKTSLHD <sub>S</sub> Q - VPTSNNSNTSVQNMGNTDSV
	<i>UniRef100_A0A3T7VN39/1-558</i>	NPYADVKTSLHD <sub>S</sub> Q - VPTSNNSNTSVQNMGNTDSV
	<i>UniRef100_A0A0K4R2K2/1-558</i>	NPYADVKTSLHD <sub>S</sub> Q - VPTSNNSNTSVQNMGNTDSV
	<i>UniRef100_O85506/1-558</i>	NPYADVKTSLHD <sub>S</sub> Q - VPTSNNSNTSVQNMGNTDSV



# Questions?

---

EMBL

