# **Supplementary information**

## Mechanism of centromere recruitment of the CENP-A chaperone HJURP and its implications for centromere licensing

Dongqing Pan, Kai Walstein, Annika Take, David Bier, Nadine Kaiser & Andrea Musacchio

Including Supplementary Figures 1 to 8.

Supplementary Data 1 (a resource table), Supplementary Data 2 (details of cross-links identified) and a Source Data file are available as separate Excel files.

### Predicted structure

human bat dolphin cow cat armadillo mous e

### Predicted structure

human bat dolphin cow cat armadillo mous e

## Predicted structure

human bat dolphin cow cat armadillo mous e

### Predicted structure

human human bat dolphin cow cat armadillo mous e

### Predicted structure

human bat bat dolphin cow cat armadillo mous e

### Predicted structure

human bat dolphin cow cat armadillo mous e

### Predicted structure

human bat dolphin cow cat armadillo mous e

### Predicted structure

human
bat
dolphin
COW
cat
armadillo
mous e

Predicted structure

numan	
oat	
dolphin	
cow	
cat	
armadillo	
mous e	

					<b></b>	
MLGTLRAMEGEDVEI	20 DOLLOKLRAS	RRRFORRMORLIEKY	NOPFEDTPVVOMA	60 TLTYETPOGLR	IWGGRLIKERN	EGEIODSSMKPADR
R F G S P C A M E G E V S G A F G C L R A M E G E V L G	DDVLMO <mark>KLK</mark> KSI EDALLOKLRDSI	RHR FÖRHMÖÖLLEKY RRR FÖRRMÖÖLIEKY	NÕPFEDAPVVÕMS NÕPFEDAPVVÕMS	TLTY <mark>E</mark> TPÒGLR TLTYETPÒGLR	IWGGRLEKEKS IWGGRLIKERN	KGQIQGSPEQTVGRF TGOIOGSPVKTDDRT
MEGEVLS		RCRFORRMOOLIEKY	NÕPFEDGPVIÕMS NÕPFEDAPI VÕMS	TLTYQTPÒGLR TLTYFTPÒGLR	IWGGGLVK-RS	TGHM
MESVCAMEŠESLEI	KAALVRKLODS	RCRFORRMOOLIEKY SSRFOTLMKRLIAKY	NÕPFEDAPLVÕMT NÕPFEDDPLVEMR	TLTYETPOGLR TLTYETPOGLR	IWGGGLIKEMN VWGGKLMKKED	TEQIQESFVELVNGI KEYTOVIDRLNGOAF
100	110	120 13	0 140	15	0 160	170
	SHRTV-LGADSI	SGEVDATSDQEESV	AWALAPAVPQSPL	KN <mark>EL</mark> RR K	ÝLTOVDILLOG	AEYFECAGNRAGRD
GSTRVPAGGHELP		SSDEDATE TOEDLA	AGNLMPAVPWSPL	KNELRRK	YLTOVDILLOD	EGCLEGASYGEGKD
JGAPVVPGTTCLL	PRLG	LISPSFQWDNTD		KNELRRK	YLTOVDYLLOD	EGCLECTGDGGGEDA
DSESSGADTSLEI	ENWPS	CSSAMREASGD	PRQRQPAVPGNTL	ETDLRR K	YLTQVDILPQD	EEYF <mark>K</mark> NAEKRG <mark>G</mark> KD1
			<b></b>			
180 19	200	210	220 2	30 2	40 25	0 260
VTLVPSLASPAVP	HGYCSRISRK	SPGDPAKPASSPREW	LCSPDTVIV	PRNDSLSLQE- PRNDSVSLQE-	TSSHSFLSSOP	FVAGDICNVTISDL
TLAPSLASPARP	QGYCKDISED	RPGGPLQPPSPREG		PRGDGISLQGG PRGDGIPLHG-	TGGGSFSSGPC	SKGEDVCNATLSDL
		NLGHPIDSASSHGEC	GPSRPRSADMALV	PRNDSLWLPG-	TRDGSFLNSQS	FEAEDTCDVTISDL
	ASUCQUAISAK.	SSUCE VSALSSKUU	GF 5 IF CF ADMA <mark>I V</mark>		-12211262268	FEVDULCIVIISULI
270 23	80 21	300	310	320	330	340 350
GMLHSMSRLLSTK	2 SSIISTKT-F	MQNWNSRRRHRYKS	RMNK TYCKGAR	RS <mark>QR</mark> SS <mark>KE</mark> NFI	PCSEPVKGTGA	LRDCKNVLDVSCRKI
GMLHSMSRLLSSK GMLHSMSCLLGVR	PSCVISTKT-F SCVISTKT-F	L V R N W S S R R R H G C K S L R Q H W S S K R R H R C K S	RMNR TYCRGGR RMDR TSCRGGG	HA <mark>RR</mark> DS <mark>QE</mark> TLV RS <mark>RR</mark> SP <mark>RE</mark> RRP	PCSEPVKEAGV PGSEPAKDVAV	L R E S Q N F R D L S G Q K A L R D R E N V L D V S G H K 1
GMLHSMSRLLGAR GMLHSMSLLLSAK	PSCIISTKT-L PSCIISTKTSF	L G O N W S F R R R H R C K G L H N W S S R R R C S – R S	RMNR TSCRGGG RVNR THCRGGR	R S <mark>R Q</mark> S P Q E R R P R P H R A S Q E R L P	PCSEPLEDAGT VRAGPGRDRGA	L R D R G N L P D V S V P K 1 L R D C E N V L D A S G Q K 1
MLHSMSQLLSAK MMHSMSRLLRSK	PSCIISIKI-L PSCIISTKT-Y	L V O NWN SRERL R G KN L N O SWK L <mark>R R R</mark> P S R <mark>K</mark> O	KMNK TYCKGGR G <mark>L</mark> H <mark>K</mark> NR <mark>T</mark> H <mark>C</mark> PRSK	PS <mark>QR</mark> SSKKRAF PS <mark>QR</mark> SA <mark>RK</mark> GPA	SCSEPGKEVEV SCSEPGKEAGI	L R D A K N L L N A S C H Q I L R D Y G N L L H V A P H K 1
360	370	380 390	400	410	420	430 440
L <mark>K</mark> LEKAFL <mark>E</mark> VN <mark>R</mark> PI L <mark>K</mark> LEKAFP <mark>E</mark> VNKL	IHKLDPSWKE	RKVTPS <mark>K</mark> -YSSLIYF LKGTPR <mark>K</mark> -LSSLTYV	DSSATYNLDEENR DSSVTHRLDRENR	FR <mark>TLKWLI</mark> SPV LM <mark>TLKWLI</mark> SPV	KIVSRPTIROG KVVSRPRMLOG	HGENROREIEIRFDC EGGNHYGAFDSKFK
_ <mark>K</mark> LGRAFL <mark>E</mark> VNQP _KSGRALLKVNKP	ĴIHAFASHW <mark>KE</mark> ĴIHTFTPPWKE	LPRTPQ <mark>K</mark> HHSSLTYL LQGMPQ <mark>K</mark> -HSSLTYL	DSNAVYHLDQENR DSKVLSPLDQENR	FM <mark>TLNWLISPV</mark> FV <mark>TLK</mark> WLISPV	KIGSRPRVPPG KIASRPRILRG	EGGNRYREIEIRFD QAGSRFKEIEIKFD
LEVNKS KLGKVFLEVNKP	VLKWGPSWKE IRKSDPSRKE	SLTPQG <mark>R</mark> SLLPYG LRLTPQ <mark>K</mark> -HSSLTYL	DSRAVCHLDQKNR R HNDDQENR	YK <mark>ALK</mark> WLISPV LM <mark>ALK</mark> WLISPV	KIVSRPRILPG KTVSNLGTRQG	KRGNHYREIEIRFDH HRGNRYREIEINFEH
_ <mark>E</mark> LKSVSL <mark>E</mark> GS <mark>K</mark> R	QVHKSSPAW <mark>KE</mark>	_QMMPQ <mark>K</mark>	DLDL <u>NRERENR</u>	VM <mark>TLQWLISPV</mark>	KVVPRPRMLPS	<u>QVEKWYRE</u> IKIK <mark>F</mark> D
	460 MCLPDSWA	470 4	80 490 GLETRRLSLPSSK	500 AKAKSLS <b>E</b> ÅFE	510 NL GKR <mark>S</mark> LEA	520 GRCLPKSDSSSLP
HQKYCPSPRKQLG HOEYCSNLRKOPC	TYLPSSSSFG TYLPSSSA	DVLRSGLASPRSPO VDVLRSGLASPRSPO	GLETHRPSRPLSK GIEAHRPSSPLRR	AKV <mark>K</mark> SLNEAFK AKAKRLSEAFE	NLGKWAIET SLGKGSIRE	GRCLPQSDSSASFS GSCPPKSDSFPSLS
HÕEYCLSPG <mark>K</mark> ÕLS HÕEYCPSPRKÕPV	. T Y P P G <mark>S</mark> A . T S L P E <mark>S</mark> S A	<mark>/ D V Y R</mark> GG P M S P G S P <mark>Q</mark> / D V Y R G G P A S P R G P R	GLEAHRLSSPLCK GFETRRLSGTFGT	SKA <mark>K</mark> RLS <mark>E</mark> ASE AEA <mark>K</mark> RFN <b>EA</b> FE	GLGRR <mark>S</mark> IRV ELGDR <mark>A</mark> RGA	GSCLPESESFPSLS GRCLQRRGSPPPLPE
HQEYCPNPGKQPS HQEYCLSSGKQPR	. TYLPS <mark>S</mark> WGI . TDPTE <mark>S</mark> WA	<mark>MDVYR</mark> GSP <mark>K</mark> VDVYRSGSKSPGSRQ	GSETORLSITFP- DVETCRPSSPFGR	<mark>K</mark> RLS <mark>E</mark> SFE EKT <mark>E</mark> RPG <mark>EA</mark> LE	NQGER <mark>S</mark> LEG DLRGNGK <mark>S</mark> VKT	GRYLLKSNSDSSLS KSCLLRSCPSP
530 540	550	560	570 58	0 590	600	610
	OLHVQGNSSGI	ΓΚΚSVSPSKTLSVPD RESAALSKATSVPG	KEVPGHG <mark>RN</mark> RYDE	IKEEFDKLHOK IREKEDOLHOO	YCLKSPGOMTV	PLCIGVSTDKASME
NSTCSPGRSEQTS		FRKSVSLNKAISVPR FSVSFSKATSVPR	VOPPGCARDRYDD	IKEKFDKLHOK TKEKEDKLHOO	YCOKSLOOTOV	
SPVQGPRRSEQTL DAAHSTSHFELTCI	Í FÓGŇ-NSLGTI NSLFOGGHPGVI	LRKSASPNRPTSVAG FRMLFSPSKAMPKPG	V O P L S C G R D R Y K E L O A L G G G R N H Y D E	IKEKFDKLHQK IKEKFNKLHOK	YCOOSPPRTKA	LLRVGACPDKASLE TLCIRHSPNKASTL
PSRSPSHSQLSS	GLQEHNSEPTG	KAVWPSTAISAPS	IĞSPĞCĞKDNYYE	LKKEFNRLYÖK	YCLS-PQRAKV	TSCGRVSPM <mark>KA</mark> AAAI
620 630	640	650	660 67	0 680	690	
NOTEAFFKKLDPD	SGLLGPRKLSS	SPEGERKSEEGSTAL SLORSIRSSRDSTTV	EPHPSPWFELAAS		KRRRLSEPUGS	GRWAESQDPSRMVG
YQKEGFSVKFNPD	SGCRGPPSLSS	SPQQSIKSPLGSNII SPQQSVKSPLASNTV	GAPPSTGFALDAS	SGHQVPA	KRCRL SDPQVC	GEGARPWSFS - PVGF
YKKDDLG-KFNPD		- POCSMKSPWGSTPV	KVLPSTRIAQASR		KRRRLSDSVLC	GOWADCWNPSSVMSF
	οι άλος τη άλος τη άλος τη άλος τη άλος τη άλος τη από τη	σεσοποκνίζυσι Αυ	1A3ILV	NUSWLPI	NNCKES IPVAC	AUVAKLUN SUNANUN
710 73						
V-RPGDQGSSSQPI	S-EERGENTS	RMEEKSDFMLEKLE	TKSV-			
A F P R P G K É V C S S P D I A I P R P G E E A G S V A T I	NE-RKRGKSTS NE-EKGKNNTS	FRMEDK FRMEEKSDFVLENLK	AENLL			
MPRP	AE-KKRCNNTS	/KMEEKSDFVFEKFK	AGEV-			
ATARAÓEEAVCPLO	DFEEKGSNLS	- QMEERNDFGLENFK	AKHL -			

# Supplementary Figure 1. Amino-acid sequence alignment of mammalian HJURP.

Multiple sequence alignment was generated using ClustalW (Larkin et al. 2007). Residues that are identical in all sequences are shaded red, and residues that only have conserved substitutions with similar properties are shaded yellow. Secondary structure prediction was performed using PSIPRED (Buchan et al., 2013). Green bars indicate the residues predicted to be in  $\alpha$ -helical conformation. Blue arrows indicate the residues predicted to be  $\beta$ -strand. Red boxes indicate the parts of this alignment presented in Figure 1G. The amino-acid sequences were retrieved from UniProt or Ensembl databases with following identifiers: human, Q8NCD3 (UniProt); bat, G1PWU4 (UniProt); dolphin, ENSTTRP00000004146 (Ensembl); cow, E1BB33 (UniProt); cat, ENSFCAP00000051395 (Ensembl); armadillo, ENSDNOP00000013322 (Ensembl); mouse, Q6PG16 (UniProt).



#### Supplementary Figure 2. Characterization of HJURP R1 and R2.

(A) Amylose-resin pull-down assays showing that HJURP<sup>555-748</sup> binds more weakly than HJURP<sup>541-748</sup> to Mis18<sup>core</sup>. (B) Sedimentation velocity AUC results of MBP-HJURP<sup>394-540</sup> and MBP-HJURP<sup>541-748</sup>.



#### Supplementary Figure 3. Protein expression and HJURP depletion in HeLa cell lines.

(A) Western blots showing doxycycline-induced expression of EGFP-HJURP variants and SNAP-CENP-A in the stable HeLa cell lines used in this study. Cells were treated with 50 ng/mL doxycycline for 24 hours before harvesting. Clear lysate (containing 50 µg protein) obtained after cell disruption and centrifugation was applied on each lane of SDS-PAGE gels and blotted using the antibodies indicated in the figure. (B) Western blotting results showing the efficient depletion of the endogenous HJURP by the HJURP Stealth siRNA. Clear lysate containing 25 µg protein was analyzed for each condition. Source data are provided as a Source Data file.



#### Supplementary Figure 4. Characterization of Mis18core variants.

(A) Sedimentation velocity AUC results confirm the hexamer state of Mis18α<sup>78-C</sup>: Mis18β<sup>65-C</sup>. (B) Amino-acid sequence alignment of Mis18α. Residues that are identical in all sequences are shaded red, and residues that only have conserved substitutions with similar properties are shaded yellow. The residues in the boxed region (blue rectangle) are shaded in red and yellow according to their conservation among the mammalian species compared in this alignment. The amino-acid sequences were retrieved from UniProt, Ensembl or NCBI Protein databases with following identifiers: human, Q9NYP9 (UniProt); dolphin, A0A2U3V0K3 (UniProt); cow, A5D7N9 (UniProt); cat, M3WWJ6 (UniProt); armadillo, ENSDNOP00000004670 (Ensembl); mouse, Q9CZJ6 (UniProt); chicken, E1BQA3 (UniProt); frog, XP\_002938488.2 (NCBI Protein). (C) Amylose-resin pull-down assays showing that deletion of the Mis18α N-terminal tail did not affect the interaction of Mis18<sup>core</sup> with M18BP1<sup>1-60</sup>.



#### Supplementary Figure 5. Mis18a<sup>56-C</sup> mutant is functional for centromere localization and CENP-A deposition.

(A) Western blotting results showing the effective depletion of the endogenous Mis18 $\alpha$  by Mis18 $\alpha$  siRNA. Clear lysate containing 100 µg protein was analyzed for each condition. (B) Western blots showing doxycycline-induced expression of EGFP-Mis18 $\alpha$  variants and SNAP-CENP-A in the stable HeLa cell lines used in this study. Clear lysate was prepared as described in the legend of Supplementary Fig. 3. Clear lysate containing 100 µg protein was analyzed by SDS-PAGE and western blotting using the antibodies indicated in the figure. (C) Representative images showing the fluorescence of SNAP-CENP-A and EGFP-NLS, EGFP-Mis18 $\alpha$  or EGFP-Mis18 $\alpha$ <sup>56-C</sup> in the fixed HeLa cells. CENP-A deposition experiments were performed as described in Figure 2A using Mis18 $\alpha$  siRNA instead of HJURP siRNA. White scale bars indicate 10 µm. (D) Quantification of the centromere fluorescence intensity of SNAP-CENP-A and EGFP signals. The bar graphs represent mean values from three replicate experiments (blue dots indicate the mean values from each experiment). Error bars indicate standard deviations. Source data are provided as a Source Data file.



Supplementary Figure 6. HJURP forms a stable complex with Mis18core and M18BP1

(A, B) SEC profiles and SDS-PAGE gels showing the results of sample preparation of Mis18 $\alpha^{56-C}$ :Mis18 $\beta$ :HJURP<sup>394-C</sup> complex or Mis18 $\alpha^{56-C}$ :Mis18 $\beta$ :MBP-HJURP<sup>394-C</sup> for AUC experiments. TEV protease was used to cleave off the MBP-tag from MBP-HJURP<sup>394-C</sup> to obtain Mis18 $\alpha^{56-C}$ :Mis18 $\beta$ :HJURP<sup>394-C</sup> in (A). Black arrows indicate the SEC fractions from which samples were taken for AUC analyses presented in Figure 5G. (C) Best-fitting results of the sedimentation velocity AUC data of the AUC analyses presented in Figure 5G. (D) SDS-PAGE gels showing the samples used for the AUC analyses presented on Figure 6F. (E) Best-fitting results of the sedimentation velocity AUC data of the AUC analyses provided as a Source Data file.



Supplementary Figure 7. UV-cross-linking experiments of HJURP incorporating Bpa or AbK and additional binding assays. (A) SDS-PAGE gel images showing the protein samples at each step of UV-cross-linking experiments. Step 1, Mis18 $\alpha^{56-C}$ :Mis18 $\beta$  was incubated with MBP-HJURP fragments (with C-terminal 8His-tags) containing Bpa- or AbK-substitution at the indicated positions. Step 2, protein samples were exposed to UV light for 30 min. Step 3, unbound fraction of the Ni<sup>2+</sup>-affinity purification was separated from the Ni<sup>2+</sup>-resin. Step 4, Ni<sup>2+</sup>-resin-bound fraction contained cross-linked products and free MBP-HJURP fragments and HJURP-bound Mis18 $\alpha^{56-C}$ :Mis18 $\beta$ . Step 5, Ni<sup>2+</sup>-resin-bound fraction after washing the beads with urea buffer contained less non-cross-linked Mis18 $\alpha^{56-C}$ :Mis18 $\beta$ . (B) Amylose-resin pull-down assays showing that HJURP<sup>394-C</sup> does not pull-down the trimeric helix-bundle of Mis18 $\alpha$ :Mis18 $\beta$ . Buffer B was used for the pull-down assays. (C) Sedimentation velocity AUC results of Mis18 $\alpha^{56-C,V211D}$ :Mis18 $\beta$ .



Supplementary Figure 8. Chemical structures related to the chemical synthesis of 3'-azibutyl-N-carbamoyl-lysine.