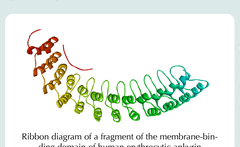
Spatial organization and distribution of linear motifs in the Ankyrin repeat protein family and its interacting partners

N. Verstraete, I.E. Sánchez and D.U. Ferreiro

Protein Physiology Laboratory, IQUIBICEN and CONICET, University of Buenos Aires, Argentina

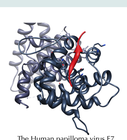
Introduction

Interactions between proteins regulate cellular physiology. Many of these interactions involve the recognition of short peptidic regions (i.e. short linear motifs, SLiMs) which can be characterized by simple sequence patterns of 3 to 11 residues, usually found in intrinsically disordered regions or in loops connecting globular or transmembrane domains. These peptide-domain interactions are typically transient and often involve folding upon binding, challenging the lock-and-key paradigm of protein recognition.



Ribbon diagram of a fragment of the membrane-binding domain of human erythrocytic ankyrin (ankyrin 1, ankyrin R).

Ankyrin-repeats domains are one of the most frequently observed protein-protein interactors in nature. These domains are composed of tandem arrays of 33 recurrent amino acids that cooperatively fold into elongated structures that mediate molecular recognition with high specificity. They are found in proteins with diverse functions such as transcription initiation, cell cycle regulation, cytoskeletal integrity, ion transport or cell–cell signaling. Many ankyrin-binding sites are either predicted or demonstrated to correspond to extended peptides mimicking SLiMs.



The Human papilloma virus E7 oncoprotein mimic of the LxCxE motif (red) bound to the host Retinoblastoma protein (dark grey)(PDB 1gux)

We present here an exhaustive analysis of linear motif identification in Ankyrin proteins and their binding partners. We searched for enriched or depleted SLiMs with respect to a random exploration of the sequence-space in the Ankyrin protein family and their partners. We also analyzed the spatial distribution of SLiMs along the protein sequences and describe how particular SLiMs are structurally distributed in the Ankyrin-containing proteins.

Methods

1. Retrieve Ankyrin protein sequences from Uniprot (1234 sequences)
2. Reduce sequence redundancy (CD-hit 0.5, 690 sequences)
3. Retrieve all experimentally verified interacting partners for each Ankyrin protein from STRING database (exp_score > 0, 15'267 sequences)

690 Ankyrin proteins
+
15267 interacting partners

4. Search for domains (SMART database) and measure enrichment
5. Search for linear motifs (ELM database) based on their regular expressions (ex : LIG_ULM_U2AF65_1 → [KR]{1,4}[KR].[KR]W.) and measure enrichment
6. Identify linear motifs enriched and conserved in Ankyrin protein homologs

Domain enrichment :
A domain D is considered enriched in Ankyrin proteins if the ratio observed_counts/expected_counts > 1,
with expected_counts = (# of times domain D is found in all Uniprot proteins / # of proteins in Uniprot) x (# of Ankyrin proteins)

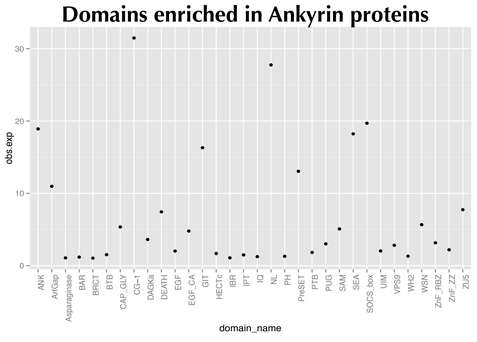
Linear Motif enrichment :
A linear motif LM is considered enriched in a sequence family if the ratio observed_counts/expected_counts > 1,
with expected_counts = probability(LM) x total length of the sequence family.
Probability(LM) is calculated according to the amino-acid frequencies of the sequence family.

Linear motifs conservation measurement :
1. Retrieve homologs (blastp) for each Ank protein
2. Check at SLiM matches positions if the Regular Expression is conserved (not the matched sequence)
3. Score SLiM conservation in each Ank protein homologs family
4. Final score : average SLiM conservation scores over all Ank protein homologs families

Results

1. Ankyrin proteins are enriched in specific domains and linear motifs (Ankyrin-associated domains and linear motifs)

Domains enriched in Ankyrin proteins

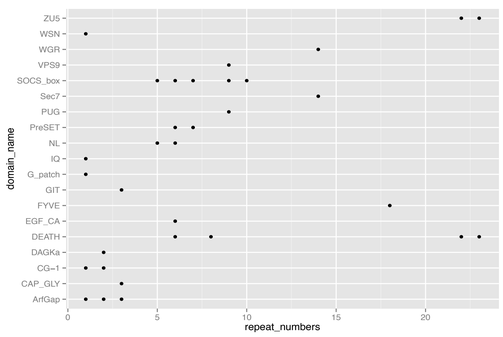


Top 10 Ankyrin-associated domains

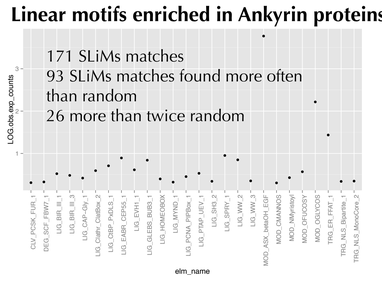
ANK	PreSET
ArfGap	PUG
CAP_GLY	SAM
CG-1	SOCS_box
DAGKa	WSN
DEATH	ZU5
EGF_CA	
GIT	
NL	

- 72 SMART domains found in Ank proteins
- 34 found more often than expected by random (obs/exp > 1)
- 16 found twice more often than expected by random (obs/exp > 2)

Domains enriched in Ankyrin proteins and repeat numbers



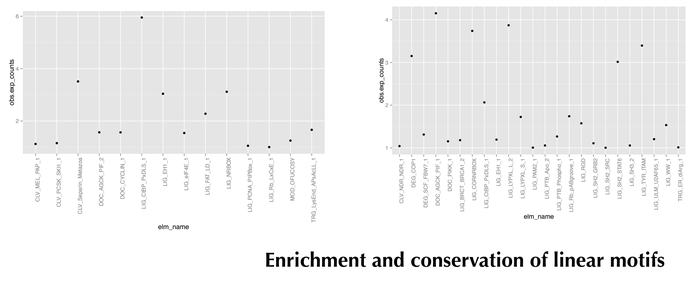
Linear motifs enriched in Ankyrin proteins



Top 10 SLiMs in **Ank proteins** → in **Ank-containing region**

1. MOD_ASX_betaOH_EGF → absent
2. MOD_OGLYCOS → absent
3. TRG_ER_FFAT_1 → absent
4. LIG_SPRY_1 → absent
5. LIG_EABR_CEP55_1 → absent
6. LIG_WW_2 → absent
7. LIG_GLEBS_BUB3_1 → absent
8. **LIG_CtBP_PxDLS_1 → enriched**
9. LIG_EVH1_1 → depleted (under-represented)
10. LIG_Clathr_ClatBox_2 → absent

Linear motifs enriched in Ankyrin repeats and linkers

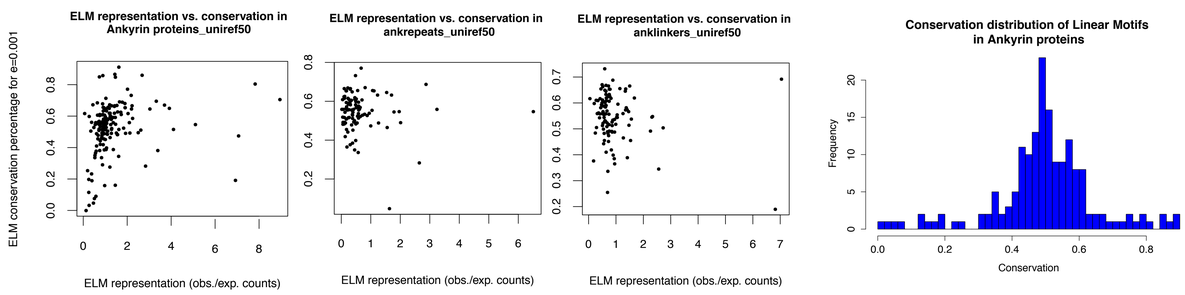


SLiMs enriched and conserved in Ank proteins	Associated domains
MOD_ASX_betaOH_EGF	Asp_Arg_Hydrox
LIG_BIR_II_1	BIR
MOD_CAAXbox	PPTA
MOD_NMyristoyl	NMT_C
LIG_PDZ_Class_1	PDZ
MOD_OGLYCOS	Galactosyl_T_2
LIG_EABR_CEP55_1	EABR
TRG_ER_FFAT_1	Motile_Sperm
TRG_LysEnd_GGAAcLL_1	VHS
LIG_SPRY_1	SPRY
LIG_BIR_III_1	BIR
TRG_ER_diLys_1	WD40
LIG_Clathr_ClatBox_2	Clathrin_propel
LIG_BIR_III_3	BIR
TRG_PTS1	TPR_1

SLiMs enriched and conserved in Ank repeats	Associated domains
LIG_NRBOX	Hormone_recep
MOD_OFUCOSY	O-FucT
TRG_LysEnd_APsAcLL_1	Clat_adaptor_s
DOC_CYCLIN_1	Cyclin_N
CLV_PCSK_SKI1_1	Peptidase_S8

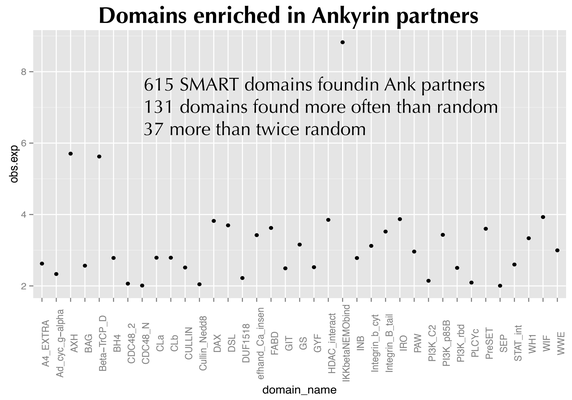
SLiMs enriched and conserved in Ank linkers	Associated domains
DOC_AGCK_PIF_1	Pkinase

Enrichment and conservation of linear motifs



2. Domains and linear motifs enriched in Ankyrin interacting partners.

Domains enriched in Ankyrin partners



Linear motifs enriched in Ankyrin partners

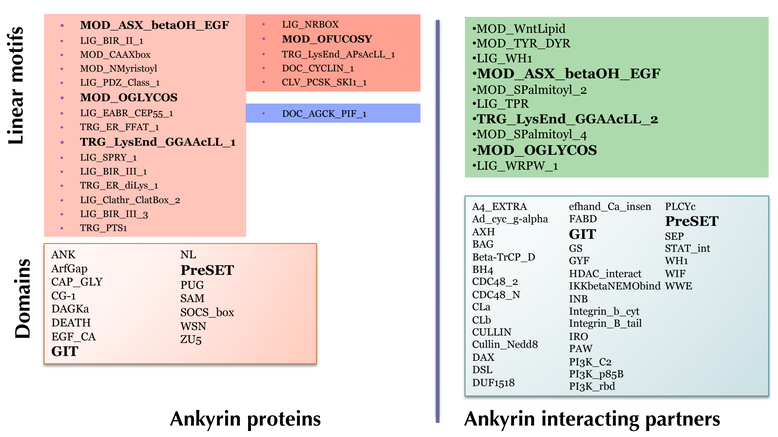
Top ten SLiMs in Ankyrin interacting partners:

- MOD_WntLipid
- MOD_TYR_DYR
- LIG_WH1
- MOD_ASX_betaOH_EGF
- MOD_SPalmitoyl_2
- LIG_TPR
- TRG_LysEnd_GGAAcLL_2
- MOD_SPalmitoyl_4
- MOD_OGLYCOS
- LIG_WRPW_1

Associated domains:

- MBOAT
- Pkinase
- WH1
- Asp_Arg_Hydrox
- zf-DHHC
- TPR_1
- VHS
- zf-DHHC
- Galactosyl_T_2
- WD40

3. Several domains and linear motifs are enriched both in Ankyrin proteins and their interacting partners.



Conclusions

We present here an extenive analysis of linear motif identification in Ankyrin proteins and their interacting partners. We report several Ankyrin-associated domains and linear motifs. We also describe how some Ankyrin-associated linear motifs distribute specifically along the protein sequence and structure.

Linear motifs are platforms for interactions and are generally recognized by globular protein domains, which we found in Ankyrin interacting partners. Conversely, we suggest that interacting partners are enriched in linear motifs which bind Ankyrin-associated domains.

Furthermore, we showed that several domains and linear motifs were enriched both in Ankyrin and their interacting partners, suggesting co-evolution and functional association between some domains and linear motifs (domain-domain, linear motif-linear motif and/or domain-linear motif functional association).

Altogether, these preliminary results suggest that combining analysis of domain-linear motifs architecture, co-occurence networks and experimental data for interactions can help to better predict functional linear motifs.