SUPPLEMENTARY DATA

Evolutionary and Functional Insights into the Ski2-like helicase Family in Archaea : Comparison of Thermococcales ASH-Ski2 and Hel308 Activities

Manon Batista[1†], Petra Langendijk-Genevaux[2†], Marta Kwapisz[1], Isabelle Canal[1], Duy Khanh Phung[3], Laura Plassart[1], , Régine Capeyrou[1], Yann Moalic[4,5], Mohamed Jebbar[4], Didier Flament[4], Gwennaele Fichant[2], Marie Bouvier[1], and Béatrice Clouet-d'Orval[1*]

[1] Laboratoire de Biologie Moléculaire, Cellulaire et du Développement, UMR5077, Centre de Biologie Intégrative (CBI), Université de Toulouse, CNRS, Université Paul Sabatier, F-31062 Toulouse, France.

[2] Laboratoire de Microbiologie et de Génétique Moléculaires, UMR5100, Centre de Biologie Intégrative (CBI), Université de Toulouse, CNRS, Université Paul Sabatier, F-31062 Toulouse, France.

[3] University College London, Gower Street, Darwin Building, London WC1E 6BT United Kingdom.

[4] Univ Brest, CNRS, Ifremer, UMR6197 Biologie et Ecologie des Ecosystèmes marins Profonds, F-29280 Plouzané, France

[5] Present address: LabISEN, Yncréa Ouest, 20 rue Cuirassé de Bretagne, F-29200 Brest, France

[†] Joint Authors

*****To whom correspondence should be addressed: Tel: 33 5 61 55 69 06; Fax: 33 5 61 33 58 86; Email: Beatrice.Clouet-dOrval@univ-tlse3.fr

CONTENT:

SUPPLEMENTARY FIGURES

Figure S1. Congruence of the archaeal species tree and ASH-Ski2 or Hel308-Ski2 trees. In the archaeal tree (left), the branches are coloured according to the taxonomy using the same colour code as in Figure 2. Reconciled ASH-Ski2 (A) and Hel308-Ski2 (B) trees have been inferred with GeneRax (54). Trees were obtained by using the cophylo function from the phytools R package (79). On the reconciled protein trees, branches for which GeneRax predicted an HGT are in blue and branches for which a duplication event was predicted are in black. The tips of the reconciled protein tree are linked to the corresponding species in the species tree. The crossing of the lines connecting the species to the corresponding protein reveals incongruence between both trees. A square indicates a group of duplicated Hel308-Ski2 proteins among several Halobacteria.

Figure S2. Conserved motifs of euryarchaeal ASH-Ski2 and Hel308 proteins visualized with WebLogo 3 (https://weblogo.threeplusone.com/create.cgi). Hel308 and ASH-Ski2 specific domains are framed in light blue and in red, respectively. (A) Ski2 SF2 core domain. RecA1 and RecA2 domains are highlighted in light and dark orange, respectively. Conserved motif involved in ATP binding and hydrolysis are shown by a black star. (B) Ski2 C-terminal domain. Winged helix and ratchet domains are indicated in light and dark green, respectively.

Figure S3. Alignment of predicted Structures of *P. abyssi* ASH-Ski2 and Hel308 and *Archeoglobulus fulgidus* Hel308 structure with double-stranded DNA (PDB 2P6R). The structures for ASH-Ski2 and Hel 308 from *P. abyssi* were predicted using alphafold (https://alphafold.ebi.ac.uk/). The models present a global pLDDT confidence score >80 corresponding to a good backbone prediction excepted for the end of the C-terminal extension of ASH-Ski2 (831 to 855aa). The SF2 cores and the C-terminal regions are highlighted in orange and in green, respectively. Hel308 (ID-H) and ASH-Ski2 (ID-A) specific insertion domains are indicated in blue and in red, respectively. The RMSD from structures aligned with pymol software is 1,620.

Figure S4. ASH-Ski2WT, ASH-Ski2K245A, ASH-Ski2ΔN and Hel308WT recombinant proteins from *P. abyssi* used in this study. A- Aliquots of recombinant proteins purified to homogeneity were visualized on 4-15% SDS-PAGE stained by Coomassie-blue. The absorbance ratio of A260/A280 of the final purification sample for ASH-Ski2 WT, K245A, ΔN & Hel308 is 0,63 ; 0,69 ; 0,70 & 0,54 respectively. B- Size exclusion chromatography (S200 increase 10/300 GL) showing that ASH-Ski2 WT is a monomeric protein with a single peak at an elution Ve of 13,43ml that corresponds to an apparent molecular mass of 77,6 kDa. Calibration of the S200 increase 10/300 GL column is shown on the right panel.

Figure S5. Nucleic acids binding filtration assay. (A) Nitrocellulose filter binding assays using a dot-blot apparatus was performed using 5nM of radiolabelled single strand RNA_{50} or DNA_{50} incubated with increasing concentrations (0, 10, 20, 30, 40, 50, 75, 100, 150 & 200 nM) of protein at 30°C for 15 minutes. Nucleic acids protein complexes are retained on the nitrocellulose membrane while unbound RNA or DNA molecules are trapped on DEAE nylon membrane. (B) Binding assays were performed using 0.5nM of radiolabelled [γ-32P]ATP RNA50 or DNA50 incubated with increasing concentrations of ASH-Ski2ΔN and ASH-Ski2K245A variants ranging from 0 to 200nM.(C) ASH-Ski2/RNA_{50} binding assays in presence of ATP. The same experiment of RNA_{50} binding is realized with in presence of 5mM

ATP. (D) Comparison of ASH-Ski2 binding assays performed in presence of double stranded RNA_{50} (dsRNA).

Figure S6. ATP phosphohydrolytic activity of ASH-Ski2 WT and Hel308 from *P. abyssi*. (A) Radiolabeled [γ-32P]ATP was incubated at 65°C with 250nM of purified recombinant protein. Aliquots taken after increasing times of incubation as indicated. The reaction products were analysed on Thin Layer Chromatography. Each datum is an average of three separate experiments. The percentage of radioactive inorganic phosphate was quantified over time. Experiments were carried with *P. abyssi* total RNA, *P. abyssi* genomic DNA or without nucleic acids. (B) ATPase assays of ASH-Ski2 and Hel308 in presence of *P. abyssi* genomic DNA. The initial rates calculated for ASH-Ski2 and Hel308 are 4.1±0.9 min-1 and 17.7±1.7 min-1, respectively.(C) ATPase assays of ASH-Ski2, ASH-Ski2ΔN or ASH-Ski2K245A in absence or presence of *P. abyssi* genomic RNA (+RNA).

Figure S7. (A) Kinetics of strand dissociation of 3'RNA_{50}/RNA_{26}, 3'DNA_{59}/DNA_{31} and 5'RNA_{50}/RNA_{26} homoduplex over time at 65°C in presence of 250nM recombinant protein as follow: ASH-Ski2 and variants, Hel308 and ASH-Ski2 in absence of ATP. A control was conducted in absence of protein (No protein) to evaluate the stability of the homoduplexes at 65 °C. The products of the reaction were separated on an 8% native acrylamide gel. (B) Kinetics of strand dissociation by ASH-Ski2 in the presence of ATP are shown for 3′ overhang (3'RNA_{50}/RNA_{26}) and 5′ overhang duplex (5'RNA_{50}/RNA_{26}). Three independent experiments were performed in each condition. (C) Kinetics of strand dissociation by protein variants ASH-Ski2ΔN or ASH-Ski2K245A are shown for 3' overhang 3'RNA_{50}/RNA_{26} duplex (upper panel). Kinetics of strand dissociation by ASH-Ski2 in presence or absence of ATP are shown for 3' overhang 3'RNA_{50}/RNA_{26} duplex (lower panel). Protein free experiments were carried out in parallel (Ø).

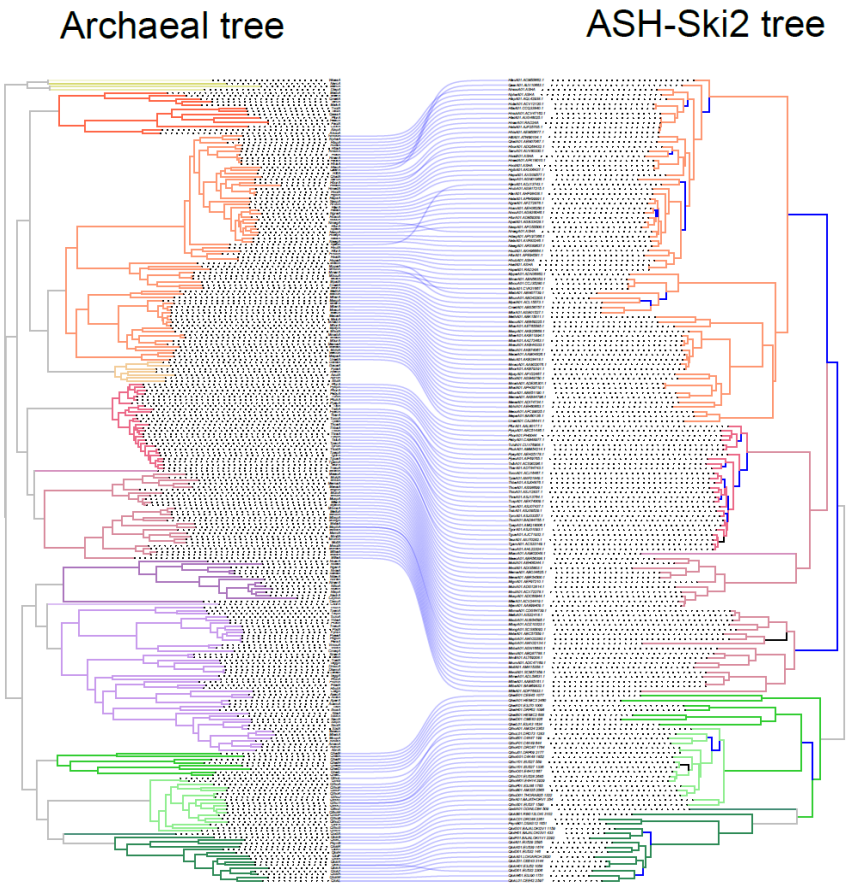
Figure S8. (A). Kinetics of strand association of 3'RNA_{50}/RNA_{26}, or 3'DNA_{50}/DNA_{26} homoduplex over time at 65°C in presence of 250nM ASH-Ski2 or Hel308 recombinant protein. A control in absence of protein (No protein) was conducted to evaluate over time the formation of the homoduplexes at 65 °C. The products of the reaction were separated on an 8% native acrylamide gel. (B) An annealing assay was performed after a 5-minute incubation period with increasing concentrations of ASH-Ski2 protein ranging from 0 to 250 nM (C) Kinetics of strand association by ASH-Ski2ΔN or ASH-Ski2K245A protein variants are shown for 3′ overhang duplexes (3'RNA_{50}/RNA_{26}).

Figure S9. Relative quantities of representative transcripts identified in RNA-seq analysis. Quantities of three significantly accumulated (left part; TERMP_RS02165, TERMP_RS07170 and TERMP_RS05490) and three down-regulated transcripts (TERMP_RS06165, TERMP_RS06170 and TERMP_RS00405), as well as unchanged transcript (TERMP_RS06995) were normalized by the quantity of reference gene (TERMP_RS03110). Quantities of wild-type (WT) and ΔASH-Ski2 transcripts are indicated by black and orange bars, respectively. Values are the means of three independent experiments, each from a different cell culture.

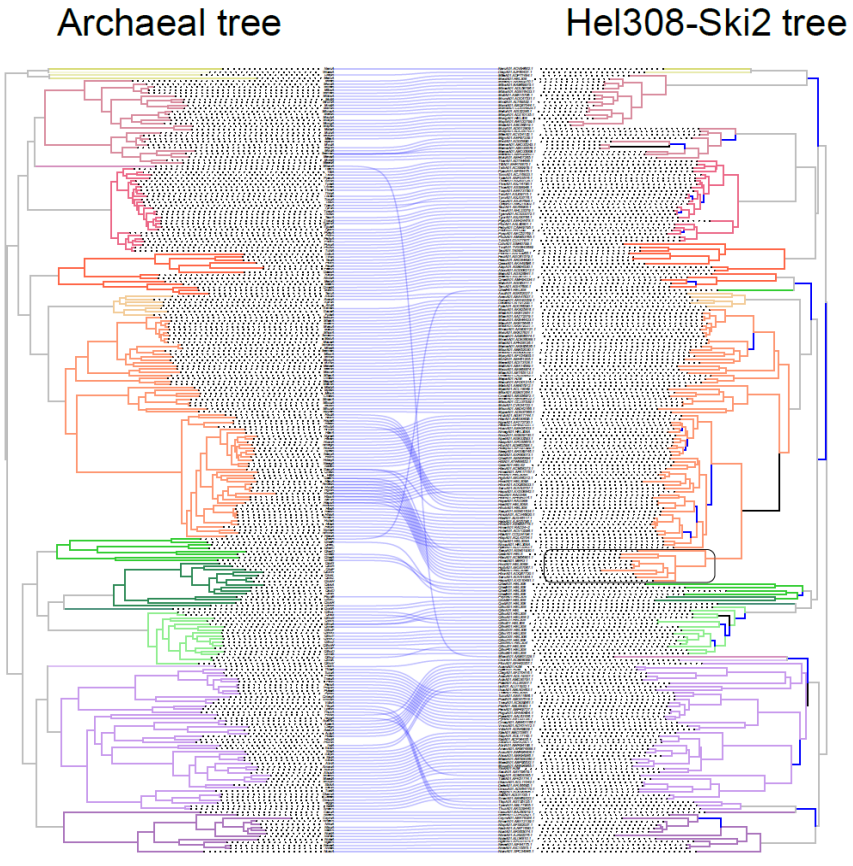
Supplementary data

Figure S1

A



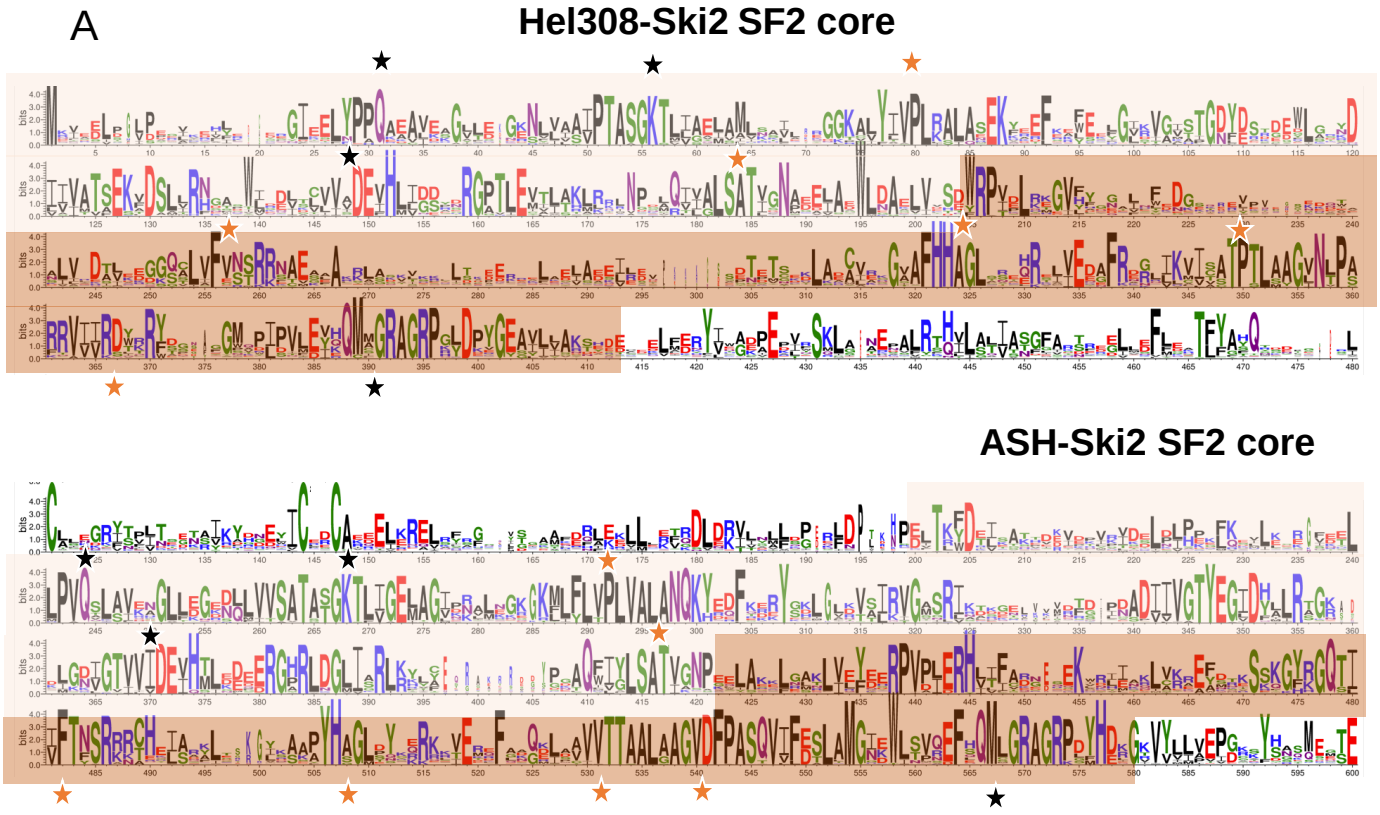
B



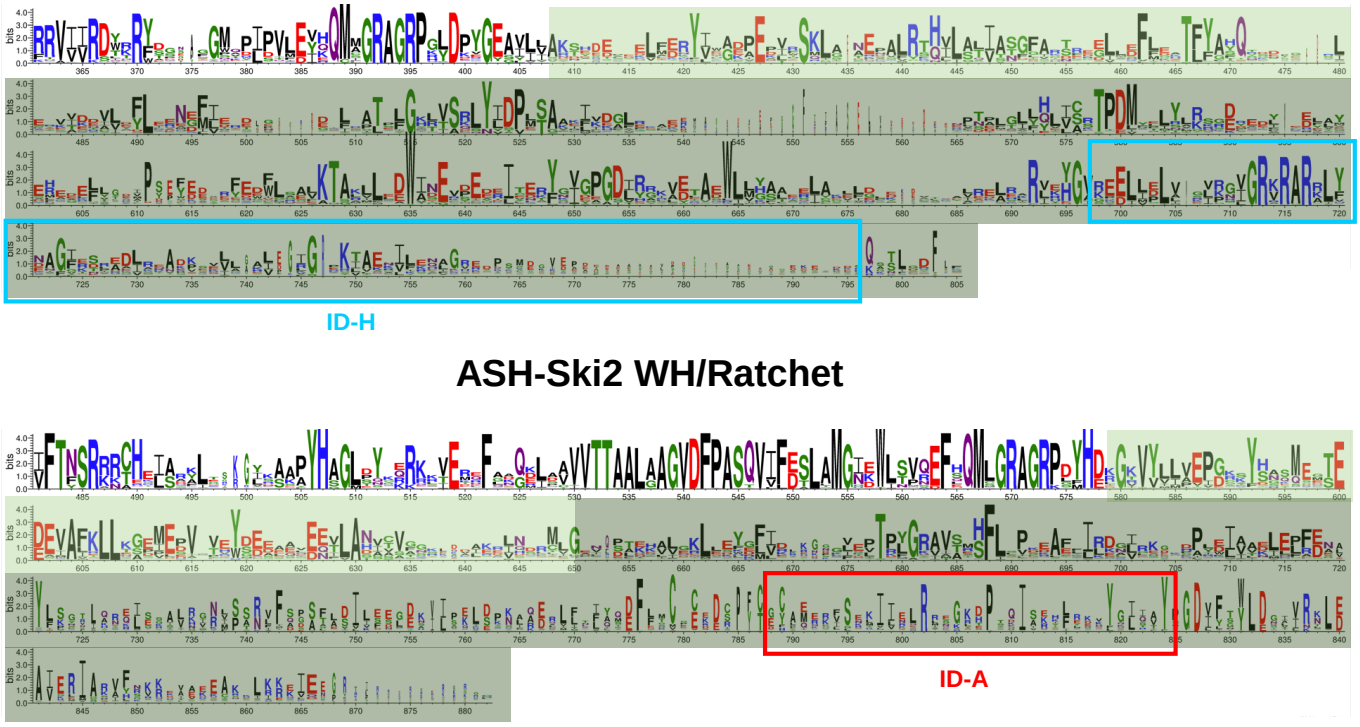
Supplementary data

Figure S2

A

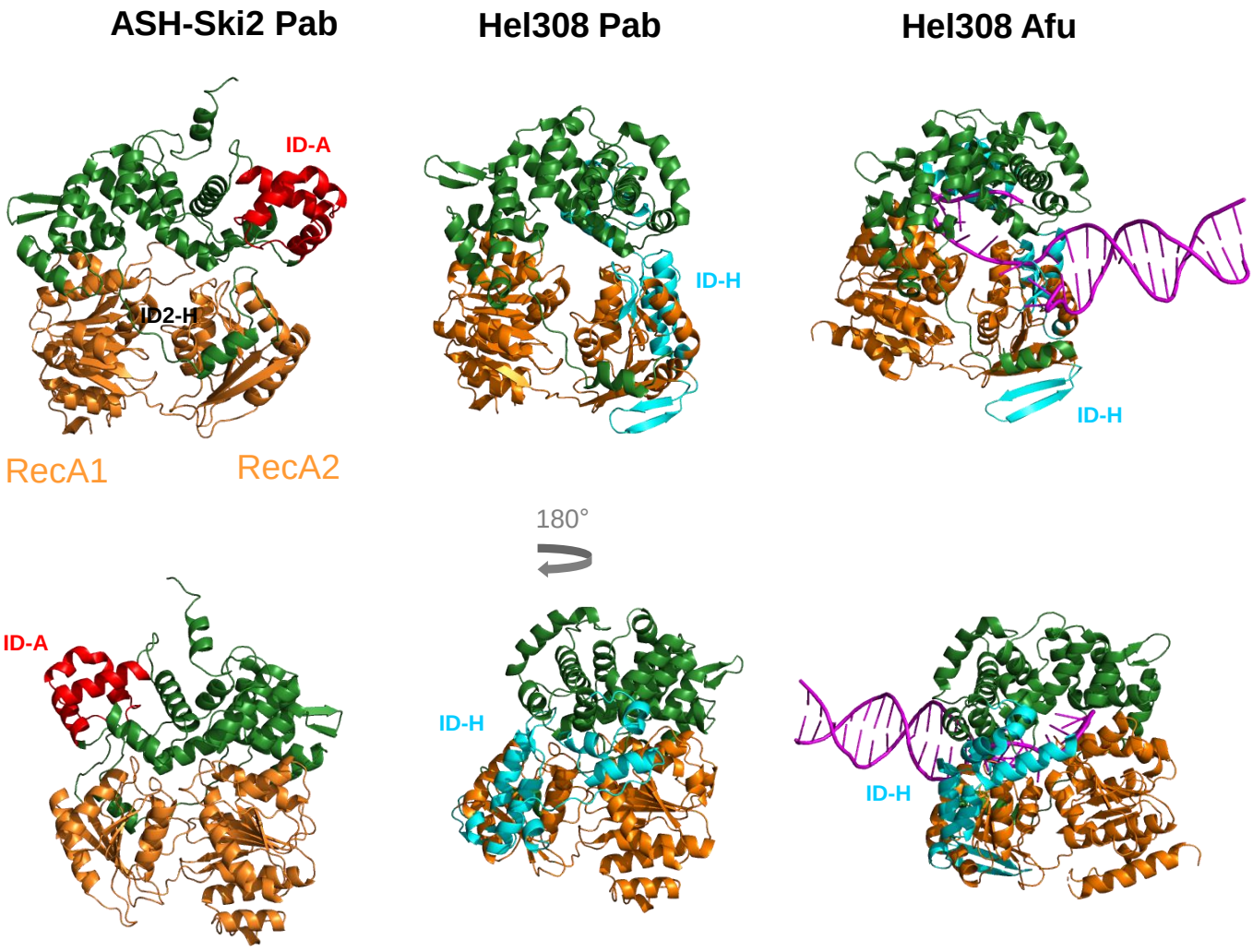


B



Supplementary data

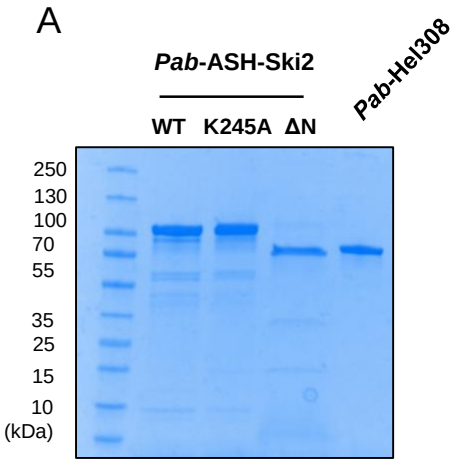
Figure S3



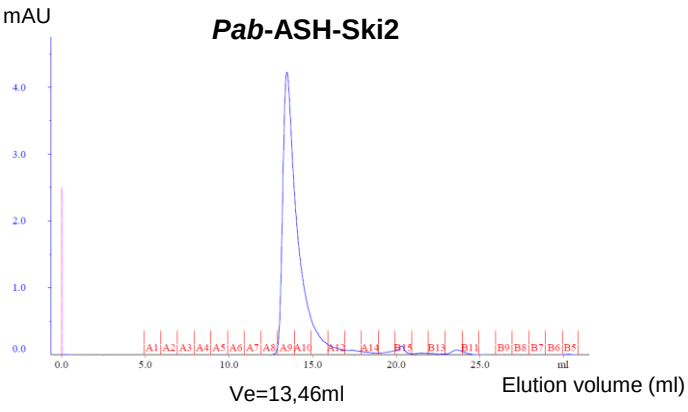
Supplementary data

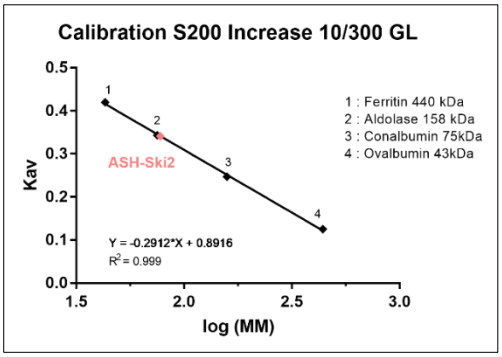
Figure S4

A



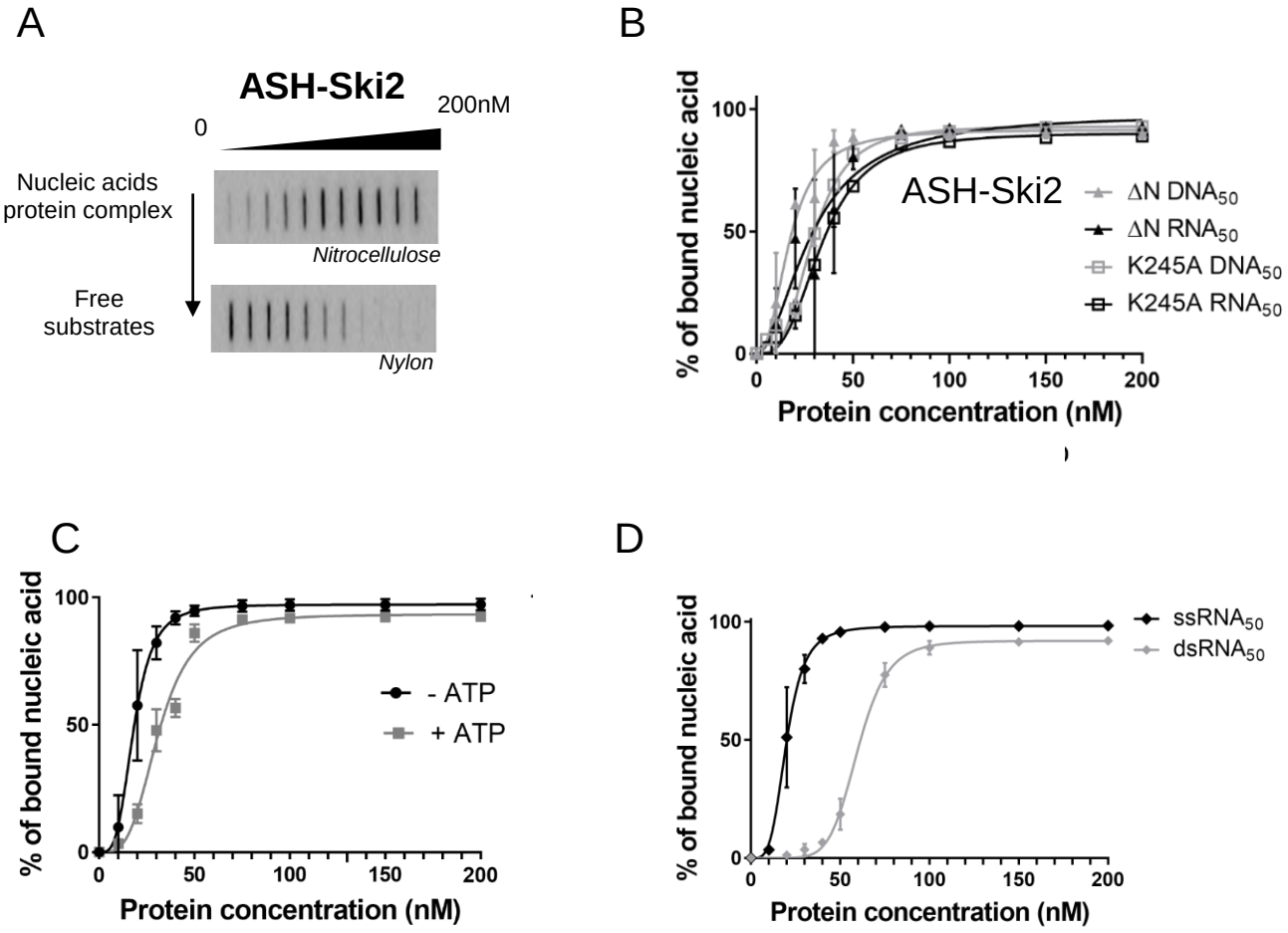
B





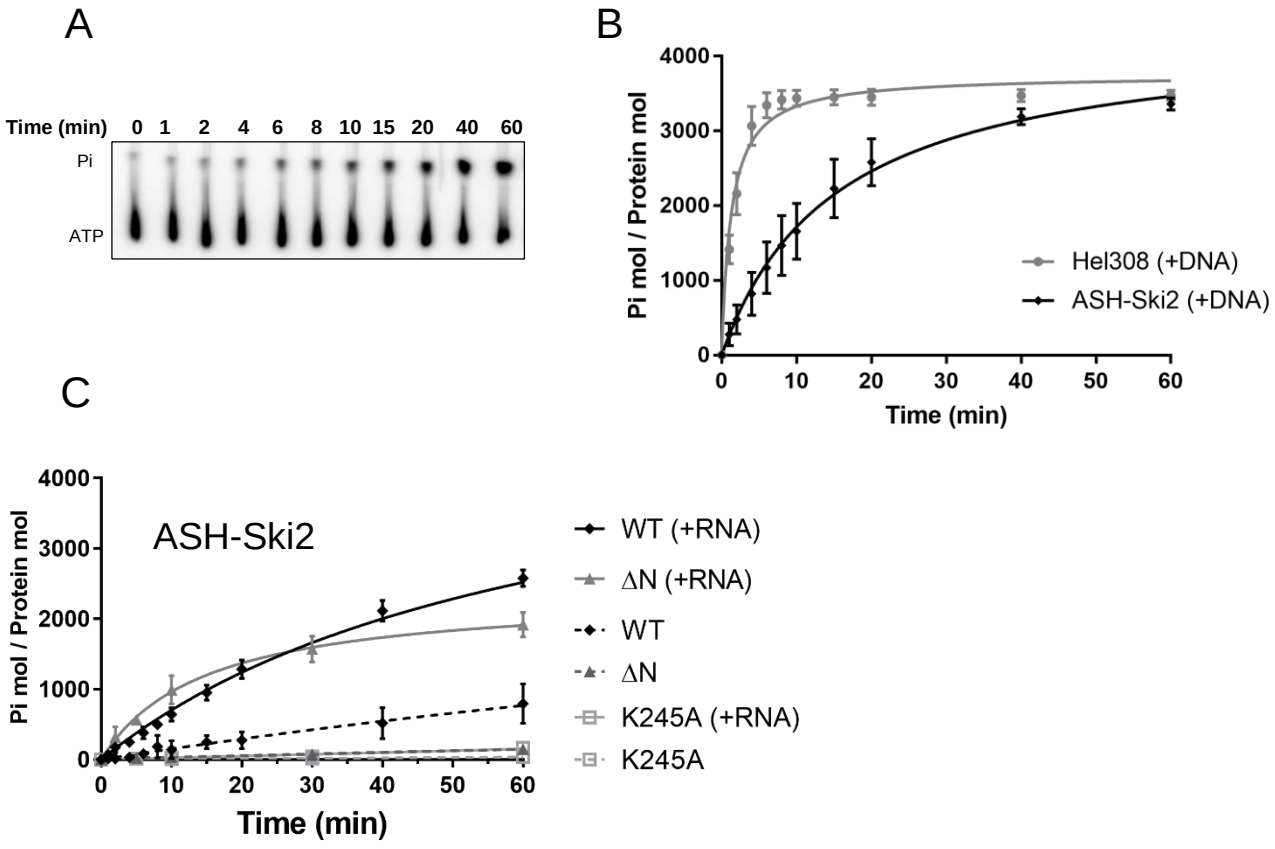
Supplementary data

Figure S5



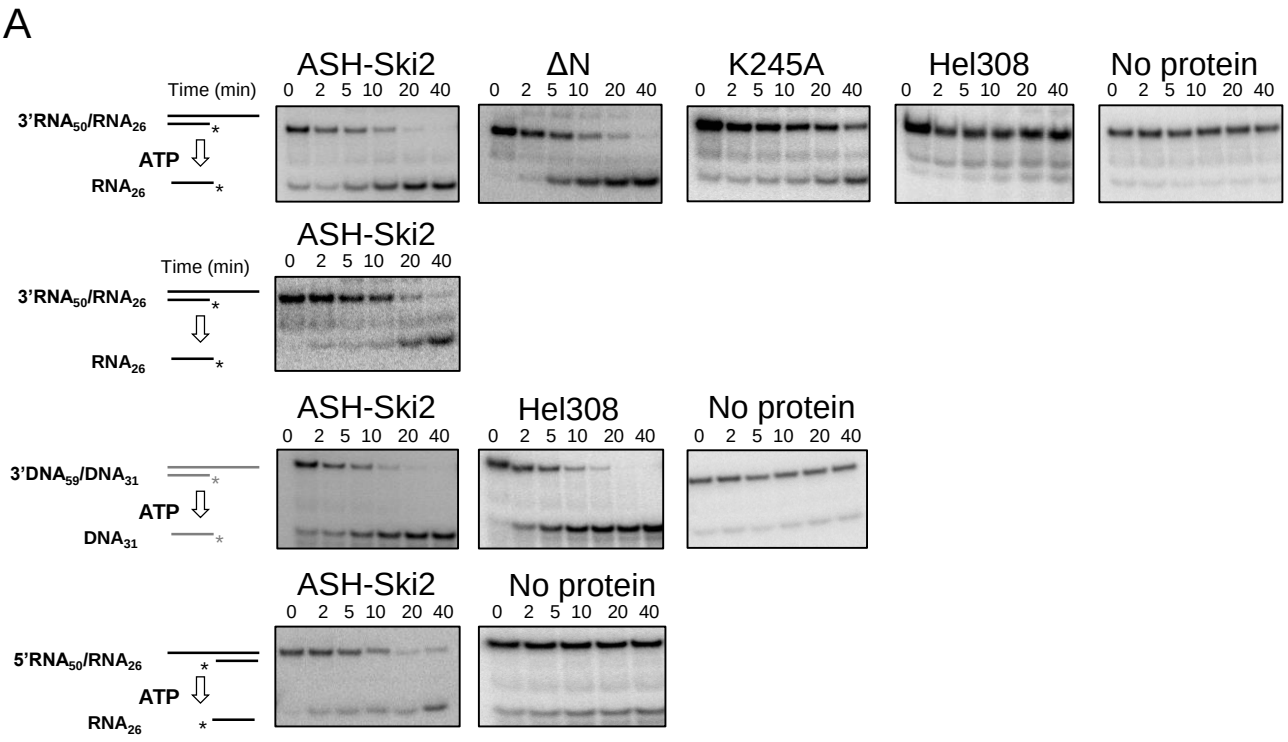
Supplementary data

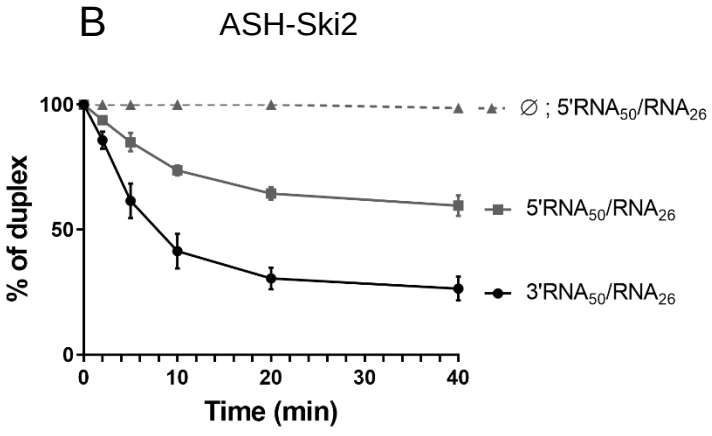
Figure S6

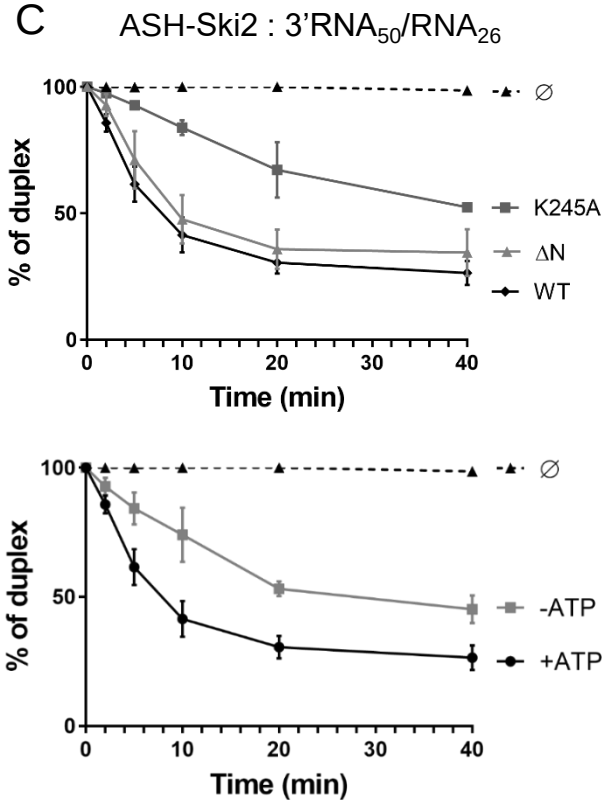


Supplementary data

Figure S7

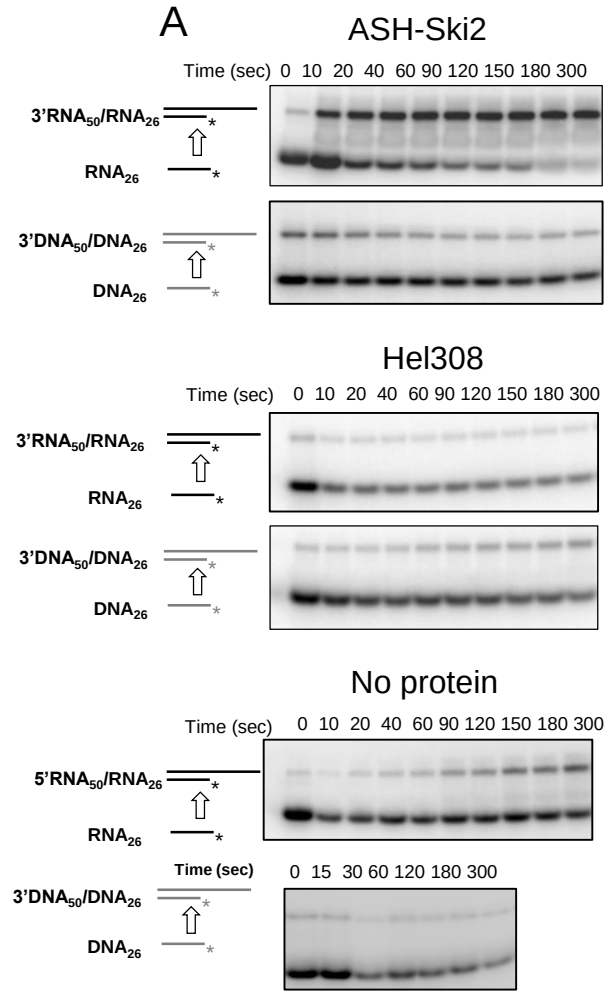


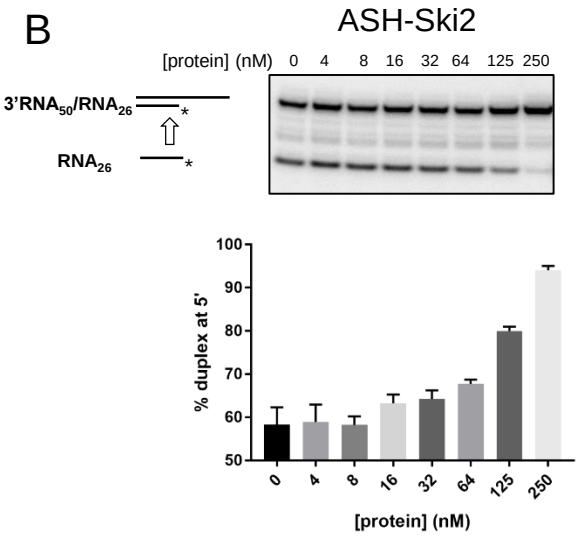


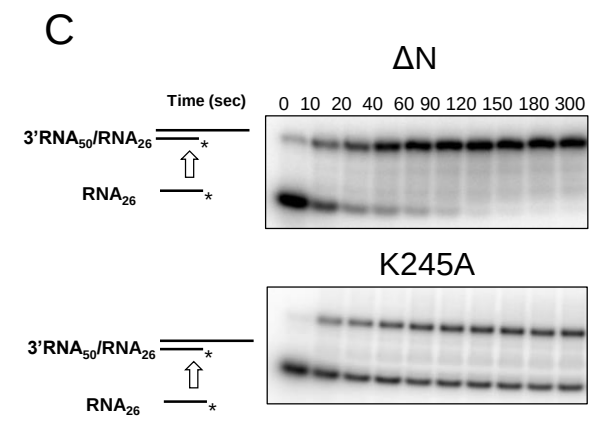


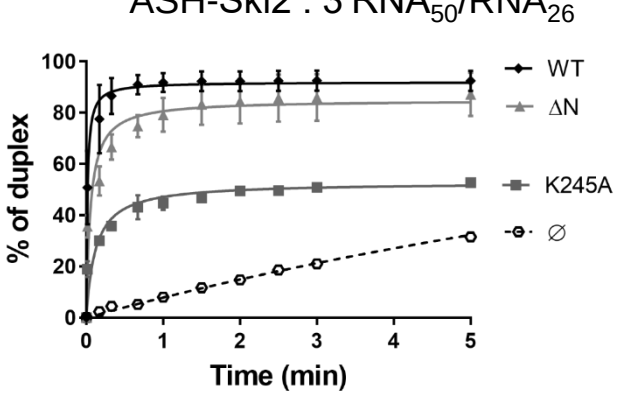
Supplementary data

Figure S8



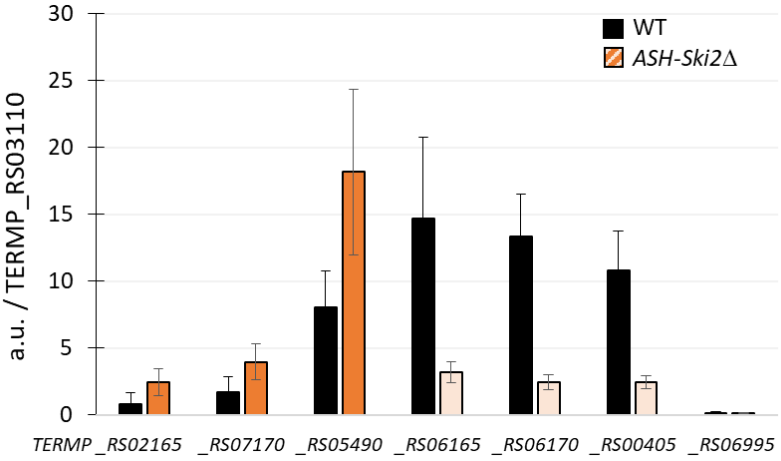






Supplementary data

Figure S9



SUPPLEMENTARY TABLES

Table S1 (see Excel file Table S1).: Organisms, locus tag (sequence identifier), Uniprot acccession numbers and taxonomic lineage of archaeal ASH-Ski2, Hel308-Ski2, BRR2-Ski2 and eukaryal Ski2-Mtr4.

Table S2. List of oligonucleotides used to construct pET11b expression vectors

Oligonucleotides	Sequences
T7 Prom	TAATACGACTCACTATAGGGGA
T7 Ter	GCTAGTTATTGCTCAGCGGT
For pET11b	CATATGTATATCTCCTTCTTAAAGTT
Rev pET11b	GGATCCGGCTGCTAACAAAGCC
For ASH-Ski2 pET11b	GGAGATATACATATGATGCTATTCGTTATTCGCCCAG
Rev ASH-Ski2 pET11b	TTAGCAGCCGGATCCTTATGGTTTTCTTTCTCTCCTCA
For ASH-Ski2ΔN-ter (194- 855) pET11b	CCGCGCGGCAGCCATATGACGATTGACGAGCTGGAT
Rev DomN (1-193) ASH-Ski2 pET11b	GGAGATATACATATGACGATTGACGAGCTGGAT
For ASH-Ski2 K245A	GCCACGGCAAGCGGTgcAACCTTGATTGGAGA
Rev ASH-Ski2 K245A	TCTCCAATCAAGGTTcgACCGCTTGCCGTGGC
For Hel308 pET11b	CGCGGCAGCCATATGATGAAAGTTGGAGAGCTAAACG
Rev Hel308 pET11b	GGAGATATACATATGATGAAAGTTGGAGAGCTAAACG

Table S3. DNA and RNA oligonucleotides substrates

Substrates	Sequences
DNA_{50}	5'-ATCGATAGTCTCTAGACAGCATGTCCTAGCAAGCCAGAATTCGGCAGCGT-3'
RNA_{50}	5'-AUCGAUAGUCUCUAGACAGCAUGUCCUAGCAAGCCAGAAUUCGGCAGCGU-3'
RNA_{26}	5'-GGACAUGCUGUCUAGAGACUAUCGAU-3'
3'DNA_{50}/DNA_{26}	5'-ATCGATAGTCTCTAGACAGCATGTCCTAGCAAGCCAGAATTCGGCAGCGT-3' 3'-TAGCTATCAGAGATCTGTCGTACAGG-5'
3'DNA_{59}/DNA_{31}	5'-GACGCTGCCGAATTCTACCAGTGCCTTGCTAGGACATCTTTGCCCACCTGCAGGTTCAC-3' 3'-CTGCGACGGCTTAAGATGGTCACGGAACGAT-5'
3'RNA_{50}/RNA_{26}	5'-AUCGAUAGUCUCUAGACAGCAUGUCCUAGCAAGCCAGAAUUCGGCAGCGU-3' 3'-UAGCUAUCAGAGAUCUGUCGUACAGG-5'
5'RNA_{50}/RNA_{26}	5'-AUCGAUAGUCUCUAGACAGCAUGUCCUAGCAAGCCAGAAUUCGGCAGCGU-3' 3'-GGAUCGUUCGGUCUUAAGCCGUCGCA-5'

Table S4 (see Excel file Table S4).List of up and down regulated transcripts in D2 *T. barophilus* strain and gene nomenclature correspondence.

Table S5: Up-regulated transcripts from genes encoding ribosomal proteins in *T. barophilus* ΔASH-Ski2 strain with 50S and 30S ribosomal proteins in green and red, respectively

Gene function	Name
50S ribosomal protein L3e (L3p)	TERMP_RS00440
50S ribosomal protein L1e (L4p)	TERMP_RS00445
50S ribosomal protein L23Ae (L23p)	TERMP_RS00450
50S ribosomal protein L8e (L2p)	TERMP_RS00455
30S ribosomal protein S15e (S19p)	TERMP_RS00460
50S ribosomal protein L17e (L22p)	TERMP_RS00465
30S ribosomal protein S3e (S3p)	TERMP_RS00470
50S ribosomal protein L35e (L29p)	TERMP_RS00475
translation initiation factor SUI1-like	TERMP_RS00480
ribonuclease P protein component 1	TERMP_RS00485
30S ribosomal protein S11e (S17p)	TERMP_RS00490
50S ribosomal protein L23e (L14p)	TERMP_RS00495
50S ribosomal protein L26e (L24p)	TERMP_RS00500
30S ribosomal protein S4e	TERMP_RS00505
50S ribosomal protein L11e (L5p)	TERMP_RS00510
30S ribosomal protein S29e (S14p)	TERMP_RS00515
30S ribosomal protein S15Ae (S8p)	TERMP_RS00520
50S ribosomal protein L9e (L6p)	TERMP_RS00525
50S ribosomal protein L32e	TERMP_RS00530
50S ribosomal protein L19e	TERMP_RS00535
30S ribosomal protein S18e (S13p)	TERMP_RS00625
30S ribosomal protein S9e (S4p)	TERMP_RS00630
30S ribosomal protein S14e (S11p)	TERMP_RS00635
50S ribosomal protein L18Ae	TERMP_RS05425
50S ribosomal protein L30e	TERMP_RS01220
30S ribosomal protein S20e (S10p)	TERMP_RS03005