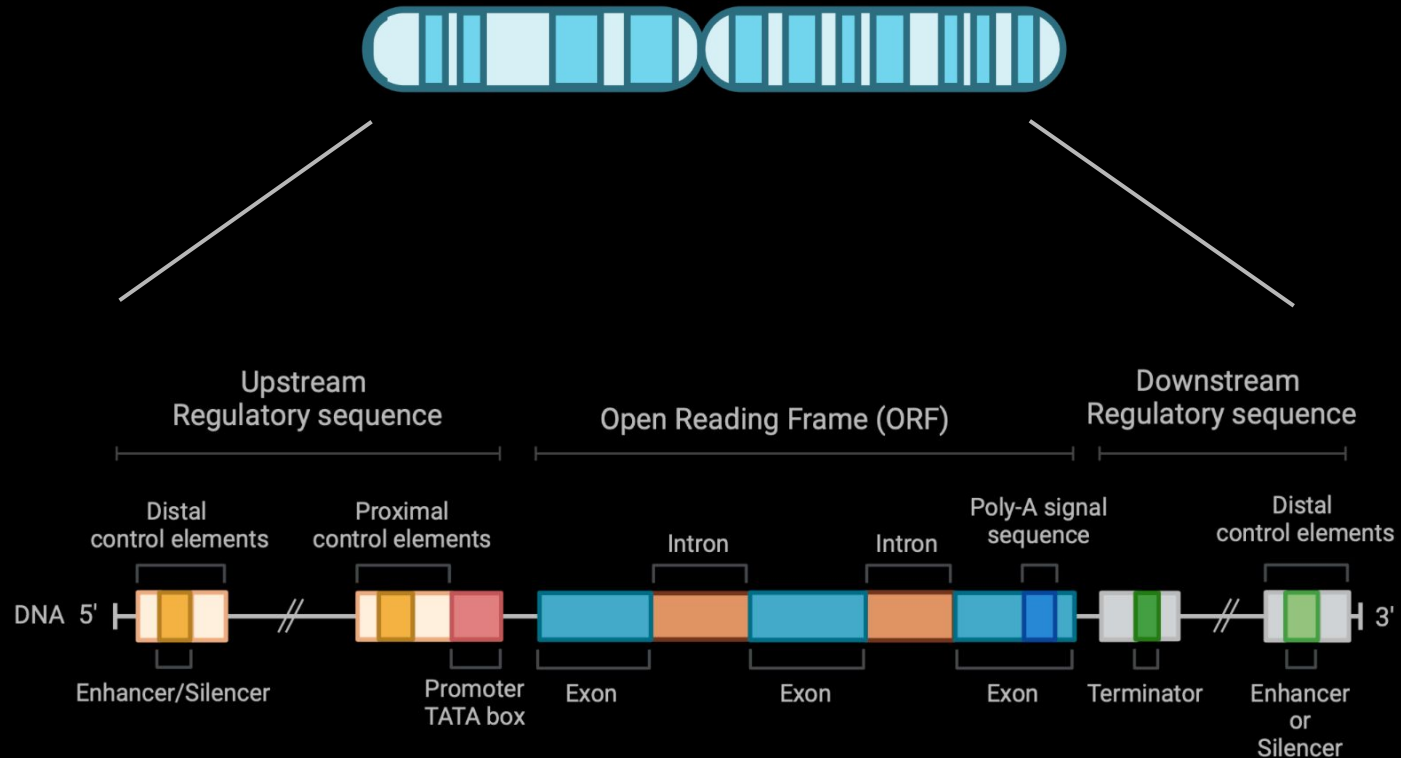
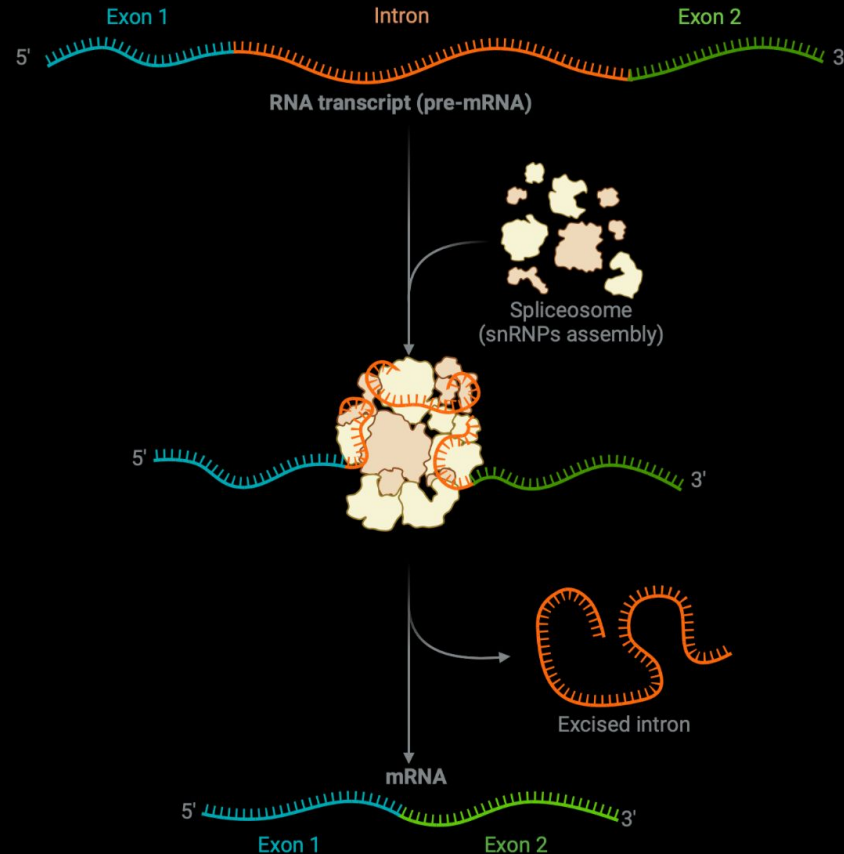


Eukaryotic gene structure



Splicing



The Spliceosome

Marc Pinós, María Reyes and Cristina Ventura

Index

1.

The Spliceosome and
their Complexes

2.

PRP28 Structure

3.

RNA unwinding by
DExD-box helicases

4.

ATP hydrolysis by
DEAD-box helicases

5.

Helicase core
conformation

6.

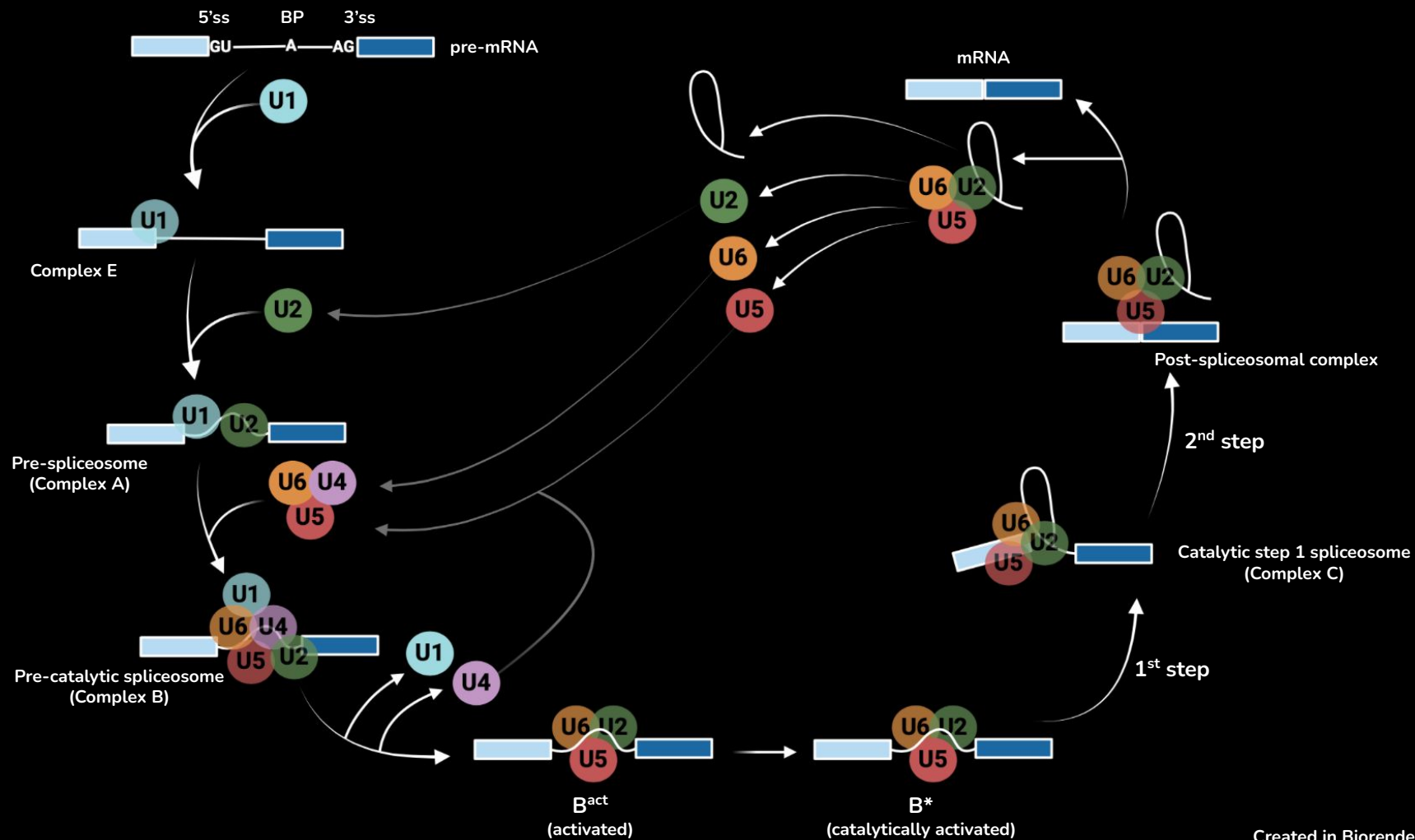
Polarity

7.

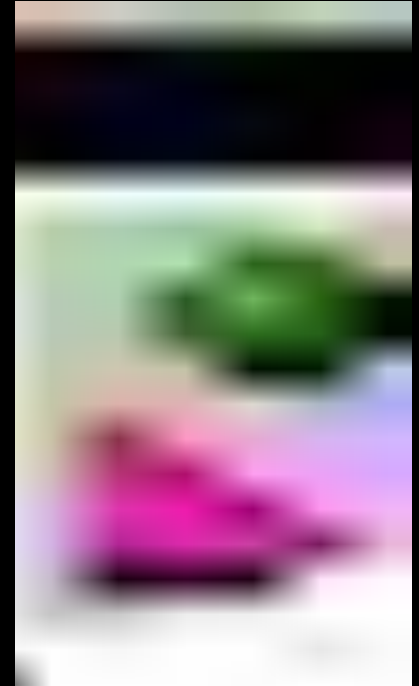
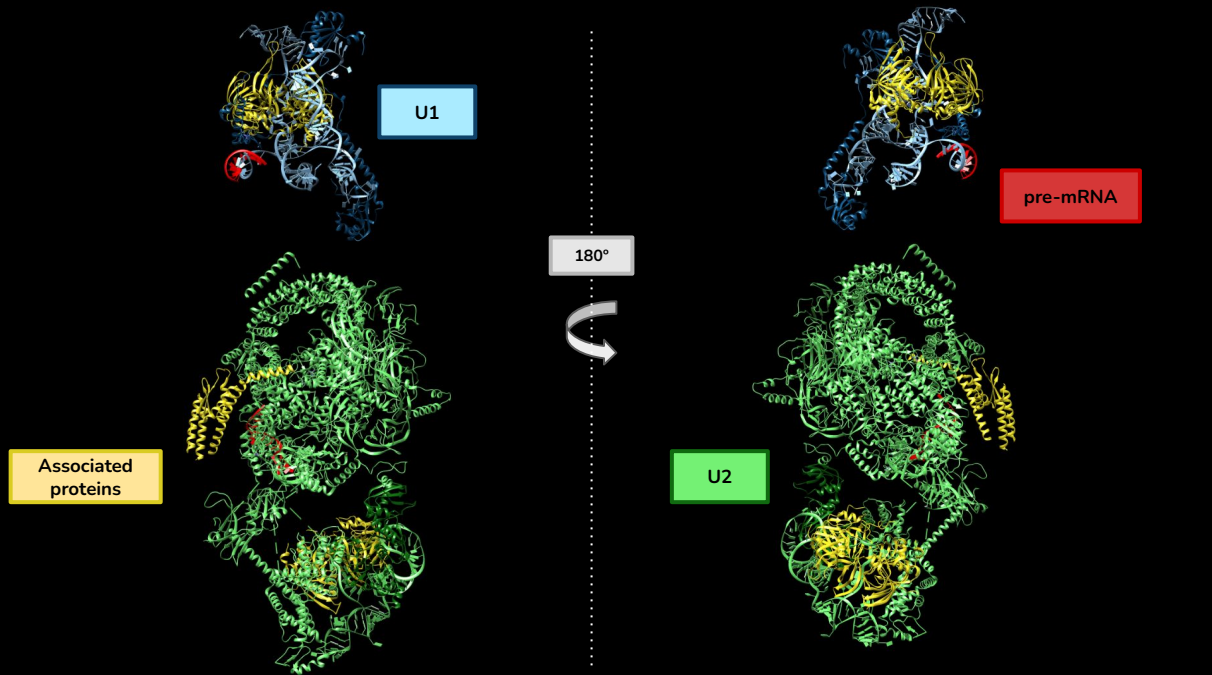
Buried Surfaces

8.

Likely pathogenic
variant

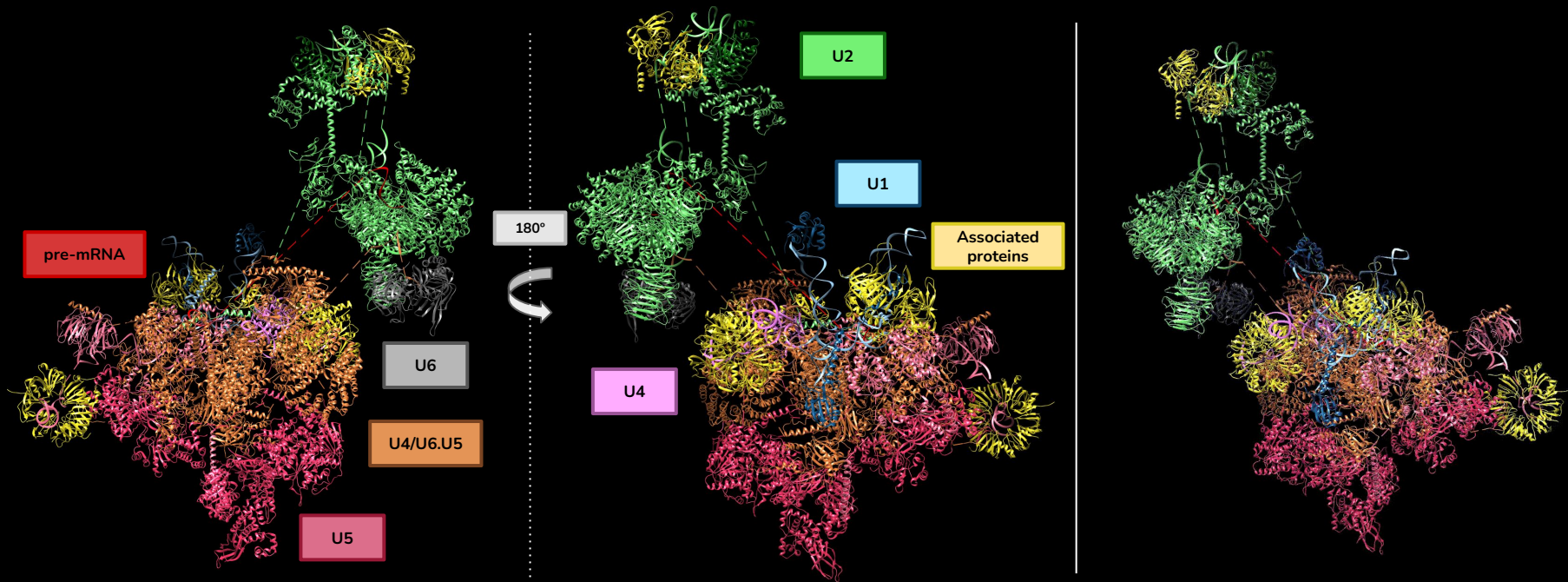


Complex A



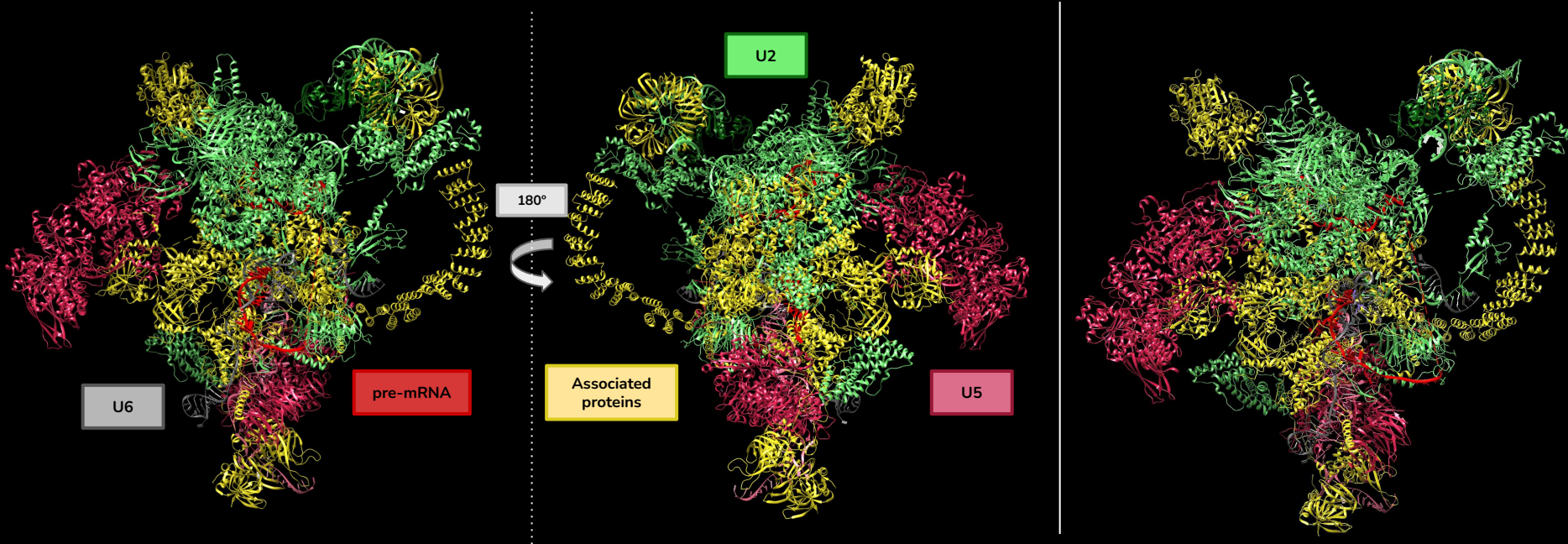
The cryo-EM structure of the human pre-A complex (7VPX): *Homo sapiens*, 3.00Å

Complex B



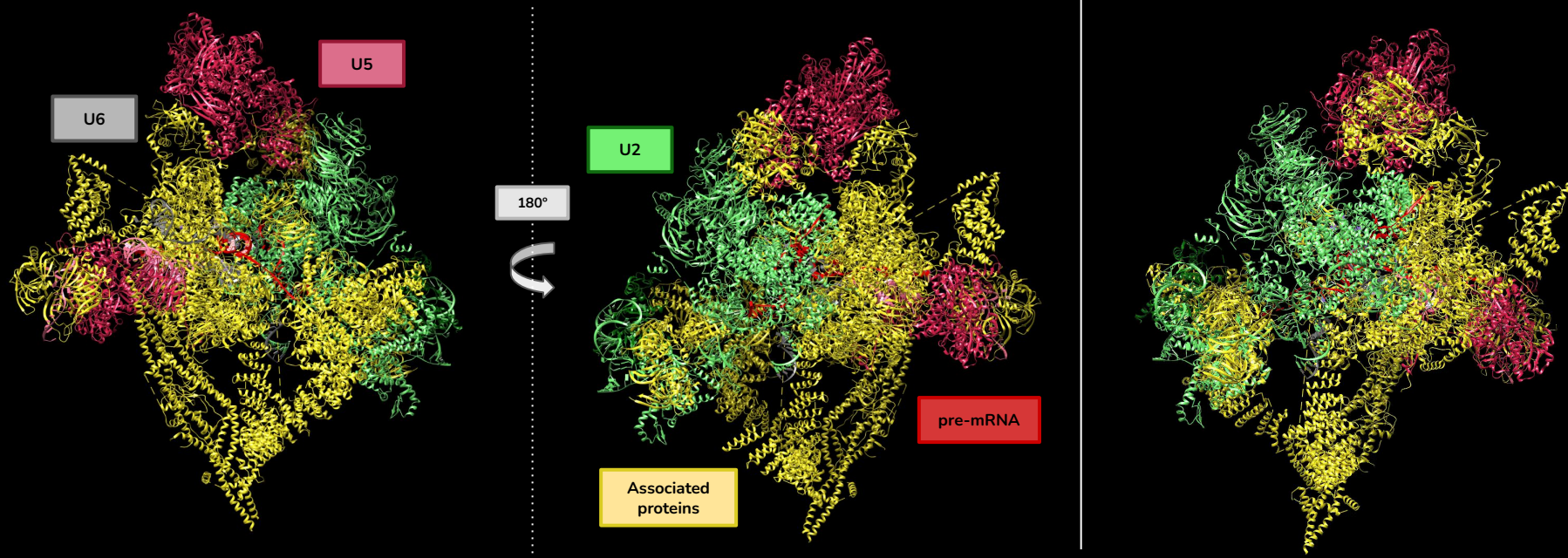
Structure of a human fully-assembled precatalytic spliceosome (pre-B complex) (6QX9): *Homo sapiens*, 3.28Å

Complex B^{act}



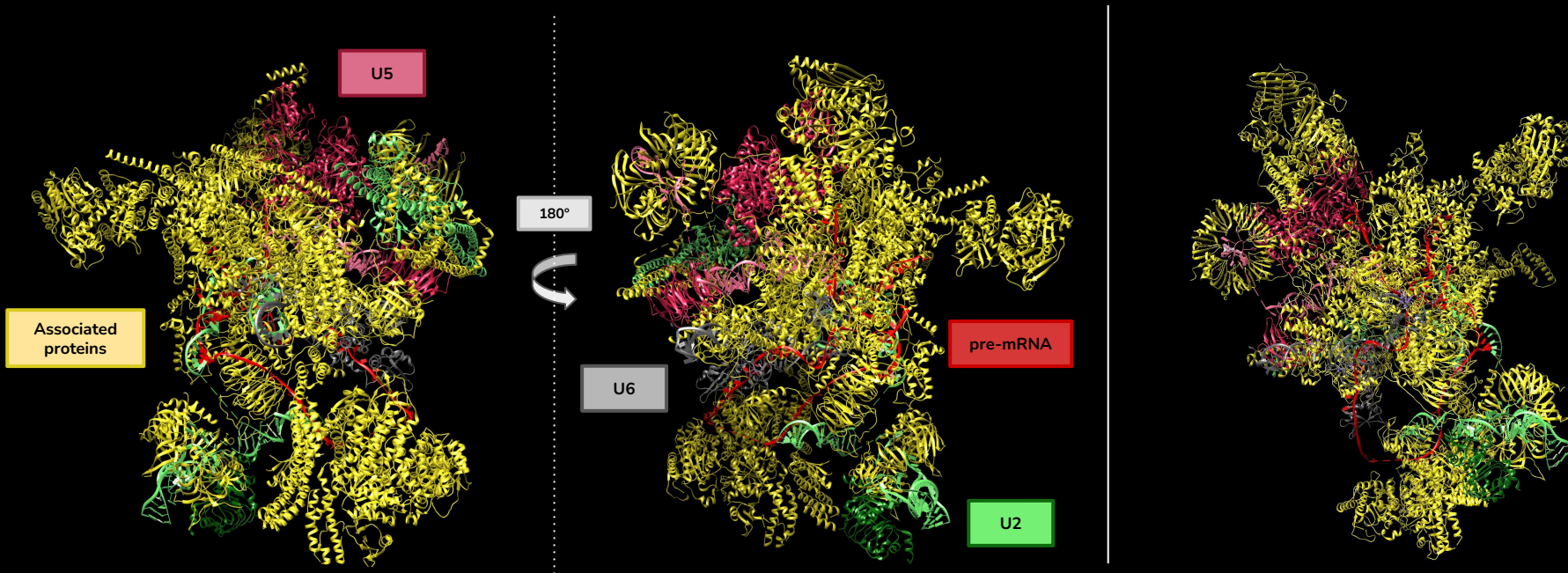
Cryo-EM structure of a human activated spliceosome (early B^{act}) (5Z58): *Homo sapiens*, 4.90 Å

Complex B*



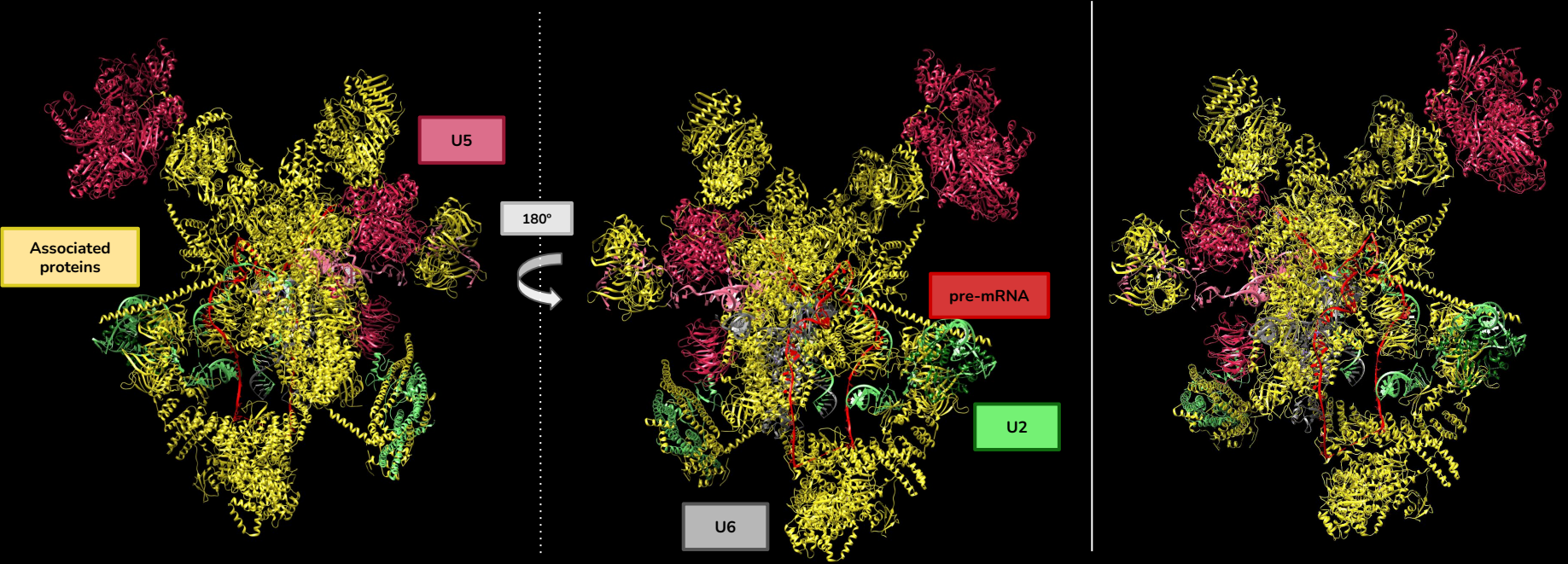
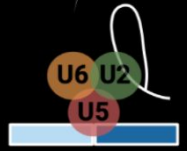
Cryo-EM structure of a human activated spliceosome (mature Bact) (5Z56): *Homo sapiens*, 5.10Å

Complex C



The cryo-EM structure of human pre-C*-I complex (7W59): *Homo sapiens*, 3.60Å

Post spliceosomal complex



Cryo-EM structure of a human post-catalytic spliceosome (P complex) (6ICZ): *Homo sapiens*, 3.00Å

Evolutionary conservation (U1 snRNP)

Consensus Conservation	1 MTQILPPNLL	11 ALFAPRdPI	21 YLPPLeKLPH	31 EKhhnqpYcG	41 IAPYireFED	51 PrDaPpPTr	61 ETREERmERk	71 RREKIerrqq	81 eVeTEIkMWd	91 PhnDpNAqGD
U1snRNP_HUMAN	MTQILPPNLL	ALFAPRdPI	YLPPLeKLPH	EKHNNQPYCG	IAPYIREFED	PRDAPPPTRA	ETREERMERK	RREKIERRQQ	EVETEIKMWD	PHNDPNAQGD
U1snRNP_MOUSE	MTQFLPPNLL	ALFAPRdPI	YLPPLeKLPH	EKHNNQPYCG	IAPYIREFED	PRDAPPPTRA	ETREERMERK	RREKIERRQQ	EVETEIKMWD	PHNDPNAQGD
U1snRNP_BOVIN	MTQFLPPNLL	ALFAPRdPI	YLPPLeKLPH	EKHNNQPYCG	IAPYIREFED	PRDAPPPTRA	ETREERMERK	RREKIERRQQ	EVETEIKMWD	PHNDPNAQGD
U1snRNP_XENOP	MTQFLPPNLL	ALFAPRdPVP	YLPPLdKLPH	EKHNNQPYCG	IAPYIREFED	PRDAPPPTRA	ETREERMERK	RREKIERRQQ	DVELEKIDW	PHNDQNAQGD
U1snRNP_DROSO	MTQYLLPPNLL	ALFAAREPIL	FMPVPVKLPH	EKKK - RGYLG	VAKFMAEDF	KDPTLPKTV	ETRQERLERR	RREKAEQVAY	KLEIREIALWD	PTEIKNATED
U1snRNP_POMBE	MAEKLPAPLL	ALFAPRPPIL	YLPMPD - VPP	EKRSTPRVSG	IAKYLKVAQS	HDQQYHPT -	ESLEEKRLRL	ADKQKQGR	RLRSMIKVWD	PDHHRHVIDG
Consensus Conservation	51 PrDaPpPTr	61 ETREERmERk	71 RREKIerrqq	81 eVeTEIkMWd	91 PhnDpNAqGD	101 aFKTLFvArv	111 NYDTtESKLr	121 REFEvYGPiK	131 RIhmVysKrs	141 GKPrGYAFIE
U1snRNP_HUMAN	PRDAPPPTRA	ETREERMERK	RREKIERRQQ	EVETEIKMWD	PHNDPNAQGD	AFKTLFVARV	NYDTTESKLRL	REFEVYGPiK	RIHMVYSKRS	GKPRGYAFIE
U1snRNP_MOUSE	PRDAPPPTRA	ETREERMERK	RREKIERRQQ	EVETEIKMWD	PHNDPNAQGD	AFKTLFVARV	NYDTTESKLRL	REFEVYGPiK	RIHMVYSKRS	GKPRGYAFIE
U1snRNP_BOVIN	PRDAPPPTRA	ETREERMERK	RREKIERRQQ	EVETEIKMWD	PHNDPNAQGD	AFKTLFVARV	NYDTTESKLRL	REFEVYGPiK	RIHMVYSKRS	GKPRGYAFIE
U1snRNP_XENOP	PRDAPPPTRA	ETREERMERK	RREKIERRQQ	DVELEKIDW	PHNDQNAQGD	AFKTLFVARV	NYDTTESKLRL	REFEVYGPiK	RIHMVYNKRS	GKPRGYAFIE
U1snRNP_DROSO	ETRQERLERR	RREKAEQVAY	KLEIREIALWD	PTEIKNATED	PFRTLFIAIRI	NYDTTESKLRL	REFEVYGPiK	KIVLIHQDES	RIIVLVNKNVT	GKSMGYAFV
U1snRNP_POMBE	HDQQYHPT -	ESLEEKRLRL	RDEKQKQGR	RLRSMIKVWD	PDHHRHVIDG	PYKTMFLSRL	SYDTKESDIE	REFTRYGPiE	RIIVLVNKNVT	GKSMGYAFV
Consensus Conservation	101 aFKTLFvArv	111 NYDTtESKLr	121 REFEvYGPiK	131 RIhmVysKrs	141 GKPrGYAFIE	151 YEHERDMHsA	161 YKHADGKKID	171 GRRVLVDVER	181 GRTVKGWRPR	191 RLGGGLGGTR
U1snRNP_HUMAN	AFKTLFVARV	NYDTTESKLRL	REFEVYGPiK	RIHMVYSKRS	GKPRGYAFIE	YEHERDMHSA	YKHADGKKID	GRRVLVDVER	GRTVKGWRPR	RLGGGLGGTR
U1snRNP_MOUSE	AFKTLFVARV	NYDTTESKLRL	REFEVYGPiK	RIHMVYSKRS	GKPRGYAFIE	YEHERDMHSA	YKHADGKKID	GRRVLVDVER	GRTVKGWRPR	RLGGGLGGTR
U1snRNP_BOVIN	AFKTLFVARV	NYDTTESKLRL	REFEVYGPiK	RIHMVYSKRS	GKPRGYAFIE	YEHERDMHSA	YKHADGKKID	GRRVLVDVER	GRTVKGWRPR	RLGGGLGGTR
U1snRNP_XENOP	AFKTLFVARV	NYDTTESKLRL	REFEVYGPiK	RIHMVYNKRS	GKPRGYAFIE	YEHERDMHSA	YKHADGKKID	GRRVLVDVER	GRTVKGWRPR	RLGGGLGGTR
U1snRNP_DROSO	PFRTLFIAIRI	NYDTTESKLRL	REFEVYGPiK	KIVLIHQDES	YEHERDMHSA	YKHADGKKID	YKASAGLMLN	SKRVLVDVER	ARTVKGWRPR	RLGGGLGGTR
U1snRNP_POMBE	PYKTMFLSRL	SYDTKESDIE	REFTRYGPiE	RIIVLVNKNVT	GKSMGYAFV	FERERDLKVA	YKASAGLMLN	GRRIVVDVER	GRTVKGWRPR	KLGGGLGG -
Consensus Conservation	151 YEHERDMHsA	161 YKHADGKKID	171 GRRVLVDVER	181 GRTVKGWRPR	191 RLGGGLGGTR	201 RGGaDVNIrH	211 SGRdDtsRyd	221 ER - - - - - ph	231 rdRdRdrERE	241 RRERsrerdK
U1snRNP_HUMAN	YEHERDMHSA	YKHADGKKID	GRRVLVDVER	GRTVKGWRPR	RLGGGLGGTR	RGGADVNIrH	SGRDDTSRYD	ERPGPSPiLPH	RDRDRDRERE	RRERSRERDK
U1snRNP_MOUSE	YEHERDMHSA	YKHADGKKID	GRRVLVDVER	GRTVKGWRPR	RLGGGLGGTR	RGGADVNIrH	SGRDDTSRYD	ERPGPSPiLPH	RDRDRDRERE	RRERSRERDK
U1snRNP_BOVIN	YEHERDMHSA	YKHADGKKID	GRRVLVDVER	GRTVKGWRPR	RLGGGLGGTR	RGGADVNIrH	SGRDDTSRYD	ERPGPSPiLPH	RDRDRDRERE	RRERSRERDK
U1snRNP_XENOP	YEHERDMHSA	YKHADGKKID	GRRVLVDVER	GRTVKGWRPR	RLGGGLGGTR	RGGADVNIrH	SGRDDTSRYD	ERPGPSPiLPH	- DRDRERERD	RRERSRERDK
U1snRNP_DROSO	YEHERDMHSA	YKHADGKKID	SKRVLVDVER	ARTVKGWRPR	RLGGGLGGTR	RGGADVNIrH	SGREDNERER	ER - - - - - YR	LERERE - DR	GPGRGGSGNG
U1snRNP_POMBE	FERERDLKVA	YKASAGLMLN	GRRIVVDVER	GRTVKGWRPR	KLGGGLGG -	- - - - - RH	YTK - - - - -	- - - - -	- - - - - ERP	RRERGSRFRG
Consensus Conservation	201 RGGaDVNIrH	211 SGRdDtsRyd	221 ER - - - - - ph	231 rdRdRdrERE	241 RRERsrerdK	251 ereRRrsrsr	261 drRRRSRSrD	271 keeRrRsReR	281 skdkdrDrkr	291 rssRsRerar
U1snRNP_HUMAN	RGGADVNIrH	SGRDDTSRYD	ERPGPSPiLPH	RDRDRDRERE	RRERSRERDK	ERERRRSRSR	DRRRRSRSRD	KEERRRSRER	SKDKDRDRKR	RSSRSRERAR
U1snRNP_MOUSE	RGGADVNIrH	SGRDDTSRYD	ERPGPSPiLPH	RDRDRDRERE	RRERSRERDK	ERERRRSRSR	DRRRRSRSRD	KEERRRSRER	SKDKDRDRKR	RSSRSRERAR
U1snRNP_BOVIN	RGGADVNIrH	SGRDDTSRYD	ERPGPSPiLPH	RDRDRDRERE	RRERSRERDK	ERERRRSRSR	DRRRRSRSRD	KEERRRSRER	SKDKDRDRKR	RSSRSRERAR
U1snRNP_XENOP	RGGADVNIrH	SGRDDTSRYD	ERPGPSPiLPH	- DRDRERERD	RRERSRERDK	ERERRRSRSR	DRRRRSRSRD	KEERRRSRER	SKDKDKDKDK	DKDKDKDKDK
U1snRNP_DROSO	RGGADVNIrH	SGREDNERER	ER - - - - - YR	LERERE - DR	GPGRGGSGNG	LDAPPRFR	AEERRRSRSE	GAVASNSRSR	SSSRERKRKR	SSSRERKRKR
U1snRNP_POMBE	- - - - - RH	YTK - - - - -	- - - - -	- - - - -	RRERGSRFRG	DSGFRGGYRG	FRFKSSGGGS	RFGRGPTRSS	HSSDYGGGRDS	SPKRRRYN -

U1 small nuclear ribonucleoprotein 70 kDa:

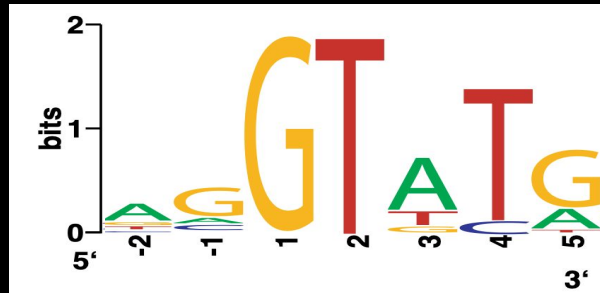
a. *Homo sapiens*, 448 aa (P08621)
b. *Mus musculus*, 448 aa (Q62376)

c. *Bos taurus*, 439 aa (Q1RMR2)
d. *Xenopus tropicalis*, 471 aa (Q66I18)

e. *Drosophila melanogaster*, 448 aa (P17133)
f. *Schizosaccharomyces pombe*, 261 aa (O13829)

Evolutionary conservation (U1 snRNA)

5' splice site

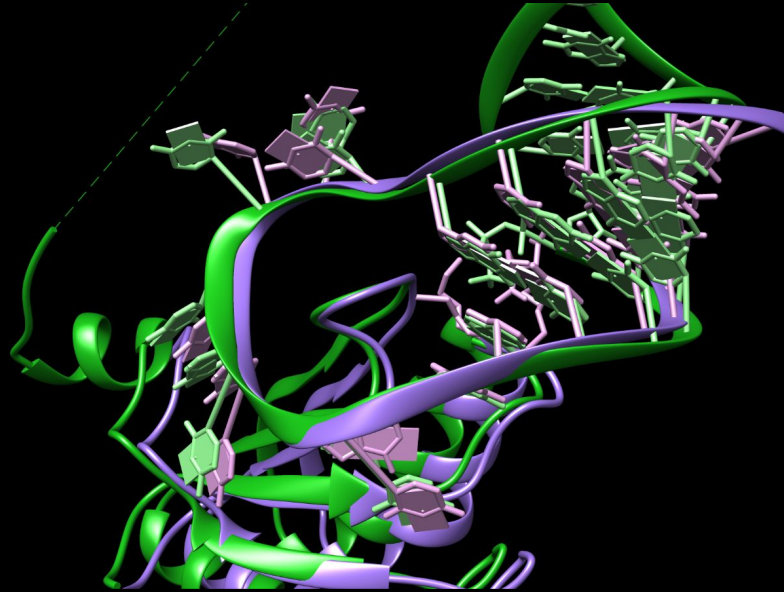


	1	11	21
Conservation			
U1snRNA_HUMAN	- - - - G G A U C	C A U U G C A C U C	C G G A U C C A G G
U1snRNA_DROSO	- - - - G G C C G	C A U U G C A C C U	C G C G G C C - - -
U1snRNA_SCERE	G C C A U U U U A G	U U U U U C A C G G	C A G A U U C G A A
	31	41	51
Conservation			
U1snRNA_HUMAN	A G A U A C C A U G	A U C A C G A A G G	U G G U U U U C C U
U1snRNA_DROSO	- - - - - - - -	- - - - - - - -	- - - - - - - -
U1snRNA_SCERE	U G A A C U U A A G	U U U A U G A U G A	A G G U A - - - - -

Adapted from: Claudio H. and Patrick J (2006)

U1-70k in complex with U1 snRNA stem-loops 1 and U1-A RRM (4PKD): *Homo sapiens*, 2.50Å
 Crystal structure of SNF/U1-SL2 complex (6F4H): *Drosophila melanogaster*, 2.00Å
S. cerevisiae U1 snRNP (6N7X): *Saccharomyces cerevisiae* S288C, 3.60Å

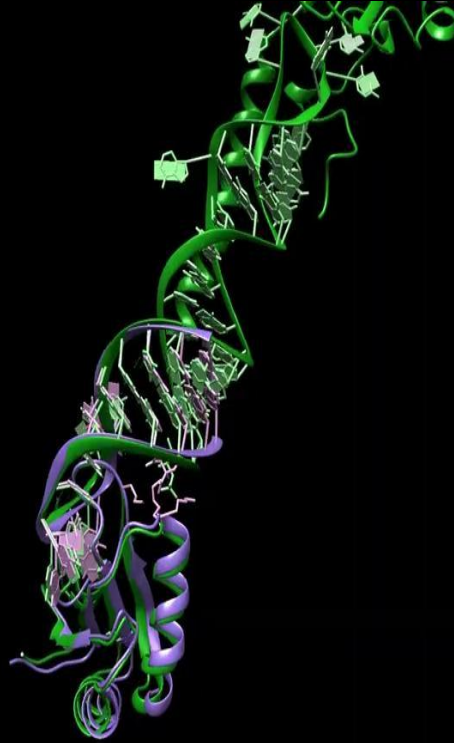
Evolutionary conservation (U1 snRNA)



Domain1	Domain2	Sc	RMS	Len1	Len2	Align	NFit
4pkd	6f4h	3.54	0.45	240	90	89	89

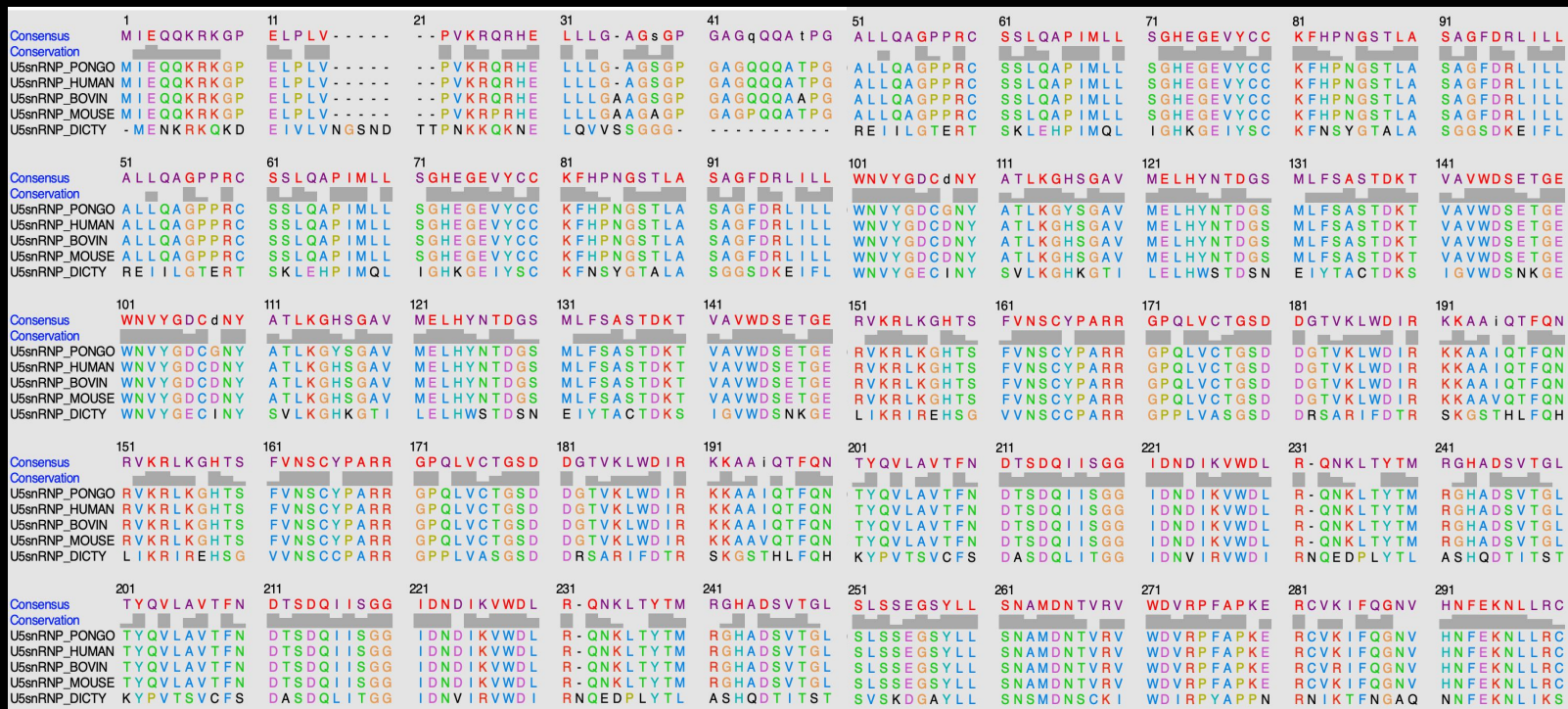
U1-70k in complex with U1 snRNA stem-loops 1 and U1-A RRM in complex with stem-loop 2 (4PKD): *Homo sapiens*, 2.50Å
Crystal structure of SNF/U1-SL2 complex (6F4H): *Drosophila melanogaster*, 2.00Å

Evolutionary conservation (U1 snRNA)



U1-70k in complex with U1 snRNA stem-loops 1 and U1-A RRM in complex with stem-loop 2 (4PKD): *Homo sapiens*, 2.50Å
Crystal structure of SNF/U1-SL2 complex (6F4H): *Drosophila melanogaster*, 2.00Å

Evolutionary conservation (U5 snRNP)



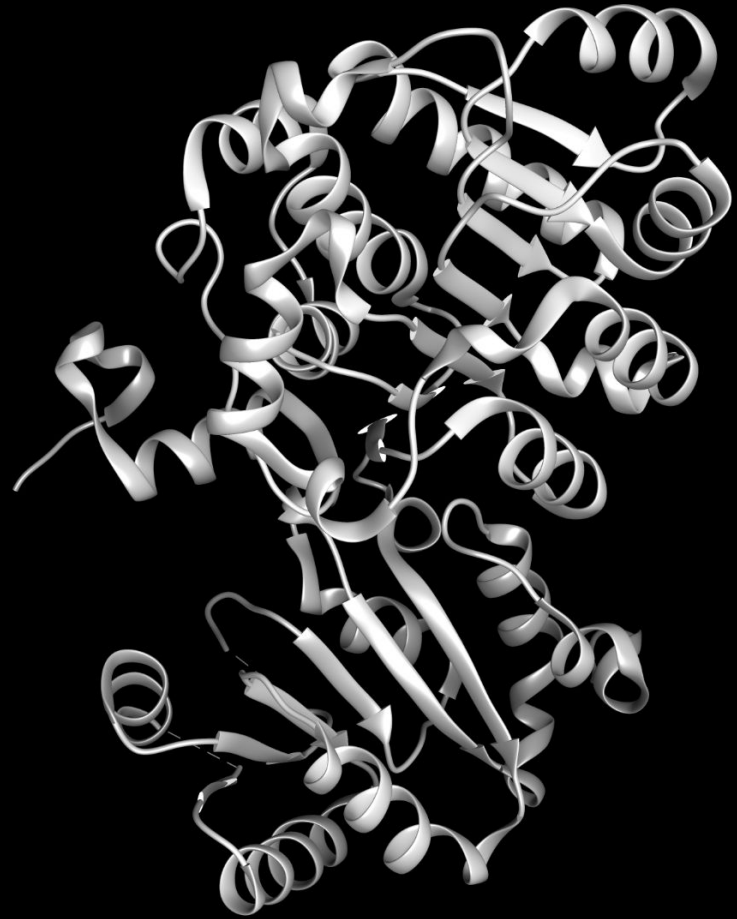
U5 small nuclear ribonucleoprotein 40 kDa:

a. *Pongo abelii*, 357 aa (Q5RF51)
b. *Homo sapiens*, 357 aa (Q96DI7)

c. *Bos taurus*, 358 aa (Q2HJH6)
d. *Mus musculus*, 358 aa (Q6PE01)

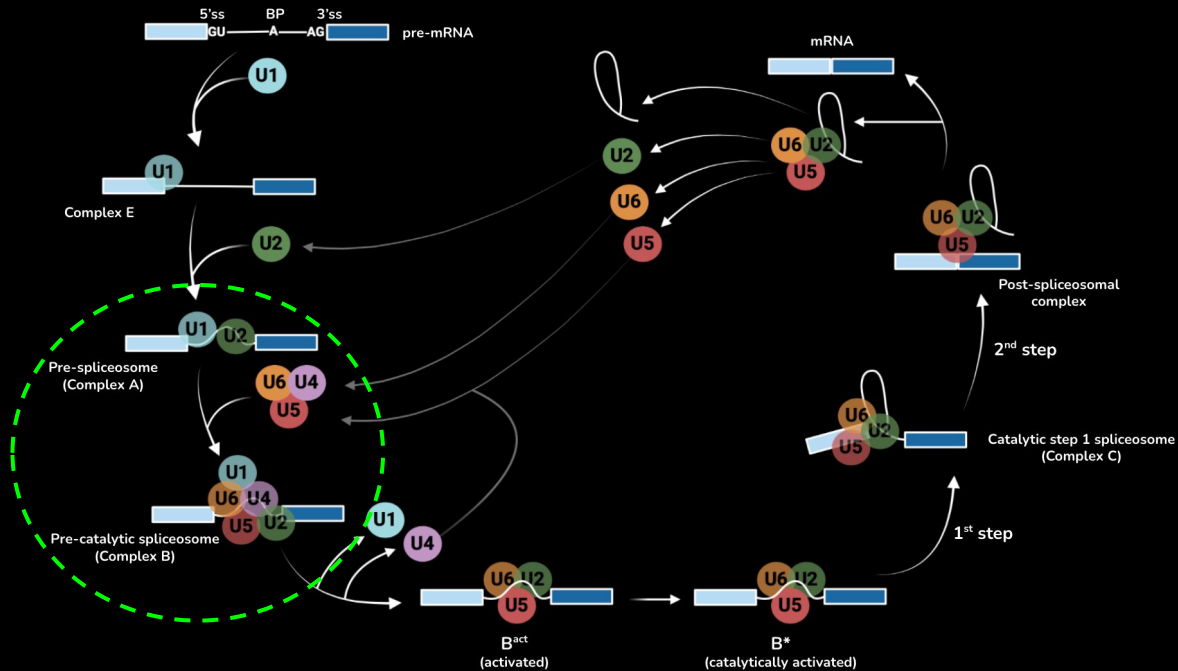
e. *Dictyostelium discoideum*, 355 aa (Q55AR8)

Pre-mRNA splicing factor 28 (Prp28)



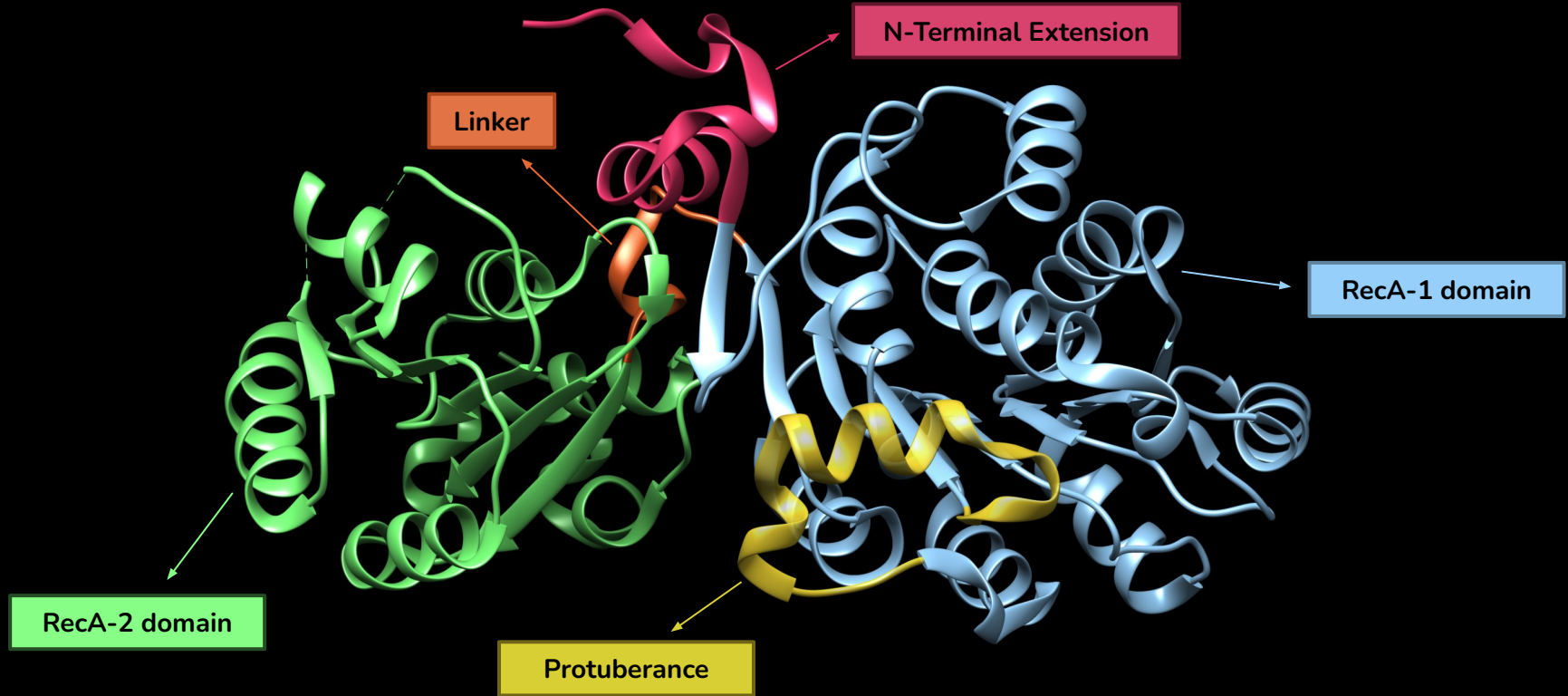
Structure of the spliceosomal DEAD-box protein Prp28
(4NHO): *Homo sapiens*, 2.00Å

Prp28 role



Structure of a human fully-assembled precatalytic spliceosome (pre-B complex) (6QX9): *Homo sapiens*, 3.28Å

Prp28 structure



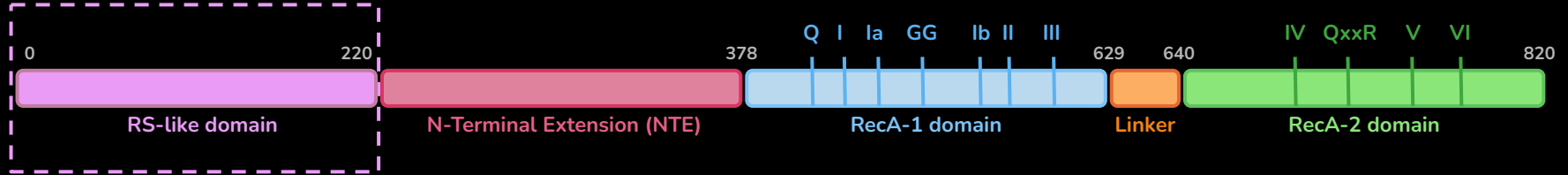
Structure of the spliceosomal DEAD-box protein Prp28 (4NHO): *Homo sapiens*, 2.00Å

Prp28 structure



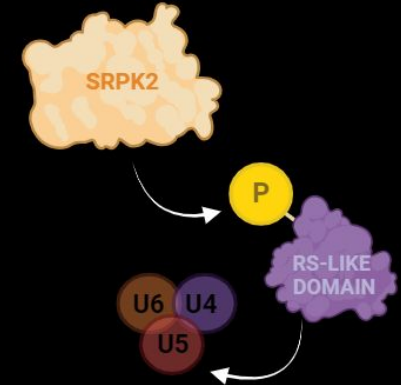
Structure of the spliceosomal DEAD-box protein Prp28 (4NHO): *Homo sapiens*, 2.00Å

RS-like domain



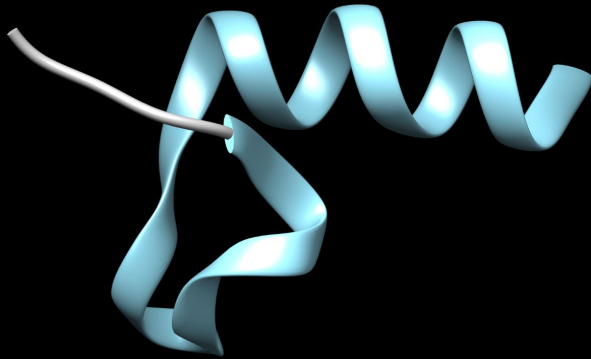
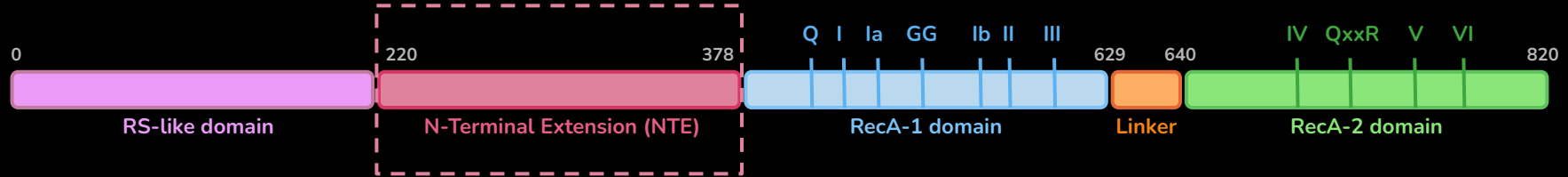
- All attempts to crystallize full-length hPrp28 = Failed
- RS-Like domain predicted to be intrinsically disordered
- Present in HIGHER eukaryotes

Rich in ARGININE/SERINE as well as ARGININE/GLUTAMATE and ARGININE/ASPARTATE



Mathew R, Hartmuth K, Möhlmann S, Urlaub H, Ficner R, Lührmann R. Phosphorylation of human PRP28 by SRPK2 is required for integration of the U4/U6-U5 tri-snRNP into the spliceosome. Nat Struct Mol Biol. 2008;15(5):435–43.

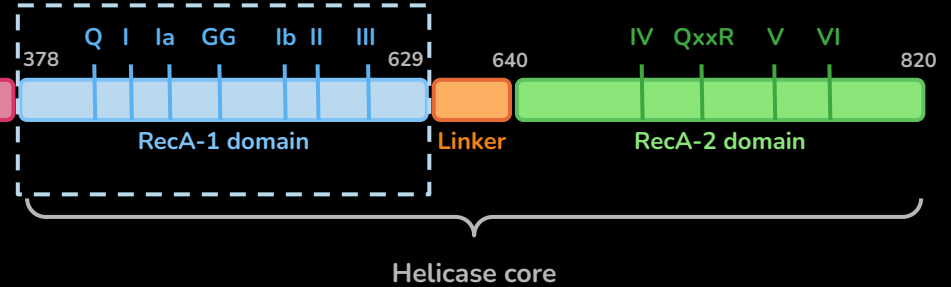
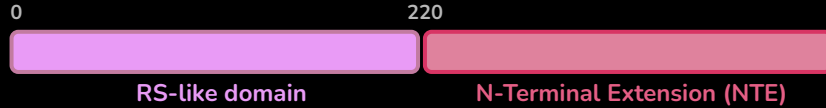
N-Terminal Extension (NTE)



- Residues 367-377 form an alpha-helix which packs against RecA-1 domain.
- The further N-terminal residues form an irregular loop including two short turns

Structure of the spliceosomal DEAD-box
protein Prp28 (4NHO): *Homo sapiens*, 2.00Å

RecA-1 domain

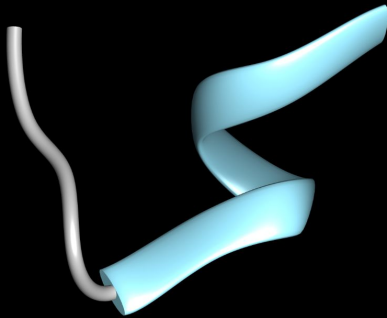
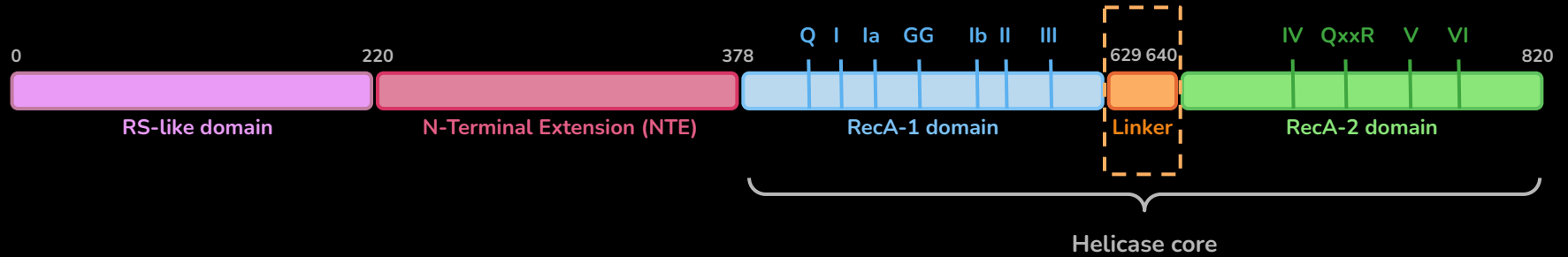


Structure of the spliceosomal DEAD-box protein Prp28 (4NHO): *Homo sapiens*, 2.00Å

- Eight stranded beta-sheet comprising seven parallel strands and one antiparallel strand
- Ten alpha helices pack against this sheet on both sides

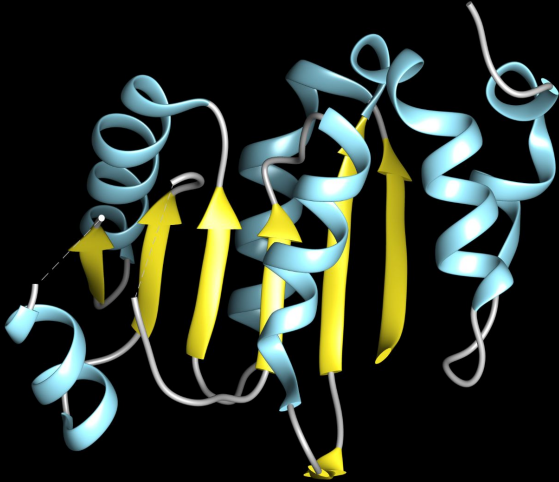
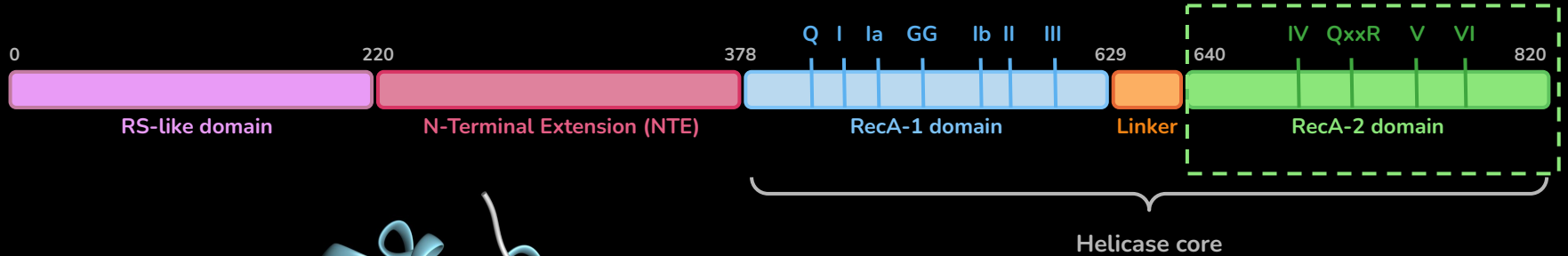
Insertion of 24 aa (576-599) which form a protuberance that consists of an alpha helix within an extended loop

Linker



Structure of the spliceosomal DEAD-box protein Prp28 (4NHO): *Homo sapiens*, 2.00Å

RecA-2 domain



Structure of the spliceosomal DEAD-box protein Prp28 (4NHO): *Homo sapiens*, 2.00Å

- Six stranded parallel beta-sheet
- Five alpha helices pack against this sheet

SCOPe Classification

Lineage for Protein: RecA protein, ATPase-domain

1. Root: SCOPe 2.08
2.  Class c: Alpha and beta proteins (a/b) [51349] (148 folds)
3.  Fold c.37: P-loop containing nucleoside triphosphate hydrolases [52539] (1 superfamily)
3 layers: a/b/a, parallel or mixed beta-sheets of variable sizes
4.  Superfamily c.37.1: P-loop containing nucleoside triphosphate hydrolases [52540] (27 families) *S*
division into families based on beta-sheet topologies
5.  Family c.37.1.11: RecA protein-like (ATPase-domain) [52670] (22 proteins)
core: mixed beta-sheet of 8 strands, order 32451678; strand 7 is antiparallel to the rest
6.  Protein RecA protein, ATPase-domain [52671] (6 species)
C-terminal domain is alpha+beta

ClustalW MSA

- ATP binding and hydrolysis: Motifs **Q**, **I** (Walker A), **II** (Walker B), **V** and **VI**
- RNA binding: Motifs **Ia**, **Ib**, **IV** and **V**
- Coupling ATP hydrolysis to the RNA-unwinding activity: Motif **III**

Motif I contains the P-loop → sequence T/S-G-T/S-G-K-T.

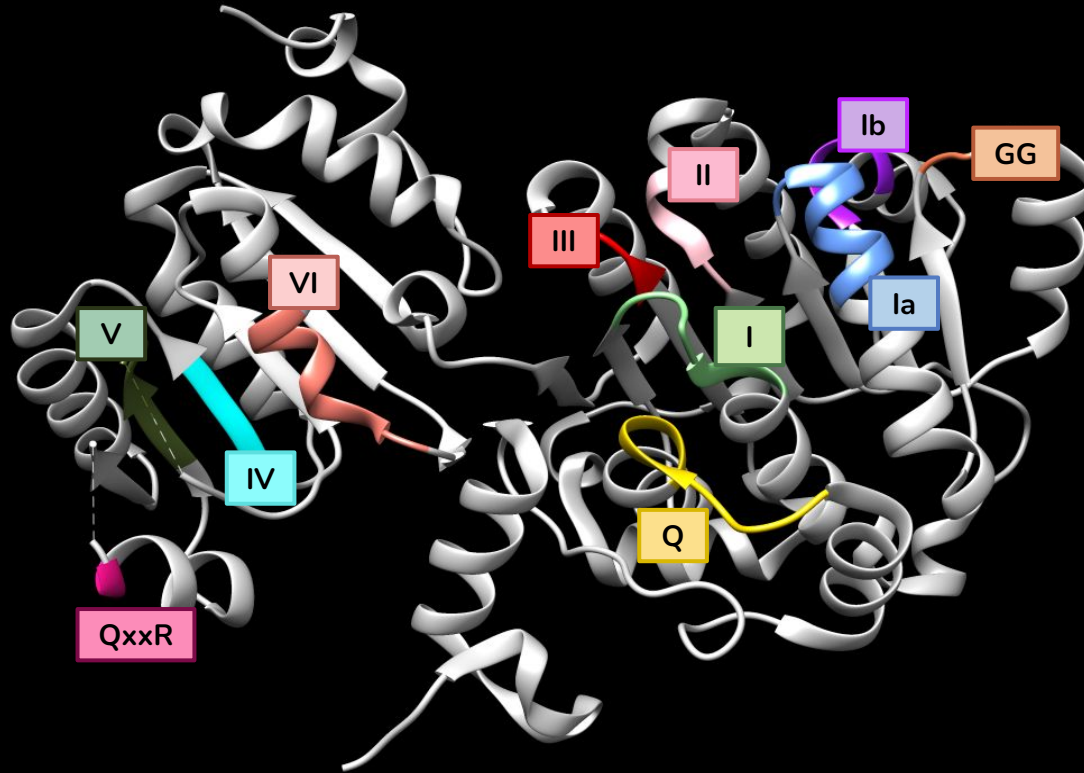
Motif II → sequence DExD/H

CLUSTAL 2.1 multiple sequence alignment



Structure of DEAD-box protein *Drosophila* Vasa (2DB3): *Drosophila melanogaster*, 2.20Å
 Structure of the spliceosomal DEAD-box protein Prp28 (4NHO): *Homo sapiens*, 2.00Å
 Crystal structure of the DEAD box protein Dhh1p (1S2M): *Saccharomyces cerevisiae*, 2.10Å

DExD/H-box protein motifs



Structure of the spliceosomal DEAD-box protein Prp28 (4NHO): *Homo sapiens*, 2.00Å

Prp28 homolog



Structure of the spliceosomal DEAD-box protein
Prp28 (4NHO): *Homo sapiens*, 2.00Å

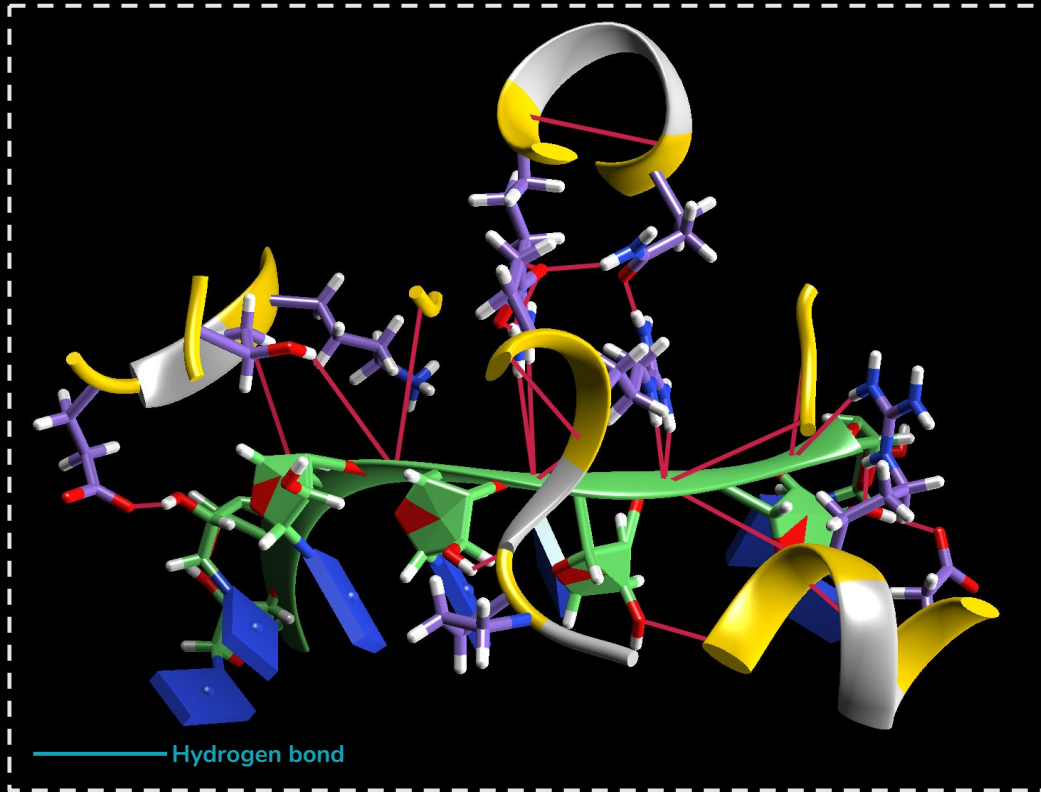


Structure of DEAD-box protein *Drosophila* Vasa (2DB3):
Drosophila melanogaster, 2.20Å



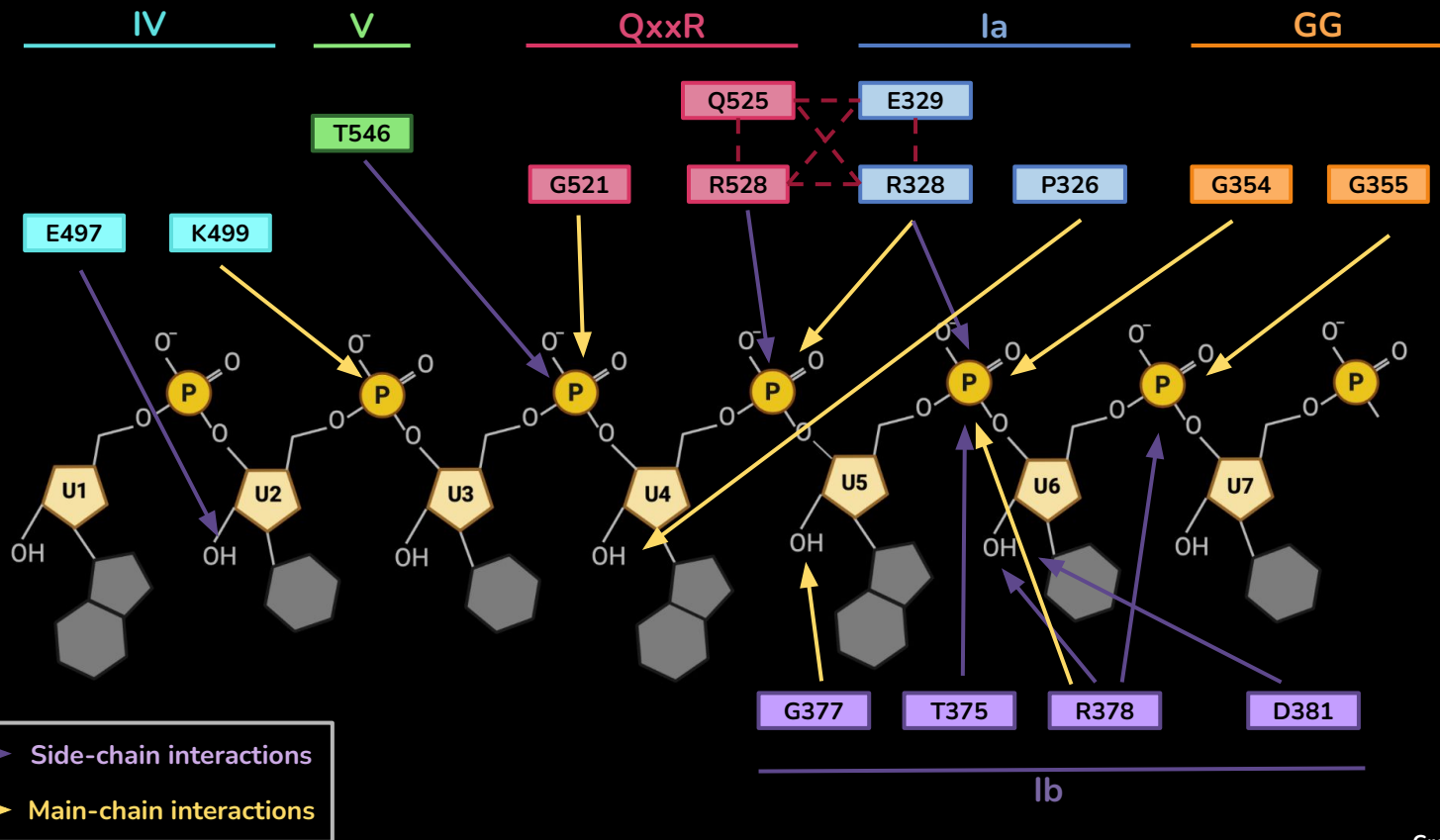
	No.	Domain1	Domain2	Sc	RMS
Pair	1	4nho	2db3	0.87	2.58

RNA binding by DExD/H-box helicases



Structure of DEAD-box protein Drosophila Vasa (2DB3):
Drosophila melanogaster, 2.20Å

RNA binding by DExD/H-box helicases



RNA binding by DExD/H-box helicases

451 Prp28_HUMAN Vasa_DROSO DHH1_SCERE	461 DRIEESDQGP LELGR - - - - P NKI - - - - -	471 YAILALPTRE QVVIVSP TRE QALIMVPTRE 326 328 329	481 LAQQIEEETI LAIQIFNEAR LALQTSQVVR	491 KFGKPLGIRT KFAFESYLKI TLGKHCGISC
501 Prp28_HUMAN Vasa_DROSO DHH1_SCERE	511 DQGFRLRMGC HQNECITRGC PDILRLNETV	521 EIVIAITPGRRL HVVIAITPGRRL HILVGTTPGRV 375 377 378	531 IDVLENRYLV LDFVDRTFIT LDLASRKVAD 381	541 LSRCTYVVLD FEDTRFVLD LSDCSLFIMD
551 Prp28_HUMAN Vasa_DROSO DHH1_SCERE	561 EPDVQKILEH SEDMRRIMTH KTIIEQILSF	571 MPVSNQKPD VTMRPEH - - - LP - - - - -	581 DEAEDPEKML - - - - - - - - - -	591 ANFESGKHKY - - - - - - - - - - P T
601 Prp28_HUMAN Vasa_DROSO DHH1_SCERE	611 PPAVERLARS PEEIQRMAGE PLTVKEFMVK	621 YLRRPAVVYI FLKNYVFVAI HLHKPYEINL	631 GSAGKPHERV GIVGGACSDV -MEELTLKGI	641 EQKVFLMS KQTIYEVNKY TQYYAFVEER
651 Prp28_HUMAN Vasa_DROSO DHH1_SCERE	661 EQGFDPPIII SEQADG - TIV SKLQINQAI	671 FVNQKKGCDV FVETKRGADF FCNSTNRVEL 497 499	681 LAKSLEKMGY LASFLSEKEF LAKKITDLGY	691 NACTLHGGKG PTTSIHGDRL SCYYSHA 521
701 Prp28_HUMAN Vasa_DROSO DHH1_SCERE	711 LKA GAKDILV FKNGSMKVLI FRQ GKVRTLV	721 ATDVAGRGID ATSVASRGLD CSDLLTRGID 546	731 IQDVSMVVNY IKNIKHVINY IQAVNVVINF	741 DMAKNIEDYI DMP SKIDDYV DFPKTAETYL

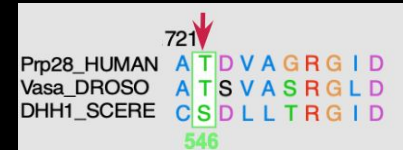
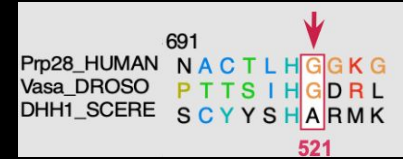
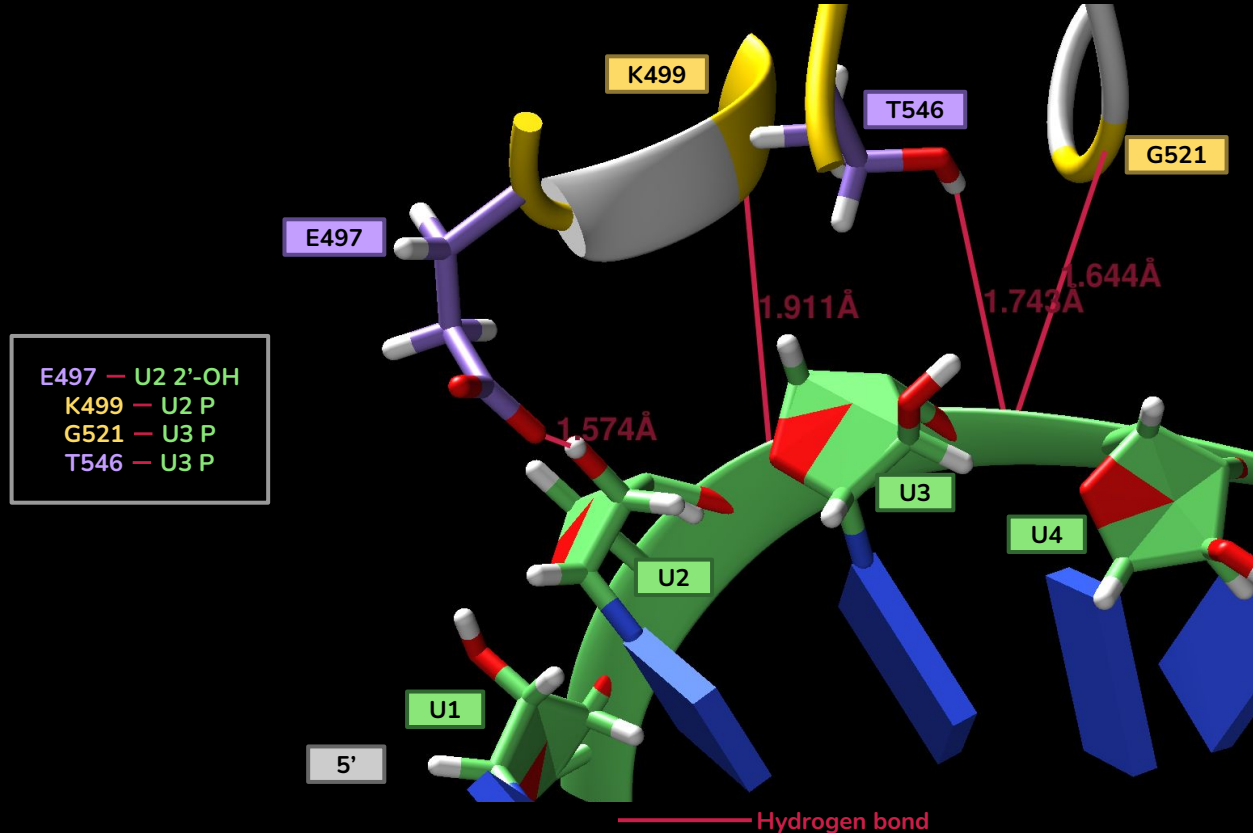
- Motif Ia
- Motif GG
- Motif Ib
- Motif IV
- Motif QxxR
- Motif V

Structure of DEAD-box protein *Drosophila* Vasa (2DB3): *Drosophila melanogaster*, 2.20Å

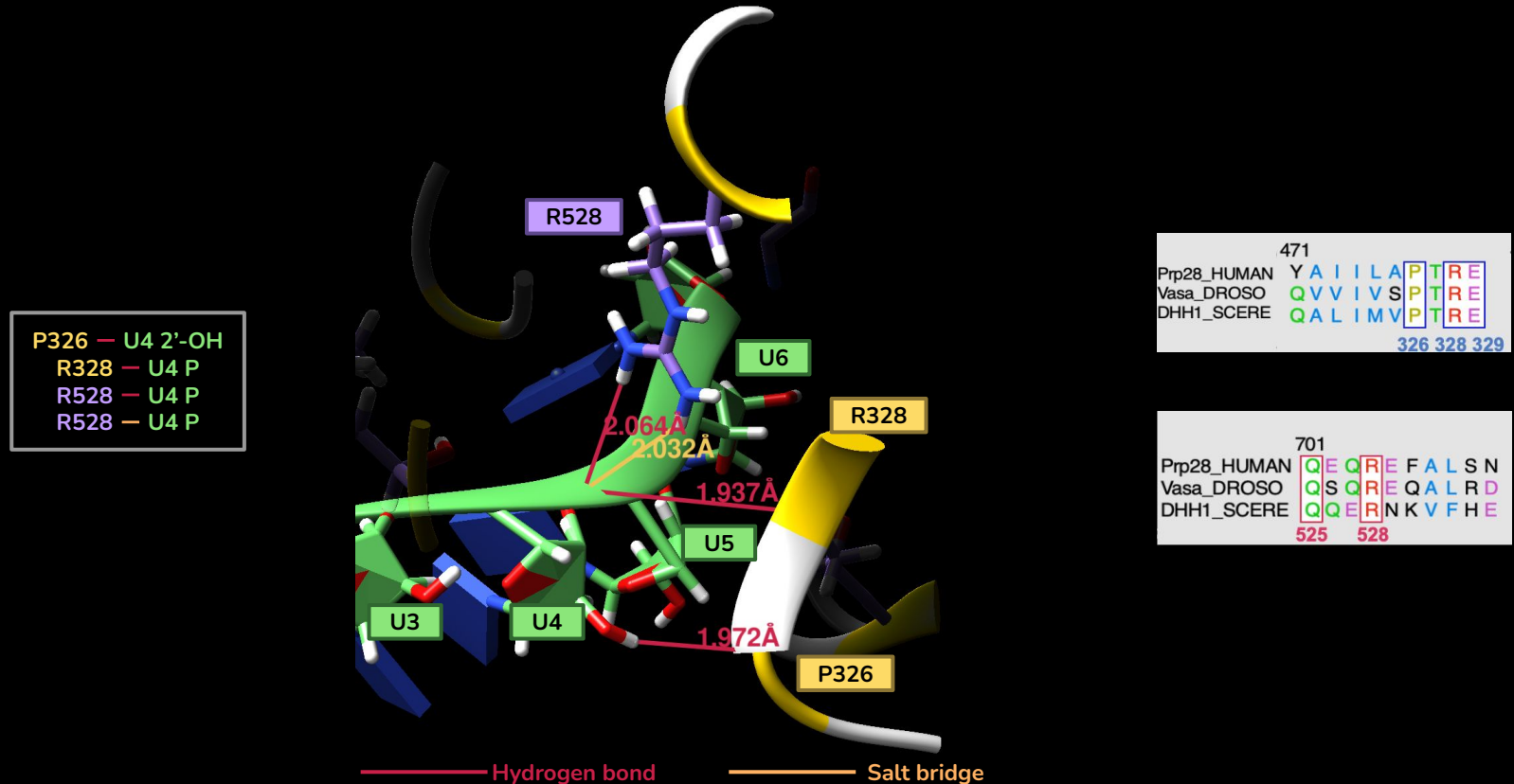
Structure of the spliceosomal DEAD-box protein Prp28 (4NHO): *Homo sapiens*, 2.00Å

Crystal structure of the DEAD box protein Dhh1p (1S2M): *Saccharomyces cerevisiae*, 2.10Å

RNA binding by DExD/H-box helicases

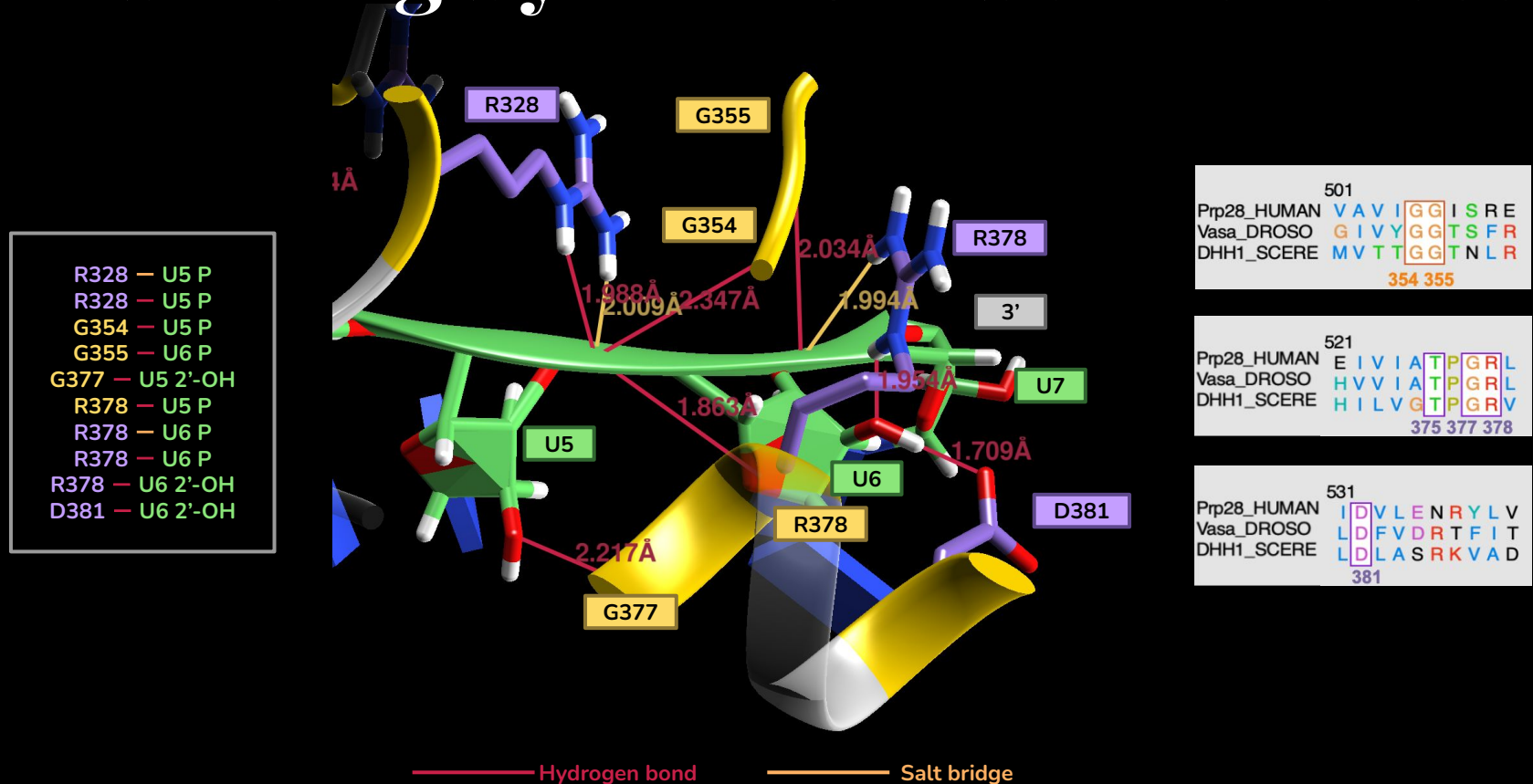


RNA binding by DExD/H-box helicases



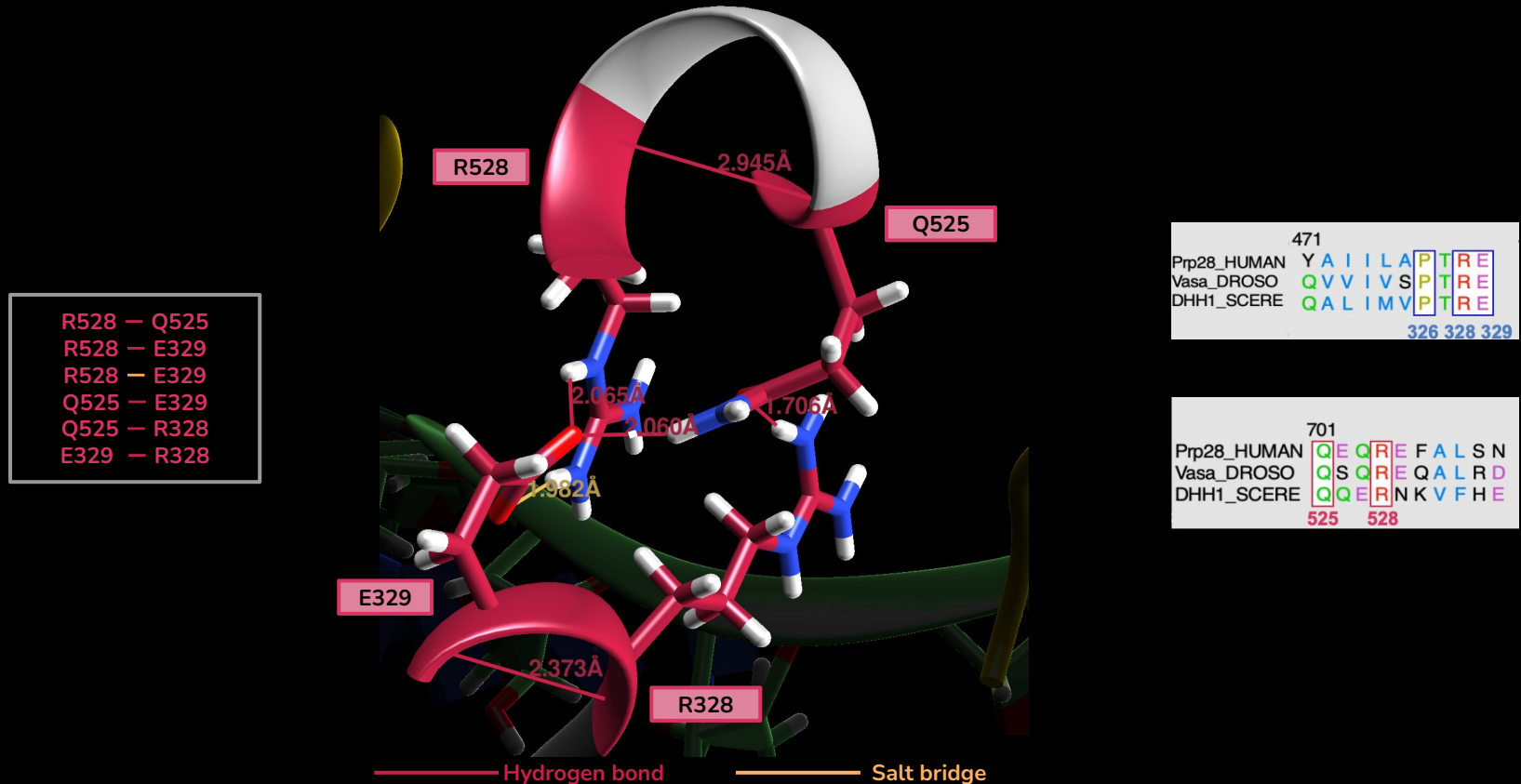
Structure of DEAD-box protein Drosophila Vasa (2DB3): *Drosophila melanogaster*, 2.20 Å

RNA binding by DExD/H-box helicases



Structure of DEAD-box protein Drosophila Vasa (2DB3): *Drosophila melanogaster*, 2.20Å

RNA binding by DExD/H-box helicases



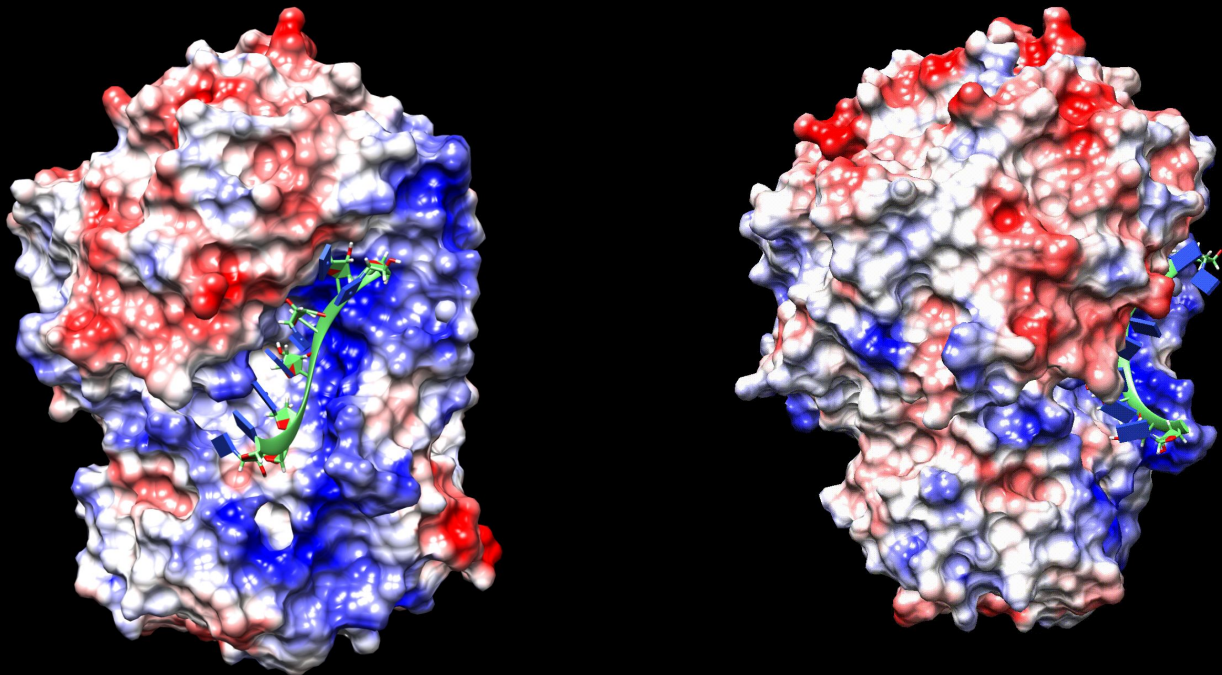
Structure of DEAD-box protein *Drosophila* Vasa (2DB3): *Drosophila melanogaster*, 2.20Å

RNA binding by DExD/H-box helicases

Electrostatic potential (Coulomb's Law)

-10 kCal/mol^e

+10 kCal/mol^e

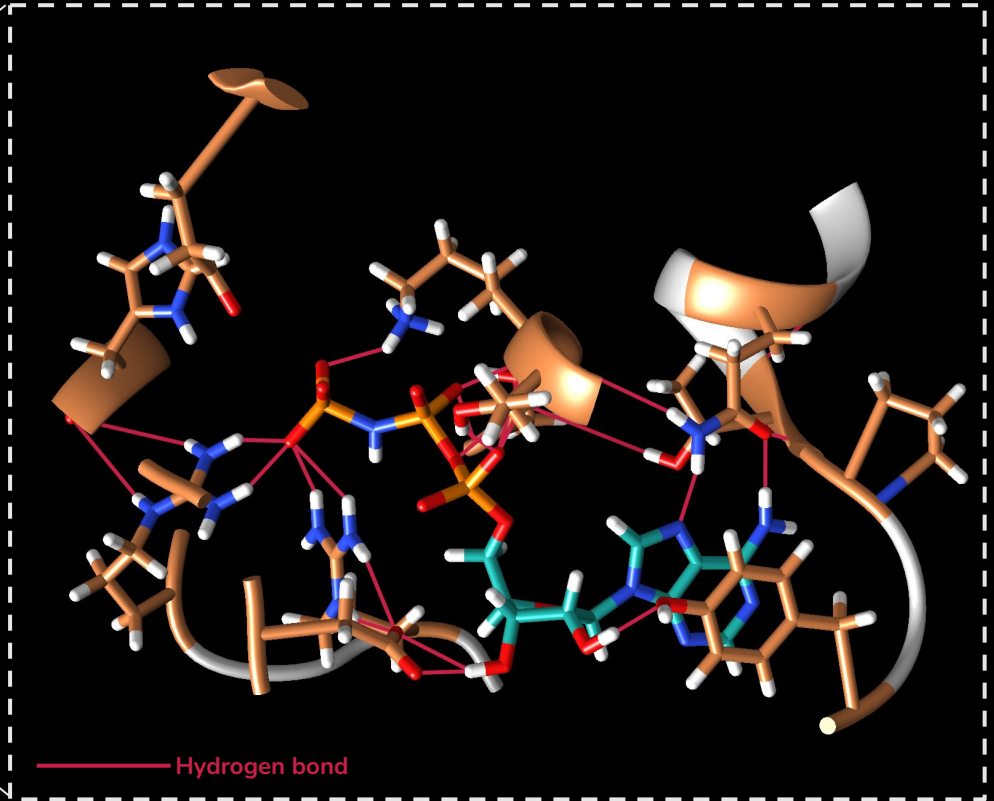


Structure of DEAD-box protein *Drosophila* Vasa (2DB3): *Drosophila melanogaster*, 2.20Å

ATP hydrolysis by DExD/H-box helicases



Structure of DEAD-box protein *Drosophila* Vasa (2DB3):
Drosophila melanogaster, 2.20Å



ATP hydrolysis by DExD/H-box helicases

401	411	421	431	441
Prp28_HUMAN	GYKEPTPIQR	QAIPIGLQNR	DIIGVAETGS	GKTA AFLIPL
Vasa_DROSO	GYKIPTPIQK	CSIPVIS SGR	DLMACAQTGS	GKTA AFLLP I
DHH1_SCERE	GFEKPSPIQE	EAIPVAITGR	DILARAKNGT	GKTA AFLV IPT
	265 268 272		293 294 296	
451	461	471	481	491
Prp28_HUMAN	DRIEESDQGP	YAIILAP TRE	LAQQIEEET I	KFGKPLGIRT
Vasa_DROSO	LSKLED PHE	QVVIVSPTRE	LAIQIFNEAR	KFAFESYLKI
DHH1_SCERE	LEKV - - KPKL	NKI - - - - -	LALQTSQVVR	TLGKHCGISC
501	511	521	531	541
Prp28_HUMAN	VAVIGG I SRE	EIV IATPGR L	IDVLE NRYLV	LSRCTYVVL D
Vasa_DROSO	GIVYGGTSFR	HVVIATPGR L	LDFVDRTFIT	FEDTRFVVL D
DHH1_SCERE	MVT TGGTNLR	HILVGT PGRV	LDLAS RKVAD	LSDCSLFIM D
651	661	671	681	691
Prp28_HUMAN	EKKRKLLA I L	FVNQKKGCDV	LAKSLEKMGY	NACTLHG GKG
Vasa_DROSO	AKRSKLIE I L	FVETKRGADF	LASFLEKEF	PTTSIHGDR L
DHH1_SCERE	QKLHCLNTLF	FCNSTNRVEL	LAKKITDLGY	SCYYSHARMK
701	711	721	731	741
Prp28_HUMAN	QEQREFALSN	ATDVAGRG I D	IQDVSMVVNY	DMAKNIEDY I
Vasa_DROSO	QSQREQALRD	ATSVASRG L D	IKNIKHVINY	DMP SKIDDYV
DHH1_SCERE	QQERNKVFHE	FRQ GKVRTLV	IQAVNVVINF	DFPKTAETYL
		552 554		
751	761	771	781	791
Prp28_HUMAN	KSGVAITFLT	KEDSAVFYEL	KQAILESPVS	SCPPELANHP
Vasa_DROSO	NNGRATSF FD	PEKDRAIAAD	LVKILEGSGQ	TVPDFLRTC G
DHH1_SCERE	HLGLA INLIN	WNRDFNLYKI	EQE - LGTEIA	AIPATIDKSL
	575 579 582			

Structure of DEAD-box protein *Drosophila* Vasa (2DB3): *Drosophila melanogaster*, 2.20Å

Structure of the spliceosomal DEAD-box protein Prp28 (4NHO): *Homo sapiens*, 2.00Å

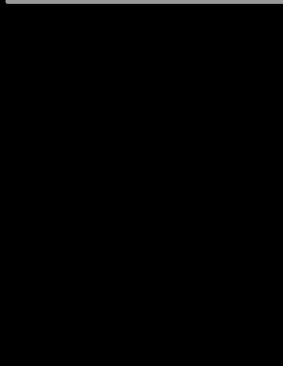
Crystal structure of the DEAD box protein Dhh1p (1S2M): *Saccharomyces cerevisiae*, 2.10Å

Y265 — 2'-OH
P268 — Adenine
Q272 — Adenine
D554 — 3'-OH

Hydrogen bond

411
Prp28_HUMAN G Y K E P T P I Q R
Vasa_DROSO G Y K I P T P I Q K
DHH1_SCERE G F E K P S P I Q E
265 268 272

721
Prp28_HUMAN A T D V A G R G I D
Vasa_DROSO A T S V A S R G L D
DHH1_SCERE C S D L L T R G I D
552 554

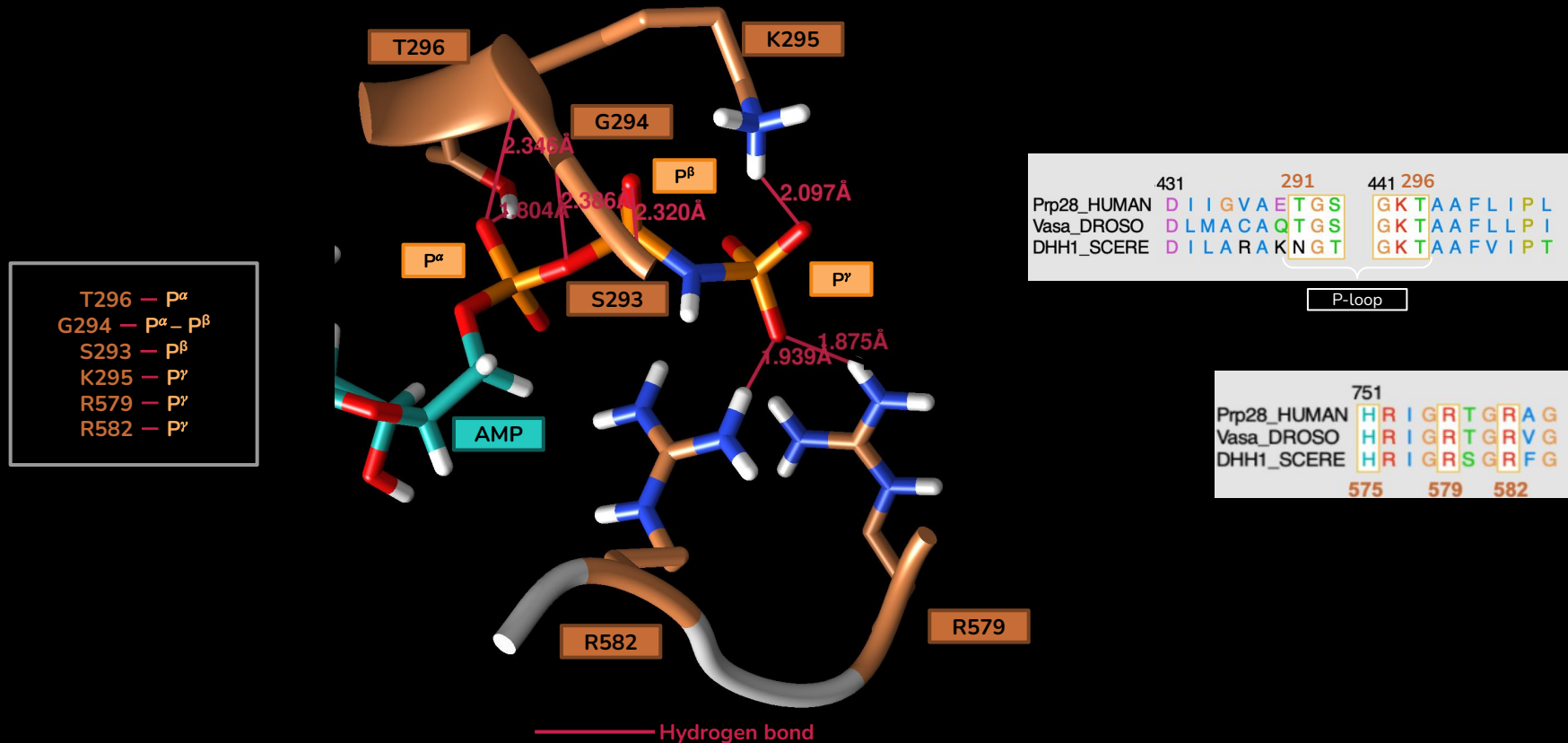


721

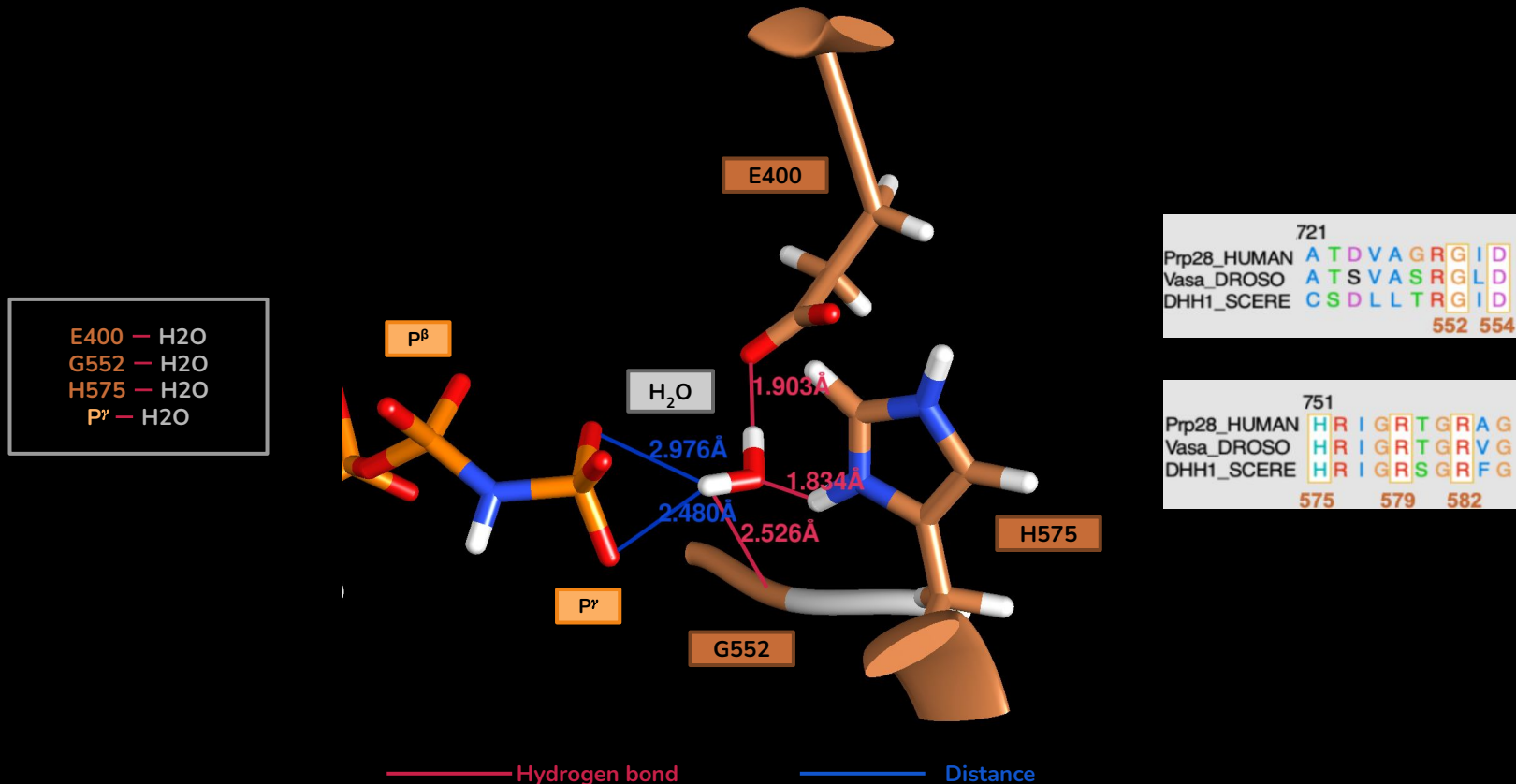
Prp28_HUMAN A T D V A G R G I D
Vasa_DROSO A T S V A S R G L D
DHH1_SCERE C S D L L T R G I D

552 554

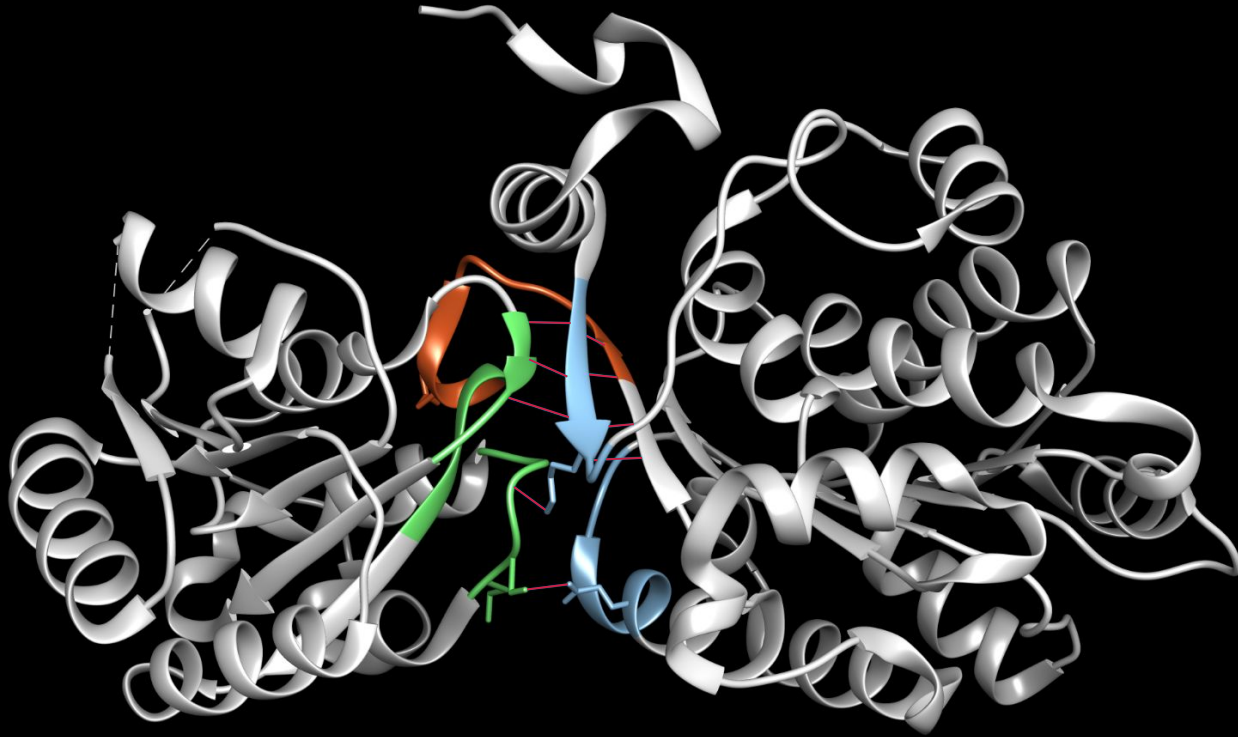
ATP hydrolysis by DExD/H-box helicases



ATP hydrolysis by DExD/H-box helicases

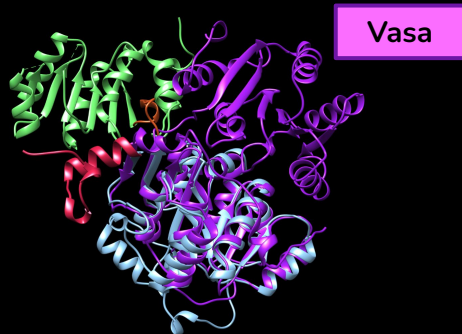


Helicase core conformation

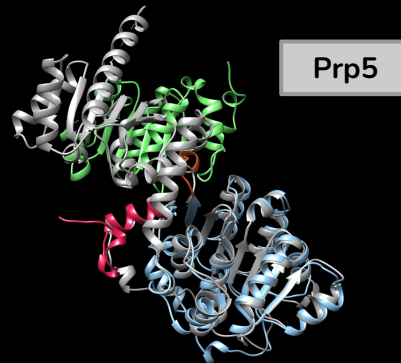


Structure of the spliceosomal DEAD-box protein Prp28 (4NHO): *Homo sapiens*, 2.00Å

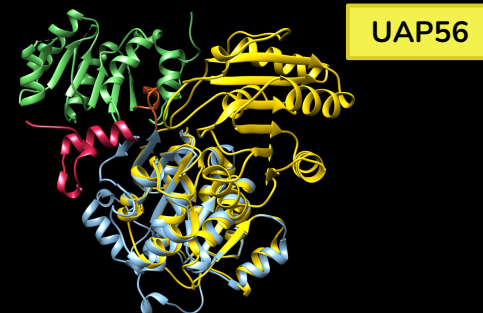
States



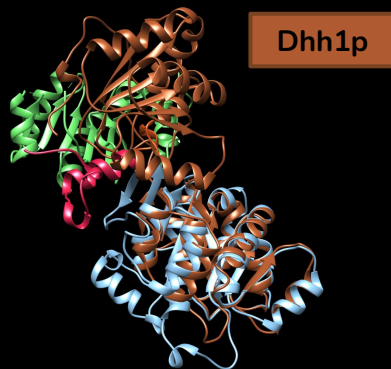
Structure of DEAD-box protein *Drosophila* Vasa (2DB3): *Drosophila melanogaster*, 2.20Å



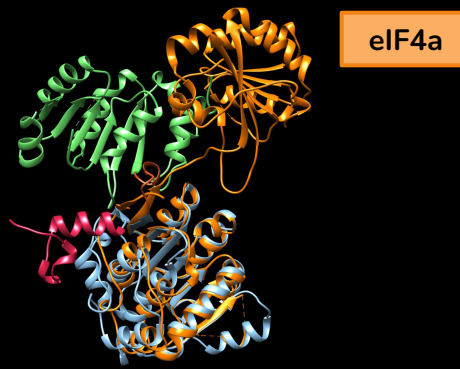
Crystal structure of RNA splicing effector Prp5 (4LK2): *Saccharomyces cerevisiae* S288C, 2.12Å



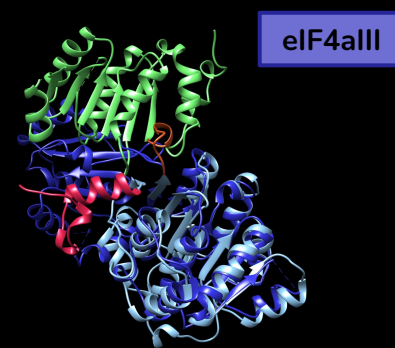
Crystal structure of the N-terminal domain of human UAP56 (1T6N): *Homo sapiens*, 1.94Å



Crystal structure of the DEAD box protein Dhh1p (1S2M): *Saccharomyces cerevisiae*, 2.10Å



Crystal structure of BLF1 in complex with human eIF4a (7PPZ): *Homo sapiens*, 2.52Å



Crystal structure of eIF4AIII-CWC22 complex (4C9B): *Homo sapiens*, 2.00Å

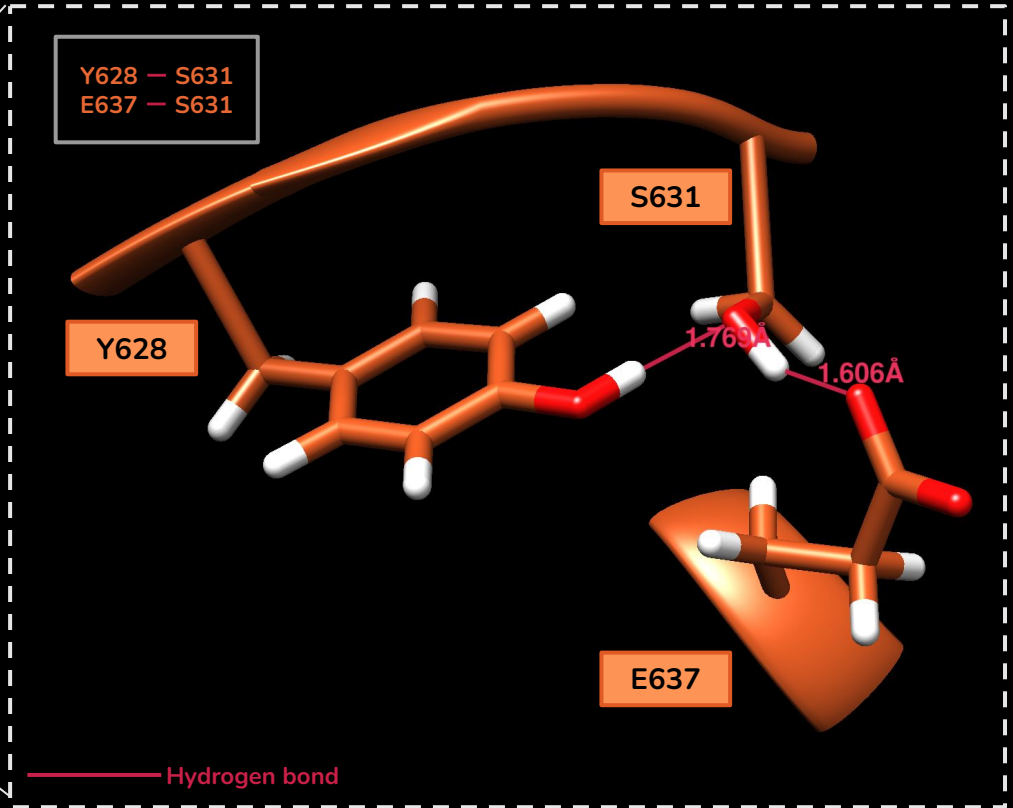
States



Linker

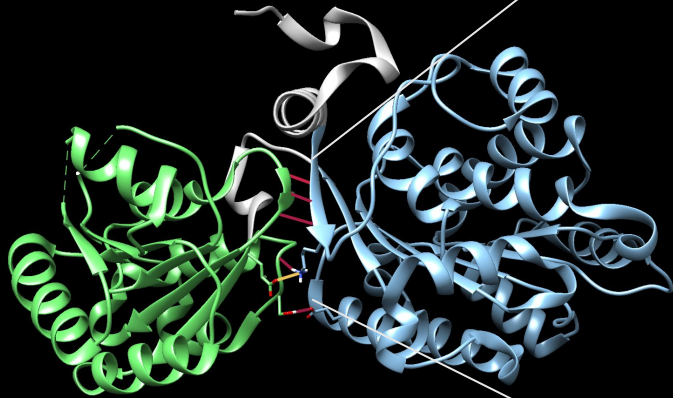


Structure of the spliceosomal DEAD-box protein Prp28 (4NHO): *Homo sapiens*, 2.00Å

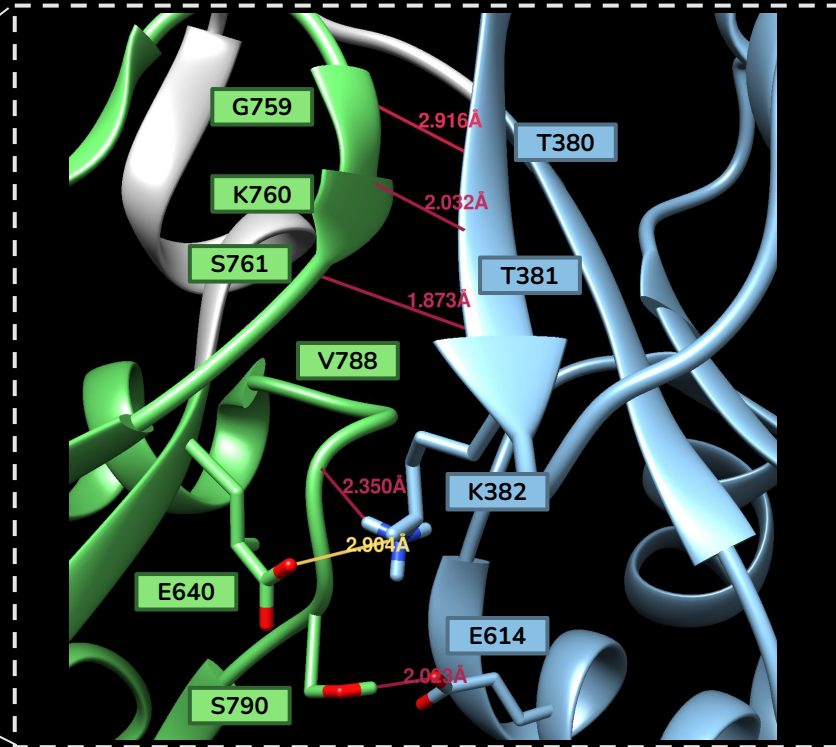


RecA-1 and RecA-2 union

G759	—	T380	E640	—	K382
K760	—	T381	V788	—	K382
S761	—	T381	S790	—	E614



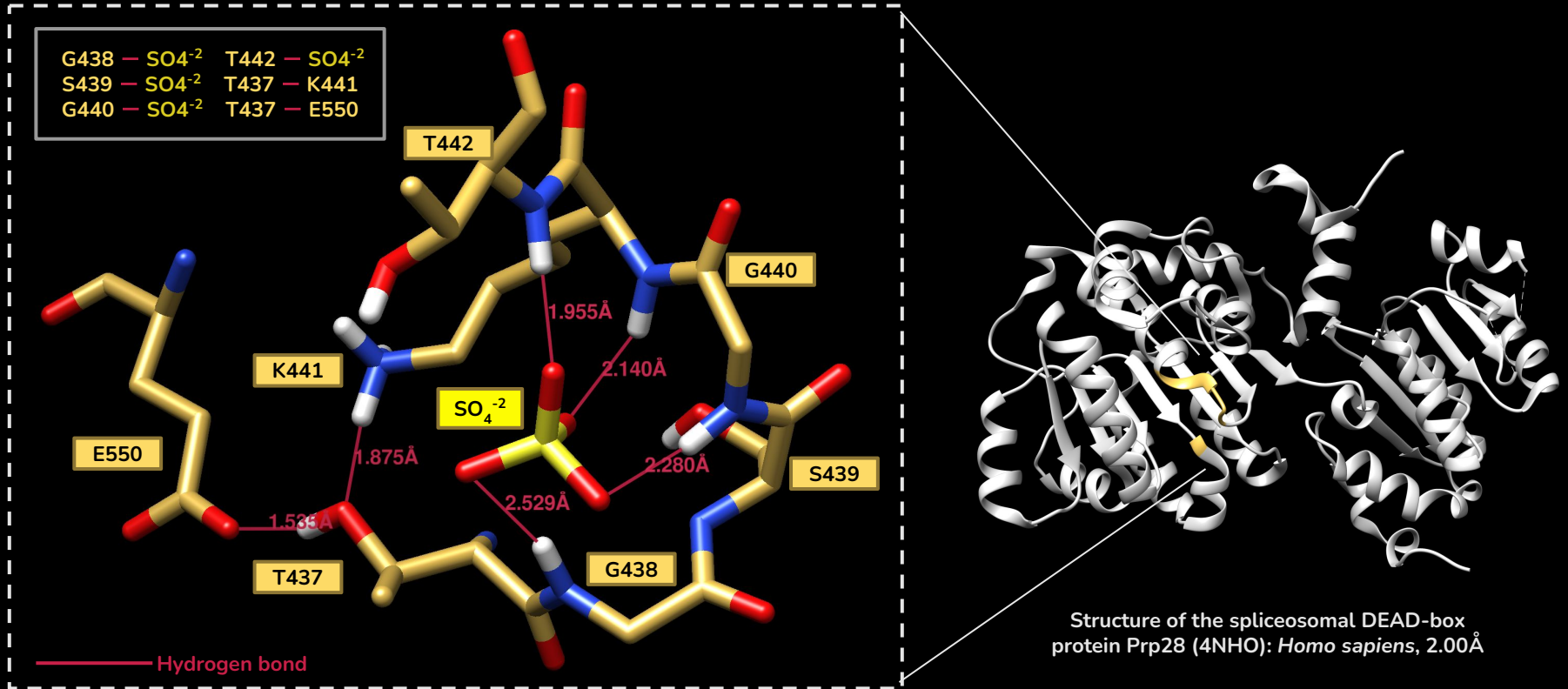
Structure of the spliceosomal DEAD-box protein Prp28 (4NHO): *Homo sapiens*, 2.00Å



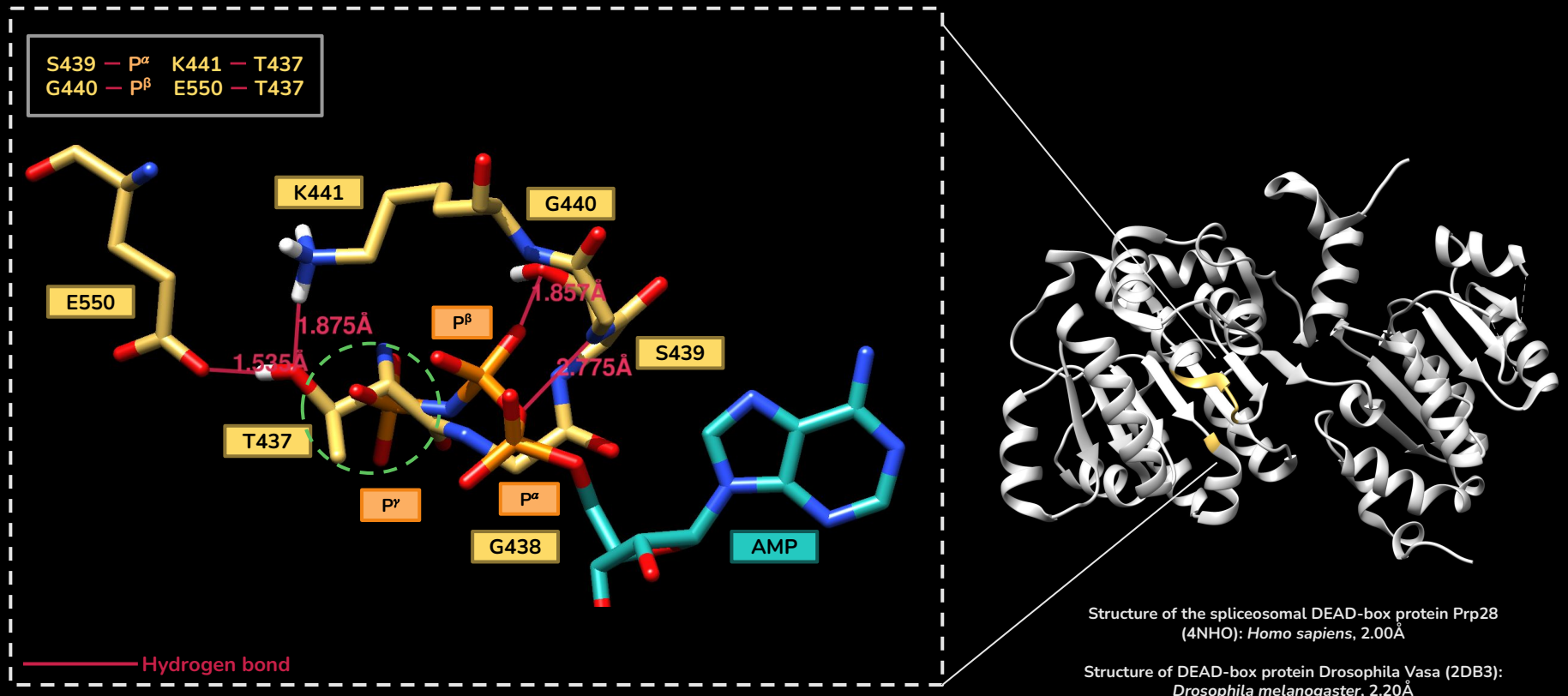
Hydrogen bond

Salt bridge

Prp28 inactive conformation



Prp28 active conformation



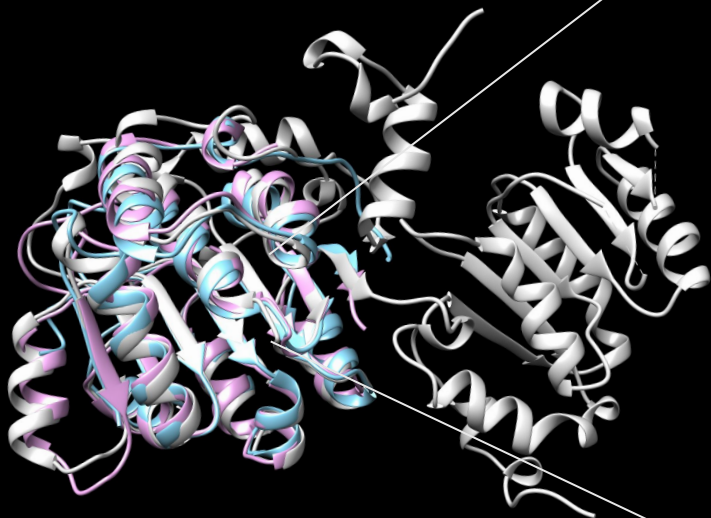
Prp28 active conformation



Structure of the spliceosomal DEAD-box protein Prp28 (4NHO): *Homo sapiens*, 2.00Å

Structural superimposition (P-loop)

	No.	Domain1	Domain2	Sc	RMS
Pair 1	4nho	2gxu	3.92	1.32	
Pair 2	4nho	1qde	3.79	1.43	
Pair 3	2gxu	1qde	7.71	1.16	

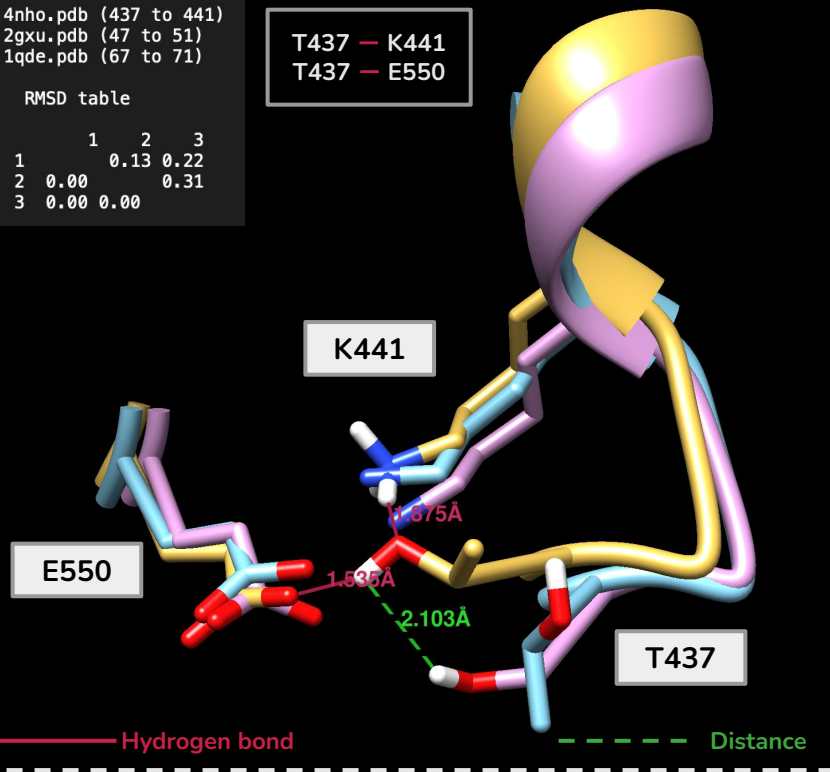


Structure of the DEAD-box protein Prp28 (4NHO): *Homo sapiens*, 2.00Å
HERA N-terminal domain in complex (2GXU): *Thermus thermophilus*, 1.67Å
Structure of the ATPase eIF4a (1QDE): *Saccharomyces cerevisiae*, 2.00Å

```
# 4nho.pdb (437 to 441)
# 2gxu.pdb (47 to 51)
# 1qde.pdb (67 to 71)
```

```
# RMSD table
#
#      1      2      3
# 1      0.13  0.22
# 2      0.00      0.31
# 3      0.00  0.00
```

T437 — K441
T437 — E550



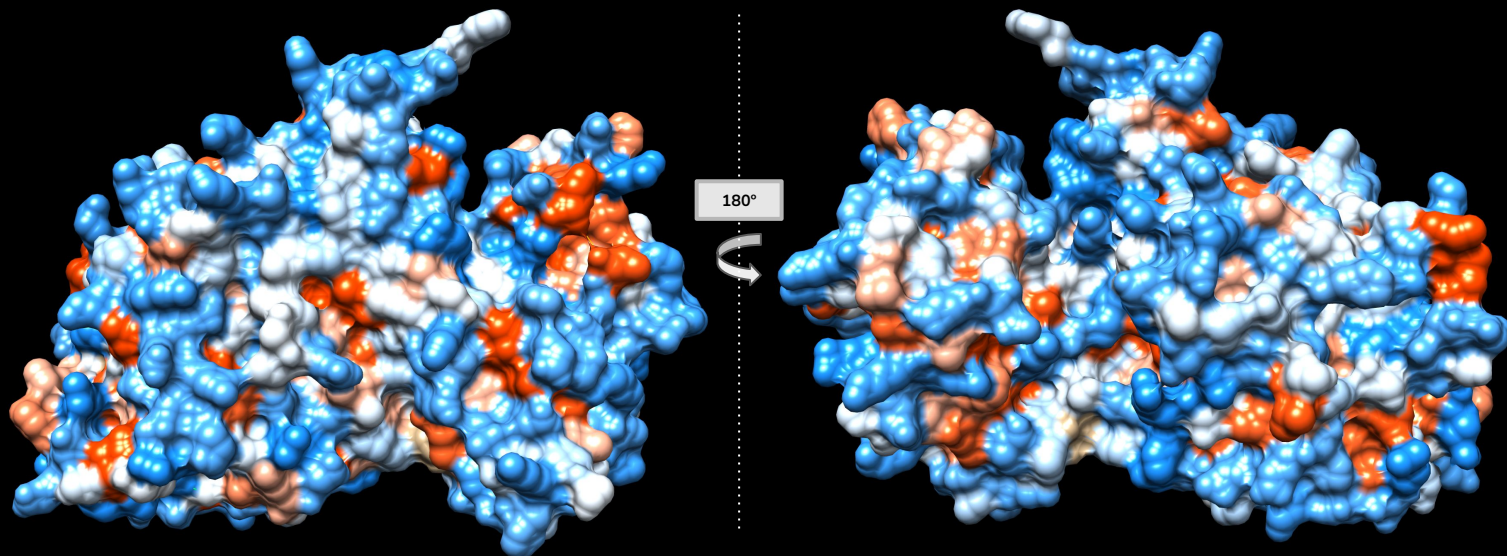
mean global backbone RMSD : 0.22 +/- 0.08 Å (0.13..0.31 Å)

Polarity

Hydrophilic

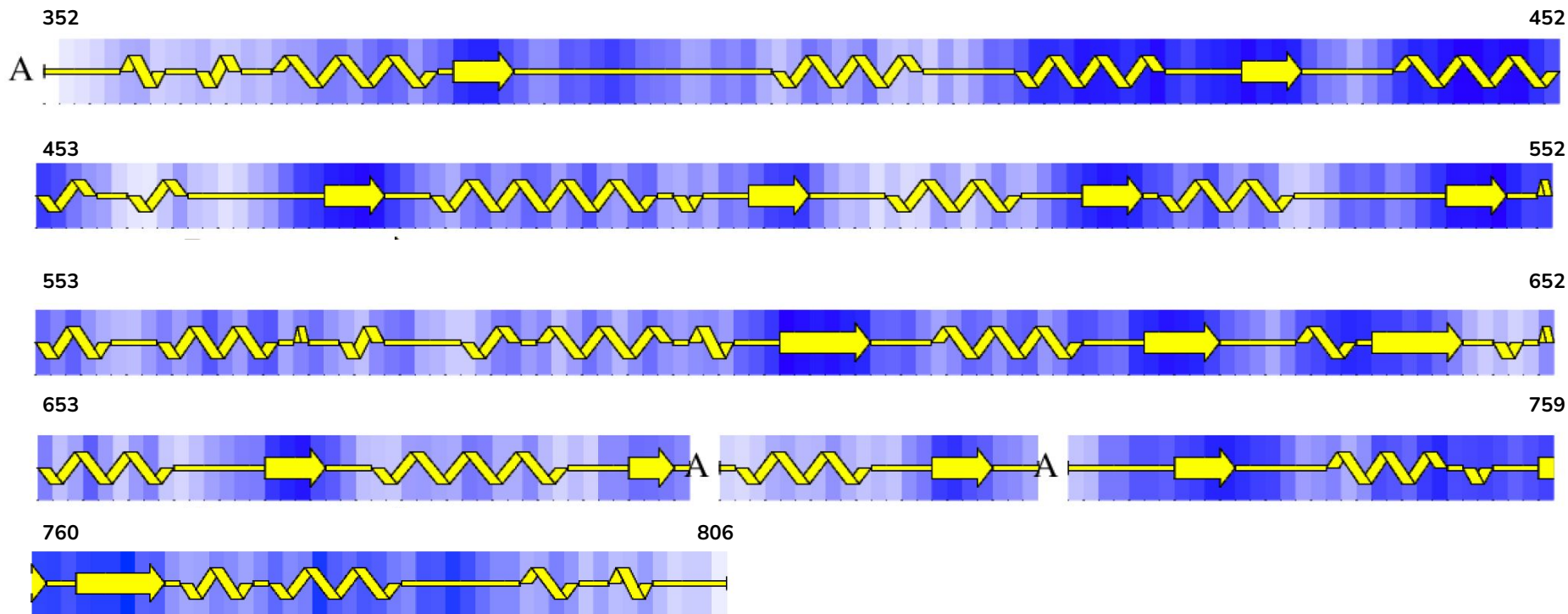


Hydrophobic



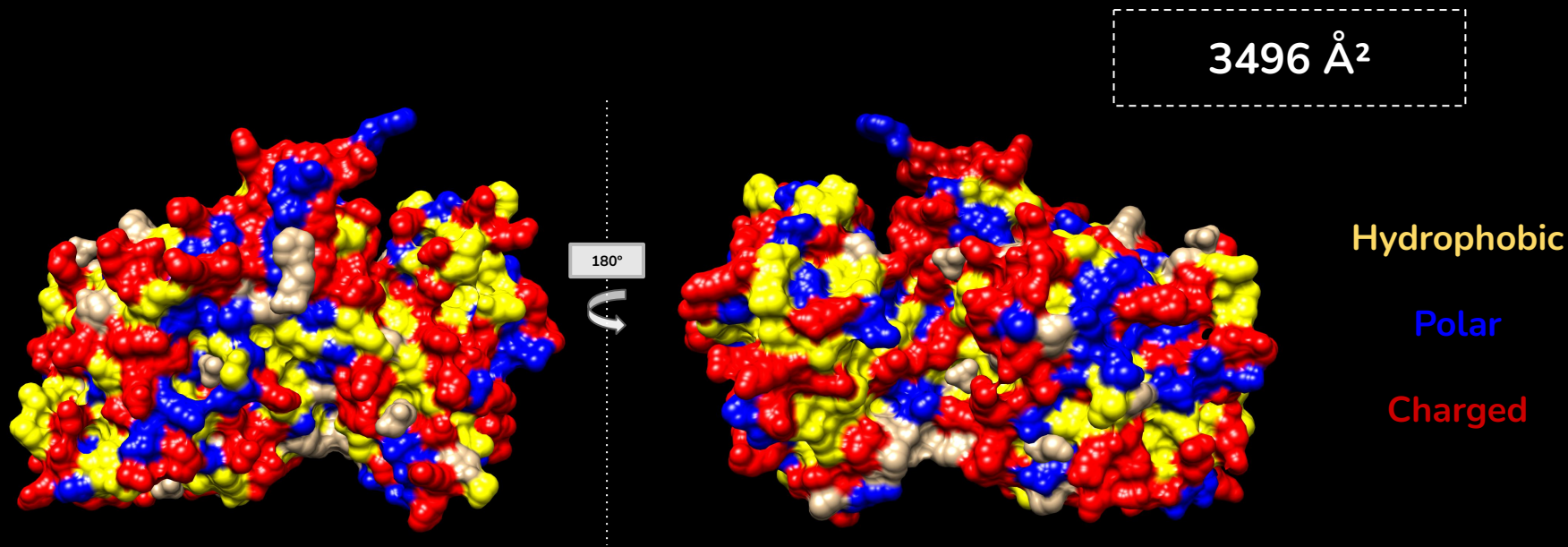
Structure of the spliceosomal DEAD-box protein Prp28 (4NHO): *Homo sapiens*, 2.00Å

Buried surfaces



Key:-  Helix  Beta strand  Random coil Accessibility shading:  Buried  Accessible

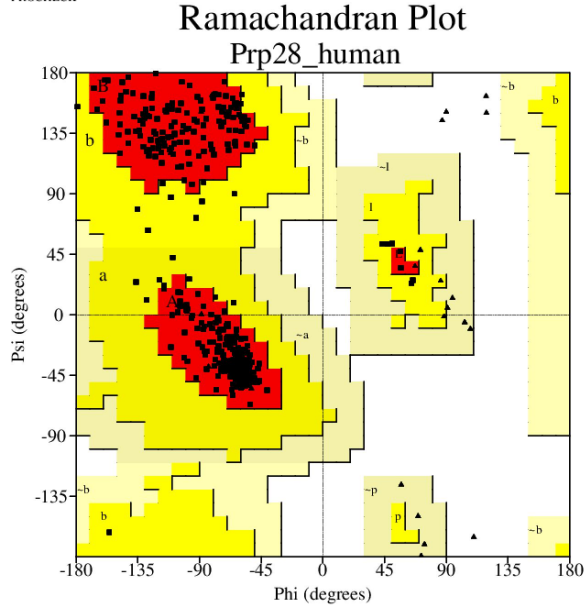
Buried surface



Structure of the spliceosomal DEAD-box protein Prp28 (4NHO): *Homo sapiens*, 2.00Å

Ramachandran plots

PROCHECK

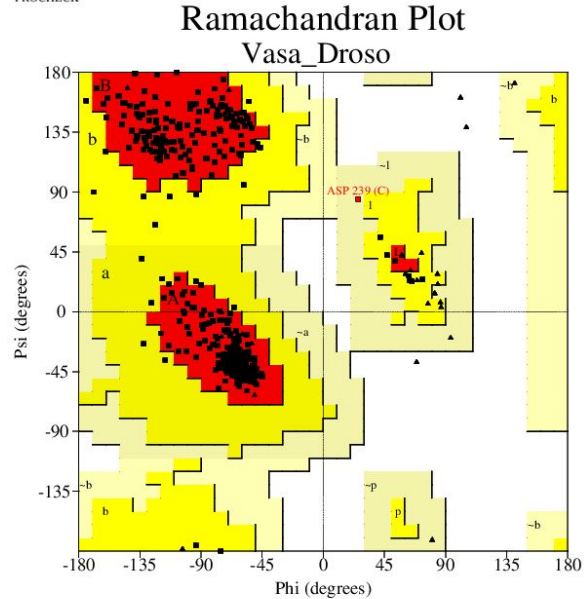


Plot statistics

Residues in most favoured regions [A,B,L]	355	92.9%
Residues in additional allowed regions [a,b,l,p]	27	7.1%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	0	0.0%

Number of non-glycine and non-proline residues	382	100.0%

PROCHECK

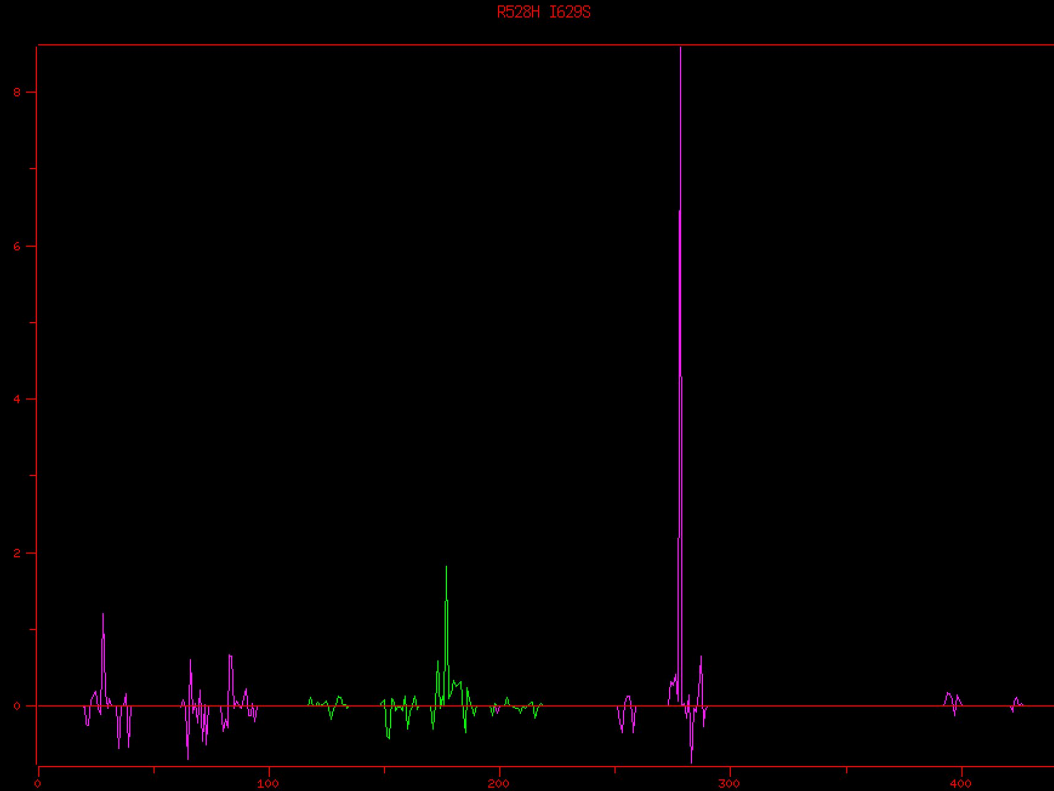


Plot statistics

Residues in most favoured regions [A,B,L]	340	92.1%
Residues in additional allowed regions [a,b,l,p]	28	7.6%
Residues in generously allowed regions [~a,~b,~l,~p]	1	0.3%
Residues in disallowed regions	0	0.0%

Number of non-glycine and non-proline residues	369	100.0%

Prp28 mutations



Prp28 - Likely pathogenic variant

R528H

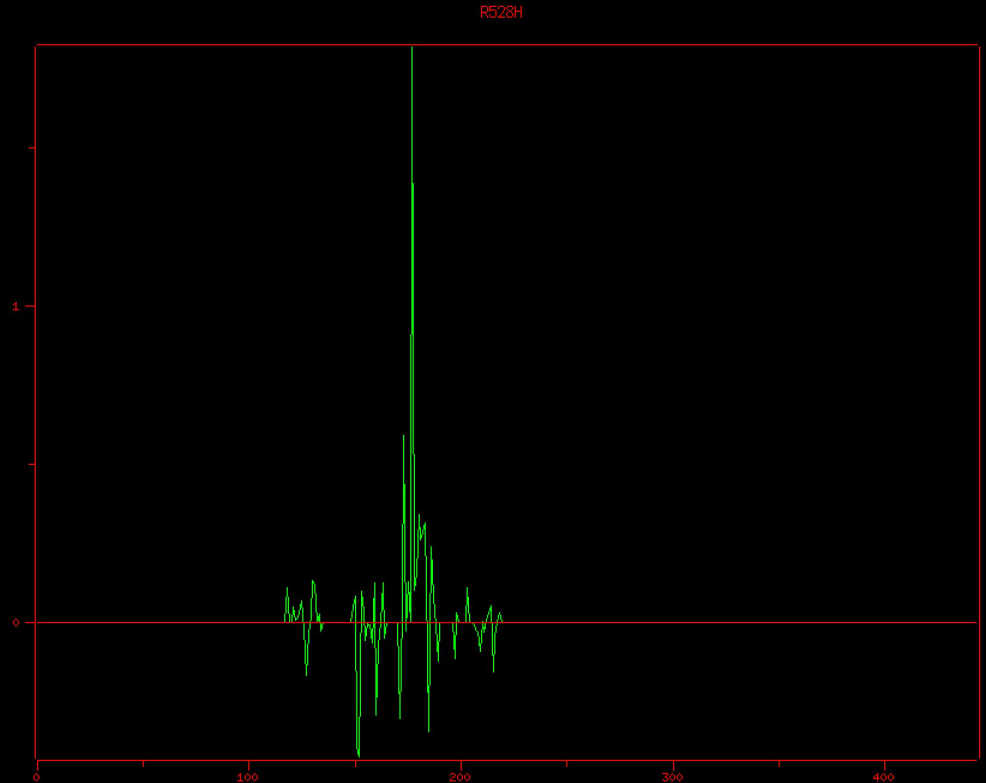
Missense mutation (C>T)

rs1938420751 (dbSNP)

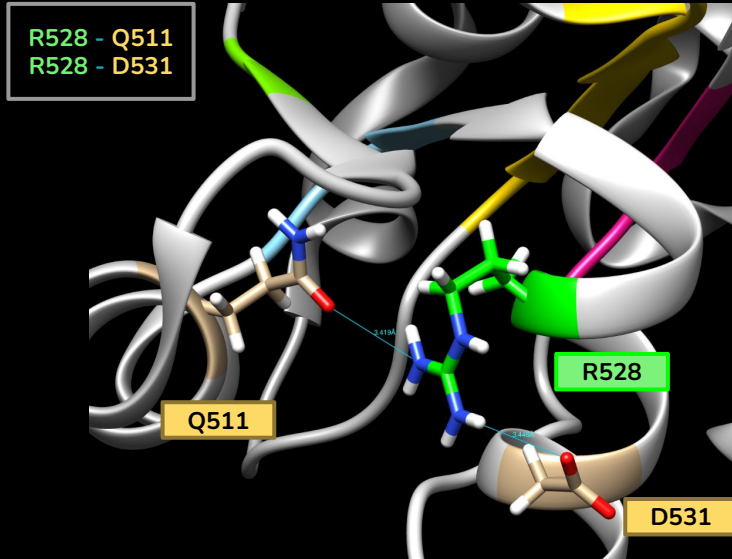
T = 0.000007 (1/140234)

Outside RecA1 region

Intellectual disability



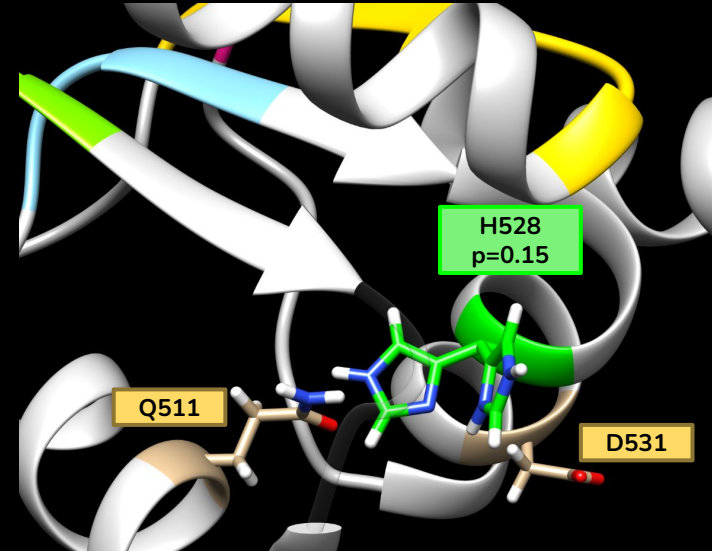
Prp28 – Likely pathogenic variant



Polar, positively charged

Guanidine group → High HB formation
↳ Easier salt bridge formation

Structure of the spliceosomal DEAD-box protein
Prp28 (4NHO): *Homo sapiens*, 2.00Å



Polar, can be positively charged

Imidazole group → Protonation state
depends on local pH value

Shapovalov and Dunbrack (2011), A smoothed backbone-dependent rotamer
library for proteins derived from adaptive kernel density estimates and
regressions structure, 19, 844-858.

Prp28 - Pathogenic variant

I629S

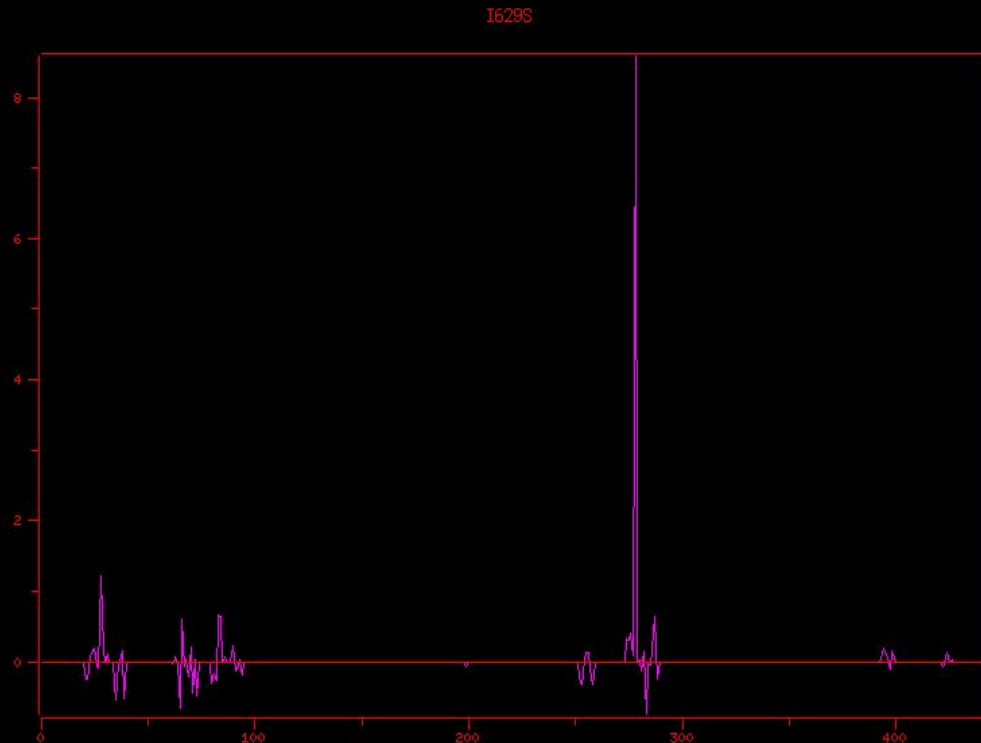
Missense mutation (A>C)

rs1938405472 (dbSNP)

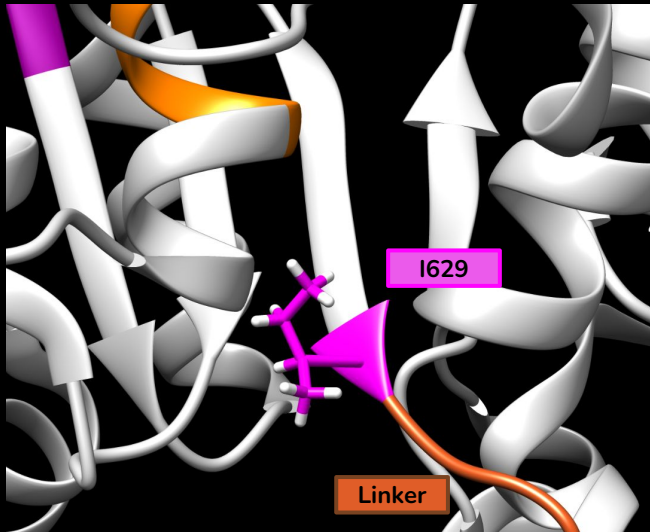
T = 1 case

Preceding linker region

Fetal growth restriction



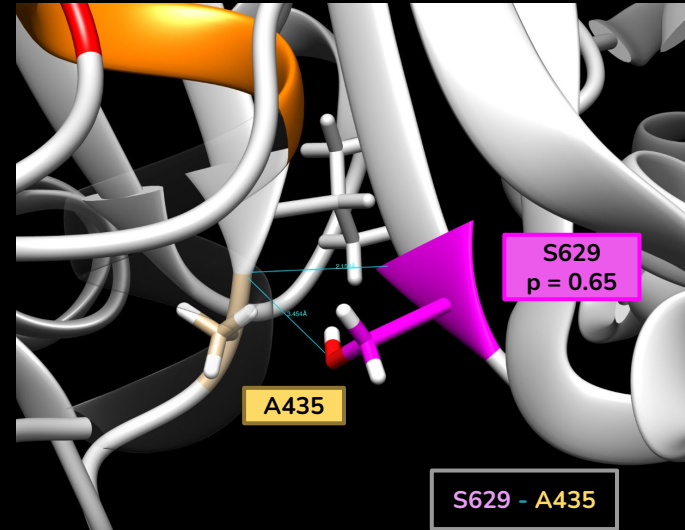
Prp28 – Pathogenic variant



Non-polar, non charged

Structural role → Hydrophobic core, stabilizing

Structure of the spliceosomal DEAD-box protein
Prp28 (4NHO): *Homo sapiens*, 2.00Å

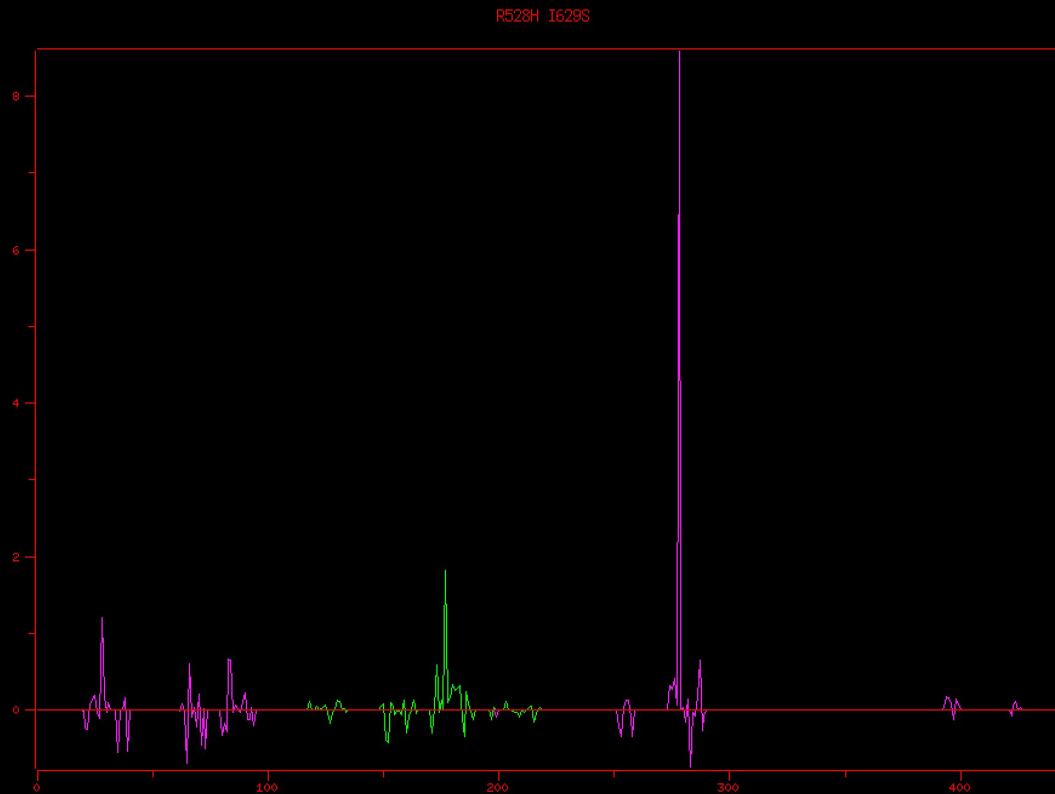


Polar, non charged

Found in the surface
Higher capability of HB formation

Shapovalov and Dunbrack (2011), A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions structure, 19, 844-858.

Prp28 variants



Remarks

1. The spliceosome is a large complex consisting of snRNPs, snRNA and over 170 associated proteins that helps to perform its function.
2. Splicing is a vital process, so the different subunits are highly conserved across eukaryotic species.
3. To carry out the splicing process, different remodelings are necessary in the RNA-RNA and RNA-protein interaction networks.
4. Prp28 is a DExD/H-box helicase, so it conserves different motifs of this family.
5. Changes in the conformation of the RecA domains are important for the activation of the helicase activity of Prp28. Fundamental to these changes is the sequence of the linker that binds them, since it determines the flexibility of the movement.
6. Depending on which region is mutated in Prp28, the destabilization of the protein and the energy peaks will be different. A mutation in the region preceding the linker will be more severe than one in the outer region of the RecA1 domain.

Bibliography

Borao, Sonia, José Ayté, y Stefan Hümmer. «Evolution of the Early Spliceosomal Complex—From Constitutive to Regulated Splicing». *International Journal of Molecular Sciences* 22, n.º 22 (18 de noviembre de 2021): 12444. <https://doi.org/10.3390/ijms22212444>

Charenton, Clément, Max E. Wilkinson, y Kiyoshi Nagai. «Mechanism of 5' splice site transfer for human spliceosome activation». *Science* (New York, N.Y.) 364, n.o 6438 (26 de abril de 2019): 362-67. <https://doi.org/10.1126/science.aa>

Galej, Wojciech P, Thi Hoang Duong Nguyen, Andrew J Newman, y Kiyoshi Nagai. «Structural Studies of the Spliceosome: Zooming into the Heart of the Machine». *Current Opinion in Structural Biology* 25 (abril de 2014): 57-66. <https://doi.org/10.1016/j.sbi.2013.12.002>

Jacewicz, Agata, Beate Schwer, Paul Smith, y Stewart Shuman. «Crystal structure, mutational analysis and RNA-dependent ATPase activity of the yeast DEAD-box pre-mRNA splicing factor Prp28». *Nucleic Acids Research* 42, n.º 20 (10 de noviembre de 2014): 12885-98. <https://doi.org/10.1093/nar/aku930>

Mathew, Rebecca, Klaus Hartmuth, Sina Möhlmann, Henning Urlaub, Ralf Ficner, y Reinhard Lührmann. «Phosphorylation of Human PRP28 by SRPK2 Is Required for Integration of the U4/U6-U5 Tri-snRNP into the Spliceosome». *Nature Structural & Molecular Biology* 15, n.o 5 (mayo de 2008): 435-43. <https://doi.org/10.1038/nsmb.1415>

Möhlmann, Sina, Rebecca Mathew, Piotr Neumann, Andreas Schmitt, Reinhard Lührmann, y Ralf Ficner. «Structural and Functional Analysis of the Human Spliceosomal DEAD-Box Helicase Prp28». *Acta Crystallographica Section D Biological Crystallography* 70, n.º 6 (1 de junio de 2014): 1622-30. <https://doi.org/10.1107/S1399004714006489>

Papasaikas, Panagiotis, y Juan Valcárcel. «The Spliceosome: The Ultimate RNA Chaperone and Sculptor». *Trends in Biochemical Sciences* 41, n.º 1 (enero de 2016): 33-45. <https://doi.org/10.1016/j.tibs.2015.11.003>

Sengoku, Toru, Osamu Nureki, Akira Nakamura, Satoru Kobayashi, y Shigeyuki Yokoyama. «Structural Basis for RNA Unwinding by the DEAD-Box Protein Drosophila Vasa». *Cell* 125, n.o 2 (abril de 2006): 287-300. <https://doi.org/10.1016/j.cell.2006.01.054>

Shapovalov and Dunbrack (2011). A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions structure. 19, 844-858.

Slamovits, Claudio H, y Patrick J Keeling. «A High Density of Ancient Spliceosomal Introns in Oxymonad Excavates». *BMC Evolutionary Biology* 6, n.o 1 (2006): 34. <https://doi.org/10.1186/1471-2148-6-34>

Tauchert, Marcel J., y Ralf Ficner. «Structural Analysis of the Spliceosomal RNA Helicase Prp28 from the Thermophilic Eukaryote *Chaetomium Thermophilum*». *Acta Crystallographica Section F Structural Biology Communications* 72, n.º 5 (1 de mayo de 2016): 409-16. <https://doi.org/10.1107/S2053230X16006038>

Zhang, Xiaofeng, Chuangye Yan, Xiechao Zhan, Lijia Li, Jianlin Lei, y Yigong Shi. «Structure of the Human Activated Spliceosome in Three Conformational States». *Cell Research* 28, n.º 3 (marzo de 2018): 307-22. <https://doi.org/10.1038/cr.2018.14>

Programs used:

- **ClustalW**: Thompson, J D, D G Higgins, y T J Gibson. «CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.» *Nucleic Acids Research* 22, n.o 22 (11 de noviembre de 1994): 4673-80.

- **HMMer**: Eddy, S R. «Profile Hidden Markov Models.» *Bioinformatics* 14, n.o 9 (1 de enero de 1998): 755-63. <https://doi.org/10.1093/bioinformatics/14.9.755>.

- **PROCHECK**: Laskowski R A, MacArthur M W, Moss D S, Thornton J M (1993). PROCHECK - a program to check the stereochemical quality of protein structures. *J. App. Cryst.*, 26, 283-291.

- **ProSa**: Sippl M.J.(1993), Recognition of errors in three-dimensional structures of proteins., *Proteins*, 17, 355-62.

- **SCOPE**: SCOPE: Structural Classification of Proteins — extended. Release 2.08 (updated 2023-01-06, stable release September 2021). From: <https://scop.berkeley.edu/sunid=52671>

- **STAMP**: R. B. Russell & G. J. Barton, Multiple protein sequence alignment from tertiary structure comparison, *PROTEINS: Struct. Funct. Genet.*,14, 309–323, 1992.

- **UCSF Chimera**: UCSF Chimera--a visualization system for exploratory research and analysis. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. *J Comput Chem.* 2004 Oct;25(13):1605-12

- **XAM**: Xia T.H. "Software determination and visual display of NMR structures of proteins" Ph.D. thesis 1992. ETH Zürich, Switzerland.

Multiple Choice Questions

1. How many subunits does the spliceosome have (considering all the complexes)?

- a) 2 subunits
- b) 3 subunits
- c) 4 subunits
- d) 5 subunits
- e) It is not known

2. How many transesterification reactions are needed to transform the pre-mRNA into mature mRNA?

- a) One reaction
- b) Two reactions
- c) It depends on the gene
- d) None
- e) No transesterification reactions are needed; indeed there are two phosphoester transferases

3. Which of the following motifs is name-giving to the family of Prp28 helicases?

- a) DExD/H
- b) QxxR
- c) GG
- d) P-loop
- e) SAT

4. Which species has a unique P-loop conformation in Prp28 characterized by a 2Å separation?

- a) *Thermus thermophilus*
- b) *Drosophila melanogaster*
- c) *Saccharomyces cerevisiae*
- d) *Homo sapiens*
- e) All of the above

Multiple Choice Questions

5. From which spliceosome subunit and complex is Prp28 associated with?

- a) U4 and complex C
- b) U1 and complex A
- c) U6 and complex B
- d) U5 and pre-complex B
- e) U5 and post spliceosomal complex

6. In the inactive Prp28 conformation (closed), P-loop is stabilized by:

- a) ADP
- b) SO_4^{2-}
- c) Mg^{2+}
- d) Ca^{2+}
- e) None of the above

7. The P-loop, required for ATP binding is located in:

- a) RecA2 domain
- b) RS-like domain
- c) RecA1 domain in the motif I
- d) RecA1 domain in the motif II
- e) N-terminal extension

8. Which post translational modification is required in RS-like domain in order to bind Prp28 with the U5 subunit?

- a) Glycosylation
- b) Phosphorylation
- c) Acetylation
- d) Protonation
- e) Solvation

Multiple Choice Questions

9. In which region of Prp28 can occur a likely pathogenic (life-compatible) mutation, creating a positive energy peak?

- a) Outer part of the RecA1 domain
- b) P-loop
- c) RNA binding site
- d) Linker between RecA1 and RecA2 domains
- e) RS-like domain

10. Which is the helicase in charge of the cleavage of U1 subunit from the 5'ss?

- a) Brr2
- b) Prp2
- c) Snu114
- d) Prp28
- e) Prp28 and Prp2