

Figure S1: Nuclear Localization Signal

gi 21265034	[Homo sapiens]	29	----PSHFQQSCLQALEPQAVSSYLSPGAPLKGRPPSPGFQRQRQRRAAGGILHL	82
gi 109109894	[Macaca mulatta]	13	----VKEWFGACEESLP-----	26
gi 76671947	[Bos taurus]	27	----PSPGCVCCKHGPRLSALP-----APSPGRPLAPLRRRAAGHVLHL	68
gi 47604978	[Mus musculus]	33	----LSDFQKSLLDLEPKDVSSYFG--HHAAPFTGHPPSHLQRLRRRRTLEDILHL	84
gi 109467944	[Rattus norvegicus]	33	----LSDFQKSFLLDLEPKDLSSSFS--RHTAPLTGHPPSHLQRLRQRRTLEDILHL	84
gi 126297969	[Monodelphis domestica]	150	KQLYIQPLKSEHWNLLPDCCRAQPHLIHKAMRAAGAPQVHRLSARISKRALGSKVHL	207
gi 118099369	[Gallus gallus]	8	-----HFAPRLQKRAAGAVKHL	25

Figure S2: Sequence alignment of CUB-2 domain homologs

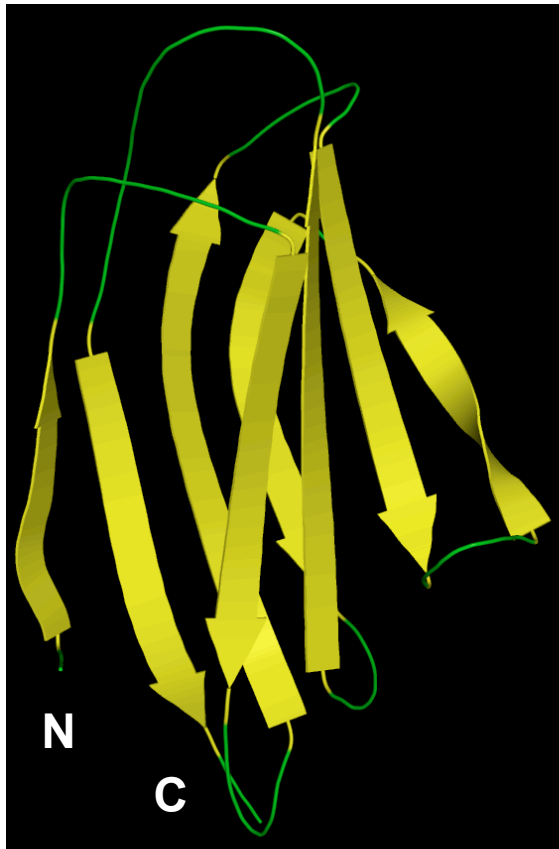
Mapped Residues		*	* * * *	* * * *	
291243359_Sko_LOC100378044	1828	-----SGVISSLNYPNPYP--INTNLTIIVRGPQDSHIIIVTF--EGCLMTGSAHCAKSLRVVNYLL--YSERREIGQIMCGDECNITTED-----	VLTVQSTLNTLFI	IEFVSHD----	FELEGFQARYRV
189236895_Tca_C3po	3059	-----EAGCGGKIVLNKTMTLEAPNLADMDCLEWTIQSQIDTTVKIHFDELNLPS-----	SCAALEVRDG--	HPGVGLIERFCG	ITDTHYS-----REITTSGNVM-----
241164816_Isc_IscW_ISCW002397	220	-----TAGAGAIASPGFAEQRLPANRKCQWEIQAEAGHQVKAFAFRLRLEPHRLGG--EAPHLVYFDG--DENTPLANL	SAVQIEEGESP-----	LNVTTVSSKLLV	LVLLTGTRL-----EDAVMYMEYAANPS
221129558_Hma_LOC100208297	399	-----NPSDDTACSELVKPTGYINVIKP--KN--CTWNIITVPVKLVEIFIIINIDMADDDPC--HSEIEVRD--	YVNAPLLFR	TCGQKSKPIL-----	TSSNTAFIQHMLNT-----SRLSFLKLFQAVD
31615702_Rno_Masp2	8	-----SKWPEPVGRLVSPGFPEKYG--NHQDRSWTLAPPGRFLRLYFTHFNLELSYRC--EYDFVKLTS--	GTKVLATL	CGQESTDTERAP----	GNDTFYSLGPSLKVTFHSDYSNEK----
290447443_Cel_Y69H2.17	35	SACECLDRIVLLHNFQDFRLISSPNYPEAYC--TGMDCWHHFVAPNAGQKVFHSTDNLDIRK----	NLDQILFYD--	FKTSQLEN	STESFRCTGDD-----LCYYSKFKQVLTIRFKTQ----
115764766_Spu_LOC756506	607	YECCDPGCGIQEPFATSATGNFLENFGDSNAAPFVSVCTWITLPEGHFIFNFTEFDVKRDVESGACVDSVEIFPTEHRWDEGRSRHVFCGQGAAYV-----	MTNVSSVT	VI	FGTGLD-----GSSGFQVYASKAV
260797195_Bfl_BRAFLDRAFT_88162	377	SKGQPLPPCGGYLEGSRNGTITSPNFPSSAHT--PLECWTIRNPKPGRHYIYVYLLQNYLSC----	QSSALTYLEY--	QSVSNL	TDSYIYSSNFNSRNDCCGSLHRIRSSKAF
115665188_Spu_LOC576545	174	EDEDEDETECKGT----GGIIMSKNYFNDYG--EDLQHTWTIITRLWTKVLFTFIEIDVYSPEDDC--	GDDYVNV	SVS--	GEHKVYCKVPGKSAVVESSG-----NVMQIEFYSDVQYTI----
21265034_Hsa_ADAMTS13	1290	LAPETFYRECDMQLFGPWGEIVSPSLSPATR--NAGGCRLFINVAPHARIAIHALATNMGAGTEGA--	NASYIL	IRD--	HSLRTTAFHGQQLVWYSES-----SQAEMFSEGFLKAK-----
296191110_Cja_ADAMTS13	1270	-APETFYRECDMQLFGPWGEIVSPSLSPATR--NAGGCRLFINVAPHARIAIHALATNMGAGTEGA--	NASYIS	IRD--	HSLRTTFRGQQLVWYSES-----SQAEMFSEAFLEAK-----
301770673_Ame_ADAMTS13	1368	-APGTPHRCDMQLFGPWGEIVSPSPSPNGS--NSGGCRIFINVAPQARIAIHALVTDVGTGTG--	ASYIS	IRD--	HSLKTMFRGQQLVWYSEG-----SQAEMFSSQGLEAH-----
76671947_Bta_ADAMTS13	1272	-ALGAFHRCDMQLFGPRGEIVSPSPSPDGR--NVGGCRIFINVAPWARIAIHALTVDSGTRAEGT--	DASYIL	IRD--	HSLRTTAFRQQLVWYSEG-----SQAEMFSSQGLEAH-----
291416170_Ocu_ADAMTS13	1094	-APQP--YRECDVQLFGPRGEIVSPSPSAHGG--RAGGCRLFISVAPQARIAIHALAAGTAPGTEGA--	HSWST	TA	FRGQQLVWYSESG-----SQAEVEFSQGLEAG-----
47604978_Mmu_ADAMTS13	1295	-APGTIFYKCDRQLFGPRGEIVSPSLSPDGR--KAGTCRVFISVAPQARIAIHALASDMGTASEGT--	NANYVS	IRD--	HSLRTTFRGQQLVWYSEG-----SEAELEFSPGLEAH-----
126297969_Mdo_LOC100018539	1413	VATENYHKECDVQLFGPRGEIVSPPRDQNR--NQGGRVFINVAPGARIAIHALAIDLGTEANQT--	SDSYIS	IRD--	HNMKMITFHGQQLVWYSEG-----SRAEIEFSQFQEDN-----
224073646_Tgu_LOC100226372	1105	-ATKYYQDCDQQLFGPQGEIVAPVER----QEGVCRFTIVAPRQIRISIRALNTDLGPEGNQT--	DFNYIL	VRD--	STMKTVVFRGQQLVWQSTG-----SQAEIEFHENVKDHR-----
118099369_Gga_ADAMTS13	822	-AIKYYQDCDQQLFGPQGEIVNPTQSPGQN--QEVVCRFTINVAPQHIIAIRGLYVDLSHESNQT--	HFNYIL	IRD--	STMKTMVFRGQQLVWQSTG-----SQAEIEFHENIKDHQ-----
268556522_Cbr_CBG008530	1074	-----TCDYTMTPSGSLVFNPDMGT----KYDKCEYHIAVHDNRRILLKMNMTLPC-----	GKSSL	SFRNGPT	ETSPFSLIPPEVCTPQVN----YMPMIRSFNSRVTIVFKSVNTEG-----
195442916_Dwi_Dwil_GK24548	1008	-----CGSNMTPGVGTIASSML-----TTNCDWHIEVNAGRKIDITIQYAKINMNIT--	CPSYGL	IY--	DGLDNAAPLMVNGKFCNRH-----DFATENYRSGS-----
3318759_Ssc_1SPPB	8	-ARINGPDECGRVIKDTSGISNTDRQ-----KNLCTWITLMPKPDQKVRMAIPYLNLC-----	GKEYVE	VFDGLL	SGPSYGKLCAGAAIVFLST-----ANTMTIKYNRISGNSS-----
1SPPB_Sec_Struct		-EEEE-----EEE--EEEEEEE-----EEEEEEEE-----EEEEEEEE-----	EEEEEE	-----	EEEEEEEEEEEEEEEE-----EEEEEEEE-----
Jpred prediction		-----EEEE-----EEEEEEEE-----EEEEEEEE-----EEEEEEEE-----	EEEEEE	-----	EEEEEEEE-----EEEEEEEE-----
consensus/95%	p.....l.....h.....l.....h.....l.....h.....l.....h.....h.....
consensus/90%	s.....ls.....s.....hph.l.s.....pl.h.h.....p.....l.....hb.....p.....s.....sph.....h.....h.....l.....hb.....p.....s.....sph.....h.....h.....
consensus/85%	ss.G.lssssb.....Cph.I.s.....pl.h.h.....p.....s.....l.....hb.....p.....s.....sphph.....p.....hb.....a.s.....s.....l.....hb.....p.....s.....sphph.....hb.....a.s.....
consensus/80%	us.G.lssssb.....Cph.Ips.....pl.h.h.....sh.....p.s.l.lbs.....p.....p.....s.....sphphb.p.....p.....hb.ba.s.....p.....s.....l.....lbs.....p.....p.....s.....sphphb.p.....hb.ba.s.....
consensus/75%	s.....us.G.lssssb.....Cph.Ipsss.+l.l.h.....sh.....p.s.l.lbs.....s.ps..bp.....hb.ppps.....sphphba.p.....p.....shsbsy.s.s.....p.....s.....l.....lbs.....s.....ps..bp.....hb.ppps.....sphphba.p.....shsbsy.s.s.....
consensus/70%	ss.....us.GpIsusbs.s.....ssCph.Ipsss.+l.lph.....shsp.....s.sbl.lb.....s.ps.sbps.....shbhsps.....sphphbapps.p.....sshsbsy.s.s.....s.....ps.sbps.....shbhsps.....sphphbapps.p.....sshsbsy.s.s.....

Legend:
 Sko: *Saccoglossus kowalevskii*, Tca: *Tribolium castaneum*, Isc: *Ixodes scapularis*, Hma: *Hydra magnipapillata*, Rno: *Rattus norvegicus*
 Cel: *Caenorhabditis elegans*, Spu: *Strongylocentrotus purpuratus*, Bfl: *Branchiostoma floridae*, Hsa: *Homo sapiens*, Cja: *Callithrix jacchus*
 Ame: *Ailuropoda melanoleuca*, Bta: *Bos taurus*, Ocu: *Oryctolagus cuniculus*, Mmu: *Mus musculus*, Mdo: *Monodelphis domestica*, Tgu: *Taeniopygia guttata*
 Gga: *Gallus gallus*, Cbr: *Caenorhabditis briggsae*, Dwi: *Drosophila willistoni*, Ssc: *Sus scrofa*

- a: Aromatic residue
- b: Bulky residue
- C: Cysteine
- G: Glycine
- h: Hydrophobic residue
- l: Leucine or Isoleucine
- p: Polar residue
- s: Small residue
- u: Uncharged residue
- Y: Tyrosine

Figure S3: Structural similarities between CUB domain, B1 domain of neuropilin-2 and Nucleoplasmins

CUB domain of major seminal plasma glycoprotein psp-II



PDB: 1spp, Chain B

B1 domain of Neuropilin-2



PDB: 2qq1, Chain A

Drosophila NLP-core (Nucleoplasmin family)



PDB: 1nlq, Chain A

Table S1: DALI structural hits for query CUB domain structures

Protein domain	PDB code	Dali Z-score
B1 domain of neuropilin-2	2qql, chain A	13.6
Class 1 collagenase	1nqd, chain A	7.4
Microvirus capsid protein Procapsid of bacteriophage phix174	1cd3, Chain G	6.2
Tumor necrosis factor ligand superfamily member 4, OX40L	2hew, Chain F	6.1
Supernatant protein factor (SPF), C-terminal domain	1o6u, Chain E	6.0
Chromatin decondensation protein 1 (Crp1, Nlp) Nucleoplasmin-like	1nlq, Chain A	5.6