

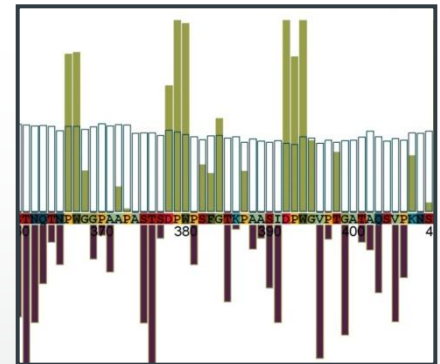
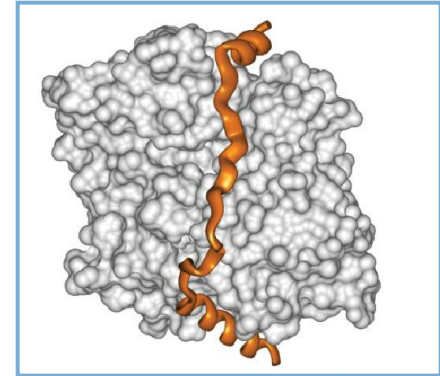
SLiMSearch:

a webserver for finding novel occurrences
of short linear motifs in proteins,
incorporating sequence context.

Dr Richard J Edwards
Bioinformatics & Molecular evolution

Talk Outline

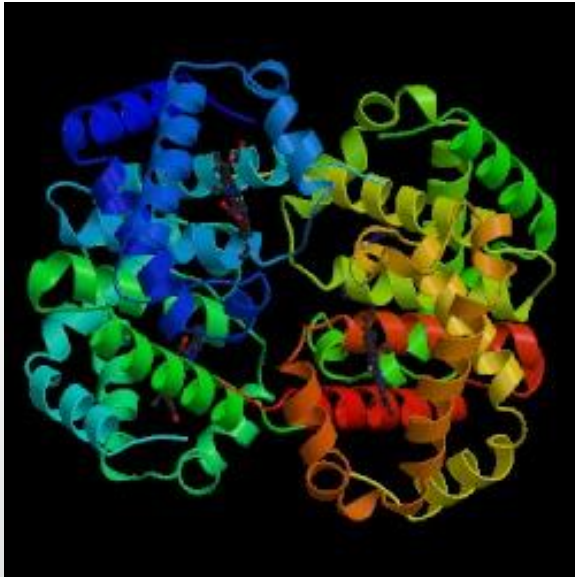
- Short Linear Motifs (SLiMs)
- The importance of (biological) sequence context
 - Relative Local Conservation
- SLiMSearch webservice
 - Application to Homeodomain ligand motif
- Future directions



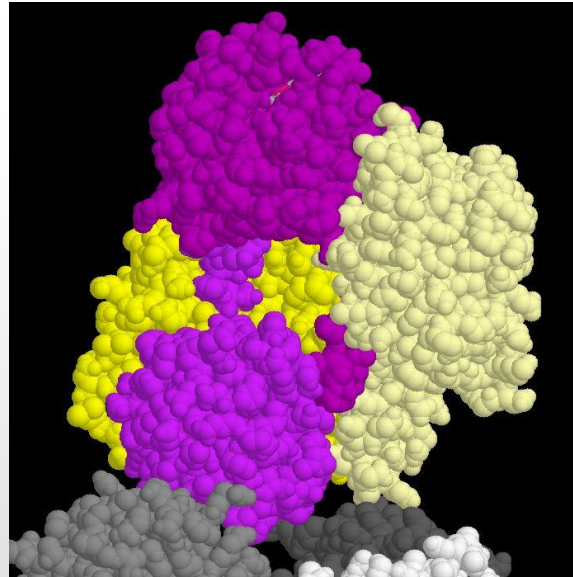
Protein-Protein Interactions

- Protein-Protein Interactions (PPI) control and regulate all fundamental cellular processes

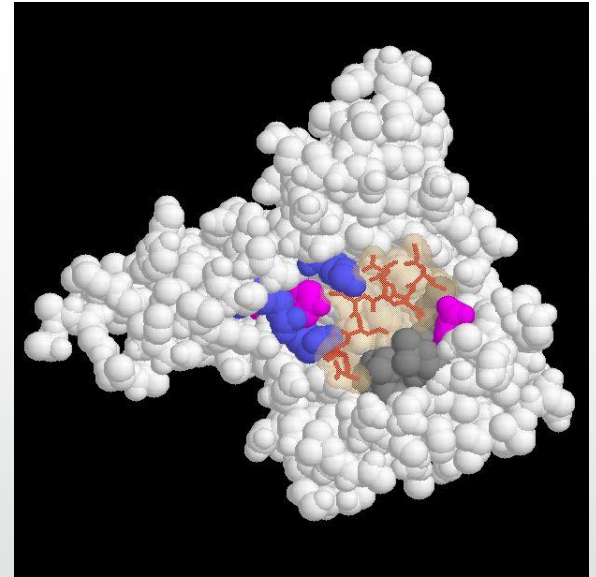
Stable complexes



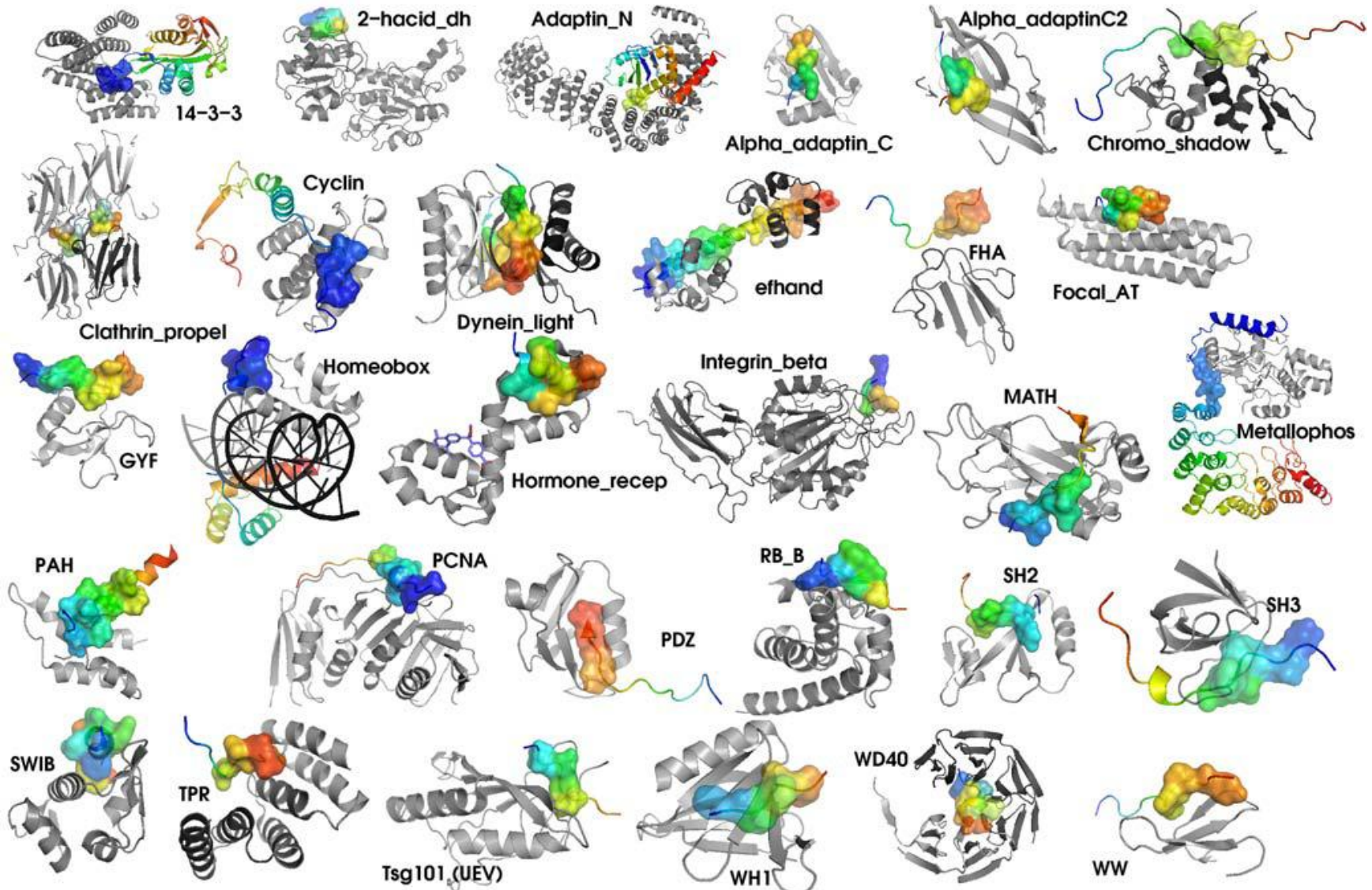
Domain-domain interactions



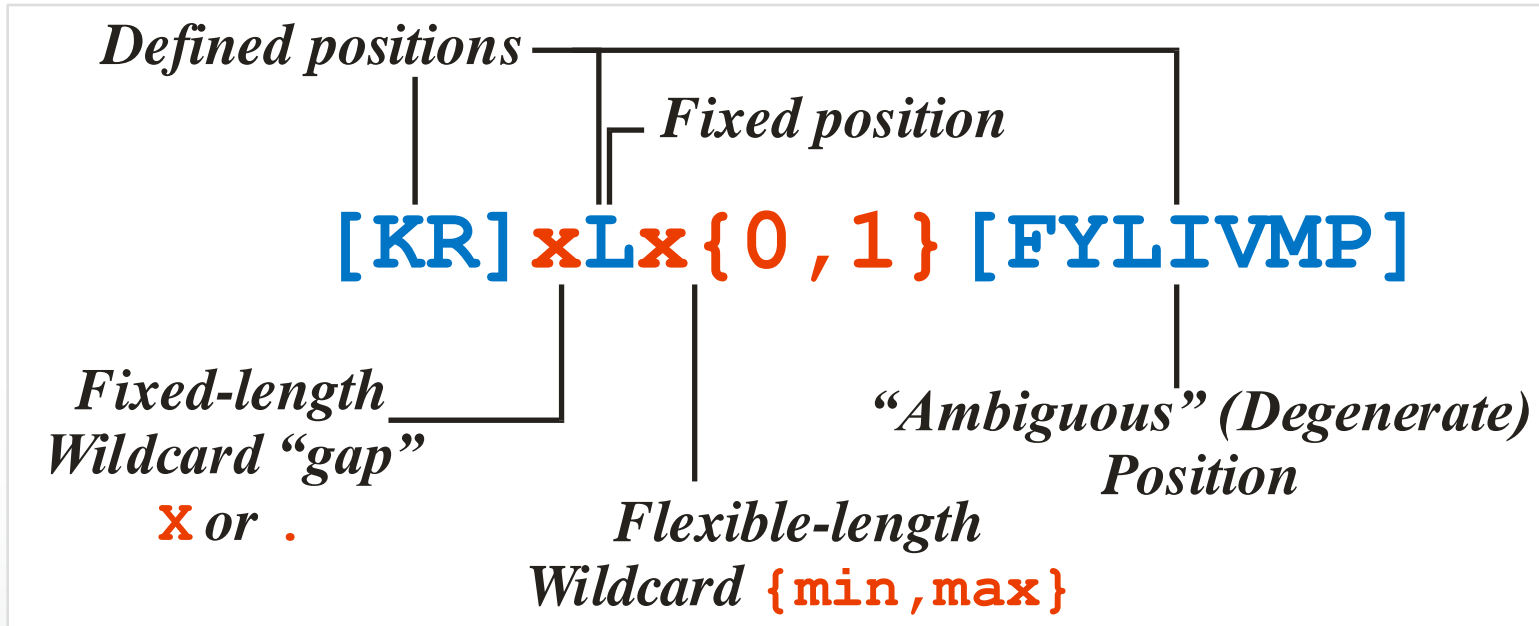
Domain-motif interactions



Short Linear Motifs (SLiMs)



Short Linear Motifs (SLiMs)



- *Short* – generally <10aa long few defined residues.
- *Linear* – adjacent amino acids in sequence.
- *Motifs* – defined sequence patterns necessary for function
 - recur in relevant proteins, allowing identification.

ELM (www.elm.eu.org)
100+ experimentally
validated eukaryotic
SLiMs

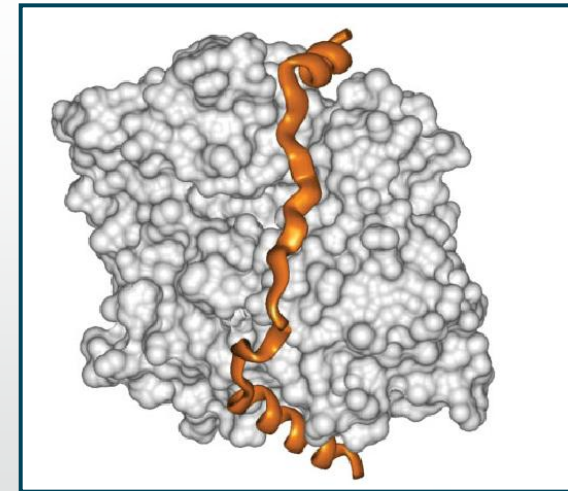
Short Linear Motifs (SLiMs)

- Regular expression searches easy
but...
 - High likelihood of matches by chance
 - Significance?
 - Observed *vs* Expected
 - **Biological** significance?
 - Context
- } **Statistical Sig. ≠ Biological Sig.**
- *Short* – generally <10aa long few defined residues.
 - *Linear* – adjacent amino acids in sequence.
 - *Motifs* – defined sequence patterns necessary for function
 - recur in relevant proteins, allowing identification.

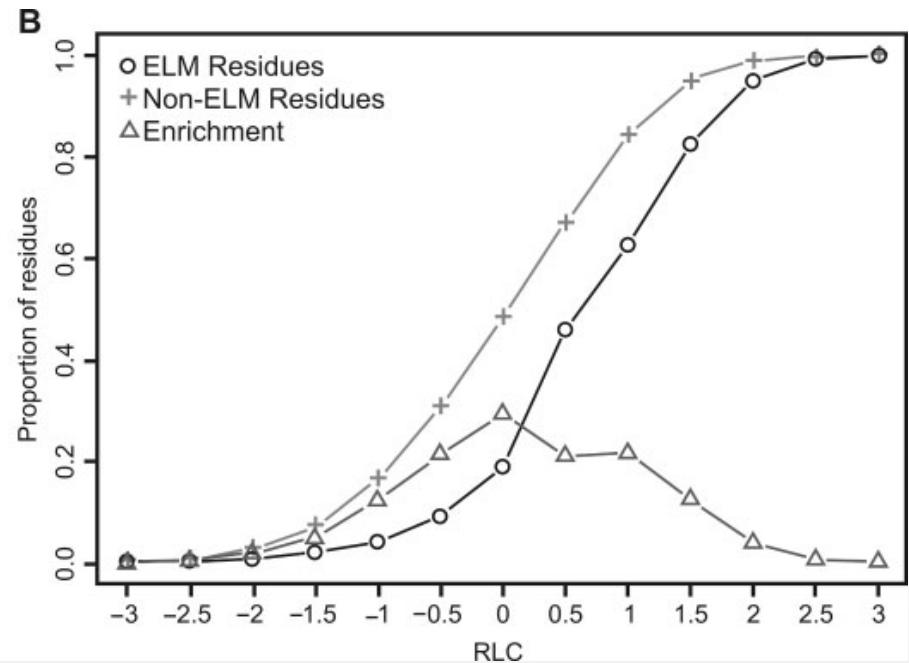
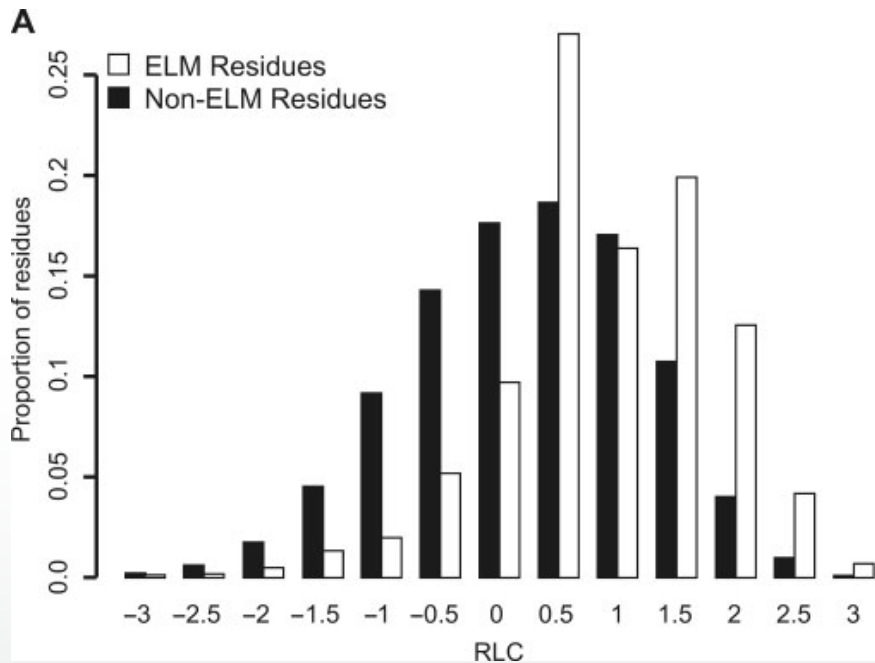
Relative Local Conservation (RLC)

- Use evolutionary conservation analysis to identify important residues
- Problem:
 - SLiMs tend to be in disordered (not globular) regions
 - Disordered regions tend to be less conserved
- Solution:
 - Normalise conservation to local environment
 - Compare within (dis)order state only

$$\text{RLC}_i = \frac{C_i - b_i}{\sigma_i} \quad b_i = \frac{1}{2N+1} \sum_{j=i-N}^{i+N} C_j$$

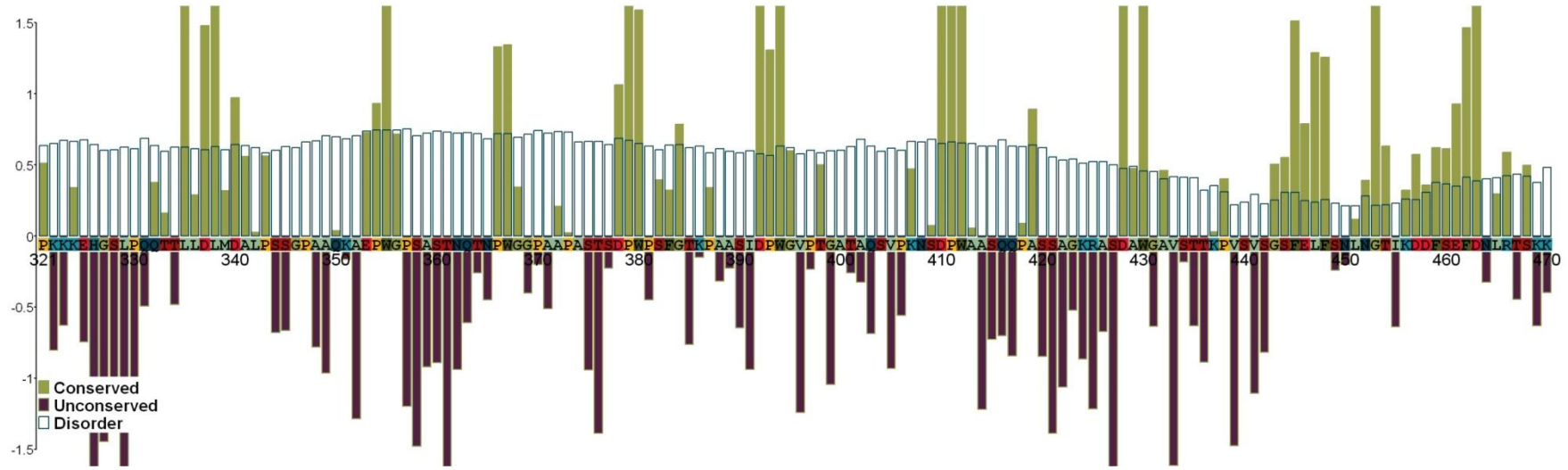
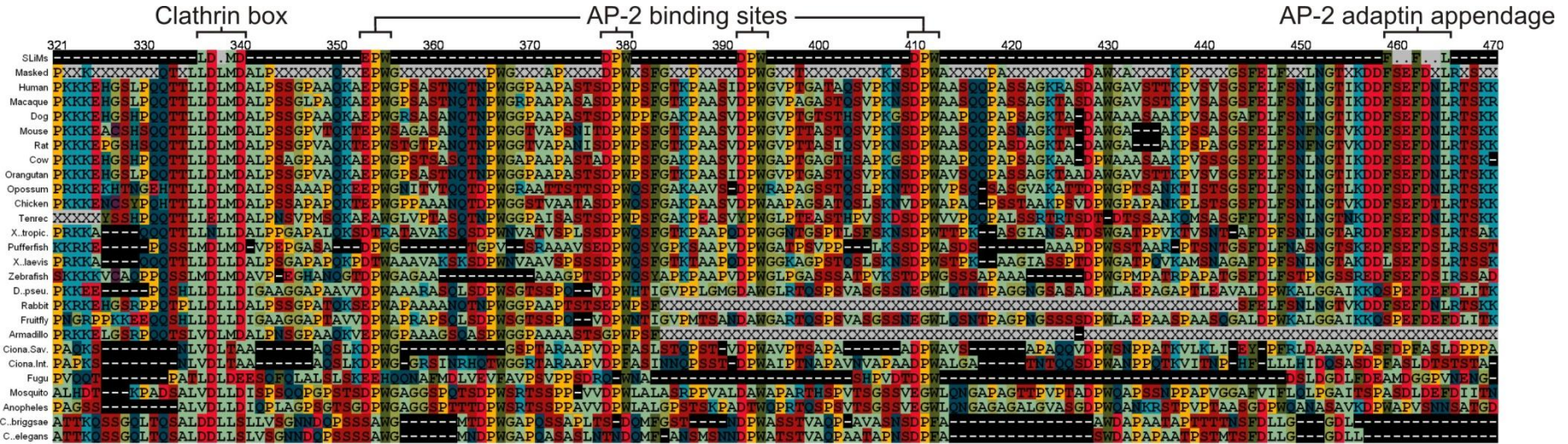


Relative Local Conservation (RLC)



□ ELM = Eukaryotic Linear Motif database

RLC: Epsin 2



Short Linear Motifs (SLiMs)

- Regular expression searches easy
but...
- High likelihood of matches by chance
 - Significance?
 - Observed *vs* Expected
 - **Biological** significance?
 - Context
- Combine patterns of:
 - Primary sequence (motif)
 - Protein structural (disorder)
 - Evolution (conservation)



SLiMSearch

SLiMSearch

Step 1: Choose a search database

Load sequence from UniProt

(Load example|Clear example)

Choose from a predefined database.

Human

- All proteins [?]
- Cytoplasmic proteins only [?]
- Nuclear proteins only [?]
- Membrane proteins only [?]

or

Step 2: Choose a motifs and options

Enter motifs

K.~~TQT~~

Options

Inclusively mask [?]

Cytoplasmic Tails [?]

Search only [?]

Disorder [?]

Relatively conserved [?]

Remove [?]

Domain [?]

Analyses are limited to 100 proteins.
For larger custom datasets, download
SLiMSearch [here](#).

submit

SLiMSearch

Results

Motif Hits

Switch table view (Motifs|Summary)
Remove motifs with IUP less than 0.3|0.5|reset
Click on headers to sort

Motif Statistics

Click to switch motif. Viewing none

Pattern	N_Occ	Seq	Desc	Pos	Len	RLC [†]	IUP	Pattern	Match
[KR].TQT	7	DC111	Cytoplasmic dynein 1 intermediate chain 1	151	628	2.03 view	0.563	[KR].TQT	KETQT
		DC112	Cytoplasmic dynein 1 intermediate chain 2	158	638	1.95 view	0.605	[KR].TQT	KETQT
		DYIN	Cytoplasmic dynein 1 intermediate chain	130	863	1.61 view	0.607	[KR].TQT	KQTQT
		SWA	Protein swallow	283	537	1.52 view	0.474	[KR].TQT	KATQT
		SWA	Protein swallow	291	548	1.4 view	0.532	[KR].TQT	KATQT
		ZMY11	Zinc finger MYND domain-containing protein 11	413	562	0.616 view	0.477	[KR].TQT	RXTQT
		B2L11	Bcl-2-like protein 11	112	198	0.518 view	0.624	[KR].TQT	KSTQT

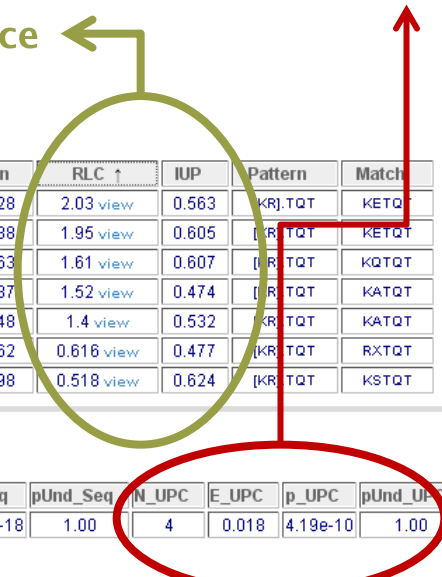
Motif Hits

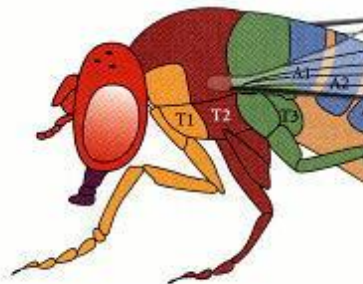
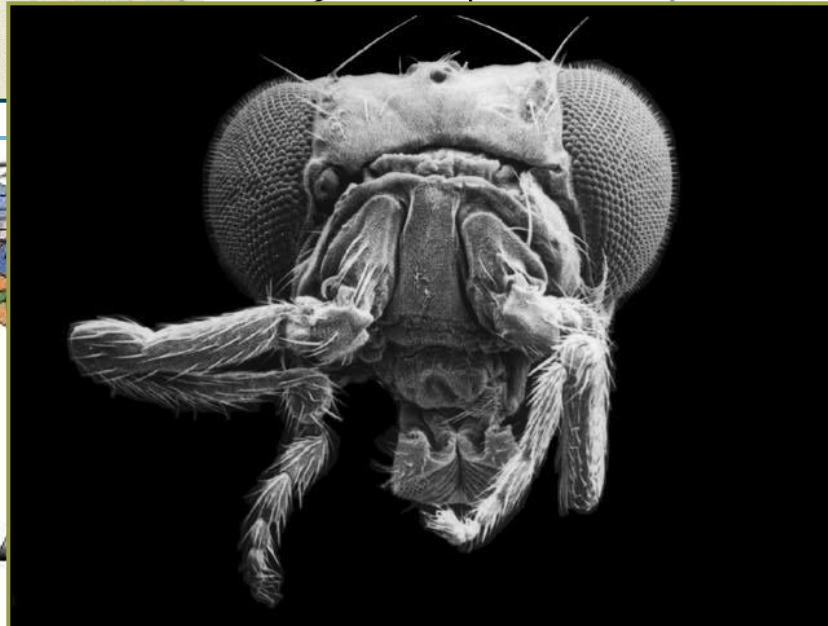
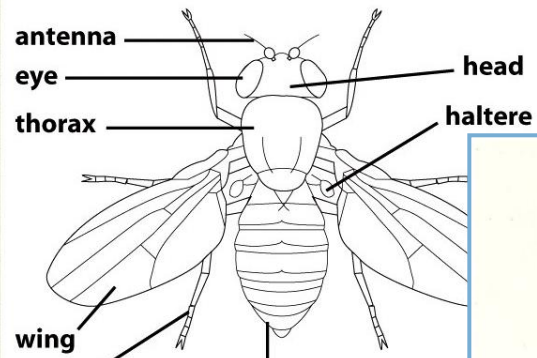
Pattern	IC	SeqNum	N_Occ	E_Occ	p_Occ	pUnd_Occ	N_Seq	E_Seq	p_Seq	pUnd_Seq	N_UPC	E_UPC	p_UPC	pUnd_UPC	Cons_mean	IUP_mean
[KR].TQT	3.77	7	7	0.024	8.60e-16	1.00	7	0.024	5.21e-18	1.00	4	0.018	4.19e-10	1.00	1.38	0.555

[Raw Data](#)

Biological Significance ←

Statistical Significance ↑





Antennapedia complex

lab Pb Dfd Scr Antp



labial (lab)



Deformed (Dfd)



Sex combs reduced (Scr)



Antennapedia (Antp)



Ultrabithorax (Ubx)



Abdominal B (AbdB)

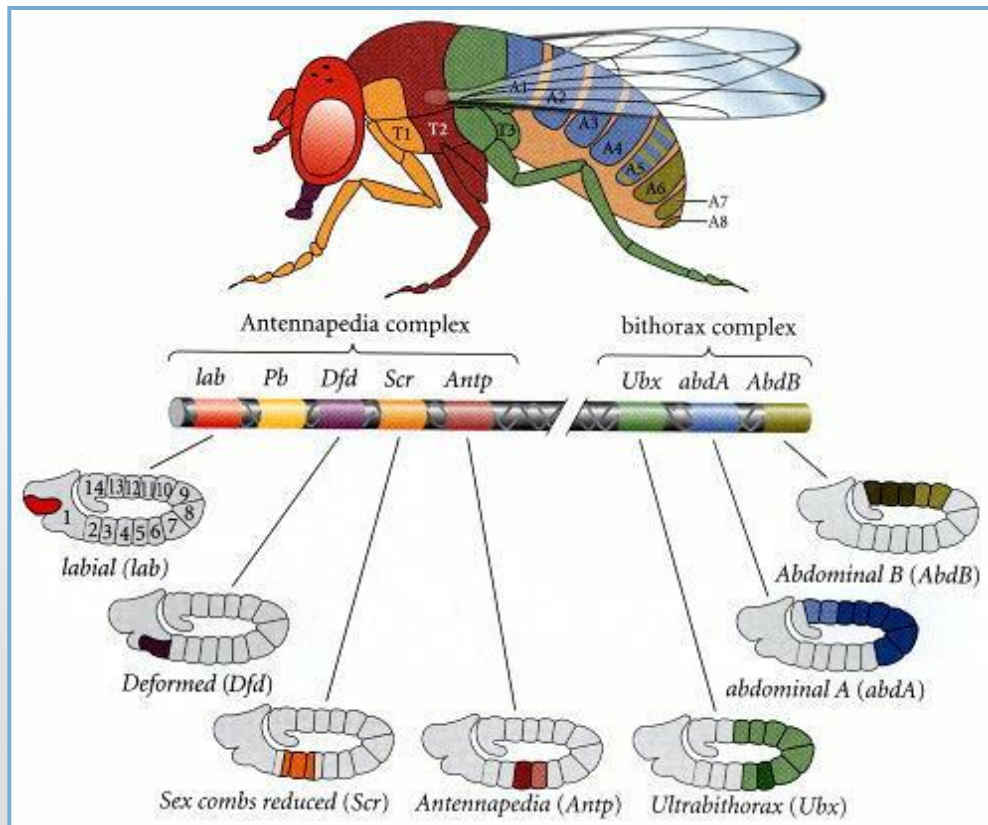


abdominal A (abdA)



Example: Homeodomain ligand

- **LIG_HOMEobox** (www.elm.eu.org)
 - [FY][DEP]WM
 - Mediates interactions between Homeobox proteins and Pbx



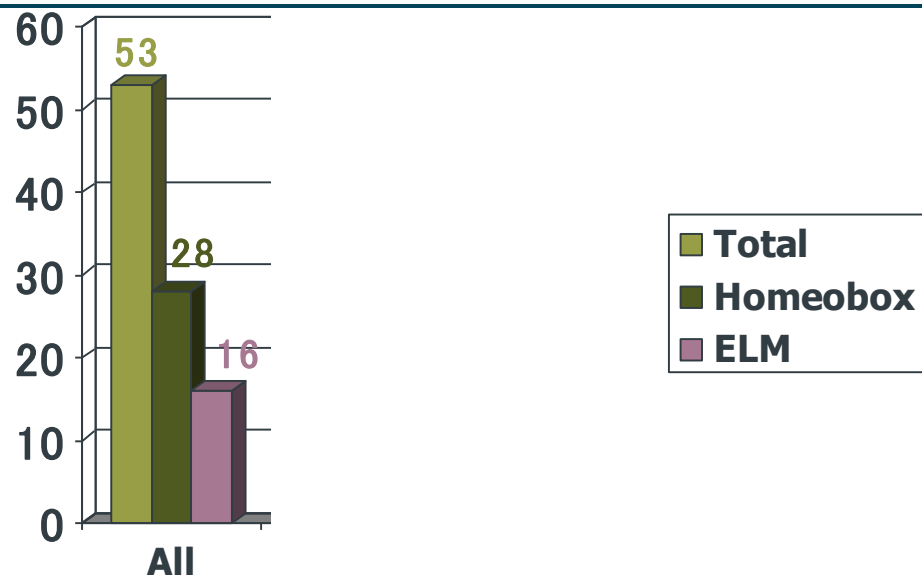
Example: Homeodomain ligand

Sequence ID	Description	Pos	ELM	RLC
HXA5_HUMAN	Homeobox protein Hox-A5	177		1.56
HXC5_HUMAN	Homeobox protein Hox-C5	141		1.496
HXB2_HUMAN	Homeobox protein Hox-B2	95	Y	1.494
HXC6_HUMAN	Homeobox protein Hox-C6	123		1.49
HXB5_HUMAN	Homeobox protein Hox-B5	177	Y	1.458
HXB3_HUMAN	Homeobox protein Hox-B3	130	Y	1.396
HXA6_HUMAN	Homeobox protein Hox-A6	137		1.351
HXA1_HUMAN	Homeobox protein Hox-A1	205	Y	1.155
HXC4_HUMAN	Homeobox protein Hox-C4	136		1.152
HXD3_HUMAN	Homeobox protein Hox-D3	161		1.096
HXD4_HUMAN	Homeobox protein Hox-D4	134	Y	1.079
HXA2_HUMAN	Homeobox protein Hox-A2	95		1.078
HXB6_HUMAN	Homeobox protein Hox-B6	128	Y	1.019
PDX1_HUMAN	Pancreas/duodenum homeobox protein 1	119	Y	0.936
HXB7_HUMAN	Homeobox protein Hox-B7	127	Y	0.894
HXB1_HUMAN	Homeobox protein Hox-B1	180	Y	0.824
HXB4_HUMAN	Homeobox protein Hox-B4	142	Y	0.807
HXD1_HUMAN	Homeobox protein Hox-D1	205		0.747
TLX1_HUMAN	T-cell leukemia homeobox protein 1	174	Y	0.711
HXA7_HUMAN	Homeobox protein Hox-A7	120		0.668
CDX1_HUMAN	Homeobox protein CDX-1	133	Y	0.598
HXB8_HUMAN	Homeobox protein Hox-B8	135	Y	0.562
HXA3_HUMAN	Homeobox protein Hox-A3	156		0.537

Accepted
(not masked)

Sequence ID	Description	Pos	ELM	RLC
HXA4_HUMAN	Homeobox protein Hox-A4	195		1.145
HXD8_HUMAN	Homeobox protein Hox-D8	186		0.978
MTMRD_HUMAN	Myotubularin-related protein 13	1594		0.975
CBPO_HUMAN	Carboxypeptidase O	57		0.902
HXC8_HUMAN	Homeobox protein Hox-C8	139	Y	0.883
GRP3_HUMAN	Ras guanyl-releasing protein 3	126		0.864
EXC6B_HUMAN	Exocyst complex component 6B	627		0.816
SAST_HUMAN	S-acyl fatty acid synthase thioesterase, medium chain	32		0.652
UBXN8_HUMAN	UBX domain-containing protein 8	218		0.539
MLTK_HUMAN	Mitogen-activated protein kinase kinase kinase MLT	170		0.469
PKDRE_HUMAN	PKD and REJ homolog	505		0.389
ZSCA1_HUMAN	Zinc finger and SCAN domain-containing protein 1	389		0.289
FOXO3_HUMAN	Forkhead box protein O3	184		0.228
RBP17_HUMAN	Ran-binding protein 17	741		0.203
XPP3_HUMAN	Probable Xaa-Pro aminopeptidase 3	204		0.197
FOXO4_HUMAN	Forkhead box protein O4	128		0.191
UBP24_HUMAN	Ubiquitin carboxyl-terminal hydrolase 24	1906		0.183
FOXO6_HUMAN	Forkhead box protein O6	115		0.182
D42E2_HUMAN	Putative short chain dehydrogenase/reductase family 42E member 2	286		0
CNOT4_HUMAN	CCR4-NOT transcription complex subunit 4	498		0
CLIP3_HUMAN	CAP-Gly domain-containing linker protein 3	537		0
FOXO1_HUMAN	Forkhead box protein O1	187		-0.027
HHAT_HUMAN	Protein-cysteine N-palmitoyltransferase HHAT	201		-0.141
FHAD1_HUMAN	Forkhead-associated domain-containing protein 1	103		-0.146
TLX2_HUMAN	T-cell leukemia homeobox protein 2	118	Y	-0.505
SRBP2_HUMAN	Sterol regulatory element-binding protein 2	532		-0.644
SFXN1_HUMAN	Sideroflexin-1	260		-0.67
TLX3_HUMAN	T-cell leukemia homeobox protein 3	125	Y	-0.764
FA13A_HUMAN	Protein FAM13A	179		-0.927
GDE5_HUMAN	Putative glycerophosphodiester phosphodiesterase 5	612		-0.932

Rejected
(masked)



**Strong enrichment
for
Biological Significance**

Example: Homeodomain ligand

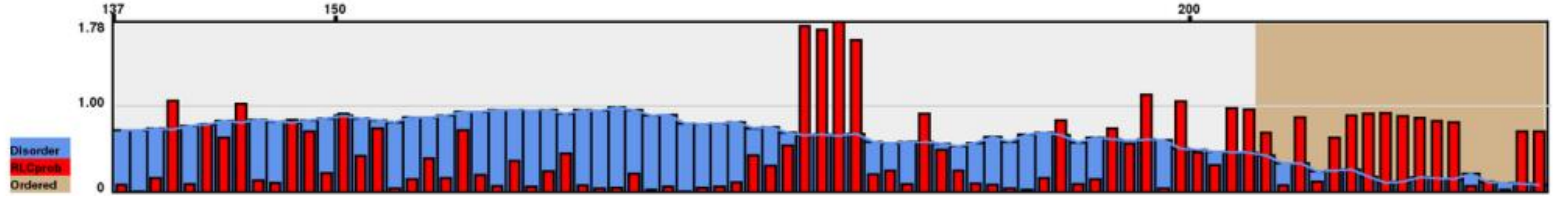
Sequence ID	Description	Pos	ELM	RLC
HXA5_HUMAN	Homeobox protein Hox-A5	177		1.56

137-220

HXA5

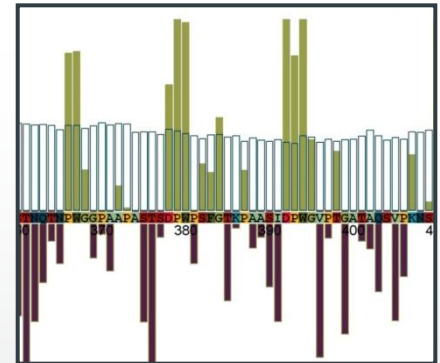
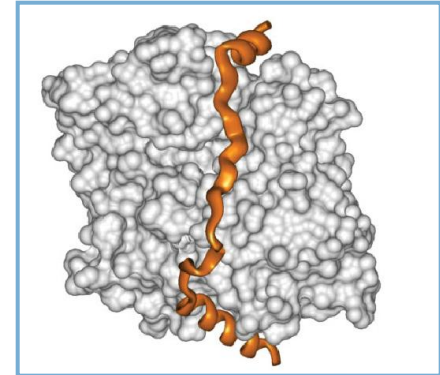


CHAIN	Homeobox protein Hox-A5, #TId:PRO_000020057.
DNA_BIND	Homeobox.



Summary/Future work

- Incorporating biological information improves biological significance
- Next:
 - SLiMSearch 2.0
 - Gene Ontology (GO) enrichment
 - Improved data integration
 - Weighting, not masking
 - Bayesian probabilities?



Acknowledgements

Bioware CompariMotif SAAMCO Gopher SLiMDisc SLiMFinder SLiMPrints SLiMSearch

Discovery@Bioware

About Bioware

Welcome to the Bioware webserver for the Clinical Bioinformatics lab. This server contains several tools for motif discovery related analysis of proteomics data.

[View queue](#)

Description

This server contains several bioinformatics tools, with an emphasis on motif discovery, hosted by the Clinical Bioinformatics group led by Prof. Denis Shields. The tools have been developed by [Rich Edwards](#) (currently at The University of Southampton) and [Norman Davey](#) (currently at EMBL Heidelberg).

This project is a collaboration between 3 institutions, [Conway Institute of Biomolecular and Biomedical Research](#) at University College Dublin (Dublin, Ireland), [School of Biological Science at University of Southampton](#) (Southampton, England) and [European Molecular Biology Laboratories](#) (Heidelberg, Germany). The server is situated in the [Conway Institute of Biomolecular and Biomedical Research](#).

Various Lab related Links

Publications

A list of recently published articles relating to the applications on this server.

Downloads

Software downloads for local versions of the motif discovery tools available on this server

License

This server may only be used under the terms of the [GNU GPL](#)

Collaborators



Prof Denis Shields
Dr Niall Haslam



Dr Norman Davey

Norman Davey (2009-2010)

Questions?